UNIVERSITY OF BRIGHTON

Case-Based Reasoning in medicine

Intelligent Systems (SWM41)

Emmanouil Lainas (12847191)

19/02/2015

1. Introduction

This paper aims to give an overview of the effective application of case-based reasoning (CBR) in the medical domain. It will provide an explanation of what CBR is and why it is considered an ideal approach for the medical sector. It will discuss how the state-of-the-art has evolved through the years since its early introduction and how CBR has been adopted and combined with other artificial intelligence approaches in order to meet certain domain requirements. In addition an investigation of the current state-of-the-art will be attempted and possible pathways of evolution will be discussed.

2. Case-based Reasoning

Case-based reasoning was firstly conceived by Janet Kolodner in the late 1970s based on Schank and Abelson's (1977) work *Scripts, Plans, Goals and Understanding* and on Schank's (1982) *Dynamic Memory*. According to Kolodner (1993) CBR is the attempt of solving new problems based on previous experiences. An example would be the attempt of a medical diagnostician to treat a patient with an unusual combination of symptoms. The diagnostician identifies symptoms and based on his experience he is trying to match the most closely related case to the symptoms observed in order to produce a diagnosis. Old diagnoses may not necessarily be correct but they may produce a plausible answer.

Kolodner (1993) suggests that CBR is a reasoning model that combines learning, understanding and problem solving, and, the three components are integrated with memory processes. Memory is represented as a set of cases which form the case-base. Kolodner's CBR cycle can be described as four stages: case retrieval, case adaptation, solution evaluation and case-base updating.

Because of its different approach towards artificial intelligence, CBR gained reputation among researchers. The motivation behind CBR is that it can utilize specific knowledge as a way of trying to provide a solution to a new problem instead of relying solely on general knowledge of a domain. For example in a rule-based system all the knowledge of the domain must be incorporated in the system, and solutions to problems are drawn based on relationships and conclusions. A CBR system on the other hand will try to solve a problem which is not in the range of its knowledge capabilities by relating it to similar problems (Aamodt & Plaza 1994).

Aamodt and Plaza (Aamodt & Plaza 1994), state in their paper that another important difference of CBR systems is the incorporation of machine learning. As mentioned earlier in the paper the fourth stage of Kolodner's CBR cycle is case-base updating. Whenever a new case is solved, it is then added to the case-base so that it can be reused for future problems thus leading in an incremental sustained learning (Aamodt & Plaza 1994). Although CBR supports incremental learning, they mention that in order to make effective use of it, a well worked set of methods is required as regards knowledge extraction, case integration and case indexing for later matching.

As CBR was spreading to the world different methods regarding organisation, retrieval, utilisation, and indexing of the knowledge were introduced. This has led to some confusion because while systems were still using the CBR paradigm as a basic underlying model, new terms were introduced to describe types of more specific approaches. In their paper, Aamodt and Plaza (Aamodt & Plaza 1994) attempt to clarify between the terms related to CBR as follows:

- CBR methods that address the learning of concept definitions are sometimes referred to as *exemplar-based*.
- Specialization of exemplar based reasoning into highly syntactic CBR approach is referred to as *instance-based*. A non-generalization approach to the concept learning problem (machine learning methods).
- *Memory-based* emphasises on a collection of cases as a large memory, and reasoning as accessing and processing this memory.
- *Case-based* approach is assumed to have a degree of richness of information and complexity with respect to its internal organization. Characterised by the ability to modify a retrieved solution when applied in different problem solving context.
- Methods that solve new problems based on past cases from different domain are referred as *analogy-based*. Major focus of study is the reuse stage of the CBR cycle.

Aamodt and Plaza (1994) introduced a descriptive framework for describing CBR methods, which consists of two complementary models listed below:

- A process model of the CBR cycle with the following steps: RETRIEVE the most similar cases, REUSE information form the retrieved case to solve the problem, REVISE the proposed solution and finally RETAIN parts of the experience which can be useful in the future (usually referred as the four R's).
- A task-method structure for case-based reasoning (see Aamodt & Plaza 1994).

In the paper written by David W. Ada "The omnipresence of case-based reasoning in science and application" (Ada 1998), the author states that in order to introduce CBR, one must identify it in a particular context. For that reason, when CBR is introduced to artificial intelligence researches a variant of the Aamodt and Plaza's process model (Aamodt & Plaza 1994) is frequently used. This model is usually referred as the five R's and consists of the following steps: RETRIEVE, REUSE, REVISE, REVIEW, RETAIN. As we can see a new REVIEW step is included in the cycle. The REVIEW step's purpose is to evaluate the outcomes when applying the constructed solution to the current problem. This seems like a natural step if we consider how we would want an autonomous intelligent agent to behave. Ideally the agent should be able to assess the effect of applying that solution to the problem. For example let us consider a prescription generation system. In many cases wrong prescriptions can lead to severe damage or even death. The agent should be aware of these effects so that it can immediately exclude such a case from the possible solutions.

3. CBR in Medicine

In 1976 Edward Shortliffe (Shortliffe 1976) produced some work on diagnosis of infection diseases and ever since Artificial Intelligence has been widely used in the medical domain. More specifically CBR was suggested as an interesting alternative for the medical domain from Phyllis Koton back in 1988 (Koton 1998) when she created the CASEY system, a heart failure detection system. Following the next year, Bareiss (Bareiss 1989) described a system named Protos capable of classifying new cases by explaining similarities to known exemplars. Protos purpose was to serve as a computational model of psychological theories of concept learning and classifications. Additionally Protos can be used as a tool for constructing knowledge-based systems (Bareiss 1989).

After that, CBR has been further established in the field of Artificial Intelligence and more specifically in the medical domain (Nilson & Sollenborn 2004).

In their paper, Nilson and Sollenborn (2004) accurately state that a point of attraction for CBR in medicine relays on the nature of the domain. In the medical domain, patients and diseases are considered as cases having several characteristics.

As other artificial intelligence approaches do, this too has its advantages and disadvantages (Gierl & Schmidt 1998). Some of the advantages of CBR in medicine are the following:

- Cognitive adequateness: CBR systems are able to mimic the reasoning approach used by diagnosticians.
- Explicit experience: CBR systems can adjust themselves to specific requirements of certain medical domains.
- Combination of both Objective and Subjective knowledge: CBR systems are built upon existing cases rather than using the subjective knowledge of domain experts.
- Automatic acquisition of subjective knowledge: Since CBR systems support incremental knowledge acquisition that can be abstracted by generalizing cases.
- System Integration: Patient records are being stored on machine readable mediums by health organisations thus case-data can be easily collected and used, although, in most cases modification of the data is required in order to integrate them with a CBR system.

Disadvantages of CBR in medicine as listed by Nilson and Sollenborn (2004):

- Adaptation: Because of the complexity of the medical domain adaptation becomes an issue. Schmidt and Gierl (2000) state that generalization and efficient feature identification methods partly deal with this issue but the problem persists.
- Unreliability: The incremental learning supported by CBR systems will not necessarily make the system more reliable. Although the system will extend the range of coverage of the problem domain reliability is not guaranteed.
- Concentration on reference: CBR systems rely upon similar cases rather than underlying diagnostic factors. If a similar case does not exist, the system cannot function as a source of previous experience.

4. Overview of advances and trends in medical CBR

Nilson and Sollenborn (2004) produced a paper in which they discuss trends and advancements in CBR systems. They based their work on a 1998 survey produced by Gierl and Schmidt (Gierl & Schmidt 1998). In their paper, Nilson and Sollenborn (2004), focus on systems reported after 1998. They divide system properties into two categories:

1. Purpose-oriented, which is characterised by the general action the system must perform.
2. Construction-oriented, which is identifying different types of construction.

Purpose-oriented properties consist of the following system types:

- Diagnostic systems: Systems that attempt to provide assistance in the diagnosing process of a medical condition, possibly to the point of autonomous diagnoses.
- Classification systems: Systems that attempt to identify a group of real world cases such us image classification systems.
- Tutoring systems: Systems that are built around the concept of learning by examples, typically by providing access to real patient cases.
- Planning systems: Systems that attempt to help in solving a process that constitutes a number of steps for example therapy support.

Construction-oriented properties consist of the following construction types:

- Hybrid systems (often referred as multi-modal): Systems that are constructed using two or more artificial intelligence paradigms in order for the one to compensate for the weaknesses of another. Examples include systems that use CBR as the main organiser of data, and data intensive techniques like artificial neural networks (ANN) to handle lower-level identifications.
- Adaptive systems: Systems capable of adapting previous solutions to solve the current problem. Schmidt and Gierl (2000) describe the adaptation problem as a result

of the enormous amounts of features in a case.

- Case library size: This deals with the amount of cases stored in the case library as well as the degree of case generalization into prototypes. In other words, the ability of the system to categorize a more generalized case by merging existing cases.
- Autonomicity (although autonomicity is not a correct English term, the term has been widely used to describe these kinds of systems. A more accurate term could be autonomy): This involves the degree of autonomy of a system. The degree of autonomy is the need of human interaction during the reasoning cycle of a system. Ideally a fully autonomous system would produce diagnoses without the need of a medical expert to evaluate them.
- Constraints: System constraints deal with reliability and self-criticality. If we recall the five R's cycle (Ada, 1998) mentioned earlier in the paper this relates to the REVIEW step of the cycle.

Nilson and Sollenborn (2004) provide a description of CBR medical systems that are created during the years 1998-2004 (see Nilson & Sollenborn, 2004). They mention that the selection of papers used is highly subjective but still, certain trends are distinctive enough to deserve mentioning. Their observations of trends in CBR at that point are the following:

1. Majority of systems are multimodal.
2. Only one system utilized adaptation.
3. Prototype generalization was rare (however the intention is to extend the systems with prototypes).
4. Most systems are dependent on some level of user interaction.
5. Few systems have been commercialised (but typically they are kept on research level).
6. Safety and reliability constraints are not to common and systems that do have safety constraints depend on operational reliability.

As regards construction-oriented trends the following has been observed:

1. Hybrid systems constitute the majority of medical CBR systems. This was probably expected since as we said earlier a combination of paradigms hinders the disadvantages of using an exclusive paradigm.
2. The autonomy of systems is relatively low. As Nilson and Sollenborn (2004) state, not relaying on complete autonomy appears to be sound if we consider the inherent problem of unreliability in CBR.
3. Prototyping through case aggregation seems to be a commonly intended extension.

5. More recent work

Bichindaritz and Montani (2011) produced a paper describing the advances in case-based reasoning in the health sciences up to 2011. In their paper they describe that early CBR systems in biomedicine were driven by the early goals of artificial intelligence to represent experts reasoning. They add that all the early systems developed in CBR in medicine tried to model medical expertise along the main medical tasks: diagnosis, treatment planning and follow-up. Later CBR was applied to a wider range of tasks consisting of: diagnosis, classification tasks, treatment planning, assessments test planning, image analysis, long-term follow-up, quality control, tutoring and research assistance.

The authors explain that one of the reasons CBR has found one of its most fruitful applications in biomedicine is because as we said earlier CBR relays mostly on cases. The fact that biology and medicine belong in the family of descriptive experimental sciences where knowledge is derived from studies of natural phenomena, patient problem situations and more results in vast amounts of cases.

Nevertheless the application of CBR in biomedicine has several complexities which are inherent from the nature of the domain. For example the high-dimensionality of cases in biomedicine, the long term follow-up, multimedia case representation and the development of appropriate CBR methods to process these, increase the rate of complexity in CBR systems. Additionally the co-occurrence of several diseases, the obscurity of diagnostic categories, sensor signals and more are other factors that extend the complexity.

Bichindaritz and Montani (2011) observed a major trend towards the widening of applications of CBR beyond the traditional diagnosis, treatment and quality control, to the applicability of CBR to novel reasoning tasks. They mention an example of a system which studies how cases can represent non-compliance instances of clinical guidelines, and eventually lead to expert refinement of these guidelines. They suggest that recent papers open new fields for application for CBR and will enhance the spread of CBR in biomedical domains which sounds like a valid argument since the application of CBR in new fields will draw the attention of new researchers.

In a more recent paper, Bichindaritz et al. (2014) discuss six outstanding research papres in CBR in health sciences (CBR-HS). The papers described present lessons learned from CBR-HS research and methodological advances. The advances discussed involve adaptation models and missing data management. In addition the papers address new approaches to dealing with time series data, medical work flows, and sensor signals.

The adaptation model discussed is introduced by Henriet et al. (2014). This method uses an interpolation tool to provide solutions to unknown problems by adapting known solutions from other problems already solved and stored in the case base. The interpolation-based adaptation model is based on ANN. The results show that the adaptation model meets the requirements of radiation protection experts. In addition the accuracy of adapted solutions mostly depends on the accuracy of the solutions furnished by the source cases. The paper identifies potential areas for improvement as future research directions.

Guessoum et al. produced the RESPIDIAG system. RESPIDIAC is a decision support system for the diagnosis of a very serious respiratory disease caused by tobacco use. The system is based on cased-based reasoning principles and gathers the experience of experts of the pulmonology department of Dorban Hospital, in Annaba, Algeria. Several methods for dealing with missing values in the case-base or the target case are proposed.

New approaches to dealing with time series data are proposed by Huang et al. (2014). Their paper studies the prediction of inpatient length of stay (LOS). LOS is valuable as an indication of hospital activity and of resource consumption. Because unexpected scenarios can easily occur in hospitals LOS predictors should deal with adaptability considering the temporal evolution of the patient. The system uses CBR to predict an inpatient LOS by using past traces of clinical treatment processes having similar medical actions. The approach was evaluated and shown to be successful using 248 patient traces from pulmonary infection clinical treatment processes extracted from Zhejiang Huzhou Central Hospital of China.

6. Conclusions

Case-Based reasoning has been a widely established field in the area of cognitive sciences and artificial intelligence. Because CBR combines experiential learning and solving in the similar way in which humans solve problems it introduces a different approach towards Artificial Intelligence which is based on Memory. As we mentioned in the paper CBR has several advantages and disadvantages. In order to remedy the issues raised within CBR, it should be better used in comparison with other AI approaches.

References

Aamodt, A. and Plaza, E. (1994) 'Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches', *AI Comminications,* 7(1), pp. 39-59.

Aha, W.D. (1998) 'The Omnipresence of Case-Based Reasoning in Science and Application', *Knowledge-based Systems,* 2(5-6), pp. 261-273.

Bareiss, E. R. 1989. *Exemplar-Based Knowledge Acquisition: A unified Approach to Concept Representation, Classification and Learning*. 300 North Zeeb road, Ann Arbor, MI 48106-1346: UMI.

Bichindaritz, I., Marling, C. and Montani, S. (2014) 'Preface for the Special Section on Case-Based Reasoning in the Health Sciences', *Expert Systems with Applications*, 41(2), pp. 247-248.

Bichindaritz, I. and Montani, S. (2011) 'Advances in Case-Based Reasoning in the Health Sciences', *Artificial Intelligence in Medicine*, 51(2), pp. 75-79.

Gierl, L., and Schmidt, R. 1998. Cbr in medicine. *Case- Based Reasoning Technology, from Foundations to Applications* 273–297.

Guessonn, S., Laskri, M. T. and Lieber, D. (2014) ‘RESPIDIAG: A Case-Based Reasoning System for the Diagnosis of Chronic Obstructive Pulmonary Disease’, *Expert Systems with Applications,* 41(2), pp.267-273.

Heuriet, J., Leni, P.E., Laurent, R. and Salomon, M. (2014) ‘Case-Based Reasoning Adaptation of Numerical Representation of Human-Organs by Interpolation’, *Expert Systems with Applications*, 41(2), pp. 260-266.

Huang, Z., Juarez, J.M., Duan, H. and Lee, H. (2014) ‘Length of Stay Prediction of Clinical Treatment Process Using Temporal Similarity’, *Expert Systems with Applications,* 41(2), pp. 274-283.

Kolodner, J. (1993) *Case-Based Reasoning*, USA: Morgan Kaufmann Publishers.

Koton, P. (1988) *Using Experience in Learning and Problem Solving*, Massachusetts: MIT Press.

Nilsson, M. and Sollenborn M. (2004) ‘*Advancements and Trends in Medical Case-Based Reasoning: An Overview of Systems and System Development*’ FLAIRS Conference.

Schmidt, R., and Gierl, L. 2000. Case-based reasoning for medical knowledge-based systems. In *German Workshop on Experience Management*, 720–725. Proceedings of MIE’00 and GMDS’00.

Shortliffe, E.H. (1976) *Computer-Based Medical Consultations: MYCIN*, New York: American Elsevier Publishing Company.