Preliminary Results of an Analysis Using Association Rules to Find Relations between Medical Opinions About the non-Realization of Autopsies in a Mexican Hospital

Elayne Rubio Delgado[1], Lisbeth Rodríguez-Mazahua[1],
Silvestre Gustavo Peláez-Camarena[1], María Antonieta Abud-Figueroa[1],
José Antonio Palet Guzmán[2], Asdrúbal López-Chau[3], Giner Alor-Hernández[1]

[1] División de Estudios de Posgrado e Investigación,
Instituto Tecnológico de Orizaba, Veracruz,
Mexico

[2] Hospital Regional de Rio Blanco (H.R.R.B.), Veracruz,
Mexico

[3] Universidad Autónoma del Estado de Mexico, Zumpango, Estado de Mexico,
Mexico

{lrodriguez, galor}@itorizaba.edu.mx, {gpelaez, mabud}@ito-depi.edu.mx,
n3nyrbd@gmail.com, palet_guzman@hotmail.com, alchau@uaemex.mx

Abstract. In the last years, a significant reduction in the number of autopsies realized in the hospitals of the world has been observed. Since medics are the closest people to this problematic, they can offer information that helps clarify why the decreasing of this practice has occurred. In this paper, data mining techniques are applied to perform an analysis of medical opinions regarding the realization of autopsies in a hospital of Veracruz, in Mexico. The opinions were collected through surveys applied to 85 medics of the hospital. The result is a model represented by a set of rules that suggests some of the factors that are related to the decrease in the number of autopsies in the hospital, according to the survey respondents.

1 Introduction

Autopsies are one of the most important exams of the medical practice. They allow obtaining vital information about diseases, their development and the effects in the human organism. Nevertheless, the realization of autopsies supposes problematic and interesting situations of acceptance or rejection in the population [1].

In the last years, a decrease in the practice of autopsies in a worldwide level has been observed [1]. In the hospital centers of the biggest cities of Mexico, and particularly in the state of Veracruz, this decreasing trend in the number of autopsies performed is

presented. However, to the best of our knowledge, there are not published studies about the causes related to this phenomenon.

In this sense, it is pertinent to directly and specifically investigate the motives for the non-realization of autopsies in the hospitals, focusing the attention on the medics and authorities trained to act on that subject.

In this paper, the preliminary results of an analysis performed to datasets obtained by the application of an instrument (survey) to medics of a hospital of Veracruz are presented. Association rules were used in such analysis because they allow finding relations between medical opinions regarding several causes for the non-realization of autopsies, as well as the level of studies, years of experience, and participation of the medics in autopsy cases, to mention a few.

For a better explanation of the object of study, the rest of the paper is structured as follows. Section 2 shows the relating to the state of the art and Section 3 describes the case of study. Finally, the conclusions and the future work are given in Section 4.

2   Related Works

With the motivation of supporting the development of the health sector in smart cities, Oviedo Carrascal, et al. [2] described a set of advances and trends of data mining in this area. Nowadays, it is known that several medicines produce adverse effects and the efficient identification of these drugs is a challenge for the experts. Wang, et al. [3] detected relations between diseases and illnesses using association rules derived from the clinical evidence and analysis of cases. The factors that influence the survival to the disease of cervix cancer and the time of survival to the disease were identified in [4] through the application of decision trees and association rules. The study presented in [5] demonstrated that by means of the application of association rules to a dataset of patient records, it is possible to obtain correlations between the different attributes, such as exams, medicines, treatments and patient profiles. In [6] it was presented the new prediction system ICU ARM-II (Intensive Care Unit Association Rule Mining), based in a set of association rules that forms a flexible model for the prediction of personalized risk. This approach assumes a classification supported by association. The risk factors correlated to Diabetes Mellitus type 2 (DM2) were identified in [7] using association rules.

Embracing other areas, in [8] the effectiveness of online collaborative learning environment and the predictors generated using data mining (DM) techniques were evaluated. The study [9] examined the pattern of student performance by using the "K-means" clustering technique. In the study [10] authors offered a general overview of the recent works undertaking the predictive and descriptive tasks in the building field, with the aim of improving building performance. On the other hand, the works [11, 12] confirmed the considerable potential of Bayesian networks to explore coastal databases. The software industry has also benefited from DM techniques as seen in [13], where Ahmed H. Yousef proposed a solution architecture that enhances software development based on data in software repositories and presented a benchmark that provides an ensemble of DM models in the defective modules prediction problem and compares the

results. In [14] an automatic real-time web usage data mining and recommendation system is presented. The system is powered by K-NN (K-Nearest Neighbor) classification model implemented with Euclidian distance method. In [15] DM methods were analyzed for detection of anomalies in social networks. The research focuses on the different types of anomalies, classification based on features along with the assumptions and the reasons for their existence, as well as techniques to prevent and detect them. The work [16] described the procedure to recognize patterns that allow evaluating the energy efficiency of buildings by applying the K-means algorithm.

Table 1. Related works.

Work	DM techniques	Area	DM Tasks
[2]	Decision trees, Neural networks, "K-means"	Health	Classification, Clustering
[3]	Association rules	Health (Pharmacology)	Association
[4]	Decision trees and association rules	Health	Classification, Association
[5]	Association rules	Health	Association
[6]	CBA (Classification based in Association)	Health	Classification, Association
[7]	Association rules	Health	Association
[8]	Decision trees	Education	Classification
[9]	Clustering based in partition	Education	Clustering
[10]	Support Vector Machines. Neural network Decision trees Association rules	Engineering	Classification, Clustering, Association
[11]	Bayesian methods	Ecology	Classification
[12]	Bayesian methods	Ecology	Classification
[13]	Neural network Decision trees Naive Bayes	Software	Classification
[14]	K-Nearest Neighbor	Internet: Online applications	Classification
[15]	K-Nearest Neighbor Support Vector Machines. Neural network	Social networks	Classification Clustering,
[16]	Decision trees	Engineering	Classification

As we can see in Table 1, the different studies mentioned demonstrate the usefulness of data mining techniques for the solution of problems in several areas of the modern reality, and the medical area is raised as a great object of study, with appropriate problems to be studied from this perspective [2-7]. Nevertheless, to the best of our

knowledge there are not works which used association rule mining to analyze the decrease in the number of autopsies performed in a hospital, therefore this determines the appropriateness, novelty and interest of this research.

3 Analysis of Relations between Medical Opinions

3.1 Association Rules

This research is aimed to perform a descriptive analysis through DM, which allows the identification of the relations existing in the survey data and obtaining a model based on rules. With this model it is possible to evaluate which are the factors that influence in the decreasing of the practice of autopsies. For this reason, association rules were considered for this study.

Association rules [17] allow the discovering of relations between attributes based on the occurrence frequency of the attributes in transactions. An association rule is an implication of X over Y representing the form X⇒Y, where X and Y are sets of disjunct elements and the rule is interpreted as a norm that when X appears, Y also tends to appear. These rules are evaluated with the measures of support, which represents the ratio of elements of the dataset that contains X and Y, and confidence, which is the probability that a transaction containing X also contains Y. The association algorithms perform automatic searches of rules that relate sets of attributes and evaluate whether those rules are statistically significant. To do this, Weka basically has the Apriori [18], FilteredAssociator [19], FPGrowth [20], GeneralizedSequentialPatterns [21], PredictiveApriori [22] and Tertius [23] algorithms.

Apriori [24] is a classic algorithm for association rule mining. Apriori generates rules by means of an incremental process that performs searches of frequent relations between attributes delimited by a minimum confidence. The algorithm can be configured to be executed under certain criteria such as upper and lower limits of coverage to accept itemsets that accomplish the minimum confidence constraint, which is a ranking criteria to show the rules, and a parameter to indicate the specific quantity of rules that we want to find. Apriori it only seeks rules between symbolic nominal attributes and generates better results than other algorithms mentioned above, as it is shown in [25].

3.2 Dataset Collection and Processing

With the objective of analyzing the possible causes of the decrease in autopsies performed in the hospital system of "Servicios de Salud de Veracruz", a survey elaborated by one of the authors of this work was used as social research technique. Such survey was applied to 85 medics of one of the hospitals of the above mentioned system. The survey fundamentally pretended to investigate the opinion of the medics facing the practice of autopsies, as well as the principles, beliefs or motives that characterize the medics. The answers of the medics were transformed to a suitable representation to apply the data mining techniques in order to significantly reduce the

complexity of the data analysis. Table 2 shows a summary of the applied survey, and the amount of categories generated by each question. Figure 1 presents the percentage of questions that belongs to each aspect explored by the survey.

The answers of the surveys applied were introduced in a binary matrix, in this matrix every row represents a survey respondent, the columns represent every answer. The value of each column is the intersection that can be read as a pair *<answer, value>*, where *value* is equal to 'Y' if the survey respondent gives such answer and 'N' otherwise. Such dataset is denominated in this article as **C**.

Table 2. Summary of the survey applied to the medical staff.

Aspects	Questions	Type of question	Generated categories
Medical training	Area	Closed	3
	Grade	Closed	3
	General Medicine Training Center	Closed	47
	Medical Specialty Training Center	Closed	47
Medical experience	Years of medical practice	Closed	5
	Participation in Autopsy cases	Closed	5
Medical opinion about the discoveries in autopsies	Cause discrepancy with the clinical diagnoses	Closed	5
	Originate in claim cases	Closed	5
	Originate in arbitration cases	Closed	5
Survey respondent opinion about the request of autopsies	Motives for autopsy acceptance	Open	26
	Motives for autopsy rejection	Open	27
	Motives for autopsy rejection by family	Closed	9
	Motives for not enough autopsies performed in the hospital	Closed	8
Medical opinion about the procedure to request an autopsy	Suitable staff to request an autopsy	Closed	6
	Efficient methods to request an autopsy	Closed	14
General aspect	Comments	Open	25

With the objective of analyzing relations between the medical experience and the medical opinion about the procedure of autopsy requests, a subset C_1 was generated from **C**. With the purpose of reducing the number of attributes to simplify the analysis of the dataset **C**, a subset C_2 was formed by **C**. For this reason, the algorithm presented in [26] was used for attribute selection, it was necessary to designate a class label attribute, *motives for autopsy rejection* was selected for this objective. A summary of the analyzed datasets is presented in Table 3. The generated datasets do not have either missing, inconsistent or out of range values.
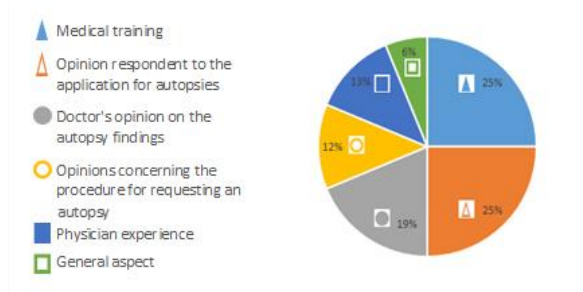
Fig. 1. Main areas explored by the survey.

Table 3. Data sets characteristics.

Set	Attributes	Objects
C	240	85
C_1	56	85
C_2	39	85

3.3 Application of the Data Mining Methods

Because of space limitations, only the association rules with greater confidence level obtained for each dataset are presented. Table 4 shows the association rules from the dataset **C**, the more frequent attributes in the rules are: *motives for autopsy acceptance*, *suitable staff to request an autopsy*, *motives for autopsy rejection by family* and *motives for not enough autopsies performed in the hospital*.

Table 4. Results of applying the Apriori algorithm to the dataset **C**.

Rule	Antecedent	Consequent	Confidence
1.	P06R01=S P09R04=S	P10R01=S	0.97
2.	P09R04=S P11R01=S	P10R01=S	0.96
3.	P11R01=S P14R01=S	P10R01=S	0.95
4.	P06R01=S P11R01=S	P10R01=S	0.94
5.	P09R04=S P10R01=S	P11R01=S	0.94
6.	P08R01=S	P10R01=S	0.93
7.	P11R05=S	P10R01=S	0.95
8.	P06R03=S	P10R01=S	0.95

The following analyzed dataset was C_1, the results are showed in Table 5 and they demonstrate the relations between attributes like: *motives for autopsy acceptance*, *suitable staff to request an autopsy*, and *efficient methods to request an autopsy*.

Table 5. Results of applying the Apriori algorithm to the dataset C_1.

Rule	Antecedent	Consequent	Confidence
9.	P06R01=S P12R03=S	P10R01=S	0.97
10.	P06R01=S P06R03=S	P10R01=S	0.94
11.	P06R01=S P11R01=S	P10R01=S	0.94
12.	P06R03=S P11R01=S	P10R01=S	0.94
13.	P06R01=S P11R05=S	P10R01=S	0.97

C_2 was analyzed with the objective of discovering interesting rules between the attributes with more correlations in the dataset and Table 6 shows the results. In this case the rules with more confidence were those that related *motives for autopsy rejection* and the *medical opinion about the discoveries in autopsies*.

Table 6. Results of applying the Apriori algorithm to the dataset C_2.

Rule	Antecedent	Consequent	Confidence
14.	P07R05=S	P10R01=S	1
15.	P04R02=S P07R03=S	P10R01=S	1
16.	P05R01=S	P10R01=S	1
17.	P07R03=S	P10R01=S	0.96
18.	P04R02=S P06R03=S P11R01=S	P11R01=S	0.95
19.	P01R05=S	P11R01=S	0.94
20.	P11R01=S P12R01=S	P10R01=S	0.96

The analysis of the data in general allowed the identification of the elements related to the autopsy according to the opinion of the medics of the hospital that was object of this study, as we can see in Table 7. The results are subject to a final interpretation of the specialist who designed the instrument used to collect the information, in order to confirm the precision of the generated models.

4 Conclusions

The prominent decrease in the number of autopsies in the hospitals around the world has raised questions about the motives for this phenomenon. The purpose of this work was analyze the possible causes of the reduction of autopsies in the hospital system of "Servicios de Salud de Veracruz" by means of association rule mining from the data that belongs to the medical opinions about such medical practice.

The analyzed data were collected through a survey that was applied to the doctors of the hospital. The survey focused on the medical opinions about the causes why the autopsies were not performed, the study level of the specialists, their years of experience, the cases of autopsies they have been involved on, among others.

Table 7. Preliminary results of the medical opinions about autopsies.

Aspects	Results
Motives for autopsy acceptance	-Establish definitive diagnoses -Wrong diagnoses -Pedagogical objectives -Interest
Efficient methods for request an autopsy	- Instrument that regulates the autopsies as obligatory for the patients who passed away in the hospital - Asking the consent to the family once the patient is admitted to the hospital - Request by the medic
Suitable staff to request an autopsy	- Medic
Motives for autopsy rejection	- Lack of interest - Refusal of the family - Known Disease
Discoveries in autopsies	- Originate in claim cases. - Originate in arbitration cases. - Cause discrepancy with the clinical diagnoses
Autopsy rejection by family	- Religious motives

The use of association rule mining techniques allowed perform a descriptive analysis of the problematic situation and find the correlations between the categorical attributes of the dataset, which formed the information obtained from the medical staff.

So, we can conclude that according to the survey respondents, the decrease in the number of autopsies in the analyzed hospital is because of the reduction in the requests, the rejection by family due to religious issues, death because of known diseases, and lack of interest for the family, as well as for the health staff in some cases.

Most of the medics identified the professional interest, the possibility of obtaining definitive diagnoses and the pedagogic usefulness as motivating elements to request an autopsy, procedure that must be initiated by the medics, this was demonstrated by several of the obtained rules. The rules emphasized the role of the medic in the request

of autopsies and identified as efficient methods for the request of an autopsy asking the consent of the family of the patient and the legalization by the institution, in a regional or state level, of the obligatory nature of the autopsy practice for the patients who die.

In this sense, as future work we suggest to study data of the clinical records of the patients who died in the hospital, to analyze with real data the trend of the causes that lead to perform autopsies in some patients and not in others. This will confirm the veracity of the results of this research. We also recommend perform similar studies in other parts of the country, to identify whether the medical opinions and the consequences of autopsies rejection differ by region. Finally, investigations that consider other data mining techniques, such as Bayesian Networks, can be seen as further work. This will establish through a comparative analysis which algorithms are more efficient for this kind of studies.

References

1. Sanz-Ortiz, J., Mayorga, M., Mart, A.: Autopsy in clinical oncology: Is it in crisis? Medicina Clínica, 137(7), pp. 317–320 (2011)
2. Oviedo Carrascal, E. A., Oviedo, A. I., Vélez Saldarriaga, G. L.: Minería de datos: aportes y tendencias en el servicio de salud de ciudades inteligentes. Revista Politécnica, 11(20), pp. 111–120 (2015)
3. Chao Wang, et al.: Exploration of the Association Rules Mining Technique for the Signal Detection of Adverse Drug Events in Spontaneous Reporting Systems. PLoS ONE, 7(7), pp. 6 (2012)
4. Timarán Pereira, R., et al.: La minería de datos aplicada al descubrimiento de patrones de supervivencia en mujeres con cáncer invasivo de cuello uterino. Revista Universidad y Salud, vol. 14, pp. 117–129 (2012)
5. Antonelli, D. et al., MeTA.: Characterization of Medical Treatments at Different Abstraction Levels. ACM Trans. Intell. Syst. Technol., no. 57, pp. 25 (2015)
6. Chih-Wen C., et al.: icuARM-II: improving the reliability of personalized risk prediction in pediatric intensive care units. In: Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB '14), ACM, New York, NY, USA, pp. 211–219 (2014)
7. Franco, A. M., Pérez, León Guzmán, E.:An approach to the risk analysis of diabetes mellitus type 2 in a health care provider entity of Colombia using business intelligence. In: Proc. of the 6th Euro American Conf. on Telematics and Information Systems (EATIS '12), Rogerio Patricio Chagas do Nascimento (Ed.). ACM, New York, NY, USA, pp. 49–56 (2012)
8. Shukor, N. A., Tasir, Z., Van der, H.: An Examination of Online Learning Effectiveness Using Data Mining. Procedia - Social and Behavioral Sciences, vol. 172, pp. 555–562 (2015)
9. Harwati, Permata, A., Ayu Wula, F.: Mapping Student's Performance Based on Data Mining Approach (A Case Study). Agriculture and Agricultural Science Procedia, vol. 3, pp. 173–177 (2015)
10. Yu, Z. J., Haghighat, F., Benjamin, C. M. F.: Advances and challenges in building engineering and data mining applications for energy-efficient communities. Sustainable Cities and Society (in press), pp. 2–6 (2015)

11. Gutierrez, B., et al.: A Bayesian network to predict coastal vulnerability to sea level rise. J. Geophys. Res. Earth Surf, vol. 116, pp. 15 (2011)

12. Bulteau, T., Baills, A., Petitjean, L., Garcin, M., Palanisamy, H.: Gaining insight into regional coastal changes on La Réunion island through a Bayesian data mining approach. Geomorphology, vol. 228, pp. 134–146 (2015)

13. Yousef, A. H.: Extracting software static defect models using data mining. Ain Shams Engineering Journal, vol. 6, pp. 133–144 (2015)

14. Adeniyi, D., Wei, Z., Yongquan, Y.: Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. Applied Computing and Informatics, vol. 12, pp. 90–108 (2016)

15. Kaur, R., Singh, S.: A survey of data mining and social network analysis based anomaly detection techniques. Egyptian Informatics Journal (in press, corrected proof), no. 1 (2015)

16. Capozzoli, A., Grassi, D., Savino Pi, M.: Discovering Knowledge from a Residential Building Stock through Data Mining Analysis for Engineering Sustainability. Energy Procedia, vol. 83, pp. 370–379 (2015)

17. Tan, P.N., Steinbach, M., Kumar, V.: Introduction to Data Mining, Addison-Wesley, pp. 71–125 (2006)

18. Liu, B., Hsu, W., Ma, Y.: Integrating Classification and Association Rule Mining. In: Int. Conf. on Knowledge Disc. and Data Mining, pp. 80–86 (1998)

19. Bathla, H., et al.: Apriori Algorithm And Filtered Associattion Rule Mining. International Journal of Computer Science and Mobile Computing, vol. 4, pp. 209–306 (2015)

20. Han, J., Pei, J., Yin, Y.: Mining frequent patterns without candidate generation. In: Proceedings of the 2000 ACM-SIGMOD, pp. 1–12 (2000)

21. Srikant, R., Agrawal, R.: Mining Sequential Patterns: Generalizations and Performance Improvements. In: Proc. of the Int. Conf. on Extending Database Technology, pp. 1–17 (1996)

22. Scheffer, T.: Finding Association Rules that Trade Support Optimally against Confidence. In: 5th European Conference on Principles of Data Mining and Knowledge Discovery (2001)

23. Flach, P. A., Lachiche, N.: Confirmation-Guided Discovery of first-order rules with Tertius. Machine Learning, no. 42, pp. 61–95 (1999)

24. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD Int. Conf. on Management of Data, Washington, DC, pp. 207–216 (1993)

25. Aher, S. B., Lobo, L.M.R.J.: A Comparative Study of Association Rule Algorithms for Course Recommender System in E-learning. International Journal of Computer Applications, vol. 39, no. 1, pp. 48–52 (2012)

26. Kira, K., Rendell, L. A.: A practical approach to feature selection. In: International Conference on Machine Learning, pp. 368–377 (1992)