Entrue Journal of Information Technology

July 2008 / Vol. 7, No. 2 (pp.33-43)

한국 주가지수 등락 예측을 위한 유전자 알고리즘 기반 인공지능 예측기법 결합모형

A Combination Model of Multiple Artificial Intelligence Techniques Based on Genetic Algorithms for the Prediction of Korean Stock Price Index(KOSPI)

이형용
Lee, Hyoungyong

Abstract

각종 인공지능 기법들을 활용하여 주식시장의 흐름을 예측하려는 연구가 지금까지 많은 인공지능 및 금융공학의 연구자들에 의해 시도되어 왔으며, 그 결과 다양한 인공지능 기법들이 예측 방법론으로 제시되어 왔다. 이런 가운데 서로 다른 예측모형들이 산출하는 예측결과를 종합하고 보완하는 결합기법에 관련된 연구가 90년대 후반부터 오늘날까지 꾸준하게 발표되고 있다. 본 연구 역시 유전자 알고리즘 기반의 새로운 인공지능 예측모형간 결합기법을 제시하고 있다. 다만, 기존의 연구모형들이 각 개별모형 결과의 상대적 가중치에 초점을 맞추고 있었다면, 본 연구의 제안모형은 등락을 판단하는데 활용되는 임계치까지 유전자 알고리즘을 이용해 동시에 최적화하도록 설계되어 있다는 점에서 차별화된다. 제안모형의 유용성을 검증하기 위해, 본 연구에서는 지난 1998년부터 2007년까지의 KOSPI 지수 등락 예측을 위해 구축된 로지스틱 회귀모형, 인공신경망, SVM모형의 결과들을 제안모형을 이용해 결합하였다. 그 결과, 예측력 향상에 본 연구의 제안모형이 기여할 수 있음을 확인할 수 있었다.

주제어 : 모형간 결합, 임계치 최적화, 유전자 알고리즘, 주식시장 예측

There have been many studies that predict the up/down patterns of the stock markets using various artificial intelligence techniques, such as logistic regression(LR), artificial neural networks(ANN), and SVM(Support Vector Machines). Also, some researchers have tried to combine the prediction results of the individual classifiers in order to improve prediction accuracy since the late of 1990s. In this paper, a novel combination model of several artificial intelligence techniques is suggested. The proposed model combines the prediction results of the individual classifiers with the optimized relative importance weights, which are generated by using genetic algorithms(GA). In addition, it also optimizes the threshold, which becomes a criterion to decide the direction of the stock market price (Up or Down). To validate the usefulness of the propose model, we built four classification models – LR, ANN, and SVM – to predict the up/down directions of KOSPI from 1998 to 2007. And, we tried to combine these individual classifiers using the proposed model. As a result, we found that our model may improve the prediction accuracy.

Keywords : Combination Model, Threshold Optimization, Genetic Algorithms, Stock Market Prediction

이형용 : 한성대학교, 경영학부, 전임강사
* 본 연구는 2008년도 한성대학교 교내 연구비 지원과제임.

1. 서 론

주식시장(Stock Market)의 예측은 자료가 방대하고, 잡음이 포함되어 있는 경우가 많은 특징이 있다. 또한 주식시장은 불확실성이 높고, 눈에 보이지 않는 제 3의 영향요인이 너무나도 많이 존재하고 있다. 예를 들어, 정치적 사건, 일반적인 경제 상황, 거래자(Trader)의 기대심리 등 다양한 요소들이 주식시장에 복잡하게 영향을 미칠 수 있다. 이런 연유로 주식시장을 예측하는 것은 가장 난제(難題)에 속하는 경영학 문제라고 할 수 있다[12].

금융공학 분야에서는 주가의 움직임이 무작위적(Random)인지, 혹은 그렇지 않은지에 대한 많은 논란이 있었다. 하지만, 많은 학술 연구들을 통해, 주가의 움직임은 무작위적이지 않으며, 설명하기는 어렵지만 고도의 비선형적이면서도, 유동적인 방식에 의해 결정되고 있음이 지속적으로 제시되고 있다[29]. 특히 '복잡계 이론(Complex Systems Theory)'을 활용하여, 주가지수 변동과 같은 거시경제적인 시계열 예측을 시도하는 행위자 기반 모형에 대한 연구들이 제시되면서, 이러한 믿음은 더욱 확고해지고 있다[10].

경제학 분야에서 이러한 복잡계 이론 등에 기반하여 현실세계의 금융시장 메커니즘을 이해하려는 연구를 시도해 온 반면, 경영공학 및 금융공학 분야에서는 실제 거래에 활용될 수 있는 주식시장을 예측 모형에 대한 연구를 활발하게 진행시켜왔는데, 이 중 상당수의 연구가 예측 방법론으로 '인공지능 기법'을 활용하고 있다. 그 중에서도 특히 LOGIT(Logistic Regression, 로지스틱 회귀분석), ANN(Artificial Neural Networks, 인공신경망), SVM(Support Vector Machines) 등의 기법들은 주식시장의 등락을 예측하는데 있어서, 가장 대표적으로 적용되어 온 인공지능 기법들이다. 하지만, 이러한 기법들은 지금까지 개별적으로 적용되거나, 혹은 소위 앙상블(Ensemble) 방법을 이용해 단일 기법을 서로 다른 설정을 통해 다양한 모형으로 구축해 결합하는 방식으로만 확장되었을 뿐, 서로 다른 이종(異種)의 인공지능 기법들이 동시에 적용되어, 종합적으로 결합되는 모형에 관한 연구는 거의 소개되지 않았다[23][26][28]. 특히 이종의 기법들을 통해 산출된 예측결과를 종합할 때, 기존 연구들은 대부분 투표(Voting) 혹은 단순 평균(Averaging)만을 통해 결합하는 것이 일반적이었다[20].

본 연구는 LOGIT, ANN, 그리고 SVM 등 3가지 기법들의 예측결과를 결합하여 보다 정확성을 높일 수 있는 주식시장 예측모형을 제안한다. 본 연구의 제안모형은 각 기법별 예측결과에 대한 가중치뿐만 아니라, 등락을 가늠하는 임계치(Threshold, Cut-off Value)까지 동시에 최적화함으로써, 보다 정밀한 예측결과를 생성할 수 있도록 설계하였다. 이러한 서로 다른 설계요소들을 동시에 최적화하기 위한 도구로서, 본 연구는 유전자 알고리즘(Genetic Algorithms, GA)을 제안하고 있다.

본 논문의 뒷부분은 다음과 같이 구성된다. 우선 2장에서는 기존 문헌 연구를 간단히 살펴보고, 3장에서는 본 연구의 제안모형인 유전자 알고리즘 기반의 새로운 이종 인공지능 기법간 결합 모형을 소개한다. 4장에서는 앞서 제시한 모형의 유용성을 검증하기 위한 실험 데이터 및 설계 내용을 설명하고, 5장에서는 실험 결과를 종합적으로 정리해 제시하도록 한다. 끝으로 마지막 장에서는 결언과 함께 본 연구와 관련한 향후 연구방향이 제시된다.

2. 이론적 배경

본 연구에서는 각 개별 통계 및 인공지능 모형들을 산출결과를 도출하고, 이를 결합하는 모형을 제시하려고 한다. 이에, 기존 문헌을 검토하게 될 본 절에서는 우선 본 연구에서 적용하는 기계학습 기법인 LOGIT, ANN, SVM에 대해 각각 간략히 살펴보고, 이어 경영분야의 예측과 관련하여 모형간 결합을 시도한 기존 연구들에 대해 간략하게 살펴보도록 한다.

2.1 로지스틱 회귀분석(LOGIT)

분류를 하는데 있어서, 가장 흔한 경우는 이분법(二分法)을 기준으로 분류하는 경우다. 예를 들어, 인터넷 쇼핑몰의 경우에도, 특정 고객이 물건을 살 것인가, 말 것인가 혹은 특정 이벤트에 고객이 반응을 할 것인가, 말 것인가 하는 문제들과 같이 이분법을 필요로 하는 분류 문제들의 해결이 종종 요구되는 것이다. 그런데, 이처럼 이항 확률을 가진 종속변수를 통계적으로 설명하고자 할 때, 일반적인 회귀 분석은 적용하기에 어려움이 있다. 왜냐하면 일반적인 회귀분석에서는 종속변수를 연속변수의 형태를 가정하는데 반해, 이러한 분류의 문제에서는 종속변수가 이항변수이므로 오차의 분포가 정규분포를 한다는 회귀분석의 일반적인 가정이 위배되기 때문이다. 또한 일반적인 회귀분석을 사용할 경우에는 예상값이 0과 1 사이에만 국한되지 않는다는 문제점 역시 발생할 수 있다. 이러한 문제들을 해결하기 위해 사용되는 기법이 바로 로지스틱 회귀분석(Logistic Regression)기법, 약자로 로짓(LOGIT)분석이다.