

## Retraction

# Retracted: Data Mining and Analysis of the Compatibility Law of Traditional Chinese Medicines Based on FP-Growth Algorithm

### Journal of Mathematics

Received 19 December 2023; Accepted 19 December 2023; Published 20 December 2023

Copyright © 2023 Journal of Mathematics. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### References

- [1] S. Zhou, "Data Mining and Analysis of the Compatibility Law of Traditional Chinese Medicines Based on FP-Growth Algorithm," *Journal of Mathematics*, vol. 2021, Article ID 1045152, 10 pages, 2021.

## Research Article

# Data Mining and Analysis of the Compatibility Law of Traditional Chinese Medicines Based on FP-Growth Algorithm

Shuchun Zhou 

Affiliated Nanhua Hospital, University of South China, Health School of Nuclear Industry, Hengyang 421002, Hunan Province, China

Correspondence should be addressed to Shuchun Zhou; 631505040131@mails.cqjtu.edu.cn

Received 26 October 2021; Accepted 23 November 2021; Published 17 December 2021

Academic Editor: Naeem Jan

Copyright © 2021 Shuchun Zhou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The compatibility law of prescriptions is the core link of TCM theory of “theory, method, prescription and medicine,” which is of great significance for guiding clinical practice, new drug development and revealing the scientific connotation of TCM theory, and is also one of the hot spots and difficulties of TCM modernization research. How to efficiently analyze the frequency of drug use, core combination, and association rules between drugs in prescription is a basic core problem in the study of prescription compatibility law. In this paper, a systematic study was made on the compatibility rules of traditional Chinese antiviral classical prescriptions and the mechanism of traditional Chinese medicine molecules. FP-growth algorithm was used to analyze association rules of 961 classical prescriptions collected and to explore the compatibility rules of traditional Chinese antiviral classical prescriptions. In terms of compatibility law of traditional Chinese antiviral prescriptions, this paper studied the compatibility law of traditional Chinese antiviral prescriptions based on the FP-growth algorithm and made exploratory research on the compatibility law information of 961 traditional classical antiviral prescriptions. Firstly, FP tree was constructed based on the classic recipe data set. Then, frequent item set rules were established, and association rules contained in FP tree were extracted. Finally, the frequency and association rules of antiviral TCM prescriptions were analyzed according to dosage forms (decoction, pill, paste, and ingot). The results show that the FP-growth algorithm adopted in this paper has excellent algorithm performance and strong generalization and robustness in the screening and mining of large-scale prescription data sets, which can provide important processing tools and technical methods for the study of the compatibility rule of traditional Chinese medicine prescriptions.

## 1. Introduction

In recent years, infectious diseases present a high incidence trend, among which more than 75% are caused by viruses. With the continuous variation of viruses and the enhancement of drug resistance, the treatment of viral diseases has become one of the world’s difficult problems. In the aspect of prevention, control, and treatment of virus infectious diseases, traditional classic antiviral prescription, such as MaXingShi sweet soup, hot disease, Fructus Forsythiae and scattered, small Bupleurum decoction, and Sang Ju Yin, because of its adjustable immunity, can interfere with viral DNA or RNA replication, suppress the virus proliferation, and have the effect that protect cells against the virus damage [1]. It has played a

pivotal role in the treatment of epidemic viral diseases such as SARS; AIDS; hand, foot, and mouth disease; and H7N9. The traditional Chinese medicine antiviral prescription plays an important role, promotes new prescription antiviral herbs and new drug research and development to dig deeper into the traditional classic antiviral prescription information, has become a new drug development of traditional Chinese medicine study the compatibility of the law, an important subject in the field of it to study the internal relations and characteristics of the prescription system, and also has the very vital significance [2, 3].

Data mining is a processing technology to search for information with special relevance hidden in a large amount of data, which has played an important role in the basic

theory of traditional Chinese medicine, traditional Chinese medicine prescription, traditional Chinese medicine philology, and clinical research of traditional Chinese medicine. As an important branch of data mining, the association rule can describe the potential relationship between data items in the database. Moreover, the said method discovers the relationship between variables of interest in the database [1]. Some important results have been obtained in the dose-effect study of traditional Chinese medicine based on association rule method, which has played an important role in promoting the study of traditional Chinese medicine formulations. For example, in terms of clustering analysis [4], such as using the *K*-means clustering analysis method for treating diabetes, analysis of prescription drugs, and learned of diabetes prescription drug law and basic medicine, Radix Rehmanniae, prepared Rhizome of Rehmannia, Radix Trichosanthis, Rhizoma Anemarrhenae, Rhizoma Alismatis, and Radix Ophiopogonis, for prescriptions of traditional Chinese medicine theory research and new drug development to provide the reference information. Reference [5] used the clustering method to automatically divide the fuzzy interval of drug dose in dot-effect analysis of drug pairs and then analyzed the association rules of drug pairs by combining the fuzzy association rules. The mined knowledge had a high accuracy. Using frequency and frequent term set method, reference [6] used frequency statistical analysis to explore the compatibility methods of reducing toxicity and increasing efficiency of toxic Chinese traditional medicine *Pinellia ternata*. It was concluded that the compatibility of soothing poison, dampening to prevent dryness, cold to make heat, and phase killing to make poison could reduce the toxicity of *Pinellia ternata*. The authors in [3] used frequency analysis and association rule method to analyze and study the compatibility rules of TCM prescriptions for the clinical treatment of senile dementia by physicians of all dynasties and obtained the compatibility rules of common prescriptions for the clinical treatment of senile dementia. Reference [7] analyzed the prescription law of Chinese herbal compound oral treatment of ulcerative colitis and excavated 60 core combination prescriptions and 23 new prescriptions. Using association rule technology, [8] explored the compatibility rule and core drug use of Chinese medicine prescription of Chai Songyan in the treatment of premature ovarian failure based on syndrome differentiation and found 45 pairs of commonly used 2 drug combinations. Data mining was conducted [9] on the compatibility rule of the drugs with the strongest correlation of TCM antiemea prescription and obtained that ginger, *Pinellia ternata* and *Poria cocos* were the most commonly used drug combinations in TCM antiemea prescription, and confirmed that *Pinellia ternata* plus *Poria cocos* decoction created by Zhang Zhongjing was the core drug group of TCM antiemea prescription. Based on the apriori association rule algorithm, [8] analyzed the real-world clinical rules of the combined application of compound *Sophora flavescens* injection and Chinese and Western drugs in the treatment of malignant tumors, providing useful reference for clinical treatment ideas and reasonable reference for the clinical application of

compound *Sophora flavescens* injection. The medicinal properties were processed [10] and efficacy of 365 flavors in Shennong Materia Medica Classic to find frequent patterns and strong association rules between qi, flavor, and efficacy, providing new methods and ideas for theoretical research on the medicinal properties of the four qi and five flavors of Traditional Chinese medicine. Liver fibrosis antiviral drugs, such as Liuwei Wupian combined with Ganoderma lucidum, have certain advantages in the antiviral treatment of chronic hepatitis B liver fibrosis virus-like meta-analysis. A systematical evaluation of the effectiveness and safety of *Astragalus* was carried out based on the TCM compound in the treatment of diabetic nephropathy in 488 patients and conducted a meta-analysis, concluding that *Astragalus*-based TCM compound may be a relatively safe and effective drug in the treatment of diabetic nephropathy.

Data mining is the extraction or “mining” of knowledge from large amounts of data. Through data mining, valuable knowledge, rules or high-level information can be extracted from the relevant data set of the database. And, display from different angles, so that the large database or data warehouse as a rich and reliable resource for decision-making services. In data mining, the discovery of rules is based on the statistical rules of large samples. When the confidence reaches a certain threshold, rules can be considered to be established. The core methods of data mining are association rules and sequential pattern mining, classification, and clustering. Association rule analysis is a very important research topic in the field of data mining, and it is also one of the most mature research methods. The purpose is to mine the association rules that meet the minimum support and minimum confidence between transaction features from the given data. Minimum support and minimum confidence are two measures reflecting the value of association rules, representing the usefulness and reliability of rules, respectively. Rules are considered meaningful only if they satisfy both minimum support and minimum confidence. We believe that there is some form of association in the compatibility of Chinese medicines. According to the theory of traditional Chinese medicine, there are the following five relationships between Chinese medicines, that is, the seven must, cause, fear, kill, and have nothing to do with each other. For example, in Buzhong and Yiqi decoction, the combination of Bupleurum and Hoshio can draw seven liters of qi tonic from ginseng, Qi qi, shu, and grass. Together to achieve the effect of beneficial qi rising trap, this combination is the role of phase. We can find meaningful combination patterns of traditional Chinese medicine from common prescriptions. The tool used in this study is the algorithm of extracting association rules from data mining—FP-growth algorithm.

Demand for new antiviral medicine of Chinese medicine research and development, using the experience of the study, in order to improve the effectiveness and accuracy of association rules analysis algorithm, this study intends to use the FP-growth algorithm to the traditional Chinese medicine classical prescription data screening for large-scale data set mining, aim to do exploratory research antiviral herbs prescription compatibility of the law, and verify the

effectiveness of the algorithm and explore the law of traditional classical antiviral prescription and potential useful information.

## 2. Association Rule Data Mining

*2.1. Mining Association Rules.* Data mining is a new research field gradually developed in recent 30 years. It is the product of the combination of multidisciplinary and technology, which is widely used in various fields such as government decision-making, enterprise management, scientific discussion, and medical research, and plays an important role in promoting the development of all aspects of society. Association rule mining is one of the most typical knowledge types in data mining. In the medical field, it has a wide range of applications.

Association rules are used to represent the association degree of many attributes (item sets) in OLTP database. They are used to find the correlation of attributes by the association algorithm using a large amount of data in the database. The problems of association rule mining are described as follows.

Let  $I = \{I_1, I_2, \dots, I_m\}$  is the set of data items,  $D = \{T_1, T_2, \dots, T_n\}$  is a transaction database, where each transaction  $T$  is a subset of data item set  $I$ , namely,  $T \subseteq I$ , and each transaction  $T$  has an identifier TID associated with it. Transaction  $T$  is said to contain item set  $X$  if a subset  $X$  of  $I$  satisfies  $X \subseteq T$ . An association rule is something like " $X \Rightarrow Y$ ." The meaning is that the occurrence of some items in a transaction leads to the occurrence of other items in the same transaction, where " $\Rightarrow$ "; called the "association" operation,  $X$  is the prerequisite for the association rule, and  $Y$  is the result of the association rule. For example, in the compatibility of Chinese medicine prescriptions, more than 90% of prescriptions using Chinese medicine A must have to use the Chinese medicine B at the same time. So, the association rule  $R$  can be expressed as  $R: A \Rightarrow B$ . Support and confidence are important concepts in association rules.

Support is similar to the percentage of prescriptions using both A and B in the total prescriptions. Confidence is the percentage of the prescriptions of all Chinese medicine A and B to the prescriptions of Chinese medicine A, which is called regular confidence. The former is used to measure the statistical importance of association rules in the whole data set, while the latter is used to measure the credibility of association rules. Their formulas are formulas (1) and (2), respectively:

$$\text{Support}(A \Rightarrow B) = P(A \cup B) = \frac{B}{A}, \quad (1)$$

$$\begin{aligned} \text{Confidence}(X = Y) &= P(Y|X) \\ &= \frac{P(X \cup Y)}{P(X)} = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)}. \end{aligned} \quad (2)$$

In practical applications, associations with high support and confidence can be used as useful association rules, which are called minimum support threshold (min\_ sup)

and minimum confidence threshold (min-conf). Min-sup indicates the lowest statistical importance of data items. Only data item sets that meet min-sup appear in association rules, which are called frequent item sets. The minimum confidence is the lowest reliability of the association rule. Rules that meet the requirements greater than min-sup and min-conf are called strong rules. The task of association rule mining is to discover all frequent item sets and dig out all strong rules in transaction database  $D$ .

Association rule mining is actually frequent pattern mining. According to the following criteria, frequent pattern mining has multiple classification methods:

*2.1.1. Classification according to the Completeness of Mined Patterns.* Given the minimum support threshold, the complete, closed, and extremely frequent item sets of frequent item sets can be mined. It is also possible to mine constrained frequent itemsets (that is, frequent itemsets that satisfy a set of constraints specified by the user), approximate frequent itemsets (that is, only the approximate support count of the mined frequent itemsets is derived), near-matched frequent itemsets (that is, itemsets that conform to the support count of close or nearly matched itemsets), and mostK frequent itemsets (that is,  $k$  most frequent itemsets for user-specified  $K$ ), and so on [12].

*2.1.2. Classify according to the Abstraction Layer Involved in the Rule Set.* Some methods of mining association rules can discover different abstraction layer rules. For example, suppose the mined association rule set contains the following rules:

$$\begin{aligned} &\text{buys}(X, \text{"computer"}) \Rightarrow \text{buys}(X, \text{"HP\_printer"}), \\ &\text{buys}(X, \text{"laptop\_computer"}) \Rightarrow \text{buys}(X, \text{"HP\_printer"}). \end{aligned} \quad (3)$$

*2.1.3. If the Item or Attribute in an Association Rule Involves Only One Dimension, It Is a Single-Dimensional Association Rule.*

$$\begin{aligned} &\text{buys}(X, \text{"computer"}) \Rightarrow \text{buys}(X, \text{"antivirus\_software"}), \\ &\text{age}(X, \text{"30...39"}) \Rightarrow \text{buys}(X, \text{"high\_resolutionTV"}). \end{aligned} \quad (4)$$

*2.2. Improving FP-Growth Algorithm.* FP-growth algorithm is a famous algorithm based on FP-growth tree proposed by Han Jiawei et al. This algorithm provides a good frequent pattern mining process without generating candidate set, and its performance is improved compared with apriori algorithm. However, FP-growth algorithm generates more and more conditional FP-trees with the deepening of recursive calls. Especially in the case of shared prefixes, FP-growth algorithm is very time-consuming. In order to solve this



problem, this paper proposes the improvement of FP-growth algorithm, FP-growth\* algorithm.

The idea of FP-growth algorithm is to reduce the time of searching shared prefixes to reduce the time of generating FP tree to improve mining efficiency. That is, if there is a shared prefix, the shared prefix is found by traversing the first child node of the node. Its mining steps are as follows.

**2.2.1. Ranking of Frequent  $L$  Item Set.** Describe the transaction database  $D$  once, generate frequent  $L$  item set and the support degree of each frequent item set, sort by descending support degree, and the result is  $L$ .

**2.2.2. Transaction Item Reordering.** The transaction database items are sorted according to the order of frequent item table  $L$  to generate transaction database  $D$ .

**2.2.3. Transaction Set Reordering.** The whole data set of  $D$  is reordered according to the order of  $L$ , that is, the first column of the transaction set is sorted according to the order of  $L$ . Then, the second columns of the transaction set are sorted in the order of  $L$ , and the final columns of the data set are analogized to get sorted data set  $D$ .

**2.2.4. Construct FP-Tree Condition.** Create root node marked with "NULL," scan  $D$ , call insert—Tree (P, T1) procedure for each transaction in it. Generate the FP tree.

**2.2.5. Mining FP Tree.** Recursively call FP-growth algorithm to mine FP tree and obtain frequent item sets.

**2.3. Research on FP-Growth Algorithm in Mining Compatibility Rules of Traditional Chinese Medicine Prescriptions.** In the fact of more than 100,000 TCM prescriptions, spleen and stomach prescriptions were selected as the data source of association mining in this paper. All prescriptions were from the clinical prescriptions of Hua Tuo Hospital of Traditional Chinese Medicine and the Database of TCM Prescriptions of Shanghai TCM Data Center. As the hometown of the magical doctor Hua Tuo, Bozhou has long been known as the "peony flowers outside the city of Xiaohuang, producing the morning clouds for miles and five miles." It is a well-known center for the planting and processing of Chinese medicinal materials in the world. There are hao peony, hao chrysanthemum, hao mulberry bark, and hao pollen in the real estate medicinal materials included in the Pharmacopoeia. With a planting area of 1 million mu, it is known as the "Chinese medicine Capital." A large number of traditional Chinese medicine resources provide natural conditions for the development of traditional Chinese medicine prescriptions. Huatuo traditional Chinese medicine hospital has a large number of clinical prescriptions: "TCM Prescription database" of Shanghai TCM data center contains 190,000 TCM prescriptions and extracts the prescriptions contained in the literature. The data items include

the name, composition, dosage, indications, and other information of prescriptions.

**2.3.1. Data Processing of TCM Prescriptions.** The original data expression of the existing prescription database is not standardized, so it is necessary to transform the descriptive language of the prescription into the data information that can be processed by the computer, so as to make it standard and standard, so as to realize the correct expression and reasonable organization of prescription data in the computer system. Using computer data to express is not only helpful for in-depth analysis and operation of data. It is also an important way to realize data normalization and standardization. The data preprocessing method in this paper is as follows:

(1) *Standardized Data.* The purpose is to standardize the semantic ambiguity and expression of the concept words, polysemy monosyllabic word, multiword monosyllabic word lexicalization. The split expression of multiconcept combination words such as dizziness refers to symptoms such as dizziness, which are different from simple dizziness, blood dizziness, motion sickness, etc., such as fever, severe fever, and night fever which are treated as a single concept of fever.

(2) *Structured Data.* The purpose is to refine and organize the original data of prescription reasonably, so as to meet the requirements of data mining and to realize the orderly arrangement of key concepts and the formation of the associated structure between data.

Prescription data have multiple associations, such as between drugs, between drugs and symptoms, and between efficacy and indications. "Syndrome, medicine and prescription" is the core, and "medicine" is the key element in the core. Their relationship is as follows: select "medicine" and "prescription" for "syndrome." "Syndrome" is composed of syndrome sets, "medicine" contains different taste and quantity, etc., and "prescription" has complex matching relations and the problem of adding or subtracting prescription.

(3) *Digitize Data.* Numbers are easy to represent the structure and mutual relationship between data, while data described by other characters or symbols is not easy to do, so as far as possible to use numbers to replace the characters or symbols containing some knowledge. If the dose is described in grams, the drug taste and virulence are also represented by numbers. If flatness is set to 0, the corresponding value of skewness is shown in Table 1.

**2.3.2. Mining Compatibility Rules of TCM Prescriptions Based on FP-Growth\* Algorithm.** A total of 106 spleen-stomach prescriptions with symptom frequency greater than 25 were screened out of 338 prescriptions collected, and each prescription was considered as a transaction with the marker code TID: T001, T002...T106, the code of each Chinese medicine in the formula is  $I_i(I=1, 2, 3, \dots)$

The collected spleen and stomach prescriptions and their components are as follows:

TABLE 1: Digital representation of data.

Very hot	Hot	Wen	Lukewarm	Flat	Cool	Slight cold	Cold	Severe cold
1.2	0.9	0.6	0.3	0	0.3	0.6	0.9	1.2

Two Chen soup flavor (F001): atractylodes, *Glycyrrhiza glycyrrhiza*, tangerine peel, *Magnolia officinalis*, *Poria cocos*, *Pinellia ternata*, rhizoma SPP

Sijunzi decoction (T002): atractylodes, glycyrrhiza, ginseng, *Poria cocos*, sweet sand

Liujunzi decoction (TO03): atractylodes, *Glycyrrhiza glycyrrhiza*, ginseng, tangerine peel, *Poria cocos*, xylobacter, *Pinellia ternata*, *Amomum amomum*, ginger

Qinlianping, Weitang (T004): liquorice, tangerine peel, magnolia bark, *Scutellaria baicalensis*, *Coptis chinensis*, *Atractylodes atractylodes*

Huoxiang Fupi drink (T005): liquorice, tangerine peel, wood incense, *Magnolia officinalis*, *Pinellia ternata*, Huoxiang, Malt

Shipi drink (T006): atractylodes, *Glycyrrhiza*, *Magnolia officinalis*, *Poria cocos*, xylobacter, papaya, grass fruit, areca nut, ginger, monkshood fruit, jujube

Guipi decoction (T007): atractylodes, licorice, ginseng, xylobacter, *Poria cocos*, angelica, *Astragalus membranaceus*, *Polygonum longan*, jujube seed

Yigong powder (T008): atractylodes, *Glycyrrhiza*, tangerine peel, ginseng, *Poria cocos*

Lizhong pill (T009): atractylodes, licorice, ginseng, ginger

Baoyuan soup (TO10): atractylodes, ginseng, angelica, cassia twig, *Astragalus membranaceus*, raw aconite

The main codes of prescriptions are as follows: atractylodes I1, *Glycyrrhiza* I2, tangerine peel I3, ginseng I4, *Magnolia officinalis* I5, aucklandoides I6, *Poria cocos* I7, codonopsis I8, *Pinellia ternata* I9, angelica I10, Quanyi peony I12, ginger I selected ginseng I14, Su stem I15 amomum I16, almond I17, *Coptis coptidis* I18, *Astragalus membranaceus* I19, cinnamon I20, jujube I21, polygonum I22, yam I23, lentil I24, *Solanum solanum* I25, *Lobelia* I26, cardamom I27, *Corni ruyu* I28, black aconite I29, *Atractylodes atractylodes* I30, nutmeg I31, papaya I32, cardamom I33, preparation of aconite I34, longan meat I34, weak, sour jujube seed I36, patchouli I37, selected ginseng I38, *Scutellaria baicalensis* I39.

The establishment of spleen and stomach agent transaction database (part) D is shown in Table 2:

FP-tree tree was constructed according to transaction database D (FP tree was omitted due to limited space), and the corresponding support degree of the frequency of the occurrence of traditional Chinese medicine in the prescription was set at the minimum of 30. FP-growth improved algorithm was used to obtain frequent sets by establishing conditional pattern library, mining all frequent item sets, and the compatibility rules of spleen and stomach prescription were found as follows.

(1) *Prescription Core Drugs*. Liquorice (97), dried tangerine or orange peel, atractylodes (93) (92), ginseng (78), thick ~ b (56), combination (48), angelica (36), the 7 of TCM to occur more often than other drugs in the prescription but also can get the ingredients are: sijunzi decoction, different work loose and sweet sand six main medicine soup, namely, is Lord of the spleen and prescription drugs.

(2) *Prescription Structure*. After the above analysis, results prove that the spleen and stomach fangfang looks complex. There is a basic structure.

The decoction of invigorating qi and invigorating spleen represented by Sijunzi Tang is the most basic prescription. The second is the combination of qi medicine + qi medicine prescription, such as Xiangsha Liujunzi Decoction, Yigong powder prescription. Replenishing qi medicine + regulating qi medicine + disease medicine (or humidification medicine) compatible prescription, such as Shenling Baizhu Powder, six gentleman decoction and other prescriptions. Supplementing qi medicine + warm medicine compatibility of prescriptions, such as Bao Yuan soup, li Zhong pills and other prescriptions.

In order to improve the efficiency of apriori algorithm, Han Jiawei et al. proposed a FP-growth algorithm based on growth tree structure to generate frequent item sets [27]. The basic idea of the algorithm is to scan the database only twice: the first time scans the number of the occurrence of a single item in the data set and filters out the items that do not meet the minimum support.

In the second scan, the frequent pattern tree (FP-tree) structure is established, and then the FP tree is recursively grown into a large item set, and then the test is carried out on the whole data set. This algorithm does not generate candidate item sets, avoids multiple scanning of the original database, and can directly compress the database to generate a FP tree, and finally form association rules. Studies have shown that FP-growth algorithm is one order of magnitude faster than apriori algorithm in finding large item sets.

### 3. Study on Compatibility Rule of Chinese Medicine Antiviral Prescription Based on FP-Growth Algorithm

#### 3.1. Data Source of Chinese Medicine Antiviral Prescription.

In order to study the compatibility rules of traditional and classical antiviral prescriptions, the research group designed and developed the TCM prescription management system in advance. The system is based on web B/S architecture mode, using Java development language and access database management software, and can run in Windows/Linux system. It adopts top-down overall planning, top-down application development strategy, standardized framework structure, and easy to operate import mode. TCM

TABLE 2: Transaction database  $D$  (Part).

TID	Items
T001	I1, I2, I3, I5, I7, I9
T002	I1, I2, I3, I7
T003	I1, I2, I3, I4
T004	I2, I3, I5, I18, I30, I39
T005	I1, I2, I3, I5, I6, I9, I37, I40
T006	I2, I3, I5, I7, I6, I15, I21, I32
T007	I1, I2, I4, I6, I7, I10, I19, I22, I35, I36
T008	I1, I2, I3, I4, I7
T009	I1, I2, I4, I15
T0010	I1, I4, I10, I19, I41, I4
.....	.....

prescription management system can meet the basic import, export, retrieval, and other operations and simple statistical functions. Through the TCM prescription management system, all the books on epidemic diseases collected in the First Part of Wen's Disease Dacheng (2007, Fujian Science and Technology Publishing House) are selected.

**3.2. Data Preprocessing.** Literature data sources of classical antiviral prescriptions are diverse, and drug names are not standardized. Therefore, according to the traditional Chinese medicine name standard in the Dictionary of Traditional Chinese Medicine, the collected prescriptions were cleaned and the names of medicines were standardized. In the process of this study, examples of traditional Chinese medicine name standardization are shown in Table 3.

**3.3. Application Process of FP-Growth Algorithm.** The following uses a specific example to illustrate the implementation process and characteristics of the FP-growth algorithm.

*Step 1.* According to the FP-growth algorithm, the sample data set was scanned first, and the traditional Chinese medicines meeting the minimum support threshold were arranged in the descending order according to the frequency of occurrence in the data set.

*Step 2.* Arrange formula data in the example in the descending order of frequency and select TCM with frequency greater than 3. According to the result of reordering, FP tree is established.

In Figure 1, root is the empty set used to build the subsequent FP tree. The structure of FP tree itself is represented by solid arrow head, and the count at the node represents the frequency of occurrence of this item in the data set. For example, Gardenia and Scutellaria in the first branch on the right of the tree correspond to the ninth prescription, while Gardenia and Scutellaria in the second branch on the right correspond to the seventh and eighth prescriptions, so the count at the node is 2. The whole FP tree can be obtained by analogy. The title table on the left of the figure shows the frequency of TCM meeting the minimum support in the data set, in the descending order. Dotted arrows connect the title table to the tree structure and join

items with the same name together for easy traversal of the tree structure. The sum of counts of items with the same name in the figure corresponds to the item support in the title table. After the FP tree is obtained, the reverse recursive processing tree can get the gradually increasing item set, and the association rules can be further calculated. It is worth noting that, in the process of establishing FP tree, the traditional Chinese medicine that does not meet the minimum support in the example will not be inserted into the FP tree. Therefore, the FP-growth algorithm can effectively remove the terms less than the support and enable multiple prescriptions to share the most frequent traditional Chinese medicine and finally achieve a high compression effect in the root of the tree. The designed experimental algorithm flow is shown in Figure 2, Algorithm flow of FP-growth is shown in Figure 3.

#### 4. Experimental Results and Analysis

Chinese medicine prescription is not a random combination of drugs but has potential compatibility law and processing technology. According to the characteristics of drugs and the needs of clinical syndrome treatment, in order to give full play to the effect of drug therapy, TCM prescriptions are often made into various dosage forms such as decoction, wine, tea, dew, pill, powder, paste, dan, tablet, ingot, glue, striping agent, and line agent for internal and external use. Due to the low quantity of some dosage forms in the research data, this study mainly analyzed the four dosage forms of decoction, pill, ointment, and spindle and obtained the core drug use and corresponding association rules of the corresponding dosage forms of antiviral prescriptions. Among them, the occurrence frequency of glycyrrhiza uralensis "with all kinds of drugs, cure all kinds of poison" was up to 480 in 961 antiviral prescriptions, and the occurrence frequency of glycyrrhiza uralensis was too high with other drugs, so that part of the analysis results were not valuable. In order to make the association rules mined more meaningful. In the experimental process, except for the ointment (only 15 pieces), the item sets with drug combinations greater than (including) 3 traditional Chinese medicines were selected for study and analysis.

**4.1. Medicinal Broth.** The top ten frequently used drugs in decoction antiviral prescription were glycyrrhiza, *Scutellaria baicalensis*, rhubarb, angelica, orange stem, ginger, shengdi, mint, Rhizoma chinensis, and *Poria cocos*, which were the main drugs in decoction antiviral prescription. Drug combination is greater than 3 taste traditional Chinese medicine (TCM), the frequency is more than 10, and confidence is more than 90% of a total of 32 groups, the association rules in the combination of the frequency highest with liquorice—Gardenia—Radix scutellariae and mint—*Platycodon grandiflorum*—licorice root, rhubarb—mint—Radix scutellariae and mint—even become warped, licorice, mint—St. John's wort, Radix scutellariae, all is one common combination of antiviral prescription. Some of them had strong association rules. For example,

TABLE 3: Sample table of drug name normalization.

The specification before	After the specification
Cicada, cicada skin, cicada slough, cicada shell	Cicada
Dried reed root, dried reed root, reed root	Reed root
Tender osmanthus, sichuan osmanthus, liugui, osmanthus tip, osmanthus tip, osmanthus section	Cassia twig
Wolfberry, sweet wolfberry, red wolfberry, western	Medlar
Thousand lily, southern lily, large lily, xuan lily, new lily, du lily, sweet lily, white lily, white flower lily, lily slice	Lily
Bright sulfur, yellow fluid, stone sulfur, ship sulfur	Sulfur
.....	.....

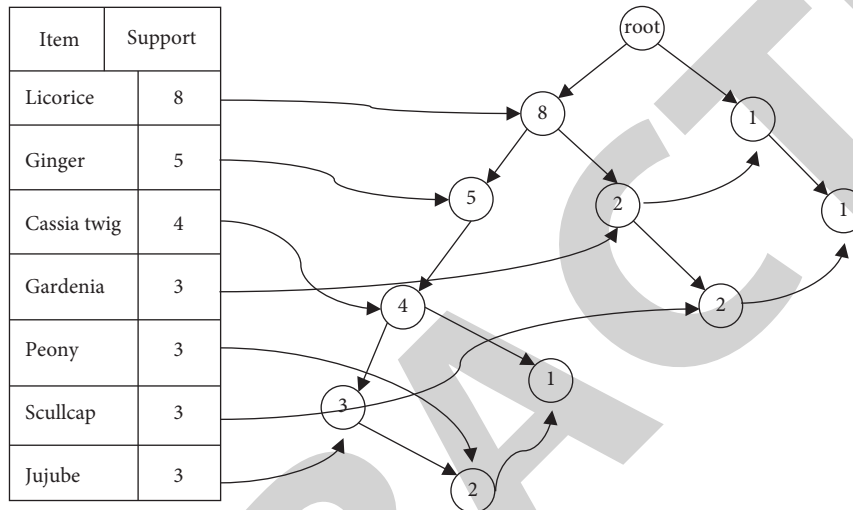


FIGURE 1: Diagram of FP tree of 9 prescriptions.

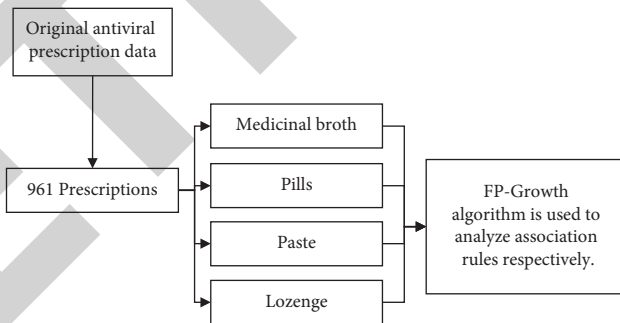


FIGURE 2: Algorithm flow of compatibility rule of TCM antiviral prescription based on the FP-growth algorithm.

when *Scutellaria baicalensis* and cicada slua appeared simultaneously, the occurrence probability of coptis was 100%. When *Scutellaria baicalensis*—silkworm—*Rhizoma coptidis* appears simultaneously in one prescription, cicada exuvium will inevitably appear in the prescription, which vividly excavate the internal relationship between the drugs in the prescription and provide the basis for clinical doctors to use medicine. Frequency and probability of the top ten drugs used in decoction is shown in Table 4.

4.2. Pills. The frequency and probability of the top ten Chinese medicines in pill antiviral prescriptions are shown in Table 5. There were 30 association rules with drug

combinations greater than 3 traditional Chinese medicines, frequency higher than 25, and confidence greater than 80%. The combinations with the highest frequency were ginger—jujube—glycyrrhiza, glycyrrhiza—*Rhizoma coptidis*—*Scutellariae*, glycyrrhiza—*forsythia*—*Scutellariae*, which were the core combinations of pill antiviral TCM prescriptions. There are strong association rules among some Chinese medicines, which can provide theoretical support for new drugs. For example, in a formula, when jujube—ginseng appeared at the same time, ginger appeared in the probability of 97.06%. When *Scutellariae* and cicada slits appeared at the same time, the probability of silkworm emergence was 97.06%.



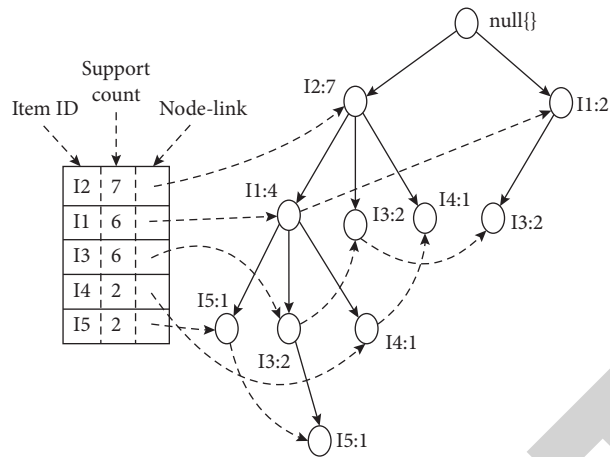


FIGURE 3: Algorithm flow of FP-growth.

TABLE 4: Frequency and probability of the top ten drugs used in decoction.

Medicines	Frequency	Probability
Licorice	81	46.55
Scullcap	49	28.16
Rhubarb	34	19.54
Angelica	27	15.52
<i>Platycodon grandiflorum</i>	27	15.52
Ginger	26	14.94
Radix rehmanniae	26	14.94
Mint	23	13.22
Rhizoma coptidis	23	13.22
<i>Poria cocos</i>	23	13.22

TABLE 5: Frequency and probability of the top ten drugs used in pills.

Medicines	Frequency	Probability
Licorice	350	52.87
Scullcap	201	30.36
Ginger	143	21.60
Radix rehmanniae	105	15.86
Rhubarb	103	15.56
Jujube	96	14.50
Ginseng	92	13.90
Peony	91	13.75
Gypsum	91	13.75
Rhizoma coptidis	83	12.54

**4.3. Medicinal Extract.** The collection of cream antiviral prescriptions is relatively small, only 15. Chinese medicines with frequencies greater than 3 were scutellaria baicalensis, licorice, mint, Sichuan rhizome, rhubarb, shengdi, and rhino horn, which were the main drugs used in ointment antiviral prescriptions. The specific frequency and probability of occurrence are shown in Table 6. Table 7 shows the association rules of TCM frequency greater than 3. It is easy to know that there are strong association rules between the drugs in the ointment, such as the prescription of scutellaria, raw ground, and rhinoceros horn at the same time any two drugs, another medicine will also appear.

**4.4. Experimental Analysis of FP-Growth\* Algorithm.** In the same computer software and hardware system, with the increase of the number of data sets, the time of FP-tree generation by the improved algorithm decreases obviously. According to the experimental analysis, when the number of data sets is large, the mining efficiency of FP-growth\* algorithm is increased by about 20%, as shown in Figure 4:

**4.5. Algorithm Analysis and Comparison.** FP-growth\* algorithm is improved on the basis of FP-growth algorithm, and it retains the efficient characteristics of FP-growth algorithm and adds support for the mining of numerical data, support for interdimensional association mining, mining maximum frequent itemsets instead of mining all frequent patterns. This method can greatly save the space and time cost of producing all frequent patterns and also meet the needs of traditional Chinese medicine mining. From the perspective of time complexity, the FP-growth\* algorithm is better than FP-growth algorithm.

- (i) FP-growth\* algorithm finally mined the maximum frequent item set, which is more than one order of magnitude different from all frequent item sets. Therefore, when FP-growth\* algorithm generates conditional pattern tree and maximum frequent item set, it takes much less time than FP-growth algorithm.
- (ii) FP-growth\* algorithm adopts an optimized search strategy, omits a certain number of item searches, and does not need to generate conditional modular basis, conditional pattern tree, and longest frequent item set for these items, saving considerable time. Performance comparison in time of  $c$  and FP-growth is shown in Figure 5.

Therefore, the FP-growth\* algorithm proposed by the author can not only deal with numerical interdimensional rules in mining function but also outperforms FP-growth algorithm in running time efficiency. Through the analysis of the mining results of this algorithm, it is obvious that the interdimensional maximum frequent item set is really meaningful for the FP-

TABLE 6: Frequency and probability of drugs with frequency greater than 3 in the ointment.

Medicines	Frequency	Probability
Scullcap	5	33.33
Licorice	4	26.67
Mint	3	20.00
Chuanlian	3	20.00
Rhubarb	3	20.00
Radix rehmanniae	3	20.00
Rhinoceros horn	3	20.00

TABLE 7: List of ointment association rules.

Pharmaceutical compositions	Frequency	Support	Confidence (%)
Chuan lian => Scutellaria	3	0.20	100.00
Rhizome of rehmannia => Scutellaria	3	0.20	100.00
Rhinoceros horn => Scutellaria	3	0.20	100.00
Rhinoceros horn => rhizome of rehmannia	3	0.20	100.00
Rhizome of rehmannia => rhinoceros horn	3	0.20	100.00
Scutellaria, rhizome of rehmannia => rhinoceros horn	3	0.20	100.00
Rhinoceros horn, rhizome of rehmannia => Scutellaria	3	0.20	100.00

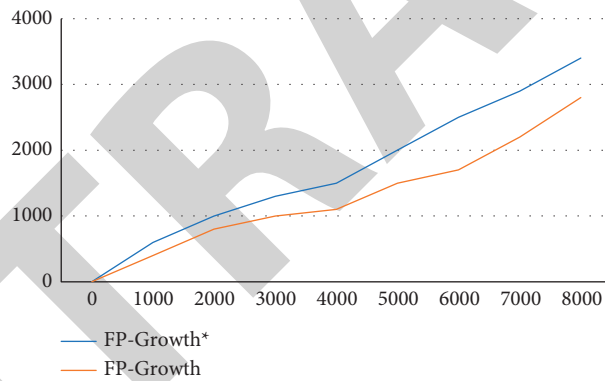


FIGURE 4: Comparison between FP-growth\* and FP-growth algorithm to generate FP tree.

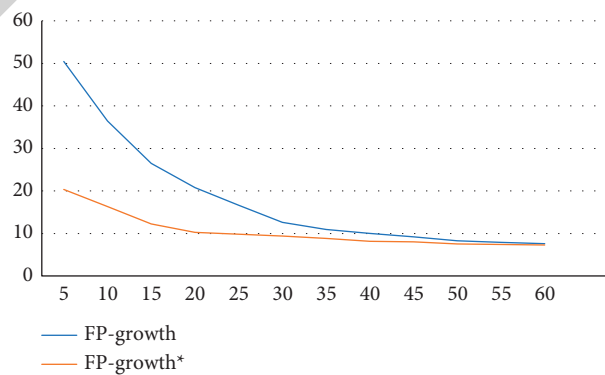


FIGURE 5: Performance comparison in time of c and FP-growth.

growth\* rule of TCM data, which is not as effective and meaningful as the FP-growth algorithm in mining.

## 5. Conclusion

The traditional Chinese medicine antiviral prescriptions related research was aimed by the current paper. This paper designed a data mining method based on FP-growth algorithm through the literature data of large-scale traditional Chinese medicine antiviral classic prescriptions, which could analyze the frequency and association rules of the literature data of highly effective antivirus prescriptions by dosage form (soup, pill, paste, and ingot). Research results show that FP-growth algorithm has good performance. The prescriptions selected from the massive dataset have strong generalization and robustness. In this experiment, there are differences in drug combination and antiviral agents among the four main drug dosage forms of Chinese medicinal soup, pills, ointment, and lozenges.

## Data Availability

The data used to support the findings of this study are available upon request to the author.

## Disclosure

This article is one of the results of the 2019 Hunan Provincial Vocational Education Reform Project: Research and implementation of the craftsmanship spirit of medical students under Chinese medical sage culture (Project no. ZJZB2019108).

## Conflicts of Interest

The author declares no conflicts of interest.

## References

- [1] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proceedings of the International Conference on Very Large Data Bases (VLDB 94)*, pp. 487–499, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, September 1994.
- [2] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *Proceedings of the ACM Special Interest Group on Management of Data*, pp. 1–12, New York, NY, USA, June 2000.
- [3] A. L. Hopkins, "Network pharmacology: the next paradigm in drug discovery," *Nature Chemical Biology*, vol. 4, no. 11, pp. 682–690, 2008.
- [4] M. A. Yildirim, K. I. Goh, M. E. Cusick, A. L. Barabasi, and M. Vidal, "Drug-target network," *Nature Biotechnology*, vol. 25, no. 10, pp. 1119–1126, 2007.
- [5] A. Friboulet and D. Thomas, "Systems biology-an interdisciplinary approach," *Biosensors and Bioelectronics*, vol. 20, no. 12, pp. 2404–2407, 2005.
- [6] J. T. Metz and P. J. Hajduk, "Rational approaches to targeted polypharmacology: creating and navigating protein-ligand interaction networks," *Current Opinion in Chemical Biology*, vol. 14, no. 4, pp. 498–504, 2010.
- [7] M. Y. Chang, *Practical Anticancer Prescription*, China Medical Science and Technology Press, Beijing, China, 1998.
- [8] T. Souho, L. Lamboni, L. Xiao, and G. Yang, "Cancer hallmarks and malignancy features: gateway for improved targeted drug delivery," *Biotechnology Advances*, vol. 36, no. 7, pp. 1928–1945, 2018.
- [9] Y. Y. . Tu, "The development of new antimalarial drugs: qinghaosu and dihydro- qinghaosu," *Chinese Medical Journal*, vol. 112, no. 11, pp. 976–977, 1999.
- [10] N. J. Cui and H. Q. Wang, "Study on the mechanism and application of antipyretic and detoxifying traditional Chinese medicine in the prevention and treatment of malignant tumors," *Gansu Journal of Traditional Chinese Medicine*, no. 3, pp. 43–44, 2005.
- [11] X. R. Yan and P. T. Zhang, "Advances in clinical research on the differentiation and treatment of cancer fever," *Guiding Journal of Traditional Chinese Medicine and Pharmacy*, vol. 24, no. 1, pp. 72–74, 2018.
- [12] L. Pan and P. F. Chen, "Progress in research on the mechanism of antitumor action of antipyretic and detoxifying traditional Chinese medicine," *Chinese Archives of Traditional Chinese Medicine*, no. 3, pp. 569–571, 2007.