OPTIMIZATION OF SYMBOLIC SIMILAITY BY GENETIC ALGORITHM FOR CASE BASED APPOACH TO CORPORATE BOND RATING

[1] Kyung-shik Shin, [2] Hyun-jung Kim
[1,2] Ewha Womans University, College of Business Administration
[1] ksshin@ewha.ac.kr, [2] charitas@empal.com

[3] Kyoung Jun Lee
[3] School of Business, Kyung Hee University, klee@khu.ac.kr

ABSTRACT

A critical issue in case-based reasoning (CBR) is to retrieve a usefully similar case to the problem. Though the integration of domain knowledge into the case indexing and retrieving process is highly recommended in building a CBR system, it is difficult to carry it out because such knowledge often cannot be successfully captured. Moreover, a symbolically represented problem we are trying to solve may not be easily verified in this way. This study investigates a hybrid approach, which is that the genetic search technique is used to symbolic case-based retrieval process in an attempt to find optimal values of the similarity matrix. We apply these values to the matching and ranking procedure for a more effective CBR system to the problem of corporate bond rating classification. The results demonstrate that this approach supports effective retrieval of symbolically represented cases and increases overall classification accuracy rate significantly.

KEYWORDS: symbolic similarity, case-based reasoning, genetic algorithms, corporate bond rating

INTRODUCTION

The development of the corporate bond rating prediction model has long been regarded as an important and widely studied issue in the academic and business community. In spite of interests in accurate quantitative prediction of corporate bond rating, due to lack of scientific credit rating methodology and sufficient data accumulation to construct the model, the traditional approach produce an internal rating on the basis of human judgment to a significant extent in the real world. Another possible reason why it is difficult to construct models that achieve performance comparable to that of highly trained human expertise is due to a failure to learn symbolic rules that could be understood and verified by experts. However, there is a need of fairly accurate and robust corporate bond rating prediction models, to deal with the information about complex human perception through linguistic representation for indiscriminate practices.

The early studies of bond rating applications tend to use statistical techniques such as multiple discriminant analysis (MDA) model [2][3][29], which is a most common means of classifying bonds into their rating categories. However, statistical methods have some limitations in applications due to the violation of multivariate normality assumptions for independent variables, which is frequently occurred in financial data. Moreover, these are not amenable to the underlying representation of the data, and so for mixed types of data such as numeric and categorical values, the analysis is more complex or even not applicable to many symbol variables that have more than two levels. Recently a number of studies have demonstrated that artificial intelligence approaches such as case-based reasoning can be alternative methodology for corporate bond rating

classification problems.

Case-based reasoning (CBR) is a problem-solving technique in that the case specific knowledge of past experiences is utilized to find the most similar solution to a new problem. The more basic principle underlying CBR is that a model articulates and restructures the knowledge acquisition framework of domain expertise, which uses analogical reasoning to solve complex problems and to learn from problem-solving experiences. CBR is preferred over rule-based systems if rules are inadequate to express the richness of the domain knowledge. Numerous applications have been reported that CBR seems best suited for experience-rich domains [5][8][13][20][23][26][27][30], including those dealing with business classification such as bond rating and bankruptcy prediction [6][7][18][19][32][33].

Building a successful CBR system mainly depends on an effective and efficient retrieval of similar cases from useful cases for the problem. A good matching and retrieving function should take into account the features of a case that are more important and scores cases for usefulness according to given tasks. For the symbolic retrieval function, to measure the degree of similarity of corresponding features in two cases is preferable since a case that matches based on more concrete value of similarity may perform better than one based on an indicator that the two cases being evaluated are identical. Our particular interest lies in the assignment of domain knowledge such as similarity values of a similarity matrix for pairs of symbolic values of each feature for retrieving the most similar case to a new problem. However, even for experts, a symbolically represented problem we are trying to solve cannot be easily determined which set of similarity values would be the most effective for the particular pair of feature values in solving a specific problem. Though the integration of domain knowledge into the case matching and retrieving processes is highly recommended, it is difficult to carry out because such knowledge often cannot be successfully and exhaustively captured and represented.

In this paper, we investigate a hybrid approach using genetic algorithms (GA) to the symbolic case-based retrieval process (GA-CBR) in an attempt to find optimal or near optimal values of the similarity matrix. This study uses GA to extract knowledge that can guide effective retrieval of useful cases. We apply these values to the matching and ranking procedure for a more effective CBR system. Compared to the existing symbolic retrieval function, which relies on an exact match counting method, in GA-CBR it is exploited a real similarity computation algorithm. Integration of CBR with GA is also important, because it helps to reap the benefits of both systems. CBR provides analogical reasoning structures for experience-rich domain while GA provides CBR with knowledge that can guide effective retrieval of more relevant cases through machine learning. Our proposed approach is demonstrated by applications to corporate bond rating.

The remainder of this paper is organized as follows. Next section contains the methodologies used in this study and a hybrid structure of CBR supported by GA. The specific information about data and experiments is described in the research development section. In the result and analysis section, empirical results are summarized and analyzed. The final section includes a conclusion and future research issues.

HYBRID APPROACH USING GA

Integration of domain knowledge into the case indexing and retrieving process is important in building a useful case based system. The central idea of the combination of GA and case-based system is that CBR transfers the burden of knowledge assignment of the indexing and retrieving process to the searching and learning capabilities of evolutionary algorithms. In this subsection, we propose a hybrid approach using GA to the symbolic case-based retrieval process in an attempt to increase the overall effectiveness of the system.

Prior Research on the Hybrid Approach Using GA

Hybridization of techniques can produce better systems if it ensures synergistic combination. GA has been increasingly applied in conjunction with other AI techniques such as neural networks, rule-based system, fuzzy theory, and CBR. The integration of GA and neural networks is a rapidly expanding area. The common problems faced by researchers and developers in using neural network techniques are optimization of input selection, network design and learning condition. Various problems of neural network design can be optimized using GA [39]. Examples include selecting relevant input variables, determining the optimal number of hidden layers, nodes and connectivity, and tuning the learning parameters [4][14][31]. Another approach of combining neural networks and GA is genetic training. GA has been used to search the weight space of a neural network without the use of any gradient information [10][15][25]. GA is also used in conjunction with fuzzy logic systems to provide an appropriate set of fuzzy IF-THEN rules for classification problems [16] and to improve fuzzy logic controller [28].

Few studies have dealt with hybridization of GA and CBR, though there exists a great potential for useful applications in this area. For enhancing matching and ranking procedure of CBR, a machine learning approach using GA is used to find an optimal or near optimal weight vector of the attributes in case indexing and retrieving [9][12][32]. On the other hand, defining similarity measures is a crucial task when developing CBR applications. In particular, when employing similarity measures that contain a lot of specific knowledge about the addressed application domain, modeling similarity measures is a complex and time-consuming task. One common element of the similarity representation is local similarity measures used to compute similarities between the values of single attributes. Stahl [35] points out some problems of the state-of-the-art procedure to defining similarity measures and propose an alternative strategy to acquire the necessary domain knowledge based on a Machine Learning approach. Stahl and Gabel [34] present an approach to learn local similarity measures by employing an evolution program a special form of a genetic algorithm. The goal of the approach is to learn similarity measures that sufficiently approximate the utility of cases for given problem situations in order to obtain reasonable retrieval results. Jarmulak et al. [17] present a tool which helps to reduce the knowledge acquisition effort for building a typical CBR retrieval stage consisting of a decision-tree index and similarity measure and use Genetic Algorithms to determine the relevance/importance of case features and to find optimal retrieval parameters. The optimization is done using the data contained in the case base. Tsatsoulis and Stephens [37] use genetic algorithms to discover methods to combine and operationalize vague selection criteria such as "recency" and "frequency." In their research, GA helps discover selection criteria for the contradictory solutions retrieved by CBR retrieval and significantly improved the accuracy and performance of the CBR system. Wang and Ishii [38] apply GA to the method of similarity metrics based on the cases being represented by structured representations.

Hybrid Structure of a GA-CBR System

CBR is one of the most promising techniques to handle many attributes that are complex, unstructured and mixed with quantitative and qualitative type information of the practical problem-solving field, because a CBR system finds a solution that is already stored in the case library to a given situation according to their similarity. Considering that similarity calculation plays a very important role for a successful CBR system, a reasonable technique to measure the degree of similarity between the reference case and the test case is necessary. There are various similarity-measuring methods depending on an attribute type, which are not always public [1][11][21][22][24][36].

Our particular interest lies in improving the effectiveness of symbolic case-based retrieval process

in a CBR system. Among several ways to assign the similarity values, the most widely used symbolic similarity measure is an only exact match counts; that is, the similarity has the maximum value (1) when the two cases being evaluated are identical and has the minimum value (0) when they are completely different. However it is difficult to give sufficient information of the similarity, since there is no real similarity calculating process to find local similarities relevant not only to their own values but also to the values of other attribute. Another way to assign similarity values is to have a human expert assign them as the case library is being built. Though the expert is expected to have the knowledge and experience required to decide which set of local similarities would be the most effective to solve a specific problem, even for experts, it is not easy to verify such similarities in a symbolically represented domain.

As an alternative approach, we introduce the notion of machine learning to learn the optimal set of local similarities from historical cases using evolutionary search technique. By evaluating the fitness of different values of the similarity matrix, we may find good solutions for the system. GA is stochastic search techniques that can search large and complicated spaces on the ideas from natural genetics and evolutionary principle. GA performs the search process in four stages: initialization, selection, crossover, and mutation. In the initialization stage, a population of genetic structures (called chromosomes) that are randomly distributed in the solution space is selected as the starting point of the search. After the initialization stage, each chromosome is evaluated using a user-defined fitness function. Chromosomes with a good performance may be chosen for replication several times whereas poor-performing structures may not be chosen at all. Such a selective process causes the best-performing chromosomes in the population to occupy an increasingly larger proportion of the population over time. GA applies crossover and mutation to generate a new population of problem solutions and select the best solution for the problem. Crossover causes to form a new offspring between two randomly selected "good parents". Crossover operates by swapping corresponding segments of a string representation of the parents and extends the search for a new solution in far-reaching direction. Mutation is a GA mechanism where we randomly choose a member of the population and change one randomly chosen bit in its bit string representation [33].

For the first step to construct the GA-CBR system, we search an optimal or near optimal values of the similarity matrix with precedent cases of each symbolic attribute for which the classification outcome has been determined. The similarity matrix expresses similarity values between each symbolic feature values and is inserted into the matching function to rank and retrieve useful cases. We set the range of the similarity values from 0 to 1 and do not apply particular constraints for this search.

The task of defining a fitness function is always application specific. In this case, the objective of the system is to retrieve more relevant cases that can lead to the correct solutions. The ability of case-based system to achieve these objectives can be represented by the fitness function that specifies how well the matching function increases the classification accuracy. We apply the classification accuracy rate of the test set to the fitness function for this study. The test set consists of known cases of which the classification outcome has been determined and is used to evaluate the fitness of different sets of similarity values. Mathematically, this fitness function is expressed as:

$$\text{Maximize} \qquad CR = \frac{1}{n}\sum_{i=1}^{n} CA_i$$

$$\text{s.t.} \qquad CA_i = \begin{cases} 1 & \text{if } O(T_i) = O(S_{j*(i)}) \\ 0 & \text{otherwise} \end{cases}$$

$$S_{j^*(i)} = \underset{j \in R}{\text{Max}}\left(\sum_{k=1}^{l} w_k * sim(T_{ik}, R_{jk}) \right)$$

$$sim(T_{ik}, R_{jk}) = \begin{cases} 1/(1+|T_{ik} - R_{jk}|) & \text{if } \kappa \text{ is numeric} \\ \text{LS}(T_{ik}, R_{jk}) & \text{if } \kappa \text{ is symbolic} \end{cases} \quad \text{for given } i \ (i=1,2,\ldots\ldots,n)$$

CR: the classification accuracy rate of test set,
CA_i: the classification accuracy of *i*th case of test set denoted by 1 and 0 ('correct' = 1, 'incorrect' = 0),
$O(T_i)$: target output of *i*th case of test set,
$O(S_{j^*(i)})$: the output of *j*th case of reference set that has the minimum distance with *i*th case of test set,
$S_{j(i)}$: the similarity between *i*th case of test set and the *j*th the case in reference set,
LS: the local similarity value of the similarity matrix,
T_{ik}: the *k*th feature of the *i*th case of test set (T),
R_{jk}: the *k*th feature of the *j*th case of reference set (R),
w_k: the importance weight of the *k*th feature of case, and
l: the number of numeric features and *n* the number of test cases.

For example, if the bond rating of *i*th case of test set and the closest case in the reference set are the same, then CA_i is 1.

The second step is to apply the derived similarity values in the first step to case indexing scheme for the case-based retrieval process and evaluate the resulting model with additional validation cases for which the outcome is also known. Since the validation cases are not used for parameter optimization, the prediction performance tested by these cases would be the closest to the current or future cases. If the project is successful, this leads to production. The last step is a production phase. In production phase, new (unclassified) data is presented in the model to solve the problem.

RESEARCH DEVELPMENT

The research data consists of 297 financial ratios and the corresponding bond rating of 1,816 Korean companies whose commercial papers have been rated from 1997 to 2000. The bond rating we employ is provided by National Information and Credit Evaluation, Inc., one of the most prominent bond rating agencies in Korea. Credit grades are classified as 5 coarser rating categories (A1, A2, A3, B, C) according to credit levels. Table 1 shows the organization of the data set.

Table 1. Number of companies in each rating

Ratings	Number of cases	%
A1	58	3.2
A2	242	13.3
A3	586	32.3
B	780	43.0
C	150	8.3
Total	1,816	100.0

The data set is arbitrarily split into three subsets; about 80% of the data is used for a reference set, 10% for a test set and 10% for a validation set. The reference data is used to learn an optimal similarity values in genetic learning and also used as a case base for retrieval. The test data is used to evaluate the weight vectors, verifying how well the indexing of CBR system works. The

validation data is used to test the results with the data that is not utilized to develop the model. The number of the reference cases, the test cases and the validation cases are 1,454, 181 and 181, respectively.

We apply two stages of the input variable selection process. At the first stage, we select 106 variables by 1-way ANOVA for the numeric type and Kruskal–Wallis test for the symbolic type between each financial ratio as an input variable and credit grade as an output variable. In the second stage, we select 9 variables using a MDA stepwise method to reduce dimensionality. We select input variables satisfying the univariate test first, and then select significant variables by the stepwise method for refinement. Two qualitative variables are coded into five types according to the yearly sign transition of ordinary profit (X8) and operating activities cash flows (X9) respectively for three-year period. The specific descriptions of selected variables for this research are shown in Table 2.

Table 2. Definition of variables

Variable	Definition	Data type	Code	Illustration
X1	Net income to total asset	Numeric		
X2	Net interest coverage ratio	Numeric		
X3	Times interest earned	Numeric		
X4	Net income to capital stock	Numeric		
X5	Equity to total asset	Numeric		
X6	Fixed assets to total asset	Numeric		
X7	Current liabilities to total asset	Numeric		
X8	Transition of ordinary profit	Symbol	1	(+,+,+)
			2	(-,+,+),(+,-,+)
			3	(-,-,+)
			4	(+,+,-)(-,+,-)(+,-,-)
			5	(-,-,-)
X9	Transition of operating activities cash flows	Symbol	1	(+,+,+)
			2	(-,+,+),(+,-,+)
			3	(-,-,+)
			4	(+,+,-)(-,+,-)(+,-,-)
			5	(-,-,-)

The CBR model in this study uses a similarity-based indexing and nearest-neighbor retrieving in which weights obtained by Pearson's correlation analysis are assigned among attributes. Correlation coefficient acquired from Pearson's correlation analysis is transformed to the weighted value of each attribute.

Table 3. The similarity matrix of a symbolic attribute in a pure-CBR

Variable	1	2	3	4	5
1	1	0	0	0	0
2		1	0	0	0
3			1	0	0
4				1	0
5					1

To compare with the symbolic case-based retrieval process using the GA-CBR approach, we also

design the similarity matrix of a symbolic attribute applied in a conventional CBR model (pure-CBR) as shown in Table 3. The following similarity matrix for the symbolic type variables is obtained on the basis of the existing symbolic retrieval function that is an exact match counting method.

The parameters consisting population size, crossover rate, mutation rate and stopping condition need to be defined first when developing the GA-based system. There had been much debate regarding the optimal controlling parameters that we should specify for experiment. For this experiment, we use 100 organisms in the population, 0.5 in the crossover rate and 0.06 in the mutation rate (GA-CBR_1) and 0.6 in the crossover rate and 0.1 in the mutation rate (GA-CBR_2). As a stopping condition, we use 1,000 trials. These processes are performed by a prototype system implemented in VB supported by Evolver Developer's Kit (EDK) version 4.0.2.

RESULT & ANALYSIS

To investigate the effectiveness of the integrated approach for the case retrieval of a symbolic attribute using GA in the context of the corporate bond rating classification problem, we set GA to search the similarity values of the particular pair of each feature values. The derived results by genetic search are summarized in Table 4. To reduce the impact of random variations in the GA search process, we replicate the experiment several times and suggest the best similarity matrix found in each model.

Table 4. The optimized similarity matrix of symbolic attributes using (a) GA-CBR_1 and (b) GA-CBR_2

(a) The optimized local similarities of attributes using GA-CBR_1

X8	1	2	3	4	5
1	1	0.93	0.98	0.96	0.75
2		1	0.32	0.88	0.63
3			1	0.93	0.93
4				1	0.97
5					1

X9	1	2	3	4	5
1	1	0.93	0.22	0.78	0.69
2		1	0.61	0.97	0.32
3			1	0.69	0.42
4				1	0.92
5					1

(b) The optimized local similarities of attributes using GA-CBR_2

X8	1	2	3	4	5
1	1	0.58	0.28	0.85	0.59
2		1	0.05	0.96	0.69
3			1	0.55	0.79
4				1	0.76
5					1

X9	1	2	3	4	5
1	1	0.78	0.81	0.93	0.22
2		1	0.27	0.99	0.98
3			1	0.97	0.25
4				1	0.52
5					1

Table 5 shows the comparison of the results of the classification techniques applied for this study. Each cell of Table 5 contains the accuracy of the classification techniques. Among the techniques, the GA-CBR_2 model has the highest level of accuracies in the given data set. Based on the empirical results, we conclude that the symbolic case-based retrieval method using the local similarity values derived from the genetic search process can improve the performance of a CBR system as GA finds optimal or near optimal solution for the specified objective function.

McNemar test results for the comparison of the predictive performance between the comparative models and the GA-CBR models for the validation cases are summarized in Table 6. The results of McNemar tests support that the GA-CBR model has higher classification accuracy than the pure-CBR model with significant levels. In addition, it appears that GA-CBR_2 outperforms more

significantly than GA-CBR_1 compared with pure CBR; however, the controlling parameters that we should specify for experiment in GA-CBR has no significant difference.

Table 5. Classification performance of the validation cases

Ratings	Pure-CBR		GA-CBR_1		GA-CBR_2	
	Number of hit /total hit	Hit ratio	Number of hit /total hit	Hit ratio	Number of hit /total hit	Hit ratio
A1	3/5	60.00	5/5	100.00	4/5	80.00
A2	13/25	52.00	14/25	56.00	15/25	60.00
A3	36/58	62.07	43/58	74.14	45/58	77.59
B	68/78	87.18	72/78	92.31	71/78	91.03
C	11/15	73.33	12/15	80.00	12/15	80.00
Total	131/181	72.38	146/181	80.66	147/181	81.22

From the overall results of experiments, we can conclude that GA-CBR performs better than pure-CBR and the integrated approach proposed for this study is effective, enhancing the classification accuracy of the CBR for the corporate bond rating application domain.

Table 6. McNemar values for the comparison of performance between models

	(Significance level)	
	GA-CBR_1	GA-CBR_2
Pure-CBR	0.017 **	0.007 ***
GA-CBR_1		1

** significant at 5%, *** significant at 1%

CONCLUSION

Both CBR and GA are artificial intelligent approaches that have received significant attention in recent years. In this study, we propose that a hybrid approach using CBR and GA enhances the effectiveness of CBR to the problem of corporate bond rating classification. In this approach, the genetic search technique is used to symbolic case-based retrieval process in an attempt to find optimal or near optimal values of the similarity matrix. The results demonstrate that this approach supports effective retrieval of symbolically represented cases and increases overall classification accuracy rate significantly.

These results support the following findings. First, the knowledge acquired by problem domain supports the retrieval of usefully similar case to solve the problem. Since the task of GA to define a fitness function is always domain specific, this hybrid approach utilizes domain knowledge and case specific knowledge simultaneously. Second, GA is an effective method of extracting knowledge for case-based retrieval. Using GA, we can obtain the near optimal values of local similarities representing the similarity of each feature values for the symbolic type.

Our study has the following limitations that need further research. First, in setting up the GA optimization problem, we must select several parameters such as stopping conditions, the population size, crossover rate, mutation rate and so on. The values of these parameters can greatly influence the performance of the algorithm. The varying parameters also generate a lot of groups for our general result. The second limitation is on the appropriateness of the importance assignment values using correlation coefficient acquired from Pearson's correlation analysis that we applied in CBR. Finally, this study only focuses on optimizing the similarity matrix for the symbolic case-based retrieval process in a CBR system. However, a GA approach could potentially be used to

optimize other specific points of the case-based reasoning process. We hereby believe that the potential is great for further research with hybrid approaches using GA and also different intelligent techniques as ways to improve the performance of the applications.

REFERENCES

1. Amen, R. and P. Vomacka, "Case-based Reasoning As a Tool for Materials Selection," *Materials and Design,* 22, 2001, pp.353-358.
2. Baran, A., J. Lakonishok, and A. R. Ofer, "The Value of General Price Level Adjusted Data to Bond Rating," *Journal of Business Finance and Accounting*, 7(1), 1980, pp.135-149.
3. Belkaoui, A., "Industrial Bond Ratings: A New Look," *Financial Management*, 9(3), 1980, pp.44-51.
4. Bishop, J. M., M. J. Bushnell, A. Usher, and S. Westland, "Genetic Optimization of Neural Network Architectures for Color Recipe Prediction," *Artificial Neural Networks and Genetic Algorithms*, Springer-Verlag, New York, 1993, pp.719-725.
5. Brown, C. E. and U. G. Gupta," Applying Case-based Reasoning to the Accounting Domain," *Intelligent Systems in Accounting, Finance and Management*, 3, 1994, pp. 205-221.
6. Bryant, S. M., "A Case-based Reasoning Approach to Bankruptcy Prediction Modeling," *Intelligent Systems in Accounting, Finance and Management*, 6, 1997, pp. 195-214.
7. Buta, P., "Mining for Financial Knowledge with CBR," *AI EXPERT*, 9(2), 1994, pp.34-41.
8. Chi, R. T., M. Chen, and M. Y. Kiang, "Generalized Case-based Reasoning System for Portfolio Management," *Expert Systems with Applications*, 6(1), 1993, pp.67-76.
9. Chiu, C., "A Case-based Customer Classification Approach for Direct Marketing," *Expert Systems with Applications*, 22, 2002, pp.163-168.
10. Deboeck, G. J., "Using GAs to Optimize a Trading System," In Deboeck, G. J. (Eds.), *Trading on the Edge*, John Wiley, New York, 1994, pp.174-188.
11. Fix, E. and J. L. Hodeges, "Discriminatory Aanalysis, Nonparametric Discrimination: Consistency Properties," *Report 4. Randolph Field, TX: USAF School of Aviation Medicine*, 1951.
12. Fu, Y. and R. Shen, "GA Based CBR Approach in Q&A system," *Expert Systmes with Applications*, 2003.
13. Hansen, J., R. D. Meservy, and L. E. Wood, "Case-based Reasoning: Application Techniques for Decision Support," *Intelligent Systems in Accounting, Finance and Management*, 4, 1995, pp.137-146.
14. Harp, S. A. and T. Samad, "The Genetic Synthesis of Neural Network Architectures," In Davis, L. (Eds.), *Handbook of Genetic Algorithms,* Van Nostrand Reinhold, New York, 1991, pp.202-221.
15. Ichikawa, Y. and Y. Ishii, "Retaining Diversity of Genetic Algorithms for Multivariable Optimization and Neural Network Learning," *Proc. of IEEE International Conference on Neural Networks*, San Francisco, 2, 1993, pp.1110-1114.
16. Ishibuchi, H., K. Nozaki, and N. Yamamoto, "Selecting Fuzzy Rules by Genetic Algorithm for Classification," *Proc. of IEEE International Conference on Fuzzy Systems*, San Francisco, 2, 1993, pp.1119-1124.
17. Jarmulak, J., S. Craw, and R. Rowe, "Self-Optimising CBR Retrieval," *Proceedings 12th IEEE International Conference on Tools with Artificial Intelligence*, 2000, pp. 376–383.
18. Jo, H., I. Han, and H. Lee, "Bankruptcy Prediction Using Case-based Reasoning, Neural Networks, and Discriminant Analysis," *Expert Systems with Applications*, 13(2), 1997, pp.97-108.
19. Kim, K. S. and I. Han, "The Clustering-indexing Method for Case-based Reasoning Using Self-organizing Maps and Learning Vector Quantization for Bond Rating Cases," *Expert Systems*

with Applications, 12, 2001, pp.147-156.
20. Lee, K., H. Kim, J. Lee, and T. Kim, "Case- and Constraint-Based Project Planning for Apartment Construction," *AI Magazine*, 19(1), 1998, pp.13-24.
21. Liao, T. W., Z. M. Zhang, and C. R. Mount, "A Case-based Reasoning System for Identifying Failure Mechanism," *Engineering Applications of Artificial Intelligence,* 13, 2000, pp.199-213.
22. McLachlan, G., *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York, 1992.
23. Mechitov, A. I., H. M. Moshkovich, D. L. Olson, and B. Killingsworth, "Knowledge Acquisition Tool for Case-based Reasoning System," *Expert Systems with Applications*, 9(2), 1995, pp.201-212.
24. Mejasson, P., M. Petridis, B. Knight, A. Soper, and P. Norman, "Intelligent Design Assistant (IDA): A Case Base Reasoning System for Material and Design," *Materials and Design*, 22, 2001, pp.163-170.
25. Montana, D. J. and L. Davis, "Training Feedforward Networks Using Genetic Algorithms," *Proc. of 11th Int. Joint Conference on Artificial Intelligence*, Detroit, 1989, pp.762-767.
26. Morris, B. W., "SCAN: A Case-based Reasoning Model for Generating Information System Control Recommendations," *Intelligent Systems in Accounting, Finance and Management,* l(3), 1994, pp.47-63.
27. O'Roarty, B., D. Patterson, S. McGreal, and A. Adair, "A Case-based Reasoning Approach to the Selection of Comparable Evidence of Retail Rent Determination," *Expert Systems with Applications*, 12(4), 1997, pp.417-428.
28. Park, D., A. Kandel, and G. Langholz, "Genetic-based New Fuzzy Reasoning Models with Application to Fuzzy Control," *IEEE Transactions on Systems, Man and Cybernetics*, 24(1), 1994, pp.39-47.
29. Pinches, G. E. and K. A. Mingo, "A Multivariate Analysis of Industrial Bond Ratings," *Journal of Finance*, 28(1), 1973, pp.1-18.
30. Riesbeck, C. K. and R. C. Schank, *Inside Case-Based Reasoning*, Lawrence Erbium Associates, Hillsdale, NJ, 1989.
31. Schffer, J. D., D. Whitley, and L. J. Eshelman, "Combinations of Genetic Algorithms and Neural Networks: A Survey of the State of the Art," *Proc. of the International Workshop on Combinations of Genetic Algorithms and Neural Networks*, Baltimore, 1992, pp.1-37.
32. Shin, K. S. and I. Han, "A Case-based Approach Using Inductive Indexing for Corporate Bond Rating," *Decision Support Systems*, 32(1), 2001, pp.41-52.
33. Shin, K.S. and I. Han, "Case-based Reasoning Supported by Genetic Algorithms for Corporate Bond Rating," *Expert Systems with Applications*, 16(2), 1999, pp.85-95.
34. Stahl, A. and T. Gabel, "Using Evolution Programs to Learn Local Similarity Measures," *Proceedings of the 5th International Conference on Case-Based Reasoning (ICCBR)*, Trondheim, Norway, 2003.
35. Stahl, A., "Defining Similarity Measures: Top-Down vs. Bottom-Up," *Proceedings of the 6th European Conference on Case-Based Reasoning (ECCBR)*, Aberdeen, Schottland, UK, 2002.
36. Syswerda, G., "Uniform Crossover in Genetic Algorithms," In Schaffer, J. D. (Eds.), *Proc. 3rd Int.Conf. Genetic Algorithms*, Morgan Kaufmann, San Maeto, CA, 1989.
37. Tsatsoulis, C. and B. Stephens, "Using Genetic Algorithms to Discover Selection Criteria for Contradictory Solutions Retrieved by CBR," *Lecture Notes in Computer Science*, 2689, 2003, pp. 567 – 580.
38. Wang, Y. and N. Ishii, "A Method of Similarity Metrics for Structured Representations," *Expert Systems with Applications*, 12(1), 1997, pp.89-100.
39. Wong, F. and C. Tan, "Hybrid Neural, Genetic and Fuzzy Systems," In Deboeck, G. J. (Eds.), *Trading on the Edge*, John Wiley, New York, 1994, pp. 245-247.