Review of Rational Drug Use Based on Apriori Algorithm

Zhangming Luo¹, Binjie Cheng¹, Sen Tian², Anding Hong¹, Jin Zhang¹, Qiang Li¹,*

¹College of Information Science and Engineering, Hunan Normal University, Changsha 410081, China
²School of Mathematics and Statistics, Hunan Normal University, Changsha 410081, China

*32246026@qq.com

Abstract: One of the important reasons for irrational drug use is that the medical industry has been facing the problem of complex and redundant information. In order to solve this problem, a multitude of researchers carried out research on rational drug use by data mining technology and tried to apply it to real medical treatment. Apriori is one of the most frequently used and valuable algorithms. Based on the Apriori algorithm, we discussed the basic theory and the improved result. We elaborated on the detail and analyzed the development of the applied scenarios in modern medicine and traditional Chinese medicine. Then we discovered the applied scenarios of the Apriori algorithm mainly included medication law, dose research, and experience inheritance. Through the analysis of the literature on the use of the Apriori algorithm in rational drug use and the current development trend, this paper pointed out the existing problems in this field and put forward the future work and research focus.

Keywords: Rational drug use; Apriori algorithm; data mining; traditional Chinese medicine; experience inheritance

1. Introduction

According to statistics, about 1/3 of the global annual deaths are due to irrational drug use [1], which poses a major threat to human life. In 1985, the World Health Organization defined rational drug use as requiring patients to receive drugs suitable for their clinical needs, dosage in line with their individual needs, adequate course of treatment, and the drug price is the most favorable for patients.[2]

It is arduous to make a completely reasonable drug prescription for every patient just by the knowledge and experience of doctors. Therefore, it is necessary to research how to make full use of these data quickly and accurately, which provides correct decision support for rational drug use. The characteristic of data mining technology is that it can extract useful knowledge effectively from a range of data which could be random, unclear, and incomplete.[3] In the early stage, researchers just used some tools to deal with medical data with the data mining algorithm, most of those are...
association rule algorithm, such as Apriori. Result of the low mining efficiency and the poor quality of rules, researchers gradually focus on improving the algorithm and combining with other data mining algorithm to achieve better performance.

This paper introduced the Apriori algorithm and the application based on the Apriori algorithm. We concluded the development status of Apriori in the field of rational drug use and put forward some opinion about the combination of rational drug use and existing algorithms or traditional Chinese medicine for researchers.

| Time | Researcher | Improvement | Effect |
|------|------------|-------------|--------|
| 2009 | Wang Hua[5] | Restricted the item's appearance position, maximum support degree selection, and minimum support selection. | $O((2^{|I|}+|I|-1))$ Useless rules will be deleted after the restrictions are added to legal rules. |
| 2010 | Cheng Yuan[6] | Apriori_Gen algorithm was optimized. | After the improvement, only $p * (|L_{K-1}| + 1)$ operations are needed, compared with $2^{|L_{K-1}|}$ operations before the improvement, the efficiency is improved. |
| 2014 | Yuhan Zhou[7] | By using the fuzzy enquery method, the most possible data could be extracted from the database at one time. | It only needs to scan the database once instead of k times of the original algorithm. |
| 2015 | Wang Feng[8] | Proposed to use confidence evaluation rule instead of a support degree. | It can be used in the case of high confidence and low support, which can effectively avoid deleting useful rules. |
| 2016 | WANG Renli[9] | Used the bit concept in mathematics to improve the fast response of strings. | Solve the problem of duplicate scanning. |
| 2018 | Noguchi, Y[10] | By using large-scale data mining association rules instead of analysis. | It has a 100% recall rate and accuracy in the data set of chronic kidney disease. In the heart disease data set, the recall rate is 87%, but the accuracy rate is 70%, which is lower than the 100% accuracy rate of the classification based on the association rules. |
| 2019 | Rjeswari[11] | By detecting the existence of outliers to improve the accuracy and prediction rate of the disease. | |

Tab.1 Main improvement direction of Apriori algorithm

2. Introduction of Apriori algorithm

2.1. The classic Apriori algorithm
R. Agrawal[4] proposed the Apriori association rule mining algorithm in 1994, which could scan the data set to calculate the support of candidate item-sets to determine whether the candidate item-sets are frequent. It is mainly used in the case of short frequent patterns and a small number of data sets. Be result in the number of times the Apriori algorithm scans the data set is related to the length of the longest frequent pattern, usually, the efficiency of the algorithm is reduced greatly when the number of data set is large and the length of the frequent pattern is long. Because in the process of generating frequent item-sets, the huge database will be scanned repeatedly. Apriori_Gen algorithm will produce candidate item-sets, which will undoubtedly increase the computational complexity of the algorithm. For example, if there are 5 items in frequent itemset 1, the number of candidate sets is 10. If there are 10000 items in frequent itemset 1, the number of candidate sets will reach $c_{10000}^2$. Therefore, the amount of calculation becomes extremely large. It can be seen that the amount of calculation will increase with the number of frequent item-sets. Appropriate improvement will make the Apriori algorithm more widely used in the field of rational drug use.

2.2. The improved Apriori algorithm.

2.2.1. Main improvement direction

As shown in Tab.1, Wang Hua[5], Cheng Yuan[6], Zhou[7], and other researchers have studied the algorithm itself or the process of generating frequent item-sets and the effective reduction of rules, trying to solve the inherent problems of Apriori algorithm. Wang[9] proposed the Apriori-BSO algorithm. This research mainly focused on the use of logical operation of bit string to solve the repeated scanning problem. In addition, Wang Feng [8] proposed that the use of support as the evaluation standard in some scenarios is not reasonable, accordingly, they proposed to use the confidence level to evaluate the rules rather than the support. Rajes wari[11] combined the Apriori algorithm with other algorithms and proposes an association classification technology based on Apriori rare to predict the unpredictable problems in the medical field. It can detect outliers well and is more predictive than the classification based on the association rules.

2.2.2. Convert transaction database to the matrix

Some researchers tried to improve with the Boolean matrix. As shown in Tab.2, Liu[12], Zhang [13], Li [14], Zhang [15], and others set out to transform the database into a multiple Boolean matrix in an attempt to improve the mining efficiency.

| Time  | Researcher  | Improvement |
|-------|-------------|-------------|
| 2012  | Liu Zhi[12] | Vapriori algorithm was proposed, that is, after a single scan, the transaction database is transformed into a Boolean matrix, and the transaction database scanning is transformed into vector operation. |
| 2019  | Zhang Chong[13] | Through a single scan, the transaction database is transformed into a Boolean matrix, and the matrix is compressed according to specific properties. |
| 2020  | LiYa[14] | In order to construct the Boolean matrix corresponding to the data, the transaction attributes should be transformed into boolean variable values (value 0 or 1) suitable for mining, and then mined the corresponding frequent item-sets. |

Tab.2 Research on the transformation matrix
| Time  | Researcher | Improvement |
|-------|------------|-------------|
| 2014  | Ji\cite{16} | Develop an association mining algorithm based on SQL. This algorithm can deal with multiple relational data tables directly. |
| 2014  | Reps.\cite{17} | By eliminating the confusion caused by the rules to refine the association, from the retention of the real adverse drug reactions corresponding to the rules. The strength of association rules is measured by minimum left support and confidence. |
| 2018  | Chen\cite{18} | Pruning, the JOIN algorithm, the searching for a frequent itemset, and frequent 1-item set of Apriori algorithms were optimized. |

Tab.3 Analysis and improvement of the ADR report mining process

According to the improved algorithm, transaction attributes need to be transformed into boolean variable values suitable for mining (value is 0 or 1), so as to construct a boolean matrix corresponding to data, and then mine corresponding frequent item-sets. Transaction i and item-set j correspond to row vector and column vector respectively. After the transaction database D is scanned, if the item j is set in the transaction i, the corresponding value of Row i, column j is set to 1, otherwise, it is set to 0. In this way, the Boolean matrix corresponding to the transaction database can be constructed.

2.2.3. Analysis of ADR reports

In addition, many researchers tried to solve the dilemma of rational drug use through the analysis of ADR reports. Ji\cite{16} has developed an association mining algorithm based on SQL, which can deal with multiple relational data tables directly. Although the algorithm proposed by the researchers seems to be efficient and scalable in the experiment, it is necessary to rely on parallel cloud computing technology. In the same year, Reps\cite{17} proposed another method of learning association rules between drugs and adverse reactions. In 2018, Wei Chen\cite{18} found association rules between adverse events and chemotherapy based on the proposed improved Apriori algorithm, which proved that it is an effective method to reveal the risk factors of adverse events during cancer treatment as shown in Tab.3.

3. Application of Apriori algorithm in rational drug use

3.1. Application in modern medicine at home and abroad

In recent years, many researchers are committed to reducing doctors' memory of complex drugs and improving the efficiency of doctors' prescribing. Chen\cite{19} studied the individualized treatment of AIDS based on patient similarity and frequent set analysis of the Apriori algorithm. Among them, cluster analysis, case-based reasoning, and association rule mining are discussed. The prediction consistency rates of the three results are 67.3%, 83.3%, and 71.4% respectively. Chen\cite{20} proposed a disease diagnosis and treatment recommendation system (DDTRS), and introduced the clustering algorithm based on DPCA (density peak clustering algorithm), and defined and analyzed the effective association rules of disease diagnosis and treatment by Apriori algorithm, so as to provide valuable diagnosis and treatment plan suggestions for doctors and patients.
3.2. Application in traditional Chinese medicine

3.2.1. Explore the law of Chinese Medicine

Zheng Kun [21] combined the Apriori algorithm and complex system entropy clustering method to study the medication law of distinguished TCM doctors. Wang [22] applied the Apriori algorithm to analyze headache of traditional Chinese medicine, which effectively helped traditional Chinese medicine decide to prescribe appropriate prescriptions in the face of various headache patients, and obtained the basic rules among common Chinese medicine, symptoms, and syndromes, which were similar to the principles of clinical practice. Ji [23] explored the characteristics and law of Chinese medicine in the treatment of diabetic thirst by analyzing the taste of Chinese medicine prescriptions and used Excel to establish a database for statistical analysis of the Apriori algorithm.

3.2.2. Study on the dosage of traditional Chinese Medicine

For the clinical prescription of traditional Chinese medicine, Yu [24] used fuzzy clustering and fuzzy association rules to cluster analysis. The algorithm mainly according to the characteristics of traditional Chinese medicine prescription dose data, automatic fuzzy grouping, mining results of high confidence. Chen Li [25] used the Apriori algorithm to analyze the related factors in adverse reaction reports, which provided data support for the occurrence mechanism of adverse reactions of traditional Chinese medicine injection.

3.2.3. Inheriting the experience of famous traditional Chinese Medicine

Xu [26] through the Apriori algorithm, he mined out the combination of famous and old Chinese medicine drugs and summarized the experience prescriptions of famous doctors. Hsieh [27] based on the association rule analysis of the Apriori algorithm to study the potential core combination of acupoints for the treatment of chronic obstructive pulmonary disease (COPD). Li [28] made statistics on the frequency of traditional Chinese medicine, nature, flavor, and meridian tropism, and conducted factor analysis. Zhu [29] used the Apriori algorithm to analyze the association rules between pulse condition and medication data table of pulse diagnosis medical records, which also has a high reference value for pulse diagnosis and medication.

3.3. In minority traditional medicine

3.3.1. Tibetan Medicine

Tibetan medicine has a unique theory on the source, nature, taste, efficacy, and medication principle of drugs. Wen [30] established the basic data framework of Tibetan medicine pharmacological mechanisms and applied it to the classic Tibetan medicine prescriptions. Wang [31] used the Apriori algorithm to find the related characteristics of disease symptoms and prescriptions, combined with the individual characteristics and disease characteristics of patients, realized the Tibetan medicine diagnosis and treatment prediction model of high-altitude stomach disease (atrophic gastritis), which can reach 80.1% accuracy. Based on the traditional Chinese medicine inheritance support system (TCMIS), Liu [32] analyzed the commonly used drugs, drug combination frequency, core drug combination, and new prescription drug combination used for the treatment of plateau disease prescriptions in relevant books.

3.3.2. Zhuang Medicine
Until now, Zhuang medicine is still one of the important health protections for Zhuang people. Wei [33] explored the compatibility rules of Zhuang medicine in the treatment of children's cough, collected 177 kinds of oral prescriptions for treating children's cough. Pang [34] and Jiang [35] used association rules to analyze the prescriptions and compatibility rules of Gudao disease in Zhuang medicine, so as to provide theoretical guidance for rational prescription and medication of Zhuang medicine.

3.3.3. Mongolian Medicine

Zhang [36] analyzed and compared the fuzzy c-means algorithm (FCM), hard c-means algorithm (HCM), and C4.5 decision tree algorithm, proposed Apriori algorithm based on the simplified binary matrix, which provided decision support and powerful protection tool for the research of Mongolian medicine treatment method.

4. Problems to be solved and development trend

Medical data has the characteristics of large amount, various categories, fast generation, and medical industry. What’s more, it has the characteristics of mass, complexity, accuracy, privacy, heterogeneity, and closeness [36]. Although the data mining technology is widely used and the algorithm performance is improving, there still exists a series of difficult problems and challenges. This section discusses the problems to be solved and the development trend in the future. At present, there are still some problems in the quality of medical data, which are embodied in: i) the authenticity of medical data. ii) the lack of unified standard specifications in the process of medical data entry. iii) the loss of medical data in the process of saving. iv) the problems of the hospital information system. The main research prospects of the Apriori algorithm in the field of rational drug use are as follows:

4.1. The combination of rational drug use and existing algorithms

Through a large number of literature, we found that researchers had made great contributions to the development of the Apriori algorithm. However, due to the particularity of the medical industry, how to better combine the improvement of the Apriori algorithm with the field of rational drug use is a problem that many researchers need to consider. Based on the current research, some researchers, especially medical practitioners, just use the Apriori algorithm embedded in the software. Therefore, we could focus on how to reasonably apply the development results of the Apriori algorithm theory to the field of rational drug use and improve the applicability of the algorithm that still need to be discussed and studied.

4.2. The combination of traditional Chinese medicine and algorithm

Through the above research, we found that subjectivity and experience lead to the development of traditional Chinese medicine, but lack objective diagnostic criteria. To improve the performance of traditional Chinese medicine, we could normalize the standards in diagnostic to give reasonable and objective explanations by using Apriori.

At present, the main defects of TCM diagnosis are: i) Lack of objective and scientific physiological standards. ii) Lack of unified information collection method and symptom information analysis method. iii) Lack of quantitative symptom standard. iv) Lack of unified inquiry method. v) limitations of the thinking mode of TCM inquiry. vi) Lack of understanding of the disease.

5. Conclusion
This paper explained the basic theory and the improved algorithm about Apriori. In the application of the data mining technology in modern medicine and traditional Chinese medicine, it mainly included data processing in both modern medicine and traditional Chinese medicine, theoretical support for disease diagnosis and medical orders giving, useful knowledge for contemporary traditional Chinese medicine from the experience of famous doctors. We discussed some problems in the development of rational drug use, mainly including, lack of applicable and standard data, the bad quality of existing medical data, few attention to the quality of medical data, the efficiency and accuracy of Apriori algorithm in rational drug use, and the combination of data mining algorithm and rational drug use. To solve those problems, we put forward some opinions on how to combine traditional Chinese medicine with data mining technology from different aspects.

Acknowledgments

This work was supported by the research projects: the Hunan Provincial Department of Education Innovation Platform Open Fund Project(15K082), the Hunan Provincial Department of Transportation Scientific and Technological Progress and Innovation Program Project(201927), the Central Military Commission Department of Equipment Development Pre-research Project(31511010105), the National Defense Science and Technology Bureau of National Defense Basic Research(WDZC20205500119), the Hunan Provincial Department of Education Project (JG2018A012, [2019]No.291, [2019]No.248, [2019]No.370; [2020]No.9, [2020]No.90, [2020]No.172), the Ministry of Education's Industry-University Cooperation and Collaborative Education Project (201901051021).

References

[1] Xiang F, Zhangwei Y and Shengxin C 2009 Application of data mining in the analysis of rational medication information J. Journal of Pharmaceutical Practice 27(06): 411-413+433

[2] Qing W, Fen L and Aili X 2003 Unreasonable medication problems and intervention research J. The Chinese Journal of Clinical Pharmacology (01): 75-78

[3] Ming F, Xiaofeng M and et al 2001Data Mining: Concepts and Techniques M. China Machine Press Beijing

[4] Rakesh A, Tomasz I, Arun S 1993 Mining association rules between sets of items in large databases SIGMOD Record: Special Interest Group on Management Data 22(2) 207-216

[5] Hua W and Xuegang H 2009 Data preprocess and algorithm apriori improvement of the medical data mining J. Computer Systems & Applications 18(09): 94-97

[6] Yuan C 2010 Application of association rule mining in disease data processing D. Chongqing: ChongQing Medical University Chongqing

[7] Yuhuan Z and Tian P 2014 An improvement of apriori algorithm in medical data mining J. Applied Mechanics and Materials 631-632 125-128

[8] Feng W, Yu W, Pengfei L and et al 2015 Construction of drug prompt system for acute upper respiratory tract infection based on association rule algorithm J. China Digital Medicine 10(10): 49-52

[9] Renli W, Yueming D and Liming D 2016 The application of Apriori-BSO algorithms in medical records data mining Proceedings of 2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference(ITNEC 2016) Chongqing China
827-832

[10] Noguchi Y, Ueno A, Otsubo, M and et al. A new search method using association rule mining for drug-drug interaction based on spontaneous report system *J. Frontiers in Pharmacology*, 2018, 9

[11] Rajeswari N, Nachammai S, Jenima P E, et al. 2019 Unexpected Health Issues Prediction In Medical Data Using Apriori Rare Based Outlier Detection Method 2019 *International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*

[12] Zhi L 2012 Research of association rule mining algorithms and their applications of coronary heart disease diagnosis and treatment with traditional chinese medicine *D. Dalian Maritime University*, Dalian

[13] Chong Z and Yunhua Z 2019 Diabetes prediagnosis system based on improved apriori algorithm *J. Computer Systems & Applications*, 28(02): 94-100

[14] Ya L, Lei Y, Li W and et al 2020 Application research of apriori algorithm based on matrix multiplication in children's drug interaction 2020 *12th International Conference on Measuring Technology and Mechatronics Automation ICMTMA*: 507-512

[15] Yadong Z, Kongfa H and Tao Y 2019 Mining effect of famous Chinese medicine doctors on Lung-cancer based on association rules 2019 *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* : 2036-2040

[16] Yangqing J, Fangyang S and John T 2014 A multi-relational association mining algorithm for screening suspected adverse drug reactions 2014 *11th International Conference on Information Technology: New Generations (ITNG)*: 407-412

[17] Jenna M. Reps, Uwe Aickelin, Jiangang M and et al 2014 Refining adverse drug reactions using association rule mining for electronic healthcare data *J. Computer Science*: 763-770

[18] Wei C, Jun Y, Huiling W and et al 2018 Discovering associations of adverse events with pharmacotherapy in patients with non-small cell lung cancer using modified apriori algorithm *J. BioMed Research International*: 1-10

[19] Jieqing C 2017 Study on data mining based personalized treatment strategy for AIDS patients *D. Capital Medical University*, Beijing

[20] Jianguo C, Kenli L, Huihui R and et al. 2018 A disease diagnosis and treatment recommendation system based on big data mining and cloud computing *J. Information Sciences* 124-149

[21] Kun Z, Jiarui W and Feifei Z 2017 Research on Chinese Medicine Treatment Rules of Coronary Heart Disease Based on Data Mining *J. China Medical Herald*, 14(36): 113-116

[22] Miao W, Lei Z, Zihao Z and et al. 2014 The application characteristics of traditional chinese medical science treatment on headache based on data-mining apriori algorithm 2014 *IEEE International Conference on Bioinformatics and Biomedicine(BIBM)* 153-157

[23] Tao J, Shulan S, Er'xin S and et al 2016 Determining the rules of traditional Chinese medicine on treatment of consumptive thirst based on association rules mining *J. China Journal of
Traditional Chinese Medicine and Pharmacy, 31(12): 4982-4986

[24] Xingwen Y, Lei Z, Yidi C and et al 2019 Study on the compatibility law of chinese medicine prescription dose based on fuzzy clustering and fuzzy association J. Chinese Archives of Traditional Chinese Medicine, 37(04): 818-822+1039

[25] Li C, Yuanda T, Yajing Z and et al 2019 Study on occurrence rules of adverse drug reactions of traditional Chinese medicine injections and their components based on association rules J. Chinese Journal of Rational Drug Use, 16(08): 1901-1907+1916

[26] Liang X 2016 Discussion on data preprocessing and data mining of medical records of distinguished veteran doctores of TCM D. Shandong University Of Traditional Chinese Medicine Jinan

[27] Hsieh P, heng ChuFang C, Chihwei W and et al 2020 Combination of Acupoints in Treating Patients with Chronic Obstructive Pulmonary Disease: An Apriori Algorithm-Based Association Rule Analysis J. Evidence-Based Complementary and Alternative Medicine, 1-7

[28] Juanjuan L, Fengyun W, lin L and et al 2019 Analysis on medication rules of national medical masters in treatment of stomachache J. Chinese Journal of Experimental Traditional Medical Formulae 25(21): 170-175

[29] Haoru Z, Jingjie H, Ziyao J and et al 2018 A study of association rules between pulse manifestation and medication in Linzheng Zhinan Yi'an based on data mining J. Guiding Journal of Traditional Chinese Medicine and Pharmacy, 24(01): 55-56+79

[30] Wen C g and Dang Z 2016 Study on data mining method of pharmacological mechanism of Tibetan Medicine D. Qinghai University, Xining

[31] Shiying W, Lei Z, Lu W and et al 2016 Research on syndrome classification prediction model of Tibetan medicine diagnosis and treatment based on data mining Signal-Image Technology & Internet-Based Systems (SITIS), 2016 12th International Conference on. Naples, Italy 497-502

[32] Huan L, Caiyun Z, Wen Z and et al 2018 Study on medication laws of Tibetan medicine in treatment of plateau disease based on data mining technology J. China Journal of Chinese Materia Medica, 43(08): 1726-1731

[33] Xing W, Min Z, Zhifei N and et al 2019 Drugs Laws and the Compatibility of Zhuang Medicine in the Treatment for Infantile Cough Based on Data mining J. Western Journal of Chinese Medicine, 32(06): 47-49

[34] Yuzhou P and Zuling J 2012 Application of association rules on study of prescription regularity in Gudao Disease of Zhuang medicine J. Chinese Journal of Experimental Traditional Medical Formulae 18(22): 1-3

[35] Zuling J, Yuzhou P and Jianying L 2015 Analysis of the prescription regulation on Shuidao disease of Zhuang minority medicine with association rule mining J. China Journal of Traditional Chinese Medicine and Pharmacy 30(01): 205-207

[36] Chunsheng Z and Ya T 2016 The design and realization of mongolian medicine prescription data mining system 2016 3rd International Conference on Information Science and Control
Engineering (ICISCE) 697-701