Contrasting temporal trend discovery for large healthcare databases

CrossMark

Goran Hrovat[a,*], Gregor Stiglic[a,b], Peter Kokol[a,b], Milan Ojteršek[a]

[a] *Faculty of Electrical Engineering and Computer Science, University of Maribor, Smetanova ulica 17, 2000 Maribor, Slovenia*
[b] *Faculty of Health Sciences, University of Maribor, Žitna ulica 15, 2000 Maribor, Slovenia*

ARTICLE INFO

ABSTRACT

With the increased acceptance of electronic health records, we can observe the increasing interest in the application of data mining approaches within this field. This study introduces a novel approach for exploring and comparing temporal trends within different in-patient subgroups, which is based on associated rule mining using Apriori algorithm and linear model-based recursive partitioning. The Nationwide Inpatient Sample (NIS), Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality was used to evaluate the proposed approach. This study presents a novel approach where visual analytics on big data is used for trend discovery in form of a regression tree with scatter plots in the leaves of the tree. The trend lines are used for directly comparing linear trends within a specified time frame. Our results demonstrate the existence of opposite trends in relation to age and sex based subgroups that would be impossible to discover using traditional trend-tracking techniques. Such an approach can be employed regarding decision support applications for policy makers when organizing campaigns or by hospital management for observing trends that cannot be directly discovered using traditional analytical techniques.

1. Introduction

Applications where trend-discovery is applied have become important tools for helping companies analyze information and providing support when making decisions mainly for the purpose of reducing business costs. In addition, more and more data is stored by healthcare organizations and available within immense data repositories on a daily basis. Extensive information is buried within large amounts of available clinical data that could be used to enhance new knowledge [1–3]. Hence, new techniques need to be applied within knowledge discovery systems in order to take advantage of such an amount of data [4]. This paper presents a novel approach that exploits larger medical datasets in order to discover trends in healthcare. Whilst fully featured medical datasets are difficult to obtain due to privacy issues, we were able to use datasets without attributes that could identify individual patients. During this study we consulted the largest dataset in this area – i.e. the Nationwide Inpatient Sample (NIS), Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality. The dataset contains hospital discharge records for a stratified sample of approximately 20% of US community hospitals. In this study, we used data from the years 2000 to 2009, where each year includes approximately 7 million discharge records.

The aim of our work was to identify significant subgroups (e.g. males aged between 18 and 50) from these large datasets, and to comprehensibly show trends for each significant subgroup. The subgroups were structured within a tree structure

where, within the leaves, graphs were generated representing the trends. From such trees the trends for rules can be easily employed by policymakers and insurance companies to target their campaigns [5,6] or activities toward different age or gender groups. The meaning of a term subgroup here is different from that in pattern mining subgroup discovery, where aim is to find subgroup for which the distribution of a single target variable varies from its distribution in the whole database [7].

2. Hospital discharge dataset

The NIS dataset [8] contains hospital discharge records for a stratified sample of approximately 20% of US hospitals. This new dataset is available every year and consists of approximately 7 million discharge records. Each record contains up to 126 attributes comprising the personal characteristics of the patient including, age, gender, race, patient's county of residence; administrative information including admission month, year, admission type; and medical information including up to 15 diagnoses and up to 15 procedures. During our experiment, records from 10 consecutive years were used, namely 2000 to 2009, where we selected only those patients aged 18 or over, resulting in a final dataset of 65,308,185 discharge records. Diagnoses and procedures were coded according to the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes. An ICD-9-CM contains codes for diagnoses, where taxonomy containing five-digit codes is applied. The first three digits represent the general diagnosis and are followed by two additional digits describing a more specific subgroup of the general diagnosis. Additionally, the ICD-9-CM contains codes for procedures, where taxonomy of four-digit codes is applied and the first two digits represent the general procedure, followed by two digits describing a more specific subgroup of the general procedure.

3. Methods

3.1. Association rule mining

Association rules were used for discovering interesting relations between diagnoses within NIS datasets. For mining association rules we used one of the more popular algorithms called Apriori [9,10]. It was developed by Agrawal and Srikant for mining rules on large datasets consisting of transactions, where each transaction contains a set of items. In our case the discharge records were treated as transactions and attributes (e.g. age, gender, admission month, diagnoses) were considered as items. During the first stage Apriori finds frequent itemsets where the user specifies a minimum support threshold. Generated and pre-defined minimum confidence was considered from these itemsets' rules. The discovered rules followed the form $X \Rightarrow Y$, where X consists of one or more items and Y of only one item. $X \cap Y = \varnothing$. X is called an antecedent or the left-hand side of the rule (LHS) whilst Y is called the consequent or right-hand side of the rule (RHS). For each rule, measures of interest are calculated, to allow for the ranking of often huge sets of discovered rules [11]. The most common measures of interest included

support [9], confidence [9], and lift [12], however many other interesting measures could be applied (e.g. all-confidence [13], χ^2 [14]). Support of a rule is defined as $\text{supp}(X \Rightarrow Y) = P(X \cup Y)$, the fraction of transactions within the dataset that contain the itemset $X \cup Y$. Confidence in a rule is defined as $conf(X \Rightarrow Y) = \text{supp}(X \Rightarrow Y)/\text{supp}(X)$. Confidence can also be estimated using $conf(X \Rightarrow Y) = P(Y|X)$. The lift of a rule is defined as $lift(X \Rightarrow Y) = P(X \cup Y)/(P(X) \cdot P(Y)) = conf(X \Rightarrow Y)/\text{supp}(Y) = conf(Y \Rightarrow X)/\text{supp}(X)$. Lift measures the dependence between X and Y. X depends on the absence of Y when lift is below 1. X and Y depend on each other when lift is greater than 1 and X and Y are independent when lift is 1. The χ^2 measure of interest regarding a rule is defined as $\chi^2 = (lift - 1)^2 \cdot \text{supp} \cdot conf/((conf - \text{supp}) \cdot (lift - conf))$.

We performed association rule mining in statistical open-source language R [15] using package arules [16] where all the functionalities of the original Apriori algorithm were implemented.

3.2. Model-based recursive partitioning

Model-based recursive partitioning, as introduced by Zeileis et al. [17], was used for building a regression tree based on linear regression and splitting with respect to partitioning variables. The regression tree was built in the following order:

1. A linear model was estimated for all observations.
2. Those parameters, estimated using the objective function, were tested for instabilities within all partitioning variables. The most significant partitioning variable, with a p value less than the chosen alpha (0.05), was selected for further computation of the split points. The empirical fluctuation process [18] was calculated, in order to assess parameter instability. In the case of numerical variables, the maximum of the squared L2 norm of the empirical fluctuation process scaled by its variance was performed, whereas χ^2 statistics was used for categorical variables [19].
3. A split point was computed and a binary node, pointing toward two child-nodes, was created.
4. Steps 1–3 were repeated for all the child nodes. Recursion stopped once the p value became greater than the alpha or the number of observations within each node was lower than the user-selected value–this step is also called a pre-pruning of the tree. From the comprehensibility point of view, it is important to choose an appropriate alpha value that directly controls the size of a tree. The same applies to the argument min-split that controls the minimal number of samples within a node.

Model-based recursive partitioning in R was applied using the party package [20].

3.3. Discovery of temporal trends

In this subsection we introduce a novel approach for discovering temporal trends. We show that association rule mining in combination with model-based recursive partitioning can be used to discover trends in healthcare.

In our experiment only age, gender, admission month, year, and (i.e. in all up to 15) diagnoses were used from the
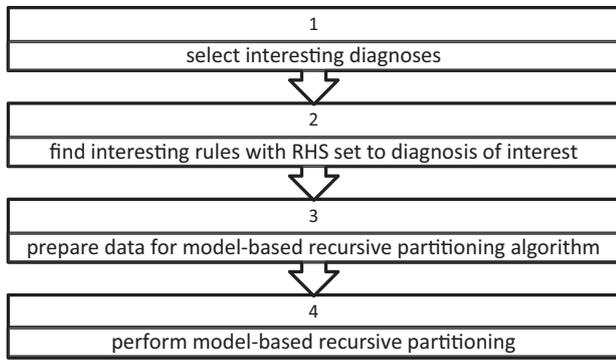
Fig. 1 – Flow-diagram of trend discovery.

Table 1 – Top 10 diagnoses with the highest support.	
Diagnose	Support (%)
Unspecified essential hypertension (401.9)	33.38
Diabetes mellitus without mention of complication, type II or unspecified type, not stated as uncontrolled (250.00)	13.21
Coronary atherosclerosis of native coronary artery (414.01)	12.98
Congestive heart failure, unspecified (428.0)	12.28
Outcome of delivery, single liveborn (V27.0)	12.15
Other and unspecified hyperlipidemia (272.4)	11.02
Atrial fibrillation (427.31)	9.66
Tobacco use disorder (305.1)	9.48
Esophageal reflux (530.81)	8.86
Anemia, unspecified (285.9)	7.59

126 features of NIS dataset. The most important features required when discovering trends are admission month, and year. Admission months, and years can also be transformed into quarters. We use age and gender as additional stratification variables due to their common use in healthcare related studies (e.g. [21–23]).

Fig. 1 presents a flow-diagram for the proposed technique of temporal trend discovery.

Our purpose was to find temporal trends for association rules including a user-defined diagnosis. We experimentally analyzed those diagnoses with the highest support. Diagnoses with the highest support are presented in Table 1. In the

following section we describe two cases – i.e. hyperlipidemia and hypertension.

Initially, association rules are built for selected diagnoses. For each diagnosis we executed Apriori with RHS set to the selected diagnosis. From these rules we selected those with the highest support and confidence. Rule examples for hyperlipidemia are presented in Table 2.

During the following step we prepared data for each rule, to be used as input for a model-based recursive partitioning algorithm. Each of the subgroups consists of discharge records filtered by time span (e.g. quarter), age (or age group) and gender. Data in Table 3 was constructed by calculating measures of interest for rule 3 in each of the subgroups.

In the last step, a regression tree was built by model-based recursive partitioning, representing temporal trends for significant subgroups of patients. Gender and age were set as partitioning variables, whereas the admission quarter or month was used as an independent and desirable measure of interest (e.g. support, confidence or χ^2) representing a dependent variable for the regression model.

4. Experiments

The introduced method was experimentally applied to the NIS dataset to discover temporal trends. We discovered trends for rules acquired by setting RHS rules to the diagnoses of hyperlipidemia and hypertension. Only diagnosis codes from the adult population were used during the rule-generation process.

For each of the rules it was necessary to prepare data for the building of regression trees. For each subgroup of the population, we calculated a rule for selective measures of interest with respect to each subgroup. Each subgroup was defined by gender, age, and admission month and quarter. All parameters (minimal support, minimal confidence, minimal and maximal lengths of rules) were chosen empirically as they produced promising results with the dataset. During all the experiments we used R version 2.13.2 with the packages party 0.9–99996, and arules 1.0–7.

We now present two examples of the proposed approach's effectiveness, where Hyperlipidemia and Hypertension-related rules were examined (Appendix A. Supplementary data).

Table 2 – Rule examples for hyperlipidemia discovered from NIS dataset.						
	LHS of rule	RHS of rule	Supp.	Conf.	Lift	χ^2
R1	Unspecified essential hypertension (401.9), Coronary atherosclerosis of autologous vein bypass graft (414.02), Esophageal reflux (530.81), Long-term (current) use of aspirin (V58.66)	Other and unspecified hyperlipidemia (272.4)	1.01×10^{-5}	0.76	6.85	5.69×10^{-5}
R2	Unspecified essential hypertension (401.9), Coronary atherosclerosis of native coronary artery (414.01), Coronary atherosclerosis of autologous vein bypass graft (414.02), Esophageal reflux (530.81), Long-term (current) use of aspirin (V58.66)	Other and unspecified hyperlipidemia (272.4)	8.35×10^{-6}	0.75	6.84	4.68×10^{-5}
R3	Unspecified essential hypertension (401.9), Esophageal reflux (530.81), Other complications due to other cardiac device, implant, and graft (996.72), Long-term (current) use of aspirin (V58.66)	Other and unspecified hyperlipidemia (272.4)	7.61×10^{-6}	0.76	6.92	4.33×10^{-5}

Table 3 – Example of the data table for rule R2.

Year	Quarter	Age group	Female	Support	Confidence	Lift	χ^2
2004	20	75–79	0	1.73×10^{-5}	1.00	6.36	9.27×10^{-5}
2005	21	50–54	0	1.94×10^{-5}	1.00	6.88	11.38×10^{-5}
2005	21	60–64	0	1.88×10^{-5}	1.00	5.22	7.91×10^{-5}
2006	25	55–59	0	5.30×10^{-5}	0.75	3.66	12.88×10^{-5}
2006	25	60–64	0	1.79×10^{-5}	1.00	4.57	6.40×10^{-5}
2006	25	70–74	0	1.71×10^{-5}	0.50	2.21	1.47×10^{-5}
2009	40	70–74	1	6.78×10^{-5}	1.00	3.42	16.44×10^{-5}
2009	40	75–79	1	3.18×10^{-5}	0.67	2.31	3.33×10^{-5}
2009	40	80–84	1	3.09×10^{-5}	1.00	3.59	8.00×10^{-5}

4.1. Hyperlipidemia

Examples of discovered rules for unspecified hyperlipidemia (11.02% prevalence) are presented in Table 2.

For the first regression tree (Fig. 2) age was grouped into five-year intervals to allow for higher support. Age was coded as a numerical attribute so that larger significant age groups could be determined during model-based recursive partitioning by joining neighboring age groups. Minimal support was set at 10^{-6}, minimal confidence at 0.75, and for the minimal and maximal lengths of the rules at 2 and 7 respectively.

An example of the first few rows of the prepared data-frame for the second rule R2 of Table 2, is presented in Table 3. This data-frame served as input to the model-based partitioning algorithm. The final outcome is represented by a regression tree that is used for exploratory trend analysis (Fig. 2). The tree in Fig. 2 was built using an R formula $\chi^2 \sim quarter \mid age + female$, where χ^2 was a dependent variable, *quater* is a regressor and the remaining attributes served as partitioning variables.

The resulting regression tree showed the temporal trend for rule R2 between the years 2004 and 2009, since this rule did not exist before 2004 due to changes in ICD9 coding. This rule represented patients who had *unspecified essential hypertension (401.9), coronary atherosclerosis of autologous vein bypass graft (414.02), esophageal reflux (530.81)* and *long-term (current) use of aspirin (V58.66)* having, in 75% of cases, *other and unspecified hyperlipidemia (272.4) also*. As seen in Table 3, 100% confidence in year 2006 occurred for a subgroup of males aged between 60 and 64 years. The generated tree in Fig. 2 shows that, for patients aged 59 and less, the trend of rule significance was positive, whilst for males aged 60 and over, the trend was negative. Higher χ^2 value indicates more significant and consequently more statistically interesting rule. Algorithm first splits dataset by age 59, leading to the two most significantly different groups, which is reflected also in density plot (Fig. 3). From the density plot is evident that patients in age group over 59 tend to have different time distribution in comparison to age group less or equal 59.

4.2. Hypertension

Another example, as shown in Fig. 4, where *unspecified essential hypertension (401.9)*, one of the most prevalent diagnoses (33% of cases), was used on the right-hand side of the rule. Minimal support was set at 10^{-6}, minimal confidence at 0.7,
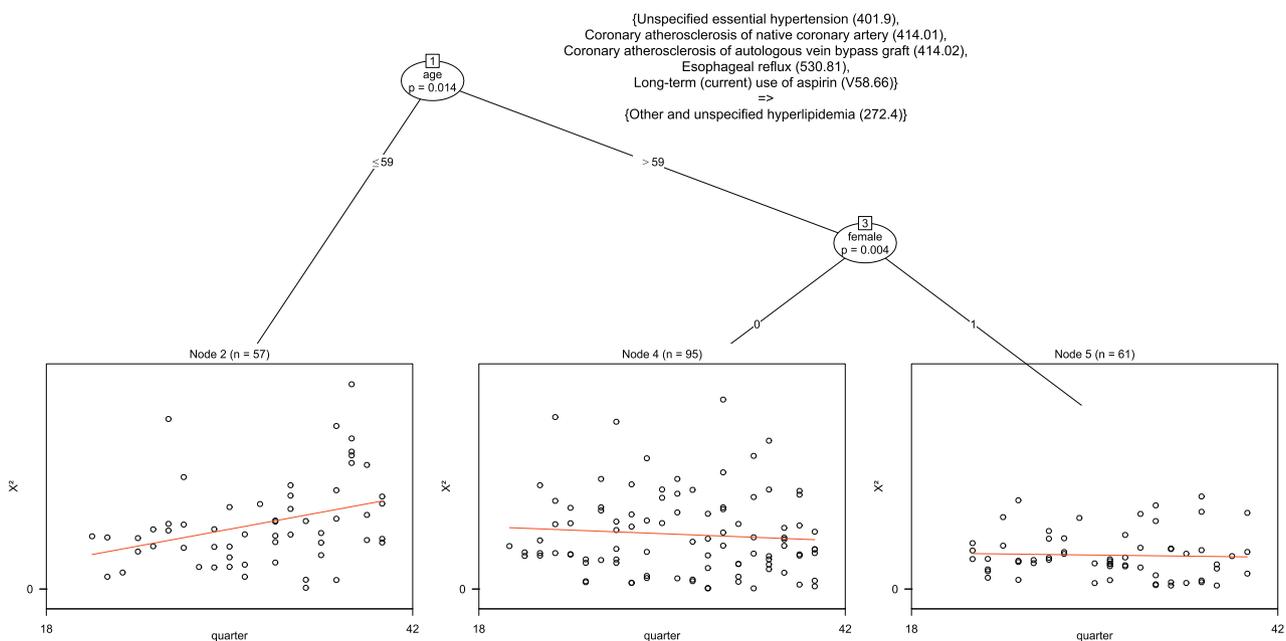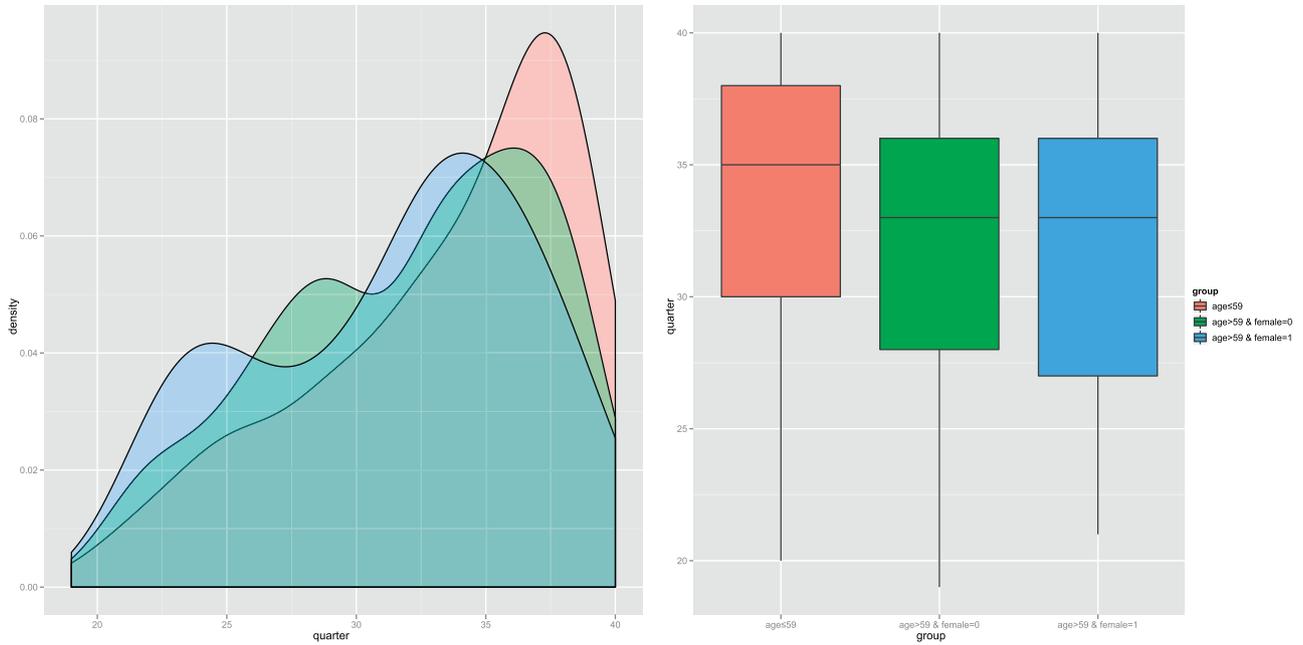

Fig. 2 – Regression tree of rule R2.

Fig. 3 – Density plot for rule R2.

and the minimal and maximal lengths of the rule to 2 and 7, respectively.

Support was used as a dependent variable with age partitioned into four intervals (18–49, 50–64, 65–74 and 75–124) and was represented as a categorical variable with 120 months used as the time points for trend analysis. The age groups were chosen accordingly to age group standards set by Center for Disease Control and Prevention (CDC) [24]. The tree represented rule R5 of Table 4, where a downward temporal trend of support could be observed for patients aged between 50 and 64, and an upward temporal trend for patients aged 65 and over.

5. Discussion

Temporal trend discovery is becoming important within different fields. A closely related field is the mining of temporal association rules [25], where temporal association rules are rules of interest only during a certain time period. Wang et al. [26] also evolved numerical attributes for temporal association rule mining.

Adomavicius and Bockstedt [27] developed a data analysis and visualization technique that presents complex multi-attribute temporal data in a cohesive graphical manner.
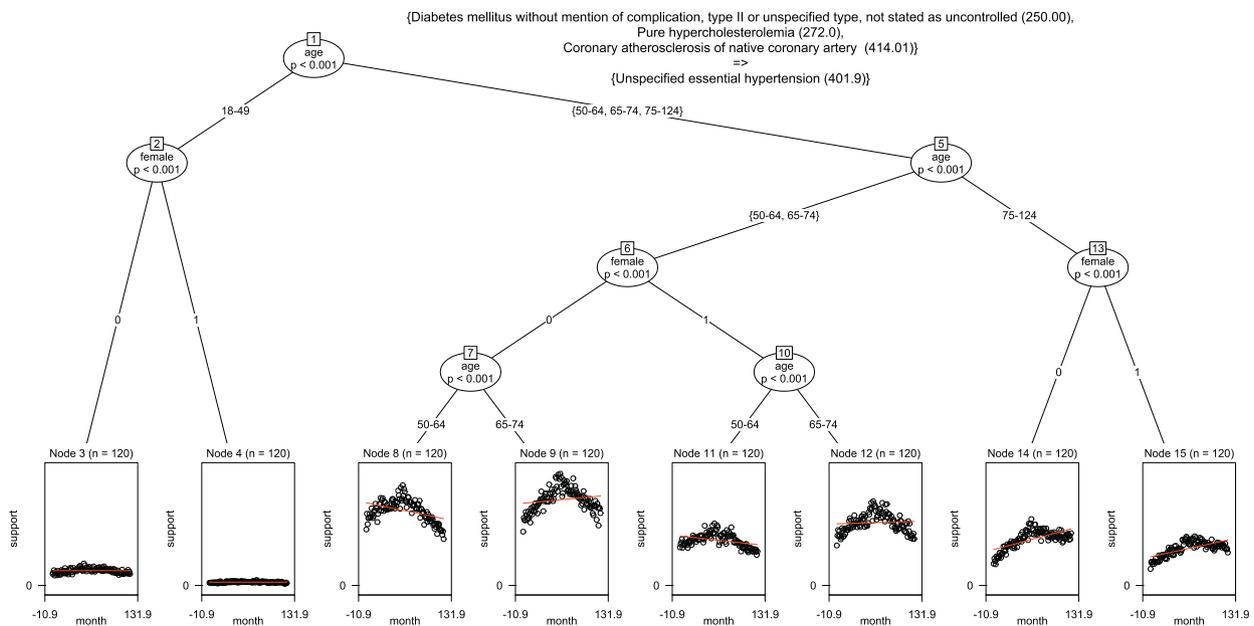

Fig. 4 – Regression tree of rule R5.

Table 4 – Rule examples for hypertension discovered from NIS dataset.

	LHS of rule	RHS of rule	Supp.	Conf.	Lift	χ^2
R4	Diabetes mellitus without mention of complication, type II or unspecified type, not stated as uncontrolled (250.00), Pure hypercholesterolemia (272.0)	Unspecified essential hypertension (401.9)	9.37×10^{-3}	0.70	2.10	8.24×10^{-3}
R5	Diabetes mellitus without mention of complication, type II or unspecified type, not stated as uncontrolled (250.00), Pure hypercholesterolemia (272.0), Coronary atherosclerosis of native coronary artery (414.01)	Unspecified essential hypertension (401.9)	3.09×10^{-3}	0.71	2.13	2.78×10^{-3}
R6	Diabetes mellitus without mention of complication, type II or unspecified type, not stated as uncontrolled (250.00), Other and unspecified hyperlipidemia (272.4), Obesity, unspecified (278.00)	Unspecified essential hypertension (401.9)	2.15×10^{-3}	0.71	2.12	1.92×10^{-3}

A recent study by Law et al. [28] extend the targeted positive/negative rule mining to a contrast mining, with the aim of extracting interesting rules between years. They identified trends in outbound tourists' behavioral patterns within a five-year historical dataset of travel behavior. Leman et al. [29] discover subgroups where a model fitted to the subgroup is substantially different from that same model fitted to the entire database. By using trend slopes instead of models, this method may also be used for trend analysis of subgroups. Similarly, Simonic et al. [30] used Disease Activity Score (DAS) trends and patterns for treatment optimization and as a measure for the quality of patient outcome.

In our case the temporal trends represent trends of interesting measures for association rules. These trends are represented as regression trees where we can identify upward or downward trends for significant subgroups. The significance of subgroups is determined using statistical tests. We identified trends in healthcare on the HCUP NIS ten-year dataset. This study presented trend discovery for selected diagnoses from hospital discharge data, however any association rule can be used to build a regression tree, representing subgroup trends for that rule. We need to keep in mind that rules with low support can lead to regression trees with only one node. For such trees we can identify trends for a whole population but not for smaller significant subgroups. In such cases large datasets are desirable, such as HCUP NIS.

6.   Conclusion

In our presented work a large healthcare HCUP NIS dataset was used for discovering temporal trends. A new technique is proposed where association rule mining is used together with model-based recursive partitioning. With this new technique temporal trends can be discovered, not only for the whole population but also for significant subgroups, where differences between subgroups for the same rule can be observed. Subgroups are defined according to patient characteristics (e.g. age, gender) or any other attributes we select. We represent trends for subgroups by observing changes in measures of interest for rules over time.

Information about temporal trends in medicine may be very useful for hospitals, insurance companies, research institutes, and other institutions trying to avoid or change negative trends for specific diagnoses of interest. Based on results from temporal trend analysis an insurance company can optimize their insurance plans and increase profits. For example, according to trend in rule R5, an insurance company could reduce price for insurance premium in plans offering insurance for hypertension, diabetes, hypercholesterolemia and coronary atherosclerosis in the 50–64 age group, due to a decrease of such cases in the last few years. Hence, we believe that in the future even more data from medical institutions will be included in large healthcare data repositories. At the moment NIS dataset consists of only 20% of discharges from US hospitals. Further research and evaluation of the method is needed with the help of physicians or other medical staff. Furthermore, other methods for building regression trees may be included and other variables may be selected for defining subgroups (e.g. procedures, charges). Finally, trends can be identified not only for rules, but also for frequent itemsets and other data mining structures where a variety of measures of interest can be applied.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cmpb.2013.09.005.

REFERENCES

[1] D.G. Lee, K.S. Ryu, M. Ezzeldin, A. Bashir, J.W. Bae, K.H. Ryu, Discovering medical knowledge using association rule mining in young adults with acute myocardial infarction, Journal of Medical Systems 37 (2) (2013).

[2] S. Tamang, S. Parsons, Unsupervised Modeling of Disease Dynamics. Association for the Advancement of Artificial Intelligence Spring Symposium, Workshop on Data Driven Wellness: From Self-Tracking to Behavior Change, 2013.

[3] N. Peek, S. Swift, Intelligent data analysis for knowledge discovery, patient monitoring and quality assessment, Methods of Information in Medicine 51 (4) (2012) 318–322.

[4] A. Holzinger, On Knowledge Discovery and Interactive Intelligent Visualization of Biomedical Data – Challenges in Human-Computer Interaction & Biomedical Informatics,

Conference on e-Business and Telecommunications (ICETE 2012), Rome, Italy, 2012, pp. IS9–IS20.

[5] B.M. Ramageri, Data mining techniques and applications, Indian Journal of Computer Science and Engineering 1 (4) (2010) 301–305.

[6] C. Rygielskia, J.C. Wang, D.C. Yen, Data mining techniques for customer relationship management, Technology in Society 24 (4) (2002) 483–502.

[7] A. Abu-Hanna, B. Nannings, D. Dongelmans, A. Hasman, PRIM versus CART in subgroup discovery: when patience is harmful, Journal of Biomedical Informatics 43 (5) (2010) 701–708.

[8] HCUP Nationwide Inpatient Sample (NIS), Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality, Rockville, MD, 2000–2009, www.hcup-us.ahrq.gov/nisoverview.jsp

[9] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: Proceedings of the 1993 ACM SIGMOD international conference on Management of data, 1993, pp. 207–216.

[10] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proceedings of the 20th International Conference on Very Large Data Bases, 1994, pp. 487–499.

[11] R.J. Bayardo, R. Agrawal, Mining the most interesting rules, in: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1999, pp. 145–154.

[12] S. Brin, R. Motwani, J.D. Ullman, S. Tsur, Dynamic itemset counting and implication rules for market basket data, in: Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data, 1997, pp. 255–264.

[13] E.R. Omiecinski, Alternative interest measures for mining associations in databases, IEEE Transactions on Knowledge and Data Engineering 15 (1) (2003) 57–69.

[14] C. Silverstein, S. Brin, R. Motwani, Beyond market baskets: generalizing association rules to dependence rules, Data Mining and Knowledge Discovery 2 (1) (1998) 39–68.

[15] R. Development Core Team: R, A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2004.

[16] M. Hahsler, B. Gruen, K. Hornik, arules – a computational environment for mining association rules and frequent item sets, Journal of Statistical Software 14 (15) (2005) 1–25.

[17] A. Zeileis, T. Hothorn, K. Hornik, Model-based recursive partitioning, Journal of Computational and Graphical Statistics 17 (2) (2008) 492–514.

[18] A. Zeiles, K. Hornik, Generalized M-fluctuation tests for parameter instability, Statistica Neerlandica 61 (4) (2007) 488–508.

[19] D.W.K. Andrews, Test for parameter instability and structural change with unknown change point, Econometrica 61 (4) (1993) 821–856.

[20] T. Hothorn, A. Zeileis, K. Hornik, Party: A Laboratory for Recursive Partytioning, 2012, http://CRAN.R-project.org/package=party

[21] D.E. Chiriboga, J. Yarzebski, R.J. Goldberg, Z. Chen, J. Gurwitz, J.M. Gore, J.S. Alpert, J.E. Dalen, A community-wide perspective of gender differences and temporal trends in the use of diagnostic and revascularization procedures for acute myocardial infarction, American Journal of Cardiology 71 (4) (1993) 268–273.

[22] J.G. Canto, W.J. Rogers, R.J. Goldberg, E.D. Peterson, N.K. Wenger, V. Vaccarino, C.I. Kiefe, P.D. Frederick, G. Sopko, Z.J. Zheng, Association of age and sex with myocardial infarction symptom presentation and in-hospital mortality, Journal of the American Medical Association 307 (8) (2012) 813–822.

[23] K.L. Moore, W.J. Boscardin, M.A. Steinman, J.B. Schwartz, Age and sex variation in prevalence of chronic medical conditions in older residents of US nursing homes, Journal of the American Geriatrics Society 60 (4) (2012) 756–764.

[24] C.A. Lexau, R. Lynfield, R. Danila, T. Pilishvili, R. Facklam, M.M. Farley, L.H. Harrison, W. Schaffner, A. Reingold, N.M. Bennett, J. Hadler, P.R. Cieslak, C.G. Whitney, Active Bacterial Core Surveillance Team, Changing epidemiology of invasive pneumococcal disease among older adults in the era of pediatric pneumococcal conjugate vaccine, Journal of the American Medical Association 294 (16) (2005) 2043–2051.

[25] Y. Li, P. Ning, X.S. Wang, S. Jajodia, Discovering calendar-based temporal association rules, Data & Knowledge Engineering 44 (2) (2003) 193–218.

[26] W. Wang, J. Yang, R. Muntz, TAR. Temporal Association Rules on Evolving Numerical Attributes, in: Proceedings of the 17th International Conference on Data Engineering, 2001, pp. 283–292.

[27] G. Adomavicius, J. Bockstedt, C-TREND. Temporal cluster graphs for identifying and visualizing trends in multiattribute transactional data, IEEE Transactions on Knowledge and Data Engineering 20 (6) (2008) 721–735.

[28] R. Law, J. Rong, H.Q. Vu, G. Li, H.A. Lee, Identifying changes and trends in Hong Kong outbound tourism, Tourism Management 32 (5) (2011) 1106–1114.

[29] D. Leman, A. Feelders, A. Knobbe, Exceptional Model Mining, in: Proceedings of the European conference on Machine Learning and Knowledge Discovery in Databases – Part II, 2008, pp. 1–16.

[30] K.M. Simonic, A. Holzinger, M. Bloice, J. Hermann, Optimizing long-term treatment of rheumatoid arthritis with systematic documentation, pervasive health, in: 5th International Conference on Pervasive Computing Technologies for Healthcare, Dublin, IEEE, 2011, pp. 550–554.