The use of Case-Based Reasoning for Financial Market Monitoring

Sung Kwon Han*, Kyong Joo Oh**, Tae Yoon Kim***

*Department of Information and Industrial Engineering, Yonsei University (hansk@yonsei.ac.kr)

**Department of Information and Industrial Engineering, Yonsei University (johanoh@yonsei.ac.kr)

***Department of Statistics, Keimyung University (tykim@kmu.ac.kr)

Abstract

This paper shows that case-based reasoning (CBR), an artificial intelligence technique, is a quite efficient tool in monitoring financial market against its possible collapse. For this purpose, daily financial condition indicator (DFCI) monitoring financial market is built on CBR and its performance is compared to DFCI on neural network. This study is empirically done for the Korean financial market.

Keywords: Financial market monitoring; Case-Based Reasoning (CBR); Neural network

1. Introduction

During the last decade, every financial market in the world more often than not had bump into sudden collapses that cause a severe blow to the market (Eichengreen et al., 1995; Frankel & Rose, 1996; Kaminsky & Reinhart, 1999). To avoid or prepare for such sudden collapse, it is necessary to have a proper tool that monitors the market efficiently. Recently Kim et al. (2004a, b) found that neural network (NN) may monitor financial market effectively thanks to its overfitting tendency. Problem with NN as a monitoring tool, however, is its difficulty in updating since it usually requires a large amount of training data for its proper functioning (Refenes, 1995). This is not desirable for monitoring financial market since modern financial market tends to undergo change of its mechanism over a short period of time and hence needs to be updated regularly with relatively small amount of data. To resolve such updating difficulty, we propose case-based reasoning (CBR) for financial market monitoring and examine its efficiency. It will be shown that CBR is more efficient than NN. Recall that CBR is known as a very useful artificial intelligence technique that could be efficiently trained on relatively small amount of data.

For our discussion, daily financial condition indicator (DFCI) is built with CBR and then compared to its counter part (DFCI with NN). Specifically DFCI's built with CBR and NN respectively are constructed for the Korean financial market and compared. This paper consists of as follows. Following Section 1, Section 2 describes DFCI construction procedure and reviews CBR briefly. Section 3 establishes and compares DFCI's for the Korean financial market. Concluding remarks are given in Section 4.

2. DFCI construction

The core of CBR is the case base which stores a collection of cases or memories from the past. If a new event or problem occurs, CBR recognizes it as a target case and retrieves multiple exemplars of the target case from the case base (i.e., source case is found). Then CBR is to produce a response or output corresponding to the new event. Inside such procedure, two key parameters, the locale k (or number of neighbors for the source case) and the weights of input variables are to be specified (Kolodner, 1991). Figure 1 describes this CBR process including its update procedure (Shin & Han, 1999; Roh et al., 2002). Note that CBR is known to have potential to exceed the performance of neural networks at some events (Kolodner, 1991, 1993; Deboeck, 1994).

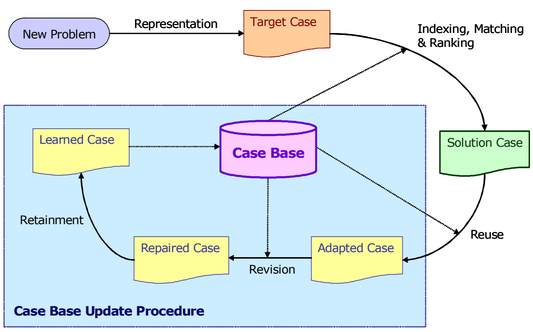


Figure 1. CBR process

DFCI is designed to monitor financial market by classifying daily financial market condition and hence is to be equipped with a set of categories that financial market condition can belong to. For this, DFCI of this paper focuses volatility of financial market since it is well known that market crash could be well defined in terms of volatility (Choudhry, 1996; Johansen, 2004). More precisely, conditions of financial market are categorized into three states according to its volatility movement: (ⅰ) stable period (SP), (ⅱ) unstable period (UP), (ⅲ) crisis period (CP). SP is marked by relatively low volatility while UP is a zone characterized by a sudden increase of volatility and rapid swings in market sentiment. At CP, the financial market experiences collapse usually indicated by relatively high volatility. These three states or periods are considered as a pattern set to be classified and hence serve as a basis of the DFCI. See also Kim et al. (2004a, b) for further references of these.

DFCI construction procedure consists of two phases where three financial markets (i.e., stock price index (SPI), interest rate (INT), foreign exchange rate (FER)) are considered separately or and then combined (see Figure 2). At Phase 1, DFCI is established for each financial market (sub-DFCI) and then, at Phase 2, sub-DFCI's are integrated into one DFCI by genetic algorithm (GA).

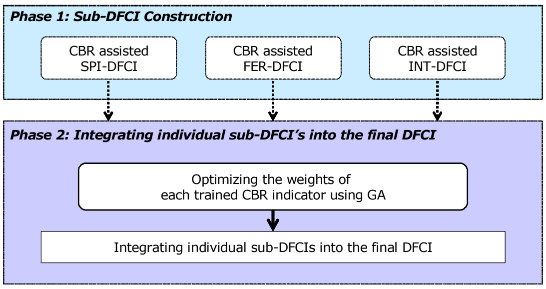


Figure 2. DFCI construction architecture

Phase 1: Sub-DFCI Construction

Input and target variables are selected with due consideration paid to the final task of DFCI. From the past period containing a market collapse, define SP, UP and CP through the selected variables, which basically establishes the case base. Transformation of the original daily financial variables and expert's opinion always proves quite useful at this step. Then the CBR-assisted sub-DFCI is ready to function when the weights of the input variables and the locale or number of neighbors k are specified. When a new case is given, indeed, it is to be retrieved on the case base to produce proper response (i.e., SP, CP or UP). This phase can be done for each financial market.

Phase 2: Integrating individual sub-DFCI's into final DFCI

At this phase, GA may be employed to combine the individual (sub-) DFCI's for each market in the following manner

$$DFCI_t = w_1 S_t + w_2 F_t + w_3 I_t \qquad (1)$$

where w_1, w_2 and w_3 are the weights given to S_t, F_t and I_t representing sub-DFCI's of SPI, FER and INT, respectively. Notice that S_t, F_t and I_t take on one among 1 (SP), 2 (UP), or 3 (CP). Finding the optimal weights in (1) is the core part of phase 2. More precisely, the objective function $E(w)$ given by

$$E(w) = \sum_{t=1}^{n} (w_1 S_t + w_2 F_t + w_3 I_t - a)^2 \qquad (2)$$

is minimized over $w_1, w_2, w_3 \geq 0$ satisfying $w_1 + w_2 + w_3 = 1$ where n is the number of data in the case base. Fixing 'a=2' at (2) indicates that the weights are optimized so that DFCI may issue as many '2' as possible. Fixing a=2 strategy is desirable if one wants to have a sensitive early warning systems (EWS) since '2' suggests that financial market enters gray zone that may proceed to either SP or CP.

3. An Empirical Study

Korea had experienced the economic crisis during 1997-1998 which had initially started as financial market collapses in late 1997. Since then much attention has been paid to efficient monitoring of financial market. Kim et al. (2004a, b) proposed using NN for stock market monitoring but found that the trained neural network is hard to update. In this section, DFCI is reconstructed by employing CBR instead of NN. Throughout this section, the variable names given in Table 1 are used.

Table 1. Input variables considered

Variable name	Numerical formula	Description
IND	x_t	Index or rate
DRF	$P_t = \dfrac{x_t - x_{t-1}}{x_{t-1}}$	Daily rise and fall rate
MA(m)	$\overline{P}_{m,t} = \sum_{i=t-(m-1)}^{t} P_i$	m -day moving average
MV(m)	$S^2_{m,t} = \sum_{i=t-(m-1)}^{t} (P_i - \overline{P}_{m,t})^2$	m -day moving variance

KOSPI 200, Korea won/U.S. dollar exchange rate, and Korea Treasury bill rate with 3-year maturity are considered as three major financial variables. In order to establish the case base for CBR, the movements of 1997 financial markets are examined closely. A rough look at the three major variables in 1997 reveals that all the three variables appear to change their movements around October 1997 (see Figure 3). To check structural changes of the three variables during the crisis, DRF P_t is calculated. The main reason for calculating P_t is that it is expected to efficiently measure increased instability or volatility of the financial markets with its frequency and amplitude. As expected, such is easy to notice in Figure 4. Around October or November of 1997, indeed, there was a strong signal of a volatility increase in each DRF movement. Note that, in December 1997, Korea was officially put under the IMF financial rescue program.

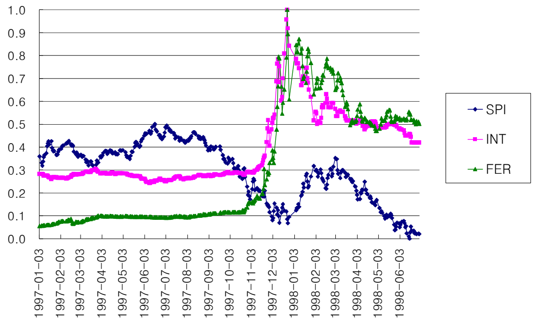
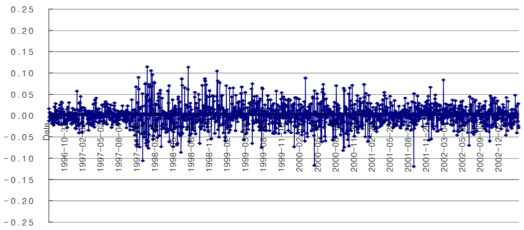
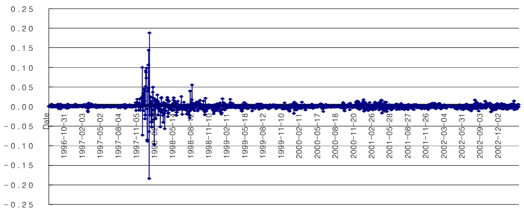


Figure 3. Korea stock price index, foreign exchange rates and interest rates of 1997, which are scaled from 0 to 1.

(a) Stock price index



(b) Foreign exchange rates



(c) Interest rates

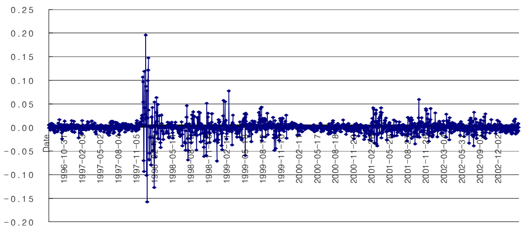
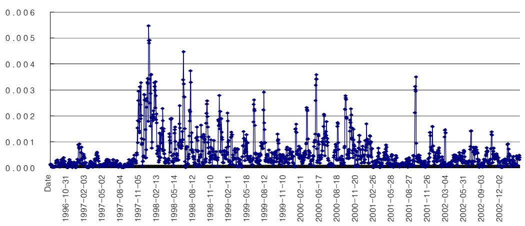


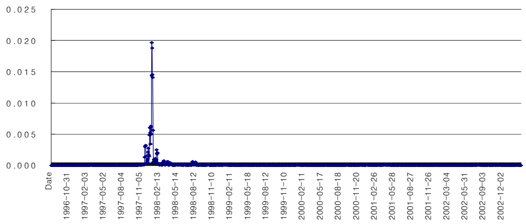
Figure 4. DRF, , from May 1996 to March 2003

To utilize DRF fully, its 5-day moving average($\overline{P_t}$ (MA(5)) and 5-day moving variance S_t^2 (MV(5), another measure of volatility) are studied for each major financial variable (see Figure 5). A rather short period of 5 days was chosen here for the moving average in order to take into account the visibly clear non-stationarity of DRF from Figure 4. Figure 5 shows that all three MV(5)'s have sudden increases in its volatility around October 1997.

(a) Stock price index



(b) Foreign exchange rates



(c) Interest rates

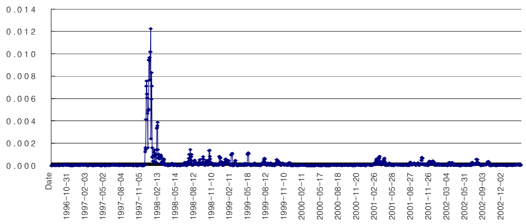


Figure 5. 5-day moving variance of daily rise and fall rate, S_t , from May 1996 to March 2003

For building the case base, the period between 1997 and early 1998 containing the crisis is considered. Since sudden increase of moving variance (S_t^2) is a signal of volatility increase, UP is first established around that signal. In fact, UP for each financial variable is established as Sep. 19 – Oct. 21 for SPI, Oct. 27 – Nov. 30 for FER, and Nov. 13 – Dec. 12 for INT, each of which is basically designed to contain the point of sudden volatility increase. SP and CP are established before and after the UP, respectively. Table 2 shows the case base for each variable. Notice that the lengths of SP, CP and UP are almost same. Input and output variables for each DFCI are selected as shown in Table 3 where additional moving averages and variances (MA(m) or MV(m), m=20, 60) are included. For activating each sub-DFCI assisted by CBR, the number of neighbors (k) is fixed as 9, and the weights of input variables are given 1.

Table 2. Specific dates of SP, UP and CP for case bases

Index	SP	UP	CP
SPI	Aug. 14, 97 – Sep. 18, 97	Sep. 19, 97 – Oct. 21, 97	Oct. 22, 97 – Nov. 20, 97
FER	Sep. 19, 97 – Oct. 26, 97	Oct. 27, 97 – Nov. 30, 97	Dec. 1, 97 – Jan. 13, 98
INT	Oct. 14, 97 – Nov. 12, 97	Nov. 13, 97 – Dec. 12, 97	Dec. 13, 97 – Jan. 22, 98

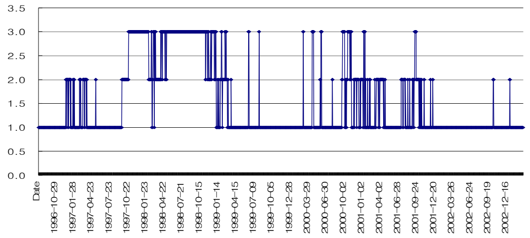
Table 3. List of input and output variables for each sub-DFCI

Indicator	Input variables	Output variables
DFCI (SPI)	IND, DRF, MA(5), MV(5)	SP: 1 UP: 2 CP: 3
DFCI (FER)	DRF, MA(5), MV(5), MA(20), MV(20), MA(60), MV(60)	
DFCI (INT)	DRF, MA(5), VA(5), MA(20), MV(20), MA(60), MV(60)	

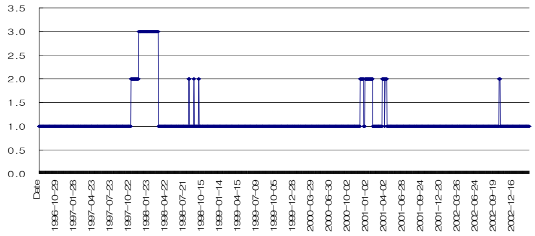
Each sub-DFCI is applied to testing period other than its own case base period. The overall classification results of each sub-DFCI are given in

Figure 6, which shows that sub-DFCI for SPI fluctuates highly than other financial variables. When the three sub-DFCI's are combined into the final DFCI by GA with a=2 in (2), its weights to each sub-DFCI and its classification results are given in Table 4 and Figure 7, respectively. In GA-optimization, crossover and mutation rate runs from 0.5 to 0.8 and from 0.05 to 0.06, respectively. One may notice that FER contributes least among the three major financial variables, which was somewhat anticipated since FER in 1997 was largely controlled by the policy authority.

(a) Stock price index



(b) Foreign exchange rates



(c) Interest rates

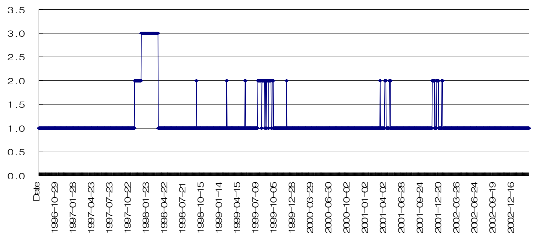


Figure 6. Classification results of individual sub-DFCI from May 96 to Mar. 03

Table 4. The weights given to three sub-DFCI's

Financial variables	Coefficients
SPI	0.5003
FER	0.0012
INT	0.4986

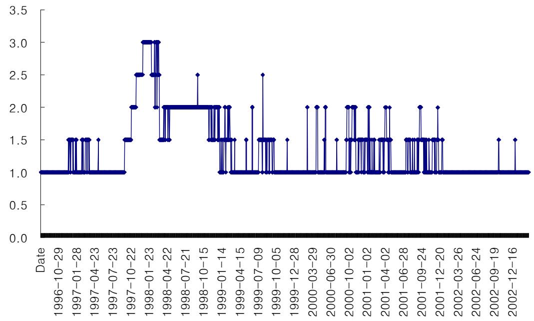


Figure 7. Classification result of the final DFCI from May 96 to Mar. 03

In order to check monitoring performance of the final DFCI, moves of DFCI are examined against the monthly chronicle of economic or financial events of Korea from Jan. 97 to Dec. 02 (see monitoring results in CBR of Table 5). Indeed frequency table of monitoring results for each month is checked against economic or financial events of the month. Recalling that December 1997 to 1998 is the period through which Korea suffered from lasting impact of economic crisis in 1997 (that period is shaded in Table 5), it is worthwhile noticing that DFCI have quite a few values jumping out of SP (=1) throughout that period. In addition those jumps appear quite often in the month when major internal or external shock hits the economy. Hence the final DFCI seems to perform pretty well in monitoring the Korean economy. One may also notice that DFCI itself may serve as an EWS if it is designed to issue a signal when it reaches a certain value, e.g.2

Table 5. Monthly chronicle of Korean economy from Jan. 97 to Dec. 02 (Training period: Aug. 97 - Jan. 98). The month which is highly fluctuated is shaded in the row of frequency of monitoring results in CBR and NN

Year	Month	Description of major events	Frequency of monitoring results in CBR					Frequency of monitoring results in NN				
			3	2.5	2	1.5	1	3	2.5	2	1.5	1
1997	01	Bankruptcy of Hanbo Group				8	13				6	15
	02					4	15				2	17
	03	Bankruptcy of Sammi Group				11	10				19	2
	04	Liquidity crisis of Jinro Group					22				4	18
	05	Liquidity crisis of Daenong and Kia Motors				1	18					19
	06						20					20
	07						22					22
	08						20					20
	09					8	11				13	6
	10	Bankruptcy of Kia Motors			8	14				8	14	
	11	Bankruptcy of Haitai Group and New Core Group		12	8				9	11		
	12	Bankruptcy of Halla Group	8	10				8	10			
1998	01	Bankruptcy of Nasan Group and Keukdong Construction Co.	17					15		2		

	02		7		11	2		9	4	6	1	
	03	Bankruptcy of Jeil Financial Co.	6		3	13		1	4	7	10	
	04				15	7				22		
	05	Bankruptcy of Donga Construction Co.			19					19		
	06				21					21		
	07	Major banks enter into workout.			22					22		
	08	Bankruptcy of major Insurance Companies.			21					21		
	09	Bankruptcy of Kukje Company			22					22		
	10	Bankruptcy of Kapul Group			20					20		
	11				11	10				21		
	12				13	6				17	2	
1999	01				2	7	11			2	13	5
	02				3	13	1			9	8	
	03	Unemployment rates reached 9%.				2	20				16	6
	04	Bankruptcy of Daehan Finance Co. and 9 minor banks					21					21
	05					2	18				9	11
	06	Bankruptcy of Samsung Motors				2	20					22
	07	Liquidity crisis of Daewoo Group		1		12	9				15	7
	08	Liquidity crisis of Daewoo Group				17	5				22	
	09	Liquidity crisis of Daewoo Group				11	9				16	4
	10						21					21
	11	6 companies of Daewoo Group enter into workout.				1	21				2	20
	12						20					20
2000	01						20				8	12
	02						20					20
	03				1		21					22
	04	Liquidity crisis of Hankuk and Daehan Trust & Investment Co.			5		13					18
	05	Saehan Group enters into workout.			1	1	18					20
	06	Liquidity crisis of Hyundai Group			2		18					20
	07	Business suspension of 3 minor banks				1	18					19
	08	Bankruptcy of Woobang Co. and Hankuk Financial Co.					22					22
	09	Liquidity crisis of Daewoo Motors			7	2	9			3	5	10
	10	Liquidity crisis of Donga Constructions Co.			8	9	4			6	9	6
	11	Liquidity crisis of Hyundai Construction Co.			1	12	9			2	17	3
	12	Default rates increase suddenly.				12	5				17	
20	01				3	2	14			1	13	5
	02	Bankruptcy of Hankuk Real Estate Trust Co.				2	18				2	18
	03	Bankruptcy of Korea Industrial Development Co.			1	13	6			1	4	15
	04	Liquidity crisis of Hyundai Construction Co.			5	10	6			16	5	
	05						22				6	16
	06						20					20
	07	Business Suspension for 18 major companies				11	10				12	9
	08					3	19				7	15
	09	9/11 terror for World Trade Tower in N.Y.			5	13	2			6	14	
	10					6	14				20	
	11					10	12				17	5
	12				1	5	13			1	7	11
2002	01					3	19				5	17
	02						16					16
	03						20					20
	04						20					20
	05						22					22
	06						18					18
	07						21					21
	08						21					21
	09						20					20
	10					2	20				1	21
	11						21					21
	12					1	18					19

Source: Kim and Kwon (2003)

For comparison, DFCI assisted by NN (i.e., monitoring results in NN of Table 5) is reconstructed by taking similar steps where the case base of CBR is used as training data set, logistic function is employed as an activation function with learning rate, momentum and initial weight given by0.1, 0.1 and 0.3, respectively, and the number of hidden layers is between 2 and7. Its monitoring results are given in Figure 8. Comparing it with Figure 7, it is easy to notice that NN suffers from too much noise, which is typical when NN is trained on relatively small amount of data. To compare two DFCI's in a more meaningful manner, noise from NN is smoothed out as follows. The classification results given by the real numbers between 1 and 3 are rounded to the nearest number among 1, 1.5, 2, 2.5, or 3, and they are given next to CBR results in Table 5. One may easily see that overall two DFCI's are consistent in monitoring financial markets. In the year 2000, however, NN rather remains silent while CBR issues some signals for the coming crisis (i.e., sudden default rate increase following the liquidity crisis of huge conglomerate Hyundai group). In Figure 9, technically more sound comparisons are made through comparing frequency distributions of monitoring results for each year. It is interesting to notice that CBR is more sensitive to market's deviation from the stable than NN since CBR on the average tends to register less tallies on "1 (stable market)" than NN. Thus apparently CBR addresses better efficiency in monitoring financial markets than NN.

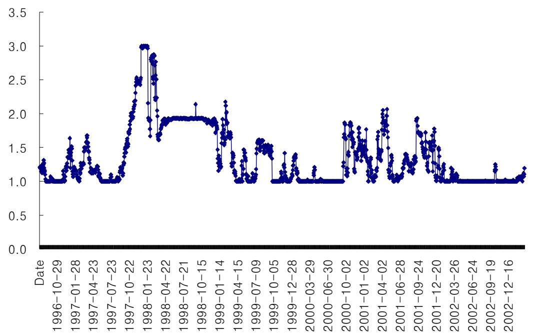
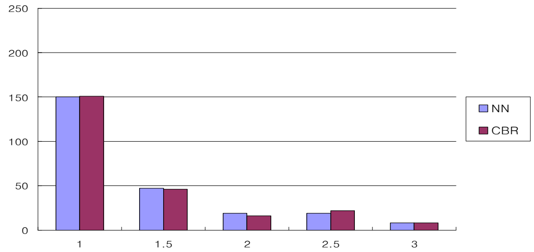
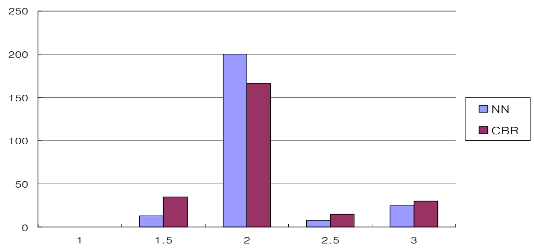


Figure 8. Classification result of the final DFCI assisted by NN from May 96 to Mar. 03

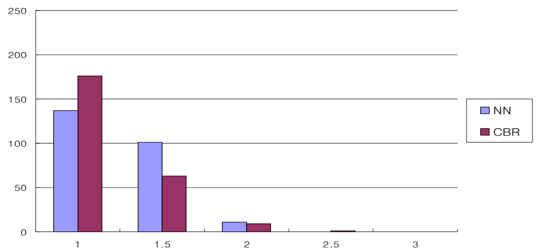
(a) 1997



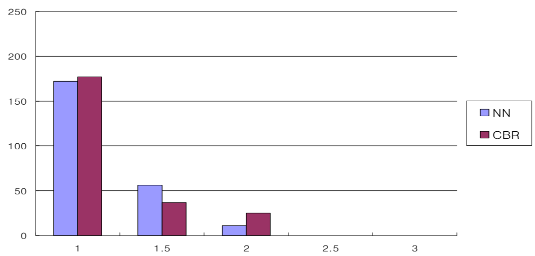
(b) 1998



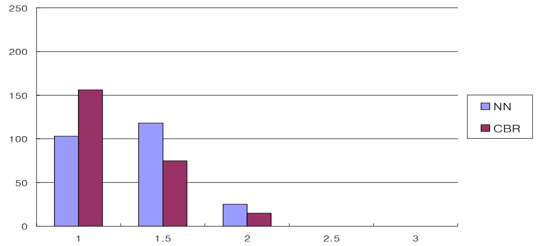
(c) 1999



(d) 2000



(e) 2001



(f) 2002

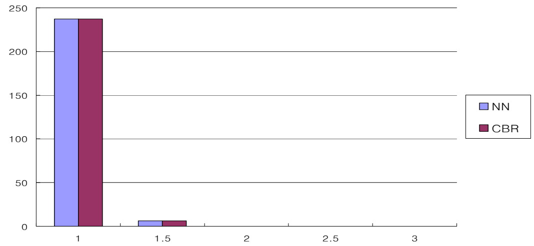


Figure 9. Frequency distributions of monitoring results (1, 1.5, 2, 2.5, or 3) from 1997 to 2002. NN and CBR in legend mean DFCI's assisted by them, respectively.

4. Concluding remarks

Modern financial market tends to be easily unstable since external or internal shocks spread easily through fast and massive electronic communication or transaction system. Recent economic or financial crises in the 1990s seem to provide typical examples of these, which emphasizes the need to have a proper monitoring tool of financial markets against its possible crash. Recently Kim et al. (2004a, b) found that NN may monitor financial market effectively thanks to its over-fitting tendency. In this paper we propose CBR as an efficient monitoring tool. Through empirical studies it is shown that CBR has competitive edge over its counterpart NN particularly when monitoring tool is to be updated with small amount of data.

References

Choudhry, T. (1996). Stock market volatility and the crash of 1987: evidence from six emerging markets. *Journal of International Money and Finance*, **25**(6), 969-981.

Deboeck, G. (1994). *Trading on the Edge*. New York: John Wiley & Sons.

Eichengreen, B., Rose, A., & Wyplosz, C. (1995). Exchange market mayhem: The antecedents and aftermath of speculative attacks, *Economic Policy*, **21**, 249-312.

Frankel, J.A., & Rose, A.K. (1996). Currency crashes in emerging markets: *An empirical treatment. Journal of International Economics*, **41**, 351-366.

Johansen, A. (2004). Origin of crashes in three US stock markets: *shocks and bubbles*. Physica A, **338**, 135-142.

Kaminsky, G., & Reinhart, C.M. (1999). The twin crises: The causes of banking and balance-of payments problems. *American Economic Review*, **89**(3), 473-500.

Kim, K., & Kwon, S. (2003). How has Korean economy changed during 5 years' financial crisis? Seoul, Korea: *Samsung Economy Research Institute*

(In Korean).
Kim, T.Y., Hwang, C., & Lee, J. (2004a). Korean economic condition indicator using a neural network trained on the 1997 crisis. *Journal of Data Science*, **2**, 371-381.
Kim, T.Y., Oh, K.J., Sohn, I., & Hwang, C. (2004b). Usefulness of artificial neural networks for early warning system of economic crisis. *Expert Systems with Applications*, **26**(4), 585-592.
Kolodner, J. (1991). *Improving human decision making through case-based decision aiding*. AI Magazine, **12**, 52-68.
Kolodner, J. (1993). *Case-Based Reasoning*. San Mateo, CA: Morgan Kaufmann.
Refenes, A.-P. (1995). *Neural Networks in the Capital Markets*. New York: Wiley.
Roh, T.H., Oh, K.J., & Han, I. (2002). The collaborative filtering recommendation based on SOM cluster-indexing CBR. *Expert Systems with Applications*, **25**, 413-423.
Shin, K.-S., & Han, I. (1999). Case-based reasoning supported by genetic algorithms for corporate bond rating. Expert Systems with Applications, **16**, 85.