servator

modelo preditivo de apoio à prospecção arqueológica

Natália Maria da Costa Botica
Universidade do Minho – Escola de Engenharia

Dissertação submetida à Universidade do Minho
para obtenção do grau de mestre em Sistemas de Informação

Orientadores:
Prof. Doutora Maribel Yasmina Santos
Doutor Francisco Sande Lemos

Universidade do Minho
2004

servatis servandis

conservando-se o que deve ser conservado

Culpa est non praevidere quod facile potestevenire

é culpa não prever o que facilmente pode acontecer

Agradecimentos

À Doutora Maribel Santos, orientadora deste trabalho, a quem muito agradeço e estimo pela sua disponibilidade constante e sábias sugestões que proferiu.

Ao Doutor Francisco Sande Lemos que desde o início acreditou neste trabalho e o apoiou incondicionalmente.

À Professora Doutora Manuela Martins pela sua preocupação constante com a formação dos seus colaboradores.

Aos meus colegas da Unidade de Arqueologia, em particular o Dr. Paulo Bernardes pela sua frequente ajuda, amizade e apoio.

À minha família, em particular ao José Pedro, à Olga e ao Hugo, por serem uma constante fonte de estímulo, amor e dedicação.

Resumo

A Descoberta de Conhecimento em Bases de Dados integra teorias, métodos e algoritmos com o objectivo de identificar relacionamentos implícitos nos dados. Por sua vez, a Arqueologia possui Bases de Dados, para as quais a aplicação dos princípios associados à Descoberta de Conhecimento em Bases de Dados para identificar relacionamentos implícitos, constitui um grande desafio.

O modelo ***servator*** representa uma nova abordagem às Bases de Dados de Arqueologia e revelou poder vir a constituir uma importante ferramenta para as actividades de investigação nesta área. A metodologia adoptada, passível de ser utilizada na identificação de novos modelos, poderá ser um instrumento de protecção do Património cultural, tão importante para prolongar no tempo a nossa memória colectiva.

Abstract

Knowledge Discovery in Databases comprises theories, methods and algorithms that aim the identification of implicit data relationships. The application of Knowledge Discovery principles in archaeological databases to identify implicit relationships, constitutes a great challenge.

The ***servator*** model proposes a new approach to archaeological databases and therefore might be considered a fundamental tool in archaeological research activities. This methodology can also be used in the identification of new models and as a valuable tool to protect cultural heritage, which is undoubtedly vital to preserve our collective memory.

Índice

Índice de Figuras

Índice de Tabelas

Siglas

a.C.	Antes de Cristo
BD	Bases de Dados
CCRN	Comissão de Coordenação da Região Norte
CD	*Compact Disk*
CD-ROM	*Compact Disk- Read Only Memory*
CRISP-DM	*CRoss Industry Standard Process for Data Mining*
DC	Descoberta de Conhecimento
d.C.	Depois de Cristo
DCBD	Descoberta de Conhecimento em Bases de Dados
DM	*Data Mining*
DSI	Departamento de Sistemas de Informação
DW	*Data Warehouse*
ODBC	*Open Database Connectivity*
OLAP	*Online Analytic Processing*
PDM	Plano Director Municipal
SDC	Sistema de Descoberta de Conhecimento
SIABRA	Sistema de Informação Arqueológica de *Bracara Augusta*
SIG	Sistema de Informação Geográfica
TI	Tecnologias da Informação
UAUM	Unidade de Arqueologia da Universidade do Minho

Capítulo 1

Introdução

Os ecossistemas humanizados, nos quais se inserem numerosas formas de Património, sofreram e continuam a sofrer sucessivas agressões. As alterações da paisagem associadas ao tempo, ao crescimento demográfico e decorrentes das actividades industriais, agrícolas e florestais, constituem linhas de desenvolvimento que têm obliterado muito Património.

Despontam, no entanto, iniciativas nas áreas do turismo cultural, ecológico ou histórico, orientadas para a valorização de espaços e vivências do passado. É uma nova economia, cujo principal recurso são as paisagens culturais e que deverão manter vivas, evitando a sua degradação e desaparecimento. O incremento destas actividades poderá contribuir de forma decisiva para que o Património seja mais estudado, preservado e valorizado.

Também as Tecnologias da Informação (TI) têm assumido uma importância crescente em todo o processo de registo e estudo dos valores patrimoniais. Começaram pelo uso de Bases de Dados (BD), para armazenar e gerir os dados arqueológicos. Ao longo dos anos têm impulsionado de forma decisiva os projectos de Arqueologia, em vários âmbitos: na gestão do processo arqueológico, na representação dos dados [Barceló *et al*., 2000], no desenvolvimento de Sistemas de Informação Geográfica (SIG) [Allen *et al.*, 1990] [Lock e Stantic, 1995], na Reconstituição Virtual de Património [Bernardes, 2002] e na criação de Conhecimento Arqueológico [Botica *et al.*, 2003a].

Com este trabalho pretende-se definir um sistema integrado de gestão de informação arqueológica, ***servatis***[1] e identificar um modelo preditivo de apoio à prospeção arqueológica, o ***servator***[2].

---

[1] O termo ***servatis*** foi extraído da frase latina ***servatis servandis*** que significa "conservando-se o que deve ser conservado".

[2] A palavra latina ***servator*** significa servo, guarda, que assegura a salvação ou conservação.

O sistema ***servatis***, cuja arquitectura e implementação se apresentam neste trabalho, dará suporte ao armazenamento de dados do processo arqueológico, desde a caracterização de estruturas e espólio, passando pelos registos fotográficos, cartografia e desenhos. Será ainda uma interface com os utilizadores, proporcionando uma diversificada visualização de dados, através da produção orientada de listagens, relatórios ou gráficos. Para além do exposto, o ***servatis*** constituirá um apoio à actividade arqueológica na da identificação de padrões nos dados, usando os princípios associados à Descoberta de Conhecimento em Base de Dados (DCBD). O sistema prevê a utilização de uma BD de modelos, sendo também uma finalidade deste trabalho a identificação de um modelo preditivo de património arqueológico, o ***servator***, aplicado a Trás-os-Montes Oriental, a disponibilizar nessa BD do ***servatis***.

O Património está inserido em contextos geográficos, sociais e culturais específicos, que podem variar de região para região. A BD de caracterização de sítios arqueológicos, disponível para este trabalho, corresponde aos sítios de Trás-os-Montes Oriental, pelo que se optou por identificar um modelo preditivo de apoio à prospecção arqueológica, para essa região.

A metodologia utilizada para identificar o ***servator*** pode ser aplicada a outras BD de arqueossítios, sendo utilizada pelo ***servatis*** para conduzir o utilizador na identificação de outros modelos.

1.1. Motivações, finalidade e objectivos

A localização periférica do território português, no continente europeu, não provocou o seu isolamento. Bem pelo contrário, graças à sua situação de charneira supra-regional, desempenhou um papel preponderante nas relações culturais entre povos e saberes. Permitiu a adopção de novidades técnicas e ideológicas, capazes de imprimirem grande vigor, tanto à génese, como ao desenvolvimento de diversos surtos civilizacionais que se sucederam na península [Gomes, 2000].

Apesar da sua pequena extensão territorial, Portugal é um país com uma alta densidade relativa de testemunhos arqueológicos. Todo este Património é um bem incontornável para estudar o passado e suscitar o desenvolvimento sócio-cultural.

Abrange todas as Idades e materializa-se das mais variadas formas, desde a arte rupestre, até aos monumentos medievais, ou mesmo mais recentes.

A consciência de que o conhecimento do passado torna mais forte a nossa memória colectiva e faz prolongar no tempo a nossa civilização, impulsiona o estudo e salvaguarda dos sítios arqueológicos. Motiva também a procura de soluções para prever a localização do extenso Património ainda por detectar [Lemos, 1991].

Partindo da necessidade de conhecer, estudar e divulgar o nosso legado patrimonial, definiu-se a arquitectura e implementação de um sistema integrado de gestão de informação arqueológica, o ***servatis***, e um modelo preditivo de apoio à prospecção arqueológica para a região de Trás-os-Montes Oriental, o ***servator***. Pretende-se que estes venham a ser uma ferramenta de trabalho na investigação arqueológica e, também, um veículo de comunicação com outras áreas de interesse.

Para atingir esta finalidade definiram-se como objectivos o estudo do domínio da Arqueologia, a compreensão dos conceitos associados à DCBD, a concepção da arquitectura de um sistema integrado de gestão de informação arqueológica, bem como, a identificação de um modelo preditivo de apoio à prospecção de Património arqueológico.

Na identificação do ***servator*** utilizaram-se os princípios associados à DCBD, aplicando-se a metodologia CRISP-DM (*Cross-Industry Standard Process for Data Mining*), que a seguir se apresenta.

1.2. Metodologia utilizada

A metodologia CRISP-DM (Figura 1.1) define um conjunto de seis etapas para o desenvolvimento estruturado e metodológico de projectos de *Data Mining* (DM) [Chapman *et al.*, 2000].

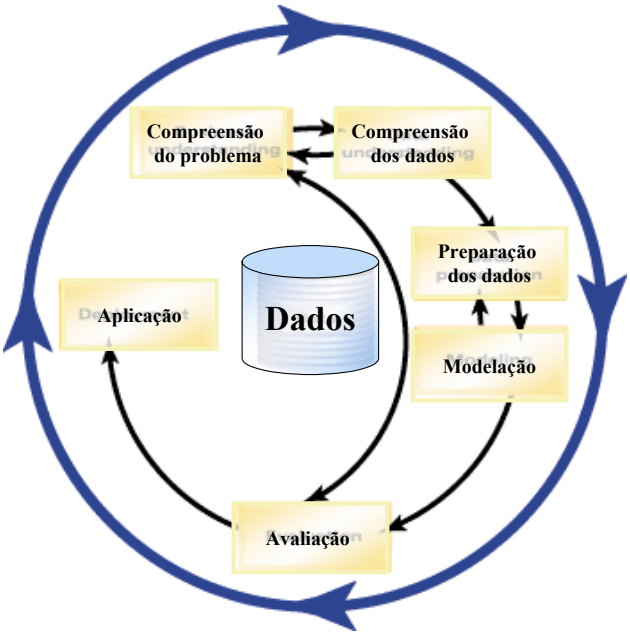


Figura 1.1 – Ciclo de vida do processo de DCBD, segundo a metodologia CRISP-DM

De acordo com esta metodologia, o modelo preditivo ***servator*** será identificado executando, de forma iterativa, as etapas a seguir apresentadas: compreensão do problema, compreensão dos dados, preparação dos dados, modelação, avaliação e aplicação.

1.2.1. Compreensão do problema

Nesta fase, procurou-se identificar as necessidades e os objectivos a atingir, convertendo este conhecimento numa tarefa de DM. Fez-se um diagnóstico das necessidades na Arqueologia, tendo-se escolhido a prospecção arqueológica, como actividade onde as TI, nomeadamente as ferramentas de DCBD, poderiam dar um contributo positivo. Assim, definiu-se como finalidade do trabalho, a identificação de um modelo preditivo para apoio à prospecção arqueológica.

O conhecimento arqueológico existente, associado à prospecção de arqueossítios e ao enquadramento dos sítios inventariados, será objecto de estudo e de sistematização, de forma a compreender os princípios em que esta actividade assenta, bem como as metodologias de trabalho e técnicas que podem ser utilizadas para a sua realização.

1.2.2. Compreensão dos dados

Nesta fase, estudaram-se os dados disponíveis para a realização do trabalho, analisando-se o seu conteúdo e escolhendo-se os dados relevantes para o estudo. Procurou-se também identificar as fontes, os procedimentos de leitura dos dados utilizados, os formatos adoptados, a descrição, a qualidade e utilidade dos mesmos, bem como o volume disponível para o trabalho.

Nesta primeira análise exploratória, identificaram-se os problemas associados aos dados e fizeram-se as primeiras descobertas, recorrendo a ferramentas de interrogação e visualização.

1.2.3. Preparação dos dados

Nesta fase de preparação executaram-se uma série de tarefas sobre os dados, com o objectivo de construir o conjunto para análise, sobre o qual serão aplicadas as técnicas de modelação. Incluem-se nesta etapa todas as actividades de extracção de dados das BD fonte, limpeza e de transformação.

São também realizadas tarefas de junção de tabelas, alteração de formatos ou agregação de valores. Pretende-se que o significado dos dados não seja alterado, mas que estejam de acordo com as necessidades dos algoritmos de DM.

1.2.4. Modelação

Na fase de modelação aplicaram-se aos dados técnicas de DM, escolhidas de acordo com objectivos pretendidos. Estas técnicas são seleccionadas e parametrizadas, de forma muito iterativa, procurando-se optimizar os resultados obtidos.

Para identificar o modelo pretendido utilizaram-se dados de um conjunto de Treino, sobre os quais se aplicaram técnicas de modelação.

1.2.5. Avaliação

Na avaliação do modelo faz-se a análise dos resultados obtidos, quando aplicado aos dados do conjunto de Testes e pela sua confrontação com os objectivos definidos na fase de compreensão do problema.

De acordo com a avaliação de resultados, são tomadas decisões sobre a continuidade do processo, ou sobre a sua revisão. A cada nova avaliação podem refazer-se algumas fases, de acordo com o que for considerado conveniente.

Todo este ciclo será repetido tantas vezes quantas as necessárias, sempre que as alterações contribuam para melhorar os resultados obtidos, ou até que os objectivos inicialmente definidos sejam atingidos.

1.2.6. Aplicação

Nesta fase, desenvolveram-se as acções necessárias à organização do conhecimento obtido, definindo-se a melhor forma de apresentação, para que possa ser entendido na área específica em que irá ser utilizado.

Devem ser apresentadas também, indicações sobre a actualização periódica a fazer aos dados, para que o modelo possa estar sempre actualizado e seja útil.

Na área de estudo em que se insere este trabalho, a actualização dos dados pode ser feita para completar a BD e não na perspectiva de reflectir alterações nos mesmos. Na Arqueologia, os dados reflectem quadros congelados do passado, onde já não é possível qualquer alteração dos mesmos. Apenas a descoberta de novos dados, ou novas interpretações feitas, pode acontecer.

1.3. Organização da dissertação

A finalidade desta dissertação é, como foi já referido, definir a arquitectura do sistema ***servatis*** e identificar o modelo preditivo de apoio à prospecção arqueológica, ***servator.***

O ***servatis*** será um sistema integrado de gestão de informação arqueológica, apoiando a Arqueologia nas suas diversas vertentes, nomeadamente na localização de sítios arqueológicos, no registo da informação resultante do processo de escavação e na interpretação arqueológica. O ***servator,*** que fará parte integrante do ***servatis,*** tem a finalidade específica de apoiar a actividade de prospecção arqueológica, isto é, apresentar indicadores sobre a localização de sítios arqueológicos, em Trás-os-Montes Oriental.

Para atingir esta finalidade, definiram-se os objectivos apresentados na secção anterior e cuja concretização será feita por etapas, reflectidas na estrutura desta dissertação.

No capítulo 1 definem-se a finalidade e objectivos do trabalho a realizar, justificados pela necessidade de utilizar ferramentas de apoio à prospecção arqueológica, que possam contribuir para um melhor Ordenamento do Território. Apresenta-se também neste capítulo a metodologia utilizada para a sua concretização e a organização desta dissertação.

Envolvendo este trabalho saberes de duas áreas distintas, reservaram-se os capítulos 2 e 3, para a revisão bibliográfica associada à Arqueologia e aos princípios associados à DCBD. Os conceitos que envolvem são estruturais para o bom desenvolvimento deste trabalho, sendo com base neles que é identificado o modelo ***servator,*** proposto neste trabalho.

No capítulo 4 define-se a arquitectura do sistema ***servatis***. Este foi concebido para ser um sistema visualizador de informação arqueológica e apoiar a Arqueologia nas actividades de gestão de informação e de investigação.

A concretização do ***servator*** é um processo moroso e complexo, pelo que lhe são dedicados os capítulos 5 e 6. No capítulo 5 são apresentadas as operações a realizar sobre os dados, reservando-se o capítulo 6 para a aplicação de algoritmos de DM e respectiva validação e avaliação do modelo gerado.

No capítulo 7 apresenta-se uma síntese da dissertação, retirando-se as conclusões sobre o trabalho desenvolvido e equacionando-se algumas questões consideradas pertinentes para futuros trabalhos. Por ultimo expõe-se as considerações finais.

Capítulo 2

Compreensão do domínio arqueológico

A Arqueologia é uma ciência que estuda o passado. Esse estudo, decorre da actividade arqueológica, que se faz basicamente através de prospecções e escavações [Martínez, 1992], tendo em vista a interpretação dos sítios arqueológicos. Na fase interpretativa procura-se compreender a estrutura e organização do sítio, bem como o modo como se formou e as alterações decorrentes da deposição de sucessivas camadas de sedimentos [Baker, 1977].

Durante a fase de prospecção desenvolve-se um conjunto de trabalhos de campo e de gabinete, com o objectivo de localizar sítios arqueológicos. Depois de localizados, os sítios poderão ser escavados, a fim de se obterem mais dados. De qualquer modo, as recomendações internacionais vão no sentido de se limitar ao mínimo o número de trabalhos intrusivos (escavações) e de privilegiar a recolha de dados por contextualização dos sítios.

Para melhor compreender do domínio arqueológico, nos aspectos mais relevantes, relacionados com os objectivos definidos para este trabalho, serão analisados os vectores espaciais e temporais em que se inserem os sítios arqueológicos inventariados. Serão também caracterizadas as técnicas utilizadas para a localização dos arqueossítios, bem como o estado actual do conhecimento arqueológico, sobre a região escolhida.

Neste capítulo começa-se por fazer uma análise do território escolhido para a realização do trabalho, uma vez que constitui o cenário natural onde se inserem os sítios arqueológicos. A seguir, caracterizam-se os vários períodos cronológicos que marcaram a ocupação humana na região. Os arqueossítios inventariados, e que constituem a base para a identificação do modelo preditivo ***servator,*** enquadram-se sempre numa dimensão espacial e temporal.

As técnicas utilizadas para a prospecção arqueológica são apresentadas numa secção deste capítulo. Considerou-se importante o seu conhecimento prévio, dado que o modelo a identificar servirá para apoiar esta actividade.

Por último, apresenta-se neste capítulo o conhecimento arqueológico sobre a região de Trás-os-Montes Oriental, através da caracterização sumária dos inventários realizados e dos estudos que permitiram identificar algumas estratégias de povoamento.

2.1. Território alvo do estudo

Portugal Continental situa-se na ponta mais ocidental do continente europeu, apresentando a configuração de um pequeno rectângulo, alongado no sentido norte-sul.

Se esta situação geográfica nos manteve muitas vezes afastados dos grandes focos civilizacionais, contribuindo para que a influência de algumas épocas só tarde e pouco intensamente se fizesse sentir, também proporcionou, noutras alturas, intensas relações estabelecidas por via marítima [Medeiros, 2000].

Os acentuados contrastes geográficos entre o norte e o sul, o litoral e o interior modelaram, ao longo dos milénios, significativas diferenças culturais, acentuadas também pelos os contributos externos, tanto de procedência mediterrânea como atlântica e continental [Gomes, 2000].

A região de Trás-os-Montes Oriental, em particular, engloba acentuados contrastes geográficos que terão influenciado a estratégia de povoamento, ao longo dos tempos. Estes serão analisados e incluídos nos dados a tratar no modelo preditivo. O modelo assenta num inventário de sítios arqueológicos de Trás-os-Montes Oriental[3], localizados a norte do rio Douro e a oriente dos primeiros contrafortes das serras do Gerês, Barroso, Alvão e Marão.

A região denominada, desde a Baixa Idade Média, por Trás-os-Montes, é um espaço que, em termos geográficos, climáticos e paisagísticos, possui unidades muito diversificadas. O seu agrupamento ou classificação nunca obteve consensos. Dada a

[3] A BD de trabalho continha inicialmente registos de sítios arqueológicos das regiões de Trás-os-Montes Ocidental e Oriental. No entanto, como será explicado mais tarde, durante o processo de Descoberta de Conhecimento a região de estudo passou a ser apenas a de Trás-os-Montes Oriental.

variedade de micro-regiões identificadas é frequente encontrar designações tão díspares como "Trás-os-Montes", "Alto Douro", "Terra Quente", "Terra de Miranda", "Terra Fria", "Beira Transmontana", "Trás-os-Montes Oriental e Ocidental", "Alto Portugal" ou ainda "Nordeste Transmontano". Independentemente de outras designações ou subdivisões existentes, irá adoptar-se a de Trás-os-Montes Oriental para designar o território alvo, objecto deste estudo.

Definidos os contornos territoriais é necessário caracterizar os factores ambientais, cuja relevância para a localização dos sítios arqueológicos é inquestionável. A importância destes factores varia de acordo com as civilizações, com a economia e com as tecnologias de que dispõem.

Assim, a localização dos *habitats* teria sido influenciada por vários factores, nomeadamente pelas características defensivas do território, visibilidade sobre os territórios envolventes, acessibilidade a recursos aquíferos, características dos solos ou ainda pelos recursos mineiros existentes.

Para melhor compreender este enquadramento, caracterizam-se a seguir as variáveis ambientais Relevo, Rede Hidrográfica, Litologia e Solos, para a região de Trás-os-Montes Oriental.

2.1.1. Relevo

No relevo de Trás-os-Montes Oriental dominam as formas resultantes de sucessivas aplanações, deslocadas e desniveladas por um complexo sistema de falhas. As serras são superfícies planalticas, soerguidas ao longo de falhas, onde pontualmente se encontram cristas quartzíticas (Figura 2.1)**,** que se destacam por serem rochas de maior dureza, ora aplanadas, ora de arestas vivas que resistiram à erosão [Ribeiro *et al*., 1991].

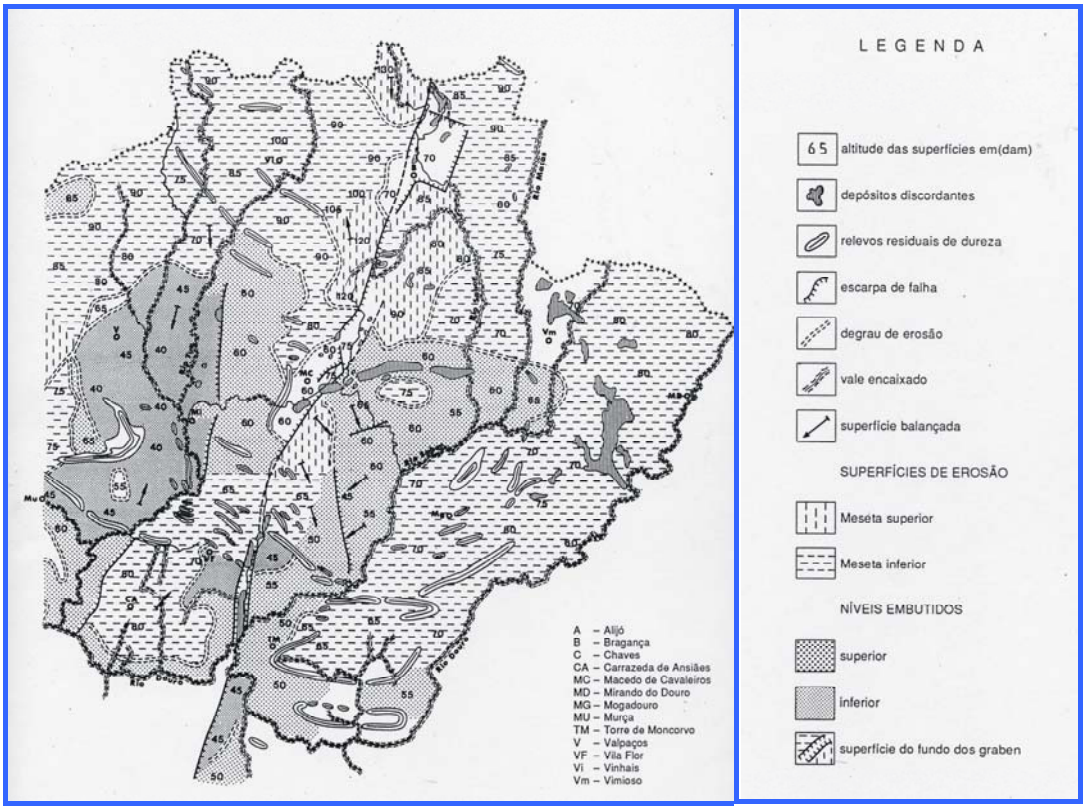


Figura 2.1 – Esboço geomorfológico de Trás-os-Montes Oriental (reproduzido de [Ribeiro *et al*., 1991])

Esta configuração é o resultado de um processo tectónico caracterizado pelo levantamento e fractura de uma antiga superfície erodida. É o caso da Serra da Padrela, de cimo plano, separado da serra do Alvão pelo fosso de Vila Pouca de Aguiar, cuja base é um troço abatido da mesma superfície [Ribeiro *et al*., 1991].

Como resultado de movimentos tectónicos podem encontrar-se relevos mais antigos, com superfícies suaves nos cumes das montanhas e nos planaltos elevados, e formações mais recentes, com formas mais agrestes, normalmente associadas aos patamares mais baixos [Lemos, 1993].

As formas de relevo mais comuns na região, referenciadas nos inventários utilizados, são:

- **Cabeço** – bloco que se destaca do resto da paisagem, formado pelo arrastamento de massas do rebordo, como resultado da drenagem de águas. Este processo de erosão confere-lhe uma configuração idêntica à de uma cabeça;

- **Castelo granítico** – ou batólito granítico é uma grande massa rochosa que sobressai dos mantos de xistos. Resulta da erosão exercida sobre rochas de resistência diferente;
- **Crista quartzítica** – relevo residual de massas de quartzito que, pela sua dureza, resistiu à erosão e se destaca das rochas envolventes;
- **Depressão tectónica** – resultante de movimentos tectónicos que provocaram o alongamento de superfícies ou o abatimento de extensas áreas. Em Trás-os-Montes Oriental a falha tectónica de Vilariça – Bragança é um exemplo deste tipo de relevo;
- **Esporão** – lomba convexa que se destaca do resto da paisagem;
- ***Inselberg*** – são montes-ilha formadas sobre os rios que emergem de uma base plana. Resultam do recuo do flanco montanhoso, que deixa para trás pedaços de rocha mais duros que resistiram à erosão;
- **Planalto** – forma alta e aplanada, levantada pelos movimentos tectónicos. Existem diversos patamares, podendo os planaltos ser cimeiros ou de transição para as depressões;
- **Serra** – em Trás-os-Montes Oriental, as serras resultaram de elevações ou fracturas de superfícies planas, provocadas por movimentos tectónicos;
- **Terraço** – superfície resultante da acumulação de materiais, quando o nível do mar estava mais elevado, ou resultante da acumulação fluvio-glaciar;
- **Vale** – nesta região existem diferentes tipos de vales criados por incisão fluvial. Uns, como os vales do Douro e do Sabor, são escarpados e jovens, formados por abatimento resultante de processos tectónicos. Outros, como o vale da Vilariça, resultaram da erosão diferencial ou regressiva.

2.1.2. Rede Hidrográfica

A rede hidrográfica é uma estrutura que, desde sempre, influenciou a localização dos povoados. Na região de Trás-os-Montes Oriental, assume um papel ainda mais determinante, dado tratar-se de uma zona de clima mediterrâneo com cariz continental,

onde o índice hídrico é bastante baixo. Para além disso, os rios constituem a principal via natural de acesso às terras mais setentrionais, interiores ou de cotas mais elevadas [Cruz, 2000]. Estes corredores naturais de circulação proporcionaram condições para que o homem escolhesse a localização dos seus *habitats,* ponderando o seu posicionamento, relativamente à rede hidrográfica. Nessa perspectiva, é importante caracterizar a rede hidrográfica de Trás-os-Montes Oriental, para que esse conhecimento seja incorporado no modelo preditivo de apoio à prospecção arqueológica.

A região de Trás-os-Montes Oriental possui duas grandes bacias hidrográficas, a do Tua e do Sabor, que desaguam directamente no Douro (Figura 2.2).

O Rio Douro corre no sentido este – oeste, mas os seus principais afluentes correm quase paralelamente entre si, com uma orientação nordeste - sudoeste.

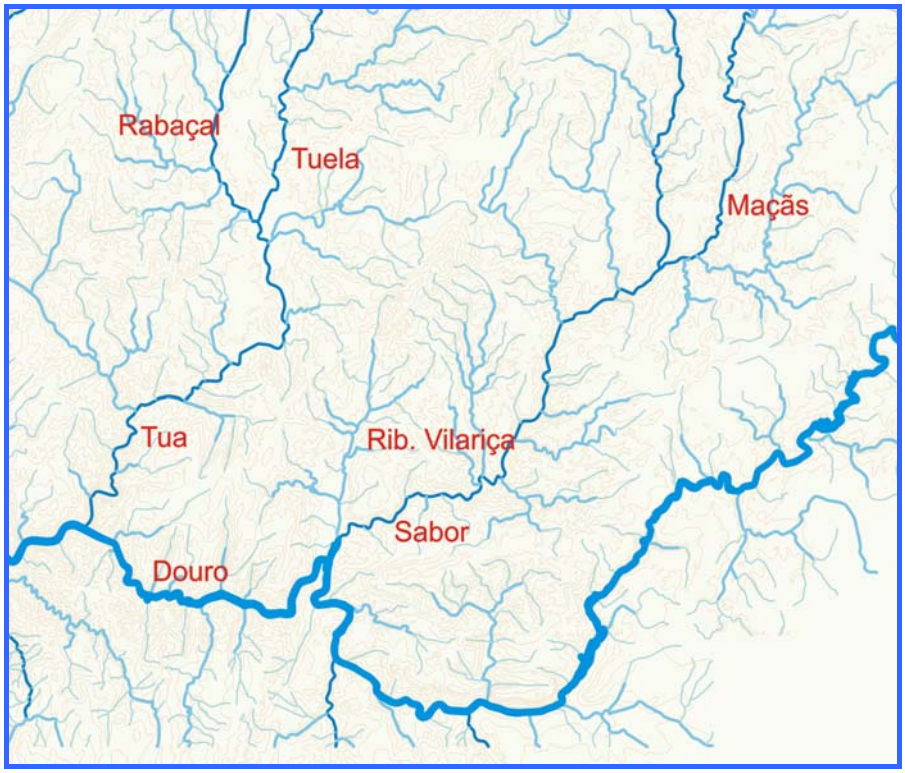


Figura 2.2 – Rede hidrográfica de Trás-os-Montes Oriental

O traçado da rede hidrográfica caracteriza-se pelo alinhamento paralelo dos grandes rios, com cursos de água seguindo na mesma direcção, ora convergindo, ora opondo as cabeceiras separadas por altas portelas [Ribeiro *et al*., 1991]. Este traçado,

resulta de deslocações que desnivelaram fragmentos do antigo planalto, ou ainda, do aparecimento de faixas de esmagamento favoráveis à erosão linear dos cursos de água. Estas faixas criaram compartimentos montanhosos cortados por vales profundos. Os campos de fractura criados deram origem a uma rede quase ortogonal, que orientou as sinuosidades dos cursos de água. A direcção mais importante dos rios é, como já se referiu, Nordeste - Sudoeste, seguindo uma linha de depressões, desde a raia de Trás-os-Montes Oriental até ao Douro.

2.1.3. Solos

Em Trás-os-Montes Oriental são pequenas as áreas de solos consideradas de boa capacidade agrícola. A maior parte do território é dominada por solos de pendor mais florestal. No entanto, deve referir-se que a capacidade de uso de solos depende do momento histórico a que se reporta, variando em função das tecnologias utilizadas pelas comunidades.

A utilização dos solos nas actividades agrícolas, pastorícia e silvicultura tem provocado nestes algumas alterações. O uso de técnicas ancestrais, como a mobilização de solos, nivelamento ou criação de patamares, muito tem contribuído para essas alterações. O exemplo mais típico é a construção de socalcos, em zonas de declives acentuados, onde os solos eram delgados e pouco produtivos. Nestes patamares passou a ser possível o aproveitamento dos solos para a agricultura [Coba, 1991].

A maior parte dos solos do nordeste formaram-se a partir de materiais resultantes da alteração e desagregação das rochas. As rochas sofreram a acção dos agentes erosivos, em condições ambientais muito heterogéneas, dando origem a solos de granulometria e espessura variáveis.

Seguindo a classificação das unidades taxionómicas de Agroconsultores e Coba [Coba, 1991], apresentam-se a seguir os tipos de solos mais comuns, na região de Trás-os-Montes Oriental, e cujas designações foram adoptadas no preenchimento da BD analisada:

- **Leptossolos** – solos com menos de 50 cm de profundidade, limitados por rocha dura, contínua e coerente. São formados por material não consolidado, muito

pedregoso. Estão associados à pastorícia ou a uma agricultura extensiva de cereal, exploração da oliveira, vinha e amendoeira;

- **Cambissolos** – solos moderadamente evoluídos, derivados de xistos ou granitos, que se localizam no fundo das encostas. São utilizados na exploração florestal e também na agricultura, embora necessitem de correcções químicas para compensar a carência de alguns nutrientes;
- **Fluvissolos** – solos pouco evoluídos, resultantes da sedimentação no fundo dos vales. Neles pode-se praticar uma agricultura de regadio intensivo ou sequeiro, pomar, olival ou vinha;
- **Luvissolos** – utilizados para o plantio da vinha, oliveira, amendoeira e cereais, desenvolvem-se a partir de xistos e rochas afins e sedimentos detríticos argilosos, em zonas relativamente quentes e secas;
- **Antropossolos –** são solos muito alterados, onde a actividade humana tem provocado profundas alterações das características originais através da remoção, cortes, enchimentos, adições seculares de materiais orgânicos e rega continuada.

2.1.4. Litologia

A estrutura geológica de Trás-os-Montes Oriental é dominada pelas rochas metamórficas, na zona central, e pelas formações de rochas graníticas, na periferia.

O quadro litológico é composto por formações xistosas e de xistos/quartzitos, ponteadas por algumas manchas graníticas tendo, cada um destes tipos, originado formas de relevo bastante díspares. Os xistos argilosos, muito mais permeáveis que o granito, cobrem-se de sulcos, por onde escorre a água das chuvas. Esta rocha parte-se e esfolheia-se, reduzindo-se a pequenos fragmentos que são levados pelas chuvas. As escarpas esbatem-se e degradam-se. A rede hidrográfica encaixa-se, formando um mar de cabeços, separados por sulcos escavados a várias alturas. No granito, as águas penetram profundamente na rocha, sem alterar a superfície. A arenização conserva uma topografia de maturidade, com vales fundos e largos, de vertentes esbatidas [Ribeiro *et al*., 1991].

No entanto, esta litologia complexa não é apenas o resultado de acções erosivas, realizadas ao longo do tempo. É também resultado de movimentos orogénicos, operados sobre uma antiga planície. Destes resultaram a formação de espaços de elevada altitude média, compostos por montanhas e planaltos, que alternam com depressões e vales encaixados.

A distribuição dos *habitats* e dos materiais recolhidos, revela que as comunidades da região tinham um bom conhecimento da sua litologia. Havia um aproveitamento de minérios e de algumas rochas, que serviam de matéria-prima para a construção e manufactura de artefactos [Cruz, 2000].

Apresentam-se a seguir alguns tipos de rochas, que constituíam um recurso económico interessante, e que estão referenciadas na BD de sítios arqueológicos, utilizadas neste trabalho:

- **Calcários** – utilizados na produção de lápides votivas e funerárias e, também, no fabrico de cal para construção;
- **Mármores e alabastros** – explorados na época romana pelo seu valor ornamental;
- **Talco** – muito utilizado na época romana para construção de estelas funerárias e peças ornamentais;
- **Granitos** – utilizados na construção de sistemas defensivos e como material de construção de colunas, bases, capitéis e cantaria em geral;
- **Xistos** – até à Idade do Ferro eram usados como suporte da arte rupestre, tendo depois sido utilizados como matéria-prima na construção muralhas e habitações;
- **Formações argilosas** – usadas no fabrico de cerâmicas;
- **Jazidas com ouro, prata, cobre, chumbo, ferro e zinco** – utilizadas para fabrico de utensílios e elementos decorativos.

2.2. Períodos cronológicos

O tempo é uma dimensão muito importante na Arqueologia. No entanto, tal como as acções que ocorreram no passado, o tempo é invisível. Podemos percebê-lo mas não

podemos tocá-lo. A percepção do tempo, para épocas passadas, não se esgota com o uso de cronologias, mas está intimamente associada a objectos, tecnologias e cultura arqueológica [Gamble, 2002].

Sendo o tempo, um dos vectores necessários para dar forma ao passado, a Arqueologia tem dedicado grande parte dos seus recursos na sua medição e na definição de conceitos temporais. As diferenças de relacionamentos entre comunidades e do Homem com a natureza, dividiram a escala cronológica em compartimentos mais ou menos estanques.

Alguns desses compartimentos são caracterizados a seguir, nomeadamente os períodos do Neolítico, Calcolítico, Idade do Bronze, Idade do Ferro, Período Romano e Idade Média [MNAE, 1989] [Oxford, 1996].

2.2.1. Neolítico

O Neolítico está associado às alterações climáticas, após o fim do último período glaciar. Na Europa os sítios mais antigos, pertencentes a este período, foram encontrados no sul e pensa-se serem de uma data próxima do ano 7000 a.C..

O Neolítico caracteriza-se pelo desenvolvimento da agricultura e pastorícia, bem como pelas inovações tecnológicas, como o polimento da pedra, o fabrico de cerâmica e a tecelagem. Passou-se de uma economia simples, baseada na caça e na recolecção, para uma economia mais complexa, onde predominam as sociedades que se dedicam à produção de alimentos e domesticação de animais.

As grutas e abrigos continuam a servir de *habitats* e necrópoles. No entanto, começam a aparecer os primeiros povoados, localizados nas cercanias de terrenos férteis e de cursos de água. Esta organização permite a produção de excedentes que incentiva o intercâmbio com outras comunidades.

2.2.2. Idade do Cobre ou Calcolítico

A Idade do Cobre terá tido início durante o quinto milénio a.C. e caracterizou-se por uma nova economia. Os animais eram utilizados para tracção e transporte. Utilizaram-se novas tecnologias que ajudaram a desenvolver a agricultura. O cobre começou a ser

utilizado no sudeste do continente Europeu, contribuindo para o aparecimento de uma nova actividade económica, a metalurgia.

Há um maior aproveitamento dos solos uma vez que, mesmo os solos mais pobres, passaram a ser utilizados. O desenvolvimento da metalurgia trouxe inovações tecnológicas na agricultura, nomeadamente com os sistemas de irrigação artificial construídos.

Esta nova economia acentua a necessidade dos pequenos grupos sociais, ou comunidades, de defenderem a propriedade e controlarem os excedentes. Para tal, construíram-se muralhas, com torres e bastiões redondos.

Os rios são muito utilizados como vias de comunicação naturais. Os povoados calcolíticos privilegiam os locais elevados, de encosta, e próximos dos vales férteis que ladeiam os principais rios.

2.2.3. Idade do Bronze

A Idade do Bronze desenvolve-se na Europa a partir de 2300 a.C. e perdurou até ao ano 800 a.C..

O período calcolítico desenvolveu a actividade metalúrgica que diversificou as actividades económicas existentes. Como resultado desta nova economia, destacam-se na Idade do Bronze, centros proto-urbanos com funções centrais, em contraponto com o mundo rural composto por pequenas e dispersas unidades agrícolas. Estas novas formas de organização económica e social contribuíram para o aparecimento dos mais antigos Estados organizados da História. A produção intensiva de utensílios de bronze terá sido a alavanca que impulsou o surgimento desta nova etapa cultural, denominada por Idade do Bronze.

2.2.4. Idade do Ferro

O início da Idade do Ferro não foi marcado pela descoberta deste metal, mas pela sua utilização em larga escala. O início deste período está associado, para grande parte do território Europeu, a uma data próxima dos 700 anos a.C..

A metalurgia do ferro e a produção de material bélico fez surgir desequilíbrios de forças e uma nova economia centrada na guerra. Acentuaram-se as desigualdades sociais, sendo frequente encontrar para este período, sepulturas que evidenciam sinais de riqueza, muitas vezes associadas a residências fortificadas.

Desenvolveram-se os centros urbanos e procuraram-se matérias-primas, em especial os metais, mesmo em territórios fora do espaço Europeu. É no período da Idade do Ferro que ocorre a ruptura definitiva entre a Proto-história e a História. Muitas das estruturas sociais, políticas e culturais Europeias, que irão perdurar até aos nossos dias, têm aqui o seu ponto de partida. Em muitos pontos da Europa, como os Balcãs e as Penínsulas Itálica e Ibérica, desenvolveram-se sociedades urbanas complexas, com sofisticadas tradições no campo das artes e da arquitectura.

2.2.5. Período Romano

O império Romano nasce por volta do ano 27 a.C., embora, nessa altura, já houvesse na Europa um extenso território sob o domínio Romano. É no tempo do imperador Trajano (117 d.C.) que o império atinge a sua maior dimensão europeia.

A romanização dos territórios conquistados introduziu alterações a nível político, económico, social e tecnológico. Estas reflectiram-se profundamente no mapa do povoamento e produziram várias discrepâncias relativamente à matriz de povoamento proto-histórico.

A força do organizado exército romano impôs um poderio sobre os territórios conquistados, tendo sido posteriormente organizados administrativamente. No entanto, a difusão e distribuição de bases militares, a política romana de encorajar as populações locais a adoptar costumes romanos, o estilo de vida urbano e ainda a intensa actividade económica com territórios conquistados, contribuiu para criar uma unidade cultural. O latim passou a ser a língua dominante, unificando-se o direito, a moeda, os padrões de pesos e medidas e o calendário.

2.2.6. Idade Média

Considera-se a Idade Média como o período da história europeia que vai desde a queda do Império Romano do Ocidente, em finais do século V, até ao século XV, quando se

dá a afirmação do capitalismo sobre o modo de produção feudal, o florescimento da cultura renascentista e os grandes descobrimentos.

A Idade Média europeia divide-se em duas fases bem distintas. A Alta Idade Média, que vai da formação dos reinos germânicos, a partir do século V, até a consolidação do feudalismo, entre os séculos IX e XII, e a Baixa Idade Média, que vai até ao século XV e se caracteriza pelo crescimento das cidades, a expansão territorial e o florescimento do comércio.

2.3. Prospecção arqueológica

A Arqueologia estuda a actividade humana das comunidades antigas, recorrendo a documentos, objectos, estruturas e outras marcas deixadas em determinados ambientes, chamados sítios arqueológicos ou arqueossítios. Essas estruturas chegam ao conhecimento dos arqueólogos fruto de descobertas ocasionais, como é o caso de muitos sítios arqueológicos, mas também, como resultado de indicações dadas por informadores locais ou documentos escritos.

A actividade de pesquisa de sinais do passado denomina-se prospecção arqueológica.

Tradicionalmente a prospecção arqueológica, ou seja, a localização de arqueossítios era feita com base em pesquisas bibliográficas, na recolha de informações orais e na observação directa do terreno. Mais tarde foram introduzidas novas metodologias das quais se destacaram as de "*fieldwalking*" ou de definição de faixas de amostragem ("*sampling survey*"), que permitem a recolha de materiais arqueológicos em amplas extensões.

No entanto, nem todos os arqueossítios são indiciados por materiais à superfície, podendo estar totalmente soterrados. Algumas estruturas foram cobertas por acção dos agentes naturais, como o vento e a chuva, que as cobrem com sedimentos. Outras porém, foram intencionalmente enterradas pelo Homem, como por exemplo os túmulos e necrópoles e mesmo alguns tesouros. Estes sítios obrigam a uma prospecção com recurso a métodos e técnicas complementares, nomeadamente os métodos de prospecção indirecta, como a observação de fotografias aéreas ou a utilização de técnicas geofísicas.

Apresentam-se a seguir algumas técnicas utilizadas para a prospecção arqueológica, tais como a documentação bibliográfica, o *fieldwalking,* estudo do terreno, fotografia aérea, resistividade eléctrica, estudos magnéticos, acústicos, análise químicas e as tecnologias de informação.

2.3.1. Documentação bibliográfica

Os documentos escritos, como textos, cartas e toponímia, são fontes onde se encontram referências a vestígios do passado, que por vezes podem até já não existir. A análise destes documentos constitui um dos pontos de partida para a prospecção arqueológica. As informações que fornecem reportam-se à época em que foram escritos, mas também a épocas passadas, uma vez que recolhem tradições orais locais, as quais constituem preciosos indicadores para a localização de arqueossítios [Fernández, 1977]. Algumas destas indicações conduzem a pistas que, muitas vezes, permitem a descoberta de importantes sítios arqueológicos, em todo o mundo. No entanto, recuando no tempo para períodos como a Idade do Ferro ou anteriores, são escassos os registos escritos que fazem alusões a sítios arqueológicos, pelo que outras metodologias são imprescindíveis para a identificar a sua localização.

2.3.2. Estudo do terreno

Os cenários naturais onde se instalaram os sítios arqueológicos, caracterizados pela geomorfologia do terreno, hidrografia e edafologia, terão naturalmente influenciado as estratégias de povoamento adoptadas pelas várias civilizações. O estudo dos terrenos é fundamental para tentar visualizar qualquer indício de ocupação passada, mas também para perceber possíveis alterações que nele se verificaram, ao longo do tempo, identificando-se assim possíveis locais onde os sítios desapareceram, devido à erosão, deposição de sedimentos ou inundações [Renfrew e Bahn, 1991].

Independentemente dos documentos escritos e vestígios existentes que reportam a locais arqueológicos, o conhecimento do terreno é fundamental para se realizar uma prospecção arqueológica.

A prospecção arqueológica implica, portanto, o estudo do terreno através da observação directa da macro e micro topografia do local, da hidrografia e da vegetação,

bem como do estudo de possíveis alterações climáticas, movimentos de terra e características dos sedimentos.

Algumas técnicas de prospecção geofísica (eléctrica, magnética, acústica ou térmica) podem complementar a análise dos terrenos, permitindo investigar o subsolo e detectar variações na sua estrutura.

2.3.2.1. *Fieldwalking*

A técnica de *fieldwalking* consiste em percorrer de uma forma sistemática os terrenos da zona a prospectar, recolhendo e registando artefactos e estruturas encontradas à superfície. Os fragmentos de cerâmicas, por exemplo, que muitas vezes aparecem dispersos no terreno são indicadores de vestígios no subsolo. Os muros dos edifícios modernos podem também fornecer importantes indicações quando possuem inscrições e fragmentos de materiais que, pelas suas características, podem ter pertencido a sítios arqueológicos proximos.

2.3.2.2. Fotografia aérea

A observação de fotografias aéreas tem sido uma metodologia utilizada para a localização de sítios arqueológicos.

Na análise dos fotogramas procuram-se anomalias no relevo, que possam indiciar a existência de um arqueossítio no local. O princípio que está na base desta técnica é o de que os povoados, ou outros tipos de sítios abandonados, foram cobertos por sucessivas camadas de sedimentos, alterando as formas normais do relevo.

Essas alterações podem traduzir-se por variações na vegetação onde se identificam diferentes índices de crescimento, ou anomalias no relevo que evidenciam alterações ao nível do subsolo.

O crescimento diferenciado da vegetação é muitas vezes provocado por alterações do solo, resultantes de actividades humanas anteriores [McGill, 1995] (Figura 2.3).

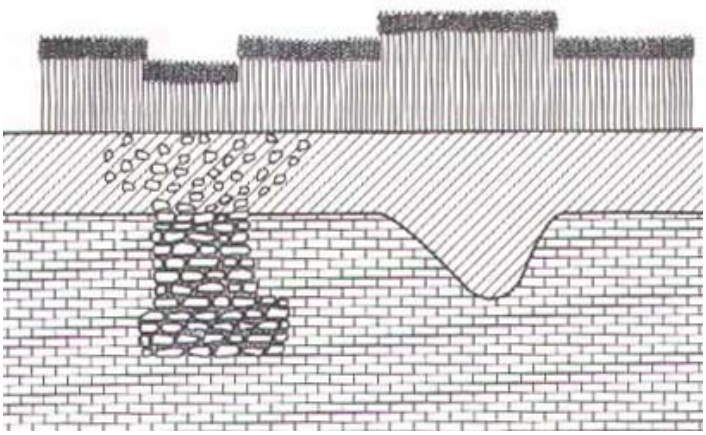


Figura 2.3 – Crescimento diferenciado dos vegetais sobre estruturas enterradas

As alterações de relevo são muitas vezes imperceptíveis quando observadas de perto, mas perfeitamente visíveis à distância, indiciando a existência de estruturas subterrâneas [Fabre, 1992] (Figura 2.4.).

Figura 2.4 – Foto aérea de estruturas arqueológicas (adaptado de [Green, 2002])

A leitura da fotografia aérea permite analisar diferenças existentes no relevo, na cobertura vegetal ou ainda na coloração dos solos.

As manchas de coloração no solo podem resultar de diferentes concentrações de humidade. As zonas onde existiram antigos canais ou poços são zonas de grande

humidade, aparecendo mais escuras no fotograma. Áreas mais claras podem indiciar a existência de estruturas no subsolo, que impedem uma maior absorção de água. Quando existe vegetação não é possível observar a coloração dos solos, mas a diferença de humidade influencia também o crescimento da vegetação.

A fotografia aérea permite ainda compreender a estrutura geográfica global de uma zona, extremamente importante para os projectos de prospecção arqueológica. Os povoados inserem-se num contexto geográfico e compreendê-lo é estar mais perto das razões que levaram os seus habitantes a optar por determinados locais, em detrimento de outros.

2.3.2.3. Resistividade eléctrica

A utilização da resistividade eléctrica como método de prospecção arqueológica apresentou-se como uma das primeiras técnicas de prospecção geofísica aplicadas à Arqueologia. Este método baseia-se no princípio da idêntica resistividade para idênticos materiais. Assim, a resistividade dos solos deveria ser mais ou menos uniforme. Medidas de resistividade díspares podem dever-se à existência no subsolo de diferentes materiais.

Para utilizar esta técnica injecta-se uma corrente eléctrica no terreno e faz-se a leitura dos valores da resistividade eléctrica encontrada para diferentes pontos [Figueiredo, 1995]. Sabendo que a resistividade da pedra é superior à da terra, podem fazer-se medições orientadas e proceder-se ao registo das oscilações de resistividade.

2.3.2.4. Estudos magnéticos

Esta técnica baseia-se na medição de valores do campo magnético. Tal como na técnica da resistividade eléctrica também aqui se fazem comparações e quando se registam anomalias, concluiu-se que devem ser o resultado do efeito de depósitos de materiais [McGill, 1995]. Se estiverem enterradas estruturas e artefactos verificam-se alterações no campo magnético que são detectadas pelas medições efectuadas.

Esta técnica deve no entanto ser complementada com outras técnicas, uma vez que as alterações de valores podem ser diferentes conforme o local, a área e o tipo de materiais que formam o subsolo.

2.3.2.5. Acústica

A prospecção acústica faz-se pela introdução no solo de uma vara metálica que, depois de agitada, funciona como um diapasão emitindo ondas sonoras [Fernández, 1977]. Essas ondas irão ser propagadas e a sua recepção diferenciada, em locais distintos, pode indiciar a existência de anomalias no terreno.

Tal como nas duas técnicas anteriormente apresentadas conclui-se que diferentes leituras de valores se podem dever à existência de diferentes materiais no subsolo. Esses materiais poderão ter resultado de estruturas soterradas ou artefactos.

2.3.2.6. Análises químicas

As metodologias de avaliação das características químicas dos solos, baseiam-se no princípio de que alterações de valores estarão relacionadas com actividades humanas ou animais.

A deposição no solo de resíduos humanos ou animais é detectada através da análise dos níveis de fosfatos. Um nível elevado de fosfatos indica a presença de grande actividade animal. A maior dificuldade deste método está associada à incapacidade de distinguir a actividade humana das actividades dos outros animais, pelo que necessitará de ser complementada com outras técnicas de prospecção.

2.3.2.7. Tecnologias de informação

Os fenómenos naturais de erosão, processos pós-deposicionais e as acções do Homem sobre o solo, produzem alterações que actuam sobre os vestígios arqueológicos, desempenhando um papel desorientador e que dificulta, muitas vezes, a prospecção de superfície.

O trabalho arqueológico com recurso às TI constitui mais uma metodologia, complementar aos métodos tradicionais, uma vez que apoia a localização, exploração, documentação e recuperação do máximo de dados, onde as metodologias tradicionais podem deixar lacunas.

Nesta área têm sido utilizadas técnicas de apoio à prospecção arqueológica destacando-se a utilização de SIG, e em fase ainda muito embrionária a DCBD [Botica *et al.*, 2003b].

Os SIG têm sido amplamente utilizados na Arqueologia com vários objectivos, nomeadamente na investigação arqueológica [Cruz e Sánchez, 1997], na gestão administrativa do território [Espiago e Baena, 1997], no tratamento e gestão de imagens [Preysler *et al.,* 1997] e como ferramenta de apoio aos trabalhos de prospecção e escavação [Wheatley e Gillings, 2002]. Na área da prospecção arqueológica têm sido desenvolvidos alguns modelos preditivos como sistemas de apoio à gestão patrimonial. Partindo de variáveis ambientais e da localização dos assentamentos arqueológicos, estabelecem-se padrões estatísticos de ocupação para a área de estudo definida [Sánchez, 2000].

Embora estas ferramentas de SIG tenham já dado valiosos contributos na área da Arqueologia, apresentam, normalmente a característica de considerarem apenas factores geográficos e temporais [Kuiper e Wescott, 1999]. Sendo esta uma área que estuda a actividade humana do passado, os factores económicos, humanos e sociais terão também influenciado a escolha dos locais de povoamento e do seu modo de vida. Se estes factores não forem considerados, a procura dos vestígios do passado será sempre incompleta. Esta é uma preocupação presente no desenvolvimento actual dos SIG [Leusen, 2002] e é uma mais valia apresentada, à partida, pelos sistemas que recorrem à DCBD, onde se procuram padrões nos dados, sejam eles ambientais, sociais ou culturais. Os achados arqueológicos já recolhidos e estudados têm uma localização, uma forma, um período de utilização, mas também um contexto. Se considerarmos que estes achados contêm implícita alguma desta informação, então estamos perante um cenário onde a DCBD poderá dar um valioso contributo.

2.4. O conhecimento arqueológico para Trás-os-Montes Oriental

Ao longo da Idade Moderna apareceram várias monografias onde se encontram referências arqueológicas relativas à região de Trás-os-Montes Oriental. As fontes documentais conjuntamente com achados arqueológicos, são sempre um instrumento de trabalho e muitas vezes um ponto de partida para o trabalho de prospecção e escavação.

Tal terá sido o caso das primeiras escavações realizadas por Henrique Pinheiro, em 1887 e sob a égide de Martins Sarmento, onde incentivado pelo interesse no passado e guiado pelas referências arqueológicas existentes, iniciou as intervenções arqueológicas em Trás-os-Montes, na esteira do povoado que teria dado origem à cidade de Bragança [Lemos, 1993].

Depois do impulso inicial dado por Martins Sarmento, foram vários os arqueólogos que, quer com intervenções pontuais, quer pelo seu trabalho mais contínuo e consolidado, colocaram a região de Trás-os-Montes Oriental no mapa da Arqueologia portuguesa. Leite de Vasconcelos é uma referência obrigatória, dado ter sido ele que após a era de Martins Sarmento, impulsionou e dinamizou o estudo da Arqueologia no Nordeste Transmontano, contagiando muitas figuras locais e nacionais que com ele colaboraram para desenvolver a Arqueologia transmontana. O Abade de Baçal foi também responsável por ter criado uma cultura arqueológica na região, tendo a sua obra impulsionado o interesse pela Arqueologia e a realização de estudos regionais [Lemos, 1993].

Embora os trabalhos iniciais de Arqueologia na região fossem fundamentalmente trabalhos de recolha bibliográfica e de registo de escavações e achados avulsos, eles constituem uma fonte de informação muito importante, tendo impulsionado a realização dos estudos arqueológicos que se seguiram. Destes destacam-se, para além da localização de dezenas de sítios e achados arqueológicos, o estudo do traçado da via romana, da distribuição espacial dos povoamentos da Idade do Ferro e do Período Romano, o estudo dos termos das comunidades e a demarcação dos territórios, o sistema defensivo dos povoados fortificados e a arte rupestre.

Na década de 80 assiste-se a uma transformação nos estudos arqueológicos, que passaram a ter um carácter menos regionalista e a haver intervenções continuadas. Neste período intensificaram-se as actividades arqueológicas, aumentou o número de investigadores a trabalhar na região, de inventários dos sítios arqueológicos, alargou-se e reviu-se todo o acervo informativo existente e ampliaram-se os projectos científicos. Foi também por esta altura que despontaram acções de emergência ou salvamentos, sob o patrocínio do Serviço Regional de Arqueologia da Zona Norte, e que puseram em

evidência novos arqueossítios, importantes para uma melhor definição dos contornos da Arqueologia na região.

Dos projectos de inventário realizados destacam-se o Projecto de Inventário de Sítios Arqueológicos do Planalto de Miranda, realizado por Domingos Marcos e concluído 1983, o Projecto de Inventário da Terra Quente Transmontana, realizado pela Universidade do Minho em 1984-85 e também o inventário de sítios da Torre de Moncorvo, realizado pela equipa do Projecto Arqueológico da Torre de Moncorvo [Lemos, 1993].

F. Sande Lemos e Maria de Jesus Sanches foram, no entanto, os responsáveis pelos únicos trabalhos de síntese da região de Trás-os-Montes Oriental, caracterizando detalhadamente o povoamento da Idade do Ferro e do Período Romano da região [Redentor, 2002].

2.4.1. Matriz de povoamento para a Idade do Ferro

As forças decorrentes do poder de quem tem superioridade de material bélico, produzido na metalurgia do ferro, lançaram uma nova economia centrada na guerra.

O povoamento é por isso organizado com base nos *habitas* fortificados, situados em locais de boas condições naturais de defesa e de visibilidade.

A Figura 2.5 representa a matriz de povoamento na região de Trás-os-Montes Oriental, para o período da Idade do Ferro.

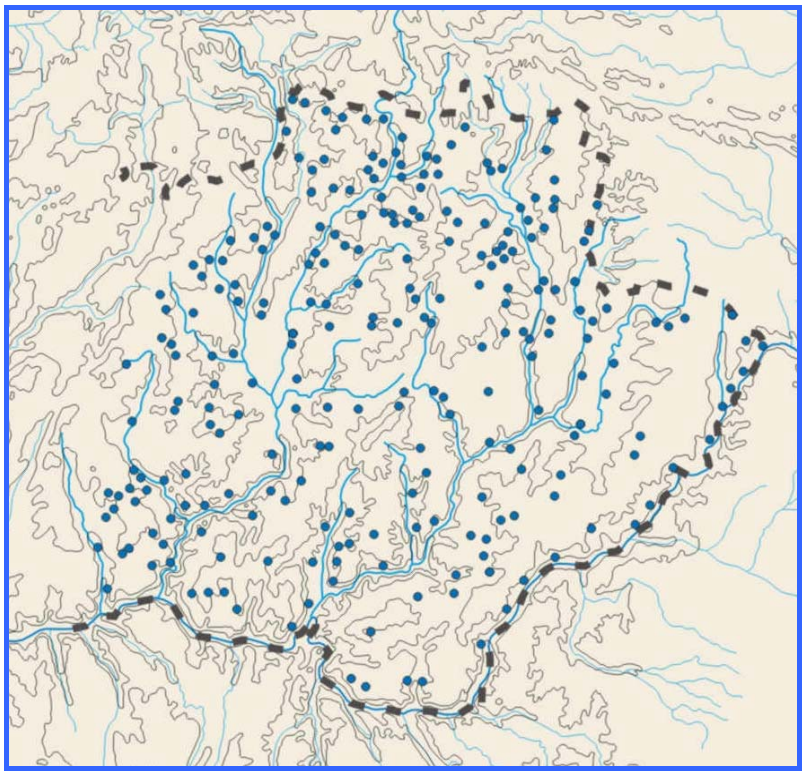


Figura 2.5 – Povoamento da Idade do Ferro em Trás-os-Montes Oriental (adaptado de [Lemos, 1993])

Observando a carta de distribuição dos povoados da Figura 2.5, pode atentar-se nas seguintes características da matriz de povoamento:

- Elevada densidade de povoamento da região, nomeadamente na área de Vinhais e Bragança, diminuindo no sentido noroeste-sudeste;
- Núcleos de maior concentração a par de áreas vazias;
- Maior regularidade e escalonamento ao longo dos cursos de água;
- Distribuição espaçada no rebordo dos planaltos.

No entanto, a matriz de povoamento de Trás-os-Montes Oriental, bastante complexa, torna-se mais simples de interpretar quando se faz o mapeamento desta carta de povoamento com as diversas cartas geográficas, de solos, pluviométricas e de jazidas de minério. Desse cruzamento, resultam algumas evidências que ajudam a caracterizar a matriz de povoamento da região. Destas características realçamos, a seguir, os aspectos que se julgaram mais relevantes para a construção do modelo preditivo [Lemos, 1993]:

- Nos planaltos aparecem assentamentos nas orlas, quando estes confinam com rios profundamente encaixados ou com montanhas;
- Nos pontos centrais dos planaltos os povoados localizam-se normalmente em zonas estratégicas associadas a cristas quartzíticas;
- As vertentes escarpadas de relevos residuais, vales encaixados e linhas de fractura foram utilizadas como locais de assentamento, aproveitando as boas condições defensivas proporcionadas;
- Nos vales encaixados os *habitats* situam-se em esporões ou arribas;
- Nos vales abertos os povoados distribuem-se ao longo dos cursos de água principais;
- A sobreposição das curvas de escoamento de água na rede hidrográfica e a matriz de povoamento revela uma forte dependência destes relativamente aos recursos hídricos;
- A sobreposição do povoamento e a carta de solos mostra que a maioria dos *habitats* estão em locais com solos aráveis, embora de fracas aptidões agrícolas. Solos mais pesados não eram aproveitados, dado não poderem ser trabalhados pelos rudimentares instrumentos agrícolas da altura [Cruz, 2000];
- Da sobreposição da carta de povoamentos da Idade do Ferro com as cartas de precipitação, denotam-se algumas associações, distribuindo-se os povoados entre as curvas de 800mm[4] e os 1 400 mm;
- A litologia parece ter condicionado a escolha e densidade de alguns locais de povoamento. As zonas graníticas são mais procuradas porque geram solos com maior retenção de água, por serem mais leves e, portanto, mais favoráveis à agricultura. As zonas de rochas básicas e ultra-básicas são também favoráveis às actividades agrícolas, porque os solos são menos ácidos e de com períodos de renovação mais curtos;

[4] mm - unidade de medida da precipitação. A precipitação pode medir-se em milímetros por metro quadrado. 800 mm significa que em média a água da chuva que cai enche um quadrado, com 1 m^2 de área, até à altura de 800 mm.

- Os recursos mineiros parecem não ter influenciado directamente a escolha dos locais de assentamento, mas terão contribuído para aumentar a densidade dos pequenos e médios povoados, já existentes nos locais próximos das jazidas. A actividade mineira deveria ser na altura uma actividade do tipo familiar e artesanal.

2.4.2. Matriz de povoamento para o Período Romano

A romanização do território introduziu alterações a nível político, económico, social e tecnológico que se reflectiram profundamente no mapa do povoamento e produziram várias discrepâncias relativamente à matriz de povoamento que a antecedeu.

Durante o período Romano alguns *habitats* do período da Idade do Ferro mantiveram-se, tendo sido romanizados, mas outros foram abandonados. Fundaram-se também novos povoados durante este período (Figura 2.6).

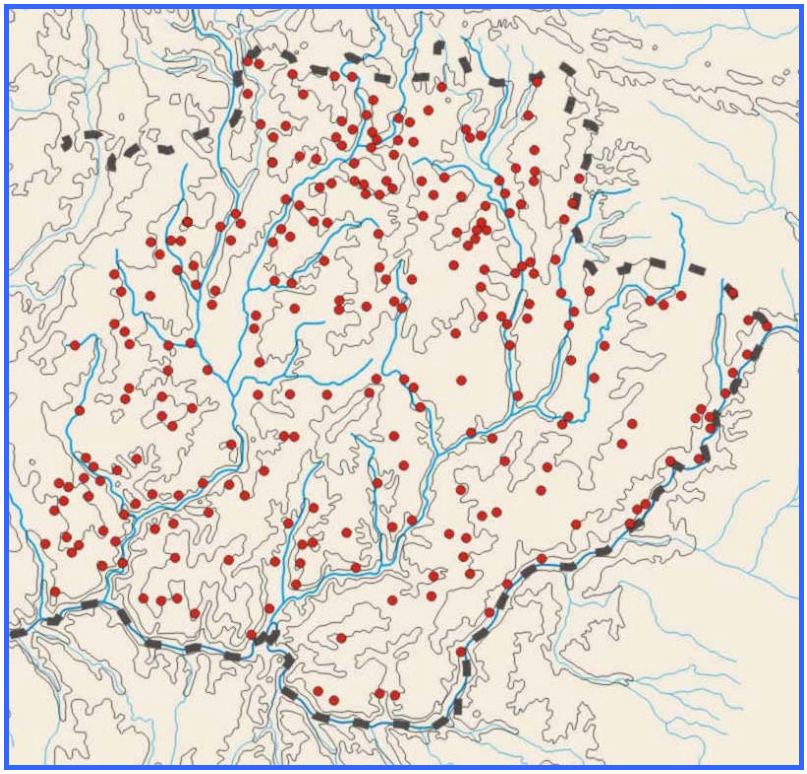


Figura 2.6 – Povoamento Romano em Trás-os-Montes Oriental (adaptado de [Lemos, 1993])

São vários os factores que, isoladamente ou em conjunto, foram responsáveis pelas alterações introduzidas durante a romanização [Lemos, 1993], dos quais destacamos os seguintes:

- O aspecto defensivo não era prioritário para a sociedade romana, tendo sido até incómodo durante os primórdios da romanização, pelo que muitos dos *habitats* da Idade do Ferro que possuíam boas condições defensivas e posicionamentos estratégicos sobre as vias de comunicação, poderão ter sido forçados ao abandono;
- O factor defensivo foi relevado em função das alterações económicas introduzidas. Passou a ter mais importância a proximidade de solos favoráveis à agricultura, pastorícia ou às actividades mineiras, do que o posicionamento em locais geo-estratégicos ou com boas condições de defesa;
- A organização política, económica e administrativa do território, a par com a crescente importância que trouxe às vias de comunicação, impôs uma nova organização territorial;
- Fundaram-se novos povoados que funcionavam como centros administrativos;
- Ao longo das vias de comunicação posicionaram-se sítios que serviriam como centros de acolhimento de pessoas e mercadorias.

Tal como aconteceu para o povoamento da Idade do Ferro, foram analisadas as cartas de povoamento Romano, em função de mapas ambientais que permitiram evidenciar algumas características [Lemos, 1993]:

- Os planaltos e o fundo das depressões tectónicas constituíram os cenários preferenciais da instalação de povoados romanos;
- Alguns *habitats* localizam-se no sopé de relevos residuais;
- A actividade mineira é muito expressiva no período Romano, notando-se uma articulação entre o mapa dos sítios romanos e os recursos mineiros. Desta actividade destaca-se a exploração mineira de ouro, prata, estanho, chumbo, ferro, mármore, pedra de talco e granitos;

- A capacidade do uso de solos é um factor que certamente terá influenciado o posicionamento de muitos *habitats.* O cultivo de cereais, vinha, oliveira foi bastante estimulada, havendo no período romano maior facilidade de trocas comerciais, dada a extensa rede viária criada. Nota-se portanto uma interdependência grande da malha de povoamento Romano e as áreas de solos de maior potencial agrícola, distanciando-se do modelo anterior onde eram valorizadas as características defensivas e a economia era baseada no aproveitamento harmonioso dos recursos naturais;
- Apesar de haver forte relação da matriz de povoamento com o tipo de solos, não se encontra uma interdependência do povoamento com a retenção de água na rede hidrográfica, os níveis de pluviosidade e as temperaturas. Tal facto pode ser atribuído à utilização de novas técnicas de cultivo, como a policultura, o aproveitamento de solos adversos ao cultivo através da criação de antropossolos e do uso de tecnologias de regadio que levaram ao aproveitamento de áreas que até aí não eram utilizadas para a agricultura.

A matriz de povoamento Romana não é, no entanto, influenciada apenas por factores ambientais. Conjugados com o quadro ambiental, sumariamente enumerado acima, articularam-se ainda outros factores, destacando-se os de ordem político-administrativa e cultural.

O cenário imposto pelo Império Romano é dominado por uma normalização dos *habitats*, criando diferentes tipologias de acordo com as principais funções desempenhadas. Num curto período de tempo passou-se da existência de apenas um tipo de *habitat*, o povoado fortificado da Idade do Ferro, para uma diversidade de povoados, na maioria abertos e que constituíam a malha de povoamento da época Romana. Desses povoados destacam-se alguns tipos, diferenciados de acordo com as actividades e funções desempenhadas, tais como os centros administrativos, os *habitats* rurais, os *habitats* articulados com a rede viária e ainda os povoados mineiros.

Complicando um pouco mais o cenário, a matriz de povoamento romana contou ainda com a interferência de factores locais de ordem cultural, uma vez que existiam no

território de Trás-os-Montes Oriental diferentes etnias, que aceitaram de forma diversa o processo de romanização.

2.5. Conclusão

A prospecção marca o início da actividade arqueológica. A Arqueologia recorre a várias técnicas de prospecção, podendo a DCBD ser uma nova ferramenta de apoio a esta actividade.

Para atingir esta finalidade, definiu-se uma metodologia de trabalho que tem como um dos objectivos a compreensão do problema e dos conceitos associados ao domínio de aplicação, onde se insere a temática abordada neste capítulo. Estudou-se o conhecimento arqueológico existente, relacionando-se as actividades de prospecção, com o estudo da região de Trás-os-Montes Oriental, como território alvo do trabalho, os factores temporais e ambientais e, ainda, com a malha de povoamento já identificada. Este estudo é fundamental para o processo de DC, permitindo que, em conjunto com o especialista de Arqueologia, se faça a necessária incorporação de conhecimento no modelo ***servator***.

No capítulo seguinte será feita uma revisão bibliográfica ao processo de DCBD, estudando-se as várias fases do processo e analisando-se as tarefas e técnicas adoptadas para a identificação do modelo de apoio à prospecção arqueológica, proposto neste trabalho.

Capítulo 3

A Descoberta de Conhecimento em Base de Dados

As BD e as TI, que lhe estão associadas, têm-se desenvolvido no sentido de permitir o armazenamento e utilização de grandes quantidades de dados. As organizações não se alhearam desta evolução e aproveitaram as potencialidades oferecidas por estas tecnologias. Ampliaram as BD que passaram a ter um papel importante no suporte à tomada de decisões. Todavia, apenas uma pequena parte desses dados é analisada e utilizada como instrumento de apoio à decisão ou na construção de conhecimento. A restante é armazenada, na perspectiva de que mais tarde pode vir a ser útil.

Com o crescimento exponencial dos dados armazenados começou a haver necessidade de desenvolver novas ferramentas de análise e organização dos mesmos. As tradicionais ferramentas de pesquisa nas BD começaram a ser manifestamente insuficientes, para obter informação legível e útil, a partir de grandes quantidades de dados. Esta necessidade impulsionou o aparecimento de ferramentas de *Online Analytic Processing* (OLAP) e de DC.

O ***servator,*** modelo preditivo de apoio à prospecção arqueológica, que constitui a finalidade deste trabalho, recorre ao processo geral de DC a partir de dados, habitualmente designado por DCBD. Este processo utiliza técnicas de DM, aplicando aos dados algoritmos de extracção de padrões, e incorporando conhecimento do domínio de aplicação, através da adequada interpretação de resultados.

Neste capítulo apresentam-se as fases do processo de DC, caracterizando-se as tarefas a desenvolver e as técnicas de DM que podem ser utilizadas na identificação de padrões nos dados.

3.1. Princípios

Na década de 80 assistiu-se a uma revolução no domínio dos equipamentos, caracterizada pela sua miniaturização, standardização de componentes, custos mais baixos e desenvolvimento de sistemas de gestão de BD, largamente difundidos e de utilização simplificada [Zorrinho, 1991]. Este avanço tecnológico, aliado a conjunturas económicas e sociais favoráveis, estimularam a utilização de TI. Estas rapidamente se tornaram num factor diferenciador e estratégico para as Organizações.

No entanto, a capacidade limitada para processar informação, começa a diminuir quando se atinge o ponto de saturação, mesmo utilizando poderosas ferramentas de pesquisa de BD. A partir deste limite, o aumento do volume de dados não se traduz num aumento da sua utilização eficaz, ou em informação útil. Poderá até equivaler a uma diminuição, conforme se pode analisar no gráfico da Figura 3.1.

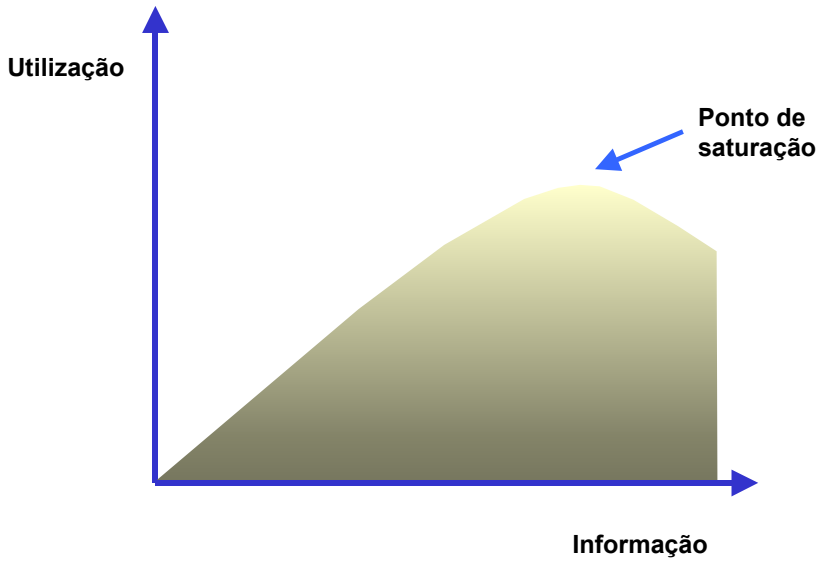


Figura 3.1 – Saturação na utilização da Informação (adaptado de [Amaral e Varajão, 2000])

Os diversos repositórios de dados construídos, vão armazenando grandes quantidades de informação. Esta pode constituir uma mais valia importante, quando dela se retira conhecimento mas, pode tornar-se impossível de digerir completamente com as tradicionais ferramentas de análise de dados. Esta noção de que os grandes repositórios de dados poderiam proporcionar informação mais útil e levar à DC (Figura

3.2), reuniu saberes de diferentes áreas, nomeadamente da Gestão de BD, Estatística, Inteligência Artificial, Aprendizagem Automática, Reconhecimento de Padrões e Visualização de Dados [Fayyad *et al*., 1996b], para desenvolver uma nova geração de tecnologias. Estas tecnologias utilizam dados, mas colocam um grande enfoque na DC que eles podem proporcionar.

Figura 3.2 – Pirâmide de dados *versus* conhecimento

O princípio que está na base desta geração de tecnologias é o de apoiar de forma inteligente e, o mais automática possível, o processo de navegar e analisar os dados para extrair informação útil, ou seja, conhecimento.

3.2. Fases do processo

A DCBD é um processo interactivo e iterativo, que deve ser desenvolvido dando especial atenção à metodologia utilizada, à incorporação de conhecimento na área e à apresentação e visualização de resultados. O conhecimento identificado deve ser apresentado de forma a ser compreendido e avaliado pelos especialistas [Han e Kamber, 2001].

A incorporação de conhecimento existente, deve funcionar como guia ao longo de todo o processo de descoberta. Este desenvolve-se em várias etapas, muitas delas executadas mais que uma vez, e onde é constante a tomada de decisões, de acordo com o conhecimento de especialistas ou competências técnicas da área dos dados a analisar [Fayyad *et al*., 1996a].

As várias fases do processo de DCBD compreendem tarefas centradas nos dados. No final de cada tarefa pode existir a necessidade de refazer tarefas anteriores, para incluir alterações identificadas. A interactividade é também uma característica deste processo. Sem a intervenção humana e a incorporação de conhecimento existente na área podem-se identificar padrões, mas estes podem não ser válidos ou úteis.

As etapas que integram o processo de DC incluem a **Selecção**, o **Tratamento** e **Pré-processamento** dos dados. Segue-se a aplicação de algoritmos de **DM** para identificar padrões implícitos nos dados, a **Interpretação e Validação** dos resultados. Todas as fases são muito iterativas, podendo em cada fase avançar-se para a fase seguinte ou recuar para qualquer uma das fases anteriores, como se ilustra na Figura 3.3.

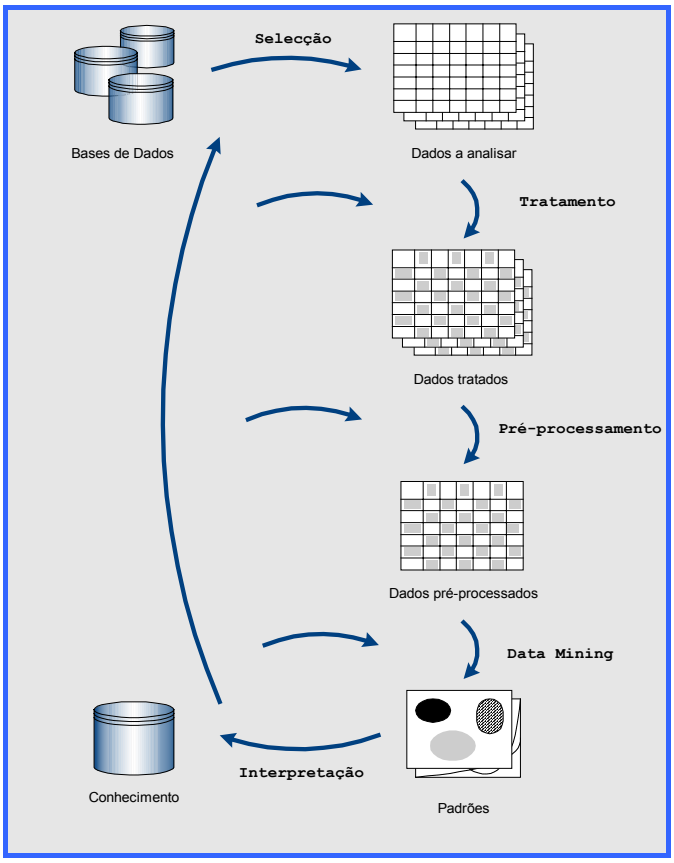


Figura 3.3 – Fases do processo de DCBD (adaptado de [Fayyad *et al.*, 1996a])

3.2.1. Selecção dos dados

A selecção de dados é a primeira tarefa do processo de DCBD, depois de definidos os objectivos a atingir, conforme a metodologia CRISP-DM, apresentada no capítulo 1. De

acordo com estes objectivos, é necessário avaliar os dados disponíveis e escolher aqueles que se julga serem relevantes para o processo. Esta selecção deve ser feita em parceria com o especialista na área que, sendo alguém muito familiarizado com os dados e com os objectivos do trabalho a desenvolver, ajudará a fazer a análise e selecção dos dados relevantes para o processo de DC [Microsoft, 2003].

Recolhidos e seleccionados os dados, passa-se ao tratamento e pré-processamento dos mesmos. Quando se aplicam técnicas de DM directamente sobre os dados recolhidos, sem que estes sejam previamente tratados, dificilmente se encontrarão bons modelos. As BD possuem normalmente uma elevada dimensionalidade, associada a uma grande imprecisão dos dados, o que compromete os resultados apresentados pelos algoritmos de DM. Para atenuar esses efeitos é muito importante realizar, de forma cuidada, as tarefas de selecção, tratamento e pré-processamento dos dados. Sempre que possível, deve recorrer-se a técnicas que ajudem a completar os dados, eliminar atributos irrelevantes, condensar informação nas amostras seleccionadas e eliminar casos irrecuperáveis [Rodrigues, 2000].

Analisam-se de seguida, alguns aspectos a considerar, quando se faz a selecção dos dados, nomeadamente quanto à sua relevância, representatividade e volume necessário. A periodicidade de recolha dos dados é também um aspecto a analisar, para garantir que o modelo identificado é actualizado, sempre que necessário.

3.2.1.1. Relevância dos dados

As fontes de dados disponíveis para o processo de DC são, normalmente, repositórios resultantes de processos operacionais das Organizações. Estes podem estar dispersos, compartimentados, muitas vezes definidos com diversos formatos e em BD de elevada dimensionalidade. Na maioria dos casos as BD possuem centenas de tabelas, com milhares de registos e onde cada registo pode ainda ter um grande número de campos associados.

É necessário eliminar atributos que não tenham interesse para o processo de DC, nomeadamente aqueles que possuem apenas um carácter meramente informativo [Fayyad *et al.*, 1996b]. Nomes ou códigos de produtos não devem ser seleccionados,

reunindo-se apenas atributos que se julguem relevantes para o processo. Deve, sempre que possível, evitar-se grandes volumes de dados, que podem até inviabilizar as tarefas de DM. Esta tarefa deve ser realizada com a colaboração de um especialista no domínio de aplicação, para que não sejam eliminadas variáveis essenciais ao processo de DC.

3.2.1.2. Representatividade dos dados

A elevada dimensionalidade das BD, aliada ao facto destas armazenarem dados recolhidos com outros objectivos que não a DC, implica que se faça uma selecção dos dados a tratar.

O conjunto de dados escolhido deve ser representativo, de acordo com os objectivos definidos no processo de DCBD e a sua representatividade analisada. As ferramentas estatísticas podem ser úteis, porque permitem aferir a distribuição dos dados, avaliando se a amostra escolhida reflecte o conjunto completo de relações existentes na população [Pyle, 1999]. Deve também ser estudado o historial da recolha de dados, no contexto da Organização, para que eventos anómalos não influenciem a amostra. Por eventos anómalos podem considerar-se, por exemplo, problemas técnicos que tenham inviabilizado a recolha de determinados dados, que deixaram de ter representatividade na amostra em determinado período de tempo, ou ainda, acontecimentos resultantes de mudanças do sistema que não caracterizam um estado do mesmo.

3.2.1.3. Volume de dados necessário

Depois de definidos os objectivos do estudo e de estarem identificadas as fontes de informação, passa-se à fase de selecção dos dados, onde se recolhem aqueles que, à partida, são relevantes para o trabalho a realizar.

Reunidos esses dados numa BD e garantida a representatividade da amostra, avalia-se o volume de dados disponível para identificar o modelo. Quanto mais atributos forem seleccionados e, quanto maior for o número de valores a eles associados, maior será o volume de dados necessário para se criarem modelos mais sólidos. Um maior número de relações entre os dados implica que mais dados sejam utilizados, para que os algoritmos integrem nos modelos as respectivas relações [Pyle, 1999].

Apesar dos pressupostos atrás apresentados não se pode concluir que, de uma forma geral, seja mais seguro desenvolver um trabalho de DCBD, com grandes volumes de dados. Mais dados parecem garantir, *a priori,* maior representatividade e maior facilidade na aprendizagem dos modelos. No entanto, constituem também algumas desvantagens [Berry e Linoff, 2000], nomeadamente:

- Maior tempo gasto para construir o modelo, para o testar e apresentar resultados. Com um menor volume de dados é possível realizar estas operações com maior rapidez, podendo o restante tempo ser aproveitado para fazer experiências de modelação com novos parâmetros, variáveis ou algoritmos;
- Muitas ferramentas de modelação trabalham com os dados em memória central, para melhorar o desempenho, o que desde logo impõe restrições ao volume de dados a trabalhar;
- Um grande volume de dados pode provocar um *over-sampling*, isto é, ocorrências pouco frequentes adquirem um peso relativo mais pequeno, quando integradas numa grande amostra. Quando é importante que os modelos apresentem estas particularidades dos dados, os resultados são melhores com uma amostra mais pequena, porque aumenta a frequência com que dados menos comuns aparecem.

Por outro lado, reduzir o volume de dados pode não ser uma tarefa inócua. A eliminação de registos deve ser feita de forma aleatória, como recomenda a estatística. No entanto, realizar esta tarefa de forma automática constitui um processo pseudo-aleatório e de difícil replicação. Operações idênticas, feitas posteriormente, podem produzir resultados bastante diferentes.

Retirar dados de acordo com determinado critério pode também ser um processo com resultados imprevisíveis, uma vez que se corre o risco de retirar representatividade à amostra.

No final de um processo de remoção de registos devem ser usadas ferramentas estatísticas para avaliar a representatividade da amostra e, para garantir que os dados da amostra final, possuem uma distribuição idêntica à da amostra inicial.

Para gerar dados em falta devem utilizar-se, sempre que possível, novas fontes de dados, integrando-os na BD de trabalho e aumentando assim o volume de dados. No entanto, a integração de dados oriundos de múltiplas BD deve ser feita com especial cuidado, para evitar que sejam introduzidas redundâncias e inconsistências no conjunto final [Fayyad *et al.*, 1996a].

Pode também fazer-se a adição manual, ou automática, de novos atributos, obtidos a partir de registos existentes nas amostras de dados. Também esta tarefa deve ser bem avaliada, para que não se acrescentem atributos relacionados com outros já existentes na BD. A inclusão de relações já representadas nos dados não contribui para facilitar a aprendizagem dos modelos.

Muitos processos de DM requerem um conjunto de dados para identificar um modelo e um outro conjunto para o testar. Um primeiro conjunto, a que se dá o nome de Treino, será utilizado pelos algoritmos, no processo de aprendizagem, para encontrar padrões nos dados. O segundo conjunto, normalmente denominado de conjunto de Testes, engloba os restantes dados e serve para fazer a validação do modelo identificado. É essencial que os dois conjuntos sejam representativos da amostra, caso contrário o modelo não reflecte as relações aí existentes. Os modelos construídos utilizando algoritmos de DM podem apresentar problemas de sobre-ajustamento (*overfitting*) ou de sub-ajustamento (*underfitting*) [Santos, 2001]. No primeiro caso, o modelo identificado memoriza características dos dados, a partir das quais realiza as actividades de previsão. No segundo caso, dada a dimensionalidade reduzida da amostra, é identificado um modelo demasiado genérico, que não contempla relações interessantes que possam existir nos dados. Em qualquer caso, são sempre modelos com reduzida capacidade de previsão.

3.2.1.4. Periodicidade de recolha dos dados

As BD crescem continuamente sendo, muitas vezes, introduzidas alterações nos dados. Estas actualizações das BD podem produzir alterações das relações entre os dados e com reflexo nas regras produzidas. Assim, os modelos identificados devem prever a forma e a periodicidade com que alterações dos dados devem ser incorporadas.

A recolha de dados relativa a um curto período de tempo, para um sistema de longos períodos, implicará, por certo, uma ausência de representatividade de determinadas características. Por outro lado, em sistemas que variam pouco, não é necessário seleccionar dados relativos a pequenos períodos de tempo, porque se estaria a recolher informação repetida, de reduzido conteúdo informativo [Rodrigues, 2000].

Para sistemas muito dinâmicos é recomendável que os dados sejam actualizados, com uma certa periodicidade. Essa actualização visa garantir a sua representatividade, o refrescamento das relações entre dados e assegurar a actualização do modelo.

3.2.2. Tratamento dos dados

Os principais problemas que se enfrentam num processo de DCBD estão relacionados com os dados. As BD têm, normalmente, estruturas inadequadas ao processo de DM, os dados estão dispersos, muitas vezes com taxas significativas de dados omissos ou inconsistentes. As operações de limpeza dos dados assumem, assim, um papel muito importante e demorado, nem sempre fácil de executar. A maior parte dos problemas surgem devido aos dados omissos ou inconsistentes [Jermyn *et al.,* 1999].

Num projecto típico de DC os trabalhos de preparação dos dados podem consumir até 80% do esforço total do projecto. Apresentam-se, no gráfico da Figura 3.4, valores aproximados para o tempo consumido nas fases de preparação, análise e modelação dos dados [Pyle, 1999].

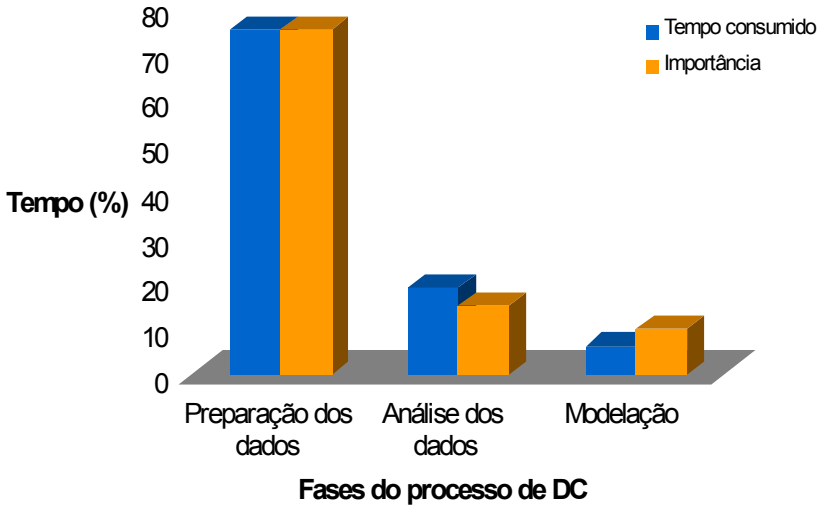


Figura 3.4 – Tempo/importância das fases do processo de DM

Sem bons dados não haverá bons resultados neste tipo de projectos, daí a grande importância desta fase do processo. Pesa ainda na relevância destas tarefas o facto de, operações de limpeza mal sucedidas, poderem provocar estragos nos dados [Pyle, 1999].

Os dados em falta, erros, registos duplicados com inconsistências são problemas muito frequentes e que exigem mais recursos temporais e humanos. A sua detecção e correcção, ou eliminação, requer quase sempre a intervenção humana e, acima de tudo, um bom conhecimento dos dados e contextos em que se inserem.

Muitos investigadores têm dedicado o seu trabalho para que as actividades de limpeza dos dados deixem de estar tão dependentes da actividade humana, e passem a ser automáticas ou semi-automáticas [Hernández e Stolfo, 1998].

No tratamento dos dados realizam-se operações de limpeza dos mesmos. Utilizando ferramentas de análise de dados procuram-se registos duplicados, ruído, erros de digitação e outras incorrecções detectadas mediante a verificação de inconsistências. Estas operações devem ser sempre feitas de acordo com os objectivos definidos para a sua utilização. Há BD consideradas "limpas", porque contêm toda a informação necessária para a utilização que se lhe quer dar, mas com informação considerada "suja", quando o objectivo passa a ser outro. Se não se definirem os

objectivos e a utilização a dar à informação, pode incorrer-se em operações de limpeza inadequadas.

As operações de limpeza mais comuns, associadas a registos duplicados, valores inválidos, omissos ou isolados, são abordadas de seguida, com mais detalhe.

3.2.2.1. Registos duplicados

A junção de duas ou mais BD provoca, normalmente, o aparecimento de registos duplicados. O tipo de duplicidade e dificuldades encontradas na sua detecção, depende das BD terem, ou não, o mesmo modelo de representação dos dados. O problema fundamental na junção de BD está no facto dos dados, recolhidos de várias fontes, poderem ter identificadores diferentes, erros de escrita, transcrição ou ainda valores alterados com propósitos fraudulentos.

Os registos duplicados podem surgir de duas maneiras. Ou aparecem porque apresentam atributos com diferentes valores, ou aparecem porque apresentam identificações diferentes, mas correspondem à mesma entidade no mundo real [Hernández e Stolfo, 1998]. O primeiro tipo de duplicidade é, normalmente, mais simples de detectar. Por vezes basta proceder à ordenação dos valores, para que se evidenciem de imediato as duplicações nos dados. No entanto, dados simbólicos com diferentes identificações têm normalmente uma detecção mais demorada. Identificar se dois registos diferentes, estão a caracterizar a mesma entidade, pode ser uma tarefa altamente complexa. Nestes casos, a selecção dos registos duplicados não é feita apenas por um atributo, como a designação ou um código, mas por uma duplicidade de atributos que no essencial convergem e correspondem à mesma entidade do mundo real.

Identificar instâncias similares de uma entidade, com representações díspares numa BD, pode fazer-se através de interrogações à BD, seleccionando-se todos os registos que satisfaçam condições de restrição estipuladas, e que representem a mesma entidade [Buckles e Petry, 1982]. Todavia, pode ser feita uma outra abordagem, subdividindo a DB por segmentos para reduzir a complexidade do problema. Com esta operação os potenciais registos duplicados ficam agrupados no mesmo segmento [Hernández e Stolfo, 1998].

3.2.2.2. Valores inválidos

Consideram-se valores inválidos aqueles dados que possuem valores semanticamente ou sintacticamente errados. Os dados semanticamente errados caracterizam-se por lhe estarem atribuídos valores de um tipo diferente do esperado. Os dados sintacticamente errados são aqueles que contêm valores fora do intervalo de parâmetros admissíveis [Rodrigues, 2000].

Estes dados não são facilmente identificáveis, sendo necessário um bom conhecimento no domínio para a sua detecção. A análise de dados e detecção de erros pode fazer-se através da ordenação dos dados ou pela sua representação em gráficos como os histogramas, ou outros.

3.2.2.3. Valores omissos

A existência de dados em falta é um dos problemas presentes na maioria dos projectos de DC, sendo a sua detecção um processo simples. O facto de nas BD aparecerem valores nulos pode ter duas explicações. A razão mais frequente tem a ver com valores que são esperados, mas que estão ausentes. No entanto, pode também acontecer que para algumas entidades, não se apliquem alguns atributos, pelo que são intencionalmente deixados em branco.

Há diferentes formas de lidar com os dados omissos. Para algumas ferramentas os registos com dados omissos são descartados durante o processo de DM, enquanto que, outras ferramentas mais especializadas em determinadas áreas, têm módulos assistidos para limpeza de dados que ajudam a lidar com este problema [Pyle, 1999].

O tratamento de valores omissos pode ser feito manualmente, analisando e procurando os valores caso a caso ou ainda, através da atribuição automática de valores médios ou estimados. A atribuição de valores por inspecção, registo a registo, é uma abordagem que só é possível quando o número de valores em falta é reduzido e onde a sua identificação seja inequívoca.

No caso de haver elevado número de valores em falta, pode recorrer-se a técnicas de atribuição automática de valores, utilizando diferentes metodologias. Pode optar-se pela atribuição de valores fixos, que desde logo se associam a valores

atribuídos artificialmente, ou pela atribuição de valores médios ou modais, ou ainda valores previstos através da utilização de modelos preditivos. Quando se utilizam modelos preditivos, que à partida podem apresentar valores mais aproximados dos reais, deve ter-se em atenção o facto dos modelos preditivos usarem outras variáveis para prever os dados em falta. Neste caso, pode chegar-se à conclusão que se está perante um atributo que é altamente relacionável com outras variáveis e, então, a melhor opção será a sua remoção da BD de trabalho.

Qualquer que seja o método escolhido para atribuir valores aos dados omissos, deve ser cuidadosamente estudado e definido de acordo com o especialista da área. Opções erradas tomadas nesta fase do processo podem introduzir ruído na BD e contribuir para que sejam identificados modelos pouco representativos.

3.2.2.4. Valores isolados

Os valores isolados podem representar dados reais ou valores errados, podendo ser detectados pela análise de gráficos de distribuição de dados. Quando se trata de um valor errado ele deve ser corrigido, ou eliminado. Tratando-se de um dado que representa um valor real, deve ser ponderada a hipótese da sua inclusão ou remoção do conjunto de dados a submeter aos algoritmos de DM. No caso de representarem conceitos e relações que não se afigurem imprescindíveis para os objectivos definidos, deve considerar-se a hipótese de retirar estes registos, para simplificar o modelo ou por não existirem exemplos suficientes para o modelar.

3.2.3. Pré-processamento dos dados

O pré-processamento prepara os dados para serem submetidos aos algoritmos de DM. Esta transformação tem o objectivo de incorporar o máximo de informação possível, reduzindo, ao mínimo, o número de linhas e colunas da amostra. Por vezes, se não se limitar o número de linhas e colunas a tratar, pode inviabilizar-se a utilização de determinados algoritmos de exploração de dados, por se tornarem menos eficientes com um grande número de variáveis.

A redução do número de variáveis e de registos pode ser conseguida pela normalização de dados, pela combinação de variáveis de entrada, pela remoção de atributos ou ainda pelo agrupamento de registos.

3.2.3.1. Normalização de dados

A normalização de dados é uma técnica de pré-processamento muito utilizada na DC. O modo como os dados são codificados/agregados tem uma grande influência nos resultados obtidos [Santos, 2001]. A normalização permite reduzir os dados a uma mesma escala e uniformizar nomenclaturas para o mesmo conceito. Desta forma, reduzem-se o número de variáveis a considerar, para cada atributo, podendo diminuir-se também o número de linhas.

A combinação de variáveis de entrada não correlacionadas é um processo que pode limitar o espaço de pesquisa. A combinação pode ser feita, por exemplo, pela combinação matemática das variáveis de origem.

3.2.3.2. Remoção de atributos

O especialista da área de aplicação do trabalho pode, à partida, avaliar o interesse e relevância dos dados, dando indicações para que sejam removidos aqueles que se julga não serem relevantes para o estudo. No entanto, podem também usar-se algoritmos de indução de árvores de decisão[5], para identificar os atributos relevantes para a identificação do modelo preditivo. Os atributos não contemplados nas árvores de decisão geradas podem ser retirados, uma vez que não são relevantes na tomada de decisão.

3.2.3.3. Agrupamento de registos

O agrupamento de registos pode ser realizado pela generalização ou discretização de atributos [Rodrigues, 2000]. A generalização é feita de acordo com as hierarquias conceptuais que representam e com o conhecimento do domínio. Na discretização os dados são agrupados em classes, definidas para o efeito.

[5] As árvores de decisão são um dos algoritmos de indução de regras utilizados na identificação do modelo ***servator*** e são apresentadas posteriormente neste capítulo.

Esta tarefa, sendo fundamental no processo de DC, é muito interactiva e iterativa. Se, por um lado, exige um trabalho estreito com os especialistas na área, detentores do conhecimento no domínio, por outro, raramente termina na primeira iteração. Esta fase é normalmente repetida várias vezes, mesmo depois de se pré-processarem os dados, a fim de melhorar os resultados.

3.2.4. Aplicação de algoritmos de DM

O DM é uma etapa do processo de DC, na qual se aplicam aos dados, um conjunto de técnicas de análise, para extrair conhecimento.

Os dados, que resultam das fases de selecção, tratamento e pré-processamento, são submetidos a algoritmos de DM para extracção de padrões. A escolha dos algoritmos depende dos objectivos definidos para a DC. Normalmente, utilizam-se várias técnicas que, combinadas entre si, podem produzir melhores resultados [Fayyad *et al.*, 1996a].

Os algoritmos de DM são analisados com mais detalhe na subsecção 3.4.3.

3.2.5. Interpretação e avaliação de resultados

Nesta etapa analisam-se os resultados obtidos, aplicando-se os modelos encontrados a novos conjuntos de dados e avaliando-se o seu desempenho, perante dados que lhe são desconhecidos. É nesta fase que se analisa a informação produzida e se verifica a validade dos resultados gerados pelos algoritmos de DM.

O conhecimento descoberto é avaliado à luz de quatro características principais, associadas à sua validade, novidade, utilidade e interesse para a área de aplicação. Espera-se que o conhecimento identificado seja:

- **Válido**, isto é, tenha um certo grau de certeza. É possível estimar o grau de certeza da previsão realizada aplicando-se o modelo aos dados de Teste;
- **Novo** para o domínio de estudo;
- **Útil**, ou seja, traga algum benefício para o utilizador. A medida de utilidade depende muito da área de estudo e pode ser avaliada, por exemplo, pelo dinheiro ou vidas poupadas por uma boa previsão. Quando se faz uma boa

recomendação de determinado medicamento podem evitar-se novas idas ao médico, nova toma de medicamentos ou até a perda de vidas;

- **Interessante** para o domínio de aplicação. A avaliação do interesse do conhecimento obtido é feita combinando os aspectos anteriores da novidade, utilidade e validade dos padrões encontrados.

3.3. Importância do conhecimento no domínio

O conhecimento no domínio consiste na informação já disponível sobre os dados, e que resultou de outros processos de descoberta ou da incorporação de conhecimento dos especialistas da área [Anand *et al.*, 1995].

Durante todo o processo de descoberta de conhecimento há necessidade de intervenção humana, introduzindo no sistema a sua flexibilidade, criatividade e conhecimento do domínio de estudo. A necessidade de intervenção humana no processo é, ainda, reforçada porque os objectivos e o grau de interesse colocado no conhecimento obtido dependem das expectativas colocadas pelo utilizador.

O conhecimento do domínio pode afectar o processo de DCBD, quer tornando os padrões mais visíveis através da generalização de valores de atributos, quer pela restrição do espaço de pesquisa, tornando mais simples as regras obtidas.

Apresentam-se a seguir as principais tarefas do processo de DC, de acordo com a metodologia CRISP-DM [Chapman *et al.*, 2000]. Em todas as tarefas se realça a extrema importância assumida pela incorporação de conhecimento já existente, para o domínio de aplicação do estudo.

3.3.1. Definição do objectivo do processo de DCBD

A DC pode partir de objectivos bem definidos, elegendo-se, *a priori,* os resultados que procuramos obter com o modelo a identificar. No entanto, pode simplesmente procurar-se conhecimento, sem estabelecer qualquer objectivo a alcançar. Em qualquer uma destas situações é fundamental a compreensão do domínio de aplicação dos dados em que se está a trabalhar.

3.3.2. Escolha do conjunto de dados

Qualquer processo de DC actua sobre um conjunto de dados**.** Não será possível descobrir conhecimento, se os dados não possuírem características relacionadas com o objectivo definido. Assim, é importante reunir uma amostra que seja representativa dos dados, na área que se pretende estudar. Esta tarefa terá que ser feita com base no conhecimento existente, para o domínio de aplicação.

3.3.3. Tratamento e pré-processamento dos dados

As operações sobre os dados devem, também, ser realizadas de acordo com conhecimento específico na área. Decisões sobre a limpeza dos dados, uniformização de nomenclaturas, opções a tomar perante dados omissos ou redundantes, devem ser acompanhadas por alguém que seja detentor de conhecimento no domínio e, mais detalhadamente, sobre os dados.

3.3.4. Interpretação do conhecimento

A interpretação do conhecimento encontrado, a sua avaliação e as decisões sobre as fases a executar novamente, são operações executadas de acordo com o conhecimento já existente. A integração dos padrões encontrados no domínio de aplicação e a, eventual, resolução de conflitos com o conhecimento já instituído, pressupõe um grande conhecimento do domínio sobre o qual se está a desenvolver o processo de DCBD.

3.4. *Data Mining*

A confusão instalada entre DC e DM foi muitas vezes palco de discussão e tema de escrita para vários artigos. A descoberta de padrões úteis nos dados foi já apelidada de DCBD, DM, Extracção de Conhecimento, Descoberta de Informação, Colheita de Informação, *Data Archaeology* ou ainda Processamento de Padrões nos Dados [Fayyad *et al.,* 1966a].

Neste trabalho associa-se a DC a um processo amplo, onde são importantes os dados, a intervenção humana e as técnicas de *DM.*

Os dados são a matéria-prima do processo. No entanto, essa matéria-prima tem que ser preparada para ser possível a identificação de padrões. O DM, que significa

"ajustar modelos a" ou "determinar relações a partir de", é uma fase da DC, onde se aplicam aos dados algoritmos de extracção de padrões.

3.4.1. Tipos de actividades

No processo de DC as técnicas de DM são utilizadas para desenvolver actividades de descrição ou de previsão, com o objectivo de analisar os dados para construir modelos que descrevam importantes classes de dados ou consigam prever valores ainda desconhecidos [Han e Kamber, 2001].

Nas actividades descritivas o processo de DCBD é utilizado para extrair padrões e fazer a sua representação de forma simples e num formato legível para o utilizador, facilitando um melhor conhecimento dos dados. Uma das formas mais utilizadas para a extracção de informação descritiva, a partir dos dados, é a utilização das técnicas de indução de regras. A principal vantagem desta técnica de DM está na forma como apresenta os resultados, numa linguagem natural, que facilita a avaliação do conhecimento e a integração no conhecimento já existente.

Numa actividade de previsão usam-se os dados para construir um modelo que consiga prever determinados atributos de interesse, em futuras instâncias.

3.4.2. Tarefas de DM

O DM, considerado como uma componente do processo global que é a DCBD, envolve a execução de um conjunto de operações para extracção de informação útil, a partir dos dados [Berry e Linoff, 2000].

O tipo de operações a desenvolver, varia de processo para processo, de acordo com os objectivos que estipulados. Quando os objectivos são definidos, em termos do conhecimento que é esperado obter, ou seja, quando é definida a variável de saída, estamos perante a chamada aprendizagem supervisionada. Nesta aprendizagem o processo de DC é orientado para a identificação de um modelo que descreva uma determinada variável de interesse, a partir de um conjunto de dados. Prever a cronologia de um determinado objecto em Arqueologia, a partir de uma série de características predefinidas, é uma operação supervisionada.

Por outro lado, quando não é definida nenhuma variável de saída está-se perante uma tarefa de aprendizagem não supervisionada. Neste tipo de aprendizagem o conhecimento obtido pode ser inesperado, uma vez que não é imposta qualquer restrição, ou orientação inicial, na procura de relações entre as variáveis.

As tarefas de DM, a seguir descritas, podem ser agrupadas em cinco operações distintas: classificar, estimar, segmentar, agrupar por afinidade e descrever e visualizar [Berry e Linoff, 2000].

Estas actividades estão normalmente associadas a tarefas supervisionadas. Apenas a segmentação se apresenta como uma operação não supervisionada.

3.4.2.1. Classificar

As tarefas de classificação são utilizadas quando queremos identificar um modelo que indique a classe a que pertence um determinado objecto.

Partindo de um conjunto de classes pré-definidas, onde todas as entradas estão classificadas, gera-se um modelo que aprende a classificar os objectos. Essa aprendizagem baseia-se nos atributos e na classe a que pertencem os dados classificados.

O modelo após este processo de aprendizagem deverá, de acordo com as regras que encontrou, fazer o mapeamento de novos dados, não classificados, com as classes a que pertencem.

Alguns autores defendem que os modelos preditivos devem aparecer como uma actividade separada de estimar ou classificar, enfatizando a diferença existente na validação de resultados [Berry e Linoff, 2000]. Na actividade de classificação é sempre possível averiguar se a classificação foi, ou não, bem feita, comparando os resultados com os valores reais. Nos modelos preditivos prevê-se comportamentos ou valores futuros, que não é possível confirmar no momento da identificação e validação do modelo. Podem acontecer, por exemplo, alterações conjunturais que tornem o conjunto de dados pouco representativo da população alvo.

As actividades de classificação têm sido utilizadas nas mais variadas áreas, das

quais se destacam as da área Banca ou Marketing e Vendas [Rodrigues *et al.,* 1998].

Os algoritmos de DM mais populares para realizar actividades de classificação são as árvores de decisão e as redes neuronais. São técnicas de DM que serão apresentadas na próxima subsecção.

3.4.2.2. Estimar

A estimativa de valores é uma actividade muito idêntica à classificação, diferindo apenas no tipo de valores a apresentar. Enquanto a classificação lida com valores discretos, as classes, a estimativa tem como resultado valores contínuos.

São exemplos destas actividades as estimativas do número de filhos por casal ou dos rendimentos por família [Berry e Linoff, 2000].

3.4.2.3. Associar

As BD armazenam valores para atributos que, muitas vezes, estão relacionados entre si. As actividades de associação têm o objectivo de determinar relações entre algumas variáveis da BD, para a mesma instância. Estas são definidas através de um conjunto de regras de associação.

Esta actividade tem sido muito usada, por exemplo, na área de vendas, onde o arranjo e localização de produtos são estudados de forma a tirar partido de determinadas correlações existentes entre eles.

Este tipo de análise é muito semelhante à sequenciação, onde também se procuram associações entre dois ou mais atributos, mas desfasados no tempo. Por exemplo, quem compra um computador poderá, mais tarde, vir a comprar uma impressora, um gravador de CDs, tinteiros e CD-ROM.

3.4.2.4. Segmentar

A segmentação, também conhecida por *clustering*, é uma actividade idêntica à classificação, mas onde as classes não são predefinidas. Na segmentação os algoritmos fazem o agrupamento dos dados por afinidades de atributos, criando desta forma um conjunto de classes. Feito este agrupamento procede-se então à classificação dos dados,

mapeando-os com as classes encontradas.

A segmentação é escolhida, em detrimento da classificação, quando se pretendem descobrir afinidades entre os dados. Nesta actividade não há classes predefinidas, nem exemplos classificados, a exemplo do que acontece na classificação, pelo que os dados são agrupados de acordo com as afinidades existentes entre eles.

Pode usar-se a esta actividade para fazer a segmentação do mercado, por tipo de clientes, dividindo, por exemplo, os consumidores por hábitos de compra, por interesses ou culturas. É depois possível estudar o impacto de algumas promoções, em cada um destes grupos [Berry e Linoff, 2000].

3.4.2.5. Descrever e Visualizar

A crescente dimensionalidade das BD torna mais difícil a exploração e análise dos dados. Por vezes, há necessidade de aumentar o conhecimento de uma BD de dimensionalidade elevada, descrevendo-a de forma resumida. Uma descrição da BD poderá, desde logo, dar pistas para encontrar algumas explicações.

O objectivo deste tipo de actividades é analisar as BD, de forma a descrever informações complexas, através de representações visuais, como os gráficos ou diagramas.

A exploração visual dos dados é a forma mais poderosa de descrição dos mesmos. Os sistemas de visualização devem apresentar os dados de uma forma simples, de leitura fácil e intuitiva. Devem assistir os utilizadores, deixando que estes participem na orientação a tomar para exploração dos dados. Devem, também, apresentar estimativas de fiabilidade dos valores e serem suficientemente versáteis, para que possam ser utilizados em diferentes meios e cenários. O acesso seguro e ubíquo deverá ser assegurado, tornando estes sistemas de visualização disponíveis onde e sempre que necessário [Keim *et al.*, 2002].

3.4.3. Técnicas de DM

As técnicas de DM consistem na aplicação de algoritmos aos dados, para detectar padrões. A escolha dos algoritmos a utilizar no processo de DCBD depende,

fundamentalmente, das tarefas a desenvolver, de acordo com os objectivos definidos para o trabalho.

Existem situações em que pelo menos duas técnicas de DM são combinadas, de acordo com as tarefas a realizar e para obter resultados com o máximo grau de confiança. A escolha e forma de combinação destes algoritmos são um processo iterativo, sendo repetido, tantas vezes quantas as necessárias, em função da análise dos resultados obtidos e das reformulações sugeridas.

Embora existam várias técnicas de DM, elas podem ser agrupadas em quatro grandes categorias: **redes neuronais**; **indução de regras**; **algoritmos genéticos** e **aproximação de vizinhanças** [Santos, 2001].

Caracterizam-se, de seguida, as técnicas referentes aos quatros grupos principais nomeados, tendo-se feito uma abordagem mais detalhada para as técnicas de redes neuronais e indução de regras, nomeadamente as árvores de decisão, por serem utilizadas na identificação do modelo ***servator***.

3.4.3.1. Redes Neuronais

As redes neuronais são modelos que simulam o funcionamento do sistema nervoso humano.

As formas mais simples de redes neuronais, do tipo perceptrão, podem ver-se como um modelo de regressão linear ou uma função. A partir dos valores de entrada é encontrado um valor de saída. A Figura 3.5 representa a rede neuronal da função Z = 5X + 3Y onde, para qualquer valor de entrada X e Y, é encontrado um valor de saída Z.

Cada elemento da rede, ou nodo, está ligado a um ou mais elementos do nível seguinte, através de ligações às quais é atribuído um peso.

O valor final ou preditivo é resultado da propagação dos valores de entrada, através dos neurónios, passando por uma função de transformação. No caso da Figura 3.5 o valor de Z é resultado da aplicação dos factores 5 e 3, aos valores de entrada, respectivamente X e Y.

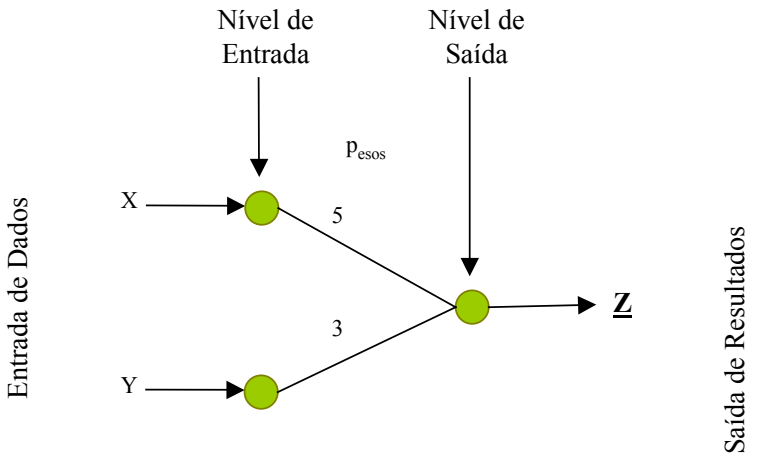


Figura 3.5 – Rede neuronal da função Z = 5X + 3Y (adaptado de [Berry e Linoff, 2000])

Embora o funcionamento das redes neuronais possa ser fácil de entender para as redes do tipo perceptrão, na realidade os modelos não são normalmente assim tão simples, apresentando níveis adicionais entre os níveis de entrada e saída. Esses níveis são conhecidos como níveis ocultos e as redes deste tipo são denominadas de perceptrão multi-nível [Santos, 2001].

Por sua vez os pesos associados aos neurónios que saem dos níveis ocultos são, eles próprios, funções, pelo que a rede neuronal deixa de ser traduzível por uma equação.

Na Figura 3.6 apresenta-se uma configuração de uma rede neuronal, com um nível oculto e com vários nodos, organizados em camadas e ligados entre si. A partir de um conjunto de elementos de entrada como o sexo e a idade, entre outros, a rede vai propagando os valores destes nodos até ao nível de saída, alterando-os através da atribuição de pesos às ligações. Este processo é repetido várias vezes e os pesos atribuídos vão sendo ajustados, em função da aprendizagem obtida em cada iteração.

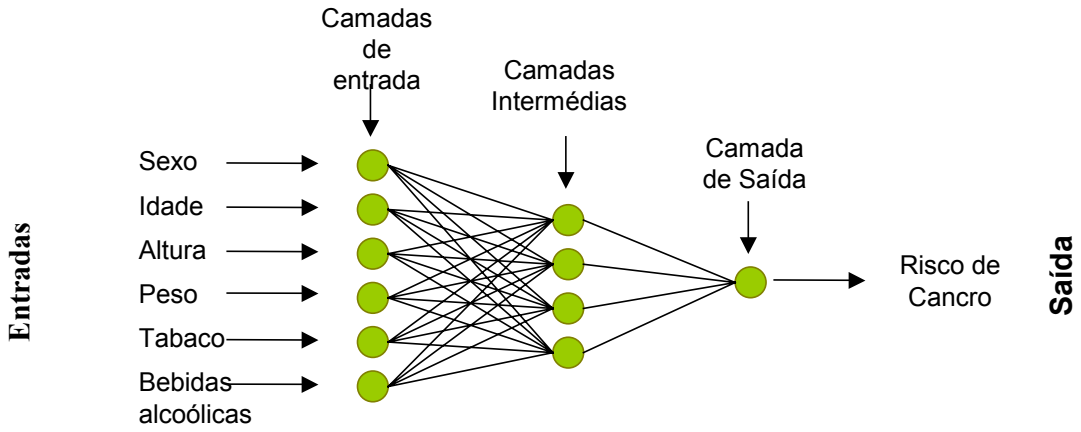


Figura 3.6 – Configuração de uma rede neuronal

As configurações possíveis para as redes neuronais são muitas. Podem variar de acordo com o número de nodos de entrada, o número de nodos de saída, com o número de níveis ocultos e na forma como os valores se propagam. Pode ainda haver várias iterações intermédias até ser apresentado um resultado final.

Grande parte dos algoritmos de redes neuronais usa o sistema de *backpropagation*, para treinar uma rede neuronal. Partindo de um conjunto de dados de treino, onde existem os valores de entrada e também os respectivos resultados, a rede atribui inicialmente pesos iguais a todas as ligações da rede. Os valores iniciais são aplicados e propagados pela rede, originando um resultado final. Comparando esse resultado com o valor que era expectável, é calculado o erro. Com base no erro calculado a rede faz a alteração dos pesos atribuídos e realiza uma nova propagação dos valores de entrada. Este processo irá repetir-se até que as alterações feitas nos pesos não produzam alterações significativas nos resultados finais. Quando tal acontece dá-se por terminado processo de aprendizagem da rede, estando esta preparada para classificar casos desconhecidos.

As redes neuronais podem produzir bons resultados quando se pretende construir um modelo preditivo [Berry e Linoff, 2000]. São genericamente utilizadas em tarefas de classificação e segmentação.

Existe uma outra configuração de redes, conhecidas como redes de *Kohonen***,** que são redes em que os nodos de entrada estão directamente ligados a todos os nodos adjacentes [Kohonen, 1989] (Figura 3.7).

São redes auto-organizáveis e de aprendizagem competitiva, cuja técnica foi inventada por Teuvo Kohonen. Cada nodo da rede tem associada a posição que ocupa numa estrutura bidimensional, que no início do processo é aleatória. À medida que avança a aprendizagem da rede, os nodos vão competindo entre si para ganhar a classificação de um dado registo. Os nodos semelhantes ao nodo classificado são agrupados a ele, em vectores, formando classes ou segmentos [Lobo e Moura-Pires, 1998]. Os pesos obtidos para as ligações permitem identificar o seu peso na definição das classes.

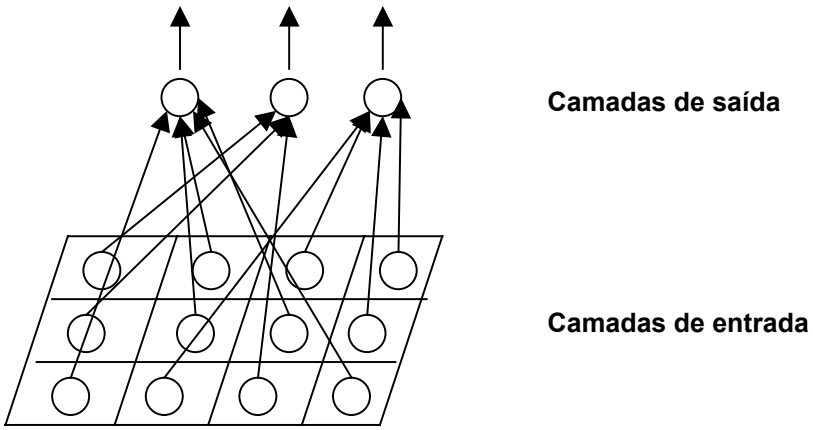


Figura 3.7 – Rede neuronal do tipo auto-organizável

A rede de *Kohonen* possui um número de nodos, calculado em função dos atributos de entrada, sendo o número de nodos de saída dependente do número de classes obtidas na aprendizagem. São algoritmos muito úteis em tarefas de segmentação, onde não existem classes pré-definidas [Santos, 2001].

As críticas apontadas aos algoritmos de redes neuronais relacionam-se com a dificuldade de utilização e de compreensão. Para serem utilizados, com sucesso, estes algoritmos exigem uma exaustiva preparação dos dados. Sem bons dados não haverá bons modelos. Por outro lado, os modelos identificados são difíceis de compreender. A

falta de transparência do processo de decisão dentro da rede, as dificuldades sentidas para interpretação do significado dos valores simbólicos associados aos pesos e, o facto de não produzirem regras, tornam estes modelos não lineares difíceis de entender. São técnicas utilizadas quando os resultados são mais importantes do que o entendimento sobre o funcionamento do modelo e dos critérios que fundamentam as decisões.

3.4.3.2. Indução de regras

Os algoritmos de indução de regras são bastante explícitos relativamente à detecção de tendências dentro dos dados. A indução de regras permite gerar árvores de decisão ou regras de associação.

As árvores de decisão apresentam-se como estruturas onde são encadeadas uma série de regras que apontam para uma classe ou valor. Em cada nodo da árvore define-se uma condição lógica sobre um atributo de uma instância.

A árvore de decisão representada na Figura 3.8 ilustra o conjunto de regras que apoiam a tomada de decisão de ir ou não para trabalho de campo, em função do estado de tempo. Os nodos da árvore contêm uma condição sobre um atributo, neste caso, sobre o estado do tempo. Cada ramo que deriva de um nodo corresponde a um valor possível do atributo considerado nesse nodo. As folhas da árvore representam as classes, isto é, as decisões possíveis.

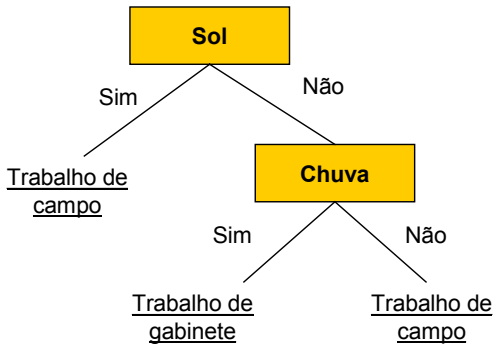


Figura 3.8 – Exemplo de árvore de decisão

As árvores de decisão expressam de uma forma simples uma lógica condicional. O seu funcionamento baseia-se na divisão da tabela de dados, em várias tabelas mais

pequenas, através da selecção de subconjuntos, com base nos valores de um dado atributo.

A figura 3.9 evidencia um conjunto de regras extraídas da árvore de decisão apresentada anteriormente.

Se Sol = "Sim"
Então Trabalho de campo
Se Sol = "Não" e Chuva = "Sim"
Então Trabalho de gabinete

Figura 3.9 - Exemplo de regras induzidas por uma árvore de decisão

As regras de associação identificam relacionamentos entre os dados, representando-os numa linguagem compreensível pelo utilizador.

Embora esta técnica de DM seja muito utilizada em processos de DCBD, por representar o conhecimento encontrado numa linguagem natural, simples e facilmente perceptível pelo utilizador, não deixa, no entanto, de suscitar alguns problemas. Quando lidam com grandes quantidades de informação, estes algoritmos podem gerar um grande número de regras, cujo controlo e monitorização é difícil e dispendioso. Por vezes, pode também acontecer que são geradas regras incompletas, que não cobrem todas as instâncias ou que não são exclusivas [Brijs *et al.*, 2000].

Este tipo de algoritmos é normalmente utilizado em tarefas de classificação, segmentação, associação, sequenciação e sumariação. Apesar dos problemas identificados, que podem aparecer neste tipo de algoritmos, eles são habitualmente utilizados por serem bastante explícitos relativamente à detecção de tendências dentro nos dados. São ainda muito utilizados quando se pretende seleccionar os atributos mais importantes, para definir as entradas de uma rede neuronal.

3.4.3.3. Algoritmos genéticos

Os algoritmos genéticos foram desenvolvidos com base nos princípios da selecção natural e genética, apresentando semelhanças com o processo evolutivo das espécies.

As informações referentes a um determinado sistema são codificadas de maneira análoga aos cromossomas biológicos. Sobre os dados iniciais são aplicados operadores de **selecção**, **cruzamento** e de **mutação**.

Partindo de uma população inicial composta por um conjunto aleatório de regras, de comprimento fixo e representadas por uma sequência de bits (Figura 3.10), os algoritmos genéticos desencadeiam um processo de **selecção**. Neste processo os elementos mais aptos são escolhidos, de acordo com uma função de avaliação, definida para o domínio. Esta função tenta associar uma maior probabilidade aos melhores, para que sejam seleccionados, garantindo assim a sua reprodução para formarem a nova geração. Os elementos seleccionados serão os pais das regras geradas a seguir.

Na operação de **cruzamento** são aleatoriamente escolhidos pares de regras, dentro do conjunto seleccionado para reprodução. Estes pares escolhem, também aleatoriamente, um ponto de cruzamento, para fazer uma troca de sub-*strings* entre eles, simulando assim combinações que podem acontecer durante a reprodução.

Na **mutação** são introduzidas alterações esporádicas a um dado gene, podendo um caracter do alfabeto ser trocado por outro. Este processo é muito similar às mutações genéticas que ocorrem no DNA.

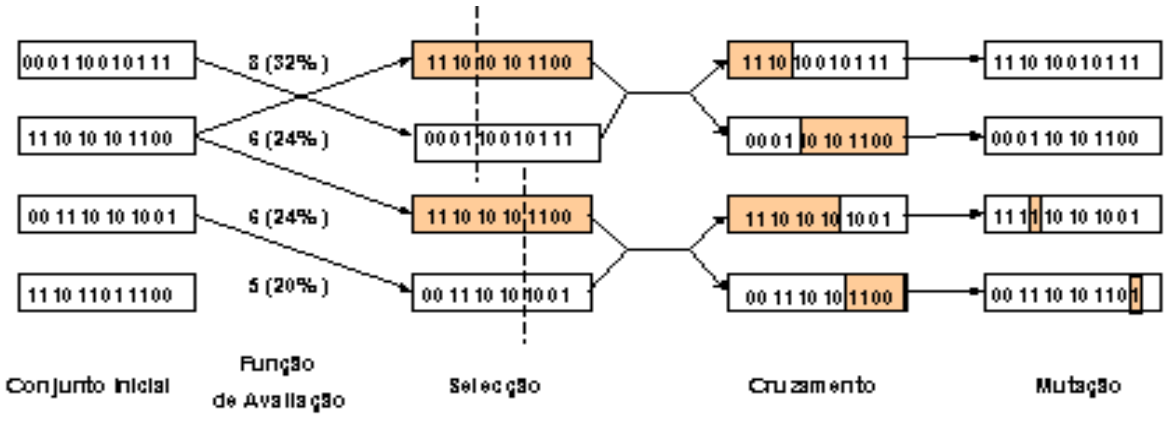


Figura 3.10 – Modo de operação dos Algoritmos Genéticos (adaptado de [Santos, 2001])

Os algoritmos genéticos têm sido utilizados em tarefas de classificação e sumariação. Estes algoritmos agrupam os dados em vectores, com mesma dimensão,

pelo que os dados de diferentes origens têm que ser tratados, tendo em conta este imposição.

3.4.3.4. Aproximação de vizinhanças

Os algoritmos de aproximação de vizinhanças são baseados no princípio de que registos semelhantes estão próximos, quando analisados numa perspectiva espacial [Santos, 2001]. Cada região define uma classe e é identificada pela proximidade de registos. Cada um é interpretado como um ponto no espaço e registos de características comuns estarão mais próximos uns dos outros.

O processo de treino dos algoritmos de aproximação de vizinhanças envolve o armazenamento e treino dos dados. Para cada saída alvo é realizado um trabalho de pesquisa em todos os dados, procurando-se aqueles que lhe são similares. A saída do alvo é então associada com o vizinho mais próximo.

A implementação desta técnica implica a distribuição dos objectos por classes. O número de classes inicial constitui um parâmetro de entrada. Cada classe pode ser representada pelo seu centro de gravidade e cada registo é transformado num ponto do espaço que contém tantas dimensões quantos os atributos em análise [Santos, 2001] (Figura 3.11).

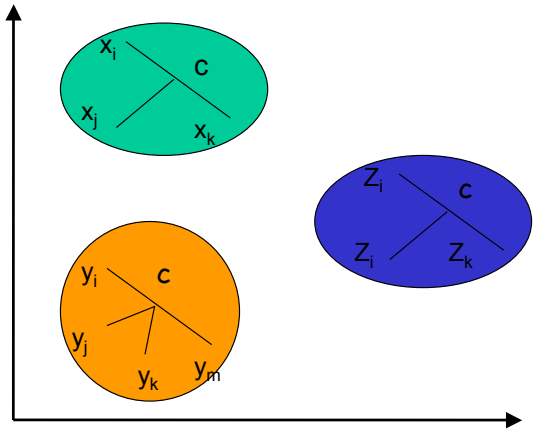


Figura 3.11 – Partição dos objectos em classes (adaptado de [Santos, 2001])

O processo de obtenção de classes começa com os centroídes distribuídos aleatoriamente. Os reposicionamentos são iterativos e feitos de acordo com as

afinidades encontradas entre as classes. A similaridade é depois medida em função da distância existente entre os objectos.

São algoritmos utilizados em actividades de segmentação ou de sumariação.

3.5. Conclusão

Neste capítulo abordaram-se os conceitos associados à DCBD, nomeadamente as tarefas e técnicas relacionados com o DM. Estes conceitos serão utilizados na definição dos módulos de análise de dados e identificação de padrões do sistema ***servatis,*** apresentado no capítulo 4 e no decorrer da identificação do modelo ***servator***, apresentado nos capítulos 5 e 6.

Capítulo 4

Sistema *servatis*

Neste capítulo será apresentada a arquitectura do sistema ***servatis,*** um sistema integrado de gestão de informação arqueológica. A sua implementação (uma parte é apresentada neste trabalho) e desenvolvimento constituem um trabalho a realizar no futuro, embora muitas premissas relativas à implementação sejam, desde já, indicadas neste trabalho.

Durante o processo arqueológico produz-se um grande manancial de informação, difícil de gerir sem recurso às TI [Badia, 1992]. Não obstante a sua utilização ser já bastante alargada, os dados continuam muito espartilhados. A divulgação dos resultados das actividades de investigação é muito restrita, muitas vezes limitada a apresentações em encontros científicos ou publicações em revistas da especialidade. O ***servatis*** foi desenhado para que a partilha de informação seja possível, de acordo com critérios e perfis de utilização a definir posteriormente. A circulação de informação entre a comunidade científica e também com outras comunidades deve ser facilitada.

A sua arquitectura foi desenhada para que o ***servatis*** possa ser construído com módulos desenvolvidos especificamente para o efeito, tal como o ***servator.*** No entanto, está prevista também a integração de módulos já desenvolvidos, tal como SIG e Modelos de Reconstituição Virtuais.

4.1. Enquadramento

A Arqueologia lida com muitos dados. As várias fases do processo arqueológico, prospecção, escavação e interpretação, produzem um numeroso e variado leque de informação que é necessário armazenar, tratar e divulgar.

As TI têm sido um valioso contributo para gerir este crescente conjunto de informação. São muitos os exemplos de aplicações informáticas que, durante todo o processo de escavação, apoiam o registo, tratamento e armazenamento dos dados. Despontam também os sistemas de apoio à decisão, nomeadamente através da utilização de SIG [Leusen, 2002], do levantamento automatizado e representação virtual dos sítios arqueológicos [Allen *et al.*, 2004] ou de análise dos dados, recorrendo à reconstituição virtual [Vote, 2001]. Não obstante este grande investimento em TI, as aplicações desenvolvidas estão muito dispersas, constituindo módulos estanques e de acesso muito restrito.

O sistema ***servatis*** tem como finalidade disponibilizar, de uma forma integrada, vários módulos de armazenamento, gestão e visualização de informação, apoiando todas as fases do processo arqueológico, particularmente a gestão e a divulgação da informação. Ao definir a finalidade do ***servatis*****,** valorizou-se a componente de divulgação de informação e de integração dos vários módulos.

A informação arqueológica é necessária, não só para as diversas tarefas e actividades no âmbito da Arqueologia, mas também nas actividades de várias Organizações que com ela se relacionam. Por exemplo, a informação de sítios arqueológicos serve de base à investigação arqueológica, assim como à elaboração de catálogos Patrimoniais e de PDM, à criação de Modelos Virtuais de sítios, à elaboração de Roteiros Culturais ou ainda de modelos preditivos de património arqueológico.

O acesso à informação arqueológica para investigadores da área, ou para organismos que trabalham com dados de Património arqueológico, ou ainda, para o público em geral, teria no ***servatis*** uma interface amigável e acessível. A elaboração de PDM, de planos de Ordenamento e Gestão Territorial ou de bolsas de informação para a emergente Indústria do Património, são actividades que precisam da informação arqueológica e, para as quais, o ***servatis*** pode constituir um válido interlocutor.

Potenciar a identificação de padrões nos dados, usando os princípios associados à DCBD, foi também uma finalidade definida. Contendo o ***servatis*** um repositório de dados arqueológicos, é desde logo, um elemento potenciador da descoberta de conhecimento, no âmbito da Arqueologia.

Procurou-se, portanto, que o ***servatis*** fosse um valioso contributo para uma ampla visibilidade da informação arqueológica existente, para as equipas de investigação, e também para outras comunidades científicas, Organizações e público em geral. O tipo de informação, o nível de agregação da mesma, bem como as tarefas a que têm acesso, dependerá dos perfis de utilizador criados.

Assume-se como pressuposto que a arquitectura do sistema terá que ser bastante flexível, para alojar novos módulos especificamente desenvolvidos para o ***servatis*** como, por exemplo, o ***servator,*** e outros que tenham já sido desenvolvidos e estejam a funcionar autonomamente.

Para atingir a finalidade estabelecida para o ***servatis,*** desenhou-se uma arquitectura e definiram-se algumas vertentes da sua implementação, apresentadas na próxima secção com mais detalhe.

4.2. Arquitectura do *servatis*

O sistema ***servatis,*** cujas finalidades e objectivos foram já enunciados, foi definido com base em duas vertentes principais: os dados e o processamento que lhes é dado.

A escavação é sempre um processo destrutivo [Green, 2002], o que torna o repositório de dados e modelos ainda mais importante na Arqueologia, bem como a visualização da informação processada com base nesses dados e modelos. O registo do processo de escavação permite que a memória dos sítios se prolongue no tempo, facilitando também a interpretação arqueológica e a reconstituição virtual dos arqueossítios e dos contextos.

Com base no conjunto de dados e modelos armazenados no repositório, o ***servatis*** disponibiliza a informação arqueológica, processada de diversas formas, nomeadamente em catálogos de sítios ou materiais, relatórios ou gráficos. A integração de dados e ferramentas proporciona, ainda, que as BD sejam usadas nos SIG, na restituição virtual de sítios e/ou materiais, bem como na identificação de modelos com base nos princípios associados à DCBD. Para cada uma destas actividades, o utilizador pode sempre receber indicações/sugestões sobre as operações a realizar a seguir – *wizards.* Estes guiões orientadores deverão estar disponíveis em cada fase do processo,

contendo ainda textos (“saber mais”) que, sempre que o utilizador pretenda, expliquem conceitos e opções possíveis em cada iteração.

O sistema ***servatis*** integra uma arquitectura cliente-servidor, de acordo com a estrutura apresentada na Figura 4.1.

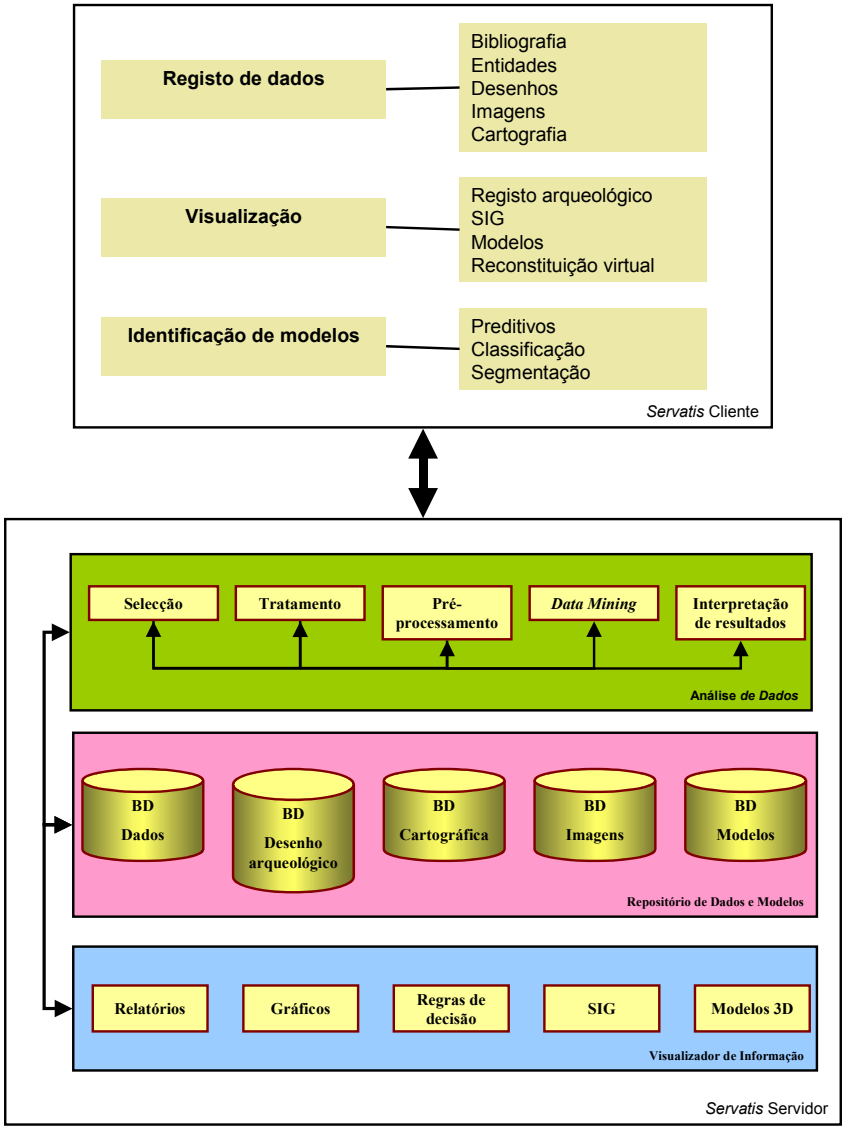


Figura 4.1 – Arquitectura do sistema ***Servatis***

A interface com o utilizador, estabelecida pelo ***servatis*** **cliente**, permite que sejam realizadas sobre o sistema três operações distintas: Registo de dados, Visualização e Identificação de modelos.

O sítio arqueológico pode ser visto como um conjunto de elementos e estruturas que constituem relíquias do passado, mas também como um conjunto de vestígios integrados num contexto geográfico, cultural, económico e social. Assim, os dados que resultam do processo arqueológico são muito diversificados, como se pode constatar nos exemplos que a seguir se apresentam:

- Bibliografia – como foi já referido no capítulo 2, muitas vezes o processo arqueológico tem início na verificação de indicações registadas em livros, revistas ou outros documentos;
- Entidades – as entidades arqueológicas são muito diversificadas. Podem ser sítios arqueológicos, estruturas, unidades estratigráficas[6] ou materiais encontrados nos arqueossítios. Como os sítios foram locais de actividade humana, as características dos seres, os seus hábitos alimentares, sociais e culturais são também entidades cuja existência é prevista no Repositório de Dados e Modelos do ***servatis***;
- Desenhos – durante o processo arqueológico, mais ainda na fase de escavação de um sítio arqueológico, desenham-se planos, alçados ou perfis, de forma a registar todas as fases do processo. O seu armazenamento em formato vectorial e a sua análise posterior levará à interpretação arqueológica, isto é, à formulação de hipóteses cognitivas sobre a formação do sítio, características, ocupação e abandono;
- Imagens – a fotografia é um processo de memorizar as características encontradas para o sítio e a evolução das escavações. Para além do interesse documental destas imagens e da visualização de várias fases de escavação, há projectos onde, a partir da fotografia digital, se faz a modelação dos sítios e a sua implantação no contexto geográfico;
- Cartografia – a cartografia assume um papel muito importante na Arqueologia, uma vez que o conhecimento do contexto geográfico onde os sítios se inserem é fundamental.

[6] Os sítios arqueológicos são muitas vezes cobertos por sucessivas camadas de terra. A cada camada pode dar-se o nome de unidade estratigráfica.

O módulo de visualização tem como finalidade fornecer informação arqueológica processada de diversas maneiras, de acordo com os objectivos do utilizador. A forma de processar a informação é muito variada, desde a elaboração de listas organizadas dos dados, catálogos de sítios e imagens, até à integração de vários dados e ferramentas como SIG, modelos 3D de representação de sítios ou objectos, ou ainda de modelos preditivos.

Apresentam-se a seguir alguns exemplos de visualização de informação a disponibilizar no ***servatis***:

- Registo arqueológico – a visualização da informação resultante do registo arqueológico pode assumir formas muito variadas, onde se destacam, os inventários de sítios arqueológicos, as listagens das estruturas associadas a determinado sítio, a sua estratigrafia, os achados encontrados. Uma vez que se definiu o ***servatis*** com um sistema integrado, utilizando BD relacionais, é sempre possível criar listagens que combinam dados oriundos de diferentes BD e diferentes formatos, por exemplo, texto com imagens digitais das entidades ou de modelos. O elevado número de listagens, que normalmente os utilizadores requerem e que podem ser definidas a partir dos dados, foi determinante para que se definisse, como requisito do sistema, um módulo assistido para o utilizador personalizar as suas listagens;

- SIG – são muitos os exemplos de SIG aplicados à Arqueologia [Sánchez, 2000]. O ***servatis*** irá integrar alguns módulos já desenvolvidos, nomeadamente o SIABRA (Sistema de Informação Arqueológica de *Bracara Augusta* [Giestal, 1998]);

- Modelos – neste módulo, o ***servatis*** irá integrar o modelo preditivo de apoio à prospecção arqueológica, o ***servator***, desenvolvido nesta dissertação e identificado para a região de Trás-os-Montes Oriental. No entanto, este modelo poderá ser aplicado a novos dados desta região, ou ainda ser adoptada a metodologia utilizada, para identificar novos modelos, a nível regional ou nacional. Para isso, o utilizador da área de

Arqueologia, pode usar o módulo identificação de modelos, também a disponibilizar no ***servatis*** e mais adiante apresentado nesta secção;

- Reconstituição Virtual – O *laser scanner*, por exemplo, pode ser utilizado para criar modelos 3D de representação de sítios [Allen *et al*., 2004]. A reconstituição virtual, pode-se obter com base nos modelos 3D ou a partir dos dados de BD, desenhos arqueológicos e cartografia [Bernardes, 2002]. Estas formas de visualização da informação não espelham uma realidade, mas constituem hipóteses de interpretação do passado, que podem contribuir para gerar novo conhecimento arqueológico e para melhorar a compreensão, valorização e conservação do nosso património cultural.

4.2.1. *Servatis* servidor

Na secção anterior apresentou-se a arquitectura do sistema ***servatis*** e as operações disponibilizadas aos utilizadores. Cada acção desencadeia uma série de operações e interacções entre o ***servatis* cliente** e o ***servatis* servidor**. Este último foi estruturado em três módulos básicos, a Análise de Dados, o Repositório de Dados e Modelos e o Visualizador de Informação, cujas finalidades se descrevem a seguir.

4.2.1.1. Análise de dados

Este módulo é utilizado quando se pretende identificar novos padrões, utilizando os princípios de DCBD. A componente Análise de Dados foi desenhada de acordo com o método CRISP-DM, onde se estabelece uma metodologia para o processo da descoberta de conhecimento, contemplando as fases de selecção de dados, tratamento dos dados, pré-processamento dos dados*, DM* e interpretação de resultados**,** já apresentadas no capítulo 1.

A incorporação do conhecimento existente na área de Arqueologia é fundamental para identificar bons modelos [Pyle, 1999]. Para isso, durante todo processo, o utilizador interage com o sistema através de funções implementadas para o efeito. A utilização de *wizards* é, mais uma vez indicada, permitindo que os especialistas de Arqueologia e de TI possam executar todas as tarefas associadas à

metodologia adoptada, para identificar modelos.

4.2.1.2. Repositório de dados e modelos

O Repositório de Dados e Modelos é a componente responsável por armazenar dados e modelos identificados, utilizando BD relacionais.

Os Dados caracterizam os sítios arqueológicos, através das estruturas encontradas, o espólio, a estratigrafia, as cronologias associadas e a informação resultante da interpretação arqueológica. Como os arqueossítios estão intimamente ligados à área geográfica onde se inserem, os dados relativos à geomorfologia do terreno, rede hidrográfica, tipo de solos e cobertura vegetal, entre outros, são também Entidades consideradas no sistema.

O Desenho arqueológico é uma ferramenta muito utilizada em arqueologia e o seu armazenamento em formato vectorial será feito na BD de Desenho arqueológico. São desenhos utilizados para fazer relatórios de escavação, também usadas nos SIG e nos modelos 3D dos sítios.

A Cartografia representa o cenário geográfico onde se inserem os sítios. O seu armazenamento na BD Cartográfica permite que, por exemplo, seja utilizada nos SIG de Arqueologia.

As imagens, nomeadamente as fotografias, constituem, como já foi referido atrás, uma importante fonte documental em Arqueologia. As possibilidades oferecidas pelo seu armazenamento numa BD relacional são de grande interesse, particularmente na elaboração de relatórios e memórias científicas.

Os modelos preditivos, de classificação ou segmentação, irão também ser armazenados na BD de Modelos, de forma a poderem ser aplicados aos Dados. Novos modelos entretanto identificados irão ser armazenados, de acordo com procedimentos a definir durante a fase de implementação.

4.2.1.3. Visualizador de Informação

A área da Arqueologia trabalha com dados multi-dimensionais bastante complexos. Esta característica levanta, muitas vezes, barreiras aos especialistas da área, dadas as dificuldades encontradas para organizar e visualizar estes dados, de uma forma normalizada, simples e intuitiva.

O sistema ***servatis*** pretende apoiar os arqueólogos na organização e gestão da informação, recorrendo a BD e a várias tecnologias que manipulam informação, tais como os SIG e os modelos os 3D. A informação passa a estar integrada, facilitando a investigação arqueológica e sua utilização em relatórios de escavação, gráficos de manipulação de dados e na elaboração de memórias científicas.

As Regras de Decisão, por exemplo, são uma forma de visualização que se revelam de grande interesse do ponto de vista da prospecção arqueológica e do ponto de vista didáctico. São apresentadas numa linguagem natural, normalizando e estruturando o conhecimento arqueológico duma forma simples e intuitiva.

A utilização de ferramentas como os SIG, ligadas a Bases de Dados e a Bases de Dados de Modelos, permite representar de uma forma gráfica o Património visível e potencial, para uma determinada área geográfica. A mais valia do uso integrado destas ferramentas reside na visão global e integrada que fornece do Património, delimitando áreas de Património inventariado e ainda zonas onde poderão existir sítios arqueológicos não detectados. Estas áreas identificadas podem ser associadas a áreas de risco, com interesse na elaboração de PDM. Utilizando estes modelos, o Ordenamento e Gestão Patrimonial passaria a ter em consideração não só o Património visível, mas também o Património que é necessário proteger, mas que ainda está oculto.

4.2.2. *Servatis* cliente

O ***servatis*** **cliente** constitui a interface com o utilizador. As operações possíveis no sistema estão, como foi já referido, associadas às tarefas de registo de dados, visualização de informação e ainda de identificação de modelos.

Este módulo proporciona acesso a informação arqueológica, com interesse para utilizadores da área de Arqueologia e de outras áreas. Estabeleceu-se como pré-requisito

essencial que todas as operações devem ter disponíveis textos *on-line*, a funcionar como guias de apoio à execução das operações em curso, ou para apresentar sugestões de tarefas a executar a seguir. No entanto, a opção de identificação de modelos, que utiliza os princípios associados à DCBD, aplicando algoritmos de DM aos dados, requer, especificamente, utilizadores com competências na área das TI e Arqueologia. Todo este processo, embora orientado pelo sistema, mantém uma interacção constante com os utilizadores. As decisões tomadas vão influenciar o curso do processo, bem como os resultados finais, pelo que é fundamental que o interlocutor seja um especialista em Arqueologia.

Foram já enunciadas as finalidades de cada um destes módulos, pelo que se irá caracterizar a seguir os vários componentes do ***servatis cliente*** e do ***servatis*** **servidor**.

4.2.2.1. Registo de Dados

Os Dados constituem os alicerces do sistema ***servatis***. No processo arqueológico são geradas grandes quantidades de dados, em diversos formatos e podendo ser estruturados do seguinte modo:

- Dados – tabelas de caracterização de entidades do processo arqueológico, nomeadamente áreas de escavação, estruturas, materiais, estratigrafia, paisagem, solos, entre outras;
- Desenho arqueológico – o desenho arqueológico, resulta como foi já referido atrás, do registo de escavação. Esses registos serão armazenados numa BD, associados aos contextos em que se inserem;
- Cartografia – todo o sítio arqueológico tem uma existência associada a um local. Armazenar a cartografia numa BD, de modo a poder relacionar, por exemplo através de um SIG, os sítios arqueológicos com o seu posicionamento geográfico e todo o contexto ambiental associado, é uma tarefa muito importante e útil à Arqueologia;
- Modelos – os modelos estão sempre associados a informação das BD (Dados) e, normalmente, inseridos num contexto geográfico. Tal como

na Cartografia, é fundamental que os modelos sejam armazenados numa BD relacional, de modo a poderem ser integrados com tabelas de Dados e Cartografia. As possibilidades de utilização que oferecem são imensas. Pode referir-se a título de exemplo, a vantagem de integrar o ***servator*** com um SIG, de modo a que, para determinado tipo de arqueossítio a prospectar, possam ser assinaladas na cartografia da região as áreas de provável localização de Património.

O armazenamento dos Dados e Modelos no sistema será feito, de acordo com o esquema apresentado na Figura 4.2., em BD relacionais.

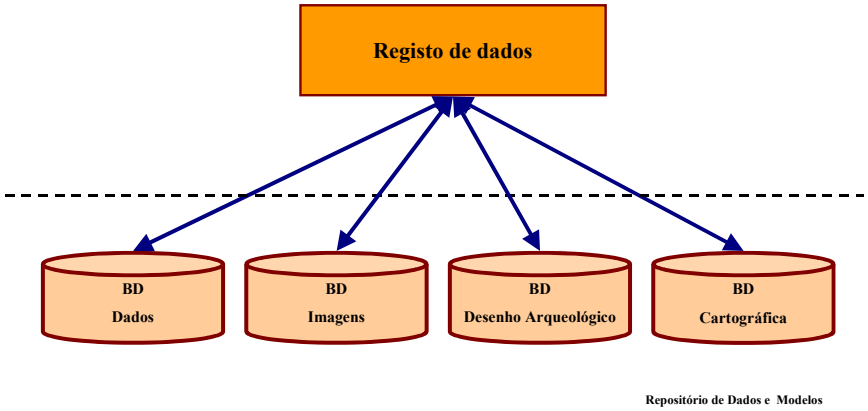


Figura 4.2 – ***Servatis*** **–** módulo de Registo de dados

A estrutura das BD e o modo como irão ser feitas as operações de carregamentos dos dados, serão definidas em futuros trabalhos, por uma equipa pluridisciplinar de especialistas da área das TI e da Arqueologia.

4.2.2.2. Visualização

A opção de Visualização permite ter acesso a um conjunto de ferramentas de elaboração de relatórios, gráficos e mapas, com base nos dados armazenados no Repositório de Dados e Modelos.

A Figura 4.3 define o relacionamento estabelecido entre os componentes do ***servatis***, durante uma chamada ao módulo de Visualização do ***servatis*** **cliente**.

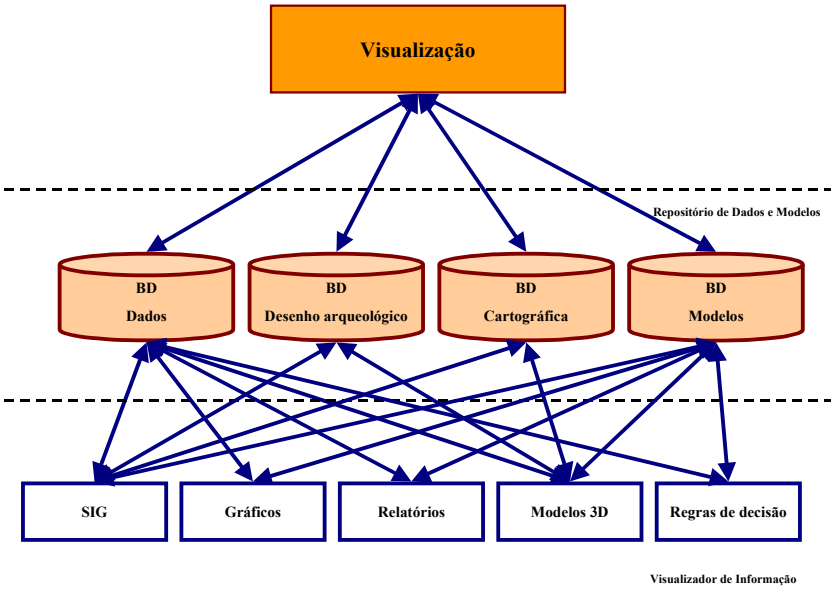


Figura 4.3 – ***Servatis*** **–** módulo de Visualização

Para cada pedido de Visualização do ***servatis*** **cliente** o sistema faz chamadas a módulos do ***servatis*** **servidor,** de acordo com a informação pretendida. Esta pode estar alojada numa, ou mais, BD do Repositório de Dados e Modelos e ser processada pelas ferramentas disponibilizadas no módulo Visualizador de Informação do ***servatis*** **servidor.**

Com este módulo podem ser elaboradas as memórias científicas dos projectos arqueológicos e também podem ser realizados relatórios e mapas cartográficos de arqueossítios, para apoio à decisão nas áreas de Planeamento de Ordenamento do Território.

A criação de roteiros culturais por parte da Indústria do Património pode também ser muito apoiada pelo ***servatis***, como fornecedor de indicadores detalhados e integrados do património de determinada região.

Os registos arqueológicos poderão servir de base à construção de modelos 3D, para visualização das estruturas identificadas durante as escavações [Clark *et al.*, 2002] e podem também servir para ante-visualizar o que não pode ser observado [Barceló, 2000], porque o património está ainda oculto. A Realidade Virtual aplicada à Arqueologia permite, também, que os dados sejam interpretados e que se façam simulações do que se pensa terem sido os sítios arqueológicos [Bernardes, 2002].

O sistema ***servatis*** permite ainda visualizar modelos, utilizados para extrair informação implícita nos dados, previamente desconhecida e potencialmente útil [Han e Kamber, 2001].

A opção de Visualização de Modelos implica o acesso às BD de Modelos, que armazenam os modelos preditivos, modelos de classificação ou segmentação.

A informação ou padrões visualizados podem ser apresentados numa linguagem natural, como por exemplo as regras de decisão, ou sob a forma de gráficos, que evidenciam relações existentes entre os dados. As BD de arqueologia podem estar associadas a indicadores geográficos, georeferenciando os sítios arqueológicos, as suas estruturas e achados. Desta forma, os modelos podem ser integrados com ferramentas SIG e/ou ligados a modelos 3D.

4.2.2.3. Identificação de modelos

O módulo de Identificação de Modelos utiliza a DCBD para identificar modelos, extraindo informação implícita dos dados. Esses modelos poderão ser de vários tipos, desde a identificação de modelos preditivos de património arqueológico, onde são identificadas áreas de provável existência de arqueossítios, ainda não detectados, até aos modelos de classificação ou modelos de atribuição de cronologias. A utilização dos padrões identificados serve de apoio à actividade arqueológica, mas pode também ser útil a outras actividades, nomeadamente à elaboração de PDM ou de Roteiros Culturais.

Tal como acontece no módulo de Visualização, também neste módulo os padrões identificados podem ser apresentados sob a forma de linguagem natural, como as regras de decisão, ou ainda, sob a forma gráfica, integrados em SIG ou modelos 3D.

Havendo indicadores geográficos associados aos dados dos arqueossítios e, as correspondentes entidades cartográficas na BD Cartográfica, pode fazer-se uma incorporação dos relacionamentos espaciais existentes entre as entidades arqueológicas geograficamente endereçadas no processo de DC [Santos, 2001], tornando os modelos mais úteis e permitindo uma apresentação gráfica mais completa.

Na Figura 4.4 apresenta-se um gráfico onde se esquematiza o processo de identificação de modelos, utilizando o sistema ***servatis.***

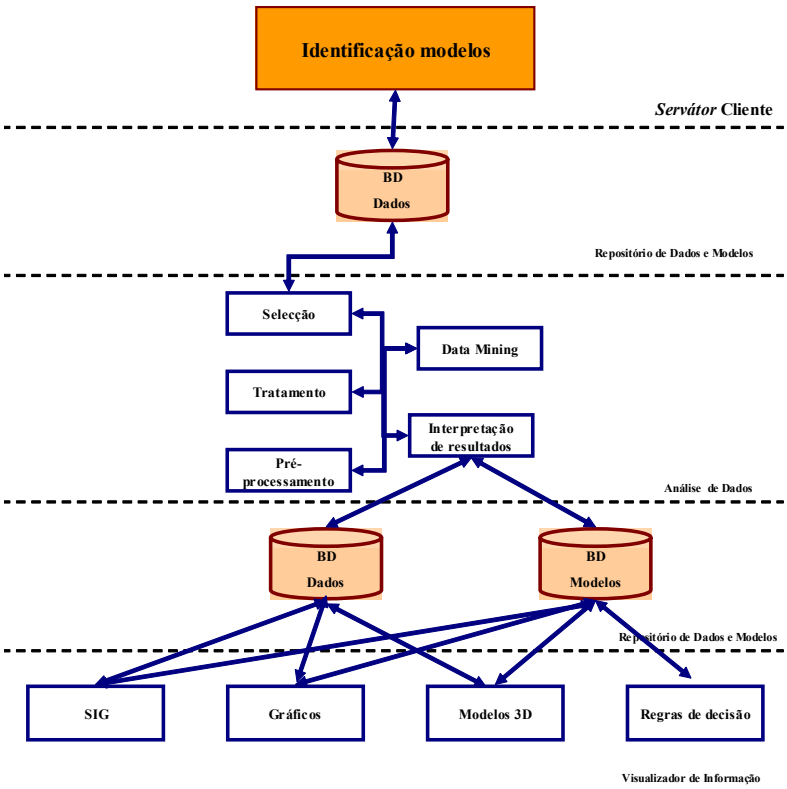


Figura 4.4 – ***Servatis*** – identificação de modelos

Como já foi referido, este módulo deve ser assessorado por utilizadores com competências na área da Arqueologia, uma vez que durante este processo o utilizador intervém em várias etapas, tomando decisões sobre opções a seguir. A qualidade dos padrões encontrados, bem como o seu grau de inovação ou utilidade para a área, é avaliada, também, pelo especialista.

Os padrões identificados podem ser, posteriormente, guardados na BD Modelos do Repositório de Dados e Modelos. A sua apresentação ao utilizador é feita utilizando as ferramentas disponibilizadas no módulo Visualizador de Informação do ***servatis* servidor**.

4.3. Implementação do *servatis*

O ***servatis*** é um sistema que se espera venha a crescer, com a sua utilização, através do registo de dados e da identificação de novos modelos. Para conseguir esse estado de maturidade é necessário que o seu desenvolvimento seja modular, parametrizado e constantemente realimentado.

Um conjunto de módulos a disponibilizar no ***servatis*** estão já implementados e podem ser integrados no sistema, nomeadamente a criação de um Sistema de Informação de Gestão do Património Arqueológico [Botica *et al*., 2003c], o Sistema de Informação Arqueológica de *Bracara Augusta* [Giestal, 1998], os modelos 3D para simulação e divulgação de Património Arqueológico [Bernardes e Martins, 2003] e o ***servator*** - modelo preditivo de apoio à prospecção arqueológica de Trás-os-Montes Oriental, cuja identificação é apresentada nos próximos capítulos.

O ***servator*** foi identificado através da utilização de algoritmos de indução de árvores de decisão, permitindo a sistematização de conhecimento arqueológico num conjunto de regras que integram uma árvore de decisão. Um pequeno conjunto dessas regras é apresentado na Figura 4.5.

```
Cronologia Período Romano
    topografia Cume -> Povoado
    topografia Talvegue -> Rede Viária
    topografia Vertente -> Povoado
    topografia 0
        geomorfologia_mac 0 -> Epigrafia
        geomorfologia_mac Bacia hidrográfica -> Tesouro
        geomorfologia_mac Depressão tectónica -> Epigrafia
        geomorfologia_mac Planalto -> Epigrafia
        geomorfologia_mac Serra -> Povoado
Cronologia Idade Média
```

Figura 4.5 – Algumas regras de decisão do modelo preditivo para Trás-os-Montes Oriental ***servator***

Pode verificar-se, por exemplo, que para o Período Romano, quando a Topografia do terreno é um **Cume**, então provavelmente irá lá encontrar-se um **Povoado**. No entanto, se a Topografia for um **Talvegue**, então o tipo de arqueossítio poderá ser uma **Rede Viária**.

Os arqueólogos, nos seus projectos, formulam hipóteses sobre a forma como se distribuem os arqueossítios numa determinada região e da sua relação com o meio. As regras de decisão permitem estruturar e normalizar esse conhecimento, apresentando-o numa linguagem natural, clara e intuitiva.

Esta forma de visualização de padrões pode, portanto, constituir-se como uma ferramenta de trabalho para apoio à prospecção arqueológica. Mas pode, também, ser um importante elemento pedagógico e potenciar a aquisição de novo conhecimento.

Um outro exemplo de um módulo a integrar no ***servatis*** é um SIG de gestão Patrimonial, onde o Património está inventariado e georeferenciado numa BD relacional (*Oracle*). Na Figura 4.6. apresenta-se uma vista parcial da sua aplicação na elaboração de Cartas Arqueológicas.

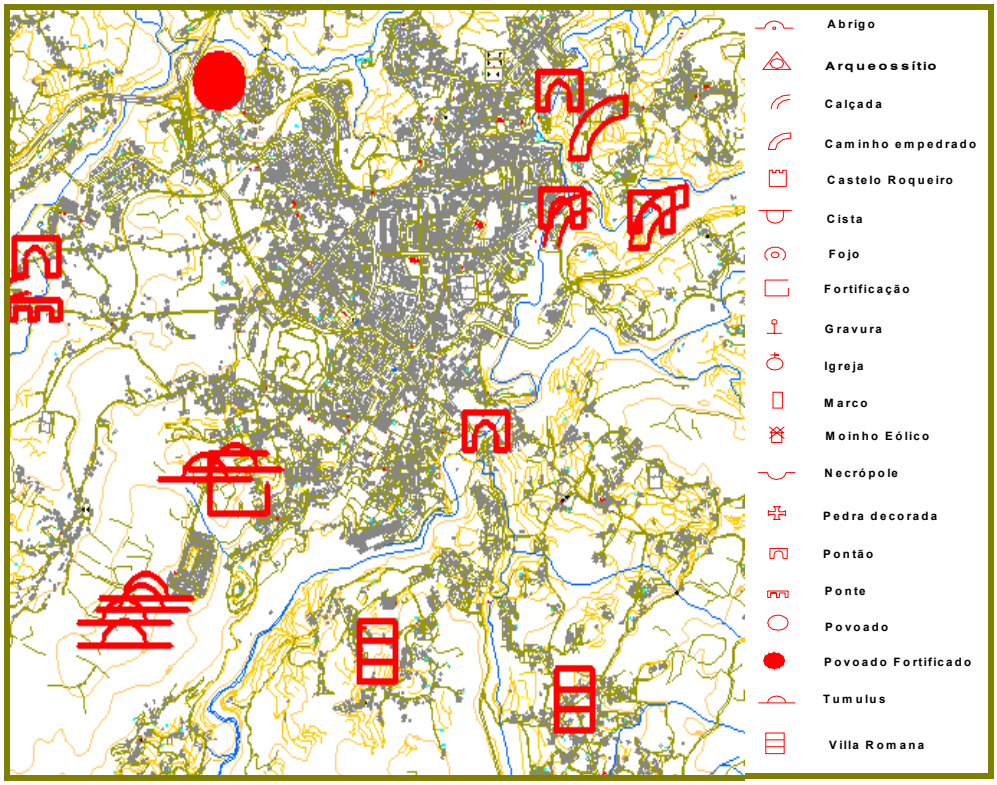


Figura 4.6 – Vista parcial de Carta Patrimonial (adaptado de [Botica *et al*., 2003c])

As tecnologias a utilizar no ***servatis*** são muito variadas, podendo referir-se, a título de exemplo, que o Sistema de Informação Arqueológica assenta numa Base de Dados *Oracle e* o SIG, apresentado na Figura 4.6, foi desenvolvido usando o software *Autodesk MapGuide*.

Apesar de serem utilizadas no sistema as mais variadas ferramentas, este deve apresentar-se com uma interface normalizada, definida de acordo com perfis de utilização a criar e a disponibilizar em *Intranets* e/ou na *Internet*.

4.4. Conclusão

O ***servatis*** foi desenvolvido a pensar nas actividades Arqueológicas e na sua divulgação. Por um lado, definiu-se um sistema que fizesse uma gestão integrada da

informação arqueológica, apoiando o desenvolvimento das actividades desenvolvidas na área. Mas, por outro lado, manteve-se sempre presente a necessidade de partilhar alguma desta informação com outras comunidades e com o público em geral.

A divulgação dos resultados de investigação, mais do que uma necessidade é uma obrigação. A Arqueologia, que lida com informação de crescente importância para as Organizações, necessita de um sistema visualizador, que divulgue o Património e incentive a sua preservação e valorização.

Respeita-se melhor o que se conhece, pelo que o acesso à informação arqueológica é muito importante e contribui para prolongar no tempo a nossa memória colectiva.

Capítulo 5

servator – preparação dos dados

O Património é definido na Carta de Malta de 1992 [Convenção, 1992] como fonte de memória colectiva e instrumento de estudo histórico e científico. Esse Património está presente nos vestígios, bens e outros indícios do passado. O trabalho dos arqueólogos tem sido inventariar e estudar este Património.

Desde o séc. XVIII, o Património inventariado tem aumentado de forma exponencial, constituindo uma fonte sempre crescente de dados. É com base nesses dados que se podem desenvolver modelos e ferramentas de apoio à investigação arqueológica para que seja, cada vez menos, uma actividade intrusiva e se evitem alterações ou danos de bens culturais. O modelo preditivo ***servator*** apoia a prospecção arqueológica e a identificação de áreas de risco onde pode existir Património. Insere-se, desta forma, na linha de acção recomendada pelas entidades internacionais para que o Património seja protegido, mesmo nos casos em que ainda não foi detectado, mas onde se admite que possa existir.

Neste capítulo será apresentada a ferramenta de trabalho utilizada e tarefas a realizar sobre os dados, com vista à identificação do modelo preditivo ***servator***, utilizando os princípios associados à DCBD, aplicada, como já foi referido, a uma BD de sítios arqueológicos de Trás-os-Montes Oriental. Este modelo consistirá num conjunto estruturado de regras que organizam a informação arqueológica e induzem a previsão dos locais onde poderão ser encontrados arqueossítios. De acordo com os valores que estão associados a determinados variáveis (altitude, topografia ou outras), o modelo irá conduzir à identificação, para esse local, de um tipo de sítio arqueológico.

A identificação do modelo preditivo será realizada no capítulo 6, aplicando-se algoritmos de DM aos dados e procedendo-se, de seguida, à validação e interpretação de resultados.

5.1. A ferramenta *Clementine*

A identificação do modelo ***servator*** é feita utilizando os princípios associados à DCBD, procurando-se padrões nos dados de inventários de sítios arqueológicos de Trás-os-Montes Oriental.

O *Clementine,* por ser uma ferramenta bastante completa, foi a ferramenta de DM escolhida para desenvolver este trabalho. Oferece um conjunto variado de funcionalidades de DM e grande flexibilidade para criar aplicações personalizadas.

Apresentam-se a seguir algumas características desta ferramenta [SPSS, 1999] [Rodrigues, 2000] que a tornam uma opção bastante atractiva:

- Abrange todo o processo de DCBD;
- Fornece uma interface gráfica intuitiva, para programação visual;
- Fornece um variado leque de algoritmos de modelação como as redes neuronais, redes de *Kohonen* e análises baseadas em árvores de decisão, regras de associação e regressão linear;
- Permite combinar técnicas e modelos;
- Trabalha com várias plataformas computacionais, correndo em sistemas Windows e Unix;
- Permite o acesso, via *Open Database Connectivity* (ODBC) às BD relacionais, trabalhando também com ficheiros de texto;
- Proporciona a avaliação da qualidade dos dados;
- Permite operações de manipulação dos dados como a filtragem, ordenação e agregação;
- Permite a exploração gráfica dos dados, incluindo as relações entre valores;
- Facilidade de elaboração de estatísticas descritivas sobre os dados.

Definidos os objectivos a atingir com o processo da DCBD, seleccionada a BD de trabalho e a ferramenta de DM a utilizar, apresenta-se a seguir a fase de preparação dos dados para identificação do ***servator***.

5.2. Preparação dos dados

Nesta secção e respectivas subsecções analisam-se os dados disponíveis para o processo de DCBD. O objectivo desta fase do trabalho consiste na análise da BD de arqueossítios de Trás-os-Montes Oriental, seleccionando os atributos que se julgam relevantes para o processo de DC, a fim de os preparar para serem submetidos aos algoritmos de DM.

As BD disponíveis não foram construídas tendo em vista a DC, pelo que os dados não estão ainda preparados para serem submetidos a este processo. É necessário proceder a operações de limpeza, tratamento de valores omissos e ainda outras operações de transformação dos dados que se julguem necessárias.

5.2.1. Selecção

Para identificar o modelo preditivo ***servator***, com base no processo de DCBD, escolheu-se a região de Trás-os-Montes Oriental como região piloto, para a qual existiam já dados de caracterização dos arqueossítios, estudos geográficos e conhecimento arqueológico. Trás-os-Montes é uma região com identidade própria e que, apesar da diversidade geomorfológica, possui uma simetria que lhe confere uma identidade comum. Existem trabalhos detalhados sobre a sua geografia e geomorfologia [Ribeiro *et al*., 1991], estudos de caracterização dos solos [Coba, 1991] e ainda um inventário bastante completo de arqueossítios [Lemos, 1993], cuja BD constitui a fonte de dados principal para este trabalho. A esta BD foram ainda acrescentados dados relativos a sítios inventariados em [Cruz, 2000] e de inventários realizados pela Unidade de Arqueologia da Universidade do Minho, no âmbito de projectos Directores Municipais (PDM).

Numa primeira fase, a fim de obter um elevado número de registos, procurou-se integrar na BD um conjunto de dados que abrangia duas áreas: Trás-os-Montes Ocidental e Oriental. Como adiante se descreve, este processo, embora tenha permitido obter uma BD bastante mais completa, acabou por apresentar dificuldades de várias ordens. Assim, ainda que o trabalho de preparação de dados aplicado a todo o conjunto, tenha sido útil como metodologia, decidiu-se limitar a BD a registos de Trás-os-Montes Oriental, uma vez que estes, embora em menor número, são mais coerentes e fidedignos.

A estrutura da BD disponível para este trabalho está representada, recorrendo a um diagrama Entidades-Relacionamentos, nas tabelas da Figura 5.1, onde são caracterizados os arqueossítios e o contexto geográfico e ambiental onde se inserem.

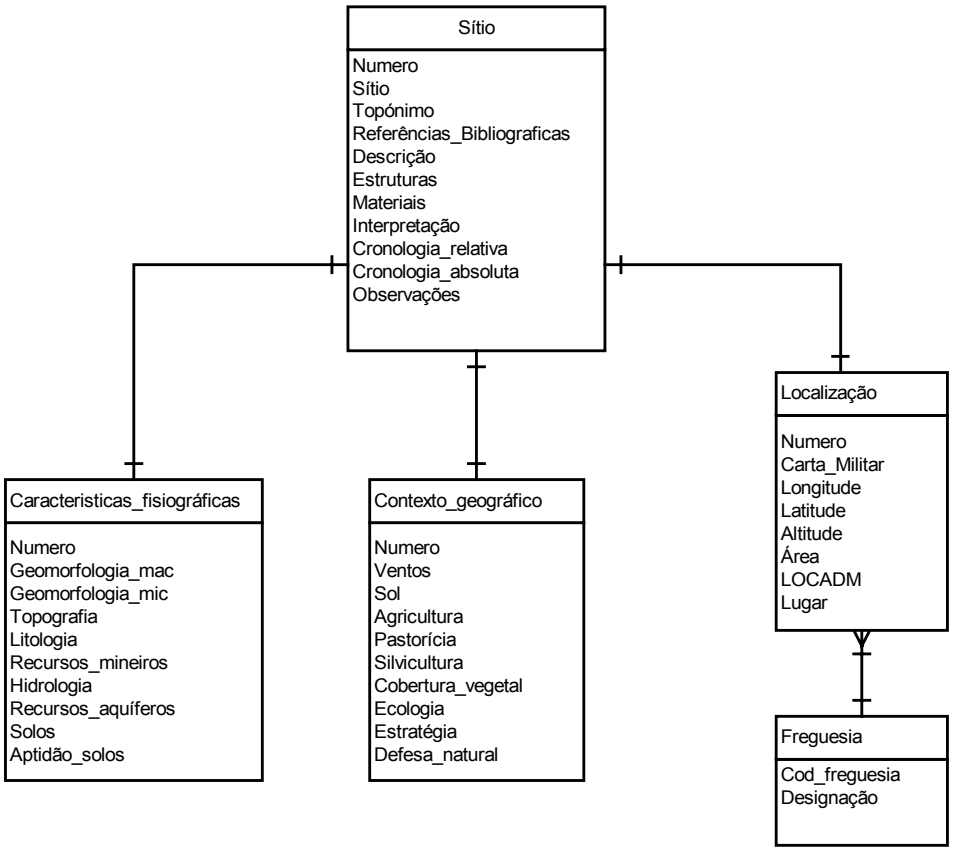


Figura 5.1 – Estrutura da BD de arqueossítios de Trás-os-Montes Oriental

A tabela **Sítio** corresponde à identificação de cerca de 2000 sítios arqueológicos e as tabelas **Localização** e **Freguesia** definem a sua localização geográfica. As características associadas ao relevo e recursos naturais, existentes na área envolvente do sítio arqueológico, são apresentadas na tabela **Características_fisiográficas.** A tabela **Contexto_geográfico** caracteriza as variáveis ambientais, tais como o grau de exposição solar do local ou de exposição aos ventos.

Para garantir que cada arqueossítio corresponde a um único registo individual na BD, de acordo com os requisitos impostos pelos algoritmos de DM [Berry e Linoff, 2000], agruparam-se todas as tabelas de dados numa única, armazenada no sistema *Microsoft Access,* sobre a qual a seguir se apresenta uma vista parcial de um conjunto de registos (Figura 5.2).

Sítio	Locadm	Altitude	Geomorfologia_mac	Estruturas	Cronologia
castelo de safres	170114	530	riba tua	não se observam fortificações	calcolítico
castelo de balsa	170117	0	vale do pinhão	referências a um castro no termo da aldeia de balsa	idade do ferro
cerca de freixo	170117	708	planalto de alijó	duas linhas de muralha de aparelho granítico de pedra picada	idade do ferro
alto da muralha	170118	630	planalto de alijó	castro com duas linhas de muralha;vestígios de romanização	idade do ferro
castelo de cheires	170112	360	vale do pinhão	castro com duas linhas de muralha e vestígios de romanização	romano
circa de casal de loivos	170104	475	vale do douro	fortificação de época indeterminada	idade do ferro
castelo de castedo	170105	510	vale do douro	não se observam estruturas;no entanto um muro que circunda o topo do rele	
alto de s, pedrinho	171217	353	vale do rabaçal	dois circuitos de muralha;um torreão no topo;vestígios de uma sepultura de l	idade do ferro
castro do cabeço	171207	850	planalto de carrazedo	duas linhas de muralha de aparelho de pedra granítica rectangular,disposto e	idade do ferro
castrilhao de coelhoso	40210	600	planalto de parada-izeda	restos de torreão e de uma linha de muralha muito destruída por trabalhos ag	idade do ferro
quatro caminhos	40205	610	depressao de bragança	local do achado de 2 estelas funerárias,uma de proculo reburrini,e outra com	romano
ponte das carvas	40205	550	depressao de bragança	ponte medieval de tabuleiro plano assente sobre três arcos em ogiva,constru	idade media
terronha de alimonde	40207	720	serra da nogueira	pequeno povoado defendido por fosso,torreão,imponente e bem comservada	idade do ferro
castro de carrazedo	40207	929	serra da nogueira	povoado com imponente torreão e muralha que delimita um espaço subcircu	idade do ferro
s, martinho de carrazedo	40207	820	serra da nogueira	ruinas de um edifício no cimo do cabeço (templo de s, martinho?)	romano
termo de carrazedo	40207	0	serra da nogueira	local de passagem da via romana,	romano
igreja de terroso	40213	880	planalto de espinhosela	duas estelas funerárias incorporadas nas paredes da igreja matriz de terroso	romano
s, tome de terroso	40213	890	planalto de espinhosela	povoado aberto;não se observam estruturas;numa rea apreciavel notam-se fi	romano
alto do carocedo	40214	880	planalto de parada-izeda	povoado médio defendido por muralha,de que restam escassos vestígios;acr	romano

Figura 5.2 – Vista parcial da tabela de dados

A Tabela 5.1. apresenta os atributos da tabela de dados, sistematizando o significado dos mesmos.

Atributos da tabela de sítios arqueológicos	
Atributo	**Significado**
Número	Número de catálogo do sítio
Sítio	Designação do sítio
Topónimo	Outro nome pelo qual o sítio é conhecido
LOCADM	Código geográfico nacional
Lugar	Aldeia em cujo termo se localiza o sítio
Descrição	Descrição do sítio
Estruturas	Estruturas arqueológicas observáveis
Materiais	Tipo de materiais observáveis
Interpretação	Interpretação do sítio
Cronologia_relativa	Cronologia relativa
Referencias_cronologicas	Outras referências cronológicas
Referencias_bibliográficas	Referências bibliográficas
Carta_militar	Carta militar 1:25 000 em que se localiza o sítio
Longitude	Valor da longitude em coordenadas Gauss
Latitude	Valor da latitude em coordenadas Gauss
Altitude	Altitude do ponto central do sítio arqueológico
Área	Área ocupada pelo sítio[7]
Geomorfologia_mac	Unidade geomorfológica em que se insere o sítio
Geomorfologia_mic	Tipo de relevo onde está instalado o sítio
Topografia	Zona do relevo ocupada pelo sítio

[7] Este atributo Área foi quase exclusivamente aplicado aos sítios do tipo **Castro** que conservaram intactas as linhas de muralhas. Os **Castros** eram povoados rodeados de muros de pedra solta, alcandorados no cimo dos montes. As casas, também de pedra solta, eram na sua maioria redondas e cobertas de giestas ou de colmo.

Litologia	Natureza do substrato rochoso
Recursos_mineiros	Recursos mineiros existentes na área envolvente
Hidrologia	Cursos de água com os quais o sítio se articula
Recursos_aquíferos	Abastecimento de água na zona envolvente
Solos	Tipo de solo da área
Aptidão_solos	Grau de aptidão do solo
Ventos	Tipo de exposição ao vento
Sol	Tipo de exposição ao sol
Agricultura	Sistema agrícola actual da área envolvente
Pastorícia	Espécies de gado associadas ao sistema agrícola
Silvicultura	Espécies arborícolas cultivadas
Cobertura_vegetal	Revestimento vegetal espontâneo
Defesa_natural	Grau de defesa natural
Ecologia	Zonação ecológica do Atlas do Ambiente
Estratégia	Grau de posicionamento geoestratégico do sítio
Observações	Observações

Tabela 5.1 – Atributos da tabela de dados

Da análise dos atributos da tabela de dados, constata-se que se procurou, não só recolher e registar dados relativos às características do arqueossítios inventariados, mas também, das variáveis ambientais que os rodeiam. Algumas correntes da Arqueologia atribuem um valor expressivo às condicionantes ambientais, tratando os sítios em conjunto com o meio envolvente. A latitude, a longitude, o relevo e o clima influenciam a vegetação, a qual, por sua vez, condiciona a vida animal [Renfrew e Bahn, 1991]. Para a identificação do ***servator***, como modelo preditivo de apoio à localização de arqueossítios, irão ser consideradas as variáveis ambientais, dado que, mesmo que não tenham sido determinantes, alguma influência podem ter tido na escolha dos locais de assentamento.

Com base na tabela de dados foi feita uma selecção dos atributos a considerar, o que constituiu a primeira fase do processo de DCBD. Tendo em mente a finalidade do trabalho, foram retirados todos os dados que, por terem carácter meramente informativo, não são relevantes para este processo. Assim, foram retirados os atributos indicados na Tabela 5.2.

Atributos com carácter meramente informativo	
Atributo	**Atributo**
Número	Descrição
Sítio	Carta Militar
Topónimos	Área
LOCADM	Referências bibliográfica
Lugar	Observações

Tabela 5.2 – Atributos retirados da tabela de dados

A inclusão destas colunas nos dados a tratar não iria ser um contributo positivo para a construção do modelo. Pelo contrário, estaria a aumentar o número de atributos, sem facultar informação relevante.

5.2.2. Tratamento

As BD, dos mais variados domínios, foram organizadas e preenchidas com objectivos diversos, que não a DC. Deste modo, estas BD apresentam-se de um modo geral muito incompletas, com dados vagos e imprecisos [Rodrigues *et al.,* 1998]. Neste contexto, as tarefas de selecção, limpeza e de transformação dos dados assumem um relevo muito especial e consomem grande parte dos recursos temporais. Para construção de modelos válidos e úteis, há necessidade de incorporação intensiva de conhecimento da área.

O modo como os dados foram recolhidos, a forma como foram introduzidos nas BD e ainda as operações de junção de várias tabelas, fizeram com que se acumulassem alguns erros, redundâncias e inconsistências nos dados. Para reduzir, ou mesmo eliminar estes problemas, serão apresentadas de seguida algumas tarefas a realizar sobre os dados na fase de tratamento, nomeadamente, a eliminação de registos duplicados e de inconsistências, o tratamento de dados omissos e de valores isolados.

5.2.2.1. Registos duplicados

A tabela de dados sobre a qual se realizou este trabalho, resultou da junção de várias tabelas dispersas. Essas tabelas foram preenchidas por pessoas diferentes, muitas vezes com objectivos diversificados, o que se reflecte na utilização de estruturas da BD e normas de preenchimento diferenciadas. A junção destas tabelas provocou a introdução

de duplicações e inconsistências. Também se constatou que, muitas vezes, são utilizadas diferentes terminologias para designar o mesmo conceito. Foram detectadas, por exemplo, designações diferentes para o mesmo sítio, o que originou algumas redundâncias.

Para tentar eliminar estas redundâncias, ordenaram-se os sítios pelos atributos **Estruturas**[8], **Materiais** e **Cronologia_relativa** e ainda pela sua localização geográfica (**Latitude e Longitude**). No caso de existirem idênticos vestígios do passado para o mesmo local, analisam-se os sítios, verificando se é uma duplicação para o mesmo sítio arqueológico.

5.2.2.2. Dados inconsistentes

Consideram-se inconsistentes ou corrompidos os dados que estão semântica ou sintacticamente errados. Dados semanticamente errados caracterizam-se por conterem valores de um tipo diferente do esperado. Por exemplo, valores de altitude, latitude ou longitude do tipo texto não são esperados, uma vez que se tratam de dados de natureza numérica. Os dados sintacticamente errados são aqueles que contêm valores fora do intervalo de valores admissíveis. Considerando como exemplo a região estudada e o sistema de coordenadas utilizado, valores negativos de altitude, latitude ou longitude não são esperados.

Alguns dados semanticamente errados foram detectados durante esta fase do trabalho. Nalguns casos, era visível a atribuição de valores a uma variável, quando na realidade eles correspondiam a atributos da coluna precedente ou seguinte. A título de exemplo, referem-se os dados do atributo **Recursos_aquíferos,** que foram encontrados na coluna de **Solos**, que lhe é adjacente.

A contribuir para uma taxa mais elevada deste tipo de erros, está também o facto destes valores terem sido introduzidos na BD sem que tenha sido desenvolvida qualquer aplicação de interface com o utilizador, de modo a validar a entrada dos dados. As BD utilizadas neste trabalho foram preenchidas recorrendo à introdução manual dos dados, directamente na BD, sem qualquer procedimento automático de validação. Destes

[8] Para facilidade de leitura irão ser apresentados os nomes dos **atributos** com o tipo de letra *Times New Roman* 12 *Bold* e os **valores** associados a atributos serão escritos a *Arial* 11 *Bold.*

processos de digitação resultam inevitavelmente erros, alguns resultantes de lapsos ortográficos, outros de leitura errada de valores e até, como já foi referido, do registo de dados no campo inadequado.

Para os dados alfanuméricos, desenharam-se gráficos que representam os diversos valores possíveis e que permitiram, em colaboração com o especialista da área, identificar os maus valores, quer estes fossem resultantes de lapsos de digitação ou de inconsistências de outra ordem.

Os erros mais frequentes estavam associados com deficiências na digitação de dados, tais como **Vretente,** em vez de **Vertente**.

Verificou-se ainda que, para o mesmo atributo, eram utilizados indiscriminadamente femininos e masculinos, bem como singulares e plurais. A todos os registos foram associados valores no mesmo género e número (singular).

Os valores numéricos dos campos **Altitude**, **Longitude** e **Latitude,** foram colocados por ordem crescente, para detectar atributos fora da ordem de grandeza esperada. Segundo o sistema de coordenadas Gauss, utilizado na BD inicial, a região a Norte do Douro tem coordenadas que variam de 135 000 a 370 000, para a **Longitude** e de 450 000 a 575 000, para a **Latitude**. Todos os registos foram comparados com estes valores de referência, tendo-se verificado que alguns saíam fora desta gama de valores.

Os valores da **Latitude** que estavam para além dos limites superiores foram corrigidos de imediato. Verificou-se resultarem de lapsos de digitação, por omissão da vírgula.

Dados inválidos de **Latitude**, abaixo dos 450 000, foram observados em conjunto com os respectivos valores de **Longitude** e verificou-se ter havido uma troca entre eles. Uma consulta das cartas geográficas confirmou esta troca, tendo sido efectuadas as respectivas correcções.

Os restantes valores foram corrigidos, procurando-se na cartografia militar os valores correctos, sempre que as indicações sobre o local eram precisas. Nos casos em que a correcção não era possível, foi-lhes atribuído um valor nulo.

Validaram-se também os valores de **Altitude,** considerando que todos terão que ser positivos e inferiores a 1350 m, não tendo sido detectados valores anómalos.

5.2.2.3. Dados omissos

A BD sobre a qual se realiza este trabalho, não tem um elevado número de registos, mas sim, um elevado número de atributos para cada registo. Os atributos são normalmente do tipo descritivo, encontrando-se muitos deles por preencher.

Este elevado número de dados omissos deve-se, em grande parte, ao facto de se congregarem dados resultantes de vários trabalhos, realizados com objectivos diferenciados. Atributos que para um trabalho eram relevantes e sistematicamente caracterizados, poderiam apresentar-se como menos relevantes noutros, ficando muitas vezes por preencher.

O tratamento de valores omissos pode ser feito manualmente, analisando-se cada valor caso a caso, ou através de um processo automático, a definir de acordo com o tipo de dados em falta. No caso presente, dado o reduzido número de registos, pouco mais de 2000, e sendo os dados maioritariamente alfanuméricos, não se optou pela atribuição automática de valores prováveis, esperados ou mesmo gerados. Corria-se o risco de introduzir no sistema uma elevada taxa de ruído.

O número de dados omissos era variável de atributo para atributo. Cada um foi analisado caso a caso, sendo feita a atribuição manual de valores, sempre que eram identificados inequivocamente. Foi o caso dos valores omissos para a **Altitude**, a **Geomorfologia_mac, Geomorfologia_mic**, **Topografia** e **Solos**, que estavam identificados com as respectivas coordenadas geográficas (**Latitude** e **Longitude**). Foram utilizadas as cartas topográficas e de solos para se encontrar os dados em falta.

Quando a percentagem de valores omissos era muito elevada (acima dos 50%) e, não era viável proceder ao seu preenchimento, optou-se por remover estes atributos da tabela de dados. Foi o caso dos dados associados às características do local, tais como, **Ventos**, **Sol**, **Recursos_mineiros**, **Agricultura**, **Pastorícia**, **Silvicultura**, **Cobertura_vegetal**, **Paisagem**, **Defesa_natural**, **Ecologia**, **Estratégia.** Estes dados só podem ser recolhidos no local, o que não era viável e saía fora do âmbito deste trabalho.

Refira-se ainda que o **Clima, Agricultura**, **Pastorícia**, **Silvicultura, Cobertura vegetal** e **Ecologia** são variáveis ambientais que sofreram alterações profundas ao longo do tempo. Os valores registados reportam-se assim a dados actuais e que, possivelmente, já não são idênticos aos existentes no momento da escolha dos locais de povoamento. Assim, a sua relação com os modelos de povoamento, para os diversos períodos cronológicos, pode ter sido alterada.

No final desta fase do trabalho, havia ainda valores cuja identificação não tinha sido possível. No entanto, os atributos que foram mantidos continham já uma percentagem mais baixa de valores omissos e apresentavam os restantes campos devidamente preenchidos.

Outros valores omissos foram mantidos porque não se tratavam de dados em falta, mas sim de valores que, para o tipo de arqueossítio em causa, não são passíveis de serem preenchidos. Por exemplo, os sítios do tipo **Rede Viária** ou **Termo**, que não se restringem apenas a um local, mas atravessam uma série de locais, não podem ser associados a uma determinada **Geomorfologia_mac, Geomorfologia_mic**, ou mesmo **Altitude.** Assim, para estes casos, há atributos que não são intencionalmente preenchidos.

5.2.2.4. Valores isolados

Os histogramas realizados sobre os dados de entrada permitiram identificar alguns valores isolados. Foi feita uma avaliação, registo a registo, de forma a distinguir os casos isolados de inconsistências ou erros de digitação. Foram identificados valores isolados, tais como **Alvéolo** e **Chã** para a coluna **Topografia**, **Paleolítico superior** para **Cronologia** ou ainda **Vertissolos** para o **Tipo de solos**.

A decisão de manter ou excluir cada valor isolado foi tomada, de acordo com o parecer do especialista em Arqueologia, conforme a sua representatividade e importância atribuída para a construção do sistema preditivo. Sempre que estes dados não eram considerados fundamentais para a construção do modelo, ou que não constituíam um número suficiente para serem modelados, procedeu-se à sua remoção, para simplificação do mesmo.

5.2.3. Pré-processamento

Terminada a fase de tratamento de dados, efectua-se o seu pré-processamento, onde os dados são transformados na sua forma final, antes de serem analisados pelos algoritmos de DM.

Os modelos identificados poderão ser bons ou maus modelos, de acordo com os dados que são utilizados. A aplicação directa de técnicas de DM sobre dados resultará num modelo, cujo grau de confiança é, provavelmente, bastante menor que o obtido com dados tratados e pré-processados.

Não sendo possível utilizar métodos totalmente automáticos para realizar as tarefas de preparação dos dados, esta fase absorve uma grande percentagem de tempo dos projectos de DC. O seu grau de importância é grande, uma vez que, como já foi referido, sem bons dados não se identificam bons modelos. O pré-processamento dos dados requer muito tempo, não só por parte de quem efectua as tarefas na área das TI, mas também no âmbito da Arqueologia, uma vez que a incorporação de conhecimento específico é fundamental.

Como foi já sublinhado, nas tarefas de pré-processamento os dados são preparados para serem submetidos aos algoritmos de DM. Esta fase tem o objectivo de incorporar o máximo de informação relevante para a identificação do modelo, com o mínimo número de linhas e colunas da amostra. Limitando-se o número de variáveis a tratar, simplifica-se, viabilizando, a utilização de alguns algoritmos de exploração de dados, que se tornam menos eficientes com um grande número de variáveis.

Partindo do conjunto inicial de dados, foram realizadas as tarefas de selecção e tratamento dos dados, obtendo-se uma tabela com 15 atributos. A Tabela 5.3 apresenta os atributos da BD sobre os quais serão realizadas as tarefas de pré-processamento.

Sítios arqueológicos		
Latitude	Topografia	Cronologia_relativa
Longitude	Litologia	Referências_cronológicas
Altitude	Hidrologia	Estruturas
Geomorfologia_mac	Recursos_aquíferos	Materiais
Geomorfologia_mic	Solos	Interpretação

Tabela 5.3 – Atributos da BD, após tratamento dos dados

A redução do número de colunas e linhas pode ser realizada recorrendo às seguintes operações:

- Normalização de dados;
- Agrupamento de registos;
- Remoção de atributos;
- Discretização de atributos;
- Combinação de variáveis;
- Eliminação de variáveis altamente correlacionadas.

No presente trabalho foram realizadas operações de combinação de atributos correlacionados**,** normalização de dados, especialização e discretização de atributos, descritas a seguir detalhadamente.

5.2.3.1. Combinação de atributos correlacionados

Analisando-se agora os dados obtidos após as operações de selecção e tratamento, verifica-se que algumas colunas contêm descritivos que não obedecem a qualquer formalismo de apresentação. É utilizada uma linguagem natural, umas vezes abreviada, outras vezes com descrições muito detalhadas.

Apresenta-se na Figura 5.3, a título de exemplo, uma vista sobre os conteúdos dos campos **Estruturas**, **Materiais** e **Interpretação**.

Estruturas	Materiais	Interpretação
achado de moeda suevica	moeda suevica de ouro cunhada em braga pelo rei suintila,	achado monetário isolado
fosso;torreão;três linhas de muralha;construções circulares	dois machados de cobres calcolíticos;cerâmicas romanas,incluindo t,	Castro Romanizado
povoado fortificado total e recentemente arrazado por obras	tégulas;imbrices;cerâmica comum romana;4 estelas funerárias citadas	Castro Romanizado
níveis da idade do ferro;castelo;cerca;habitações medievais;	cerâmica calcolítica;cerâmica carenada do bronze;cerâmica do ferro;te	povoado pre-hist;Castro Castelo Me
talude que circunda o topo do cabeço;nao se observam par	tégulas;imbrices;tijoleiras;cerâmica comum;moedas de bronze;moeda	povoado romano;vicus
local onde se encontra actualmente o touro de parada	atalaia zoomorfica;touro	Escultura zoomórfica; Ara
Fortificações;Ruínas de construções medievais	Cerâmica da Idade do Ferro; Cerâmica Medieval	Castro; Povoado Medieval
não se observam estruturas;no entanto um muro que circun	fragmentos de tégulas;imbrices;tijolos;dolios;cerâmica comum romana	Povoado Romano fortificado
local de passagem da via romana	2 marcos miliários,um anepigrafe,o outro de constâncio	rede viária romana
alicerces de um antigo templo;sepulturas de lages	1 pequeno cavalo de bronze;fragmentos de terra sigillata hispanica alto	necrópole romana;templo e necrópol

Figura 5.3 – Vista parcial de atributos da tabela de dados

O conteúdo destas colunas deverá ser normalizado, devendo avaliar-se os relacionamentos entre os dados. Por exemplo, o atributo **Materiais** descreve quais os achados encontrados no local. Essa informação apoia a interpretação arqueológica que é feita e leva ao preenchimento da coluna **Interpretação.** Por exemplo, para o segundo

registo da Figura 5.3., como estruturas identificaram-se "fosso; torreão; três linhas de muralha; construções circulares". Estes dados, conjugados com o registo dos materiais "dois machados de cobre calcolíticos; cerâmicas romanas,incluindo sigillata clara; uma fibula; um alfinete de cabeça", ajudaram o processo de interpretação, que levou à identificação de um sítio do tipo "Castro Romanizado".

Assim, para o objectivo deste trabalho, a coluna **Materiais** possui conteúdos relacionados com os dados da coluna **Interpretação.** Analisando os atributos, na perspectiva de avaliar os conteúdos e as relações entre eles, verificou-se que atributos como **Hidrologia** e **Recursos_aquíferos**, **Cronologia_relativa** e **Referências_cronológicas**, bem como **Estruturas**, **Materiais** e **Interpretação**, são colunas que contêm informação relacionada entre si. São campos do tipo texto, com várias descrições relacionadas com o mesmo conceito e que, para aplicação de algoritmos de DM, há todo o interesse em que sejam normalizados e combinados, de forma a reduzir o número de variáveis e tornar mais simples o modelo.

Antes mesmo de realizar o processo de normalização destes atributos, decidiu-se agrupar dados de variáveis correlacionadas, e só depois realizar as tarefas inerentes à sua normalização.

As colunas **Cronologia_relativa** e **Referências_cronológicas** atribuem uma ou várias cronologias ao sítio. As colunas **Hidrologia** e **Recursos_aquíferos** caracterizam a proximidade do arqueossítio aos cursos de água que os rodeiam e os atributos **Sítio**, **Estruturas**, **Materiais** e **Interpretação** contêm dados relativos ao sítio arqueológico e ao material encontrado no local, bem como a interpretação arqueológica. Estas colunas têm uma existência separada, importante para os trabalhos arqueológicos, mas que para efeitos da aplicação de algoritmos de DM pode ser combinada.

Partindo dos dados referentes às **Cronologia_relativa** e **Referências_cronológicas,** criou-se uma nova coluna **Cronologia,** onde foram atribuídos valores relativos à cronologia dos sítios arqueológicos.

As colunas **Hidrologia** e **Recursos_aquíferos** pretendem relacionar os sítios arqueológicos com a proximidade ao importante recurso que é a água. Na **Hidrologia** identifica-se o curso de água mais próximo do local, bem como os cursos de água onde

desagua, directa ou indirectamente. Nos exemplos da Figura 5.4, pode ver-se que o primeiro sítio arqueológico mostrado se situa nas proximidades do **Rio Torno**, que é afluente do **Rio Louredo**, que desagua no **Rio Tâmega** e que, por sua vez, desagua no **Rio Douro**. Na coluna **Recursos_aquíferos** caracteriza-se o tipo de recurso, podendo ser uma **Nascente, Linha de água**, **Fonte, Ribeira** ou **Rio**.

Hidrologia	Recursos aqualiferos
Rio Torno > Rio Louredo > Rio Tâmega > Rio Douro	Nascentes; Rio Torno
Rio Fervença > Rio Sabor	Ribeira; poços
Rio Douro	Rio Douro
Ribeira de Terroso > Rio Baceiro	Fonte de S, Tomé
Rib Castro > Ribeira Vale Pereiro > Rib Zacarias > Sabor	Ribeiro
Linhas de água > Ribeiro > Rio Pinhão > Rio Douro	
Linhas de água > Ribeira > Rio Pinhão > Rio Douro	Linhas de água
Linhas de água > Ribeira de S. Mamede > Rio Tua	Linhas de água
Linhas de água > Ribeira de S. Mamede > Rio Tua	
Linhas de água > Ribeira das Canadas > Rio Douro	
Linha de água > Rio Calvo > Rio RaBaçal > Rio Tua	Ribeira
Divide bacias do Sabor e Tuela	
Divide bacias do Sabor e Tuela	
Corgo do Carvalhal > Rio Tâmega > Rio Douro	Linhas de água
Corgo do Bidoedo > Rio Tâmega > Rio Douro	Nascentes; Regatos

Figura 5.4 – Vista sobre os atributos **Hidrologia** e **Recursos_aqualíferos**

Para efeitos de DM, decidiu-se criar uma nova coluna, **Hierarquia_hidrográfica**, à qual foram atribuídos valores de 1 a 6, de acordo com o posicionamento do sítio arqueológico relativamente ao recurso hídrico mais próximo.

A rede hidrográfica de Trás-os-Montes Oriental, caracterizada no capítulo 2 (subsecção 2.1.2), mostra que todos os cursos de água desaguam no "rio Douro", considerado curso de água principal da região. A todos os sítios localizados próximo deste rio atribuiu-se o valor **1** na **Hierarquia_hidrográfica.**

Os sítios localizados próximos dos rios que desaguam no **Rio Douro**, nível **1,** tais como o **Rio Tua** ou o **Rio Sabor**, têm o valor **2** na coluna **Hierarquia_hidrográfica,** isto é, mais 1 que o nível anterior.

O nível de determinado curso de água corresponderá sempre ao valor numérico, mais um, resultante do somatório do número de cursos de água onde desagua, até chegar ao **Rio Douro**. Por exemplo, o **Rio Tuela** tem associado o valor **3** na **Hierarquia_hidrográfica.** Desagua no **Rio Tua**, que por sua vez é afluente do **Rio Douro**, o que contabiliza dois cursos de água.

De acordo com os critérios apresentados, atribuíram-se valores à coluna **Hierarquia_hidrográfica,** associando-a ao nível do curso de água que está mais próximo do sítio arqueológico (Figura 5.5).

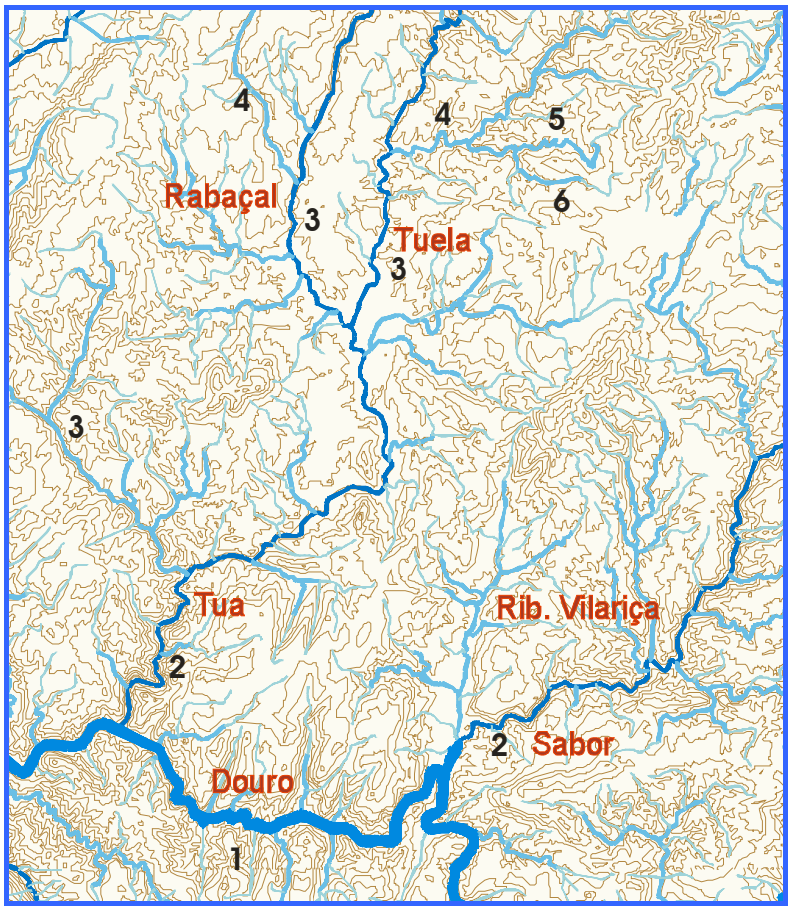


Figura 5.5 –Valores de **Hierarquia_hidrográfica**

As colunas **Estruturas**, **Materiais** e **Interpretação** caracterizam os arqueossítios, descrevendo as estruturas encontradas, os achados aí localizados e a interpretação arqueológica. Toda esta informação permite identificar o sítio.

Para classificar o tipo de sítio, foi criada uma nova coluna designada por **Tipologia**, à qual foram atribuídos valores em função dos conteúdos das colunas **Estruturas**, **Materiais** e **Interpretação**.

O atributo **Tipologia** assume um papel de particular importância neste processo, uma vez que este será o atributo que se espera que o modelo ***servator*** venha a prever.

A figura 5.6 mostra as alterações produzidas na tabela de dados, destacando as colunas que foram combinadas para dar origem a uma nova coluna.

Estruturas
Interpretação
Latitude
Longitude
Altitude
Geomorfologia_mac
Geomorfologia_mic
Topografia
Litologia
Recursos_aquíferos
Solos
Cronologia_relativa
Referências_cronológicas

Tipologia
Latitude
Longitude
Altitude
Geomorfologia_mac
Geomorfologia_mic
Topografia
Litologia
Hierarquia_hidrográfica
Solos
Cronologia

Figura 5.6 – Combinação de atributos correlacionados

Após a criação das novas colunas **Hierarquia_hidrográfica**, **Cronologia** e **Tipologia**, as colunas de dados que lhes deram origem foram eliminadas, ficando a tabela com 11 atributos.

Os sítios arqueológicos, depois de interpretados, são associados a um tipo, de acordo com as funções a que se destinavam. Como já foi acima referido, este é o atributo a prever pelo modelo ***servator***. Assim, a tabela de dados tem no seu campo **Tipologia** os valores associados ao tipo de sítio, e que são: **Abrigo**, **Povoado**, **Castro, Necrópole, Fortificação, Rede Viária, Limites, Santuário, Tesouro, Estatuária, Arte Rupestre, Epigrafia, Mineração** e **Achado Isolado.**

Como resultado da interpretação feita sobre os sítios arqueológicos, para além do tipo de sítio, associa-se também o período cronológico da sua origem.

Na figura 5.7 apresenta-se a distribuição dos arqueossítios inventariados de acordo com a cronologia que lhes está associada.

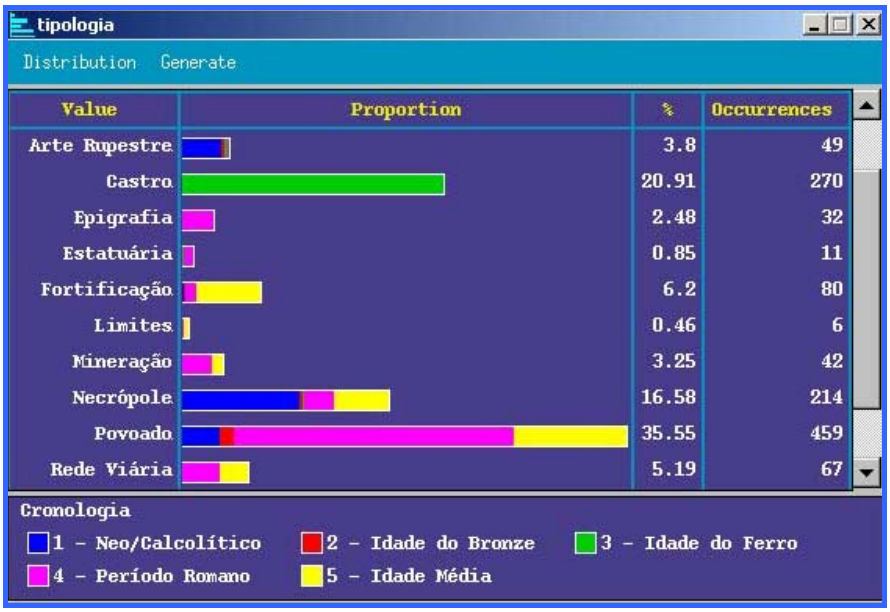


Figura 5.7 – Distribuição de tipos de sítios arqueológicos por **Cronologia**

De seguida será feita uma caracterização sumária dos tipos mais representativos encontrados em cada período cronológico:

- **Neo/Calcolítico** – os vestígios mais antigos da ocupação humana, registados na BD de trabalho, apontam para este período cronológico (**Neolítico** e **Calcolítico**). Os sítios que na maioria são do tipo **Necrópole**, denominam-se também por mamoas, antas ou *dólmens* e estão associados a uma função funerária. São pouco diversificados os outros tipos de arqueossítios inventariados, reduzindo-se a alguns **Povoados** e **Arte Rupestre**;

- **Idade do Bronze** – são poucos os sítios da **Idade do Bronze** inventariados, sendo a maioria do tipo **Povoado**. Para além deste tipo há ainda um número muito limitado do tipo **Necrópole, Achado Isolado, Arte Rupestre** e **Fortificação**;

- **Idade do Ferro –** nos sítios arqueológicos referentes a este período dominam os **Castros**. Estes são povoados fortificados, associados a este período cronológico. Para além dos **Castros,** registam-se também sítios do tipo **Santuário** e **Tesouro**, embora em escasso número;

- **Período Romano** – neste período, o tipo de arqueossítios com maior número de registos é o **Povoado**. Embora com menos representatividade, encontram-se ainda sítios do tipo **Rede Viária**, **Epigrafia**, **Necrópole**, **Mineração**, **Santuário**, **Tesouro**, **Fortificação** e **Estatuária**;
- **Idade Média** – no inventário de sítios arqueológicos, o tipo mais representado neste período cronológico é o **Povoado.** No entanto, também existem alguns registos de tipos como **Fortificação, Rede Viária, Necrópole** e **Santuário**.

5.2.3.2. Normalização dos dados

A normalização de dados permite uniformizar nomenclaturas, devendo ser aplicada sempre que sejam utilizadas diferentes designações para o mesmo conceito. Desta forma os dados são reduzidos à mesma escala. Com a normalização dos dados pode limitar-se o número de variáveis, simplificando o modelo e tornando mais simples a sua interpretação.

Estas operações nem sempre são simples de realizar, nomeadamente na área da Arqueologia, onde se utilizam diferentes vocábulos para exprimir idênticos conceitos. É necessário um acompanhamento constante do especialista da área, para que seja possível realizar esta tarefa de forma eficaz e sem introduzir erros no sistema.

Analisaram-se os campos da tabela de dados **Geomorfologia_mac**, **Geomorfologia_mic, Topografia**, **Litologia**, **Solos**, **Cronologia** e **Tipologia** a fim de serem normalizados, procurando desta forma reduzir o número de valores possíveis a serem utilizados na aplicação de algoritmos de DM.

O relevo da região está caracterizado na BD através de três atributos distintos: a **Topografia**, a geomorfologia macro (**Geomorfologia_mac**) e a geomorfologia micro (**Geomorfologia_mic**).

Para o atributo **Geomorfologia_mac** definiram-se as classes de valores possíveis, as quais estão representadas no gráfico da Figura 5.8.

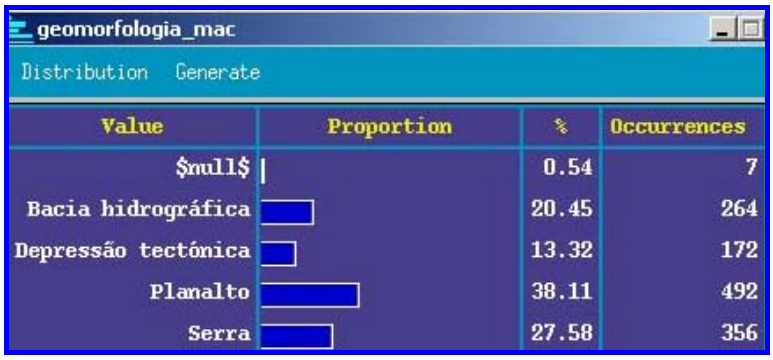


Figura 5.8 – Classes de valores para **Geomorfologia_mac**[9]

Estas classes foram identificadas de acordo com as classificações geomorfológicas estabelecidas para a região [Ribeiro *et al.*, 1991] e todos os valores da coluna **Geomorfologia_mac** da tabela de dados foram reclassificados de acordo com esses critérios.

Assim, valores como **Depressão de Macedo**, **Depressão de Freixo**, **Depressão de Bragança**, **Depressão Tectónica,** foram todos substituídos pelo valor **Depressão tectónica.**

Para atributos como **Vale da Vilariça** ou **Vale do Douro**, fez-se a correspondência, caso a caso, a uma das classes pré-definidas. Para os exemplos apresentados substituiu-se **Vale da Vilariça** por **Depressão Tectnónica** e **Vale do Douro** por **Bacia Hidrográfica.**

Da mesma forma, os atributos que continham a designação de **Planaltos** e **Serras** foram simplificados e normalizados, utilizando-se o valor da classe de **Geomorfologia_mac** correspondente (**Planalto** e **Serra**).

Para alguns registos foram encontrados diferentes designativos para representar o mesmo conceito de **Geomorfologia_mac**, pelo que se optou pela sua normalização. Foi o caso de **Lombada,** que se fez equivaler a **Planalto** e de **Transbaceiro** que se fez a equivalência com **Bacia Hidrográfica**.

[9] As distribuições são pouco homogéneas, no entanto correspondem a classes utilizadas em arqueologia e associadas aos sítios arqueológicos da BD. Sempre que existam classes de valores já definidas ou adoptadas em Arqueologia, elas serão utilizadas. Apenas nos casos em que esta predefinição não exista, será feita uma distribuição de valores por classes, utilizando outros critérios, nomeadamente os da distribuição mais ou menos uniforme dos dados. É o caso das classes de valores criadas para a **Latitude** e **Longitude**, apresentados mais adiante nesta rubrica.

Para o campo **Geomorfologia_mic** foram também criadas classes de valores possíveis, considerando as classificações geomorfológicas existentes para a região [Ribeiro *et al.*, 1991]. De acordo com as classes de valores criadas, transformaram-se todos os valores da coluna **Geomorfologia_mic,** cuja distribuição se apresenta na Figura 5.9.

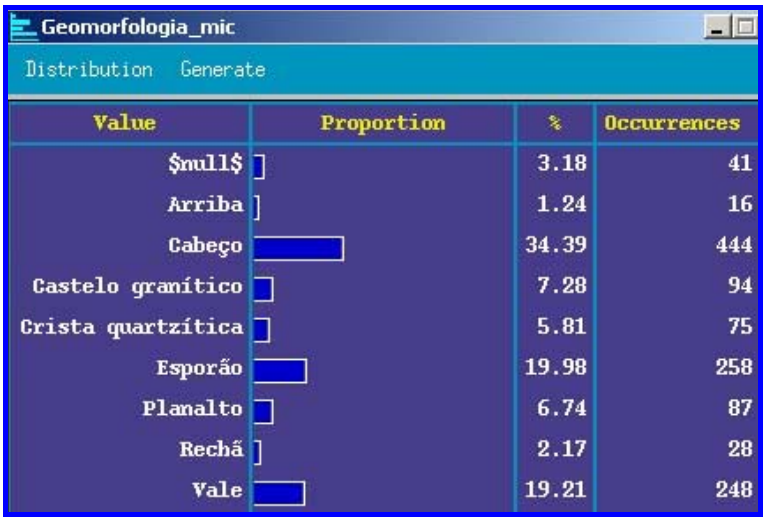


Figura 5.9 – Distribuição dos valores de **Geomorfologia_mic**

Durante a normalização de **Geomorfologia_mic**, fez-se, por exemplo, a substituição dos valores de **Cerro**, **Outeiro** e **Lombeiro** por **Cabeço**. Valores de **Crista, Crista Quartzítica** e **Sinclinal** foram reduzidos a um único valor (**Crista Quartzítica**).

Para o atributo **Topografia** procedeu-se de forma idêntica ao realizado para a **Geomorfologia_mic** e **Geomorfologia_mac**, isto é, definiram-se primeiro as classes de valores possíveis (Figura 5.10) e realizaram-se todas as equivalências necessárias entre os valores alocados e as classes definidas.

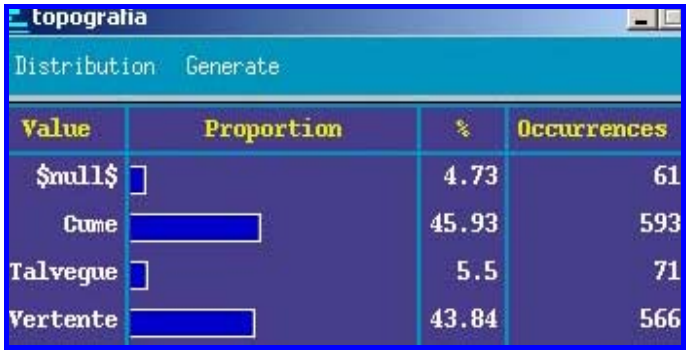


Figura 5.10 – Distribuição de valores de **Topografia**

O campo **Litologia**, que assinala o tipo de massas rochosas associadas a determinado local, sofreu também um processo de normalização, definindo-se as classes de valores de **Litologia** existentes para a região de Trás-os-Montes Oriental (Figura 5.11).

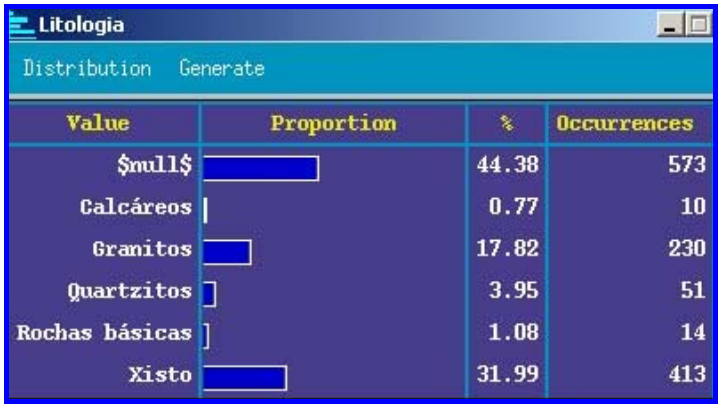


Figura 5.11 – Distribuição de valores de **Litologia**

Ao nível da **Litologia** a Região de Trás-os-Montes Oriental possui essencialmente rochas do tipo **Xisto** e, esporadicamente, algumas zonas de **Granito** ou nichos de outros tipos de rochas, embora sem grande expressão. Depois da definição das classes de valores seguiu-se todo um processo, minucioso e demorado, de associar a cada arqueossítio um destes valores pré-definidos. No entanto, a elevada percentagem de valores nulos que ainda existente, irá contribuir para que este atributo venha a ser retirado da DB de trabalho.

O atributo **Solos** caracteriza o tipo de solos da área onde se posicionam os arqueossítios. A região de Trás-os-Montes Oriental tem uma estrutura geológica com diversas zonas de contacto entre **Granitos** e **Xistos** e um complexo sistema de falhas que deu origem a um tipo de solos variado [Lemos, 1993] (Figura 5.12).

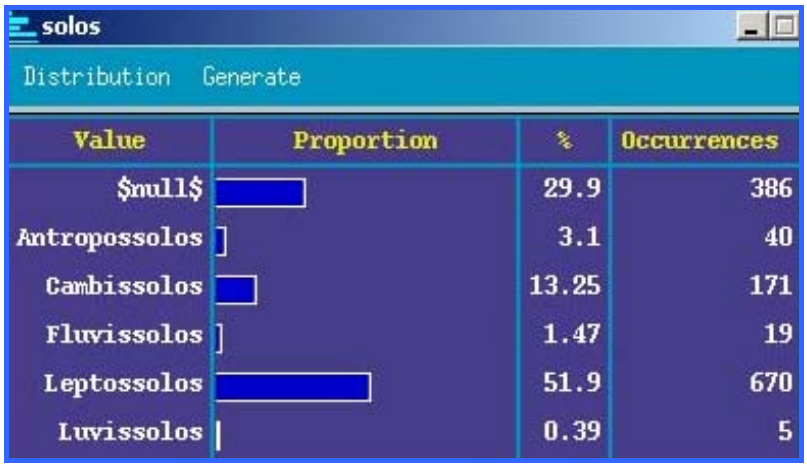


Figura 5.12 – Distribuição de valores de tipo de **Solos**

No inventário de sítios utilizado neste trabalho, a caracterização do tipo de solos foi feita utilizando uma carta de solos 1:1 000 000, usando a classificação das unidades edafológicas que lhe estava associada. No entanto, os trabalhos no âmbito da elaboração de PDM, foram utilizadas outras classificações, pelo que foi preciso normalizar este atributo.

Por outro lado, a existência de cartas de solos mais actualizadas, realizadas a uma escala de 1:100 000, fez com que as operações de normalização fossem complementadas com operações de actualização, à luz destas novas cartas de solos [Coba, 1991]. Para alguns registos não foi possível associar qualquer valor ao tipo de **Solos**, por não ter sido viável a sua localização na carta de solos.

O campo da **Cronologia** relativa contém as referências a épocas/datas associadas aos sítios arqueológicos. Neste campo da tabela encontraram-se diferentes designações para o mesmo período cronológico. Por exemplo, para o **Período Romano**, foram encontrados termos como: **Romanização, Romano** e **Tardo-romano**. Sobre estes dados foi feita uma normalização, tendo sido adoptada sempre a mesma designação, para cada classe cronológica (Figura 5.13).

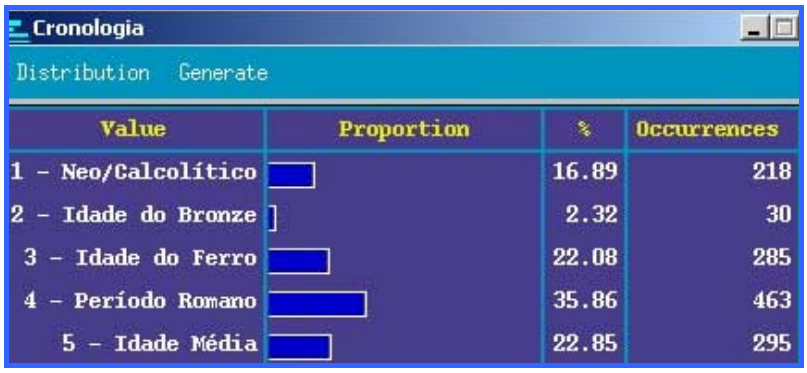


Figura 5.13 – Distribuição de valores de **Cronologia**

As classes cronológicas criadas foram definidas de acordo com as classificações apresentadas no capítulo 2 (secção 2.2), diferindo apenas na associação dos períodos **Neolítico** e **Calcolítico**, por se julgar ser o mais adequado para este trabalho.

O campo **Tipologia** resultou, como já foi referido atrás, da combinação das colunas **Estruturas**, **Materiais** e **Interpretação,** que continham informação relacionada com o tipo de arqueossítio inventariado (Figura 5.14).

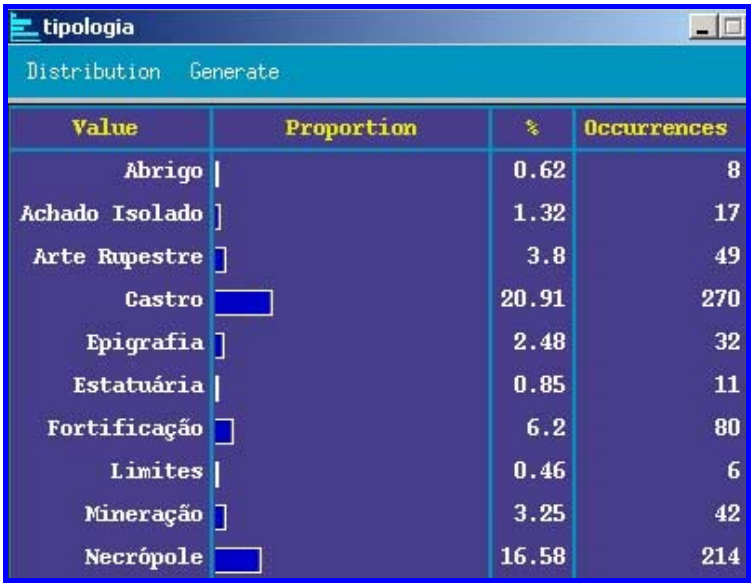


Figura 5.14 – Distribuição de valores para o atributo **Tipologia**

A normalização dos dados implicou uma uniformização de designativos, tendo-se procurado todas as equivalências de termos com as classes de **Tipologias** criadas. Por exemplo, termos como *habitat* e **Povoado**, que representam o mesmo conceito, passaram a ter a designação única de **Povoado**.

5.2.3.3. Especialização de atributos

Durante a fase de pré-processamento dos dados verificou-se que alguns registos continham múltiplos valores para um dado atributo. Esta situação foi detectada nas colunas **Tipologia** e **Cronologia,** onde alguns registos tinham associados dois ou mais tipos de sítios, ou períodos cronológicos**.** Por exemplo, para a coluna **Tipologia** é frequente encontrar valores como **Castro Romanizado.** Um **Castro** é um sítio que corresponde a um **Povoado Fortificado** da **Idade do Ferro.** Quando os **Castros** foram ocupados durante o **Período Romano,** então passaram a designar-se **Castros Romanizados**. De acordo com as classes de valores criadas para a **Tipologia**, fez-se o desdobramento de **Castro Romanizado** em dois registos, correspondentes a dois tipos de sítios arqueológicos (**Castro** e **Povoado**), ocupados durante dois períodos cronológicos diferentes (**Idade do Ferro** e **Período Romano**).

Todos os registos que continham no campo **Cronologias** várias designações do tipo **Neo/Calcolítico, Bronze, Ferro, Período Romano** ou **Idade Média,** foram desdobrados em tantos registos, quantas as diferentes cronologias que representavam.

Foram também encontrados no campo **Tipologia** designações do tipo "*vicus;* **Necrópole**", ou "tesouro monetário; habitat romano". Cada atributo deste tipo corresponde a duas classes diferentes de **Tipologia,** pelo que foi feito o respectivo desdobramento. Para os valores de "*vicus;* **Necrópole**" fez-se o desdobramento em dois registos, correspondendo um a **Povoado** e outro a **Necrópole**. Os valores de "Tesouro monetário; *habitat* romano" passaram a estar representados em dois registos distintos: **Tesouro** e **Povoado**.

5.2.3.4. Discretização de atributos

Sobre os valores numéricos associados às colunas **Altitude**, **Longitude** e **Latitude** fez-se a discretização dos dados, transformando os valores contínuos em valores discretos, para serem agrupados por classes. O objectivo desta operação é reduzir o espaço da pesquisa, limitando o número de valores distintos para um dado atributo. No final, obtêm-se regras menos complexas e mais legíveis.

O campo **Altitude** contém os valores da altitude média associados aos locais onde foram encontrados os arqueossítios (Figura 5.15). Como são valores contínuos, há

toda a vantagem em associar estes valores, para que sejam agrupados em classes, simplificando o modelo a identificar.

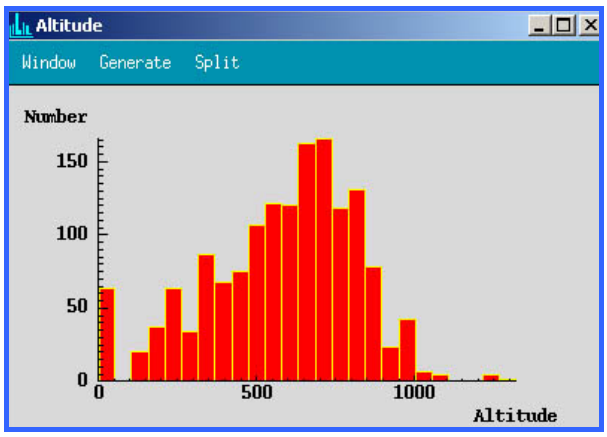


Figura 5.15 – Valores de **Altitude**

Para este atributo foram criadas classes, deduzidas do mapa geomorfológico de Trás-os-Montes Oriental [Ribeiro *et al.,* 1991], tendo-se realizado a correspondência entre os valores numéricos e as respectivas classes (Figura 5.16).

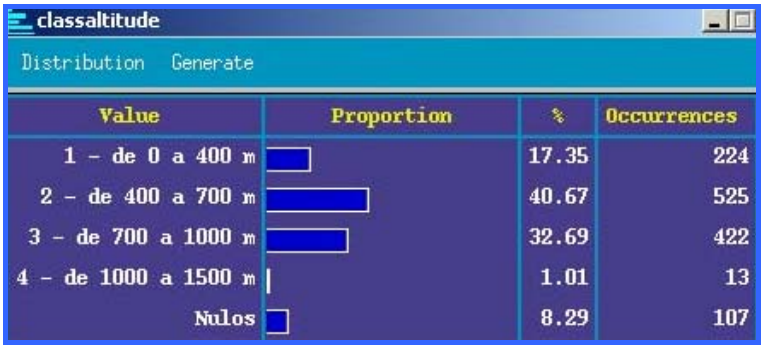


Figura 5.16 – Valores de **Altitude**, distribuídos por classes[10]

Para a **Longitude** e **Latitude** foram também atribuídos valores por classes**.** Não havendo outros critérios pré-definidos para estes valores, foram usados critérios de classificação centrados na distribuição uniforme dos dados (Figura 5.17).

[10] As distribuições de altitude são pouco homogéneas porque resultaram da criação de classes de acordo com o mapa do esboço geomorfológico de Trás-os-Montes, normalmente utilizado em inventários arqueológicos.

Figura 5.17 – Classes de valores e distribuição por classes para a **Longitude** e **Latitude**

Para a Latitude foram criadas quatro classes. A primeira (Lat 472) agrupa todos os valores de **Latitude** inferiores a 472 000, a segunda (Lat 498) agrupa valores superiores a 472 000 e inferiores a 528 000, a terceira valores entre 498 000 o 528 000 e, a quarta e última classe, agrupa valores de **Latitude** superiores a 528 000.

De igual forma se procedeu para agrupar os valores de **Longitude**, tendo-se criado quatro classes, com uma distribuição mais ou menos uniforme. A primeira classe agrupa valores inferiores a 228 000 (Long 228), a classe Long 273 agrupa valores entre os 228 000 e os 273 000, a classe Long 308, reune valores entre os 273 000 e os 308 000 e a Long 308 agrupa valores superiores a 308 000.

5.2.4. Análise de relações entre os dados

Terminadas as fases de tratamento e pré-processamento, passou-se à análise dos dados e das relações entre eles. Deve referir-se que as fases anteriores implicaram a execução de tarefas muito demoradas e recursivas. As análises aos dados, que vão ser a seguir apresentadas, fizeram parte de um processo também muito interactivo. Constantemente

se apelou ao saber da área de Arqueologia para analisar os resultados obtidos e tomar decisões sobre alterações a fazer de novo na preparação dos dados.

As decisões tomadas para formação das classes de valores criadas para o atributo **Tipologia,** por exemplo, são representativos da grande interactividade e iteractividade destas etapas. Algumas acções encetadas durante a fase de preparação dos dados foram posteriormente alteradas, em função da análise das relações entre os dados. Para o campo **Tipologia** foram criadas classes, onde inicialmente se incluía o **Castro,** que corresponde a um povoado fortificado. Fez-se, então, a duplicação dos registos com este atributo, para a **Tipologia Povoado** e **Fortificação**. Mais tarde, reviu-se esta decisão, uma vez que se perdia a associação entre os dois tipos, **Povoado** e **Fortificação**. Há sítios que são **Povoado**, outros que são **Fortificação** e outros, como os **Castro**, correspondem às duas tipologias juntas. Com o desdobramento passou a não ser visível esta associação. Assim, decidiu-se incluir na classe de valores para **Tipologia** o tipo **Castro**, distinto de **Povoado** e **Fortificação** isolados. Após a fase de análise de relações entre os dados e mesmo da aplicação de algoritmos de DM, procedeu-se então à reclassificação de alguns atributos, de acordo com as novas classes criadas.

As tarefas de análise dos dados e das relações entre eles foram apoiadas pela utilização intensiva de gráficos *Web Node,* usando a ferramenta *Clementine*. São tarefas realizadas com a colaboração do especialista da área, sendo uma mais valia a sua apresentação gráfica, de leitura fácil e intuitiva. Os *Web Nodes* permitem identificar relações interessantes entre dois ou mais atributos simbólicos, ligando-os uns aos outros por linhas, indicativas da frequência com que aparecem juntos.

As ligações são expressas graficamente, através de pontos, linhas e linhas sombreadas. As relações mais fortes são desenhadas a traço contínuo mais carregado, passando a linha tracejada quando estamos perante relações fracas. Dados não ligados indicam que não foi identificada qualquer relação entre eles.

Nesta fase procede-se a uma avaliação da qualidade e representatividade dos dados com que se está a trabalhar, desenhando-se já algumas hipóteses acerca das relações entre eles.

A seguir irá ser apresentada a tabela de dados a analisar, bem como as relações detectadas entre os atributos **Tipologia** e **Coordenadas geográficas, Tipologia** e **Altitude, Tipologia, Geomorfologia_mac, Geomorfologia_mic** e **Topografia, Tipologia** e **Hierarquia_hidrográfica,** bem como **Tipologia** e **Cronologia.**

5.2.4.1. Tabela de dados

A Tabela 5.4 apresenta os atributos resultantes da selecção, tratamento e pré-processamento, operadas sobre o conjunto inicial de dados.

Referiu-se já que este trabalho teve início com um conjunto alargado de dados, correspondendo aos sítios inventariados para a Região de Trás-os-Montes. Durante a preparação dos dados, foram desencadeadas várias operações de limpeza e validação, bem como o preenchimento e actualização de valores. No final, constatou-se que os dados, correspondentes à área de Trás-os-Montes Ocidental, tinham uma elevada taxa de valores omissos e apresentavam maior dificuldade no preenchimento e validação dos dados. Assim, pelo exposto, após terem sido sucessivamente retirados dados, a BD a utilizar para identificação do modelo ***servator***, ficou restrita à região de Trás-os-Montes Oriental, passando a denominar-se de TMO.

Atributos de TMO
Latitude
Longitude
Altitude
Geomorfologia_mac
Geomorfologia_mic
Topografia
Hierarquia_hidrográfica
Solos
Cronologia
Tipologia

Tabela 5.4 – Tabela de dados TMO

Sobre este conjunto de dados TMO, compilados numa tabela com 10 atributos e 1291 registos, irão ser aplicadas as técnicas de DM, para identificação do modelo

preditivo de sítios arqueológicos. Será feita de seguida a análise de relações entre os dados da tabela TMO.

5.2.4.2. Tipologia e Coordenadas geográficas

Para encontrar relações entre os atributos **Tipologia** e as coordenadas geográficas, **Latitude** e **Longitude,** elaboraram-se os *Web Nodes* representados na Figura 5.18.

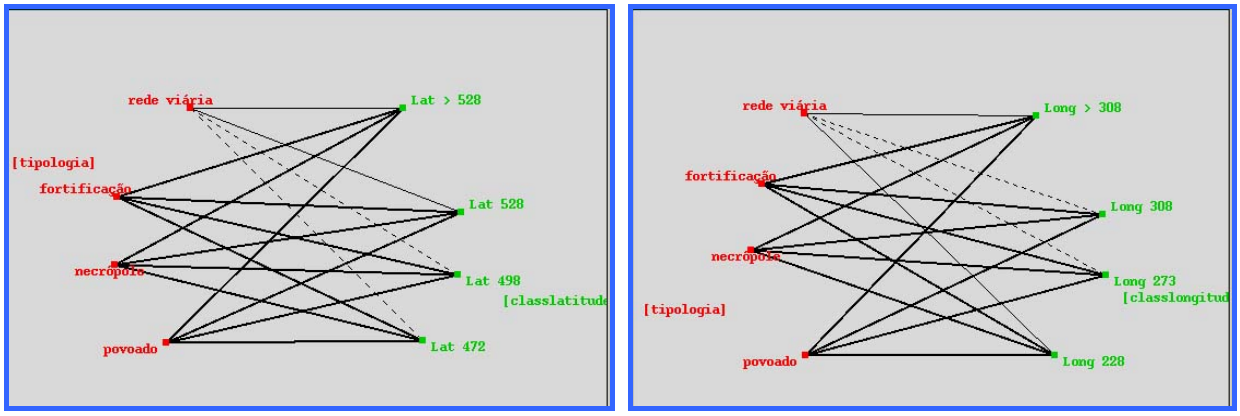


Figura 5.18 – *Web Node* que relaciona **Tipologia** com a **Latitude** e **Longitude**

A região de estudo para este trabalho, Trás-os-Montes Oriental, como já foi referido, está confrontada pelos valores de **Latitude** de 135 000 até 370 000 e de **Longitude** entre os 450 000 e os 575 000, para o sistema de coordenadas Gauss. Depois de agrupados os valores por classes, elaboraram-se os *Web Nodes*, para procurar as relações entre estes e o tipo de sítios arqueológicos. Cruzando os resultados dos gráficos da **Latitude** com a **Longitude** o resultado é bastante interessante, uma vez que mostra uma distribuição idêntica dos tipos **Povoado**, **Necrópole** e **Fortificação**, por todo o território. No entanto, para o caso da **Rede Viária** encontraram-se, claramente, umas manchas onde a **Rede Viária** é mais densa. É nos corredores definidos pela Latitude superior à classe 498 (498 000) e Longitude até 228 e superior a 308 (> 308 000), onde a rede de caminhos é mais intensa.

É muito interessante comparar este resultado com o traçado das vias romanas principais (Figura 5.19), onde desde logo se evidencia um grande eixo, que atravessa a região de Trás-os-Montes Oriental e que ligava duas importantes urbes. Este eixo, denominado de Via XVII, ligava *Bracara Augusta* e *Asturica,* por *Aquae Flaviae*,

enquadrando-se nos parâmetros de valores de **Latitude** e **Longitude** evidenciados pelas relações mais fortes dos ***Web Nodes***.

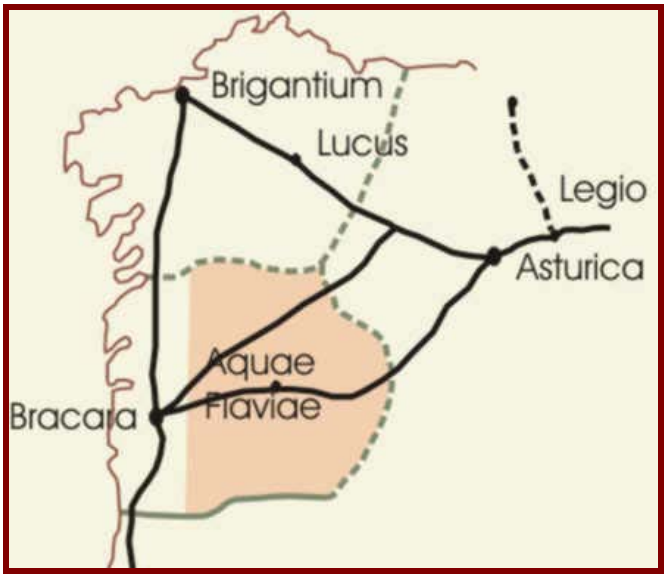


Figura 5.19 – **Rede Viária** romana do Noroeste da Península Ibérica (adaptado de [Lemos, 2002])

Esta via, já com raízes na época da **Idade do Ferro,** dirigia-se para a Meseta Norte no sentido transversal e constituía um corredor natural de circulação do Norte Peninsular [Lemos, 2002].

5.2.4.3. Tipologia e Altitude

A região de Trás-os-Montes Oriental teve origem numa superfície aplanada que sofreu enrugamentos e falhas. Destes movimentos resultou a actual configuração, caracterizada pelos planaltos, vales e serras, de altitudes médias bastante díspares. Para analisar as relações existentes entre as classes de altitude e os tipos de sítios arqueológicos elaborou-se o gráfico apresentado na Figura 5.20.

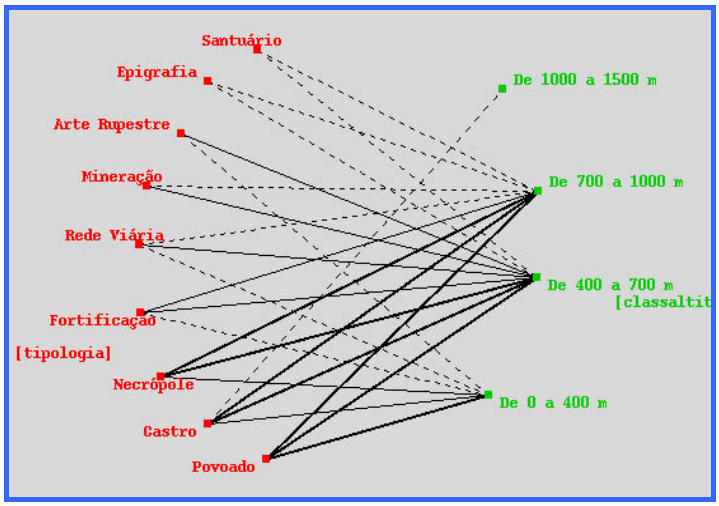


Figura 5.20 – *Web Node* que relaciona as variáveis **Tipologia** e **Altitude**

O *Web Node* acima mostra que, apesar de haver grandes variações de **Altitude** numa região relativamente pequena, quase todos os patamares sofreram uma ocupação no passado. Pode também concluir-se que as regiões de patamares mais elevados seriam menos atractivas para localização dos sítios do tipo **Povoado**, evidenciando-se apenas uma relação com o tipo **Castro.** Os **Castros,** eram povoados fortificados, o que revela que os aspectos defensivos eram muito importantes. Os locais de maior altitude oferecem, normalmente, boas condições de visibilidade e boas características defensivas.

A distribuição evidenciada para os outros tipos de sítios, tais como os **Santuários**, **Epigrafia** e **Mineração,** mostra que aparecem com maior incidência nas plataformas dos 400 aos 1000 m, enquanto que a **Arte Rupestre** aparece mais em zonas de **Altitude** abaixo dos 700 m. Estes indicadores confirmam o conhecimento existente, por exemplo, em relação à **Arte rupestre,** encaixada no fundo dos vales, fora das zonas de cotas mais elevadas.

A **Rede Viária** desenvolve-se em patamares de cotas intermédias, evitando grandes desníveis de altitude. Para esta configuração contribuiu, certamente, o bom conhecimento do terreno, por parte dos engenheiros romanos, que numa região de numerosas alterações súbitas de altitude [Ribeiro *et al*., 1991], conseguem orientar as vias, evitando as cotas muito elevadas e grandes variações de altitude.

As diferentes relações existentes, entre os *habitats* e a **Rede Viária** com a **Altitude,** revelam que a localização de cada um destes tipos obedecia a estratégias diferenciadas. Pensa-se que a **Rede Viária** não era definida para unir povoados, mas orientada para estabelecer a comunicação entre os maiores centros habitacionais e políticos. O traçado articulava as estratégias definidas de acordo com o estudo prévio da geomorfologia dos terrenos [Botica *et al.*, 2003a].

Para tentar perceber se esta estratégia terá sido a mesma durante períodos cronológicos distintos, elaborou-se um novo gráfico que relaciona as **Cronologias** com a **Altitude** (Figura 5.21).

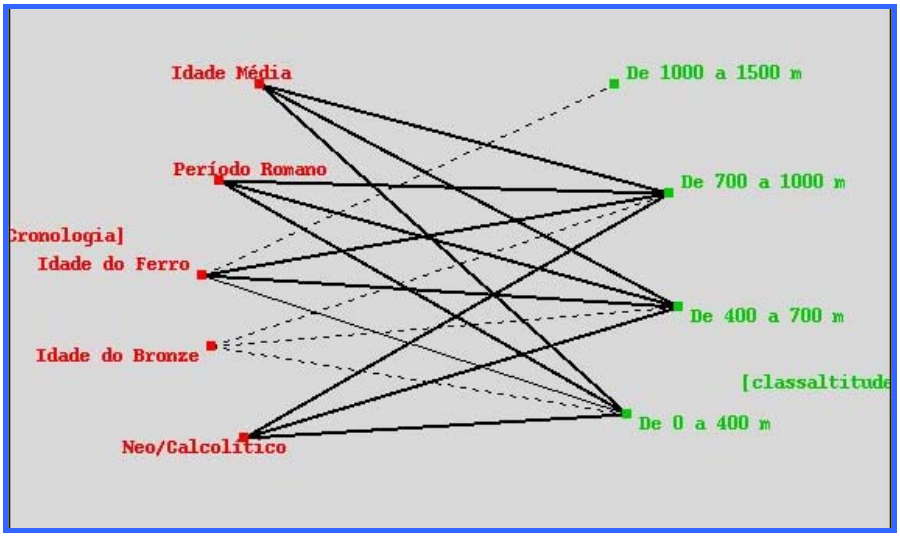


Figura 5.21 – *Web Node* que relaciona a **Cronologia** com os valores de **Altitude**

Pode agora verificar-se que durante o período da **Idade do Ferro** a ocupação do território de Trás-os-Montes Oriental era mais abrangente. Comparando as estratégias de povoamento deste período com o **Período Romano**, por exemplo, constata-se que obedeciam a princípios distintos.

De acordo com o conhecimento arqueológico existente, na **Idade do Ferro,** havia um grande equilíbrio com o contexto ambiental. Os recursos naturais eram aproveitados e ditavam a estratégia de povoamento a adoptar. Os pontos mais altos do território eram também ocupados, por oferecerem boas condições de visibilidade e de defesa naturais. Estes aspectos eram muito importantes e ponderados na escolha dos locais de assentamento. Já no **Período Romano,** a estratégia de povoamento foi

orientada não apenas de acordo com os recursos naturais, mas também, influenciada por uma nova rede de caminhos e uma nova economia [Botica *et al.*, 2003b]. Assim, parte dos povoados que estavam posicionados em patamares mais elevados foram abandonados, por não se enquadrarem nesta nova política de ocupação do território.

5.2.4.4. Tipologia, Geomorfologia_mac, Geomorfologia_mic e Topografia

A **Topografia** e a macro e micro geomorfologia (**Geomorfologia_mic** e **Geomorfologia_mac**) são o cenário natural onde se inserem os povoados e as tipologias resultantes das actividades humanas. É dentro deste contexto que as escolhas dos locais são feitas, em consonância com as estratégias de cada época.

Algumas decisões iniciais de agregar informação, para simplificação do modelo, foram revistas após esta fase do processo. Foi o caso dos atributos **Geomorfologia_mac**, **Geomorfologia_mic** e **Topografia,** onde inicialmente se pensou utilizar apenas um campo para a **Geomorfologia_mic**. Tal decisão partia do princípio de que este atributo caracterizava o terreno, de modo detalhado, e que seria suficiente para a identificação do modelo. No entanto, após algumas simulações verificou-se que estes dados eram muito importantes para o processo de prever a localização dos sítios. Teria que ser estudada a hipótese de utilizar também os atributos **Geomorfologia_mac** e **Topografia**. Voltou-se atrás no processo e foram considerados, em separado, os três atributos de caracterização geográfica do sítio.

O gráfico da Figura 5.22, relaciona a **Topografia** com os valores de **Tipologia**, e mostra que há uma utilização diferenciada dos terrenos da região, em função da **Topografia**.

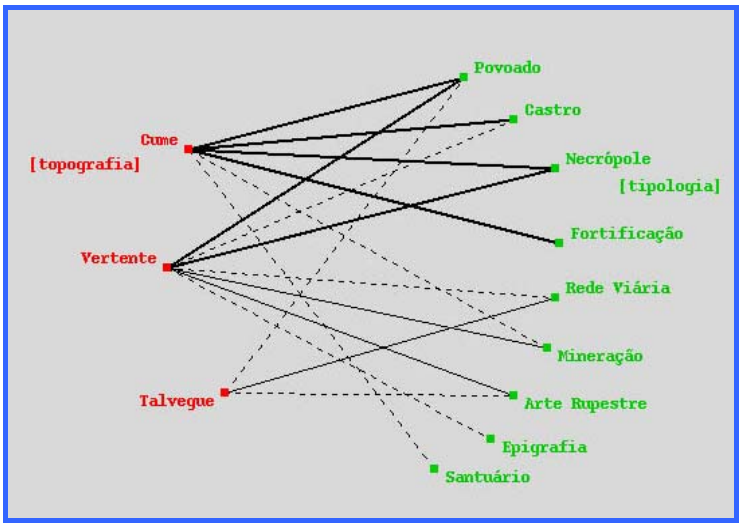


Figura 5.22 – *Web Node* que relaciona a **Topografia** com a **Tipologia** dos sítios

O tipo **Povoado** posiciona-se preferencialmente na **Topografia** do tipo **Cume** e **Vertente** havendo, no entanto, alguns povoamentos em zonas de **Talvegue**, ou seja junto aos leitos dos rios.

Tal característica poderá causar alguma perplexidade, uma vez que os rios constituem uma importante via natural de comunicação. De acordo com o conhecimento arqueológico existente, este deficiente relacionamento pode explicar-se pela circunstância dos rios, nomeadamente o **Rio Douro**, terem vertentes muito escarpadas e de difícil acesso. Esta característica pode ter impedido uma ocupação mais densa destes locais. Por outro lado, poderá também ter havido uma prospecção pouco intensiva desta área [Silva, 1995], dadas as dificuldades de acesso impostas pela sua configuração natural, aliada a circunstâncias arqueológicas várias [Botica *et al.*, 2003a].

O tipo **Fortificação** não apresenta qualquer relação com a **Topografia Talvegue** ou **Vertente**, embora existam aí **Povoados**. Esta ausência de relacionamento sistemático com os **Povoados** poderá confirmar uma corrente de pensamento, dentro da Arqueologia, de que as fortificações não tinham apenas um papel defensivo, pois nesse caso apareceriam associadas a **Povoado.** As **Fortificações** teriam assim um carácter multifuncional [Botica *et al.*, 2003c], podendo servir para:

- Controle e delimitação do território envolvente;
- Estratégias de defesa;

- Aspecto simbólico de valor arquitectónico.

A Figura 5.23 define as relações encontradas pelos *Web Nodes*, entre os atributos de **Tipologia** e as características da Geomorfologia macro e micro.

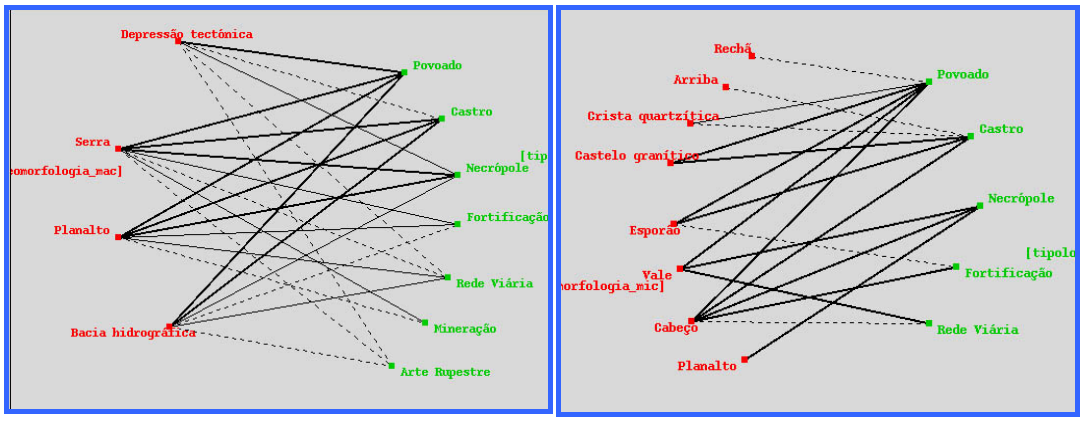


Figura 5.23 – *Web Nodes* que relacionam a **Tipologia** com a **Geomorfologia**

Analisando os gráficos da Figura 5.23, pode apenas concluir-se que as comunidades que povoaram a região de Trás-os-Montes Oriental dispunham de um conhecimento detalhado da geomorfologia da região. Esta conclusão decorre dos **Povoados** terem sido implantados em todas as classes de **Geomorfologia_mac**, bem como na maioria das classes de **Geomorfologia_mic**. Regista-se um aproveitamento sistemático da diversidade territorial. Os **Povoados** instalaram-se em **Serra**, **Depressão tectónica**, **Planalto** ou **Bacia hidrográfica**.

5.2.4.5. Tipologia e Hierarquia_hidrográfica

O gráfico da Figura 5.24 relaciona os tipos **Castro**, **Povoado**, **Fortificação** e **Rede Viária** com as classes de valores da **Hierarquia_hidrográfica**. Esta operação reforça o que foi já evidenciado nos gráficos anteriores. Os tipos **Castro**, **Povoado** e **Fortificação** encontram-se disseminados por todo o território, com uma menor densidade nas zonas mais afastadas dos cursos de água principais. A **Rede Viária** está instalada com maior incidência nos patamares 2, 3 e 4 da **Hierarquia_hidrográfica**, o que corresponde aos níveis intermédios de **Altitude**, evitando as grandes variações de cotas.

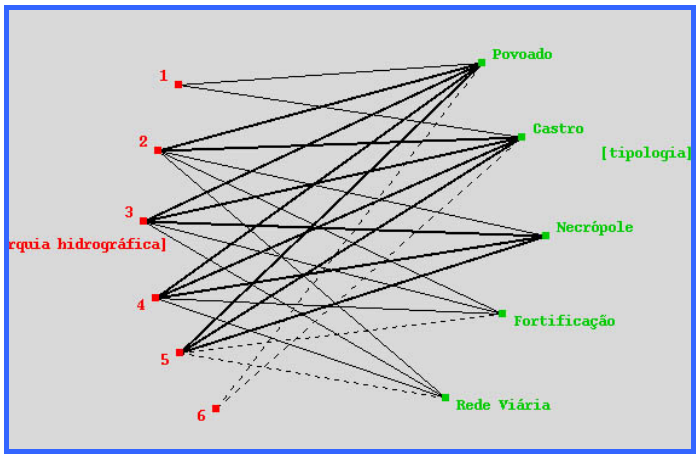


Figura 5.24 – *Web Node* que relaciona a **Tipologia** com a **Hierarquia_hidrográfica**

5.2.4.6. Tipologia e Solos

Os resultados iniciais da análise das relações entre os dos **Solos** e a **Tipologia** eram muito ambíguos, talvez devido à elevada taxa de valores omissos. Decidiu-se estudar uma maneira de poder obter melhores resultados, dado ser uma questão bastante debatida na Arqueologia. Todos os valores alocados a estes dois atributos foram revistos, agrupados de novo em classes, procedendo-se a uma nova recolha detalhada de dados, com base nas cartas de solos da região, bem como os catálogos dos sítios arqueológicos inventariados, com referências a estes atributos [Lemos, 1993] [Cruz, 2000].

A capacidade do uso de solos foi desenvolvida ao longo dos tempos. A sua utilização mais intensiva progrediu à medida que eram descobertas novas técnicas agrícolas. De acordo com esta perspectiva desenhou-se um gráfico onde pudessem estar visíveis as relações entre o tipo de solos e os sítios do tipo *habitat*, isto é, os **Abrigo**, **Castro** ou **Povoado** (Figura 5.25). Apenas foram seleccionados estes tipos de sítios por serem as Tipologias directamente ligadas à actividades de uso de solos.

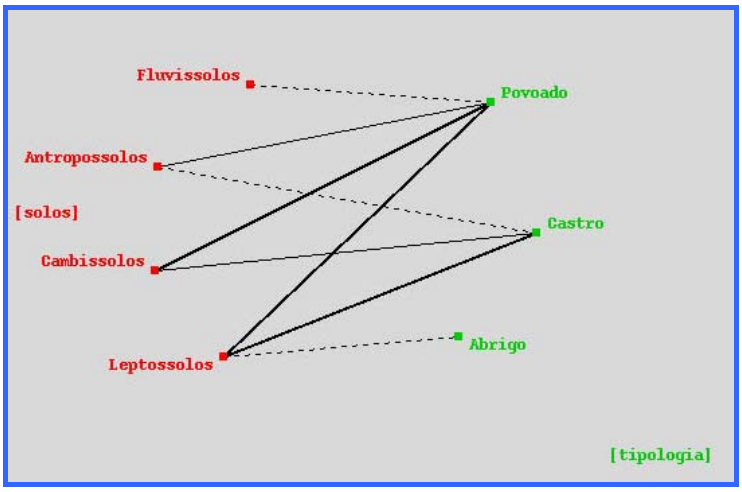


Figura 5.25 – *Web node* que relaciona os sítios do tipo *habitat* com o tipo de **Solos**

Pode verificar-se que os sítios do tipo **Castro** e **Povoado** utilizavam intensamente os solos designados por **Leptossolos** e menos os **Cambissolos**, sendo fraca a relação com os **Antropossolos** e **Fluvissolos**. Os sítios do tipo **Abrigo** apenas se relacionavam com o tipo **Leptossolos**.

Estes relacionamentos têm que ser interpretados à luz das capacidades naturais dos solos, mas também, do ponto de vista das tecnologias que eram utilizadas para trabalhar os solos. Assim, pode ver-se que os **Leptossolos,** sendo solos com boas aptidões agrícolas, eram os mais procurados**.** No caso do tipo **Abrigo** eram até os únicos solos utilizados, uma vez que esse período antecedeu o aparecimento de técnicas e tecnologias mais sofisticadas.

Os **Solos** do tipo **Leptossolos** têm boas aptidões agrícolas, para a cultura de cereais, oliveira e vinha. Já os solos do tipo **Cambissolos** têm características boas para o aproveitamento florestal, podendo ser aproveitados para a agricultura, quando usadas compensações químicas para colmatar algumas carências de nutrientes.

As tecnologias mais evoluídas existentes no **Período Romano,** permitiram um aproveitamento maior dos **Solos,** mesmo para os mais pobres. Por esse motivo é mais forte a relação do **Período Romano** com os **Solos** do tipo **Cambissolos**, do que a relação existente entre o mesmo tipo de **Solos** e a **Idade do Ferro.** Esta conclusão é reforçada quando se analisa a relação entre o

Povoado e os **Solos** do tipo **Antropossolos**. Estes **Solos** sofreram alterações profundas resultantes da actividade humana. São **Solos** utilizados na agricultura porque foram feitos enchimentos, cortes, adições de matéria orgânica e rega continuada. Estas alterações foram mais fortes no **Período Romano,** do que na **Idade do Ferro,** onde as tecnologias eram menos evoluídas e não permitiam um aproveitamento tão intensivo.

Embora o gráfico anterior apenas representasse as relações entre o tipo de **Solos** e o tipo de sítio, mostra também implicitamente algumas as relações interessantes entre o aproveitamento dos **Solos** e os períodos cronológicos. Tal só é possível porque o tipo de sítio **Castro** é específico da **Idade do Ferro,** enquanto que o **Povoado** aparece nos outros períodos cronológicos.

Para melhor analisar as relações entre o tipo de **Solos** e a **Cronologia,** traçou-se o gráfico *Web Node,* apresentado na Figura 5.26.

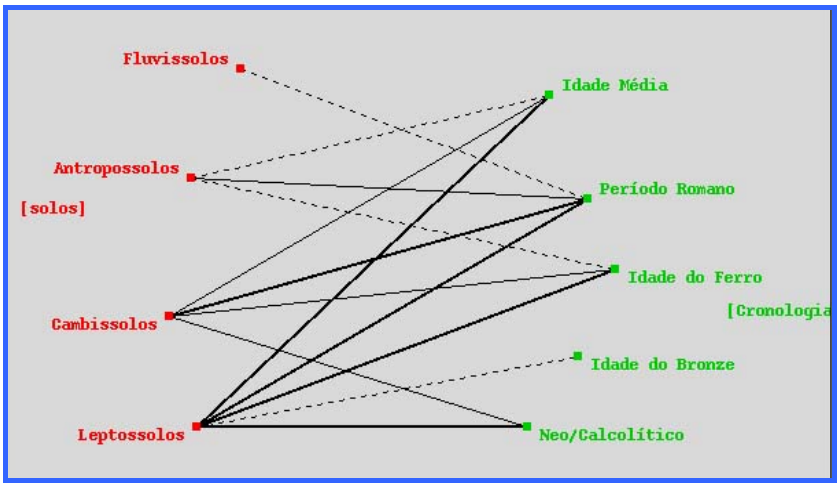


Figura 5.26 – *Web Node* que relaciona a **Cronologia** com o tipo de **Solos**

A análise do gráfico evidencia uma utilização mais intensiva dos **Solos**, de vários tipos, durante o **Período Romano**, sendo menos abrangente essa utilização à medida que se vai recuando no tempo. Épocas anteriores a este período tinham uma menor actividade agrícola, estando dotados de menos instrumentos e sendo mais rudimentares as técnicas utilizadas.

5.2.4.7. Tipologia e Cronologia

O relacionamento dos principais tipos de sítios com a **Cronologia** (Figura 5.27) evidencia algumas relações interessantes entre os dados.

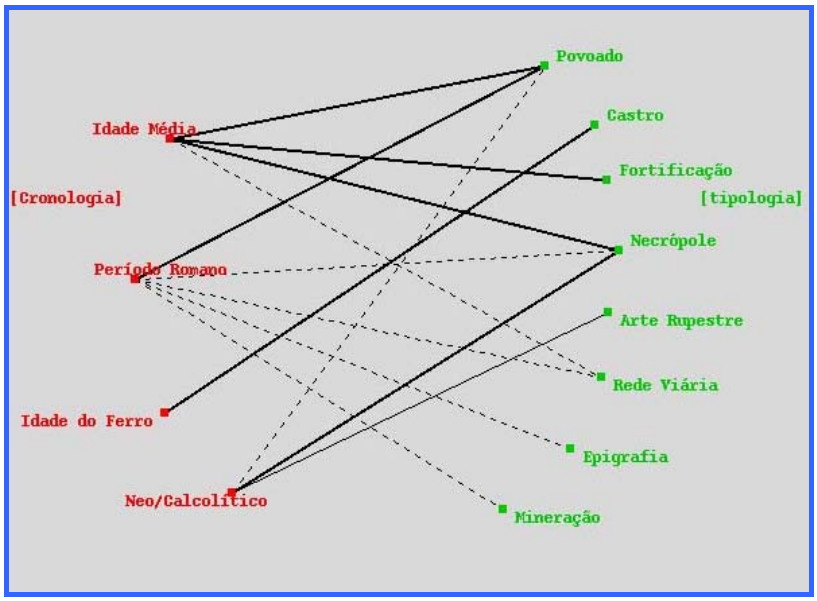


Figura 5.27 – *Web Node* que relaciona a **Tipologia** com a **Cronologia**

Este *Web Node* (Figura 5.27) reflecte, por um lado, os modelos de *habitat* de cada período cronológico e, por outro, o grau de conhecimento arqueológico.

Relativamente ao tipo **Fortificação**, é uma estrutura que começa a ter grande expressão na **Idade do Ferro**, rodeando os **Povoados** e obtendo a denominação de **Castros**. A sua importância e utilização foram reduzidas no **Período Romano**, voltando a assumir um papel especial na **Idade Média**. As razões foram já apontadas quando se analisou o gráfico que relaciona a **Tipologia** com a **Altitude**. Para a **Idade do Ferro** e período Medieval a estratégia de povoamento passava pela ocupação de locais de boas características de defesa natural e boa visibilidade. O **Período Romano** possuía uma estratégia diferenciada, onde a defesa não era tão importante para a localização dos locais de povoamento, mas antes a acessibilidade, a proximidade à **Rede Viária** e aos corredores naturais de circulação e a proximidade a **Solos** favoráveis ao aproveitamento agrícola.

A **Rede Viária** tem a sua máxima expressão no **Período Romano**, estando inserida numa das prioridades ao nível das estruturas a construir. A acessibilidade é para

o Império Romano um factor muito importante, sendo inúmeros os testemunhos que chegaram até aos nossos dias. Os caminhos, as pontes e os marcos miliários formam um conjunto associado à **Rede Viária** que marcou este período cronológico. A **Epigrafia** está também muito ligada ao **Período Romano**, aparecendo muitas vezes associada a inscrições votivas ou funerárias.

Já a **Arte Rupestre** foi uma actividade que teve a sua máxima expressão no período **Neo/Calcolítico**.

A mineração começa a aparecer na **Idade do Ferro**, mas ainda como uma actividade muito residual, pelo que não aparece qualquer relação no gráfico da Figura 5.27. É no **Período Romano** que esta actividade se desenvolve e adquire um peso significativo.

5.3. Conclusão

Este capítulo, foi dedicado às tarefas de selecção, tratamento e pré-processamento dos dados, seguidas de uma análise dos mesmos. Foram tarefas demoradas, desenvolvidas sempre em estreita colaboração com o especialista da área e sujeitas a constantes alterações e reformulações.

As análises dos dados serviram por um lado, para apoiar as decisões tomadas no tratamento e pré-processamento dos dados e, por outro, evidenciaram relações interessantes entre vários atributos, em consonância com o conhecimento existente na área.

No final, considerou-se que os dados estavam preparados para se proceder à aplicação de algoritmos de DM. A descrição deste processo é apresentada no capítulo seguinte.

Capítulo 6

servator – identificação do modelo

Para a identificação do modelo ***servator*** serão utilizados os princípios associados à DCBD, tendo como objectivo extrair conhecimento a partir de BD.

O capítulo anterior foi dedicado à preparação dos dados que agora serão submetidos a algoritmos de DM, para proceder à identificação de um modelo preditivo que apoie a actividade arqueológica.

Neste capítulo descreve-se o processo de identificação do modelo ***servator,*** sendo de destacar a aplicação de técnicas de modelação avançada aos dados, a interpretação e avaliação dos padrões encontrados, bem como as dificuldades encontradas durante todo o processo.

6.1. Aplicação de algoritmos de DM

Após terem sido realizadas as tarefas de selecção, tratamento e pré-processamento dos dados, procede-se agora à aplicação de técnicas de modelação avançada, de modo a identificar o modelo preditivo de apoio à prospecção arqueológica, designado ***servator***.

O conjunto de dados TMO para a modelação contempla 1291 registos, com 10 atributos cada. Estes registos foram subdivididos em dois subconjuntos, aos quais se deu o nome de Treino e Testes.

O modelo ***servator*** é treinado utilizando dados pré-classificados do subconjunto de Treino, usando-se o subconjunto Testes para refinar o modelo. Com esta operação pretende-se confirmar a generalização do modelo, avaliando-se os resultados obtidos, para que as suas previsões não sejam uma memorização de resultados do conjunto de Treino [Berry e Linoff, 2000].

Pode dividir-se a BD em três conjuntos, usando-se o terceiro conjunto de dados, normalmente chamado de Avaliação**,** para estimar o desempenho do modelo. No

entanto, dado o reduzido número de registos disponível para identificação do ***servator,*** irão apenas ser utilizados dois conjuntos de dados. O conjunto de Treino**,** usado para a identificação de padrões e, o conjunto de Testes**,** para avaliar a generalização do modelo e o seu desempenho.

Na primeira fase da aplicação de técnicas de DM ao conjunto de dados, preparados nas fases anteriores, subdividem-se os dados nos conjuntos de Treino e Testes. A figura 6.1 apresenta a *stream* que ilustra esta operação, utilizando a ferramenta *Clementine*[11].

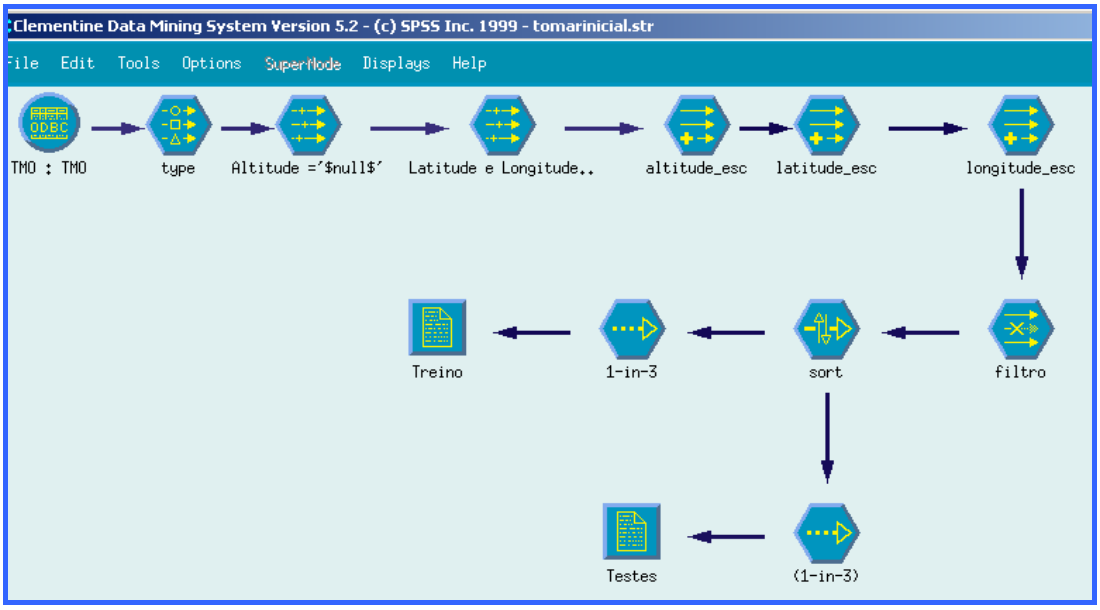


Figura 6.1 – ***servator*** - conjunto de dados de Treino e de Testes

Para a divisão dos conjuntos utilizou-se a regra de 1 em 3, isto é, por cada registo colocado no conjunto de Treino, são reservados 2 registos para conjunto de Testes. Considerando os 1291 registos da BD, reservaram-se 430 para Treino e 861 para o conjunto de Testes.

Dado o carácter previsional do modelo que se pretende identificar, decidiu-se aplicar ao conjunto de Treino um algoritmo de indução de regras e, posteriormente, um algoritmo de redes neuronais, com o objectivo de melhorar os resultados e aumentar a confiança nos modelos.

[11] O *Clementine* é uma ferramenta de programação visual, onde todas as operações realizadas sobre os dados são representadas por um símbolo gráfico ou nodo. As setas que ligam os nodos indicam o fluxo dos dados [Clementine, 1998].

Da aplicação de algoritmos de indução de árvores de decisão resultam regras que descrevem, numa linguagem natural e de fácil leitura, a forma como são tomadas decisões. Para além da compreensão, quase intuitiva das regras apresentadas, este algoritmo tem a vantagem de não considerar os atributos que não são importantes para a tomada de decisões. Assim, só serão submetidos ao algoritmo de redes neuronais os atributos utilizados pela árvore de decisão.

O algoritmo de indução de árvores de decisão C5.0 foi usado para produzir regras, cuja informação pode levar à construção de conhecimento arqueológico, bem como à identificação dos atributos mais relevantes para a previsão do tipo de arqueossítio**.** Em função dessa análise, serão submetidos, a um algoritmo de redes neuronais, os atributos que se revelarem importantes na tomada de decisão, retirando-se todos os outros das entradas da rede.

6.1.1. Algoritmo de Indução de Regras

O algoritmo de indução de árvores de decisão C5.0, quando aplicado aos dados, encadeia uma série de regras, que conduzem a uma determinada classe ou valor. Na vista parcial, apresentada na Figura 6.2, podem observar-se algumas regras obtidas. Por exemplo, para a **Cronologia Período Romano** se a **Topografia** for um **Cume**, então possivelmente irá encontrar-se nesse local um **Povoado**. No entanto, se a **Topografia** for um **Talvegue,** provavelmente haverá uma **Rede Viária.** Embora estas regras possam parecer interessantes para o **Período Romano** são, no entanto, pouco específicas quando se reportam à **Idade do Ferro**.

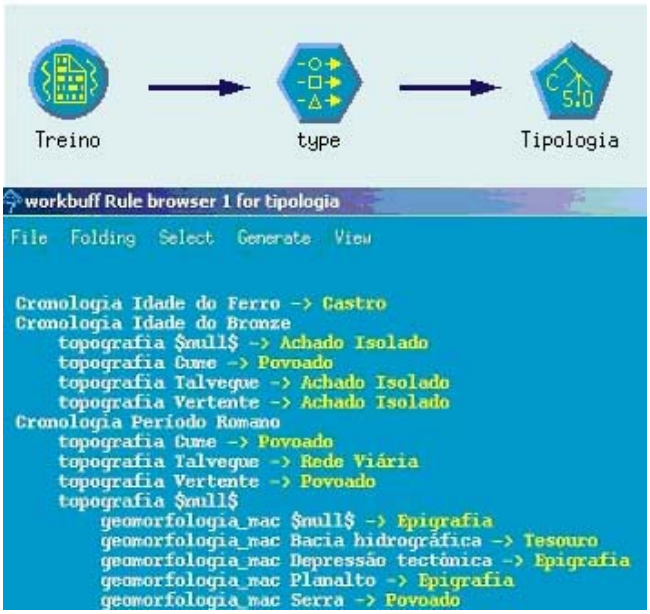


Figura 6.2 – Aplicação de algoritmo de indução de árvores de decisão

Para a **Cronologia Idade do Ferro** o modelo aponta apenas para um tipo de arqueossítio, embora nos dados existam outros tipos de sítios arqueológicos associados a este período cronológico. A diversidade de sítios inventariados não se reflecte no modelo identificado, indicando este a localização indiferenciada de apenas um tipo – o **Castro**. O apoio à localização de arqueossítios para a **Idade do Ferro,** a partir do modelo identificado, não se mostraria útil, uma vez que para o tipo **Castro,** não há qualquer indicador que possa restringir no terreno as áreas mais prováveis de localização. Para além disso, é omisso quanto à localização de outros tipos de arqueossítios.

A pouca especificidade das regras de decisão apresentadas tornam o modelo identificado pouco esclarecedor e, portanto, pouco útil no apoio à prospecção de sítios arqueológicos.

Para melhorar o modelo, experimentaram-se outras divisões dos dados, nomeadamente usando uma proporção de 50% e ainda de 60%, para o conjunto de Treino e Testes. No entanto, os modelos identificados, a partir destes valores, não apresentavam maior especificidade nas regras elaboradas. As regras só passaram a ter um nível de detalhe maior quando foram construídas com base no conjunto total dos dados (TMO).

Concluiu-se que o número de registos da BD teria que ser aumentado. A primeira hipótese que se colocou foi tentar encontrar mais inventários de sítios da região de Trás-os-Montes Oriental. No entanto, essa solução revelou-se inviável, uma vez que os inventários disponíveis, mais completos e com rigor científico, tinham já sido incorporados na BD ([Lemos, 1993] e [Cruz, 2000]).

Decidiu-se então proceder ao aumento do número de registos da BD fazendo uma clonagem balanceada dos existentes (*Boost*), com base no princípio de que alguns algoritmos de DM, nomeadamente as árvores de decisão, apresentam melhores desempenhos com registos balanceados [SPSS, 1999]. A opção de *Reduce* não foi usada, uma vez que o problema residia no reduzido número de registos. Mesmo assim, fizeram-se alguns testes que confirmaram esta decisão, mostrando que a especificidade das regras não ser melhorada.

Deste modo, o número de registos foi ampliado, usando a opção de *Generate - Balance Node (Boost)*, disponibilizada pelo *Clementine,* para operações sobre os dados.

A escolha do atributo sobre o qual irá recair o balanceamento dos dados foi feita com base na avaliação dos resultados experimentais obtidos, quer ao nível da análise das regras de decisão, por aplicação do algoritmo C5.0 aos dados, quer com base na análise do desempenho dos modelos obtidos.

Os atributos utilizados para o balanceamento foram a **Tipologia**, **Geomorfologia_mac** e **Geomorfologia_mic, Topografia** e **Cronologia,** tendo-se aplicado um algoritmo de indução de regras a cada um dos conjuntos de dados resultantes.

Apresentam-se, de seguida, os resultados obtidos para cada um dos tipos de balanceamento realizado.

6.1.1.1. Tipologia

O balanceamento por **Tipologia,** realizado através da operação *Generate – Balance Node (boost)* atribuiu por defeito os factores apresentados na Tabela 6.1, para cada tipo de arqueossítio.

Tipologia	Factor de balanceamento
Abrigo	57.4
Achado isolado	27
Arte Rupestre	9.3
Castro	1.7
Epigrafia	14.3
Estatuária	41.7
Fortificação	5.7
Limites	76.5
Mineração	10.9
Necrópole	2.1
Povoado	1
Rede Viária	6.8
Santuário	18.3
Tesouro	41.7

Tabela 6.1 – Factores de balanceamento gerados por **Tipologia**

Aplicando estes factores ao conjunto de dados TMO, obtiveram-se 2144 registos para o conjunto de Treino e 4273 para o conjunto de Testes.

As regras produzidas são apresentadas no Anexo I e representam já um conjunto estruturado e normalizado de regras, cuja informação está de acordo com o pensamento arqueológico existente. Analisando todas as regras identificadas, a partir do balanceamento feito em função dos vários atributos atrás referidos, verificou-se que estas eram mais detalhadas, apontando os indicadores geográficos como referência para localização de arqueossítios.

Fez-se, também, uma avaliação do desempenho do modelo (**Tipologia**), identificado pelo algoritmo C5.0, quando aplicado ao conjunto de dados de Treino (Figura 6.3), para decidir qual o atributo a escolher para a operação de *boost*.

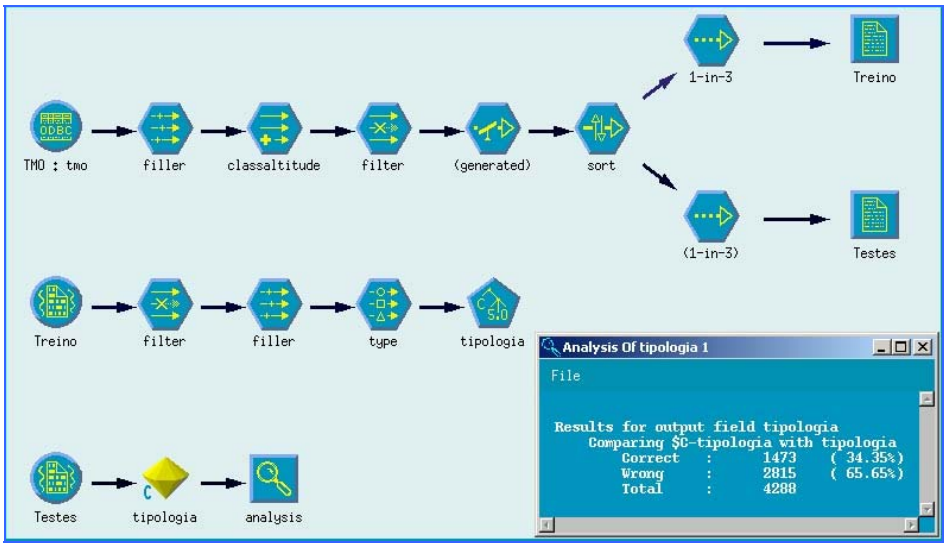


Figura 6.3 – Análise de desempenho do modelo (**Tipologia)**

A avaliação do desempenho do modelo construído é no entanto bastante baixa, apresentando apenas cerca de 34% de valores correctos na previsão do tipo de arqueossítio.

Para avaliar se a opção de *boost* por **Tipologia** é a mais indicada, será repetido o processo de *boost* usando outros atributos, tais como, a **Geomorfologia_mic, Geomorfologia_mac, Topografia** e **Cronologia.** O conjunto de dados resultante será subdividido em dois (Treino e Testes). Ao conjunto de Treino irá ser aplicado o algoritmo C5.0 e as regras produzidas analisadas pelo especialista em Arqueologia. Esta avaliação conjugada com a avaliação automática de desempenho, aplicada ao conjunto de Testes, será determinante para escolher o atributo sobre o qual irá recair a opção de *boost*, para aumentar o número de registos e melhorar a performance do modelo.

6.1.1.2. Geomorfologia_mic

O balanceamento pelo atributo **Geomorfologia_mic** atribuiu, por defeito, os factores apresentados na Tabela 6.2, para cada tipo de arqueossítio.

Geomorfologia_mic	Factor de balanceamento
null	10.8
Arriba	27.7
Cabeço	1
Castelo	4.7
Crista	5.9
Esporão	1.7
Planalto	5.1
Rechã	15.8
Vale	1.7

Tabela 6.2 – Factores de balanceamento gerados para **Geomorfologia_mic**

Não havendo qualquer interesse em aumentar o número de registos com valores **Geomorfologia_mic** nulos, foi alterado o respectivo factor de balanceamento de 10.8 para 0.

Aplicando os factores ao conjunto de dados TMO, obtiveram-se 1789 registos para o conjunto de Treino e 2390 para o conjunto de Testes. A avaliação do desempenho do modelo **Geomorfologia_mic** é aferida na Figura 6.4.

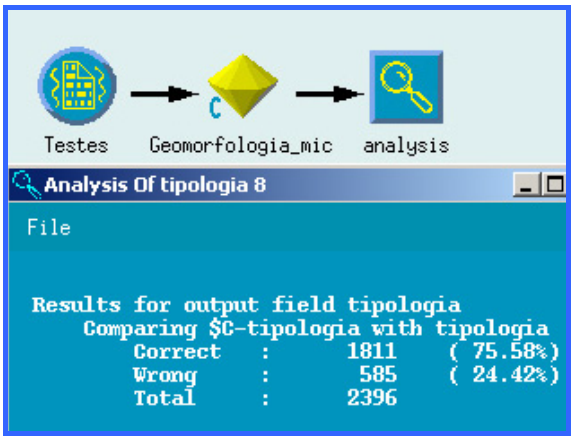


Figura 6.4 – Análise de desempenho do modelo **Geomorfologia_mic**

O desempenho do modelo foi bastante melhorado, apresentando-se as regras de decisão obtidas no Anexo II.

6.1.1.3. Geomorfologia_mac

O balanceamento pelo atributo **Geomorfologia_mac** atribuiu por defeito os factores apresentados na Tabela 6.3, para cada tipo de arqueossítio.

Geomorfologia_mac	Factor de balanceamento
null	70.3
Bacia	1.9
Depressão	2.9
Planalto	1.0
Serra	1.4

Tabela 6.3 – Factores de balanceamento gerados para **Geomorfologia_mac**

Tal como no *boost* por **Geomorfologia_mic** não se considerou ser útil o incremento dos registos com valores nulos, pelo que se colocou este factor a zero.

Aplicando os factores de balanceamento ao conjunto de dados TMO, foram obtidos 1232 registos para o conjunto de Treino e 1638 para o conjunto de Testes.

As regras produzidas são apresentadas no Anexo III e a avaliação do desempenho do modelo **Geomorfologia_mac** é mostrada na Figura 6.5.

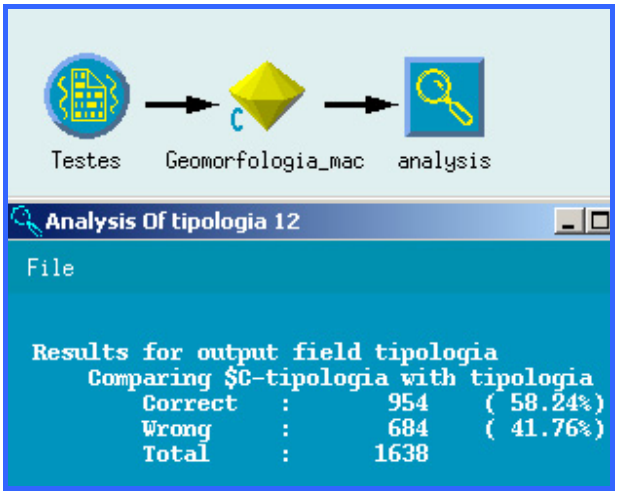


Figura 6.5 – Análise de desempenho do modelo **Geomorfologia_mac**

6.1.1.4. Topografia

O balanceamento pelo atributo **Topografia** atribuiu, por tipo de arqueossítio, os seguintes factores de balanceamento (Tabela 6.4).

Topografia	Factor de balanceamento
null	9.7
Cume	1.0
Talvegue	8.4
Vertente	1.0

Tabela 6.4 – Factores de balanceamento gerados por **Topografia**

O factor de balanceamento para o valor *null* foi retirado, tendo-se obtido 920 registos para o conjunto de Treino e 1224 para o conjunto de Testes.

A avaliação global do desempenho do modelo **Topografia** é mostrada na Figura 6.6 e as regras de decisão são apresentadas no Anexo IV.

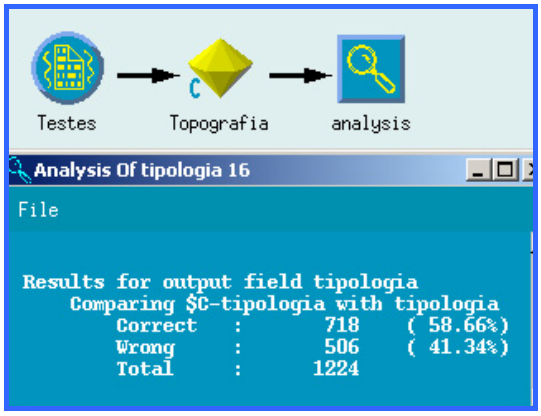


Figura 6.6 – Análise de desempenho do modelo **Topografia**

6.1.1.5. Cronologia

O balanceamento pelo atributo **Cronologia** atribuiu por defeito os factores de balanceamento, por tipo de arqueossítio que se apresentam na Tabela 6.5.

Topografia	Factor de balanceamento
Neo/Calcolítico	2.1
Idade média	1.6
Idade do Bronze	15.4
Idade do Ferro	1.6
Período Romano	1.0

Tabela 6.5 – Factores de balanceamento gerados por **Cronologia**

Aplicados estes factores de balanceamento ao conjunto de dados TMO,

obtiveram-se 1390 registos para o conjunto de Treino e 1848 para o conjunto de Testes.

As regras de decisão gradas por aplicação do algoritmo de árvores de decisão aos dados de Treino apresentam-se no Anexo V e a respectiva avaliação do desempenho do modelo **Cronologia** é mostrada na Figura 6.7.

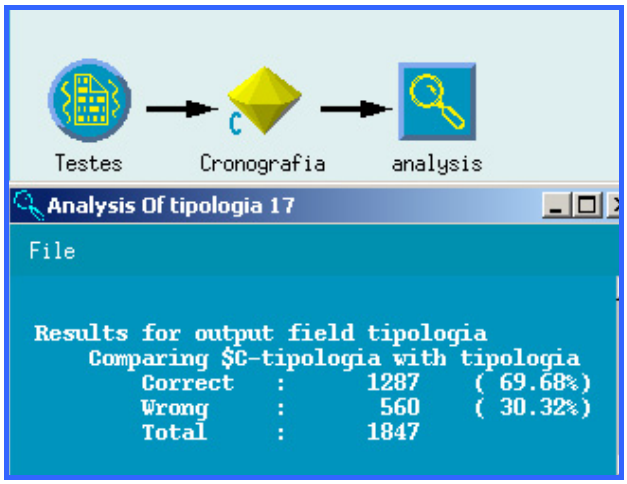


Figura 6.7 – Análise de desempenho do modelo **Cronologia**

Em função dos resultados obtidos optou-se pelo aumento de registos através da operação de *boost* por **Geomorfologia_mic**. A avaliação do desempenho do modelo obtido foi a mais elevada (cerca de 75%), quando comparada com o desempenho dos modelos obtidos. Consequentemente, é também maior a percentagem de valores correctos na previsão dos tipos de arqueossítios.

Decidida a forma como se irá incrementar o número de registos da tabela de dados irá avançar-se na identificação do modelo ***servator***, aplicando aos dados de Treino o algoritmo árvores de decisão C5.0.

As regras resultantes da aplicação do algoritmo C 5.0, ao conjunto de dados de Treino**,** estão organizadas numa estrutura em árvore, onde cada folha representa o tipo de arqueossítio que se prevê encontrar, quando se verificam as condições que conduzem a essa folha da árvore. Essas regras, já apresentadas no Anexo I, organizam o conhecimento arqueológico, evidenciando a influência dos atributos relacionados com o cenário natural onde se inserem os arqueossítios. A previsão da sua localização por **Tipologia** está quase sempre associada a factores como a **Geomorfologia**, micro ou

macro, **Topografia** e **Hierarquia_hidrográfica**. Atributos como a **Latitude**, **Longitude** e **Solos** não aparecem como elementos que influenciem a decisão.

De acordo com o pensamento arqueológico, os **Solos** seriam atributos que deviam ter tido pouca influencia na localização dos arqueossítios. Se por um lado, a BD apresenta para estes atributos uma taxa relativamente elevada de valores omissos, por outro, há uma corrente na arqueologia que defende que a localização dos *habitats* se tornou cada vez mais independente dos tipos de **Solos**, na medida que as técnicas agrícolas evoluíram, permitindo o aproveitamento dos vários tipos de **Solos** [Lemos, 1993].

Em função desta avaliação prévia retiraram-se os atributos **Latitude**, **Longitude** e **Solos** do conjunto de dados a submeter a um algoritmo de treino de redes neuronais (Figura 6.8).

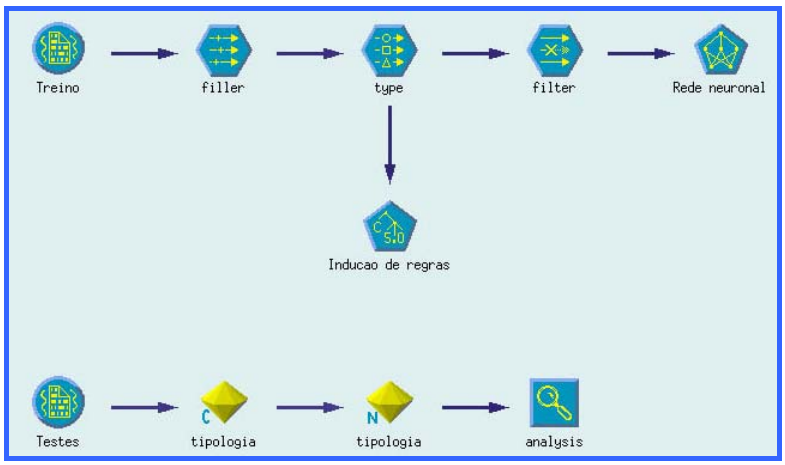


Figura 6.8 – Aplicação de algoritmos de indução de árvores de decisão e redes neuronais ao conjunto de Treino e Testes

6.1.2. Algoritmo de Rede Neuronal

Os algoritmos de indução de regras produziram uma árvore que integra regras de decisão onde, em função de um conjunto de atributos, se prevê o tipo de sítio arqueológico, isto é, o atributo **Tipologia** associado a determinado local.

O facto destas regras serem apresentadas numa linguagem natural, facilita a sua confrontação com o conhecimento já existente na área. De certa forma este está já organizado, na mente dos arqueólogos, de modo idêntico ao aqui apresentado.

Uma outra vantagem deste método está no facto de eliminar atributos que não são importantes para a tomada de decisão, enquanto que as redes neuronais, por exemplo, incluem todos as entradas consideradas [SPSS, 1999]. Assim, como já foi referido, decidiu-se aplicar primeiro um algoritmo de indução de regras e, em função dos resultados obtidos, eliminar os atributos que não são importantes para a tomada de decisão. Os restantes dados são então submetidos a um algoritmo de redes neuronais.

Para o treino da rede neuronal foram utilizados os métodos de treino da rede, disponibilizados pelo *Clementine*, nomeadamente os métodos *Quick*[12], *Dynamic*[13], *Multiple*[14], *Prune*[15] e *RBFN*[16].

A avaliação qualitativa dos resultados, quando se aplica cada um dos referidos métodos ao conjunto de Testes, está representada nos quadros da Figura 6.9.

Das experiências realizadas, a que produziu melhor resultado foi a aplicação do método *Prune* ao conjunto de dados de aprendizagem (Treino) e, quando aplicado ao conjunto de Testes**,** apresenta uma percentagem superior a 80% de valores correctos.

O método *Multiple* mostrou ser pouco apropriado, apresentando um desempenho extremamente fraco.

[12] O processo de treino da rede tem início com a melhor previsão da saída pretendida, sendo normalmente mais rápido que os outros métodos [Watkins, 2000].
[13] O método *Dynamic* tem início com duas camadas intermédias. É útil na resolução de problemas complexos ou susceptíveis de *overtrainig* (memorização das especificidades dos dados). Normalmente apresenta melhores resultados que o método *Quick,* embora seja mais demorado [Watkins, 2000].
[14] Permite treinar em paralelo diferentes saídas. Requer bastante tempo para apresentar resultados [Watkins, 2000].
[15] Este método é recomendado para treinar redes neuronais com um grande número de nodos de entrada e de camadas intermédias. À medida que a rede é treinada os níveis e camadas em excesso vão sendo removidos. Este método consome bastante tempo, mas apresenta normalmente bons resultados [Watkins, 2000].
[16] A função de base radial (*RBFN*) é uma técnica utilizada para a predição e classificação supervisionadas. O *RBFN* olha para o conjunto dos dados como um sistema espacial, tendo como requisito a predefinição das saídas da rede, para que os modelos possam ser aplicados aos dados. Neste método os dados são treinados numa única passagem, pelo que apresenta mais rapidamente resultados [Watkins, 2000].

Método	Avaliação qualitativa de resultados
Quick	Results for output field tipologia Comparing $N-tipologia with tipologia Correct : 1901 (79.34%) Wrong : 495 (20.66%) Total : 2396
Dynamic	Results for output field tipologia Comparing $N-tipologia with tipologia Correct : 1819 (75.92%) Wrong : 577 (24.08%) Total : 2396
Multiple	Results for output field tipologia Comparing $N-tipologia with tipologia Correct : 153 (7.14%) Wrong : 1991 (92.86%) Total : 2144
Prune	Results for output field tipologia Comparing $N-tipologia with tipologia Correct : 1740 (81.16%) Wrong : 404 (18.84%) Total : 2144
RBFN	Results for output field tipologia Comparing $N-tipologia with tipologia Correct : 1673 (69.82%) Wrong : 723 (30.18%) Total : 2396

Figura 6.9 – Aplicação de vários métodos de redes neuronais ao conjunto de Teste

Das experiências realizadas, a que produziu melhores resultados foi a aplicação do método *Prune* ao conjunto de dados de aprendizagem (Treino) e que, quando aplicado ao conjunto de Testes**,** apresenta uma percentagem superior a 80% de valores correctos.

O modelo ***servator*** será identificado pela aplicação, ao conjunto de dados de Treino**,** de um algoritmo de Indução de Regras (C5.0), seguido de um algoritmo de treino de rede neuronal (Método *Prune*), de acordo com a *stream* apresentada na Figura 6.8.

6.2. Análise e interpretação de resultados

O modelo obtido pela aplicação do algoritmo de indução de regras C5.0, apresenta uma *Predicted Accuracy* de 86.52%, enquanto que o modelo obtido pela aplicação de um algoritmo de rede neuronal consegue um valor de *Predicted Accuracy* de 81.16%, concordando os dois modelos em 87.41% das previsões realizadas individualmente (Figura 6.10).

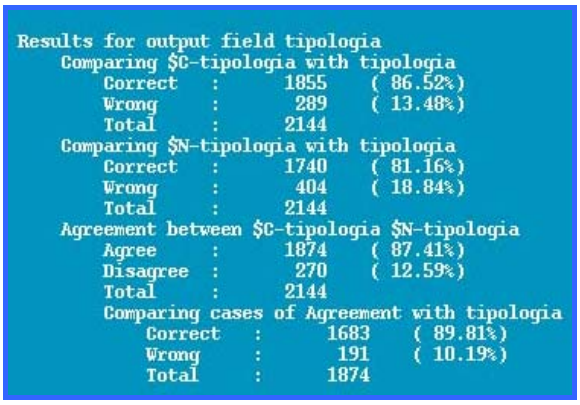


Figura 6.10 – Análise qualitativa da aplicação dos algoritmos de indução de árvores de decisão e rede neuronal aos dados de Teste

O atributo **$C-tipologia** representa a saída da árvore de decisão, enquanto que o campo **$N-tipologia** diz respeito à previsão realizada pela Rede Neuronal gerada.

A Figura 6.11 contém os campos gerados pelos modelos. As novas colunas aparecem nesta tabela com as designações de **$C-tipologia**, **$CC-tipologia**, para as geradas pelo algoritmo de indução de árvores de decisão e por **$N-tipologia** e **$NC-tipologia** para as colunas geradas pela rede neuronal. Os valores previstos pelo modelo para cada registo e o grau de confiança dessa previsão são apresentados na tabela apresentada.

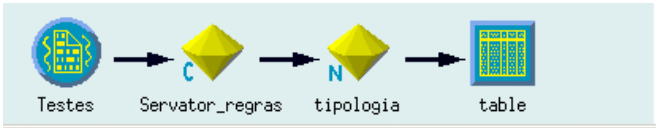


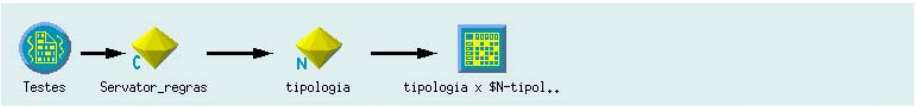
table (2386 records)

Table Generate

tipologia	$C-tipologia	$CC-tipologia	$N-tipologia	$NC-tipologia
Rede Viária	Povoado	0.611	Povoado	0.881
Rede Viária	Rede Viária	0.913	Rede Viária	0.863
Rede Viária	Rede Viária	0.913	Rede Viária	0.904
Rede Viária	Rede Viária	0.833	Rede Viária	0.765
Rede Viária	Povoado	0.6	Rede Viária	0.015
Rede Viária	Rede Viária	0.913	Rede Viária	0.4
Rede Viária	Rede Viária	0.913	Rede Viária	0.961
Rede Viária	Rede Viária	0.913	Rede Viária	0.904
Rede Viária	Rede Viária	0.913	Rede Viária	0.904
Rede Viária	Rede Viária	0.913	Rede Viária	0.318
Rede Viária	Rede Viária	0.913	Rede Viária	0.318
Rede Viária	Rede Viária	0.913	Rede Viária	0.958

Figura 6.11 – Novos campos gerados pelos algoritmos C5.0 e de redes neuronais

A Figura 6.12 apresenta uma outra forma de comparar os valores previstos com os valores reais. Cada linha está associada a um tipo de sítio arqueológico. Nas colunas da tabela estão os vários sítios possíveis, representando cada célula a previsão feita pelo modelo.



tipologia x $N-tipol.. : Percentages Across

Matrix Generate

	Achado Isolado	Arte Rupestre	Castro	Estatuária	Fortificação	Mineração	Necrópole	Povoado	Rede Viária
Abrigo	0.0	44.444	0.0	0.0	0.0	0.0	38.889	16.667	0.0
Achado Isolado	46.154	0.0	3.846	0.0	3.846	0.0	15.385	19.231	11.538
Arte Rupestre	0.0	30.0	5.0	0.0	0.0	13.333	25.0	25.0	1.667
Castro	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Epigrafia	0.0	0.0	0.0	0.0	3.509	0.0	12.281	84.211	0.0
Estatuária	0.0	0.0	7.143	64.286	0.0	3.571	0.0	25.0	0.0
Fortificação	0.0	0.0	0.0	0.0	57.282	0.0	0.0	42.718	0.0
Limites	0.0	0.0	0.0	0.0	18.182	0.0	0.0	45.455	36.364
Mineração	0.0	0.0	0.0	3.704	3.704	20.37	5.556	59.259	7.407
Necrópole	0.0	0.0	0.218	0.218	3.275	1.31	78.603	15.284	1.092
Povoado	0.0	0.116	0.116	0.465	5.459	0.929	1.278	89.431	2.207
Rede Viária	0.0	0.0	0.0	0.971	0.0	0.971	0.0	28.155	69.903
Santuário	0.0	0.0	23.077	0.0	26.923	3.846	3.846	42.308	0.0
Tesouro	0.0	0.0	12.5	12.5	0.0	0.0	0.0	62.5	12.5

Figura 6.12 – Comparação dos valores conhecidos e os previstos pela rede neuronal

Pode ver-se que o modelo não consegue prever a localização da **Tipologia Abrigo**, mas apresenta já uma boa percentagem de valores correctos para os tipos

Povoado, **Fortificação**, **Necrópole** ou mesmo **Rede Viária,** tendo até feito uma previsão correcta para a totalidade do tipo **Castro**.

As dificuldades apresentadas pelo modelo para uma previsão acertada de alguns tipos de arqueossítios, podem estar associadas ao modelo preditivo em si, mas podem também ocorrer por haver localizações coincidentes para diferentes tipos de sítios. Muitas vezes se encontra um **Povoado** rodeado por uma **Fortificação,** podendo ainda existir no mesmo local uma **Necrópole** ou uma **Rede Viária**. Como os atributos da BD estão muito limitados a dados relativos à localização geográfica dos sítios, a previsão feita dependerá portanto das características do local. A inclusão na BD de outros atributos, nomeadamente associados à história dos povos e dos locais, poderia contribuir para melhorar o modelo.

De uma forma geral os arqueossítios mais representados na BD e representativos para a actividade arqueológica, são os **Povoado, Fortificação, Necrópole** e **Rede Viária,** para os quais o modelo apresenta as taxas mais elevadas de valores correctos.

Na figura 6.13 apresentam-se algumas regras do ***servator,*** para a cronologia da **Idade do Ferro**

Cronologia Idade do Ferro

Topografia Cume
  Classaltitude [de 1000 a 1500 m] -> **Tesouro**
  Classaltitude de 0 a 400 m
    Geomorfologia_mac [Depressão tectónica, Planalto] -> **Santuário**
    Geomorfologia_mic Esporão -> **Povoado**
Topografia Talvegue -> **Rede Viária**
Topografia Vertente
  Geomorfologia_mic Arriba -> **Epigrafia**
  Geomorfologia_mic Castelo granítico -> **Povoado**

Figura 6.13 – Algumas regras de decisão do ***servator***

Como se pode verificar a leitura e interpretação destas regras é muito simples e acessível para o utilizador. Quando se prevê a localização de um sítio, apresenta-se o raciocínio seguido até chegar e essa decisão. Para além de fornecer esses indicadores, valiosos no apoio à localização de arqueossítios, podem ainda ser utilizadas noutras actividades, nomeadamente para estruturar o conhecimento arqueológico, apoiar o

ensino e aprendizagem na Arqueologia e ainda percepcionar a distribuição de sítios arqueológicos numa região e não apenas num determinado local.

6.3. Avaliação do *servator*

O DM pode gerar inúmeros padrões, no entanto, nem todos os padrões gerados são interessantes ou representam conhecimento. A avaliação dos modelos é uma componente fundamental no processo de DCBD.

Para cada padrão gerado por um sistema de DM deverá colocar-se a questão da sua qualidade e também do seu interesse para a área de conhecimento onde serão aplicados.

Um padrão de conhecimento interessante deverá ser compreensível, válido, novo e potencialmente útil.

Será feita de seguida uma análise do modelo obtido do ponto de vista do seu desempenho, da qualidade de conhecimento produzida e também da sua utilidade para a Arqueologia. Havendo padrões de conhecimento válidos e úteis será também analisada a forma como este conhecimento poderá ser integrado com o conhecimento arqueológico.

6.3.1. Avaliação do desempenho

A avaliação é o ponto-chave de qualquer processo de DM e serve dois propósitos: a previsão de como funcionará o modelo no futuro, ou mesmo se deverá ser usado, e como parte integrante de vários métodos de aprendizagem, que ajudarão a encontrar o modelo que melhor represente os dados de Treino [Souza *et al.,* 2002].

Apesar do número de registos da BD ser relativamente pequeno, optou-se fazer o *boost* dos dados, o que permitiu dividir o conjunto inicial em dois subconjuntos (Treino e Testes).

O modelo foi treinado usando os dados pré-classificados do conjunto de Treino, onde os algoritmos encontram padrões usados para prever o atributo pretendido. O modelo é testado usando um outro subconjunto de dados (Testes). A utilização de dados diferenciados, para identificar os padrões e realizar as previsões, assegura que o modelo

não dá respostas acertadas por ter memorizado os dados, mas é mais generalista e pode ser aplicado a dados desconhecidos [Berry e Linoff, 2000].

A avaliação é muito importante, uma vez a identificação de um modelo envolve normalmente a identificação de vários, escolhendo-se o que apresenta melhor desempenho. O modelo ***servator*** foi desenvolvido aplicando-se algoritmos de indução de árvores de decisão e de redes neuronais aos dados. Utilizando os mecanismos de avaliação da ferramenta *Clementine,* foi seleccionando o modelo com melhor grau de confiança, de acordo com os resultados apresentados na rubrica anterior.

Foi também feita uma avaliação qualitativa do modelo, confrontando-se os resultados obtidos pela leitura das regras de decisão, com o conhecimento existente. A Figura 6.14 apresenta algumas regras de decisão do modelo ***servator*** que confirma a localização, por exemplo, da **Arte Rupestre** nas áreas junto aos Rios (**Talvegue** , **Vertente**) ou para a **Idade do Ferro** a localização de **Santuários** em áreas onde os valores de **Hierarquia_hidrográfica** são maiores que 4, isto é, mais próximos do cimo dos montes. De acordo com as descobertas arqueológicas já realizadas, é frequente encontrarem-se **Santuários** no cimo dos montes, associados à divinização dos mesmos. Cita-se, a título de exemplo, o santuário do Deus Larouco, onde foi encontrada uma ara a evocar este monte como divindade.

Cronologia Idade do Bronze

Topografia Cume
  Geomorfologia_mic [Arriba, Castelo granítico, Crista quartzítica, Planalto, Rechã] -> **Povoado**
  Geomorfologia_mic [Cabeço, Vale] -> **Necrópole**
  Geomorfologia_mic Esporão -> **Fortificação**
Topografia [Talvegue, Vertente]-> **Arte Rupestre**

Cronologia Idade do Ferro

Classaltitude de 400 a 700 m
  Hierarquia hidrográfica > 4 -> **Santuário**

Figura 6.14 – Algumas regras de decisão do ***servator***

Os relacionamentos entre variáveis, as regras obtidas e as previsões do modelo identificado, suscitaram muitas questões interessantes para a Arqueologia. Algumas vêm confirmar o conhecimento consensual outras, porém, vêm reforçar correntes de pensamento abordadas neste trabalho, e cuja discussão arqueológica não está ainda

pacificada. É o caso da função das **Fortificações**, da importância do tipo de **Solos** na localização dos **Povoados**, ou ainda da identificação de áreas pouco prospectadas ou de reduzida identificação de sítios (margens do **Rio Douro**).

6.3.2. Avaliação da utilidade

Os modelos preditivos são construídos a partir de um conjunto de dados disponível. O grande desafio da DCBD é conseguir identificar modelos que tenham um comportamento estável, quando aplicados a novos conjuntos de dados.

O objectivo definido para este trabalho foi a identificação de um modelo preditivo de apoio à prospecção arqueológica, em Trás-os-Montes Oriental. A escolha dos locais de assentamento depende de diversos factores, nomeadamente geográficos, económicos, sociais e culturais, que variam de região para região, de comunidade para comunidade e com o tempo. Assim, o modelo preditivo será útil para aplicar à região de Trás-os-Montes Oriental. A sua utilização mais alargada a outras áreas do país será possível, a partir do momento em que se inclua na BD sítios arqueológicos dessas regiões e se identifique o modelo preditivo, utilizando a metodologia seguida neste trabalho.

O sistema ***servatis,*** apresentado no capítulo 4, pretende ser um elemento impulsionador para a identificação de novos modelos aplicados à Arqueologia.

O modelo ***servator,*** cuja área de aplicação é a região de Trás-os-Montes Oriental, permitiu identificar um modelo cujo grau de confiança apresenta bons resultados para os dados de Teste. Este modelo poderá ser aplicado sempre que se pretenda prospectar um arqueossítio nesta região. Essa utilização pode estar ligada a iniciativas arqueológicas ou de Gestão e Ordenamento do Território, nomeadamente quando se pretende fazer intervenções ao nível do subsolo, para edificações ou construção de rede viária.

Ao procurar vestígios de uma **Rede Viária,** os locais de prospecção serão necessariamente diferentes dos seleccionados para prospectar **Povoados** ou **Fortificações**. Assim, um sistema como o ***servator*** é fundamental para fazer uma primeira triagem das zonas a prospectar.

O modelo encontrado, resultante da aplicação de algoritmos de indução de

árvores de decisão, apresenta um conjunto de regras escritas em linguagem natural e com uma organização em árvore, de fácil leitura. Estas estruturam e normalizam informação que pode conduzir à construção de conhecimento arqueológico. Estas regras podem ainda funcionar como um elemento pedagógico e como uma ferramenta de trabalho em campo.

No âmbito do ensino da Arqueologia será possível, através da leitura destas regras, perceber as características diferenciadas da malha de distribuição de sítios arqueológicos, dentro do cenário fisiográfico onde se enquadram e de acordo com o período temporal a que se reportam.

Do ponto de vista de quem ensina ou tem responsabilidades na selecção das zonas a prospectar, o modelo ***servator*** pode revelar-se um importante instrumento de apoio à sua actividade. Para além de ajudar a compreender as especificidades de cada sociedade, na escolha dos locais de assentamento, permite ainda que se estudem as áreas onde o sistema revele insuficiências de construção de conhecimento. Referimos dois exemplos, detectados durante o processo de identificação do modelo. O primeiro está associado com as áreas junto ao **Rio Douro,** onde se verificou que rareavam os sítios arqueológicos, nomeadamente os do tipo **Povoado**. Este tipo de constatações pode levar os investigadores a fazer as suas reflexões e pesquisas, podendo dar-se o caso de serem zonas pouco prospectadas ou de, circunstâncias arqueológicas várias, terem eliminado estes locais das áreas escolhidas para localização de sítios arqueológicos. A primeira hipótese poderá levar à realização de prospecções arqueológicas nestes locais.

Uma outra situação evidenciada, durante a identificação do modelo, está associada com os poucos **Povoado** ou **Abrigo,** para o período **Neo/calcolítico** havendo, no entanto, sítios do tipo **Necrópole** e **Arte rupestre**. Havendo vestígios de actividade humana era natural que nesses esses locais também existissem **Abrigos** ou **Povoados.** Estamos perante uma evidência que poderá merecer a atenção dos arqueólogos. No presente caso, esta situação tinha já sido detectada e constituído objecto de estudo de alguns arqueólogos. Muitas vezes de tem associado este défice ao reduzido número de estudos, realizados a nível micro-regional, e orientados para prosseguir a busca de povoados relacionáveis com as sepulturas megalíticas [Silva, 1995].

O sistema pode portanto revelar insuficiências de conhecimento que, por sua vez apontem áreas de estudo, contribuindo dessa forma para programar as actividades associadas à investigação arqueológica.

O sistema ***servator*** para além de potenciar a aquisição de novo conhecimento, na área da Arqueologia, na medida em que permite traçar áreas geográficas de excelência, para localização de determinados tipos de arqueossítios permitirá, também, que a Indústria do Património possa elaborar roteiros culturais, com base no traçado geográfico evidenciado por este tipo de modelos.

6.4. Dificuldades encontradas no processo de DCBD

Os modelos preditivos de património arqueológico são, de acordo com o saber existente na área, influenciados por aspectos geográficos, sociais e culturais. Embora Portugal seja um país pequeno tem, no entanto, características geográficas bem diferenciadas, pelo que os modelos encontrados devem reflectir essas especificidades. Certamente que a matriz de povoamento da Região de Trás-os-Montes Oriental terá sido influenciada por factores diferentes dos existentes, por exemplo, no Alentejo ou Algarve. Assim, o modelo preditivo de apoio à prospecção arqueológica de determinada região, terá certamente de ser identificado com dados representativos da região em causa.

As BD de sítios arqueológicos passíveis de serem utilizadas no processo de DCBD, relativas à região de Trás-os-Montes Oriental, continham um número relativamente pequeno de arqueossítios, o que veio a revelar-se como um factor limitativo, implicando até o aumento do número de registos através de uma operação de *boost* dos dados.

Os problemas encontrados durante o processo de DCBD relacionam-se não só com o número de registos, mas também com o conteúdo dos dados. As BD de Arqueologia são, de uma forma geral muito descritivas, não existindo qualquer normalização, nem na organização dos dados, nem nos seus conteúdos. Os atributos têm em geral uma grande variedade de valores atribuídos e, também, uma grande diversidade de designações para caracterizar o mesmo conceito. A validação da informação é, de uma forma geral, insuficiente. Permite que se acumulem erros de digitalização de dados, com erros de interpretação e se usem diversos valores, com

diferente interpretação semântica, associados ao mesmo atributo. Os valores omissos foram também uma dificuldade, dado terem aparecido em percentagens muito elevadas e nos mais variados atributos.

A Arqueologia começa a preocupar-se já com a normalização dos dados. A partir do momento em que se definam e utilizem meta-dados e hierarquias conceptuais, as BD passarão a ser mais consistentes, permitindo que se junte e cruze mais informação e, que esta possa ser utilizada para a identificação modelos abrangentes, com informação normalizada.

Constatou-se portanto que as maiores dificuldades encontradas durante o processo de DCBD se relacionaram com os dados e não com a aplicação da metodologia adoptada.

6.5. Conclusão

O modelo ***servator,*** cuja identificação se descreveu neste capítulo, mostrou que é válido e tem utilidade, quando aplicado quer à investigação arqueológica, quer a áreas ligadas à Gestão de Património, Territorial ou Cultural.

Para permitir o acesso ao ***servator,*** definiu-se um sistema onde este modelo será disponibilizado, bem como um conjunto de ferramentas para que novas BD possam ser utilizadas e novos modelos sejam identificados. Esse sistema, denominado de ***servatis***, foi já descrito no capítulo 4.

Capítulo 7

Conclusão e trabalho futuro

Ao longo dos anos as sociedades adoptaram estratégias diferenciadas em relação ao espaço e ao meio ambiente. O homem passou de uma economia recolectora, e de aproveitamento dos recursos naturais, para atitudes mais intervencionistas que têm causado grande impacto na paisagem. No entanto, talvez como consequência das rupturas decorrentes da industrialização e da urbanização têm-se desenvolvido políticas conservacionistas dos valores patrimoniais.

O Património Arqueológico, por ser um dos recursos culturais mais sensíveis e não renováveis, tem sido objecto de acções de salvaguarda, nomeadamente através da adopção de convenções legislativas, criação de museus e de parques culturais.

No sistema legislativo e nas recomendações, nomeadamente da Convenção Europeia, ressaltam algumas linhas de acção quanto à necessidade de se proceder ao inventário dos valores arqueológicos e inserção dos sítios registados nos Planos de Ordenamento, a fim de serem considerados no planeamento urbano e do território. No entanto, a forte pressão urbanística da Idade Contemporânea provoca, a um ritmo exponencial, o achado ocasional de sítios que estavam ocultos e que, muitas vezes, são irremediavelmente destruídos.

A investigação arqueológica e a divulgação da informação são, hoje em dia, um imperativo, contribuindo para salvaguardar os vestígios do passado e da nossa memória colectiva.

Este trabalho pretendeu, por um lado, apoiar a investigação arqueológica, através da criação de uma ferramenta de apoio à detecção de património oculto e, por outro, definir um sistema que assegurasse a visibilidade deste e de outros modelos.

O ***servator*** apresenta-se como um contributo positivo para a Arqueologia, como:

- Modelo Preditivo de Apoio à Prospecção de Património Arqueológico - servindo como instrumento de apoio à Arqueologia na detecção de Património, em Trás-os-Montes Oriental, permitindo orientar e definir estratégias de prospecção;
- Elemento que estrutura e normaliza informação que potencia a construção de conhecimento arqueológico - na identificação do modelo ***servator*** foram utilizados algoritmos que produzem um conjunto de regras, apresentadas numa linguagem natural e facilmente legíveis. Estas regras estruturam e normalizam o conhecimento arqueológico. A sua leitura permite compreender as relações existentes entre os vários sítios arqueológicos e os factores geográficos e temporais que lhe estão associados;
- Potenciador de aquisição de novo conhecimento ou revelador de insuficiências de construção do conhecimento - o ***servator*** apresenta indicadores sobre a localização de arqueossítios. Se, para determinado local, se confirmam as previsões realizadas, então está-se perante um caso de sucesso. Porém, quando o modelo aponta para a existência provável de um tipo de arqueossítio e se constata que o tipo de sítio encontrado é outro, pode ser uma previsão errada ou uma excepção que leve à aquisição de novo conhecimento;
- Elemento pedagógico e ferramenta de trabalho de campo - a utilização do ***servator*** no ensino/aprendizagem da Arqueologia, nomeadamente na pesquisa dos arqueossítios, pode revelar-se um instrumento bastante enriquecedor. Proporciona uma leitura fácil das regras com informação que permite a sua utilização para encontrar locais de existência provável de sítios arqueológicos.

Para a Arqueologia, os padrões identificados pelo ***servator*** podem considerar-se:

- Válidos – porque apresentam previsões com um considerável grau de

certeza e que revelam uma lógica coerente com as descobertas de sítios arqueológicos realizadas;

- Úteis – porque pode ser aplicado na investigação arqueológica, no ensino/aprendizagem da Arqueologia, no Ordenamento e Gestão Territorial e ainda na Indústria de Turismo Cultural. O tempo e custo de realização de cada uma destas actividades poderão ser bastante reduzidos;
- Interessantes – dada a forma como o conhecimento arqueológico é estruturado e o modo como são apresentados os relacionamentos entre variáveis, pode funcionar como elemento impulsionador de novos estudos e descobertas.

O contributo do ***servator*** na área das TI, pode ser sistematizado nos seguintes pontos:

- Aplicação dos princípios da DCBD na área da Arqueologia. Os princípios foram aplicados, com sucesso, na identificação de um modelo preditivo de apoio à localização de Património arqueológico, para a região de Trás-os-Montes Oriental. As pesquisas realizadas apontam para que este trabalho seja o primeiro a utilizar a DCBD aplicada à Arqueologia;
- Definição de uma metodologia a adoptar para identificação de novos modelos aplicados à Arqueologia, de acordo com a metodologia seguida para o ***servator.***
- Aplicação de técnicas de DM na organização de instrumentos pedagógicos que podem ser utilizados no ensino/aprendizagem da Arqueologia. O pensamento arqueológico para a localização de arqueossítios, na região estudada, foi estruturado e apresentado numa linguagem natural, facilmente interpretada pelos utilizadores. As regras de decisão e os relacionamentos identificados nos dados, através de gráficos elaborados no ***servator***, tiveram um forte impacto, junto dos

especialistas de Arqueologia e pode constituir guias de apoio no ensino/aprendizagem da Arqueologia.

- Confirmação de que as maiores dificuldades na aplicação do processo de DCBD estão associadas ao tratamento e preparação dos dados. Estes são armazenados para gestão do processo arqueológico e não com o objectivo da DC. A inexistência de normas de representação dos atributos e processos de validação dos mesmos constitui ainda uma dificuldade ao desenrolar do processo de DCBD.
- Identificação de modelos que confirmam o conhecimento arqueológico existente na área e que validam todo o trabalho realizado neste projecto.

O ***servator*** apoia a localização de património visível e ainda não detectado. Esta característica torna este modelo interessante, não só para a comunidade académica, mas também para as entidades ligadas à Gestão Patrimonial, Ordenamento de Território, Turismo Cultural e público em geral.

A disponibilização alargada de modelos preditivos a vários utilizadores, com perfis de utilização pré-definidos, de acordo com as várias áreas de interesse, foi definida no sistema ***servatis.*** Este, fará a interface entre o conhecimento arqueológico existente e os mais variados domínios de utilização, permitindo alojar e tratar novos dados, bem como desenvolver outros trabalhos, tais como, definição de áreas de Património a proteger, definição de roteiros culturais, apoiar na reconstrução de antigos traçados de caminhos, aquedutos, entre outros.

O sistema ***servatis*** apresenta-se como um valioso contributo, na área das TI e da Arqueologia, nomeadamente na:

- Definição de uma arquitectura para implementação de um sistema integrado de informação arqueológica;
- Constituição de um repositório de dados e modelos no âmbito da Arqueologia;
- Definição de um sistema que integra diferentes TI, a disponibilizar aos arqueólogos de modo normalizado, amigável e apoiado por tutores *on-*

line, para gestão da diversificada informação arqueológica (dados, fotografias, cartografia e desenhos vectoriais);

- Aplicação à Arqueologia de metodologias e TI para identificação de novos modelos preditivos, aplicados a BD de âmbito regional ou nacional;
- Interacção criada entre a comunidade de arqueólogos e outras comunidades, nomeadamente Organizações que precisam de aceder à informação Patrimonial para realizar tarefas de Ordenamento do Território, elaboração de roteiros culturais pela Indústria do Património, ou simplesmente de divulgação ao público em geral;

Só com um conhecimento profundo do passado poderemos orientar de forma mais eficiente e eficaz a prospecção de sítios arqueológicos mas, acima de tudo, estabelecer uma política de conservação, cujos critérios não entrem em ruptura com o sentido histórico da paisagem. O modelo ***servator*** caracteriza-se por abrir um conjunto de oportunidades na criação de conhecimento arqueológico, para além da inventariação dos bens e da criação de itinerários turísticos ou pedagógicos, ou ainda a delimitação de zonas de protecção do património inventariado.

Como considerações finais pode afirmar-se que este trabalho aplicou os princípios da DCBD à área da Arqueologia, identificando com sucesso um modelo preditivo útil à prospecção arqueológica e no ensino/aprendizagem da Arqueologia. O ***servator*** apresenta-se bastante consistente nas suas previsões e coerente com o conhecimento arqueológico.

A definição da arquitectura de um sistema que permita a utilização de diferentes TI, num ambiente amigável e que disponibilize ao utilizador o acesso a guiões que orientam o utilizador na prossecução das suas tarefas, constitui um importante contributo para a gestão integrada de valores e recursos Patrimoniais.

Será interessante ver como trabalho futuro a aplicação destas TI e princípios da DCBD a novas BD e a um leque mais diversificado de variáveis, na identificação de novos modelos preditivos a integrar no ***servatis***.

O conhecimento gerado com base no saber já existente poderá levar-nos um pouco mais além no conhecimento das matrizes de povoamento passadas, dos critérios a que obedeciam e dos fenómenos de mudança e continuidade. Desta forma, se poderá pensar as estratégias de estudo, conservação e valorização do passado, em harmonia com as actuais estratégias de planeamento e ordenamento do território.

Bibliografia

[Allen *et al.*, 1990] Kathleen M. S. Allen, Stanton W. Green,e Ezra B. W. Zubrow
Interpreting space: GIS and archaeology. Taylor & Francis, London, 1990.

[Allen *et al.*, 2004] P. Allen, S. Feiner, A. Troccoli, H. BenkoE, Ishak, B. Smith
Seeing into the Past: Creating a 3D Modeling Pipeline for Archaeological Visualization.
2nd International Symposium on 3D Data Processing,Visualization and Transmission (3DPVT'04) . Grécia, 2004.

[Amaral e Varajão, 2000] Luís Amaral, e João Varajão
Planeamento de Sistemas de Informação.
FCA - Editora de Informática Lda, 2000.

[Anand *et al.*, 1995] Sarabjot S. Anand, David A. Bell,e John G. Hughes
Role of Domain Knowledge in Data Mining.
The CIK ' 95, Baltimore USA. ACM, 1995.

[Badia, 1992] Antoni Saborido i Badia
La utilización de los medios informáticos en arqueología.
Ciencias, metodologías y técnicas aplicadas a la Arqueología.
Universitat Autònoma de Barcelona
Barcelona, 1992.

[Baker, 1977] Philip Baker.
The Techniques of Archaeological Escavation.
Batsford Ltd, London, 1977.

[Ballart, 1997] Josep Ballart
El Patrimonio histórico y arqueológico: valor y uso.
Editorial Ariel, S.A., Barcelona, 1997.

[Barceló, 2000] Juan A. Barceló
Virtual Archaeology and Artificial Intelligence
VAST - International Symposium on Virtual Reality, Archaeology and Cultural Heritage, Italy - Arezzo, 2000.

[Barceló *et al*., 2000] Juan A. Barceló, Maurizio Forte e Donald H. Sanders
British Archaeology Reports.
Archaeopress, Oxford, 2000.

[Bernardes, 2002] Paulo José Correia Bernardes
Arqueologia Urbana e Ambientes Virtuais: um Sistema para Bracara Augusta.
Universidade do Minho, Braga, Tese de Mestrado, 1997.

[Bernardes e Martins, 2003] Paulo José Correia Bernardes e Manuela Martins
Computação gráfica e Arqueologia Urbana: o caso de Bracara Augusta.
12ª Encontro Português de Computação Gráfica, Porto, 2003.

[Berry e Linoff, 2000] Michael J. A. Berry, e Gordon Linoff
Mastering Data Mining – The Art and Science of Customer Relationship Management.
Wiley Computer Publishing, New York, USA, 2000.

[Botica *et al.*, 2003a] Natália Botica, M. Santos, e Francisco Sande Lemos
Modelo Preditivo de Património Arqueológico.
4ª Conferência da Associação Portuguesa de Sistemas de Informação, Porto, 2003. Edição CD-ROM.

[Botica *et al.*, 2003b] Natália Botica, M. Santos, e Francisco Sande Lemos
Data Mining e Património Arqueológico.
Conferência IADIS Ibero-Americana WWW/Internet 2003, Algarve, 2003.

[Botica *et al.*, 2003c] Natália Botica, Francisco Sande Lemos, e M. Santos
Desenvolvimento Sustentado - Património Arqueológico e Tecnologias de Informação.
1º Congresso Internacional de Investigação e Desenvolvimento Socio-Cultural, Cabeceiras de Basto, 2003. Edição CD-ROM.

[Brijs *et al.*, 2000] Tom Brijs, Koen Vanhoof, e Geert Wets
Reducing Redundancy in Characteristic Rule Discovery by Using IP-Techniques.
Limburg University Centre, Belgium, 2000.

[Buckles e Petry 1982] J. P. Buckles e Frederick E. Petry
A fuzzy representation of data for relational databases.
Fuzzy sets and systems Journal, 7, Pág. 213-226 1982.

[Chapman *et al.*, 2000] Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, Rudiger Wirth
CRISP -DM – Step-by-Step data mining guide,
Version 1.0, SPSS, 2000.

[Clark *et al.*, 2002] Jeffrey T. Clark, William Perrizo, James E. Landrum, Richard Frovarp, Aaron Bergstrom, Sanjay Ramaswamy e William Jockheckizo
Digital Archive Network for Anthropology.
Journal of Digital Information, 2002.

[Clementine, 1998] *User Guide*. Versão 5.
Integral Solutions Limited, 1998.

[Coba, 1991] Agroconsultores Coba
Carta dos solos, carta do uso actual da terra e carta de aptidão da terra do nordeste de Portugal-memórias.
Universidade de Trás-os-Montes e Alto Douro, 1991.

[Convenção, 1992] Convenção Europeia para a Protecção do Património Arqueológico.
La Valetta, Malta, 16 de Janeiro de 1992, Diário da República I Série A, nº 289/97, 16-12-1997, Pág, 6624 a 6638.

[Cruz e Sánchez, 1997] J. C. Martín de la Cruz, e J. Bermúdez Sánchez
La utilidad de los SIG en la investigación y Gestión del patrimonio Arqueológico de la Campina de Córdoba – Los Sistemas de Información Geográfica como Tecnología Aplicada a la Arqueología y a la Gestión del Patrimonio.
Ediciones de la Universidade Autónoma de Madrid, Cantoblanco, Madrid, 1997.

[Cruz, 2000] Carlos Manuel Simões Cruz
A Paisagem e o Povoamento na Longa Duração. O Nordeste Transmontano – Terra Quente.
Universidade do Minho, Braga, Tese de Mestrado, 2000.

[Espiago e Baena, 1997] Javier Espiago, e Javier Baena
Los SIG y el análisis espacial en Arqueologia – Los Sistemas de Información Geográfica como Tecnología Aplicada a la Arqueología y a la Gestión del Patrimonio.
Ediciones de la Universidade Autónoma de Madrid, Cantoblanco, Madrid, 1997.

[Fabre, 1992] Georges Fabre
La fotografía aérea a baja altura y su utilización en arqueología, con especial atención al norte de Francia.
Ciencias, metodologías y técnicas aplicadas a la arqueología.
Ediciones Bellaterra, S.A., Barcelona, 1992, pág 139 - 153.

[Fayyad *et al*., 1996a] U. M.Fayyad, G. Piatetsky-Shapiro, Padhraic Smythe R. e Uthurusamy
Advances in Knoledge Discovery and Data Mining.
The MIT Press, Massachusetts, 1996.

[Fayyad *et al*., 1996b] U. M.Fayyad, G. Piatetsky-Shapiro, e Padhraic Smyth.
From Data Mining to Knowledge Discovery in Databases
American Association for Artificial Intelligence, pág. 37-54, 1996.

[Fernández, 1977] Rafael Ramos Fernández.
Arqueologia – Metodos e Tecnicas
Ediciones Bellaterra, S.A., Barcelona, 1977.

[Figueiredo, 1995] F. P Figueiredo
Métodos de Resistividade aplicados ao estudo de monumentos megalíticos: o Dólmen de Picoto do Vasco.
Revista do Centro de Estudos Pré-históricos da Beira Alta, Vol. III, Viseu, 1995.

[Fowler, 1994] Peter Fowler
Archaeology in a Matriz - Archaeological Resource Management in the UK. An introduction.
Allan Sutton Publishing Inc, 1994.

[Gamble, 2002] Clive Gamble
Arqueología básica.
Ariel Prehistória,Barcelona, 2002.

[Giestal, 1998] Carlos Dantas Giestal
Sistema de Informação Geográfica para a Arqueologia Urbana: O Caso de Bracara Augusta.
Universidade do Minho, Braga, 1998, Tese de Mestrado.

[GIS, 2002] *Archaeology – the Archaeological Applications of GIS*
Taylor & Francis, Inc., London, 2002.

[Gomes, 2000] Mário Varela Gomes
Arte Rupestre em Portugal – perspectiva sobre o último século.
Arqueologia e História - Arqueologia 2000 – Balanço de um século de investigação arqueológica em Portugal. Nº 54, pág. 139-194. Associação dos Arqueólogos Portugueses. Lisboa, 2002.

[Green, 2002] Kevin Green
Archaeology: an introduction.
Routledge, fourth edition, 2002.

[Han e Kamber, 2001] Jiawei Han, e Micheline Kamber
Data Mining: Concepts and Techniques.
Morgan Kaufmann Publishers, 2001.

[Hernández e Stolfo, 1998] Mauricio A. Hernández, e Salvatore J.Stolfo
Real-world Data is Dirty:Data Cleasing and The Merge/Purge Problem.
Data Mining and Knowledge Discovery Journal, Vol. 2, pp. 9-37, 1998.

[Hynst *et al.*, 2002] Stefan Hynst, Michael Gervautz, Markus Grabner e Konrad Schindler
A Work-flow and data model for Reconstruction, Management and Visualization of Archaeological Sites.
International Symposium on Virtual Reality, Archaeology and Cultural Heritage. Grécia. 2001.

[Jermyn *et al.,* 1999] Jermyn, Maurice Dixon e Brian J. Read Paul
Preparing Clean Views of Data for Data Mining.
Conference Proceedings of ERCIM Workshop on Database Research, Amsterdam, 1999.

[Keim *et al.*, 2002] Daniel A. Keim, Wolfgang Muller e Heidrun Scumann
Visual Data Mining.
EUROGRAPHICS 2002. Eurographics Digital Library. 2002.

[Kohonen, 1989] Teuvo Kohonen
Self-Organization and Associative Memory.
Springer-Verlag, Berlin, 3ª edição, 1989.

[Kuiper e Wescott, 1999] James A. Kuiper e Konnie L. Wescott
A GIS for Predicting Prehistoric Site Locations.
90ª Annual ESRI Conference. California, EUA, 1999.

[Lemos, 1991] Francisco Sande Lemos
Ordenamento da Paisagem e Conservação do Património Arqueológico.
FORUM , Nº9/10, Universidade do Minho, 1991.

[Lemos, 1993] Francisco Sande Lemos
Povoamento Romano de Trás-os-Montes Oriental.
Universidade do Minho, Braga, 1993, Tese de Doutoramento.

[Lemos, 2002] Francisco Sande Lemos
Bracara Augusta – a grande plataforma viária do Noroeste da Hispania,
FORUM , Nº31, Universidade do Minho, 2002.

[Leusen, 2002] Peter Martjin van Leusen
Pattern to process: methodological investigation into the formation and interpretation of spatial patterns in archaeological landscapes.
Rijksuniversiteit Groninggen, 2002, Tese de Doutoramento.

[Lobo e Moura-Pires, 1998] Vitor Sousa Lobo e Fernando Moura-Pires
Som – Kohonen's Self-Organizing Maps. International Summer School on. Knowledge Discovery in Databases and Data Mining: Methods and Applications, Volume II. Caminha, Portugal, 1998.

[Lock e Stancic, 1995] Gary Lock, e Zoran Stancic
Archaeology and Geographical Information Systems: a European Perspective.
1995.

[Martínez, 1992] Víctor M. F. Martínez
Teoría e Método de la Arqueología.
Editorial Sínteses, Madrid, 1992.

[McGill, 1995] Greg McGill
Building on the Past- a guide to the archaeology and development process.
E&FN Spon, 1995.

[Medeiros, 2000] Carlos Alberto Medeiros
Geografia de Portugal – Ambiente Natural e Ocupação Humana : Uma Introdução.
Editorial Estampa, Lisboa, 2000.

[Microsoft, 2003] Microsoft
SQL Server 2000 Resource Kit- Effective Strategies for Data Mining.
2003.

[MNAE, 1989] *Portugal das Origens à Época Romana*.
Museu Nacional de Arqueologia e Etnologia, Lisboa, 1989.

[Oxford, 1996] The Oxford Companion to Archaeology, Brian M. Fagam. Oxford University Press, 1996.

[Preysler *et al.,* 1997] J. Baena Preysler, C. Blasco Bosqued e, L. Ramos Gómez
Aplicación de los SIG al Tratamiento de las Imágenes– Los Sistemas de Información Geográfica como Tecnología Aplicada a la Arqueología y a la Gestión del Patrimonio.
Ediciones de la Universidade Autónoma de Madrid, Cantoblanco, Madrid, 1997.

[Pyle, 1999] Dorian Pyle
Data Preparation for Data Mining.
Morgan Kaufmann Publishers, Inc., San Francisco, California, 1999.

[Redentor, 2002] Armando Redentor
Epigrafia romana da região de Bragança.
Trabalhos de Arqueologia, Nº 24, Instituto Português de Arqueologia, Lisboa, 2002.

[Renfrew e Bahn, 1991] Colin Renfrew e Paul Bahn
Archaeology. Theories, Methods and Practice.
Thames and Hudson Lta, London,1991.

[Ribeiro *et al*., 1991] Orlando Ribeiro, Hermann Lautensach, e Suzanne Daveau
Geografia de Portugal – I. A Posição Geográfica e o Território.
Edições João Sá da Costa, Lisboa,1991.

[Rodrigues, 2000] Maria de Fátima C. Rodrigues
Arquitectura Heterogénea para Extracção de Conhecimento a partir de Dados. Universidade do Minho, Braga, 2000, Tese de Doutoramento.

[Rodrigues *et al.* 1998] Maria de Fátima C. Rodrigues, Carlos Ramos e Pedro Rangel Henriques
Extracção de Conhecimento em Sistemas de Informação Imprecisos.
EEI'98, 1998.

[Sanches e Santos, 1987] Maria de Sanches e Branca C. T. Santos
Levantamento Arqueológico do Concelho de Mirandela.
Portugália, 2ª Série 8. Porto, 1987.

[Sánchez, 2000] Jesús Bermúdez Sánchez
La aplicación de los sistemas de información geográfica a la arqueología.
Universidad Autónoma de Madrid, Facultad de Filosofía y Letras, 2000, Tese de Doutoramento.

[Sánchez-Marrè *et al*., 2002] Miquel Sánchez-Marrè, Karina Gibert, Ignasi Rodríguez-Roda, Eva Bueno, Lidia Mozo, Aleix Clavell, Mario Martín, e Philippe Rougé
Development of Intelligent Data Anayisis System for Knowledge Management in Environmental Data Bases.
Procceedings da International Environmental Modelling and Software Society, Switzerland, 2002.

[Santos, 2001] M. Santos
PADRÂO – Um sistema de Descoberta de Conhecimento em Bases de Dados Georeferenciadas.
Universidade do Minho, 2001, Tese de Doutoramento
.

[Silva, 1995] António José Marques da Silva
O Megalitismo na Beira Alta.
Cyberarqueólogo Português, Coimbra, 1995.

[Souza *et al*., 2002] Jerffeson Souza, Stan Matwin e Nathalie Japkowicz
Evaluating Data Mining Models: A Pattern Language.
Procceedings da 9th Conference on Pattern Language of Programs, Illinois, 2002.

[SPSS, 1999] SPSS, Clementine, *User Guide*,Versão 5.2, SPSS Inc., 1999.

[Vote, 2001] Eileen Louise Vote
A New Methodology for Archaeological Analysis: Using Visualization and Interaction to Explore Spatial Links in Excavation Data.
Brown University, Providence - Rhode Island, 2001, Tese de Doutoramento.

[Watkins, 2000] David Watkins
Neural Networks MasterClass, 2000.
www.spss.com/clementine/clug/ppts/NNM.ppt

[Wheatley e Gillings, 2002] David Wheatley e Mark Gillings
Spatial Technology and Archaeology - the Archaeological Applications of GIS.
Taylor & Francis, Inc, London, 2002.
.

[Yaginuma, 2000] Yoshinori Yaginuma
High-performance Data Mining System.
Fujitsu Dci. Tech. J., 36, 2, pp201-210. 2000.

[Zorrinho, 1991] Carlos Zorrinho
Gestão da Informação.
Editorial Presença, 1991.

Anexos

Anexo I

Regras de decisão após balanceamento por Tipologia

Cronologia Neo/Calcolítico

- Geomorfologia_mic Arriba -> Arte Rupestre
- Geomorfologia_mic Planalto -> Necrópole
- Geomorfologia_mic Castelo granítico
  - classaltitude De 0 a 400 m -> Povoado
  - classaltitude De 400 a 700 m -> Arte Rupestre
  - classaltitude De 700 a 1000 m -> Povoado
  - classaltitude De 1000 a 1500 m -> Povoado
- Geomorfologia_mic Rechã
  - Hierarquia hidrográfica =< 2 -> Achado Isolado
  - Hierarquia hidrográfica > 2 -> Necrópole
- Geomorfologia_mic Crista quartzítica
  - Hierarquia hidrográfica =< 2
    - Hierarquia hidrográfica =< 1 -> Fortificação
    - Hierarquia hidrográfica > 1 -> Abrigo
  - Hierarquia hidrográfica > 2 -> Arte Rupestre
- Geomorfologia_mic Esporão
  - Hierarquia hidrográfica =< 2
    - geomorfologia_mac [Serra] -> Abrigo
    - geomorfologia_mac Bacia hidrográfica -> Abrigo
    - geomorfologia_mac Depressão tectónica -> Arte Rupestre
    - geomorfologia_mac Planalto -> Necrópole
  - Hierarquia hidrográfica > 2
    - classaltitude De 0 a 400 m -> Povoado
    - classaltitude De 400 a 700 m -> Arte Rupestre
    - classaltitude De 700 a 1000 m -> Povoado
    - classaltitude De 1000 a 1500 m -> Arte Rupestre
- Geomorfologia_mic Cabeço
  - topografia Talvegue -> Santuário
  - topografia Cume
    - classaltitude De 0 a 400 m -> Santuário
    - classaltitude De 400 a 700 m -> Arte Rupestre
    - classaltitude De 700 a 1000 m , De 1000 a 1500 m -> Santuário
  - topografia Vertente
    - geomorfologia_mac [Depressão tectónica] -> Arte Rupestre
    - geomorfologia_mac Bacia hidrográfica -> Arte Rupestre
    - geomorfologia_mac Planalto -> Necrópole

```
        geomorfologia_mac Serra
            Hierarquia hidrográfica =< 1 -> Arte Rupestre
            Hierarquia hidrográfica  > 1 -> Necrópole
    Geomorfologia_mic Vale
        topografia Cume -> Necrópole
        topografia Talvegue
            classaltitude De 0 a 400 m
                geomorfologia_mac [0 Planalto Serra] -> Arte Rupestre
                geomorfologia_mac Bacia hidrográfica -> Abrigo
                geomorfologia_mac Depressão tectónica -> Arte Rupestre
            classaltitude De 400 a 700 m -> Necrópole
            classaltitude [De 1000 a 1500 m, De 700 a 1000 m] -> Arte Rupestre
        topografia Vertente
            classaltitude De 0 a 400 m -> Arte Rupestre
            classaltitude De 400 a 700 m -> Arte Rupestre
            classaltitude De 700 a 1000 m
                geomorfologia_mac ['Bacia hidrográfica' 'Depressão tectónica'] -> Arte
Rupestre
                geomorfologia_mac Planalto -> Necrópole
                geomorfologia_mac Serra
                    Hierarquia hidrográfica =< 4 -> Arte Rupestre
                    Hierarquia hidrográfica  > 4 -> Necrópole
            classaltitude De 1000 a 1500 m -> Arte Rupestre

Cronologia Idade do Bronze

    topografia Talvegue -> Arte Rupestre
    topografia Vertente -> Arte Rupestre
    topografia Cume
        Geomorfologia_mic ['Crista quartzítica' Planalto Rechã] -> Povoado
        Geomorfologia_mic Arriba -> Povoado
        Geomorfologia_mic Cabeço -> Necrópole
        Geomorfologia_mic Castelo granítico -> Povoado
        Geomorfologia_mic Esporão -> Fortificação
        Geomorfologia_mic Vale -> Necrópole

Cronologia Idade do Ferro

    topografia Talvegue -> Castro
    topografia Vertente
        Hierarquia hidrográfica =< 2 -> Santuário
        Hierarquia hidrográfica  > 2
            geomorfologia_mac Depressão tectónica -> Castro
            geomorfologia_mac Serra -> Castro
```

geomorfologia_mac Planalto
classaltitude ['De 0 a 400 m' 'De 1000 a 1500 m' 'De 400 a 700 m'] -> Arte Rupestre
classaltitude De 700 a 1000 m -> Castro
geomorfologia_mac Bacia hidrográfica
classaltitude De 0 a 400 m
Hierarquia hidrográfica =< 3 -> Arte Rupestre
Hierarquia hidrográfica > 3 -> Estatuária
classaltitude De 400 a 700 m -> Arte Rupestre
classaltitude ['De 1000 a 1500 m' 'De 700 a 1000 m'] -> Estatuária
topografia Cume
Geomorfologia_mic [Planalto Rechã Vale] -> Castro
Geomorfologia_mic Arriba -> Castro
Geomorfologia_mic Crista quartzítica -> Castro
Geomorfologia_mic Esporão -> Castro
Geomorfologia_mic Castelo granítico
classaltitude De 0 a 400 m -> Castro
classaltitude De 400 a 700 m
Hierarquia hidrográfica =< 4 -> Castro
Hierarquia hidrográfica > 4 -> Santuário
classaltitude De 700 a 1000 m -> Castro
classaltitude ['De 1000 a 1500 m'] -> Castro
Geomorfologia_mic Cabeço
Hierarquia hidrográfica =< 3
Hierarquia hidrográfica =< 1 -> Castro
Hierarquia hidrográfica > 1
Hierarquia hidrográfica =< 2 -> Santuário
Hierarquia hidrográfica > 2
classaltitude De 0 a 400 m -> Castro
classaltitude De 400 a 700 m -> Castro
classaltitude De 700 a 1000 m
geomorfologia_mac ['Depressão tectónica'] -> Santuário
geomorfologia_mac Bacia hidrográfica -> Castro
geomorfologia_mac Planalto -> Castro
geomorfologia_mac Serra -> Santuário
classaltitude De 1000 a 1500 m -> Castro
Hierarquia hidrográfica > 3 -> Castro

Cronologia Período Romano

topografia Talvegue
classaltitude De 0 a 400 m -> Rede Viária
classaltitude De 400 a 700 m -> Rede Viária
classaltitude ['De 1000 a 1500 m' 'De 700 a 1000 m'] -> Rede Viária

topografia 0
classaltitude De 0 a 400 m -> Tesouro
classaltitude De 400 a 700 m -> Tesouro
classaltitude De 700 a 1000 m -> Povoado
classaltitude De 1000 a 1500 m -> Epigrafia
topografia Cume
classaltitude De 0 a 400 m
geomorfologia_mac [0 Planalto] -> Santuário
geomorfologia_mac Depressão tectónica -> Santuário
geomorfologia_mac Bacia hidrográfica
Geomorfologia_mic ['Crista quartzítica' Planalto Rechã Vale] -> Estatuária
Geomorfologia_mic Arriba -> Estatuária
Geomorfologia_mic Cabeço -> Fortificação
Geomorfologia_mic Castelo granítico -> Limites
Geomorfologia_mic Esporão -> Povoado
geomorfologia_mac Serra
Geomorfologia_mic [Arriba Cabeço 'Crista quartzítica' Planalto Rechã] -> Fortificação
Geomorfologia_mic Castelo granítico -> Fortificação
Geomorfologia_mic Esporão -> Fortificação
Geomorfologia_mic Vale -> Mineração
classaltitude De 400 a 700 m
Geomorfologia_mic [Arriba 'Castelo granítico' Rechã] -> Santuário
Geomorfologia_mic Crista quartzítica -> Mineração
Geomorfologia_mic Planalto -> Rede Viária
Geomorfologia_mic Vale -> Rede Viária
Geomorfologia_mic Cabeço
geomorfologia_mac Planalto -> Fortificação
geomorfologia_mac Serra -> Mineração
geomorfologia_mac Bacia hidrográfica
Hierarquia hidrográfica =< 2 -> Santuário
Hierarquia hidrográfica > 2 -> Fortificação
geomorfologia_mac Depressão tectónica
Hierarquia hidrográfica =< 3 -> Santuário
Hierarquia hidrográfica > 3 -> Povoado
Geomorfologia_mic Esporão
geomorfologia_mac ['Depressão tectónica'] -> Santuário
geomorfologia_mac Bacia hidrográfica -> Fortificação
geomorfologia_mac Serra
Hierarquia hidrográfica =< 4 -> Mineração
Hierarquia hidrográfica > 4 -> Santuário
geomorfologia_mac Planalto
Hierarquia hidrográfica =< 4
Hierarquia hidrográfica =< 2 -> Fortificação

```
                Hierarquia hidrográfica  > 2 -> Mineração
              Hierarquia hidrográfica  > 4 -> Santuário
      classaltitude De 700 a 1000 m
        Geomorfologia_mic [Arriba 'Castelo granítico' 'Crista quartzítica' Rechã] ->
Epigrafia
        Geomorfologia_mic Planalto -> Rede Viária
        Geomorfologia_mic Vale -> Mineração
        Geomorfologia_mic Esporão
          Hierarquia hidrográfica =< 4 -> Epigrafia
          Hierarquia hidrográfica  > 4 -> Mineração
        Geomorfologia_mic Cabeço
          Hierarquia hidrográfica =< 4
            geomorfologia_mac ['Bacia hidrográfica'] -> Epigrafia
            geomorfologia_mac Depressão tectónica -> Estatuária
            geomorfologia_mac Planalto
              Hierarquia hidrográfica =< 1 -> Epigrafia
              Hierarquia hidrográfica  > 1 -> Estatuária
            geomorfologia_mac Serra
              Hierarquia hidrográfica =< 1 -> Mineração
              Hierarquia hidrográfica  > 1 -> Epigrafia
          Hierarquia hidrográfica  > 4 -> Santuário
      classaltitude De 1000 a 1500 m -> Tesouro
    topografia Vertente
      Geomorfologia_mic [Arriba] -> Epigrafia
      Geomorfologia_mic Castelo granítico -> Povoado
      Geomorfologia_mic Rechã
        Hierarquia hidrográfica =< 2 -> Povoado
        Hierarquia hidrográfica  > 2 -> Rede Viária
      Geomorfologia_mic Crista quartzítica
        Hierarquia hidrográfica =< 1 -> Epigrafia
        Hierarquia hidrográfica  > 1
          classaltitude ['De 0 a 400 m' 'De 1000 a 1500 m' Nulos] -> Mineração
          classaltitude De 400 a 700 m -> Mineração
          classaltitude De 700 a 1000 m -> Povoado
      Geomorfologia_mic Planalto
        classaltitude ['De 0 a 400 m' 'De 1000 a 1500 m'] -> Epigrafia
        classaltitude De 400 a 700 m
          Hierarquia hidrográfica =< 2 -> Necrópole
          Hierarquia hidrográfica  > 2 -> Rede Viária
        classaltitude De 700 a 1000 m -> Epigrafia
      Geomorfologia_mic Esporão
        geomorfologia_mac Depressão tectónica -> Mineração
        geomorfologia_mac Planalto -> Povoado
        geomorfologia_mac Serra -> Povoado
```

```
    geomorfologia_mac Bacia hidrográfica
      Hierarquia hidrográfica =< 1 -> Epigrafia
      Hierarquia hidrográfica  > 1
        Hierarquia hidrográfica =< 3 -> Rede Viária
        Hierarquia hidrográfica  > 3 -> Povoado
Geomorfologia_mic Cabeço
    geomorfologia_mac Depressão tectónica
      classaltitude De 0 a 400 m -> Necrópole
      classaltitude ['De 1000 a 1500 m' Nulos] -> Santuário
      classaltitude De 700 a 1000 m -> Santuário
      classaltitude De 400 a 700 m
        Hierarquia hidrográfica =< 3 -> Rede Viária
        Hierarquia hidrográfica  > 3 -> Tesouro
    geomorfologia_mac Bacia hidrográfica
      classaltitude De 0 a 400 m
        Hierarquia hidrográfica =< 2 -> Mineração
        Hierarquia hidrográfica  > 2 -> Estatuária
      classaltitude De 400 a 700 m
        Hierarquia hidrográfica =< 3
          Hierarquia hidrográfica =< 2 -> Epigrafia
          Hierarquia hidrográfica  > 2 -> Mineração
        Hierarquia hidrográfica  > 3 -> Rede Viária
      classaltitude ['De 1000 a 1500 m' 'De 700 a 1000 m'] -> Rede Viária
    geomorfologia_mac Planalto
      classaltitude ['De 0 a 400 m' 'De 1000 a 1500 m'] -> Epigrafia
      classaltitude De 400 a 700 m
        Hierarquia hidrográfica =< 2
          Hierarquia hidrográfica =< 1 -> Epigrafia
          Hierarquia hidrográfica  > 1 -> Mineração
        Hierarquia hidrográfica  > 2 -> Epigrafia
      classaltitude De 700 a 1000 m
        Hierarquia hidrográfica =< 2 -> Epigrafia
        Hierarquia hidrográfica  > 2 -> Povoado
    geomorfologia_mac Serra
      classaltitude De 0 a 400 m -> Santuário
      classaltitude De 400 a 700 m
        Hierarquia hidrográfica =< 1 -> Epigrafia
        Hierarquia hidrográfica  > 1
          Hierarquia hidrográfica =< 2 -> Santuário
          Hierarquia hidrográfica  > 2 -> Rede Viária
      classaltitude De 700 a 1000 m
        Hierarquia hidrográfica =< 3 -> Mineração
        Hierarquia hidrográfica  > 3
          Hierarquia hidrográfica =< 4 -> Epigrafia
```

```
                Hierarquia hidrográfica  > 4 -> Mineração
            classaltitude De 1000 a 1500 m -> Estatuária
        Geomorfologia_mic Vale
            geomorfologia_mac Serra
                Hierarquia hidrográfica =< 3 -> Mineração
                Hierarquia hidrográfica  > 3 -> Tesouro
            geomorfologia_mac Depressão tectónica
                classaltitude De 0 a 400 m -> Epigrafia
                classaltitude ['De 1000 a 1500 m' 'De 700 a 1000 m'] -> Epigrafia
                classaltitude De 400 a 700 m
                    Hierarquia hidrográfica =< 4 -> Mineração
                    Hierarquia hidrográfica  > 4 -> Rede Viária
            geomorfologia_mac Planalto
                classaltitude ['De 0 a 400 m' 'De 1000 a 1500 m'] -> Mineração
                classaltitude De 700 a 1000 m -> Epigrafia
                classaltitude De 400 a 700 m
                    Hierarquia hidrográfica =< 2 -> Rede Viária
                    Hierarquia hidrográfica  > 2 -> Mineração
            geomorfologia_mac Bacia hidrográfica
                Hierarquia hidrográfica =< 3
                    Hierarquia hidrográfica =< 1 -> Rede Viária
                    Hierarquia hidrográfica  > 1
                        Hierarquia hidrográfica =< 2 -> Arte Rupestre
                        Hierarquia hidrográfica  > 2 -> Povoado
                Hierarquia hidrográfica  > 3 -> Epigrafia

Cronologia Idade Média

    topografia Talvegue -> Rede Viária
    topografia Cume
        Geomorfologia_mic [Arriba Rechã] -> Fortificação
        Geomorfologia_mic Esporão -> Fortificação
        Geomorfologia_mic Planalto -> Limites
        Geomorfologia_mic Castelo granítico
            geomorfologia_mac ['Bacia hidrográfica'] -> Fortificação
            geomorfologia_mac Depressão tectónica -> Povoado
            geomorfologia_mac Planalto -> Fortificação
            geomorfologia_mac Serra -> Fortificação
        Geomorfologia_mic Vale
            geomorfologia_mac ['Depressão tectónica'] -> Rede Viária
            geomorfologia_mac Bacia hidrográfica -> Rede Viária
            geomorfologia_mac Planalto -> Rede Viária
            geomorfologia_mac Serra -> Limites
        Geomorfologia_mic Crista quartzítica
```

```
      geomorfologia_mac Bacia hidrográfica -> Fortificação
      geomorfologia_mac Depressão tectónica -> Povoado
      geomorfologia_mac Serra -> Fortificação
      geomorfologia_mac Planalto
        Hierarquia hidrográfica =< 3
          Hierarquia hidrográfica =< 1 -> Fortificação
          Hierarquia hidrográfica  > 1 -> Santuário
        Hierarquia hidrográfica  > 3 -> Fortificação
    Geomorfologia_mic Cabeço
      geomorfologia_mac Bacia hidrográfica -> Fortificação
      geomorfologia_mac Depressão tectónica -> Fortificação
      geomorfologia_mac Serra -> Fortificação
      geomorfologia_mac Planalto
        Hierarquia hidrográfica =< 2 -> Fortificação
        Hierarquia hidrográfica  > 2
          classaltitude ['De 0 a 400 m' 'De 1000 a 1500 m' Nulos] -> Santuário
          classaltitude De 400 a 700 m -> Santuário
          classaltitude De 700 a 1000 m
            Hierarquia hidrográfica =< 3 -> Santuário
            Hierarquia hidrográfica  > 3 -> Fortificação
topografia Vertente
    Geomorfologia_mic [Arriba 'Castelo granítico'] -> Santuário
    Geomorfologia_mic Crista quartzítica -> Necrópole
    Geomorfologia_mic Esporão -> Rede Viária
    Geomorfologia_mic Rechã -> Povoado
    Geomorfologia_mic Planalto
      Hierarquia hidrográfica =< 1 -> Mineração
      Hierarquia hidrográfica  > 1 -> Rede Viária
    Geomorfologia_mic Vale
      Hierarquia hidrográfica =< 4
        classaltitude De 0 a 400 m -> Povoado
        classaltitude De 1000 a 1500 m -> Povoado
        classaltitude De 700 a 1000 m -> Povoado
        classaltitude De 400 a 700 m
          Hierarquia hidrográfica =< 3 -> Necrópole
          Hierarquia hidrográfica  > 3 -> Fortificação
      Hierarquia hidrográfica  > 4
        geomorfologia_mac ['Bacia hidrográfica' Planalto] -> Limites
        geomorfologia_mac Depressão tectónica -> Rede Viária
        geomorfologia_mac Serra -> Limites
    Geomorfologia_mic Cabeço
      geomorfologia_mac Bacia hidrográfica -> Necrópole
      geomorfologia_mac Depressão tectónica -> Santuário
      geomorfologia_mac Serra
```

```
    Hierarquia hidrográfica =< 1 -> Mineração
    Hierarquia hidrográfica  > 1
      classaltitude ['De 0 a 400 m' 'De 1000 a 1500 m'] -> Rede Viária
      classaltitude De 400 a 700 m -> Mineração
      classaltitude De 700 a 1000 m -> Rede Viária
geomorfologia_mac Planalto
    Hierarquia hidrográfica =< 2
      classaltitude ['De 0 a 400 m' 'De 1000 a 1500 m'] -> Santuário
      classaltitude De 400 a 700 m -> Santuário
      classaltitude De 700 a 1000 m
        Hierarquia hidrográfica =< 1 -> Necrópole
        Hierarquia hidrográfica  > 1 -> Santuário
    Hierarquia hidrográfica  > 2
      classaltitude De 0 a 400 m -> Necrópole
      classaltitude ['De 1000 a 1500 m'] -> Fortificação
      classaltitude De 400 a 700 m -> Fortificação
      classaltitude De 700 a 1000 m -> Necrópole
```

Anexo II

Regras de decisão após balanceamento por Geomorfologia_mic

Cronologia Neo/Calcolítico

- Geomorfologia_mic Arriba -> Necrópole
- Geomorfologia_mic Planalto -> Necrópole
- Geomorfologia_mic Castelo granítico
  - topografia Cume -> Povoado
  - topografia Talvegue -> Povoado
  - topografia Vertente -> Povoado
- Geomorfologia_mic Esporão
  - topografia [0 Talvegue] -> Arte Rupestre
  - topografia Cume -> Povoado
  - topografia Vertente -> Arte Rupestre
- Geomorfologia_mic Cabeço
  - topografia 0 -> Arte Rupestre
  - topografia Cume -> Povoado
  - topografia Talvegue -> Arte Rupestre
  - topografia Vertente
    - geomorfologia_mac [0 'Depressão tectónica' Planalto] -> Arte Rupestre
    - geomorfologia_mac Bacia hidrográfica -> Arte Rupestre
    - geomorfologia_mac Serra -> Necrópole
- Geomorfologia_mic Crista quartzítica
  - topografia [0 Talvegue] -> Arte Rupestre
  - topografia Cume -> Povoado
  - topografia Vertente
    - Hierarquia hidrográfica =< 2 -> Abrigo
    - Hierarquia hidrográfica > 2 -> Arte Rupestre
- Geomorfologia_mic Rechã
  - Hierarquia hidrográfica =< 2
    - geomorfologia_mac [0 'Bacia hidrográfica' Serra] -> Povoado
    - geomorfologia_mac Depressão tectónica -> Achado Isolado
    - geomorfologia_mac Planalto -> Povoado
  - Hierarquia hidrográfica > 2 -> Necrópole
- Geomorfologia_mic Vale
  - classaltitude De 0 a 400 m -> Arte Rupestre
  - classaltitude De 1000 a 1500 m -> Arte Rupestre
  - classaltitude De 700 a 1000 m -> Necrópole
  - classaltitude Nulos -> Necrópole
  - classaltitude De 400 a 700 m

```
        Hierarquia hidrográfica =< 2 -> Arte Rupestre
        Hierarquia hidrográfica  > 2 -> Necrópole
   Geomorfologia_mic 0
      geomorfologia_mac Bacia hidrográfica -> Necrópole
      geomorfologia_mac Planalto -> Necrópole
      geomorfologia_mac Serra -> Necrópole
      geomorfologia_mac Depressão tectónica
         classaltitude ['De 0 a 400 m' 'De 1000 a 1500 m' 'De 700 a 1000 m'] ->
Necrópole
         classaltitude De 400 a 700 m
            Hierarquia hidrográfica =< 2 -> Abrigo
            Hierarquia hidrográfica  > 2 -> Arte Rupestre

Cronologia Idade do Bronze


   topografia Cume -> Povoado
   topografia Talvegue -> Achado Isolado
   topografia Vertente -> Achado Isolado

Cronologia Idade do Ferro

   topografia Cume -> Castro
   topografia Talvegue -> Castro
   topografia Vertente -> Castro


Cronologia Período Romano

   topografia Vertente -> Povoado
   topografia Talvegue
      geomorfologia_mac Bacia hidrográfica -> Rede Viária
      geomorfologia_mac Depressão tectónica -> Povoado
      geomorfologia_mac Planalto -> Rede Viária
      geomorfologia_mac Serra -> Rede Viária
   topografia 0
      classaltitude De 0 a 400 m -> Tesouro
      classaltitude De 1000 a 1500 m -> Epigrafia
      classaltitude De 400 a 700 m -> Tesouro
      classaltitude De 700 a 1000 m -> Povoado
   topografia Cume
      classaltitude De 1000 a 1500 m -> Povoado
      classaltitude Nulos
         Hierarquia hidrográfica =< 2 -> Epigrafia
```

Hierarquia hidrográfica > 2 -> Povoado
classaltitude De 0 a 400 m
Geomorfologia_mic ['Crista quartzítica' Planalto Rechã] -> Estatuária
Geomorfologia_mic Arriba -> Estatuária
Geomorfologia_mic Cabeço -> Povoado
Geomorfologia_mic Esporão -> Povoado
Geomorfologia_mic Vale -> Povoado
Geomorfologia_mic Castelo granítico
geomorfologia_mac [Planalto] -> Fortificação
geomorfologia_mac Bacia hidrográfica -> Limites
geomorfologia_mac Depressão tectónica -> Santuário
geomorfologia_mac Serra -> Fortificação
classaltitude De 400 a 700 m
Hierarquia hidrográfica =< 4 -> Povoado
Hierarquia hidrográfica > 4
geomorfologia_mac ['Depressão tectónica'] -> Rede Viária
geomorfologia_mac Bacia hidrográfica -> Fortificação
geomorfologia_mac Planalto -> Rede Viária
geomorfologia_mac Serra -> Santuário
classaltitude De 700 a 1000 m
Geomorfologia_mic [Arriba 'Crista quartzítica' Rechã] -> Povoado
Geomorfologia_mic Cabeço -> Povoado
Geomorfologia_mic Castelo granítico -> Povoado
Geomorfologia_mic Esporão -> Epigrafia
Geomorfologia_mic Vale -> Mineração
Geomorfologia_mic Planalto
Hierarquia hidrográfica =< 4 -> Povoado
Hierarquia hidrográfica > 4 -> Rede Viária

Cronologia Idade Média

topografia Talvegue -> Rede Viária
topografia Cume
classaltitude De 0 a 400 m -> Povoado
classaltitude De 400 a 700 m -> Povoado
classaltitude De 700 a 1000 m -> Fortificação
classaltitude De 1000 a 1500 m -> Fortificação
topografia null
Geomorfologia_mic [Arriba 'Castelo granítico' Esporão Planalto Rechã Vale] -> Povoado
Geomorfologia_mic Cabeço -> Necrópole
Geomorfologia_mic Crista quartzítica -> Necrópole
Geomorfologia_mic 0
Hierarquia hidrográfica =< 4 -> Povoado

```
        Hierarquia hidrográfica  > 4 -> Achado Isolado
topografia Vertente
    classaltitude De 1000 a 1500 m -> Povoado
    classaltitude De 400 a 700 m -> Povoado
    classaltitude De 0 a 400 m
        Geomorfologia_mic [Arriba 'Crista quartzítica' Planalto] -> Necrópole
        Geomorfologia_mic Cabeço -> Santuário
        Geomorfologia_mic Castelo granítico -> Povoado
        Geomorfologia_mic Esporão -> Rede Viária
        Geomorfologia_mic Rechã -> Necrópole
        Geomorfologia_mic Vale -> Povoado
    classaltitude De 700 a 1000 m
        Geomorfologia_mic [Arriba 'Castelo granítico' Esporão Rechã] -> Necrópole
        Geomorfologia_mic Cabeço -> Necrópole
        Geomorfologia_mic Crista quartzítica -> Povoado
        Geomorfologia_mic Planalto -> Necrópole
        Geomorfologia_mic Vale -> Povoado
    classaltitude Nulos
        geomorfologia_mac ['Depressão tectónica'] -> Mineração
        geomorfologia_mac Bacia hidrográfica -> Povoado
        geomorfologia_mac Planalto -> Necrópole
        geomorfologia_mac Serra -> Mineração
```

Anexo III

Regras de decisão após balanceamento por Geomorfologia_mac

Cronologia Neo/Calcolítico

- Geomorfologia_mic Arriba -> Necrópole
- Geomorfologia_mic Planalto -> Necrópole
- Geomorfologia_mic Rechã -> Necrópole
- Geomorfologia_mic 0
  - classaltitude De 0 a 400 m -> Necrópole
  - classaltitude ['De 1000 a 1500 m' 'De 700 a 1000 m'] -> Necrópole
  - classaltitude De 400 a 700 m -> Arte Rupestre
- Geomorfologia_mic Castelo granítico
  - topografia Cume -> Povoado
  - topografia Talvegue -> Povoado
  - topografia Vertente -> Povoado
- Geomorfologia_mic Crista quartzítica
  - topografia [Talvegue] -> Povoado
  - topografia Cume -> Povoado
  - topografia Vertente -> Arte Rupestre
- Geomorfologia_mic Cabeço
  - geomorfologia_mac Bacia hidrográfica -> Arte Rupestre
  - geomorfologia_mac Planalto -> Arte Rupestre
  - geomorfologia_mac Serra -> Necrópole
  - geomorfologia_mac Depressão tectónica
    - classaltitude De 0 a 400 m -> Santuário
    - classaltitude ['De 1000 a 1500 m' 'De 700 a 1000 m'] -> Necrópole
    - classaltitude De 400 a 700 m -> Achado Isolado
- Geomorfologia_mic Esporão
  - geomorfologia_mac Bacia hidrográfica -> Abrigo
  - geomorfologia_mac Planalto -> Necrópole
  - geomorfologia_mac Serra -> Arte Rupestre
  - geomorfologia_mac Depressão tectónica
    - topografia [Talvegue] -> Povoado
    - topografia Cume -> Povoado
    - topografia Vertente
      - Hierarquia hidrográfica =< 3 -> Arte Rupestre
      - Hierarquia hidrográfica > 3 -> Necrópole
- Geomorfologia_mic Vale
  - topografia Cume -> Necrópole
  - topografia Vertente -> Necrópole

topografia Talvegue
classaltitude ['De 1000 a 1500 m' 'De 700 a 1000 m'] -> Arte Rupestre
classaltitude De 400 a 700 m -> Necrópole
classaltitude De 0 a 400 m
geomorfologia_mac [0 Planalto Serra] -> Arte Rupestre
geomorfologia_mac Bacia hidrográfica -> Necrópole
geomorfologia_mac Depressão tectónica -> Arte Rupestre

Cronologia Idade do Bronze

topografia Cume -> Povoado
topografia Talvegue -> Achado Isolado
topografia Vertente -> Arte Rupestre

Cronologia Idade do Ferro

topografia Cume -> Castro
topografia Talvegue -> Castro
topografia Vertente -> Castro
topografia 0
geomorfologia_mac ['Bacia hidrográfica' 'Depressão tectónica' Serra] -> Tesouro
geomorfologia_mac Planalto -> Achado Isolado

Cronologia Período Romano

topografia Cume -> Povoado
topografia Vertente -> Povoado
topografia Talvegue
classaltitude De 0 a 400 m -> Povoado
classaltitude ['De 1000 a 1500 m' 'De 700 a 1000 m'] -> Povoado
classaltitude De 400 a 700 m -> Rede Viária

Cronologia Idade Média

topografia Talvegue -> Rede Viária
topografia 0
geomorfologia_mac ['Bacia hidrográfica' 'Depressão tectónica'] -> Necrópole
geomorfologia_mac Planalto -> Povoado
geomorfologia_mac Serra -> Necrópole
topografia Cume
Geomorfologia_mic [Arriba Rechã] -> Fortificação
Geomorfologia_mic Cabeço -> Fortificação
Geomorfologia_mic Castelo granítico -> Povoado
Geomorfologia_mic Crista quartzítica -> Fortificação

Geomorfologia_mic Esporão -> Povoado
Geomorfologia_mic Planalto -> Necrópole
Geomorfologia_mic Vale -> Povoado
topografia Vertente
Geomorfologia_mic [Arriba] -> Povoado
Geomorfologia_mic Castelo granítico -> Povoado
Geomorfologia_mic Crista quartzítica -> Povoado
Geomorfologia_mic Esporão -> Rede Viária
Geomorfologia_mic Planalto -> Mineração
Geomorfologia_mic Rechã -> Povoado
Geomorfologia_mic Vale -> Povoado
Geomorfologia_mic Cabeço
geomorfologia_mac Bacia hidrográfica -> Povoado
geomorfologia_mac Depressão tectónica -> Necrópole
geomorfologia_mac Planalto -> Necrópole
geomorfologia_mac Serra
Hierarquia hidrográfica =< 1 -> Mineração
Hierarquia hidrográfica > 1 -> Povoado

Anexo IV

Regras de decisão após balanceamento por Topografia

Cronologia Neo/Calcolítico

- Geomorfologia_mic Arriba -> Necrópole
- Geomorfologia_mic Esporão -> Arte Rupestre
- Geomorfologia_mic Planalto -> Necrópole
- Geomorfologia_mic Rechã -> Necrópole
- Geomorfologia_mic Cabeço
  - topografia Cume -> Povoado
  - topografia Talvegue -> Arte Rupestre
  - topografia Vertente -> Necrópole
- Geomorfologia_mic Castelo granítico
  - topografia Cume -> Povoado
  - topografia Talvegue -> Necrópole
  - topografia Vertente -> Povoado
- Geomorfologia_mic Crista quartzítica
  - topografia [Talvegue] -> Povoado
  - topografia Cume -> Povoado
  - topografia Vertente -> Abrigo
- Geomorfologia_mic Vale
  - Hierarquia hidrográfica =< 1
    - classaltitude De 0 a 400 m -> Abrigo
    - classaltitude ['De 1000 a 1500 m' 'De 700 a 1000 m'] -> Abrigo
    - classaltitude De 400 a 700 m -> Achado Isolado
  - Hierarquia hidrográfica > 1
    - classaltitude De 1000 a 1500 m -> Arte Rupestre
    - classaltitude De 400 a 700 m -> Necrópole
    - classaltitude De 700 a 1000 m
      - Hierarquia hidrográfica =< 3 -> Povoado
      - Hierarquia hidrográfica > 3 -> Necrópole
    - classaltitude De 0 a 400 m
      - Hierarquia hidrográfica =< 3
        - geomorfologia_mac [0 Planalto Serra] -> Arte Rupestre
        - geomorfologia_mac Depressão tectónica -> Arte Rupestre
        - geomorfologia_mac Bacia hidrográfica
          - Hierarquia hidrográfica =< 2 -> Povoado
          - Hierarquia hidrográfica > 2 -> Arte Rupestre
      - Hierarquia hidrográfica > 3 -> Necrópole

Cronologia Idade do Bronze

topografia [Vertente] -> Povoado
topografia Cume -> Povoado
topografia Talvegue -> Arte Rupestre

Cronologia Idade do Ferro

topografia Cume -> Castro
topografia Talvegue -> Castro
topografia Vertente -> Castro
topografia null
Hierarquia hidrográfica =< 1
geomorfologia_mac ['Bacia hidrográfica' 'Depressão tectónica' Serra] -> Tesouro
geomorfologia_mac Planalto -> Achado Isolado
Hierarquia hidrográfica > 1 -> Castro

Cronologia Período Romano

topografia Cume -> Povoado
topografia Vertente -> Povoado
topografia Talvegue
classaltitude ['De 1000 a 1500 m' 'De 700 a 1000 m'] -> Rede Viária
classaltitude De 400 a 700 m
Hierarquia hidrográfica =< 1 -> Povoado
Hierarquia hidrográfica > 1 -> Rede Viária
classaltitude De 0 a 400 m
Hierarquia hidrográfica =< 3
geomorfologia_mac [Planalto] -> Povoado
geomorfologia_mac Bacia hidrográfica -> Rede Viária
geomorfologia_mac Depressão tectónica -> Povoado
geomorfologia_mac Serra -> Povoado
Hierarquia hidrográfica > 3 -> Rede Viária
topografia null
Geomorfologia_mic [Arriba 'Crista quartzítica' Planalto Rechã] -> Epigrafia
Geomorfologia_mic Cabeço -> Tesouro
Geomorfologia_mic Castelo granítico -> Epigrafia
Geomorfologia_mic Vale -> Povoado
Geomorfologia_mic Esporão
geomorfologia_mac ['Depressão tectónica' Serra] -> Rede Viária
geomorfologia_mac Planalto -> Rede Viária
geomorfologia_mac Bacia hidrográfica
Hierarquia hidrográfica =< 2 -> Tesouro

Hierarquia hidrográfica > 2 -> Epigrafia
Geomorfologia_mic null
classaltitude De 0 a 400 m -> Tesouro
classaltitude De 1000 a 1500 m -> Epigrafia
classaltitude De 400 a 700 m -> Tesouro
classaltitude De 700 a 1000 m -> Povoado

Cronologia Idade Média

topografia Talvegue -> Rede Viária
topografia Cume
Hierarquia hidrográfica =< 4 -> Fortificação
Hierarquia hidrográfica > 4 -> Povoado
topografia null
Geomorfologia_mic [Arriba 'Castelo granítico' Esporão Planalto Rechã Vale] -> Povoado
Geomorfologia_mic Cabeço -> Necrópole
Geomorfologia_mic Crista quartzítica -> Necrópole
Geomorfologia_mic null
Hierarquia hidrográfica =< 4 -> Povoado
Hierarquia hidrográfica > 4 -> Achado Isolado
topografia Vertente
geomorfologia_mac Bacia hidrográfica -> Necrópole
geomorfologia_mac Depressão tectónica -> Povoado
geomorfologia_mac Serra
Hierarquia hidrográfica =< 1 -> Mineração
Hierarquia hidrográfica > 1 -> Povoado
geomorfologia_mac Planalto
Geomorfologia_mic [Arriba 'Castelo granítico' Planalto Rechã] -> Povoado
Geomorfologia_mic Crista quartzítica -> Necrópole
Geomorfologia_mic Esporão -> Necrópole
Geomorfologia_mic Vale -> Povoado
Geomorfologia_mic Cabeço
classaltitude De 0 a 400 m -> Necrópole
classaltitude De 400 a 700 m -> Povoado
classaltitude De 700 a 1000 m -> Necrópole
classaltitude De 1000 a 1500 m -> Necrópole

Anexo V

Regras de decisão após balanceamento por Cronologia

Cronologia Neo/Calcolítico

- Geomorfologia_mic Arriba -> Necrópole
- Geomorfologia_mic Planalto -> Necrópole
- Geomorfologia_mic Vale -> Necrópole
- Geomorfologia_mic Castelo granítico
  - topografia Cume -> Povoado
  - topografia Talvegue -> Povoado
  - topografia Vertente -> Povoado
- Geomorfologia_mic Rechã
  - Hierarquia hidrográfica =< 2 -> Achado Isolado
  - Hierarquia hidrográfica > 2 -> Necrópole
- Geomorfologia_mic null
  - geomorfologia_mac Bacia hidrográfica -> Necrópole
  - geomorfologia_mac [Planalto Serra] -> Necrópole
  - geomorfologia_mac Depressão tectónica
    - classaltitude ['De 0 a 400 m' 'De 1000 a 1500 m' 'De 700 a 1000 m'] -> Necrópole
    - classaltitude De 400 a 700 m -> Abrigo
- Geomorfologia_mic Cabeço
  - geomorfologia_mac 0 -> Arte Rupestre
  - geomorfologia_mac Depressão tectónica -> Achado Isolado
  - geomorfologia_mac Planalto -> Arte Rupestre
  - geomorfologia_mac Serra -> Necrópole
  - geomorfologia_mac Bacia hidrográfica
    - topografia [Talvegue] -> Arte Rupestre
    - topografia Cume -> Necrópole
    - topografia Vertente -> Arte Rupestre
- Geomorfologia_mic Crista quartzítica
  - topografia [Talvegue] -> Povoado
  - topografia Cume -> Povoado
  - topografia Vertente
    - Hierarquia hidrográfica =< 3 -> Abrigo
    - Hierarquia hidrográfica > 3 -> Arte Rupestre
- Geomorfologia_mic Esporão
  - classaltitude De 0 a 400 m -> Povoado
  - classaltitude De 400 a 700 m
    - geomorfologia_mac Bacia hidrográfica -> Arte Rupestre

geomorfologia_mac Depressão tectónica -> Arte Rupestre
geomorfologia_mac Planalto -> Necrópole
geomorfologia_mac Serra -> Arte Rupestre
classaltitude De 700 a 1000 m -> Povoado
classaltitude ['De 1000 a 1500 m' Nulos] -> Arte Rupestre

Cronologia Idade do Bronze

topografia Talvegue
Hierarquia hidrográfica =< 1 -> Achado Isolado
Hierarquia hidrográfica > 1 -> Arte Rupestre
topografia Cume
Geomorfologia_mic [Planalto Rechã] -> Povoado
Geomorfologia_mic Arriba -> Povoado
Geomorfologia_mic Castelo granítico -> Povoado
Geomorfologia_mic Crista quartzítica -> Povoado
Geomorfologia_mic Esporão -> Povoado
Geomorfologia_mic Vale -> Necrópole
Geomorfologia_mic Cabeço
classaltitude ['De 0 a 400 m' Nulos] -> Povoado
classaltitude De 1000 a 1500 m -> Povoado
classaltitude De 400 a 700 m -> Povoado
classaltitude De 700 a 1000 m -> Necrópole
topografia Vertente
classaltitude De 1000 a 1500 m -> Achado Isolado
classaltitude De 400 a 700 m -> Povoado
classaltitude De 700 a 1000 m -> Necrópole
classaltitude De 0 a 400 m
Geomorfologia_mic [Arriba 'Castelo granítico' 'Crista quartzítica' Planalto Rechã Vale] -> Arte Rupestre
Geomorfologia_mic Cabeço -> Arte Rupestre
Geomorfologia_mic Esporão -> Achado Isolado

Cronologia Idade do Ferro -> Castro

Cronologia Período Romano

topografia Talvegue -> Rede Viária
topografia Vertente -> Povoado
topografia Cume -> Povoado

Cronologia Idade Média

topografia Talvegue -> Rede Viária
topografia Cume
Geomorfologia_mic [Arriba Rechã] -> Fortificação
Geomorfologia_mic Cabeço -> Fortificação
Geomorfologia_mic Castelo granítico -> Fortificação
Geomorfologia_mic Crista quartzítica -> Povoado
Geomorfologia_mic Planalto -> Necrópole
Geomorfologia_mic Esporão
Hierarquia hidrográfica =< 2 -> Fortificação
Hierarquia hidrográfica > 2 -> Povoado
Geomorfologia_mic Vale
classaltitude De 0 a 400 m -> Povoado
classaltitude De 700 a 1000 m -> Limites
classaltitude ['De 1000 a 1500 m' 'De 400 a 700 m'] -> Povoado
topografia Vertente
Hierarquia hidrográfica =< 1
geomorfologia_mac Bacia hidrográfica -> Povoado
geomorfologia_mac Depressão tectónica -> Necrópole
geomorfologia_mac Planalto -> Necrópole
geomorfologia_mac Serra -> Mineração
Hierarquia hidrográfica > 1 -> Povoado