

Application Association Rule Mining in Medical-biological Investigations: a Survey

Xenia Naidenova, Vyacheslav Ganapolsky, Alexandre Yakovlev, Tatiana Martirova

Military Medical Academy, Saint Petersburg, Russia
ksennaidd@gmail.com

Abstract. In this chapter, a survey is proposed related to the application of Apriori-like algorithms in medical and biological investigations for discovering frequent sets of attributes values in data and then extracting logical rules in the form of implication dependencies between values of observed and measured attributes and diagnostic parameter

Keywords: Apriori algorithm, association rule mining, medical investigation

1 Introduction

Intelligent data processing is now an integral part of biomedical research. Revealing different dependencies in the data (implicative, functional, correlational, etc.) helps in diagnosis, treatment's planning, predicting the course of diseases and in identifying new factors that expand the understanding of specialists about specific diseases and their combinations.

The purpose of many biomedical studies is to highlight associative rules in a given data set. The association rule is the rule in the form $X \Rightarrow Y$, where X and Y are non-intersecting sets of distinct literals called items. In general case, we can consider a set of items as a set of all attributes' values that can appear in descriptions of some objects or situations (transactions) in a data base.

Let $I = \{i_1, i_2, \dots, i_N\}$ be a set of items. A database (TDB) is a set of transactions, where transaction $\langle \text{tid}, X \rangle$ contains a set of items (i.e., $X \subseteq I$) and is associated with a unique identifier tid . A non-empty itemset $Y \subseteq I$ is a q -itemset if it contains q items.

A transaction $\langle \text{tid}, X \rangle$ is said to contain itemset Y if $Y \subseteq X$. The number of transactions in TDB containing itemset X is called the support of X , denoted as $\text{sup}(X)$: $\text{sup}(X) = |\{\text{tid} \mid (\text{tid}, Y) \in \text{TDB}, X \subseteq Y\}|$, where $|s|$ denotes the cardinality of s . Giving a minimum support threshold, min-sup , an item Y is frequent if $\text{sup}(Y) \geq \text{min-sup}$.

In frequent itemsets, association rules are extracted in the form of implications, for which the value of support is an important characteristic: $\text{Sup}(\text{rule}) = \text{sup}(X \Rightarrow Y) = \text{sup}(X \cup Y)$. The rule has a measure of reliability called confidence and defined as follows: $\text{conf}(X \Rightarrow Y) = \text{sup}(X \cup Y) / \text{sup}(X)$. Confidence is defined as the part of all transactions in TDB containing X and Y , among those transactions that contain X .

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The third main metric of association rules is Lift defined as follows: $\text{lift}(X \Rightarrow Y) = \text{conf}(X \Rightarrow Y) / \text{sup}(Y)$. This measure quantifies the predictive power of rule $X \Rightarrow Y$.

The traditional purpose of data analysis is to find all associative rules (ASRs) that have support and confidence above the specified minimum values.

In this paper, we consider the applications of the classical Apriori algorithm for mining ASRs in biomedical data. The choice of this algorithm is justified by the following arguments: it is the most understandable and easy mastered by specialists in different applied fields, it is universal, in the sense that ASRs containing in frequent itemsets can be extracted with the help of this algorithm [30]; it is constantly improving based on many progressive techniques in ASR mining.

The idea of the classical Apriori algorithm [2] is based on the following consideration: q -itemset can be frequent if and only if all its proper sub-itemsets are frequent.

At the first step of the algorithm, all the items are considered (values of attributes, elements of transactions) and among these are separated those satisfying the condition of minimal support. Then the separated items are used to form itemsets of two items (candidates for frequency). For them, the support is calculated and those that do not meet the minimum support are removed. The remaining itemsets are used to form ones of three items. The process is going on iteratively, as long as it is possible to generate a new set of candidates for frequency. The Apriori algorithm uses an inductive method of constructing sets of the cardinality $(q+1)$ ($(q+1)$ -sets) from their subsets of the cardinality q (q -sets). The method of forming $(q+1)$ -sets from q -sets and calculating their supports are the main sub-processes of the Apriori algorithm determining its computational complexity.

The paper is organized as follows. The Section 2, 3 contain brief surveys of applying the Apriori algorithm in medical studies. The perfection of the Apriori algorithm is considered in Section 4 followed by a small conclusion section.

2 The Apriori Algorithm in Medical Studies

The Apriori algorithm is widely used in medical research. These studies cover: cardiovascular disease [23, 29, 41]; lung cancer [18]; oral cancer [39]; infectious diseases (Ebola virus) [13]; type 2 diabetes [26, 34, 37, 45, 46]; Alzheimer's disease [7]; liver cancer [32]. In [41], some other diseases are enumerated for the study of which were used the Apriori algorithm or its modifications: asthma, impotence, lupus, obesity, whooping cough, pregnancy, phenomenon Raynaud syndrome. The problems solved are also varied: searching for unknown trends in disease; determining the nature of disease based on a prediction method; diagnosis (detection) of disease; predicting a patient's response to drug [38, 44]; early diagnosis and prevention of disease [37]; prediction of illness's progress (course of disease); predicting the outcome of disease [23]; identification of disease risk factors [34]; identification of relationships between different medical operations, appointments, analyses and diagnoses [35]; extracting diagnostic patterns (sets of features, symptoms) and association rules in electronic medical database [1, 11, 19, 22] and many others.

The papers [3, 5, 41, 51] give detailed reviews of ASR mining based on Apriori algorithm and its modifications. In [39], early detection and prevention of oral cancer is considered and the Apriori algorithm is used to assess the chance of patients' survival. This is achieved by extracting a set of significant rules among various laboratory tests and investigations like FNAC (Fine Needle Aspiration Cytology) of neck node, LFT (Liver Function Tests), Biopsy, USG (Ultra-SonoGraphy), CT scan or MRI (computer tomography), and survivability of the oral cancer patients. Liver Function Tests (LFT) give information about the state of a patient's liver. Most liver diseases cause only mild symptoms initially, but it is vital that these diseases be detected early. Biopsy is important, as it is the only sure way to know if the abnormal area is cancer. USG is an ultrasound based diagnostic imaging technique to visualize intra-abdominal structures. The extracted ASRs clearly show that if FNAC of neck node, USG and CT scan/ MRI is positive, then chance of survival is reduced. However, if LFT is normal, the probability of survival is high. All the generated rules hold the highest confidence level.

Nahar J. et al. [31] extract the significant prevention factors for particular types of cancer: bladder, breast, cervical, lung, prostate and skin cancer. The algorithms Apriori, Predictive Apriori, and Tertius are used to discover most of the significant prevention factors against these specific types of cancer. Predictive Apriori tries to maximize expected accuracy rather than confidence in Apriori, Tertius is a top-down rule discovery system employing well known decision tree algorithm.

The article [6] describes the methods to detect the comorbidity of professional diseases caused by pathogenic factors, namely, by ionizing radiation. Apriori algorithms implemented in the environment of the SQL Server Analysis Services Data Mining. The following diseases are considered: chronic radiation sickness of degree 1, 2 and 3, residual phenomena of chronic radiation sickness, exposure to ionizing radiation, malaise and fatigue, and vegetative nervous system disorders.

Numerous papers deal with the use of the Apriori algorithm to study coronary heart disease. Karaolis et al. [20] developed a data analysis system to search for associations to assess the heart disease risk factors with WEKA tools. In [29], the objective of the study was to effectively predict possible heart attacks from a patient dataset. The Data Base consisted of 209 records (instances) collected from a hospital in Iran and 8 attributes. Apriori algorithms were implemented in WEKA 2016 (version 3.9.0) and MATLAB R2013a software. The preprocessing of data consisted of previous purifying data through the Discretization Unsupervised Filter and a discretizing method to change numeric data into nominal data. The algorithm implemented in MATLAB showed the best results related to the diagnosis prediction accuracy by the use of the obtained ASRs.

It should be noted that there are still very few works in the domestic literature on the use of the Apriori algorithm in biomedical research. In addition to the work of Bi-ryukov A. and Dumansky S. [6], we can cite the article [5], which proposes a new effective algorithm AprioriScale to build ASRs. The algorithm is applied to the problem of detecting children's diseases: obesity and metabolic syndrome. Two problems are solved: 1) the possibility of distinguishing between obesity and metabolic syndrome, and 2) identifying a combination of risk factors determined in the early stages

of a child's development that may indicate the onset of the disease in adolescence. 12 rules with the reliability above 0.75 were obtained. Let's take a closer look at two of these rules:

- Toxicosis, TIP, WeakLD → MS frequency = 0,16, probability = 0,83;
- Toxicosis, TIP, ExtraGM, PPH → MS frequency = 0,09, probability = 0,93,

where TIP – threat of interruption of pregnancy, MS – metabolic syndrome, WeakLD – weak labour delivery, and ExtraGM – extragenital diseases of mother, PPH is post-prandial hyperlipemia. These rules show that if factors such as toxicosis, TIP, ExtraGM, later accompanied by factors such as WeakLD, PPH are observed in the early stages of a child's development, the risk of metabolic syndrome is high, the probability of its manifestation is between 0.83 and 0.93.

3 Analysis of Biological and Genetic Data Based on Association Rule Extracting

A topic dealing with the analysis of patient biological data is now becoming particularly relevant. The biological data analysis is connected with the identification of previously unknown hidden patterns (frequent itemsets), associative structures in the large number of biological sequences. These sequences include gene sequences, amino acid sequences, protein composition, and other data that display the structure, localization, interaction or functioning of proteins and genes in cells. Amino acids are the building material of proteins. The shape and other properties of proteins are associated with the exact sequence of amino acids contained in them. The chemical properties of amino acids determine the biological activity of proteins.

Many diseases have biological nature: obesity, high blood cholesterol, diabetes, insomnia, arthritis, and many others. Analysis of gene information, including Apriori algorithms, [1, 10, 21, 27] helps to study the nature of disease, optimize its treatment, predict the course of disease. An overview of some methods of extracting knowledge from biological (DNA) sequences is given in [8]. The comparison of the Apriori algorithm with other algorithms in the mutation analysis is produced in [28]. In [17], a model is proposed for finding a dominant sequence of amino acids to block the growth of cancer cells based on protein clustering.

In [16], a comparative analysis of classifiers in cancer prediction using multiple data mining techniques is given. The dataset contained a total of 844 records and nine features: Clump thickness, Uniformity of cell size, Uniformity of cell shape, Marginal adhesion, Single epithelial cell size, Bare nuclei, Bland chromatin, Normal nucleoli and Mitoses. The analysis of data consisted of two stages. The first stage applies the algorithm Apriori to reduce the number of input features. In the second stage, six classifiers have been applied and validated through a k-fold cross-validation scheme. Some variables named as Marginal adhesion and Bare nuclei have been removed as the noise data. Thus, a new subset of seven features has been provided for solving the problem of classification. Six predictive algorithms were chosen: decision tree (DT), support vector machine (SVM), k-nearest neighbour (KNN), naïve Bayes (NB), ran-

dom forest (RF) and neural network (NN). For the experiments, the R statistical environment was chosen, as it is an open source scripting language specifically designed for data analysis. The classifiers were evaluated based on performance metrics including accuracy, sensitivity, and specificity. The SVM classifier achieved a classification accuracy of 0.9372 with a sensitivity of 0.9332 and a specificity of 0.9226, so it performs better than all the remaining classifiers.

A similar method of two-stage data processing is used in [44] to predict a patient's response to a drug in the treatment of cancer. For prediction of drug response based on molecular profiles of multiple cancer cell types, it was generated a large-scale pharmacogenomics dataset for 1001 cancer cell lines and 251 anti-cancer drugs. The authors performed the feature selection in the form of ASRs and utilized the selected features to train the state-of-the-art Deep Learning Neural Networks (DLNNs) to predict pharmacological response in a blind (control) set. The ASRs are treated as a novel meta-dataset. Specifically, the Apriori algorithm was applied to generate a rule-set, containing all tissue-to-gene, tissue-to-drug, gene-to-drug and drug-to-drug associations.

The study shows that type 2 diabetes is a genetic disease and evidence of a statistical interaction among several Single Nucleotide Polymorphisms (SNPs) has been reported. In [26], the algorithm Apriori-Gen has been applied to SNP data of type 2 diabetes for association study. The obtained associations are measured through risk rate (RR) and odds ratio (OR) proposed by the authors. The obtained results allow to assess with high accuracy and statistical reliability the interaction of nucleotide polymorphisms with disease complexes. An analysis of diabetic ASR based on the Apriori algorithm is given in [45].

The volume of biological knowledge is rapidly increasing in the form of gene expression databases (GEO, Arrayexpress, etc.), information on microarray experiments (spotted probes, data processing protocols, etc.), molecular databases (GenBank, Embl, Unigene, etc.), semantic sources as thesaurus, ontologies or semantic networks (UMLS, GO, etc.), bibliographical databases (Medline, Biosis, etc.) and gene/protein related specific sources (KEGG, OMIM, etc.) [27].

4 Perfection of the Apriori Algorithm

The popularity of the Apriori algorithm for medical diagnostic tasks is due to its simplicity, however, its application for large data sets requires the development of more efficient modifications in terms of reducing its computational complexity. And such work to improve this algorithm is being carried out all over the world [4]. In particular, we may be able to choose the following directions in improving the Apriori algorithm: 1) developing new algorithms; 2) data management; 3) constraint-based ASR mining; 4) incremental mode of ARS mining.

Developing New Algorithms. The algorithms FP-GROWS and ECLAT are attributable to the first direction. The main drawbacks of the Apriori algorithm is scanning the database several times. The algorithm FP-GROWTH [14, 32] uses a frequent-pattern tree structure (FP-Tree), which stores all the database. This structure

can compress the data up to 200 times, and it is stored to the computer's memory. Then, frequent itemsets are directly extracted from the FP-Tree using the divide-and-conquer method. This algorithm is used for analyzing risk factors of Type 2 diabetes in [46].

The algorithm "Equivalence Class Transformation" (ECLAT) mines frequent itemsets in a vertical data format [12]. The algorithm builds the TID set of all items in the transaction database. The ASR mining on vertically partitioned data is used in [15].

The article [45] gives a general overview of effective processing the Apriori algorithm on medical data. The authors propose a modification of the Apriori algorithm, in which the amount of support for many candidates for frequency multiple attributes is calculated only for transactions, the length of which is longer or equal to the cardinality of the candidates in question. The study in [9] aims to see the effect of the k-means clustering algorithm on the Apriori algorithm by combining these two algorithms. A logical combinatorial neuron-like network is advanced for optimization of the Apriori algorithm in [30].

Data Management. The following articles can be attributed to the category of data management. In [37], the ASRs are extracted to predict the co-diseases in diabetic mellitus patients. The peculiarity of this work is in selecting rules via their testing on a sample of data not used in data processing. The efficiency of the Apriori algorithm was increased with the help of a prefixed-itemset-based data structure [49]. In paper [38], a new method and a statistic test on rules were introduced to mine ASR over multiple databases. In [42], a method for ASRs extraction based on ontology semantics is proposed. The medical dataset is transformed into an ontology in the form of triples (subject, object, predicate), and SPARQL (Query Results XML Format) is used to query the generated ontology.

Constraint-Based Association Rules Mining. Currently, a lot of studies appeared in which the Apriori algorithm is optimized from the point of view of obtaining not all the possible set of ASRs but only some of its subset satisfying a given essential property – interesting rules, non-redundant, negative, maximal rules, association rules generated through the questions of the end users and some others.

The authors of [1] introduce a Query-constraint based ASR Mining (QARM) approach for exploratory analysis of multiple, diverse clinical data sets in the National Sleep Research Resource (NSRR). Top-k Non-Redundant (TNR) ASR mining algorithm is used in this work. Non-redundant ASR is rule deleting from which at least one item implies the change of rule's support for the worse. The work [36] also proposes an algorithm to generate non-redundant ASRs.

Both positive and negative rules were generated in [27] to analyze which diagnosis types require or not require Laboratory Diagnostic Tests (LDTs) for patients. The negative rules are generated from infrequent itemsets.

Maximal ASRs are extracted from maximal frequent patterns (itemsets). A pattern X is a maximal frequent in data set D if X is frequent, and there exists no super-pattern x such that $X \subset x$ and x is frequent in D . The maximal frequent pattern mining algorithms is, for example, Maximal Frequent Itemset Algorithm (MAFIA) [14, 21].

Currently there is considerable interest in the methods restricting the extraction of rules to the specific type of the most interesting rules for the users. In [33], it is ad-

vanced a method of Rank Based Weighted ASR Mining (RANWAR)) elaborated for biological data processing. Two new measures of interestingness are considered based on ranging gens.

In [43], a method for generating efficient rules for associative classification is advanced. Associative classification is a technique that integrates classification and ASRs mining for classifying unseen data. Associative classification gives more accurate and easier to understand rules than it is possible to obtain by using the traditional classifiers. In [37, 40], mining ASRs is also combined with extracting classification rules.

Existing ASRs mining algorithms rely on frequency-based rule evaluation methods failing to provide sound statistical or computational measures for rule evaluation, and often suffer from many redundant rules. In [37], the authors propose predictability-based an ASRs mining algorithm based on cross-validation with a new rule evaluation step. A training dataset is partitioned into inner training sets and inner test sets and then candidate rules' predictive performance is evaluated.

Incremental Mode of Association Rules Mining Traditional static ASRs mining cannot solve real-world problems with dynamically changing data. When the size of the transaction database increases, then an initially frequent item may become an infrequent one, and an initially infrequent item may become a frequent one. Incremental ASRs mining algorithms help to copy with the drawbacks of the classical Apriori algorithm. The authors [52] combine the Fast Update Pruning (FUP) algorithm with a compressed Boolean matrix [25] and propose a new incremental ASRs mining algorithm, named FBCM. This algorithm requires only a single scan of both the database and incremental database. An incremental algorithm for mining interesting ASRs has been developed in [48]. The papers [50] summarize the methods for incremental ASRs mining.

Conclusion

The paper provides an overview on mining associative rules from data in biomedical research. This review is based on a study of the work from 2013 to 2020 and shows the widespread use of the Apriori algorithm and its modifications in medicine. The review includes also the methods to improve the Apriori algorithm to mining more effective associative rules adapted for various research tasks.

References

1. Abeysinghe, R. and Cui, L.: Query-constraint-based mining of association rules for exploratory analysis of clinical datasets in the National Sleep Research Resource. *Medical Informatics and Decision Making* 18 (Suppl 2), 89-100 (2018)
2. Agrawal, R., Imieliński, T., and Swami, A.: Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 22(2), 207-216 (1993)
3. Altaf, W., Shahbaz, M., and Guergachi, A.: Applications of association rule mining in health informatics: A survey. *Artificial Intelligence Review*, 47(3), 313-340 (2017)

4. Bhende, D., Kasarker, U., and Gedam, M.: Study of various improved Apriori algorithm. *IOSR Journal of Computer Science Engineering*, e-ISSN:2278-0661, pp. 55-58 (2016)
5. Billig, V. A., and Ivanova, O.V.: Building association rules in a task of medical diagnosis. *Software Products and Systems*, 2(114), 146-157 (2016) (in Russian).
6. Birukov, A. P. and Dumansky, S. M.: Revealing the comorbidity of professional diseases caused by pathogenic factors with the help of associative algorithms on the examples of a cohort of victims of ionizing radiation. *Medicine of Extreme Situations*, 2, 13-24 (2016) (in Russian).
7. Chavez, R., Górriz, J. M., Ramírez, J., Salas-Gonzalez, D., and Gómez-Rio, M.: Efficient mining of association rules for the early diagnosis of Alzheimer's disease. *Phis Med Biol.*, 56, 6047-6063 (2011)
8. Das, N. N. et al.: Brief survey on DNA sequence mining. *International Journal of Computer Science and Mobile Computing*, 2(11), 129-134 (2013)
9. Dharshini, N. P., Azmi, F., Fawwaz, I., Husein, A. M., and Siregar, S. D.: Analysis of accuracy k-means and Apriori algorithms for patient data clusters. *Journal Physics.: Conference Ser.* 1230, 1-9. IOP Publishing (2019)
10. Dhumale S.: Predicting patterns over protein sequences using Apriori algorithm. *International Journal of Engineering and Computer Science* 4(7), 13011-13016 (2015)
11. Doddi, D., Marath, A., Ravi, S.S., and Torney, D.C.: Discovery of association rules in medical data. *Medical Informatics and Internet in Medicine* 26(1), 25-33 (2001)
12. Domadia, N. and Rao, U.P.: Privacy-preserving distributed association rule mining approach on vertically partitioned healthcare data. *Procedia Computer Science*, 148, 303-312 (2019)
13. Go, E., Lee, S., and Yoon, T.: Analysis of ebolavirus with decision tree and Apriori algorithm. *International Journal of Machine Learning and Computing*, 4(6), 543-548 (2014).
14. Han, J., Pei, J., and Yin, Y.: Mining frequent patterns without candidate generation. In: *Proceedings of ACM SIGMOD International Conference on Management of Data*, pp. 1-12 (2000)
15. Harahap, M., Husein, A.M., Aisyah, S. et al.: Mining association rules based on diseases population for recommendation of medicine need. *Journal of Physics: Conference Series*, 1007, 1-12 (2018)
16. Jalali, S. M., Moro, S., Mahmoudi, M. R., Ghaffary, K. A., Maleki, M., and Alidoostan, A.: A comparative analysis of classifiers in cancer prediction using multiple data mining techniques. *International Journal of Business Intelligence and Systems Engineering*, 1(2), 166-178 (2017)
17. Kalaiyarasi, R. and Prabasri, S.: Predicting the Lung Cancer from Biological Sequences. *International Journal of Innovation in Engineering and Technology (IJJET)*, 5(1), 106-111 (2015)
18. Kanageswari, S. and Gladis, D.: Generation of association rules of data mining for lung cancer by air pollution. *International Journal of Engineering and Advanced Technology (IJEAT)*, 9(3), 2874-2880 (2020)
19. Kang'ethe, S. M. and Wagacha, P. W.: Extracting diagnosis patterns in electronic medical records using association rule mining. *International Journal of Computer Applications* 108(15), 19-27 (2014)
20. Karaolis, M., Moutris, J.A., Papaconstantinou, L., and Pattichis, C. S.: Association rule analysis for the assessment of the risk of coronary heart events. In: *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6238-6241. Minneapolis, MN: USA (2009)

21. Kavakiotis, I., Tzanis, G., and Vlahavas, I.: Mining frequent patterns and association rules from biological data. In: Mourad Elloumi and Albert Y. Zomaya (eds.), *Biological Knowledge Discovery Handbook: Preprocessing, Mining, and Postprocessing of Biological Data*, First Edition, Chapter 34, pp. 737-762. John Wiley & Sons, Inc. Published (2014)
22. Kumar, R. P. R., Jayakumar, R., and Sankaridevi, A.: Apriori-based frequent symptomset association mining in medical databases. *International Journal of Recent Technology and Engineering (IJRTE)*, 7(5C), 65-68 (2019)
23. Lakshmi, K.R., Krishna, M.V., and Kumar, S.P.: Performance comparison of data mining techniques for predicting of heart disease survivability. *International Journal of Scientific and Research Publications*, 3(6), 1-10 (2013)
24. Lee, J.Y. and Kim, K.-Y.: Semantic and association rule mining-based knowledge extension for reusable medical equipment random forest rules. *Journal of Integrated Design and Process Science*, 22(4), 55-81 (2018)
25. Li, T. and Luo, D.: A new improved Apriori algorithm based on compression matrix. In: *Proceedings of International Conference on Advanced Data Mining Application.*, vol. 8933, pp. 1-15. Berlin, Germany: Springer-Verlag (2014)
26. Mao, W. and Mao, J.: The Application of Apriori-Gen algorithm in the association study in type 2 diabetes. In: *Proceedings of 3rd International Conference on Bioinformatics and Biomedical Engineering (ICBBE)*, 4(1), 126-140 (2009)
27. Martinez, R., Pasquier, C., and Pasquier, N.: GENMINER: Mining informative association rules from genomic data. In: *Proceeding of the IEEE International Conference on Bioinformatics and Biomedicine*, pp. 15-22 (2007)
28. Mayilvaganan, M. and Hemalathe, R.: Performance comparison of FSA Red and Apriori algorithm's in mutation analysis. *International Journal of Computer Trends and Technology (IJCTT)*, 17(4), 205-209 (2014)
29. Mirmozaafari, M., Alinezhad, A., and Gilanpour, A.: Data mining Apriori algorithm for heart disease prediction. *International Journal of Computing, Communication & Instrumentation Engineering (IJCCTE)*, 4(1), 20-23 (2017)
30. Naidenova, X., Parkhomenko, V., and Shvetsov, K.: Application of a logical-combinatorial network to symbolic machine learning tasks. In: Naidenova, X., Shvetsov, K., and Yakovlev, A. (eds.), *Machine Learning in Analysis of Biomedical and Socio-Economic Data*, pp. 371-409. Saint-Petersburg, Russia: Polytech Press (2020)
31. Nahar, J., Tickle, K., Ali, A.B.M.S., and Chen, Y.P. Significant cancer prevention factor extraction: An association rule discovery approach. *Journal of Medical System* 35(3), 353-367 (2011)
32. Pinheiro F.M.R.: Applying the Apriori and FP-Growth association algorithms to liver cancer data. A Thesis Submitted in Partial Fulfillment of the Requirement for the Degree of Master of Science, 172 pp. University of Victoria (2013)
33. Premalatha S. and Nandhini, U. C.: Efficiently generating the rank based weighted association rule mining using Apriori algorithm in High Biological Database. *International Research Journal of Engineering and Technology (IRJET)*, 2(9), 2143-2147 (2015)
34. Ramezankhani, A., Pournik, O., Shahrabi, J., Azizi, F., and Hadegh, F.: An application of association rule mining to extract risk pattern for type 2 diabetes using tehran lipid and glucose study database. *International of Endocrinol Metabolism*. 13(2) e25389 (2015)
35. Saryer, G. and Taşar, C. Ö.: Highlighting the rules between diagnosis types and laboratory diagnostic tests for patients of an emergency department: use of association rule mining. *Health Informatics Journal*, 2(6), 1-17 (2019).
36. Séverac, F., Sauleau, E. A., Meyer, N., Lefèvre, H., Nisand, G., and Jay, N.: Non-redundant association rules between diseases and medications: an automated method for

- knowledge base construction. *BMC Medical Informatics and Decision Making* 15(1), 2-7 (2015)
37. Shahebaz, Ah. Kh. and Jabbar, M. A.: Improved classification techniques to predict the co-disease in diabetic mellitus patients using discretization and Apriori algorithm. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(11), 730-733 (2019)
 38. Shang, E., Duan, J., Fan, X., Tang, Y., and Ye, L.: Association rule mining and statistic test over multiple datasets on TCM drug pairs. *International Journal of Biomedical Data Mining* 6(1), 2-6 (2017)
 39. Sharma, N. and Om, H.: Early detection and prevention of oral cancer: association rules mining on investigation. *WSEAS Transaction on Computers*, 13, 1-8 (2014)
 40. Song, K. and Lee, K.: Predictability-based collective class association rule mining. *Expert System with Application*, 79, 1-7 (2017)
 41. Swathi, P. and Prajna, B.: The effective procession of Apriori algorithm prescribed data mining on medical data. *International Journal of Computer Science and Technology (IJCST)*, 7(3), 22-26 (2016)
 42. Thamer, V., El-Sappagh, S., and El-Shishtawy, T.: A semantic approach for extracting medical association rules. *International Journal of Intelligent Engineering Systems*, 13(3), 280-293 (2020)
 43. Thanajiranthorn, Ch. and Songram, P.: Generation of efficient rules for associative classification. In: *Proceedings of the 13th International Conference MIWAI*, pp. 109-120 (2019)
 44. Vougas, K., Krochmal, M., Jackson, T., Polyzos, A. et al: Deep learning and association rule mining for predicting drug response in cancer. A personalized medicine approach. Current version available <https://www.biorxiv.org/content/10.1101/070490v3.full> (2017)
 45. Wang, X., Su, K., and Liu, Zh.: Analysis of diabetic association rules based on Apriori algorithm. In: Ch. Huang, Yu-Wei Chan, and Neil Yen (eds), *Data Processing Techniques and Applications for Cyber-Physical System*, pp. 553-563. Springer (2020).
 46. Wei, Z., and Guangjian, Ye.: The research on analyzing risk factors of type 2 diabetes mellitus based on improved frequent pattern tree algorithm. In *Proceedings of the 2015 Internal Conference on Materials Engineering and Information Technology Applications*, pp. 459-463. Atlantic Press (2015)
 47. Win, S.L., Htike, Z.Z., Yusof, F., and Noorbata, I.A.: Gene expression mining for survivability of patients in early stages of lung cancer. *International Journal of Bioinformatics and Biosciences*, 4(2), 1-9 (2014).
 48. Yafi, E., Al-Hegami, A., Alam, A., and Biswas, R. YAMI: Incremental Mining of Interesting Association patterns. *The International Arab Journal of Information Technology* 9(6), 504-5011 (2012)
 49. Yu, S. and Zhou, Y.: A prefixed-itemset-based improvement for Apriori algorithm. Cornell University, arXiv preprint, arXiv: 1601.01746 (2016)
 50. Zhan, F., Zhu, X., Zhang, L., Wang, X., Wang, L., and Liu, Ch.: Summary of association rules. *IOP Conf. Series: Earth and Environment Science (ESMA)*, vol. 252, pp. 12 (2019).
 51. Zhang, B.Z., Jiang, K. Q., and Zhang, Y. Z.: Survey on incremental association rule mining research. *Journal of China Computer. Systems*, 37(1), 18-23 (2016)
 52. Zhou, D., Ouyag, V., Kang, Zh., Li, Zh., Zhou, J.P., and Cheng, X.: Incremental ARM based on matrix compression for edge computing. *IEEE Access, Section on Innovation and Application of Intelligent Processing of Data, Information and Knowledge as Resources in Edge Computing*, vol. 7, 173044-173053 (2019)