장바구니분석을 이용한 주식투자전략 수립 방안

주 영 진*

A Trade Strategy in Stock Market using Market Basket Analysis

Joo, Young Jin

Abstract

We propose a new application method of the datamining technique that might help building an efficient trade strategy in the stock market, where the analysis of the huge database is essential. The proposed method utilizes the association rules among the price changes of individual stock from the market basket analysis (a datamining technique typically used in the Marketing field) in building the strategy. We also apply the proposed method to the daily stock prices in Korean stock market, from Jan. 2000 to Dec. 2001. The application results show that the proposed method gives an significantly higher yield rate than the actual stock chage rate.

* 배재대학교 전자상거래학과 전임강사

1. 서론

정보기술의 발달은 다양한 분야에서 대규모의 데이터를 생성하게 되었으며, 이에 따라 전통적인 통계기법을 이용하여 분석되어 온 축적된 데이터에 대한 다양한 태스크들을 보다 편리하고 효과적으로 처리하기 위한 방안으로 개발된 것이 데이터마이닝기법이다. 즉, 데이터마이닝은 대규모 데이터베이스로부터 탐색과 분석을 통하여 의미있는 패턴이나 규칙을 찾아내는 과정으로, 여기에는 데이터의 군집화(clustering), 분류(classification), 예측(prediction), 규칙발견(rule discovery) 등이 포함된다[조성준 2001].

효과적인 데이터마이닝기법을 적용한 문제해결방식은 많은 분야에서 통계기법 등 기존 방법론에 의한 문제해결방식의 대안이나 보완적 수단으로 사용되어 오고 있으며, 이러한 데이터마이닝기법을 적용한 문제해결방식은 경영학분야에 있어서도 마케팅, 재무관리 등 각 분야에서 널리 사용되어 오고 있다. 이는 기존의 이론적 모형들이 엄격한 통계적 가정을 기초로 만들어져 실증연구에서 부딪히는 상황의 변화나 복잡한 환경변수들의 영향을 고려하지 못하는 것을 데이터마이닝기법이 보다 쉽게 제공하기 때문이다[이건창 2000].

마케팅 분야의 경우 일찍이 케이블TV를 통한 고객 개개인의 광고에 대한 노출정보와 이와 연계한 고객의 구매행태에 관한 패널자료는 상당한 규모의 고객 데이터를 제공하고 있으며, 최근 폭발적인 확산을 보이고 있는 전자상거래는 이러한 고객 데이터베이스를 보다 방대한 규모로 만들었다. 이러한 고객 데이터를 정보시스템과 연계하여 일대일 마케팅을 위한 수단으로 개인화 기술 등에 응용하고 있다[Allen et al. 1998, 이건창 외 1인 1999, 김종우 외 1인 2000].

재무관리 분야에서도 재무관리 이론에 근거한 주가결정모형이 실제 주식시장의 매우 복잡하고 다양한 변화를 효과적으로 설명하기 어려운 점을 보완하고자 데이터마이닝기법을 사용하여 왔다. 특히, 재무관리 분야의 경우에는 주가예측 등을 위해 신경망(Neural Network)모형을 기본으로 한 다양한 데이터마이닝기법들을 많이 사용하고 있다[오경주 외 2인 2001, 김경태 외 2인 2001].

이렇듯 다양한 분야에서 응용되고 있는 데이터마이닝 기법들은 정보기술의 발달에 따라 각 분야에서 축적되고 있는 데이터들의 규모가 증가할수록 더 큰 관심을 모을 것이다. 또한, 데이터마이닝 기법을 사용하여 다양한 분야의 방대한 자료들을 보다 효과적으로 분석하기 위해서는 지속적인 관련 응용 문제에 대한 파악과 이를 위한 데이터마이닝 적용 기법을 개발할 필요가 있다고 할 수 있다.

이러한 관점에서, 본 연구는 이제까지 주로 마케팅 분야에서 연관성 규칙을 추출하기 위해 주로 사용되어온 장바구니분석기법을 주식투자전략 수립에 활용하기 위한 방안을 제안하고, 실증분석을 통해 그 유용성을 검증하고자 한다. 이를 위해 본 연구는 2장에서 장바구니분석기법 및 활용, 주가예측을 위한 데이터마이닝기법 등에 대한 기존 연구들을 검토한 후, 3장에서 장바구니분석기법을 주식투자전략에 활용할 수 있는 방안을 모색하여 제시하고 있다. 제시된 전략에 대한 실증분석의 결과는 4장에 정리되고 있으며, 5장은 본 연구의 결론 및 향후 관련된 추가 연구진행방향을 제시하고 있다.

2. 관련 연구

2.1. 장바구니분석 기법 및 활용

장바구니분석기법이란 자동화된 연관성 규칙 추출 기술로 교차판매에 활용되는 대표적인 데이터마이닝 기법으로, 상품 X가 상품 Y와 함께 판매될 확률이 높다는 연관성 규칙을 추출하여 이를 상품 진열 레이아웃, 상품 패키징 등에 활용하는 분석기법이다[강현철 외 4인 1999, 김종우 외 1인 2000, Berry & Linoff 1997].

상품 X와 상품 Y간의 연관성($X \Rightarrow Y$)이 활용가치가 있기 위해서는 먼저, 전체 거래 중에 상품 X와 상품 Y의 거래를 포함하는 거래가 많아서 그러한 상품 X, Y가 함께 거래되는 것이 일회적인 것이 아니라 빈번히 발생할 수 있음을 보일 수 있어야 하며, 특정 거래가 상품 X를 포함하고 있다면 그 거래가 상품 Y의 구매도 포함하는 경우가 많아야 하고, 임의로 상품 Y를 구매한 경우보다 상품 X를 구매한 경우에 상품 Y를 구매한 경우가 많아야 한다. (식 1)은 이러한 높은 연관성을 위한 기준이 되는 계량적 지표들인 Support, Confidence, Lift 등을 나타내고 있다.

$$\begin{aligned} Support &= \frac{n(X \cap Y)}{N} \\ Confidence &= \Pr(Y \mid X) \\ Lift &= \frac{\Pr(Y \mid X)}{\Pr(Y)} \end{aligned} \qquad \text{(식 1)}$$

(식 1)에서 $n(X \cap Y)$는 상품 X, Y에 대한 거래를 모두 포함하는 거래의 수를, N은 전체 거래수를 의미하며, $\Pr(Y \mid X)$는 상품 X를 구매한 경우 상품 Y도 구매하였을 조건부확률을, $\Pr(Y)$는 상품 Y를 구매할 확률을 의미한다.

연관성 규칙은 이들 3가지 지표를 기준으로 추출하게 되는데, 어떤 연관성 규칙이 의미있는지를 판단하기 위한 Confidence, Support, Lift값의 기준이 이론적으로 결정되어 있는 것은 아니며, 일반적으로 분석자의 판단과 경험에 의해 결정되는 경우가 많다.

전통적으로 상품 진열 레이아웃, 상품 패키징 등에 활용되던 연관성 규칙에 의한 장바구니분석기법은 인터넷상점에서 축적된 고객구매자료를 분석하여 의미있는 인터넷마케팅 전략으로 발전될 수 있다. 이와 관련하여 Wal-Mart에서는 1980년대부터 축적된 방대한 고객 데이터베이스를 분석하여 고객들이 한번의 방문에서 함께 구매하는 물품들을 탐색하고 이를 상품진열에 활용하고 있다[Stair & Reynolds 1999, Zellner 1998]. 또한, 김종우 외[2000]에서는 장바구니분석기법을 이용하여 추출된 인터넷상점에서의 상품구매에 대한 연관성 규칙을 인터넷상점을 이용하는 개인에 대한 차별화된 개인화 광고 추천에 적용하여 의미있는 결과를 보이고 있다.

2.2. 주가예측을 위한 데이터마이닝 기법

주가를 설명하기 위한 기본모형인 APT(Arbitrage Pricing Theory) 모형[Ross 1976]에 따르면 개별 주식의 수익률은 해당 주식의 수익률 변동에 영향을 주는 몇개 요인들에 의해 설명이 된다. 이에 따라 주가예측을 위한 많은 모형들에서는 개별 주식의 수익률을 의미있게 설명할 수 있는 요인들을 효과적으로 추출하기 위한 것으로 이해될 수 있으며, 순수하게 재무관리 이론에 근거한 모형들은 이론적 뒷받침은 뛰어날 수 있으나 주식시장의 복잡하고 다양한 요인들을 효과적으로 설명하기에는

부족할 수 있다.

이에 따라 주식시장에서 주가의 변화를 예측하는 문제는 이제까지 데이터마이닝 기법이 가장 활발하게 적용되어 온 분야들 중 하나이다. 즉, 데이터마이닝 기법이 재무관리 이론에 근거한 주가결정모형에 비해 주식시장의 매우 복잡하고 다양한 변화를 보다 쉽고 효과적으로 설명하기 때문이다. 더욱이 정보기술의 발달과 인터넷에 의한 주식거래의 급증으로 인하여 주식시장에서 거래되는 각종 주식들에 대한 시간대별, 분별, 초별 자료 등이 방대한 데이터베이스를 형성하고 있는 현재, 데이터마이닝 기법에 의한 다각적인 주가예측은 보다 높은 관심을 끌고 있다고 할 수 있다.

이제까지 주가예측에 있어서는 단변량 시계열분석 또는 다변량 시계열분석에서 내재된 시계열이 갖는 고도의 비선형성을 설명하기에 적합한 신경망(Neural Network)모형이 많이 사용되어 왔다.

이러한 연구들 중 초창기 연구의 하나인 Kimoto et al. [1990]에서는 신경망을 이용하여 다양한 변수들간의 관계를 도출하여 도쿄주식시장에서의 지수예측을 수행하였고, 국내에서도 이와 유사하게 수행된 연구로는 박종엽 외[1995], 지원철 외[1996] 등이 있다. 초기연구에서 신경망을 이용한 주가예측모형은 일반적으로 주가의 상승 및 하락 등의 패턴을 예측하고자 시도된 것들과 직접적으로 주가를 예측하고자 시도된 것으로 나눌 수 있다.

이후 신경망을 이용한 주가예측연구는 주가변동상황을 보다 정확히 모형에 반영함으로써 신경망 모형의 예측력을 보다 높일 수 있는 방안에 관한 연구[오경주 외 2인 2001]와 신경망모형을 유용성을 바탕으로 각종 재무관련 지표들이 주가에 영향을 주는지를 검증하기 위한 연구[김경태 외 2인 2001] 등으로 확대 발전되었다. 즉, 오경주 외[2001]에서는 일반적으로 경기변동론에서 많이 논의되었던 상승/하락 국면전환에 대한 이론을 적용하여 주가의 전환점을 추정하여 신경망모형에 결합함으로써 보다 높은 예측력을 보이는 신경망모형에 의한 주가예측모형을 제시하였으며, 김경태 외[2001]에서는 주가와 주가에 영향을 줄 것으로 예상되는 변수들 간의 신경망모형을 이용하여 회계정보가 주가를 유의적으로 설명하고 있음을 보이고 있다.

3. 장바구니분석에 의한 주식투자전략

3.1. 장바구니분석의 주식투자전략 활용 가능성

일반적으로 많은 경제/경영 시계열들은 공통의 내재요인을 가지고 있는 것으로 알려져 있다[Engle & Kozicki 1993]. 즉, 많은 경제/경영 시계열들은 해당 시계열들이 영향을 받는 환경을 공유함에 따라 추세 등의 내재요인이 함께 움직이는 것으로 나타나게 된다.

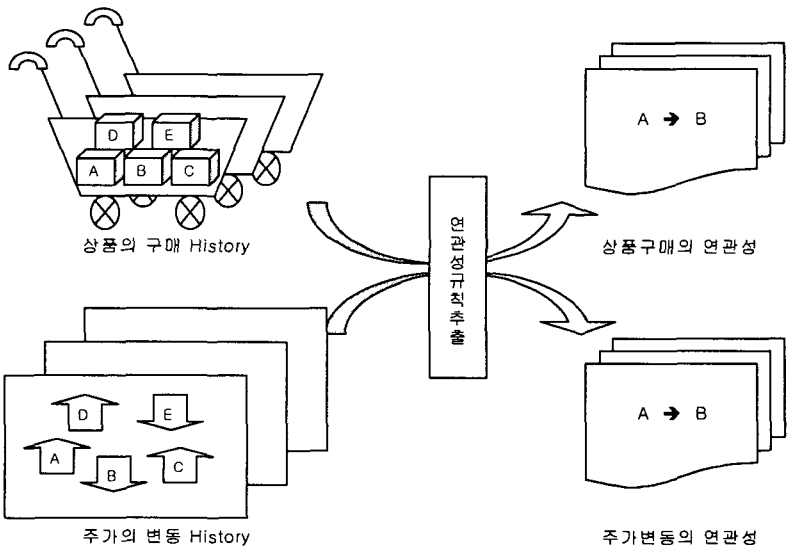
주식시장에서의 개별주식의 가격도 각각의 주식이 속한 산업군, 기업군 등 다양한 공통적 환경에 따라 공통의 내재요인을 갖고 있다고 가정할 수 있다. 즉, 특정산업군에 속한 대다수 기업이나, 특정기업군에 속한 대다수 계열사의 주가의 변화가 개별적 요인이 미미한 경우 대체로 비슷한 움직임을 보일 수 있다.

본 연구는 이와 같이 개별주식들의 주가변화가 공통된 내재요인에 따라 비슷한

움직임을 보일 수 있음과 이러한 공통된 내재요인이 모든 주식에 동시에 반영되기 보다는 관련 정보가 시장에 전파됨에 있어 약간씩의 시차가 발생될 수 있을 것이라는 가정에 근거하여, 이러한 공통된 내재요인의 변화를 데이터마이닝 기법 중 연관성 규칙 추출을 위한 장바구니분석으로 찾아내서 이를 주식투자전략에 활용할 수 있는 방안을 제시하고자 한다.

앞의 2장에서 살펴본 바와 같이 연관성 규칙에서는 Support, Confidence, Lift 등의 기준을 통해 특정 상품(군)의 구매가 발생할 때 동시에 구매가 발생하게될 또 다른 특정 상품(군)을 추출하게 된다. 개별 주식들의 주가의 변동에 있어서도 공통된 내재요인의 영향에 따라 특정 주식(군)의 주가 변동이 또 다른 특정 주식(군)의 주가 변동과 높은 연관성을 갖게 된다. (그림 1)은 상품 구매시 발생되는 연관성 규칙과 주가 변동에서의 연관성 규칙간의 개념적 대응관계를 나타내고 있다.

(그림 1)에서 상품 구매와 관련된 연관성 규칙은 한번의 구매에 포함되는 상품들에 대한 정보를 바탕으로 추출된다. 예를 들어 상품 A와 상품 B가 동시에 구매되는 빈도가 상대적으로 높다면, 상품 A의 구매와 상품 B의 구매간의 연관성 규칙이 추출될 수 있다. 이제, 주가 변동과 관련된 연관성 규칙은 동일 시점(특정일 또는 특정 시점)의 거래결과 발생된 개별 주식들의 이전 시점 대비 주가변동에 대한 정보를 바탕으로 추출된다. 이 경우 개별 주식의



(그림 1) 상품구매 및 주가변동의 연관성 규칙

주가변동은 연속적인 값이 아닌 상승/하락 등과 같이 변동폭에 대한 구간을 고려한 이산적인 등급값을 의미한다. 이제 주식 A와 주식 B의 주가변동에 있어 주식 A의 주가가 상승하는 것과 주식 B의 주가가 상승하는 것이 동시에 발생하는 빈도가 상대적으로 높은 것과 같이 개별 주가변동이 같는 특정 변동 구간에 대한 이산적 등급값이 함께 발생되는 빈도가 상대적으로 높다면, 이러한 정보를 바탕으로 주가변동의 연관성 규칙이 추출될 수 있다.

한편, 상품 구매와 관련하여 추출된 연관성 규칙들은 상품의 진열 레이아웃, 패키징, 개인화된 광고추천 등의 마케팅전략에 활용될 수 있다. 즉, 상품 A와 상품 B의 연관성이 높은 경우, 상품 A, B를 함께 진열한다거나, 상품 A를 구매한 고객에게 상품 B에 대한 광고를 개인화하여 제공하는 전략을 구사할 수 있다.

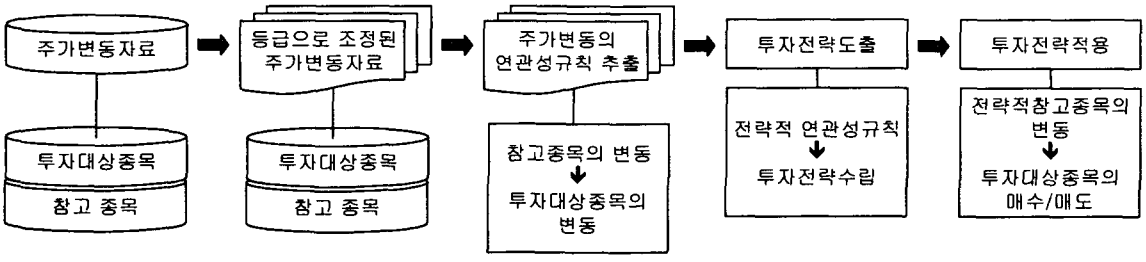
주가변동의 연관성 규칙을 의미있게 도출한다면, 이러한 주가변동의 연관성 규칙은 특정 종목에 대한 매수 및 매도에 대한 추천전략으로 활용될 수 있다. 즉, 주식 A와 주식 B의 연관성이 높은 경우, 주식 A의 주가변동으로 주식 B의 매수 및 매도 신호를 포착하여 이를 주식 B에 대한 투자전략으로 활용할 수 있다. 이 때, 실증적 유용성을 고려하여 주식 A와 주식 B의 관찰시점에 차이를 두어 과거 또는 현재에 발생한 주식 A의 변동과 미래에 발생할 주식 B의 변동간의 연관성 규칙을 도출하여 전략화 하는 것이 요구된다.

3.2. 주식투자전략의 도출

(그림 2)는 장바구니분석을 통해 도출된 연관성 규칙을 주식투자전략에 활용하기 위한 절차를 나타내고 있다.

먼저, (그림 2)에서와 같이 투자대상종목 및 참고종목에 대해 수집된 연속적인 주가변동자료를 연관성 규칙의 추출에 앞서 주가변동의 크기에 따라 등급으로 구분된 이산적 자료로 전환한다. 이 때 투자대상종목이라함은 실제 주식투자전략의 운용과정에서 매수 및 매도의 대상이 되는 종목을 의미하며, 참고종목은 투자대상종목과 사전적으로 연관성이 높을 것으로 예상되는 종목들을 의미한다. 실증분석시 참고종목은 사전적 정보가 없다면, 무작위 다수종목을 포함하여 적용할 수 있다. 또한, 실증적 의미를 높이기 위해서 투자대상종목은 참고종목에 비해 관측시점이 적어도 한 시점 이상 지연된 자료를 사용할 수 있다.

이제, 장바구니분석에 의한 연관성 규칙의 추출을 위한 기본 자료로는, 주가의 일별 자료를 사용할 경우, 매일 매일의 투



(그림2) 장바구니분석에 의한 주식투자전략 적용 절차

자대상종목 및 참고종목들의 주가변동의 등급값들이 한번의 거래에서 동시에 구매되는 상품들과 같이 인식되어 적용되며, 이를 이용하여 장바구니분석을 수행하면 투자대상종목 및 참고종목들간의 연관성 규칙(들)이 추출될 수 있다.

추출된 투자대상종목 및 참고종목들간의 연관성 규칙들로부터 주식투자전략을 도출하기 위해서는 우선 이들 종목들간에 "$X \Rightarrow Y$" 형태로 발생된 연관성 규칙들 중 참고종목들이 X에 표현되고 투자대상종목이 Y에 표현되는 연관성 규칙들만을 1차로 분류한다. 이에 따라 투자대상종목 및 참고종목들을 통해 얻어진 모든 "$X \Rightarrow Y$" 형태의 연관성 규칙들 중 Y에 투자대상종목이 포함된 연관성 규칙들만 남게되며, Y에 투자대상종목이 포함되지 않는 연관성 규칙들은 폐기한다.

다음으로 1차 분류된 연관성 규칙들에 대해 (식 1)에서와 같이 정의되는 연관성 규칙 선정 기준인 Support, Confidence, Lift값을 고려하여 평가함으로써 최종 주식투자전략으로 사용하게 될 규칙들만을 남기는 과정을 거친다. 최종 주식투자전략은 투자대상종목 변동의 이산적 등급 각각을 의미있게 설명할 수 있는 참고종목의 변동에 관한 연관성 규칙들로 정리되어, 이를 근거로 참고종목의 변동으로부터 투자대상종목의 매수 및 매도 의사결정을 지원하게 된다.

마지막으로, 주식투자전략 적용구간에서 매시점 매시점 관찰되어지는 참고종목의 변동이 주식투자전략에 포함된 연관성 규칙과 같이 발생되는지를 확인하여 투자대상종목에 대한 매수 및 매도 신호를 포착함으로써 주식투자전략을 적용할 수 있다.

4. 실증분석

4.1. 개요

본 연구에서는 앞에서 제시된 장바구니분석에 의한 주식투자전략 수립의 실증적 유용성을 검증하고자, 실제 주식변동 자료를 이용하여 연관성 규칙을 추출하고 그에 따른 주식투자전략을 수립·운영하였다. 본 연구에서 제시된 방법을 투자자들이 실제 주식투자전략에 활용하기 위해서는 증권거래소 및 코스닥 상장 주식 전체에 대한 가능한 많은 시계열을 이용하여 가장 효과적인 연관성 규칙을 추출하여 주식투자전략에 적용하는 것이 타당하나, 본 연구에서는 제시된 방법의 유용성을 검증하고자 하는 차원에서 실험적 수준의 가장 일반적인 실증분석 중 하나를 선정하고자 한다.

이에 따라, 본 연구의 실증분석에 이용된 자료는 2000년 1월부터 2001년 12월까지 2년동안 우리나라 증권거래소에서 발생된 일별 주가변동자료(종가 기준)이다. 투자대상종목으로는 종합주가지수를 참고종목으로는 2000년 1월이후 일관되게 주가자료를 관측할 수 있는 증권거래소 상장기업들 중 시가총액상위 10개종목(2002년 3월 기준)들을 선정하였다. 투자대상종목으로 종합주가지수를 선정한 것은 종합주가지수 편입종목들에 대해 구성비로 가중치를 부여한 포트폴리오로서 구성된 인덱스투자를 가정한 것이며, 참고종목에는 삼성전자, SK텔레콤, 한국통신공사, 한국전력, 포항제철, 현대자동차, LG전자, 삼성전기, 기아자동차, 조흥은행 등 10개종목이 포함되었다.

실증분석은 수집된 자료들을 분기별로 나누어 총 8개 구간으로 구분하여 각 분기마다 독립적으로 연관성 규칙을 추출하여 투자전략으로 적용하는 형태로 반복적으로

수행되었다. 이 때 분기로 구분된 8개 구간별로 앞의 2개월 자료는 연관성 규칙을 추출하기 위한 추정구간으로 사용하였으며, 나머지 1개월에 대하여 추정구간에 적용된 장바구니분석 결과 도출된 투자전략을 이용한 가상의 투자를 수행하는 것으로 하였다.

수행 결과 얻어진 주식투자전략의 운영 실적을 동일 구간의 주가변동율과 비교하였다.

4.2. 분석 절차

투자대상종목인 종합주가지수와 참고종목인 10개종목들에 대해 장바구니분석을 위해 각 종목별 일별종가가 전날의 종가에 비해 상승했는지, 하락했는지, 변동이 없는지 등 3개 등급으로 구분하여 기본자료로 사용한다. 이에 따라 장바구니분석을 위해 고려되는 기본입력자료를 전형적인 상품의 구매형태로 표현하면 <표 1>에서와 같이 나타난다.

<표 1>의 장바구니분석을 위한 입력자료 형태에서 종합주가지수의 변동은 실증적 의미를 높이기 위해 참고종목의 변동에 비해 관측시점을 한 시점 뒤로 설정하였다. 즉, <표 1>의 거래번호 "2000. 1. 5."의 동시구매품목이 "Y-, X1-, X2+, …"로 표시된 것은 종목 X1과 X2의 2000년 1월 5일 종가가 각각 2000년 1월 4일 종가에 비해 하락 및 상승하였으며, 종목 Y의 2000년 1월 6일 종가는 2000년 1월 5일 종가에 비해 하락하였음을 의미한다. 이에 따라 특정일에 발생된 참고종목의 변동이 다음날 발생할 종합주가지수의 변동과 연계된 연관성 규칙이 추출되게 된다.

<표 2>는 2000년 1분기의 추정구간(2000년 1, 2월)에 대해 <표 1>과 같이 입력된 자료로부터 추출된 연관성 규칙의 예를 보이고 있다. 실증분석에서 이 과정은 데이터마이닝 도구인 SAS Enterprise Miner의 연관성 분석(장바구니분석)을 이용하여 수행되었다.

<표 2>의 첫번째행에 나타난 연관성 규칙인 "X9- & X7+ ==> Y+"은 "종목 X9와 X7의 오늘 종가가 어제 종가에 비해 상승한 경우, 내일 종합주가지수의 종가는 오늘에 비해 상승할 것으로 기대된다."라고 해석될 수 있으며, 이와 같이 참고종목의 주가변동과 종합주가지수의 상승 및 하락에 대한 연관성 규칙들로 부터 종합주가지수에 대한 매수/매도 신호를 도출하여 이

<표 1> 주가변동에 대한 장바구니분석을 위한 입력자료 형태

거래번호	동시구매품목
2000. 1. 4.	Y-, X1+, X2+, X3+, X4+, X5-, X6-, X7-, X8-, X9-, X10-
2000. 1. 5.	Y-, X1-, X2+, X3-, X4+, X5+, X6+, X7-, X8-, X9-, X10+
. . .	. . .

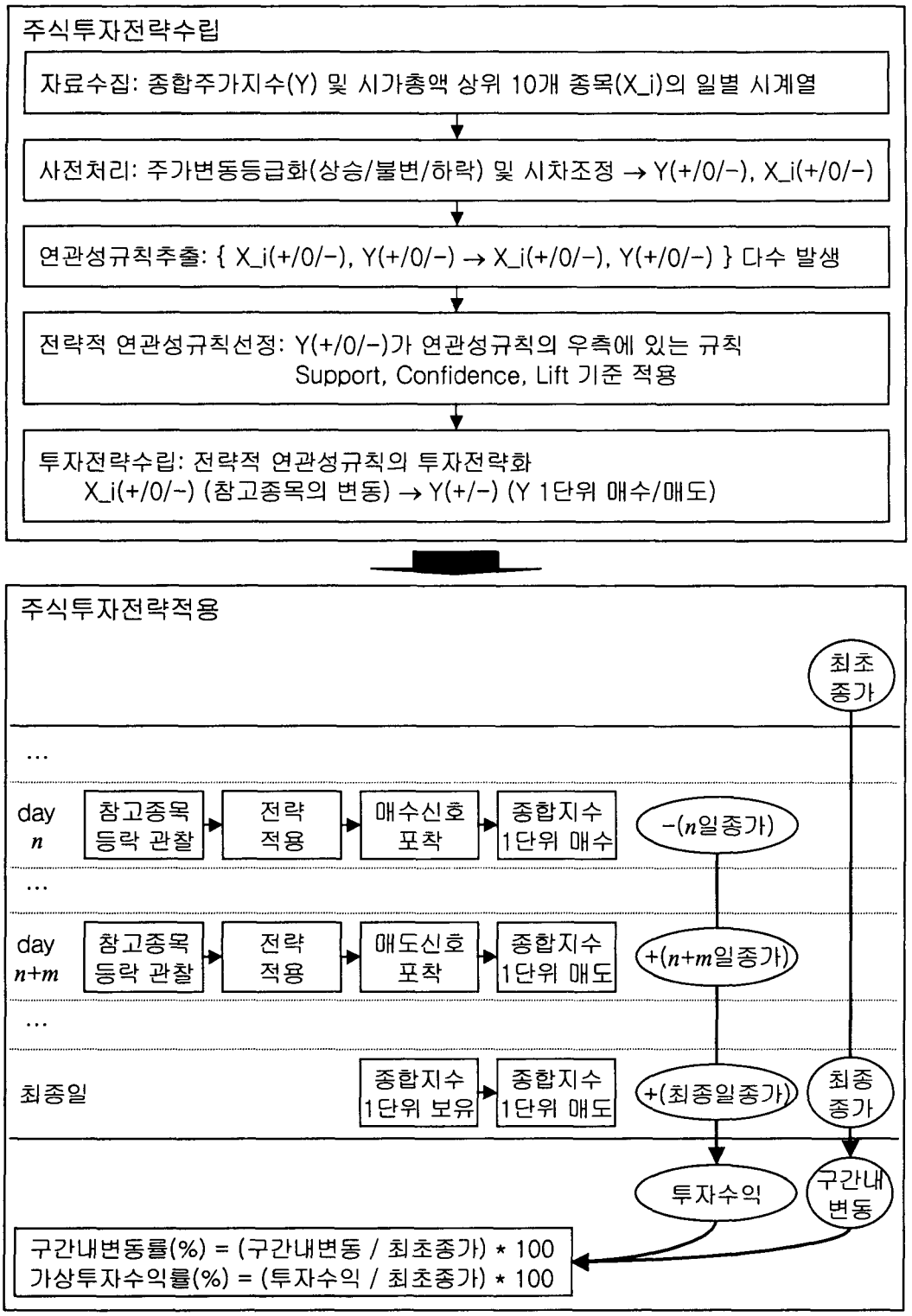
주) * 거래번호는 주가변동 해당일을 의미함
* 동시구매품목 중 Xi+, Xi0, Xi-는 각각 종목i의 해당일 종가가 전일 종가 대비 상승, 불변, 하락함을 의미함
* 동시구매품목 중 Y+, Y0, Y-는 각각 종합주가지수의 익일 종가가 해당일 종가 대비 상승, 불변, 하락함을 의미함

<표 2> 주가변동 연관성 규칙의 예

SUPPORT	CONF	LIFT	RULE
17.50	100.00	2.50	X9- & X7+ ==> Y+
17.50	100.00	2.50	X10- & X4- & X2- ==> Y+
12.50	100.00	2.50	X2- & X7+ ==> Y+
12.50	100.00	2.50	X10- & X9- & X7+ ==> Y+
12.50	100.00	2.50	X10- & X6- & X7+ ==> Y+
12.50	100.00	2.50	X10- & X5- & X7+ ==> Y+
12.50	100.00	2.50	X10- & X4- & X7+ ==> Y+
12.50	100.00	2.50	X10- & X2- & X7+ ==> Y+
12.50	100.00	2.50	X8+ & X7+ & X3+ ==> Y+
20.00	88.89	2.22	X10- & X7+ ==> Y+
17.50	87.50	2.19	X4- & X2- ==> Y+
17.50	87.50	2.19	X8+ & X7+ ==> Y+
15.00	85.71	2.14	X6- & X7+ ==> Y+
15.00	85.71	2.14	X5- & X7+ ==> Y+
15.00	85.71	2.14	X4- & X7+ ==> Y+

를 종합주가지수에 대한 투자전략을 활용하게 된다. 즉, 앞의 3.2.절에서 제시한 바와 같이 1차로 종합주가지수 및 10개의 참고종목들을 통해 얻어진 모든 "$X \Rightarrow Y$" 형태의 연관성 규칙들 중 Y에 종합주가지수가 포함된 연관성 규칙들만 남기고, 이어서 남겨진 연관성 규칙들의 유용성을 고려하여 투자전략에 활용할 연관성 규칙을 선정한다. 이 때 남겨진 연관성 규칙들 중에서 투자전략에 활용할 연관성 규칙은 연관성 규칙을 평가하는 기준으로 사용되는 (식 1)의 Confidence, Support, Lift값들을 고려하여 선정한다. 이때 투자전략에 활용할 최적의 연관성 규칙을 도출하기 위한 이론은 아직 개발되지 않았고, 다만 앞의 2장에서 언급한 바와 같이 분석자의 판단과 경험에 크게 의존하게 되며, 일반적으로 Confidence값을 최우선적으로 고려한다. 본 연구에서는 추출된 연관성 규칙들 중에서 Support와 Lift값이 일정한 값 이상인 연관성 규칙들 중에 Confidence값이 가장 높은 연관성 규칙들을 우선적으로 고려하여 종합주가지수 상승 및 하락에 관련된 연관성 규칙을 각각 3개씩 투자전략에 활용하였고, 이 과정에서 Confidence값이 같은 연관성 규칙들간의 우선순위는 다시 Support와 Lift값이 고려되었다.

이제, 이러한 과정을 거쳐 도출된 최종 투자전략은 가상투자구간에 다음과 같이 적용되었다. 가상투자구간의 각 종목별 자료가 매일매일 관측되는 것으로 가정하고, 종합주가지수 1단위를 미보유(보유)한 상태에서 매일매일 투자전략으로 채택된 연관성 규칙들을 적용하여, 그 결과 다음날 종



(그림 3) 연관성 규칙을 이용한 주식투자전략의 적용 사례

합주가지수의 상승(하락)에 대한 신호가 발생되면 종합주가지수 1단위를 매수(매도)한다. 이 때, 연관성 규칙의 각 참고종목의 주가와 종합주가지수의 매수 및 매도 가격은 해당일의 종가를 기준으로 한다. 이는 현재 우리나라 주식시장에서 적어도 폐장 1시간전에 종가가 확정된 이후 그가격으로 폐장때까지 거래가 지속되는 점을 반영한 것이다.

(그림 3)은 이상의 실증분석절차를 요약하여 제시하고 있다.

(그림 3)의 주식투자전략적용 시 "day n" 시점은 투자전략적용구간 중 최초로 매수 신호가 발생한 시점을 의미하며, "day n" 이전에 발생한 매도 신호는 무시된다. 이제 투자자는 "day n" 시점에 발생한 매수 신호에 따라 "day n" 시점의 종가를 기준으로 종합주가지수 1단위를 매수하여 "-(n일 종가)"의 현금흐름이 발생한다. 다음 매수이후 최초의 매도신호가 발생하는 "day $n+m$" 시점까지 추가적인 매수 신호는 무시되며, 투자자는 "day $n+m$" 시점의 매도신호에 따라 "day $n+m$" 시점의 종가를 기준으로 종합주가지수 1단위를 매도하여 "+($n+m$일 종가)"의 현금흐름이 발생한다. 투자전략적용구간의 최종일까지 이러한 매수 및 매도는 다수 반복될 수 있으며, 최종일 기준으로 종합주가지수 1단위를 보유하고 있다면, 이는 최종일 종가로 매도됨을 가정하여 "+(최종일 종가)"의 현금흐름이 발생한다. 이와 같이 발생된 현금흐름의 합으로 적용된 투자전략에 의한 투자수익을 계산할 수 있다.

4.3. 분석 결과

<표 3>은 본 연구에서 제시된 장바구니분석을 활용한 주식투자전략의 운영 실적을 나타내고 있다. <표 3>에서는 8개 구간으로 나뉘어 적용된 가상투자결과 발생된 구간별 수익률을 해당 구간에서의 종합주가지수 변동률과 함께 나타내고 있다. <표 3>의 p값은 이들 가상투자수익률과 구간내변동률이 통계적으로 유의한 차이를 보이고 있는가를 검증하기 위한 분산분석(ANOVA)의 결과이다.

<표 3>에서 보이는 바와 같이 실증분석을 수행한 총 8개 구간에서의 종합주가지수의 평균 변동률은 -2.38%인데 비해 본 연구에서 제시한 장바구니분석을 이용한 주식투자전략을 활용한 가상투자의 평균 수익률은 9.71%로 나타났으며, 가상투자수익률과 구간내변동률에 대한 분산분석 결과 얻어진 p값도 0.0018로 나타남으로써 본 연구의 결과가 통계적으로도 유의하게 평균적인 주가변동 이상의 투자수익을 발생시키고 있다고 할 수 있다.

<표 3>에서 구분된 8개 구간 각각에 대해서도 종합주가지수가 해당 1개월동안 급격히 증가한 2000년 6월을 제외하고는

<표 3> 연관성 규칙에 의한 주식투자전략의 적용 결과

구분	2000년도				2001년도				평균	p값
	3월	6월	9월	12월	3월	6월	9월	12월		
구간내 변동률(%)	-3.77	11.20	-11.41	-1.91	-6.47	-1.97	-11.47	6.74	-2.38	0.0018
가상투자 수익률(%)	10.22	5.10	10.01	7.93	6.12	8.33	12.72	17.22	9.71	

본 연구에서 제시한 투자전략의 수익률이 해당 구간의 변동률을 크게 상회함을 알 수 있다. 2000년 6월의 경우에는 해당구간 초기에 발생한 종합주가지수의 급격한 상승시에 매수 신호 포착을 위해 사용된 3개의 연관성 규칙 모두가 유효한 신호를 발생시키지 못한 때문이다.

5. 결론

본 연구에서는 데이터마이닝 기법 중 주로 마케팅 분야에서 널리 사용되고 있는 장바구니분석기법을 주식투자전략 수립에 활용하기 위한 방안을 개발하여 제시하고, 제시된 방안을 우리나라 주식시장의 실제 자료에 적용하여 실증적 유용성을 검증하였다. 이를 위해 본 연구에서는 3장에서 상품구매과정과 주가변동과정에서의 장바구니분석기법 및 그 결과 생성되는 연관성 규칙의 개념을 비교함으로써 장바구니분석기법이 주식투자전략에 활용될 수 있는 가능성을 탐색하고, 구체적으로 장바구니분석기법을 주식투자전략에 활용하기 위한 절차를 제시하였다. 이어서 4장에서는 2000년 1월부터 2001년 12월까지 우리나라 주식시장에서 관측된 종합주가지수 및 시가총액 기준 상위 10개 종목의 주가자료를 이용하여 3장에서 제시된 장바구니분석기법을 이용한 주식투자전략의 유용성을 검증하여 의미있는 결과를 얻었다. 부분적으로 본 연구는 주식투자전략을 수립함에 있어 참고종목들의 주가변동에 대해 상승·하락 등 간략한 이산적 정보만을 활용하여 효과적인 전략 수립을 지원하고 있음도 의미있는 결과이다.

또한, 본 연구는 정보기술의 발달은 다양한 분야에서 생성되고 있는 대규모의 데이터를 이용하여 의미있는 분석을 도출하고자 개발되어 사용되고 있는 데이터마이닝 기법의 새로운 효과적인 응용 방안을 제시하고 있다는 점에서도 의미를 찾을 수 있다. 즉, 본 연구에서는 이제까지 주로 인과관계탐색을 위한 주식시장에서의 데이터마이닝 연구에 인공신경망모형이 널리 사용되어 오고 있는데 반해, 일반적으로 마케팅 분야에서 구매상품들간의 연관성을 파악하기 위해 사용되고 있는 장바구니분석기법을 주식투자전략에 연계함으로써, 데이터마이닝 기법으로서의 장바구니분석기법의 응용 분야를 넓힐 수 있었으며, 주식시장 분석에 유용한 새로운 방안을 제시하고 있는 것이다. 이러한 관점에서 본 연구에서는 제시된 장바구니분석을 이용한 주식투자전략의 실증분석에 있어 간단한 전형적인 가상투자 결과의 투자수익률을 주가변동률과 비교하여 의미있음을 보였다. 이에 따라, 신경망모형 등 기존에 활용되어온 다른 데이터마이닝기법 및 재무관리이론에 근거한 이론적 모형 등 잠재적 비교 대상 모형과의 상대적 성과차이의 비교, 본 연구에서 제시한 방법 중에서도 최적의 투자전략에 연계될 연관성 규칙의 엄밀한 탐색 등은 향후의 연구과제로 미루고자 한다.

또한, 본 연구에서는 연관성 규칙의 개념을 이제까지 주로 인과관계에 기초하여 분석되어 온 주가변동예측문제에 적용하고 있다. 본 연구의 실증분석에서는 적용된 연관성 규칙 "$X \Rightarrow Y$"에서 X가 Y보다 관측시점이 앞선 상황에서 데이터 X, Y들의 패턴을 주식투자전략에 활용한 것이다. 본 연구에서 설정한 문제와 유사한 문제에 연관성 규칙을 보다 타당하게 적용하기 위해서는 데이터마이닝 기법으로서의 연관성 규칙과 통계적 또는 재무관리이론적 기법으

로서의 인과관계에 대한 체계적 개념 정립이 요구된다.

한편, 본 연구에서는 제시된 장바구니 분석을 이용한 주식투자전략을 보다 일반화시키기 위해서는 다양하게 발생되는 연관성 규칙들에서 전략수립에 적용할 최적의 규칙들을 선정하기 위한 기준, 서로 유사한 연관성 규칙들을 투자전략에 활용하기에 앞서 통합하여 반영하기 위한 방법 등에 관한 연구들이 추가되어야 할 것이다. 아울러, 본 연구에서는 제시된 장바구니분석을 이용한 주식투자전략을 종합주가지수에 대한 일별 투자전략에 응용하고 있으나, 향후 연구에서는 데이트레이딩 투자전략 등 보다 다양한 시점을 기준으로 하는 주식투자전략 및 종합주가지수 이외의 다양한 포트폴리오 및 개별 종목들에 대한 투자전략에도 그 가능성을 살펴볼 수 있을 것이다.

참고문헌

[1] 강현철, 한상태, 최종우, 김은석, 김미경, 「SAS Enterprise Miner를 활용한 데이터마이닝」, 자유아카데미, 1999.

[2] 김경태, 이장형, 최현호, "신경망에 의한 주가수익률 예측모형", 「기업경영연구」, 제14집, 2001, pp.127-142.

[3] 김종우, 이경미, "인터넷 상점에서 개인화 광고를 위한 장바구니 분석 기법의 활용", 「경영과학」, 제17권, 제3호, 2000, pp.19-30.

[4] 박종엽, 한인구, "인공신경망을 이용한 종합주가지수 예측", 「한국전문가시스템 '95 추계학술대회논문집」, 1995, pp.359-371.

[5] 오경주, 김경재, 한인구, "주가지수 예측에서의 변환시점을 반영한 이단계 신경망 예측모형", 「경영정보학연구」, 제11권, 제4호, 2001, pp.99-111.

[6] 이건창, 정남호, "데이터마이닝 기법과 지능형 에이전트 기법을 결합한 인터넷 쇼핑몰의 설계 및 구현에 관한 연구", 「정보기술응용연구」, 제1권, 제2호, 1999, pp.113-137.

[7] 이건창, 「PC를 활용한 회계정보시스템」, 무역경영사, 2000.

[8] 조성준, "데이터마이닝과 산업공학", 「IE매거진」, 제8권, 제1호, 2001, pp.58-60.

[9] 지원철, 박시우, 신현정, 신홍섭, "Cascade-Correlation Network를 이용한 종합주가지수 예측", 「한국경영과학회/대한산업공학회 '96 춘계공동학술대회논문집」, 1996, pp.745-748.

[10] Allen, C., D. Kania and Y. Beth, *Internet World Guide to One-to-One Web Marketing*, John Wiley & Sons Inc., New York, 1998.

[11] Berry, M.J.A. and G. Linoff, *Data Mining Techniques for Marketing, Sales, and Customer Support*, Wiley Computer Publishing, New York, 1997.

[12] Engle, R.F. and S. Kozicki, "Testing for Common Features," *Journal of Business & Economic Statistics*, Vol.11, No.4, 1993,

pp.369-380.

[13] Kimoto, T., K. Asakawa, M. Yoda, and M. Takeoka, "Stock market prediction system with modular neural network," *Proceedings of the I. Joint Conference on Neural Networks*, San Diego, California, 1990, pp.1-6.

[14] Ross, S. A., "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, December 1976, pp.343-362.

[15] Stair, R.M. and G.W. Reynolds, *Principles of Information Systems*, ITP, 1999.

[16] Zellner, W., "A Grand Reopening for Wal-Mart," *BusinessWeek*, February 9 1998.

■ 저자소개

주영진

연세대학교 경제학과에서 학사, 한국과학기술원 경영과학과에서 석·박사를 수여받고, 한국전자통신연구원 선임연구원/팀장을 거쳐 2001년 3월부터 배재대학교 경영정보학부에 재직중이다.