Using Domain Knowledge to Improve Intelligent Decision Support in Intensive Medicine – A Study of Bacteriological Infections

Rui Veloso[1], Filipe Portela[1], Manuel Filipe Santos[1], Álvaro Silva[2], Fernando Rua[2], António Abelha [1], José Machado [1]
[1]*Algoritmi Centre, University of Minho, Guimarães, Portugal*
[2]*Serviço Cuidados Intensivos, Centro Hospitalar do Porto, Hospital Santo António, Porto, Portugal*
ruiveloso51046@gmail.com, {cfp, mfs}@dsi.uminho.pt, moreirasilva@gmail.com, fernandorua.sci@hgsa.min-saude.pt, {abelha, machado}@di.uminho.pt

Keywords: Antibiotics, Therapies, Infections, Bacteria, Intensive Care Units, Heuristics, Artificial Intelligence, INTCare, Decision Support

Abstract: Nowadays antibiotic prescription is object of study in many countries. The rate of prescription varies from country to country, without being found the reasons that justify those variations. In intensive care units the number of new infections rising each day is caused by multiple factors like inpatient length of stay, low defences of the body, chirurgical infections, among others. In order to complement the support of the decision process about which should be the most efficient antibiotic it was developed a heuristic based in domain knowledge extracted from biomedical experts. This algorithm is implemented by intelligent agents. When an alert appear on the presence of a new infection, an agent collects the microbiological results for cultures, it permits to identify the bacteria, then using the rules it searches for a role of antibiotics that can be administered to the patient, based on past results. At the end the agent presents to physicians the top-five sets and the success percentage of each antibiotic. This paper presents the approach proposed and a test with a particular bacterium using real data provided by an Intensive Care Unit.

1 INTRODUCTION

Infections are the principal cause of mortality in Intensive Care Units (ICU) both in Europe (Angus et al., 2001) and USA (Vicent et al., 2006). According to the medical community this type of problem is most common in patient with more than five days of stay.

Vicent et al. (2009) conducted a study about infections in the ICU. In this study it was evaluated a set of patient where 50% of the patients had an infection. From the infected patients, in 70% of the cases were prescribed therapeutics associated with antibiotics and microbiological cultures being in generally obtained positive results.

The antibiotic prescription varies pretty much from country to country, being the reason unknown (Lindbaek, 2006).

In many cases the patient body resists to the administered antibiotic, being necessary to test and prescribe another antibiotic. The main purpose of this work is to explore a complementary approach in order to support the decision process related to infections. This work aims to answer to the question *"Is it possible to use an intelligent approach in order to support the infection decision process?"*

This paper introduces an algorithm based in the heuristic concept to provide information about alternative antibiotics that can have success in control of a particularly infection.

The algorithm will be embedding in an intelligent decision support system for intensive medicine – INTCare system. The system is based on agents responsible to collect and process the data in real-time. The heuristic aims for scanning new infections or patterns of infections. Whenever an infection is found the system automatically runs the algorithm returning to the clinical staff a list of possible antibiotics that can be administered to the patient.

The algorithm is responsible for finding the best treatments for a specific infection. The system searches all the treatments performed in the past and based on the patient clinical data, admission variables like age and sex, and shows some alternative treatments that can achieve success in the combat of the infection. This research also considers factors like the cost, successful cases and expected time to the antibiotic produce effect.

The system returns then a set of treatments successfully applied in the past using input variables similar to the patient data (case based approach).

To test this system they were used real data provided by Intensive Care Unit of Centro Hospitalar do Porto (CHP), Porto, Portugal. All of this work is inserted in the INTCare research project.

This article is divided into five sections. The first one consists in the introduction of the paper. The main concepts and related work are described in background that is the section number two of this paper. On the third section it is described the data analysis, the data transformations and it is presented the heuristic algorithm and achieved results with the application of the heuristic. Then in the section four it is analysed and discussed the results and on the last section they are presented some considerations as well the future work.

2 BACKGROUND

2.1 Infection, Bacteria and Antibiotics in Healthcare

Among bacteria, viruses, fungi and parasites, a very small amount of these organisms are linked to the development of infections, usually called pathogens. An infectious disease is normally characterized by symptoms like fever, pain and swelling and others more specific to the organs that the infection is affecting. The diagnosis of an infection is done recurring to the isolation of a pathogen and grown of them in artificial cultures. So these pathogens can be identified and then it is administrated to the patient the most convenient antibiotic (Ryan et al., 2004).

In Europe in each year about 4 million of patients acquire an infection associated to healthcare and approximately 37000 of them dies from the contracted infection. The most common infections in healthcare are: respiratory, urinary, surgical, gastrointestinal and bloodstream (ECDC, 2012).

Bacteria are microorganisms; the smallest among the living cells, without nucleus but all of them have the nucleic acid and protein synthesis (Ray et al., 2004).

The most common bacteria in healthcare are *Escherichia Coli*, *Pseudomonas aeruginosa*, *Candida* and *Enterococcus* species (ECDC, 2012). The *Pseudomonas aeruginosa* is a common bacteria from the Gram-negative class, it is a common pathogen in inpatients with a length of stay superior to one week. The infections provoked by this pathogen are many and can be located on respiratory track (pneumonia), heart (endocarditis), bloodstream (bacteraemia) among others (Ray et al., 2004).

2.2 Heuristics

The term heuristic comes from the Greek and means “to find out, discover”. Romanycia and Pelletier (1985) created a survey of definitions and concluded that a heuristic is a device that is used in problem solving. It can be a strategy, knowledge, computer program, data structure and have to guarantee the supply of a solution. The heuristic developed in this work is based on the medical domain knowledge about antibiotics and infections in intensive medicine.

2.3 INTCare

This study is being developed under the research project INTCare. INTCare is an Intelligent Decision Support System (IDSS) that is in constant development and testing. It is deployed in the ICU of the Centro Hospitalar do Porto (CHP). This IDSS is based on intelligent agents (Santos et al., 2011) and aims to predict clinical events as patient organ failure (cardiovascular, respiratory, renal, hepatic, neurological and hematologic systems), possible readmissions, diseases, infections and patient outcome (Portela et al., 2013). The system is able to suggest procedures, treatments and therapies. This system is based in four subsystems (data acquisition, knowledge management, inference and interface) that recur to intelligent agents in order to take actions (Portela et al., 2013).

3 STUDY DESCRIPTION

3.1 Data Understanding

As above referred the administration of antibiotics is an important issue because there is a great number of particularities that should be analysed in the treatment of infections. The system developed for this study tries to eliminate some of those issues. In order to analyse and identify treatment patterns it is necessary to provide input variables for the system. After some meetings with the medical staff of the CHP it was defined a fixed variables package.

Independently the type of infection, it should be always considered four type of variables. They are age, sex, leucocytes and days of internment. The remaining variables are related with the organic system that the infection is affecting: cardiovascular, respiratory, hepatic, renal, neurological and hematologic. For the study it was considered infections provoked by the Pseudomonas aeruginosa a bacteria that typically affects the respiratory system. To induce the models they were used four variables related with this system: SPO2, PCO2 and PaO2.

Table 1 shows the used variables for this study on the first part of the heuristic algorithm and a brief description of each one.

Table 1: Variables Used for Search

Variable	Description
age	The patient age in years
sex	Information about the patient sex. Can be male or female.
leuc	Leucocytes – quantity of leucocytes referred in the laboratory analysis with the closest date of the date in where the infection appears.
doi	Length of stay – the number of days that the patient is hospitalized in the ICU.
SPO2	Peripheral capillary oxygen saturation – is an estimation of the oxygen saturation level. Normal values are considered 95% to 100%. Between 90% and 95% the oxygen saturation is low but not necessarily represents a health issue and below 90% is considered that a patient is in hypoxemia.
PaCO2	Partial pressure of carbon dioxide – Represents the partial pressure of CO2 in arterial blood and expresses the effectiveness of the alveolar ventilation. Normal values are between 35 to 45 mm Hg. If the value is higher than 45 mmHg the patient is in hypercapnia.
PaO2	Oxygen partial pressure – refers to the measurement of oxygen in the arterial blood. The normal range is between 75-100 mm Hg. If the value is below that the patient is not receiving enough oxygen.
PaO2/FIO2	Ratio of the oxygen partial pressure and fraction of inspired oxygen – compares the level of oxygen level in the blood with the oxygen that is breathed. It's very useful do check if exists problems with how lungs transfer the oxygen to the blood. If the ratio is below to 250 mm Hg is one of the criteria of evaluation for pneumonia.

The variables described in table 1 belong to the search variables group. In order to perform the heuristic algorithm it is necessary to have a set of data containing results of past treatments. The variables represented in this set, are the variables mentioned in table 1 plus the variables: culture, antibiotic, antibiotic result, days of treatment and cost. Table 2 shows the variables present in the set of data related with the past results.

Table 2: Variables related with past results

Variable	Description
tage, tsex, tleuc, tdoi, tSPO2, tPaCO2,tPaO2, tPaO2/FIO2	The variables represent the same results as the ones from Table 1, but it was added the prefix *t* for treatment.
culture	Represents a positive case for some bacteria. For example if the field as the text "*pseaer*" it means the data from that row represents a positive case of *pseaer* bacteria.
antibiotic	This field represents the antibiotic given to a patient in order to combat bacteria.
antibiotic result	The antibiotic result assumes two results: positive and negative. If the value is positive it means that the antibiotic succeeds and the infection is being stopped. If the value is negative, it means the antibiotic did not produce curative effects and it was necessary to try with another antibiotic.
Treatment days	Is the number of days that the antibiotic was administrated until makes effect on the combat to the infection. If the value is null it means the antibiotic was not suited for the bacteria.
cost	Represents the cost of the antibiotic administration. This value is obtained multiplying the unitary cost of the dose by the number of days of treatment.

In order to test the proposed algorithm it was used a dataset of historic treatments. This dataset only contains 275 rows referring to 42 inpatients treatments using 5 different antibiotics. The dataset size is small because at the moment there are some laboratory results in a closed format which means that we do not have access to the results of all the bacteriological analysis. However to test this approach the dataset is enough because it represents a complete sample of an infection.

3.2 Heuristic Approach

The heuristic searches for new infections and patterns that can indicate an infection. When an infection is detected it will wait for the results of cultures analysis. When the culture results denote the presence of a bacterium a search is performed for looking for possible treatments.

Next, for a better understanding, an example will be given using a real data for *pseaer* bacterium. Despite of this representation, the heuristic developed can be used in another type of infection / bacteria, by only changing the input data and output target.

An algorithm was developed for searching possible treatments that best suit the case. The historic treatments and domain knowledge are used to find the list of best treatments possible.

The system needs a dataset. These data is represented by the following groups:

Variables searched = {doi + leuc + age + sex + SPO2 + PaCO2 + PaO2 + PaO2/FIO2}

Treatment history = {tdoi + tleuc + tage + tsex + tSPO2 + tPaCO2 + tPaO2 + tPaO2/FIO2 + culture + antibiotic + antibiotic result}

An intelligent agent is constantly monitoring the respiratory variables and when a set of results occur the system emits a signal alerting for a possible infection of the patient. Then the agent waits for the microbiological cultures results to identify the bacteria.

If the bacteria verified in the exam it is *pseaer*, the agent invokes the function *treatments* - responsible to found the possible treatments. This function compares the *biological variables* from the search with the variables from *past treatments* (history).

The algorithm returns results where the search and history variables are closer using priorities in the calculation of the difference. In this part of the algorithm it can be returned the treatment data, considering other factors as the days of treatment until the antibiotic makes effect and/or the cost of the treatment.

The variables chosen to compare the results and their priorities were defined meeting with medical staff of CHP and they are:

1. SPO2;
2. doi;
3. leuc;
4. PaO2/FIO2 ratio;
5. PaO2;
6. PaCO2;
7. Age;
8. Sex;

This heuristic is represented by the following algorithm:

Algorithm - *Treatment Alternatives*

Requires: search variables, history variables

Function Evaluate Infection
  Get search variables
  If SPO2 < 90 and (PaO2 < 75 or PaCO2 > 75) and PaO2/FIO2 < 250 **then**
    For patient **do**
  There is an infection? - Wait for Cultures Results
      If *Pseaer* is positive **Then**
        Function Treatments
      Else
        Wait for positive culture results
      End if
  End if
End Function

Function Treatments
  Get search variables, history of treatment
    If antibiotic result = success
      Treatments = results ordered by minimum difference between variables
    Else
      No successful treatment
    End if
  Return Treatments
End function

For other bacteria the process of finding the best treatment is the same. For these types of infections the algorithm will change heuristic remains the same. There will be only differences on the variables used (but the two groups are the same) and on the first *if clause* of the heuristic algorithm.

3.3 Results

As above referred, to test this heuristic it was used a dataset with past treatments applied to inpatients. Since this study pretend to test the viability of this concept it was given an input so the system can search for the results which best suits the input variables. In future the agent will do this step in real-time and using online-learning.

Table 3 represents the inputs given to the algorithm – infected patient data.

Table 3: Inputs given to the algorithm

Variable	Description
doi	7
leuc	13
age	65
sex	M
SPO2	90
PaCo2	41
PaO2	71
PaO2/FIO2	180

To exemplify the algorithm let's consider a male inpatient characterized by: Age = 65 years: Length of stay=7 days; Leucocytes count = 13x10^9/; SPO2 = 90%; PaCO2=41 mm Hg; PaO2=71 mm Hg; and a PaO2/FIO2=180 mm Hg.

Given these inputs, the algorithm found 23 possible treatments. Table 4 indicates the result from the search (top 5 results). This table presents the historical variables values for each suggestion (R1-R5), the antibiotic administered, the result achieved (success or failure), the expected days for the treatment and the treatment cost.

Table 4: Algorithm output

Variable	Results				
	R1	R2	R3	R4	R5
doi	8	12	7	4	12
leuc	14.5	14.2	13.2	15.2	16.1
age	70	60	80	65	60
sex	M	M	M	M	F
SPO2	90	91	87	85	91
PaCo2	39.3	39.6	40.0	39.3	39.6
PaO2	71.5	73.3	70.8	57	55
PaO2/FIO2	181	194	196	198	198
Antibiotic	Merop	Colist	Colist	Fluco	Pipera
Antibiotic Result	Suc	Suc	Suc	Suc	Suc
Treatment Days	7	11	6	3	11
Cost	210	52	28	8	31

Along with these results the algorithm also gives information about the support level of the antibiotics. It is expressed in form of a percentage and gives an overview of the antibiotics used in the past and their success. For example, the output about the antibiotics application for this infection is:

- ✓ Meropenem (Merop) – Used successfully 5 times (26.3%); Used Unsuccessfully 14 times (73.7%)
- ✓ Colistin (Colist) – Used successfully 16 times (52%); Used Unsuccessfully 14 times (48%)
- ✓ Fluconazole (Fluco) – Used successfully 10 times (20%); Used Unsuccessfully 40 times (80%)
- ✓ Piperacillin (Pipera) – Used successfully 2 times (12.5%); Used Unsuccessfully 14 times (87.5%)

4 DISCUSSION

Analysing the obtained results (table 4) it is possible to observe that the algorithm suggest some treatment options in this case the usage of colistin, *meropenem*, *fluconazole* and *piperacillin*. So based on this information the medical staff can decide on which antibiotic should be prescribed using as base the suggestions made by the algorithm.

The approach developed will be improved and embedded into the INTCare System. The results will be available anywhere and anytime by whom have access privileges.

With this algorithm it is possible to observe a list of treatments that can produce effects in the combat of *pseaer* bacteria. By analysing this information the clinical staff can have an idea about which is the more adequate antibiotic.

The system proposes some scenarios about treatments, eliminating at the start past treatments which not produce any effect or that have very low taxes of success.

The physicians are always responsible by the final decision. The heuristic only is used to help them to take the better decision in the patient best interest.

5 CONCLUSIONS AND FUTURE WORK

This study explored an approach using the heuristic concept to develop a searching algorithm for combating infections in order to help the doctors taking better decisions and saving lives. The domain knowledge combining literature and empirical knowledge of the medical staff was crossed with patient and antibiotic data in order to obtain the rules. Additionally, this approach introduces the cost factor in the decision criteria.

Preliminary results demonstrate de utility of the approach and encourage further work in a wider scope and the integration of the heuristic in INTCare system in order to complement data mining predictive models.

The algorithm is represented by a set of rules that can be easily stored in a knowledge base and adapted to other infections.

Future work will include the development of an intelligent agent to search and use:

- other infections and bacteria;
- other patient data to improve the heuristic.

ACKNOWLEDGEMENTS

This work has been supported by FCT – Fundação para a Ciência e Tecnologia in the scope of the project: Pest-OE/EEI/UI0319/2014 and PEst-OE/EEI/UI0752/2014.

The authors would like to thank FCT (Foundation of Science and Technology, Portugal) for the financial support through the contract PTDC/EEI-SII/1302/2012 (INTCare II).

REFERENCES

Angus, D.C., Linde-Zwirble, W.T., Lidicker, J., Clermont, G., Carcillo, J., Pinsky, M.R., 2001. *Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care*. Crit. Care Med. 29, 1303–1310.

ECDC, 2012. *Surveillance of healthcare-associated infections in Europe 2007*. European Centre for Disease Prevention and Control. Surveillance Report from ECDC, Stockolm.

Lindbaek, M., 2006. *Prescribing antibiotics to patients with acute cough and otitis media*. Br. J. Gen. Pract. 56, 164–165.

Portela, F., Pinto, F., Santos, M.F., 2012. *Data mining predictive models for pervasive intelligent decision support in intensive care medicine*.

Portela, F., Santos, M.F., Machado, J., Abelha, A., Silva, Á., 2013. *Pervasive and Intelligent Decision Support in Critical Health Care Using Ensembles*, in: Bursa, M., Khuri, S., Renda, M.E. (Eds.), Information Technology in Bio- and Medical Informatics, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 1–16.

Santos, M.F., Portela, F., Vilas-Boas, M., 2011. *INTCARE : multi-agent approach for real-time intelligent decision support in intensive medicine*. SciTePress.

Ray, C. George, 1934- & Ryan, Kenneth J. (Kenneth James), 1940- & Sherris, John C & ebrary, Inc 2004, *Sherris medical microbiology: an introduction to infectious diseases*, 4th ed, McGraw-Hill, New York

Romanycia, M.H. J., Pelletier, F. J., 1985. *What is a Heuristic?*. Computational Intelligence, 1, 47-58.

Vincent, J.-L., Rello, J., Marshall, J., Silva, E., Anzueto, A., Martin, C.D., Moreno, R., Lipman, J., Gomersall, C., Sakr, Y., Reinhart, K., EPIC II Group of Investigators, 2009. *International study of the prevalence and outcomes of infection in intensive care units.* JAMA 302, 2323–2329. doi:10.1001/jama.2009.1754

Vincent, J.-L., Sakr, Y., Sprung, C.L., Ranieri, V.M., Reinhart, K., Gerlach, H., Moreno, R., Carlet, J., Le Gall, J.-R., Payen, D., S*epsis Occurrence in Acutely Ill Patients Investigators, 2006. Sepsis in European intensive care units: results of the SOAP study*. Crit. Care Med. 34, 344–353.