

주식 투자자의 의사결정 지원을 위한 데이터마이닝 도구

Data Mining Tool for Stock Investors' Decision Support

김성동

한성대학교 컴퓨터공학과

Sung-Dong Kim(sdkim@hansung.ac.kr)

요약

주식시장에는 많은 투자자들이 참여하고 있으며 점점 더 많은 사람이 주식투자에 관심을 가지고 있다. 주식시장에서 위험을 회피하고 수익을 얻기 위해서는 다양한 정보를 바탕으로 정확한 의사결정을 해야 한다. 즉 수익을 얻을 수 있는 종목 선택, 적절한 매수-매도 가격의 결정, 그리고 적절한 보유기간 등을 결정해야 한다. 본 논문에서는 개인 주식 투자자의 의사결정 지원을 위한 데이터마이닝 도구를 제안한다. 즉, 개인 투자자가 직접 기계학습 방법을 적용하여 주가예측 모델을 생성할 수 있게 하고, 적절한 매수-매도 가격과 보유기간 등을 결정하는 것을 도와주는 도구를 제안한다. 제안하는 도구는 과거 데이터를 이용하여 투자자 자신의 성향에 맞는 투자에서의 의사결정을 할 수 있도록 지원하는 도구로서 주가데이터 관리, 기계학습 적용을 통한 주가예측 모델 생성, 투자 시뮬레이션 등의 기능을 제공한다. 사용자는 스스로 주가에 영향을 미칠 수 있다고 판단하는 기술적 지표를 선정하고 이를 이용하여 주가예측 모델을 만들고 테스트 할 수 있으며, 적절한 예측모델을 적용하여 시뮬레이션을 수행해 봄으로써 실제로 어느 정도 수익을 얻을 수 있는지 평가하고 적절한 매매 정책을 수립할 수 있다. 제안하는 도구를 이용하여 주식 투자자는 기존의 감정적 판단에 의한 투자가 아닌 객관적 데이터에 의해 검증을 거친 주가예측 모델과 매매정책에 따라 주식투자를 할 수 있어 이전 보다 나은 수익을 기대할 수 있다.

■ 중심어 : | 의사결정 지원 | 데이터마이닝 | 주가예측 모델 | 신경망 학습 | 결정트리학습 |

Abstract

There are many investors in the stock market, and more and more people get interested in the stock investment. In order to avoid risks and make profit in the stock investment, we have to determine several aspects using various information. That is, we have to select profitable stocks and determine appropriate buying/selling prices and holding period. This paper proposes a data mining tool for the investors' decision support. The data mining tool makes stock investors apply machine learning techniques and generate stock price prediction model. Also it helps determine buying/selling prices and holding period. It supports individual investor's own decision making using past data. Using the proposed tool, users can manage stock data, generate their own stock price prediction models, and establish trading policy via investment simulation. Users can select technical indicators which they think affect future stock price. Then they can generate stock price prediction models using the indicators and test the models. They also perform investment simulation using proper models to find appropriate trading policy consisting of buying/selling prices and holding period. Using the proposed data mining tool, stock investors can expect more profit with the help of stock price prediction model and trading policy validated on past data, instead of with an emotional decision.

■ keyword : | Decision Support | Data Mining | Stock Price Prediction Model | Neural Networks | Decision Tree |

* 본 연구는 한성대학교 교내연구비 지원과제임.

접수번호 : #111207-003

접수일자 : 2011년 12월 07일

심사완료일 : 2012년 01월 06일

교신저자 : 김성동, e-mail : sdkim@hansung.ac.kr

1. 서론

주식 투자자들의 공통적인 목표는 지속적으로 높은 수익을 얻는 것이다. 주식시장에는 많은 종목이 거래되고 있으며 또한 수많은 정보가 넘쳐나고 있고 투자자들은 많은 정보를 활용하여 투자 대상을 선정한다. 개별 주식 종목에 대한 분석 방법으로는 기술적 분석(technical analysis)과 가치 분석(fundamental analysis) 방법이 있다. 기술적 분석은 주식의 가격 및 거래량에 의해 계산되는 다양한 기술적 지표들을 분석하는 것으로 주로 주식의 차트를 이용한 분석을 의미하며 단기 투자를 위해 적합하다. 이에 비해 가치 분석은 회사의 재무제표를 이용한 분석으로 회사의 현황 및 미래에 대한 전망 등을 분석하는 것으로서 장기투자를 위해 적용될 수 있다. 그런데, 일반 개인 투자자들은 기술적 분석이나 가치 분석을 직접 수행할 만한 데이터를 가지고 있지 않아 대부분 증권회사에서 제공하는 분석 결과나 전망을 토대로 감정적 판단에 의한 투자를 하고 있다. 따라서 대량의 정보를 보유하고 정보의 분석에 의한 투자를 수행하는 외국인이나 기관투자자들에 비해 수익을 얻지 못하는 현실이다. 또한 많은 개인 투자자들은 장기적 투자보다는 단기 투자를 주로 하고 있다. 이에 본 논문에서는 매일매일의 주가 데이터를 이용하여 기술적 지표(technical indicators)를 계산하고 이를 이용한 기술적 분석을 수행할 수 있는 도구를 제안하여 일반 개인 투자자들이 직접 기술적 분석을 수행하고 데이터에 기반한 객관적 판단을 통해 주식 투자를 할 수 있도록 지원하고자 한다. 그런데 주가가 오를 것이라 예상되는 종목을 선정한다 하더라도 적절한 매매정책(trading policy) 없이 주식매매를 한다면 실제로 수익을 얻기가 매우 어렵다. 따라서 적절한 매매정책 수립을 위해 과거의 데이터에 대해 투자 시뮬레이션을 수행할 수 있도록 하는 기능도 제공하고자 하였다.

본 논문에서는 주가 상승에 영향을 준다고 판단되는 다양한 기술적 지표를 계산하고 이를 기계학습 기법(machine learning technique)에 적용하여 주가예측 모델(stock price prediction model)을 생성하고, 실제로 수익을 얻을 수 있었던 매매정책을 수립하도록 함으로

써 주식 투자에서 수익을 얻을 수 있게 지원하는 데이터마이닝 도구를 제안한다. 주식 종목의 가치 분석을 제외하고 오직 기술적 분석을 수행하므로 제안하는 시스템은 단기 주식 투자자의 의사결정을 지원하는데 적합하다. 주식 투자자들은 현재 상황에서 특정한 종목의 가격이 앞으로 상승할 것인지, 상승한다면 얼마만큼의 수익을 기대할 수 있는지 대한 의사결정을 주로 주관적, 감정적으로 수행한다. 그런데, 제안하는 데이터마이닝 도구는 이러한 의사결정을 과거의 데이터를 이용하여 검증하도록 함으로써 투자자들이 보다 객관적이고 신뢰성 있는 의사결정을 할 수 있도록 지원한다.

[1]에서는 대량의 데이터로부터 의사결정을 지원하는 정보를 추출하는 데이터마이닝(data mining)이 많은 주가 데이터를 분석하여 수익을 얻을 수 있는 종목을 추천하는 문제에 적절하게 적용될 수 있다고 하였다. 또한 적절한 매매정책을 수립함에 있어서도 대량의 데이터를 활용하면 보다 신뢰성 있는 정책 수립이 가능하다. 데이터마이닝은 데이터의 분석을 통해 사용자에게 유용한 정보를 제공하는 데 널리 활용되고 있다. 예를 들어 [2]에서는 자동차산업의 고객분류와 이를 통한 타겟 마케팅에 데이터마이닝을 적용하였으며 [3]에서는 콘텐츠 추천 시스템을 구축하는데 데이터마이닝을 적용하였다. 이와 같이 데이터마이닝은 실생활에서 널리 적용되고 있으며 본 논문에서는 주식 투자자에게 주식 투자에서의 의사결정을 지원하는 데이터마이닝 도구를 제안한다. 제안하는 데이터마이닝 도구는 주가 데이터를 효과적으로 관리하여 사용자가 손쉽게 이용할 수 있도록 한다. 제안하는 도구를 이용하여, 투자자는 스스로 직접 기술적 지표를 선정하여 주가예측 모델을 생성하고 검증할 수 있으며, 다양한 매매정책을 투자 시뮬레이션을 통해 검증하여 투자자 성향에 맞는 적절한 매매정책을 수립할 수도 있다.

다양한 기술적 지표 등을 활용하여 주가지수 및 개별 종목 주가의 단기적 움직임을 포착하려는 여러 가지 연구가 진행되어왔으며, 최근에는 인공지능을 이용한 지능형 트레이딩 시스템에 관한 연구가 활발하게 이루어지고 있다. [4]에서는 결정트리 학습을 이용한 단기주가 예측 에이전트 시스템을 개발하였다. 실시간으로 관심

종목의 주가를 수집하여 결정트리를 생성하고 사용자에게 매수/매도 시점을 제시하는 시스템을 제안하였다. 그리고 [5]에서는 support vector machine과 유전자 알고리즘(genetic algorithm)을 이용하는 트레이딩 시스템을 제안하였다. 또한 [6]에서는 다양한 인공지능기법을 결합한 의사결정 지원 모형에 대해 연구하였다. TradeStation[7]이라는 도구는 기술적 분석을 수행하는 소프트웨어로 많은 증권회사 등에서 적용되고 있으며 매매연동까지 포함하고 있는 시스템이다. 이러한 시스템들은 개인이 사용하기에는 비용도 비싸고 사용법이 간단하지 않다. 또한 일반 개인 투자자들은 이러한 도구들이 제시하는 결과만을 보고 의사결정을 할 수 밖에 없다.

제안한 데이터마이닝 도구는 일반 개인 투자자들이 자신만의 주가예측 모델을 스스로 만들어 종목을 선정할 수 있으며, 적절한 매수 가격, 매도 가격, 보유 기간 등을 결정하는 데 있어서도 자신이 직접 검증한 결과를 이용하므로 개인 투자자가 의사결정을 객관적으로 명확하게 할 수 있도록 지원한다. 제안하는 도구는 크게 세 가지 구성요소를 가진다. 첫째, 주가 데이터 관리 모듈이다. 매일매일의 주가 데이터를 이용해 종목별 주가 데이터를 생성하고 이를 데이터베이스화 하여 효과적으로 관리한다. 둘째, 주가예측 모델 생성 모듈이다. 주가 예측에 영향을 줄 수 있다고 판단되는 기술적 지표를 선택하고 기계학습을 위한 데이터 생성, 기계학습 기법 적용 등을 한다. 기계학습 방법으로는 신경망(neural networks)[8-10]과 결정트리(decision trees)[11] 방법을 이용할 수 있도록 하였다. 셋째, 투자 시뮬레이션 모듈이다. 생성한 주가예측 모델에 대해 백테스팅(back-testing)을 수행하여 적절한 매매정책을 수립하는 것을 지원한다.

2장에서는 본 논문에서는 제안하는 데이터마이닝 도구의 구조와 기능을 설명하고 3장에서 도구를 사용하여 주가예측 모델을 생성하고 시뮬레이션을 수행하는 과정을 설명한다. 그리고 4장에서는 제안하는 도구를 사용함으로써 얻을 수 있는 기대효과를 제시하고 5장에서 앞으로의 과제를 제시하며 논문을 마무리한다.

II. 데이터마이닝 도구의 구조와 기능

본 논문에서 제안하는 데이터마이닝 도구는 기술적 분석에 의해 수익을 얻을 수 있는 주식 종목을 추천하고 실제 시장에서 적용할 수 있는 적절한 매매정책 수립 기능을 제공함으로써 주식 투자자의 의사결정을 지원한다. 기술적 분석만을 적용하여 주가예측을 시도하기 때문에 단기 주식 투자자를 위한 의사결정지원에 활용될 수 있다. 이를 위해 증권거래소에서 제공하는 매일매일의 주가 데이터를 관리하여 개별 주식의 기술적 지표를 계산해야 하며, 기계학습을 위한 학습/테스트/검증 데이터 생성과 기계학습 연동, 그리고 투자 시뮬레이션을 수행한다. [그림 1]은 제안하는 데이터마이닝 도구의 논리적 구조를 보여준다.

개별 주식의 기술적 분석을 위해서는 주가 데이터를 이용해 여러 가지 기술적 지표 값을 계산해야 한다. 이를 위해 매일매일의 주가 데이터를 이용하여 개별 주식의 주가 데이터를 생성하고 이를 관리한다. 당일 주가 데이터는 주식시장에서 거래되는 모든 종목의 주가 데이터를 포함한다. 이로부터 각 종목의 데이터를 추출하여 개별 종목을 위한 데이터를 생성하여 종목별로 데이터를 관리하는 주가 데이터베이스를 구축한다. [그림 1]의 “1”에 해당하는 부분이 주가 데이터 관리 모듈에 해당한다. 주가예측 모델을 만들기 위해 주가예측에 필요한 기술적 지표를 선정한다. 주가 데이터베이스로부터 선택한 기술적 지표를 계산하여 주가예측 모델 구축에 필요한 데이터와 투자 시뮬레이션을 위한 데이터를 생성한다. 생성된 데이터에 대해서 기계학습 기법을 적용하여 주가예측 모델을 생성한다. [그림 1]의 “2”에 해당하는 부분이 주가예측 모델 생성을 위한 모듈이다. 그리고 “3”에 해당하는 시뮬레이션 모듈에서는 주가예측 모델을 백테스팅을 수행 하여 다양한 매매정책에 대한 투자 결과를 제시한다. 사용자는 투자 결과를 분석하여 수익을 얻을 수 있는 적절한 매매정책을 수립한다. 즉, 주식 투자자는 스스로 생성한 주가예측 모델과 매매정책을 만들고 이를 이용하여 주식 투자에서의 의사결정을 할 수 있다.

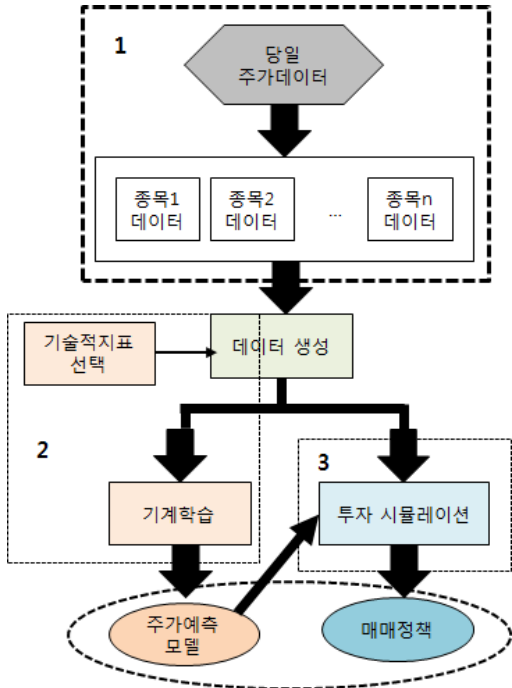


그림 1. 주식 투자자의 의사결정 지원을 위한 데이터마이닝 도구의 논리적 구조

1. 주가 데이터 관리

1.1 주가 데이터베이스

개별 주식의 기술적 분석을 위해서는 주가 데이터를 이용해 여러 가지 기술적 지표 값을 계산해야 한다. 주식 거래일의 주가 데이터는 주식시장에서 거래되는 모든 주식 종목의 주가 데이터를 포함한다. 각 종목의 미래 주가를 예측하기 위해서는 각 개별 종목별 데이터 관리가 필요하다. 제안하는 도구는 한국증권거래소에서 제공하는 주식 거래일의 데이터를 이용하여 각 개별 종목별 데이터를 추출하고 필요한 기술적 지표 값을 계산하여 각 종목별 데이터를 구축할 수 있도록 하였다.

한국증권거래소에서 제공하는 거래일의 주가 데이터는 종목 코드, 종목 이름, 시가, 종가, 고가, 저가, 거래량 등의 정보를 포함한다. 이로부터 생성되는 주가 데이터베이스의 각 종목별 데이터는 [표 1]과 같다.

표 1. 개별 주식 종목을 위한 데이터의 종류

데이터	설명
종목코드, 이름	개별 주식 종목 식별자
시가, 고가, 저가, 종가	당일 주식 가격
거래량	당일 주식 거래량
이동평균	5, 10, 20, 60, 120일간의 종가의 평균
가격 변동률	어제 가격대비 변동률
가격추세	5, 10, 20일 간의 가격상승/하락 여부
이동평균선 추세	이동평균선의 상승/하락 여부, 추세전환 여부, 정/역배열 여부
Typical price	당일의 가격변동 정도를 나타내는 지표
RSI	가격 상승/하락압력 간의 상대적 강도
MFJ	자금 유입/유출을 측정하는 모멘텀지표
투자심리도	시장의 과열/침체 지표

1.2 주가 데이터베이스 구축

일반 개인 사용자가 쉽게 데이터를 관리할 수 있도록 고려하여 [그림 2]와 같이 주가 데이터 관리 화면을 구성하였다. “Stock data from KSE”의 경로에 증권거래소에서 제공하는 거래 당일의 주가 데이터를 포함하는 폴더를 설정한다. 그리고 “Create/Update” 버튼을 클릭하면 폴더내의 모든 데이터 파일을 이용하여 각 개별 종목별 데이터를 생성/갱신할 수 있다. 이미 있는 종목의 경우 당일 주가 데이터를 이용하여 각 기술적 지표 값을 계산하여 종목별 데이터가 갱신되며, 새로운 종목의 경우에는 새로운 종목의 데이터가 새로 추가된다. “종목 코드” 목록에서 하나의 종목을 선택하면 그 종목에 대한 데이터를 원본 화면에서 확인함으로써 올바르게 종목별 데이터가 생성되었는지 확인할 수 있다.

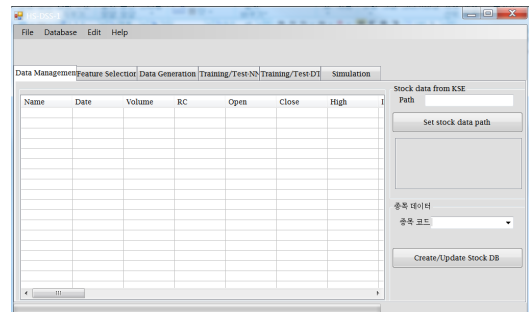


그림 2. 주가 데이터 관리 화면

1.3 주가 데이터베이스 인터페이스

논문에서 제안하는 데이터마이닝 도구는 주가 데이터베이스로부터 필요한 데이터를 생성하는 인터페이스 부분을 [그림 3]과 같이 구성하였다.

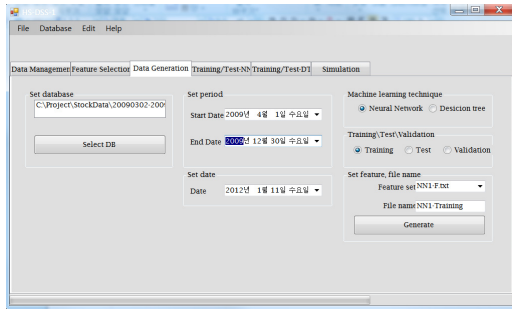


그림 3. 주가 데이터베이스 인터페이스 화면

인터페이스를 이용하여 주가예측 모델 생성을 위한 기계학습에 사용되는 학습(training), 검증(validation), 테스트(test) 데이터를 손쉽게 생성할 수 있다. 즉, 기간을 설정하기만 하면 필요한 기술적 지표값을 계산하여 설정한 기간 동안의 데이터를 생성하며 사용자가 이를 학습, 검증, 테스트용으로 지정하면 된다. 여기서 테스트는 투자 시뮬레이션을 의미하며 과거의 일정한 기간 동안의 주가예측 모델의 예측 결과를 생성하여 시뮬레이션을 위해 사용된다. 또한 실제 투자에 적용하기 위해서 필요한 당일 데이터를 생성할 수 있다. 당일 데이터는 주가예측 모델에서 사용하는 기술적 지표 값을 포함하는데, 이를 주가예측 모델에 적용하면 미래 주가에 대한 예측결과가 생성되는데 결과 값은 실제 투자에 적용하는데 활용된다. 거래량이나 가격 변동이 다른 종목에 비해 매우 비정상적인 종목들은 제외하였는데, 이들 종목의 데이터는 기계학습에서 잡음(noisy)이 될 수 있어 올바른 학습을 방해할 수 있기 때문이다.

2. 주가예측 모델 생성

논문에서 제안하는 데이터마이닝 도구는 기계학습 기법을 적용하여 주가예측 모델을 생성함으로써 투자 의사결정을 지원한다. 여기서는 인공신경망과 결정트리 등의 학습 기법을 적용할 수 있도록 하였다. 사용자

는 미래 주가예측에 필요한 기술적 지표들을 선택하고, 학습, 검증, 테스트를 위한 데이터를 생성하며, 기계학습 기법을 선택하여 주가예측 모델을 생성, 평가함으로써 최적의 주가예측 모델을 생성하는 일련의 과정을 직접 수행할 수 있다. 이를 통해 사용자에 의한 고유의 주가예측 모델 생성을 지원한다.

2.1 기술적 지표 선택

사용자가 미래의 주가에 영향을 미친다고 판단하는 기술적 지표를 선택할 수 있도록 [그림 4]와 같이 기술적 지표 선택 화면을 구성하였다.

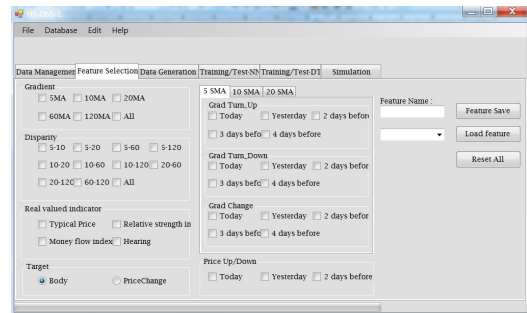


그림 4. 기술적 지표 선택 화면

실제로 어떤 기술적 지표가 미래 주가에 주요한 영향을 미치는지 알 수 없으며, 어떤 지표가 더 영향을 많이 미치는지 또한 알 수 없으므로 사용자는 자신이 중요하다고 판단하는 기술적 지표를 선택하고 기계학습을 통해 각 지표가 미래 주가에 미치는 영향을 고려하는 주가예측 모델을 생성하고 테스트하는 과정을 통해 자신의 판단을 검증할 수 있다. 사용자가 선택한 기술적 지표들을 저장하고 이를 다시 로드하여 다시 활용할 수 있도록 하여 다양한 기술적 지표를 적용하여 주가예측 모델을 생성할 수 있도록 하였다. 주가 데이터베이스 인터페이스 부분에서는 선택된 기술적 지표를 이용하여 적절한 데이터를 생성한다.

현재 제공되는 기술적 지표들은 크게 실수 값(real value) 유형과 이진 값(binary value) 유형으로 구분된다. 실수 값 유형의 지표들로는 이동평균선의 기울기 (gradient), 이동평균선간 이격(disparity), RSI, 투자심

리도 등이 있다. 그리고 이진 값 유형의 지표에는 이동평균선의 상승전환, 하락전환 여부, 가격상승 여부, 이동평균선 정배열/역배열 여부 등이 있다. 가격의 이동평균 값은 일반적으로 5일, 10일, 20일, 60일, 120일 등의 5가지 값을 이용하므로 5가지 이동평균 값 및 각각의 관계에 대해 여러 가지 기술적 지표를 고려할 수 있으므로 이들을 선택할 수 있도록 화면을 구성하였다. 그리고 학습을 통해 예측하고자 하는 결과가 봉차트(candle chart)에서의 봉의 몸체(body = (종가 - 시가)/시가)인지 또는 주가 등락율(rate of price change)인지를 선택할 수 있도록 하였다.

2.2 기계학습 방법 적용

주가예측(stock price prediction)은 일종의 시계열 분석(time series analysis) 문제이며 시계열 분석을 위해서 인공신경망이 널리 활용되고 있다. 인공신경망은 관찰 데이터로부터 함수를 추론하는데 유용하며 특히 실수 값 데이터를 입/출력으로 하는 함수 추정에 널리 활용되고 있다. 기술적 분석을 통한 주가예측을 위해 다양한 기술적 지표가 이용되며 이들은 주로 실수 값을 가지므로 인공신경망을 이용한 주가예측 모델을 생성하는 것이 적절하다고 판단하여, 제안하는 도구에서는 인공신경망 학습을 지원하는 기능을 제공한다.

신경망을 이용하여 예측모델을 생성하기 위해서는 학습, 검증, 테스트 과정을 수행해야 한다. 이를 위해 먼저 [그림 4]를 이용하여 학습을 위한 특성(feature)을 선택하고 [그림 3]의 주가 데이터베이스 인터페이스를 이용하여 필요한 데이터를 생성한다. 신경망을 적용하는 것은 학습과 검증 데이터를 이용하여 신경망의 가중치(weight)를 계산하는 것이다. 이를 위해 학습/검증 데이터 및 가중치 데이터를 저장할 파일을 선택하는 기능을 제공한다. 학습을 통해 생성된 예측모델의 테스트를 위해 테스트 데이터와 신경망의 가중치 파일을 선택할 수 있으며, 예측모델의 예측치 값을 저장할 파일을 선택할 수 있다. 신경망의 구조 및 학습을 위한 인자(입력 개수, 출력 개수, 학습 횟수, 은닉 노드 개수 등)를 설정하고, 저장하고, 나중에 다시 이용할 수 있도록 저장된 인자 값을 로드하는 기능도 제공하여 다양한 신경망 구조에

대해 학습을 수행할 수 있도록 하였다. 신경망에 대해 잘 모르는 사용자도 신경망을 이용할 수 있도록 기본적인 인자는 자동적으로 세트되도록 하였으며, 필요에 따라 인자를 변경하여 학습하는 것도 가능하도록 기능을 제공한다. 학습을 수행할 때 학습 에러로그(error log)를 볼 수 있도록 하여 가장 학습이 잘된 추가예측 모델을 생성할 수 있도록 한다. [그림 5]는 신경망 학습/테스트를 위한 화면의 모습을 보여준다. 왼편이 학습을 위해 필요한 데이터 파일을 세팅하는 부분이고 오른편이 테스트를 위한 부분이다. 그리고 아래 부분은 신경망 인자를 사용자가 세트할 수 있는 부분이다.

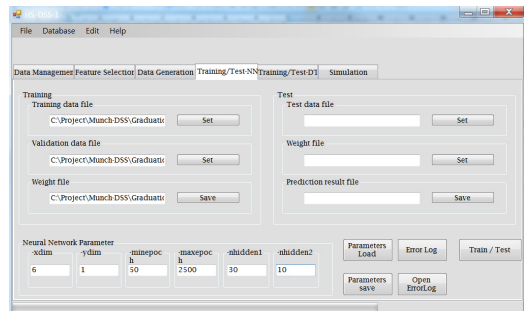


그림 5. 신경망 학습/테스트를 위한 화면

주가예측을 위해 추세를 분석하기도 하는데, 이를 위해 가격의 상승여부, 이동평균선의 변화 여부 등 다양한 이진 값 지표를 고려할 수 있다. 기계학습 방법 중 결정트리 방법은 이진 값으로 구성되는 입력에 대해 예측을 수행하기에 적절한 방법이다. 따라서 제안하는 도구에서는 C4.5 결정트리 학습 기법을 적용하는 기능을 제공한다. 결정트리 학습 방법은 신경망 학습에 비해 학습 결과에 대한 해석이 용이하다. 즉, 학습 결과가 if-then-else 규칙으로 생성될 수 있어 추가예측 모델을 생성할 때 추가적인 정교화(tuning, refinement)를 할 수 있는 장점을 가진다. [그림 6]은 결정트리 학습/테스트를 위한 화면의 모습이다. 왼편이 학습을 위해 필요한 파일을 세팅하는 부분이고 생성된 if-then-else 규칙을 이용하여 테스트를 수행하기 위해 필요한 파일을 세팅하는 부분이 오른편에 나타난다.

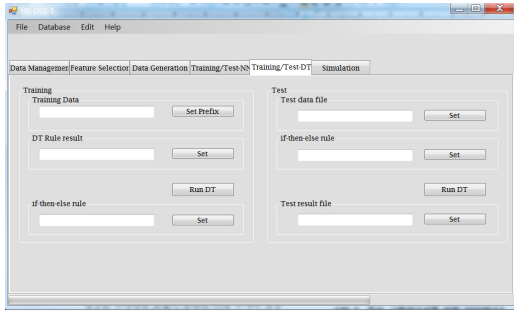


그림 6. 결정트리 학습/테스트를 위한 화면

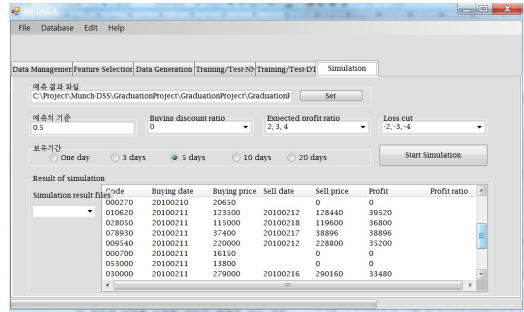


그림 7. 투자 시뮬레이션을 위한 화면

3. 투자 시뮬레이션

생성된 주가예측 모델이 종목의 미래 주가의 방향에 대한 예측을 잘한다 하더라도 적절한 매매정책 없이 주식거래를 한다면 실제로 수익을 얻기가 어렵다. 즉 주가가 오른다고 예측된 종목을 단순히 매수한다고 수익을 얻는 것이 아니고 적절한 가격에 사서 적절한 가격에 팔아야 하며 예측 실패의 경우 그 위험을 최소화해야 궁극적으로 수익을 올릴 수 있다. 이를 위해 매수 할인율(buying discount rate: BDR), 기대 이익 실현률(expected profit ratio: EPR), 손절률(loss cut ratio: LCR), 보유기간(holding period: HP) 등을 적절하게 선정해야 하며 이들 4가지 요소를 매매정책으로 정의한다.

투자 시뮬레이션 기능은 과거 데이터에 대해 매매정책의 4가지 요소를 다양하게 변화시켜 어떤 결과를 얻는지를 파악하고 그 결과를 가지고 사용자가 원하는 매매정책을 결정할 수 있도록 한다. 일반적으로 주식 투자자들은 높은 수익을 지속적으로 얻기를 원하나 실제로 두 가지를 모두 만족시키는 예측모델이나 매매정책을 찾는 것은 매우 어려운 일이다. 따라서 시뮬레이션 결과를 시뮬레이션 전 기간의 수익률 순서로 제공하되, 월별로 수익률을 보여줄 수 있도록 한다. 월별 수익률과 함께 월간 수익률의 표준편차, 거래 횟수, 평균 보유기간 등의 정보를 함께 제공하여 선호하는 매매정책을 선택할 수 있도록 한다. 투자 시뮬레이션 기능은 사용자가 설정한 매매정책이 과거에 어떠한 결과를 보였는가를 과거 데이터를 통해 확인하게 하는 것이며 이를 통해 주식 투자자가 실제로 매매할 때 필요한 의사결정을 지원한다.

```

Invest_Simulation(예측값)
{
    if (예측값 >= 예측치 기준) {
        buying_price = closing_price * BDR;
        if (buying_price >= open_price of next day) {
            bought_price = open_price of next day;
            profit_price = bought_price * EPR;
            loss_price = bought_price * LCR;
            for (day=next_day; day<=next_day+HP; day++) {
                if (profit_price <= day's highest_price) {
                    sold_price = profit_price;
                    sold = TRUE; break;
                } else if (loss_price >= day's lowest_price) {
                    sold_price = loss_price;
                    sold = TRUE; break;
                }
            } // for
            if (sold == FALSE)
                sold_price = day's closing_price;
            profit = sold_price - bought_price;
            profit_ratio = (sold_price - bought_price) / bought_price;
        } // if
    } // if
} // Invest_Simulation
    
```

그림 8. 한 종목에 대한 투자 시뮬레이션 알고리즘

투자 시뮬레이션을 위한 화면은 [그림 7]과 같다. 화면상에서 사용자는 자신만의 매매정책을 검증하기 위해 먼저 매매정책의 각 요소를 설정한다(BDR, EPR, LCR, HP). 그리고 기계학습을 통해 생성한 주가예측 모델이 예측한 각 종목별 예측결과가 저장된 파일을 선택한 후("예측결과 파일"), 매수 기준이 되는 기준 값을 설정하여("예측치 기준") 시뮬레이션을 실행한다. 설정한 각 구간에서 개별적인 매수할인율, 기대수익률, 손절

를, 보유기간 등을 적용하여 시물레이션이 수행되며 시물레이션이 종료되면, 시물레이션 결과 파일들 목록에서 특정 시물레이션 결과, 즉 다양한 매매정책별로 시물레이션 결과를 제시한다. [그림 8]에 예측결과 파일에 포함된 결과 중 하나의 예측값에 대한 시물레이션 수행 알고리즘을 제시하였다. 이 알고리즘을 파일에 포함된 모든 예측결과에 대해 수행하여 결과를 생성한다.

위의 알고리즘은 다음과 같이 수행된다. 예측치가 예측치 기준 이상인 경우, 설정한 매수할인율(BDR)을 이용하여 매수가격(buying_price)를 계산하고 이 가격이 다음날 시가(open_price of next day) 이상이면 시가에 매수한다고 가정한다. 다음에 설정한 기대수익률(EPR)과 손절률(LCR)을 이용하여 이익가격과 손절가격을 계산한다. 그리고 설정한 보유기간(HP)동안 이익가격과 손절가격이 나타난 경우에는 이익 또는 손절 가격에 매도하고 그렇지 않은 경우는 보유기간의 마지막 날의 종가(closing_price)에 매도한다. 마지막으로 매수가와 매도가를 평가하여 이익(profit)과 이익률(profit_ratio)를 계산한다.

III. 실험

본 절에서는 제안한 데이터마이닝 도구를 사용하여 주가예측 모델을 생성하고, 매매정책 개발을 위한 시물레이션 수행 과정을 설명한다. 이를 위해 한국증권거래소에서 제공하는 2008년 1월 2일부터 2010년 3월 17일까지의 주가 데이터를 수집하였다. 여기서는 KOSPI의 68개 종목에 대해서 주가 데이터베이스를 구축하고 이들로부터 데이터를 생성하여 주가예측 모델과 투자 시물레이션을 수행하고 결과를 제시한다.

주가예측 모델 생성을 위해 먼저 [그림 2]의 화면을 통해 주가 데이터베이스를 구축한다. 수집한 주가 데이터가 저장된 폴더를 지정하면 폴더 내의 모든 데이터 파일들에 대해서 종목별 주가 데이터를 생성하여 데이터베이스화 한다. 그리고 주가예측 모델에서 사용하는 기술적 지표를 [그림 4]의 화면에서 선택하고 저장한다. 다음에 [그림 3]의 화면에서 신경망 학습을 위한

학습, 평가 데이터를 생성한다. 이 때 앞 단계에서 저장한 기술적 지표들을 선택하여 이들 값을 계산하여 데이터를 생성한다. 생성된 데이터를 이용하여 [그림 5]의 화면에서 신경망 학습을 수행하여 주가예측 모델을 생성한다. 여기서 생성한 모델을 가지고 시물레이션을 진행하기 위해 [그림 3]의 화면을 통해 테스트 데이터를 생성한다. [표 2]는 [그림 3]의 화면을 통해 생성한 데이터에 대한 정보를 보여준다.

효과적인 매매정책을 수립하기 위해 [그림 7]에서 매매정책 요소를 변화시켜가며 투자 시물레이션을 수행한다. [표 2]의 시물레이션 데이터에 대해서 주가예측 모델이 예측값을 계산하면 이 값을 기준으로 매수를 결정하고, 매매정책의 요소에 따라 매수, 매도를 수행한다. 여기서는 [표 3]과 같은 매매정책에 대해서 시물레이션을 수행하였다. 각 요소에 대해서 시물레이션을 수행하므로 모두 108 가지의 결과가 생성되었으며 그 중 일부를 [표 4]에 제시한다.

표 2. 주가예측 모델과 매매정책 수립을 위해 생성한 데이터

	기간	데이터 수
학습 데이터	2009.4.1 ~ 2009.12.30	24,548
평가 데이터	2008.4.1 ~ 2008.5.31	5,440
시물레이션 데이터	2010.2.1 ~ 2010.3.16	2,312

표 3. 시물레이션을 수행한 매매정책 요소

요소	값
보유기간 (일), H	1, 3, 5
매수 할인율 (%), B	0
기대 이익률 (%), E	2, 3, 4, 5, 6, 7
손절률 (%), L	-2, -3, -4, -5, -6, -7

표 4. 시물레이션을 결과 예

H, B, E, L	이익횟수	손실횟수	총 수익 (원)
1, 0, 2, -2	94	66	485,080
1, 0, 5, -5	95	64	556,135
3, 0, 2, -2	116	50	1,158,701
3, 0, 5, -5	122	40	2,257,588
5, 0, 2, -2	127	42	1,512,810
5, 0, 5, -5	142	22	3,765,845

1. 신경망학습을 위한 특성이라는 의미로 "NNI-feature.txt"라는 파일로 저장하였다.

[표 4]에 제시한 결과들 중 0.9 이상의 예측치 값은 이상 값(strange value)으로 간주하여 제외하고 예측값이 0.5 이상 0.9 미만인 것을 대상으로 하였다. 시뮬레이션 기간인 2010년 2월 1일부터 2010년 3월 16일까지 KOSPI는 1606.44에서 1648.01로 약 2.6% 상승하였으며 시뮬레이션에서는 모두 169번의 주식 매수가 발생하였고 각각에 대해서 100만원씩 투자한다고 가정하여 수익을 계산하였다. 표의 첫 번째 줄은 보유기간, 매수할 인율, 기대수익률, 손절률이 각각 1일, 0%, 2%, -2%인 경우이며 이익이 난 경우가 94회, 손실이 난 경우가 66회, 총 수익이 485,080원임을 보이고 있다. 매매정책에 따라 이익 횟수, 총 수익에서 큰 차이를 보이고 있음을 표 4를 통해 알 수 있으며, 다양한 시뮬레이션을 통해 시장 수익율을 상회하는 적절한 매매정책을 수립할 수 있음을 알 수 있다.

IV. 기대효과

주식시장 참여자들은 어떤 주식을 언제 어느 가격에 사고 어느 가격에 팔아야 하는지 등 다양한 의사 결정을 수행해야 하며 이러한 의사결정을 위해서 많은 정보를 수집하고 분석하여 시장에서의 행동을 결정한다. 그런데 일반 개인 투자자들은 정보에 의한 객관적 판단이 아니라, 소문이나 불확실한 정보 등에 의지해 주관적, 감정적 판단에 의한 투자를 하고 있어 기관투자자나 외국인에 비해 낮은 수익을 얻고 있는 상황이다. 본 논문에서 제안하는 데이터마이닝 도구는 개인 투자자 스스로 추가데이터를 관리하고 필요한 기술적 지표를 이용하여 추가예측 모델을 만들 수 있도록 한다. 즉, 다양한 분야에서 활용되고 있는 신경망과 결정트리 등의 기계 학습 기법을 적용하여 사용자 자신만의 추가예측 모델을 만들 수 있도록 한다. 추가예측 모델은 추가 상승이 예상되는 주식을 특정 시점에 추천하여 매수 종목을 결정하는 데 도움을 준다. 주식 투자자는 매수 종목 결정 후에, 어떤 가격에 사야 할 것인지, 어느 정도의 수익을 기대할 것인지, 위험은 어느 정도 부담할 것인지 등에 대한 결정이 내려야 한다. 제안하는 데이터마이닝 도구

는 이러한 의사결정의 지원을 위해 사용자 스스로 과거 데이터를 이용하여 투자 시뮬레이션을 수행할 수 있게 한다. 시뮬레이션 결과를 다양한 기준에 의해 제시함으로써 사용자는 적절한 매수가격과 매도가격을 설정하고 어느 정도의 이익을 기대하고 어느 정도의 손실을 감수하면서 어느 정도 기간을 보유해야 하는지에 대한 의사결정을 할 수 있다. 결과적으로 데이터에 바탕한 객관적 판단을 통해 위험을 회피하고 보다 안정적으로 수익을 얻을 수 있는 의사결정을 가능하게 한다.

추가예측 모델을 구축하기 위해서 사용자는 추가에 영향을 줄 것이라 판단하는 기술적 지표를 스스로 선택하여 예측모델을 만들고 투자 시뮬레이션을 통해 예측 모델을 실제로 투자에 활용할 수 있는지 검증할 수도 있다. 제안하는 데이터마이닝 도구는 하나의 환경에서 후보 주식종목 군 선택, 학습을 위한 특성 선택, 기계학습 방법 선택, 기계학습의 실행, 학습결과 분석, 매매정책 결정 등 모든 과정을 수행할 수 있어 사용자는 자신의 투자성향에 부합하는 추가예측 모델과 매매정책을 효율적으로 개발할 수 있다. 따라서 본 논문에서 제안하는 시스템은 주식 투자자들이 주식시장에서 반드시 수행해야 하는 여러 가지 의사결정에 많은 도움을 줄 것이라고 기대된다.

V. 결론

본 논문에서는 추가데이터의 관리, 기술적 분석에 의한 추가 예측을 수행하는 기계학습 방법 기반의 추가예측 모델의 생성, 매매정책 수립을 위한 투자 시뮬레이션 등의 기능을 제공하여 단기 주식 투자자들의 의사결정을 지원하는 데이터마이닝 도구를 제안하였다. 제안하는 도구의 개발을 통해 개별 종목의 추가예측 및 매매정책 검증 등을 위해 주식시장에서 매일매일 생성되는 추가데이터를 효과적으로 관리하는 체계를 구축하였으며, 이를 통해 지속적으로 추가데이터를 확장, 유지할 수 있다. 따라서 항상 현재까지의 데이터에 기반하여 주식 투자자의 적절한 의사결정을 지원할 수 있다. 대량의 데이터로부터 유용한 정보를 추출하는데 널리

활용되고 있는 데이터마이닝 기법을 활용할 수 있도록 인공신경망과 결정트리 등의 기계학습 기법을 손쉽게 적용할 수 있도록 하였다. 투자자는 종목 선정 의사결정에 활용되는 추가예측 모델을 스스로 생성하고 과거 데이터를 이용하여 성능을 검증할 수 있다. 실제 매매에 있어서 필요한 적절한 매수가격, 이익률, 손절률, 보유 기간에 대한 의사결정을 지원하도록 과거 데이터에 대한 투자 시뮬레이션을 수행하고 그 결과를 다양한 기준에 따라 제시함으로써 투자자의 투자성향에 부합하는 매매정책을 수립하는 것을 지원한다. 결과적으로 본 논문에서 제안하는 데이터마이닝 도구는 주식시장 참여자들이 데이터에 기반하여 객관적으로 의사결정을 하도록 돕는다.

그동안 다양한 기계학습 기법들을 추가예측에 적용하는 연구가 많이 진행되어왔다. 주로 신경망과 Support Vector Machine을 이용하여 추가예측을 더 잘하기 위한 방법을 연구해 왔다. 그리고 최근에는 기계학습 방법을 결합한 추가예측 모델을 생성하는 연구도 활발해지고 있다. [12]에서는 추가예측을 위해 필요한 특성을 유전자 학습을 이용해 선택하는 방안을 제안하면서 선택한 특성을 Support Vector Machine을 이용하여 학습하여 개선된 예측 성능을 얻을 수 있음을 보이고 있다. [13]에서는 신경망과 결정트리 학습 방법을 결합하여 한 가지 방법만을 적용하였을 때 보다 개선된 성능을 보임을 제시하고 있다. 또한 [14]에서는 추가예측에 활용되는 다양한 기계학습 기법들을 정리해서 제시하고 있다. 여기서는 Support Vector Machine과 Boosting을 적용한 방법이 가장 좋은 예측 성능을 보인다고 하였다. 따라서, 현재 신경망과 결정트리 학습 방법을 제공하는 있는 제안하는 도구를 확장하여 Support Vector Machine 기법을 적용할 수 있도록 기능을 확장하고, 제공하는 기계학습 기법들을 결합하여 추가예측 모델을 생성할 수 있는 기능을 제공하는 것이 필요하다. 또한 사용자가 기술적 지표를 스스로 정의할 수 있도록 하여 더 다양한 추가예측 모델을 만들 수 있도록 할 필요가 있다. 이는 기계학습을 위한 특성 선택 기능의 확장을 통해 가능하다. 투자 시뮬레이션에서 수익을 계산하는 부분에 대한 개선이 필요하다. 현재는 일정한

금액을 매수하고 이에 대한 투자 결과를 제시하고 있거나 자산 관리(asset management)를 통해 투자금을 적절히 배분하는 문제도 함께 고려해야 실제와 유사한 시뮬레이션 결과를 얻을 수 있을 것이다. 또한 더 많은 요소를 고려하여 매매정책을 수립할 수 있도록 매매정책의 변수를 추가할 필요가 있으며 이를 사용자가 선택하여 정책을 수립하도록 할 필요가 있다. 현재는 추가예측 모델을 매수종목 선정에 이용하는데, 매도시기 선정에도 활용하도록 하면 보다 바람직한 의사결정을 지원할 수 있을 것이다. 즉, 현재는 매매정책에서 설정된 보유기간 또는 수익률, 손절률에 의해 매도시기를 판정하지만 종목별 추가예측을 통해 종목별로 다른 매도시점을 결정하는 것이 수익률 향상에 더 유용할 수 있다는 판단이다. 수익률 향상, 위험 회피라는 주식 투자자의 목적을 달성하기 위한 의사결정을 지원하기 위해, 제안하는 데이터마이닝 도구는 지속적으로 개선되어야 할 것이다.

참 고 문 헌

- [1] K. S. Kannan, P. S. Sekar, M. M. Sathik, and P. Arumugam, "Financial Stock Market Forecast using Data Mining Techniques," Proceedings of the International MultiConference of Engineers and Computer Scientists, Vol.1, 2010.
- [2] 이병엽, 박용훈, 유재수, "자동차산업의 고객분류 및 타겟 마케팅 모델", 한국콘텐츠학회논문지, 제9권, 제4호, pp.313-322, 2009(4).
- [3] 송창우, 김중훈, 정경용, 류중경, 이정현, "시맨틱 웹에서 개인화 프로파일을 이용한 콘텐츠 추천 검색 시스템", 한국콘텐츠학회논문지, 제8권, 제1호, pp.318-327, 2008(1).
- [4] 서동환, "의사결정트리 학습을 이용한 단기주가 예측 에이전트 시스템 개발", 명지대학교 대학원, 산업공학 석사학위논문, 2000.
- [5] 김선웅, 안현철, "Support Vector Machines와 유전자 알고리즘을 이용한 지능형 트레이딩 시스템

개발”, 지능정보연구, 제16권, 제1호, pp.71-92, 2010.

- [6] 안현철, 이형용, “투자의사결정 지원을 위한 유전자 알고리즘 기반의 다중인공지능기법 결합 모형”, e-비즈니스 연구, 제10권, 제1호, pp.267-288, 2009.
- [7] <http://www.tradestation.com/>
- [8] M. C. Chan, C. C. Wong, and C. C. Lam, “Financial Time Series Forecasting by Neural Network Using Conjugate Gradient Learning Algorithm and Multiple Linear Regression Weight Initialization,” *Computing in Economics and Finance*, 2000.
- [9] R. J. Frank, N. davey, and S. P. Hunt, “Time Series Prediction and Neural Networks,” *Journal of Intelligent and Robotics System*, Vol.31, Issue1-3, pp.91-103, 2001.
- [10] G. Zhang, B. E. Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks: The state of the art,” *International Journal of Forecasting*, 1998.
- [11] J. R. Quinlan, *CA5: Programs for machine Learning*, Morgan Kaufmann Publishers, 1993.
- [12] R. Choudhry and K. Garg, “A Hybrid Machine Learning System for Stock Market Forecasting,” *World Academy of Science, Engineering and Technology*, Vol.39, pp.315-318, 2008.
- [13] C. F. Tsai and S. P. Wang, “Stock Price Forecasting by Hybrid Machine Learning Techniques,” *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Vol.1, pp.755-760, 2009.
- [14] Vatsal H. Shah, “Machine Learning Techniques for Stock Prediction,” *Technical Report*, New York University, 2007.

저 자 소 개

김 성 동(Sung-Dong Kim)

정회원



- 1991년 2월 : 서울대학교 컴퓨터 공학과(공학사)
- 1993년 2월 : 서울대학교 컴퓨터 공학과(공학석사)
- 1992년 8월 : 서울대학교 컴퓨터 공학과(공학박사)

▪ 2001년 3월 ~ 현재 : 한성대학교 컴퓨터공학과 교수
 <관심분야> : 데이터마이닝, 기계번역, 자연언어처리