A CBR Diagnostics System Applied in the Brazilian Public Health System

Márcia Regina Moss Júlio[1], Gilberto Nakamiti[2]

[1]R. Clariano Peixoto, 280 – Limeira – SP - BRAZIL
Communitary Faculty of Limeira
lmjulio@linkway.com.br
[2]Rod. D. Pedro I, Km 136– Campinas – SP - BRAZIL
Catholic University of Campinas / Paulista University
g_nakamiti@uol.com.br

Abstract. Case-based systems constitute an Artificial Intelligence technique, which is well-known for their ability to reuse and to adapt past experiences to solve new problems. To achieve this, they store and retrieve past experiences, based on their attributes, adapt them, and apply the adapted solution to the new problem. After analysing the solution performance, it may store the new case in the case-based, so that the system can improve its performance over time. The main feature of such systems is the retrieving mechanism, which is responsible for comparing and identifying similar, relevant information in the stored cases, through similarity metrics. This work presents an approach for handling multivalued attributes in well-known similarity metrics in order to retrieve the most relevant cases for a specific class of application. A health area application concerning lateral epicondilitis, which is an elbow tendonitis, was developed, and it illustrates its use. This work contains real data from the Brazilian public health system, where the work was developed and is been prepared to be used.

1 Introduction

When we face a new problem, we often remind similar problems that we solved in the past, and use them to solve the new one.

Case-Based Reasoning (CBR) may be seen as a problem solving approach that uses and adapts similar past solutions for new problem situations.

The mechanism used to retrieve the best and most similar past situations is of utmost relevance in such systems, because the situations retrieved will be used as the basis for the new problem solution.

This work presents an approach to handle multivalued attributes in well-known similarity metrics, and it shows its use in a health area case-based system,. This application had a strong demand on the Brazilian public health system, in particular in Araras and Limeira regions, in the state of Sao Paulo, Brazil, due to the lack of personnel in some public health facilities.

Section 2 presents an overview of the main case-based concepts and similarity metrics used in this work.

© A. Gelbukh, S. Torres, I. López (Eds.)
Advances in Computer Science and Engineering
Research in Computing Science 19, 2006, pp. 225-234

Section 3 describes the application, and shows how the multivalued attributes were handled, besides the results obtained by the system.

Finally, we present the conclusions of this work.

2 Case-Based Systems

Case-Based Systems are an Artificial Intelligence (AI) technique that reproduces human cognition aspects to solve specific problems that usually are solved by human experts in a specific knowledge area. They simulate the human act of remembering a past case to solve a new one through the identification of affinities among them [2] [6] [7].

In a Case-Based System, a case represents a complete description of a problem in that application domain, with a solution already applied. It is an abstraction of an experience, described by valued attributes. The attributes describe the experience content and context. After a similar case is retrieved based on its attributes, it is adapted to fit the new problem, offering a better solution [3] [12] [13].

The CBR cycle proposed by Aamondt e Plaza, illustrated by Figure 1, is very useful to understand the process, being composed by four tasks:

- Retrieve the most similar cases;
- Reuse (adapt) the solution for the new problem;
- Revise the proposed solution; and
- Retain (learn) the experience representing the new case for further use.

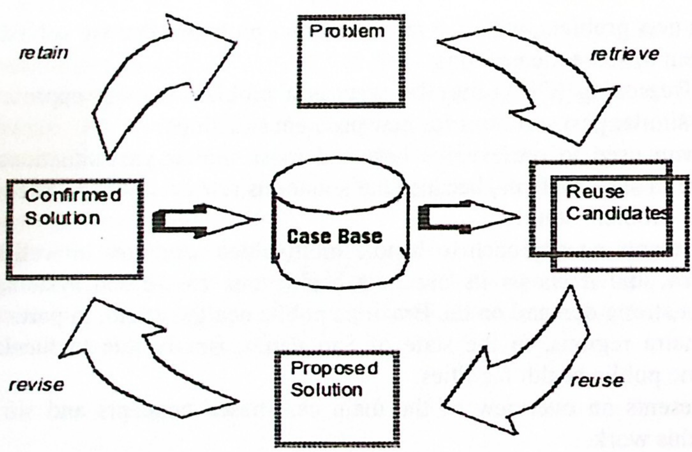

Fig. 1 The CBR Cycle

Similarity metrics are of utmost importance in case-based systems because they are responsible to identify the most similar stored solutions to be used in the new problem situation. These solutions will serve as basis for the whole cycle [10] [11].

Similarity may be seen as an intuitive concept used to describe the perception of a human observer about the common aspects existing in two objects [14].

A very accepted formalization of the similarity concept in computer systems resides in the definition of a numeric measure of distance or similarity. We can understand the similarity measure as a formalization of a specific similarity judgment through a concrete mathematical model [4].

A similarity metric synthesizes the similarity through a measure of importance of each attribute when compared with the other ones, and this process can be conducted in a variety of ways [5] [8] [9].

3 The FisioSmart System

In order to develop the application system, we first conducted extensive interviews with health area professionals from public health facilities in Araras and Limeira regions, in the state of Sao Pauloi, Brazil, to better understand the application domain. These professionals needed an automated system to help their work and of their assistants to take care of the poor people of those regions. The first focus was to identify the relevant cases attributes, and their correlations among themselves and the possible diagnoses.

3.1 Application

FisioSmart is a Diagnostics Evaluation System for Lateral Epicondilitis, which is an elbow tendonitis. The application developed uses CBR and serves as a basis to test the efficacy of similarity metrics, in particular the ones based on distance handling multivalued attributes.

The system was developed in Delphi, using a Firebird Database.

When a physiotherapist interviews a patient with lateral epicondilitis, he or she formulates a series of questions to aid the choice of the best procedures and conducts to treat the patient. With the patient answers, the system will be able to retrieve the most similar cases in the case-base, in order to provide a good similar case to be reused.

3.2 Implemented Metrics

The similarity metrics implemented in FisioSmart were:

- City-Block (Manhattan);
- Characteristics Count;
- Euclidian;

- Square Euclidian;
- Weighted Square Euclidian;
- Closest Neighbor – Query Insert Function;
- Closest Neighbor – Intersection Function;
- Closest Neighbor – Linear Function.

When we applied some of the cited similarity metrics, in particular when dealing with multivalued attributes, some adjustments were performed, as they will be described in the next sections.

3.3 Dealing with Monovalued and Multivalued Attributes

In order to test and to compare the similarity metrics implemented, we started from 40 real cases, taken from public health centers in Araras region, state of Sao Paulo, Brazil.

From the 40 cases collected, 25 cases fed the case-base, and 15 cases were used as new cases. The system had to propose treatments for these new cases. Each test set consisted of inserting the 25 cases in the case-base, and simulating the 15 new ones. This test set was performed for each of the similarity metrics implemented.

At this point, some considerations should be made:

1. Some monovalued attributed were classified as Boolean, and the Boolean metrics was used to handle them, independently of which metrics was been used with the other attributed [10]

2. Each multivalued attribute could use a set of predefined values.

Moreover, to apply the similarity metrics some adjustments in the original metrics were performed:

1. Usually, the similarity metrics use the descriptors distance to calculate the attributes similarity value. This strategy is very useful when handling monovalued attributes, but it is not usable in the original form to handle multivalued attributes. The first proposed adaptation consists in analyzing the set of valued, as presented in Table 1.

Table 1. Adjustment 1

	New Case		Case in the Case-Base	
	Attribute – Main Symptom	**Descriptor value**	**Attribute – Main Symptom**	**Descriptor value**
01	Elbow pain	0.7	Elbow pain	0.7
02	Itching	0.8	Itching	0.8
			Fist pain	0.6

To compare the multivalued attribute Main Symptom, we used a Cartesian product among the stored attributes' values and the new cases ones because, for this specific application area, the symptoms are not independent, as they reflect in one another. Due to this interdependency, they were rated based on their mutual proximity, so that the closer the symptoms, the closer their values. We can observe below a model of how this strategy is conducted:

1-(L1N – L1B) → 1- |0.7 – 0.7| = 1
1-(L1N – L2B) → 1- |0.7 – 0.8| = 0.9
1-(L1N – L3B) → 1- |0.7 – 0.6| = 0.9
1-(L2N – L1B) → 1- |0.8 – 0.7| = 0.9
1-(L2N – L2B) → 1-|0.8 – 0.8| = 1
1-(L2N – L3B) → 1-|0.8 – 0.6| = 0.8

Average → (1+0.9+0.9+0.9+1+0.8)/6 = 0.91
Maximum → 1

Where, in Table 01:

L1N → Row 01 of the New Case
L2N → Row 02 of the New Case
L1B → Row 01 of the Case in the Case-Base
L2B → Row 02 of the Case in the Case-Base
L3B → Row 03 of the Case in the Case-Base

2. When comparing two multivalued attributes with identical values, the Cartesian product may lead to a misleading result. Table 2 shows an example.

Table 2. Adjustment 2

	New Case		Case in the Case-Base	
	Attribute – Main Symptom	**Descriptor value**	**Attribute – Main Symptom**	**Descriptor value**
01	Elbow pain	0.7	Elbow pain	0.7
02	Itching	0.8	Itching	0.8

Since the two attributes are identical, the similarity measure between them should be 1. This simple adjustment only turns to 1 the similarity value measured between identical multivalued attributes.

1-(L1N – L1B) → 1- |0.7 – 0.7| = 1
1-(L1N – L2B) → 1- |0.7 – 0.8| = 0.9
1-(L2N – L1B) → 1- |0.8 – 0.7| = 0.9
1-(L2N – L2B) → 1- |0.8 – 0.8| = 1

Average → (1+0.9+0.9+1)/4 = 0.95
Maximum → 1

Where, in Table 2:

L1N → Row 01 of the New Case
L2N → Row 02 of the New Case
L1B → Row 01 of the Case in the Case-Base
L2B → Row 02 of the Case in the Case-Base

Table 3 shows an example of how the global case similarity function is calculated in a case example, with monovalued, multivalued and Boolean attributes:

Table 3. Calculating the similarity of a case

Attribute	Weight	New Case	Case in the Case-Base	Local Similarity
Main Symptom	10	Alteration of Sensibility of the superior member (0,4) Pain in the superior Member (0,5)	Elbow pain (0.7)	0.50
Other symptoms	1	Decrease of force in the extending movement	Decrease of force in the extending movement	1.00
Initial Detection	2	Repenting	Progressive	0.00
Concomitant with which activities	4	Tennis (1)	Sports (0.5)	0.25
Improves with which activities	1	To immobilize the superior member (0,6)	Rest during the work (1)	0.30
Worsen with which activities	1	Physical exercises (0.7)	Carrying weight (0.9)	0.40
Familiar antecedents	5	No	No	1.00
Personal antecedents	1	No	No	1.00
Associated pathologies	6	No	No	1.00
Previous treatments	7	No	Physiotherapy (1)	0.00
Work activities	5	Tennis teacher (0.8) Swimming teacher (0.1)	Repetitive movements (1)	0.30
Home activities	1	No	No	1.00
Sports	1	Swimming (0.3) Tennis (1)	Swimming (0.3)	0.43
Emotional suffering	1	No	No	1.00

Palpation Osteo Tendinous	10	Positive	Positive, with irradiation	0.50
Resistive Movement when extending the Fist	5	Positive	Positive	1.00
Elbow passive movements	10	Negative	Positive	0.00
Fist passive movements	9	Negative	Positive	0.00
Passive Prolongation when Extending the Fist	6	Positive	Positive	1.00
Muscular Palpation	8	Positive	Positive	1.00
Resistive Movement	10	Positive	Positive	1.00

Global similarity: (10*0.50 + 1*1.00 + 2*0.00 + 4*0.25 + 1*0.30 + 1*0.40 + 5*1.00 + 1*1.00 + 6*1.00 + 7*0.00 + 5*0.45 + 1*1.00 + 1*0.43 + 1*1.00 + 10*0.50 + 5*1.00 + 10*0.00 + 9*0.00 + 6*1.00 + 8*1.00 + 10*1.00) / (10 + 1 + 2 + 4 + 1 + 1 + 5 + 1 + 6 + 7 + 5 + 1 + 1 + 1 + 10 + 5 + 10 + 9 + 6 + 8 + 10) = 58.38/104 = 0.56

4 Results

For each similarity metric implemented, we initialized the case-base with the same 25 cases, and 15 new cases were used to test the system. This way, the first test case was compared with the 25 initial case-based cases and, after its application, it was inserted in the case-base. The second test case was compared with 26 cases, and so on.

The similarity metrics which retrieved the most similar and useful cases, according with the area specialists, were the Weighted Square Euclidian, which returned the best cases in 8 out of the 15 tests, and the Closest Neighbor (Linear Function), which returned the best case in 7 out of 15 tests. The criteria used to select the best cases was to select the ones with the highest similarity degrees, since their similarity degrees reached at least 0.5.

Figure 2 illustrates the results of the best metrics for this Lateral Epicondilitis system. It illustrates how the retrieved cases match the cases, which were indicated by the area specialists as adequate to be taken into account in the new case. The more cases considered very similar to a particular case they retrieved, the highest their values.

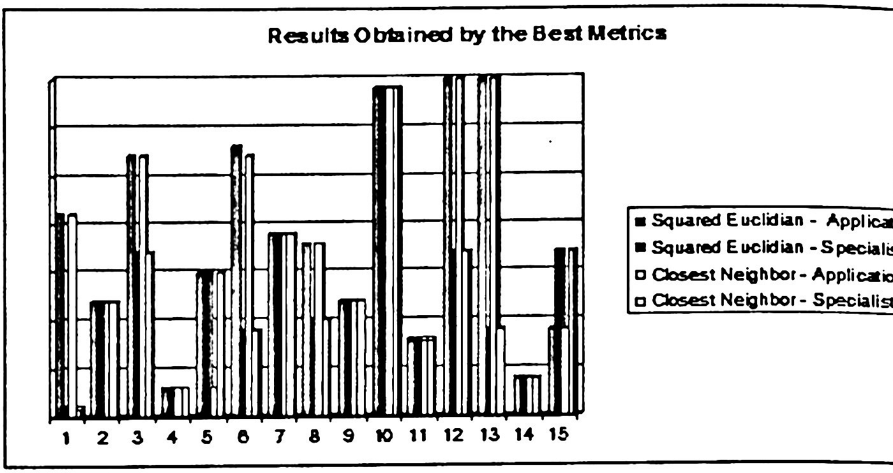

Fig. 2 Similarity metrics results

A reason why these two similarity metrics were particular efficient in the proposed application domain can due to the fact that their calculation included weights for the attributes and values for all the sets of descriptors.

5 Final Remarks

Case-Based Reasoning (CBR) is an Artificial Intelligence technique that simulates the reasoning of a specialist. Its processing is based on reusing past experiences to analyze and propose solutions for a current case.

In such systems, the main knowledge source is the case-base, and the reasoning basically consists in retrieving cases based on their similarity with the current problem, and adapt them to the new situation.

This work presented a Case-Based system applied to Lateral Epicondilitis treatment, which is an elbow tendonitis. The system included the implementation of a set of similarity metrics, which could be tested and compared in the application context. Techniques to deal with multivalued attributes were also proposed and tested. The data concerned real cases collected from public health facilities in Araras and Limeira regions, in the state of Sao Paulo, Brazil. The system will be applied in these regions to help taking better care of the poor people who cannot pay for private health professionals. Currently, the system is been monitored in public health facilities in order to incorporate new cases and to observe and to make the necessary adjustments for its application.

References

1. Aamodt, A., E. Plaza. Case-Based Reasoning: Foundational Issues, methodological Variations, and System Approaches, AICOM, Vol. 7, No. 1 (1994)
2. Camargo, K. Artificial Intelligence Applied to Nutrition. Master Thesis. Santa Catarina Federal University (1999)
3. Defense Advanced Research Projects Agency (DARPA), Case-Based Reasoning, Proceedings of a Workshop on Case-Based Reasoning, Florida (1989)
4. Gresse, C. REMEX - A case based approach for reuse of software measurement experienceware. Proceedings of the 3rd Int. Conference on Case-Based Reasoning, Germany (1999)
5. Julio, M. A Study on Similarity Metrics in Case-Based System Applied to the Health Area. Master Thesis. Univesity of Campinas (2005)
6. Kolodner, J. Case-Based Reasoning. Morgan Kaufmann (1993)
7. Lagemann, G. Client Supporting Using CBR: The Datasul Case. Master Thesis. Santa Catarina Federal University (1997)
8. Lee, R. Intelligent Jurisprudential Research. PhD Thesis. Santa Catarina Federal University (1998)
9. Martins, A. Case-Based Computing: Methodological Contributions to Models of Indexing, Evaliation, Ranking, and Similarity. PhD Thesis. Paraíba Fedreal University (2000)
10. Mello, M. Applying the 7 Project Control Metrics, Rational Software Latin América. White Paper (2002)
11. Osborne, H., Bridge, D. Similarity metrics: A Formal Unification of Cardinal and Non-Cardinal Similarity Measures. Technical Report, Department of Computer Science, University of York (1997)
12. Schank R. Dynamic memory – A theory of reminding and learning in computers and people. Cambridge University Press (1982)
13. Sbrocco J., Freitas, R. Nakamiti, G., Failure Detection on Data Servers Using Intelligent Agents. Proceedings of the VI Brazilian Symposium on Intelligent Automation, Bauru (2003)
14. Turban, E., Jay E. A. Decision Suport Systems and Intelligent Systems, Prentice-Hall, New Jersey (1998)