Decision Support System Using CBR for Lung Diseases

Hla Chan Phyu, Khin Sandar
University of Computer (Mawlamyine)
lmn.lay @gmail.com

Abstract

Expert Systems imitate the reasoning process of experts for solving specific problems and employ human knowledge captured in a case based memory. The principal method use in the memory is case-based reasoning method which can provide the solving new problem by adapting previous solution to similar problems. In this system, CBR's cyclical process is used to support enhancing a process's performance of an expert. This method retrieves the appropriate cases from a larger set of cases. If the similar between a new case and the retrieved case are very high, the previous solution to that case is returned to users. But if the cases are not exactly equal, the system gives the possible cases using nearest neighbor retrieval method. This system is tested on Postoperative Patient data.In this paper we consider the patient who suffers lung and other symptoms which may or may not be serious.The system can diagnose four types of lung disease.

1. Introduction

People has been accumulating knowledge from thousands of years stored either in books or held by living experts in their brain. The study of how to reuse valuable knowledge to shorten the cycle of problem solving has led to many different approaches. One of these is the research of knowledge-based systems.

CBR is an Artificial Intelligence (AI) methodology that provides the foundations for technology of intelligent systems [7].The rapidly growing medical knowledge and new patient cases make the diagnosis process even more difficult. Updating the medical knowledge incrementally in a traditional medical diagnosis system to cope with this is not that easy. It can be made easier and reliable if supplied with a system that contains and provide recommendation from the past diagnosis cases of different morbidity. So clinician can benefit a lot from these prior cases. This implies that The CBR approach is appropriate to the problem [3]. Certainly, one of the intuitively attractive features of CBR in medicine is that the concepts of patient and disease lend itself naturally to a case representation.

In this paper we consider the patient who suffers lung and other symptoms which may or may not be serious. The system can diagnose four types of lung diseases: Pneumonia, Tuberculosis, Lung Abscess, and Simple Pneumoconiosis. To diagnose these diseases, the patient should be asked to describe current symptoms .This information can help the physician to identify the cause of disease.

This paper is organized as follows: In section 2, we present the related work. Section 3 describes the motivation. In section 4, we express the background theory. Section 5 presents the proposed system architecture. In section 6, we describe the proposed system for CBR in diagnosis of lung diseases. In section 7, we express the experimental results of our system. Finally, we conclude the paper in section 8.

2. Related Work

In medicine, CBR has mainly been applied for diagnostic and partly for therapeutic tasks. FM-Ultranet[2] is a medical CBR project implemented with CBR works. It detects malformations and abnormalities of foetus through ultrasonographical examination.

Jaulent et.al, [5] is diagnosing histopathology in the breast cancer domain. Their system uses cases that are derived from written medical reports. The features are also weighted for importance. Cases are compared for structural, surface and feature similarity.

Salem A.B.M et. al, [9] presented the CBR-based expert system prototype for diagnosis of cancer diseases. The system aid the young doctor to check their diagnosis and it provides recommendation for controlling pain and providing symptom relief in advanced cancer.

3. Motivation

CBR has been investigated in improving human decision making and has received much attention in developing knowledge-based system in medicine[8]. In CBR, cases represent specific knowledge of the domain. New cases can be inserted into a knowledgebase without making changes to the preexisting knowledge. Incremental learning comes natural to the case-based reasoning. The more cases are available the better the domain knowledge will be presented. Therefore, the accuracy of a case-based system can be enhanced throughout its operation, as new cases become available. So, we use the advantages of CBR method to build the system that can support the decision making in medical diagnosis.

4. Background Theory

Case-based Reasoning is a problem solving paradigm that in many respects is fundamentally different from other major AI approaches. It is based on two observations of real world problems. The first is that similar problems tend to have similar solutions. The second is that the types of problems encountered tend to reoccur over time. When the two of this observation hold, it is worthwhile to remember and reuse prior cases.

4.1. Case-based Reasoning Cycle

The processes involved in CBR have been described as a cyclical process.

- Retrieve the most similar case(s).
- Reuse the information and knowledge in that case to solve the problem.
- Revise the proposed solution.
- Retain the parts of this experience likely to be useful for future problem solving [1].

The following figure shows the case-based reasoning cyclical process.

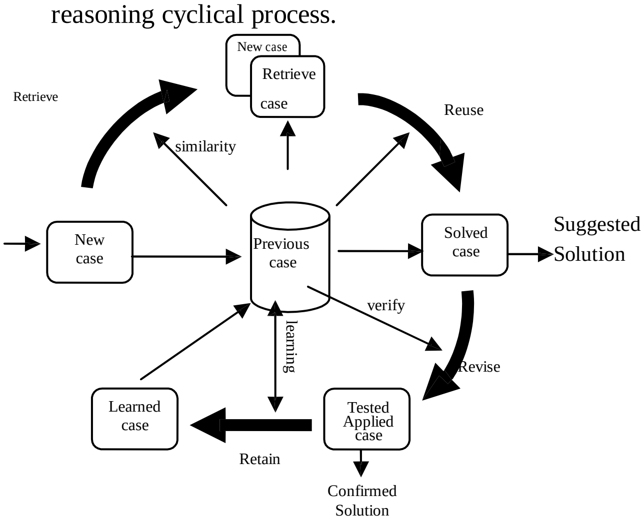

Figure1. Case-based reasoning cyclical process

An initial description of a problem defines a new case. The new case is used to RETRIEVE a case from the collection of previous cases. The retrieved case is combined with the new case through REUSE into a solved case. Through the REVISE process the solution is tested for success and repaired for future reuse and the casebase is updated by a new learned case, or by modification of some existing cases.

5. Proposed System Architecture

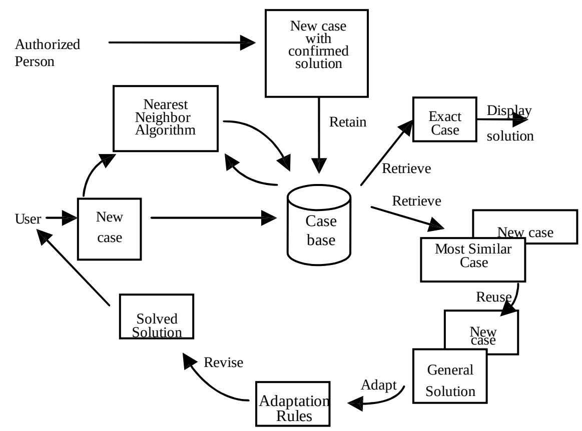

Figure 2. The proposed system architecture

The above figure shows the architecture of the proposed system. The system stores the patients' information in the form of index features in casebase. The system can be used by two types of users depending on their roles. For authorized user to maintain the casebase, he/she can saved new case with their corresponding confirmed solution into the casebase. For users, he/she can use the system to support the diagnosis of their input case by entering the case features. To give the solution of the new case, the system uses the nearest neighbor algorithm for matching the new case with the existing cases in the casebase.

In this phase, the system can retrieve either the exact match with the new case or the nearest match which has the greatest similarity with the new case. If the exact match found, the system gives the exact case solution to the user directly. Otherwise, the system reuses the solution of the most similar case as the general solution. Then it revises that solution by using adaptation rules and gives the solved solution to the user.

6. Case-based Reasoning in Diagnosis of Lung Diseases

Case-based Reasoning, a method of AI, has been used in this diagnostic system. As in all CBR based system, the system attempts to produce a solution to new problem by making use of 4R's: Retrieve, Reuse, Revise and Retain.

The diseases that the system used for diagnosis are Pneumonia, Tuberculosis, Lung Abscess, and Simple Pneumoconiosis. The system uses the signs and symptoms of each disease as the features of the case to store in the casebase. The proposed system is implemented as the following steps.

6.1. Defining Case Attributes

In a CBR problem solving cycle, the case attributes (along with its values) allow to evaluate the similarity between cases in order to retrieve appropriate information. The selection of the attributes for a Case-based Decision Support System is a domain dependent activity. Attributes must be predictive in a useful manner, influencing the outcome of the process and describing the circumstances in which a case is expected to be retrieved in the future. For practical applications attributes should be chosen using expert criteria.

In this system, each case contains 15 attributes (symptoms) which can be used safely for a good diagnosis of our chosen lung diseases. Table 1 shows the index-features of cases which can be used for case retrieval.

Table 1. Index-feature table

Features	Data Type
Feature	Boolean
Cough	Boolean
s breath	Boolean
chest pain	Boolean
Fatigue	Boolean
Fever	Boolean
Malaise	Boolean
weight loss	Boolean
Tachycardia	Boolean
Cyanosis	Boolean
night sweat	Boolean
anemia	Boolean
Chills	Boolean
chestdiscomf	Boolean
restless sleep	Boolean
palenail	Boolean

Table 2 shows the unindex-feature of cases which can be used for keeping a record of the patient.

Table 2. Unindex-feature table

Features	Data type
Name	String
Gender	character
Age	Integer
Address	String

6.2. Assigning Importance Values to Case Features

Features weights for most problem domains are context dependent. The weight assigned to each feature of the case tells how much attention to pay to matches and mismatches in the field when computing the distance measures of a case. One way to assign importance values is to have a human expert assign them as the case library is being built. The system uses this way to assign the importance values to different features in the casebase.

6.3. Retrieval using Nearest-neighbor Technique

The final goal of a CBR system is to find the case which has the maximum similarity to the input case. This system is intended to support the diagnosis for a new patient case, given all the patient's report attributes (symptoms or features). The features of the input case are assigned as indices characterizing the case. These indices are used to retrieve a similar past case(s) from case memory. The system uses the nearest-neighbor algorithm that finds the closest matches of the cases already stored in the database to the new case. The key thing in nearest-neighbor algorithm is the calculation of an attribute's comparison value for a feature between the previously stored cases and the input case. The pseudo code of this algorithm [6] can be written as follow:

For each feature in the input case:

- Find the corresponding feature in the stored case.
- Compare the two values to each other and compute the degree of the match.
- Multiply by a coefficient representing the importance of the feature to the match.
- Add the results to derive an average match score.
- This number represents the degree of match of

the old case to the input.

- A case can be chosen by choosing the item with the largest score.

A typical algorithm for calculating nearest neighbor matching is:

$$\text{Similarity}(T,S) = \frac{\sum_{i=1}^{n} f(T_i,S_i) \times w_i}{\sum_{i=1}^{n} w_i} \quad (1)$$

Where,
T is the target case.
S is the source case.
n is the number of attributes in each case.
i is an individual attribute from 1 to n.
f is a similarity function for attribute i in cases T and S.

w is the importance weighing of attribute i.

So, the weight is introduced in the case retrieval and the similarity between cases is considered to be the weighted summation of the similarity between attributes. Although each case contains 15 attributes, showing an example of how to calculate the similarity between the new case and the old cases by using nearest neighbor technique with only 15 attributes.

Table 3. Examples for comparison of cases

Features	Weight	New Case	Old Case1	Old Case2
Cough	2	Yes	Yes	Yes
s breath	2	Yes	Yes	Yes
chest pain	3	Yes	Yes	Yes
Fatigue	3	Yes	Yes	Yes
Fever	3	No	Yes	Yes
Malaise	3	No	No	No
weight loss	3	No	Yes	Yes
tachycardia	5	Yes	Yes	No
Cyanosis	5	Yes	Yes	No
night sweat	5	No	No	Yes
Anemia	5	No	No	Yes
Chills	5	No	No	No
chestdiscomf	5	No	No	No
restless sleep	5	No	No	No
Palenail	5	No	No	No
$\sum_{i=1}^{n} w_i$	59			

$$\sum_{i=1}^{n} f(T_i, S_i) \times w_i = 53$$

Therefore, Similarity (T,S)=0.898305

$$\sum_{i=}^{n} f(T_i, S_i) \times w_i = 33$$

Therefore, Similarity (T,S)=0.559322

Since, old case 1 has greater similarity than old case 2, old case 1 is selected as the nearest neighbor of the new case.

6.4. Reusing the Solution

When the system successfully retrieves the case after careful comparison and matching, the next step is to “reuse” them and produce a result. If the problem completely resembles an existing case, and exact match has been found, the result of the existing case will be used as the final diagnosis without any change. If an exact match is not found, the system reuses the solution of the nearest match as the general solution for the new case.

6.5. Revision of the General Solution

After getting the general solution, the system revises that solution by using adaptation rules. To do this, implement the adaptation rules first. Since, we use adaptation rules only for the revision of the solution produced by the nearest neighbor technique, produce the adaptation rules from the case base by using decision tree induction algorithm (ID_3). The decision tree induction algorithm, which was used in this paper is summarized as follows [4].

Input: set of examples from the casebase.
Output: decision tree, which represents the examples
Tree Construction:

a) Based on the information-theoretic measure, identify the attribute that can provide most information about the examples considered.
b) Start with the attribute that provides most information as a root-node of the tree.
c) Separate the examples into subsets based on the root-node attribute.
d) If all examples in a branch belong to the same class, the node is labeled with that class and algorithm terminates in this leaf.
e) Recursively apply the algorithm for branch of examples until a leaf node is reached.

The formula that has been widely used to identify the attribute is as follows:

The amount of information needed to classify an example from the set of examples is:

$$E(C) = -\sum_{i=1}^{N} P_i \log_2 P_i \qquad (2)$$

Where p_i is the probability for the example belong to the ith class. If an attribute A with Nc possible values is chosen for partitioning the examples, the amount of information needed for grouping an example from the set can be found from:

$$E(C \setminus A) = -\sum_{j=1}^{Nv} p_j \sum_{i=1}^{Nc} \frac{p_{ij}}{p_j} \log_2 \frac{p_{ij}}{p_j} \qquad (3)$$

Where p_j is the probability for the example to have the jth values of the attribute A, and p_{ij} the probability for the instance to both belong to the ith class and have the jth value of the attribute A. Therefore the information that can be gained by selecting attribute A is:

Information Measure (A) = E (C) – E (C\A) (4)

The attribute with the most value is chosen as a root-node for further decision tree until all the examples are exhausted can be developed using the above theory.

Further, generate IF-THEN rules by following the paths from top root node to each of the leaf nodes. Moreover, we asked the expert the demands of the novel situation in which symptoms occur in each disease. So that the system gains the adaptation rules which represents which symptoms should appear together to conclude the corresponding disease.

Then revise the general solution by using these rules. In revision phase, the backward chaining method which is a goal-driven approach is used. So, we can start with our expectation (general solution) of what is to happen and then seek evidence that support our expectation.

If all the features in the input case that satisfy the rule which has the goal (our general solution), the revision process is success and can use the general solution as the solved solution.

6.6. Saving New Case

This system is designed to successfully retain novel problems. However, saving new case can only be made by the person who has the authority to maintain the database (casebase). Any case that has the confirmed solution successessfully used for future retrieval and reuse by the system.

.

6.7. Verification of the System

The verification of the system for case retrieval using nearest neighbor technique include:

1) Check case retrieval accuracy: If the casebase is queried with one of its cases, the system should give the same case with distance measure equals 100%.
2) **Check case retrieval consistency:** If exactly the same search has been performed twice, the same source case should be retrieved.
3) **Check for case duplication:** A case should exactly match itself, but should not be identical to other cases.

7. Analysis of data

Our proposed system supports the diagnosis of four types of diseases concerning lung namely Pneumonia, Tuberculosis, Lung Abscess, and Simple Pneumoconiosis. The system has trained with 50 cases and tested with set of 20 cases for above stated lung diseases. Firstly, a data set is used to train the system. The system is presented with a set of inputs that have known outputs. By comparing output of the system with the known outputs, we can examine the accuracy of the system. The accuracy of the system for diagnosis of each disease is shown in Table 4.

Table 4. The accuracy for each disease

	Disease	Accuracy
1	Pneumonia	88.24%
2	Tuberculosis	97.62%
3	Lung Abscess	95%
4	Simple Pneumoconiosis	80%

8. Conclusion

CBR approach appears to have some advantages concerning system development if compared with other knowledge-based method. CBR diagnosis can be used as an aid to humans to guide them and make sure that they do not overlook important possibilities. Typically, a CBR system consists of a database of past cases and their solutions, a set of indices for retrieving previous cases and storing new cases, and a set of rules for adapting recalled case solutions.

The proposed system supports medical diagnosis by using the corresponding symptoms of the lung diseases that the patient suffers. This system helps the medical practitioner to take a decision with the limited amount of information he has about the patient's disease and prevent delay in the commencement of medical treatment. However, the aim of this system is not to replace a specialist but to reduce the time consumed in carrying out lengthy lab tests. Hence, the system acting as an assistant for medical diagnosis and can be used in any other related domain.

9. References

[1] Alamodt. & Plaza, *Case-based Reasoning: Foundational Issues, Methodological Variations, and System Approaches,* AI Communications, Vol.7 Nr.1, March 1994, pp 39-59.

[2] Balaa, Z.E.; Strauss, A.; Uziel, P.; maximini, K.; and Traphner, R. 2003. "Fm-Ultranet: A Decision Support System Using Case-Based Reasoning, Applied to ultrasonography", Workshop on CBR in Health Sciences, ICCBR"03, pp 37-44.

[3] Chien-Chang Hsu, Li-We Pan, and Cheng-Seen Ho, "Using Induction Trees to Do Case Adaptation in Case-based Reasoning".

[4] Filipic, B. and Junkar, M., *Using Inductive Machine Learning to Support Decision Making in Machine Learning Process,* Computer in Industry,43, 2000, pp.31-41.

[5] Jaulent, M.-C.; Bozec, C.L.; Zapletal, E.; and Degoulet, P.1997. *A Case-based Reasoning Method for Computer Assisted Diagnosis in Histography*, Artificial Intelligence in Medicine, AIME'97, pp 239-242.

[6] J.L.Kolodner (Ed.), Case-Based Reasoning, Morgan Kaufmann Publishers: California, 1993.

[7] Rom: Satria Wahono and Behrouz Homayoun Far, "Reasoning with Cases in the CBR System: A Case Study for Applying OOExpert System", *IECI Japan Workshop,* 2000.

[8] S. Quaglini, P. Barahona and S. Andressen (Eds), "Artificial Intelligence in Medicine", *Proceeding of 8th Conference on Artificial Intelligence in Medicine in Europe,* AIME 2001, Cascais, Portugal, Springer, July 2001.

[9] Salem A.B.M, Roushdy M., and EI-Bagoury, B.M., "An Expert System for Diagnosis of Cancer Disease", *Proceedings of the 7th International Conference on Soft- Co*