Expert Systems with Applications 42 (2015) 8006–8013

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Trustworthiness evaluation and retrieval-based revision method for case-based reasoning classifiers

Aijun Yan [a,b,c], Dianhui Wang [d,*]

[a] College of Electronic Information & Control Engineering, Beijing University of Technology, Beijing 100124, China
[b] Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China
[c] Beijing Key Laboratory of Computational Intelligence & Intelligent System, Beijing 100124, China
[d] Department of Computer Science and Information Technology, La Trobe University, Melbourne, VIC 3086, Australia

ARTICLE INFO

Article history:
Available online 2 July 2015

Keywords:
Case-based reasoning classifiers
Classification accuracy
Case evaluation
Case revision

ABSTRACT

To achieve better classification performance using case-based reasoning classifiers, we propose a retrieval-based revision method with trustworthiness evaluation for problem solving. An improved case evaluation method is employed to evaluate the trustworthiness of the suggested solution after the reuse step, which will divide the target cases and its suggested solutions into a trustworthy set and an untrustworthy set in accordance with a threshold value of trustworthiness. The attribute weights are adjusted by running a genetic algorithm and are used in the second round of retrieval of the untrustworthy set to obtain the classification results. Experimental results demonstrate that our proposed method performs favorably compared with other methods. Also, the proposed method has less computation complexity for the trustworthiness evaluation, and enhances understanding on thinking and inference for case-based reasoning classifiers.

1. Introduction

Case-based reasoning (CBR) is a problem solving and learning method originating from cognitive psychology, which imitates the cognition process of human beings. The CBR process mainly includes case retrieval, case reuse, case revision and case retention steps (Aamodt & Plaza, 1994). So far, CBR has been successfully applied to engineering design (Guo, Peng, & Hu, 2013), fault diagnosis (Yan, Wang, Zhang, & Zhao, 2014), bankruptcy prediction (Chuang, 2013), control and decision making (Yan, Chai, Yu, & Xu, 2012), emergency preparedness (Liao, Mao, Hannam, & Zhao, 2012) and more, demonstrating good potential for real world applications. However, in solving pattern classification problems, most CBR-based work has been done on case retrieval, case reuse and case retention, with little attention being given to the case revision step, which renders the learning process of CBR a negative learning method and will affect its performance. Moreover, according to the cognition mechanism of human beings, there must be an evaluation standard before the revision step, which comes down to the case evaluation problem. So far, many studies on case retrieval, reuse and retention in the CBR pattern classification tasks have been conducted, whereas case evaluation and revision have not been addressed sufficiently in the existing work. Therefore, an in-depth study on evaluation and revision of the case solutions is necessary.

In the existing studies, the case evaluation strategies mainly focused on the confidence assessment of the suggested solution (Chua & Tischer, 2004; Fan, Li, & Zhang, 2015; Garcia, Oroaco, Gonzàlez, & Arcos, 2007; Liao et al., 2012; Muhlbaier, Topalis, & Polikar, 2009). Chua and Tischer (2004) proposed a trustworthy region concept to realize the result evaluation. Garcia et al. (2007) proposed a reuse strategy in conducting evaluation by calculating the trustworthiness of the target case belonging to each class. Muhlbaier et al. (2009) introduced a consult and vote strategy to assess the confidence of the classification result. Fan et al. (2015) let multiple project managers or experts analyze the related information and give their judgments according to their knowledge and experience. In addition, according to the cognition mechanism of human beings, case revision is usually processed by evaluating the suggested solution, reanalyzing the differences between the target case and the source case, then adjusting the solution repeatedly. For example, Kim, Lee, Woo, and Shin (2012) used the results of the multiplication for each attribute to revise the construction cost of the previous case retrieved with its case similarity score. Qi, Hu, and Peng (2012) proposed a new adaptation method for the solution feature values of the retrieved cases

* Corresponding author.
E-mail addresses: yanaijun@bjut.edu.cn (A. Yan), dh.wang@latrobe.edu.au (D. Wang).

http://dx.doi.org/10.1016/j.eswa.2015.06.027
0957-4174/

by introducing the adaptability value to improve the adaptation performance. In addition, other revised techniques based on classical methods have been developed, such as expert experience (Fan et al., 2015; Petrovic, Mishra, & Sundar, 2011; Yan and Wang et al., 2014; Yan et al., 2012), genetic algorithm (GA) (Liao et al., 2012), multiple regression analysis (Jin, Cho, Hyun, & Son, 2012), interpolation tool (Henriet, Leni, Laurent, & Salomon, 2014), regression revision model using support vector machine (Han & Cao, 2015), grey relational analysis (Hu, Qi, & Peng, 2015) and so on. In the problem solving process using CBR, to some extent, these methods have made significant contributions to the evaluation and revision of the suggested solutions for the target cases from different perspectives. However, it can be seen that the existing methods have some limitations in practical applications. For example, (a) the assessment and/or revision method by experts' knowledge is not suitable in data-based pattern classification tasks; (b) in the trustworthiness evaluation strategy, there is no follow-up process for the case results considered unsuitable or bad, which does not improve the overall performance of the CBR classifier effectively; (c) this kind of evaluation strategy also has a high computational complexity; and (d) the repeated revision strategy will cause a huge loss in the reasoning efficiency of the CBR, especially where case revision is barely taken into consideration due to the difficulties in constructing the revision link (Kaedi & Ghasem, 2012). However, from the cognitive science perspective, case revision reflects the ability of logical thinking and creative thinking. In the pattern classification problem, if there is no revision stage in the problem solving process using CBR, the goal of classification accuracy is usually difficult to achieve. As stated above, to improve the performance of the CBR classifier, it is necessary to further study the trustworthiness evaluation strategy and case revision.

The objective of this paper, taking case revision as the main problem, is to develop a second round retrieval-based revision method for the suggested solution for the target case. In the proposed method, firstly, a case evaluation link is added between case reuse and case revision, and an improved trustworthiness evaluation strategy is used to assess the trustworthiness of the suggested reuse solution. Subsequently, the target cases are divided into a trustworthy set and an untrustworthy set according to a threshold value and the trustworthiness of the suggested solution. Then, the retrieval-based case revision strategy will be deployed on the untrustworthy set, that is, the GA is used to adjust the attribute weights and second round retrieval is used to revise the suggested solution in the untrustworthy set. Finally, the classification results are obtained.

The rest of this paper is organized as follows. Section 2 provides a brief overview of the CBR system; Section 3 details the improved case evaluation strategy and the revision strategy. The experiments and results are given in Section 4 and the conclusions of this study and the directions for future research are presented in Section 5.

2. Related work

2.1. Case-based reasoning classifiers

In order to construct a CBR classifier $\Phi: F \rightarrow C$, F represents the feature space and C is the category set. According to the CBR model introduced by Aamodt and Plaza (1994), the CBR classifier could be constructed using the following steps:

(i) Represent and retrieve: suppose the source case in the case base could be expressed in the two-tuple form:

$$H_i : \langle X_i; Y_i \rangle, \quad i = 1, 2, \ldots, m \tag{1}$$

where m denotes the total number of source cases, X_i, Y_i represent the description and the classification result of the ith source case, X_i is represented as follows (Bergmann, Kolodner, & Plaza, 2005):

$$X_i = (x_{i1}, x_{i2}, \ldots, x_{ij}, \ldots, x_{in}) \tag{2}$$

where n is the number of the attributes and x_{ij} represents the value of the jth attribute in the ith source case. Recording the target case as $T = (t_1, t_2, \ldots, t_j, \ldots, t_n)$, t_j denotes the value of the jth attribute in target case T. Let the classification result corresponding to target case T be Y_T. In order to get Y_T, a similarity calculation strategy based on Euclidian distance is adopted to calculate the similarities between target case T with every source case in the case base (Liao, Zhang, & Mount, 1998):

$$\begin{cases} s_i = s(T, X_i) = 1 - D(T, X_i), \quad i = 1, 2, \ldots, m \\ D(T, X_i) = \sqrt{\sum_{j=1}^{n} \omega_j (t_j - x_{ij})^2} \\ s.t. \sum_{j=1}^{n} \omega_j = 1, \quad \omega_j > 0 \end{cases} \tag{3}$$

where $s_i \in [0, 1]$ is the similarity between X_i and T, $D(T, X_i)$ is the Euclidian distance between X_i and T, t_j denotes the value of the jth attribute in target case T, $\omega_j (j = 1, 2, \ldots, n)$ represents the weight of the jth attribute, basically the mean weight, which regards every attribute as having the same importance. From (3), m similarities s_1–s_m could be achieved. These are sorted in descending order and the K-nearest neighbor (KNN) method (Lin & Chen, 2011) is adopted to get the first K corresponding source cases and the classification results.

(ii) Reuse and retain: according to the K similar cases, the most similar reuse strategy ($K = 1$) or the majority solution reuse strategy ($K > 1$) could be utilized to achieve the suggested solution for target case T (Lin & Chen, 2011). If $K = 1$, reuse the solution for the case that has the closest similarity to the suggested solution; if $K > 1$, then adopt the majority solution for the K similar cases as the suggested solution for target case T. In the process of transferring the suggested solution into classification solution Y_T, the traditional CBR classifier usually does not undertake the revision step, but directly uses the suggested solution as classification solution Y_T. Therefore, restore target case T and classification solution Y_T in the case base and, in the end, the number of the case bases will increase from m to $m + 1$, which realizes the incremental learning process of the CBR classifier.

2.2. Problem statement

From the description of the CBR classifier above, when the suggested solution is directly reused as the classification result of the target case without the revision step, the two following situations may exist that could affect the classification accuracy of a CBR classifier.

(a) Target case T is closest to the source case X, but the distance between the two cases $D(T, X)$ is still too far and the similarity between the two cases is small, which would lead to an untrustworthy reuse solution.
(b) Target case T lies in the overlapping zone of the source case with different categories, meaning that the reuse solution would be untrustworthy.

If any of the situations above occurs, the suggested solution is untrustworthy and should be revised. Before conducting the case revision, an evaluation discipline is needed to judge the reliability of the solution so as to guide the direction of the following revision step. So far, the commonly used evaluation strategies include: trustworthy region partition-based evaluation (TRE) (Chua & Tischer, 2004), trustworthiness evaluation (TE) (Garcia et al., 2007), confidence assessment (Muhlbaier et al., 2009), etc. Of