Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 220 (2023) 880–885

Procedia Computer Science

www.elsevier.com/locate/procedia

The 4th International Workshop on Hospital 4.0 (Hospital), March 15-17, 2023, Leuven, Belgium

Rule-based System for Effective Clinical Decision Support

Beatriz Silva[a], Francini Hak[a,*], Tiago Guimarães[a], Maria Manuel[b], Manuel Filipe Santos[a]

[a]ALGORITMI/LASI Research Center, University of Minho, Guimarães, Portugal
[b]Centro Hospitalar Universitário do Porto, Porto, Portugal

Abstract

Clinical Decision Support Systems (CDSS) are being increasingly requested and are an important role in health units. Due to the high number of data produced daily, it is necessary that these data are stored and manipulated in order to acquire knowledge to assist the decision-making processes. Representing knowledge in knowledge-based systems is one of the main tasks for achieving an effective CDSS. In this way, this narrative literature review article intends to identify different approaches to represent knowledge for rule-based CDSS. Four models are described, namely decision tables, decision trees, bayesian network and nearest neighbors, emphasizing the first two.

Keywords: Clinical Decision Support; Rule-based Systems; Knowledge Representation; Decision Table; Decision Tree.

1. Introduction

Knowledge is developed by humans through the acquisition of their thoughts and is represented through data that is not processed, analyzed, or organized [8]. The birth of the concept of Knowledge Representation occurred in the field of Artificial Intelligence, but lately, it is not the only area that has been exploring this topic [2].

Clinical Decision Support Systems (CDSS) are a way to support individual patient decisions in the healthcare environment to ensure quality in the services provided. These systems are organized into two types, knowledge-based systems, and non-knowledge-based systems. The use of CDSS in health information systems brings advantages in reducing clinical errors and preventing problems involving patients [1]. Rule-based systems are a type of knowledge-based systems that store and manipulate data extracted from other systems in order to acquire contextual knowledge.

However, the literature review shows that many systems are not considered effective due to a lack of structure and standardization in the knowledge base of decision support systems [6]. It is important to understand the different ways of representing knowledge and to understand the choices that rule-based systems have. Therefore, this article aims to

* Corresponding author.
E-mail address: francini.hak@algoritmi.uminho.pt

10.1016/j.procs.2023.03.119

review four models of knowledge representation applied to an example of clinical rules of diabetes in order to make a comparative analysis of the characteristics of each one.

The structure of this document begins with a brief introduction to the contextualization of the theme. Next, a background is presented as a way to support the definition of the key concepts of this paper. The third section contains the literature review of Knowledge Representation as well as its associated types, giving more emphasis on decision tables and decision trees. In the penultimate section, a view is given on the information gathered through the literature review and proposed future work. Finally, the conclusion represents the last part of the paper that aims to conclude based on what has been developed in the paper.

2. Background

According to Marcos et al. (2013) [7], Clinical Decision Support Systems (CDSS) can be defined as ”any computer program designed to assist health professionals in clinical decision making”. In this program, information on each patient's characteristics is matched to a knowledge base, resulting in patient-specific recommendations that health professionals can use to make a decision [14]. Therefore, CDSS needs to communicate with health information systems that contain electronic records so that they can share the necessary patient data for the decision process [7].

Decision Support Systems are classified into two types: knowledge-based and non-knowledge-based systems [14]. In general, non-knowledge-based systems utilize techniques and algorithms of machine learning or artificial intelligence for learning and prediction in decision-making. In these systems, the data do not require clinical knowledge for clinical decision processing [14]. The techniques use the patient's historical data to identify patterns [1].

On the other hand, knowledge-based systems usually have a knowledge base capable of generating rule statements through the data collected by information systems [1]. Rule-based systems or expert systems are an approach to create a knowledge-based CDSS, where data is represented in a rule format and is evaluated, obtaining a result at the end [14]. Similar to first-order logic, rule-based systems capture knowledge with conditions in logical systems, as can be organized, for example, in ”IF/THEN” statements [1]. In healthcare domain, rule-based CDSS are a positive strategy as the result that comes from the rule can be used as a recommendation or an alert, so that the health professional can make a decision according to that result.

3. Knowledge representation

The definition of knowledge can have various concepts depending on the context it is applied [2]. Mohajan (2016) [8] define Knowledge as ”a collection of experiences, appropriate information, and expertise that provides a framework for estimating and integrating new experiences and information”.

Knowledge Representation (KR) can be related to different techniques that are interconnected, including logic, ontology, and computation [2]. The classification of KR has several perspectives depending on the authors that present this concept. According to author Freitas (2014) [5], KR classification can be fit into five types, these being decision tables, classification rules, decision trees, Bayesian Network classification, and Nearest Neighbors algorithm.

However, the authors Chandrasegaran et al. (2013) [2] group KR classification into five categories: Pictorial, Symbolic, Linguistic, Virtual, and Algorithmic. After that, they organize the aforementioned forms of representation by categories, as shown in Table 1. In this article some types of KR will be presented next, but the main focus is on decision tables and decision trees.

Table 1. Examples of each category of Knowledge Representation [2].

Pictorial	Symbolic	Linguistic	Virtual	Algorithmic
Charts	Decision Tables	Condition-based Design Rules	Simulation	Computer Algorithms
Sketches	Assembly Tree	Customer Requirements	Virtual Reality	Mathematical Equations

3.1. Decision Tables

Decision Tables are a form of KR that contains conditions associated with one or more decisions, thus being considered a special table [16]. A set of conditions is associated with one or more actions, these being the decisions [15]. This table has columns with various attributes being the conditions and its rows represent objects, that is, the condition value [16].

In healthcare, this type of table helps to answer questions from health professionals, since these tables contain clinical knowledge and experience, and also support the diagnosis of clinical guidelines [11]. Therefore, clinical procedures can be represented by Decision Tables and can appear in different structures [10]. In Figure 1 it is presented an example of a Decision Table structure.

		Rules			
		R1	R2	...	Rn
Conditions	C1				
	C2				
	...				
	Cn				
Actions	A1				
	A2				
	...				
	An				

Fig. 1. Structure of Decision Table.

The Decision Table can be divided into two parts. The first part refers to Conditions, the second part to Actions, and the junction of the two parts represents the set of rules. The two parts of the Decision Table are similar and what only changes is the context of the part [12]. According to Graham Witt (2012) [15], Decision Tables can have two types: limited-entry and extended-entry, as shown in Figure 2.

		Rules			
		R1	R2	R3	R4
Conditions	Age > 30	Y	Y	Y	N
	Age <= 30	N	N	N	Y
	Insulin >= 112	Y	N	N	-
	Insulin < 112	N	Y	Y	-
	Glucose >= 128	-	Y	N	-
	Glucose < 128	-	N	Y	-
Actions	Diabetes = positive			X	
	Diabetes = negative	X	X		X

		Rules			
		R1	R2	R3	R4
Conditions	Age	> 30	> 30	> 30	<= 30
	Insulin	>= 112	< 112	< 112	-
	Glucose	-	>= 128	< 128	-
Actions	Diabetes	negative	negative	positive	negative

Fig. 2. Two types of Decision Table: limited-entry and extended-entry, respectively.

The type limited-entry, on the first part of the table, presents the conditions in full in the first column (Glucose >= 128) and in the following columns their Boolean values, being true (Y), false (N), or not applicable (-). In the second part, actions are symbolized by an X if that action is performed according to the above conditions [15]. On the other hand, in the case of the second type, extended-entry, the condition is broken down in the table, that is, the condition variable is presented in the first column and in the following columns its associated value. The same case happens in actions, the first column presents the variable and in the following columns the value [15].

3.2. Decision Trees

Another way to represent knowledge is by using Decision Trees. Through the presentation of a graphical structure, it is possible to recognize the structure of a Decision Tree. For this, it should have variables relevant to the analysis and its hierarchy should be in a tree format [5].

For Jijo Mohsin Abdulazeez (2021) [3], this form of representation is a model based on a tree structure with a hierarchy of knowledge relationships. Thus, the representation of the data starts at the root of the tree and the data is separated until it reaches a node at the end of the tree, presenting a result. In Figure 3 a) it is demonstrated a Decision Tree structure.

There are three possible node types within the tree's structure [13]:

- Root Node, which represents the first step in creating the tree and checking the condition of this node can generate one or more subdivisions.
- Decision Node or Internal Node, which can have a link to a parent node represented above the node and links to one or more child nodes that are below the node.
- Leaf Node, these being considered as the last nodes of the link and which form a combination of conditions ultimately originating the results.

The connection of nodes in the tree is represented by branches, which can connect root nodes to decision nodes as well as decision nodes to leaf nodes. The path the Decision Tree follows depends on the decision rules that are contained in the branches of the tree [13], as shown in Figure 3 b).

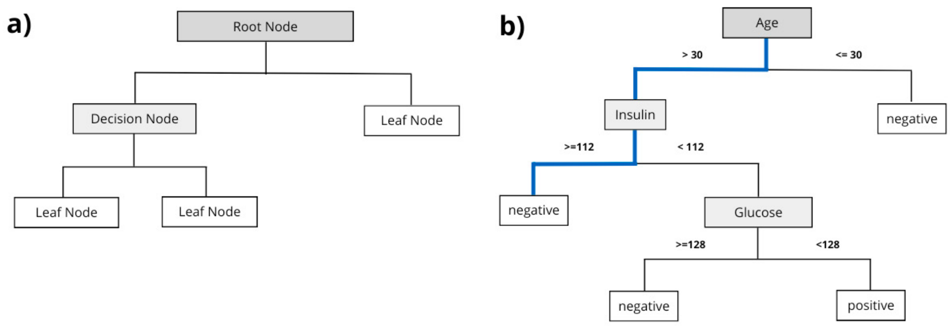

Fig. 3. a) Structure of Decision Tree; b) Example of a path in the Decision Tree.

3.3. Bayesian Network

The Bayesian Network classification has as its key point the probability based on Naive Bayes as a way to analyze all attributes instead of a set of attributes as it is in the case of the decision tree [5]. The attributes that this algorithm uses are data from past experiences, and in case there is missing information to be analyzed, it also uses data with expert knowledge [9]. In Figure 4 a) it is demonstrated an example of Bayesian Networks.

This classification is based on the deep belief network model to "combine static network with temporal information" [9]. The implementation of the Bayesian Network classification may arise limitations with regard to network connection and probability estimation [9]. Although this limitation exists, the use of this classification brings an advantage to medicine, that being, the identification of "natural" decisions due to the use of the information that is made available for this process [5].

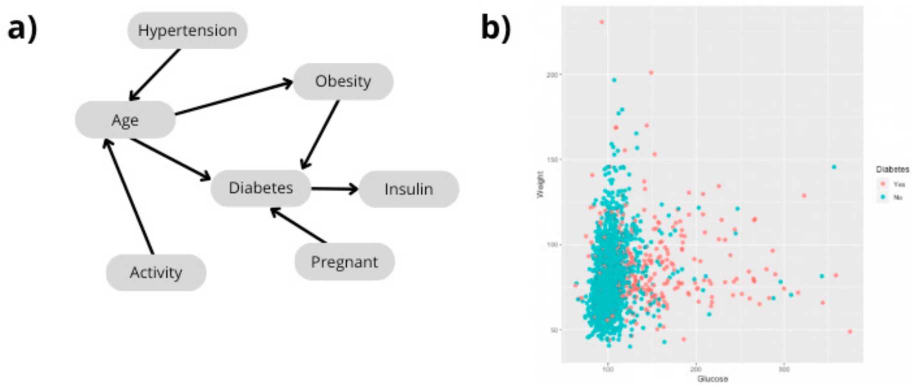

Fig. 4. a) Example of a Bayesian Network; b) Example of a Nearest Neighbors (taken from [4]).

3.4. Nearest Neighbors

The Nearest Neighbor algorithm is based on explaining the classification of an instance, that is, it uses nearest neighbor attributes to determine the class of the new instance. As opposed to decision trees, it is presented several different results when running this algorithm [5]. In Figure 4 b) it is demonstrated an example of Nearest Neighbors.

Briefly, this algorithm is related to Machine Learning and uses the shortest distance criterion to identify the class of the new instance. In addition, points are assigned to the neighboring attributes and the one containing the most points represents the class of that instance [9]. A weakness of this algorithm lies in the fact that a large amount of training data cannot be used because computational complexity is increased due to this factor [9].

4. Discussion and Future Work

Based on the literature review, in recent years Knowledge Representation for rule-based systems has been a strong topic in the healthcare, with representations being the most widely used: ontology, semantic web, decision tables, decision rules, logic and probability methods inserting here Bayesian Network classification and decision trees [11].

After reviewing the four techniques mentioned above based on an example of a clinical case about diabetes, we believe that representing rules in the form of Decision Tables or Decision Trees is an advantage when it comes to expressing their information through Knowledge Representation. Through the examples, the use of the two models are more favorable at the moment of direct contact with the patient because they are easier representations and the result becomes more intuitive. Thus, this paper focused on these two methods due to the ease and simplicity of Knowledge Representation.

Furthermore, since the goal is to understand how rule-based systems are embedded in Knowledge Representation, these models are examples that use conditions as a way of representation. In healthcare, as clinical concepts contain dependent variables, the most suitable Knowledge Representation structures would be the use of decision tables and decision trees. For future work, with ontology being a form of representation, we seek to focus on this representation of knowledge more properly through open standards models so that they are semantically readable.

5. Conclusions

The adoption of Clinical Decision Support Systems is increasingly requested in all sectors, especially in the health field. The amount of data produced daily requires processes capable of storing and managing these data so that they are manipulated in order to produce knowledge. Through a narrative literature review, this article allowed for the identification of possible models for knowledge representation, with a focus on knowledge-based systems referring to rule-based systems.

In this article, only four types of Knowledge Representation were presented. The focus was on the detailed description in the decision table and decision tree models, as these are the techniques that are most closely linked to rule-based systems. These two techniques encompass sets of conditions and actions to represent one or several actions. We believe that these two approaches are essential for building a rule-based CDSS due to the possibility of creating independent rules, rule sets, and chained rules in the same context. Thus, it is for future work to explore these two models applied to case studies in different clinical contexts, in order to evaluate the performance of each one.

Acknowledgements

Francini Hak thanks the Fundação para a Tecnologia (FCT) for the grant 2021.06230.BD.

References

[1] Eta Berner and Yang Gong. *Clinical Decision Support Systems: Theory and Practice*. Jan. 2016. ISBN: 978-3-319-31911-7. DOI: 10.1007/978-3-319-31913-1.

[2] Senthil K. Chandrasegaran et al. "The evolution, challenges, and future of knowledge representation in product design systems". In: *Computer-Aided Design* 45.2 (2013). Solid and Physical Modeling 2012, pp. 204–228. ISSN: 0010-4485. DOI: https://doi.org/10.1016/j.cad.2012.08.006. URL: https://www.sciencedirect.com/science/article/pii/S0010448512001741.

[3] Bahzad Charbuty and Adnan Abdulazeez. "Classification Based on Decision Tree Algorithm for Machine Learning". In: *Journal of Applied Science and Technology Trends* 2 (01 Mar. 2021), pp. 20–28. DOI: 10.38094/jastt20165. URL: https://jastt.org/index.php/jasttpath/article/view/65.

[4] Anisa Dhana. *K-nearest neighbor for prediction of diabetes in NHANES*. 2019. URL: https://datascienceplus.com/k-nearest-neighbor-for-prediction-of-diabetes-in-nhanes/ (visited on 02/24/2019).

[5] Alex A. Freitas. "Comprehensible Classification Models: A Position Paper". In: *SIGKDD Explor. Newsl.* 15.1 (Mar. 2014), pp. 1–10. ISSN: 1931-0145. DOI: 10.1145/2594473.2594475. URL: https://doi.org/10.1145/2594473.2594475.

[6] F. Hak, T. Guimaraes, and M. Santos. "Towards effective clinical decision support systems: A systematic review". In: *PLoS ONE* 17 (8 August 2022). ISSN: 19326203. DOI: 10.1371/journal.pone.0272846.

[7] Mar Marcos et al. "Interoperability of clinical decision-support systems and electronic health records using archetypes: A case study in clinical trial eligibility". In: *Journal of Biomedical Informatics* 46.4 (2013), pp. 676–689. ISSN: 1532-0464. DOI: https://doi.org/10.1016/j.jbi.2013.05.004. URL: https://www.sciencedirect.com/science/article/pii/S1532046413000701.

[8] Haradhan Mohajan. "Knowledge is an Essential Element at Present World". In: *International Journal of Publication and Social Studies* 11 (Nov. 2016), pp. 31–53. DOI: 10.18488/journal.135/2016.1.1/135.1.31.53.

[9] Norazwan Md Nor, Che Rosmani Che Hassan, and Mohd Azlan Hussain. "A review of data-driven fault detection and diagnosis methods: applications in chemical process systems". In: *Reviews in Chemical Engineering* 36.4 (2020), pp. 513–553. DOI: doi:10.1515/revce-2017-0069. URL: https://doi.org/10.1515/revce-2017-0069.

[10] David Riaño, Richard Lenz, and Manfred Reichert. *Knowledge Representation for Health Care: HEC 2016 International Joint Workshop, KR4HC/ProHealth 2016, Munich, Germany, September 2, 2016, Revised Selected Papers*. Jan. 2017. ISBN: 978-3-319-55013-8. DOI: 10.1007/978-3-319-55014-5.

[11] David Riaño, Mor Peleg, and Annette ten Teije. "Ten years of knowledge representation for health care (2009–2018): Topics, trends, and challenges". In: *Artificial Intelligence in Medicine* 100 (2019), p. 101713. ISSN: 0933-3657. DOI: https://doi.org/10.1016/j.artmed.2019.101713. URL: https://www.sciencedirect.com/science/article/pii/S0933365719304531.

[12] Pavel V. Senchenko and Yury P. Ekhlakov. "Use of Decision Tables in Monitoring of Performance Discipline". In: *2019 IEEE 13th International Conference on Application of Information and Communication Technologies (AICT)*. 2019, pp. 1–4. DOI: 10.1109/AICT47866.2019.8981795.

[13] Yan-Yan Song and Ying Lu. "Decision tree methods: applications for classification and prediction". In: *Shanghai archives of psychiatry* 27 (Apr. 2015), pp. 130–135. DOI: 10.11919/j.issn.1002-0829.215044.

[14] Reed T Sutton et al. "An overview of clinical decision support systems: benefits, risks, and strategies for success". In: *npj Digital Medicine* 3 (1 2020), p. 17. ISSN: 2398-6352. DOI: 10.1038/s41746-020-0221-y. URL: https://doi.org/10.1038/s41746-020-0221-y.

[15] Graham Witt. "Chapter 3 - A brief history of rules". In: *Writing Effective Business Rules*. Ed. by Graham Witt. Boston: Morgan Kaufmann, 2012, pp. 25–63. ISBN: 978-0-12-385051-5. DOI: https://doi.org/10.1016/B978-0-12-385051-5.00003-3. URL: https://www.sciencedirect.com/science/article/pii/B9780123850515000033.

[16] Yiyu Yao and Xianyong Zhang. "Class-specific attribute reducts in rough set theory". In: *Information Sciences* 418-419 (2017), pp. 601–618. ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2017.08.038. URL: https://www.sciencedirect.com/science/article/pii/S002002551730885X.