

딥러닝분석과 기술적 분석 지표를 이용한 한국 코스피주가지수 방향성 예측

이우식¹

¹안양대학교 정보통계학과

접수 2017년 1월 17일, 수정 2017년 3월 2일, 게재확정 2017년 3월 10일

요약

2016년 3월 구글 (Google)의 바둑인공지능 알파고 (AlphaGo)가 이세돌 9단과의 바둑대결에서 승리한 이후 다양한 분야에서 인공지능 사용에 대한 관심이 높아지고 있는 가운데 금융투자 분야에서도 인공지능과 투자자문 전문가의 합성어인 로보어드바이저 (Robo-Advisor)에 대한 관심이 높아지고 있다. 인공지능 (artificial intelligence)기반의 의사결정은 비용 절감은 물론 효과적인 의사결정을 가능하게 한다는 점에서 큰 장점이 있다. 본 연구에서는 기술적 분석 (technical analysis) 지표와 딥러닝 (deep learning) 모형을 결합하여 한국 코스피 지수를 예측하는 모형을 개발하고 제시한 모형들의 예측력을 비교, 분석한다. 분석 결과 기술적 분석 지표에 딥러닝 알고리즘을 결합한 모형이 주가지수 방향성 예측 문제에 응용될 수 있음을 확인하였다. 향후 본 연구에서 제안된 기술적 분석 지표와 딥러닝모형을 결합한 기법은 로보어드바이저서비스에 응용할 수 있는 일반화 가능성을 보여준다.

주요용어: 기술적지표, 딥러닝, 로보어드바이저, 인공지능, 핀테크.

1. 서론

한국 금융시장에서 예금과 적금 이자율이 2.5%를 넘는 상품은 거의 사라져 찾아보기 힘들고, 계속되는 금리 인하에 현재 기준금리는 1.25%로 역대 사상 최저 수준으로 낮아졌다. 연 1%대의 저금리 시대에 목돈 마련과 노후 대비를 위해서 금융투자가 필수가 된 상황이다 (Jung와 Park, 2016).

금융투자를 위한 일반적인 주식예측방법으로 기본적 주가분석과 기술적 주가분석이 존재한다. 기본적 주가분석은 기업의 내재가치에 영향을 미치는 모든 요소들을 분석하는 방법이지만 개별 기업의 내재가치 정의 및 측정방식 등을 기반으로 미래의 주가를 예측하는 것은 한계가 있다. 기술적 주가분석은 ‘주가는 수요와 공급의 원리에 따라 결정되며 수급의 결과는 그래프로 나타난다’는 이론을 배경으로, 그래프 분석을 통해 미래 주가를 예측하는 방법이다. 컴퓨터 알고리즘의 발달과 더불어 기본적 주가분석과 기술적 주가분석을 통한 미래의 주가를 예측하는 방법들이 진화하고 있지만 직간접적으로 서로 복잡하게 얽혀있는 수많은 변수들에 의해서 불규칙적으로 변화하기 때문에 정확히 주가를 예측하는 것은 한계가 있을 뿐 아니라 매우 어렵고 복잡하다.

2016년 3월 구글 (Google)의 바둑인공지능 알파고 (AlphaGo)가 이세돌 9단과의 바둑대결에서 승리한 이후 금융업계에서는 인공지능과 투자자문 전문가의 합성어인 로보어드바이저 (Robo-Advisor)에 대한 관심이 높아지고 있고 이에 KB국민은행, 우리은행, KEB하나은행, 신한은행, NH농협은행 등 국내

¹ (14028) 경기도 안양시 만안구 삼덕로 37번길 22, 안양대학교 정보통계학과, 외래교수.
E-mail: woosiklee@hotmail.com

주요 은행과 삼성증권, 신한금융투자, 미래에셋대우, 한국투자증권, 현대증권 등 증권사들은 자체적으로 시스템을 개발하거나 로보어드바이저 스타트업체와 제휴하는 방식으로 추진하고 있다 (Ko, 2016). 현재 일반 투자자를 대상으로 저비용 자산관리 서비스를 제공하는 로보어드바이저에 의해서 운용되는 자금은 전 세계적으로 약 20조원이고 2020년에는 약 2,000조원으로 예상하고 있다 (Chang, 2016).

우리나라의 경우 2016년 9월부터 정부 차원의 로보어드바이저 서비스에 대한 검증을 금융위원회 (financial supervisory commission)에서 실시하고 있고 안정적으로 운영된 로보어드바이저 업체들에 대해 투자에 대한 자문을 물론 투자를 대행할 수 있는 권한까지 준다는 계획이다 (Ko, 2016). 이런 정책은 로보어드바이저시장이 활성화될 수 있는 영향력을 미치며 인공지능의 영역을 넓혀가고 있지만 딥러닝 (deep learning)과 금융 (finance)이 연계된 연구들은 많이 이뤄지지 않고 있는 상태이다.

기존 연구 중 딥러닝의 적용 가능성을 제시한 Lee와 Chun (2016)은 GARCH (generalized autoregressive conditional heteroskedasticity)모형과 딥러닝을 통합한 DL-GARCH (deep learning-generalized autoregressive conditional heteroskedasticity)모형을 가지고 중국 위안화 (Chinese Yuan) 변동성 (volatility)예측을 비교 실험과 분석을 하였고 그 결과 딥러닝모형이 높은 성과를 보인 것으로 나타났다.

본 연구에서 인공지능을 포함한 기계학습을 통한 한국 코스피 지수의 예측력을 분석하고자 한다. 분석기간은 2000년 1월 4일부터 2016년 2월 12일까지 일별 지수 증가 값을 기초로 한다.

2. 이론적 배경 및 연구방법

2.1. 이론적 배경

2.1.1. 의사결정나무모형

의사결정나무모형 (decision tree)은 자료 내에 존재하는 관계 및 규칙 등을 찾아내어 그 의사결정규칙을 나무 구조로 도표화하여 탐색적 절차에 유용할 뿐 아니라 예측 및 분류를 수행할 수 있다는 장점을 가지고 있다 (Kwak과 Rhee, 2016). 의사결정나무모형은 분석의 목적과 자료구조에 따라 적절한 분리기준과 정지규칙을 지정하고 의사결정가지치기와 같은 특징선택을 통해 분류에 가장 필요한 특징들만을 추출함으로써 원자료에 비해 줄어든 자료를 얻을 수 있으며, 분류의 기준이 되기에 기여도가 떨어지는 잡음, 중복자료 그리고 규칙을 제거할 수 있다.

의사결정나무모형을 만드는 데는 카이제곱통계량을 분리 기준으로 하는 카이 제곱 자동 상호 작용 검출모형, 지니계수를 분리 기준으로 하는 분류회귀나무모형, 그리고 엔트로피 지수를 분리 기준으로 하는 C4.5등 다양한 모형이 있다. 본 연구에서는 지니계수를 분리 기준으로 하는 분류회귀나무모형이 고려되었다.

2.1.2. 서포트벡터머신모형

서포트벡터머신 (support vector machine)은 서로 다른 두 집단으로 분류시키는 최적의 초평면을 찾는 이진 분류기이다. 두 집단을 나누는 방법, 즉 마진 (margin)을 최대화하는 문제는 최적화 문제로 바꿀 수 있다. 서포트벡터머신의 대표적인 커널함수 (Kernel function)는 선형커널, 다항식커널, RBF커널, 시그모이드커널이 있다. 입력이 x_i 이고 출력이 $y_i \in \{1, -1\}$ 인 N 개의 자료 $\{(x_i, y_i)\}_{i=1}^N$ 라 할 때, 결정평면은 가중벡터가 w 이고 바이어스가 b 인 $wx + b = 0$ 로 표시된다. 이 때 두 집단의 서포트벡터 사이의 분리 마진은 $2/\|w\|$ 가 되며, 이를 최대화하기 위해 $\|w\|$ 를 최소화해야 한다. 여기서 결정평면과 목표클래스의 출력의 곱이 양수라는 제약조건을 적용하여 비용함수가 최소가 되는 w 와 b 를 계산하면 최적화된 결정평면을 얻을 수 있다.

$$\Phi(w) = \frac{1}{2} \|w\|^2$$

두 클래스에 포함된 자료들이 선형적으로 완전 분리가 어려운 경우 오류에 대한 허용 변수를 적용한 비용함수를 통해 최적화가 가능하다. 이 때 학습자료들을 이용하여 마진의 최대화와 분류 허용 에러에 대한 조절 인자인 코스트의 적절한 값을 경험적으로 결정해야 한다.

$$\Phi(w) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i$$

최적화 문제는 라그랑지승수 (Lagrange multiplier) 방법을 도입하여 해를 찾을 수 있는데, 분류하고자 하는 두 클래스가 비선형 분리면을 갖는 경우에는 커널함수를 도입하여 입력벡터, 즉 입력변수를 더 높은 차원의 특징 공간으로 사상하여 그 공간에서 초평면을 얻음으로써 선형적으로 분리할 수 있다. 커널함수를 이용할 경우 최종적으로 얻어지는 서포트벡터머신분류기를 이용하여 두 개의 비선형 분리면을 갖는 클래스로 분류할 수 있다.

2.1.3. 딥러닝모형

인공신경망 (artificial neural network) 모형은 층, 연결강도, 전이 함수, 학습 알고리즘 등으로 이루어진 데이터처리 시스템으로서 입력 자료 값과 해당 출력 자료 값들을 통해 가중치들이 반복적으로 조정되어 결국 입력 및 출력자료간의 관계가 학습되는 구조이다.

자료에 대한 표현 능력을 크게 증가시키기 위해 은닉뉴런수를 증가시키거나, 은닉층을 추가하여 사용하였지만 역전파 (backpropagation) 알고리즘이 에러 정보를 출력층에서 입력층 방향으로 전달되면서 점점 사라진다는 문제가 발생하였고, 이에 2006년 힌튼교수가 사전학습이 포함된 딥러닝 (deep learning) 모형을 제시하였다 (Choi와 Min, 2015). 딥러닝의 경우 기존 인공신경망과 달리 마지막 은닉층까지 충분히 전파가 가능하여 딥러닝 분석을 이용한 금융자료예측에 적용하고 있다 (Lee와 Chun, 2016).

2.2. 연구 모형

본 연구는 기술적 주가분석기법을 의사결정나무모형, 서포트벡터머신모형 그리고 딥러닝모형에 결합함으로써 한국 코스피지수의 상승 또는 하락을 예측 하고자 한다.

2.2.1. 기술적 분석 지표

본 연구에서는 상대강도지수, 모멘텀, 변화율지표, 지수이동평균교차등의 기술적 주가분석 지표들을 이용하여 한국 코스피 지수의 상승 또는 하락에 대한 예측 성능을 비교한다.

상대강도지수는 주가의 상향 이동과 하향 이동을 비교하는 모멘텀 오실레이터 (momentum oscillator)로, 0에서 100사이의 값을 가진다. 상대강도지수값이 0에 가까우면 가격이 상향 이동하고 상대강도지수값이 100에 가까우면 가격이 하향 이동함을 나타낸다 (Jung과 Min, 2013).

모멘텀은 주가추세의 속도나 강도를 파악하는데 사용되는 오실레이터로 주가가 상승하거나 하락하고 있을 때 얼마나 가속하게 될지를 예측하는 기술적 분석 지표이다.

변화율지표 (rate of change indicator)는 과거 일정시점의 가격과 현재가격을 비교하여 주가추세의 반전을 알려주어 주식시장특성 분석에 사용된다. 변화율지표의 움직임이 0선 위쪽에서 나타나면 주식가격은 상승 추세에 있다고 보고 반대로 변화율지표의 움직임이 0선 아래쪽에서 나타나면 주식가격은 하락 추세에 있다고 볼 수 있다.

지수이동평균 (exponential moving average)교차에서 지수 이동 평균을 주식의 추세 (trend)를 파악하기 위해 기간 내의 최신 주가에 더 많은 가중치를 부여하고 이전 주가에 적은 가중치를 부여하는 기술적 분석 지표이고 교차는 단기 이동평균이 장기 이동평균을 뚫는 시점을 매때시점으로 잡아 단기 이동평

균이 장기 이동평균을 상향돌파할 경우 매수, 단기 이동평균이 장기 이동평균을 하향돌파 할 경우 매도 하는 방법으로 최근의 추세가 기존의 흐름을 전환시키려고 한다는 것을 반영하는 기술적 분석기법이다. 위에서 설명한 기술적 주가분석 지표들을 제외한 다른 기술적 지표들은 본 연구에서 의미 있는 자료가 아니기 때문에 추가설명을 생략한다.

2.2.2. 기계학습모형

본 연구에서는 기계학습을 통한 한국 코스피 지수 예측 능력을 실험하기 위해서 교차 검증 (cross validation)을 거친 대표적인 3종류의 기계학습 모형을 이용한다.

첫째, 분류회귀나무 알고리즘은 의사결정나무를 형성하는데 있어 지니계수 (Gini coefficient)를 이용하여 불순도를 측정하며, 부모마디로부터 자식마디가 2개만 형성되는 이진분류 기반 알고리즘이다. 지니계수는 n 개의 입력변수 중에서 임의로 2개를 추출하였을 때, 그 2개가 서로 다른 군에 속할 수 있는 확률을 의미한다. 분류회귀나무 알고리즘의 첫 단계는 각 마디에서 도수가 가장 많은 목표변수의 오분류 확률을 계산하며, 다음으로 이진분리를 실시한 하위 마디에서 지니 계수를 계산한 후 그 값이 최소화 되는 값을 최상의 분리기준으로 판단한다. 지니계수의 감소량이 계산되면 분류회귀나무 알고리즘의 마지막 과정으로 지니 계수를 가장 감소시켜 주는 분류변수와 최적 분리를 자식 마디로 선택한다.

본 연구에서는 분류회귀나무 알고리즘을 활용하여 의미 있는 요인을 선택하였고 선택된 유의미한 요인을 이용하여 한국 코스피의 방향성을 예측할 수 있는 의사결정모형을 도출하였다. 둘째, 서포트벡터머신 (support vector machine)을 이용하여 한국 코스피지수를 예측하는데 있어 분류회귀나무모형에서 추출된 유의미한 예측변수들은 서포트벡터머신모형의 입력변수로 사용된다. 셋째, 딥러닝 (deep learning)모형을 이용하여 한국 코스피지수를 예측하는데 있어 분류회귀나무모형에서 추출된 유의미한 예측변수들은 딥러닝모형의 입력변수로 사용된다. 딥러닝 모형에는 쌓아 올린 오토인코더를 사용하였고, 이것은 비지도 사전학습 알고리즘을 사용하여 선형적인 관계 및 비선형 관계를 표현할 수 있고 이를 통해 잡음 제거에 탁월할 뿐 아니라 자료 분포 패턴 등을 추정할 수 있다는 장점이 있다.

위 모형들의 성능을 평가하기 위해 정확도, 오류율, 민감도, 특이도 그리고 정밀도를 가지고 비교한다 (Choi와 Lim, 2013). 목표변수가 상승 또는 하락예측과 같은 이진 분류인 경우, 판별 정확도를 검증하기 위해 정오분류표를 이용하여 성능 비교를 실시한다. 정확도는 실제 목표변수의 클래스 (class)를 제대로 분류한 전체 비율을 나타낸다. 또한 모형을 평가하기 위해 정확도와 함께 가장 많이 사용되는 척도로는 오류율이 있다. 오류율은 정확도의 반대개념으로 (1-정확도) 표현된다. 판단기준으로는 정확도는 높을수록, 오류율은 낮을수록 분류모형의 성능이 좋다고 할 수 있다.

Table 2.1 Cross tables for actual class and classified class

Actual class	Classified class	
	Class1	Class2
Class1	True positive (TP)	False negative (FN)
Class2	False positive (FP)	True negative (TN)

Table 2.2 Performance measures for classification model

Measure	Expression
Accuracy	$(TP+TN)/(TP+FP+TN+FN)$
Sensitivity	$TP/(TP+FN)$
Specificity	$TN/(FP+TN)$
Precision	$TP/(TP+FP)$

3. 실증분석

본 연구에서 사용할 분석기간은 2000년 1월 1일부터 2016년 2월 12일까지 한국 코스피 주가지수에 대한 일별 지수 증가 값이며 이는 블룸버그에서 수집했다.

Table 3.1에서 한국 코스피 주가지수 일별 지수 증가에 대한 평균, 표준편차, 왜도 및 첨도에 대한 기초통계량을 제시하였다.

Table 3.1 Statistic summary

	KOSPI	KOSPI return
mean	1399.368	0.000266
median	1454.6	0.000650
max.	2228.96	0.119457
min.	468.76	-0.120188
s.d.	538.297	0.015956
skewness	-0.242489	-0.382179
kurtosis	-1.494653	5.594851

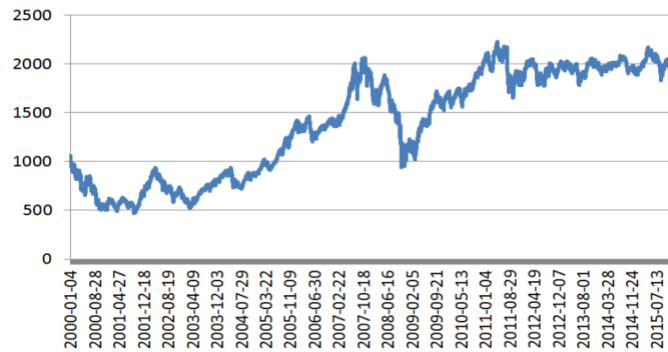


Figure 3.1 Time series of KOSPI

주가 및 수익률의 왜도가 음으로 값으로, 이는 부정적인 극단 현상이 나타날 가능성이 정규분포보다 높음을 의미하고, 주가는 음의 첨도로서 꼬리가 정규 분포보다 얇은 반면, 수익률의 경우는 첨도가 정규 분포보다 크다는 것을 나타낸다.

자료를 두 부분으로 나누는 방법으로 전체자료의 80%를 통하여 기계학습 모델을 훈련시키고 나머지 20%는 학습된 모델에 입력변수를 투입시킴으로써 추출되는 값과 실제 값과의 비교를 통해 검증하여 최종적으로 모델을 선택하였다. 평가의 정확도를 높이기 위해 10겹 교차 검증을 1회 반복하였다. 즉 자료를 훈련 자료와 검증 자료로 나누어 모델링 및 평가하는 작업을 10회 반복하였다. 그리고 교차 검증을 통해 최선의 모델을 가지고 검증자료에 적용해 성능을 평가하였다.

한국 코스피 주가지수 방향에 크게 영향을 미치는 기술적 주가분석 지표가 무엇인지를 알고자 학습 결과 if-then-else 규칙으로 해석이 용이할 뿐만 아니라 예측 모델을 생성할 때 추가적인 정교화를 할 수 있다는 장점을 가지기 때문에 분류회귀나무모형을 이용하여 분석하였다. 더불어 분류유형을 할 위험이 높거나 부적절한 규칙 그리고 과적합을 고려하여 본 연구에서는 의사결정 가지치기를 적용하였다. 이에 대한 결과는 Table 3.2에 제시되었다.

Table 3.2 Accuracy of the training and testing in decision tree

Decision tree		
Accuracy	Total variable	Selected variable
Training	0.761	0.782
Testing	0.761	0.778

분류회귀나무모형에 의한 분석 결과 (Table 3.2), 전체 기술적 분석지표에 대한 학습 및 검증자료의 정확도는 각각 76.12%와 76.05%로 과적합 없이 잘 분류되었고 가지치기에 의해 선정된 기술적 주가분석 지표변수에 대한 학습 및 검증자료의 정확도는 각각 78.17%와 77.82%로 예측성능이 더 향상되었음을 알 수 있다. 즉 전체변수를 사용할 때의 최적모델과 선택변수를 사용하여 만든 최적모델의 정확도를 비교해보면 전체변수 사용 시에는 76%이고 선택변수를 사용할 시에도 77%의 정확도를 보여줌으로써 전체 변수를 사용하지 않고 선정된 변수 (상대강도지수, 모멘텀, 변화율지표, 3일 지수이동평균 그리고 5일 지수이동평균) 만으로도 한국 코스피 주가지수 방향을 예측하는데 큰 차이를 보여주지 않음을 알 수 있다. 이 때 최적의 의사결정나무를 도출하기 위하여 가지 수 변화에 따른 교차검증 에러값이 가장 작은 기준에 따라 선택된 복잡 매개변수 (complex parameter)는 0.001이다.

본 연구에서는 전체 및 선정된 기술적 주가분석 지표를 각각 서포트벡터머신모형과 딥러닝모형에 결합하여 사용하였다.

Table 3.3 Parameter in support vector machine : Total variable

σ	0.0005	0.001	0.005	0.01	0.1
C=20	0.786	0.786	0.782	0.778	0.728
	0.768	0.777	0.773	0.763	0.710
C=40	0.787	0.788	0.784	0.776	0.706
	0.774	0.774	0.776	0.758	0.700
C=60	0.787	0.786	0.781	0.780	0.693
	0.774	0.773	0.769	0.757	0.687
C=80	0.784	0.786	0.784	0.780	0.690
	0.769	0.769	0.766	0.755	0.682

Table 3.4 Parameter in support vector machine : Selected variable

σ	0.0005	0.001	0.005	0.01	0.1
C=20	0.781	0.783	0.785	0.787	0.787
	0.764	0.764	0.768	0.767	0.771
C=40	0.783	0.785	0.786	0.787	0.789
	0.766	0.769	0.764	0.769	0.768
C=60	0.783	0.785	0.784	0.788	0.787
	0.766	0.768	0.764	0.767	0.771
C=80	0.783	0.783	0.785	0.788	0.789
	0.769	0.767	0.766	0.766	0.771

RBF (radial basis function) 커널을 이용한 서포트벡터머신모형결과 (Table 3.3), 시그마가 0.001이고 코스트가 20인 경우 전체 기술적 분석지표에 대한 검증자료의 정확도가 77.67%로 가장 우수한 결과를 나타내었다. 시그마가 0.1인 코스트가 60인 경우 선정된 기술적 분석지표에 대한 검증자료의 정확도는 77.06%로 가장 우수한 결과를 나타내었다. 즉 전체변수를 사용할 때의 최적모델과 선택변수를 사용하여 만든 최적모델의 정확도를 비교해보면 두 모형 약 77%의 정확도를 보여줌으로써 전체 변수를 사용

하지 않고 선정된 변수만으로도 한국 코스피 주가지수 방향을 예측하는데 큰 차이를 보여주지 않음을 알 수 있다.

Table 3.5 Optimal deep learning model

Optimal Network	Total Variables	Selected Variables
Hidden Layer1	32-4-2	5-5-2
Hidden Layer2	32-5-5-2	5-5-4-2
Hidden Layer3	32-5-5-5-2	5-5-5-3-2

본 실험에 사용한 딥러닝모형은 5개의 입력뉴런과 2개의 출력뉴런 그리고 1~3개까지의 은닉층개수의 변화와 1~5개의 은닉뉴런수의 변화를 주었다. 더불어 딥러닝에는 쌓아 올린 오토인코더 학습방법을 이용하였다 (Table 3.5). 오토인코더는 학습하는 과정에서 적은 수의 은닉노드에 핵심 특성에 대한 압축된 표현을 저장한다. 본 실험에서는 오토인코더의 이러한 특성을 이용하여 기술적 분석 지표에 따른 주가 방향성을 학습하고 예측에 반영하도록 하였다.”

Table 3.6 Accuracy of the training and testing in deep learning

Accuracy		Deep learning	
		Total variable	Selected variable
Training	Hidden layer1	0.649	0.771
	Hidden layer2	0.663	0.783
	Hidden layer3	0.726	0.790
Testing	Hidden layer1	0.548	0.748
	Hidden layer2	0.505	0.754
	Hidden layer3	0.750	0.769

Table 3.6은 전체표본을 이용하여 추정된 결과 딥러닝 알고리즘의 경우 은닉층 (hidden layer)과 은닉뉴런 (hidden neuron)수가 많을수록 오류율이 줄어들어 학습효과를 높임을 알 수 있었다.

Table 3.7 Comparison of models

Model	Decision tree		SVM		Deep learning	
	Total	Selected	Total	Selected	Total	Selected
Accuracy	0.761	0.778	0.777	0.771	0.750	0.769
Sensitivity	0.781	0.781	0.756	0.790	0.749	0.792
Specificity	0.740	0.776	0.797	0.752	0.751	0.742
Precision	0.756	0.783	0.785	0.757	0.797	0.797

본 논문이 제안한 딥러닝 (deep learning)모형과 기술적 분석지표를 결합한 한국 코스피주가지수 방향성예측의 성능을 평가하기 위해 C++를 사용 분류회귀나무모형, 서포트벡터모형과의 비교실험을 수행하였고 그 결과 (Table 3.7) 세 모형의 예측력이 비슷했다. 이는 수많은 금융시장 변수들이 직간접적으로 서로 복잡하게 얽혀 불규칙적으로 변화하기 때문에 딥러닝모형을 이용하더라도 기술적 지표 (technical indicators)만으로 미래 주가지수의 움직임을 예측하는 것은 한계가 있다고 사료된다.

4. 결론 및 시사점

현재 미국 내 Wealthfront, Betterment등 200여개 로보어드바이저회사가 이미 존재하고 우리나라도 2016년 3월 구글의 바둑인공지능 알파고와 이세돌 9단과의 바둑대결에서 승리한 이후 금융업계에서는

인공지능과 자문 전문가의 합성어인 로보어드바이저 (Robo-Advisor)에 대한 관심이 높아지고 있다. 현재 일반 투자자를 대상으로 저비용 자산관리 서비스를 제공하는 로보어드바이저에 의해서 운용되는 자금은 전 세계적으로 약 20조원이고 2020년에는 약 2000조원으로 예상하고 있다.

우리나라의 경우 2016년 9월부터 정부 차원의 로보어드바이저 서비스에 대한 검증을 금융위원회 (financial supervisory commission)에서 실시하고 있고 안정적으로 운영된 로보어드바이저 업체들에 대해 투자에 대한 자문을 물론 투자를 대행할 수 있는 권한까지 준다는 계획이다 (Ko, 2016). 이런 정책은 로보어드바이저시장이 활성화될 수 있는 영향력을 미치며 인공지능의 영역을 넓혀가고 있다.

본 연구에서는 기술적 주가분석기법을 의사결정나무모형, 서포트벡터머신모형 (support vector machine) 그리고 딥러닝 (deep learning)모형에 결합한 새로운 기법을 제안하였고, 한국 코스피 주가지수의 방향성 예측에 대해 비교 분석하였다. 분석 결과는 아래와 같다. 첫째, 전체 변수를 사용하지 않고 선정된 변수만으로도 한국 코스피 주가지수 방향을 예측할 수 있다는 결과를 얻었다. 둘째, 제시한 모형들의 비교실험을 수행한 결과, 세 모형의 방향성에 대한 예측력이 비슷하다는 결과를 얻었다. 이는 수많은 금융시장 변수들이 직간접적으로 서로 복잡하게 얽혀 불규칙적으로 변화하기 때문에 딥러닝모형을 이용하더라도 기술적 지표 (technical indicators)만으로 미래 주가지수의 움직임을 예측하는 것은 한계가 있다고 사료된다. 이에 향후 몇 가지 보완할 점이 필요하다. 우선 기술적 분석지표 외에 기본적 주가 분석 또는 금융시장 상황에 영향을 주는 변수들을 추가한 연구가 필요하다. 더불어 본 연구에서 비교대상으로 제시하였던 방법론 외 다양한 딥러닝 (deep learning) 알고리즘과 비교해보고 더 우월한 모형을 제시, 연구해 볼 수 있다.

References

- Chang, W. (2016). The rise of Robo Advisors. *Forbes*, **11**, 214.
- Choi, H. and Lim, D. (2013). Bankruptcy prediction using ensemble SVM model. *Journal of the Korean Data & Information Science Society*, **24**, 1113-1125.
- Choi, H. and Min, Y. (2015). Introduction to deep learning. *Korea Information Processing Society Review*, **22**, 7-21.
- Hinton, G. and Salakhutdinov, R. (2006). Reducing the dimensionality of data with neural networks. *Science*, **313**, 504-507.
- Jung, J. and Min, D. (2013). The study of foreign exchange trading revenue model using decision tree and gradient boosting. *Journal of the Korean Data & Information Science Society*, **24**, 161-170.
- Jung, S. and Park, S. (2016). Examination of possible financial market risk accumulations due to prolongation of low interest rates. *BOK Financial Stability Report*, **27**, 132-143.
- Ko, Y. (2016). A study on the measures to activate the Introduction of the Robo-Advisor in Korea. *Korea Science & Art Forum*, **25**, 19-33.
- Kwak, M. and Rhee, S. (2016). Finding factors on employment by adult life cycle using decision tree model. *Journal of the Korean Data & Information Science Society*, **27**, 1537-1545.
- Lee, W. and Chun, H. (2016). A deep learning analysis of the Chinese Yuan's volatility in the onshore and offshore markets. *Journal of the Korean Data & Information Science Society*, **27**, 327-335.

A deep learning analysis of the KOSPI's directions

Woosik Lee¹

¹Department of Information Statistics, Anyang University

Received 17 January 2017, revised 2 March 2017, accepted 10 March 2017

Abstract

Since Google's AlphaGo defeated a world champion of Go players in 2016, there have been many interests in the deep learning. In the financial sector, a Robo-Advisor using deep learning gains a significant attention, which builds and manages portfolios of financial instruments for investors. In this paper, we have proposed the a deep learning algorithm geared toward identification and forecast of the KOSPI index direction, and we also have compared the accuracy of the prediction. In an application of forecasting the financial market index direction, we have shown that the Robo-Advisor using deep learning has a significant effect on finance industry. The Robo-Advisor collects a massive data such as earnings statements, news reports and regulatory filings, analyzes those and recommends investors how to view market trends and identify the best time to purchase financial assets. On the other hand, the Robo-Advisor allows businesses to learn more about their customers, develop better marketing strategies, increase sales and decrease costs.

Keywords: Artificial intelligence, deep learning, FinTech, Robo-Advisor, technical analysis

¹ Adjunct faculty, Department of Information Statistics, Anyang University, Gyeonggi-do 14028, Korea.
E-mail: woosiklee@hotmail.com