Engineering Applications of Artificial Intelligence 65 (2017) 212–219

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

An attribute difference revision method in case-based reasoning and its application

Aijun Yan [a,b,c,*], Kuanhong Zhang [a,c], Yuanhang Yu [a,c], Pu Wang [a,c,d]

[a] School of Automation, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[b] Beijing Key Laboratory of Computational Intelligence & Intelligent System, Beijing 100124, China
[c] Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China
[d] Beijing Laboratory for Urban Mass Transit, Beijing 100124, China

ARTICLE INFO

Keywords:
Case-based reasoning
Attribute difference
Case revision
Support vector regression

ABSTRACT

To improve the learning performance of the revision stage in case-based reasoning (CBR), an attribute difference revision method (ADR) is proposed in this paper. First, the suggested solution of the target case is obtained through the case retrieval and case reuse; then, the revision value of the suggested solution and output results of the CBR model are obtained by using the support vector regression (SVR) model, which is based on the difference between the target case and similar cases; finally, the target case and its correct solutions are stored. Experiments and applications shows that the ADR method is effective and the fitting error of the ADR-based CBR (ADRCBR) model is significantly lower than other typical regression methods, indicating that ADR can improve the learning performance of the CBR model and has the advantage of application.

1. Introduction

Case-based reasoning (CBR), which was proposed by Aamodt and Plaza (1994), is a type of intelligent reasoning method that guides action based on past experience. The process conducted by CBR can be described as a classic cycle model, namely, case retrieval, case reuse, case revision and case retention (4R). This method has been used in different areas, such as product design (Hu et al., 2015), pattern classification (Yan and Wang, 2015; Fan et al., 2014), regression prediction (Yan et al., 2015; Han and Cao, 2015a), intelligent control (Xing et al., 2012) etc., and has achieved remarkable application success. The goal of CBR application research is to improve the learning performance (Wang and Yang, 2012; Fan et al., 2015) and the accuracy of the problem solving (Zhong et al., 2015). In order to achieve these goals, the task of case revision is to correct the inaccurate suggested solution obtained from the case reuse phase. The case revision method is both the difficulty and the key in the CBR model, so a successful case revision method not only can improve the performance of CBR, but also is meaningful for artificial intelligence to solve practical engineering problems (Kaedi and Ghasem-Aghee, 2012).

From the perspective of cognitive science, the revision link reflects the logical and creative thinking of human beings. When the suggested solution is not suitable for the target case, the modifying process is called case revision, which has always been a difficult issue (Shiu et al., 2001). Therefore, many researchers tend to avoid this part when conducting application research (Jalali and Leake, 2016). Until now, expert experience and machine learning are the main methods for case revision. For instance, the expert experience method when used for case revision is to a certain extent subjective and unsuitable for a data-driven model (Fan et al., 2015; Petrovic et al., 2011; Yan et al., 2012); a genetic algorithm (GA) is used to assign weights and integrate them into the revised formula (Kim et al., 2012). GA has a certain learning ability, but there is the possibility of premature convergence in the training process. By using this method, it is difficult to deal with and optimize high dimensional problems, while the stability and reliability are also not good enough. Some research using a multiple regression model (MRA) (Jin et al., 2012) and group decision making (Yan et al., 2014) has shown some achievements in applications. When SVR is introduced into case revision the reuse phase gets the revision model by training similar cases, puts the target case into the model and outputs the suggested solution, which has a very good regression ability (Han and Cao, 2015b). However, it is not clear what the difference is between the target case and the similar case when using all these methods, which will affect the objectivity of the revision solution and the learning ability of the CBR model; in addition, from the cognitive perspective of CBR, it is difficult

* Corresponding author at: School of Automation, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China.
E-mail address: yanaijun@bjut.edu.cn (A. Yan).

http://dx.doi.org/10.1016/j.engappai.2017.07.015
Received 21 August 2016; Received in revised form 24 March 2017; Accepted 16 July 2017

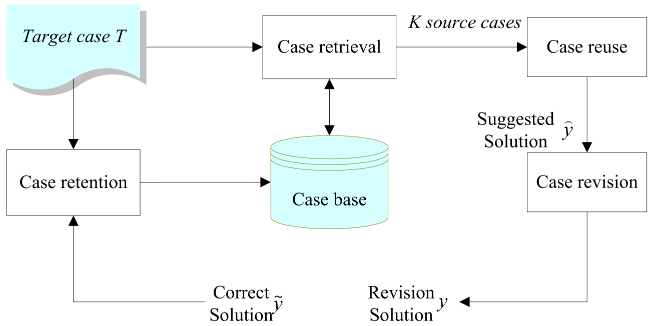

Fig. 1. Framework of traditional CBR model.

to improve the creative thinking ability of CBR if ignoring the differences in conditions of problem solving. Therefore, it is very necessary to carry out further research on the case revision link.

In order to improve the learning ability of revision method, this paper proposes an attribute difference revision method (ADR) and an ADR-based CBR (ADRCBR) model. The case base of attribute difference revision is built by the difference between a new case made by case retrieval and a number of source cases, and is trained by SVR to obtain a revision model. Based on that, the feature attribute difference between the new case and the target case is entered into the SVR revision model and the revision value of the suggested solution can be found. Finally, the paper uses the suggested solution plus the revision value to get the revision solution. The results of comparison experiments that include the typical regression data set and the concentration prediction of biochemical oxygen demand (BOD) were able to prove that the proposed method is effective.

The rest of this paper is organized as follows. Section 2 introduces the traditional CBR regression models and problems analysis. Section 3 describes the structure and algorithm steps of ADRCBR, and the algorithm convergence is analyzed. Section 4 presents the performance test on the proposed algorithm, and the comparison experiments are also designed. Section 5 introduces an example of application. In the final section, the conclusions and future research are presented.

2. Traditional CBR model

2.1. Solution process of CBR

The structure of a traditional CBR model is shown in Fig. 1, including case representation, case retrieval, case reuse, case revision and case retention. A brief introduction to the functions of each step is given as follows.

(1) Case representation: Common methods of case representation include the property characteristic values description method, the frame representation method and the object-oriented method (Bergmann et al., 2005). Of these, the property characteristic values description method has been widely used (Wang and Yang, 2012; Fan et al., 2015). Each source case C_k ($k = 1, 2, \ldots, p$) can be expressed as follows Aamodt and Plaza (1994):

$$C_k : \langle X_k ; Y_k \rangle, \quad k = 1, 2, \ldots, p \tag{1}$$

where p is the total number of cases; X_k denotes the feature attributes set in the kth source case record; Y_k denotes the solution of the decision attribute in the kth source case record. Assuming that there are n feature attributes in each source case, X_k can be expressed as follows:

$$X_k = \left(x_{1,k}, \ldots, x_{i,k}, \ldots, x_{n,k}\right) \tag{2}$$

where $x_{i,k}$ is the normalized value of the ith feature attribute in the kth record.

(2) Case retrieval: traditional CBR retrieval uses KNN (k-nearest neighbor), and then the K most similar historical cases to the target case are obtained (Cover and Hart, 1967).

(3) Case reuse: The suggested solution $\hat{y}$ is obtained by calculating the average value of the K history case solutions.

(4) Case revision: Revise the suggested solution $\hat{y}$ and obtain a revised solution y (Aamodt and Plaza, 1994; Shiu et al., 2001).

(5) Case retention: The target case and the correct solution $\tilde{y}$ are combined into a new case to be stored in the case base.

2.2. Problem analysis

In the traditional CBR model, there is no general revision method, so the revision phase of the suggested solution has been ignored. It is obvious that case revision can improve the performance of the model. Research on case revision is mainly concerned with expert experience and machine learning methods.

Some researchers use expert experience to revise the suggested solution, which is suitable for a situation where it is easier to obtain revision rules (Fan et al., 2015; Petrovic et al., 2011; Yan et al., 2012). However, for multi-feature attribute cases, using this method may mean that revision rules are difficult to define and a combined explosion could occur.

Kim uses GA to get the weight W_i of feature attributes (Kim et al., 2012). Then the weights are used in the revision formula as follows:

$$y = \hat{y} + \hat{y} \sum_{i=1}^{n} \left\{ W_i \left(\frac{V_{new\_case}}{V_{retrieved\_case}} - 1 \right) \right\} \tag{3}$$

where y is the revision solution, $\hat{y}$ is the suggested solution, V_{new_case} is the feature attribute of the target case, $V_{retrieved\_case}$ is the feature attribute of similar cases, and n is the number of feature attributes. GA has a good learning performance, but the training process may lead to premature convergence, and encoding process of GA is complicated. In addition, it is difficult to deal with and optimize for high dimensional problems when using GA, and the stability and reliability of GA are also not good enough.

The multiple regression analysis is carried out according to Jin et al., and then the regression coefficient as the revision coefficient is used in the equation as follows (Jin et al., 2012):

$$y = \hat{y} + \sum_{i=1}^{n} UC_i AE_i \tag{4}$$

where y is the revision solution, $\hat{y}$ is the suggested solution, UC_i is a regression coefficient, AE_i is the feature attribute difference between the target case and similar case. This method takes into account the differences between the target case and the retrieved case, but it cannot be used to build a revision model based on the differences. In this method, the revision value is obtained by the vector product of UC_i and AE_i. It is obvious that the coefficients obtained from the multiple regression analysis may not be completely applicable to the revision link, which could reduce the generalizability of the method.

A regression model is obtained by using SVR to train the n similar cases obtained by using the KNN strategy (Han and Cao, 2015b). The target case serves as the input of the model, and the output is the suggested solution. This method takes advantage of SVR for small samples. However, by combining the case reuse with the case revision into a single step, this method does not consider the difference between the target case and the similar cases, and ignores the learning ability of case revision.

In previous studies, the group decision model was introduced into the revision link (Yan et al., 2014). The equation is as follows:

$$y = \frac{1}{m} \sum_{j=1}^{m} \frac{y_j}{\lambda_j} \tag{5}$$

where y is the revision solution, m is the number of similar cases, y_i is the solution of the similar case, and λ_i is expert authority. This method