Enhancement of Diagnosis System for Tuberculosis Using Case-based Reasoning

Dim Lam Cing , Khin Lay Thwin
Computer University (Kalay)
ciicing@googlemail.com, khinlaythwin@gmail.com

Abstract

Case-based Reasoning (CBR) is a recent approach to problem solving and learning that has got a lot of attention over the last few years. CBR is an Artificial Intelligence method based on a plausible cognitive model of human reasoning. People take a great interest in computer and then computer-based methods are increasingly used to improve the quality of the medical services. CBR is considered established method for building medical diagnosis systems. Traditional expert systems model human problem solving as a deductive process. They construct a solution by applying general rules to the description of a problem. It becomes apparent that human experts rely heavily on memory of past cases when solving problems. In this paper, a system is presented using CBR and decision tree algorithm for medical diagnosis. It is implemented for the efficient diagnosis of tuberculosis. This system is proposed as a development tool. The main feature of the proposed system is to provide a simple and integrated tool for designing diagnostic applications. This system also provides for helping the Tuberculosis disease and controlling the treatments for patients. So that, every user can do the diagnosis as a physician.
Keywords: Case-Based Reasoning, Decision Tree Algorithm

1. Introduction

Today's CBR-based applications are used very commonly in many different fields. Case-based Reasoning (CBR) has been used to create numerous applications in a wide range of domains including financial analysis, medical diagnosis, design, classification of objects, Help desk and decision support system. The basic idea of CBR is to solve new problems, which have already been solved in the past. CBR systems are simply defined as a collection of cases that are representing of typical situations with past experience and possible solutions. This system is an approach for developing case-based medical decision support systems based on the technology of case-based reasoning (CBR) [3].

Many classification and prediction methods are proposed by researchers in machine learning, expert systems, statistics and neurobiology. Some other approaches of classification, such as K-nearest neighbor classifiers, Case-based reasoning, genetic algorithms, and rough set and fuzzy logic techniques are introduced [3].

A clinician, he must evolve a decision making pathway or serial steps of deductive thoughts which will give him a correct diagnosis in most instances. The main goals are to reduce the time required to come to a decision, to share the experiences among different sites, to support human for diagnosis and treatment, to reduce the pressure of work on one or two experts, to complete human reasoning and problem solving, to provide reasoning techniques such that easing knowledge acquisition and increasing problem solving efficiency. Therefore it becomes important for doctors to be able to take a good history and do a proper physical examination. Not only that, but pays proper emphasis to deserving clues obtained by history and physical examination. This would mean the identification and correct interpretation of symptoms and signs [3].

This paper describe an approach for developing knowledge-based medical decision support systems based on the rather new technology of Case-based reasoning (CBR). CBR is an approach for solving problems based on solutions of similar past cases. A case consists at least of a problem description (e.g., symptoms) and a solution (e.g., a diagnosis). Cases are stored in a database of cases called case base. To solve an actual problem is used to retrieve similar cases from the case base. The solutions of these found similar cases are used as starting points for solving the actual problem at hand.

The rest of the paper is organized as in section 2 described related works. In section 3 Background Theory, section 4 described about the proposed system, section 5 described classify accuracy, section 6 about experimental result and section 7 described the conclusion.

2. Related works

The research in medical CBR is concentrated to Europe and US, as with CBR in general. The medical domain of CBR is generally focused on producing systems for specific tasks, such as diagnosing a specific symptom. These systems are rarely in every-day-use and they are seldom commercially exploited [5]. Limitations of CBR in medical systems are addressed by Schmidt and Gierl [6] where they approach the difficult task of case adaptation. Atzmulluer et al [4] approach the issue of handling multiple faults, or diagnoses, in a medical system. Their solution is to decompose the problem part of a case into several smaller ones and find solutions to them instead. When the system comes up with solutions to the smaller problems they combine all solutions to solve the original problem. Some CBR issues are especially interesting for the medical domain.

3. Background Theory

3.1 Case-Based Reasoning (CBR)

CBR is to solve a new problem by remembering a previous similar situation and by reusing information and knowledge of that situation. CBR is a cyclic and integrated process of solving a problem and learning from this experience. CBR method can be divided into four steps model, such as retrieve, reuse, revise and retain. The cycle is illustrated in Figure 1[1].

The retrieve part tries to find the most similar case (or cases) comparing the case to the library of past cases. In the reuse case, the retrieve cases (the information and knowledge) are sent to the reuse part to solve the problem. Reuse is to find the difference between the new and the old case and find what part of the old case that can be used in the new case. In simple classification, the difference between the old case and new case are abstracted away, and the solution is simply copied from the old case. And then either the solution itself can be reused, or the past method that produced the solution can be used. It can be adapted. The CBR system will then suggest to evaluate the solution for the new problem. The solved case is then sent to a revision part, on the condition the evaluator has not objected to the solution. The solved case is then verified for correctness by the revision. If the solution generated by the last phase is not correct, the system can learn from its failures. The solution proposed by the reuse phase tries in the real environment and evaluate it. If the solution evaluate badly, that find the errors or flows of the solution and generate explanation them. The last step is in corporate what is useful to learn from the problem solving experience into the existing knowledge. If the solution is valid, it will be presented as a confirmed solution to the problem. The new case (the final solution) as part of a new case can be added to the case library if the case contain new experience not previously capture in the case library. This process takes place in the retain step [1].

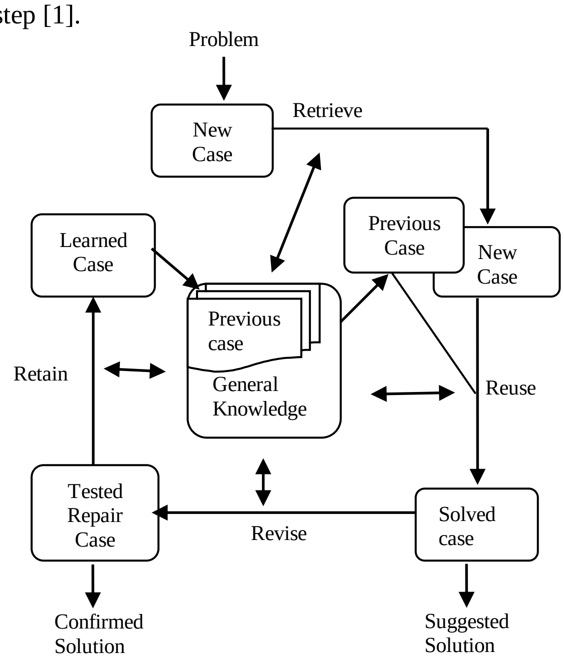

Figure 1. The Life Cycle of Case-based Reasoning

3.2 Decision Tree Algorithm

A decision tree is a flow-chart-like tree structure that employs a top down. Each internal node denotes a test on an attribute, each branch represents an outcome of the test and each leaf node represents class or class distribution. To select the test attribute at each node in the tree, the information gain measure is used. The decision tree is built using the attribute selection measure equation and decision tree algorithm [2].

Algorithm

Generate a decision tree for training data.

Input: The training samples, samples, represented by discrete-valued attributes; the set of candidate attributes, attribute-list.
Output: A decision tree
Method:

1) *Create a node N;*
2) *if samples are all of the same class, C then*
3) *return N as a leaf node labeled with the class C;*
4) *if attribute-list is empty then*
5) *return N as a leaf node labeled with the most common class in samples;*
6) *select test-attribute, the attribute among attribute-list with the highest information gain;*
7) *label node N with test-attribute;*
8) *for each known value a_i of test-attribute*
9) *grow a branch from node N for the condition test-attribute = a_i;*
10) *let s_i be the set of samples in samples for which test-attribute = a_i;*
11) *if s_i is empty then*
12) *attach a leaf labeled with the most common class in samples;*
13) *else attach the node returned by Generate-decision-tree (s_i, attribute-list-test-attribute)[2];*

Figure 2. Decision Tree Algorithm

4. Proposed System

4.1 System Design

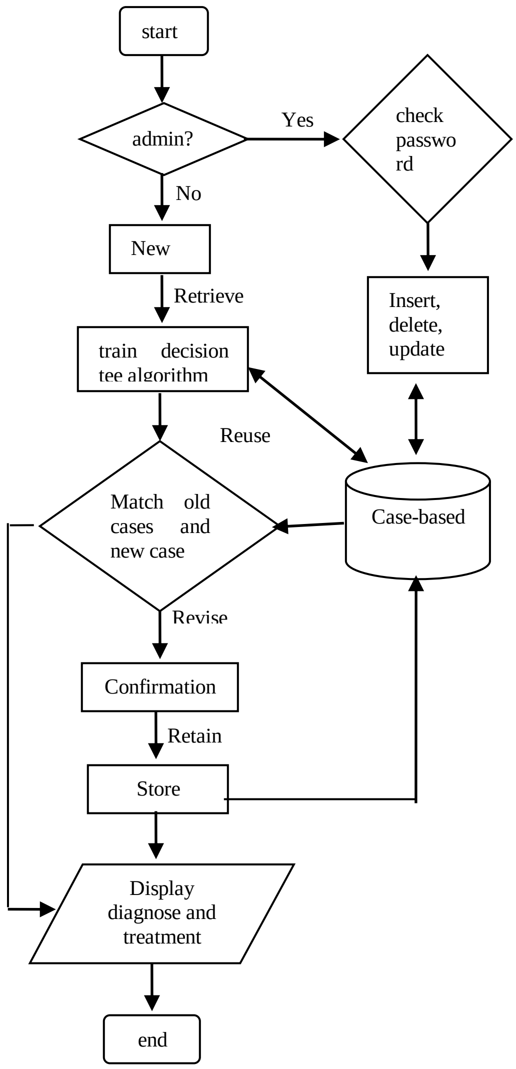

Figure 3. System Design

In this system, user can choose administrator (or physician) / user (or patient). If physician can insert the new case about TB to the database by updating the existing data and deleting useless data. If the user is patient, the system will display the common symptoms of Tuberculosis (TB) and diagnosis to him.

The symptoms and diagnosis are shown as attribute and, its related symbols are shown in Table1.

Table 1. Notation Table

No	Attribute	Symbol
1.	Age(<=12 or >12)	Age
2.	Cough over (3) weeks	C
3.	Hemoptysis	H
4.	Loss Appetite	LA
5.	Weight Loss	WL
6.	Afternoon Fever and Sweating	FS
7.	Chest Pain	CP
8.	Expectoration	E
9.	Lymphadenopathy	L
10.	Malnutrition	M
11.	Breathlessness	B
12.	New Smear Positive (Category 1)	NSP
13.	New Smear Negative with extensive chest X-ray abnormalities (Category 1)	NSNEX
14.	Previously Treated smear-positive pulmonary TB Relapse (Category 2)	PTR
15.	Previously Treated smear-positive pulmonary TB who Failure treatment in Category 1 and 3 (Category 2)	PTF
16.	Previously Treated smear-positive pulmonary TB who return after Default (Category 2)	PTD
17.	Smear Negative, TB Less Severe than Category 1 (Category 3)	SNLS
18.	Chronic and Multidrug (Smear positive after supervised retreatment) (Category 4)	CM

If the user wants to know the diagnosis for Tuberculosis (TB), he selects "case-based reasoning" from the system menu and enters the symptoms in new case problem as shown in Figure 4.

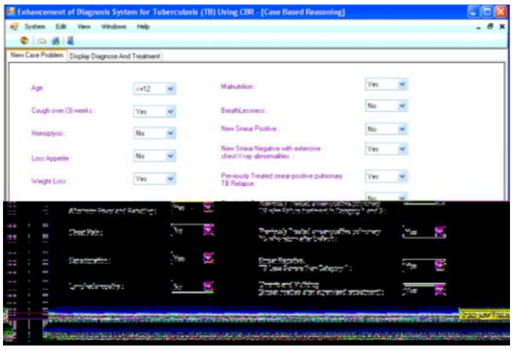

Figure 4. New Case Problem

After that, "Input New Problem" button can be clicked. It do compare the new case with old case. The system retrieves the most similar case to that problem from database using decision tree algorithm. If the case is similar, the system will reuse the existing solution in the case-based and then display the diagnosis and treatment. If the case is not similar with the old case, the system will save the symptoms and constructs decision tree and then display the diagnosis and treatments.

The case study for notation table is described in Table 2.

Table 2. Case Study

pid	Age	C	H	LA	WL	FS	CP	E	L	M	B	NSP	NSNEX	PTR	PTF	PTD	SNLS	CM	Class
1.	<= 12	Y	N	N	N	Y	N	N	Y	Y	N	N	Y	N	N	N	N	N	TB0
2.	<= 12	Y	N	Y	Y	N	N	N	Y	Y	N	N	N	N	N	N	Y	N	TB0
3.	<= 12	N	N	N	N	Y	N	N	Y	N	Y	N	Y	N	N	N	Y	N	TB0
4.	<= 12	Y	N	Y	Y	Y	N	Y	N	Y	N	N	N	N	N	N	N	N	TB0
5.	<= 12	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N	N	N	N	N	N	TB0
6.	>12	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	Y	TB4
7.	>12	N	N	Y	Y	Y	N	N	N	Y	N	N	N	N	Y	N	Y	N	TB3
8.	>12	Y	Y	N	N	Y	N	Y	N	N	Y	N	N	Y	N	N	N	N	TB2
9.	>12	Y	N	N	Y	N	Y	Y	Y	N	N	N	Y	N	N	N	N	N	TB1
10.	>12	N	N	Y	Y	N	Y	N	N	Y	Y	Y	N	N	N	N	N	N	TB1
11.	>12	Y	N	Y	Y	Y	N	Y	N	Y	Y	N	N	N	N	Y	N	N	TB2
12.	>12	Y	N	N	N	Y	N	N	Y	N	N	Y	N	N	N	N	N	N	TB1
13.	>12	Y	N	Y	Y	Y	N	Y	Y	Y	Y	N	N	N	N	N	N	Y	TB4
14.	>12	Y	Y	N	N	Y	Y	Y	Y	N	N	N	N	N	N	Y	N	N	TB2

There are five main classes in this case study: that are TB0, TB1, TB2, TB3 and TB4.

Children under 12 suffering TB are defined as TB0. Adults over 12 suffering TB are also classified according to medical check. Patients with new smear positive or new smear negative with extensive chest X-ray abnormalities are defined as TB1. Patients previously treated smear positive pulmonary TB who relapse or failure treatment or return after default are defined as TB2. Patients with smear negative and TB less severe chest X-ray abnormalities are defined as TB3 and smear positive after supervised retreatment are defined as TB4.

Decision Tree according to case study is described in Figure 5.

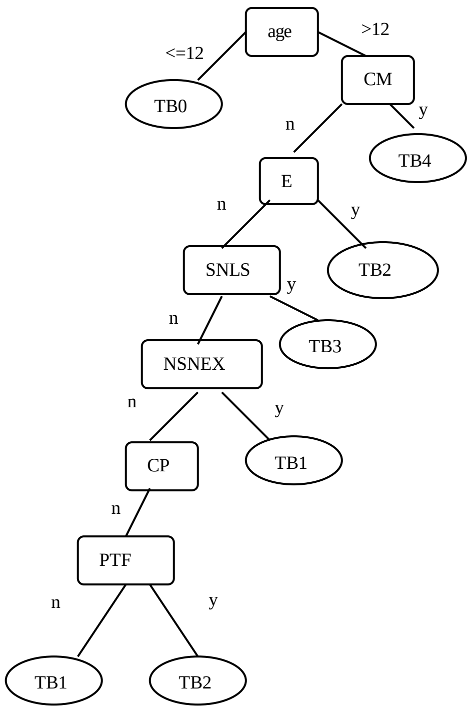

Figure 5. Decision Tree

5. Classifier Accuracy

"k-fold" cross-validation is a generalization partition data into disjoint validation subsets of size n/k train, validate, and average over the k partitions. It is approximately k times computationally more expensive than just fitting a model to all of the data. This is called k-fold cross-validation. Often the subsets are stratified before the cross-validation is performed. The error estimates are averaged to yield an overall error estimate.

Sensitivity is the probability of a positive test among patients with disease. Specificity is the probability of a negative test among patients without disease.

$$sensitivity = \frac{t\_pos}{pos}$$

$$specificity = \frac{t\_neg}{neg}$$

$$accuracy = sensitivity\frac{pos}{(pos + neg)} + specificity\frac{neg}{(pos + neg)}$$

6. Experimental Results

To determine the Tuberculosis (TB) Diagnosis Systems performance, the system was tested with different amount of training data. The system's classifier accuracy increased with the number of training data. The result of our test is described with the following table.

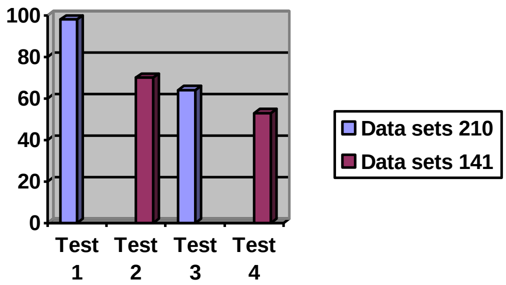

Figure 6. Comparison of accuracy

In Figure 6, Test 1 and Test 3 use the number of classes which are approximately equal. In Test 2 and Test 4, they use the number of classes almost which are TB1.

By comparison with these four results, the accuracy is increased depending on the data sets and the uses of class. In Test 1, the accuracy will reach almost 100%.

7. Conclusion

Case-based reasoning seems to be a suitable technique for medical knowledge based systems. The development of Tuberculosis (TB) case-based system will directly benefit to the users and physicians. Therefore, this medical diagnosis system will play an increasing important role in health care. By using this system, the physicians can check easily the patients' detailed Tuberculosis (TB) information and can easily know which kind of TB diagnosis and how to give the treatment for this patient. This system can diagnose and display the treatment for pulmonary TB. Furthermore, this system can be classified with other TB.

8. References

[1] A.Aamodt, E. Plaza (1994) “Case-based Reasoning: Foundational Issues, Methodological Variations, and System” Approaches AI Communications. IOS Press, Vol. 7: 1, pp.39-59.

[2] Jiawei Han, Micheline Kamber “ Data Mining Concepts and Techniques” Simon Fraser University

[3] Khin Thidar Lynn “Case-Based Reasoning for Evaluation of Symptom in Bronchogenic Carcinoma Patients” University of Computer Studies, Yangon Proceedings of The Third International Conference on Computer Application 2005 Yangon, Myanmar

[4] M.Atzmueller, W. Shi, J.Baumeister, F. Puppe, and J.A. Barnden, “Case-based approaches for diagnosing multiple disorders”, in Processings of the 17^{th} International FLAIRS Conference, FloridaArtificial Intelligence Research Society , May 2004.

[5] N.Markus and M. Sollenborn. “Advancements and trends in medical case-based reasoning: An overview of systems and system development”, in Proceedings of the 17^{th} International FLAIRS Conference, Florida Artificial Intelligence Research Society , May 2004

[6] R.Schmidt and L.Gierl. “Case-based reasoning for medical knowledge-based systems”, XVI International Congress of the European Federation for Medical Informatics, 2000.