

패턴 매칭과 자동 규칙 생성에 기반한 2단계 주식 트레이딩 시스템

이 종 우[†] · 김 유 섭^{††} · 김 성 동^{†††} · 이 재 원^{††††} · 채 진 석^{†††††}

요 약

일반적인 동적 매매 환경에서의 금융 예측 시스템은 주어진 목적을 최적으로 만족시키는 매매 형태를 찾고자 한다. 본 논문은 수익률을 극대화시키기 위하여 추출과 여과라는 두개의 단계로 구성된 새로운 형태의 주식 매매 시스템을 제안한다. 주식 추출 단계에서는 특정 시계열 패턴에 부합하는 주식을 추출하는데, 이러한 시계열 패턴은 기술 지표 값들의 조합으로 표현된다. 그리고 여과 단계에서는 추출된 주식 집합에 여과 규칙들을 적용하여 실제 매매 대상이 되는 주식들을 골라내는데, 여과 규칙은 과거 주가 데이터로부터 자동으로 유도되었다. 이를 위하여, 우리는 먼저 방대한 과거 일별 주가 데이터로부터 기술 지표 값들을 계산하였다. 계산된 기술 지표 값들은 시계열 패턴을 추출하는데 사용되고 이 값들의 이산화 구간들의 분포가 양성 및 음성 데이터들에 대하여 계산하였다. 본 논문에서는 독특한 분포를 보이는 구간에 존재하는 기술 지표 값들이 주가의 향후 움직임을 예측하는 데 도움을 준다는 가정을 하였다. 그리고 여과 규칙은 바로 이런 독특한 분포를 보이는 구간 내의 데이터 값들로부터 자동으로 유도되었다. 우리는 시뮬레이션을 통해, 본 논문에서 제시한 트레이딩 시스템이 시장 평균 수익률을 상회한다는 사실을 확인함으로써 위의 가정에 대한 검증할 수 있었다.

A Two-Phase Stock Trading System based on Pattern Matching and Automatic Rule Induction

Jong-Woo Lee[†] · Yu-Seop Kim^{††} · Sung-Dong Kim^{†††}
Jae-Won Lee^{††††} · Jin-Seok Chae^{†††††}

ABSTRACT

In the context of a dynamic trading environment, the ultimate goal of the financial forecasting system is to optimize a specific trading objective. This paper proposes a two-phase (extraction and filtering) stock trading system that aims at maximizing the rates of returns. Extraction of stocks is performed by searching specific time-series patterns described by a combination of values of technical indicators. In the filtering phase, several rules are applied to the extracted sets of stocks to select stocks to be actually traded. The filtering rules are automatically induced from past data. From a large database of daily stock prices, the values of technical indicators are calculated. They are used to make the extraction patterns, and the distributions of the discretization intervals of the values are calculated for both positive and negative data sets. We assumed that the values in the intervals of distinctive distribution may contribute to the prediction of future trend of stocks, so the rules for filtering stocks are automatically induced from the data in those intervals. We show the rates of returns when using our trading system outperform the market average. These results mean rule induction method using distributional differences is useful.

키워드 : 주식 트레이딩 시스템(Stock Trading System), 패턴 매칭(Pattern Matching), 자동 규칙 생성(Automatic Rule Generation), 금융 예측 시스템(Finance Prediction System)

1. 서 론

투자자의 궁극적인 목적은 매매 목적을 만족시키는 척도

를 찾아서 최적화하는데 있는데, 매매 목적으로는 이익의 최대화, 경제활동 수단의 구축, 그리고 리스크를 고려한 이익 실현의 최적화 등이 있다. 따라서 지금까지, 이러한 목적들을 최적화하기 위해 금융 예측 모형에 관한 많은 연구 노력이 있어왔는데, 본 논문에서는 주식 시장에서의 최적의 금융 예측 시스템을 구축하는데 있어 발생할 수 있는 문제들에 대하여 언급한다.

금융 예측과 관련된 대부분의 연구는 예측 모형의 구축

* 본 연구는 2003년도 한성대학교 교내연구비 지원과제 임.

† 정 회 원 : 광운대학교 컴퓨터공학과 교수

†† 정 회 원 : 한림대학교 정보통신공학부 교수

††† 정 회 원 : 한성대학교 컴퓨터시스템공학과 교수

†††† 정 회 원 : 성신여자대학교 컴퓨터정보공학부 교수

††††† 종신회원 : 인천대학교 컴퓨터공학과 교수

논문접수 : 2003년 1월 27일, 심사완료 : 2003년 5월 29일

에 집중되어 왔다. 예측 모형들은 다양한 형태의 설명 변수, 즉 과거 가격 시퀀스에 기반한 기술 변수, 회사의 수익성에 관한 척도들과 같은 미시적 관점에서의 주식 관련 변수, 그리고 경기 사이클과 관련된 거시 경제 변수 등을 사용하였다[1]. 일반적으로 주식 시장을 예측하기 위해서는 두 가지 유형의 분석 방법론이 사용된다. 첫째, 기술 분석(technical analysis)은 일별 주가 데이터에 기반하여 이루어지는데, 일반적으로 단기적 예측 모형에서는 이 방법이 보다 효과적으로 사용될 수 있는 것으로 알려져 있다. 둘째, 가치 분석(fundamental analysis)은 각 회사의 경제적 활동과 재무제표에 관련된 정보를 주로 고려하는 방식으로 중장기 예측 모형에 유리한 것으로 알려져 있다. 금융 예측은 시계열 데이터를 대상으로 하는 문제인데 특히 여기서의 금융 데이터는 엄청난 잡음을 내포하고 있다. 그래서 잡음에 가장 효과적인 해법으로 알려진 인공 신경망(Artificial Neural-Net)으로 금융 예측과 같은 시계열 예측 문제를 해결하고자 하는 시도가 많은 연구자들에 의하여 고려되어 왔다[2]. 또한 잡음이 많은 시계열 분석에 신경망을 사용함으로써 금융 예측에 사용될 수 있는 규칙을 자동으로 추론하기도 하였다[3]. 그러나 위와 관련된 대부분의 연구에서는 복수 개의 단일 종목의 주가 예측이 아닌 주가지수의 예측에 그 초점이 맞추어져 있었다. 그래서 매매에 있어서 예측 시스템을 사용한다는 의미는 매매에 직접 예측 결과를 적용한다는 의미 보다는 여러 금융 상황정보와 해당 지식의 하나로써 예측정보를 통합한다는 것을 의미하였다. 한편, 동적인 주식 매매 환경에서 의사 결정을 하는데 있어서의 주된 접근 방식은 매매에 직접 적용되는 시스템을 구축하는 것을 의미하는데, 이 매매 시스템은 예측 모형은 물론 예측을 실제 매매 행위로 변환시키는 결정 규칙까지 포함해야 한다[4].

본 논문에서 우리는 한국 주식 시장을 구성하는 KOSPI와 KOSDAQ[5]에 상장된 여러 주식들을 고려하는 실제적인 2-단계 개별 주식 거래 시스템을 개발하였다. 실제 매매를 위하여 본 시스템은 추출과 여과 단계로 구성되며, 이 시스템의 궁극적인 목적은 매매 수익률을 최대화하는데 있다. 이 시스템은 또한 이익실현율, 손절매율, 그리고 거래 기간 등을 명시하는 등 거래 행위와 관련한 실제 인자들을 포함한다. 본 시스템의 분석과정에서는 기술적 지표들만을 사용하기 때문에, 본 시스템은 단기적인 주식 매매에 활용될 수 있다고 볼 수 있다. 먼저 주식의 추출은 패턴 매칭에 의해 수행되며 우리가 정의한 의사 결정 규칙을 사용하여 여과가 이루어진다. 여기서 우리가 사용하는 패턴은 개별 종목이 가지고 있는 기술 지표 값들을 자질 값으로 구성한 조합이다. 그리고 의사 결정 규칙(여과 규칙)들은 분포 차이를 이용하여 과거 데이터로부터 유도하였는데, 이 규칙들은 추출

된 주식들의 상태를 분류하여 투자 행위의 유무를 결정하는 규칙들이다. 우리의 주식 거래 시스템에서, 매수는 추출과 여과 결과에 기반하며, 매도는 거래 정책에 의거한다.

본 논문에서 사용되는 기술 지표 값들은 일별 주가 데이터를 저장하고 있는 데이터베이스를 조작함으로써 계산된다. 우리는 이 값들을 이용하여 우리의 추출 패턴을 정의하고 패턴에 부합하는 주식들을 수집하여 훈련 데이터로 사용하였다. 의사 결정 규칙을 유도하기 위해 훈련 데이터를 양성과 음성으로 분류하였는데 양성 데이터는 플러스 수익률을 보인 데이터를 말하며 음성 데이터는 마이너스 수익률을 보인 데이터를 말한다. 그리고 연속적인 값을 가지는 각 데이터의 자질 값들은 몇 개의 구간들로 나누어 이산화하였는데 이와 관련한 자세한 이야기는 2.4에서 설명하고 있다. 그리고 값 구간의 분포는 양성과 음성 데이터 집합들로부터 독립적으로 계산되었는데, 몇몇 구간들은 포함하고 있는 데이터들이 양성과 음성이 뚜렷이 구별되는 분포를 보였고 그 구간에 해당되는 자질 값들은 각 주식의 기대 수익률을 명확히 평가하는데 기여할 수 있을 것으로 가정하였다.

2절은 주식 거래 시스템 구축에 필요한 환경을 설명하고 본 논문에서 제안된 주식 거래 시스템의 전체적인 구조는 3절에서 설명한다. 그리고 3절에서 추출 패턴, 의사결정 규칙의 유도(induction) 과정, 그리고 거래 정책을 기술한다. 4절에서는 3개의 거래 시스템을 사용한 가상적인 투자의 결과를 보여주고, 5절에서는 결론과 후후 연구 과제를 제시한다.

2. 시스템의 배경 지식

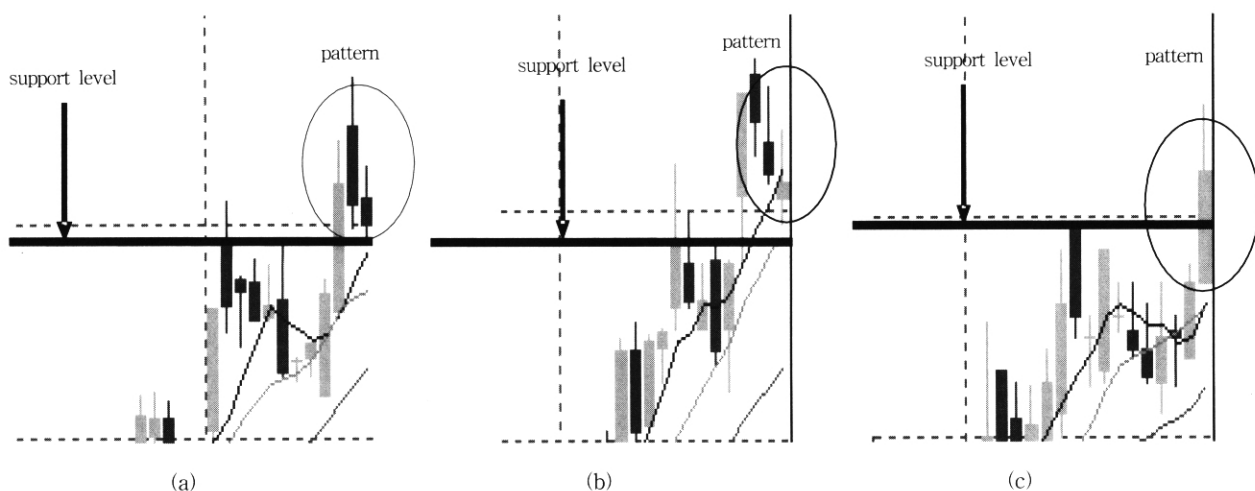
2.1 주식 데이터베이스

본 논문에서는 시스템의 구현을 위하여 우선 일별 주가로 구성된 주식 데이터베이스를 구성하였다. 주식 데이터로 KOSPI와 KOSDAQ에 상장된 모든 주식(현재 약 1700종목)에 대한 데이터를 수집하고, 이들 데이터는 일별 데이터로써 1990년 1월부터 2001년 12월까지 12년에 걸쳐 수집하였다. 각각의 원시 데이터(raw data : RD)는 7개의 필드로 구성되며 다음과 같이 표현된다.

$$RD = (name, date, p_o, p_h, p_l, p_c, v)$$

여기서 *name*은 회사명, *date*는 해당 일자, p_o 는 시가, p_c 는 종가, p_h 는 고가, p_l 은 저가이며, v 는 거래량을 나타낸다.

이 원시 데이터로부터 가격 이동평균이나 상대 강도 지수(Relative Strength Index : RSI)[6]와 같은 기술 지표들을 계산한다. 그리고 이러한 지표들의 값으로부터 시계열 패턴을 구성한다.



(그림 1) 각 패턴 집합에서의 패턴 예

2.2 시계열 패턴

금융 예측에 관한 기존의 연구를 살펴보면, 자기 조직화 맵(Self Organizing Map : SOM)[7], 신경망(Neural Network)[8]이 시계열 패턴을 발견하는데 사용되었다. 또한 최근에는 데이터 마이닝 기법을 통해 “변화 지점(change point)”을 식별해 내려는 노력이 있어 왔다[9]. 그러나 이들 연구에서는 주가를 시계열 데이터로 보고 추가적인 지식이 없이 순수하게 자동 처리하기만 하였다.

그러나 본 논문에서는 주식 전문가의 주관적인 지식을 사용하여 시계열 패턴을 구성하였다. 따라서 패턴은 시계열 데이터로부터 자동적으로 획득될 뿐 아니라 여기에 직관적이고 경험적인 도메인 지식이 추가된다. 여기서는 세 가지의 패턴 집합인 R 과 $S1$, 그리고 $S2$ 가 지지대 및 저항대[6]에 기반하여 구축되었다 : 첫 번째 집합 R 은 저항대를 돌파하는 패턴들을 위한 것이며, 나머지들($S1$, $S2$)은 지지대와 관계된 패턴들을 위한 것이다. (그림 1)은 이들 패턴 집합의 예를 보이고 있는데, (그림 1)(a)가 R 을, (그림 1)(b)가 $S1$ 을, 그리고 (그림 1)(c)가 $S2$ 를 보여주고 있다.

각각의 패턴($pattern_i$)은 각 주식 데이터로부터 계산된 기술 지표 값들의 조건($cond_i$)을 구성하여 표현할 수 있는데 n 개의 조건으로 이루어진 패턴은 다음과 같이 정의할 수 있다.

$$pattern_i = (cond_1, cond_2, \dots, cond_n)$$

다음은 4개의 조건으로 이루어진 시계열 패턴의 구체적인 예이다.

$$pattern_1 = (Grad5_0 \geq 1.23, RC_0 \geq 0.04, VL_1 < VL_0, Disp5-10_1 < Disp5-10_0)^{1)}$$

이 같은 패턴들은 3절에서 제시하는 주식 거래 시스템의 추출 단계에 사용된다.

2.3 훈련 데이터

훈련 데이터는 본 논문의 시스템에서 여과 단계에 사용될 의사결정 규칙을 만드는데 사용된다. 2.1절에 기술한 주식 데이터베이스로부터, 우리는 2.2절의 패턴에 부합하는 주식들의 데이터를 추출하였다. 이를 위하여, 1998년 1월부터 2000년 6월까지 30개월 동안의 데이터중에서 패턴 매칭으로 해당 주식들을 추출하였고 이들 주식과 관련한 데이터들을 수집하였다. 수집된 데이터($D(s)$)는 특정 패턴을 기술하는 기술 지표(technical indicator : ti) 값들과 기술 지표의 3일간의 추세를 반영하는 트라이그램(trigram : tri) 값들로 구성된다. 기술 지표 값으로는 주가 및 거래량 이동평균선의 위치, 이동평균선간 간격 등 수백가지의 지표 중에서 기술적 분석에서 주로 활용되는 수십개의 지표가 사용되었다. 다음 식은 n 개의 기술지표 값과 m 개의 트라이그램 값으로 이루어진 종목 s 의 개별 데이터를 보여준다.

$$D(s) = (ti_1, \dots, ti_n, tri_1, \dots, tri_m)$$

각각의 트라이그램 데이터는 원래의 실수 값(v)에 “signum” 함수를 적용하여 이산화된 3-항 심볼들 ($b_1, b_2, b_3 \in \{-1, 0, 1\}$)로 구성된다.

$$b_i = sign(v) = \begin{cases} -1, & \text{if } v < 0 \\ 0, & \text{if } v = 0 \\ 1, & \text{if } v > 0 \end{cases}$$

1) 이 패턴에서 아래첨자 0은 오늘을 의미하고, 1은 어제를 의미한다. $Grad5$ 는 5일 이동평균선의 기술기 값을 의미하고, RC 는 가격 변화율을, VL 은 거래량을, 그리고 $Disp5-10$ 은 5일 이동평균선과 10일 이동평균선간의 이격을 의미한다.

패턴을 기술하는데 필요한 조건들은 각 패턴에 따라서 모두 다른 기술 지표들을 사용하기 때문에, 각 패턴에 따라 필요한 주식들의 $D(s)$ 요소(element)들은 달라진다.

수집된 데이터는 다음과 같은 개별 주식 s 의 1일 수익률 $Ret_1(s)$ 에 기반하여 양성과 음성 데이터 집합으로 분류된다.

$$Ret_1(s) = \frac{\text{종가} - \text{시가}}{\text{시가}}$$

여기서 가격들은 추출일 다음날의 가격들이다. 그리고 수집된 주식(s)들이 속하게 될 클래스($C(s)$)는 다음과 같이 구할 수 있다:

$$C(s) = \begin{cases} \text{양성, if } Ret_1(s) \geq 0 \\ \text{음성, if } Ret_1(s) < 0 \end{cases}$$

<표 1>은 실험에 사용된 각 패턴 집합 별 추출 주식 수와 학습 데이터의 통계를 보인다.

<표 1> 학습 데이터의 통계

패턴 집합	합 계	양 성	음 성
R	12658	5324	7334
S1	5572	2628	2944
S2	6624	3000	3624

2.4 이산화

과거의 연구들을 살펴보면, 실수 값으로 되어 있는 금융 시계열 데이터의 기술지표 값들을 심볼 스트림(symbol stream)으로 이산화하고, 이 스트림에 예측 모델을 적용하면 많은 성능 향상이 가능하다는 것을 알 수 있었다[10, 11]. 그러나, 실수 값을 특정 구간으로 이산화할 때, 각각의 지표 값들이 변환될 이산화 구간의 가능한 갯수를 결정하고 동시에 실수로 이루어진 값을 각 구간에 매핑시키는 이산화 방법에 관한 문제는 대부분 임기응변적(ad hoc)인 방식에 의해 다루어졌다. 예를 들어 [10]은 5가지 통화(currency)들 간의 환율(exchange rate)을 조정함으로써 얻어지는 일간 수익을 9 구간으로 정량화하여 구분하였고, [11]에서는 수익을 최대 7 구간으로 나누어 정량화 하였다. 실수 값의 시계열 데이터를 정량화하기 위한 데이터-기반 인자 체계(data-driven parametric scheme)는 [12]에서 처음으로 소개되었다.

본 논문에서는 구간내에 존재하는 데이터들의 표준 편차가 최대가 되도록 구간의 갯수를 정함으로써 최적의 심볼 스트림을 얻고자 하였다. 이는 각 구간별로 내포하고 있는 데이터들의 분포의 차이가 크다는 것은 그만큼 해당 구간과 다른 구간의 데이터들이 매우 다른 성향을 가지고 있다는 의미로 볼 수 있으며, 주어진 데이터가 특정 구간으로

분류될 때 해당 데이터의 예측력이 커질 수 있다는 사실에 기반한 것이다. 특정 기술 지표(ti)의 구간 i (int_i)에서의 분포는 다음 공식으로 정의된다.

$$p_n(int_i) = \frac{\text{freq of values in } int_i}{\text{total freq of values}}$$

여기서 $freq$ 은 값들이 나타난 빈도를 의미한다. 본 연구를 위하여, 우리는 구간의 개수를 3개에서 10개 사이로 고려하였다. 그리고, 특정 기술지표(ti)를 위한 구간 수(n_{ti})는 다음과 같은 계산식을 통하여 결정하였다.

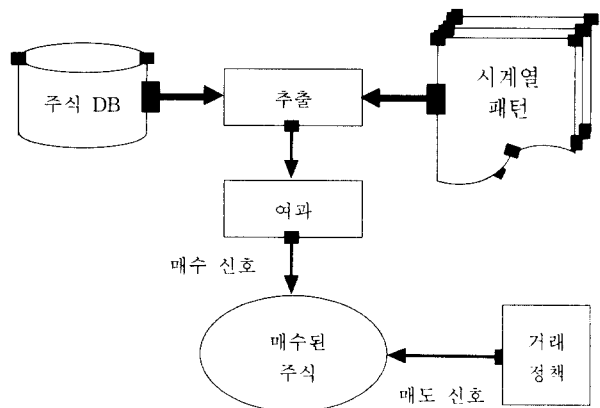
$$n_{ti} = \text{argmax}_{3 \leq k \leq 10} \sigma(p_{ti}^k)$$

여기서 $\sigma(p_{ti}^k)$ 는 구간 수가 k 일 때 분포의 표준 편차이다. 이때 모든 구간의 크기는 같도록 하였다. 그리고 각 구간의 위치를 결정하기 위해서, 개별 기술 지표 값(v)들은 먼저 내림차순으로 정렬되고 i 번째 위치(pos_i)는 다음과 같이 결정된다.

$$pos_i = \max(v) - (\max(v) - \min(v)) \times i$$

3. 매매 시스템

3절에서는 추출과 여과 단계로 구성되고 인자화된 매매 정책을 사용하는 주식 매매 시스템을 제시하고 있다. (그림 2)는 이 거래 시스템의 전체적인 구조를 보여주고 있다.



(그림 2) 주식 거래 시스템의 구조

3.1 추출 단계

이 단계에서는, 하나의 주식은 기술 지표 값들의 집합으로 표현되는데 이들 값들은 2.1절에서 설명한 바대로 주식 데이터베이스로부터 추출한 데이터를 계산함으로써 얻어진다. 또한 2.2절의 구성된 패턴들과 관련한 조건들은 2.3절에서 표현된 주식 기술 공식을 가지고 검사한다. 실제 주식

시장에는 다수의 주식들이 있기 때문에, 우리는 단일 주식에 대한 예측 대신 패턴-기반 추출을 수행한다. 패턴들이 다중 시계열 데이터에 적용되고 특정 패턴에 대한 모든 조건들을 만족시키는 지표 값들을 가진 주식들이 추출된다. 패턴 집합들(C_p)에 기반한 매수 후보들의 집합은 다음과 같이 표현된다.

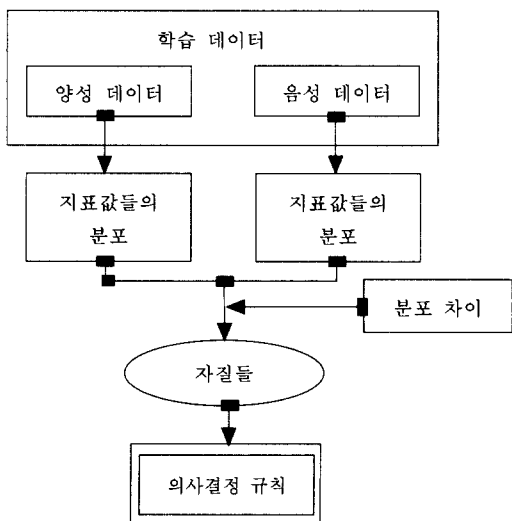
$$C_p = \{cs \mid \forall ti_i \text{ of } cs, cs(v_{ti_i}) \vdash cond_i \in pattern_i (\in P)\}$$

여기서 P 는 패턴 집합이며, cs 는 후보 주식(candidate stock : cs), $cs(v_{ti_i})$ 는 cs 의 기술 지표 값 ti_i , 그리고 ' \vdash '는 조건에 대한 만족을 의미한다.

3.2 여과 단계

최적의 매매를 위해서는, 3.1절에서 추출된 후보 주식들을 대상으로 하여 가장 최적의 수익률을 보일 수 있을만한 주식들을 선택하여야 한다. 즉, 동일 패턴에 해당하는 것으로 인정되어 추출된 주식들이라도, 개별 주식들의 기술 지표의 값들은 얼마든지 서로 다를 수 있다. 따라서 매수를 위한 의사 결정을 돕기 위해서는 이들 주식들을 대상으로 하고 기술 지표 값들을 이용한 분류(classification) 규칙을 구축할 필요가 있다.

본 논문에서는 추출된 주식들을 **수익-기대, 손실-기대**의 두 가지 범주로 분류하는 규칙들을 2.3절에서 구축된 훈련 데이터를 사용하여 자동으로 유도하였다. 물론 여기서는 수익-기대 주식들만이 여과될 것이다. (그림 3)은 규칙 유도의 개괄적인 과정을 보이고 있다.



(그림 3) 규칙 유도의 개괄적 과정

여기서 기술 지표들의 이산화 구간들의 분포들은 2.4절의 공식을 이용하여 양성 데이터 집합과 음성 데이터 집합에

서 각기 따로 계산하였다. 즉, 이산화 구간들을 랜덤 변수로 간주하며 이 변수들의 확률 분포를 구축한 것이다. 그리고 하나의 주식을 기술하는데 사용된 모든 기술 지표들에 대하여, 구간들의 분포들을 비교하였다. 또한 분포 차이를 사용하여, 몇 개의 구간들을 분류 규칙들을 구축하기 위한 자질로 선택하였다. 선택된 자질들의 집합(FS_p)은 다음과 같이 표현된다.

$$FS_p = \{f_{ti_i}, f_{tri_i} \mid p_{pos}(f_{ti_i}) > p_{neg}(f_{ti_i}) + \epsilon \text{ or } p_{pos}(f_{tri_i}) > p_{neg}(f_{tri_i}) + \epsilon\}$$

여기서 f_{ti_i} 는 ti_i 의 이산화 구간을, f_{tri_i} 는 트라이그램 값을 각각 나타낸다. 또, $p_{pos}(f_{ti_i})$ 는 f_{ti_i} 의 양성 데이터 집합에서의 확률을 의미하고, $p_{neg}(f_{ti_i})$ 는 음성 데이터 집합에서의 확률을 의미한다. 상수 ϵ 는 두드러진 자질들을 결정하기 위해 사용된다.

그리고 집합의 각 원소에 대하여, 다음의 규칙(r_i)을 생성하였다.

$$\text{if } v_{ti_i} \in f_{ti_i}, \text{ then } RS(r_i) = p_{pos}(f_{ti_i}) - p_{neg}(f_{ti_i}) \\ \text{else } RS(r_i) = 0$$

여기서, 각 규칙에 대하여 구하여지는 규칙 점수 $RS(r_i)$ 는 확신도 요소로 간주할 수 있다.

매수 신호(bs)는 주식의 점수를 사용하여 생성되는데 매수 신호 함수는 아래와 같이 정의할 수 있다.

$$f_{bs}(cs) = \begin{cases} 1, & \text{if } \sum_{i=1}^n RS(r_i) > \lambda \\ 0, & \text{otherwise} \end{cases}$$

여기서 cs 는 3.1절의 식에서의 C_p 의 원소이고 λ 는 임계값이다. 마지막으로, 여과된 주식 집합(BS_p)은 아래와 같이 정의되며 이 집합의 주식들에 대해 매수 행위가 수행된다.

$$BS_p = \{s \mid f_{bs}(s) = 1 \text{ and } s \in C_p\}$$

3.3 주식 거래 정책

주식 거래 정책(STP)은 목표 수익률(target profit ratio : tpr), 손절률(stop loss ratio ; slr), 그리고 최대 보유 기간(maximum holding period : mhp)를 다음과 같이 지정함으로써 결정하였다.

$$STP = (tpr, slr, mhp)$$

이때 STP 는 결국 매도 행위를 결정하게 된다.

여과 단계에서의 매수 신호는 주식 시장이 종료된 후에 생성될 수 있으므로, 실질적인 매수는 신호가 생성된 다음 날에 이루어지게 된다. 그리고 매수는 시가에 이루어지는 것으로 가정한다. 주식 거래 정책 $STP = (\alpha, \beta, \gamma)$ 가 주어지면, 목표 수익가(target profit price : TPP)와 손절가(stop loss price : SLP)를 구할 수 있다.

$$TPP = bp \times (1 + \alpha), \quad SLP = bp \times (1 - \beta)$$

여기서 bp 는 매수가이다. 지정된 최대 보유 기간 내에 주가가 목표 수익가나 손절가에 도달하면 매도가 이루어진다. 그러나 이 기간 동안 두 가지 가격이 모두 나오지 않으면, 최대 보유 기간의 마지막 날의 종가에 주식을 매도한다.

4. 실험

<표 2> 거래 주식의 통계

패턴 집합(P)	C_p 수	BS_p 수
R	1829	668
S1	1116	604
S2	1288	895

본 논문에서는 제안된 거래 시스템을 사용하여 2001년 6월부터 12월까지 6개월간의 데이터에 대하여 주식 거래를 시뮬레이션 하였다. <표 2>는 2.2절의 세 가지 패턴 집합을 사용하여 추출 및 여과한 주식들의 수를 보이고 있다. 두 가지 거래 정책 $STP_1 = (0.05, 0.05, 1)$ 과 $STP_2 = (0.1, 0.05, 5)$ 를 사용하여 각 패턴 집합의 수익률을 계산하였다. 즉, 일 단위 수익률과 주 단위 수익률에 대한 예측력을 평가한 것이다. BS_p 에 포함된 주식들에 대하여 매수를 수행하였고 3.3절에 기술한 방식으로 매도를 수행하였다. 수익률 $Ret(s)$ 는 2.3에서와는 조금 다르게 다음과 같이 정의하였다.

$$Ret(s) = \frac{\text{매도가} - \text{매수가}}{\text{매수가}} \times 100$$

또한 거래 시스템의 평가 척도로 거래 당 수익율(profit per trade : PPT)을 사용하였다.

$$PPT = \frac{\sum_{i=1}^n Ret(s)}{n} (\%)$$

$$s \in BS_p, \text{ and } n = |BS_p|$$

<표 3>은 위 실험의 평가 결과를 보여주고 있다. STP_2 의 경우 평균 보유 기간은 R, S1, S2에 대하여 각각 2.78,

2.88, 2.78일 이었다. PPT 는 거래세를 감안하지 않고 계산하였다. 본 시스템의 성능을 보다 자세히 비교하기 위하여, 여과 단계에 결정 트리(Decision Tree)와 신경망(Neural-Net)을 사용한 결과를 <표 3>에 함께 제시하였다. 여기서 결정 트리를 위하여 C4.5 알고리즘[13]을 사용하였다. 또한 신경망은 백프로퍼게이션(back-propagation) 알고리즘에 기반한 다층 전방향(multi-layered feedforward) 구조를 사용하였으며, 2.3절에 제시한 것과 동일한 학습 데이터를 사용하여 학습을 수행하였다. <표 3>에 제시한 결과에 의하면, 본 논문에서 제안한 분포 차이를 이용한 자동 규칙 유도 방식이 더 나은 성능을 보이는 것을 알 수 있다. 또한, 몇 가지 경우에는 거래세를 극복하여 실질적인 수익을 얻을 수 있는 것으로 나타났다. 시물레이션 기간 동안, KOSPI는 약 16.8% 상승하였고 KOSDAQ은 약 -6.4% 하락하였다. S1 패턴 집합에 대한 거래의 단순 누적 수익률은 170.328%²⁾과 714.532%³⁾였다. 물론 이 결과가 실제 거래 상황에서의 결과와 약간의 차이가 있을 수도 있겠지만, 거래세를 감안해서도 시장 평균을 상회하는 수익률을 기대할 수 있을 것으로 보인다.

<표 3> 평가 결과(PPT)

패턴 집합	추출	여과	의사 결정 트리	신경망	
R	STP 1	-0.322	0.107	-0.17	-0.01
	STP 2	-0.58	-0.275	0.35	0.03
S1	STP 1	0.365	0.782	0.757	0.505
	STP 2	0.937	1.683	1.65	1.165
S2	STP 1	-0.03	0.318	0.04	0.145
	STP 2	0.085	0.265	0.344	0.273

5. 토론과 결론

우리는 본 논문에서 패턴 매칭과 자동 규칙 유도에 기반한 추출과 여과로 이루어진 2-단계 주식 거래 시스템을 제안하였다. 이를 위하여 단일 스톡 시계열에 대한 예측 모델을 구축하는 대신, 우리는 실제 주식 시장의 다중 종목에 적용될 수 있는 시계열 패턴을 구성하였다. 본 논문에서의 거래 시스템은 주식들에 대한 기술 분석에 기반하여 시계열 패턴을 구축하고, 주식들을 여과하기 위한 분류 규칙을 자동으로 유도하는 데에 기술 지표들을 사용하였다. 의사결정(여과) 규칙의 자동 획득에 있어서 분포 차이를 이용하는 비교적 단순한 방법을 사용하였지만 이 방식이 경쟁력 있

2) $(0.782 \cdot 0.5) \times 604 = 170.328$.

3) $(1.683 \cdot 0.5) \times 604 = 714.532$.

는 성능을 보인다는 것을 확인할 수 있었고, 규칙들의 빠른 유도를 돕는다는 것도 알 수 있었다. 이 방법은 독립적으로 그리고 결합적으로 예측력이 있는 자질들을 사용하는데, 이것은 자질간 상호작용을 놓칠 수 있다. 수치 값으로 기술하기 힘든 기본(fundamental) 데이터와 같은 다른 많은 요소들이 주식들의 미래 추세에 영향을 미치므로, 자질들간의 상호 연관성을 찾는 것은 복잡한 작업이다. 여기서 우리는 어떤 명확한 자질들을 찾고자 하였다. 그리고 분포 차이를 이용하는 자질 선택 방법은 신경망 학습을 위한 인자들을 줄이는 것을 도와 줄 수 있다. 또한 패턴은 전문가의 지식에 의해 모형화된다. 본 매매 시스템은 과거 데이터를 사용하여 그 지식을 검증하고 개선하는 역할을 하며 동시에 본 시스템 구조는 전문가의 예측 모델을 구축하는데 유용할 수 있을 것이다.

참 고 문 헌

[1] J. Ghosn and Yoshua Bengio, "Multi-Task Learning for Stock Selection," Advances in Neural Information Processing Systems, 9, M. C. Mozer, M. I. Jordan and T. Petsche editor, The MIT Press, 1997.

[2] A. Refenes, "Neural Networks in the Capital Markets," John Wiley and Sons, 1995.

[3] C. L. Giles, S. Lawrence and A. C. Tsoi, "Rule Inference for Financial Prediction using Recurrent Neural Networks," Proc. of IEEE/IAFE Conference on Computational Intelligence for Financial Engineering, pp.253-259, 1997.

[4] N. Towers and A. N. Burgess, "Optimisation of Trading Strategies using Parameterised Decision Rules," Proc. of IDEAL 98, Perspectives on Financial Engineering and Data Mining, L. Xu et al editor, Springer-Verlag, 1998.

[5] "Korea Stock Exchange : KOSPI & KOSPI 200," <http://www.kse.or.kr>, 2001.

[6] "Optima Investment Research : Interpreting Technical Indicators," Fourth Edition, Neural Networks in the Capital Markets, John Wiley and Sons, 1995.

[7] T. Fu, F. Chung, V. Ng and R. Luk, "Pattern Discovery from Stock Time Series Using Self-Organizing Maps," Workshop on Temporal Data Mining, pp.27-37, 2001.

[8] J. Baek and S. Cho, "Left Shoulder Detection in Korea Composite Stock Price Index Using an Auto-Associative Neural Network," Proc. of IDEAL 2000, Data Mining, Financial Engineering and Intelligent Agents, 2000.

[9] V. Guralnik and J. Srivastava, "Event Detection from Time Series Data," Proc. of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.

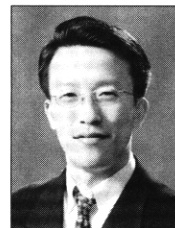
33-42, 1999.

[10] C. P. Papageorgiou, "High frequency time series analysis and prediction using Markov models," Proc. of IEEE/IAFE Conference on Computational Intelligence Financial Engineering, pp.182-185, 1997.

[11] C. L. Giles, S. Lawrence and A. C. Tsoi, "Noisy time series prediction using a recurrent neural network and grammatical inference," Machine Learning, 44, pp.161-183, 2000.

[12] P. Tino, C. Schittenkopf and G. Dorffner, "Volatility trading via temporal pattern recognition in quantized financial time series," Pattern Analysis and Applications, 2001.

[13] R. Quinlan, "C4.5 : Programs for Machine Learning," Morgan Kaufmann, 1992.



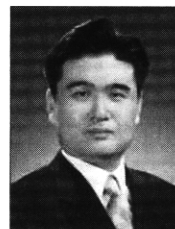
이 종 우

e-mail : jwlee44@daisy.kw.ac.kr

1990년 서울대학교 컴퓨터공학과(학사)
 1992년 서울대학교 컴퓨터공학과(석사)
 1996년 서울대학교 컴퓨터공학과(박사)
 1996년~1999년 현대전자산업 선임연구원
 1999년~2002년 한림대학교 정보통신공학부 조교수

2002년~현재 광운대학교 컴퓨터공학과 조교수

관심분야 : 전산금융, 군집 컴퓨팅, 분산 병렬 시스템, 시스템 소프트웨어



김 유 섭

e-mail : yskim01@hallym.ac.kr

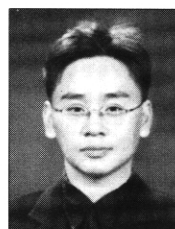
1992년 서강대학교 전자계산학과(학사)
 1994년 서울대학교 컴퓨터공학과(석사)
 2000년 서울대학교 컴퓨터공학과(박사)
 2000년~2001년 서울대학교 컴퓨터신기술 공동연구소 전문연구요원

2001년 (주)아이시티 연구소장

2001년~2002년 이화여자대학교 과학기술대학원 연구전임강사

2002년~현재 한림대학교 정보통신공학부 조교수

관심분야 : 전산금융, 자연언어처리, 기계번역, 데이터마이닝, 기계학습

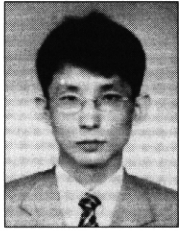


김 성 동

e-mail : sdkim@hansung.ac.kr

1991년 서울대학교 컴퓨터공학과(학사)
 1993년 서울대학교 컴퓨터공학과(석사)
 1999년 서울대학교 컴퓨터공학과(박사)
 2001년~현재 한성대학교 컴퓨터시스템공학과 조교수

관심분야 : 전산금융, 자연언어처리, 기계번역, 데이터마이닝, 기계학습

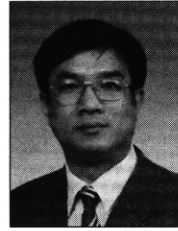


이 재 원

e-mail : jwlee@cs.sungshin.ac.kr

- 1990년 서울대학교 컴퓨터공학과(학사)
- 1992년 서울대학교 컴퓨터공학과(석사)
- 1998년 서울대학교 컴퓨터공학과(박사)
- 1999년~현재 성신여자대학교 컴퓨터정보공학부 전임강사

관심분야 : 전산금융, 인공지능, 기계학습, 자연어처리, 컴퓨터뮤직



채 진 석

e-mail : jschae@incheon.ac.kr

- 1990년 서울대학교 컴퓨터공학과(학사)
- 1992년 서울대학교 컴퓨터공학과(석사)
- 1998년 서울대학교 컴퓨터공학과(박사)
- 1992년~1997년 서울대학교 공학연구소 조교

1997년~1998년 한국연구정보센터 선임연구원

1998년~현재 인천대학교 컴퓨터공학과 조교수

관심분야 : 인터넷 소프트웨어, 마크업 언어, 전자 도서관