Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 35 (2008) 262–272
Expert Systems with Applications
www.elsevier.com/locate/eswa

Two stages of case-based reasoning – Integrating genetic algorithm with data mining mechanism

Heng-Li Yang *, Cheng-Shu Wang

Department of Management Information Systems, National Cheng-Chi University, 64 Section 2, Chihnan Road, Mucha Dist., 116 Taipei, Taiwan

Abstract

Case-based reasoning (CBR) is a paradigm, concept and instinctive mechanism for problem solving. Recently, CBR has been widely integrated with some AI algorithms and applied to various kinds of problems. The ill-defined and unstructured problems are suitably solved by CBR. This research proposes a hybrid CBR mechanism including two stages. In stage I, the genetic algorithm is adopted to improve efficiency of case retrieving process. Compared to traditional CBR, the proposed mechanism could reduce about 14% case evaluations, but still achieved 90% satisfactory results. In stage II, the knowledge discovering and data mining (KDD) processes are implemented to produce the refined information from the retrieved cases. Because these retrieved cases and target problem satisfy similar or even same conditions, the outcome of KDD would be more valuable for reference. In addition to retrieved cases of stage I, the proposed mechanism provides direction and relevant knowledge for decision makers in their decision supporting and revising processes in stage II. The proposed CBR mechanism also deals with efficiency and outcome quality issues of traditional CBR.

Keywords: Case-base reasoning; Case retrieve; Genetic algorithm; KDD

1. Introduction and related works

Case-based reasoning (CBR) is a paradigm, concept and instinctive mechanism for problem solving. Similarly to human problem solving process, CBR retrieves past experiences to reuse for target problem. Of course, the solutions of past cases may need to be revised for applying. The successful problem-solving experiences are then retained for further reusing (Aamodt & Plaza, 1994). These are well-known 4R processes (retrieve, reuse, revise, and retain) of traditional CBR.

For recently decade, CBR has been applied to various kinds of problem solving. The ill-defined and unstructured problems are suitably solved by CBR (Belecheanu, Pawar, Barson, Bredehorst, & Weber, 2003). Some applications are as follows:

- Construction industry: Cirovic and Cekic (2002) applied CBR to support construction project in preliminary design phase. They retrieved pervious cases in historical IPDB (Integrated Project DataBase), and then stored in CKB (Construction Knowledge Base) for quality designing. Juan, Shin, and Perng (2006) integrated CBR with genetic algorithm (GA) for housing designing. They devoted to customization housing layout design.
- Diagnosis system: Guiu, Ribe, Mansilla, and Fabrega (1999) and Golobardes, Llora, Salamo, and Marti (2002) both used CBR for breast cancer diagnosis from Mammary Biopsy Image and micro calcifications separately. Chang (2005) applied CBR to screen children with developmental delay. In practice, these diseases can be early detected from symptoms and increase the cure rate.
- SCM/CRM supporting: Choy and Lee (2002) used CBR to select suppliers for outsourcing. They evaluated cases (i.e., suppliers) according to technical capability, quality

* Corresponding author. Tel.: +886 2 29387651; fax: +886 2 29393754.
E-mail addresses: yanh@nccu.edu.tw (H.-L. Yang), 93356506@nccu.edu.tw (C.-S. Wang).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.06.027

assessment and organizational profile. Further, Choy, Fan, and Lo (2003) extended the above research to CRM and developed intelligent customer-SRM system (ISRMS) by CBR. Cheng and Cheng (2004) predicted yarn qualities from cotton fibers properties by CBR. In past, these problems usually solved by mathematic model, e.g., regression. But, there are too many hypotheses to apply in real world. CBR seems to be another good option to deal with these kinds of problems.

- Strategic planning: Changchien and Lin (2005) produced marketing plans for marketing personnel of Telecommunication Company. Through CBR they analyzed situations, e.g., external environmental audit, industry audit, etc., and produced key marketing policies, including product, price, place, and promotion. Lim, Ahn, and Lee (2005) applied CBR to formulate strategies for competing stakeholder management. After stakeholder analyzing, RDAP (reactive, defensive, accommodative, or proactive) strategies are retrieved, revised, and implemented by CBR mechanism.
- Other applications: Belecheanu et al. (2003) referred to past records to improve uncertain information, e.g., industrial requirements, for new product development especially in the concurrent engineering approach. Gwen (2003) suggested the application of CBR to knowledge management. Choobineh and Lo (2005/2006) intended to improve database conceptual design by CBR.

For all above mentioned applications, CBR is easier to understand, implement and outperform than others compared algorithms.

In practices, there is some well-developed software to facilitate CBR concept implementation, including the commercial packages (such as KATE, Spotlight, and ESTEEM) and free software (such as CBR-Works and CASPIAN). For illustration, in the CBR-Work 4.0 user Interface as Fig. 1, decision makers can set conditions of target problem and then submit to system. CBR-Work will provide similar case(s) for the references of decision makers.

However, traditional CBR has some problems. Traditional CBR has to evaluate all cases in the case-base to return the most similar case(s). The efficiency of CBR is obviously negatively related to the case-base size. Thus, a number of approaches have devoted to decrease the effort for case evaluation. For illustration, K-means is the most popular. K/CBR, CBR integrated with K-means approach, clusters all cases first and only evaluates the case of the most similar cluster for case retrieving. In addition, Chang and Lai (2005) attempted self-organizing map (SOM). In their research, SOM/CBR outperformed than K/CBR. Both K/CBR and SOM/CBR improve the efficiency of CBR. However, the performances of these two revised CBR mechanisms are closely related to case representation and indexing approaches (Shin & Han, 1999). Hence, their superior performances are not stable and can not be promised.

CBR has been widely integrated with other mechanisms and applied to various territories. Marling, Sqalli, Rissland, Hector, and Aha (2002) surveyed various CBR integrations. Some systems, e.g., ADIOP, CADRE, CADSYN, CHARADE, COMPOSER, IDIOM, JULIA, and Weigel's, integrated CBR with constraint satisfaction problem (CSP) algorithm. Some systems, e.g., ANAPRON, AUGUSTE project, CABARET, CAMPER, GREBE, GYMEL and SAXEX, combined CBR with rule-based reasoning (RBR) mechanism. There are also hybrids of CBR and model-based reasoning (MBR), e.g., CASEY, FORMTOOL and SOPHIST. These integrations could improve and enhance traditional CBR.

There are some researches integrating CBR with other AI techniques. Genetic algorithm (GA) seems to be a good option for CBR efficiency improving. As mentioned before, Juan et al. (2006) combined CBR with GA for housing layout customization. In their proposed approach, GA facilitates the solution revision. In addition, Shin and Han (1999) used GA to support CBR for enhancing classification accuracy. In their research, a weight vector is determined by GA to improve case indexing and retrieving process. Finally, Kim and Han (2001) applied GA to determine case representation and indexing for case

Attributes	Query (Vacation)	1	2
Accommodation	?	TwoStars	TwoStars
Code	?	1	2
Destination	?	Egypt	Egypt
Duration	?	14	21
HolidayType	?	Bathing	Bathing
Hotel	?	Hotel White House, Egypt	Hotel White House, Egypt
Persons	?	2	3
Price	?	2498.0	3066.0
Season	n.n.	April	May
Transportation	?	Plane	Plane

Fig. 1. The user interface of CBR-works.