INT. J. BIOAUTOMATION, 2018, 22(3), 263-274 doi: 10.7546/ijba.2018.22.3.263-274

Research on Exploring the Patients' Hiding Disease Based on Symptom Weighted Clustering Technique

Yingying Peng, Gang Yi*

School of Management and Information Engineering
Hunan University of Chinese Medicine
Changsha 410208, Hunan, China
E-mails: 31581677@qq.com, 6636244@qq.com

**Corresponding author*

Received: November 7, 2017 **Accepted: August 15, 2018**

Published: September 30, 2018

Abstract: *The research regards the diagnostic data of the patients' disease as the mining data source. Each diagnostic data includes patients' symptom and sickness. The paper regards a certain patient as the mining target, considering the symptom weighted situation, using the clustering method in the data mining to dig the tendency of patients' disease. In addition, the paper combines a group center formed by the patients' symptom and designs a symptom weighted clustering method to satisfy the diagnostic data of minimal symptom similarity which belongs to the clustering. Later, the disease item whose number is the maximum can be found out in the clustering and the tendency of patients' sickness. The methods proposed in the paper design and build a diagnostic system of patients' sickness. The mining results of system can offer some useful referent information for those people check the sickness tendency of patients' disease or those medical staff whose clinical experience is not enough confirms the disease diagnosis.*

Keywords: *Data mining, Clustering, Symptom weighted, Disease.*

Introduction

With the development of information and technology, the diagnostic data of the patients stored in the hospital has been changed from the traditional paper records into electronic medical records. According to the description of the Computer-based Patient Record Institute (CPRI), In regard to a human's lifelong healthy state and the electronic information of medical care, the electronic medical record will replace the paper record. This way can conform to the clinic application, the public administration, the medical education, the research investigation and other legally needed mainly medical data source. The relevance between the external symptoms and diseases can be found as the reference information of the medical diagnosis from the past patients' diagnostic data. In addition, it can improve the accuracy and efficiency of the medical diagnosis, reducing the delay of the procedure of diagnosing the disease and is the main research topic in the field of applying the diagnostic data.

The data mining is to dig the potential useful information and knowledge from a number of data and is regarded as the referent information of the decision analysis. The data mining has been applied in many fields [7]. The research regards the patients' diagnostic data as the mining data source, and each diagnostic data records the symptoms and diseases of the patients. The patient's symptom is as the mining target and the clustering is used to analyze that those symptoms and the diseases have the highest relevance. In this way, the patients' symptoms can be mined that they may have the tendency of the diseases. For example, the patients' symptoms include chest discomfort, chest distress, chest heavy, chest pain and others; they may suffer from the cardiovascular disease. The paper regards a patient's symptom as the mining target and makes the sunset of the symptom be the cluster's central point. Considering the situation of the symptom weighted, a clustering is designed to adjust the symptom similarity between the diagnostic data and the central points. It can also satisfy the diagnostic data of minimal symptom similarity which belongs to the clustering and find out the tendency of suffering from the disease changing from the patients' symptoms. The diagnostic data of a medical center in the south is as an example; the disease diagnostic system of a patient is designed and built. The mining results can offer some useful guiding information for the tendency of patients' sickness and can provide some useful efficiency of receiving the disease treatment.

The relevant research

The data mining can find out some potentially useful information and knowledge from a number of data, it can finish the following tasks: association rules, clustering, classification, sequential pattern analysis, forecasting and others [3, 15]. The application information technology in the medical field can develop medical informatics, and its purpose is to build the medical knowledge and find out the medical guidance of all diseases with the support of the information technology, regarding the diseases as the center and the medical problems as the oriented diagnostic models. If the information technology can be effectively used in the disease diagnosis, the relevance between the disease symptoms and possible diseases can be analyzed so that it can offer great help to the treatment of the disease and the prevention of the diseases.

Many researches have shown that the data mining can be effectively applied in the medical diagnosis, and its relevant researches are as follows: Ye et al. [14] take the insomnia and the cardiovascular and cerebrovascular diseases as an example, the concurrence and the causal relation among the diseases can be analyzed with the use of data mining. Vani and Shimabukuro [13] regard the standard health data as the source of the system data with the use of the data mining technology. It can develop a special data mining system in the medical field to explore the relations among different diseases so that the data can be used as the reference for the future prevention treatment. Ji et al. [5] adopt the data mining technology to construct the knowledge management system for the classification of the hospital diseases. Molina et al. [10] use the data mining technology to conduct the medical forecast for the patients who suffer from the tuberculosis. Regan et al. [11] use the association rules to find out the probability of the diseases. They used to adopt the association rules with composite items to discuss the relevance between the symptoms and the diseases.

The clustering can cluster its objects in terms of the similarity. The research of the clustering is mainly as follows: the partitioning, the partitioning, the grid-based, the density-based, the model-based and others [9].

Berry and Linoff [2] used to describe that if people start to analyze the data, know about the meaning of the data and describe the best utility pattern, the cluster analysis is a better way. The research will modify the method of the partitioning clustering and then regard the modified method as the basis of the clustering diagnostic data.

Some more famous methods of the partitioning clustering are partitioning around medoids (PAM) [4], *k*-means [8, 12] and CLARANS [1]. The *k* clusters $k \geq 1$ appointed by the users can be clustered and the partitioning method can make each object be assigned to the most similar clustering. The clustering steps about the PAM algorithm are as follows: Kaufman and Rousseeuw [6] propose the PAM algorithm which can make the whole objects be clustered into *k* clusters. The PAM can make each cluster decide the representative objects, and representative objects can be called as *k*-medoids. If *k*-medoids are chosen, the objects which do not belong to the medoid can be decided which cluster they belong to according to the similarity. Its similarity can be represented by the Euclidean distance. $d\left(O_a,\ O_b\right)$ represents the distance between O_a and O_a.

For example, O_i is the medoid and O_j is not the object of the medoid, if

$$d\left(O_j,\ O_i\right) = \min\{d(O_j, O_e)\}$$

and O_e represents all medoids, O_j belongs to the O_i clustering.

As for any O_j which does not belong to medoid, if the medoid O_i is replaces by O_h which does not belong to medoid, the definition of the changing cost C_{jih} is as follows:

$$C_{jih} = d\left(O_j,\ O_m\right) - d\left(O_j,\ O_n\right). \tag{1}$$

O_m shows that the medoid which has the maximum similarity (the shortest distance) with O_j after the O_h replaces the O_i. O_n shows that the medoid which has the most similarity (the shortest distance) with O_j before the O_h replaces the O_i.

$$TC_{ih} = \sum C_{jih}. \tag{2}$$

If $TC_{ih} > 0$, the total distance is larger than before, the O_i can not be replaced by the O_h. TC_{ih} is used as the measuring basis, the explanation of the PAM algorithm is as follows:

1. Randomly choose k objects as medoids.
2. As for all the combination between O_i and O_h, the TC_{ih} can be calculated. O_i represents any medoid, and O_h represents any objects which do not belong to the medoid.
3. Choose the O_i whose TC_{ih} is the minimum value to combine with the O_h. If TC_{ih} < 0, O_h can replace the O_i and become the medoid, later it can jump to the step 2.
4. Otherwise the execution should be stopped and the clustering is finished.

The research regards the patients' each diagnostic data as the data mining source with the situation of the symptom weighted. The clustering technology is used to discuss the tendency of the diseases from the patients' symptoms. The form of the diagnostic data is $\{S,\ D\}$, S is the set which includes one or more symptoms, D is the set which includes one or more diseases. Each diagnostic data include the patients' symptoms and the diseases. For example, a diagnostic data is $T_1 = \{abc,\ X\}$, and its showing symptom is *abc*, its disease is *X*, that is, $\{a,\ b,\ c\} \subseteq S \setminus X \subseteq D$.

Mining the tendency of the diseases from the patients' symptoms

During the process of diagnosing the patients' disease, the doctors hope to know about the relevance between the symptoms and disease. As for the uncommon or unique symptoms, the weighted value of these symptoms can be improved so that it can easily become the affected symptoms. Later the relevance between the symptoms and the diseases can be found out which can be the basis on mining the tendency of the diseases from the patients' symptoms. The paper uses the patients' each diagnostic data as the data mining source and each diagnostic data include the patients' symptoms and the diseases. A patient's symptom is as the target under the situation of the symptom weighted, the clustering method can be used to dig the tendency of the diseases from the patients' symptoms.

The clustering method

The paper defines the diagnostic data and the symptom similarity among central points, the similarity is as the basis on whether the diagnostic data belong to the clusters.

The symptom's number of the symptom similarity

$$s = \{\text{diagnostic data} \cap \text{clustering center points}\} \text{ / The symptom's number } \{\text{diagnostic data} \cup \text{clustering center points}\}.$$

For example, if each diagnostic data is $\{abc, X\}$ and the clustering center point is $\{acde\}$, $\{a, b, c, d, e\}$ is the set of the symptoms, $\{X, Y\}$ is the set of the diseases, so the symptom similarity is 40%.

If the patients' symptom is P, P is the item set which contains a or more symptoms considering the situation of the symptom weighted. If P contains k symptoms, $s_1, s_2, \ldots, s_j, \ldots, s_k$ respectively, $k \geq 1$ represents and the corresponding weighted value is shown by $w_1, w_2, \ldots, w_j, \ldots, w_k$. The calculating method of the P's symptom weighted similarity is as follows:

$$w = 0; j = 0; \tag{3}$$

$$\text{for } (i = 1;\ i \leq k;\ i{+}{+}) \tag{4}$$

$$\text{if } \ s_j \in \{\text{ diagnostic data} \cap \text{clustering center points}\} \tag{5}$$

$$\text{then } \{w = w + w_i;\ j = j + 1;\} \tag{6}$$

$$w = \frac{w}{j}; \tag{7}$$

$$\text{the symptom weighted similarity } ws = \sum_{i=1}^{k} w \times s. \tag{8}$$

For example, P includes 3 symptoms *abc*, the weighted value of the *abc*, is respectively represented by 1, 2 and 3. $ws = 2.5 \times s$ is because that {diagnositic data $\cap$ clustering center points} = {*bc*}, *j* = 2, *w* = 5/2 = 2.5. The paper assumes a minimum symptom similarity and P is the patient's symptom, the subset of P is represented by

$$S = \{s_1, s_2, \ldots, s_j, P\},\ S_i \subseteq P.$$

 doi: 10.7546/ijba.2018.22.3.263-274

A simple and rapid clustering method is designed in the paper, s_i, is the central point of a cluster and is the same cluster whose diagnostic data T_j is satisfied with the minimum symptom similarity. It is called S_i-cluster. $1 \le j \le m$, represents the m diagnostic data.

The process of the clustering is as follows:

Clustering(P){ (9)

Calculate the subset P: $S = \{s_1, s_2, \ldots, s_j, \ldots, P\}$ (10)

for ($j = 1$; $j \le m$; $j++$) (11)

while ($S = \varnothing$) { (12)

Select the next s_i and set the cluster as the central points;

Calculate the symptom similarity s between the diagnostic data T_i and the central points of the clusters.

According to the weighted values of all symptoms in P, the weighted symptom similarity ws are calculated.

If $ws \ge$ the minimum symptom similarity, then $T_j \in S_i$-cluster (13)

}

}

If the mining symptom is ab, the weighted value of a, b is respectively 2 and 1. $S = \{a,\ b,\ ab\}$, a, b and ab is respectively the central points of the cluster. If the diagnostic data T_1 is $\{acd,\ XY\}$ and the symptom is acd, the diseases XY.

If the similarity of the minimum symptoms is 50%, T_1 belongs to a-cluster. According to the above clustering steps, the diagnostic data which satisfies with the minimum symptom similarity to the S_i belongs to S_i-cluster.

The calculations are as follows: The ratio value in all diseases is equal to the number of diseases happened in the S_i-cluster/the number of the diagnostic data included in the S_i-cluster.

The above calculation is used as the basis on digging the tendency of the diseases from the symptom P. For example, the maximum ratio values in the S_i-cluster can be found out. The tendency of the diseases from the symptoms P is defined as follows: the diseases whose ratio values are the maximum can be found out in each S_i-cluster. The number of these diseases can be accumulated. The disease whose number is the maximum is the tendency of the diseases from the symptom P. As for the diseases which have the maximum numbers, they can be the most possible tendency of the diseases. Other diseases of the ratio values or other accumulated number of diseases can be adopted. Both of them can be used as the referent basis on knowing about the tendency of the diseases suffering from the symptom. In this way, the extension of suffering from the diseases can be increased.

The explanation of the examples

The example explains the process of digging the tendency of the diseases from the symptom with the clustering based on the patients' weighted symptoms. Table 1 present the diagnostic data D, including 6 diagnostic data, where

- $\{a, b, c, d, e\}$ is the set of the symptoms,
- $\{X, Y, Z\}$ is the set of the diseases and
- $\{T_1, T_2, T_3, T_4, T_5, T_6\}$ is the set of the diagnostic data.

If the mining patients' symptom is being, the symptom's weighted value is 1 and 2 and the minimum symptom similarity is 60%.

Table 1. The diagnostic data D

The number of the diagnostic data	**The symptoms**	**The diseases**
T_1	*abd*	*X*
T_2	*abce*	*Y*
T_3	*bce*	*XY*
T_4	*ab*	*YZ*
T_5	*ae*	*X*
T_6	*abe*	*XY*

At first, the subset of the patient's symptom be is $\{b, e, be\}$, b, e and be is the central points of the cluster respectively. After calculating the algorithm Clustering(*be*), the following cluster can be obtained:

b-cluster = $\{\varnothing\}$, (14)
e-cluster = $\{T_3, T_5, T_6\}$, (15)
be-cluster is $\{T_2, T_3, T_5, T_6\}$. (16)

The maximum diseases in all clusters are as follows:

b-cluster: $\varnothing$, (17)
e-cluster: disease X, (18)
be-cluster: disease X, Y. (19)

The number of all diseases is as follows:

The number of the disease X is 2, the number of the disease Y is 1 and the tendency of the disease from the patients' symptom is X.

The calculation without the symptom weighted can obtain that the be-cluster is $\{T_3, T_5\}$ and the tendency of the disease from the patients' symptom be is X or Y. Therefore, the symptom weighted mining by the medical profession can feature the tendency of the disease from the patients' symptom without the weight.

The establishment of the diagnostic system for the patients' diseases

The research designs and builds the diagnostic system for the patients' diseases and the Table 2 is the development platform of the system.

Table 2. The development platform of the system

The operating system	Windows XP Professional Edition
CPU	Intel Core i7 CPU 3GHz
The main memory	3 GB SDRAM
The designing language	ASP, VB Script, Java Script
The data base	Microsoft Access2007

The patients in a hospital each medical diagnostic data for example, diagnostic data from 2016/7/1 to 2016/7/7, a total of 6350 cases, as the system data sources, the diagnostic data in front of 6000 pen as the training data, and in the last 350 diagnostic data as verification data for mining computation. Fig. 1 is the original data of the diagnostic data, the original data in patients of each medical treatment for a record store, every diagnostic data includes medical treatment "type" and "symptoms" and "disease" data.

The symptom description and the names of the diseases in the diagnostic data must be encoded respectively, the name of the disease can use the ICD-10-CM code (The International Classification of Disease, 10th Revision, and Clinical Modification) to encode. Some more important words of the symptoms can be chosen out from the symptom descriptions and be encoded, and it can be respectively encoded with the S0001, S0002, S0003 and others. In Fig. 2, the disease code and the symptom code after being coded replace the name of the diseases and the symptom description in the original diagnostic data.

The research regards 6000 diagnostic data as the mining training data so that it can find out the tendency of the diseases from the symptom. The implementation of digging the training data is as follows: the digging symptoms can be input in the segments of the patients' symptoms in the Fig. 3 and the weighted value can be set, the minimum symptom similarity is 60%. The segment in the tendency of the diseases can show the mining results as shown in the Fig 4.

The research regards the remaining 355 diagnostic data as the digging verifying data, evaluating the efficiency of the digging results in the former training data. The paper adopts the digging to execute the picture from the remaining 355 diagnostic data. The symptoms appearing in these diagnostic data can be input and the tendency of the diseases from the symptom can be found out. Later, check whether the diseases in these diagnostic data are the tendency of the disease. If so, the definition of the diagnostic data can respond the tendency of the diseases. If no, the definition of the diagnostic data can not respond the tendency of the diseases.

The Fig. 5 can evaluate the number of the diagnostic data with the tendency of the diseases under the different number of the diagnostic data. The system can respond the tendency of the diseases under the high ratio value. If the medical professor sets the symptom weight, the responded number can be improved.

Code	Department	Symptom	Disease
2204	General surgery	Rt.inguinal mass for one week	K40.9Inguinal hernia
2205	General surgery	Dysuria for one day Abd.pain	N30.9Cystitis
2206	General surgery	Bil.mastalgia for noe week	D36.9Benign breast tumor
2207	General surgery	.HCC,Rt.lobe,S/P general weakness Cough	K27Peptic ulcer C97 Liver cancer
2208	General surgery	Gall bladder polyp was told for one week	K80.1Cholecystolithiasis
2209	General surgery	Rt.thyroid.ca.S/P Oral ulcer	193Thyroid carcinoma
2210	General surgery	Rt.breast mass for one week	D36.9Benign breast tumor
2211	General surgery	Ca.of D-colon with obstruction,S/P,ADJUV	C18.9Carcinoma of colon
2212	General surgery	Ac.app.S/P	K35.9 Acute appendicitis
2213	General surgery	Lt.breast nass for one week	C50.9 Benign breast tumor、I84.9Hemorrhoids
2214	General surgery	Fall stones with jaundicems/p Wd,infection	K80.1 Cholecystolithiasis
2215	General surgery	CBD stones with ob.jaundice,S/P PTCD T.	K80.1 Cholecystolithiasis
2216	General surgery	Anal painful sensation for one weeek	I84.9Hemorrhoids
2217	General surgery	Ac.app.S/P	K35.9Acute appendicitis
2218	General surgery	HAD STOMACH CA WITH LIVER METAS	C18.9Carcinoma of colon
2219	General surgery	Lt.breast mass for one week	D36.9 Benign breast tumor
2220	General surgery	Lt.breast mass for one week	D36.9 Benign breast tumor
2221	General surgery	Ca.of retum,S/P Rt.armnumbness Skin ra;	C20 Rectal cancer
2222	General surgery	Rt.buttock mass for one week,S/P	D10-D36 Benign tumors
2223	General surgery	Rt.breast mass for one week	D36.9 Benign breast tumor
2224	General surgery	Rt.breast mass for one week	D36.9Benign breast tumor
2225	General surgery	Rt.middle finger trauma for one month Ingn	T11.0Multiple wounds of the upper limb
2226	General surgery	Lt.breast ca.S./P	C50.9Mastocarcincma
2227	General surgery	BilIHD with ob.cholangitis,S/P PTCD	K80.1 Cholecystolithiasis
2228	General surgery	A skin defect with granulation tissue for one:	D36.9Benign breast tumor
2229	General surgery	Bil.MTNG,S/P	D34 Benign thyroid tumor
2230	General surgery	Ac.cholecystist,S/P PTGBD	K80.1Cholecystolithiasis

Fig. 1 Original diagnostic data

Code	Department	Recoding	Symptom code
194	physician	S0011,S0255	K73.9Hronic hepatitis
195	physician	S0011	I80.3Venous thrombosis、D64.9anaemia、M25.0-M25.9Other joint diseases
196	physician	S0057,S0011	L03.9 Foot cellulitis、C20rectal cancer、I10Hypertension
197	physician	S0034,S0011,S0142	D64.9Chronic hepatitis、K70.2Alcoholic cirrhosis、R51Headache
198	physician	S0034,S0011,S0142	L03.9Foot cellulitis、L08.0-L08.8Atopic dermatitis、K25Gastric ulcer
199	physician	S0218,S0092	I51.6Cardiovascular disease、I80.3Thrombophlebitis
200	physician	S0011	D64.9Chronic hepatitis、K70.2Alcoholic cirrhosis、
201	physician	S0057,S0070	G20Parkinson's disease、F51.0-F51.9Sleep disorder、J18.9Pneumonia、J40Bronchitis
202	physician	S0057,S0070	A41.9Septicemia、J18.9Pneumonia、J40Bronchitis
203	physician	S0473	M40.0-M40.5Spinal curvature、M06.9Rheumatoid arthritis、M86.9Chronic osteomyelitis
204	physician	S0013,S0042,S0109	J42Chronic bronchitis、E27.9Insufficiency of adrenocortical function、K29.0Acute gastritis
205	physician	S0008,S0023,S0075,S0296	K25Gastric ulcer and bleeding、M06.9Rheumatoid arthritis
206	physician	S0473,S0092,S0237	K92.29Acute gastroenteritis、R31Hematuria
207	physician	S0437	G62.93Peripheral neuritis、N18.9Chronic renal failure
208	physician	S000,S0437	G20Parkinson's disease、G40.9Epilepsy、F51.0-F51.9Sleep disorder
209	physician	S0008	L03.9Foot cellulitis、C55 Fibroid、I10Hypertension
210	physician	S0218,S0092	M00.9 Suppurative arthritis、L03.9 Foot cellulitis、M06.9Rheumatoid arthritis
211	physician	S0034,S0011,S0142	J39.9Abnormality of respiration、I10Hypertension
212	physician	S0011	J18.9Pneumonia、A41.9Septicemia
213	physician	S0011	A05.9Food poisoning, unindicated
214	physician	S0034,S0011,S0142	G40.9Epilepsy、G62.9Peripheral neuritis、I61.9 Intracerebral Hemorrhage
215	physician	S0057,S0070	N18.9Chronic renal failure、J11.8Influenza、K76.7Nephrotic Syndrome
217	physician	S0210	M86.9Acute osteomyelitis、F51.0-F51.9Sleep disorder、M00.9Suppurative arthritis
218	physician	S0008	J11.8Influenza
219	physician	S0473	I70.9Atherosclerosis、I10Hypertension

Fig. 2 Coded diagnostic data

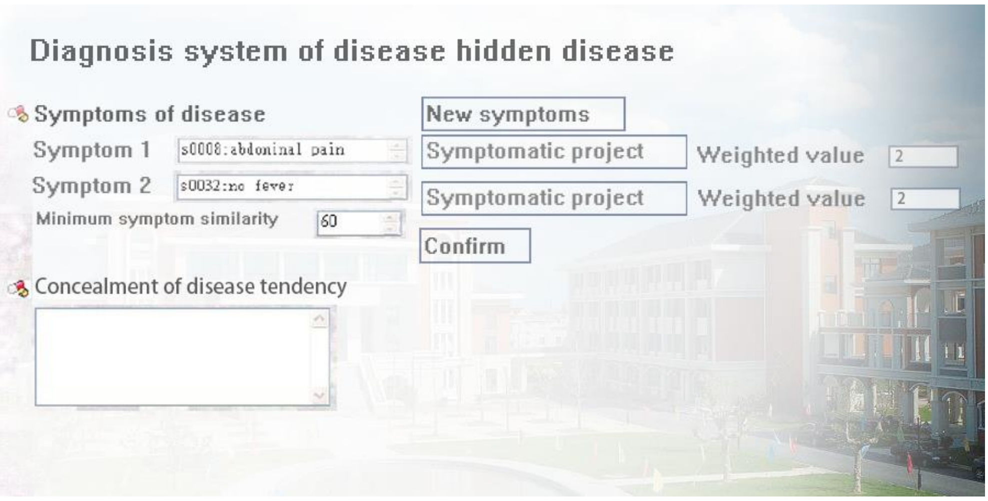


Fig. 3 Execution interface of data mining
disease symptom concealment disease tendency

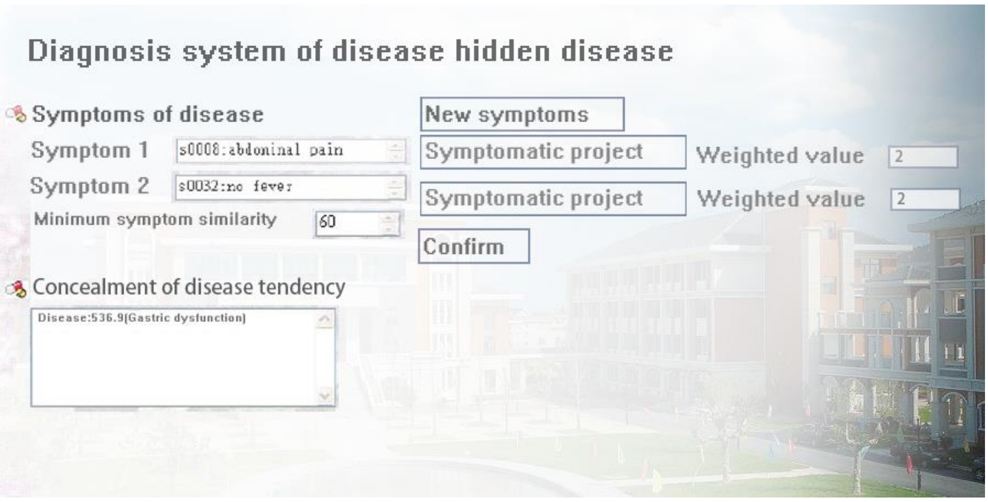


Fig. 4 Result interface of data mining
disease symptom concealment disease tendency

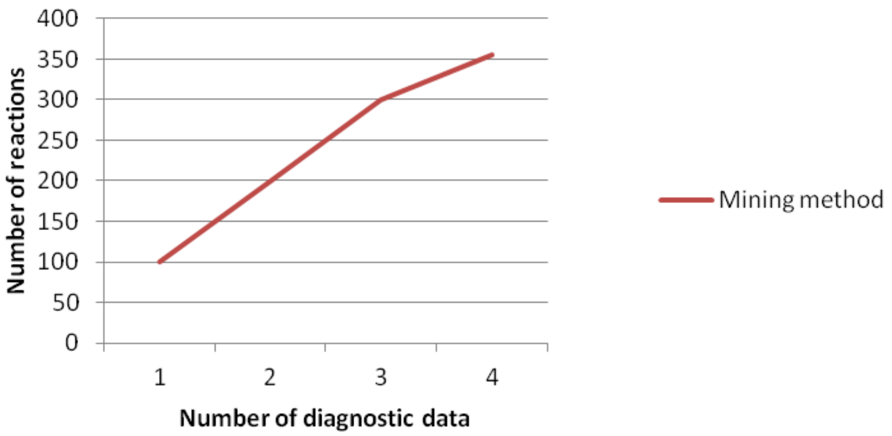


Fig. 5 The number of diagnostic data that is prone to disease progression

Conclusion and future research directions

The medical data can record the clinic departments, the symptom statements of the patients, the Chinese name of the diseases, its corresponding ICD-10-CM code, prescription drugs and other data when the patient sees the doctor. The clinic data hide the diagnostic capacity, experience and knowledge of the diseases. If they can be managed and applied, the people can check the symptom has the disease and some useful referent information can be offered. The paper regards the diagnostic data as the data mining source, a patient's symptom as the mining target and designs a clustering method. The mining results can receive the efficiency of treatment and can offer some useful referent information and efficiency. The research uses the PAM algorithm as the main reference and the modified methods. The following research directions are as follows:

(1) Discuss the difference and efficiency between the research and other clustering methods, such as *k*-means, CLARANS and others.
(2) Discuss the feasibility and effectiveness of the topic with the use of other data mining technologies, such as the association rules.
(3) Discuss the effectiveness of a clinic department.
(4) Improve the reliability and kindness of the operation in the system and promote the usefulness of the actual application.
(5) Improve the mining methods designed in the research and promote the executing efficiency of the mining.

Acknowledgements

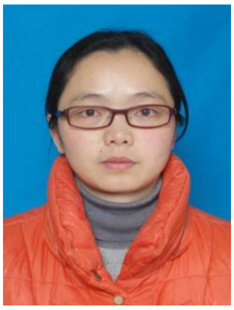
This work is supported by the Planned Science and Technology Project of Hunan Province of China (2013FJ3022).

References

1. Armano G. M., R. Farmani (2016). Multiobjective Clustering Analysis Using Particle Swarm Optimization Expert Systems with Applications, 55(C), 184-193.
2. Berry M. J. A., G. S. Linoff (1997). Data Mining Techniques: For Marketing, Sales and Customer Support, John Wiley & Sons.
3. Bhoi A. K. (2017). Classification and Clustering of Parkinson's and Healthy Control Gait Dynamics Using LDA and K-means, International Journal Bioautomation, 21(1), 19-30.
4. Forero D. A. (2008). Data Mining in Genomics, Retour Au Numéro, 28(1), 145-166.
5. Ji Y., H. Ying, J. Tran, P. Dews, R. M. Massanar (2016). Integrating Unified Medical Language System and Association Mining Techniques into Relevance Feedback for Biomedical Literature Search, BMC Bioinformatics, 17(9):264.
6. Kaufman L., P. J. Rousseeuw (1990). Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley & Sons.
7. Kim J., M. Han, Y. Lee, Y. Park (2016) Futuristic Data-driven Scenario Building: Incorporating Text Mining and Fuzzy Association Rule Mining into Fuzzy Cognitive Map, Expert Systems with Applications, 57, 311-323.
8. Li Q., P. Wang, W. Wang, H. Hu, Z. Li, J. Li (2014). An Efficient K-means Clustering Algorithm on MapReduce, In: Bhowmick S. S., C. E. Dyreson, C. S. Jensen, M. L. Lee, A. Muliantara, B. Thalheim (Eds), Database Systems for Advanced Applications. (DASFAA 2014), Lecture Notes in Computer Science, 8421, 357-371.
9. Mahapatra B., S. Patnaik (2016). Self Adaptive Intrusion Detection Technique Using Data Mining Conceptinan Ad-hoc Network, Procedia Computer Science, 92, 292-297.
10. Molina M. E., A. Perez, J. P. Valente (2016). Classification of Auditory Brainstem Responses through Symbolic Pattern Discovery, Artificial Intelligence in Medicine, 70, 12-30.

11. Regan D. K., R. I. Davidovitch, S. Konda, A. Manoli 3rd, P. Leucht, K. A. Egol (2017). Presence of Failed Fracture Implants in Association with Lower Extremity Long Bone Nonunion Does Not Portend Worse Outcome after Nonunion Repair, Journal of Orthopedic Trauma, 31(5), e143-e147.
12. Silva J. D. A., E. R. Hruschka, J. Gama (2017). An Evolutionary Algorithm for Clustering Data Streams with a Variable Number of Clusters, Expert Systems with Applications, 67, 228-238.
13. Vani B. C., M. Shimabukuro (2016). Visual Exploration and Analysis of Ionospheric Scintillation Monitoring Data: The ISMR Query Tool, Computers and Geosciences, 104, 125-134.
14. Ye T., S. G. Zhou, H. F. Wei (2016). Professor XIA Xiang's Formula Rules for Treating Cardiovascular and Cerebrovascular Diseases with Qi-supplementing and Blood-activating Method, Chinese Journal of Experimental Traditional Medical Formulae, 18, 177-183.
15. Zou G. (2016). Ant Colony Clustering Algorithm and Improved Markov Random Fusion Algorithm in Image Segmentation of Brain Images, International Journal Bioautomation, 20(4), 505-514.

Yingying Peng, M.Sc.
E-mail: 31581677@qq.com



Yingying Peng was born in Hunan Province, China, in 1982. She received her B.Sc. degree of Engineering in Computer Science and Technology from Hunan Normal University in 2004 and M.Sc. degree of Engineering from Hunan University in 2010. Her research interests include networks security and digital-watermarking.

Gang Yi, M.Sc.
E-mail: 6636244@qq.com

Gang Yi was born in Hunan Province, China, in 1976. He received her B.Sc. degree of Engineering in Computer Science and Technology from Hunan Normal University in 2000 and M.Sc. degree of Engineering from Hunan University in 2011. His research interests include internet of things technology and network engineering.