Artificial Intelligent support for Medical Diagnosis

Vladimir Kurbalija
Department of Mathematics and Informatics, Faculty of Science,
Trg D. Obradovića 4, 21 000 Novi Sad, Yugoslavia
e-mail:kurba@im.ns.ac.yu

Mirjana Ivanović
Department of Mathematics and Informatics, Faculty of Science,
Trg D. Obradovića 4, 21 000 Novi Sad, Yugoslavia
e-mail: mira@im.ns.ac.yu

ABSTRACT

Multiple Sclerosis (MS) is a chronic disease of unknown cause that affects the nervous system. The process of determining this disease is very complicated because the final diagnosis depends on many parameters and, additionally, the dependencies between parameters are unknown. In these kind of problems, computer aided diagnosis could help medical experts to recognize all important aspects of a problem. The research at our department is focused on the special area of artificial intelligence – Case Based Reasoning (CBR). CBR is an approach where every new problem is solved by adapting the solutions of the previously successfully solved problems. This approach is equivalent to human process of solving problems by reminding the similar situations from the past. This paper is mainly focused on the memory organization and the retrieval algorithm used for this domain and application type.

1. INTRODUCTION

Case-Based Reasoning (CBR) has become successful technique for knowledge-based systems in different domains. This promising technique is based on use of previous experience in form of cases to better understand and solve new problems in particular domain. The main idea is an assumption that, for many particular domains, similar problems usually have similar solutions.

Generally speaking, the CBR systems have not only to provide solutions to problems but also to take care of other tasks occurring when it is used in practice. However, CBR system can do a several tasks depending on the intended use of reasoning:

- Explain current situation according to previously experienced similar situation,
- Critique current situation based on old cases,
- Reason from precedents to understand a current situation,
- Combine old solutions in order to solve a current problem,
- Learn from the successfully solved new problems.

To summarize, mentioned aspects could be divided into two types: interpretative and problem solving CBR. In interpretative CBR systems the essence is to achieve whether or not a current situation should be treated like previous ones based on similarities or differences among them. In problem solving CBR systems the main task is to propose a solution to a current situation based on the adaptation of solutions to past cases. However in practice, many problems have components of both types of CBR and certainly combination of both methods is employed.

Medicine is rather suitable domain for application of CBR because the knowledge of experts consists of mixture of textbook (objective) knowledge and experience (subjective) knowledge, which mostly consists of cases. The problem of updating the changeable subjective knowledge can be easily solved by incrementally incorporating new up-to-date cases. Usually two sorts of knowledge can be clearly separated, and represented in an appropriate manner.

Objective knowledge can be represented in forms of rules or functions, while subjective knowledge is contained in cases. So, there are several obvious facts for usage of case-oriented methods in medicine [12]:

- Reasoning with cases corresponds to the decision making process of physicians,
- Incorporating current cases obtains automatically updated parts of the knowledge,
- Two distinguished types of knowledge can be clearly separated,
- Integration into clinic communication system is easy (cases are routinely stored).

Multiple sclerosis (MS) is a chronic disease of unknown cause that affects the nervous system (brain, spinal cord and peripheral nerves) in form of multifocal lesions of myelin nerve sheet. It causes damage of vision, muscle strength, sensation, coordination, speech, and bladder control and may affect cognitive functions. MS usually appears in young adulthood (more in females) and Vojvodina is region with high incidence of disease.

In this paper, architecture of the database and corresponding retrieval algorithm for CBR diagnostic system is presented. The rest of the paper is organized as follows. In section 2 foundations of CBR technology, are mentioned. Section 3 briefly describes the realized system focusing on the memory organization and retrieval algorithm, while section 4 concludes the paper.

2. FOUNDATIONS OF CASE BASED REASONING TECHNOLOGY

Case-Based Reasoning is a relatively new and promising area of artificial intelligence and it is also considered a problem solving technology (or technique). This technology is used for solving problems in domains where experience plays an important role [1,3,10].

Generally speaking, Case-Based Reasoning is applied for solving new problems by adapting solutions that worked for similar problems in the past. The main supposition here is that similar problems have similar solutions. The basic scenario for mainly all CBR applications looks as follows. In order to find a solution of an actual problem, one looks for a similar problem in an experience base, takes the solution from the past and uses it as a starting point to find a solution to the actual problem.

In CBR systems experience is stored in form of cases. The case is a recorded situation where problem was totally or partially solved, and it can be represented as an ordered pair *(problem, solution)*. The whole experience is stored *in case base*, which is a set of cases and each case represents some previous episode where the problem was successfully solved.

The main problem in implementing almost every CBR system is to find a good similarity measure – the measure that can tell in what extent the two problems are similar. In the functional way similarity can be defined as a function:

$$sim : U \times CB \rightarrow [0 , 1]$$

where *U* refers to the universe of all objects (from a given domain), while *CB* refers to the case base (objects which were examined in the past and saved in the case memory). The higher value of the similarity function means that these objects are more similar [10].

The case-based reasoning system has not the only goal of providing solutions to problems but also of taking care of other tasks occurring when it is used in practice. The main phases of the case-based reasoning activities [1] are described in the *CBR-cycle* (Figure 1).

In the *retrieve* phase the most similar case (or *k* most similar cases), to the problem case, is retrieved, while in the *reuse* phase some modifications to the retrieved case is done in order to provide better solution to the problem (case adaptation). As the case-based reasoning only suggests solutions, there may be a need for a correctness proof or an external validation, so that system will stay consistent in regard to environment. That is the task of the phase *revise*. In the *retain* phase the knowledge, learned from this problem, is integrated in the system by modifying some knowledge containers.

The main advantage of this technology is that it can be applied to almost any domain. CBR system does not try to find rules between parameters of the problem; it just tries to find similar problems (from the past) and to use solutions of the similar problems as a solution of an actual problem. So, this approach is extremely suitable for less examined domains – for domains where rules and connections between parameters are not known.

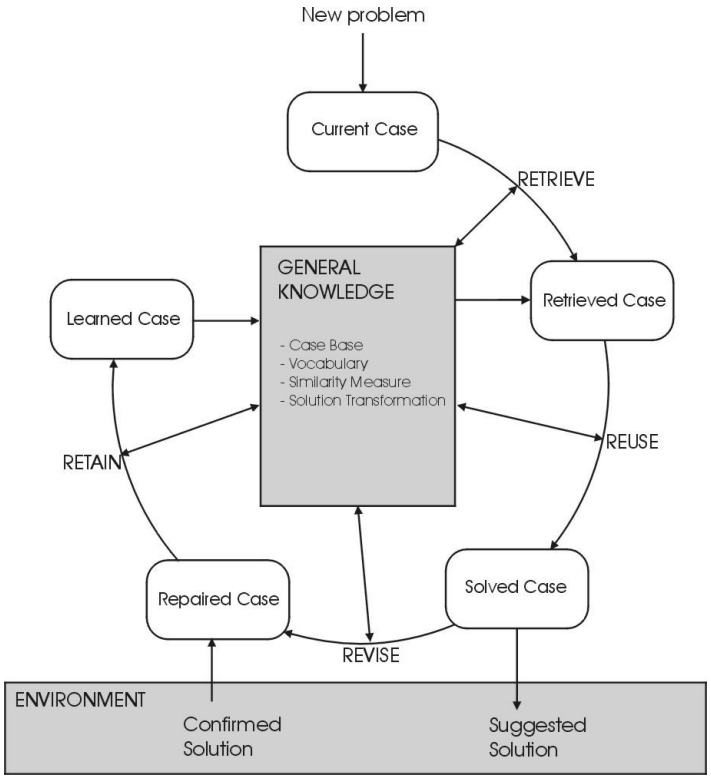

Figure 1. The *CBR-Cycle* after Aamodt and Plaza (1994)

In some extent – medicine is one such domain, especially the diagnoses of not so examined domains, like multiple scleroses. Furthermore, in the more examined domains integration of CBR in classical rule-based reasoning systems brings some efficiency. The second very important advantage is that CBR approach to learning and problem solving is very similar to human cognitive processes – people take into account and use past experience to make future decisions.

3. "CBG" – COMPUTER AID DIAGNOSIS SYSTEM BASED ON CBR

Since summer, 2000 a small group at the Department of Mathematics and Informatics in Novi Sad has been investigating in the area of CBR [2, 5, 6, 7]. During that research one natural demand came out: it would be very useful to have some basic, core system (framework) that can produce systems in different domains especially having in mind application in neurology domain. "CBG" (Casebase Generator) [8] was direct result of these intentions.

3.1. Memory Organization

In this system a special memory structure, called Case Retrieval Net (CRN) [9, 10, 11], is used. Case Retrieval Nets are memory model that has been developed for efficient retrieval in large case bases. This net has nodes for each case and nodes for each value of every feature - these values are called *information entities*. There exists an acceptance arc between every two

corresponding information entities - information entities of the same feature. Also there exists a relevance arc from information entity to the case, if that information entity is relevant for that case. All the arcs in the net are weighted with appropriate values.

At the beginning of the retrieval process, physician has to enter his observations in form of the query, which consists of the values of existing features. For every feature *i* the special numeric value α_i called *importance* is defined. This value contains the information how much is the corresponding feature important for the diagnosis. High values indicate high importance, while lower values indicate lower importance. At the beginning, all importance values have the same default values. Of course, the physician can increase value α_i if he believes that feature *i* has higher importance then some of the others. The physician will decrease the default value of α_i if he considers that the value of the feature *i* is not precise or that this feature is not just so important.

Retrieval process consists of evaluating *acceptance* values for each case. Acceptance values are computed by spreading activation process in the net as follows: Information entity nodes are initially activated by α_i. The computation is performed by propagating activations along the acceptance arcs to further information entity nodes and from all activated information entity nodes over relevance arcs to case nodes. The acceptance value for every case is obtained by summing the activations gained from all information entities trough relevance arcs.

3.2. Characteristics of "CBG"

"CBG" system was completely implemented in Java – JDK 1.3. Java programming language was chosen mainly because it supports all concepts of object-oriented technology, but also because it's main characteristic – platform independence. The system was realized as an application, but the small modifications are necessary in order to make an applet or servlet. Of course, *javax.swing* components are used for creating a graphical user interface (GUI).

The system reads the data from two input files. In the first input file (which is called the "Case Pattern File") the description of the database i.e. the description of the case is stored. Case pattern file contains the information about the structure of every case. There, the list of the attributes, containing the name and the type of the attribute is listed. The type of the attribute can be: *int, float* or *string*. Boolean type can be simulated, for example, with the integer type where only two values (0/1) are allowed. The number of the attributes is arbitrary, but all the attributes in the correct order must be listed.

The second file (which is called the "Case Base File") contains the list of all already solved cases. Every case is described with the values of its attributes and with the final solution of that case. The values of the attributes must be sorted in the correct way according to the order of the attributes listed in the case pattern file. The final solution of the case is always listed at the end of the case (after values of all attributes). Type of the solution can also be : *int, float* or *string* and it is determined dynamically, when all cases are parsed. At this moment it is assumed that the case base file is textual file where every case is listed in one line, and the values of the attributes and solution are separated with commas.

Together with the reading of the case base file, system creates case retrieval net (CRN). The broad structure of CRN is given in the figure 2. Two main parts of the CRN are array of attributes and list of solutions. The array of attributes is created using case pattern file, while the list of the solutions is created from case base file – solution is the last value from every case. Every attribute consists of its name and the list of values. Every value for every attribute represents one information entity because the information entity is an ordered pair *(attribute, value)*. Every value (or information entity) contains the list of arcs to the solution nodes. The arc is given with its weight and a pointer to the solution node; just the arcs whose weights are different from zero are saved. A weight of the arcs between the information entity node e to the solution node c represents the value of the relevance function between *e* and *c*. These weights are calculated as a number of cases (from the case base file) that contain the information entity *e* and whose solution is *c*.

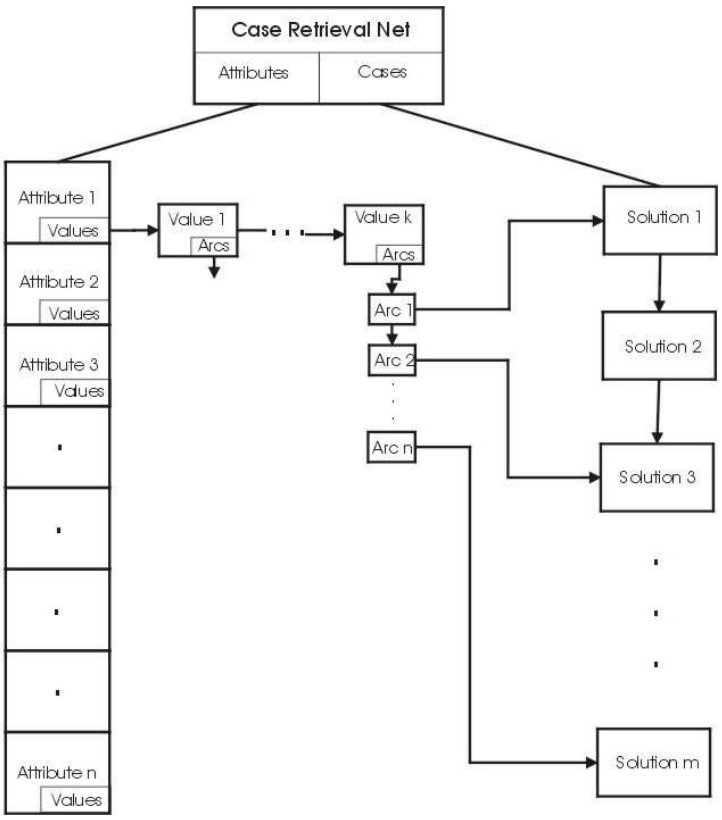

Figure 2. Structure of CRN

After creating the CRN the system expects from the user to enter the current problem in form of a query. Since the query and the case have the same structure (a set of the information entities) the user has to enter the values of some (or all) attributes in a form. In order to better describe the problem, the user should enter all known values of the attributes although it is not necessary. The form contains one more field for every

attribute – importance. The importance is the value from the interval (0,1), and describes how much is the user sure in the validity of the value of the attribute he is entering. Value 1 means that he is 100% sure that the data are valid, while the value 0 means that he doesn't know the value of that attribute at all.

After entering the query, the system searches for the possible solution in the following manner:

> *The information entities (attribute, value) that occur in the query are initially activated with the value of importance. The activation is propagating trough the arcs to the solution nodes, by multiplying the value of the activation of the information entity node and the weight of the corresponding arc. Final activation of the solution nodes is calculated by summing all gained activations.*

3.3. The Application of "CBG" in Multiple Sclerosis Diagnosis

Implemented "CBG" is general tool which can be applied in different domains in order to improve various activities and actions. But our initial intention was to apply it in medical domain, especially in MS diagnoses.

When the "CBG" system was completed, next problem in its application was extracting the data from previous diagnoses of MS. This problem is very complex because the data must be retrieved from the medical histories of the patients from the last 10 years.

By analyzing the diagnoses problem of MS disease the appropriate case structure was defined: Every case is characterized by 72 different features representing the most important observations in the diagnoses process of MS disease. The last feature represents the correct diagnosis for the corresponding cases and can be one of the four values: *Definitive Multiple Sclerosis, Probable Multiple Sclerosis, Possible Multiple Sclerosis* or *No Multiple Sclerosis.*

At this moment the database is modeled, most of the patient files are examined and necessary data are extracting. However, the system is tested [4] using some randomly generated data, but the structure of the data is the same as in the complete version.

4. CONCLUSIONS AND RELATED WORK

"CBG" system was initially designed to be a core system for some future intelligent systems based on CBR technology. Fortunately, very soon it turned out that this system could be successfully used in medical domain. CBR technology is very convenient for diagnosis of multiple sclerosis disease because it suggests solutions just on the basis of some past diagnosis, and not on the basis of some rules, which, in fact, are unknown for this disease.

The core of this system is the Case Retrieval Net. Its good characteristics proved in many domains and areas. Good memory organization and fast retrieval algorithm are the main reasons why it was chosen.

References

[1] Aamodt, A., Plaza, E.: Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches, *AI Commutations*, pp. 39-58. 1994.

[2] Budimac Z., Kurbalija V., *Case-Based Reasoning – A Short Overview*, Proceedings of Second International conference on Informatics and Information Technology "Molika 2001", Molika, Macedonia, December 20-23, 2001, pp. 222-234.

[3] Case-Based Reasoning at AIAI, http://www.aiai.ed.ac.uk/project/cbr/

[4] Ivanović M., Kurbalija V., Budimac Z., Semnic M., *Role of Case-Based Reasoning Technology in Multiple Sclerosis Diagnosis,* draft version.

[5] Ivanović M., Kurbalija V., Budimac Z., Semnic M., *Role of Case-Based Reasoning in Neurology Decision Support,* Proceedings of the Fifth Joint Conference on Knowledge-Based Software Engineering "JCKBSE 2002", Maribor, Slovenia, September 11-13, 2002, pp. 255-264.

[6] Kurbalija, V.: On Similarity of Curves – project report, Humboldt University, AI Lab, Berlin, 2003.

[7] Kurbalija V., Ivanović M., *Case-Based Reasoning – A Powerfull Artificial Intelligence Approach,* Journal of Electrical Engineering 12/s, Slovak Centre of IEE, Vol. 53, 2002, pp.20-24.

[8] Kurbalija V., Ivanović M., *CBG – A Framework for Intelligent Decision Support Systems*, draft version.

[9] Lenz M.: Case Retrieval Nets Applied to Large Case Bases, draft, Lenz's home-page.

[10] Lenz M., Bartsh-Sporl B., Burkhard HD., Wess S.: Case-Based Reasoning Technology: From Foundations to Aplications, Springer Verlag, October 1998.

[11] Lenz M., Burkhard H.D., Bruckner S.: Applying Case Retrieval Nets to Diagnostic Tasks in Technical Domains, Lenz's home-page.

[12] R. Schmidt, L. Gierl, Case-based Reasoning for Medical Knowledge-based Systems, Internet source.