

## SVM 기반의 재무 정보를 이용한 주가 예측

SVM based Stock Price Forecasting Using Financial Statements

---

저자 (Authors)	허준영, 양진용 Junyoung Heo, Jin Yong Yang
출처 (Source)	<a href="#">정보과학회 컴퓨팅의 실제 논문지 21(3)</a> , 2015.03, 167-172(6 pages) <a href="#">KIISE Transactions on Computing Practices 21(3)</a> , 2015.03, 167-172(6 pages)
발행처 (Publisher)	<a href="#">한국정보과학회</a> KOREA INFORMATION SCIENCE SOCIETY
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06204183">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06204183</a>
APA Style	허준영, 양진용 (2015). SVM 기반의 재무 정보를 이용한 주가 예측. 정보과학회 컴퓨팅의 실제 논문지, 21(3), 167-172
이용정보 (Accessed)	한양대학교 166.***.140.13 2019/11/22 08:08 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

# SVM 기반의 재무 정보를 이용한 주가 예측 (SVM based Stock Price Forecasting Using Financial Statements)

허 준 영 <sup>\*</sup>  
(Junyoung Heo)

양 진 용 <sup>\*\*</sup>  
(Jin Yong Yang)

**요 약** 기계 학습은 컴퓨터를 학습시켜 분류나 예측에 사용되는 기술이다. 그 중 SVM은 빠르고 신뢰할 만한 기계 학습 방법으로 분류나 예측에 널리 사용되고 있다. 본 논문에서는 재무 정보를 기반으로 SVM을 이용하여 주식 가격의 예측력을 검증한다. 이를 통해 회사의 내재 가치를 나타내는 재무정보가 주식 가격 예측에 얼마나 효과적인지를 평가할 수 있다. 회사 재무 정보를 SVM의 입력으로 하여 주가의 상승이나 하락 여부를 예측한다. 다른 기법과의 비교를 위해 전문가 점수와 기계 학습방법인 인공신경망, 결정트리, 적응형부스팅을 통한 예측 결과와 비교하였다. 비교 결과 SVM의 성능이 실행 시간이나 예측력 면에서 모두 우수하였다.

**키워드:** 기계 학습, SVM, 주가 예측, 재무 정보

**Abstract** Machine learning is a technique for training computers to be used in classification or forecasting. Among the various types, support vector machine (SVM) is a fast and reliable machine learning mechanism. In this paper, we evaluate the stock price predictability of SVM based on financial statements, through a fundamental analysis predicting the stock price from the corporate intrinsic values. Corporate financial statements were used as the input for SVM. Based on the results, the rise or drop of the stock was predicted. The SVM results were compared with the forecasts of experts, as well as other machine learning methods such as ANN, decision tree and AdaBoost. SVM showed good predictive power while requiring less execution time than the other machine learning schemes.

**Keywords:** machine learning, SVM, stock prediction, financial statement

## 1. 서 론

주식 투자자들은 미래주가의 예측을 위해 기본적 분석과 기술적 분석을 사용하여 왔다. 기본적 분석은 주가에 영향을 주는 기업 내·외부의 여러 요인들을 분석하여 주식의 내재가치를 추정하고자 하는 방법이다. 장기적으로 주가는 주식의 내재가치에 수렴한다는 믿음을 기반으로 향후 주가흐름을 예측한다. 반면 기술적 분석은 주가의 움직임에서 미래주가를 예측할 수 있는 패턴을 찾아내고자 하는 방법이다. 기술적 분석에서는 기본적 분석과 달리 주식의 내재가치 요인은 고려하지 않고 주가나 거래량의 움직임 같은 과거자료를 이용하여 미래주가를 예측하고자 한다. 이러한 기술적 분석은 1900년대 초 다우이론(Dow theory)[1]을 시작으로 수많은 방법들이 개발되었다. 대부분의 기술적 분석은 도표를 그려 분석하기 때문에 도표 분석 방법이라고도 한다.

하지만 효율적 시장가설(efficient market hypothesis:

· 본 연구는 한성대학교 교내학술연구비 지원과제임

<sup>\*</sup> 종신회원 : 한성대학교 컴퓨터공학과 교수(Hansung Univ.)  
jyheo@hansung.ac.kr  
(Corresponding author)

<sup>\*\*</sup> 비 회 원 : 한성대학교 컴퓨터공학과  
jyang0112@gmail.com

논문접수 : 2014년 10월 2일

(Received 2 October 2014)

논문수정 : 2014년 12월 17일

(Revised 17 December 2014)

심사완료 : 2015년 1월 16일

(Accepted 16 January 2015)

Copyright©2015 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회 컴퓨팅의 실제 논문지 제21권 제3호(2015. 3)

EMH)) 하에서는 주가에 영향을 줄 수 있는 정보가 정확하고 신속하게 주가에 반영되므로 기본적 분석이나 기술적 분석을 통해 미래 주가를 예측하는 것이 불가능해진다[2]. 효율적 시장가설이 성립하는 시장에서는 주식의 내재가치를 결정짓는 기본요인에 대한 정보가 이미 현재의 주가에 반영되어 있으므로 기본적 분석을 통해 초과수익을 얻을 수 없으며, 주가의 움직임은 임의 보행에 가깝기 때문에 어떤 특정 패턴을 찾아내어 초과 수익을 얻는 것도 불가능하다[3]. 이러한 효율적 시장가설의 실증분석을 위해 사건 연구, 수익률의 예측가능성에 관한 연구들이 이루어져 왔다[4]. 전자는 새로운 정보가, 후자는 과거정보가 시장가격에 잘 반영되고 있는지를 검증하는 것이다.

새로운 정보에 대한 시장가격의 반응을 통한 시장효율성검증, 즉 사건 연구(event study)는 Fama와 Fisher, Jensen, Roll에 의해 시작되었다[5-7]. 이들의 연구결과는 효율적 시장가설을 지지하지만 이후 다른 연구자들의 연구결과는 엇갈리고 있다. Ball, Brown의 연구는 지지하는 결과를, Rendleman, Ibbotson 등의 연구는 부정하는 결과를 보여주고 있다[8].

이런 통계에 기반한 연구들의 결과와 함께 컴퓨터 기술의 눈부신 발전에 힘입어 기계학습을 기반으로 한 주가 예측에 대한 연구가 많이 이루어지고 있다. 기계학습 기법은 다차원 공간에 복잡한 비선형 분류를 할 수 있다는 특징이 있어 다양한 실험적 시도가 가능하다는 장점이 있다. 대표적으로 인공신경망, 유전알고리즘, 퍼지 이론, SVM, 결정트리, 적응형부스팅 등 다양한 기계학습 방법이 알려져 있다[9-16].

Hadavandi는 퍼지와 인공신경망을 이용하여 주식 가격 시계열 데이터를 입력으로 하여 미래 가격을 예측하였다[9]. Kim은 시퀀스 정렬 알고리즘을 응용한 주식 가격 예측 기법을 제안하였다[10]. Pai는 주식 가격 시계열 데이터를 입력으로 하여 전통적 시계열 예측인 ARIMA를 사용한 후 SVM을 사용하여 주가 방향에 대한 예측을 수행하였다[11]. Wu는 결정트리를 이용하여 주식 가격 예측 방법을 제안하였다[12]. Han은 다른 기법들과 달리 재무 정보를 기반으로 주가 방향을 SVM을 통해 예측하였다[13]. Kazem은 Support Vector Regression을 이용하여 주식 가격에 대한 예측 모델을 제안하였다[14]. Wen은 SVM을 사용하여 주가 방향에 대한 예측과 자동으로 매수/매도 결정을 내리는 시스템을 제안하였다[15]. Zhiqiang은 주식 가격 시계열 데이터를 입력으로 하여 SVM을 이용한 주가 예측 모델을 제안하였다[16]. 다른 연구와 달리 Zhiqiang은 PSO (particle swarm optimization)을 이용하여 SVM 파라미터를 결정하였다.

본 논문에서는 Han이 연구한 방법[13]과 유사하게 SVM을 사용하여 기본적 분석을 통한 기업 주식 가격 등락을 예측한다. 즉, 기업의 재무정보를 SVM의 입력으로 사용하여 기본적 분석, 즉 기업의 내재 가치를 나타내는 재무 정보에 따른 분석을 하고 이 결과로 주식의 향후 등락을 예측하는 것이다. Han의 방법[13]과 차이는 효율적 시장가설의 실증 분석을 위한 사건 연구로서 기업의 재무 정보 발표를 사건으로 보고 그로 인한 주식 가격 등락을 예측한다. 이를 위해 재무 정보 발표 후 1개월과 2개월 후의 가격을 예측하였다.

기본적 분석을 위한 재무 정보로 자산과 이익에 대한 정보를 활용하였다. 재무 정보에서 자산과 이익에 대한 정보는 대표적으로 해당 기업의 재무 상태를 설명해줄 수 있는 지표이다[17]. 본 논문은 이 지표를 따라 주가의 등락을 예측할 수 있는지, 그리고 그 예측이 어느 시점까지 가능한지 기계학습 기법인 SVM을 사용하여 평가하였다. 본 논문에서 사용한 재무 정보는 상대가치지표로 주당순이익(EPS), 주당순자산(BPS)과 성장성을 고려해 볼 수 있는 순이익증가율(NPGR)이다[13,18].

실제로 주가는 학술적 연구뿐 아니라 증권 애널리스트들에 의해 분석되고 예측되고 있다. 애널리스트들의 예측은 주식 시장에서 그나마 신뢰받는 정보로, 이들의 분석은 어느 정도 검증된 분석, 예측 기술에 바탕을 둔 것으로 볼 수 있다. 따라서, 전문가의 예측과 기술적 예측 방법의 비교가 필요하다. 본 논문에서는 이런 전문가의 예측과 본 논문에서 제안하는 방법의 예측을 비교하여 전문가의 예측보다 더 우수함을 보였다.

본 논문은 다음과 같이 구성된다. 2장에서는 본 논문에서 해결하려는 문제를 설명한다. 3장에서는 논문에서 사용한 데이터에 대한 기술을 하고, 4장에서는 SVM을 이용한 실험 방법과 결과를 설명하고, 5장에서 결론을 맺는다.

## 2. 문제 기술

분기별로 발표되는 기업의 재무 정보를 바탕으로 주식 가격 상승/하락 여부를 기계학습을 이용하여 예측한다. 기업의 분기별 재무 정보 발표가 주식에 얼마나 영향을 미치게 되며 시간이 지남에 따라 그 영향이 얼마나 차이가 있는지 알아보기 위함이다.

기업의 재무 정보에서 기업의 주식 가격과 밀접한 주당순이익과 주당순자산, 당기순이익을 입력으로 하여 한 달과 두 달 후 주식 가격이 상승/하락할지 예측한다. 주당순이익과 주당순자산, 당기순이익은 기존 연구에 따르면 일반적으로 주식 가격 예측에 유의한 것으로 알려져 있다[13,19]. 예측의 성능을 비교하기 위해 전문가 투자 의견점수와 비교하여 재무정보를 활용한 기계학습 예측

력을 평가한다. 그리고 한 달과 두 달 후의 주식 가격 상승여부 예측을 통해 재무 정보에 따른 주식 예측이 기간에 따라 얼마나 차이가 나는지 평가한다.

주당순이익과 주당순자산의 경우 기업의 규모에 따라 상대적으로 결정되는 값이지만, 당기순이익의 경우 기업 규모에 따라 매우 큰 차이를 보인다. 따라서 당기순이익을 그대로 입력으로 사용하지 않고 직전 분기의 당기순이익과 비교하여 비율을 사용한다. 즉, 당기순이익증가율을 사용한다.

### 3. 데이터 설명

본 논문에서 사용한 기업 데이터는 2013년 기준으로 KOSPI 200에 속한 200개 기업에 대해 2010년 1분기부터 2013년 3분기까지 분기별로 수집한 재무 데이터이다. 이 재무 데이터에서 주당순이익(EPS, Earning Per Share)과 주당순자산(BPS, Book-value Per Share), 당기순이익을 선택하고, 증권회사에서 발표하는 투자자의견 점수(5점부터 1점)를 수집하였다. 그리고 각 분기마다 분기 마지막 일을 기준으로 한달 후, 두달 후 주식 가격을 수집하였다. 위 데이터는 모두 FnGuide에서 제공하는 DataGuide에서 조회하였다. 200개 회사에 대해 총 15분기의 정보를 조회하면 총 3000개의 데이터가 나와야 하나, 경우에 따라 필요한 정보가 빠져 있는 경우가 있어서 총 2913개의 샘플만 확보할 수 있었다. 표 1은 삼성전자의 샘플을 예로 보인 표이다.

EPS와 BPS는 기업의 규모에 상관없이 주식 가격을 기준으로 이익과 자산을 계산한 것이므로 기계학습에 사용하는데 문제가 없다. 하지만 당기순이익의 경우 기업의 규모에 따라 매우 큰 차이를 보이기 때문에 당기순이익율(NPGR, Net Profit Growth Rate)로 계산하여 사용하였다. 당기순이익율은 재무 데이터에서 당기순이익을 분기별로 수집하고 분기마다 증가/감소율을 계산하였다.

$$\text{당기순이익율(\%)} = (\text{당기순이익} - \text{직전분기당기순이익}) / \text{직전분기당기순이익} \times 100$$

표 1 삼성전자의 재무 정보와 주가(단위: 원)  
Table 1 Financial statements and stock price of Samsung Electronics (unit: Won)

Quarter	2010-03-31	2010-06-30
Net Profit	3,167,036,000	3,155,614,000
EPS	21,370.00	21,125.00
BPS	445,444.69	460,047.68
Stock Price	814,000	792,000
1M Later	825,000	827,000
2M Later	778,000	776,000
Experts(5-1)	4 points	4 points

표 2 삼성전자 데이터 전처리 결과

Table 2 Pre-processing results of Samsung Electronics data

Quarter	2010-03-31	2010-06-30
EPS	21,370.00	21,125.00
BPS	445,444.69	460,047.68
NPGR	3.71	-0.36
1M Later	+1	+1
2M Later	-1	-1
Experts(5-1)	+1	+1

표 3 학습/테스트 데이터

Table 3 Training/Test data

EPS	Earnings Per Share
BPS	Book-value Per Share
NPGR	Net Profit Growth Rate
1M Later (Target_1)	+1 if the price rises after one month later. -1 if drops.
2M Later (Target_2)	+1 if the price rises after two month later. -1 if drops.
Experts	+1 if experts' score is 3~5. -1 if experts' score is 1~2.

분기 한달 후와 두달 후 주식 가격은 증가/감소 여부에 따라 +1과 -1로 계산하였다. 예를 들어 2010년 1분기 한달 후(4월 말일) 주가가 1분기(3월 말일)보다 상승하였다면 +1로 하고, 두달 후(5월 말일) 주가가 1분기보다 감소하였다면 -1로 하였다. 투자자의견점수는 기계학습 예측과 비교하기 위한 비교군으로 5점과 4점, 3점을 주가 상승(+1)으로 하고, 나머지는 주가 하락(-1)으로 가정하였다. 이렇게 나눌 때 가장 좋은 예측 결과가 나왔기 때문이다. 표 2는 표 1의 삼성전자 데이터를 전처리한 데이터를 보여준다.

정리하면, 200개 기업에 대한 표 3과 같이 재무 데이터와 주식 가격을 수집하여 SVM의 학습과 테스트에 사용하였다.

### 4. 실험

수집한 2913개의 샘플을 무작위로 섞은 후 절반을 나누어 학습용(1457개)과 테스트용(1456개)으로 사용하고, 이 둘을 바꾼 후에 다시 학습과 테스트를 수행하였다. 그리고 이 과정을 10회 반복하여 평균 예측률을 계산하였다. 데이터를 무작위로 섞고, 상호검증을 반복적으로 함으로써 학습과 테스트 데이터의 우연에 의해 나타날 수 있는 SVM 예측 결과를 상쇄하도록 하였다.

각 기업 분기 재무 정보에서 수집한 EPS, BPS, NPGR 중 어떤 조합이 예측력이 더 우수한지 비교하기 위해 모든 가능한 조합에 대하여 실험을 하였다. 즉,

표 4 파라미터에 따른 SVM 예측 정확도(%) 입력 = {EPS, BPS}

Table 4 SVM prediction results based on the parameters ( $\gamma$ , C) - prediction accuracy (%) features={EPS, BPS}

	$\gamma = 1$	0.1	0.5	0.01	0.001
C=1	57.3	57.4	57.4	57.3	57.02
5	57.3	57.3	57.4	<b>57.5</b>	56.64
10	57.3	57.3	57.36	57.32	57
Average=57.30%, Standard Dev.=0.22					

{EPS}, {BPS}, {NPGR}, {EPS, BPS}, {EPS, NPGR}, {BPS, NPGR}, {EPS, BPS, NPGR} 각각을 입력으로 하는 예측 모델을 만들었다.

SVM의 비선형 분류를 위해 커널 함수를 RBF(Radial Basis Function)를 사용하였다. SVM 구현은 LibSVM [20]을 파이썬(Python)에서 쓸 수 있도록 만든 mlpy (machine learning py)를 사용하였다[21]. 이때 파라미터  $\gamma$  ( $\gamma = \frac{1}{2\sigma^2}$ )는 0.01, C는 5를 사용하였다. 이 파라미터를 선정하기 위해 다양한  $\gamma$ 와 C값에 대해 표 4와 같은 실험 결과를 구하였다. 그 중 가장 좋은 결과를 낸 파라미터 값을 선택하였으나, 예측 정확도가 파라미터에 따라 차이가 크지 않았다. 비교를 위한 인공신경망은 neurolab을 사용하였고[22], 결정트리와 적응형 부스팅은 scikit learn 모듈을 사용하였다[23]. 인공신경망은 입력 층의 노드 수는 입력 변수의 수와 동일하게 하고, 출력 층의 노드 수는 1개, 히든 층의 노드 수는 10개로 하였다. 적응형부스팅의 약분류기(weak classifier)로는 결정트리를 사용하였다.

#### 4.1 실험 결과

표 5는 실험 결과를 나타낸 표이다. 재무 정보 발표 한달 후 주가 예측(SVM(1M))은 두 달 후 예측(SVM(2M))에 비해 어떠한 입력을 사용하더라도 예측력이 높음을 알 수 있다. 즉, 시간이 지남에 따라 재무 정보에 따른 주가 예측력이 떨어짐을 확인할 수 있다.

표 5에서 SVM(1M)과 Expert에 해당하는 결과를 비교해보면 쉽게 한 달 후 예측이 전문가 점수보다 뛰어난 것을 볼 수 있다. 다만, 전문가 점수가 증권회사 등에서 발표하는 전문가 점수를 단순히 평균한 것이므로 이 결과만으로 SVM의 예측이 전문가보다 뛰어나다고 단정 짓기는 어렵다.

표 5에서 SVM(1M)과 ANN, D.Tree, Adaboost 결과를 비교해보면 SVM의 예측 결과가 다른 기계학습 방법에 비해 우수함을 알 수 있다. SVM의 예측력과 가장 근접한 결과를 보이는 인공신경망(ANN)의 경우 학습 시간을 최대 10분으로 하였고, 매번 학습 시간을 초

표 5 예측 결과 - 예측 정확도(%) 평균과 표준편차  
Table 5 Prediction results - prediction accuracy (%) and standard deviation

	eps, bps, npgr	eps, npgr	eps, bps
svm(1m)	57.1 (0.12)	54.6 (0.02)	57.5 (0.11)
svm(2m)	50.1 (0.21)	50.6 (0.22)	50.4 (0.23)
expert	48.6	48.6	48.6
ANN(1m)	57.0 (0.30)	53.0 (0.25)	56.5 (0.31)
D.Tree(1m)	53.9 (0.25)	51.0 (0.22)	55.4 (0.31)
Adaboost(1m)	56.0 (0.32)	52.5 (0.33)	55.9 (0.41)

	bps, npgr	eps	bps	npgr
svm(1m)	56.1 (0.34)	53.7 (0.44)	55.2 (0.53)	56.7 (0.45)
svm(2m)	49 (0.55)	49 (0.53)	47.7 (0.49)	52.9 (0.37)
expert	48.6	48.6	48.6	48.6
ANN(1m)	54.5 (0.33)	53.7 (0.23)	54.7 (0.45)	54.9 (0.43)
D.Tree(1m)	54.9 (0.25)	49.7 (0.21)	51.7 (0.19)	53.5 (0.22)
Adaboost(1m)	55.0 (0.30)	51.7 (0.25)	53.6 (0.23)	53.9 (0.31)

과하여 학습이 종료되었다. 즉, 본 실험에서 사용한 데이터와 같이 인공신경망의 노드 가중치가 빠르게 수렴되지 않는 경우 학습 시간이 너무 오래 걸리는 단점이 있다. SVM은 빠른 학습시간에도 불구하고 인공신경망보다 나은 예측 결과를 보였다.

표 5에서 SVM(1M)의 결과를 중심으로 입력 변수에 따른 결과를 비교해보면 하나의 입력 변수로는 {NPGR}이 제일 예측력이 우수하지만, 둘 이상을 조합하는 것이 예측력 향상에 도움이 됨을 알 수 있다. 본 실험에서 사용한 데이터의 경우 {EPS, BPS}만으로 된 입력 조합이 가장 우수한 결과를 보여준다. NPGR까지 포함한 {EPS, BPS, NPGR} 입력 조합 보다 더 우수하다. 이는 입력 변수가 많다고 항상 좋은 결과가 나오지 않음을 의미한다.

예측 정확도가 57.5%로 기존의 주가 방향 예측 연구들과 유사한 성능을 나타내고 있다. 50% 초과 예측 정확도는 승률게임에서는 매우 유용할 수 있다. 특히 오랜 기간에 걸쳐 여러 번 게임이 이루어지는 경우 복리효과가 쌓이기 때문에 51%의 승률도 의미를 가질 수 있다.

#### 4.2 결과의 통계적 검증

SVM의 예측 정확도에 대한 통계적 검증을 위해 다음 두 가지 귀무가설에 대해 p-value를 계산하고자 한다.

이때 표 5에서 결과가 가장 좋은 입력이 {EPS, BPS}인 경우만을 고려한다.

1번 귀무가설: SVM의 예측 정확도는 무작위 예측에 의해서도 우연히 발생할 가능성이 있다.

2번 귀무가설: SVM의 예측 정확도와 다른 기계학습 방법(ANN, D.Tree, Adaboost)의 정확도의 차이는 우연히 발생할 가능성이 있다.

먼저 1번 귀무가설이 발생할 확률(p-value)을 계산하기 위해 무작위로 주가 예측을 시도한다고 가정한다. 이때 예측이 맞을 확률은 반반, 즉 0.5이다. 실험에서 총 1456개의 테스트 데이터에 대해 예측을 하기 때문에 SVM(1M) 만큼 예측 정확도가 나오기 위해서는 57.5% (837개)에 대해 예측이 성공해야 한다. 이러한 무작위 예측은 베르누이 시행으로 볼 수 있다. 즉, 다음과 같이 계산된다.

$$P(X = 837) = \binom{1456}{837} 0.5^{1456}$$

여기에서 57.5% 이상의 정확도가 나올 확률을 계산하면 1번 귀무가설에 대한 p-value를 구할 수 있다.

$$\begin{aligned} P(X > 836) &= 1 - P(X < 837) \\ &= 1 - F_X(836) \\ &= 4.5 \times 10^{-9} \end{aligned}$$

1번 귀무가설에 대한 p-value는 0에 가까운 아주 작은 값으로 귀무가설을 기각할 수 있고, 대립가설인 SVM(1M)의 예측 정확도는 통계적으로 유의미한 결과라고 할 수 있다.

2번 귀무가설에 대한 검증을 위해 예측 정확도의 평균이 중심극한정리에 의해 정규분포에 근사함을 보인다. 10번의 실험 결과(각각에 대한 확률 변수를  $X_n$ 이라 함)를 더한 것은  $Y = X_1 + X_2 + \dots + X_n$  이고, 중심 극한 정리에 의해 정규분포  $N(10\mu, 10\sigma^2)$ 에 근사하게 따른다고 할 수 있다. 평균을 위해 10으로 나누게 되면, 즉 Y에 0.1을 곱하면 정규분포 성질에 따라 평균에 대한 확률변수 0.1Y는 정규분포  $N\left(\mu, \frac{\sigma^2}{10}\right)$ 를 근사하게 따른다.

동일한 분포에서 샘플링된 두 그룹의 평균 차이에 대한 분포는 정규분포 성질에 따라  $Y_1 - Y_2 \sim N\left(0, \frac{\sigma^2}{5}\right)$ 라고 할 수 있다. 이 결과를 이용하여 2번 귀무가설에 해당하는 확률 p-value를 구한다.

2번 귀무가설에 따라 SVM(1M)과 ANN의 결과가 동일 분포에서 나왔다면 두 측정된 평균의 차이는 앞에서 구한 평균 차이 분포( $Y_1 - Y_2$ )에 따라 정규분포  $N(0, 0.022)$ 에 근사한다고 볼 수 있다. 이때 표준편차  $\sigma$ 는 SVM(1M)과 ANN의 결과 전체에 대해 계산한 값(0.33)을 사용한다. SVM(1M)과 ANN의 결과 차이, 즉

평균 차이는 1이다. 평균 차이 1이상이 발생할 확률은 다음과 같이 구할 수 있다.

$$\begin{aligned} P(|Y_1 - Y_2| \geq 1) \\ &= F_{Y_1 - Y_2}(-1) + (1 - F_{Y_1 - Y_2}(1)) \\ &= 3.5 \times 10^{-22} \end{aligned}$$

2번 귀무가설에 대한 p-value는 0에 가까운 아주 작은 값으로 귀무가설을 기각할 수 있고, 따라서 SVM(1M)의 결과가 ANN의 결과 차이가 통계적으로 유의미한 결과라고 할 수 있다. 마찬가지로 D.Tree와 Adaboost의 결과에 대한 p-value도 0에 가까운 값이 나오기 때문에 SVM(1M)의 결과가 이 방법들에 비해 우수하다고 볼 수 있다.

## 5. 결론

본 논문에서 SVM과 같은 기계학습을 통하여 재무정보를 기반으로 주식 가격의 변화를 예측하였다. 기업의 재무정보를 SVM의 입력으로 사용하여 기본적 분석을 하고 이 결과로 주식의 향후 등락을 예측하는 것이다. 기본적 분석을 위한 재무 정보로 자산과 이익에 대한 정보를 활용하였다. 재무 정보에서 자산과 이익에 대한 정보는 대표적으로 해당 기업의 재무 상태를 설명해줄 수 있는 지표이다. 이 지표에 따라 주가의 등락을 예측할 수 있는지, 그리고 그 예측이 어느 시점까지 가능한지 SVM을 사용하여 평가하였다.

실험 결과 재무 정보를 활용한 SVM의 주가 예측력은 전문가 예측에 비하여 우수한 예측력을 보여주며, 기간이 지날수록 예측력이 떨어지게 된다. 이는 재무 정보를 기반으로 한 예측이 단 기간에는 우수하지만 일정 기간 후에는 재무 정보와 주식 가격의 부조화가 합리적 투자자들에 의하여 상쇄된다고 볼 수 있다. 또한 SVM은 다른 기계학습 방법인 인공신경망이나 결정트리, 적응형부스팅에 비해 수행 속도나 예측력이 더 뛰어난 것을 확인하였다.

## References

- [1] Hamilton, W. P., *The Stock Market Barometer: A Study of its Forecast Value Based on Charles H. Dow's Theory of the Price Movement*, Barrons, New York, 1922.
- [2] Malkiel, B. G. (1987). Efficient market hypothesis, *The new palgrave: A dictionary of economics*, 2, 120-23.
- [3] Conrad, J., & Kaul, G. (1989). Mean reversion in short-horizon expected returns, *Review of Financial Studies*, Vol. 2, No. 2, pp. 225-240.
- [4] Timmermann, A. and C. WJ Granger, "Efficient market hypothesis and forecasting," *International*

- Journal of Forecasting*, Vol. 20, No. 1, pp. 15-27, 2004.
- [5] Fama, E. F. (1991). Efficient capital markets: II, *The journal of finance*, Vol. 46, No. 5, pp. 1575-1617.
- [6] Lim, K.-P. and R. Brooks, "The evolution of stock market efficiency over time: a survey of the empirical literature," *Journal of Economic Surveys*, Vol. 25, No. 1, pp. 69-108, 2011.
- [7] Sewell, M., "History of the efficient market hypothesis," *RN* Vol. 11, No. 04, 2011.
- [8] Rendleman Jr, R. J., Jones, C. P., & Latane, H. A. (1982). Empirical anomalies based on unexpected earnings and the importance of risk adjustments, *Journal of Financial Economics*, Vol. 10, No. 3, pp. 269-287.
- [9] Hadavandi, E., H. Shavandi, and A. Ghanbari, "Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting," *Knowledge-Based Systems*, Vol. 23, No. 8, pp. 800-808, 2010.
- [10] H. Kim, and H. Cho, "Developing Stock Pattern Searching System using Sequence Alignment Algorithm," *Journal of KIISE : Computer Systems and Theory*, Vol. 37, No. 6, pp. 354-367, Dec. 2010.
- [11] Pai, P.-F. and C.-S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, Vol. 33, No. 6, pp. 497-505, 2005.
- [12] Wu, M.-C., S.-Y. Lin and C.-H. Lin, "An effective application of decision tree to stock trading," *Expert Systems with Applications*, Vol. 31, No. 2, pp. 270-274, 2006.
- [13] Han, S., and R.-C. Chen, "Using SVM with Financial Statement Analysis for Prediction of Stocks," *Communications of the IIMA*, Vol. 7, No. 4, pp. 63-72, 2007.
- [14] Kazem, Ahmad, et al., "Support vector regression with chaos-based firefly algorithm for stock market price forecasting," *Applied Soft Computing* 13.2, pp. 947-958, 2013.
- [15] Wen, Qinghua, et al., "Automatic stock decision support system based on box theory and SVM algorithm," *Expert Systems with Applications* 37.2, pp. 1015-1022, 2010.
- [16] Zhiqiang, Guo, Wang Huaqing, and Liu Quan. "Financial time series forecasting using LPP and SVM optimized by PSO," *Soft Computing* 17.5, pp. 805-818, 2013.
- [17] Chung, C. H. and S. K. Kim, "An Investigation on the Stock Return Predictability of Dividend Yield and Earning-Price Ratio," *The Korean Journal of Financial Engineering*, Vol. 9, No. 3, pp. 61-87, 2010.
- [18] Song, D.-S., "A Study on the Relation Between the Financial Ratio and Earnings Quality," *Korea International Accounting Review*, Vol. 40, pp. 135-156, 2011.
- [19] Kim, K. Y. and Y. B. Kim, "Testing the Predictability of Stock Return in the Korean Stock Market," *Korean Journal of Industrial Economic*, Vol. 17, No. 4, pp. 1255-1271, 2004.
- [20] Chang, C.-C. and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001. (Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.)
- [21] Albanese, D., R. Visintainer, S. Merler, S. Riccadonna, G. Jurman and C. Furlanello. *mlpy: Machine Learning Python*, 2012.
- [22] NeuroLab, <https://pythonhosted.org/neurolab/>, 2011.
- [23] scikit learn, <http://scikit-learn.org/>, 2010.



허 준 영

1998년 서울대학교 컴퓨터공학과 졸업(학사). 2009년 서울대학교 컴퓨터공학부 졸업(박사). 2009년~현재 한성대학교 컴퓨터공학과 조교수. 관심분야는 운영체제, 무선 센서 네트워크, 임베디드 시스템, 기계 학습, 금융 공학



양 진 용

2013년 한양대학교 경제학과 졸업(박사) 2014년~현재 한성대학교 컴퓨터공학과 박사과정. 현재 자유 연구가로 활동. 관심분야는 계량금융, 계산금융, 경제물리