Information Sciences 287 (2014) 50–60

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Memory and forgetting: An improved dynamic maintenance method for case-based reasoning

Aijun Yan *, Limin Qian, Chunxiao Zhang

College of Electronic Information & Control Engineering, Beijing University of Technology, Beijing 100124, China

A R T I C L E I N F O

Article history:
Received 21 March 2014
Received in revised form 23 July 2014
Accepted 25 July 2014
Available online 4 August 2014

Keywords:
Case-based reasoning
Memory
Forgetting
Dynamic maintenance

A B S T R A C T

The problem solving performance of a case-based reasoning (CBR) system is closely related to the quantity and quality of the cases stored in the case base. With the continuous growth of the size of the case base, the so called "swamping problem" may occur when the time cost of retrieval exceeds the benefit of the accuracy. From the perspective of cognitive science, a dynamic maintenance method improved by selective memory and intentional forgetting for CBR is proposed, which can imitate the memory function of the human brain to selectively save new cases, update the *forgotten* values and intentionally delete the old cases. The experiments show the effectiveness of the proposed method. The selective memory and international forgetting policy can significantly reduce the time and space complexity, and retain or improve the accuracy of the CBR classifier, thus improving the performance of CBR.

1. Introduction

Based on the cognitive hypothesis of "similar problems have similar solutions", CBR can solve new problems by retrieving similar cases from the case base [1]. It has been applied in forecasting [6,25], business management [4], and classification [19], etc. The traditional wisdom in case-based system has been that, the more cases stored in the case base, the more likely that similar cases may be retrieved, and the more accurate the problem solving becomes. However, the so called "swamping problem" may occur [16] when the time cost of retrieval exceeds the benefit of the accuracy. In addition, harmful and redundant cases existing in the case base may also affect the overall performance of the CBR system [17]. Therefore, in order to improve the overall performance of the system, the establishment of a suitable case base maintenance method is urgently needed which can control the growth rate of case base while maintaining a high accuracy of problem solving.

Many case base maintenance methods have been proposed, and these can be divided into efficiency-oriented methods [3,18] and performance-oriented [11,22–24] methods. For the former, a simple policy is *random deletion*. Once the size of the knowledge base exceeds a predefined limit, a random item is removed from the knowledge base. Surprisingly, in certain circumstances, the apparently naive random deletion can work very well and can be as effective as more sophisticated methods [3]. One more sophisticated approach is Minton's utility metric. He used a utility metric to delete the invalid knowledge, thereby controlling the growth of the knowledge base indirectly [18]. These efficiency-oriented methods can prevent efficiency degradation in problem solving, however the neglect of the quality of problem solving may result in the removal of useful cases, which may leading to the failure of the problem-solving. For the latter, Smyth and others [23,24] established

* Corresponding author.
E-mail address: yanaijun@bjut.edu.cn (A. Yan).

http://dx.doi.org/10.1016/j.ins.2014.07.040
0020-0255/

a case competence model, in which according to their different effects on the solving ability and efficiency of the CBR system, the cases were classified into pivots, auxiliary, spanning and supporting, and the partition of case competence can be used to guide the selection of cases for deletion, which can guarantee a smaller system size and prevent the degradation of its ability to solve problems. Inspired by the case competence model, some researchers use the clustering methods to guide the case addition and case deletion [11,22], which can reduce the cost of case retrieval as well as maintain the competence of the case base. However, these performance-oriented methods may cost a lot of time to establish the case competence model or to partition the case base, and incorrect classification of the cases will directly affect the effectiveness of deletion strategy, therefore an overall performance of CBR becomes limited. In order to improve the solving efficiency of CBR, Fayed and Atiya applied a new template redundant approach [12] to maintain the case base off line, and with the reduced case base, the classification time can be reduced. But similar to most of the above methods, this approach also spends much additional time on reducing the case base in advance, and thus it is hard to satisfy the real-time requirements. Therefore, it is necessary to further study the maintenance strategy for CBR.

Based on the correlation between CBR and cognitive science, this paper introduces selective memory and intentional forgetting into the CBR cycle. Based on the *forgotten* value of the cases, the improved CBR system can selectively save the new cases, and intentionally remove the old cases in real time. The dynamic storage and deletion can guarantee the learning and updating of case base meanwhile control its growth rate, so it will play a positive role in improving the accuracy and efficiency of the whole CBR system in the long run.

The structure of this paper is as follows. Section 2 introduces the improved CBR model with memory and forgetting functions. The details of the dynamic maintenance method based on selective memory and intentional forgetting are given in Section 3. The effectiveness of the proposed method is analyzed in Section 4 through the contrast experiments using the UCI classification data sets. The final section includes the conclusion and an outline for future work.

2. Traditional CBR cycle

There are many models which strive for a better description of CBR, among which the most widely used is the "*4R*" cycle [1] proposed by Aamodt and Plaza as shown in Fig. 1. The traditional "*4R*" includes the following four steps:

Retrieve the most similar cases.
Reuse the retrieved conclusion to solve the new problem.
Revise the suggested conclusion.
Retain the new problem and the corrected solution as a new case.

Since CBR is an incremental learning method, with the continual saving of new cases, the size of the case base is increasing too. Although the possibility of retrieving similar cases is increased, the subsequent "swamping problem" can lead to lower efficiency of case retrieval, and at the same time, redundant cases and harmful cases in the case base may not only increase the complexity of the case retrieval space, but may also damage the accuracy of case retrieval. Therefore, it is necessary to find a construct case base maintenance method to improve the overall performance of CBR.

3. Memory and forgetting strategies

Since cognitive science forms one the pillars of CBR, in this section we study the learning process of CBR from the perspective of memory and forgetting, the main content is to obtain an expanded CBR cycle by improving the traditional

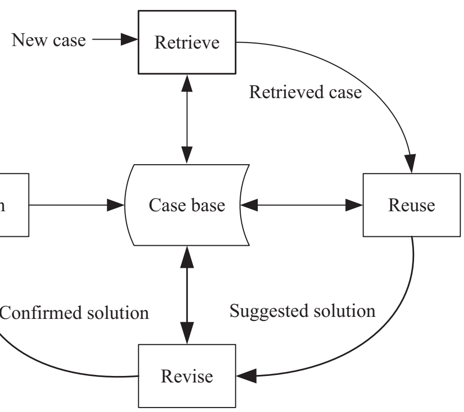

Fig. 1. Traditional CBR cycle.