The Multi-Level Case Retrieval Model by Integrating Case-Based Reasoning, Grey Relational Degree and Decision Support System

Hui Li* and Jie Sun

School of Business Administration, Zhejiang Normal University, 91 Subbox in P.O. Box 62, YingBinDaDao 688, Jinhua City 321004, Zhejiang Province, P.R. China

Abstract

Under the condition of multiple levels of management case, a new retrieval model of case-based reasoning (CBR) was proposed by integrating the classical mechanism of case retrieval, grey relational degree, and decision support system. Basic working flow of the new case retrieval model in CBR was firstly analyzed. On the precondition of multiple levels of management case structure, stratification retrieval strategy and delaminating structure of cases were then set up. On their basis, multi-level case retrieval and selection model in decision support system was finally constructed. In the retrieval model, layer similarity is employed to calculate the integrated similarity between a pair of cases. Grey relational degree is employed in the approach to improve performance of case retrieval. At last, multi-level case similarity of decision group could be reckoned. Empirical experiment results indicated that the new retrieval model achieves a little better performance than traditional models based on Manhattan distance and Euclidean distance.

Key Words: Case-Based Reasoning, Multi-Level Case Structure, Grey Relational Degree, Decision Support System

1. Introduction

In 1982, case-based reasoning (CBR) was firstly proposed by Schank, which is more like a methodology [1,2]. Thus, various methods or techniques, e.g., decision techniques and grey system theory, can be absorbed into CBR. CBR has got broad applications in various domains, e.g., solution of general questions, legal cases, diagnosis and medication [3,4]. CBR is very applicable in situations without solid theories and models while there is enough domain knowledge. When CBR is applied into management area, there always exits the concept of delaminating structure in feature space of management cases. Taking the case of CBR-based marketing plans system in Ref. [5] as an example, a concept case is composed of case name, case company, status analysis, objective and marketing mix. Meanwhile, there are three sub-features in the feature of case company, i.e. company name, business target, and market position, four sub-features in the feature of market mix, i.e. price, product, channel, and promotion, respectively. In fact, there are often two levels of cases when CBR is applied into management area.

Whatever area CBR is applied into, the commonly employed algorithm for case retrieval is the technique of Euclidean distance or Manhattan distance. For example, Jo et al. and Yip respectively used the classical Euclidean distance to calculate case similarity when they applied CBR into the area of financial distress prediction [6,7]. Lee also employed the classical CBR algorithm with rule-based reasoning to solve the problem of internal audit of bank [8]. Waheed & Adeli utilized the classical Euclidean distance to carry out similarity calculation when they applied CBR into the area of steel bridge en-
gineering [9]. Actually, Watson has already argued that CBR is a methodology not a technology [10]. Hence, any technique or approach, including decision technique and grey system theory, could be absorbed into CBR to enhance its performance. We wonder whether or not the classical case retrieval method could be enhanced by combining other techniques into it, even with just a little improvement. On the basis of the similarity calculation, we attempt to build up a multi-level case retrieval model to solve management problems by integrating the technique of decision support system (DSS) and grey relational degree in grey system theory. Grey system theory was put forward by the Chinese researcher of Prof. Deng J.-L. [11]. One of the core techniques of grey system theory is the calculation of grey relational degree. The contribution of this research is that the classical CBR, grey relational degree, decision support system are combined to build up a multi-level case retrieval model by taking into account the characteristic of delaminating structure in feature space of management cases. To some extent, it is a kind of soft CBR which is described in Ref. [12].

The breakdown of rest of the paper is organized as follows. Section 2 outlines processes and basic working flow of CBR by integrating it with decision support system. On the foundation of the introduction of the assumption of case delaminating structure, stratification retrieval strategy, layer similarity, integrated similarity and grey relational degree, Section 3 constructs multi-level case retrieval strategy and the hybrid model of CBR in management decision. Delaminating structure of cases is firstly set up, on the basis of which, layer similarity and integrated similarity are calculated, and the similarity of decision group can finally be reckoned. Grey relational degree is integrated into similarity computation between a pair of cases. Section 4 carries out a comparative empirical experiment between the new enhanced retrieval models and the traditional models based on Manhattan distance and Euclidean distance, followed by a brief discussion in Section 5.

2. Basic Flow of CBR by Integrating It with Decision Support System

In general, complex decision processes are practiced without adequate information. H. A. Simon’s theory of bounded rationality [13] described the limitation of human’s subjective cognition and circumstance’s complexity. That means it is difficult for decision makers to obtain rounded information at the very beginning. By taking expertise as a guide, the imperfection of information is gradually cleaned up. At the same time, more information is continuously fed back into the decision process to obtain a satisfactory solution. By combining the calculation of grey relational degree into the CBR system, a new case retrieval model can be constructed. The new hybrid CBR model is attempted to be used in decision support problems in management area. The basic flow of CBR by integrating it with decision support system is shown as Figure 1. Thus, decision technique is integrated with CBR to form the new hybrid case retrieval model.

Step 1. Through the intelligent man-machine interface, decision makers could draw out descriptions on target cases. Whether the drawing out of eigenvector is reasonable or not, it will directly influ-

![Figure 1. Basic flow of CBR-based decision support system in management area.](image-url)
ence the retrieval result of stored cases. Thus, this procedure is very important for the output of the system.

Step 2. Depending on eigenvectors of the target case, the system based on CBR will compute the grey relational degrees of target case and each stored case. Similarities, including layer similarity and final similarity, between target case and each stored case are calculated. A comparison of corresponding numerical values between similarities calculated and stored thresholds in knowledge base is made. As a result, corresponding codes gather of stored cases which are similar to target case is formed. The system based on CBR can retrieve out solutions from result base (RB) to make an adjustment of the result. After this procedure, the final solution of target case is formed. Case retrieval strategy and models are the key factors.

Step 3. If the final solution of target case meets the requirement of decision makers, then turn to Step 4, else turn to Step 1. In this situation, decision makers can make new descriptions on target case through the intelligent man-machine interface. Thus, the solution of target case can be achieved by another round of retrieval.

Step 4. Through the intelligent man-machine interface, final solution of target case can be fed back to decision makers to support decision processes in real-world management problems.

Step 5. According to case’s learning strategy and models stored in knowledge base, corresponding target case can be updated into case base, result base and knowledge base. Thus, a self-learning process is realized.

3. Multi-Level Case Retrieval Strategy and Models

When applying CBR to support decision problems, case retrieval models and strategy are very crucial. To solve the current decision problem, it needs to make a description on target case’s eigenvector. In this sector, we focus on case retrieval process. In the area of management science, corresponding cases are usually complex. Thus, case delaminating is necessary. Delaminating structure of cases and stratification retrieval strategy of similar case are respectively proposed in this section. On the basis of them, layer similarities are set up to compute integrated similarity of target case and each stored case. Grey relational degree is integrated in similarity calculation approach. Finally, similar cases are retrieved based on similarities of decision makers. This is the way to construct the new hybrid multi-level retrieval model of CBR.

3.1 Case Formalization

Domain experts’ knowledge and experiences can be denoted by characteristics and attributes. A multi-element group denotes as $A = \{E, C, W, R, O\}$. In this group, $E$ is a limited gather denoting description of a case, e.g., name of a case, and type of a case. $C$ is a non-empty limited gather denoting characteristic information of a case. $W$ is a non-empty limited gather denoting expert’s preference on case’s eigenvector. $R$ is a non-empty limited gather denoting results of cases. $O$ is a limited gather denoting relevant knowledge of cases. When applying CBR to support decision problems, one of the key factors is similarity calculation between cases. Assume that there are $n$ attributes of a case, which denotes $C = \{c_i\} (i=1, 2, \ldots, n)$. A case’s eigenvector is divided into key attribute $d_j$ and common attribute $f_k$. The eigenvector gather of a case denotes $F = \{d_j, f_k\} (j=1, 2, \ldots, m; k=1, 2, \ldots, t)$. $G(c_i, d_j)$ and $G(c_i, f_k)$ respectively denote the corresponding membership degree of $d_j$ and $f_k$ of a case $c_i$. Then, the corresponding eigenvector of a case $c_i (i=1, 2, \ldots, n)$ denotes as $H_{c_i} = \{G(c_i, d_j), G(c_i, f_k)\}$. Target case’s eigenvector denotes as $Q_{c_0} = \{G(c_0, d_j), G(c_0, f_k)\}$. All gathered cases for decision-making can be stored in case base according to the established storing strategy.

3.2 Case Delaminating Structure of Common Attributes

There are often two levels of cases when CBR is applied into management area. Common attributes of a case can be classified as $p$ types, and each type is a sub-group of the gather of case attributes, denoting as $E_k (k = 1, 2, \ldots, p)$. Assume that there are $q_k$ common attributes of a case in $E_k$. The delaminating structure of common attributes of a case is shown as Figure 2. Please refer to Ref. [5] for a comprehensible instance of case’s delaminating structure of common attributes. In this research, we mainly focused on concept model of case’s
3.3 Stratification Retrieval Strategy of Similar Cases

The retrieval strategy of similar cases has an important influence on case retrieval. Main retrieval policies of CBR are parallel retrieval strategy and serial retrieval strategy. Currently, the common used retrieval methods are nearest neighbor algorithm, induction methods and knowledge guided indexing. Each retrieval method has its own merits. It is not an ideal way to use just one single method. We believe it is more appropriate to use a stratification retrieval strategy combined by nearest neighbor algorithms and knowledge guided indexing. That is the foundation of the multi-level case retrieval model by integrating the classical CBR, grey relational degree and decision support system. The model of the case’s stratification retrieval strategy is shown as Figure 3.

Primary retrieval

Knowledge guided indexing is employed in primary retrieval. According to key attributes appointed by domain specialists, system can make primary retrieval. For example, decision type is a key attribute appointed by domain specialists, which includes investment, financing, dividend distributing, early warning etc. If the facing decision type is early warning, all cases that relates to early warning should be retrieved in the process of primary retrieval. Cases of other decision type should be ignored automatically to form a case gather of early warning.

Advanced retrieval

The algorithm of k nearest neighbors is used in advanced retrieval. On the basis of the definition of grey relational degree, grey relative relational coefficient, and similarity of target case and each stored case, the final case gather can be achieved. Based on case delaminating structure of common attributes, we use layer similarities and integrated similarities to describe similarity between a pair of cases. Detailed contents are described in the following section. The new multi-level case retrieval model is used in the procedure of advanced retrieval.

3.4 Layer Similarity and Integrated Similarity

After primary retrieval, case eigenvector can be simplified as

\[ H_{ci} = \{ G(c_i, f_k) \} \ (i = 1, 2, \ldots, n; k = 1, 2, \ldots, t). \]

All profitable attributes are divided by maximum values of the column to get a relative membership degree. At the same time, all cost attributes are firstly divided by maximum values of the column, and then subtracted by 1. It denotes

\[ G(c_i, f_k) = \begin{cases} \frac{G(c_i, f_k)}{max(G(c_i, f_k))} & \text{if } G(c_i, f_k) \text{ is profitable attributes} \\ 1 - \frac{G(c_i, f_k)}{max(G(c_i, f_k))} & \text{if } G(c_i, f_k) \text{ is cost attributes} \end{cases} \]

After the process, primary case gather denotes as \( H_{ci} = \{ G'(c_i, f_k) \} \). Target case denotes as \( Q_{c_0} = \{ G'(c_0, f_j) \} \). Grey relational efficient of target case and stored cases can be calculated by the following way.

\[ s_{dk} = \inf(G(c_0, f_k) - G(c_i, f_k)) + k \sup(G(c_0, f_k) - G(c_i, f_k)) \]

\[ |G(c_0, f_k) - G(c_i, f_k)| + k \sup(G(c_0, f_k) - G(c_i, f_k)) \]

In which, \( s_{dk} (i = 1, 2, \ldots, n; k = 1, 2, \ldots, t) \) is grey relat-
tional efficient for the $k$th attribute between target case and the $i$th stored case in primary case gather. $\inf (G'(c_0, f_k) - G'(c_i, f_k))$ is the minimum distance value between the $k$th attribute between target case ($c_0$) and all stored case in primary case gather. $\sup (G'(c_0, f_k) - G'(c_i, f_k))$ denotes a maximum distance value. $|G'(c_0, f_k) - G'(c_i, f_k)|$ is the distance on the $k$th attribute between target case ($c_0$) and the $i$th stored case. $k$ is in the range of [0, 1]. In general, the value of 0.5 is often taken. Thus, grey relational degree is integrated in the calculation of case similarity.

Obviously, $s_k$ is in the range of [0, 1]. Weights of case attributes are denoted as $W = \{w_k\}_{k=1}^t$. In the sub-gather attributes of cases, $E_k' (k' = 1, 2, \ldots, p)$, there are $q_k$ common attributes. Weights of case attributes in $E_k'$ can be denoted as $W_k' = \{w_k\}_{k=1}^{q_k}; k' = 1, 2, \ldots, p$. Layer similarities between target case ($c_0$) and the $i$th stored case can be defined based on distance formula.

\[
sim_{i} = 1 - \left\{ \sum_{k'=1}^{p} \left( w_{k'} \left( 1 - s_{k'} \right) \right)^{\frac{1}{h}} \right\}^{\frac{1}{h}}
\]

\[
\begin{align*}
\text{if } h &= 1 \text{ then Hamming distance is employed} \\
\text{if } h &= 2 \text{ then Euclidean distance is employed}
\end{align*}
\]

(3)

Based on delaminating structure of common attributes, layer similarities can be filled into a cube, which denotes as Figure 4. This figure gives a further description of the way the integration of case delaminating and CBR.

4. Comparative Empirical Experiment and Discussion

Empirical experiment was carried out under the environment of financial early warning decision. Total sample number is 216 and seven common attributes are selected from various financial ratios in this experiment. Data is collected from open information of Chinese listed companies from the Shenzhen Stock Exchange and Shanghai Stock Exchange. Each time one case is regarded as target case and the others are regarded as stored cases. This is consistent with the strategy of leave-one-out cross-validation. It means that leave-one-out cross-validation was employed. A comparative experiment was carried on for comparison between the new model, which integrates grey relational degree, and decision technique into the calculation of similarity in CBR, and traditional models based on Manhattan distance and Euclidean distance. Traditional models denote as follows.

\[
SIM_{\text{TM}}^{(w)}(c_p, c_q) = \frac{1}{1 + d_{pq}^{(w)}}
\]

\[
= \sqrt{\left( 1 + \alpha \sum_{j=1}^{n} w_j^p (x_{p,j} - x_{q,j})^2 \right)^{\frac{1}{2}}}
\]

\[
\begin{align*}
\text{if } h &= 1 \text{ then Hamming distance is employed} \\
\text{if } h &= 2 \text{ then Euclidean distance is employed}
\end{align*}
\]

Among which, $sim_i$ is grey relational coefficient, which we name as integrated similarity, of target case ($c_0$) and case ($c_i$); $h$ is a distance parameter. The bigger the value of $sim_i$, the more similar target case is to stored cases. We can find that grey relations degree is combined into the calculation of case similarity.

In advanced retrieval based on grey relational degree, case delaminating structure of common attributes, layer similarities, and integrated similarities, the system firstly compares $sim_i$ with threshold appointed by domain experts and then retrieves all cases above threshold to form a candidate case gather.

4. Comparative Empirical Experiment and Discussion

Empirical experiment was carried out under the environment of financial early warning decision. Total sample number is 216 and seven common attributes are selected from various financial ratios in this experiment. Data is collected from open information of Chinese listed companies from the Shenzhen Stock Exchange and Shanghai Stock Exchange. Each time one case is regarded as target case and the others are regarded as stored cases. This is consistent with the strategy of leave-one-out cross-validation. It means that leave-one-out cross-validation was employed. A comparative experiment was carried on for comparison between the new model, which integrates grey relational degree, and decision technique into the calculation of similarity in CBR, and traditional models based on Manhattan distance and Euclidean distance. Traditional models denote as follows.

\[
SIM_{\text{TM}}^{(w)}(c_p, c_q) = \frac{1}{1 + d_{pq}^{(w)}}
\]

\[
= \sqrt{\left( 1 + \alpha \sum_{j=1}^{n} w_j^p (x_{p,j} - x_{q,j})^2 \right)^{\frac{1}{2}}}
\]

\[
\begin{align*}
\text{if } h &= 1 \text{ then Hamming distance is employed} \\
\text{if } h &= 2 \text{ then Euclidean distance is employed}
\end{align*}
\]

Among which, $sim_i$ is grey relational coefficient, which we name as integrated similarity, of target case ($c_0$) and case ($c_i$); $h$ is a distance parameter. The bigger the value of $sim_i$, the more similar target case is to stored cases. We can find that grey relations degree is combined into the calculation of case similarity.

In advanced retrieval based on grey relational degree, case delaminating structure of common attributes, layer similarities, and integrated similarities, the system firstly compares $sim_i$ with threshold appointed by domain experts and then retrieves all cases above threshold to form a candidate case gather.

4. Comparative Empirical Experiment and Discussion
When the first case acts as target case, similarities calculated by the four models respectively based on original Euclidean distance, hybrid Euclidean distance, original Manhattan distance, and hybrid Manhattan distance are as Figures 5–8.

The coefficient, $k$, for calculating grey similarity was set as 0.5. In order to obtain optimal similarity thresholds for the four case retrieval mechanisms, the technique of grid search was employed to search optimal similarity thresholds. Finally, similarity thresholds of the four models are respectively set as 98%, 97%, 96% and 94% of the maximum similarity, which are obtained by exhaustive searching in the range of [0.6, 1] with the step of 0.01. There are different mechanisms of similarity calculation in Euclidean distance, hybrid Euclidean distance, Manhattan distance, and hybrid Manhattan distance. Thus, the optimal similarity thresholds are not the same for the four models. After 216 times of calculations for each model under the assessment of leave-one-out cross-validation, the number of correct predictions, wrong predictions and their corresponding accuracies are listed in Table 1.

From Table 1, it can be found that CBR algorithms which integrate grey relational degree in the calculation of similarity produce a little better performance than traditional models based on Manhattan distance and Euclidean distance on the whole. It is indicated that hybrid Euclidean method improves CBR’s performance by 0.46% than traditional Euclidean method. Meanwhile, hybrid Manhattan method improves CBR’s performance by 1.85%. We also find that this type of improvement is under the assessment of leave-one-out cross-validation,
which is unbiased in performance estimation. Though there are no significant improvements on performance of the CBR system between the classical retrieval methods and the new grey retrieval methods, we believe the new ones are still acceptable since there are also no significant differences in computing complexity between classical ones and new ones. Taking into consideration both the characteristic of unbiased estimation of the assessment and there are no apparent difference in computing complexity between traditional ones and the new ones, it is acceptable that the two new methods only produce a little better performance than traditional ones.

Obviously, accuracies of models based on Euclidean distance and Manhattan distance are different values, because there are different calculating mechanisms employed in the two approaches. It means that there may be opposite judgment on a target case if both models based on Manhattan distance and Euclidean distance are employed to make a contrastive analysis. On the contrary, accuracies of models based on hybrid Euclidean distance and hybrid Manhattan distance are almost the same value in this experiment. It means that new algorithms based on grey relational degree are more stable than traditional algorithms. This may indicate that it is a feasible way to support decision problems using the multi-level case retrieval model by integrating decision technique and grey relational degree with CBR.

5. Conclusion

The methodology of CBR is a developing reasoning method accompanying the development of cognitive psychology. It has solved a lot of real-world problems that some other reasoning methods are not good at. In non-structured decision problems, CBR has shown strong advantages. In this paper, we mainly focus on management area. A new multi-level case retrieval model of CBR is constructed by integrating decision technique and grey relational degree with CBR under the circumstance of multi-level case structure. Basic working flow when integrating CBR with decision support system is analyzed. Multi-level case retrieval & selection models on the foundation of case delaminating structure are built up. Layer similarity, integrated similarity, and grey relational degree are fundamental issues of the hybrid model. Experiment indicates that the new hybrid case retrieval model can achieve acceptable performance in real-world application.

Acknowledgments

This research is partially supported by the National Natural Science Foundation of China (#70801055), and the Zhejiang Provincial Natural Science Foundation of China (#Y6070111). The authors gratefully thank anonymous referees for their useful comments and editors for their work.

References


Manuscript Received: Aug. 8, 2007
Accepted: Oct. 4, 2008