

KOSPI 예측을 위한 NEWFM 기반의 특징입력 및 퍼지규칙 추출[☆]

Extracting Input Features and Fuzzy Rules for Forecasting KOSPI Stock Index Based on NEWFM

이 상 홍* 임 준 식**
Sang-Hong Lee Joon S. Lim

요 약

본 논문은 가중 퍼지소속함수 기반 신경망(Neural Network with Weighted Fuzzy Membership Functions, NEWFM)을 사용하여 생성된 퍼지규칙과 비중복면적 분산 측정법에 의해 추출된 최소의 특징입력을 이용하여, 1일 후의 KOSPI 예측을 하는 방안을 제안하고 있다. NEWFM은 KOSPI의 최근 32일 동안의 CPPn,m(Current Price Position of day n for n-1 to n-m days)을 이용하여 1일 후의 KOSPI 상승과 하락을 예측한다. 특징입력으로써 CPPn,m과 최근 32일간의 CPPn,m을 웨이블릿 변환한 38개의 계수들 중 비중복면적 분산 측정법을 적용하여 추출된 5개의 계수가 사용되었다. 제안된 방법으로 1991년부터 1998년까지의 실험군을 사용한 결과 평균 67.62%의 예측율을 나타내었다.

Abstract

This paper presents a methodology to forecast KOSPI index by extracting fuzzy rules based on the neural network with weighted fuzzy membership functions (NEWFM) and the minimized number of input features using the distributed non-overlap area measurement method. NEWFM classifies upward and downward cases of KOSPI using the recent 32 days of CPPn,m (Current Price Position of day n for n-1 to n-m days) of KOSPI. The five most important input features among CPPn,m and 38 wavelet transformed coefficients produced by the recent 32 days of CPPn,m are selected by the non-overlap area distribution measurement method. For the data sets, from 1991 to 1998, the proposed method shows that the average of forecast rate is 67.62%.

☞ keyword : Fuzzy Neural Networks, Rule Extraction, Wavelet Transforms, NEWFM, KOSPI

1. 서 론

패턴 분류나 진단, 예측 등을 위해 신경망 및 퍼지집합 이론을 결합한 적응적 의사결정지원 툴(adaptive decision support tool)인 퍼지신경망(Fuzzy Neural Network, FNN)이 제안되어 왔다[2,6,8-10,12,15,19,21-22]. 여러 가지 구조의 FNN이 학습(learning),

적응(adaptation), 규칙추출(rule extraction)을 위한 알고리즘과 더불어 제시된 바 있다[5,8,13,16,18]. 주어진 일련의 학습 데이터로부터 지식 추출을 위하여 자기 조직화 시스템(self-organizing system) 기반 퍼지신경망[7,20-21]이 개발되었다. Setnes[15]는 유전자 알고리즘을 이용하여 간단하고 정확도가 높은 퍼지규칙 기반 모델을 얻는데 초점을 맞추었다.

이러한 인공지능형 시스템을 이용하여 금융 예측을 위한 응용이 다양하게 이루어지고 있다. 과거의 통계적 지표를 기반으로 금융추세를 예측하는 비선형 시계열 예측 방법[4], 손익 결정 시스템

* 정 회 원 : 경원대학교 일반대학원 전자계산학과 박사과정
shleedosa@kyungwon.ac.kr

** 정 회 원 : 경원대학교 소프트웨어학부 교수
jslim@kyungwon.ac.kr(교신저자)

[2007/08/07 투고 - 2007/08/17 심사 - 2007/09/02 심사완료]

☆ 본 논문은 2007년도 정부재원(교육인적자원부)으로 한국 학술진흥재단의 BK21 지원을 받아 수행된 연구임

템을 이용한 주식거래[1], 경기전환점 예측[3,17] 등이 연구되고 있다.

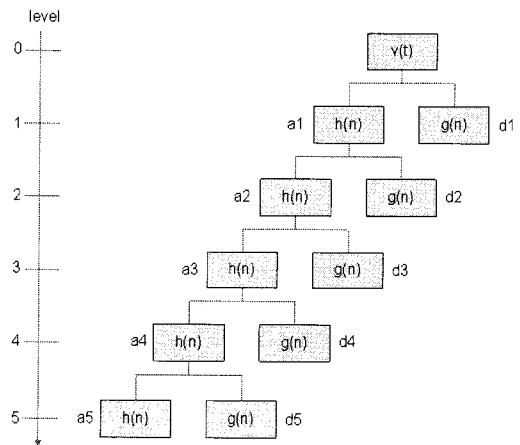
본 논문은 가중 퍼지소속함수 기반 신경망 (Neural Network with Weighted Fuzzy Membership Functions, NEWFM)[13]기반의 KOSPI 예측을 위한 퍼지규칙을 추출하고, 비중복면적 분산 측정법을 사용하여 특징입력수를 최소로 하는 방안을 제안하고 있다. NEWFM은 각 입력에 대해 퍼지 이론의 경계합(bounded sum)을 이용하여 분류를 위한 가중 퍼지소속함수들을 생성한다. 또한 비중복면적 분산 측정법을 이용하여 중요도가 낮은 특징입력을 제거하면서 최소의 특징입력을 추출한다. 이와 같이 단순화된 최소의 특징입력에 대한 가중 퍼지소속함수들로 간단한 퍼지규칙을 만들 수 있다. 또한 예측이나 분류의 효율을 낮추는 특징입력을 제거함으로써 더욱 효과적인 예측이나 분류 결과를 줄 수 있다.

NEWFM을 이용한 KOSPI 예측성능 비교를 위해 [11]의 실험군과 결과를 사용하였다. [11]에서는 KOSPI 예측성능 향상을 위해 잡음이 있는 데이터(noisy data) 제거를 위한 유전자 알고리즘을 제안하였다. [11]에서는 전체 실험군 중 약 30%~40%의 사례선택(instance selection)을 한 후 신경망을 이용하여 평균 65.47%의 1일 후 예측율을 보여주었다.

본 논문은 NEWFM기반의 비중복면적 분산측정법[14]을 사용하여 1일 후의 KOSPI 상승과 하락의 분류를 위한 중요도가 가장 높은 5개의 특징입력을 추출하고, 이 특징입력들에 대한 학습된 가중 퍼지소속함수를 제시하고 있다. 추출된 5개의 특징입력들이 다양한 실험군에 적용될 수 있음을 검증하기 위해 1991년부터 1998년까지의 실험군[11]들을 사용하였다. 그 결과 각각 72.41%, 62.07%, 62.72%, 61.02%, 65.52%, 79.31%, 67.24%, 70.69%로 평균 67.62%의 예측율을 보여주었다.

2. 웨이블릿 변환(Wavelet Transforms)

웨이블릿 변환은 신호처리에 있어서 시간상의 국소적인 특정 지점에서의 주파수 특성을 분석함으로써 전역적 주파수 특성 정보를 주는 푸리에 분석의 단점을 보완해 준다. 비연속 웨이블릿 변환은 시간-주파수 신호를 다양한 스케일의 비연속 신호로 분리해 준다. 그림 1은 이분 비연속 웨이블릿 분리 구현을 위한 필터뱅크를 보여주고 있다. detail이라고 하는 $g(n)$ 은 웨이블릿 계수와 관련된 FIR(finite impulse filters) 고대역필터(high-pass filters) 계수이고, approximation이라고 하는 $h(n)$ 은 스케일 함수 계수와 관련된 FIR 저대역필터(low-pass filters) 계수이다. 각 필터를 지나 그 길이가 반으로 감소한 $h(n)$ 신호로 다음 스케일 레벨에서 반복하여 변환 시킨다. 웨이블릿 변환으로 추출된 웨이블릿 계수는 웨이블릿 모함수(mother wavelet)에 대한 유사도로써 이는 스케일에 의해 주어진 시간에 따른 주파수 신호를 나타낸다. 그림 1에서 d_i 와 a_i 는 각각 스케일 레벨 i 의 detail과 approximation 계수들을 의미한다.



(그림 1) 스케일 레벨 5인 이분 비연속 웨이블릿 변환

본 논문에서는 1일 후의 예측을 위한 기술지표로 다음과 같이 $CPP_{n,m}$ (Current Price Position)을 제안하고 있다. $CPP_{n,m}$ 은 기준일 n 의 종가(closing price)가 $n-1$ 일부터 $n-m$ 일까지의 이동평균(moving

average)에 대해 어느 위치에 있는지를 나타내는 지표이다. 다음 식에서 C_n 은 기준일인 n 일의 종가이고, $MA_{n-1,n-m}$ 는 $n-1$ 일부터 $n-m$ 일까지의 이동평균을 가리킨다.

$$CPP_{n,m} = ((C_n - MA_{n-1,n-m}) / MA_{n-1,n-m}) \times 100$$

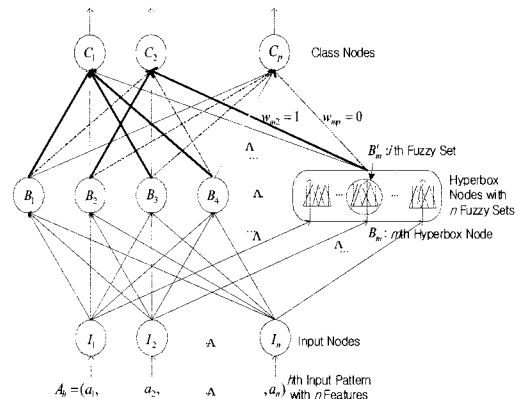
이 지표를 이용한 특징입력 추출을 위해 기준일 n 의 $CPP_{n,5}$ 로부터 31일 전인 $CPP_{n-31,5}$ 까지의 32개 $CPP_{n,5}$ 에 대해서 그림 1에서와 같이 스케일 레벨 5인 이분 비연속 Haar 웨이블릿 변환을 수행하여 38개의 계수를 생성하였다. 이들은 d1에서 16개, d2에서 8개, a3와 d3에서 각각 4개, a4와 d4에서 각각 2개, a5와 d5에서 각각 1개의 계수들로 구성되어 있다. 이들 38개의 계수에 $CPP_{n,5}$ 값을 포함한 총 39개가 초기 특징입력이 된다. 최소의 특징입력 추출을 위해 NEWFM기반의 비중복면적 분산추정법[14]을 사용하여 39개의 초기 특징입력 중 중요도가 가장 높은 다음과 같은 5개의 특징입력을 추출하였다.

- d1의 16개 계수들 중에서 첫 번째 계수
- d2의 8개 계수들 중에서 첫 번째 계수
- a3의 4개 계수들 중에서 첫 번째 계수
- a5 계수
- $CPP_{n,5}$

3. 가중 퍼지소속함수 기반 신경망 (Neural Network with Weighted Fuzzy Membership Function, NEWFM)

NEWFM은 입력으로부터 학습된 가중 퍼지소속함수의 경계합을 이용하여 클래스 분류를 하는 지도학습(supervised) 퍼지 신경망이다. 가중 퍼지소속함수의 경계합은 가중치를 갖는 대, 중, 소의 3개 퍼지소속함수의 경계합을 하나의 퍼지소속함수로 합성한 것이다. NEWFM의 구조는 그림 2에

서 보는 바와 같이 입력, 하이퍼박스, 클래스의 3개 계층으로 구성되어 있다. 입력 계층에는 n 개의 입력노드들로 구성되며 각 입력노드는 하나의 특징입력을 입력받게 된다. 하이퍼박스 계층은 m 개의 하이퍼박스 노드로 구성되고, l 번째 하이퍼박스 노드 B_l 은 단 하나의 클래스 노드에 연결되며 n 개의 퍼지집합을 갖는다.



(그림 2) 가중 퍼지소속함수 신경망(NEWFM)의 구조

B_l 의 i 번째 퍼지집합은 B_l^i 로 표시되는데 대, 중, 소로 구성된 3개의 가중 퍼지소속함수를 갖는다. 출력 계층은 p 개의 클래스 노드들로 구성되며 각 클래스 노드는 하나 혹은 하나 이상의 하이퍼박스 노드에 연결되어 있다. 입력노드로 입력되는 h 번째 입력 패턴은 $I_h = \{A_h = (a_1, a_2, \dots, a_n), class\}$ 로 기록되며 여기서 $class$ 는 분류값이고 A_h 는 n 개 특징입력으로 구성된다.

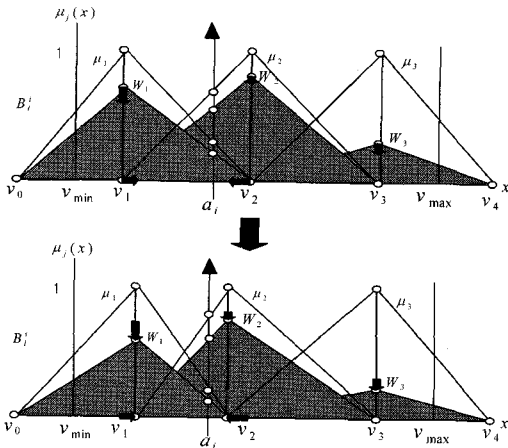
하이퍼박스노드 B_l 과 클래스노드 C_i 사이의 연결 가중치는 초기값으로 연결이 안 된 상태에서는 $w_{hi}=0$ 이 주어지나 연결 시 $w_{hi}=1$ 로 설정된다. B_l 과 클래스노드 C_i 를 연결하기 위해서 입력 $I_h = \{A_h = (a_1, a_2, \dots, a_n), class\}$ 에 따른 각 하이퍼박스의 $Output(B_l)$ 값 중 최대값을 갖는 B_l 과 $i=class$ 인 C_i 를 선택한다. 함수 $Output(B_l)$ 은 다음과 같다.

$$Output(B_i) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^3 B_i^j(\mu_j(a_i)) W_j$$

그림 3의 v_1, v_2, v_3 는 대, 중, 소 퍼지소속함수의 중심위치이고 학습 수행 중에 좌우로 조정된다. v_0 와 v_4 는 일정한 값으로 고정되어 있다. 입력 a_i 는 그림 2의 $v_{min} \sim v_{max}$ 범위 안에 있다.

μ_j 는 B_i^j 의 가중 퍼지소속함수를 나타내며 $j(j=1,2,3)$ 는 대, 중, 소 가중 퍼지소속함수의 인덱스를 각각 나타낸다. 각 가중 퍼지소속함수 μ_j 의 모양은 $(v_{j-1}, 0), (v_j, 0), (v_j, W_j)$ 의 3 점으로 표현되는 삼각형으로 되어있다. 여기서 W_j 는 소속함수의 강도(strength)를 나타내는 소속함수 가중치 ($0 \leq W_j \leq 1$, 초기값: $0.45 \leq W_j \leq 0.55$)를 의미한다.

$Adjust(B_i)$ 은 h 번째 입력패턴인 $A_h = (a_1, a_2, \dots, a_n)$ 에 의해 B_i 내의 대, 중, 소 퍼지소속함수와 그 가중치를 조정하는 함수이다. 입력 a_i 를 받는 B_i^j 의 v_j 와 W_j 는 a_i 값에 의하여 그림 3과 같이 입력값에 수렴하는 방향으로 조정 된다($j=1,2,3$).

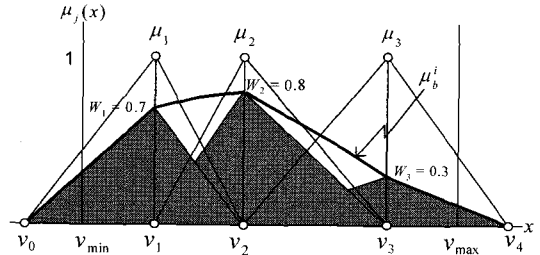


(그림 3) B_i^j 의 $Adjust(B_i)$ 연산 전후의 예

학습을 거친 NEWFM은 입력패턴을 분류하기 위한 퍼지규칙으로 사용될 수 있다. 하이퍼박스노드

B_i 의 각 퍼지집합에는 3개의 가중 퍼지소속함수 (weighted fuzzy memberships, WFM, 그림 3의 회색 부분)가 만들어진다. $\mu_b^i(x)$ (그림 4의 굵은 선) 로 표시된 WFM의 경계합(bounded sum of WFM, BSWFM)은 다음 식과같이 정의되며 그림 4에서와 같이 3개의 WFM의 퍼지 특성을 통합하게 된다. 학습된 BSWFM $\mu_b^i(x)$ 은 i 번째 입력에 대한 퍼지규칙이 된다.

$$\mu_b^i(x) = \sum_{j=1}^3 B_i^j(\mu_j(x))$$



(그림 4) 3개 가중 퍼지소속함수의 경계합(BSWFM) 예

이와 같이 n 개의 특징입력에 의해 생성된 n 개의 BSWFM들은 다시 비중복면적 분산 측정법[14]에 의하여 중요도가 낮은 특징입력을 제거함으로써 BSWFM의 수를 최소화할 수 있게 된다.

4. 실험 결과(Experimental Results)

본 논문은 1일 후의 KOSPI 상승과 하락의 분류성능 평가를 위해 표 1에서와 같이 [11]의 1991년도부터 1998년까지의 KOSPI 실험군을 사용하였다. 표 1의 GANN(the conventional Artificial Neural Network with the Genetic Algorithm)은 1991년부터 1998년까지의 년도 별로 전체 데이터의 약 80%를 사용하여 실험을 하였고, GAIS(the Genetic Algorithm approach to Instance Selection for Artificial Neural Network)는 유전자 알고리즘

기반 사례선택(instance selection)기법을 이용하여 전체 데이터 중에서 선택된 일부의 데이터로 실험을 하였다[11].

(표 1) KOSPI 실험군(11)

Set	Year								Total
	1991	1992	1993	1994	1995	1996	1997	1998	
Training instances for GANN	234	236	237	237	235	235	234	234	1882
Selected instances for GAIS	74	71	87	66	93	86	93	85	655
Holdout instances for GANN & GAIS	58	58	59	59	58	58	58	58	466

표 2는 [11]의 실험에서 사용한 12개의 특징 입력들과 NEWFM에서 사용한 특징 입력을 비교하고 있다. 본 논문에서는 NEWFM에 의하여 생성된 가중퍼지 소속함수의 경계합(BSWFM)을 이용하여 입력값에 대한 분류강도를 구하게 되고, 이러한 분류강도에 의해서 중요도가 낮은 특징입력을 제거하면서 5개 특징입력으로 최소화하였다. 이들 5개 특징입력으로 1일 후의 KOSPI의 상승과 하락을 예측하는데 사용하였다.

(표 2) [11]과 NEWFM에서 사용된 특징입력 비교

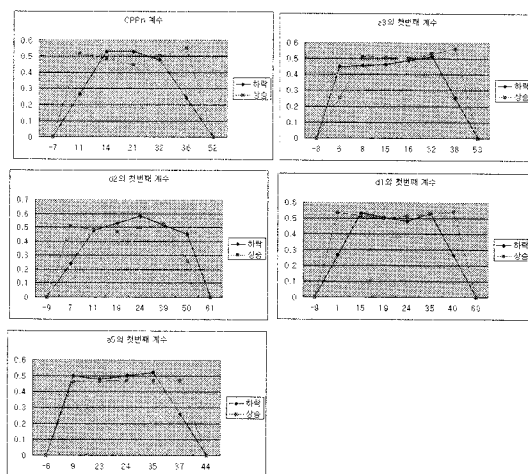
특징입력([11])	특징입력(NEWFM)
CCI, RSI, Stochastic 등 주가지표 12개	CPPn 및 추출된 최근 32일간의 CPPn 웨이블릿 계수 4개 등 5개

표 3에서는 GANN, GAIS에 대한 NEWFM 분류성능 비교를 보여준다. 표 1에서 Training instance for GANN에 사용된 각 년도 별 데이터로 실험한 경우 NEWFM이 모든 년도와 평균에서

GANN보다 향상된 성능을 보여주고 있다. GAIS는 전체 데이터 중 선택된 데이터만을 사용하여 평균 65.47%를 나타냈으며 NEWFM은 이 보다 향상된 67.62%를 보여주고 있다.

(표 3) [11]에 대한 NEWFM 분류성능 비교

Year	GANN(%) [11]	GAIS(%) [11]	NEWFM
	Accuracy (%) (전체분류율)	Accuracy (%) (전체분류율)	Accuracy (%) (전체분류율)
1991	53.45	72.41	72.41
1992	56.90	58.62	62.07
1993	59.32	59.32	62.72
1994	57.63	61.02	61.02
1995	65.52	67.24	65.52
1996	65.52	77.59	79.31
1997	58.62	58.62	67.24
1998	56.90	68.97	70.69
평균	59.23	65.47	67.62



(그림 5) 1998년도 실험군에서 생성된 5개 계수에 대한 가중 퍼지소속함수의 경계합(BSWFM)의 예

그림 5는 1998년의 KOSPI 실험군에서 사용된 5개의 특징입력에 대한 퍼지소속함수를 보여주고 있다. 이는 그림 4에서 보여준 가중 퍼지소속함수의 경계합(BSWFM)을 의미한다. 이를 통해 5개의 특징입력에 대한 1일 후의 KOSPI 상승과 하락 요인의 차이점을 가지적으로 볼 수 있으며 이에 따른 특성 분석을 가능하게 한다.

5. 결 론

본 논문은 NEWFM과 비중복면적 분산 측정법을 이용하여 1일 후의 KOSPI 상승과 하락에 대한 예측을 위한 가중 퍼지소속함수를 제안하고 있다. 또한 예측을 위한 주가지표로써 $CPP_{n,m}$ (Current Price Position of day n)을 제안하였다. 또한 비중복면적 분산 측정법을 이용하여, 최근 32일간의 $CPP_{n,m}$ 값을 웨이블릿 변환한 38개 계수와 $CPP_{n,m}$ 값 중 5개 특징입력을 추출하였다. 비중복면적 분산 측정법은 불필요하거나 분류 결과에 악영향을 주는 특징입력을 제거함으로써 특징입력의 수를 최소화 시켜준다.

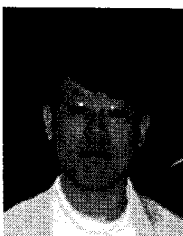
실험 결과 1일 후의 KOSPI 상승과 하락에 대한 67.62%의 예측율을 보여주었다. 향후 이번 결과를 활용하여 매도 매수 시점에 대한 예측 및 주식매매 자동화 시스템에 대한 연구가 진행될 것이다.

참 고 문 헌

- [1] K. K. Ang and C. Quek, "Stock Trading Using RSPOP: A Novel Rough Set-Based Neuro-Fuzzy Approach," *IEEE Trans. Neural Networks*, Vol. 17, No. 5, pp.790-802, Sept. 2006.
- [2] G. A. Carpenter, S. Grossberg and J. Reynolds, "ARTMAP : Supervised real-time learning and classification of nonstationary data by a self-organizing neural network," *Neural Networks*, Vol.4, pp.565-588, 1991.
- [3] S. H. Chai and J. S. Lim, "Economic Turning Point Forecasting Using Fuzzy Neural Network and Non-Overlap Area Distribution Measurement Method," *The Korean Economic Association*, Vol. 23, No. 1, pp. 111-130, Summer 2007.
- [4] T. V. Gestel et al., "Financial Time Series Prediction Using Least Squares Support Vector Machines Within the Evidence Framework," *IEEE Trans. Neural Networks*, Vol. 12, No. 4, pp.809-821, July 2001.
- [5] H. Ishibuchi and T. Nakashima, "Voting in Fuzzy Rule-Based Systems for Pattern Classification Problems," *Fuzzy Sets and Systems*, Vol.103, pp.223-238, 1999.
- [6] R. Jang, "ANFIS : Adaptive network-based fuzzy inference system," *IEEE Trans. Syst., Man, Cybern.*, Vol.23, pp.665- 685, May-June 1993.
- [7] C. F. Juang and C. T. Lin, "An On-Line Self-Constructing Neural Fuzzy Inference Network and Its Applications," *IEEE Trans. Fuzzy Systems*, Vol.6, No.1, pp.12-32, 1998.
- [8] N. Kasabov, *Foundation of Neural Networks, Fuzzy Systems and Knowledge Engineering*, The MIT Press, Cambridge, MA, 1996.
- [9] T. Kasuba, "Simplified Fuzzy ARTMAP," *IEEE AI Expert*, pp.19-25, Nov., 1993.
- [10] Ho J. Kim, T. W. Ryu, T. T. Nguyen, J. S. Lim, and Sudhir Gupta, "A Modified Fuzzy Min-Max Neural Network for Pattern Classification," *Computational Science and Its Applications ICCSA 2004 (LNCS 3046)*, pp.792-798, Springer-Verlag, 2004.
- [11] Kyoung-jae Kim, "Artificial neural networks with evolutionary instance selection for financial forecasting," *Expert System with Applications* 30, pp.519-526, 2006.
- [12] H.-M. Lee, K.-H. Chen and I.-F. Jiang, "A Neural Networks with Disjunctive Fuzzy Information," *Neural Networks*, Vol.11, pp. 1113-1125, 1998.
- [13] J. S. Lim, "Finding Fuzzy Rules by Neural Network with Weighted Fuzzy Membership Function," *International Journal of Fuzzy Logic and Intelligent Systems*, Vol. 4, No.2, pp.211-216, September, 2004.
- [14] J. S. Lim and S. Gupta, "Feature Selection

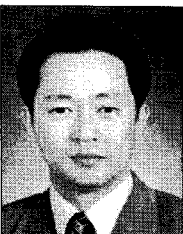
- Using Weighted Neuro-Fuzzy Membership Functions," The 2004 International Conference on Artificial Intelligence(IC-AI'04), June 21-24, 2004, VOL. 1, pp. 1301-1315, Las Vegas, Nevada, USA.
- [15] C. T. Lin and C. S. George Lee, "Neural-network-based fuzzy logic control and decision system," IEEE Trans. Computers, Vol. 40, No. 12, Dec. 1991.
- [16] D. Nauck and R. Kruse, "A Neuro-Fuzzy Method to Learn Fuzzy Classification Rules from Data," Fuzzy Sets and Systems, Vol.89, pp.277-288, 1997.
- [17] Min Qi, "Predicting US recession with leading indicators via neural network models," International Journal of Forecasting, Vol. 17, pp.383-401, 2001.
- [18] M. Setnes and H. Roubos, "GA-Fuzzy Modeling and Classification : Complexity and Performance," IEEE Trans., Fuzzy Systems, Vol.8, No.5, pp.509-522, 2000.
- [19] P. Simpson, "Fuzzy min-max neural networks-Part 1 : Classification," IEEE Trans., Neural Networks, Vol.3, pp. 776-786, 1992.
- [20] K. Tanaka, M. Sano and H. Watanabe, "Modeling and Control of Carbon Monoxide Concentration Using a Neuro-Fuzzy technique," IEEE Trans., Fuzzy Systems, Vol.3, pp.271-279, June, 1995.
- [21] J. S. Wang and C. S. G. Lee, "Self-Adaptive Neuro-Fuzzy Inference System for Classification Applications," IEEE Trans., Fuzzy Systems, Vol.10, No.6, pp.790-802, 2002.
- [22] C. Z. Ye, J. Yang, D. Y. Geng, Y. Zhou, N. Y. Chen, Fuzzy Rules to Predict Degree of Malignancy in Brain Glioma, Medical and Biological Engineering and Computing, Vol.40, 2002.

● 저 자 소 개 ●



이 상 홍 (Sang-Hong Lee)

1999년 경원대학교 공과대학 전자계산학과 졸업(공학사)
 2001년: 2001년 경원대학교 일반대학원 전자계산학과 졸업(공학석사)
 2006년~현재 경원대학교 일반대학원 전자계산학과 박사과정
 관심분야 : Neuro-Fuzzy System을 이용한 주가예측, 전문가 시스템
 E-mail : shleedosa@kyungwon.ac.kr



임 준 식 (Joon S. Lim)

1986년: 인하대학교 전자계산학과(학사)
 1989년: University of Alabama at Birmingham, Dept. of Computer Science (M.S.)
 1994년: Louisiana State University, Dept. of Computer Science (Ph.D.)
 1995년~현재 : 경원대학교 소프트웨어학부 교수(현)
 관심분야: Neuro-Fuzzy Systems, Biomedical Systems
 E-mail : jslim@kyungwon.ac.kr