Advances in Informatics, Management and Technology in Healthcare
J. Mantas et al. (Eds.)
© 2022 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI220757

Discovering Association Rules in Antimicrobial Resistance in Intensive Care Unit

Aikaterini SAKAGIANNI[a,1], Georgios FERETZAKIS[b,c], Dimitris KALLES[b], Evangelos LOUPELIS[d], Zoi RAKOPOULOU[e], Ilias DALAINAS[e] and Georgios FILDISIS[f]
[a] *Sismanogleio General Hospital, Intensive Care Unit, Marousi, Greece*
[b] *School of Science and Technology, Hellenic Open University, Patras, Greece*
[c] *Sismanogleio General Hospital, Department of Quality Control, Research and Continuing Education, Marousi, Greece*
[d] *Sismanogleio General Hospital, IT department, Marousi, Greece*
[e] *Sismanogleio General Hospital, Administration, Marousi, Greece*
[f] *School of Nursing, National and Kapodistrian University of Athens, Athens, Greece*

Abstract. Multidrug resistant infections in intensive care units represent a worldwide problem with adverse health effects and negative economic implications. As artificial intelligence techniques are increasingly applied in diagnosing, treating, and preventing antimicrobial resistance, in this study, we explore the use of association rule mining in the antibiotic resistance profile of critically ill patients suffering from multidrug resistant infections.

Keywords. Antimicrobial resistance, Machine Learning, Association rules

1. Introduction and background

Nosocomial multidrug infections in the intensive care unit environment constitute the main cause of morbidity and mortality in critically ill patients [1-2]. As antimicrobial resistance (AMR) is growing the treatment options become more limited, and increased rates of adverse outcomes are observed. Serious economic implications are also induced from the evolving phenomenon of AMR.

Artificial intelligence (AI) techniques are increasingly applied in biomedical research, gradually altering medical practice [3]. Antimicrobial resistance prediction is one of the fields where these innovative methods are gaining ground.

Several recent studies present the use of machine learning in diagnosing, treating, and preventing AMR [4–6]. In a previous study [7], we evaluated five ML techniques to predict antibiotic resistance using data from the clinical microbiology laboratory, available in the Laboratory Information System (LIS) of the hospital.

---

[1] Corresponding Author, Aikaterini Sakagianni, MD, MSc, PhD(c), Sismanogleio General Hospital, Intensive Care Unit, Marousi, Greece; E-mail address: sakagianni@sismanoglio.gr.

One of the main concepts of machine learning is association rule mining (ARM), a procedure that seeks to identify repeatedly ensuing patterns, correlations, or relationships within large datasets stored in different sorts of databases. Medical diagnosis and protein sequencing are some examples where association rule mining has been employed with successful results [8-9]. Infection control surveillance is another domain where data mining techniques could be useful [10].

In this study, we explore the use of association rule mining in the antibiotic resistance profile of critically ill patients suffering from multidrug resistant infections.

2. Methods and Materials

Data from the clinical Microbiology Laboratory on patients admitted to the intensive care unit (ICU) of a Greek public tertiary hospital over a one-year period (2019) were studied. The 6,086-instance data set contains the attributes of gender (binary), bacterial species (categorical - *Acinetobacter baumannii*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*), antibiotics (categorical- 30 antibiotics), specimen (categorical - Respiratory secretions, Pus, Peritoneal (ascitic) fluid, Blood, Urine, Tissue, Catheter tip, Pleural fluid), and the binary class attribute (antimicrobial susceptibility - Resistant, Sensitive) obtained by the culture process. The following table (Table 1) summarizes the descriptive statistics of the data set used in this study. The Institutional Review board of the Sismanogleio General Hospital approved the research protocol.

Table 1. Summary statistics of the data set.

Gender	**Bacterial species**	**Specimen**			**Class**
Male (63.9%)	*A.baum.*(39.16%)	Resp.(76.75%)	Asc.(0.64%)	Blood(8.38%)	Resistant (71%)
Female (36.1%)	*K.pneum.*(35.03%)	Pus (3.91%)	Ur.(7.33%)	Cath.(2.61%)	Sensitive (29%)
	P.aerug. (25.81%)	Pleur.(0.23%)	Tissue (0.15%)		

For detecting patterns in large datasets, association rule mining is a popular unsupervised machine learning technique. Since market basket analysis is a most common application of association rules, convention has it that data is usually structured into rows and columns with one 'transaction' per row and one 'item' per column. In our case, the 'items' correspond to the categorical variables: gender, bacterial species, type of sample, antibiotics, and antimicrobial susceptibility. Each 'transaction' represents a susceptibility testing of an isolate.

A formulation of the problem of discovering association rules is as follows [11]. Let $I = \{i_1, i_2, \dots i_m)$ a set of items and $D = \{T_1, T_2, \dots T_m)$ a set of transactions. Each transaction T is a set of items such that $T \subseteq I$. If a transaction T contains X, this is denoted by $X \subseteq T$. An association rule is an “if-then” rule stating that X associates with Y, denoted by $X \Rightarrow Y$, where $X, Y \subseteq I$ and $X \cap Y = \emptyset$. The left-hand side, LHS, of a rule is called antecedent and the right-hand side, RHS, is called the consequent. Given the rule $X \Rightarrow Y$, N: the total number of transactions, the support ($supp$), the confidence ($conf$) and the $lift$ are defined as follows:

$$supp(X \cup Y) = \frac{Frequency(X, Y)}{N}$$

$$conf(X \Rightarrow Y) = \frac{Frequency(X, Y)}{Frequency(X)}$$

$$lift(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X) \times supp(Y)}$$

The percentage of transactions that contain both parts of the rule is called *support*, whereas *confidence* is the number of cases in which the rule is correct compared to the number of cases to which it applies. At the very least, an interesting rule must have *support* and *confidence* values greater than some (user-specified) minimum criteria, *minsup* and *minconf*. The *lift* value of an association rule is the ratio of the *confidence* of the rule and the *expected confidence* of the rule, which, in turn, is defined as the product of the *support* values of X and Y divided by the *support* of X. A *lift* value greater than one indicates that X and Y appear more often together than expected; this means that the occurrence of X has a positive effect on the occurrence of Y.

In our analysis, we used the R programming language, specifically the '*arulesViz*' and '*arules*' packages, as well as the Apriori algorithm [11]. The Apriori algorithm, which mines frequent itermsets and interesting associations in transaction databases, is not only the first but also the most popular association rule mining tool.

3. Results and Discussion

The association rules that can be deduced based on the given antimicrobial susceptibility data set can be parametrized according to the knowledge that clinicians want to extract from the database. For example, by setting *minsup* and *minconf* thresholds to 0.001 and 0.30, respectively, the first five rules sorted by decreasing order in respect of confidence are presented in the following sample output (Table 2).

Table 2. The first five rules sorted by decreasing order in respect of confidence, by setting *minsup* and *minconf* thresholds 0.001 and 0.30.

		LHS		**RHS**		**Measures**	
Gender	men	Bacteria	Antibiotic	Susceptibility	Support	Confidence	Lift
F	p.	*K. pneumoniae*	Cefta/Avi	S	0.0013	1.0000	3.4482
M	p.	*K. pneumoniae*	Cefta/Avi	S	0.0023	1.0000	3.4482
F	ne	*P. aeruginosa*	Colistin	S	0.0012	1.0000	3.4482
F	p.	*P. aeruginosa*	Colistin	S	0.0071	1.0000	3.4482
M	p.	*P. aeruginosa*	Colistin	S	0.0127	1.0000	3.4482

M: male, F: female, S: sensitive, Cefta/Avi: ceftazidime/axibactam, RHS: right-hand side, LHS: left-hand side

For example, the first row of the above table can be formulated as follows: IF the patient is female AND the sample type is Respiratory secretions AND the bacteria detected is *K.pneumoniae* AND the antibiotic used is Cefta/Avi, THEN the susceptibility testing will show that this bacteria is sensitive to Cefta/Avi. By changing *minsup* and *minconf* thresholds, we get a different set of rules.

The full collection of rules from the preceding example, along with their scatter plots, the network graphs of the fifteen most significant rules sorted by confidence, and the accompanying R code used, can be found in the online appendix [12].

The AMR problem has already been assessed from the point of view of supervised ML, in previous studies [2-3,5,7]. In the present study, we strive to diagnose antibiotic resistance against certain pathogens by searching strong association rules among variables retrieved from the antibiotic susceptibility dataset and possible determinants of resistance.

4. Conclusion

Association rule mining is an evolving technology linking statistics, artificial intelligence, and machine learning to pull out significant information from large amounts of accumulated data in databases and discover interesting correlations among them. In this paper we applied ARM in an antibiotic susceptibility dataset, aiming to uncover the most meaningful rules in the antibiotic resistance profile of critically ill patients suffering from multidrug resistant infections and to offer those rules for evaluation and reflection to the medical experts.

References

[1] Feretzakis G, Loupelis E, Sakagianni A, et al. A 2-Year Single-Centre Audit on Antibiotic Resistance of Pseudomonas aeruginosa, Acinetobacter baumannii and Klebsiella pneumoniae Strains from an Intensive Care Unit and Other Wards in a General Public Hospital in Greece. Antibiotics (Basel). 2019 May 15;8(2):62. https://doi: 10.3390/antibiotics8020062.

[2] Feretzakis G, Loupelis E, Petropoulou S, et al. Using Microbiological Data Analysis to Tackle Antibiotic Resistance of Klebsiella Pneumoniae. Stud Health Technol Inform. 2019 Jul 4;22:180-183. https://doi: 10.3233/SHTI190047.

[3] Rodríguez-González A, Zanin M, Menasalvas-Ruiz E. Public Health and Epidemiology Informatics: Can Artificial Intelligence Help Future Global Challenges? An Overview of Antimicrobial Resistance and Impact of Climate Change in Disease Epidemiology. Yearbook of Medical Informatics. 2019;28(01):224–31.

[4] Feretzakis G, Loupelis E, Sakagianni A, Kalles D, et al. Using Machine Learning Techniques to Aid Empirical Antibiotic Therapy Decisions in the Intensive Care Unit of a General Hospital in Greece. Antibiotics (Basel). 2020 Jan 31;9(2):50. https://doi: 10.3390/antibiotics9020050.

[5] Feretzakis G, Loupelis E, Sakagianni A, et al. Using Machine Learning Algorithms to Predict Antimicrobial Resistance and Assist Empirical Treatment. Stud Health Technol Inform. 2020 Jun 26;272:75-78. https://doi: 10.3233/SHTI200497.

[6] Martínez-Agüero S, Mora-Jiménez I, Lérida-García J, Álvarez-Rodríguez J, Soguero-Ruiz C. Machine Learning Techniques to Identify Antimicrobial Resistance in the Intensive Care Unit. Entropy. 2019;21:603

[7] Feretzakis G, Sakagianni A, Loupelis E, et al. Using Machine Learning to Predict Antimicrobial Resistance of Acinetobacter Baumannii, Klebsiella Pneumoniae and Pseudomonas Aeruginosa Strains. Stud Health Technol Inform. 2021 May 27;281:43-47. doi: 10.3233/SHTI210117.

[8] Sarıyer G, Öcal Taşar C. Highlighting the rules between diagnosis types and laboratory diagnostic tests for patients of an emergency department: Use of association rule mining. Health Informatics Journal. June 2020:1177-1193. doi:10.1177/1460458219871135.

[9] Kumar AV, Ali RF, Cao Y, Krishnan VV. Application of data mining tools for classification of protein structural class from residue based averaged NMR chemical shifts. Biochim Biophys Acta. 2015;1854(10 Pt A):1545-1552. doi:10.1016/j.bbapap.2015.02.016.

[10] S. E. Brossette, A. P. Sprague, W. T. Jones, and S. A. Moser, “A data mining system for infection control surveillance”, Methods of information in medicine, vol. 39, pp. 303-10, 2000.

[11] Agrawal R, Srikant R. “Fast algorithms for mining association rules”, Proceedings of 20th International Conference on Very Large Data Bases (VLDB), Santiago de Chile, Chile, September 12-15, 1994.

[12] Feretzakis G, Discovering Association rules and data mining in Antimicrobial Resistance in Intensive Care Unit. Available online: www.learningalgorithm.eu/AssociationRules_ICU_ICIMTH2022.html (accessed on 25 March 2022).