

한국종합주가지수 변동 경향에 대한 확률적 예측 시스템

강병우^o 한동수
한국정보통신대학교
{pitagi^o, dshan}@icu.ac.kr

A Probabilistic Forecasting System on the Tendency of Variation of Korea Composite Stock Price Index

Byeongwoo Kang^o Dongsoo Han
Information & Communications University

요 약

본 논문에서 기술하는 연구는 한국종합주가지수(KOSPI)의 장기적 변동 경향에 대한 확률적 예측 시스템을 제안한다. 제안된 방법론은 이미 단백질 상호작용 예측 시스템과 스트레스 확률 예측 시스템 등에 적용되어 유효성이 입증된 방법으로, 이미 알려진 데이터를 바탕으로 다양한 요인들의 가능한 모든 조합에 대한 경우의 수를 고려한 학습 결과에 기반하여 새로이 주어진 대상의 요인들을 분석해서 학습시 사용된 특정 군(class)에 속할지의 여부를 확률적으로 나타내준다. 이 방법론을 구현하기 위해 실제 과거 주가지수 데이터를 수집하여 CI(Combination Interrelation)행렬을 구현하였으며, 현재 진행중인 검증작업에 대해서도 기술하였다.

1. 서 론

주식시장은 자본주의 사회의 상징이라고 할 수 있다. 수많은 기업들이 주식 시장에서 자금을 확보하여 영업 활동을 통해 이익을 내고 있으며, 많은 투자자들 또한 주식 시장을 통해 시세차익이나 배당 이익을 얻기를 바라고 투자하고 있다. 본 논문은 한국 주식시장의 대표적 지수인 한국종합주가지수(KOSPI: Korea Composite Stock Price Index)의 변동 경향, 즉 저점과 고점의 예측을 시도하는 통계적 시스템을 제안한다.

주가지수는 수많은 사람들의 관심사이기 때문에 여러 분야에서 주가지수를 예측하는 방법을 개발하려는 연구와 노력이 있었다. 대부분이 단기적 예측을 목표로 했으며, 2장에서 이중 몇가지를 설명한다. 본 논문이 제안하는 변동 경향 예측 시스템은 수개월 이상의 기간을 두고 보는 장기적인 예측을 목표로 하고 있으며, 3장에서 이론적 배경과 확률적 방법론을 설명하고 4장에서는 수집한 데이터의 모델링 방법과 구축중인 학습시스템에 대한 설명 및 향후 검증 계획에 대해 기술한다. 5장은 종합적인 결론이다.

2. 관련 연구

약 5~6년 전부터 이른바 재테크에 대한 대중들의 관심이 커지면서 부동산과 함께 주식시장이 각광을 받게 되었고, 그에 따라 주가지수에 대한 관심이 높아지고 있다. 학계에서는 주가지수를 예측하는 방법에 대한 연구가 경제학, 수학, 통계학, 전산학 등 여러 분야에서 연구되어왔다. 본 연구에서는 관련된 연구내용으로 국내에서 한국종합주가지수 예측을 시도하거나 그에 관한 방법론을 기술한 자료를 주로 찾아보았는데, 여러 가지 수학적 예

측 모형에 대한 비교 연구[1]와, 순환 신경망 및 카오스 이론을 도입한 연구[2] 등을 찾을 수 있었다. 특히, 부산대학교 석사논문으로 나온 김해균의 연구[3]는 지식정보에 기초한 퍼지시스템(Fuzzy System)을 한국주가지수 200(KOSPI200) 예측에 도입하였으며, 이를 다층 신경회로망(MLP)을 통해 예측한 결과와 비교하기도 하였다.

이러한 연구들은 신경망(Artificial Neural Network), 카오스(Chaos) 이론 등 기존에 활발히 연구되어온 수학적, 전산학적 방법들을 주가지수 예측에 적용시켜 보았다는 데에 의의가 있다고 할 수 있다. 이들은 주로 일별 수익률 등 단기적인 주가 변동을 예측하는 것을 목표로 하였는데, 여러 가지 방법들을 도입하여 정확성을 다소 높인 시도들도 있었으나 전반적으로 예측의 정확성은 좋지 못하였다.

3. 확률적 예측 시스템

3.1 이론적 배경 및 기본 가정 사항

주가지수 변동이나 상승-하강 패턴을 분석하는 방법론에는 여러 가지가 있으나, 널리 알려진 이론으로 한시대를 풍미했던 유럽의 투자 전문가인 앙드레 코스톨라니(Audre Kostolany)의 '코스톨라니의 달걀 이론'이 특히 유명하다. 이 이론에 따르면 주식 시장은 장기적인 관점에서 보면 언제나 상승과 하강을 반복하며, 그 패턴을 분석해보면 6단계의 국면으로 나눌 수 있다고 한다. 이 이론은 각 구면별로 지수와 거래량의 변동을 명시하여 기술적인 분석으로 향후 지수가 어떻게 변동할지를 가능할 수 있게 한다. 본 연구는 이 코스톨라니의 달걀 이론을 그대로 따르지는 않았지만, 각 국면에 따라 지수와 거래량의 패턴이 다르므로 기술적 분석이 가능하다는 근거로 삼았다.

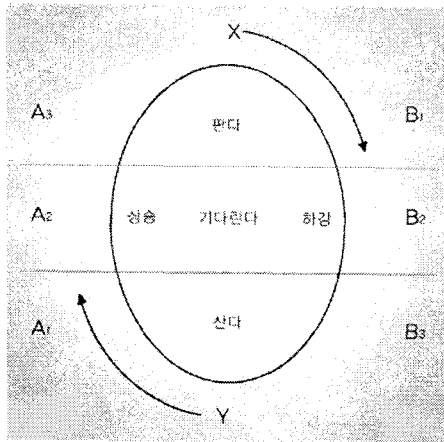


그림 1. 코스톨라니의 달걀 이론 모형 (화살표는 지수의 상승과 하강을 나타냄)

표 1. 그림1의 각 국면별 거래량의 변화

A1(수정국면)	거래량 적응.
A2(동행국면)	거래량이 증가함.
A3(과장국면)	거래량이 폭증함.
B1(수정국면)	거래량이 감소함.
B2(동행국면)	거래량이 증가함.
B3(과장국면)	거래량이 증가함.

본 연구에서 실험적으로 구축해본 시스템은 앞서 설명한 이론에 따라 지수 변동과 거래량에 관한 정보를 분석한다. 지수 변동은 널리 알려진 삼중 교차법(Triple Cross Method)를 이용하여 분석하였고, 거래량은 투자자별 순매수량을 분석하였다. 또한, 단기적인 예측이 아니라 최소 수개월 이상의 추이를 보는 장기적인 예측을 목표로 하였다.

3.2 전체적 구조

본 연구에서의 확률적 예측 시스템은 저자들이 속해있는 연구실에서 수행해온 스트레스 확률 예측 시스템 구축에서 사용했던 방식과 아이디어를 상당부분 차용하였다. 이 기법은 또한 도메인 정보에 기반한 단백질 상호작용 예측 시스템에도 적용되어 큰 효과를 보았다.[4]

본 시스템은 크게 두 단계로 나뉜다. 첫 번째 단계에서는 기존의 과거 지수 데이터를 입력으로 삼아 학습을 수행하고, 두 번째 단계에서는 학습한 내용에 기반하여 예측을 수행한다. 저점과 고점을 예측할 수 있다면 투자 수익을 기대할 수 있기 때문에 그 두 시점에 대한 데이터를 학습하고 고점을 예측하는 것을 목표로 하였다.

학습단계는 다시 두 가지 세부 과정으로 나뉜다. 첫 과정에서는 앞 절에서 설명한 이론에서 A1/B3국면과 A3/B1국면, 즉 저점인 국면과 고점인 국면의 데이터를 분석하여 지수변동과 투자자별 거래량간의 상호관계 패턴의 출현 빈도수를 구해서 CI(Combination Interrelation) 행렬에 저장한다. 따라서 이 과정에서는 고점 데이터와 저점 데이터의 행렬 즉, 두 가지의 CI 행렬

이 준비된다. 두 번째 과정에서는 두 가지의 CI 행렬을 이용하여 SCCA (Situation Characteristic Combination Appearance) 행렬을 생성한다. SCCA 행렬의 각 원소들은 고점일 때에 특정한 지수 추이(이동평균선 간의 상호 위치)와 투자자별 거래 특성이 출현하는 확률을 나타낸다. 지수 추이와 투자자별 거래 특성을 어떻게 나타내는지는 4장에서 설명하겠다.

예측단계에서는 SCCA 행렬을 기반으로 하여 확률식이 고안된다. 확률 식은 고점 데이터에서 얻어진 지수 추이와 거래 특성을 바탕으로 하여 데이터가 주어질 한 대상이 고점에 속하는 확률을 계산한다.

3.3 학습단계

학습단계에서 SCCA행렬 생성을 위하여 IA(Interrelation Appearance) 행렬과 CI 행렬들이 우선적으로 생성되어야 한다.

3.3.1 IA(Interrelation Appearance) 행렬

실제 시스템에서는 IA행렬을 만들 필요가 없으나, CI 행렬에 이해를 돕기 위하여 IA행렬에 대해서 먼저 설명한다. IA행렬에서 열(column)에서는 각 지수별 추이(이동평균선 간의 상대적 위치)의 종류를 나타내고 행(row)에서는 투자자별 거래행동특성을 나타낸다. n 가지의 서로 다른 지수들 $\{i_1, i_2, \dots, i_n\}$ 이 있고, 각각의 지수가 k 개의 서로 다른 추이를 가질 수 있다고 할 때, IA행렬의 열에 있는 원소들은 $\{i_{11}, i_{12}, \dots, i_{1k}, i_{21}, i_{22}, \dots, i_{2k}, \dots, i_{n1}, i_{n2}, \dots, i_{nk}\}$ 와 같이 표현된다. (KOSPI를 예측하기 위해 산업별 지수들도 함께 분석하며, 4장에서 보다 자세히 설명할 것이다.) 같은 방법으로, m 가지의 서로 다른 투자자들이 있고, 각각의 투자자가 l 개의 서로 다른 행동(매수, 매도, 관망)을 할 수 있다고 할 때, IA행렬의 행에 있는 원소들은 $\{t_{11}, t_{12}, \dots, t_{1l}, t_{21}, t_{22}, \dots, t_{2l}, \dots, t_{m1}, t_{m2}, \dots, t_{ml}\}$ 와 같이 표현된다. 결과적으로, 위의 예에서의 IA행렬의 크기는 $(n \times k) \times (m \times l)$ 이 된다.

한가지 예로 고점 데이터와 저점 데이터의 IA행렬을 생성하면서 설명하겠다. 먼저 IA행렬 생성을 위하여 각각 두 가지의 지수 $\{i_1, i_2\}$ 와 투자자 $\{t_1, t_2\}$ 이 있다고 가정하고, 지수와 투자자는 각각 두 가지의 종류(지수의 경우 추이, 투자자의 경우 행동)가 있다고 본다. 따라서, $\{i_{11}, i_{12}, i_{21}, i_{22}\}$ 과 $\{t_{11}, t_{12}, t_{21}, t_{22}\}$ 이 IA행렬의 열 원소와 행 원소로써 구성된다. 결과적으로, 4×4 IA행렬이 생성이 된다.

더 나아가서 $\{p_1, p_2, p_3\}$ 를 고점 데이터에 포함되는 대상들의 집단이고, $\{p_4, p_5, p_6\}$ 를 저점 데이터에 포함되는 대상들의 집단이라고 가정한다. 또한, 각 대상들을 다음과 같은 <지수, 투자자> 쌍(Pair)들로 표현된다고 보면 고점 데이터의 대상들은

- Pair (p_1) = $\{ \langle i_{11}, t_{11} \rangle, \langle i_{11}, t_{22} \rangle, \langle i_{22}, t_{11} \rangle, \langle i_{22}, t_{22} \rangle \}$
- Pair (p_2) = $\{ \langle i_{11}, t_{11} \rangle, \langle i_{11}, t_{21} \rangle, \langle i_{21}, t_{11} \rangle, \langle i_{21}, t_{21} \rangle \}$
- Pair (p_3) = $\{ \langle i_{12}, t_{12} \rangle, \langle i_{12}, t_{22} \rangle, \langle i_{22}, t_{12} \rangle, \langle i_{22}, t_{22} \rangle \}$

와 같은 쌍들을 가지고, 저점 데이터의 대상들은
 Pair (p_4) = {< i_{12} , t_{12} >, < i_{12} , t_{21} >, < i_{21} , t_{12} >, < i_{21} , t_{21} >}
 Pair (p_5) = {< i_{12} , t_{12} >, < i_{12} , t_{21} >, < i_{22} , t_{12} >, < i_{22} , t_{21} >}
 Pair (p_6) = {< i_{11} , t_{11} >, < i_{11} , t_{22} >, < i_{21} , t_{11} >, < i_{21} , t_{22} >}
 와 같은 쌍들을 가진다고 가정한다.

또한 $IA(K)_{ij}$ 원소는 집단 K의 지수와 투자자간의 쌍 < i_j , t_j >을 나타낸다. 다시 말하면, 행렬의 각 원소들은 집단 K에서의 쌍들에 출현 빈도수를 값으로 가진다. 여기서 K는 고정 데이터 또는 저점 데이터를 뜻한다. 그림 2에서는 이 예로부터 얻어진 IA행렬을 보여준다.

		지수 추이						지수 추이			
		i_{11}	i_{12}	i_{21}	i_{22}			i_{11}	i_{12}	i_{21}	i_{22}
투자자별 거래행동	t_{11}	2	0	1	1	투자자별 거래행동	t_{11}	1	0	1	0
	t_{12}	0	1	0	1		t_{12}	0	2	1	1
	t_{21}	1	0	1	0		t_{21}	0	2	1	0
	t_{22}	1	1	0	2		t_{22}	1	0	1	1

그림 2. 고정점과 저점 데이터에 대한 IA행렬

3.3.2 CI(Combination Interrelation) 행렬

IA행렬은 고정 데이터와 저점 데이터에서 나타나는 지수 추이와 투자자별 거래 행동 간의 쌍들의 정보를 저장한다. CI행렬은 보다 세밀하고 정확한 확률적 예측을 위해 생성되는데, IA행렬을 확장하여 각 지수 추이들의 조합과 투자자별 거래행동특성들의 조합의 개념을 도입한다. 고정 또는 저점인 상황에서 당연히 여러 지수들이 다양한 종류의 추이를 나타내고 각 투자자들도 다양한 종류의 행동을 보일 것이다. 그러한 종류의 모든 조합을 고려하기 위해 CI행렬을 만드는 것이다. IA행렬에서 열(column)에 지수추이 조합을, 행(row)에 투자자별 거래행동특성 조합을 각각 추가하여 CI행렬을 생성한다.

예를 들어, 지수추이 조합 { i_{11} , i_{21} }, { i_{11} , i_{22} }, { i_{12} , i_{21} }, { i_{12} , i_{22} }는 열에 추가하고, 거래행동특성 조합 { t_{11} , t_{21} }, { t_{11} , t_{22} }, { t_{12} , t_{21} }, { t_{12} , t_{22} }는 행에 추가할 때, 각 원소들은 IA행렬에서와 같은 방법으로 값이 결정된다. 결과적으로 8 × 8 CI행렬이 생성된다. 그림 3에서는 CI행렬을 보여준다.

3.2.3 SCCA(Situation Characteristic Combination Appearance) 행렬

두 데이터 집합의 CI행렬이 생성이 되면 SCCA행렬의 생성이 가능하다. SCCA행렬의 각 원소는 특정 지수추이

	i_{11}	i_{12}	i_{21}	i_{22}	{ i_{11} , t_{21} }	{ i_{11} , t_{22} }	{ i_{12} , t_{21} }	{ i_{12} , t_{22} }
t_{11}	1	0	1	0	0	0	0	0
t_{12}	0	2	1	1	0	0	1	1
t_{21}	0	1	0	1	1	0	1	1
t_{22}	1	0	1	1	0	0	0	0
{ t_{11} , t_{21} }	1	0	1	0	1	0	0	0
{ t_{11} , t_{22} }	1	0	0	1	0	1	0	0
{ t_{12} , t_{21} }	0	0	0	0	0	0	0	0
{ t_{12} , t_{22} }	0	1	0	1	0	0	0	1

그림 3. 고정점과 저점 데이터에 대한 CI행렬

조합과 거래행동특성 조합 간의 쌍 고정 데이터 집합에서 출현하는 가능성에 대한 확률 값을 가진다. SCCA행렬의 각 원소의 값은 다음의 식에 의해서 계산이 된다.

$$SCCA_{ij} = \frac{CI(\text{고점})_{ij}}{CI(\text{고점})_{ij} + CI(\text{저점})_{ij}}$$

SCCA행렬의 각 열과 행은 CI행렬과 동일하며, 앞에서 예를 든 경우를 이용하면 다음과 같은 SCCA행렬을 얻게 된다.

	i_{11}	i_{12}	i_{21}	i_{22}	{ i_{11} , t_{21} }	{ i_{11} , t_{22} }	{ i_{12} , t_{21} }	{ i_{12} , t_{22} }
t_{11}	0.67	0	0.5	1	1	1	0	0
t_{12}	0	0.33	0	0.5	0	0	0	0.5
t_{21}	1	0	0.5	0	0.5	0	0	0
t_{22}	0.5	1	0	0.67	0	1	0	1
{ t_{11} , t_{21} }	1	0	1	0	1	0	0	0
{ t_{11} , t_{22} }	1	0	0	1	0	1	0	0
{ t_{12} , t_{21} }	0	0	0	0	0	0	0	0
{ t_{12} , t_{22} }	0	1	0	1	0	0	0	1

그림 4. 그림3에서의 CI행렬에서 나온 SCCA행렬

3.4 예측단계

예측 단계의 첫 번째 과정은 각 대상들을 위한 가중치(weight)를 계산하고 SCAP행렬과 가중치를 이용하여 각 대상이 고정점에 속하는 정도를 판단하는 확률식을 고안한다.

3.4.1 대상별 가중치

본 시스템에서 각 지수추이들과 각 거래행동특성들은 균일한 가중치를 가진다. 예를 들어, n개의 지수가 있고 각 지수별 k종류의 추이가 있으면 면 각 지수추이별 가중치는 1/nk이 되고, m가지의 투자자 종류들이 있고 각 투자자 종류별로 l개의 거래행동특성이 있으면 각 거래행동특성별 가중치는 1/ml이 되는 것이다.

SCCA행렬의 원소가 행렬의 열과 행의 가중치로 weight(x), weight(y)의 값을 가질 때, 원소 $SCCA_{xy}$ 의 가중치 값은 weight(x)와 weight(y)의 평균값이 된다.

3.4.2 고점 여부 판단을 위한 확률 식

이번 과정에서는 한 대상이 고점 데이터 집합에 속하는 확률을 계산하는 확률 식을 제안한다. 확률 식은 앞에서 설명하였던 SCCA행렬과 행렬의 각 원소별 가중치를 기반으로 생성된다. 따라서, $p \times q$ SCCA행렬이 준비되어 있다고 가정하고 $ic(p_i) = \{ic_1, ic_2, \dots, ic_p\}$ 를 p_i 라는 대상에서 얻은 지수추이 조합이라고 보고

$$tc(ic_j) = \{tc_{j,x_1}(1), tc_{j,x_1}(2), \dots, tc_{j,x_1}(k)\}$$

$$(1 \leq j \leq p, 1 \leq x_j(k) \leq q)$$

를 p_i 라는 대상에서 얻은 거래행동특성 조합이라고 본다. 여기서 SCAP 행렬과 WA행렬은 같은 행과 열의 원소를 공유한다는 것을 유념해야 한다. 그러면, 한 대상 p_i 가 조합 쌍 $CP(p_i)$ 를 가질 때 고점 판단 확률 식을 다음과 같이 정의할 수 있다.

$P(p \in \text{고점 집합} | CP(p)) =$

$$\sum_{u=1}^n \left[\frac{\text{weight}(ic_u)}{\sum_{v=1}^n \text{weight}(ic_v)} \cdot \left(\sum_{i=x_u(1)}^{x_u(k)} \frac{\text{weight}(tc_u) \cdot SCCA_{ic_u tc_u}}{\sum_{j=x_u(1)}^{x_u(k)} \text{weight}(tc_u)} \right) \right]$$

4. 시스템 구축 및 검증 계획

4.1 데이터 수집

경제 정보 사이트 이데일리(Edaily)[5]에서 제공하는 과거 주가지수 그래프 자료를 이용하여 주봉을 기준으로 저점, 고점이라 할 수 있는 부분들을 찾는 작업을 하였다. 아래 그림은 부분적인 예로, 저점과 고점에 해당하는 부분을 동그라미로 표시했다.

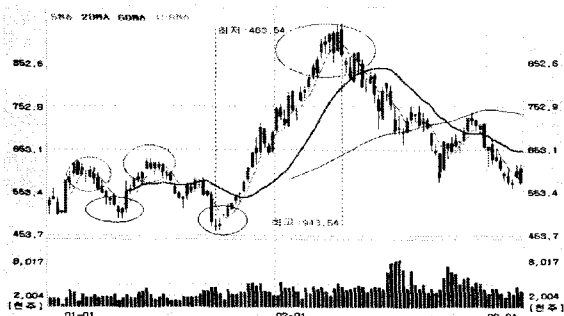


그림 5. 주봉 기준 KOSPI 지수 변화 그래프

위와 같은 방법으로 2000년 12월 초부터 현재(2006년 8월말)까지의 그래프를 분석하여 총 8개의 저점과 8개의

고점을 찾아내었다. 찾아낸 저점에서 증가가 가장 낮은 날과 고점에서 증가가 가장 높았던 날을 기준으로 그 날을 포함한 이전 10일간(거래일 기준)의 데이터를 증권 선물거래소(KRX: Korea Exchange) 홈페이지[6]에서 제공받아 DB에 입력하였다. 상세한 저점 및 고점의 일자 는 표 2에 기술하였다. 총 160일(8×10×2)의 데이터를 얻을 수 있었다.

표 2. 추출해낸 저점과 고점 일자

저점	고점
2000-12-11~2000-12-22	2001-02-07~2001-02-20
2001-03-27~2001-04-10	2001-05-25~2001-06-08
2001-09-12~2001-09-25	2002-04-10~2002-04-23
2002-09-27~2002-10-11	2002-11-20~2002-12-03
2003-03-04~2003-03-17	2004-04-09~2004-04-23
2004-07-26~2005-08-06	2005-02-25~2005-03-11
2005-04-18~2005-04-29	2006-01-03~2006-01-16
2006-02-23~2006-03-09	2006-04-14~2006-04-27

4.2 데이터 모델링 및 학습

주가지수 추이에 관한 데이터로는, KOSPI 지수 및 KOSPI200의 5가지 산업별 지수(제조업, 전기 및 통신업, 건설업, 서비스업, 금융업) 등 총 6가지 지수의 5일, 20일, 60일 이동 평균선을 DB에 포함된 날짜의 해당일 별로 계산하였다. 이동평균선(Average Moving Line)을 계산한 이유는 3중교차법(Triple Cross Method)을 적용하기 위함인데, 3중 교차법에 따라 다음과 같은 6가지 경우의 수가 나온다.

표 3. 3중 교차 모델 (*MA: moving average)[3]

상승	하락
20MA > 5MA > 10MA	10MA > 5MA > 20MA
5MA > 20MA > 10MA	10MA > 20MA > 5MA
5MA > 10MA > 20MA	20MA > 10MA > 5MA

따라서, 각 6가지 지수는 위 6가지 중 하나의 지수 추이를 가지게 된다.

투자자의 종류는 외국인, 기관, 개인, 기타 등 4종류로 나누었으며, 순매도 규모가 100만주 이상이면 매수, -100만주 미만이면 매도, 그 외에는 관망으로 거래행동 특성을 구분하여 역시 DB에 포함된 날짜의 해당일별로 정해두었다. 따라서, 각 4종류의 투자자는 3가지(매수, 매도, 관망) 중 하나의 거래행동을 가지게 된다.

각 지수별 추이들은 3장에서 설명한 이행렬과 SCCA 행렬의 열(column)을 구성하고, 각 투자자별 거래행동특성은 행(row)을 구성한다.

4.3 검증 계획 및 예상 결과

이행렬에서 6가지 지수가 각각 6가지 경우의 수를 가지므로 6⁶개의 열이 생기며, 4종류의 투자자가 각각 3가지 경우의 수를 가지므로 4³개의 행이 생긴다. 그러므로 이행렬의 총 원소 수는 6⁶ × 4³ 개가 되는 것이다.

이행률이 엄청나게 많은 원소를 가지게 되는 까닭에 당초 예상했던 것보다 학습에 오랜 시간이 소요되어 아직 시스템에 대한 검증은 마치지는 못하였다. 현재 계획에 따르면, 고점 8개중 7개 및 저점 8개중 7개의 데이터를 학습을 위해 사용하고, 남은 저점 1개와 고점 1개의 데이터를 검증을 위해 사용할 예정이다. 3장에서 기술한 고점 판단 확률 식에 따라 확률이 60% 이상으로 나오면 고점으로, 40% 이하로 나오면 저점으로, 40%~60% 사이의 값으로 나오면 고점도 저점도 아닌 것으로 판단할 예정이다.

검증을 위해 사용할 20일(2×10)간의 데이터 중에 최소 60% 이상이 정확히 분류될 것으로 현재 기대하고 있다.

5. 결론

본 연구에서는 장기적인 종합주가지수 변동 경향 예측에 대한 통계적인 방법론 및 시스템을 제안하였다. 시스템은 크게 학습단계와 예측단계로 나뉘며, 학습단계에서는 저점과 고점에서 지수의 변동 추이 및 투자자별 거래 행동경향의 조합을 고려한다. 예측단계에서는 학습된 내용을 바탕으로 주어진 대상이 고정일 확률이 얼마나 되는지를 계산하여 결과로 보여준다. 이러한 확률적 예측 기법은 저자들의 소속 연구실에서 이미 진행한 바 있는 스트레스 확률 판단 시스템 및 단백질 상호작용 예측 시스템에서 많은 부분을 차용하여 주가지수 변동 경향 예측에 응용해본 것이다.

아쉽게도 본 시스템의 검증은 아직 진행 중에 있어 확실히 끝나지 않았다. 4장에서 제시한 검증 계획 또한 충분하다고는 할 수 없을 것이다. 이미 DB에 있는 데이터 외에 다른 데이터를 이용한 검증 또한 필요할 것이다. 검증에서 일정 수준 이상의 유효성을 확인하게 되면 항목별로 가중치를 다르게 주는 등의 최적화를 실시하여 예측 확률을 높일 수 있는 연구도 시도해볼만하다.

6. 참고문헌

- [1] 박희숙, 한국종합주가지수 예측모형에 대한 비교연구, 포항공과대학교 대학원 수학과(통계학전공) 석사학위논문, 2003
- [2] 정규원, 카오스 특징을 이용한 주가지수 예측에 관한 연구, 광운대학교 전산대학원 컴퓨터공학과 석사학위논문, 1999
- [3] 김해균, 지식정보에 기초한 퍼지시스템을 이용한 주가지수 예측, 부산대학교 대학원 전기공학과 석사학위논문, 2001
- [4] Dong-Soo Han *et al.*, "PreSPI: a domain combination based prediction system for protein-protein interaction", *Nucleic Acids Research*, Vol.32, No.21, 2004.
- [5] 이데일리, <http://www.edaily.co.kr>
- [6] 증권선물거래소(KRX: Korea Exchange) 홈페이지, <http://www.krx.co.kr>