THE FUZZY CLUSTERING ALGORITHM AND SELF-ORGANIZING NEURAL NETWORKS TO IDENTIFY POTENTIALLY FAILING BANKS

이기동*

목 차

Abstract

Using 1991 FDIC financial statement data, we develop fuzzy clusters of the data set. We also identify the distinctive characteristics of the fuzzy clustering algorithm and compare the closest hard-partitioning result of the fuzzy clustering algorithm with the outcomes of two self-organizing neural networks. When nine clusters are used, our analysis shows that the fuzzy clustering method distinctly groups failed and extreme performance banks from control (healthy) banks. The experimental results also show that the fuzzy clustering method and the self-organizing neural networks are promising tools in identifying potentially failing banks.

* 인천대학교 경영학부, 인천 남구 도화동 177, kdlee@incheon.ac.kr

I. INTRODUCTION

The risk of a major banking crisis similar to the savings and loans debacle of the eighties suggests that we investigate the means of identifying banks with potential problems even before these banks face a severe liquidity or solvency crisis. Furthermore, the interdependencies between financial markets and financial institutions have the potential of national and international contagion (Benston & Kaufman, 1995).

This study uses a fuzzy clustering approach and a self-organizing neural network approach to identify potentially failing banks. The fuzzy clustering approach uses probability information in assessing banks rather than a dichotomous scale of 'failed' or 'healthy' banks. To provide a meaningful interpretation for the fuzzy clustering results, we compare the results of the closest hard-partitioning of the fuzzy clustering with the outcomes of two self-organizing neural networks.

The purpose of this study is to test whether these two approaches (fuzzy clustering and self-organizing neural networks) can be utilized as classification tools. In other words, we test whether or not these methods can distinguish healthy banks from failed or problem banks. These novel approaches have not been applied previously in the banking literature. Previous studies have used regression analysis (Meyer & Pifer, 1970), multiple discriminant analysis (Sinkey, 1977), multivariate logit or probit analysis (Bonvenzi et al., 1983; Espahbodi, 1991; Gilbert et al., 1990; Hanweek, 1977; Martin, 1977), robust regression analysis (Booth, 1982; Booth, et al., 1989), Cox proportional hazards model (Lane et al., 1986), factor-analytic approach (West, 1985), and back propagation neural network (Tam & Kiang, 1992).

Our study is organized as follows. Section two describes the data set and methodology employed. The methodology section includes a brief description for the fuzzy clustering algorithm and self-organizing neural networks. The experimental results are presented in section three followed by limitations and conclusions.

II. DATA AND METHODOLOGY

Data Set and Variables

The data for this study is obtained from the FDIC Call and Income Reports available on magnetic tapes. The data set contains

93 observations (including 3 failed) in 1991 and 248 observations (including 8 failed) in 1992. The following five variables were identified, after reviewing the banking literature, to develop fuzzy clusters and neural networks.

1) net income to total assets (NITA)

2) net loan losses to adjusted assets (NLLAA)

3) non-performing loans to total assets (NPLTA)

4) net loan losses to total loans (NLLTL)

5) net loan losses and provision for loan losses divided by net income (NLLPLLNI)

The last variable NLLPLLNI serves only as a marker to identify extreme performance banks. It is not used in fuzzy clustering or neural networks. We define extreme performance bank as those whose NLLPLLNI value is higher than +3 or less than −3. We also define problem banks as failed or extreme performance banks.

Fuzzy Clustering

Zadeh (1965) initiated the theory of fuzzy sets. His pioneering work led to the development of possibility theory. Groups of objects that do not have clear boundaries often appear in reality. The imprecision of such classes is expressed in the possibility that an element may belong to an indeterminate grade, rather than to a certain group.

The objective of cluster analysis is to find a convenient and valid organization of data. Clustering of data leads to fuzziness when grouping of elements into classes does not result in sharply defined boundaries leading to fuzziness. Fuzzy clustering is a special type of cluster analysis. It is 'capable of describing ambiguity in the data, such as the existence of points that lie between two clusters' (Rousseeuw et al., 1987, p. 1). Fuzzy cluster analysis provides probability information on the likelihood of failure through the use of the probability consistency principle. The membership coefficients developed by the fuzzy clustering algorithm are possibilities of group membership in the sense of Zadeh (Kandel, 1982). One such procedure to develop fuzzy clusters which we use is called FANNY (Kaufman & Rousseeuw, 1990).

Self-Organizing Neural Networks

Kohonen (1987) has often been cited with developing self-organizing neural networks. Self-organizing neural networks are particular types of neural networks in that they can learn to detect relationships among their inputs independently. Interest in self-organizing neural networks has grown significantly due partly to their unsupervised learning capabilities. Recent research studies of self-organizing neural networks have appeared in many fields, for example, classification (Corridoni et al., 1996; Schonweiler et al., 1996; Deschenes & Noonan, 1995), pattern recognition (Xin-Hua & Hopke, 1996), clustering (Murtagh, 1995), and forecasting (Der Voort et al., 1996).

There are in large two types of learning algorithms in self-organizing neural networks: a competitive algorithm vs. a self-organizing learning algorithm. A Competitive Neural Network (CNN) updates the winner's weight only over time. However, in a Self-Organizing Neural Network (SONN), the weights of the winner and its neighbors get updated through learning iteration. Accordingly, SONN learns to classify not only input vectors but also the features of their input vectors (Demuth & Beale, 1997).

CNN and SONN are used to provide a benchmark to fuzzy clustering algorithm. By comparing the results of the closest hard-partitioning of the fuzzy clustering with the outcomes of these two self-organizing neural networks, we attempt to identify unique features and characteristics of the fuzzy clustering algorithm and to recognize the usefulness of fuzzy clustering compared with these self-organizing nets.

Ⅲ. EXPERIMENTAL RESULTS

We run both fuzzy cluster analysis and self-organizing neural networks with 93 observations for 1991 bank failures. With respect to 248 observations from 1992 bank failures, we divided the sample into three sub-groups, containing 100 observations each in the first two groups, and 48 observations in the third group. This procedure allowed us to test and compare results from different samples.

For sake of brevity, we present results for one subgroup of the 1992 failures containing 100 observations. The data sets for which results are presented include the following failed banks: Bremen (010), American (041), North American (075). Note that the numbers within parentheses are bank identification numbers. This data set also includes 97 banks containing 13 extreme performance and 84 control (healthy) banks. The results for the other data sets for 1992 and for 1991 failures are qualitatively similar.

Results of Fuzzy Clustering

Table 1 shows the reproduction of a portion of the membership coefficients of fuzzy clustering result for 3 failed banks, 17 extreme performance banks, and the remaining 80 healthy (control) banks in the 1992 data set.

Table 1. Membership Coefficients for bank

ID	Cluster #								
	1	2	3	4	5	6	7	8	9
001	.15	.89	.15	.14	.13	.14	.11	.09	.00
002	.07	.24	.07	.04	.11	.04	.18	.24	.00
003	.16	.09	.16	.14	.13	.14	.10	.09	.00
004	.05	.29	.06	.03	.09	.03	.16	.29	.00
005	.15	.09	.15	.14	.13	.14	.11	.09	.00
M	M	M	M	M	M	M	M	M	M

Dunn's partition coefficient = .17(its normalized version = .07)

The first column of Table 1 represents bank identification number and the following columns indicate cluster identification. First, the fuzzy clustering program produces desired results with nine clusters. Second the membership coefficients for any bank sums to 1.0. Third, the table values give membership coefficients which represent the degree of belonging to each cluster, ranging from 0 to 1. For instance, looking at the first row, we see that bank 001 belongs 15 percent to cluster 1, 14 percent to cluster 4, and so on. Fourth, the normalized version of the Dunn's coefficient is 0.07 suggesting clustering to be fuzzy. The normalized version of the Dunn's coefficient varies from 0 (entirely fuzzy) to 1 (hard cluster), and is independent of the number of clusters. Fifth, the membership coefficients of clusters 2, 7, and 8 are of a similar nature but quite distinct from values in clusters 1, 3, 4, 5, and 6. Sixth, the membership coefficients in cluster 9 are generally smaller than any other cluster.

We then construct the closest hard-partitioning of the sample banks by the fuzzy clustering algorithm. By the closest hard-partitioning, we mean that a bank belongs to a cluster where its membership coefficient is highest. All the failed banks are in clusters containing extreme performance banks (Clusters 1, 3, 4, and 5). Further examination of membership coefficients in Cluster 6 reveals that these values are close to the values in clusters containing failed and extreme performance banks (Clusters 1, 3, 4, and 5). Therefore, it is concluded that these identified clusters can be divided into two major groups: a healthy-bank group (containing clusters 2, 7, 8, and 9) and a problem bank group (containing clusters 1, 3, 4, 5 and 6).

Results of Self-Organizing Neural Networks

Since the fuzzy clustering has identified 9 distinctive clusters in the data set, we also design both the CNN and the SONN with 9 output neurons (or clusters) for comparison purpose. Table 2 summarizes the clustering results of these three algorithms -- the closest hard-partitioning of the fuzzy clustering, CNN, and SONN where it shows the number of healthy banks identified for each algorithm.

Table 2. Comparison among the three algorithms

Closest-hard-fuzzy clustering, CNN, and SONN

Algorithm	Closest-Fuzzy	CNN	SONN
# of healthy banks	60	65	62
identified by each method	012 026 048 061	011 035 042 055 062 074 081 085 100	011 042 062 074 085 100
Identified by All three methods	002 004 008 013 014 015 016 018 019 020 021 022 023 024 025 028 029 030 031 032 033 034 036 037 039 040 043 044 046 047 049 050 051 052 053 054 057 058 059 063 065 068 069 070 071 073 077 079 083 086 088 089 090 093 097 098		

The closest hard-partitioning of the fuzzy clustering method identifies 60 healthy banks out of 100 banks while CNN and SONN correctly classify 65 and 62 healthy banks, respectively. The last row of Table 2 shows all healthy banks identified by all three methods. The gray area across the columns contains 56 healthy banks. The only difference in performance among the three algorithms is shown just above the gray area in each column. For example, just above the gray area in column 2, the closest hard-partitioning fuzzy clustering method classifies banks 012, 026, 048, and 061 as healthy banks in additions to the healthy ones in the gray area.

Likewise, CNN identifies 9 healthy banks additionally while SONN also identifies 6 more healthy banks besides the banks in the gray area. CNN and SONN show better results than the closest hard-partitioning of the fuzzy clustering does. Notice that we present the two neural networks, CNN and SONN, to provide a performance benchmark for the closest hard-partitioning results of the fuzzy clustering, not with the fuzzy clustering itself. The usefulness of fuzzy clustering is its flexibility in clustering. As mentioned earlier, a membership coefficient represents the degree of belonging of a bank to a cluster. We presented the closest hard-partitioning result in order to provide an example of practical interpretation of the membership coefficients in Table 1. We may utilize other clustering schemes, for example, the second-closest hard-partitioning, the third-closest hard-partitioning, or any combination of these. These various partitioning possibility reflects flexibility of the fuzzy clustering in that it identifies ambiguity in data by capturing information about the data structure between clusters.

Ⅳ. LIMITATIONS AND CONCLUSIONS

The purpose of this paper is to use fuzzy clustering and self-organizing neural networks to identify banks that are likely to fail. The empirical results are based upon 3 failed banks and 13 extreme performance banks in 1991 and 8 failed banks and 35 extreme performance banks in 1992. We report the experimental results for a sub-sample containing 100 observations from the 1992 data. Some limitations of this paper are as follows.

First, the sample data set is somewhat limited. In this paper, we attempt to identify banks likely to fail using only one year-prior financial statement. It may be useful to use two or three years prior financial statement data. Doing so would not only identify potentially failing banks but may also pinpoint when the financial condition of a bank has begun to deteriorate. Second, in order to provide a practical interpretation of the membership coefficients of a bank, we built the closest hard-partitioning case of the fuzzy clustering result and compared that case with the results of two self-organizing neural networks. Using other partitioning schemes, i.e., the second-closest, the third-closest hard-partitioning, or any combination of these partitioning schemes would be useful for bank failure prediction literature to explore. Also comparison among different algorithms (i.e., statistical or other network models) would be desirable to test the robustness of this study.

The large number of bank failures in the eighties and nineties is of concern to the regulators and policy makers. However, the challenge for the regulators is to identify these banks promptly before they go down the path of failure. This study makes a case that fuzzy clustering could be a useful and a powerful tool for bank regulators in identifying potentially failing banks. We demonstrate the clustering capability of fuzzy clustering by comparing the closest hard-clustering result with those of the two self-organizing neural networks. The two self-organizing neural networks outperform the closest hard-partitioning of fuzzy clustering in identification of potentially failed banks. The normalized version of the Dunn's coefficient of 0.07 suggests that we have been successful in obtaining appropriate number of fuzzy clusters. Our results also show that the clustering capabilities of fuzzy clustering and of self-organizing neural networks look promising for identifying potentially failing banks.

REFERENCES

1. Benston, G. J. & Kaufman, G. G. (1995). Is the Banking Payments System Fragile Journal of Financial Services Research, 9, 209-240.
2. Booth, D. E. (1982). The Analysis of Outlying Data Points Using Robust Regression: A Multivariate Problem-Bank Identification Model. Decision Sciences, 13(1), 71-81.
3. Booth, D. E., Alam, P., Ahkam, S. N., & Osyk, B. (1989). A Robust Multivariate Procedure for the Identification of Problem Savings and Loan Institutions. Decision Sciences, 20, 320-333.
4. Bonvenzi, J. F., Martin, J. A., & McFadden, J. (1983). Commercial Failure Prediction Models. Federal Reserve Bank of Atlanta Economic Review, 68(11), 14-26.
5. Corridoni, J.M., Del Bimbo, A., & Landi, L. (1996). 3D Object Classification Using Multi-Object Kohonen Networks. Pattern Recognition, 29(6), 919-935.
6. Demuth, H. & Beale, M. (1997). MTLAB: Neural Network Toolbox, User's Guide, Version 2, Math Works, Inc., Chapter 8.
7. Der Voort, M., Dougherty, M., & Watson, S. (1996). Combining Kohonen maps with arima time series models to forecast traffic flow. Transportation Research Part C: Emerging Technologies, 4(5), 307-318.
8. Deschenes, C. J., & Noonan, J. (1995). Fuzzy Kohonen Network for the Classification of Transients Using the Wavelet Transform for Feature Extraction. Information Sciences, 87(4), November, 247-266.
9. Espahbodi, P. (1991). Identification of Problem Banks and Binary Choice Models. Journal of Banking and Finance, 53-71.
10. Federal Deposit Insurance Corporation (1993, 1995). Annual Report Washington, D. C.
11. Gilbert, L. R., Menon, K., & Schwartz, K. B. (1990). Predicting Bankruptcy for Firms in Financial Distress. Journal of Business Finance & Accounting, 17(1), 161-171.
12. Hanweek, G. (1977). Predicting Bank Failures. Research Papers in Banking and Financial Economics. Financial Studies Section, Washington, D.C.: Board of Governors of the Federal Reserve System.
13. Kandel, A., (1982). Fuzzy Techniques in Pattern Recognition. New York: John Wiley & Sons.
14. Kaufman, L. & Rousseeuw, P. T. (1990). Finding Groups in Data: An Introduction to Cluster Analysis. New York: John Wiley.
15. Kohonen, T. (1987). Self-Organizing and Associative Memory. 2nd Edition,

Berlin: Spring-Verlag.

16. Lane, W. R., Looney, S. W., & Wansley, J. W. (1986). An Application of the Cox Proportional Hazards Model to Bank Failure. Journal of Banking and Finance, 10, 511-531.

17. Martin, D., (1977). Early Warning of Bank Failure: A Logit Regression Approach. Journal of Banking and Finance, 11, 249-277.

18. Meyer, P. A. & Pifer, H. W. (1970). Prediction of Bank Failures. Journal of Finance, 25, 853-868, September.

19. Murtagh, F. (1995). Interpreting the Kohonen self-organizing feature map using contiguity-constrained clustering. Pattern Recognition Letters, 16(4), 399-408.

20. Rousseeuw, P. J., Derde, M. P., & Kaufman, L. (1987). Principal Components of a Fuzzy Clustering. Trends in Analytical Chemistry, 8, 249-250.

21. Sinkey, J. F., Jr. (1977). Identifying Large Problem/Failed Banks: The Case of Franklin National Bank of New York. Journal of Financial and Quantitative Analysis, 12, 779-800.

22. Tam, Y. K. & Kiang, M. Y. (1992). Managerial Applications of Neural Networks: The Case of Bank Failure Predictions. Management Science, 38, 7, 926-947.

23. West, R. C. (1985) A Factor-Analytic Approach to Bank Condition. Journal of Banking and Finance, 9, 253-266.

24. Xin-Hua, S. & Hopke, P.K. (1996). Kohonen neural network as a pattern recognition method based on the weight interpretation. Analytica Chimica Acta, 334(1-2), 7-66.

25. Zadeh, L. A. (1965). Fuzzy Sets."Information and Control, 8, 338-353.