

시계열 네트워크에 기반한 주가예측*

박강희** · †신현정***

Stock Price Prediction Based on Time Series Network

Kanghee Park** · †Hyunjung Shin***

■ Abstract ■

Time series analysis methods have been traditionally used in stock price prediction. However, most of the existing methods represent some methodological limitations in reflecting influence from external factors that affect the fluctuation of stock prices, such as oil prices, exchange rates, money interest rates, and the stock price indexes of other countries. To overcome the limitations, we propose a network based method incorporating the relations between the individual company stock prices and the external factors by using a graph-based semi-supervised learning algorithm. For verifying the significance of the proposed method, it was applied to the prediction problems of company stock prices listed in the KOSPI from January 2007 to August 2008.

Keywords : Semi-Supervised Learning(SSL), Stock Price Prediction, Data Mining, Technical Indicators(TIs)

논문접수일 : 2010년 11월 24일 논문게재확정일 : 2011년 02월 23일

* 본 논문은 2010년도 한국경영과학회 추계학술대회 경쟁부문(응용)에 제출하여 우수논문상을 수상한 논문임. 논문의 내용을 일부 보완, 확장한 논문이며, 소정의 심사과정을 거쳐 게재 추천되었음.

본 연구는 아주대학교 Post BK 21 연구비지원 및 한국연구재단(2010-0007804)의 연구비 지원으로 수행되었음을 사사하며, 이에 감사의 뜻을 표한다.

** 아주대학교 산업공학과

*** 아주대학교 산업정보시스템학부

† 교신저자

1. 서론

주가지수는 기업내부의 영향뿐만 아니라 외부영향 즉, 타기업들의 영향, 국제유가, 환율, 금리, 타국가들의 주가지수 및 경제상황에 상호연관성을 가지므로 이들의 등락에 민감하게 영향을 받는다. 주가지수의 등락은 정부 및 기업 정책 등 경제 전반에 중대한 영향을 미치므로, 다양한 분석 모델들을 이용하여 보다 정확하게 주가지수 예측을 하고자 하는 노력들이 지속적으로 이루어져 왔다.

주가 데이터는 변동성을 가진 시계열 데이터 특성을 가지므로 전통적 시계열 분석기법들이 주된 도구로 사용되어 왔다. 예를 들어 Jeantheau[5]는 ARCH model을 사용하여 주가를 예측하였고 Amilon[1]와 Liu et al.[9]은 중국 주식시장에 Skewed-GED Distribution를 이용한 GARCH 모델을 사용하여 주가 예측을 제시하였다[1, 5, 9]. 이 방법들은 과거의 주가 변동성을 기초로 한 시계열 분석방법으로 예측을 한 것이기 때문에 미래의 주가가 과거와 유사하게 변동할 것이라는 가정을 토대로 한다. 자연현상으로부터 얻는 태양의 흑점 수, 강우량, 기온 등과 같은 시계열 데이터는 이러한 가정을 잘 따르므로 전통 시계열 분석방법으로 좋은 결과를 얻을 수 있다. 그러나 주가예측은 이러한 가정만을 토대로 모델을 만들기에는 다소 무리가 있다. 즉, 주가 자체의 움직임뿐만 아니라 주가 변동에 영향을 미치는 기업내부적 요인, 외부요인(타기업주가 및 상황, 국제유가, 환율, 금리, 타국가들의 주가지수 및 경제상황 등)의 상호 영향 관계가 주가를 형성하는데 더 많은 영향을 준다고 인지되고 있다[4]. 그러나 앞선 연구사례에서 사용되어왔던 시계열 분석방법만으로는 이러한 상호 관계 및 인과관계를 명시적으로 모델에 포함하고 정형화하는 것에 방법론적으로 한계가 있다[8, 12]. 기계학습분야에서도 주가예측에 대한 많은 연구가 이루어져 왔다. 대표적인 모델로는 artificial neural network(ANN)와 support vector machine(SVM)이 자주 이용되고 있다. Tay and Cao[13]는 금융시계열 데이터를 SVM에 도입

하는 방법을 제안하였고 Kanas[6]은 ANN 모델로 S&P500 지수 예측을 시도하였다[6, 13, 15]. 또한 Yang et al.[15]은 ANN을 이용해 commercial bank loan risk의 조기경보시스템을 제안하였고 Bekiros and Georgoutsos[2]는 호재와 악재를 구분하기에 불명확한 뉴스가 NASDAQ 지수에 어떤 영향을 미치는지 ANN 모델을 이용하여 분석하였다[2, 15, 17]. ANN과 SVM을 이용한 방법들은 주가 및 외부요인과의 상호연관성 및 인과관계가 비교적 명시적으로 모델링에 포함되기는 하나, 다소 부족한 면이 있다. 즉 주가(종속변수)의 변동이 외부요인(독립변수)들과 어떻게 작용하는지에 대한 상호연관성은 모델에 표현이 되나 외부요인들간의 상호연관성을 정형화하지는 못한다. 예를 들면 ANN과 SVM은 환율, 금리, 유가 등의 등락이 주가의 등락에 대한 일차 상호 연관성은 표현이 가능하다. 그러나 금리의 하락이 환율에 어떻게 영향을 미쳤고 이어 환율이 어떻게 변했으며 이러한 연차적인 인과관계가 주가에 어떻게 영향을 미치는지에 대한 2차, 3차, ..., 고차적인 상호 연관성에 대한 표현이 어렵다. 또한 이러한 일련의 과정 끝에 변동된 주가가 다시 이들 외부요인에 어떻게 다시 영향을 주는지에 대한 관계를 조망하기도 어렵다. 이러한 한계점을 극복하기 위해 본 논문에서는 최근 기계학습분야에서 주목 받고 있는 semi-supervised learning (SSL)을 주가 예측에 적용하는 방법론을 제안하고자 한다[11, 17].

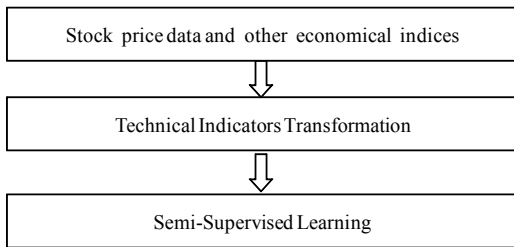
본 논문에서는 시계열 데이터를 SSL에 접목시켜 여러 경제지표들의 상호 연관성 및 다차원적인 인과관계를 고려하여 주가를 예측하는 모델을 제안한다. 제안한 모델은 2007년 1월부터 2008년 8월의 KOSPI에 상장되어 있는 개별회사 주가예측에 적용, 검증되었다.

2. 방법론

주식가격 예측 모델은 국내 주식시장, 타국가의 주식시장 및 경제지표들을 고려하고 그에 따라 주

가의 개별등락 및 주식시장의 종합적인 현황을 파악한 정보가 포함되어야 한다. 이에 본 논문에서는 주가예측에 주로 이용하는 KOSPI에 상장되어 있는 회사들의 주가 및 주가관련정보(자본금, PER 등)와 환율(원/달러), 금리, 유가(West Texas Intermediate 원유가격, WTI), 타국가들의 주가종합지수 등을 이용하여 모델을 설계하였다. 제안하는 모델은 다음과 같은 절차를 따른다.

첫째, 시계열 데이터는 잡음(oscillatory noise)을 포함하므로 이를 줄이고 추세 및 구조적 변동요인을 얻기 위하여 기술적 지표(technical indicators, TI)를 이용한다. 둘째, 기술적 지표로 변환된 변수들의 상호연관성을 그래프로 표현하고, semi-supervised learning(SSL)을 이용하여 주가 예측모델을 만든다. <그림 1>은 위 절차를 간략하게 나타낸다.



<그림 1> 주가예측모델 절차

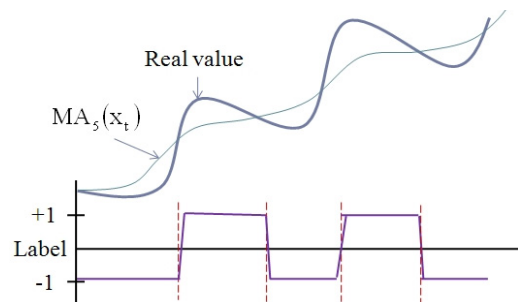
다음 각 절에서는 본 연구에서 제안하는, 시계열 예측에 적합하게 재설계된 그래프 기반 semi-supervised learning에 대한 설명에 이어, 기술적 지표(technical indicators)에 대하여 소개한다.

2.1 Graph-based Semi-Supervised Learning (SSL) for Time Series Network

그래프 기반 semi-supervised learning에서는 그래프를 이용하여 유사도가 높은 데이터들을 클러스터링 하고 이를 토대로 레이블이 없는 데이터의 레이블을 예측한다[16]. 즉, 전체 주가 및 경제지표 $n(n = 1+u)$ 개 중에서 l 개의 레이블이 있는 데이터

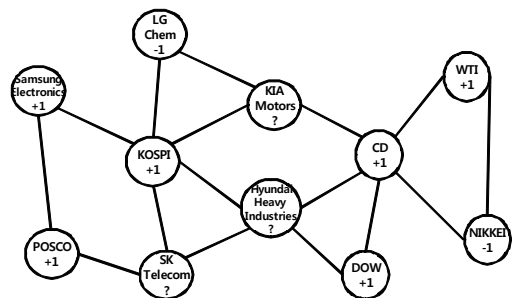
$\{(X_1, Y_1), \dots, (X_l, Y_l)\}$ 와 u 개의 레이블이 정해지지 않은 데이터를 사용한다. 데이터는 그래프의 노드로 표현되고, 각 노드들은 주가 및 외부요인들 즉 환율, 금리, 유가 등의 경제지표들을 의미한다. 노드들은 레이블이 주어지는데, 시계열 데이터로 이루어진 각 지표들은 연속형의 값을 가지므로 레이블을 설정하는데 변형이 필요하다. 따라서 이분형의 레이블로 설정하기 위해 식 (1)의 방법으로 해당지표의 등락을 이용하였고 의미는 <그림 2>와 같다.

$$y = \text{sign}(X_t - MA_5(X_t)) \tag{1}$$



<그림 2> 시계열 데이터의 레이블 설정 방법

t 시점의 데이터를 X_t 라 할 때 레이블의 의미는 실제값과 5일 이동평균값을 비교하여 실제값이 높을 때 +1, 이동평균값이 높을 때 -1로 설정된다. 노드에 레이블이 설정된 그래프기반 semi-supervised learning은 <그림 3>과 같은 기본 모형으로 표현된다.



<그림 3> Time series 예측을 위해 재설계된 graph-based semi-supervised learning의 주가 예측 모형

각 노드들 사이에는 엣지가 형성된다. i노드와 j 노드 사이의 엣지의 연결강도는 유사도가 증가할 수록 커지는데, 다음의 식 (2)에 의해 결정된다.

$$w_{ij} = \begin{cases} \exp\left(-\frac{(x_i - x_j)^T(x_i - x_j)}{\sigma^2}\right) & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

각 지표들간의 상관성 및 인과관계를 나타내는 것은 유사행렬 W를 사용한다. 유사행렬 W를 만드는 방법으로는 주로 k-nearest neighbor(KNN) 또는 일정 반경 내에서의 유클리드 거리를 활용한 방법이 주로 사용된다($\|x_i - x_j\|^2 < r$). SSL 알고리즘은 식 (3)의 이차목적함수를 최소화시켜 레이블이 없는 노드들에 대한 예측값 f를 출력한다[3, 16].

$$\text{Min}_f (f - y)^T(f - y) + \mu f^T L f \quad (3)$$

여기서 y는 목표값으로서 $y = (y_1, \dots, y_1, 0, \dots, 0)^T$ 로 설정되며 출력값(예측값)은 $f = (f_1, \dots, f_1, f_{1+1}, \dots, f_{n-1+n})^T$ 로 표현된다. L은 graph laplacian matrix이며 $D = \text{diag}(d_i)$, $L = D - W$ 로 정의된다. 알고리즘은 두 가지 학습조건인, (a) 출력값 f는 레이블이 된 노드에서는 노드의 목표 값 y와 비슷해야 하고, (b) 연결된 두 노드 I와 j의 출력값은 크게 달라지면 안 된다는 제약을 만족시키는 해를 산출한다. 식 (3)의 파라미터 μ 는 이러한 두 가지 조건이 학습에 미치는 영향을 조절하는 학습 파라미터로서 사용자에게 의해 정의된다. 식 (3)로부터 다음의 식 (4)을 유도하여 출력값 f를 계산할 수 있다.

$$f = (I + \mu L)^{-1} y \quad (4)$$

I는 단위행렬을 의미한다.

SSL은 어떻게 유사행렬 W를 만드는가가 모델의 예측성능을 좌우할 수 있다. 그런데 시계열 데이터는 고유한 잡음(oscillatory noise)을 가지고 있다. 예측성능이 높은 유사도 매트릭스를 구성하기 위해서는 시간의 연관성을 보존하면서 잡음을 줄이는 방법이 필요하다. 따라서 다음 장에서 시계열 데이

터의 시간 연관성을 보존하면서도 잡음을 줄일 수 있는 변환방법을 제시한다.

2.2 기술적 지표(Technical Indicators, TI)

기술적 분석은 금융분석에서 자주 사용하는 방법인데 기술적 지표화의 장점은 시계열 데이터가 고유하게 갖고 있는 잡음(oscillatory noise)을 제거하고 그 밑에 존재하는 큰 구조, 즉 추세 및 구조적 변동요인들을 이끌어 내는 역할을 한다[7]. 주가데이터와 경제지표들은 변수 특성상 시계열 데이터로 존재한다. 이 시계열 데이터는 시간에 따른 연관관계를 형성하고 식 (5)와 같이 표현된다.

$$X_t = \{x_1, x_2, \dots, x_t, \dots, x_t\} \quad (5)$$

<표 1> 기술적 지표

	지표	설명
s ₁	$MA_z(X_t) = \frac{1}{z}(x_t) + \frac{z-1}{z}MA_z(X_{t-1})$	z-이동평균값 (평활값)
s ₂	$BIAS_z(X_t) = \frac{x_t - MA_z(X_t)}{MA_z(X_t)}$	현재값과 이동평균값 차이의 상대비율
s ₃	$OSC_{j,z}(X_t) = \frac{MA_j(X_t) - MA_z(X_t)}{MA_z(X_t)}$	이동평균의 최고 최저점 비율
s ₄	$ROC_z(X_t) = \frac{x_t - x_{t-z}}{x_t}$	현재 값과 z시점 이전 값과의 상대 차이
s ₅	$K_n^z = \frac{C_n - \text{Min}_{i=n-z-1}^n(L_i)}{\text{Max}_{i=n-z-1}^n(H_i) - \text{Min}_{i=n-z-1}^n(L_i)}$	Cn : 구간 내 최중가, Ln : 구간 내 최저가, Hi : 구간 내 최고가
s ₆	$D_n^z = MA_3(K_n^z)$	K_n^z 의 3일 이동평균선
s ₇	$RSI_t^z = \frac{\sum_{i=t-z-1, x_i > x_{i-1}}^t (x_i - x_{i-1})}{\dots}$	상승압력과 하락압력간의 상대적인 강도

여기서 x_t 는 t 를 전체 시간으로 할 때 i 번째 시점의 데이터를 의미한다. x_t 가 시계열 데이터인 특성상 SSL에 직접 적용하기에는 몇 가지 문제점이 존재한다. 예를 들어 <그림 3>에서 DOW 노드도 시계열 데이터 x_t^{DOW} 가 있고 SK-Telecom 노드 또한 고유의 시계열 데이터 $x_t^{SK-Telecom}$ 가 있다. 문제는, 이들 간의 유사도를 두 시계열 데이터들로부터 직접적으로 도출하기가 어렵다는데 있다. 따라서 각 시계열 데이터들을 구조상의 요인들 즉 추세, 변동 요인들로 변환하고 이로부터 두 노드, DOW와 SK-Telecom 간의 유사도를 측정한다. 다음 <표 1>은 본 연구에서 시계열 데이터를 7가지 TI 타입 데이터로 변환시키는데 사용한 기술적 지표들을 정리한 것이다.

3. 실험

3.1 데이터

본 실험에서 사용한 데이터는 Kospie 상장된 개별 회사주가지수 일별 가격을 예측하기 위해서 국가들의 종합주가지수, 경제지표, Kospie 상장된 개별 회사주가지수 총 216개 종류에 대한 시계열 데이터가 사용되었다. 기간은 2007년 1월부터 2008년 8월까지 총 403개의 일별 주가데이터로 이루어져 있다.

<표 2> 데이터 및 변수

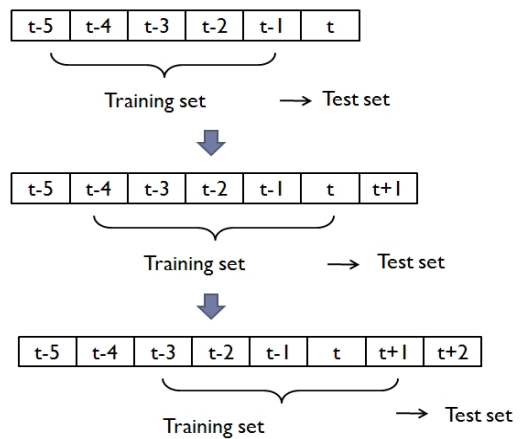
Time	Data	Input variable
Jan, 2007 ~ Aug, 2008	DOW, NASDAQ, NIKKEI, HSI, SSE, TSEC, FTSE, DAX, CAC, BSE_SENSEX, IBOVESPA, AORD, KOSPI, Exchange_rate(KRW-USD), WTI, CD(Certificate of Deposit)	Closing Price, Trading Volume.
	Kospi200_individual company	Closing Price, Trading Volume, Trading Value, PER, Dividend Yield, Capital stock, Listed shares.

동일기간에 대하여 다양한 변수들이 고려 되었다. 개별회사들에서는 종가(시작가, 고가, 저가 포함), 거래량, 거래대금, PER, 배당수익률, 자본금, 상장주식수의 변수들을 선정하였고 국가들의 종합주가지수와 경제지표에서는 종가(시작가, 고가, 저가 포함), 거래량을 선정하였다. <표 2>에서는 본 연구에서 사용한 변수들을 보여준다.

3.2 결과

3.2.1 정확성 비교 결과(AUC)

실험에 사용한 데이터 셋은 다음과 같이 설정하였다. 2007년 1월부터 2007년 5월까지의 79개의 일별 데이터를 학습데이터셋(training set)으로 설정하였고 <그림 4>과 같이 rolling-base forecasting 방법을 이용하여 모델을 만들었다. 2007년 6월부터 2008년 8월까지의 300개의 일별 데이터를 실험데이터셋(test set)으로 정하고, 이기간에 대해 예측을 시도하였다(총 403개의 시점으로부터 기술적 지표생성에 필요한 24개의 시점이 제외되었으므로 총 379개의 벡터화된 데이터를 얻음). 기술적 지표의 파라미터는 가장 많이 사용되고 있는 5일(MA5, BIAS5, RSI5 등)로 설정하였다. t 시점을 예측하기 위해서 SSL의 유사행렬 W 는 $t-1$ 시점에서 계산되었다.



<그림 4> 예측 모델의 실험 구조

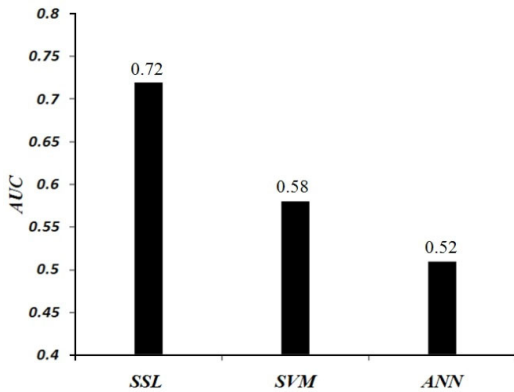
실험에서는 본 연구에서 제안한 SSL 모델을 인공신경망(artificial neural network, ANN), RBF커널함수를 사용한 SVM(support vector machine, SVM_{RBF})모델들과 비교하였다. 각 모델들의 예측 정확도는 ROC 커브의 면적인 AUC(area under the ROC curve)의 값으로 나타낸다[14]. 본 연구에서 제안한 SSL모델의 파라미터 값은 다음과 같다.

$$\{k, \mu\} \in \{2, 3, 4, 5\} \times \{0.01, 0.1, 0.3, 0.5, 0.7, 1, 10, 100\}$$

k, μ 각각은 식 (1)의 이웃노드 개수와 식 (2)의 loss-smoothness tradeoff를 의미하고 cross-validation을 통하여 최적의 조합으로 선정하였다.

비교모델로 사용된 ANN, SVM 모델들 각각에 대하여도 유사한 방법으로 최적의 학습파라미터를 설정하였다.

다음의 <그림 5>는 3개 모델들에 테스트셋 전체를 AUC 결과 값이다.



<그림 5> SSL과 다른 예측모델과의 AUC 결과

SVM과 ANN은 각각 0.58과 0.51의 AUC를 보여 제안한 SSL 모델의 0.72 AUC에 크게 못 미침을 알 수 있다. 본 연구에서 제시한 SSL 모델이 ANN과 SVM의 예측력보다 뛰어난 이유는 SSL의 경우 식 (2)에 의해 독립변수와 종속변수의 일차적 상관성뿐만 아니라 독립변수들간의 영향, 종속변수들간의 영향까지 고려할 수 있는데 반면 ANN과 SVM

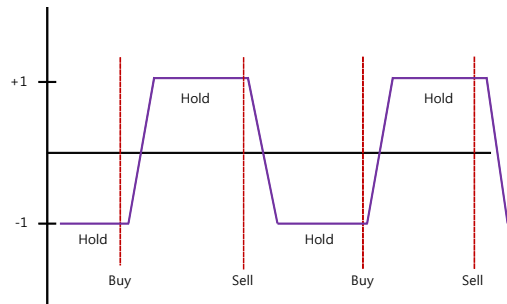
은 독립변수와 종속변수간의 관계를 일차적 상관성만을 고려하여 예측을 하기 때문에 독립변수들간의 영향, 종속변수들간의 영향을 고려하는데 한계가 있어 예측률이 상대적으로 떨어진다고 해석된다. 즉, 현대자동차주가를 예측한다고 가정했을 때 SSL 모델은 현대자동차와 다른 회사들의 주가와 일대일 관계뿐 아니라 다른 회사들끼리의 내재적 상호 관계까지 고려함으로써 더 정확한 예측이 가능하다고 분석된다.

3.2.2 수익성 비교 결과

주가예측모델은 정확도와 함께 실제 투자 시 수익률이 안정적으로 보장되어야 한다. 따라서 제안한 모델의 수익분석 결과를 제시한다. 다음의 <표 3>은 제안된 모델이 실제로 트레이딩이 되었을 때 수익률을 비교한 테이블이다. 매수와 매도의 방법은 <그림 6>과 같이 기본적인 이동평균 기법을 따라 실제가격이 5일 이동평균보다 상승으로 예측될 때 매수하고, 5일 이동평균보다 하락으로 예측될

<표 3> SSL vs. 다른 예측모델 수익률 비교

모델 (2007년 6월 ~2008년 8월)	SSL	SVM	ANN	Buy and Hold
매수/매도 횟수	27회/27회	28회/28회	30회/30회	--
매수/매도 수수료율	0.015%	0.015%	0.015%	0%
수익률	+10.83%	-15.81%	-5.02%	+6.06%



<그림 6> Buy, Hold, Sell의 방법

매 매도를 하였다. 종목은 Kospi의 대표종목 중 하나인 현대자동차를 선정하였고 기간은 테스트 셋의 기간인 2007년 6월부터 2008년 8월까지 300개의 일별 증가 데이터를 사용하였다.

제안된 SSL 모델은 SVM, ANN 모델과 시작시점에 매수 후 매매 없이 유지한 방법(Buy and Hold)과 수익률을 비교하였다.

각 모델들은 테스트 기간 중 실제가격과 5일 이동평균선이 각각 27회, 28회, 30회씩 교차하는 것으로 나타났다. 매매수수료는 온라인 최저수수료를 가정하여 매매(매수/매도) 결과 동일기간에 SSL 모델은 +10.83%의 수익률, SVM 모델은 -15.81%, ANN 모델은 -5.02%, Buy and Hold는 +6.06% 수익률을 나타내었다. 실험 결과에서 보듯이 독립변수들의 내재적 상호 관계까지 고려할 수 있는 SSL모델기법이 다른 모델들에 비해 정확도와 수익성에서 향상된 성능을 보임을 알 수 있다.

5. 결 론

본 연구는 시계열 데이터를 벡터 타입의 데이터로 변형 시킨 후 데이터마이닝(data mining) 및 기계학습(machine learning)분야에서 최신 알고리즘으로 알려지기 시작한 semi-supervised learning을 이용하여 Kospi에 등록된 개별회사주가를 예측하는 방법을 제안하였다[10]. 본 연구에서 제안하는 방법은 다음과 같은 이점이 있다. 첫째, 기존의 모델들처럼 개별회사주가의 시계열적 특성만을 고려하여 예측하는 방법이 아닌 주가의 등락에 변동을 미치는 타기업의 등락 및 경제지표를 네트워크를 이용하여 예측을 가능하게 함으로써 보다 정확한 주가예측을 가능하게 하였다. 기술적으로는 특히 SSL을 사용함으로써 독립변수가 종속변수에 미치는 영향뿐만 아니라 독립변수 간의 상호영향도 모델링에 포함되게 하여 예측력을 향상시킬 수 있다. 둘째, 시계열 데이터를 기술적 지표화함으로써 시계열 데이터의 잡음을 제거하고 동시에 저변에 존재하는 추세 및 구조적 변동요인을 반영하게 할 수

있다. 본 연구에서는 위의 두 가지 기법들이 조합된 시너지 효과로서 AUC 0.72의 정확도와 다른 모델보다 상대적으로 뛰어난 수익성을 도출하였다. 현재 제안된 연구는 기본적인 투자방법인 Buy & Hold and Sell의 방법을 이용하여 수익성을 비교하였지만 좀 더 다양한 투자방법을 도입한다면 수익성을 더 높일 수 있을 것이다. 또한 본 연구에서 제안된 모델을 여러 종목을 이용한 최적 포트폴리오 구성방법과 연관시킨다면 안정성과 수익성을 동시에 높일 수 있으리라 기대된다.

참 고 문 헌

- [1] Amilon, H., "GARCH estimation and discrete stock prices : an application to low-priced Australian stocks," *economics letters*, Vol.81 (2003), pp.215-222.
- [2] Bekiros, S.D. and D.A. Georgoutsos, "Direction-of-Change Forecasting Using a Volatility-Based Recurrent Neural Network," *Journal of Forecasting*, Vol.27(2008), pp.407-417.
- [3] Belkin, M., I. Matveeva, and P. Niyogi, "Regression and Regularization on Large," *In : Shave-Taylor, J., Singer, Y. (eds.) COLT 2004, LNCS (LNAI)*, Vol.3120(2003), pp.624-638.
- [4] Chen, N.-F., R. Roll, and S.A. Ross, "Economic Forces and the Stock Market," *Journal of Business*, Vol.59(1986), pp.383-403.
- [5] Jeantheau, T., "A link between complete models with stochastic volatility and ARCH models," *Finance Stochastics*, Vol.8(2004), pp. 111-131.
- [6] Kanas, A., "Non-linear Forecasts of Stock Returns," *Journal of Forecasting*, Vol.22(2003), pp.299-315.
- [7] Kim, K.-J., "Artificial neural networks with

- evolutionary instance selection for financial forecasting," *Expert Systems with Applications*, Vol.30(2006), pp.519-526.
- [8] Kim, K.-J., "Financial time series forecasting using support vector machines," *NEUROCOMPUTING*, Vol.55(2003), pp.307-319.
- [9] Liu, H.C., Y.H. Lee, and M.C. Lee, "Forecasting China Stock Markets Volatility via GARCH Models Under Skewed-GED Distribution," *Journal of Money, Investment and Banking*, (2009), pp.5-15.
- [10] Park, K. and H. Shin, "Stock Price Forecasting using Semi-Supervised Learning," in *KORMS 2010 Fall Conference*, Seoul National University, (2010), pp.110-116.
- [11] Shin, H., N.J. Hill, A.M. Lisewski, and J.-S. Park, "Graph Sharpening," *Expert Systems with Applications*, Vol.37(2010), pp.7870-7879.
- [12] Shin, H., T. Hou, and K. Park, "Oil Price Prediction From Influence Propagation," in *Proc. of Annual Meeting of Institute for Operations Research and the Management Sciences(INFORMS 2009)*, San Diego, USA, (2009), p.59.
- [13] Tay, F.E.H. and L. Cao, "Application of support vector machines in financial time series forecasting," *The International Journal of Management Science*, Vol.29(2001), pp.309-317.
- [14] Vuk, M. and T. Curk, "ROC Curve, Lift Chart and Calibration Plot," *Metodološki zvezki*, Vol.3(2006), pp.89-108.
- [15] Yang, B., L.X. Li, H. Ji, and J. Xu, "An early warning system for loan risk assessment using artificial neural network," *Knowledge-Based Systems*, Vol.14(2001), pp.303-306.
- [16] Zhou, D., O. Bousquet, T.N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," *Advances in Neural Information Processing Systems*, Vol.16(2004), pp.321-328.
- [17] Zhu, X., "Semi-Supervised Learning with Graphs," *Ph.D. dissertation, Carnegie Mellon University*, 2005.