

퍼지 모델을 이용한 일별 주가 예측

황 희 수[†]

요 약

본 논문에서는 주가의 일별 시가, 종가, 최고가, 최저가를 예측하기 위한 퍼지모델을 제안한다. 주가는 시장의 여러 경제 변수에 의존하므로 주가예측 모델의 입력변수를 선택하는 것은 쉽지 않은 일이다. 이와 관련하여 많은 연구가 있지만 정답이 있는 것은 아니다. 본 논문에서는 이를 해결하기 위해 주가 움직임 자체에 주목하는 스틱차트의 기술적 분석에 이용되는 정보를 퍼지규칙의 입력변수로 선택한다. 퍼지규칙은 사다리꼴 멤버십함수로 이루어진 전진부와 비선형 수식의 후진부로 구성된다. 최적의 퍼지규칙으로 구성된 퍼지모델을 찾아내기 위해 차분진화가 사용된다. 본 논문에 제안된 방법은 수치 예를 통해 다른 방법과의 비교로 타당성이 검토되며 KOSPI(Korea composite Stock Price Index) 일별 데이터를 사용, 주가예측 퍼지모델을 구축하고 신경회로망 모델과 비교, 검토된다.

키워드 : 퍼지 모델, 시계열 데이터 예측, 차분 진화, 비선형 모델, 주가 예측

Daily Stock Price Prediction Using Fuzzy Model

Heesoo Hwang[†]

ABSTRACT

In this paper an approach to building fuzzy model to predict daily open, close, high, and low stock prices is presented. One of prior problems in building a stock prediction model is to select most effective indicators for the stock prediction. The problem is overcome by the selection of information used in the analysis of stick-chart as the input variables of our fuzzy model. The fuzzy rules have the premise and the consequent, in which they are composed of trapezoidal membership functions, and nonlinear equations, respectively. DE(Differential Evolution) searches optimal fuzzy rules through an evolutionary process. To evaluate the effectiveness of the proposed approach numerical example is considered. The fuzzy models to predict open, high, low, and close prices of KOSPI(Korea composite Stock Price Index) on a daily basis are built, and their performances are demonstrated and compared with those of neural network.

Keywords : Fuzzy Model, Time Series Prediction, Differential Evolution, Nonlinear Model, Stock Prediction

1. 서 론

주식시장에서 종합주가지수나 종목별 지수를 예측하는 것은 쉬운 일이 아니지만 주가 예측 모델을 만들기 위한 다양한 방법들이 시도되어 왔다. 이들은 미시적인 관점의 주식 관련 변수나 거시적인 관점의 국내외 경제 관련 변수들을 모델에 도입하였으며[4], 크게 4가지 부류로 나누어질 수 있다: i) 기술적 분석(technical analysis), ii) 기본적 분석(fundamental analysis), iii) 전통적인 시계열 분석(time series analysis)과 iv) 기계학습(machine learning). 기술적 분석은 대개 차트 분석으로 불리며 시장의 과거의 데이터를 기술하는 차트에서 만들어지는 특정한 패턴을 이용하여 주가를 예측한다[4]. 기본적 분석은 주식의 내재 가치를 분석하여 현재 가격이 높은지 낮은지를 결정한다. 시계열 분석은 과거

의 데이터에서 패턴을 추적할 수 있는 선형 예측 모델을 만들려고 한다[6,13]. 최근에 많이 사용되는 기계학습 방법은 일련의 샘플 데이터 집합을 사용하여 데이터에 근원이 되는 함수를 근사적으로 모방하는 모델을 만드는 것이다. 모델은 전문가시스템[5], 인공지능망[8-11], 결정트리[14], SVM(Support Vector Machine)[12]이나 퍼지규칙으로[7,20] 구성된다. 데이터마이닝 기법[15,16]을 사용한 경우도 있다. 이런 기법들은 과거의 수학적 모델이 복잡성으로 인해 사용이 쉽지 않은데 비해 단순하면서도 주목할만한 성능을 보이고 있다. 특히 신경회로망은 다수의 입력변수를 이용하면서도 간단하고 사용이 쉬워 인기가 있다. 그러나 신경회로망은 일반적으로 학습 정도와 중간 층(layer) 및 그 층에서의 노드 수에 매우 민감한 것으로 알려져 있다. 과도한 학습(over-fitting)은 오히려 모델의 신뢰성을 저하시키므로 적절한 종료조건과 적합한 중간 층과 층의 노드 수를 결정하는 일은 반복적인 시행착오를 요하는 단점이 있다.

본 논문에서는 퍼지모델 기반의 주가 예측 방법을 제안한다.

[†] 정 회 원 : 한라대학교 전기전자과 부교수
논문접수: 2008년 8월 26일
수정일: 1차 2008년 10월 1일
심사완료: 2008년 10월 2일

퍼지규칙을 이용한 모델링은 크게 두 부류로 나눌 수 있다: 주관적 모델링(subjective modeling)과 객관적 모델링(objective modeling). 주관적인 모델링은 시스템에 대한 사전 지식을 필요로 하며 그 같은 지식은 전문가의 경험에 의해 추출된다. 반면에 객관적인 모델링은 시스템에 대한 지식을 추출하기 위해 과거의 데이터를 사용하며 퍼지 모델링이라 불린다. 퍼지 모델링 방법은 기존의 수학적 방법에 비해 비선형 시스템을 기술할 수 있는 능력이 뛰어난 것으로 입증되었다[18,19]. 또한 다른 기계학습 방법에 비해 모델링 결과가 언어적으로 표현되기 때문에 그 자체로 시스템에 대한 이해를 높일 수 있다는 장점이 있으며 신경회로망처럼 과도 학습의 문제에 민감하지 않다. 본 논문에서 사용한 퍼지 모델은 주가의 변동성을 감안하여 사다리꼴 형태의 멤버쉽 수를 갖는 전건부와 비선형적인 후건부로 이루어진 퍼지 규칙들로 구성된다. 퍼지모델링에서 접하는 어려움 가운데 하나는 데이터에 내재하는 특성을 근사적으로 묘사할 수 있는 신뢰성 있는 퍼지규칙을 찾아내는 것인데, 차분진화(Differential Evolution)를 통해 최적의 퍼지규칙을 찾는 것으로 이를 해결한다. 제안된 방법의 타당성을 분석하기 위해 수치 예제가 다루어지고 성능이 비교, 검토된다. 제안된 방법으로 KOSPI(Korea composite Stock Price Index)의 일간 시가, 종가, 최고가, 최저가를 예측하기 위한 퍼지모델을 구축하고 학습에 사용되지 않은 데이터를 통해 모델의 성능을 검증한다.

2. 퍼지모델

MISO(Multi-Input and Single Output) 시스템의 퍼지모델을 구성하는 퍼지규칙 형식으로 식(1)을 사용한다. TSK형 퍼지모델의 후건부는 입력변수와 상수가 선형적으로 결합된 식으로 구성된다[18,19]. 주가 데이터는 변동성이 매우 큰 경우가 많으므로 변동성을 잘 추종할 수 있도록 하기 위해 본 논문에서는 후건부를 비선형식으로 구성한다.

$$R^i: \text{if } x_1 \text{ is } A_1^i(x_1) \text{ and } x_2 \text{ is } A_2^i(x_2) \dots x_k \text{ is } A_k^i(x_k) \\ \text{then } y^i = a_0^i + a_1^i \cdot x_1^{b_1^i} + \dots + a_k^i \cdot x_k^{b_k^i} \quad (1)$$

여기서 R^i 는 i 번째 규칙($1 \leq i \leq r$), $x_j(1 \leq j \leq k)$ 는 입력변수이고 y^i 는 i 번째 규칙의 출력변수이다. $A_j^i(x_j)$ 는 퍼지변수로 식(2)와 같이 정의된다. y^i 는 상수 a_j^i 와 입력변수 x_j 의 b_j^i 승을 선형 결합한 것이다. 퍼지규칙에서 a_k^i 와 b_k^i 는 차분진화를 통해 최적화 과정에서 찾아야 할 매개변수이다.

$$A_j^i(x_j) = \begin{cases} \frac{(x_j+d_{ij}+d_{ij}^i-c_{ij})}{d_{ij}^i} & \text{if } x_j > (c_{ij} - d_{ij} - d_{ij}^i) \\ 1 & \text{if } (c_{ij} - d_{ij}) \leq x_j \leq (c_{ij} + d_{ij}) \\ \frac{(-x_j+d_{ij}+d_{ij}^i+c_{ij})}{d_{ij}^i} & \text{if } (c_{ij} + d_{ij}) \leq x_j \leq (c_{ij} + d_{ij} + d_{ij}^i) \\ 0, & \text{그렇지않으면} \end{cases} \quad (2)$$

여기서 A_j^i 는 사다리꼴 형태의 멤버쉽함수이지만 최적화를 통해 얻어진 결과에서 $d_{ij} = 0$ 이면 대칭이거나 비대칭의 삼각형 형태로 변화될 수 있다. 퍼지규칙에서 c_{ij} , d_{ij} , d_{ij}^i 와 d_{ij}^i 는 차분진화를 통해 찾을 매개변수이다.

퍼지규칙에서 다음과 같은 추론 절차를 통해 출력이 계산된다.

- ① $X = \{x_1, \dots, x_i, \dots, x_n | x_i = (x_{i1}, x_{i2}, \dots, x_{ik}, y_{i1}, y_{i2}, \dots, y_{iq})\}$ 의 입력력 데이터가 주어지면 i 번째 퍼지규칙의 전건부 추종도인 w^i 는 식(3)과 같이 계산한다.

$$w^i = A_1^i(x_{i1}) \times A_2^i(x_{i2}) \times \dots \times A_k^i(x_{ik}) \quad (3)$$

여기서 x 는 최소값 연산자를 의미한다.

- ② w^i 와 관련하여 y^i 의 가중 평균치인 y_*^i 를 식(4)에서와 같이 계산한다.

$$y_*^i = \frac{\sum_{j=1}^r w^j \times y^j}{\sum_{j=1}^r w^j} \quad (4)$$

여기서 r 은 퍼지규칙 수이고 y^i 는 $a_0^i + a_1^i \cdot x_{i1}^{b_1^i} + \dots + a_k^i \cdot x_{ik}^{b_k^i}$ 이다.

3. 퍼지모델의 인식

퍼지모델의 인식(identification)은 입력력 데이터가 주어지면 그 데이터의 특성을 가장 잘 묘사할 수 있는 최적의 퍼지규칙을 찾아내는 것이다. 본 논문에서의 인식은 차분진화(DE)[17]를 통해 이루어진다. 진화를 시작하기 전에 퍼지규칙의 수를 먼저 결정하여야 한다. 차분진화에서는 매개변수의 값인 실수를 그대로 사용하므로 이진-유전알고리즘에서와 같이 개체 구성을 위해 이진수로 코딩하고 재결합과 돌연변이 연산 후 새로이 만들어진 개체를 평가하기 위해 다시 디코딩 하는 절차가 필요 없다. 이는 그 만큼 빠른 탐색을 할 수 있음을 의미한다. 차분진화에서는 탐색할 매개변수들에 대응해서 개체가 만들어진다. 무작위 값으로 다수의 개체를 만들어 집단을 구성하며, 이 때 개체 수를 집단 크기라 부른다. 세대를 거치면서 이 집단이 진화하며, 이 집단 가운데 적합도가 가장 높은 개체는 유지된다. 진화가 종료되면 적합도가 가장 높은 개체가 퍼지규칙을 구성하는 최적의 매개변수 값으로 사용된다. 퍼지규칙 수가 정의된 초기 퍼지모델로부터 최적의 퍼지모델로 진화하는 과정은 다음과 같다.

[단계 1] 무작위 값으로 초기 집단을 구성하고 세대 수 $t=0$ 으로 설정한다.

[단계 2] 이전 세대의 개체를 사용하여 새로운 개체로 구성된 집단을 형성한다. 이를 위해 식(5)에서와 같이 무작위로 선택된 3개의 개체($X_{r1,t}$, $X_{r2,t}$, $X_{r3,t}$)로 부터 다음 세대를 위한 차분벡터 $V_{i,t+1}$ 를 먼저 만든다.

$$V_{i,t+1} = X_{r1,t} + F \times (X_{r2,t} - X_{r3,t}) \quad (5)$$

여기서 $X_{i,t}$ 는 t 번째 세대의 집단에서 i 번째($i=1,2,\dots, NP$) 개체이다. r_1, r_2 와 r_3 는 $1 \sim NP$ (집단 크기) 사이에서 무작위로 선택되며 동일한 값을 갖지 않는다. 개체 X 는 퍼지규칙을 구성하는데 필요한 전건부와 후건부 매개변수들로 구성된다. F 는 두 개체의 차이를 조정하는 상수로 통상 $0 \sim 2$ 사이의 값을 갖는다.

새로이 만들어질 개체 $X_{i,t+1}$ 는 차분벡터 $V_{i,t+1}$ 와 $X_{i,t}$ 의 균일 교배를 통해 만들어진다. 균일교배는 개체를 구성하는 각각의 매개변수에 대해 무작위 수를 발생시켜 이 값이 교배율 보다 작으면 $X_{i,t}$ 의 매개변수를, 그렇지 않으면 $V_{i,t+1}$ 의 매개변수를 갖도록 한다.

[단계 3] 새로이 만들어진 집단의 개체에 대해 적합도를 평가하고 이전 세대의 개체 보다 적합도가 높으면 생존하며, 그렇지 않으면 이전 개체가 생존한다.

[단계 4] 종료조건을 만족하지 않으면 세대 수를 증가하고($t=t+1$) 단계 2로 복귀한다. 종료조건으로 적합도 목표치를 사용하거나 최대 진화 세대 수를 사용한다.

각 개체의 적합도를 평가하는데 사용되는 적합도 함수로 식(6)이나 (7)의 역수를 사용한다. 따라서 적합도를 최대로 하는 것은 MSE(Mean Squared Error)와 MAPE(Mean Absolute Percent Error)가 최소인 것과 같다. 진화가 끝나면 그 과정에서 적합도가 최대인 개체가 최적해로 결정된다.

$$MSE = \frac{1}{n} \sum_{i=1}^n (a_i - m_i)^2 \quad (6)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|a_i - m_i|}{a_i} \quad (7)$$

여기서 a_i 는 데이터 값이고 m_i 는 퍼지모델의 출력 값이다. n 은 사용된 데이터 수이다.

4. 비선형 시스템 사례

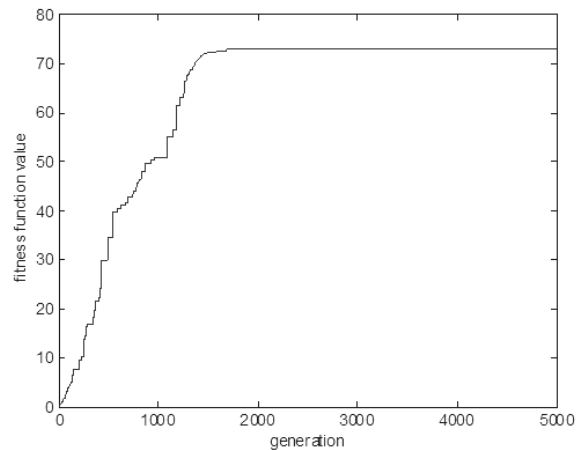
제안된 방법의 타당성을 검토하기 위해 벤치마크로 사용되는 식(8)의 비선형 수식 출력에 무작위 잡음이 혼합된 데이터를 사용한다. 25개의 데이터는 퍼지모델 학습에 사용되며 다른 25개 데이터는 성능 검증에 사용된다[18,19].

$$y = (1 + x_1^{-2} + x_2^{-1.5})^2, 1 \leq x_1, x_2 \leq 5 \quad (8)$$

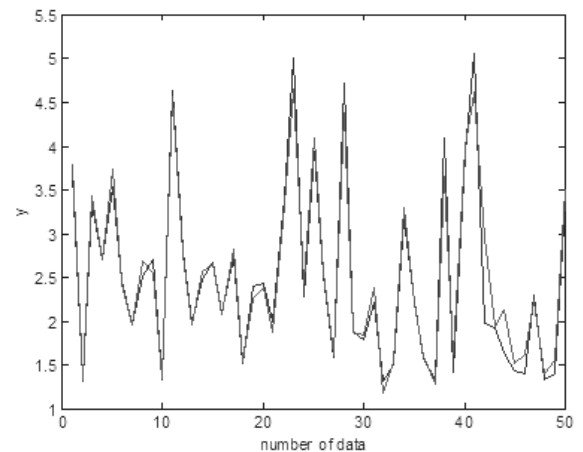
퍼지 모델링에서 사용된 차분진화의 제어변수의 값으로 최대 진화 세대 수=5000, 집단 크기=7, $F=0.5$, 교배율=0.5를 사용하였다. 최대 적합도 평가 횟수는 집단 크기×진화 세대 수로 결정된다. 비선형성이 크지 않으므로 후건부는 선형식을 사용하였고(식(1)에서 $b_j^i=1$ 로 설정) 규칙 수는 3으로 하였다. 차분진화를 통해 탐색할 퍼지규칙의 매개변수 개수는 총 33개 이다. 퍼지규칙 전건부에 2(입력변수 개수)×4(멤버

쉽함수 매개변수 개수)×3(규칙 수)=24개와 후건부에 3(입력변수 개수+1)×3(규칙 수)=9개.

(그림 1)은 진화 과정에서 식(6)의 역수인 적합도 값을 보인 것으로 진화가 이루어지면서 개체의 적합도가 증가한다. (그림 2)는 모델의 출력 값과 데이터 값을 표시한 것으로 거의 차이가 없다. <표 1>은 다른 모델링 방법과의 결과를 비교한 것으로 본 논문에서 제시한 방법이 가장 적은 수의 규칙으로도 뛰어난 성능을 보였다. <표 1>에서 α 는 GA-하이브리드 탐색에서 유전알고리즘에 이어 Complex 방법에 의한 탐색 횟수를 의미하며, MSE1과 MSE2는 학습 및 평



(그림 1) 진화 과정에서 적합도 변화



(그림 2) 실제와 모델 출력 값의 비교

<표 1> 비선형 시스템 모델의 성능 비교

참고 문헌	MSE1	MSE2	규칙 수	적합도 평가 횟수
[18]	0.010		6	
[2]	0.073	0.09	4	$50 \times 200 + \alpha$
[1]	0.048	0.22	5	70×3000
[3]	0.017	0.13	5	70×3000
본 논문	0.014	0.08	3	7×5000

<표 2> 최적 퍼지모델의 매개변수

	입력변수	매개변수	규칙(i=1)	규칙(i=2)	규칙(i=3)
전 건 부	x_1	c_{i1}	2.508	2.976	1.074
	x_2	c_{i2}	3.454	2.736	2.108
	x_1	d_{i1}^l	0.823	1.064	0.0
	x_2	d_{i2}^l	1.284	0.666	2.0
	x_1	d_{i1}^r	1.729	1.686	0.095
	x_2	d_{i2}^r	1.157	1.765	1.986
	x_1	d_{i1}^r	1.236	1.933	1.998
	x_2	d_{i2}^r	1.815	1.937	1.841
후 건 부		a_0^i	0.654	2.321	5.0
	x_1	a_1^i	0.021	0.253	-0.544
	x_2	a_2^i	0.280	-0.508	0.312

가 데이터에 대한 오차를 나타낸다. <표 2>는 진화탐색으로 찾은 퍼지모델에서 규칙을 구성하는 매개변수를 보인 것이다.

5. 주가 예측

주식 시장은 매우 복잡하며 그와 같은 복잡성을 해결하는 것은 증권거래사의 꿈이기 때문에 의사결정을 지원할 강력한 도구에 대한 필요는 항상 존재해왔다. 1990년대 이후 다양한 기계학습 방법들을 사용한 예측 모델이 개발되고 있는데, 이들 모델은 재무 지표, 경제 관련 인자들이나 과거의 데이터를 사용하였다. 주가 예측 모델을 만들 때 우선적으로 고려해야 할 것은 주가 변동에 영향을 미치는 주요 입력변수를 찾아내는 것이다. 주가는 경제나 정치적인 많은 요인에 의해 영향을 받는데, 이들 요인 모두를 예측 모델에 반영하는 것은 불가능하다. 주가 예측에 대한 많은 연구가 이루어졌지만 어떤 방법과 어떤 변수를 사용하는 것이 객관적으로 더 좋다는 연구 결과는 없다. 이는 연구마다 사용한 방법, 변수 및 데이터의 차이로 인해 객관적인 비교가 어렵기 때문일 것이다.

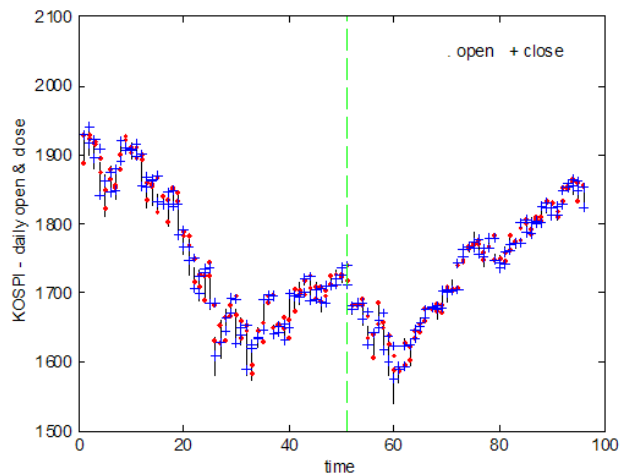
본 논문에서는 시장을 움직이는 모든 주요 인자를 고려하는 대신(고려할 수도 없지만), 주가의 움직임 자체에만 주목한다. 즉, 과거의 주가 움직임에서와 유사한 상황이 되면 주가는 동일한 방식으로 움직일 것이란 가정을 적용한다. 이런 가정은 과거의 경험에 기초한 것으로 캔들스틱(candlestick) 차트와 같은 기술적 분석에서 사용된다. 캔들스틱은 오래되었지만 주가를 가시화하는데 여전히 인기 있는 방법으로 특정한 캔들스틱 패턴을 통해 의사결정을 할 수 있다 [6]. 1992년 1월부터 1997년 6월까지 5.5년 동안의 주가를 분석한 결과 캔들스틱 차트는 시간과 분야에 독립적임이 입증되었다[5]. 이는 과거의 주가 변동 패턴이 그 때와 유사한 상황이 되면 현재나 미래에 반복될 수 있다는 것을 의미한다. 캔들스틱은 단위 시간(일, 주 또는 월) 동안 주가의 시가, 종가, 최고가와 최저가에 기초해서 만들어진다. 캔들스틱 패턴은 단일, 이중 또는 삼중의 캔들스틱에 의해 형성된다.

연구결과에 의하면 의미 있는 캔들스틱 패턴의 75%가 단일 캔들스틱이고 25%가 이중 캔들스틱이며 삼중 캔들스틱 패턴은 발생 빈도가 0.002%로 무시할 정도이다[5]. 패턴의 99.8%는 두 개의 가장 최근 시간 단위의 캔들스틱 정보로 표현될 수 있기 때문에 본 논문에서는 이들 정보를 퍼지 예측 모델의 입력변수로 사용한다.

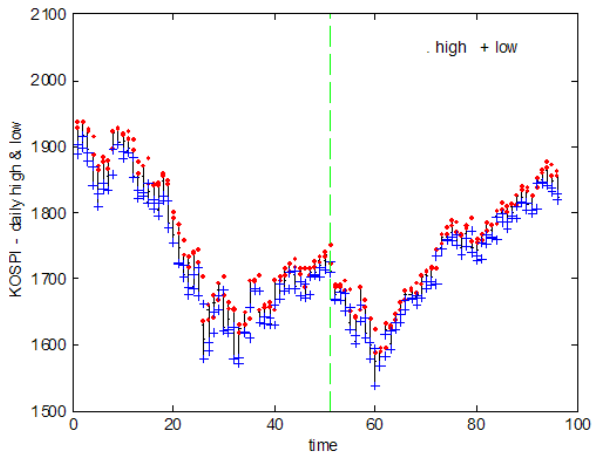
퍼지 예측 모델에서 t 는 시장의 가장 최근 시간 단위를 나타낸다. 시간 단위가 일이라면 t 는 오늘, $t-1$ 은 전날, $t+1$ 은 내일을 말한다. 예측 모델은 4개의 MISO 퍼지모델로 구성된다. 입력변수는 당일 주가의 시가, 최고가, 최저가와 종가인 $open(t)$, $high(t)$, $low(t)$, $close(t)$ 와 전일 주가의 시가, 최고가, 최저가와 종가인 $open(t-1)$, $high(t-1)$, $low(t-1)$ 와 $close(t-1)$ 로 구성된 8개 변수로 4개 모델 모두에서 동일하다. 출력변수는 각각 내일 주가의 시가, 최고가, 최저가와 종가인 $open(t+1)$, $high(t+1)$, $low(t+1)$ 와 $close(t+1)$ 이다. 즉, 퍼지모델은 어제와 오늘의 시가, 최고가, 최저가와 종가를 입력으로 받아 내일의 시가, 최고가, 최저가 및 종가를 예측한다.

퍼지모델 학습에는 2006년 12월부터 2008년 2월까지의 일별 KOSPI 데이터가, 모델 평가에서는 2008년 3월부터 6월까지의 데이터가 사용되었다. 차분진화 제어변수 값으로 진화 세대 수=20000, 집단 크기=20, $F=0.5$, 교배율=0.5를 사용하였다. 모델이 비선형 특성을 잘 추종할 수 있도록 퍼지규칙의 후건부를 비선형식으로 하였기 때문에 퍼지규칙의 수는 2로 설정한다. 차분진화를 통해 탐색할 퍼지규칙의 매개변수 수는 총 98개 이다. 퍼지규칙 전건부에 8(입력변수 개수) \times 4(멤버십함수의 매개변수 개수) \times 2(규칙 수)=64개, 후건부에 [9(입력변수 개수+1)+8(입력변수 개수)] \times 2(규칙 수)=34개.

(그림 3)과 (그림 4)는 KOSPI의 일별 시가, 종가, 최고가 및 최저가 데이터와 모델 출력의 일부를 표시한 것이다. 그 그래프 중앙의 수직 점선을 기준으로 왼쪽은 학습 데이터에 대한 결과를, 오른쪽은 평가용 데이터에 대한 결과를 표시한다. 각 그래프에서 원래 데이터는 수직선으로 표시되며,



(그림 3) KOSPI 일별 시가 및 종가와 모델 출력의 비교



(그림 4) KOSPI 일별 최고가 및 최저가와 모델 출력의 비교

수직선의 위쪽 끝은 최고가, 아래쪽 끝은 최저가, 수직선의 왼쪽 턱은 시가, 오른쪽 턱은 종가를 표시한다. 모델 출력의 시가와 최고가는 •로, 종가와 최저가는 +로 표시된다. 따라서 (그림 3)에서 수직선의 위쪽 끝은 •와, 아래쪽 끝은 +와, (그림 4)에서는 수직선의 왼쪽 턱은 •와, 오른쪽 턱은 +와 같을 수도 모델은 정확한 것이다.

<표 3>은 퍼지 모델과 신경회로망 모델의 성능을 비교한 것으로 오차는 식(7)에 의해 계산되며, MAPE1과 MAPE2는 학습용과 평가용 데이터에 대한 오차를 나타낸다. 퍼지 모델에서 KOSPI의 시가, 종가, 최고가 및 최저가 모두 퍼지규칙의 후건부가 비선형식인 경우가 선형식인 경우보다 KOSPI의 시가, 종가, 최고가 및 최저가 모두에 대해 학습 및 평가 오차에서 우수하였고 예측 오차 MAPE는 1% 이하였다. 동일 기준에서 모델링 능력을 비교하기 위해 신경회로망 모델을 구축하였다. 사용한 신경회로망 모델은 중간층이 하나와 5개의 중간층 노드로 구성된다. 중간층은 전달함수로 탄젠트 시그모이드 함수를, 출력층은 선형 함수를 사용한다. 학습 알고리즘은 Levenberg-Marquard를 사용한다. 학습은 신경회로망의 과도학습 문제를 완화시키기 위해 모델링용 데이터를 학습과 시험 데이터로 나눈 후, 학습 오차가 감소해

도 시험 오차가 증가하면 학습이 종료되도록 한다. 그럼에도 불구하고 학습 오차에 비해 평가용 데이터에 대한 오차가 퍼지 모델에 비해 큼을 볼 수 있다.

본 논문에서 사용한 평가용 데이터는 금년 3월부터 6월말까지로 미국 금융 위기가 국내 주가에 본격적으로 반영되기 시작한 시기이다. 평가용 데이터에 대한 모델 오차가 학습용 데이터의 경우 보다 약간 크지만 일반적인 모델링에서 보이는 수준임을 감안하면 금융 위기가 반영된 시기의 주가도 모델링 수준의 평균 오차 범위에서 예측이 가능함을 알 수 있다. 다만, 최대 오차가 130.2로 매우 큰 경우도 있었다. 이는 돌발적인 이벤트의 발생으로 인해 기존 데이터 패턴의 움직임과 반대의 현상이 나타나 오차가 평균 보다 매우 커진 경우로 돌발적인 이벤트가 발생한 당일의 예측은 오차가 매우 큼을 알 수 있다. 본 논문에서 제시한 모델은 돌발적인 이벤트 등으로 주가의 변동 폭이 큰 날에는 사용할 수 없지만 주가가 박스권에 있으면서 당일 변동 폭(고가 - 저가)이 평균(사용한 데이터에서 하루 주가 변동 폭의 평균은 27.9) 이하인 경우에는 저가, 고가와 종가 예측을 기준으로 매매 시점에 대한 의사결정에 활용될 수 있다고 판단된다.

6. 결 론

본 논문에서는 캔들스틱 패턴 분석에 사용되는 주가 정보를 입력으로 사용하는 퍼지 예측 모델 구축 방법을 제안하였다. 주가의 변동성을 잘 추종할 수 있도록 하기 위해 퍼지규칙의 후건부를 비선형식으로 하였는데, 시뮬레이션 결과에서도 비선형식을 사용한 경우가 학습 및 평가 오차 모두에서 선형식을 사용한 경우 보다 좋게 나왔다. 퍼지규칙은 무작위 값으로 시작하여 차분진화 과정을 거치면서 학습 데이터의 내재적 특성을 잘 묘사할 수 있는 최적의 규칙으로 진화되었다. 제안된 방법의 성능을 비교, 분석하기 위해 모델링에서 벤치마크 가운데 하나로 사용되는 비선형 수식 예를 고려하였다. 결과적으로 적합도평가 횟수가 적으면서도 정확한 모델을 구축하였는데, 이는 전역 해를 빠르고 효과적으로 찾을 수 있는 차분진화의 뛰어난 탐색능력 때문으로 보인다. 본 논문에서 제안된 방법은 차트의 기술적 분석에 사용된 정보를 사용하여 예측 모델을 만들었기 때문에 차트의 패턴이 시간과 종목에 독립적인 이상 KOSPI나 개별 종목의 일, 주 및 월 단위의 지수 예측에도 적용될 수 있을 것이다.

참 고 문 헌

- [1] H. S. Hwang, "Automatic design of fuzzy rule base for modeling and control using evolutionary programming," IEEE Proc-Control Theory Appl., 146, (1), pp.9-16, 1996.
- [2] H. S. Hwang and K. B. Woo, "Linguistic fuzzy model identification," IEEE Proc. Control Theory Appl., 142, (6), pp.537-544, 1995.

<표 3> 일별 주가 예측 모델의 성능 평가

모델	모델 출력	MAPE1	MAPE2	규칙 수	
퍼지	비선형	open	0.675	0.647	2
		high	0.740	0.664	2
		low	0.905	0.861	2
		close	1.026	0.924	2
	선형	open	0.851	0.984	2
		high	0.834	0.746	2
		low	1.022	1.028	2
		close	1.140	0.930	2
신경 회로망	open	0.625	0.899	중간층 노드수: 5	
	high	0.616	0.915		
	low	0.760	1.078		
	close	0.877	1.140		

[3] S. J. Kang, C. H. Woo, H. S. Hwang and K. B. Woo, "Evolutionary design of fuzzy rule base for nonlinear system modeling and control," IEEE Trans. Fuzzy Systems, 8, (1), pp.37-44, 2000.

[4] B. M. Louis, "Trend forecasting with technical analysis," Marketplace BOOKS, 2000.

[5] K. H. Lee and G. S. Jo, "Expert system for predicting stock market timing using a candlestick chart," Expert System With Applications, 16, pp.357-364, 1999.

[6] L. C. H. Leon, A. Liu and W. S. Chen, "Pattern discovery of fuzzy time series for financial prediction," IEEE Trans. Knowledge and Data Engineering, 18, (5), pp.613-625, 2006.

[7] P. C. Chang and C. H. Liua, "A TSK type fuzzy based system for stock price prediction," Expert Systems with Applications, 34, (1), pp.135-144, 2008.

[8] Coupelon, "Neural network modeling for stock movement prediction - A state of the art," 2007(http://olivier.coupelon.free.fr/Neural_network_modeling_for_stock_movement_prediction.pdf)

[9] K. Nygren, "Stock Prediction: A neural network approach," Master Thesis, Royal Institute of Technology, KTH, March, 2004.

[10] Y. Tang, F. Xu, X. Wan and Y. Q. Zhang, "Web-based fuzzy neural networks for stock prediction," Computational intelligence and applications, pp.169-174, 2002.

[11] G. Armano, M. Marchesi and A. Murru, "A hybrid genetic-neural architecture for stock indexes forecasting," Information Sciences, 170, 2005.

[12] Fan and M. Palaniswami, "Stock selection using support vector machines," Proc. of Int. Joint Conf. on Neural Networks, 2001.

[13] S. M. Kendall and K. Ord, "Time Series," Oxford, 1997.

[14] J. L. Wanga and S. H. Chanb, "Stock market trading rule discovery using two-layer bias decision tree," Expert Systems with Applications 30, (4), pp.605-611, May, 2006.

[15] M. Noor and R. H. Khokhar, "Fuzzy Decision Tree for Data Mining of Time Series Stock Market Databases," Critical Assessment of Microarray Data Analysis 2004 Conference, November, 2004.

[16] Giudici P. Applied Data Mining. Statistical Methods for Business and Industry. Wiley, 2003.

[17] R. Storn, "Differential evolution, a simple and efficient heuristic strategy for global optimization over continuous spaces," Journal of Global Optimization, 11, (4), pp.341-359, 1997.

[18] M. Sugeno and T. Yasukawa, "A fuzzy logic based approach to qualitative modeling," IEEE Trans. Fuzzy Syst., 1, (1), pp.7-31, 1993.

[19] T. Takagi and M. Sugeno, "Fuzzy identification of systems

and its application to modeling and control," IEEE Trans. Syst. Man. Cybern., 15, pp.116-132, 1985.

[20] M. H. Zarandi, E. Neshat, I. B. Turksen and B. Rezaee, "A type-2 fuzzy model for stock market analysis," Fuzzy System Conf., FUZZ-IEEE 2007, pp.1-6, July, 2007.



황희수

e-mail : hshwang@halla.ac.kr

1986년 연세대학교 전기공학과(학사)

1988년 연세대학교 대학원 전기공학과(공학석사)

1993년 연세대학교 대학원 전기공학과(공학박사)

1993년~2000년 한국고속철도기술개발사업단 수석연구원

2001~현 재 한라대학교 전기전자과 부교수

관심분야 : 철도 시스템의 에너지 효율 향상, 예측, 최적화, 광음향 측정, 에너지 진단