

AI에 기반한 중국 철광석가격 조기경보시스템 구축

- AI 기법의 탐색적 적용 -

남대엽*

<목 차>

- 1. 서론
- 2. 조기경보시스템 선행연구
 - 2.1 신호접근법
 - 2.2 AI 기법
- 3. AI 모델 실증 분석
 - 3.1 SVM
 - 3.2 랜덤포레스트
- 4. 결론

1. 서론

우리는 아직 철기시대에 살고 있다. 현대 인류는 석기 및 청동기 시대를 거쳐 철기시대의 중간을 지나가고 있다. 건물을 지을 때 사용되는 철근에서부터 전기 자동차의 외판까지 현대 문명의 주요한 발명품들은 아직 철강을 주소재로 사용하고 있다.¹⁾ 과거에는 철강이 제조공법의 한계로 생산량이 제한되어 군사무기 등 일부 제품에만 사용되었다. 하지만, 산업혁명 이후 용광로를 통한 철강의 대량 생산이 가능해지면서 철강 소재의 적용 범위가 확대되고 철강은

* 포스코경영연구원 글로벌연구센터 首席研究員

1) 최근 들어 알루미늄, 탄소섬유 등 다양한 신소재가 등장하고 있지만, 채굴 및 제조 원가, 강도, 가공 용이성 등 철강 고유의 경쟁우위와 매장량, 전지구적 고른 분포 등을 고려했을 때 최근 부각되는 신소재가 철강을 완전히 대체할 가능성은 매우 낮음.

국가 제조업 경쟁력의 핵심요인으로 작용해왔다. 철강 소재의 원가 경쟁력과 품질이 국가 제조업 경쟁력의 발전과 매우 밀접하게 연관되어 있기 때문이다. 따라서, 근대 이후 강대국의 역사를 살펴보면 영국, 미국, 일본 등 철강을 기반으로 제조업 경쟁력을 보유한 국가가 세계 최강국으로 발돋움할 수 있었다. 강대국의 역사에 유독 관심이 많았던 중국의 마오쩌둥은 이를 간파하고 1949년 신중국 설립 이후 '대약진운동'과 '문화대혁명'을 거치며 철강산업의 발전을 매우 강력하게 추진한바 있지만 모두 실패하였다.

이 같은 기조는 중국의 개혁개방 이후에도 이어졌다. 1980년대 후반 중국 정부는 덩샤오핑의 강력한 리더십 하에 개혁개방을 추진하며 제조업의 근대화를 중요한 목표로 설정했다. 그리고 이를 뒷받침할 고품질의 철강소재를 생산하기 위해 1978년 바오산강철 등을 설립하며 현대식 철강사의 발전을 적극 지원했다. 비록 토지 및 건물이 국가 소유로 귀속되어 철강의 주요 소비처인 건설산업의 발전이 뒤처져 철강업의 발전도 늦었지만, 일반 시민간의 부동산 거래가 정식 허용되면서 최근 20년간 중국 건설업의 호황과 함께 철강산업도 매우 빠르게 성장했다. 중국이 세계 철강 생산량에서 차지하는 비중은 2000년 15%에서 2017년 50%까지 증가하며 글로벌 철강생산 대국으로 확고하게 자리매김했다.

하지만, 중국 철강산업은 철강의 주요 소재인 철광석의 공급 안정성을 확보하지 못했다는 문제점을 안고 있다. 오히려, 중국 철강산업의 급속한 성장으로 국내 철광석 매장량이 감소하면서 소재의 수입의존도는 지속 증가하고 있다.²⁾ 특히, 2000년 이후 건설된 대형 고로에 적합한 고품질의 철광석 대부분은 호주와 브라질에서 수입하고 있다.³⁾ 이에 따라 중국은 이미 글로벌 철광석 시장에서 최대 수입국가로 부상했다.

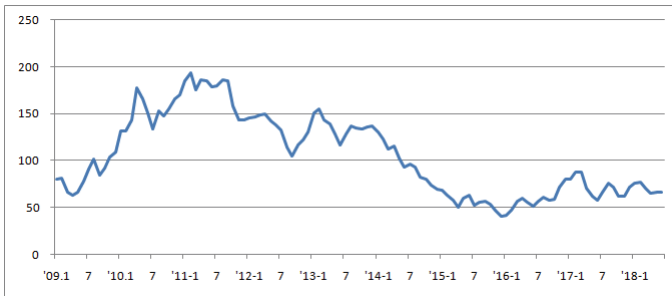
여기에서 문제점은 최근 들어 중국의 철광석 수입가격이 다시 오르고 있다는 점이다. 2011년 톤당 200달러에 근접했던 중국의 철광석 수입가격은

2) 중국의 철광석 수입의존도는 2000년 36.5%에서 지속 증가해 2015년 이후 80% 이상 유지.

3) 2017년을 기준으로 중국의 철광석 수입량 중 호주와 브라질의 비중은 각각 62%와 21%를 차지해 전체의 83% 비중 점유.

2016년 40달러까지 하락한 이후 글로벌 경기 회복 및 유가 등 원자재 가격의 상승에 힘입어 반등하기 시작했다. 철강 제조원가에서 철광석 등 원료가 차지하는 비중은 약 70% 이상으로 철강산업 및 철강사의 수익성에 절대적인 영향을 미친다. 특히, 중국, 한국과 같이 제조업과 수출을 기반으로 성장하면서 철광석의 수입 비중이 높은 국가들은 철강과 같은 기초 소재의 원가 상승이 국가의 제조업 경쟁력에 적지 않은 부정적인 영향을 미칠 수 있다. 기업의 입장에서 원료의 가격 급등에 매우 민감하기 때문에 이를 조기에 예측해 재고 조정 및 금융파생상품(futures, swap) 등을 활용해 미래의 제조원가 상승을 효율적으로 억제하고자 하는 노력이 필요하다.

(그림-1) 중국 주요 항구 철광석 수입가격 추이 (\$/t)



(출처: Mysteel)

따라서 본 논문에서는 중국의 철광석 수입 가격을 중심으로 원자재의 가격 급등가능성을 사전에 경보하는 조기경보시스템(EWS, Early Warning System) 구축에 관하여 탐색해 보았다. 기존 경제학계에서 주로 활용되었던 신호접근법(Signal Approach)과 AI(Artificial Intelligence) 기법의 예측 성능(Accuracy)과 각 기법의 장단점에 관하여 비교 분석하였다. 특히, 본 논문에서는 2016년 알파고의 등장 이후 모든 산업 및 학계의 뜨거운 주제로 부상한 AI 기법에 초점을 맞추어 중국 철광석 수입가격에 대한 조기경보시스템 구축에 관하여 연구를 진행하였다.

다음 2장에서는 신호접근법과 AI에 기반해 조기경보시스템을 구축했던 연구 결과들에 대해 살펴보겠다. 그리고 3장에서는 중국 철광석수입 가격을 대상으로 AI기반 조기경보시스템을 적용하여 실증분석하였다. AI 모델로는 범용적으로 가장 많이 사용되는 SVM(Support Vector Machine), 랜덤포레스트(Random Forest) 2가지 모델을 사용하였다. 마지막으로 4장에서는 신호접근법과 AI 기법의 정확도 및 장단점을 비교하였다.

2. 조기경보시스템 선행 연구

2.1 신호접근법(Signal Approach)

신호접근법에 따른 조기경보시스템에 관한 연구는 1998년 Kaminsky가 일국의 외환위기를 대상으로 방법론을 제시한 이후 넓은 분야로 적용이 확대되었다. 동 방법론의 기본 체계는 외환위기 등 조기경보시스템을 구축하고자 하는 대상(종속변수)를 결정하고 이에 영향을 미칠 것으로 예상되는 다수의 설명변수들을 선정한다. 그리고 설명변수들의 변동폭이 일정 수준의 임계치를 벗어날 때 이를 신호로 인지하여 종합위기지수를 산출하고 동 지수가 특정 임계치를 상회했을 때를 위기 구간으로 설정해 모델의 예측력을 검증하는 방식이다.

이는 기본적으로 특정 시계열 또는 회귀 모델에 기반하지 않고 연구자가 선택한 임의의 다수 설명변수들이 내보내는 특정 신호를 종합해 위기를 예측한다는 점에서 학계에 폭넓게 받아들여졌다. 또한, 이후의 연구자들은 예측 대상을 원유, 부동산, 국제곡물, 비철금속 등으로 확대하고 설명변수들을 폭넓게 확대하며 비교적 높은 수준의 예측력을 보유한 모델을 도출할 수 있었다.⁴⁾

4) 서승환(2004), 김지환(2009), 김태훈(2009), 康藝之(2014), 羅軍(2012) 등.

이 같은 선행 연구를 기반으로 남대엽(2016)은 중국 철광석 수입가격(종속 변수)를 대상으로 신호접근법에 기반한 조기경보시스템을 연구하였다. 동 논문은 종속변수의 변동폭이 커지기 시작한 2002년부터 2014년 2분기까지의 데이터를 사용하여 특정 시점으로부터 1년 후의 위기(가격 급등)을 예측하는 모델을 구축했다. 설명변수로는 종속변수에 영향을 미칠 수 있는 거시경제, 원자재 경기, 철강 경기, 철강 수급 등 4가지 부문에서 13개 지표를 수집해⁵⁾ 위기종합지수를 산출하고 샘플 내 기간에서 조건부확률과 절단확률을 결정하고 이를 샘플 외(테스트) 기간에 적용해 10개의 위기 구간 중 7번의 확률(70%)로 1년 후 위기의 발생 유무를 비교적 정확하게 예측하였다.

(표-1) 중국 철광석 수입가격 조기경보시스템 설명 변수

거시경제	원자재 경기	철강업 경기	철강수급
중국 GDP 증가율 OECD 경기지수	BDI 지수 스크랩가격 원유가격 동가격	세계철강가격 중국철강가격 중국 주요 철강사 주가	중국 조강생산 세계 조강생산 중국 철광석 생산 중국 철광석 수입

(출처: 남대엽(2016))

2.2 AI 기법

AI 기법이란 의사결정트리(Decision Tree), SVM(Support Vector

5) 거시경제로는 세계 최대 철강 생산 및 소비국이자 신흥 경제국을 대표해 중국의 경제성장률을 선택했으며, 이와 함께 세계 주요 경제 국가(34개)의 경기를 종합적으로 알려주는 OECD 선행지수를 선택함. 원자재 경기에 대해서는 대표성이 높은 구리가격 및 원유가격을 데이터로 수집하고 철광석의 대체재로 스크랩 가격과 국제 원자재 경기와 민감한 벌크선 운임지수(BDI)를 수집함. 철강 경기로는 철강전문 컨설팅 기관인 SBB에서 발표하는 세계 철강가격 및 중국 철강가격과 중국 주요 철강기업의 주가를 사용함. 철강 수급으로는 세계 및 중국의 조강 생산량과 중국의 철광석 생산 및 수입량을 사용함. 또한, 중국은 세계 철광석 해운 물동량의 70% 이상을 차지하기 때문에, 중국의 철광석 수입 및 생산량 역시 철광석 수급을 결정하는 매우 중요한 요인으로 선택함.

Machine), Random Forest, Deep Learning 등과 같이 컴퓨터 알고리즘 상의 학습을 통해 모델의 계수(weight)를 조정하며 정확도, 즉 조기경보시스템에서는 예측력을 높이는 기법을 말한다. 일반적으로 AI 모델은 y 값(종속변수)의 유무에 따라 지도학습과 비지도학습으로 구분하며, 지도학습의 경우 y 값의 형태(continuous number, discrete class)에 따라 회귀(regression)와 분류(classification) 문제로 구분된다.

기존 경제학에서 주로 사용했던 시계열, 회귀 모델들도 AI 모델이라고 설명할 수 있다. 모두 과거의 데이터를 기반으로 모델 내 설명변수의 계수를 결정하고 정확도를 측정하기 때문이다. 차이점은 기존의 모델이 OLS(Ordinary Least Square) 또는 MLE(Maximum Likelihood Estimation) 등을 활용해 계수 값을 결정하는 반면, AI 모델에서는 주로 SGD(Stochastic Gradient Descent) 등과 같은 기법을 통해 계수값을 학습시켜 나간다.

또한, 두 기법은 목적에서도 큰 차이점을 보인다. 기존의 경제학적 모델은 사회적 현상을 계량적 모델로 설명하기 위해 모델 전체의 정확도(R-square) 보다는 각 설명변수의 유의수준(p-value)과 해석에 초점을 맞추었다. 하지만, AI 기법은 모델 내 각 설명변수에 대한 해석 보다는 모델 자체의 설명력(정확도)를 제고하는데 중점을 두고 있다. 이를 위해 학습기간(train_set)에서 결정된 모델이 시험기간(test_set)에서도 높은 정확도를 유지하기 위해 모델의 과적합(over-fitting)을 억제하는 목적의 일반화(regularization) 등의 기법을 적용하기도 한다.

이와 같은 이유로 조기경보시스템 구축에 관하여 AI 기법을 활용하는 연구가 국내외적으로 점차 증가하고 있다. 우선 국내적으로 윤원중(2009)은 코스피 지수의 1일 변동성을 데이터로 SVM 모델을 활용해 한국의 금융위기 구간을 검증하였다. Ahn(2011)은 조기경보시스템을 분류(classification) 문제로 전환해서 금융위기를 예측하는 방법론을 제시하였다. 한국농촌경제연구원은 2014년 신호접근법, 순위프로빗모형, 인공신경망 등을 종합해 국제곡물 가격에 대한 조기경보지수를 개발하였다.⁶⁾ 허준영(2015), 서봉군(2017) 등도 주

가를 예측하거나 중소기업의 기술 유출에 대한 조기경보시스템을 구축하는데 SVM과 같은 AI 기법을 적용하였다.

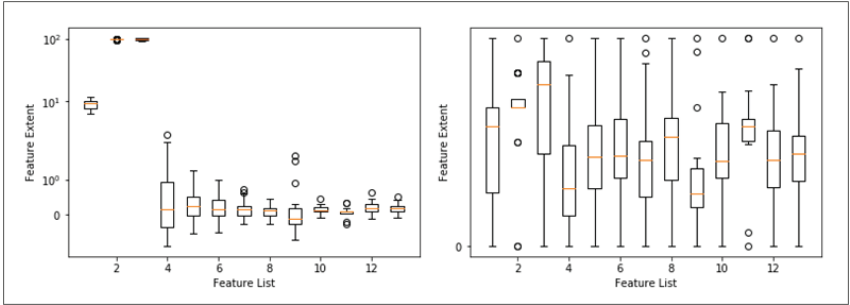
이 같은 추세는 중국에서 더욱 활발하게 연구되고 있다. Lin(2010)은 SVM을 기반으로 곡물가격에 관한 조기경보시스템을 구축했다. 곡물가격에 영향을 미치는 영역을 크게 원가, 공급, 수요, 시장, 통화, 정책, 국제정세 등으로 구분하고 각 영역에서 대표적인 변수들을 수집해 곡물가격의 급등을 예측하였다. Li(2014)는 AI 기법을 활용해 금융위기에 관한 조기경보시스템 구축을 위한 적절한 변수선택(feature selection)에 대한 방법론을 제시했다. 회귀분석의 한 종류인 Lasso-regression을 활용해 금융위기에 영향력이 높은 변수들을 선택할 수 있었다. 한편, Zhang(2015)는 보다 정확한 원유 가격 예측을 위해 전통적인 GARCH 모델과 AI 기법을 조합해서 적용하였다.

본 논문에서는 기존 AI를 활용한 조기경보시스템 관련 연구들에서 많이 사용되었던 SVM 모델 외에 랜덤포레스트 모델을 사용해 각 모델간 정확도를 비교한 후 신호접근법과의 차이 및 특징을 살펴보겠다. 데이터 기간은 조기경보시스템의 목적에 맞는 가격 급등 기간이 포함되고 기존의 연구결과와 예측력을 비교하기 위해 기존 남대엽(2016) 연구와 같은 2002년부터 2014년 2분기 구간으로 설정하였다. 한편, 설명변수도 기존 연구와의 비교를 위해 동일하게 선택하였다. 하지만 AI 기법의 필요에 따라 SVM 모델에서는 13개 변수에 MinMax 표준화 과정을 거쳤다.⁷⁾

6) 인공지능망이란 AI 기법 중 Deep-learning에 해당하는 모델로 2016년 알파고 등장 이후 AI 기법 중 매우 각광 받고 있는 모델임.

7) MinMax 표준화는 데이터 중 가장 큰 값과 작은 값을 '1', '0'로 설정하고 그 외 데이터는 비례적으로 0~1 사이의 값으로 표준화하는 기법임.

(그림-2) 데이터 MinMax 표준화 과정 전후 scale 변화⁸⁾



(출처: 저자 정리)

3. AI 기법 실증 분석

3.1 SVM

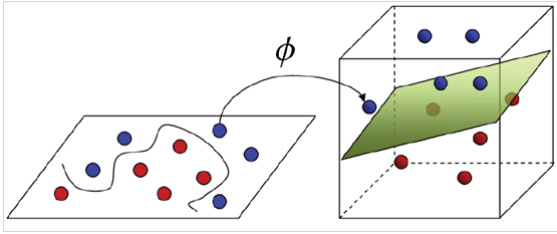
SVM은 Vapnik(1998)이 제시한 이후 지도학습 Machine Learning 분야에서 가장 폭넓게 적용되고 있는 방법론이다. SVM이 적용된 분류 문제의 기본적인 원리는 종속변수를 설명하는 다수의 변수들을 Feature Map에 투영한 후, 각 Feature 사이의 간격을 가장 넓게 만들어 주는 최적의 초평면(hyperplane)을 찾는 방식이다.⁹⁾ 여기에서 중요한 점은 2차원의 평면에서는 선형으로 구분하기 어려운 조합을 kernel 기법을 사용해 초평면으로 전환시킴으로써 선형 분리가 가능하게 만들어 준다는 점이다. SVM은 최근 각광 받고 있는 Deep-learning에 비해 적은 수의 데이터만으로도 높은 정확도를 구현할 수 있다는 장점이 있다.

www.kci.go.kr

8) 데이터 번호는 (표-1)의 왼쪽 위에서부터 순서대로 입력.

9) 윤원중(2009), Zhang(2011) 참고.

(그림-3) SVM의 기본 원리 이해도



(출처: <https://ratsgo.github.io/machine%20learning/2017/05/30/SVM3/>)

이진분류(binary classification)에서 SVM 모델은 다차원의 공간에서 데이터의 두 클래스가 양쪽으로 구분되는 가장 가까운 데이터 사이의 거리를 최대화 시키는 초평면을 찾는 문제이다. 즉, (그림-3)에서 파란색과 빨간색 데이터를 구분 지으면서 그 사이의 거리가 최대화되는 녹색 초평면을 찾아야 한다. 이를 위해 우선 N개의 점으로 이루어진 주어진 학습데이터의 집합 D를 수식 (1)과 같이 정의할 수 있다.

$$D = \{(x_i, y_i) | x_i \in R^p, y_i \in \{-1, 1\}\}^n, i = 1 \text{ ————— (1)}$$

x_i 는 p차원의 실수 벡터이고, y_i 는 x_i 가 속해있는 클래스의 값으로 -1 또는 1을 가진다. 위의 학습용 데이터 집합이 y_i 값에 따라 선형적으로 분리될 수 있을 때, 데이터 집합을 분리하는 것을 초평면이라고 하며 이는 수식(2)로 표현된다.

$$y = w_0 + w_1x_1 + w_2x_2 \text{ ————— (2)}$$

w_i 는 머신러닝 기반의 학습 알고리즘을 통해 업데이트해야 할 계수(parameter)가 된다. 이 때 두 클래스와의 거리를 최대화시키는 초평면(y)을

지지벡터(support vector)라고 부르며 수식(3)과 같다.

$$y = \sum \alpha_i y_i x_i \cdot x + b \quad \text{—————} \quad (3)$$

지지벡터를 찾기 위해서는 두 클래스 간의 거리(margin) 최대화시키는 w_i 값을 구하는 문제, 즉 수식(4)의 목적식을 최소화 하는 문제로 풀 수 있다.

$$Loss(w) = \min \left(\frac{1}{2} \| \vec{w} \|^2 \right) \quad \text{—————} \quad (4)$$

여기에서 만약 두 클래스가 선형 분리가 되지 않을 경우에는 수식(5)와 같은 커널기법($K()$)을 사용하며 본 논문에서 사용한 Gaussian RBF 커널은 수식(6)과 같다.¹⁰⁾

$$f(x) = \sum \alpha_i y_i K(x_i, x_y) + b \quad \text{—————} \quad (5)$$

$$K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2) \quad \text{—————} \quad (6)$$

비록 SVM이 다른 AI 모델들에 비해 매개변수(hyper-parameter)가 적다고 하지만, RBF kernel을 활용한 SVM은 'C'와 'gamma' 두 개의 매개변수를 최적으로 조정해 주어야 한다. 'C'는 일정 수의 데이터가 다른 클래스로 구분되는 것을 허용하는 숫자를 결정하는 매개변수이다. 예를 들어, 'C'를 3으로 설정하면 최대 3개의 샘플이 다른 클래스에 속하는 것을 허용한다. 즉, 동 매개변수를 너무 낮게 설정하면 특이치(outlier)들에 의해 과소적합 될 가능성이 높아지며, 그 반대의 경우에는 과대적합 될 가능성이 높아지기 때문에 학습기간에서의 예측 정확도(accuracy)를 근거로 최적의 'C'값을 선정해야 한다.

10) 박선(2011), Ahn(2011), Lin(2010), wikipedia 참조.

gamma는 비선형 SVM 모델에서 가장 많이 사용되는 RBF kernel을 사용할 때 설정이 필요한 매개변수이다. 이는 Gaussian 분포의 표준편차와 관련되어 있으며, gamma가 클수록 한 데이터의 영향력 범위가 넓어지는 반면, 작을수록 좁아진다.

(표-2)는 이상과 같은 매개변수 조정에 따른 학습기간 및 시험기간의 정확도 결과이다. 그 결과 C값이 4이고 gamma값이 1일 때 시험기간에서 89.5%의 비교적 높은 수준의 정확도를 확인할 수 있었다.¹¹⁾ 한편, gamma가 10 이상일 때는 모든 학습기간에서 정확도가 100%가 되지만 시험기간에서는 오히려 낮아지는 것을 확인할 수 있다.

(표-2) 매개변수 조정에 따른 SVM 모델 정확도

구분		Data	Gamma					
			0.01	0.1	1	10	100	10000
C	1	train	0.714	0.714	0.786	1	1	1
		test	0.789	0.789	0.789	0.789	0.789	0.789
	2	train	0.714	0.714	0.964	1	1	1
		test	0.789	0.789	0.789	0.789	0.789	0.789
	3	train	0.714	0.714	0.964	1	1	1
		test	0.789	0.789	0.842	0.789	0.789	0.789
	4	train	0.714	0.714	1	1	1	1
		test	0.789	0.789	0.895	0.789	0.789	0.789
	5	train	0.714	0.750	1	1	1	1
		test	0.789	0.789	0.895	0.789	0.789	0.789
	6	train	0.714	0.786	1	1	1	1
		test	0.789	0.789	0.895	0.789	0.789	0.789

(출처: 저자 정리)

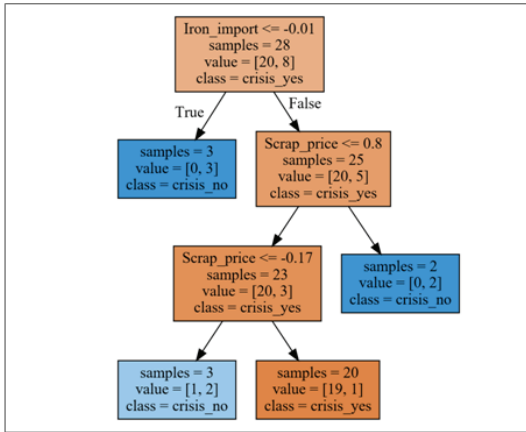
www.kci.go.kr

11) 일반적으로 다수의 C 값에서 같은 정확도가 나왔을 경우에는 가장 작은 값을 최적 파라미터로 결정.

3.2 랜덤포레스트

랜덤포레스트는 의사결정트리에서 발전된 모델이다. 의사결정트리는 (그림-4)와 같이 위기 발생의 유무를 체계적으로 구분할 수 있는 최적의 변수와 임계치를 기준으로 나누어 나가는 방식이다. 예시에서는 우선 철광석 수입증가율 -0.01%을 기준으로 그룹을 구분하고 다시 스크랩가격 증가율 0.8%를 기준으로 다시 가지를 생성해 나간다. 하지만 이처럼 모든 데이터를 계속 나누어 나갈 경우 모델의 최적 변수와 임계치가 학습기간에 과적합 될 가능성을 안고 있다.

(그림-4) 의사결정트리 예시



(출처: 저자 정리)

이와 같은 문제를 해결하기 위해 만들어진 것이 랜덤포레스트 모델이다. 동 모델은 주어진 학습데이터 사이에 중복을 허용하고 변수 선택에 임의성 (random)을 더하여 다수의 의사결정 트리 모델을 생성한 후 각 모델의 예측을 종합하는 앙상블(ensemble) 모델이다. 랜덤포레스트 모델은 다수의 모델에서 나온 결과를 종합하기 때문에 단일한 의사결정 트리에 비해 높은 정확성

과 안정성을 확보할 수 있다.¹²⁾

랜덤포레스트 모델에서 과적합을 예방하기 위해 설정하는 주요 매개변수로는 최대 트리 깊이, 노드 당 최소 샘플 개수, 생성할 트리의 수 등이 있다. 본 논문에서는 데이터의 양과 복잡도가 크지 않기 때문에 생성할 총 트리의 개수 ($N_{estimators}$)를 10~60개로 조정하고 최대 트리 깊이(Max_depth)는 1~6개로 변경해가며 모델의 정확도를 측정하였다.¹³⁾ 예상과 같이 최대 트리 깊이가 커질수록 학습데이터의 정확도가 높아지다가 3 이상일 때부터 100%의 정확도가 나타나기 시작했다. 하지만, 이는 앞에서 지속 언급한 과적합 문제일 가능성이 높다. 한편, 시험기간에서는 최대 트리 깊이가 2~5, 생성할 총 트리의 개수가 40~60개일 때 84.2%의 가장 높은 정확도를 나타낸 것을 확인할 수 있었다.

(표-3) 매개변수 조정에 따른 랜덤포레스트 모델 정확도

구분		Data	N_estimators					
			10	20	30	40	50	60
Max Depth	1	train	0.786	0.893	0.964	0.964	0.964	0.964
		test	0.737	0.737	0.737	0.737	0.789	0.789
	2	train	0.786	0.929	0.964	1	1	1
		test	0.789	0.789	0.789	0.789	0.842	0.842
	3	train	0.750	0.964	0.929	1	1	1
		test	0.789	0.789	0.789	0.789	0.842	0.842
	4	train	0.750	0.964	0.964	1	1	1
		test	0.789	0.789	0.789	0.842	0.842	0.842
	5	train	0.750	0.964	1	1	1	1
		test	0.789	0.789	0.789	0.842	0.789	0.789
	6	train	0.750	0.964	1	1	1	1
		test	0.789	0.789	0.789	0.789	0.789	0.789

(출처: 저자 정리)

12) 김성진(2016).

13) 이 외에도 분석에 사용할 최대 변수 갯수(max_features), 노드 당 최소 샘플 개수(min_samples_split) 등도 변경해가며 시험해 보았지만 유의미한 결과를 얻지 못함.

4. 결론

본 논문은 중국 철강산업의 안정적 발전에 중요한 영향을 미치는 철광석 수입가격에 관한 조기경보시스템에 대하여 살펴보았다. 우선 기존 연구결과들을 토대로 경제학에서 전통적으로 사용하던 신호접근법과 AI 기법의 특성과 장단점 그리고 예측 정확력에 대하여 분석하였다.

우선 신호접근법은 종속변수에 영향을 미칠 것으로 예상되는 다수의 모든 변수를 선택 및 적용할 수 있으며, 설명변수와 종속변수 사이의 영향 관계를 파악할 수 있다는 장점이 있다. 반면, 단점으로는 AI 기법 대비 낮은 정확도를 나타냈다. 아래 표와 같이 신호접근법과 AI 기법의 정확도를 비교한 결과 AI 기법에서 보다 높은 정확도를 나타낸 것을 확인할 수 있다. 특히, 시험기간에서 신호접근법은 70%의 정확도에 그친 반면 3개의 AI 모델은 모두 80% 이상의 정확도를 확보했다. 이는 논문의 서두에서 밝혔듯이 AI 기법은 모델 자체의 정확도를 높이기 위한 방법론이 주를 이루고 있기 때문이다. 한편, AI 모델 중에서도 데이터의 특성에 따라 정확도가 다를 수 있으며, 본 논문의 경우에는 SVM 모델이 89.5%의 가장 높은 정확도를 나타냈다.¹⁴⁾

반면, AI 기법은 설명변수와 종속변수 사이의 영향력을 설명하기 어렵다는 단점이 있다. 본 논문에서 적용한 SVM과 랜덤포레스트는 각 변수들의 계수 값과 p-value 등을 알려주지 않는다.

14) AI 기법이 보다 높은 정확도를 얻은 원인은 AI 모델이 신호접근법 대비 수학적으로 정교하면서도 매우 다양한 매개변수들은 시험 반복적으로 사용했기 때문으로 추정되지만, AI 기법의 특성상 구체적인 이유를 설명하기는 어려움.

(표-4) AI 기법과 신호접근법 정확도 비교

모델	AI 기법			신호접근법
	SVM	Random Forest	Gradient Boosting ¹⁵⁾	Signal Approach
학습기간	1	1	1	0.923
시험기간	0.895	0.842	0.842	0.7

(출처: 저자 정리)

본 논문은 보다 높은 정확도의 조기경보체계를 구축하기 위한 방법론을 비교 분석하였다. 특히, 최근 본래의 컴퓨터 공학 분야를 넘어서 다수 학계의 주요 방법론으로 부상하고 있는 AI 기법과 기존 경제학계에서 조기경보체계 구축에 보편적으로 사용하였던 신호접근법의 정확도 및 각 기법의 장단점을 살펴보았다. 그 결과 모델의 목적에 따른 방법론의 채택이 중요하며, 조기경보체계와 같이 정확도가 중요한 연구에서는 AI 기법이 연구의 목적에 부합하다는 사실을 알 수 있다. 따라서, 기업들은 보다 높은 예측력의 AI 기법을 활용하여 철광석 가격의 급등을 사전에 파악하고 재고를 비축하거나 파생금융상품 등을 활용해 헷징함으로써 원료가격 급등이 경영실적에 미치는 리스크를 줄일 수 있다.

한편, AI 기법의 대다수 모델은 선택한 변수의 설명력 또는 영향력 (p-value)을 알 수 없다는 단점에 따라 본 논문에서 사용한 변수 중 어느 변수가 가장 중요한 영향력을 행사했는지 알 수 없다. 따라서 이후 연구에서 기존의 변수 이외에 다른 중요한 변수들을 추가하거나 데이터 기간 연장, 최신의 AI 기법을 반복 시험적으로 선택하면서 모델의 정확도를 높일 수 있을 것으로 예상된다. 또한, 동 논문에서는 가격의 급등만을 예측하는 모델을 구축했지만, 향후에는 가격이 급락하거나 또는 가격 상승률을 보다 세부적으로 구분하여 예측하는 모델도 연구될 수 있다.

15) Gradient Boosting 기법은 랜덤포레스트와 마찬가지로 의사결정트리를 기본 모델로 사용하지만 샘플을 조합하는 방식이 아닌 이전 트리의 오차를 보완하는 방식으로 모델의 정확도를 높임. 자세한 설명은 scikit-learn 참조. (<http://scikit-learn.org/0.15/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html>)

〈參考文獻〉

- 서승환, <주택시장의 조기경보체계에 관한 연구>, 《지역연구》 20(3): 17-36, 2004.
- 김지환 · 이상호, <신호접근법을 이용한 비철금속 상품가격변동 예측모형 연구>, 《자원환경 지질》 42(2): 143-152, 2009.
- 김태훈 · 승준호, <신호접근법을 이용한 국제곡물가격 조기경보시스템>, 《농촌경제》 32(3): 71-84, 2009.
- 윤원중 · 강용진 · 선철용 · 박규식, <금융위기 조기경보시스템에 한 연구>, Proceedings of KIIS Fall Conference, Vol. 19, No. 2, 2009.
- 박선 · 김경준 · 이진석 · 이성로, <신경망과 SVM을 이용한 적조 발생 예측>, 《전자공학회 논문지》 48(5): 39-45, 2011.
- 김종진 · 김지연 · 공민지 · 최선우 · 이동주 · 채주호, <국제곡물 조기경보시스템 구축>, 한국농촌경제연구원, 2014.
- 허준용 · 양진용, <SVM 기반의 재무 정보를 이용한 주가 예측>, 《정보과학회 컴퓨터의 실제》 21(3): 167-172, 2015.
- 남대엽, <신호접근법에 따른 철광석가격 조기경보시스템 연구>, 《중국학연구》 76: 131-149, 2016.
- 김성진, 안현철, <기업신용등급 예측을 위한 랜덤 포레스트의 응용>, 《산업혁신연구》 32(1): 187-211, 2016.
- 서봉균 · 박도형, <중소기업 기술 유출에 대한 조기경보시스템 개발에 대한 연구>, 《지능정보연구》 23(1): 143-159, 2017.
- Kaminsky, Lizondo, Reinhart, "Leading Indicator of Currency Crises", *IMF staff Paper* Vol.45 No.1, 1998.
- LIN Wen, HOU Yuguo, Dai Wenting, HOU Yunxian, "Early-warning Model Of Grain Price Based on Support Vector Machine In China", *International Conference on Future Information Technology and Management Engineering*, 2010.
- Seoung Hwan Suh, Kabsung Kim, JaeSik Jeon, "Housing Market Early Warning System: The Case of Korea", *European Journal of Scientific Research* Vol. 56 No.4 pp.539-547, 2011.
- Jae Joon Ahn, Kyong Joo Oh, Tae Yoon Kim, Dong Ha Kim, "Usefulness of support vector machine to develop an early warning system for financial

- crisis”, *Expert Systems with Applications*, Vol.38, pp.2966-2973, 2011.
- Jingxiang Li, Yichen Qin, Danhui Yi, YangLi and Ye Shen, “Feature Selection for Support Vector Machine in the Study of Financial Early Warning System”, *Quality and Reliability Engineering International*, Vol.30, pp.867-877, 2014.
- Jin-Liang Zhang, Yue-Jun Zhang, Lu Zhang, “A Novel hybrid method for crude oil price forecasting”, *Energy Economics*, Vol 49. pp.649-659, 2015.
- 羅軍·方偉·林偉君·萬忠, <農產品價格波動與情報預警研究綜述>, 《廣東農業科學》第11期, 2012.
- 康藝之·方偉·林偉君, <我國重要農產品價格異常波動情報預警分析>, 《廣東農業科學》第7期, 2014.

< Abstract >

A Study on Early Warning System of China Iron Ore Price Based on AI Models

Nam, Daeyub

AI(Artificial Intelligence) has been attracting much attention since the coming out of Alpha Go in 2016. Recently, Economic studies have been actively conducted using AI-based machine-learning and deep-learning models instead of classical time series or regression models. In this paper, I applied the AI technique to the EWS(Early Warning System) for Chinese iron ore import price, which has an important influence on the sound development of Chinese steel industry. The work presented in this paper aims to examine the comparison analysis of the prediction accuracy between the classical signal approach which are mainly used in the contemporary economics and the AI-based SVM(Support Vector Machine) and Random Forest models. And I described the pros and cons

of each technique. The results show that the AI models have achieved higher accuracy than signal approaches. So, it is reasonable to think that AI models are more suitable than the traditional models for studies which focus on model's accuracy, such as EWS. And these models can help steel companies prevent volatility in financial performance by increasing iron ore inventory or using of financial derivatives.

Key Words: Iron Ore Price, Early Warning System, AI, SVM, Random Forest

원고접수일	심사일정	1차수정	게재확정	출간
2018. 7. 22	2018. 8. 10	2018. 8. 13	2018. 8. 28	2018. 9. 30