

# Case-Based Reasoning Applied to Medical Diagnosis and Treatment

Xiomara Blanco, Sara Rodríguez, Juan M. Corchado, and Carolina Zato

Departamento Informática y Automática, Universidad de Salamanca  
Plaza de la Merced s/n, 37008, Salamanca, Spain  
{xiopepa, srg, corchado}@usal.es

**Abstract.** The Case-Based Reasoning (CBR) is an appropriate methodology to apply in diagnosis and treatment. Research in CBR is growing and there are shortcomings, especially in the adaptation mechanism. In this paper, besides presenting a methodological review of the technology applied to the diagnostics and health sector published in recent years, a new proposal is presented to improve the adaptation stage. This proposal is focused on preparing the data to create association rules that help to reduce the number of cases and facilitate learning adaptation rules.

## 1 Introduction

The case-based reasoning (CBR) is a methodology for reasoning on computers that tries to imitate the behavior of a human expert and learn from the experience of past cases.

The CBR has demonstrated to be a systems development methodology appropriate smart when you want to apply in unstructured domains. Therefore, the choice of this methodology is appropriate for the development of diagnostic support systems in multidisciplinary medical services (1).

Nilsson and Sollenbom analyzed the CBR published between 1999 and 2003 (2), and concluded that hybrid systems using other techniques of Artificial Intelligence (AI) are increasing. This is because the application domain is increasingly complex, the potential use of these systems in the clinical area is high, but much work is still needed. More recent studies of CBR applied to health sciences conclude that there are many opportunities to work in diagnostic support for disabled and elderly (3-5). (5) emphasizes three areas for improvement: (i) reducing the search space in case recovery; (ii) maintaining knowledge always valid, embedding knowledge in cases to assist in reviewing the conclusions; (iii) work adaptation methods that consider allowing local constraints. Another paper about review published in 2011, notes that the CBR has been applied to various medical tasks such as diagnosis, classification, treatment and tutorials (6). For future work these

authors (6) suggest using probabilistic and statistical computing due to data systems every time are more and more large, complex and uncertain in clinical environment. Another conclusion is that the automatic adaptation is a weak point and a big challenge, especially in the medical field (6).

This paper focuses on medical CBR systems created between 2008 to 2011. The aim is to report the results of a systematic review of CBR applied to the health sector. We follow the methodology for systematic review of literature for software engineering researchers to Kitchenham (11)..

Another objective is to make a proposal to improve the adaptation stage. We focused on preparing the data to create association rules that help to reduce the number of cases and make learning easier around adaptation rules. This article is organized as follows: section 2 is an overview of the CBR applied to the health sector. Section 3 is shown the steps to implement the revised methodology used in this work. In section 4, the results achieved with the review and new proposals for the adaptation stage are described. In section 5 some conclusions and future work are presented.

## 2 Case-Based Reasoning Applied to the Health Sector

An early exploration of CBR in the medical field was conducted by Koton (7) and Bareiss (8) in the late 1980s. The CBR is inspired by human reason, i.e. to solve a problem by applying previous experiences adapted to the current situation.

The domain of health sciences offers the scientific community challenges that are difficult to solve with other methods and approaches. In the medical context, the symptoms represent the problem, and diagnosis and treatment are the solution. Aamodt and Plaza (9) have described a life cycle with four main steps (retrieve, reuse, revise and retain) (6).

Learning occurs naturally, if a case is successful the experience is retained. If the case is unsuccessful, the reason for the failure is identified and remembered.

In the 90's many efforts were made to create adaptive algorithms in medicine, but always was specific rules for the CBR built. In the literature there are solutions aimed at:

- Avoid the problem of adaptation.
- Solve the problem: Implementing rules that experts propose

Adaptation by composition, this is one of his biggest successes in implementing influenza early warning. The solution of multiple cases are combined to produce a new solution, It is efficient if there are few conflicts between different components so a change in one component does not have many effects on others (10).

Adaptation by abstraction of cases was performed in diagnosis and therapy, each medical case contains between 40 and 130 symptoms and syndromes. Cases prototype were abstracted and then adapted, so few restrictions have to control through rules (22).

In Adaptation by abstraction, cases are stored at many levels of abstraction, the adaptation is performed from top to bottom taking the statistical mode. Initially, the solution is adapted to the highest level of abstraction (omitting less relevant details), then the solution is gradually refines and the required details are added. This adaptation can reuse a single case or it can reuse the different cases for different levels of abstraction or refinement of various details of the solution (10).

Unfortunately, the technique is general, but the content of these rules are domain specific. Especially for complex medical tasks generating adaptation rules is often tedious and sometimes even impossible.

However, for therapeutic tasks some adaptation rules can be applied, for example, by substitution adaptation for calculating dose (22).

Views general lines that are included in the case-based reasoning applied to the health sector, the following section summarizes the methodological work done to make a review of the work done in this field to date.

### **3 Literature Review Methodology**

The methodology used is that described by Dr. Barbara Kitchenham in 2007 (11). A systematic review of literature is a secondary study that uses a well-defined methodology to identify, analyze and interpret all available data relating to a specific question. Systematic reviews seek to summarize the existing evidence on a specific topic. You can also identify gaps in current research to suggest areas for future research. A well defined methodology makes it less likely that the results are partial literature.

The steps associated with the planning of the review are: Identify the need for a review. Assign a review. Specify the research question, (it is the most important of any review). Develop a review protocol. Evaluate the review protocol. Perform the review: identify research, selection of primary studies, study quality assessment, data extraction and control and data synthesis. In this point, it is important to identify whether the results of the studies are consistent with each other (ie, homogeneous) or incompatible. Finally, report.

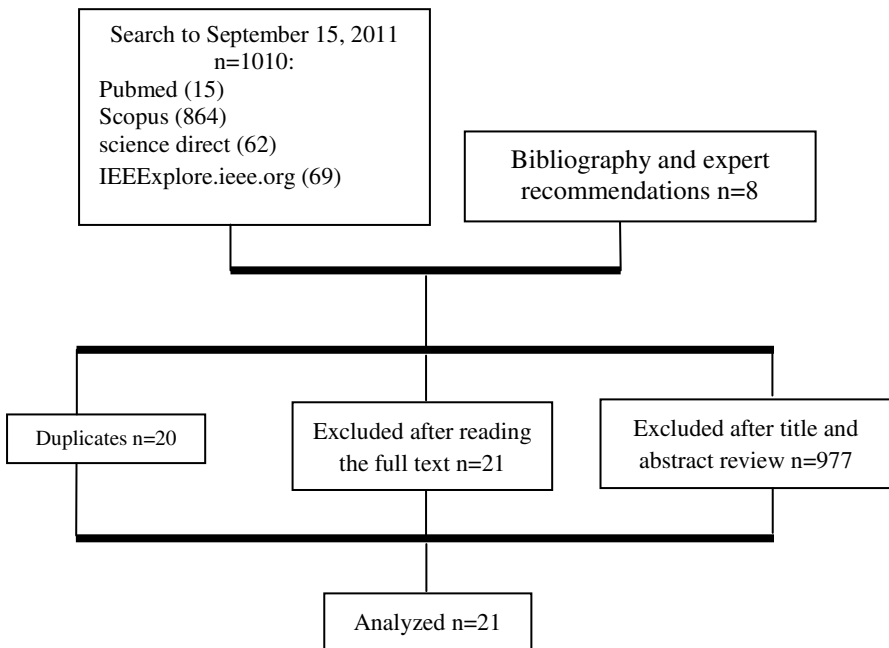
We conducted a systematic review of studies of CBR in medicine between September 2008 and September 2012, We considered only studies that validated their results. Research that were excluded: : CBR applied to tutorials, studies including summaries, systematic reviews and studies of non-human.

## 4 Results

This section shows the results more significant obtained after the methodological review:

### 4.1 Results: Review

Out of a total 1018 references were initially identified: 1010 of electronic databases and bibliographies 8 authors and suggestions. See Figure 1 for details. 21 articles were selected.



**Fig. 1** Table summarizes the flow of study selection

### 4.2 Results: Trends in Medical CBR

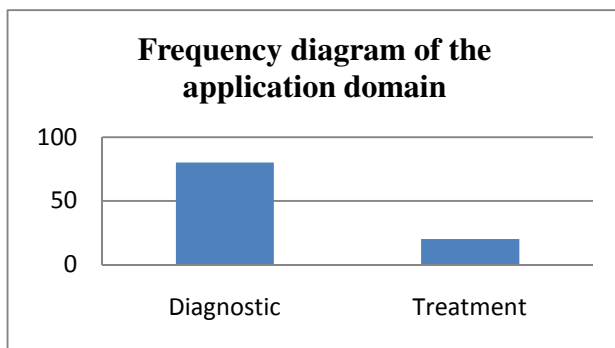
The results of the review are summarized in the following tables and figures.

Table 1 shows the techniques of CBR or artificial intelligence used in each phase of the CBR (retrieval, reuse, revision and retention or learning), and the success rate in system validation.

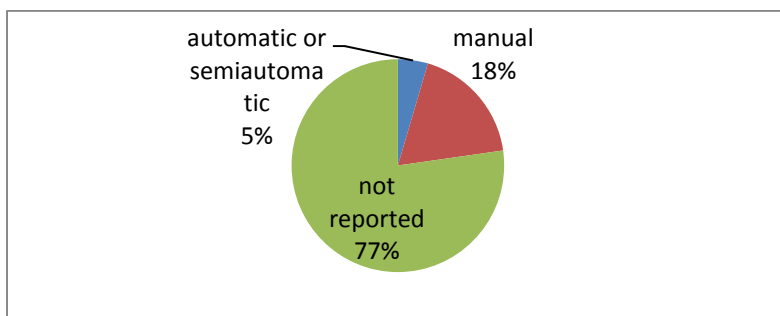
**Table 1** Techniques used for the modules of the CBR and accuracy rate

Ref.	Retrieve	Reuse	Revise	Retain	Analysis	Rate
(1)	Euclidian distance	-	-	-	ten-fold cross-validation	89
(2)	nearest neighbourhood	Dempster-Shafer	Annealing weight	Annealing weight	leave-one-out strategy	81,94
(3)	similarity function	Manual	manual	manual	statistical frequency	93
(4)	Euclidean distance	-	manual	stored case automatical-bility	conditional probability	58
(5)	nearest-neighbor	-	-	Manual	Goodness-of-fit (R2)	81
(6)	Genetic algorithms	-	-	-	3-fold cross validation	92,4
(7)	Global Similarity	-	-	-	statistical frequency	94
(8)	weight set ranked by decision tree	-	-	-	Area under the curve	
(9)	Fuzzy Mathematics and rules		manual	manual	statistical frequency	65,3
(10)	nearest neighborhood	Manual	-	-	statistical frequency	77,5
(11)	probabilistic	-	-	-	local grading	80
(12)	RMA	hierarchical	manual	manual	5-fold cross-validation	
(13)	nearest neighborhood	-	-	-	statistical frequency	95,58
(14)	nearest neighborhood	Neural net	-	-	correlation	96
(15)	cluster	fuzzy	-	-	statistical frequency	83,15
(16)	Bayesian network	-	-	-	holdout	92.17
(17)	Similarity Measurements	Copy	-	-	correction score	76
(18)	weighted clustering method	A stepwise regression	-	-	statistical frequency	98,9
(19)	similarity measure nearest neighborhood	-	-	-	5-fold cross-validation	94,57
(20)	nearest neighborhood	-	-	-	statistical frequency	
(21)	nearest neighborhood	-	-	-	statistical frequency	

Figure 2 shows a frequency diagram of the application domains of the articles analyzed. It is possible to observe that most were made for medical diagnosis. Figure 3 shows a frequency diagram that indicates what and how many techniques used in the review stage, i.e. what adaptation techniques were used in different CBR.



**Fig. 2** Frequency diagram of the application domain



**Fig. 3** Percentages of the different techniques used in the revision phase

### 4.3 Results: Proposal Medical CBR

We see that there are shortcomings in the adaptation mechanism, most studies report that it is manually. In this section, we present an adaptation mechanism to apply in medical CBR.

The proposal includes three phases: pre-processing data applying fuzzy sets, this in order to discretize continuous variables and fill empty spaces. This helps us to create better association rules. Then generate association patterns for cases and make easier adaptive learning process. Finally matching patterns search through unsupervised neural networks.

#### 4.3.1 Pre Processing Data

*Applying fuzzy sets:* The domain knowledge is represented by the values of  $m$  discrete attributes whose value is discretized in many cases, in many others apply fuzzy, so we reduce the search space (5).

*Fill empty spaces:* Another problem when working with medical data are incomplete data. This complicates the recovery and adaptation process. One solution is to apply the algorithm proposed by Henry Tirry Petri Myllymäki and (33).

The domain is represented by the values of  $m$  discrete attributes  $A_1, \dots, A_m$ . An attribute  $A_i$  has  $n_i$  possible values  $a_{i1}, \dots, a_{in}$ . Cases are coded as a vector:

$$C_k = P_k(A_1), \dots, P_k(A_2), \dots, P_k(A_m)$$

Where  $P_k(A_i)$  expresses the probability distribution for attribute values  $A_i$  when the case is  $C_k$ .  $P_k(a_{ij}) = P(A_i = a_{ij} | C_k = 1)$

To calculate these probabilities We can estimate the proportional number of instances of the class  $C_k$  in our case are medical diagnostics. Similarly  $P_k(a_{ij}) = P(A_i = a_{ij} | C_k = 1)$  can be estimated by the occurrences of values  $a_{ij}$  within the class  $C_k$ .

Once filled the empty spaces, association rules are constructed. An adaptation problem in medicine is the extreme specificity of individual cases, the generalized abstract prototypes or classes can support the adaptation process.

Each event has a list of features, which usually contains between 40 and 130 symptoms and syndromes, this means that there are many differences between similar cases, the adaptation cannot consider all. Prototypes are created, which contains the most frequently observed characteristics (34).

Structural adaptation is to change the structure of the solution during the adaptation. The proposal aims to implement such adaptation through the construction of association rules.

### 4.3.2 Construction of Association Rules Using the ART Algorithm

Association Rule Tree Algorithm (ART) builds decision lists that can be viewed as decision trees, makes use of techniques for efficient and scalable association rule mining (35).

Once made the association rules with a certainty factor value above a defined threshold, the association rules are our base cases for training of the neural network that evaluates adaptation.

Considering the algorithm and the suggestions in (36) which is trained back propagation neural network for the adaptation process and the suggestion to use another type of neural networks, our algorithm continues with the recovery by near neighbors and further training of a neural network.

Through near neighbors algorithm compute the similarity between the new case and association rules. After having recovered association rules, we train a neural network to learn adaptation knowledge.

Extracting adaptation knowledge, i.e. the differences between the characteristics of each new case and the characteristics of each retrieved rule and see how these differences affect the diagnosis. The neural network to be used is the Adaptive Resonance Theory that arise in response to two major problems with supervised neural networks: to learn and retain new patterns learned.

### 4.3.3 Adaptive Resonance Theory

The networks that use unsupervised learning approach has not been used for the design of CBR systems (37). ART stands for Adaptive Resonance Theory English, developed by Stephen Grossberg and Gail Carpenter is a network of three layers: input layer, without performing any pre-processing of the input data. Hidden layer. Output layer neurons is a competitive layer.

The input layer and hidden layer always have the same number of neurons. Each neuron in the input layer is a single neuron in the hidden layer, therefore also spoken of as networks of networks ART two layers instead of three. In our case the entrance are the differences in clinical variables  $P(a_{ij})$  and the output are the probabilities of different diagnoses  $P(C_k)$ .

## 5 Conclusions

In this paper a research study on primary studies of CBR in the medical field is presented. Moreover, It is demonstrated once again that the CBR is still applied in many medical situations for various tasks such as diagnosis and treatment (4-6).

Research in CBR applied health sector is growing, but most systems are prototypes and not yet available on the market as commercial products, as perceived in (6). The study confirms that using hybrid CBR with AI techniques makes more easily manage the complexities inherent in the data used in the studies analyzed, and better results are obtained (4).

Moreover, we observe that most have very well defined their case retrieval system, but there is a failure in the mechanism of adaptation (see Figure 3). Most studies don't report adaptation technique and others report that they do it manually. As a result of the review we suggest a proposal to improve adaptation phase; we propose solutions to each of the problems reported to the review of CBR. First fuzzy sets and applying Bayes as part of a data preprocessed. After association rules to have fewer cases. Subsequently using neural networks that learn adaptation rules. The next step is to implement each of the phases of this proposal on data from medical diagnosis or treatment, validate results and suggest improvements. Previous studies have evaluated recovery techniques cases, but studies as the present have not been found in the literature. We hope that other studies like this are made to confirm or refute the results of our analysis.

**Acknowledgments.** This work has been partially supported by the MICINN project TIN 2009-13839-C03-03 (FEDER Support).

## References

1. Chuang, C.L.: Case-based reasoning support for liver disease diagnosis. *Artif. Intell. Med.* 53(1), 15–23 (2011)
2. Petrovic, S., Mishra, N., Sundar, S.: A novel case based reasoning approach to radiotherapy planning. *Expert Systems with Applications* 38(9), 10759–10769 (2011)



3. Ocampo, E., MacEiras, M., Herrera, S., Maurente, C., Rodríguez, D., Sicilia, M.A.: Comparing Bayesian inference and case-based reasoning as support techniques in the diagnosis of Acute Bacterial Meningitis. *Expert Systems with Applications* 38(8), 10343–10354 (2011)
4. Ting, S.L., Kwok, S.K., Tsang, A.H.C., Lee, W.B.: A hybrid knowledge-based approach to supporting the medical prescription for general practitioners: Real case in a Hong Kong medical center. *Knowledge-Based Systems* 24(3), 444–456 (2011)
5. Ahmed, M.U., Begum, S., Funk, P., Xiong, N., von Scheele, B.: A multi-module case-based biofeedback system for stress treatment. *Artificial Intelligence in Medicine* 51(2), 107–115 (2011)
6. Hsu, K.H., Chiu, C., Chiu, N.H., Lee, P.C., Chiu, W.K., Liu, T.H., et al.: A case-based classifier for hypertension detection. *Knowledge-Based Systems* 24(1), 33–39 (2011)
7. Agwil, R.O., Shrivastava, D.P.: Integrated Thallassaemia Decision Support System. *WSEAS Transactions on Computers* 9(8), 857–867 (2010)
8. Huang, M.L., Hung, Y.H., Lee, W.M., Li, R.K., Wang, T.H.: Usage of Case-Based Reasoning, Neural Network and Adaptive Neuro-Fuzzy Inference System Classification Techniques in Breast Cancer Dataset Classification Diagnosis. *Journal of Medical Systems* (in press)
9. Gu, D.X., Liang, C.Y., Li, X.G., Yang, S.L., Zhang, P.: Intelligent technique for knowledge reuse of dental medical records based on case-based reasoning. *Journal of Medical Systems* 34(2), 213–222 (2010)
10. Marling, C., Shubrook, J., Schwartz, F.: Toward case-based reasoning for diabetes management: A preliminary clinical study and decision support system prototype. *Computational Intelligence* 25(3), 165–179 (2009)
11. Toward translational incremental similarity-based reasoning in breast cancer grading. Image Perception, Access and Language IPAL (UMI CNRS 2955, UJF, NUS, I2R), Singapore National University Hospital National University of Singapore Politehnica University of Timisoara, Romania University of Besançon, France Medical Informatics Service, University Hospital of Geneva, Sweden University of Applied Sciences, Western Switzerland, Sierre, Sweden (2009)
12. Rodríguez, S., De Paz, J.F., Bajo, J., Corchado, J.M.: Applying CBR systems to micro array data classification 49, 102–111 (2009)
13. Ahn, H., Kim, K.: Global optimization of case-based reasoning for breast cytology diagnosis. *Expert Systems with Applications* 36(1), 724–734 (2009)
14. Obot, O.U., Uzoka, F.M.: A framework for application of neuro-case-rule base hybridization in medical diagnosis. *Applied Soft Computing* 9(1), 245–253 (2009)
15. Fazel Zarandi, M.H., Zarinbal, M., Izadi, M.: Systematic image processing for diagnosing brain tumors: A Type-II fuzzy expert system approach. *Applied Soft Computing* 11(1), 285–294 (2011)
16. Cruz-Ramírez, N., Acosta-Mesa, H.G., Carrillo-Calvet, H., Barrientos-Martínez, R.E.: Discovering interobserver variability in the cytodiagnosis of breast cancer using decision trees and Bayesian networks. *Applied Soft Computing* 9(4), 1331–1342 (2009)
17. A Reuse-Based CBR System Evaluation in Critical Medical Scenarios (2009)
18. Fan, C.-Y., Chang, P.-C., Lin, J.-J., Hsieh, J.C.: A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. *Applied Soft Computing* 11(11), 632–644 (2009)
19. Lin, R.H., Chuang, C.L.: A hybrid diagnosis model for determining the types of the liver disease. *Computers in Biology and Medicine* 40(7), 665–670 (2010)

20. Influenza Forecast: Case-Based Reasoning or Statistics? (2007)
21. Floyd Jr., C.E., Lo, J.Y., Tourassi, G.D.: Case-Based Reasoning Computer Algorithm that Uses Mammographic Findings for Breast Biopsy Decisions. *American Journal of Roentgenology* 175(5), 1347–1352 (2000)
22. Schmidt, R., Vorobieva, O., Gierl, L.: Case-based Adaptation Problems in Medicine, pp. 267–274 (2003)