

# 시계열 자료 코스피200의 패턴분류를 위한 퍼지 서포트 벡터 기계

## Fuzzy Support Vector Machine for Pattern Classification of Time Series Data of KOSPI200 Index

이수용<sup>\*</sup> · 손소영<sup>\*\*</sup> · 김철음<sup>\*\*\*</sup> · 이일병<sup>\*</sup>  
S.Y. Lee, S.Y. Sohn, C.E. Kim, and Y.B. Lee

\* 연세대학교 컴퓨터정보공학부 및 BERC

\*\* 연세대학교 컴퓨터산업공학부

\*\*\* 연세대학교 응용통계학과

### 요 약

주식시장에서 KOSPI200지수의 상승 또는 하락으로 분류 및 예측하는 정보는 선물 및 옵션시장에서 포트폴리오를 설계할 때 의사결정을 위해 중요한 기준이 된다. 경제지표인 시계열 패턴들의 향후 추세는 가장 최근의 경제패턴에 매우 종속적이기 때문에 최근의 패턴들을 가장 우선적으로 학습해야 할 필요가 있다. 본 논문에서는 시계열분석, 신경회로망, 그리고 다양한 분야에서 각광을 받고 있는 SVM(Support Vector Machine)과 Fuzzy SVM 모형의 분류 및 예측성능을 비교하였다. 특히 학습 DB에 따라 시계열성 속성을 갖는 퍼지소속함수에 가장 적합한 차원을 제시함으로써 Fuzzy SVM이 우수함을 입증하였다.

### Abstract

The Information of classification and estimate about KOSPI200 index's up and down in the stock market becomes an important standard of decision-making in designing portfolio in futures and option market. Because the coming trend of time series patterns, an economic indicator, is very subordinate to the most recent economic pattern, it is necessary to study the recent patterns most preferentially. This paper compares classification and estimated performance of SVM(Support Vector Machine) and Fuzzy SVM model that are getting into the spotlight in time series analyses, neural net models and various fields. Specially, it proves that Fuzzy SVM is superior by presenting the most suitable dimension to fuzzy membership function that has time series attribute in accordance with learning Data Base.

**Key Words** : SVM, Fuzzy SVM, Time Series Analysis, KOSPI200, Stock Prediction, Pattern Classification

### 1. 서 론

SVM (Support Vector Machine: SVM)은 V. N. Vapnik [12]에 의해 제안된 통계적 학습이론이며 학습과정에서 얻어진 확률분포를 이용하여 판별함수를 추정한다. SVM은 패턴 분류의 문제에서 일반성이 높기 때문에 최근들어 다양한 분야에서 좋은 평가를 받고 있다.

SVM의 기본 원리는 훈련 패턴들을 고차원의 특징공간으로 사상(mapping) 시킨 후 두 분류 사이의 여백(margin)을 최대화 시키는 결정함수(hyperplane)를 찾는 것이다. 하지만 SVM은 기존의 신경회로망처럼 패턴분류와 함수추정 등을 효과적으로 처리하지만 학습과정에서 모든 훈련패턴들을 동일적으로 동등하게 처리하는 문제가 있다. 즉 시계열 경제지

표를 훈련집합으로 선택하여 향후 경제의 추세를 예측 및 분류할 때 최근의 경제패턴들이 이전의 다른 패턴들보다 더 결정적인 영향력을 가질 수 있다. 따라서 최근 훈련패턴이 우선적으로 올바른 분류가 되도록 훈련되어야 할 필요가 있다. 과거의 불필요한 잡음(noise)이나 이상치(outlier)들을 학습시키기 위해 소모되는 불필요한 훈련시간을 단축하며 오분류율(misclassification rate)을 줄일 수 있는 의사결정함수를 찾는 것이 효과적이다. 이에 C. F. Lin[7]은 시계열 데이터에서 최근 추세가 반영되도록 조화함수를 응용한 퍼지소속함수를 오분류(misclassification)의 축소인 완화변수(slack variable)와 결합하여 결정곡면의 기울기를 조정할 때 퍼지소속함수의 속성이 반영되도록 제안했다.

본 논문에서는 금융시장의 시계열자료인 KOSPI200 지수의 상승 또는 하락으로의 패턴분류를 위해 시계열성이 반영되도록 조화함수의 차원을 다양하게 확장하여 적용하여 퍼지소속함수와 결합으로 확장된 FSVM이 학습 시간을 단축시켰으며 오분류률도 크게 낮추었다.

본 논문의 구성은 2장에서 SVM을 이용한 국내외의 기존 연구에 대해 정리하였다. 3장에서는 SVM과 FSVM(Fuzzy Support Vector Machine)을 설명하였으며, 4장에서 본 논문

접수일자 : 2003년 10월 15일

완료일자 : 2003년 11월 28일

감사의 글 : 본 연구는 과기부 뇌신경정보학사업과 KOSEF 생체인식연구센터(BERC)로부터 부분적인 지원을 받아 수행되었음."

에서 제안하는 확장된 퍼지소속함수를 이용한 FSVM의 검증 을 위해 KOSPI200 지수의 증가를 목표변수로 한 모의실험 한 결과들을 보였다. 이 때 학습을 위해 사용한 데이터베이스는 2000년 1월부터 2003년 8월까지 국내외 24개 변수들의 일별 데이터로 구성하였다. 그리고 5장에서 본 연구의 결과를 정리 하였다.

## 2. 기존 연구

국내에서 금융지수에 대한 기존 연구는 주로 기업도산, 주가, 이자율, 환율의 예측 및 분류 등이다. 특히 현물시장과 선물시장 그리고 옵션시장을 통합한 투자 포트폴리오를 구축 하려는 시도가 많았으며 이를 위해 현물시장 지수의 예측 및 분류에 대한 연구가 주된 내용이다. 기존 연구에서 주로 사용한 모델은 신경회로망, 유전자 알고리즘, 사례기반추론, 퍼지추론 시스템, 카오스, 자기회귀모형, 다변량 판별분석, 다중 회귀분석, 이동평균법 등이다. SVM이 국내 금융지표에 응용된 사례는 아직까지 없었다. 반면에 SVM을 응용한 연구가 문자인식, 영상인식, 홍채인식, 그리고 문서분류 등의 분야에서 국내 연구가 활발히 진행되고 있다.

외국의 경우 Alan[1] 은 SVM을 오스트리아 주식시장의 주식종목을 선택하는데 응용하여 포트폴리오 설계시 좋은 결과를 보여 주었으며, Chang[4]은 SVM을 이용한 시계열 분석을 하였고, Lin[7]은 순차 데이터에 Fuzzy 소속함수를 도입하여 학습의 유연성을 제안했으며, Tay[9]는 미국 금융지표의 예측을 위해 BP와 SVM을 적용한 결과 SVM이 우수함을 입증했다.

그 외 응용 분야에서는 Yang[13]이 SVM의 분류 성능을 이용하여 남녀 사진을 분류하였고, Hadzic[14]은 영상인식에 적용했다. Ana[14]는 SVM의 처리결과를 기하적으로 해석하는 방법을 제시했으며, Takuya[10]는 polyhedral pyramid를 퍼지소속 함수로 정의하여 FSVM 분류기의 성능을 benchmark 데이터들과 비교 분석하여 FSVM이 SVM보다 우수함을 입증하였으며, Jeng[6]은 퍼지신경회로망을 위한 퍼지추론시스템을 개선하는 SVM을 제안하여 소속함수가 Gaussian function일 때 SVM이 퍼지추론시스템의 규칙수를 결정할 수 있음을 보였다.

Suykens[14]는 이원분류문제의 해결을 위해 Least Square SVM을 제안했으며, Roobaert[8]는 간단한 학습 알고리즘으로 Direct SVM을 제안하였고, Mangasarian[14]은 데이터마 이닝에서 SVM의 기능에 대해 연구하였다. 또한 SVM의 빠른 학습시간을 위한 연구가 Anguita[2]와 Yang[13] 에 의해 이루어 졌다.

그 외에 SVM의 다양한 연구결과 및 연구과제들은 참고 문헌 [14]에서 찾을 수 있다.

## 3. Fuzzy Support Vector Machine

Vapnik[3,11]이 제안한 Support Vector Machine의 학습 알고리즘은 훈련 패턴들을 서로 다른 두개의 Class로 이원 분류할 때 기준이 되는 결정함수를 학습 알고리즘을 이용하여 판별함수를 찾는 것이다.

SVM의 기본 원리는 훈련 패턴들을 고차원의 특징공간으로 사상(mapping) 시킨 후 두 분류 사이의 여백(margin)을 최대화 시키는 결정함수(hyperplane)를 찾는 것이다. 하지만

사상에 대한 정보가 없어도 SVM은 특징공간에서 커널(kernel)이라는 내적함수를 활용하여 원하는 최적의 의사결정함수를 찾는다.

최적의 의사결정함수는 지지벡터(support vector)라는 몇 개의 입력 벡터들의 결합으로 나타낸다. 다음에 설명되는 부호화된 학습집합  $S = (y_i, x_i) \mid i = 1, \dots, l$ 가 주어졌을 때, 각 훈련데이터  $x_i \in R^N$ 는 두개로 부호화된 부분 중 반드시 한 곳에 속하게 되며, 이 때 부호는  $y_i \in -1, +1, i = 1, \dots, l$ 이다. 입력공간에서 적절한 결정함수의 탐색이 쉽지 않으므로 입력공간의 차원보다 더 높은 차원의 특징공간으로 입력공간을 사상(mapping)시키면 최적의 의사결정함수를 탐색할 수 있게 된다.

$z = \phi(X)$ 이  $R^N$ 에서 특징공간  $Z$ 로의 사상일 때,  $W \cdot Z + b = 0$ 를 만족하는  $(W, b)$ 를 결정함수라 한다. 이 때  $W \in Z, b \in R$  이고,  $X_i$ 는 다음 함수에 의해 분리 된다.

$$f(x_i) = \text{sign}(W \cdot Z_i + b) = \begin{cases} +1 & \text{if } y_i = 1 \\ -1 & \text{if } y_i = -1 \end{cases} \quad (1)$$

선형분리 되지 않는 데이터들을 처리하기 위하여, 완화변수(slack variable)  $\xi_i \geq 0$ 를 오분류 척도(measure)로 정의하면 결정함수는 다음같이 수정된다.

$$y_i(W \cdot Z_i + b) \geq 1 - \xi_i, i = 1, \dots, l \quad (2)$$

여기서  $\xi_i$ 는 의사결정함수를 만족하지 않는  $X_i$ 에 대한 오분류(misclassification) 척도이므로  $\sum_{i=1}^l \xi_i$ 는 훈련집합  $S$ 에 대한 오분류의 척도가 된다. 따라서 최적의 의사결정함수(hyperplane)는 다음과 같이 표현 된다.

$$\text{minimize } \frac{1}{2} W \cdot W + C \sum_{i=1}^l \xi_i \quad (3)$$

$$\text{subject to } y_i(W \cdot Z_i + b) \geq 1 - \xi_i, \xi_i \geq 0$$

라그랑지 승수  $a = (a_1, \dots, a_l)$ 를 도입하고,  $\phi: R^N \rightarrow Z, \phi(X) = Z$ 일 때, 커널다항식  $K(x_i, x_j) = (1 + x_i \cdot x_j)^d$ 가 를 만족하면 최적의 결정함수(hyperplane)은 Mercer의 정리에 의해 다음 두식을 만족하는

$$\text{max } W(a) = \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l a_i a_j y_i y_j K(x_i, x_j) \quad (4)$$

$$\text{subject to } \sum_{i=1}^l y_i a_i = 0, 0 \leq a_i \leq C, i = 1, \dots, l$$

$f(x) = \text{sign}(W \cdot Z + b) = \text{sign}(\sum_{i=0}^l a_i y_i K(x_i, x_j) + b)$ 이다.

시계열 데이터들이 추세를 형성할 때 가장 최근 데이터들의 패턴에 영향을 많이 받는다면 순차적 성질을 갖는 퍼지소속함수를 정의하여 각 훈련패턴에 적용하면 학습할 때 모든 훈련 데이터들이 확일적으로 취급되지 않고 순차 데이터들의 패턴에 영향을 받도록 학습시킬 수 있다. 따라서 SVM기법은 순차적 속성을 반영하는 퍼지소속함수와 결합하여 확장된 SVM 패턴분류기 모델의 설계가 필요하다. Lin[7]은 퍼지소속함수를 SVM의 완화변수(slack variable)에 적용하는 FSVM을 제안했다.

FSVM의 특징은 SVM이 비선형 분류문제를 해결할 때 퍼지소속함수와 결합한 훈련데이터를 사용함으로써 오분류

(misclassification)의 단위인 완화(slack) 변수들이 퍼지소속 함수의 영향을 받아 결정곡면의 기울기를 조정할 때 유연성을 지니도록 조정하는 것이다. 즉,  $S = (y_i, x_i, s_i) | i = 1, \dots, l, y = +1, -1$ 라 하자. 단  $\sigma$ 는 임의의 작은 수이며,  $X_i \in R^N$ 는 훈련 데이터이고,  $y_i \in -1, +1$ 는 부호화된 목표 데이터이며,  $S_i = s_i \in R | 0 \leq s_i \leq 1$ 는 퍼지소속 함수값이다.

퍼지소속 함수값  $s_i$ 는 벡터  $X_i$ 가 한 Class에 속하는 정도를 표시한 속성이고,  $\xi_i$ 는 SVM에서 오분류에 대한 오차(error)의 척도(measure)이므로  $s_i \xi_i$ 는 서로 다른 가중치를 갖는 새로운 오차의 척도로 변환된다. 따라서 FSVM에서 최적의 결정함수(hyperplane)는 다음식을 만족하는 해이다.

$$\begin{aligned} & \text{minimize } \frac{1}{2} WW + C \sum_{i=1}^l s_i \xi_i \\ & \text{subject to } y_i (W \cdot Z_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \end{aligned} \quad (5)$$

Lin[7]은 시계열 데이터에 대한 퍼지소속함수로 함수  $s_i = f(t_i)$ 의 차원을  $n = 1, 2$ 인 경우로 제안했지만 본 논문의 실험에서는  $n \geq 3$ 을 적용했다. 여기서 퍼지소속정도의 하한(lower bound)으로  $\sigma > 0$ 을 선택했으며, 시간  $t_1 \leq t_i \leq t_l$ 에 관한 퍼지소속정도의 함수를 정의한다. 시계열 패턴의 마지막  $X_l$ 이 가장 최근의 추세를 반영함으로 소속정도를  $s_l = f(t_l) = 1$ 이라 하고, 시계열 패턴의 처음  $X_1$ 의 소속정도는  $s_1 = f(t_1) = \alpha$ 라면

$$s_i = f(t_i) = (1 - \alpha) \left[ \frac{t_i - t_1}{t_l - t_1} \right]^n + \alpha \quad (6)$$

를 시계열 패턴의 학습을 위한 퍼지 소속함수로 정의한다. 본 논문에서는 소속함수 차원을  $n \geq 3$ 차원으로 다양하게 확장하여 KOSPI200지수의 분류 예측문제에 적용하였다. 시뮬레이션 결과에 의하면  $n = 1, 2$ 보다  $n \geq 3$ 일 때 오분류율이 줄일 수 있었고, 학습 시간도 크게 단축시킬 수 있는 유용한 결과를 얻었다.

## 4. 실험

### 4.1. 실험 환경

데이터베이스는 2000년 1월부터 2002년 8월까지 24개 변수의 일별 데이터(국내금융지수:14, 해외금융지수:10)들로 구성하였고, 일별 데이터를 단순이동평균법에 의해 주별 데이터 137개로 변환하여 표준 정규분포를 따르도록 전처리했다. 24개 변수의 종류는 KOSPI(시가, 종가, 고가, 저가, 거래량), KOSPI200(시가, 종가, 고가, 저가, 거래량, 거래대금), CBY(3년만기), 원/달러환율 등의 국내 지표와 Dow-Jones(시가, 종가, 고가, 저가, 거래량, 거래대금), Nasdaq(시가, 종가, 고가, 저가, 거래량, 거래대금) 등의 미국 지표들로 구성되었다.

실험에서는 로짓회귀 분석을 이용하여 KOSPI200 지수의 등락에 영향을 있는 변수들을 선택했는데 KOSPI200(고가, 저가), KOSPI(시가, 고가), 원달러환율, DOW(시가, 저가), NASDAQ(고가, 저가), 회사채(3년 만기) 등 10개 변수이다.

시스템의 운영체제는 Windows 2000/sever이고, CPU는 Pentium IV (Dual CPU : 1+1 GHz)이며, RAM은 1,024

Mb 이다. 이런 사양에서 C++로 구현된 SVM 및 FSVM을 실험에 사용했다.

### 4.2 실험 내용

#### 4.2.1 시계열 분석(ARIMA)

KOSPI200의 시가, 저가, 고가, 종가에 대한 시계열 데이터의 전체적 변동폭에 대한 그래프는 아래와 같다.

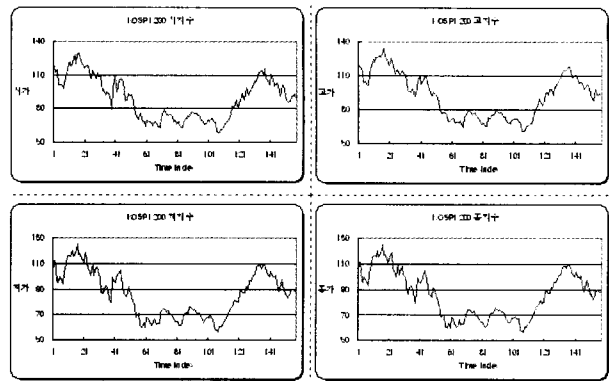


그림 1. KOSPI200 시지수, 고지수, 저지수, 종지수

위의 그림으로부터 KOSPI200 지수는 시간의 흐름에 따라 평균이 변화하는 비정상성 시계열 데이터임을 알 수 있다. 정상성을 만족시키기 위해 비정상 시계열데이터를 1차 차분하여 ARMA모형으로 적용시킬 수 있는 정상시계열데이터로 변환시켜 사용하였다. 정상시계열 모형의 식별을 하기 위해 Box-Jenkins 방법과 AIC(Akaike Information Criterion)과 BIC(Bayesian Information Criterion)를 이용하였다.

시계열 데이터들은 ARIAMA(0,1,0)인 확률보행모형(Random Walk Model)로 식별되었다. 이 모형은 현 시점의 값은 바로 전시점 자료에 오차가 더해지는 모형이므로 up/down 분류를 random하게 예측하므로 좋은 예측률을 기대할 수 없다.

#### 4.2.2 시계열 데이터에 대한 FSVM의 벤치마킹

FSVM에서 정의한  $n$  차원 퍼지소속함수의 유용성을 검증하기 위해 병원의 당뇨병 환자에 대한 시계열 데이터로 benchmark를 위해 제시되어있는 실험 결과(표 1)에 의하면 SVM이 83.33% 이었다. 반면에 Lin이 제안한  $n = 2$ 차원 퍼지소속함수를 이용한 FSVM에서는 70.00%이었고, 본 연구에서 제안하는 퍼지소속함수의 다양한 차원을 사용했을 때  $n \geq 3$ 차원 이상의 소속함수를 선택하면 SVM과 동일한 결과를 보였다.

당뇨병 환자들의 시계열 자료의 벤치마킹을 위한 SVM의 결과와 본 실험에서 제시한 FSVM의 결과가 동일했지만 다양한 파라미터들의 변화에 대해 FSVM이 SVM보다 일관성 있는 결과를 제시하여 안정된 분류를 하였다.

표 1. 인디언 당뇨병환자의 시계열자료

classifier (RBF kernel)		hit rate (%)
SVM		83.33
FSVM	$n = 1$	70.00
	$n = 2$	70.00
	$n \geq 3 (n = 189)$	83.33

경제지표인 시계열 데이터 KOSPI200 지수에 대한 실험(표 2)에서는 FSVM의 소속함수가  $n \geq 3$  일 때 80.00%로 기존의 SVM의 70.00%보다 우수함을 보였다.

표 2. KOSPI200 증가 상승/하락 이원분류

classifier (RBF kernel)		hit rate (%)
SVM		70.00
FSVM	$n = 1$	70.00
	$n = 2$	60.00
	$n \geq 3 (n = 725)$	80.00

대학병원의 심장병에 대한 시계열 데이터의 분류실험(표 -3)에서도 SVM이 60.00% 이고 제안하는 FSVM은  $n \geq 3$  일 때 63.00%를 나타냈다. 또한 FSVM의 경우 커널함수는 RBF커널 함수가 Linear 커널 함수와 Polynomial 커널 함수보다 약간 우수하였다.

시계열 데이터들에 대한 실험결과(표 1, 표 2, 표 3)에 의해 SVM과 퍼지소속함수를 결합된 FSVM ( $n \geq 3$ )이

표 3. 심장병 환자에 대한 시계열자료

classifier	kernel	hit rate (%)
SVM	RBF	60.00
FSVM ( $n \geq 3$ )	Polynomial	61.00
	RBF	63.00

기존의 SVM과 신경회로망의 BP알고리즘 보다 오분류율을 분류적중률(hit rate)이 높음을 알 수 있었다. 실험에서 사용한 FSVM의 Kernel 함수는 RBF를 이용 하였다.

4.2.3 FSVM의 학습시간 단축

표 4. Polynomial 커널의 학습시간 (sec)

d	c	hit rate (%)		time (sec)		sv	
		SVM	FSVM	SVM	FSVM	SVM	FSVM
1	100	73.33	73.33	0.4	0.2	111	111
	500	73.33	73.33	2.5	0.6	113	111
	1000	73.33	73.33	4.9	1.2	113	111
	1500	73.33	73.33	7.3	1.8	113	111
	2000	73.33	73.33	6.6	2.3	113	111
2	100	66.67	80.00	3.5	2.6	94	82
	500	66.67	70.00	23.9	4.5	98	66
	1000	66.67	56.70	43.4	5.1	98	59
	1500	66.67	66.67	70.3	10.3	98	59
	2000	66.67	60.00	89.7	10.9	98	58
average		70.00	70.00	25.2	4.0	104.90	88.00

표 4에서 d는 차수(dgree), c는 타협점(trade-off), 그리고 sv는 support vector 수를 의미한다. SVM의 학습 시간이 가장 많이 소요됨을 보였다. 파라미터들에 관한 실험 결과에 의하면 SVM과 FSVM이 동일한 적중률을 나타내는 ( $d=2, c=1,500 ; sv=98:59$ )의 경우에 6.8배, SVM이 좋은 적중률을 나타내는 ( $d=2, c=1,000 ; sv=98:60$ )의 경우에 8.5배, 그리고

FSVM이 좋은 적중률을 나타내는 ( $d=2, c=500 ; sv=98:66$ )의 경우 5.3배의 훈련 시간 차이가 있음을 알 수 있다. 결과를 통해 퍼지소속함수가 완화변수(slack variable)와 결합된 FSVM에서는 support vector 수가 적게 됨으로 학습시간이 SVM보다 상대적으로 단축된다.

데이터 크기별로 학습시간을 조사한 실험(표 5)에서는 데이터들은 1997년 1월 2일부터 2002년 8월 31일까지 일별 데이터수 중에서 임의로 92, 127, 158, 200, 249, 293, 629, 1,810개로 구분하여 학습시간에 대한 실험을 하였다.

표 5. 훈련집합 데이터 수에 따른 학습시간 (sec)

#(data)	Linear		RBF		Polynomial	
	SVM	FSVM	SVM	FSVM	SVM	FSVM
92	5	5	1	1	6	6
127	20	15	1	1	20	15
158	17	11	1	1	13	11
200	23	21	1	1	22	20
249	25	20	1	1	30	20
293	32	7	1	1	31	6
629	145	3	2	1	117	2
1,810	6,853	31	2,632	2,118	179,957	14,949

\* BPN : 2.973초 (데이터수 1,810개의 경우)

가장 학습시간이 빠른 Kernel 함수는 RBF이며 신경망의 BP알고리즘보다 SVM은 12% 를, FSVM은 40% 를 단축하였다. 특히 Polynomial 함수의 경우 SVM(49.98시간=179,957초)이고 FSVM(4.15시간=14,94초)이 소요됨으로 FSVM을 이용할 경우 45.84시간을 단축하여 SVM의 학습을 위한 시간보다 약 12배의 단축효과를 얻었다. SVM의 빠른 학습시간을 위한 연구에서 Anguita[2]은 gradient를 빠르게 계산하기 위해 block-Toeplitz 행렬을 이용했으며, Yang[13]은 학습 수에 따른 학습시간의 상계(upper bound)를 제시하는 알고리즘을 발표 하였다.

4.2.4 유용성을 검증을 위한 실험

유용성 검증을 위해 학습시간 및 테스트 데이터의 기간 이동방법을 적용하였다. 즉 2000년 1월 첫주부터 2002년 6월 둘째주까지 학습한 후 6월 셋째주를 테스트하고, 2000년 1월 둘째주 부터 2002년 6월 셋째주까지 학습한 후 2002년 넷째주를 테스트하는 방법으로, 2002년 9월 둘째주까지 시물레이션한 결과(표 6)에 의하면 RBF 커널 함수를 사용했을 때 신경회로망의 BP알고리즘은 분류적중률이 70%, SVM은 77.78%, 그리고 FSVM은 88.89% 를 보였다.

표 6. KOSPI200 기간이동 블레이션

classifier	NN	SVM	FSVM
algorithm / kernel	BP	RBF	RBF
hit rate (%)	70.00	77.78	88.89

표 7의 큰 폭의 추세 전환점에 대한 정분류 모의실험에 의하면 신경회로망의 BP알고리즘은 64%, SVM은 66.67%, 그리고 FSVM은 73.33%의 적중률을 보였다.

표 7. 추세전환점 직전에서 분류적중률

classifier	NN	SVM	FSVM
algorithm / kernel	BP	RBF	RBF
hit rate (%)	64.00	66.67	73.33

### 5. 결 론

시계열 자료인 주식시장의 KOSPI200 증가 지수의 시계열 분석 결과 ARIMA(0,10)인 확률보행모형으로 식별되었다. 신경회로망과 SVM에 비해 퍼지소속함수를 결합한 FSVM은 퍼지소속함수의 차원을 Lin 제안한 2차원 이상의 차원으로 적용했을 때 훈련시간과 오분류률(misclassification rate)을 줄일 수 있음을 보였다. 퍼지소속함수의 차원은 주어진 훈련패턴들에 적합한 최적의 차원을 실험적 방법에 의해 찾을 수 있다.

본 연구의 실험에서 사용한 퍼지소속함수의 차원은  $n = 187$  이었다. 퍼지소속함수의 차원을 최적화한 FSVM은 SVM을 비롯한 기존의 패턴분류기에 비해 분류성능이 우수했으며 특히 Polynomial 커널을 사용한 FSVM의 경우에는 학습시간이 크게 단축되었다.

### 6. 향후 연구 계획

학습할 때 패턴의 잡음 및 이상치(outlier)들에 대한 영향력을 줄이기 위해 적합하며 주어진 데이터베이스를 따르는 다양한 퍼지소속함수의 연구가 필요하다.

### References

[1] F. Alan, Stock Selection using Support Vector Machines, www.kernel-machines.org  
 [2] D. Anguita, Fast Training of Support Vector Machines for Regression, http://www.kernel-machines.org  
 [3] N. Cristianini, An Introduction to Support Vector Machines, Cambridge University Press, 2000.  
 [4] M. W. Chang, Analysis of non stationary time series using support vector machines, www.kernel-machines.org  
 [5] C. W. Hsu, "A simple Decomposition Method for Support Vector Machines", Machine Learning, Vol. 46, pp. 291-314, 2002.  
 [6] J. T. Jeng, Support Vector Machines for the Fuzzy Neural Networks, http://kernel-machines.org  
 [7] C. F. Lin, "Fuzzy Support Vector Machines", IEEE Transactions on Neural Networks, Vol.13, No.2, March 2002.  
 [8] D. Roobaert, Direct SVM : "A Simple Vector Machine Perceptron", J. of VLSI Signal Processing 32, pp. 147-156, 2002.  
 [9] F. E. H. Tay, "Application of support vector machines in financial time series forecasting",

Omega 29, pp. 309-317, 2002.

[10] I. Takuya, "Fuzzy Support Vector Machines for Pattern Classification", www.kernel-machine.org  
 [11] Z. Weida, "Linear programming support vector machines", J. Pattern Recognition Society, pp.1-10, 2001.  
 [12] V. N. Vapnik, Statistical Learning Theory, Wiley-Interscience Pub., 1998.  
 [13] D. Yang, Provably Fast Training Algorithms for Support Vector Machines, http://www.kernel-machines.org  
 [14] http://www.kernel-machines.org  
 [15] 김대수, 신경망 이론과 응용(I), 하이테크정보, 1989.  
 [16] 김대수, 신경망 이론과 응용(II), 하이테크 정보, 1993.  
 [17] 이광형, 오길록, Fuzzy 이론 및 응용 1권(이론), 홍릉과학출판사, 1991.  
 [18] 이광형, 오길록, Fuzzy 이론 및 응용 2권(응용), 홍릉과학출판사, 1997.

### 저 자 소 개

#### Sooyong Lee

hosu@hosu.yonsei.ac.kr  
 Assistant Professor of Information and Technology, Yonsei University. Ph.D.(candidate, 2004) in Department of Computer Science, Yonsei University. Ph.D. in Department of Mathematics, Kyunghee university (1992.2).

#### Soyoung Sohn

sohns@yonsei.ac.kr  
 Professor of Industrial Systems Engineering, Yonsei University. Ph.D in Department of Industrial Engineering, University of Pittsburgh (1989.8).

#### Chuleung Kim

cekim@yonsei.ac.kr  
 Associate Professor of Applied Statistics, Yonsei University. Ph.D in Carnegie-Mellon University (1989.8).

#### Yillbyung Lee

yblee@csai.yonsei.ac.kr  
 Professor of Information & Industrial Engineering, Yonsei University. Ph.D. in Department of Computer and Information Science, University of Massachusetts (1985.2).