

# Evolutionary Fuzzy Case-Based Reasoning for Financial Performance Ranking

Sheng-Tun Li<sup>1,2</sup> Hei-Fong Ho<sup>1,3</sup> Yi-Chung Cheng<sup>2,4</sup>

<sup>1</sup>Institute of Information Management, National Cheng Kung University, Taiwan, ROC

<sup>2</sup>Department of Industrial and Information Management, National Cheng Kung University, Taiwan, ROC

<sup>3</sup>Department of Business Administration, Chang Jung Christian University, Taiwan, ROC

<sup>4</sup>Department of International Trade, Tainan Woman's College of Arts & Technology, Taiwan, ROC

## Abstract

In this paper, we propose a hybrid decision model for supporting the ranking financial status of corporations using case-based reasoning augmented with genetic algorithms and the fuzzy nearest neighbor method. An empirical experimentation on 746 cases was conducted that shows that the average accuracy of the ranking is about 92% and 80% for the first order and the second order, respectively. This confirms that the proposed approach is very effective and can make a significant contribution to the decision-making of ranking.

**Keywords:** Case-Based Reasoning, Genetic Algorithms, Fuzzy Nearest Neighbor algorithm, financial statement analysis, financial performance ranking.

## 1. Introduction

As far as the knowledge of the underlying financial performance is concerned, financial statement analysis provides an ultimate method for evaluating real cases. The accounting and financial ratios indicate the financial status of an enterprise at a period of time. The common ways for the analysis include trend analysis, financial statement comparison and financial ratio analysis, etc. Among them, the financial ratio analysis was widely accepted because it could make the message amount of the financial statement lowest. But the representative and the meaning of the related ratios are always interpreted by the expert, in other words, the analysis is quite subjective.

The nature of financial performance ranking processes is unstructured and subjective, which depends on the domain-dependent expertise and the analyst's perception. However, the resulting ranking is crucial to the manager because, based on this information, he/she can realize the present

financial status and make the decision for the future plan. And the investor can refer to the analysis results to predict the running potential and the financial status of the company in the future and modify his investment. For this reason, an intelligent decision support system is needed to make ranking process effective and objective.

The computer technology has made a great progress recently. In order to make the decision-making more efficient, the Artificial intelligent (AI) has become revived to explore the problems in the commercial applications [1], e.g., Neural Networks [2], Rule-Based System [3] and Case-Based Reasoning (CBR)[4,5, 6].

For the traditional AI, because there is a lacking of a good interfaces and mechanisms between the user and machine, it is hard for it catching the expert knowledge, and even more difficult to realize the fuzzy knowledge. In this investigation, the combination of CBR with fuzzy set theory would promote the system flexibility. In working the attribute of the testing case, the application of membership function will make AI more efficient in the integration of the system. In this paper, we propose a hybrid system using evolutionary fuzzy CBR for analyzing the financial performance from financial statements and can explore the efficiency of the system in assisting the decision-making.

This paper is organized as follows. Section 2 presents the hybrid approach integrating CBR with GA and Fuzzy NN. Section 3 describes the experimental procedure. Section 4 depicts experimental results. In the final section, the conclusion is presented.

## 2. GA Fuzzy CBR approach

In essence, an effective retrieval of useful prior case plays a central role in developing a CBR system [6]. However, the design of an appropriate case-matching process in the retrieval step is still

in challenging. The CBR community has proposed several approaches for choosing indexes for retrieving particular cases: nearest neighbor, inductive, and knowledge-guided. [7]

Inductive approach is useful when a single case feature is required as a solution and is more appropriate when the case is well-defined while nearest neighbor is preferred when the retrieval goal is subjective. Inductive indexing requires large volume of cases and time for maintaining an optimal induction tree. [7, 8] Concerning those factors, we adopt the nearest neighbor retrieval approach for the proposed CBR system.

For synergizing the nearest matching resulting from individual case attributes, an useful weight vector for differentiating one case from others is critical. Furthermore the relevant weights associated with selected attributes have to be assigned before case matching process. The weights are usually determined by subjective judgment. We exploit a GA algorithm approach to automatically construct an optimal or near-optimal weight vector by learning the historical cases, which then compared with weight vector defined by an expert.

We apply the classification accuracy rate of the test case set to the fitness function for the proposed system. This fitness function is expressed mathematically as follows:

$$\max \quad CAR = \frac{1}{n} \sum_{i=1}^n CA_i, \dots$$

where  $CA_i = \begin{cases} 1 & \text{if } O(T_i) = O(S_{j^*(i)}) \\ 0 & \text{otherwise} \end{cases}$

For a given  $i$  ( $i = 1, 2, \dots, n$ ) test case,  $S_{j^*(i)} = \max(\text{Similarity}(T_i^R, T_j^I))$ , where  $S_{j^*(i)}$  is the most similar retrieved case  $j$  with testing case  $i$ .  $CA_i$  is classification accuracy for  $i$ th test case, denoted by 1 for correct, otherwise 0. And  $CAR$  is the total classification accuracy rate of the test case set. It ranges from 0 to 1. The higher the  $CAR$  the closer optimal solution this chromosome is. If  $CAR$  is equal to one, then classification accuracy rate is 100% and the corresponding chromosome is optimal.

Nearest-neighbor technology provides a simple nonparametric procedure for the assignment of a class label to the input vector based on the class labels represented by the closest neighbor of the vector. But one of the problems encountered in using nearest-neighbor classifier is that normally each of the sample vectors is considered equally important in the assignment of the class label to the input vector. Another problem is that there is no indication of its "strength" of membership in that class. For problems mentioned above, we

incorporated fuzzy membership into the classical nearest neighbor similarity function. The inverse distances from the nearest samples to input vectors served to weight the nearest vectors' class memberships more if they are closer to the vector under consideration. Then the degree of membership of input vector in each class can be specified rather than just the crisp result. It is noted that a vector's memberships in the resulting classes must sum to one. Analogous fuzzy nearest prototype algorithm [9] has adopted in the proposed system for not only the computational simplicity but also the desirable membership assignments.

### 3. Experimental procedure

Attributes of each historical case are codified into fuzzy values by expert. Then the historical case base is divided into training set and testing set. GA is employed to generate the best set of weights that are able to promote the association consistency among the cases. The fitness function is defined to find the maximum total sum of classification accuracy ratio (CAR) as mentioned above. Each solution (weight of case) calculate the classification accuracy rate using ten-fold cross validation. Finally, the statistic average of accuracy rate and variation will be computed.

The experimental details are described as follows.

#### (1) Establishment of case base

Financial statement analysis needs specific knowledge and the accumulated experience, therefore in this research, we had an accountant to help us. This accountant is a forty-year-old male and an expert with seventeen-year experience of the financial statement analysis. After discussing with the expert, we decided to adopt six main attributes to evaluate the corporate financial performance. These attributes include Current Ratio (CR), Collection Period for Account Receivable (AR), Days to Sell Inventory (INV), Net Operation Cycle (NOC), Sales (S), and Operation Income (OI).

When the financial attributes were determined, the expert starts to analyze financial data of 746 publicly-held companies in Taiwan. In defining the scope of attribute values, the expert used linguistic values for the attributes, shown in Table 1.

Table 1 the scope of attribute values

worst	bad	good	better	best
-------	-----	------	--------	------

#### (2) Evaluation of financial performance

After judging the linguistic value for each attribute and company in turns, the expert makes the final ranking according to financial performance values of each company. The judging

result of one publicly-held company by the expert was shown in Table 2. Its financial performance ranks is the worst. The fuzzy case-based reasoning was used to justify this result as system feedbacks later on.

Table 2 judged result by Expert

attributes	CR	AR	INV	NOC	S	OI	RATE
judgment	worst	better	good	good	worst	bad	worst

(3) Define the similarity function of membership of attributes

In this research, each case is composed of six attributes. We compare six attributes with the corresponding attributes of sample case and get six similarity degrees of the attributes. Then similarity values of cases can be obtained via the matching function (such as Euclidean Distance function) assigned by the expert. The similarity function of the membership is defined as follows.

$$Similarity(x) = \begin{cases} 1 & x < 0 \\ -x + 5 & 0 \leq x \leq 4 \\ 0.2 & 4 < x \end{cases}$$

(4) Genetic Algorithm Approach

GA is used to adjust the weight of each attribute. We select six attributes for evaluating, the values were set from 0 to 1, (precision setting is three digits after the decimal) which reflect the importance of each attribute. Each solution is called a chromosome, which is composed of six genes and each gene includes seven bits in the form of a binary string. The roulette wheel selection was used in the research, and the higher the fitness value, the higher the pick-out probability is. Two-point crossover and single-point mutation were adopted in this research. The crossover rate was set to 0.8 and the mutation rate was set to 0.01.

(5) The measure of the importance of the attributes in cases

When the CBR is proceeding during the experiment, the most important work is to find the accurate weight values of classification in cases and apply it to the CBR, namely, this is the measure of the importance of the attributes. The simplest way is to invite expert to set up these values according to their professional knowledge and experience, but when expert tries to quantify the importance of the decision, usually there would be somewhat vague and it cannot be well defined. Therefore, the experiment of this research was divided into two parts: one is to invite the expert to setup the weighted value of attributes, and the other is to have GA find the weighted values of the optimal or near-optimal. Finally, to explore if the weight values found by GA could be consistent with those of the decision made by the expert via the comparison of these two parts.

We begin to proceed the first phase experiment. First, to explain the manipulation of CBR to expert, and to invite the expert to give the weight values of the importance of each attribute: the most important attribute is 1 and the least is 0 according to his professional knowledge and experience. Table 3 shows the weight vector set by the expert and experimental results.

Table 3 the weight vector set by the expert and experimental results.

Attributes	CR	AR	INV	NOC	S	OI
Weight values	1	0.2	0.2	0.8	0.7	0.4
Accuracy rate	82.50%					

Thereafter, we begin to proceed the second-phase experiment of Fuzzy CBR. We adopt 10-fold cross validation and divided 746 cases into ten equal-sized parts ( $n_1, n_2, \dots, n_{10}$ ) arbitrarily. First, we take  $n_1$  as the test cases, and the rest nine parts as the training cases. The purpose of the training cases is to get a set of nearly optimized weight values from these nine parts and further to validate the accuracy of the weight value of the test case  $n_1$ . Second,  $n_2$  is taken as the test cases, and the others as the training cases. Analogously, after proceeding ten times, there would be ten sets of the weight values of near-optimal and accuracy. Finally, we take the most accurate weight values of these ten sets as the most optimum solution of the cross validation as shown in Table 4. The experimental results show that the average value of 10-fold cross validation is 92.36% and the variance is 13.17.

Table 4 the optimum solution using GA Fuzzy CBR

Attributes	CR	AR	INV	NOC	S	OI
Weight values	1	0.01	0.15	0.57	0.29	0.15
Accuracy rate	97.33%					



Figure 1 Evolutionary Fuzzy CBR

Figure 1 shows the main menu of the system. The up-left corner of Figure 1 is the setup domain of the parameters, in which the parameters of membership function can be selected and the weight values of the cases can be set up. The right hand of Figure 1 is a function domain where the types of the membership function of attributes, similarity function of case, and the adoption of the suggested weighting values of the gain module

can be selected. The down-left corner of Figure 1 shows the membership degree of each solution cases where the expert can examine the various solution cases and make decision.

## 4. Experimental results

In this investigation, the parameters used in the system and the empirical results are shown in Table 5, which shows that the accuracy rate is the worst for the traditional CBR but the best for the proposed Evolutional Fuzzy CBR.

Table 5 Accuracy rate of classification in CBR

The method that parameter set up	CAR
Traditional Case-Base Reasoning	75.64%
CBR by expert to set parameter	82.50%
GA Case-Based Reasoning [10]	90.61%
GA Fuzzy Case-Based Reasoning	92.36%

Furthermore, Table 6 lists the comparison of the analyzing difference by GA Fuzzy CBR and expert.

Table 6 differences between GA Fuzzy CBR and expert.

	1	2	3	4	5	total
1	184	9	1	0	0	194
2	13	114	10	1	0	138
3	0	5	204	4	0	213
4	0	0	7	130	3	140
5	0	0	0	4	5	61
total	197	128	222	139	6	746

When the expert resolves the problems of performance ranking, there is the presence of some vagueness. However, the results of the decision could be accepted only if its difference is restricted in one order, which would lead investor not to make a wrong judgment.

In comparing the related reports in the literature, the error of the results is quite low. Thus, the results produced by the add-in module proposed in this investigation have been proved to be better. In addition, because the classified results of the GA Fuzzy CBR have membership degree, we put the membership degree of the cases of the wrong judgment and the original data together in order for the expert to check the arguable data and try to modify the original data, which in turn not only would lend support for the GA CBR system to have more accurate judgment but also can provide the expert proper feedback, namely, the expert can review his decision strategy.

## 5. Conclusions

In this study, the GA Fuzzy CBR system has been established, which enhances the capability of the individual expert of the financial statement to

analyze financial performance of an enterprise. The decision model of the information technology expert not only can help the decision of the expert but also improve the quality of the decision.

It was found that the GA Fuzzy CBR can lend support to the decision of an expert, which reveals a better accuracy about 92.36%. In the aspect of the interpretation of the decision, because the fuzzy membership degree in the suggested solution cases can be ranked, it can provide more friendly decision. Finally, in the aspect of the decision of the practice, if two fuzzy membership degrees are selected in the hypothesis for judgment, the total accuracy would be 98%, which is very practical for managers and investors.

## 6. References

- [1] Liao, S. H., "Expert system methodologies and applications—a decade review from 1995 to 2004," *Expert Systems with Applications*, 28:93-103, 2005
- [2] Kwon, Y.S., I.G. Han, and K.C. Lee, "Ordinal pair wise partitioning (OPP) approach to neural networks training in bond rating," *Intelligent Systems in Accounting Finance and Management.*, 6: 23-40, 1997.
- [3] Nedovic, L., Devedzic, V., "Expert systems in finance--a cross-section of the field", *Expert Systems with Applications*, 23:49-66, 2002.
- [4] Chang, P. C., Lai, C. Y., and Lai, K. R., "A hybrid system by evolving case-based reasoning with genetic algorithm in wholesaler's returning book forecasting", *Decision Support Systems*, in press, 2006.
- [5] Jeng, B. C., and Liang, T. P., "Fuzzy indexing and retrieval in case-based systems," *Expert Systems with Applications*, 8(1):135-142, 1995.
- [6] Chun, S. H., and Park, Y. J., "A new hybrid data mining technique using a regression case based reasoning: Application to financial forecasting", *Expert Systems with Applications*, 31:329-336, 2006.
- [7] Barletta, R., "An introduction to case-based reasoning", *AI Expert*, 6(8): 42-49, 1991.
- [8] Kolodner, J. *Case-based reasoning*. San Mateo. CA: Morgan Kaufmann, 1993.
- [9] Keller, J.M., M.R. Gray, and J.A. Givens Jr., "A fuzzy k-nearest neighbor algorithm", *IEEE Transactions on Systems*, 15(4): 258-263, 1985.
- [10] Shin, K.S. and I. Han, "Case-based reasoning support by genetic algorithms for corporate bond rating", *Expert Systems with Applications*, 16: 85-95, 1999.