Web-Based Medical Diagnosis of Renal Diseases Using Case Based Reasoning

Tin Tin Yu, Dr. Win Lai Lai Phyu
University of Computer Studies, Yangon, Myanmar.
baybayyu@gmail.com , winlei@gmail.com

Abstract

As the web technology grow rapidly, many people take a great interest web-based methods and it is increasingly used to improve the quality of the medical services. Case Based Reasoning (CBR) is the process of solving a given new problem on the similar cases and new solution gained from the past solutions. CBR is the effective technique in the field of medical diagnosis. The primary goal of this paper to develop a case-based system where the user can diagnose four renal diseases. Nearest neighbor algorithm is used to search similar problem of the new problem in the case-base. If the new problem does not match with old problem in the case- base, this solution will be solved using adaptation rules that are taken from specialist. Then, the new solution will be introduced in the case base and it will be used to solve new problems in the future.

Keywords: Case Based Reasoning, Nearest Neighbor Algorithm, Adaptation rules.

1. Introduction

World Wide Web (WWW) is a way of accessing information over the medium of internet. This paper is based on the Web. World Wide Web is a collection of information held in hypermedia on the internet. This information is stored at locations, called Web Sites in the form of Web Pages. These Web Pages are single document and are usually prepared using Hypertext Markup Language (HTML). Special software called browser software is needed to take full advantages of the WWW. Popular browsers are Internet Explorer and Netscape Navigator.

Case-based reasoning (CBR) has been used to create numerous application in a wide range of domains including financial analysis, risk assessment, technical maintenance, process control, quality control, medical diagnosis, software support system, forecasting, planning, design, classification of objects, photo-interpretation, electronic commerce customer support, knowledge management and software engineering. CBR is a recognized and well established method for building medical diagnosis systems. In this paper, we describe an approach for solving problems based on case-based reasoning (CBR) approach. CBR is an approach for problem solving based on similar previous cases [1]. A cognitive model of reasoning and problem solving based on past experiences. It can retrieve-reuse-revise and retain. Understanding when CBR is a good solution. It basics of similarity calculation [3]. A case consists of problem description (common symptoms and specific symptoms) and solution (diseases). Cases are stored in a database of cases called case base. In this paper, the user can diagnose four renal diseases namely, Glomerulonephritis , Chronic Renal Failure, Anemia with Renal Failure and Renal Failure with Viper Bite.

2. Related works

Case-based reasoning (CBR) is an approach to problem solving that emphasizes the role of prior experience during future problem solving (i.e., new problems are solved by reusing and if necessary *adapting* the solutions to similar problems that were solved in the past). It has enjoyed considerable success in a wide variety of problem solving tasks and domains [5].

FM-Ultranet [8] is a medical CBR project implemented with CBR–works ;FM- Ultranet detects malformations and abnormalities of foetus through ultrasonographical examinations. The detection, or diagnosis, uses attributes derived from scans of the mother's uterus, and identifies abnormal organs and extremities. Cases are arranged in a hierarchical and object oriented structure. The attributes consists of anatomical features, medical history and general domain knowledge. Similarity between attributes in the concepts (objects) are mathematically calculated or compared through a look up table, depending on the attribute type. A report of the system's findings is generated when the detection (CBR) process is completed.

A.B.M Salem et.al, [2] 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.

Nilsson et. al,[6] addresses the domain of psychophysiological dysfunctions, a form of stress . The system is classifying physiological measurements from sensors. Features are extracted from the first and second set, and patient specific data, are used as a case. The cases are classified with a k-nearest neighbor match.

In this paper, to retrieve the cases, nearest neighbor cases matching retrieval method is used. For cases similarity, weighted sum similarity is used.

3. Case- Based Reasoning (CBR)

Case-based reasoning (CBR) is a problem solving techniques that solves new problem by reusing or adapting solutions that were used to solve similar problems in the past. The previous problems or past experiences are encoded as cases, known as a case-base, is the knowledge base of experience used to solve new problems. Case-based reasoning is memory based, reflecting human of remembered problems and solution as a starting point of new problem solving.

The processes involved in CBR can be represented by a schematic cycle. CBR is typically as a cyclical process comprising the four REs [7]:

Retrieving: Similar previously experienced cases whose problems are considered to be similar are selected.

Reusing: The cases are reused by either copying or integrating the solution from the retrieved cases.

Revising: The retrieved solution are revised or adapted to try to solve the new problem.

Retaining: The new solution is stored after being confirmed and validated.

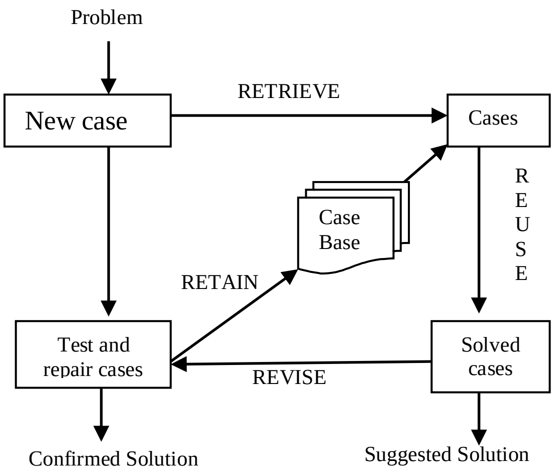

Figure1.1: CBR cycle

3.1. Case Representation

Case is a specific problem that has previously been encountered and solved. **Case base** is the set of historical records of solved problems and the observed solution is empirically justified. In other domains: a "case" is a idealized, typically way of solving a problem or set of problems: the "case base" is a collection of hypothetical problems and corresponding solutions.

A case can be regarded as three features:

1. Its description
2. Its associated solution
3. The justification of its solution

3.2. Case Retrieval

Case retrieval is a process of finding cases which are closet to current case. For efficient case retrieval, there should be selection criteria which will judge a case. The major research area of CBR is case retrieval. Given a description of a problem, a retrieval algorithm, using the indices in the case-memory, should retrieve the most similar cases to the current problem or situation. The retrieval algorithm relies on the indices and the organization of the memory to direct the search to potentially useful cases. The issue of choosing the best matching case has been addressed by research into analogy. Among well known methods for case retrieval are:

1. Nearest Neighbor Retrieval
2. Induction Retrieval
3. Knowledge Guided Induction
4. Template Retrieval

3.2.1: Nearest neighbor Retrieval

This approach involves the assessment of similarity between stored cases and the new input case, based on matching a weighted sum of features. The biggest problem there is to determine the weights of the features. The limitation of this approach includes problems in converging on the correct solution and retrieval times. In general the use of this method leads to the retrieval time increasing linearly with the number of cases. Therefore this approach is more effective when the case base is relatively small. Several CBR implementations have used this method to retrieve matching cases.

3.2.2: Induction Retrieval

Inductive retrieval algorithm is a technique that determines which features do the best job in discriminating cases and generates a decision tree type structure to organize the case in memory. This approach is very useful when a single case feature is required as a solution, and when that case feature is dependent upon others.

3.2.3: Knowledge guided induction

This method applies knowledge to the induction process by manually identifying case features that are known or thought to affect the primary case feature. This approach is frequently used in conjunction with other techniques, because the explanatory knowledge is not always readily available for large case bases.

3.2.4: Template retrieval

Similar to SQL like queries, template retrieval returns all cases that fit within certain parameters. This technique is often used before other techniques, such as nearest neighbor, to limit the search space to a relevant section of the case base.

3.3. Case Adaptation

The reuse process in the CBR cycle is responsible for proposing a solution for a new problem from the solutions in the retrieved case. Reusing the retrieved case can be as easy as returning the retrieved solution, as the general solution for new problem. If there are significant differences between the new problem and the retrieved case's problem, reuse become more difficult .In these situations, the retrieved solution is needed to adapt. In general, there are two kinds of adaptation in CBR.

(i) Structural adaptation in which adaptation rules are applied directly to the solution stored in cases.

(ii) Derivational adaptation that reuses the algorithms, methods or rules that generated the original solution to produce a new solution to the current problem.

3.4. Case Retain

The retain process in CBR is the process of incorporating whatever is useful from the new case into the case base. This is also the process of incorporating what is useful to retain from the new problem solving episode into the existing knowledge. The learning from success or failure of the proposed solution is triggered by the outcome of the evaluation and possible repair. It involves selecting which information from the case to retain, in what form to retain it, how to index the case for later retrieval from similar problem, and how to integrate the new case in the memory structure.

4. Nearest Neighbor Algorithm

Nearest neighbor algorithms are the most widely used technology in CBR. This approach involves the assessment of similarity between stored cases and the new input case, based on matching a weighted sum of features. The procedure of this algorithm is as follow: For each feature in the input case:

1 Find the corresponding feature in the stored case.
2 Compare the two values to each other and compute the degree of match.
3 Multiply by a coefficient representing the important of the feature to the match.
4 Add the results to derive an average match score.
5 This number represents the degree of match of the old case to the input case.
6 A case can be chosen by choosing the item with the largest score.

The main distance measure equation is:

$$\frac{\sum_{i=1}^{n} w_i \times Sim(fi^L, fi^R)}{\sum_{i=1}^{n} wi}$$

wi = the importance of the feature (slot) *i*
Sim(,) = the similarity function
fiL, fiR = the values for feature fi in the source and target cases, respectively.
n = the number of attributes in each case.

The similarly function is :

$$Sim(fi^L, fi^R) = 1 - (|\, fi^L, fi^R \,| / |\, f\max - f\min \,|)$$

If feature fi s numeric
0 if features fi is symbolic and fiL ≠ fiR
1 if features fi is symbolic and fiL = fiR

So, the weight is introduced in the case retrieval and the similarity between cases is considered to the weighted summation of the similarity between attributes.

5. System Implementation

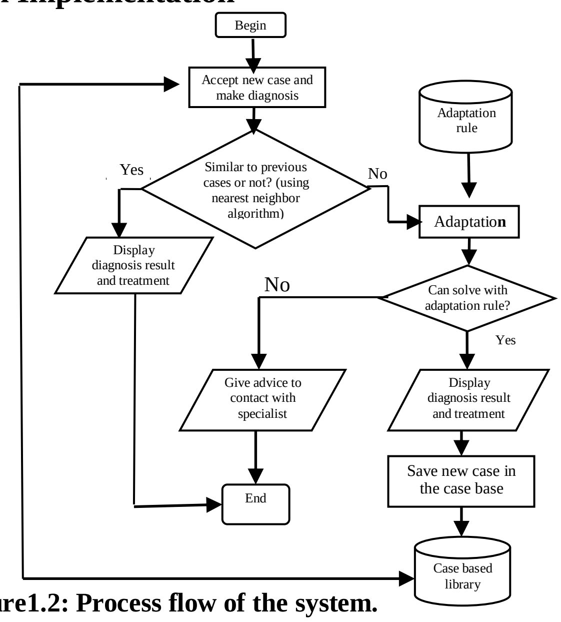

Figure1.2: Process flow of the system.

The system can diagnose four renal diseases .In this system, 400 cases of renal diseases from Sanpya hospital are used as shown in Table 1.1 and each case consists of 35 attributes as shown in Table 1.2.

Table 1.1 Showing the diseases and number of cases

No.	Diseases	Number of cases
1.	Glomerulonephritis	100
2.	Chronic Renal Failure	100
3.	Anemia with Renal Failure	100
4.	Renal Failure with viper bite	100

Table1.2 Showing the attributes and values of renal diseases

No.	Symptoms	Values
1.	Generalized Oedema	1=no, 2=mile, 3=both arms, 4=whole body
2.	Proteinuria	1=clear, 2=trace, 3=1+(10%), 4=2+(20%), 5=3+(30%), 6=4+(40%), 7=solid(>40%)
3.	Fluid Overload	1=no, 2=<200cc, 3=200cc-500cc 4=500cc-1000cc, 5=>1000cc
4.	Drawziness	1=no, 2=yes
5.	Ascites	1=no, 2=yes
6.	Tender	1=no, 2=yes
7.	Hypertension	1=normal (<=120/80) 2=mild abnormal(upper 130-150, lower 90-110) 3=severe abnormal(>150/110)
8.	Anuria	1=no, 2=yes
9.	Abdominal distension	1=no, 2=yes
10.	Puffiness of faces	1=no, 2=yes
11.	Dehydration	1=no, 2=slightly ,3=severe
12.	Pallor	1=no, 2=yes
13.	Dysponea	1=no, 2=slightly, 3=severe
14.	Heptaspenemag-uly	1=no, 2=yes
15.	Anemia	1=no, 2=yes
16.	Palpitation	1=no, 2=yes
17.	Vomitting	1=no, 2=mild, 3=several time
18.	Neckstiffing	1=no, 2=yes
19.	Ankle swelling	1=no, 2=yes
20.	Polycythemia	1=no, 2=slightly, 3=severe
21.	Haematuria	1=no, 2=slightly, 3=severe
22.	Bowel sound	1=no, 2=yes
23.	Chest pain	1=no, 2=yes
24.	Headaches	1=no, 2=yes
25.	Fever	1=normal (98.6F) 2=low grade fever(99-101F) 3=too high grade fever(>101F)
26.	Pruritus	1=no, 2=yes
27.	Insomia	1=no, 2=yes
28.	Constipation	1=no, 2=yes
29.	Coma	1=no, 2=yes
30.	Appetite loss	1=no, 2=yes
31.	Oliguria	1=no, 2=yes
32.	Being bitten by insects	1=no, 2=yes
33.	Fatigue and weakness	1=no, 2=yes
34.	Nausea	1=no, 2=yes
35.	Back pain	1=no, 2=yes

5.1. Assigning weight value to feature

Features weights for most problem domains are content dependent. The weight assign to each feature of the case tells how much attention to pay to matches and mismatches in the field when computing the distance measure of a case. There are two ways of assigning weight values to case features.

(1) One way to assign importance value is to have a human expert assign them as the case library is being built. The expert might have some feeling about which dimension and combination of dimension make good predictors.

(2) Another way to assign to important values to do a statistical evaluation of a know corpus of cases to determine which dimension predict different outcomes and/or solutions best. Those that are good predictor are then assigned higher important for matching [4].

The system uses **the first method** to assign the important value to different feature in the case base as shown in Table 1.3.

Table 1.3 Showing the weigh values of the attributes

No.	Symptoms	Important weight value
1.	Generalized Oedema	1
2.	Proteinuria	1
3.	Fluid Overload	1
4.	Drawziness	0.8
5.	Ascites	0.8
6.	Tender	0.8
7.	Hypertension	0.8
8.	Anuria	1
9.	Abdominal distension	0.4
10.	Puffiness of faces	0.4
11.	Dehydration	0.8
12.	Pallor	0.8
13.	Dysponea	0.4
14.	Heptaspenemaguly	0.4
15.	Anemia	0.8
16.	Palpitation	0.2
17.	Vomitting	0.4
18.	Neckstiffing	0.4
19.	Ankle swelling	1
20.	Polycythemia	0.4
21.	Haematuria	1
22.	Bowel sound	0.4
23.	Chest pain	0.2
24.	Headaches	0.2
25.	Fever	0.2
26.	Pruritus	0.2
27.	Insomia	0.2
28.	Constipation	0.2
29.	Coma	0.8
30.	Appetite loss	0.2
31.	Oliguria	1

32.	Being bitten by insects	1
33.	Fatigue and weakness	0.2
34.	Nausea	0.2
35.	Back pain	0.8

5.2. Case retrieval for proposed system

When a new case arrives, the proposed system will retrieve the most similar case in the case base by calculating nearest neighbor matching. 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 35 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 10 attributes. Table 1.4 shows the example of ten attribute of new case and two old cases.

Table1.4 Showing the example of ten attributes of old cases and new case

No.	Symptoms	Old case1	Old case 2	New case
1.	Generalized Oedema	4	2	4
2.	Proteinuria	7	1	7
3.	Fluid Overload	3	5	5
4.	Drawziness	2	1	2
5.	Ascites	2	2	2
6.	Tender	2	2	2
7.	Hypertension	2	2	2
8.	Anuria	1	1	1
9.	Abdominal distension	1	1	1
10.	Puffiness of faces	1	2	1

For old case 1,

$Sim\,(f_1^l, f_1^R) = 1$, $Sim\,(f_2^l, f_2^R) = 1$

$Sim\,(f_3^l, f_3^R) = 0$, $Sim\,(f_4^l, f_4^R) = 1$

$Sim\,(f_5^l, f_5^R) = 1$, $Sim\,(f_6^l, f_6^R) = 1$

$Sim\,(f_7^l, f_7^R) = 1$, $Sim\,(f_8^l, f_8^R) = 1$

$Sim\,(f_9^l, f_9^R) = 1$, $Sim\,(f_{10}^l, f_{10}^R) = 1$

$$\frac{\sum_{i=1}^{n} w_i \text{ x Sim } (f_i^l, f_i^R)}{\sum_{i=1}^{n} w_i} = \frac{7.0}{8.0} = 0.85$$

For old case2,

$Sim\,(f_1^l, f_1^R) = 0$, $Sim\,(f_2^l, f_2^R) = 0$

$Sim\,(f_3^l, f_3^R) = 1$, $Sim\,(f_4^l, f_4^R) = 0$

$Sim\,(f_5^l, f_5^R) = 1$, $Sim\,(f_6^l, f_6^R) = 1$

$Sim\,(f_7^l, f_7^R) = 1$, $Sim\,(f_8^l, f_8^R) = 1$

$Sim\,(f_9^l, f_9^R) = 1$, $Sim\,(f_{10}^l, f_{10}^R) = 0$

$$\frac{\sum_{i=1}^{n} w_i \text{ x Sim } (f_i^l, f_i^R)}{\sum_{i=1}^{n} w_i} = \frac{4.8}{8.0} = 0.6$$

In this proposed system, the predefined threshold value is 0.7. That is because the nearest case's solution solved by this value is almost identical to the real world renal disease. This solutions are recommended by the Dr. Phyu Phyu (Sanpya Hospital). Therefore, old case 1 is chosen as the nearest case of the new case according to the Nearest Neighbor algorithm.

5.3. Case adaptation for proposed system

If the retrieved case is less than the threshold, the system needs to adapt the case. In the proposed system, adaptation rules are used to transform the solution which is produced by the nearest neighbor techniques. It acquires adaptation rules form specialists. These rules are stored in the knowledge base. The proposed system saves the input cases and its solution as a new case in the case base after getting the solved solution which is produced by adaptation rules.

6. Conclusion

CBR is a suitable methodology for most of the medical domains and tasks for the following reasons; cognitive adequateness, explicit experience, duality of objective and subject knowledge and system integration. The proposed system supports medical diagnosis by using the symptoms of renal diseases that the patients suffer. Although the system cannot replace by a specialist, the system helps medical staff to get correct diagnosis in time in emergency cases and prevent delay in the commencement of medical treatment.

References

[1] A.Aamodt, and E.Plaza, 1994. "*Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches*". AI Communications. IOS Press 7(1):39-59.

[2] A.B.M Salem, M.Roushdy, and B.M.EI-Bagoury, "An Expert System for Diagnosis of Cancer Disease", Proceedings of the 7th *International Conference on Soft-Computing, MELDEI, (2001)*, pp. 300-305

[3] Funk, Peter,"*Case-Based Reasoning*".

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

[5] Mantaras, Raman Lopex De and Mcsherry, David, "*Retrieval, Reuse, Revision, and Retention in Case- Based Reasoning*"

[6] M.Nilsson, P.Funk, and M.Sollenborn, "Complex measurement classification in medical applications using a case-based approach" Workshop on CBR in the Health Sciences, ICCBR'03, pp. 63-72.

[7] M. U. Ahmed, P. Funk, N. Xiong *"Case-based Reasoning for Diagnosis of Stress using Enhanced Cosine and Fuzzy Similarity"*, Technical Report, Osaka University,2003

[8] Z.E Balaa, A.Strauss P.Uziel, K.Maximini, and R. Traphner, 2003. "*Fm-ultranet; A decision support system using case-based reasoning applied to ultrasonography*." Workshop on CBR in the Health Sciences, ICCBR'03, pp37-44