Mining compact predictive pattern sets using classification model

Matteo Mantovani[1], Carlo Combi[1], and Milos Hauskrecht[2]

[1] Department of Computer Science, University of Verona, Italy
{matteo.mantovani|carlo.combi}@univr.it
[2] Department of Computer Science, University of Pittsburgh, Pittsburgh, USA
milos@pitt.edu

Abstract. In this paper, we develop a new framework for mining predictive patterns that aims to describe compactly the condition (or class) of interest. Our framework relies on a classification model that considers and combines various predictive pattern candidates and selects only those that are important for improving the overall class prediction performance. We test our approach on data derived from MIMIC-III EHR database, focusing on patterns predictive of sepsis. We show that using our classification approach we can achieve a significant reduction in the number of extracted patterns compared to the state-of-the-art methods based on minimum predictive pattern mining approach, while preserving the overall classification accuracy of the model.

1 Introduction

Past decade has witnessed an explosion in the number of medical and healthcare datasets available to researchers and healthcare professionals. However, the analyses and utilization of these datasets still lack the data collection efforts. This prompts the development of appropriate data mining techniques and tools that can automatically extract relevant information from data and consequently provide insight into various clinical behaviors or processes captured by the data. Since these tools should interact with medical experts, it is important that all the extracted information is represented in a human-friendly way, that is, in a concise and easy-to-understand form.

One way to present knowledge to humans is to use if-then rules, that relate a condition defining a subpopulation of instances (or patients) with observed outcomes. The strength of this relation can be expressed using various statistics, such as precision and support. This human-friendly form facilitates the exploration, discovery and possible utilization of these patterns in healthcare. For example, consider a rule mining algorithm that identifies a subpopulation of patients that respond better to a certain treatment than the rest of the patients. If the rule clearly and concisely defines this subpopulation, it can be validated and potentially utilized to improve patient management and outcomes.

Many strategies to mine 'if-then' rules from the data exist. One is association rule mining [1, 2]. It gained a lot of popularity in data mining research [14], including medical data mining [8, 18]. The key strength of association rule

mining is that it searches the space of rules completely by examining all patterns that occur frequently in the data. Its disadvantage is that the number of association rules it finds and outputs is often very large. This may hinder the discovery process and the interpretability of the results. Hence, it is desirable to reduce the mined rule set as much as possible while preserving the most important relations (rules, patterns) found in the data. Various rule interestingness statistics and constraints based on such statistics have been proposed to address this problem [13].

The objective of this work is to study new ways of improving association rule mining that can lead to a smaller set of rules that are sufficient to capture the essential underlying patterns in the data. This requires analyzing relations among the mined rules and defining criteria for assessing the importance of individual rules w.r.t. other rules. The key principle studied and applied in this work for filtering the rules is rule redundancy. Our approach builds upon the minimum predictive pattern mining idea proposed by Batal and Hauskrecht [6] to eliminate spurious and highly redundant rules, and attempts to improve it by reducing the set of mined minimum predictive rules using an auxiliary classification model that combines the rules into one model. Since in general the search for the optimal set of rules is equivalent to the optimal subset selection problem [17], we propose and experiment with a more efficient greedy rule selection algorithm that avoids the need to explore and evaluate all possible rules subsets.

We have tested our method on data from MIMIC-III [15] EHR database. More specifically, our goal is to discover patterns that are associated with sepsis and its treatments. We compare our method to the original one [6] and show that the number of rules found by our method is significantly smaller than the original set. Moreover we show that the performance of the classification model that is based upon our rule set is close or better than classification models built by Batal's rule sets.

2 Related work

Association rule mining [1, 2] is a method for identifying strong relations in a dataset based on some measure of interestingness (e.g., confidence/precision, support or lift [13]). Typically, such relations are expressed in terms of if-then rules consisting of different rule antecedents (conditions) and consequents (targets). The majority of association rule mining algorithms rely on Apriori algorithm [2]. The algorithm searches the pattern space defining the condition of the rule by starting with more general patterns with the highest support before inspecting more specific patterns with a lower support. The process is bottomed-out by the minimum support parameter.

When the rule mining process is focused on a specific target class, we refer to it as to predictive pattern (rule) mining [16]. The task of identifying all important predictive patterns from a large pool of frequent patterns is similarly to association rule mining time-consuming, and may lead to a huge number predictive rules. One important contribution in limiting the size of the rule set is the

minimal predictive rule mining approach proposed in [6] to eliminate spurious predictive patterns. Briefly, a pattern is called spurious when it is predictive when evaluated alone, but is redundant given one of its subpatterns. Spurious patterns may be formed by adding irrelevant items to other simpler predictive patterns. Approach in [6] eliminates spurious patterns using statistical test based on binomial distribution. Later the same authors proposed a more robust Bayesian criterion to perform the spurious pattern elimination [3]. The minimum predictive rule mining approach has been successfully adapted and applied to mine temporal clinical data [4, 5, 7]

Predictive pattern mining process can be used for knowledge discovery when the goal is to extract a set of rules describing patterns that are important for a specific target class. Alternatively, it can be used to define a classifier [6]. In such a case, predictive patterns can be viewed as nonlinear features helping to improve overall performance of a classification algorithm. This complementary use of predictive patterns raises an interesting question. Is it possible to reduce the set of extracted predictive rules with the help of a classification model? That is, are there any rule redundancies that can be eliminated when we combine the rules into a classification model? Research in this work is centered around this interesting question. More specifically, we use mined set of minimum predictive rules to define features of the linear classification model based on Support Vector Machines (SVM). Then, feature selection methods are applied to further reduce the rule set, aiming to extract the set that optimizes the classification performance of the classification model.

3 Method

3.1 Definitions

Assume a dataset with only categorical features (attributes): all numeric features should be first discretized. Each (feature,value) pair is mapped to a distinct item in $\Sigma = \{I_1, ..., I_l\}$. A pattern is a conjunction of items: $P = I_{q_1} \wedge ... \wedge I_{q_k}$ where $I_{q_j} \in \Sigma$. If a pattern contains k items, we call it a k-pattern (an item is a 1-pattern). Assume an item $I = (fea, val)$, where fea is a feature and val is a value. Given a data instance $\boldsymbol{x}$, we say that $I \in \boldsymbol{x}$ if $fea(\boldsymbol{x}) = val$ and that $P \in \boldsymbol{x}$ if $\forall I_j \in P : I_j \in \boldsymbol{x}$.

Given a dataset $D = \{\boldsymbol{x_i}\}_{i=1}^{n}$, the instances that contain pattern P define a group $D_P = \{\boldsymbol{x_j} | P \in \boldsymbol{x_j}\}$.If P' is a subpattern of P ($P' \subset P$), then $D_{P'}$ is a supergroup of D_P ($D_{P'} \supseteq D_P$). Note that the empty pattern Φ defines the entire population. The support of P is defined as: $sup(P) = |D_P|/|D|$.

In this paper we are interested in mining patterns that are predictive of class c. So for pattern P, we can define a predictive pattern (or a rule) R: $P \Rightarrow c$ with respect to class label c. The confidence of R is the precision (or posterior probability of c in group D_P). Note that confidence of $\Phi \Rightarrow c$ is the prior probability of c. We say that rule R': $P' \Rightarrow c'$ is a subrule of rule R: $P \Rightarrow c$ if $c' = c$ and $P' \subset P$.

Let $\Omega = \{P_1, ..., P_m\}$ be a set of patterns predictive of c. Given a dataset $D = \{\boldsymbol{x_i}, y_i\}_{i=1}^{n}$ defined in d-dimensional feature space and a set of patterns Ω the instances in D can be mapped into a new m-dimensional binary array D_Ω as follows:
$\boldsymbol{x_i} \rightarrow \{b_{i,1}, ..., b_{i,m}\}$ where $b_{i,j} = 1$ if $P_j \in \boldsymbol{x_i}$ and $b_{i,j} = 0$ if $P_j \notin \boldsymbol{x_i}$.
We refer to new $D_\Omega = \{\boldsymbol{x'_i}, y_i\}_{i=1}^{n}$ as to the pattern induced projection of the dataset D based on patterns in Ω. The pattern induced dataset D_Ω and its instances can be used to define and also learn a binary classification model $f : \boldsymbol{x'_i} \rightarrow y_i = c$ that distinguishes instances with the target class c from other classes. Effectively, this classification model combines a set of patterns predictive of c into a unified model for predicting the same class.

3.2 Problem

Our objective is to identify a small set of predictive patterns (rules) for the target class c from the data. To achieve this we propose a new two-step pattern mining process.

First, the number of predictive rules one can define by considering just the rule support and its precision can be enormous and may include a large number of spurious patterns. Hence we restrict our attention only to non-spurious rules. We mine these rules using Apriori algorithm proposed by [6] that includes binomial test when selecting more specific rules.

Second, to further limit the number of predictive rules we combine the minimal predictive patterns into a unified classification model to search for the optimal minimal pattern set Ω^* predictive of the target class c. We define the optimal pattern set to be the minimal pattern set that leads to the best combined generalization performance discriminating class c from the rest of the classes.

In the following we first describe the idea behind the minimum predictive patterns, and the unified classification models. After that we propose a greedy search algorithm that combines the two ideas into one search mechanism for identifying small sets of predictive patterns.

3.3 Minimum predictive patterns

Our solution builds upon the concept of minimum predictive patterns (MPRs) proposed by Batal and Hauskrecht [6].

Definition: A predictive pattern $R : A \rightarrow c$ is a called minimal, if and only if, R predicts class c significantly better than all its subpatterns.

The gist of this definition is that every item in the condition of the predictive pattern R is an important contributor to its prediction, that is, removal of any of the items in the condition would cause a significant drop in its predictive performance. The significance of the pattern R is determined using a statistical test derived from the binomial distribution. Let us assume we are interested in testing the significance of rule $R : A \rightarrow c$. Assume that pattern

A consists of N instances, out of which N_c instances belong to class c. Let P_c represents the highest probability achieved by any subpattern of R, that is, $P_c = \max_{(} A' \subset A) Pr(c|A')$. To test, if the pattern R is significantly different, we hypothesize (null hypothesis) that N_c is generated from N according to the binomial distribution with probability P_c. If we cannot reject the hypothesis at some significance level, then, R is not significantly different from the subpattern with P_c. However, we say that pattern R is significantly different when we can reject the above hypothesis and show that the probability that generated N_c class x instances out N is significantly higher than P_c. We can perform this test using a one sided significance test and calculate its p-value. If this p-value is significant (smaller than a significance level α), we conclude that R significantly improves the predictability of c over all its simplifications, and hence R is a MPR. The mining algorithm to mine minimal predictive patterns relies on the Apriori algorithm that uses a minimum support parameter. The algorithm generates all patterns starting from more general patterns to more specific that satisfy the minimum support, but only the patterns that satisfy the binomial test (the minimality condition) are retained. As shown by studies in [6] such an algorithm retains significantly smaller subset of predictive patterns.

3.4 Combining predictive patterns via classification model

Our second solution attempts to reduce the number of minimum patters mined by considering their combinations. Briefly, we are interested in retaining only a subset of minimum predictive patterns that are critical for predictive performance of the classification model defined on the pattern induced dataset.

There are many classification models one can define on the binary dataset induced by the predictive patterns. In this work, instead of considering all possible classification models, we restrict our attention to linear support vector machines (SVM) models with shared discriminant functions (discriminating class c from the rest of the classes) that are defined by a linear combination of predictive patterns. To judge and compare the quality of such models across many features we use the area under the ROC curve (AUROC) statistic.

In general the problem of finding the optimal subset of minimum predictive patterns that leads to the best performing classification model is intractable. In order to make the search more efficient we resort to greedy pattern search approach. To make the choices of patterns we rely on the wrapper approach that tests, and selects patterns by considering the internal validation approach. That is, in order to compare two distinct sets of patterns Ω and Ω', we use the internal train and test splits of the data to evaluate the AUROC performance of the two sets in combination with the SVM model. The model and its patterns set with better AUROC performance is preferred. In the following we describe the specific algorithm we use to search a subset of minimum predictive patters to identify the best set.

3.5 Greedy pattern subset selection algorithm

Our approach starts by splitting dataset D into the training and test sets. All pattern selection and learning is always done on the training set. We use the test set only for the final evaluation.

Since our algorithm searches and compares many different subsets of predictive patterns, we use internal validation process to measure their quality and choose better subsets. Briefly, in order to evaluate and compare the goodness of a specific set of patterns Ω to other candidate sets, we use a classification model based on the linear SVM that is run on the data induced by Ω. We use multiple internal validation splits of the training data to make the comparison. The training dataset is divided as follows: first we randomly pick 30% of the data rows and use them as the test set, the remaining rows are reshuffled 10 times and for every reshuffle 80% of the data are used as the internal training set and the remaining 20% as the internal validation set. The goodness of Ω is then estimates by averaging the AUROC score for all internal splits obtained through reshuffling.

While our ultimate goal would be to find a set of predictive patterns that are optimal in terms of the quality of the predictive performance of a classifier that combines them, the full search is infeasible. To avoid the full pattern subset search, we adapt a greedy approach that generates, examines and selects the patterns level-wise, where a level k covers all k-patterns. More specifically, our method uses a two-stage procedure. First, using an Apriori algorithm with the minimum support threshold and the binomial test proposed by Batal et al, we generate a set of minimum predictive patterns for each level k. Second, we use these minimum level-wise patterns to construct greedily the final set of patterns. We implemented two procedures to conduct the greedy search. One that searches and constructs the subset of patterns starting from the most general (level 1) patterns and gradually adds new more specific (higher level) patterns. We refer to this procedure as the top-bottom greedy procedure. The other procedure starts from the most specific patterns (the highest level minimum predictive patterns) and greedily adds to the set more general patterns of lower complexity. We refer to this method as to the bottom-up greedy procedure.

Let us assume that Ω' is our current set of patterns (selected in the previous steps). Our greedy search algorithm on level k works by first trying each minimum pattern on level k in combination with Ω'. Each of these combinations are ranked in terms of the AUROC score based on the internal validation. This order defines a greedy order in which all k-level minimum patterns are sequentially tried and if successful (in terms of AUROC improvement) they are added (one-by-one) to the resulting set of patterns. The same procedure for greedily adding the patterns on level k is applied whether we build the patterns in the top-down fashion (from level 1 patterns) or from the bottom-up (from highest level patterns). The reason for using the bottom-up greedy search process is that it tends to retain a greater number of the more specific patterns.

4 Experiments

4.1 Data

To test and validate our method, we analyze clinical data derived from MIMIC III dataset [15] with the goal of identifying patterns predictive of sepsis diagnosis. Briefly, MIMIC is a publicly available database that contains EHR data for patients treated in intensive care units between 2001 and 2012. The data are de-identified and associated with 46000+ patients and $\sim$ 60000 admissions. The data consist of multiple clinical data sources: measurements of hourly vital signs (heart rate, blood pressures, oxygen saturation, and so on), administered drugs, labs and diagnosis for every patient. However, before analysis, it is necessary to transform the MIMIC-III raw data in a form that we could mine. This was accomplished through an E.T.L. (Extract, Transform, Load) process. One source of our data was CHARTEVENTS, that is the vital signs table. We used it to extract specific measurements of hearth rate, diastolic and systolic blood pressure, white blood cells, and body temperature across the admission. For each of these variables we created two attributes, one containing its maximum value during the hospitalization of a patient, and the other one containing its minimum value. Instead of numerical values, all these measurements were discretized to low, normal and medium ranges, using the thresholds shown in Table 1. Other information we selected from the records came from "PROCEDURES_ICD" table which we used to determine whether a patient had a procedure or not (true\false attribute is created) during the hospitalization. We applied the same transformation to table "DIAGNOSES_ICD" to identify all the patients diagnosed with sepsis. "INPUTEVENTS_MV" table consists of medication administration records. We used it to extract some medications administered to the patient, such as vancomycin, piperacillin/tazobactam, ciprofloxacin, epinephrine, norepinephrine, vasopressin, dopamine, metoprolol, potassium chloride, phenylephrine, omeprazole (prilosec), and pantoprazole (protonix). Let us note that while some of these medications are commonly used for treating patients with sepsis, whereas other medications such as metoprolol, potassium chloride, phenylephrine, omeprazole (prilosec), and pantoprazole (protonix) are more general. These were included to test the effectiveness of our method when mining patterns related to sepsis. At the end of the E.T.L. process we obtain data for 21880 patients, 2806 of them with sepsis.

4.2 Results

Table 2 shows the results we obtained on MIMIC-III data for the minimum predictive rule mining approach by Batal and Hauskrecht [6], and two versions

Table 1. Thresholds used to discretize the considered vital signs in low, medium, high.

	Heart Rate	**Diastolic BP**	**Systolic BP**	**White Blood Cells**	**Body Temperature**
Low	< 60	< 60	< 90	< 4.0	< 36.0
High	> 90	> 90	> 140	> 12.0	> 38.0

Table 2. Comparison between the results for our method and Batal et al's predictive pattern mining method.

Method	AUROC	Number of patterns
MPR (Batal et al)	0.8580	85
Our method (bottom-up)	0.8643	33
Our methods (top-down)	0.8635	19

of our greedy classification model driven subset selection approach. The main statistics we use to evaluate the quality of the predictive rule set is the area under the ROC (Receiver Operating Characteristics) curve (AUROC) [12]. All AUROC statistics listed in the table are obtained on the test data. In addition to AUROC performances, we list the number of patterns found by the different methods. For example, the minimum predictive pattern (MPR) baseline used 85 patterns and reached AUROC performance of 0.8580. As we can see, both greedy methods outperformed (in terms of the AUROC classification performance) the baseline. Moreover this improvement is accompanied by a significant reduction in the total number of patterns used in the set compared to the baseline. We note that while there is nearly no difference in the AUROC performance among the two versions of our greedy method, the number of patterns found and used by the two is significantly different. In particular, we observe that the majority of the patterns in the bottom-up approach are more complex patterns while the majority of patterns in the top-down approach are 1-patterns. This shows that bottom-up approach tends to keep more detailed patterns compared to the top-down approach.

5 Discussion

Sepsis is the systemic response to infection, and there are many conditions that would indicate its occurrence during the admission or hospital stay, such as: temperature $> 38\,^{\circ}$C or $< 36\,^{\circ}$C; heart rate > 90 beats per minute; systolic blood pressure < 90 mm Hg, and white blood cell count $> 12{,}000$/cu mm or $< 4{,}000$/cu mm [9]. Moreover, patients with sepsis are usually treated with antibiotics such as vancomycin, piperacillin / tazobactam, ciprofloxacin, and drugs treating episodes of hypotension such as epinephrine, norepinephrine, phenylephrine, vasopressin, and dopamine [11].

Table 3 lists all minimal predictive patterns that we mined using the bottom-up greedy procedure. The table entries include the absolute weight the rule was assigned by the final classification model, the rule support and the rule precision. By analyzing the results with respect to sepsis symptoms and treatments we see 21 patterns (out of 33) that match exactly sepsis related symptoms and/or treatments, and 9 more with the sepsis related patterns but in conjunction with Pantoprazole (Protonix). Pantoprazole is a proton pump inhibitor (PPI) and, even though it is not used to treat sepsis, PPIs are used for stress-related mucosal damage (SRMD). SRMD is an erosive gastritis of unclear pathophysiology, which can occur rapidly after a severe insult such as trauma, surgery, *sepsis* or burns

[10]. In other words, it is still reasonable to mine patterns with Pantoprazole, because it is weakly related to sepsis. Finally, we have only 3 patterns, indicated in Table 3 in italic, that include items we would consider to be weakly related to sepsis: 2 patterns have *MaxDiastolicBloodPressure = high* and one that includes *PotassiumChloride = true*. This demonstrates our algorithm is able to select a much smaller subset of patterns compared to MPR method and that the majority of the patterns predictive of sepsis are reasonable.

Table 3. The mined set of minimal predictive patterns with their absolute weight, support and precision

Pattern	**Rule Weight**	**Support**	**Precision**
Norepinephrine = true	0.4667	0.1453	0.4933
Norepinephrine = true & Vancomycin = true	0.2628	0.1114	0.5615
Piperacillin/Tazobactam = true	0.2476	0.143	0.3991
MaxSystolicBloodPressure = low	0.1981	0.4311	0.1571
Ciprofloxacin = true	0.1750	0.1157	0.2970
Vancomycin = true	0.1619	0.3890	0.2612
Pantoprazole(Protonix) = true & MaxSystolicBloodPressure = low	0.1418	0.1437	0.2801
Norepinephrine = true & Piperacillin/Tazobactam = true	0.1159	0.0566	0.6482
MaxWhiteBloodCells = high	0.1017	0.5669	0.1699
MinWhiteBloodCells = low & MaxHeartRate = high	0.0870	0.0653	0.2780
Vancomycin = true & MinWhiteBloodCells = high	0.0856	0.0940	0.2773
PotassiumChloride = true & MaxWhiteBloodCells = high	0.0738	0.3275	0.1793
Vancomycin = true & MaxHeartRate = high	0.0628	0.3015	0.2871
MinWhiteBloodCells = low	0.0601	0.0913	0.2312
MinDiastolicBloodPressure = low & MaxWhiteBloodCells = high	0.0533	0.3102	0.1847
Vancomycin = true & MaxWhiteBloodCells = high	0.0527	0.2821	0.2797
Pantoprazole(Protonix) = true & Piperacillin/Tazobactam = true	0.0512	0.0662	0.4438
MaxWhiteBloodCells = high & MaxSystolicBloodPressure = low	0.0500	0.3255	0.1792
Piperacillin/Tazobactam = true & MaxWhiteBloodCells = high	0.0471	0.1106	0.4238
Pantoprazole(Protonix) = true & Ciprofloxacin = true	0.380	0.0574	0.3436
MinTemp = low	0.0369	0.0618	0.1706
Vancomycin = true & *MaxDiastolicBloodPressure = high*	0.0324	0.1089	0.3023
Ciprofloxacin = true & MaxHeartRate = high	0.0241	0.0930	0.3251
Ciprofloxacin = true & MaxWhiteBloodCells = high	0.0117	0.0873	0.3263
Pantoprazole(Protonix) = true & Norepinephrine = true	0.0094	0.0679	0.5496
MaxWhiteBloodCells = high & MaxHeartRate = high	0.0075	0.4129	0.1974
Pantoprazole(Protonix) = true & Vancomycin = true	0.0058	0.1433	0.3423
Pantoprazole(Protonix) = true & *PotassiumChloride = true*	0.0053	0.1616	0.2442
Pantoprazole(Protonix) = true & MinDiastolicBloodPressure = low	0.0047	0.1369	0.2882
PotassiumChloride = true & MaxDiastolicBloodPressure = high	0.0047	0.1320	0.2184
MaxDiastolicBloodPressure = high & MaxHeartRate = high	0.0001	0.1451	0.2278
MaxSystolicBloodPressure = low & MaxHeartRate = high	0.0015	0.3109	0.1942
Pantoprazole(Protonix) = true & MaxWhiteBloodCells = high	0.0001	0.1851	0.2534

6 Conclusion

In this work we have developed and tested a new framework for mining predictive patterns that compactly describe a class of interest. It uses a greedy algorithm to mine the most predictive patterns level-wise and including only those that improve the overall class prediction performance. We tested our approach on intensive care data from MIMIC-III EHR database, focusing on patterns predictive of sepsis. The results preserve the overall classification quality of state-of-the-art methods based on minimum predictive pattern mining approach, but with a significant reduction in the number of extracted patterns.

Acknowledgement This work was supported by NIH grant R01-GM088224. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of SIGMOD (1993)
2. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: Proceedings of VLDB (1994)
3. Batal, I., Cooper, G., Hauskrecht, M.: A bayesian scoring technique for mining predictive and non-spurious rules. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 260–276. Springer (2012)
4. Batal, I., Cooper, G.F., Fradkin, D., Harrison, J., Moerchen, F., Hauskrecht, M.: An efficient pattern mining approach for event detection in multivariate temporal data. Knowledge and information systems **46**(1), 115–150 (2016)
5. Batal, I., Fradkin, D., Harrison, J., Moerchen, F., Hauskrecht, M.: Mining Recent Temporal Patterns for Event Detection in Multivariate Time Series Data. In: Proceedings of the international conference on Knowledge Discovery and Data mining (SIGKDD) (2012)
6. Batal, I., Hauskrecht, M.: Constructing classification features using minimal predictive patterns. In: Proceedings of the 19th ACM international conference on Information and knowledge management. pp. 869–878. ACM (2010)
7. Batal, I., Valizadegan, H., Cooper, G.F., Hauskrecht, M.: A Temporal Pattern Mining Approach for Classifying Electronic Health Record Data. ACM Transaction on Intelligent Systems and Technology (ACM TIST), Special Issue on Health Informatics (2012)
8. Bellazzi, R., Zupan, B.: Predictive data mining in clinical medicine: current issues and guidelines. International journal of medical informatics **77**(2), 81–97 (2008)
9. Bone, R.C., Balk, R.A., Cerra, F.B., Dellinger, R.P., Fein, A.M., Knaus, W.A., Schein, R.M., Sibbald, W.J.: Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Chest **101**(6), 1644–1655 (1992)
10. Brett, S.: Science review: the use of proton pump inhibitors for gastric acid suppression in critical illness. Critical care **9**(1), 45 (2004)
11. Dellinger, R.P., Levy, M.M., Rhodes, A., Annane, D., Gerlach, H., Opal, S.M., Sevransky, J.E., Sprung, C.L., Douglas, I.S., Jaeschke, R., et al.: Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock, 2012. Intensive care medicine **39**(2), 165–228 (2013)
12. Fawcett, T.: An introduction to roc analysis. Pattern recognition letters **27**(8), 861–874 (2006)
13. Geng, L., Hamilton, H.J.: Interestingness measures for data mining: A survey. ACM Computing Surveys (CSUR) **38**(3), 9 (2006)
14. Han, J., Pei, J., Kamber, M.: Data mining: concepts and techniques. Elsevier (2011)
15. Johnson, A.E., Pollard, T.J., Shen, L., Lehman, L.w.H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L.A., Mark, R.G.: MIMIC-III, a freely accessible critical care database. Scientific data **3** (2016)
16. Jovanoski, V., Lavrac, N.: Classification rule learning with apriori-c. In: EPIA (2001)
17. Koller, D., Sahami, M.: Toward optimal feature selection. Tech. rep., Stanford InfoLab (1996)
18. Lavrač, N.: Selected techniques for data mining in medicine. Artificial intelligence in medicine **16**(1), 3–23 (1999)