WILEY | Hindawi

Research Article

Algorithm and Simulation of Association Rules of Drug Relationship Based on Network Model

Hui Teng [iD],[1] Yukun Ma,[2] and Di Teng[1]

[1]*Basic Medical Science College, Qiqihar Medical University, Qiqihar, Heilongjiang 161006, China*
[2]*Institute of Medical Sciences College, Qiqihar Medical University, Qiqihar, Heilongjiang 161006, China*

Correspondence should be addressed to Hui Teng; tenghuisd77@qmu.edu.cn

Studying drug relationships can provide deeper information for the construction and maintenance of biomedical databases and provide more important references for disease treatment and drug development. The research model has expanded from the previous focus on a certain drug to the systematic analysis of the pharmaceutical network formed between drugs. Network model is suitable for the study of the nonlinear relationship of the pharmaceutical relationship by modeling the data learning. Association rule mining is used to find the potential correlations between the various sets of massive data. Therefore, based on the network model, this research proposed an algorithm for drug interaction under improved association rules, which achieved accurate analysis and decision-making of drug relationship. Meanwhile, this research applied the established association rule algorithm to discuss the relationship between Chinese medicine and mental illness medicine and conducted the algorithm research and simulation analysis of the association relationship. The results showed the association rule algorithm based on the network model constructed was better than other association algorithms. It had reliability and superiority in decision-making in improving the drug-drug relationship. It also promoted the rational use of medicines and played a guiding role in pharmaceutical research. This provides scientific research personnel with research basis and research ideas for disease-related diagnosis.

1. Introduction

With the continuous increase in the types of drugs in drug therapy and the emergence of new varieties of drugs, combination drugs have been widely used, and the interaction between drugs has also attracted more and more attention. At present, in intravenous or oral and surgical medications, there is almost no situation where only a single type of medication is used at a time during the treatment of patients. Combination medication is to enhance the therapeutic effect of the drug, while reducing adverse drug reactions and dosage [1]. However, due to the increasing number of combination drugs currently used, the frequency of adverse reactions between drugs is also rising. Drug interaction in combined administration is a very important factor leading to adverse drug reactions. Therefore, the research on drug interactions and drug relationships has important guiding and practical significance.

Data mining is different from traditional data processing methods. It can analyze data in a linear and nonlinear manner, and it can also integrate high-dimensional knowledge. It is especially good at dealing with fuzzy and nonquantified data. Its advantage is that it can reflect the mutual mapping relationship between multidimensional data through data mining methods such as clustering, classification, discrimination, and association. This is consistent with the characteristics of multilevel correlation factors in medical records, which meets the requirements of drug dosage and dose-effect research and can analyze and process drug dosage and compatibility data under the premise of retaining the logic of medical theory, so it can be used as an effective method for researching medical record data information. Its feasibility and applicability in the field of medicine have been recognized by many researchers. Especially in the study of the dose-effect relationship of drug pairs, fuzzy association rules and cluster analysis are very

consistent with the characteristics of prescriptions, and the knowledge excavated has a high accuracy rate.

Based on the three capabilities of pattern recognition, prediction, and simulation, and with its unique simulation, learning, and classification capabilities, network models have been widely used in drug analysis and testing, intelligent drug design, drug controlled release system research, and traditional Chinese medicine efficacy classification. The association rule algorithm in data mining is to discover interesting correlations between item sets in a large amount of data. The premise of the association rule algorithm is to assume that every item set in the database has the same importance. In the actual application of drug efficacy, drug analysis, and drug preparations, different data have different degrees of influence on these aspects, and sometimes, they cannot provide the required prior knowledge. This requires the network to be able to learn by itself [2]. Compared with the common clustering methods of pharmacy laws such as factor analysis and complex system entropy clustering, the process of drug clustering (that is, finding cluster communities) by using association rule algorithm based on the complex network model of drug relationship can make the same drugs appear in multiple clustering results to provide data results from different perspectives. The complex network algorithm based on the network model describes the system structure through the network perspective and focuses on exploring the internal mechanism and connotation of the system, which is more helpful to discover the complexity and implicitness of pharmacy [3].

A key step in drug development and evaluation is to discover potential side effects of drugs in clinical trials. However, large-scale clinical trials are costly and technically complex, and the size of the test population is limited, and some minor or rare side effects are not easy to detect. Many side effects are only discovered many years after the drug is on the market, which is a major loss for both patients and drug developers. Therefore, analyzing and predicting the side effects of drugs through bioinformatics methods have attracted the attention of the academic community.

Early-stage drug relationship decision-making is mainly limited to determining models and methods, and the relationship between drug efficacy and pharmacokinetics is random. The information analysis method using association rules can improve the efficiency of drug interaction, drug use, and drug development [4]. However, the traditional association rule method has the defects of low efficiency and large error in processing the massive data information in the current large-scale drug development process.

Therefore, based on the combination of the network model, this research proposes an algorithm for drug interaction in the base of initiative improved association rules to achieve accurate analysis and decision-making of drug interaction, drug development, and drug use information. At the same time, this study also carried out the research and simulation analysis of the drug-drug relationship in the treatment of Chinese medicine and chemical medicine for the association rule algorithm. The application of association rules in data mining to the analysis of drug interactions promotes the rational use of drugs and thus plays a certain guiding role in pharmaceutical research and rational use of drugs. This has also played a certain role in revealing complex drug preparation research, drug analysis, and pharmacological mechanism and promoting disease treatment and prevention.

2. Modeling of Association Rule Algorithm for Drug Relationship under Network Model

In the actual application of current network models, most of them use BP network based on error backpropagation algorithm. In the research of pharmaceutical preparations, it is difficult to grasp and accurately simulate the complex relationships between them using traditional methods, while the network model is very suitable for handling such complex multivariate nonlinear relationships. Zhao et al. prepared allicin chitosan microspheres by emulsification and cross-linking method and optimized the preparation process of microspheres using a network model. As a result, the prepared microspheres were suitable in size, spherical in shape, and well dispersed, which could meet the requirements of lung targeting microspheres [5]. In drug analysis and testing, due to multiple components and mutual interference between components, quantitative analysis is more difficult. To a certain extent, the network model can simultaneously or separately quantitatively determine the components, which has a certain reference value for determining the content of drugs with overlapping absorption spectra. Wawer et al. applied this method to the multi-component nonseparated determination of the content of t-test of precursor ketone with overlapping ultraviolet absorption spectra, with accurate results and good performance [6]; in intelligent drug design, Xiaorui Hu et al. used 8-6-1 BP neural network and C language to write programs and selected drugs with known structure-activity relationships for training and prediction. The results showed that, with the increase of training modes, the intelligence of the network model was gradually enhanced, and the network model could be used for intelligent drug design [7]. Association rules are the main method of data mining, and the core of the problem of mining association rules is to discover the largest set of items. Nowadays, there are many serial algorithms for finding the largest item set. The usual serial algorithm is to first generate candidate item sets and then calculate their support to generate the largest item set. In this way, the time and space overhead required to generate the largest item set is often very large, and large-scale databases are generally very large (in GB or even TB) [8]. Parallel data mining can obviously improve efficiency. Therefore, data mining has become an important research direction.

This part is based on the association rule algorithm and self-organizing competition network model, modeling and improving the association rule algorithm of the drug relationship, in order to obtain a better response to the drug-drug relationship association rule and further provide services for pharmaceutical research and rational drug use.

2.1. Prototype of Association Rule Algorithm. The association rule algorithm is used for Boolean association rule mining of

large data. Its principle is an item set $L = \{L_1, L_2, \ldots, L_m\}$, which is composed of m Boolean attribute data. There is a database consisting of n transactions, that is, $f = \{f_1, f_2, \ldots, f_n\}$, using a set of binary vectors to represent f. Let A be a subset of L and let $A \subseteq f$, then this Boolean association rule is expressed as $A \Rightarrow B$. Then $A \subseteq L, B \subseteq L$, and $A \cap B = \varnothing$. Tup represents the support of Boolean association rules $A \Rightarrow B$, which includes the ratio of the total number of transactions between A and B and the number of transactions [9]:

$$\text{Tup} = \frac{\text{Tup}(A \cap B)}{n}. \tag{1}$$

Among them, $\text{Tup}(A \cap B)$ is the number of transactions supporting $A \cap B$; the total number of transactions is n. Cont represents the confidence of the Boolean association rule $A \Rightarrow B$, which includes the ratio of the number of transactions of A and B to the number of transactions of A:

$$\text{Cont} = \frac{\text{Tup}(A \cap B)}{\text{Tup}(A)}. \tag{2}$$

Tree projection algorithm, FP-growth algorithm, and Apriori algorithm are all typical Boolean association rule mining algorithms. FP-growth algorithm and tree projection algorithm do not generate candidate sets, and only Apriori algorithm is the method of candidate set generation and testing [10]. This article uses the mining algorithm of the Apriori algorithm and improves on its original algorithm.

2.2. Optimization of Drug Interaction Network Model. The Apriori algorithm is also called a multilevel algorithm. Each layer depends on the results of the previous layer, but the data volume of each layer will be lower than the previous layer. As for the result, it is generated by merging the results of each layer, that is, as the layer deepens and gradually expands, each layer will not be searched again after searching, so this type of algorithm is also a wide search algorithm. The proposed Apriori algorithm makes it possible to conduct rule mining on 0 and 1 data.

Network-based computing methods can analyze the entire heterogeneous network (such as the drug-disease-gene network) by decomposing the useful information in the literature into small subnet models. This model is called network motifs (NMs). NMs are statistically significant repetitive structural patterns, the smallest unit with basic functions and conservative evolution in the pharmaceutical research network, and also an important subnet pattern. It represents the backbone of the network and is an important part of nodes (such as genes and drugs), and these NMs can also form a large summary module and use the associations formed in overlapping NMs to realize specific functions and mine implicit associations. Visualize these complex network models, and define the similarity between expressions based on the evaluation of relevance, thereby forming a network paradigm for data analysis, which is conducive to the analysis of complex systems and high-dimensional data of the interaction between network nodes. In the self-organizing neural network, the neurons in the area around the

"win" neurons are excited to varying degrees, while the neurons outside the area are inhibited. The learning process of the network is the process of self-adaptation and self-organization of the connection rights of the network according to the training samples [11]. After a certain number of trainings, the network can map similar input samples to similar output nodes in the topological sense.

As shown in Figure 1, the self-organizing competitive network is divided into input layer and output layer. The input layer is composed of N neurons, and the output layer is composed of M neurons. The connection weight of the network is k_{ij}, $i = 1, 2, \ldots, M$, and $\sum_{i=1}^{N} k_{ij} = 1$. The state of neuron j in the output layer is derived from the following equation:

$$P_j = \sum_{i=1}^{N} k_{ij} X_i. \tag{3}$$

Among them, X_i is the *i-th* element of the input sample vector, and the neuron f with the largest weight in the competition layer wins the competition, and the output is as follows:

$$a_f = \begin{cases} 1, & P_f \geq P_j, \forall, j \neq f, \\ 0, & \text{others.} \end{cases} \tag{4}$$

The weight after the competition is modified according to equation (5). For all i, there is

$$k_{ij} = k_{ij} + \left(a \frac{X_i}{m} - k_{ij}\right). \tag{5}$$

Among them, a is the learning parameter, generally 0.01 ~ 0.03. m is the number of neurons whose output is 1 in the input layer. When $X_i = 1$, the weight increases. When $X_i = 0$, the weight decreases [12].

The extraction of drug relationships is a multi-classification task that does not distinguish the relationship direction between entities [13]. In this regard, on the basis of the above network model, this research proposes a drug relationship network extraction model that integrates the dependency information attention mechanism, which uses the attention mechanism to fuse the original sentence information and the shortest dependency path SDP information. This method can measure the importance of each part of the original sentence to the relationship between entities from the perspective of syntactic structure. The model architecture is shown in Figure 2.

(1) Input layer: the model is a multi-input model, and the shortest dependency path SDP information represents the dependency relationship between two entities. It can contain the main information expressed in the sentence while discarding redundant noise information.

(2) Embedding layer: convert the input sentence into the corresponding word vector sequence by querying the trained word vector table. In order to obtain the syntactic and semantic information of words, the
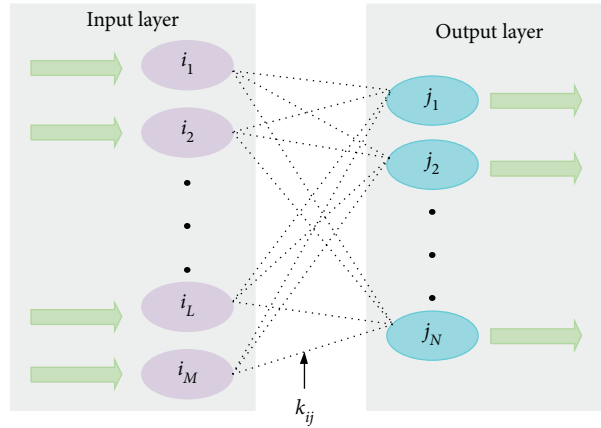
FIGURE 1: Self-organizing competition network model in drug relationship research.
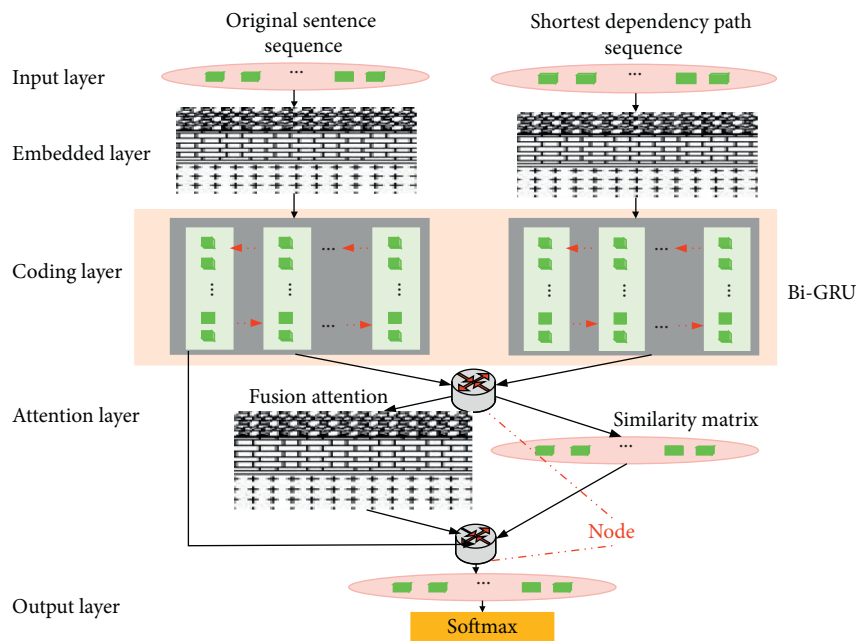

FIGURE 2: Schematic diagram of the framework of the drug relationship network model based on the attention mechanism in this study.

words in the input sequence need to be converted into word vectors. The word vector input representation of the original sentence sequence is $A = \{a_1, a_2, \ldots, a_i, a_n\}$, $a_i \in T^d$; $(i \in [1, n])$ is the word vector representation of the i-th word in the sentence, n is the sentence length, and d is the word vector dimension. The word vector of the shortest dependent path SDP sequence is represented as $B = \{B_1, B_2, \ldots, B_j, B_m\}$, $B_j \in T^d$; $(j \in [1, m])$, which is the vector of the j-th word in the shortest path, and m is the length of the SDP sequence. Finally, the sentence word vector represents A and the shortest dependent path word vector represents B as the input of the model coding layer.

(3) Coding layer: use two independent Bi-GRUs to learn the abstract semantic representation of the original sentence and SDP sequence, respectively. Recurrent

neural network (RNN) is a commonly used network model for natural language processing, which has unique advantages in dealing with time series and language text sequence problems. In order to obtain the contextual information of the sequence, this research uses bidirectional GRU (Bi-GRU) to obtain abstract semantic representation. As shown in equation (6), the final hidden layer representation $m_t(x)$ of each sentence is obtained by concatenating the outputs of the forward and reverse networks:

$$m_t(x) = \left[\overrightarrow{m_t}(x), \overleftarrow{m_t}(x) \right]. \tag{6}$$

In this study, the original sentence input sequence X and SDP input sequence S were used to learn their

semantic information and contextual information through two independent Bi-GRU models, as shown in the following equation:

$$q_{\text{sen},t} = \text{Bi} - \text{GRU}_{\text{sen}}(q_{\text{sen}}, t - 1, x_t),$$
$$q_{\text{sdp},t} = \text{Bi} - \text{GRU}_{\text{ssp}}(q_{\text{ssp}}, t - 1, x_t). \tag{7}$$

The output sequence of Bi-GRU in the original sentence is $M_{\text{sdp}} = \{q_{\text{sen},1}, q_{\text{sen},2}, \ldots, q_{\text{sen},l}\}$ ($M_{sen} \in W^{mX2l}$). The output sequence of Bi-GRU in SDP is $M_{\text{sdp}} = \{q_{\text{sen},1}, q_{\text{sen},2}, \ldots, q_{\text{sen},l}\}$ ($M_{sen} \in W^{mX2l}$), where l is the number of hidden layer units in Bi-GRU.

(4) Attention layer: calculate the similarity matrix between the original sentence and the SDP, fuse the SDP information to get the attention weight, and add the weight of the original sentence to get the final sentence representation. This research proposes a mechanism for fusion of dependent information [14].

Figure 3 shows the process of information fusion. Among them, each unit of P_{sen} and γ represents a vector, and the other units represent a value. Using the obtained attention weight β_i and the original sentence P_{sen} to perform a weighted summation, as shown in equation (8), the sentence that has been fused with SDP information is expressed as $\gamma \in W^{2l}$:

$$\gamma = \sum_{i}^{n} \beta_i \bullet P_{\text{sen}}(i). \tag{8}$$

(5) Output layer: use the output of the attention layer, that is, the sentence representation fused with SDP information for classification prediction. The output layer of the model sends the output γ of the attention layer as the final classification feature to the fully connected layer for classification. The probability $T(y = c)$ that the candidate drug-drug relationship pair y belongs to type c ($c \in C$) is shown in the following equation:

$$T(y = c) = \text{soft} \max(b + W_\gamma * \gamma). \tag{9}$$

Among them, W and b are the weight matrix and bias, respectively. The activation function of the fully connected layer is softmax, and c is a collection of DDI type labels, $c =$ {negate, efect, mechhanism, adview, int}. Finally, the following equation is used to calculate the category label with the highest probability, which is the relationship type of candidate drug-drug pair [15]:

$$\hat{y} = \arg \max H_y \in C (T(y = c)). \tag{10}$$

2.3. Association Rule Algorithm of Drug Relationship Based on Network Model.

Due to the large number of professional vocabularies in biomedical and pharmaceutical literature, if text (such as abstract) is to be mined, data must be cleaned firstly. The co-occurrence method can determine the association between two drug concepts. If they appear in the same article, they can be considered to be related. At present, the most important algorithm to find the implicit relationship between two persons based on co-occurrence is the ABC theory. The basic idea is as follows: assuming that both A and C are related to B, there may be a relationship between A and C, and this relationship may not even be discovered. The process of confirming and quantifying the association between the drug vocabulary hidden in the massive biomedical and pharmaceutical literature through the hypothesis of the association of drug vocabulary A and C is called the "closed exploration ()" process; in this process, if there is a correlation between A and C, then a shared drug concept B is discovered in the literature to support this hypothesis. This process is called the "open exploration ()" process (see Figure 4 for details) [16].

2.3.1. Improvement of Association Rules.

According to the "prototype of association rule algorithm" mentioned above, when measuring the association of biological entities, because the support and confidence are not enough to filter out the useless association rules, the correlation measurement can be used to expand the association rule framework, as shown in the following equation:

$$B[\text{Support, confidence, correlation}] \Rightarrow A. \tag{11}$$

We use *lift* as the correlation measure of correlation. *lift* is defined as follows: if $Q(A \cup B) = Q(A)P(B)$, then the appearance of item set A is independent of the appearance of item set B; otherwise, the events of item sets A and B are interdependent and related. By definition, *lift* can evaluate whether a predictive model is effective, reflecting the importance of the set $\{A\}$ to $\{B\}$. If the value is 0, it means that there is no correlation between $\{A\}$ and $\{B\}$; if the value is positive, the probability of $\{B\}$ increases; if the value is negative, the probability of $\{B\}$ decreases:

$$\text{lift}(A, B) = \frac{(Q(A \cup B))}{Q(A) * P(B)} = \frac{(\alpha * W)}{(\alpha + \beta) * (\alpha + \beta)}. \tag{12}$$

If the value is 1, it means that A and B are independent and have no connection; if the value is less than 1, it means that A and B are negatively correlated, and the appearance of A may cause the absence of B; if the value is greater than 1, then A and B are positively correlated, which means that every occurrence of A implies the occurrence of B, and the greater the value, the greater the probability of occurrence. That is, the appearance of A "boosts" the degree of appearance of B. It is generally believed that the higher the value of lift, the more valuable its association rules [17]. In this study, considering that the entity may be mentioned occasionally or in comparison in the literature, not the
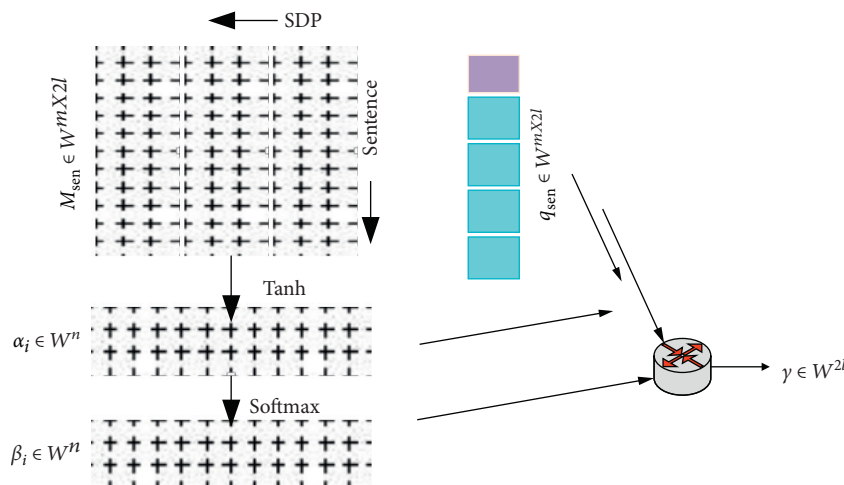
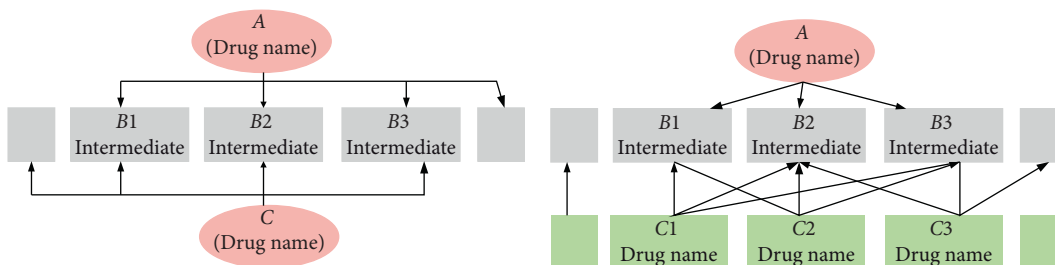FIGURE 3: Schematic diagram of the attention mechanism used in this study.

FIGURE 4: ABC principle of hidden associations between drug vocabularies in biomedical and pharmaceutical literature.

research content, the life threshold is set to 3, so the results obtained may be more meaningful.

2.3.2. Threshold Setting.

Using drug-related network-related databases (such as DrugBank, TDD, and KEGG DRUG) to filter the "*lexical item set*" that has been standardized. When setting the threshold, it should be taken into account that part of the "term set" may only be mentioned occasionally in the literature or just a comparative introduction, without specific research. Therefore, in practical applications, the threshold of *Support_count* can be set to be greater than or equal to 3; use the database to perform a full-text search on the downloaded and cleaned up related documents, and set the drug's *Support_count* threshold to be greater than or equal to 3.

2.3.3. Algorithm of the Network Model.

Based on the above theory, a network model of drug entities can be constructed. Its topological structure contains different subnet modes, which have the same type of network-specific processing tasks. In the associated network, all connected subnet nodes are organized into a homogeneous pattern and a counting method of the frequency of the usage pattern. In summary, the algorithm for constructing a drug network framework in this article is as follows: firstly, given the minimum support threshold, calculate all item sets greater than or equal to *support* (this article mainly refers to the "term set" left after

filtering documents), and obtain the item set of a single *item*; then, based on the correlation metric, the correlation between the items in the *item* set is calculated, and the items that do not meet the minimum *lift* threshold are filtered out; finally, based on the second step and the ABC theory to generate new item sets and their associations, filter out those that do not meet the minimum *lift* value, and obtain the network model data set [18].

2.3.4. Network Topology and Community Analysis of Drug Action.

The drug action network can be directly compiled from the SIDER2 database. This drug action network is obviously a binary graph, that is, its nodes can be divided into two parts, drug and action, and each edge is from a node in one part to a node in another part.

Inspired by the theory that a binary network can be mapped to a univariate undirected graph, we further improved to construct a drug side effect network. If two side effects often occur together, then they are likely to have some kind of internal connection. We ignore the inner molecular biological mechanism of this connection and only gather these connections into a network.

The most basic idea is to count the number of times two side effects occur together as the degree of association between the two side effects. This approach is natural. But there is obviously a problem, that is, those common side effects are more related, while rare side effects are less related. On the

one hand, this is not in line with the actual situation. On the other hand, it is not our intention to ignore rare side effects.

2.4. Design of Analysis System for Drug Interaction Developed by Mining Association Rules. The drug center is a medical institution that designs a dispensing environment in accordance with the requirements of international standards and the characteristics of drugs and has rigorously trained pharmacy technicians to conduct concentrated scientific and rational use of some drugs in accordance with the operating procedures. It serves the rational use of drugs and clinical treatment [19]. Therefore, applying the association rule algorithm in data mining to drug interaction has very important practical and guiding significance. Through the improvement of the previous algorithm, the flowchart of mining association rules for drug interaction is shown in Figure 5.

2.5. R Language Implementation and ROC Curve. This research uses R language, an open-source data analysis system, as the main research tool. It has very powerful analysis and graphing capabilities for specific statistical problems and is suitable for data cleaning, statistical analysis, and network model visualization operations in this research. Meanwhile, for the ROC curve is now widely used in the evaluation of the performance of medical diagnostic experiments, it is also suitable for the diagnosis effect of the discriminant model [20], so this study uses the ROC curve to judge the performance of the algorithm.

3. Simulation of Association Rule Algorithm of Drug Relationship under Network Model

3.1. Establishing an Experimental Environment. This study uses 1–8 computers as slaver and 1 HP server as master. The entire network is connected by an Ethernet switch, and the computers are all independent except for the network connection. The programming language uses Visual C++ 2017, and the database uses SQL Server 2018. The messaging library is standard MPI. MPI is a library based on message passing, which defines the naming method, calling sequence, and library functions that can be called by C/C++ programs. The MPI program design of the master-slave mode adopts the idea of first local and then global [21]. The characteristics of the calculation process can be described as follows:

(1) Data division: the *master* process broadcasts the data table of the data warehouse to the slave processes by random sampling and divides each *slave* process into recording events according to the update time difference

(2) Each slave process i independently calculates the local large item set X with a support degree of sup and the sublocal large item set Y with a support degree between $Tup/n \sim Tup$ on the local data

(3) The slave process uses *MPI Send* to exchange the calculated local large item set X with other processes, so that each process has X

(4) Each process exchanges the count on X with other processes, and the sublocal large item set Y is directly used during the exchange, thus avoiding another scan of the database and finally getting a global count everywhere

(5) The master process collects various frequent sets of *MPI Recv* from the slave process of *Slaver* and checks whether the minimum support is reached, and the final large item set S is obtained

3.2. Performance Analysis of the Algorithm. In this study, through the improvement of the network model and association rules, the speedup ratio $S = Q/(1 + K_m/K_P)$ and efficiency $E = S/Q$ are introduced to analyze the performance of the algorithm. Among them, Q is the number of word nodes, K_p is the time of k frequent sets and $k+1$ candidate sets are generated when each word node executes the algorithm of this research, K_m is the time for each node to send and receive frequent sets of k items [22]. Table 1 shows the statistical information of the main process.

As can be seen from Table 1, as the number of word nodes increases, the speedup ratio of the algorithm increases, and the total execution time decreases, but the efficiency decreases. Therefore, the improved algorithm has a certain degree of scalability and can improve computational efficiency.

3.3. Comparison of Association Rule Algorithm Based on Network Model with Other Algorithms. In order to verify the comparison between the algorithm of this study and other algorithms to reflect its performance superiority, we use a random sampling method, using 8 computers as workstations, and generate test databases on 8 computers, each with 6000 records information. The results are shown in Figure 6.

According to the test results, the association rule algorithm based on the network model in this study was effective. Under the same degree of support, the number of database scans and execution time were reduced; it was superior to other algorithms in terms of calculation method and synchronization times. And with the decrease of support, the performance advantage of the association rule algorithm became more obvious.

4. Simulation Analysis of Drug Relationship Association Rule Algorithm under Network Model

4.1. Algorithms and Simulations of the Law of Chinese Medicine Use and the Relationship between Chinese Medicines under the Network Model. Data mining is a powerful tool for analyzing and processing Chinese medicine information. It can analyze and sort huge data to realize the rational use of effective information. In particular, the research and application in the prescription compatibility law is also a hot issue of Chinese medicine data mining [23]. Data mining can provide a reliable method for scientific, reasonable, and efficient analysis of the compatibility law contained in the
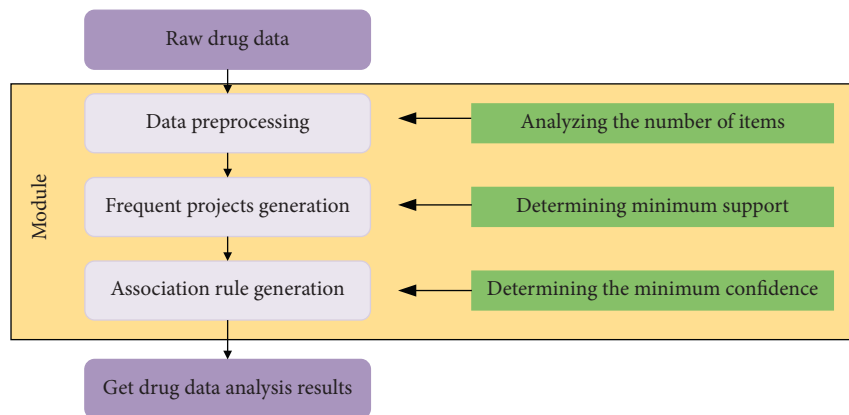
FIGURE 5: Flowchart of association rules for drug relationship analysis.

TABLE 1: Main process statistics in the performance analysis of the improved correlation specification algorithm based on the network model.

	$Q = 4$	$Q = 6$	$Q = 8$	$Q = 10$
Comm.	3.15	5.68	8.68	12.59
Total time	68.51	41.59	31.97	25.61
Speedup	2.08	3.52	5.09	5.81
Comp.	61.53	40.51	35.16	21.67
Efficiency	0.98	0.91	0.82	0.75

target prescription data from multiple angles and levels, such as the frequency of prescriptions and drug combinations, and the results provide important guidance for clinical drug use and new drug development [24].

This part uses the "Chinese Traditional Medicine Retrieval System," "China Knowledge Network," and "Yaozhi Data" three databases to collect prescriptions for treating spleen deficiency, excluding prescriptions containing only single medicine, and a total of 186 effective prescriptions are collected. Using the association rule algorithm under the network model constructed above, the association rule mining analysis was carried out on the 186 prescriptions for the treatment of spleen deficiency [25]. The names of the medicines in the prescriptions were processed in accordance with the "Pharmacopoeia of the People's Republic of China"(2020 edition) and then simulated and analyzed by *Matlab* [26, 27].

4.1.1. Frequency Analysis of Drugs in the Prescription Library. The frequency statistics of a total of 359 traditional Chinese medicines in the 186 prescriptions were conducted. There were 151 traditional Chinese medicines with a frequency of ≥3 times. The names of some medicines are shown in Table 2.

The results of the excavation of the frequency of medication showed that among the prescriptions used by physicians in the past dynasties to treat spleen deficiency, the most commonly used drugs were Poria, Rehmannia, Yam, Eucommia, Aconite, etc. Poria and Chinese yam are calming and sweet in nature, which can replenish yin and yang and also invigorate qi and essence. The meridian shows that they

can act on the lung, spleen, and kidneys. Therefore, these two traditional Chinese medicines can be combined with other drugs to treat various types of spleen deficiency.

4.1.2. Analysis of the Nature, Flavor, and Meridian of the Drugs in the Prescription Library. Sixty-seven traditional Chinese medicines (frequency ≥5) with the highest frequency in the prescription library were selected, the types of traditional Chinese medicines were calculated according to their properties, five flavors, and the distribution of meridians, and the frequency and proportion were calculated, respectively. The results are shown in Figure 7. According to the content of medicinal properties in the first part of the Pharmacopoeia of the People's Republic of China (2020 edition), the meridian distribution of the properties and flavors of each Chinese medicine is summarized.

The results of the analysis of the nature and flavor of the high-frequency drugs in the prescription library showed that the properties of the drugs for treating spleen deficiency are mostly "warm," "ping," and "cold"; the medicinal taste is mostly "sweet" and "pungent"; the main meridian was "kidney" and "spleen". According to the relationship between drug combination and efficacy and according to the characteristics of drug return, it was concluded that most of the drugs for treating spleen deficiency directly act on the spleen or the spleen and kidney; according to the characteristics of the drug's nature and flavor, it was concluded that most prescriptions for treating spleen deficiency used "pungent" medicine to nourish the kidney and used "sweet" medicine to nourish qi and benefit the kidney.

4.1.3. Simulation of Association Rule Algorithm. This research uses the association rule algorithm based on the network model constructed as described above, setting the support degree ≥0.1 and the confidence degree ≥0.5 and analyzing the association rules of the prescription database for the treatment of spleen deficiency [28–30]. The results of the mining are shown in Figure 8(a). At the same time, *Matlab* software is used to visualize the drugs in the strong association rules, as shown in Figure 8(b).
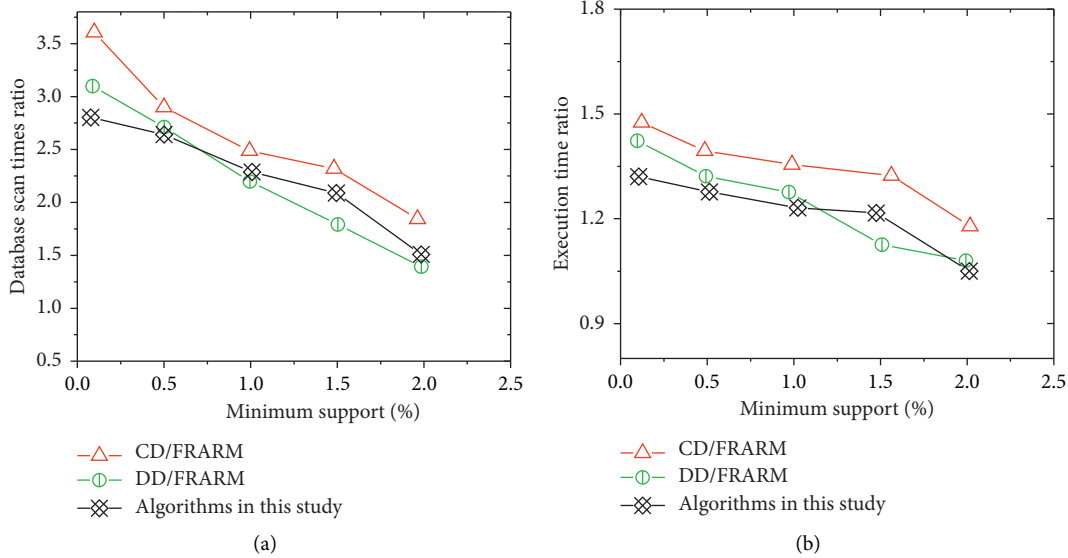
FIGURE 6: The comparison result of the association rule algorithm based on the network model and other algorithms in this study.

TABLE 2: Examples of some Chinese medicines with a frequency of ≥3 times.

Name	Frequency
Poria	69
Rehmannia	65
Yam	54
Eucommia	48
Aconite	35
White peony	29
Ginger	18
Ginseng	16
Licorice	12
Atractylodes	8
Astragalus	5
Xiangfu	5

The results in Figure 8 show that Poria is at the core of the prescription library for treating spleen deficiency. Although Poria does not have the effect of invigorating the kidney, it can be combined with many drugs (such as yam, Rehmannia glutinosa, dogwood, Schisandra, Achyranthes, and Ginseng) which constitute the core medicine for spleen strengthening. Poria and Chinese yam can strengthen the spleen and stomach, nourish the spleen and stomach, invigorate the kidney, and astringe the essence, especially suitable for patients with deficiency of spleen and kidney. The compatibility of Poria and Cornus has the effects of warming and replenishing liver and kidney, astringing, and solidifying and is suitable for treating patients with deficiency of kidney essence and kidney yang. Using the correlation specification algorithm of this study, fully excavating the treasure house of traditional Chinese medicine culture, and systematically and deeply studying the prescriptions of ancient doctors will have very important guiding significance for clinical treatment. Exploring the traditional Chinese medicine prescriptions used by physicians to treat spleen deficiency in medical books, and summarizing the rules of prescriptions, can provide a reference for clinical treatment of spleen deficiency.

4.2. Association Rule Algorithm and Simulation of Mental Illness Drug Relationship under the Network Model. This part takes the medicine for the treatment of schizophrenia as an example. First, the biomedical literature related to schizophrenia is obtained from the PubMed database, and the correlation between the medicine and the medicine for the treatment of schizophrenia is obtained through data cleaning. ABC theory and association rules are used to quantify whether there is an association between schizophrenia drugs and the degree of association, and then the visualization of the network model is realized, finally, the network node association and model structure are analyzed, and the ROC curve is used to verify the reliability of the algorithm in this study.

4.2.1. The Association between Drugs for the Treatment of Schizophrenia. Based on the aforementioned network pattern association rule algorithm, the *lift* value of all schizophrenia genes is calculated. In order to better analyze the drugs with high correlation, we set the threshold of *lift value* to 9 and obtained 85 drugs and 154 high correlations between them and generated a schizophrenia drug network model (Figure 9). After deduplication, a total of 109 schizophrenia-related drugs and their support values were obtained. See Table 3 for information on some related drugs.

In the drug association network in Figure 9, most of the nodes have a small degree, and a few nodes have a large degree. This conforms to the power law distribution and belongs to the scale-free network. This kind of network is characterized by its robustness to random failures and its vulnerability to targeted deliberate attacks. In the field of biomedicine, this feature illustrates the importance of key nodes. Among the key nodes are sulpiride, tiapride, etc.
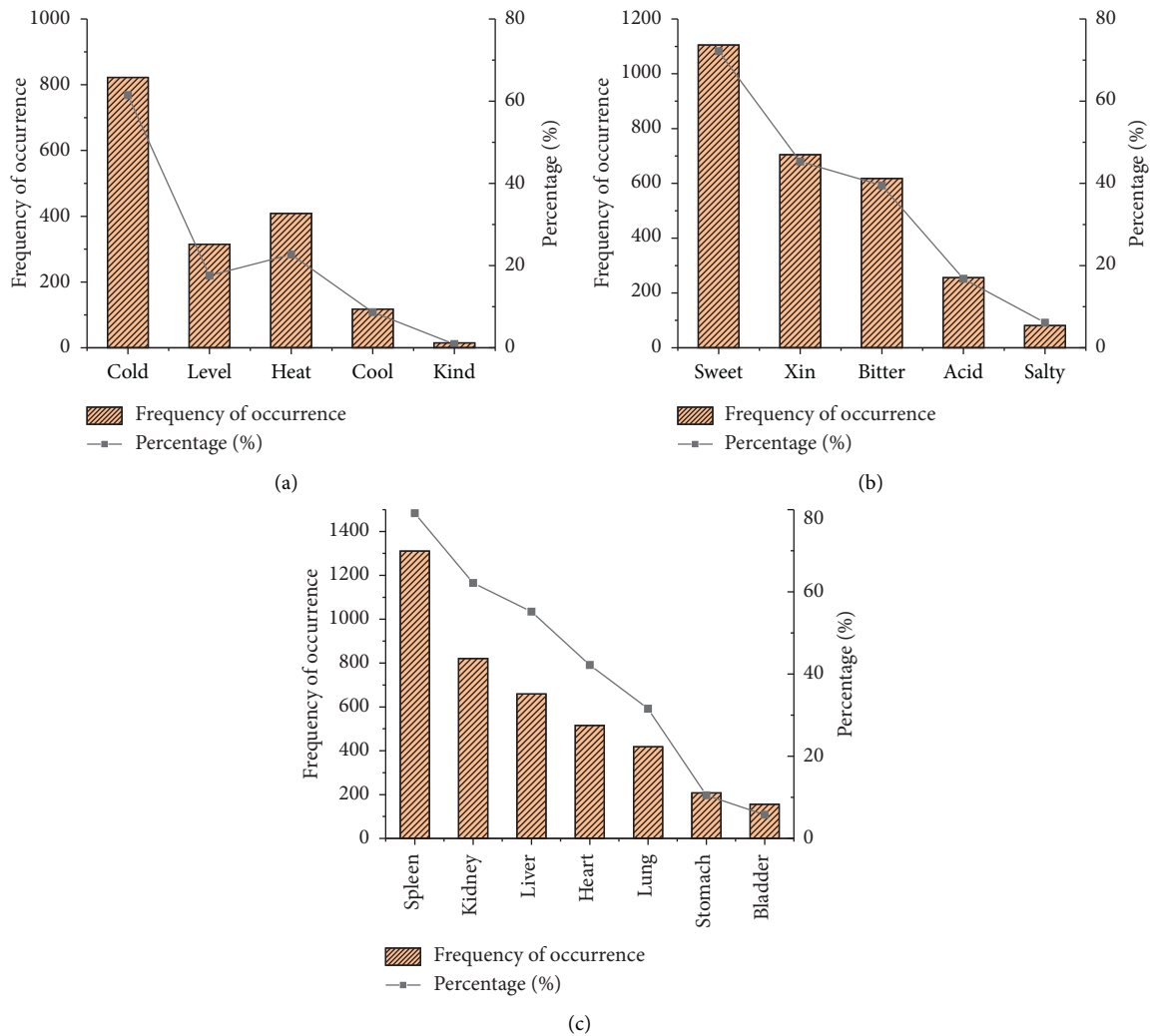
(a)

(b)

(c)

FIGURE 7: Results of the nature (a), flavor (b), and meridian (c) of the drugs in the prescription library.

These key nodes are research hotspots in schizophrenia drug research and may interact with many other drugs. In Figure 9, the isolated nodes of 21 drugs (fluphenazine, etc.) have been deleted, leaving 88 drugs. It can be found that the two drugs, aspirin and diclofenac, are the most special, with only a single correlation, and no correlation with other drugs, and the correlation is the highest.

4.2.2. ROC Curve Evaluation. This part verifies all the correlation results between the abovementioned drugs for the treatment of schizophrenia and uses the ROC curve to judge the performance of the algorithm in the SPSS 22 environment (Figure 10). It can be obtained that the area under the ROC curve is 0.895, the correlation accuracy is moderately high, the corresponding standard error is 0.049, the $P \le 0.001$, and the 95% confidence interval is (0.730, 0.996).

It can be seen from the results of the ROC curve that the management rule algorithm based on the network model constructed in this study is better than other biological entity association extraction algorithms. This provides researchers with research basis and research ideas for future schizophrenia-related diagnosis and treatment, disease candidate gene screening, targeted drugs, drug repositioning, and personalized medicine. Similarly, this
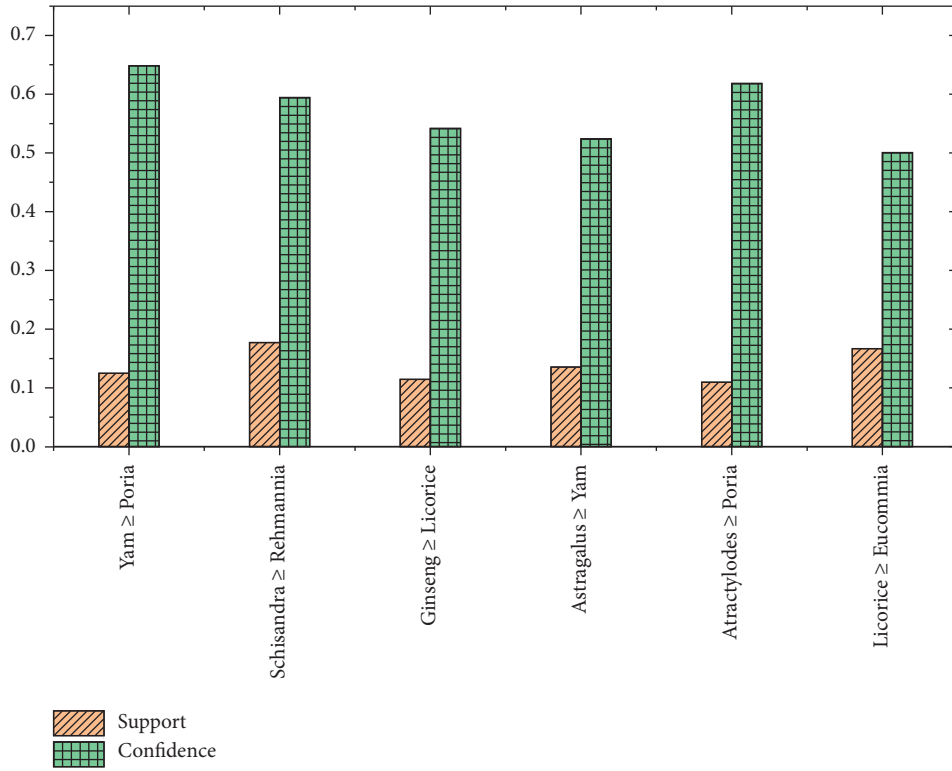
(a)
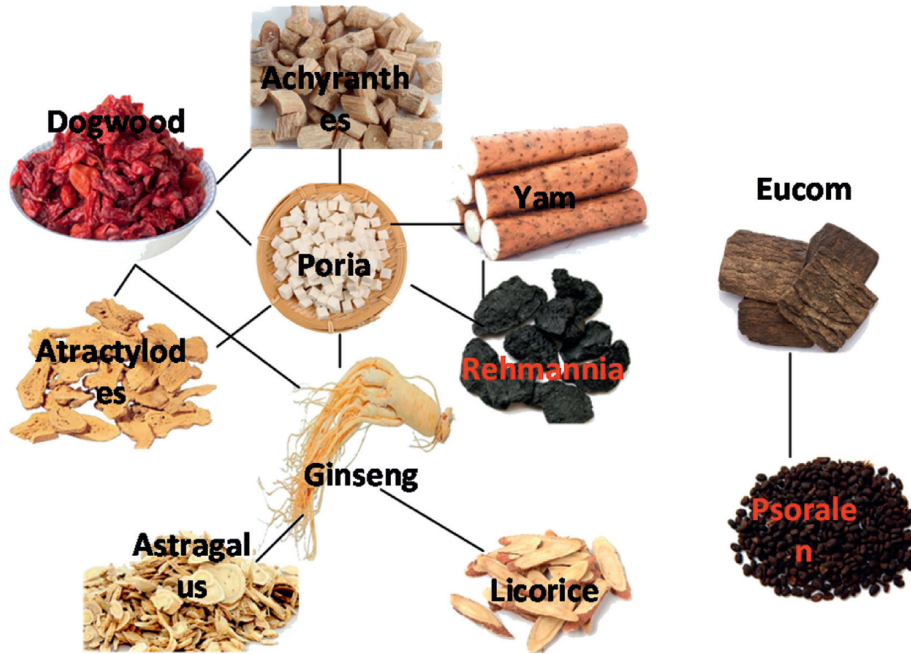
(b)

FIGURE 8: Analysis results and display of association rules of Chinese medicine. (a) The results of the mining. (b) Visualization of the drugs in the strong association rules.
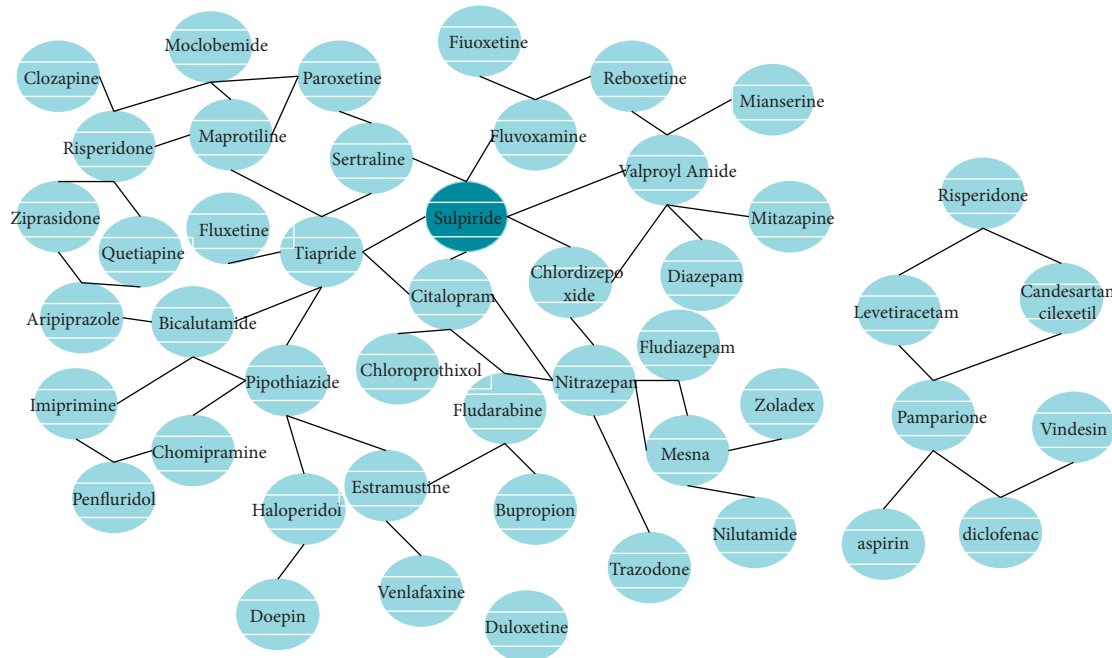
FIGURE 9: Network model of schizophrenia drugs.

TABLE 3: Examples of some schizophrenia-related drugs and their support values.

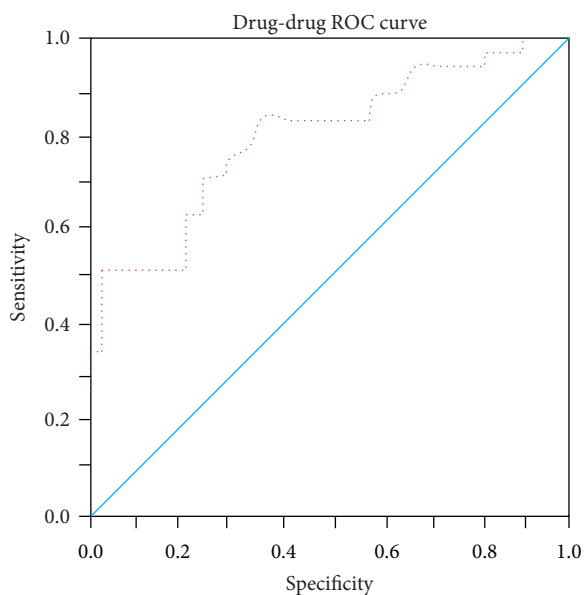| DrugBank ID	Drug name	Support count	Support (%)
DB00391	Sulpiride	515	12.58
DB13025	Tiapride	459	10.16
DB00623	Fluphenazine	325	9.49
DB00679	Thioridazine	256	7.18
DB00734	Risperidone	194	5.19
DB01202	Levetiracetam	147	4.78
DB0502	Haloperidol	75	2.16
DB13791	Penfluridol	51	1.57

FIGURE 10: ROC curve performance evaluation of all the results of the drug-drug treatment for schizophrenia.

algorithm model can also be used to analyze other clinical diseases.

5. Conclusion

In the massive biomedical literature, there are a large number of associations between biological entities. The systematic analysis of these heterogeneous data brings unprecedented opportunities to biologists, enabling them to infer the degree of association between different biological entities in the context of personalized medicine and translational medicine. However, these associations are very complex and sparse, and the amount of calculation for direct query is very challenging. The construction of the network model broadens the thinking for researchers to reposition drugs. From the traditional one disease corresponding to one target, it has developed from a holistic perspective to present the complex biological process of disease occurrence and the law of action of drugs in the body. This is a systematic analysis that integrates the whole, dynamic, and comprehensive. It has reached a new height in mining the hidden information in the obtained data resources. The association rule mining is used to find the potential correlation between the various sets in the massive data. It is an important type of algorithm in data mining. It has been widely used in various industries in recent years. Therefore, based on the combination of the network model, this research proposed an algorithm for drug interaction based on improved association rules, which realized accurate analysis and decision-making of drug interaction, drug development, and drug use information. At the same time, this research applied the established association rule algorithm to the relationship between Chinese medicine and mental illness treatment drugs and conducted algorithm research and

simulation analysis of related association relationships. The results showed that the association rule algorithm based on the network model constructed in this study was better than other association extraction algorithms. In improving the decision-making in the drug-drug relationship, it has high reliability and intelligence, which promotes the rational use of drugs, and thus plays a certain guiding role in pharmaceutical research. This also provides research basis and research ideas for scientific researchers on the repositioning of disease-related diagnosis and treatment drugs and personalized medicine.

Data Availability

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

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, vol. 72-73, pp. 303–315, 2016.

[2] L. Jing, T. Wang, M. Zhao, and P. Wang, "An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox," *Sensors*, vol. 17, no. 2, p. 414, 2017.

[3] I. N. M. Shaharanee and F. Hadzic, "Evaluation and optimization of frequent, closed and maximal association rule based classification," *Statistics and Computing*, vol. 24, no. 5, pp. 821–843, 2014.

[4] A. Gupta, A. Giridhar, G. V. Reklaitis, and V. Venkatasubramanian, "Intelligent alarm system applied to continuous pharmaceutical manufacturing," *Computer Aided Chemical Engineering*, vol. 32, pp. 499–504, 2013.

[5] Z. Zhao, Z. Yang, L. Luo et al., "Drug drug interaction extraction from biomedical literature using syntax convolutional neural network," *Bioinformatics*, vol. 32, no. 22, pp. 3444–3453, 2016.

[6] M. J. Wawer, D. E. Jaramillo, V. Dančík et al., "Automated structure-activity relationship mining," *Journal of Biomolecular Screening*, vol. 19, no. 5, pp. 738–748, 2014.

[7] X. Hu and C. Lin, "A preliminary study on targets association algorithm of radar and AIS using BP neural network [J]," *Procedia Engineering*, vol. 15, pp. 1441–1445, 2011.

[8] N. A. Rajesh, "Arul lawrence selvakumar. Association rules and deep learning for cryptographic algorithm in privacy preserving data mining," *Cluster Computing*, vol. 22, no. 1, pp. 119–131, 2019.

[9] Y. Ji, H. Ying, J. Tran, P. Dews, A. Mansour, and R. Michael Massanari, "A method for mining infrequent causal associations and its application in finding adverse drug reaction signal pairs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 4, pp. 721–733, 2013.

[10] S. Tanvir Habib and Z. Ansari, "An analysis of MapReduce efficiency in document clustering using parallel K-means algorithm," *Future Computing and Informatics Journal*, vol. 3, no. 2, pp. 200–209, 2018.

[11] J. Tang, D. Wang, Z. Zhang, L. He, J. Xin, and Y. Xu, "Weed identification based on K-means feature learning combined with convolutional neural network," *Computers and Electronics in Agriculture*, vol. 135, pp. 63–70, 2017.

[12] J. R. Horn, K. F. Gumpper, J. C. McDonnell, S. Phansalkar, and C. Reilly, "Clinical decision support for drug-drug interactions: improvement needed," *American Journal of Health-System Pharmacy*, vol. 70, no. 10, pp. 905–909, 2013.

[13] Y. Ji, H. Ying, D. Peter et al., "A potential causal association mining algorithm for screening adverse drug reactions in postmarketing surveillance," *International Conference of the Ieee Engineering in Medicine and Biology Society*, vol. 15, no. 3, pp. 428–437, 2011.

[14] S. Liu, K. Chen, Q. Chen et al., "Dependency-based convolutional neural network for drug-drug interaction extraction," in *Proceedings of the 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Shenzhen, China, 2016.

[15] C. Zheng and R. Xu, "Large-scale mining disease comorbidity relationships from post-market drug adverse events surveillance data," *BMC Bioinformatics*, vol. 19, no. 17, p. 500, 2018.

[16] W. Wu Guo, X. Zhang, B. Yang et al., "Analysis on principle of treatment of cough of yan zhenghua based on Apriori and clustering algorithm," *China Journal of Chinese Materia Medica*, vol. 39, no. 4, p. 623, 2014.

[17] S. R. Yogita, J. W. Sangma, S. R. N. Anal, and V. Pal, "Clustering-based hybrid approach for identifying quantitative multidimensional associations between patient Attributes, drugs and adverse drug reactions," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 12, no. 3, pp. 237–251, 2020.

[18] C. Upasana, W. S. Jerry, V. Pal et al., "Data-driven extraction of quantitative multi-dimensional associations of cardiovascular drugs and adverse drug reactions," in *Proceedings of the International Conference on Practical Applications of Computational Biology & Bioinformatics*, pp. 70–77, Shenzhen, China, 2019.

[19] A. Bate, M. Lindquist, I. R. Edwards et al., "A Bayesian neural network method for adverse drug reaction signal generation," *European Journal of Clinical Pharmacology*, vol. 54, no. 4, pp. 315–321, 1998.

[20] H. Leung, "Neural network data association with application to multiple-target tracking," *Optical Engineering*, vol. 35, no. 3, pp. 693–700, 1996.

[21] J. Xing and Y. Wu, "Study on main drugs and drug combinations of patient-controlled analgesia based on text mining," *Pain Research & Management*, vol. 2020, Article ID 8517652, 7 pages, 2020.

[22] Q. Sun, D. Shaw, and C. H. Davis, "A model for estimating the occurrence of same-frequency words and the boundary between high- and low-frequency words in texts," *Journal of the American Society for Information Science*, vol. 50, no. 3, pp. 280–286, 1999.

[23] Y. Zhang, W. Zheng, H. Lin, J. Wang, Z. Yang, and M. Dumontier, "Drug-drug interaction extraction via

hierarchical RNNs on sequence and shortest dependency paths," *Bioinformatics*, vol. 34, no. 5, pp. 828–835, 2018.

[24] S. Sunil Kumar and A. Anand, "Drug-drug interaction extraction from biomedical texts using long short-term memory network," *Journal of Biomedical Informatics*, vol. 86, pp. 15–24, 2018.

[25] W. Cedeno and D. Agrafiotis, "A comparison of particle swarms techniques for the development of quantitative structure-activity relationship models for drug design," in *Proceedings of the 2005 IEEE Computational Systems Bioinformatics Conference - Workshops (CSBW'05)*, pp. 322–331, Stanford, CA, USA, 2005.

[26] G. Agapito, P. H. Guzzi, and M. Cannataro, "Parallel extraction of association rules from genomics data," *Applied Mathematics and Computation*, vol. 350, pp. 434–446, 2019.

[27] K. Shimada, S. Hasegawa, S. Nakao et al., "Adverse event profiles of ifosfamide-induced encephalopathy analyzed using the food and drug administration adverse event reporting system and the Japanese adverse drug event report databases," *Cancer Chemotherapy and Pharmacology*, vol. 84, no. 5, pp. 1097–1105, 2019.

[28] M. Pham, F. Cheng, and K. Ramachandran, "A comparison study of algorithms to detect drug-adverse event associations: frequentist, bayesian, and machine-learning approaches," *Drug Safety*, vol. 42, no. 6, pp. 743–750, 2019.

[29] B. Hernández, R. B. Reilly, and R. A. Kenny, "Investigation of multimorbidity and prevalent disease combinations in older Irish adults using network analysis and association rules," *Scientific Reports*, vol. 9, no. 1, pp. 1–12, 2019.

[30] J. Rauch, "Expert deduction rules in data mining with association rules: a case study," *Knowledge and Information Systems*, vol. 59, no. 1, pp. 167–195, 2019.