Clinical Databases used for Selecting Drugs based on Pathology Consideration and Data Mining Methods

S.Mo, L.Y.Roger

Abstract — **Since my previous research " clinical databases used in hospital " finished, I have thinking about whether a clinical database can be more exactly when used to select drugs for patients. Two things I aware that should be helpful on proving my clinical database. Pathology, one factor should be considered when selecting drugs, is a new but very important Attribute of the database. The predicting of diagnosing is another aspect which happens right before using the clinical database.**

Index Terms—**clinical databases, pathology, data mining, classification**

I. Introduction

The clinical database I do research on is for doctors using to select drugs for patients. According to the diseases a patient has, the age range of the patient, gender of the patients, and other factors (eg: pregnant, side effect, use conflict, and etc), this clinical database helps doctors to search a list of drugs used for a specific disease.

However, in my previous research, I did not consider the pathology may effect the result of a searching. It is a very important factor when making decisions on using drugs. Pathology is "the form of medical science and specialty practice concerned with all aspects of disease, but with special reference to the essential nature, causes, and development of abnormal conditions, as well as the structural and functional changes that result from the disease processes"[3]. It means, different pathology may lead to the same disease or same symptoms. It may because of something different inside, the result of using drugs become different at all. In order to complete the clinical database, pathology is taken as a factor when selecting drugs.

Another aspect I take into consider is about diagnose. How does a doctor diagnose more convincing by data? This step is right before the searching on the clinical database, but is also very helpful with the whole process of selecting drugs for a patient.

In this paper, the researchers will show how they take pathology as a new attribute in the clinical database, and how they use data mining methods on diagnosing. Both of those are great improvement of the process of selecting drugs by a clinical database.

Manuscript received July 10, 2014; revised August 16, 2014.

S. Mo is with Central Michigan University, Michigan, U.S.A(phone:989 430 5346 e-mail: sun2m@cmich.com 405783097@qq.com).

L.Y.Roger is with Central Michigan University, Michigan, U.S.A(e-mail:lee1ry@cmich.edu).

II. Methods

A. Add the pathology (reasons causing diseases) as an attribute into the clinical database

a. Why we add the pathology as an attribute into the clinical database

In order to select a more fittable drug for a patient, pathology is one of the most important elements should be considered. In addition to the patient's age, the death rate of the diseases which the patient get, and all the drugs for the particular diseases use rate, pathology also has effect on selecting drugs.

For example, "hypertension (high blood pressure) due to an identifiable cause such as kidney diseases or endocrine diseases"[2].That means, even if patients are diagnosed with hypertension, they may have different diseases. While, some drugs are for the patients with hypertension caused by kidney diseases, some drugs are for the patients with hypertension caused by endocrine diseases.

As the discussion above, there will be a different result of the drugs recommended by the database after adding the pathology.

b. Which table the pathology should be added in

I. Patient history
II. Disease information
III. Drug information

After adding the pathology, the input is still patient's symptom in the system, but the output will be the result of when both the disease and the pathology is match in patient history table, disease information table, and drug information table.

B. Use predicting and classification method of data mining in the database

a. Predicting method

"Predict the risk of a specific disease is a very common application used in medical field"[1],[4],[6].

Decision tree shows the risk of a specific disease telling us the relationship between the disease and the factors may affect the result of diagnose. Predicting methods used in medical field makes the result more visible, and it not only predict the risk of a disease, it also provides convenient for doctors' diagnosing.

The risk of a specific disease is given in the original database. Decision tree gives more exact value on the risk of patients having the same diseases but with different

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

symptoms. From the decision tree, each path is going to be a possibility of the risk of the same specific diseases. It means the table of the risk would be expend by those paths.

b. Classification method

Classification provides a more reliable method to diagnose the disease just before selecting drugs, which increases the accuracy of the result of drugs on an indirect way. It provides a more reliable result of the specific disease of a patients. With such a diagnosis, the result of selecting drugs can be more reliable.

Classification method is used before selecting drugs. We use k-NN (one of the classification methods) to diagnosis. "Instance based learning is use specific instances to perform classification tasks, and the nearest neighbor (NN) algorithm is one of the best known classification algorithms"[5]. For example, k-NN method takes the k nearest training examples and determines which class the term belongs to.

There is a small model of k-NN:

TABLE I
THE SYMPTOM OF PEOPLE WHO GOT FEVER OR NOT

cough	temperatures	Running nose
Y	high	Y
Y	high	N
Y	normal	Y
Y	normal	N
N	high	Y
N	high	N
N	normal	Y
N	normal	N

There is a possibility of getting a fever of each row of the symptom. When see a new patient, find the nearest k items with his/her symptom, and get the final possibility of getting a fever.

The process of classification is taking out the history of all the patients having the same disease and compare all the items with the patients' symptoms. And then, we know every possibility of each permutation and combination of symptoms diagnosed with the disease finally.

III. Result

A. new tables after adding pathology

We add pathology as an attribute into three tables: patients' history table, disease table, and prescription table. The schema of these tables are shown below:

TABLE II
THE SCHEMA OF PATIENTS' HISTORY TABLE AFTER ADDING PATHOLOGY

attribute	type	constraint
Patient ID	char	primary key
Patient name	char	not null
Patient age	int	not null
Patient sex	char	not null
symptom	char	not null
pathology	char	not null

TABLE III
THE SCHEMA OF DISEASE TABLE AFTER ADDING PATHOLOGY

attribute	type	constraint
disease	char	primary key
pathology	char	

TABLE IV
THE SCHEMA OF PRESCRIPTION TABLE AFTER ADDING PATHOLOGY

attribute	type	constraint
name	char	primary key
function	string	not null
warning	string	
side effect	sring	
disease	char	foreign key
pathology	char	not null

B. The result of searching drugs after adding the pathology

After adding the pathology, the drug list of one searching will not cover drugs whose pathology are not match with the patient have.

Result before and after adding the pathology:(The shadow area means the result of a search.)

The shadow part in fig2 plus the part labeled 1 equals to the

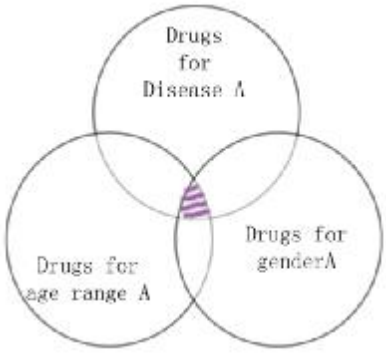

Fig. 1. Result before adding the pathology

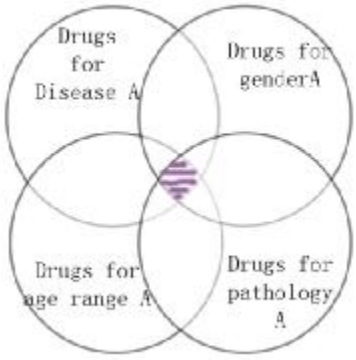

Fig. 2. Result after adding the pathology

shadow part of fig1. The part labeled 1 is what the new database take into consider when selecting drugs for patients.

C. Adding predicting and classification part in the clinical database

In this part, input is: symptom1, symptom2, symptom3, symptom4, symptom5,etc. And then, the system will search all the patients who has at least one of the symptoms on their history. Finding the closest patients' history of the current case, which helps the doctor the diagnose what disease the patient has the most possibility to have.

IV. Conclusion and Limitation

After adding pathology on the process of selecting drugs, and adding a new part of diagnosing before selecting drugs,

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

the clinical database for selecting drugs becomes more completely on both result and the premise.

However, there are still limitations of the research. Is the result of diagnose perfectly match with the disease a patient really has? If the result of diagnose is not correct, let alone the result of drug list.

V. Future works

Since selecting drugs, diagnose can get answers by databases, is there any possibilities to let a clinical database become an database who has learning ability with those history of patients. If so, the process of diagnose, selecting drugs, picking up drugs, and all the process about clinic can be done by computers, at the same time decrease the amount of people works.

References

[1] Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: a comparison of three data mining methods. Artificial intelligence in medicine, 34(2), 113-127.
[2] The introduction of hypertension. Available: http://en.wikipedia.org/wiki/Hypertension
[3] The definition of the keyword "pathology". Available: http://www.drugs.com/dict/pathology.html
[4] Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. Journal of Healthcare Information Management—Vol, 19(2), 65.
[5] Lavrač, N. (1999). Selected techniques for data mining in medicine. Artificial intelligence in medicine, 16(1), 3-23.
[6] Mullins, I. M., Siadaty, M. S., Lyman, J., Scully, K., Garrett, C. T., Greg Miller, W., ... & Knaus, W. A. (2006). Data mining and clinical data repositories: Insights from a 667,000 patient data set. Computers in biology and medicine,36(12), 1351-1377.

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)