

HTM 기반의 주식가격 연속 예측 시스템 개발

서대호[†], 배선갑^{**}, 김성진^{***}, 강현석^{****}, 배종민^{*****}

요 약

주식 가격은 연속적으로 변화하는 스트림 데이터이다. 이러한 데이터의 특성상 시간의 흐름에 따라 주식 가격의 동향이 달라질 수 있기 때문에 주식 가격 동향의 예측은 가격이 갱신될 때 마다 연속적으로 이루어져야 한다. 본 논문은 HTM 모델을 이용하여 원하는 종목의 주식 가격 동향을 설정된 구간 간격에 따라 연속적으로 주식 가격 동향을 예측하는 새로운 방법을 제안한다. 이를 위해 먼저 정규화 과정을 거친 후 그 결과를 스트림 센서로 전달하는 선처리기와 연속적인 입력 데이터를 효과적으로 처리할 수 있는 스트림 센서를 제시한다. 또한, 각 레벨별 예측 결과를 저장하여 상위 단계로 전달하는 선 예측 저장 노드를 고안하고 이를 이용하여 주식 가격 동향을 예측하는 HTM 네트워크를 제시한다. 그리고 본 시스템을 실제 주식 가격으로 실험하여 그 성능을 제시한다.

Development of a Continuous Prediction System of Stock Price Based on HTM Network

Dae-Ho Seo[†], Sun-Gap Bae^{**}, Sung-Jin Kim^{***},
Hyun-Syug Kang^{****}, Jong-Min Bae^{*****}

ABSTRACT

Stock price is stream data to change continuously. The characteristics of these data, stock trends according to flow of time intervals may differ. therefore, stock price should be continuously prediction when the price is updated. In this paper, we propose the new prediction system that continuously predicts the stock price according to the predefined time intervals for the selected stock item using HTM model. We first present a preprocessor which normalizes the stock data and passes its result to the stream sensor. We next present a stream sensor which efficiently processes the continuous input. In addition, we devise a storage node which stores the prediction results for each level and passes it to next upper level and present the HTM network for prediction using these nodes. We show experimented our system using the actual stock price and shows its performance.

Key words: Stock Price Prediction(주식가격 예측), Continuous Prediction(연속예측), HTM

※ 교신저자(Corresponding Author): 배종민, 주소: 경남 진주시 가좌동 경상대학교 컴퓨터과학과(660-701), 전화: 055)772-1383, FAX: 055)772-1389, E-mail: jmbae@gnu.ac.kr

접수일: 2011년 3월 14일, 수정일: 2011년 7월 10일

완료일: 2011년 8월 4일

[†] 준회원, 경상대학교 컴퓨터과학과

(E-mail: chorokh@naver.com)

^{**} 준회원, 경상대학교 컴퓨터과학과

(E-mail: bsgap@hanmail.net)

^{***} 정회원, 연암공업대학 컴퓨터공학과

(E-mail: sjkim@yc.ac.kr)

^{****} 종신회원, 경상대학교 컴퓨터과학과

(E-mail: hskang@gnu.ac.kr)

^{*****} 정회원, 경상대학교 컴퓨터과학과

※ 본 연구는 2009년도 경상대학교 학술진흥지원사업 연구비에 의해서 수행되었음.

1. 서 론

주식 시장에서 주식 가격은 연속적으로 변화하는 스트림 데이터로서 계속해서 변화하는 일련의 값들로 나타난다[1]. 그림 1은 주식 가격 스트림 데이터를 x축은 시간으로 y축은 시세 값을 나타낸 것이다. 붉은 색의 사각형으로 표시된 부분은 모니터링 하고자 하는 시간 간격을 나타내며 시간의 흐름에 따라 설정된 간격에서의 주식 가격 추세의 변화를 보인다. 그림 1의 (가)에서 (라)는 5분 간격의 시간 구간이 1분 간격으로 주식 가격 추세가 변화하는 모습을 나타낸다. (가)와 (나) 구간에서는 계속 상승하는 경향을 보이고 있고 (다) 구간에서는 상승에서 하강으로 변하고 있으며, (라) 구간에서는 하강하는 모습을 보이고 있다. 이처럼 시간 구간의 변화에 따라 주식 가격 동향의 예측이 달라질 수 있기 때문에 주식 가격 동향의 예측은 연속적으로 가격이 갱신될 때 마다 이루어져야 하고, 예측된 결과를 연속적으로 알려 주어야 한다.

본 논문은 HTM 모델을 이용하여 원하는 종목의 주식 가격을 설정된 구간 간격에 따라 계속하여 연속적으로 주식 가격 동향을 예측할 수 있는 새로운 방법을 제안한다. HTM은 Jeff Hawkins가 제안한 이

론으로 기존의 인공신경망에 비해 보다 인간의 두뇌 구조를 가깝게 흉내 낸 모델이다. 이 이론을 기반으로 한 HTM 네트워크는 실세계에 대한 시공간적인 패턴을 계층형으로 분산 기억함으로써, 계층형의 시공간적인 특성을 갖는 실세계의 응용에 적합한 것으로 알려져 있다[1,2]. HTM 모델의 기본 센서인 벡터 센서의 경우 벡터 값들의 목록을 입력 받아 한꺼번에 처리를 한다. 이러한 벡터센서의 특성상 매 분당 연속적으로 값을 입력받아 그것을 연속으로 처리해야 하는 연속 예측에는 효과적이지 못하다. 효과적인 연속예측을 위해 본 시스템은 주식 가격 데이터의 연속적인 입력을 효과적으로 처리할 수 있는 스트림 센서(Stream sensor)를 제시한다. 또한, HTM 네트워크의 각 노드 레벨별 예측 결과를 저장하여 상위 단계로 전달하는 선 예측 저장 노드를 고안하고 이를 이용하여 주식 가격 동향을 예측할 수 있는 HTM 네트워크를 구성한다. 이렇게 구성된 HTM 네트워크를 이용하여 연속적으로 주식 가격 동향을 예측하는 방법을 제시한다.

본 논문은 6개의 장으로 구성된다. 2장에서는 관련 연구로서 주식 가격 동향예측 기법과 연속적 예측 개념에 대하여 논한다. 3장에서는 HTM 네트워크와

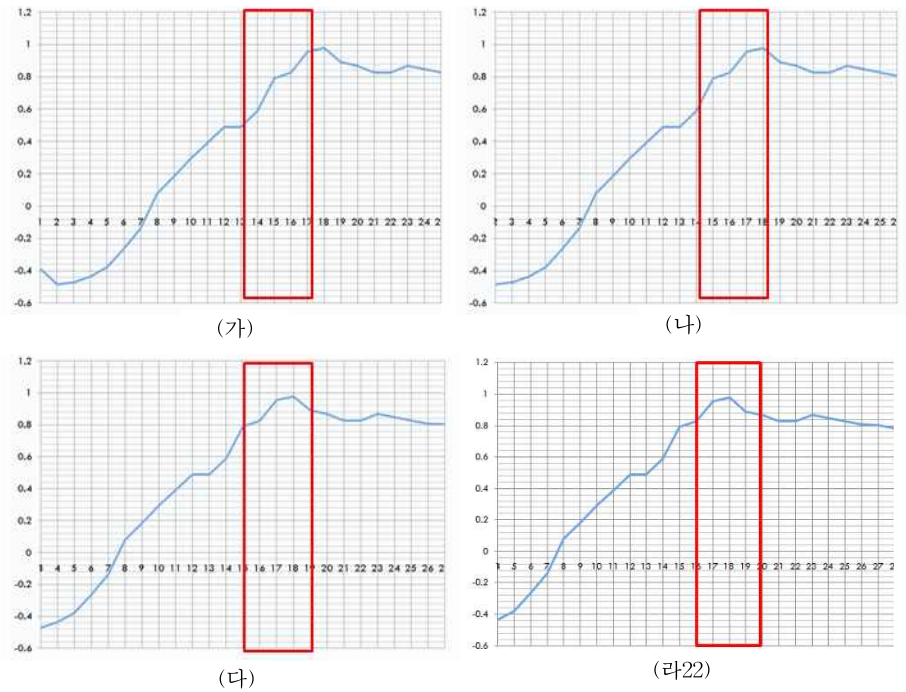


그림 1. 연속 예측의 개념

이를 이용한 주식 예측방법을 소개하고 문제점을 알아본다. 4장에서는 이를 극복하기 위해 새롭게 개발한 HTM 기반의 주식 가격 동향 연속 예측 시스템의 설계 내용을 기술한다. 그리고 5장에서는 실제 주식 데이터를 이용해 구현 결과와 성능을 분석한다. 6장에서는 개발된 시스템의 구현 결과를 설명한다. 마지막으로 7장에서는 결론과 향후 과제를 논한다.

2. 관련연구

주식 투자자의 목적은 수익률을 극대화 하는 것이다. 이를 위해서는 주식 데이터의 분석을 통해 지수의 변동, 주가의 변화 시점, 거래 시세 등을 예측함으로써 주식의 매매 시점을 잘 선택할 수 있어야 한다. 정보 시스템의 도움으로 적절한 주식 매매 시점을 선택하려는 노력들이 많이 있어 왔다. 이들 연구들은 대부분 예측 모형의 구축에 집중되어 왔다. 그러한 모형들은 과거 주식 가격 시퀀스에 기반한 미시적 관점의 변수와 경제 변수와 같은 거시적 관점의 변수들을 이용한 것[3], 그리고 과거의 주식 시세를 시계열 데이터로 이용하여 주식 가격의 미래를 예측하는 기법들이다[4].

그 중에서 시계열 데이터를 기반으로 한 주식 가격 예측 방법들은 과거의 주식 가격 데이터의 패턴을 추적할 수 있는 선형 예측 모델을 제안하는데, 전문가 시스템[5], 인공신경망[6], 결정 트리[7], SVM (Support Vector Machine)[8], 퍼지 규칙[9], 데이터 마이닝 기법[10] 등을 사용한다. 이 기법들은 대부분 복잡한 수학적 모델을 기초로 그동안 주목할 만한 성과를 보여 왔다. 그 중에서도 특히 인공신경망을 이용한 예측 기법들이 많이 연구되어 왔다.

그런데 인공신경망을 이용한 주식 시세에 대한 예측 연구들은 일반적으로 상호 연결된 뉴런에 의해 연상 기억, 필터, 변환, 분류, 최적화 등의 기능을 수행하여 예측을 할 수 있게 한다. 하지만 주식 시세 자료가 가지는 시계열 모형의 특징을 기억 구조로서 인공신경망의 학습 구조에 도입한 연구는 매우 드물고, 네트워크 학습을 위해서는 많은 시간을 필요로 하는 단점을 가지고 있다. 일반적으로 시계열 데이터의 분석은 시간 영역 분석과 주파수 영역 분석으로 구분된다. 시간 영역 분석은 현재 시점의 값이 과거 시점의 값들에 의해 결정되는 회귀 모델을 기반으로 하는 방식이다. 주파수 영역 분석은 정적인 시계열

데이터를 이용하여 월별, 계절별, 년도별 등의 거시적인 변화를 예측하는 방식이다[11]. 그러나 이러한 기법들은 주식의 변화 예측을 위한 투자자의 동적인 요구를 제대로 반영할 수 없으며, 단시간 내의 주식 가격 변화를 예측하기가 어렵다는 단점이 있다. 본 논문에서는 이러한 문제를 극복하기 위해 HTM 기능을 이용하여 주식이 가격의 동향을 연속으로 예측하는 방법을 제시한다.

3. HTM 네트워크를 이용한 주식가격 예측

3.1 HTM 이론

HTM은 미국의 Numenta사에서 2007년 중반에 개발된 기술로서 시스템을 일정한 공통 연산을 수행하는 메모리 노드들의 계층 네트워크(Hierarchy network)로 구성한다[2]. 이렇게 구성된 HTM 네트워크는 시간적 패턴 정보와 공간적 패턴 정보를 이용하여 효율적으로 판단을 할 수 있게 한다. HTM은 일종의 메모리 시스템으로 문제마다 각자 다른 알고리즘을 수행하는 것이 아니라 문제를 해결하는 방법을 학습한다. 그리고 학습 과정에서 감각 신호의 특성들 사이의 시공간적 관계성이 계층적 메모리 구조에 형성된다. 이러한 학습이 끝나면 새로운 입력 패턴을 인식할 수 있게 된다[3]. 이러한 HTM의 인식능력을 활용하여 부품의 양·불량 판별 시스템[12], 다양한 형태의 한글 인식[13] 등의 논문들이 발표되어 왔다.

3.2 HTM 네트워크의 구성

HTM에서 네트워크는 노드들의 계층으로 구성된다. 학습 단계에서는 네트워크가 훈련 패턴에 노출되어 범주(Category)들에 패턴들을 사상시키는 모델을 구축하고, 학습이 완료되면 네트워크는 추론 단계로 넘어간다. 추론 단계에서는 네트워크가 이전에 본 일이 없는 새로운 패턴에 대해 범주들에 대한 간단한 분포를 생성한다.

HTM 네트워크는 Zeta1, Zeta1Top, Data Sensor, Effector 등 몇 가지 유형의 노드들로 구성된다[2]. 우선 데이터는 계층의 바닥으로부터 데이터 센서(Data Sensor)를 통해 입력되고, 입력된 데이터는 노드들의 계층 구조를 통해 위로 올라가면서 처리된다. 다음으로 그 중에서 Zeta1 노드는 계층의 대부분을

구성하게 되며 학습과 추론을 담당한다. Zeta1 노드는 그림 2에서와 같이 정보를 공간적 풀러(Spatial pooler)와 시간적 풀러(Temporal pooler)의 두 가지 구성 요소로 처리하게 된다. 공간적 풀러는 데이터의 공간적 패턴을 분류하는 역할을 수행한다. 즉, 입력 벡터들의 분산된 표현들을 분류하는 방법으로 학습한다. 시간적 풀러는 데이터의 동시발생(Coincidence)을 기초로 시간적 패턴을 그룹화 하는 알고리즘을 수행한다. 이는 추론 결과가 이전 입력 값뿐만 아니라 최근의 입력 값을 기초로 하여 계산하는 시간 기반의 알고리즘이다. 그리고 Zeta1TopNode는 공간적 풀러와 통제 사상기로 이루어져 있다. Zeta1-TopNode의 공간적 풀러는 Zeta1Node와 같은 역할을 하며, 통제 사상기는 범주 정보를 이용하여 최종적으로 판단을 하게 된다. 마지막으로 Effector 노드는 Zeta1TopNode의 판단 결과를 외부로 출력 해주는 역할을 담당한다[14].

3.3 HTM 네트워크에서의 학습과 추론

HTM 네트워크의 동작은 크게 학습과 추론 단계로 이루어진다[7,8]. 학습은 HTM 네트워크에서 아래로부터 레벨별로 이루어진다. 즉, 학습 단계에서 공간적 풀러는 입력 패턴을 정량화하여 이들의 동시 발생들을 찾는다. 따라서 공간적 풀러의 출력은 동시 발생들의 목록들이고, 이들을 시간적 풀러에 보내게

된다. 시간적 풀러는 이 목록들을 받아서 시간적으로 동시에 발생하는 패턴들을 그룹으로 묶음으로써 학습을 완료한다. 레벨별로 학습이 완료되면 위 레벨로 보낼 데이터를 결정하기 위해 입력 데이터들에 대해서 추론하는 추론(Inference) 단계로 이행된다. 추론은 아래로부터의 입력 패턴에 대해 학습 단계에서 학습한 동시발생에 대한 그룹들에 대한 범주의 분포를 발생시킨다.

이렇게 하여 네트워크의 모든 노드들이 전체적으로 학습이 완료되면 훈련이 끝난 HTM 네트워크가 완성된다. 이제 이를 이용하여 새로운 입력 데이터를 인식할 수 있게 되는데, 이것은 최상위 레벨에서 추론을 하는 것을 의미한다. 최상위 레벨의 추론 결과는 입력 데이터의 패턴 범주들에 대한 확률 분포를 나타낸다[14].

3.4 HTM을 이용한 주식 가격 동향 예측

이러한 HTM 기술을 이용하여 주식 가격 동향을 예측할 수 있는 HTM 네트워크를 구성할 수 있다. 본 절에서는 Numenta 사에서 제시한 주식 가격 동향 예측을 위한 HTM 기술의 적용 방법을 알아본다. 여기서 제시된 HTM 네트워크는 어떤 특정한 주식 종목을 선택한 후, 사용자가 임의로 예측 구간을 설정하면 과거의 주식 가격을 바탕으로 하여 설정된 시간 간격에 기초하여 가격 경향을 예측한다.

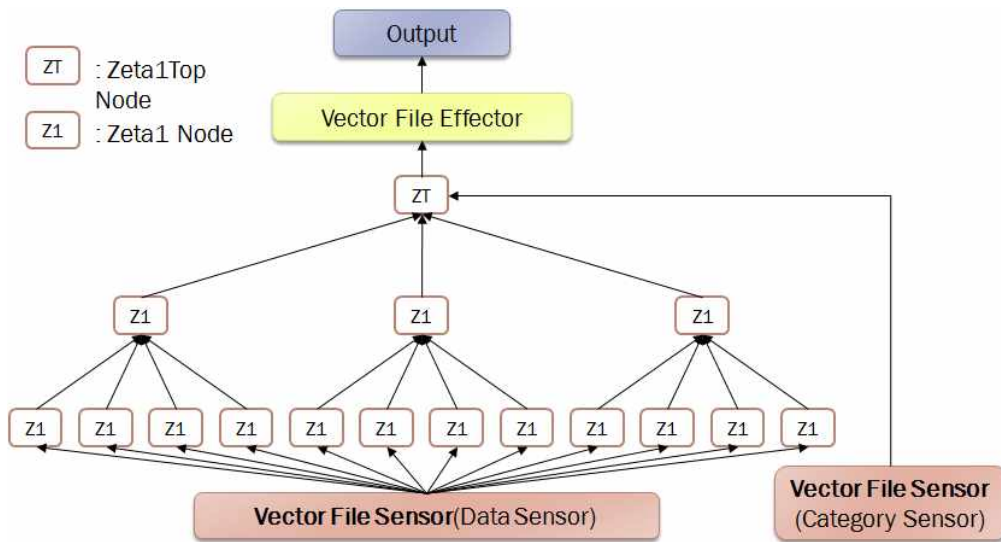


그림 2. 주식 가격 동향 예측을 위한 HTM 네트워크

그림 2는 이러한 예측을 위한 HTM 네트워크의 구조이다. 벡터 파일 센서를 통해 1분 간격의 주식데이터 60개를 입력받는다. 입력받은 데이터는 레벨1 각각의 Zeta1Node에 5개씩 입력이 되고 추론한 결과를 상위 레벨로 보낸다. 이로써 60분 간격의 주식 데이터를 기초로 주식 가격 동향을 예측한다. 그러나 이는 정적인 예측으로 연속적으로 매 분마다 같은 시간 간격을 기초로 예측하는데 사용할 수는 없다.

4. HTM 기반의 연속 예측 시스템

4.1 연속 예측을 위한 HTM 네트워크의 구조

주식 가격 동향 연속 예측을 위한 HTM 네트워크는 그림 3과 같이 3 레벨로 구성된다. 그림 3에서 최하위에는 레벨1 Zeta1Node로 입력되는 데이터를 생성하는 스트림 센서 노드가 있고 레벨1에는 1개의 Zeta1Node 노드가 있으며 그 상위에 레벨 1의 Zeta1Node의 추론 결과를 저장하는 sQUEUE₁가 있다. 그리고 레벨2에도 1개의 Zeta1Node가 있으며 그 상위에 레벨2의 Zeta1Node의 추론 결과를 저장하는 sQUEUE₂가 있다. 최상위 레벨은 한 개의 Effector

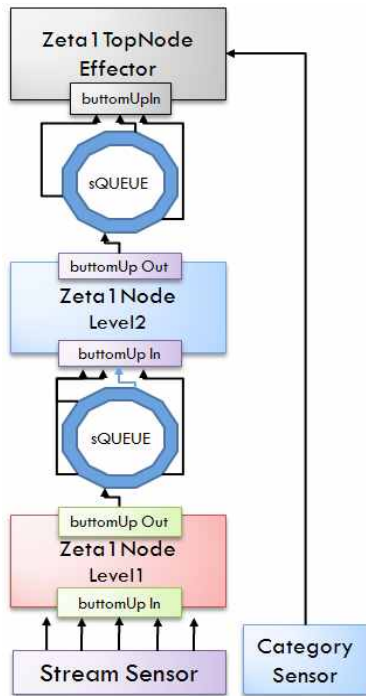


그림 3. 연속 예측을 위한 HTM 네트워크

노드로 구성된다. 스트림 센서 노드로 부터의 입력 데이터는 선 처리기에서 처리 과정을 거친 실수 값으로 Zeta1Node의 입력값의 개수와 같은 5개의 실수 값이다. 이 실수 값이 스트림 센서 노드에서 레벨1 Zeta1Node로의 입력이다. 입력을 받은 각 Zeta1-Node는 공간적 패턴과 시간적 패턴을 학습한다[15].

이러한 연속 예측을 위한 전체 시스템의 처리 흐름은 그림 4와 같다. 주식 가격 정보가 저장되어 있는 주식 가격 보관 데이터베이스로부터 주식 가격 스트림 데이터를 읽어 주식 거래 정보창과 선처리기로 보낸다. 선처리기는 주식 가격을 정규화하여 HTM 엔진으로 보낸다. 그리고 HTM 엔진에서는 HTM 네트워크를 이용하여 예측한 후 그 결과를 주식 거래 정보창으로 전달한다. 그러면 주식 거래 정보창은 주식 데이터와 HTM 엔진으로 부터 받은 결과를 종합하여 사용자에게 계속해서 보여준다.

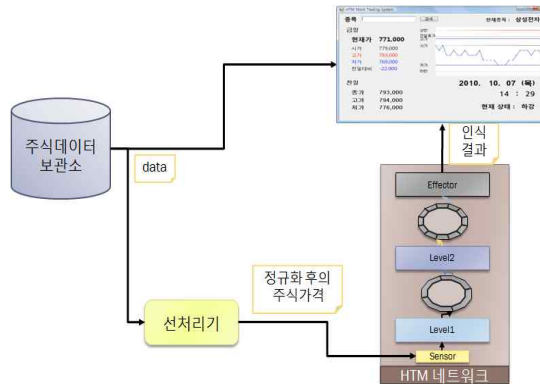


그림 4. 연속 예측 시스템의 처리 흐름

4.2 새로운 노드 타입들

4.2.1 정규화(선 처리기) 노드

입력되는 주식 가격 데이터는 주식의 종목에 따라 가격이 다르고 그 변동의 폭 차이도 많다. 그럼에도 불구하고 종목에 관계없이 동일한 시스템으로 처리하기 위해서는 변동 폭 차이를 일정한 비율로 나타나도록 하기위한 정규화가 필요하다. 주식 가격의 정규화는 수식 (1) 에 의하여 계산한다.

$$r = (CV - PV_c) \times \frac{PV_h - PV_l}{2} \tag{1}$$

여기서 CV는 현재 체결가, PV_c는 전일 중간값,

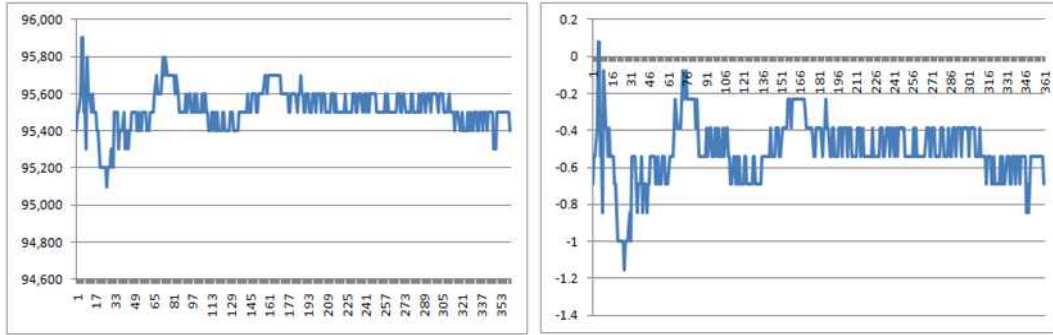


그림 5. 정규화 전·후의 데이터 분포

PV_i 는 전일 최고가, PV_i 는 전일 최저가, r 은 스트림 센서로의 입력 벡터 값으로 나타낸다. 수식 (1)의 계산식에 의한 정규화 결과는 그림 5와 같다. 정규화 이전의 패턴이 흐트러지지 않은 상태에서 값의 분포가 약 -1과 1사이에 존재한다. -1은 전일의 최저가에 근사한 값을 1은 전일의 최고가에 근사한 값을 의미한다. 정규화 결과 수천에서 수십만의 최고가와 최저가의 차이를 약 2로서 고정을 시킴으로서 여러 종목의 주식을 하나의 시스템으로 사용할 수 있다.

4.2.2 스트림 센서 노드

스트림 센서 노드는 연속적으로 입력되는 주식이격 데이터를 선 처리기에서 정규화 한 후, 자리 이동된(shifted) 벡터 형태로 계속해서 입력 데이터를 생성하여 HTM 네트워크에 입력시키는 노드이다. 스트림 센서 노드는 이를 위해 스트림 데이터를 받아서 버퍼에 누적하여 일정 크기의 벡터 데이터가 되면 HTM 네트워크의 바다 레벨의 Zeta1Node의 입력이다. 그림 6은 스트림 센서 노드의 동작 원리이다. 점선으로 표시된 부분이 스트림 센서이다. 여기서는 스트림 센서가 다섯 개의 버퍼를 가지고 있는 일반적인 큐의 모양으로 다섯 개의 값이 레벨1의 Zeta1Node의 입력 값이다. 시스템이 동작되면 r_0 가 입력됨과 동시에 레벨1의 Zeta1Node의 공간적 풀리에 값이 입력된다. 다음으로 $r_1, r_2, r_3...$ 가 입력될 때 마다 레벨1의 Zeta1Node의 공간적 풀리에 입력 값으로 전달하게 되는데, 처음 4개, 즉 r_0 부터 r_3 가 입력될 때까지는 버퍼의 초기 값과 r_i 의 입력 값이 같이 Zeta1Node의 입력 값이며, 다섯 번째부터 즉, r_4 가 입력되는 순간부터는 모두가 주식이격을 정규화한 데이터가 정상적으로 레벨1의 Zeta1Node의 입력 값으로 보내지게

된다. 그 다음 동작은 처음과 같이 연속적으로 하나의 r_i 의 입력임과 동시에 레벨1의 Zeta1Node의 입력 값이 된다.

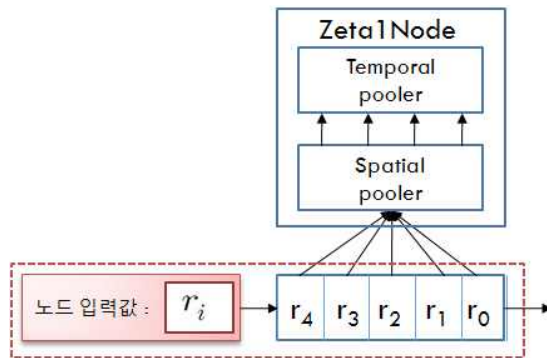


그림 6. 스트림 센서 노드의 동작

4.2.3 선 예측 저장 노드

주식이격 동향의 연속적인 예측을 효과적으로 수행하기 위하여 HTM 네트워크의 각 레벨에서 앞 단계 하위 레벨에서 처리된 결과를 저장하여 나중에 사용할 수 있도록 하는 선 예측 저장 노드를 제시한다. 즉, 스트림 데이터의 특성을 활용하여 이전 단계의 예측 처리에서 계산된 결과를 나중에 필요시 다시 처리하는 부담을 없애기 위하여 이전 결과를 활용할 수 있는 방법을 제시한다. 그리고 이를 구현하기 위해 특수한 형태의 환형 큐(Specialized circular Queue; sQUEUE)를 제시한다. 즉, 주식이격의 연속성과 HTM 네트워크의 특성을 고려하여 그림 7에 보인 것과 같이 환형으로 큐를 구성한다. HTM 네트워크의 각 레벨마다 sQUEUE가 하나씩 존재한다. 그리고 sQUEUE는 전통적인 환형 큐와 다르게 삽입

포인트와 현재 포인트의 두 개의 포인터를 가지고 있다. 또한, 고정된 크기의 버퍼를 가지고 있으나 앞과 뒤의 포인터는 없다. 대신 삽입 데이터의 위치를 나타내는 포인터만 가지고 있다. sQUEUE의 장점은 큐에서 여러 개의 데이터를 읽어 낼 때, 포인터의 위치를 변경하는 것 외에 새로운 데이터를 버퍼에 기록하고 읽어 내는 처리 부담이 버퍼의 수에 관계가 없다는 점이다. 따라서 빠른 시간 내에 기록과 패치가 가능하게 되어 연속적인 예측이 신속하게 이루어질 수 있다.

sQUEUE의 크기는 HTM 네트워크의 각 레벨마다 다르다. 레벨 i 에서 sQUEUE의 크기(S_i)는 다음 수식(2)과 같이 계산된다.

$$S_i = S_{i-1} \times n_i \quad (2)$$

여기서 S_{i-1} 은 레벨 $i-1$ 의 sQUEUE의 크기이며 S_0 의 크기는 1이다. n_i 는 레벨 $i+1$ 의 한 노드의 자식 노드개수이다. 예를 들어, 그림 3과 같이 세 개의 레벨을 갖는 HTM 네트워크의 경우에 $i+1$ 노드인 레벨2의 노드의 자식 노드개수가 4개이므로 $S_1 = S_0 \times n_1 = 1 \times 4 = 4$ 가 된다. 그리고 레벨3의 노드의 자식 노드 개수가 3개이므로 sQUEUE의 크기 $S_2 = S_1 \times n_2 = 4 \times 3 = 12$ 가 된다.

그림 7은 레벨2에서의 sQUEUE 노드의 구조와 저장과 읽기 알고리즘을 나타내었다. sQUEUE의 동작은 삽입과 리드로 이루어진다. 먼저 저장 동작에서는 하위 레벨 노드로부터 출력된 값을 입력받아 큐에 차례로 저장하게 된다. 그리고 읽기 동작에서는 현재 포인트 값과 하위 레벨 노드의 큐 크기만큼 떨어져 있는 포인트의 값들을 상위 레벨의 자식 노드 개수만큼 상위 레벨로 전달한다. 그 후 현재 포인트는 현재

포인트의 다음 포인트로 옮겨간다.

보다 구체적으로 현재 시스템에서 구성된 HTM 네트워크에서 첫 번째 레벨의 큐는 레벨1의 각 노드의 예측 결과를 저장한다. sQUEUE₁q1은 레벨1의 첫 번째 노드의 예측 결과이며 sQUEUE₁q2는 두 번째 노드의 예측 결과가 들어간다. 이러한 방식으로 sQUEUE₁의 각 노드에는 레벨1의 각 노드가 예측한 결과를 순차적으로 저장하게 된다. 상위 노드로의 출력은 예를 들어 현재 포인트가 sQUEUE₁q1이라고 하면 현재 레벨이 최하 레벨이며 상위노드의 자식노드의 개수가 4개이므로 현재 노드부터 이웃한 4개의 노드 즉, sQUEUE₁q1, sQUEUE₁q2, sQUEUE₁q3, sQUEUE₁q4의 내용을 상위 레벨로 출력한다. 그리고 현재 포인트를 다음 포인트인 sQUEUE₁q2로 옮긴다.

두 번째 레벨의 큐에서는 레벨2의 예측 결과를 저장한다. sQUEUE₂q1은 레벨1의 1, 2, 3, 4번 노드의 예측 결과를 저장, sQUEUE₂q2는 레벨1의 2, 3, 4, 5번 노드의 예측 결과를 저장한다. 이러한 방식으로 sQUEUE₂q12의 노드에는 레벨1의 9, 10, 11, 12노드의 결과가 저장된다. 상위 노드로의 출력은 현재 포인트가 sQUEUE₂q1을 가르킨다고 할 때 하위 레벨 큐의 개수 4 그리고 상위 레벨의 자식노드의 개수가 3이므로 sQUEUE₂q1, sQUEUE₂q5, sQUEUE₂q9의 세 노드에 저장된 결과를 상위 레벨로 전달하게 된다. 이때 sQUEUE₂q1에 저장된 결과 값은 레벨1의 1, 2, 3, 4번 노드의 예측 결과를 토대로 레벨2의 노드가 예측한 결과이며 마찬가지로 sQUEUE₂q5 노드는 5, 6, 7, 8번 노드의 예측 결과, sQUEUE₂q9는 9, 10, 11, 12번 노드의 예측 결과에 대한 레벨2의 예측 결과이다.

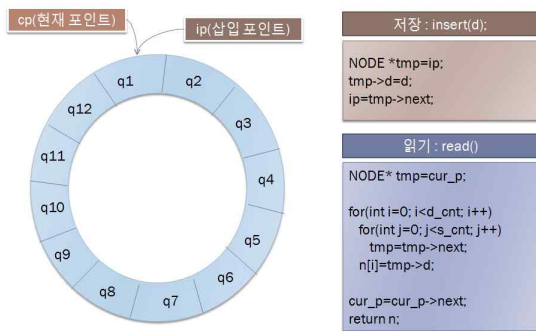


그림 7. sQUEUE 노드의 구조와 저장과 읽기 알고리즘

4.3 HTM 네트워크의 동작

제안된 HTM 네트워크에서의 예측은 각 레벨별로 이루어진다. 레벨1의 예측 결과를 레벨1의 sQUEUE₁로 보낸다. 레벨1의 sQUEUE₁는 바로 레벨2로 4개의 출력값을 보낸다. 첫 입력 단계에서는 레벨1의 Zeta1Node의 결과값 하나와 sQUEUE₁의 초기값 3개를 레벨2로 출력을 보내게 된다. 그리고 네 번째 레벨1의 Zeta1Node의 결과값이 레벨1의 sQUEUE에 입력이 되면 정상적으로 레벨1의 결과값들이 레벨2로 올라간다.

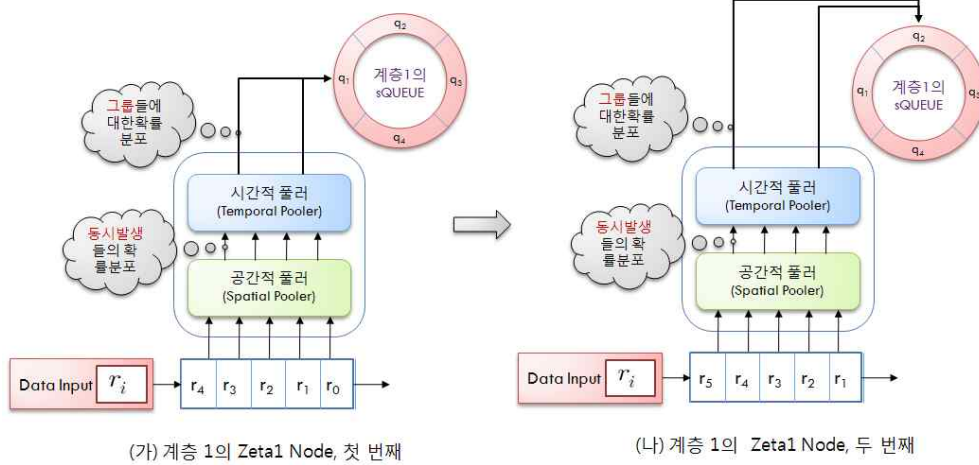


그림 8. 레벨1의 Zeta1Node의 동작

레벨2에서는 역시 1개의 Zeta1Node가 존재한다. 레벨2에서의 예측 결과는 레벨2의 sQUEUE₂에 보낸다. 레벨2의 sQUEUE₂는 즉시 3개의 출력값을 레벨3의 Effector인 Zeta1TopNode로 보낸다. 처음 동작을 레벨1의 sQUEUE₁와 같이 레벨2의 Zeta1Node의 결과값 하나와 sQUEUE₂의 초기값 2개를 Zeta1TopNode로 출력을 보낸다. Zeta1TopNode는 레벨2의 sQUEUE₂의 입력에 대하여 카테고리별 상대적 확률 분포를 발생시켜 추론한다.

4.3.1 Level1의 Zeta1Node 동작

연속 주식 가격 동향 예측 네트워크의 예측은 레벨별로 각 Zeta1Node안에서 이루어진다. 그림 8의 (가)에서 보는 바와 같이 레벨1에서의 예측은 레벨1의 Zeta1Node에 의하여 이루어진다. 입력된 주식 가격 데이터를 스트림 센서 노드가 실수 벡터로 변화하여 Zeta1Node의 공간적 풀러에 입력이 된다. 그리고 Zeta1Node의 출력은 레벨1의 sQUEUE의 q_1 에 입력된다. 그림 8의 (나)는 처음 5개의 입력 데이터에 대한 추론이 끝난 다음 또 다른 하나의 주식 가격 데이터의 입력이 있으면 바로 추론이 이루어지면서 그 결과가 레벨1의 sQUEUE의 q_2 에 입력되는 것을 보여준다. 이러한 동작이 연속적으로 이루어지면서 계속해서 변화하는 주식 가격 동향에 대한 연속적인 예측이 이루어진다.

4.3.2 Level2의 Zeta1Node의 동작

그림 9는 레벨2에서의 Zeta1Node의 동작을 보여

준다. 레벨1에서의 예측이 끝나고 레벨1의 sQUEUE₁에 그 결과가 저장이 됨과 동시에 레벨1의 sQUEUE₁는 레벨2의 Zeta1Node로 출력을 보낸다. 레벨2의 Zeta1Node는 입력된 값을 토대로 결과를 발생시킨다. 그리고 그 결과를 레벨2의 sQUEUE₂에 저장하게 된다. 레벨2의 sQUEUE₂는 결과를 저장함과 동시에 최상위 레벨의 Zeta1TopNode에 저장된 값들을 보낸다.

4.3.3 Zeta1TopNode의 동작

그림 10는 최상위 레벨의 Zeta1TopNode의 동작을 나타낸 것이다. Zeta1TopNode는 레벨2의

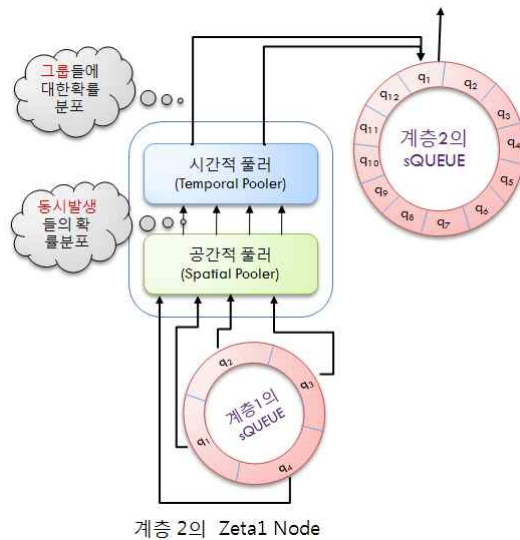


그림 9. 레벨2의 Zeta1Node의 동작

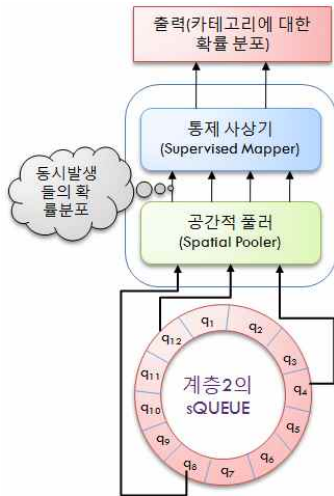


그림 10. Zeta1TopNode의 동작

sQUEUE₂의 출력을 입력으로 받아 범주들에 대한 확률 분포를 발생시켜 확률적으로 높은 범주에 사상시킨다. 즉, 현재 입력된 주식 가격 동향에 대한 상승, 유지, 하강 중의 하나의 결과를 발생시킨다. 이로써 주식 가격의 변동을 예측하여 그 결과를 발생시킨다.

5. 실험 및 분석

5.1 실험 데이터

5.1.1 훈련 데이터

훈련 데이터는 과거 일정 기간(2010. 10. 4~2010. 10. 8)에 KOSPI 시장에서 수집된 대기업 및 중소기업 12개사의 주식 가격 데이터를 각 1분당 체결가를 기준으로 추출하였다.

훈련을 위해 추출된 데이터는 레벨1의 노드 개수가 12개이고 각 노드 당 5분간의 5개의 데이터가 들어갈 수 있도록 60분 간격으로 끊어서 패턴을 만들었다. 만들어진 패턴은 -1에서 1사이의 값을 가지도록 정규화하여 상승, 유지, 하강 3단계의 카테고리로써 구성하였다. 훈련 데이터의 예는 표 1과 같다. 패턴 1은, 상승과 하강을 반복하지만, 60분 동안의 전체적인 경향은 상승으로 판정될 수 있는 패턴이다. 패턴 2은 상승과 하강을 반복하지만 60분 동안의 전체적인 경향은 하강으로 판정될 수 있는 패턴이다. 패턴 3은 상승과 하강을 반복하지만 60분 동안의 전체적인 경향은 로 판정될 수 있는 패턴이다.

표 1. 훈련 데이터의 예

구분	패턴 예	비 고
패턴1		사용자가 상승으로 판정한 훈련 데이터 패턴의 예의 일부
패턴2		사용자가 하강으로 판정한 훈련 데이터 패턴의 예의 일부
패턴3		사용자가 유지로 판정한 훈련 데이터 패턴의 예의 일부

5.1.2 테스트 데이터

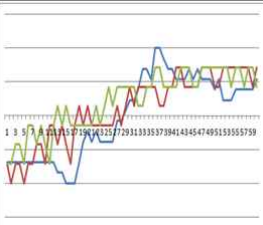
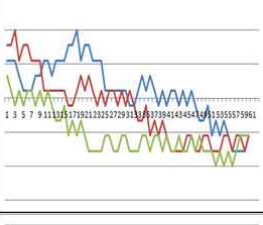
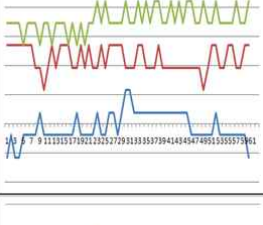
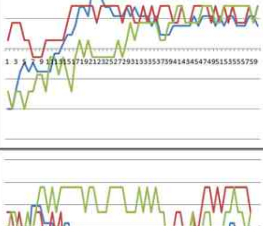
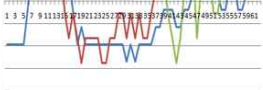
테스트 데이터는 수집된 데이터 안에서 훈련에 사용되지 않은 데이터 중 추출하였다.

테스트 데이터는 다섯가지 패턴으로 준비하였다. 테스트 데이터의 예는 표 2와 같다. 첫 번째 패턴인 패턴1'은, 상승과 하강을 반복하지만, 60분 동안의 전체적인 경향은 상승으로 판정될 수 있는 패턴이다. 패턴 2'은, 상승과 하강을 반복하지만, 60분 동안의 전체적인 경향은 하강으로 판정될 수 있는 패턴이다. 패턴 3'은, 상승과 하강을 반복하지만, 60분 동안의 전체적인 경향은 유지로 판정될 수 있는 패턴이다. 패턴 4'은, 상승과 하강을 반복하지만, 60분 동안의 전체적인 경향은 상승에서 유지로 변경된다고 판정될 수 있는 패턴이다. 패턴 5'은, 상승과 하강을 반복하면서 60분 동안의 전체적인 경향이 상승, 하강, 유지가 혼합된 패턴이다.

5.2 실험 결과

훈련이 끝난 HTM 네트워크의 훈련에 사용된 데이터를 통한 인식 정확도는 약 97.3%이며 각 테스트

표 2. 테스트 데이터의 예

구분	패턴 예	비 고
패턴 1'		상승으로 판정할 가능성이 높은 테스트 데이터 패턴의 예
패턴 2'		하강으로 판정할 가능성이 높은 테스트 데이터 패턴의 예
패턴 3'		유지로 판정할 가능성이 높은 테스트 데이터 패턴의 예
패턴 4'		상승에서 유지로 변경되는 테스트 데이터 패턴의 예
패턴 5'		상승, 유지, 하강이 혼합된 테스트 데이터 패턴의 예

유형에 따른 데이터의 인식 결과는 표 3과 같다.
 상승, 유지, 하강의 테스트 데이터 형태가 패턴 1'에 속하는 경우는 약 96%의 패턴에 대해서 상승으로 판정하였고, 패턴 2'에 속하는 경우는 약 91%의 패턴에 대해 하강으로 판정하였으며, 패턴 3'에 속하는 경우는 약 94%의 패턴에 대해서 유지로 판정하였다.
 상승, 유지, 하강의 테스트 데이터 형태가 패턴 4'에 속하는 경우는 약 53%의 패턴에 대해서 상승으로 판정하였고, 약 47%의 패턴에 대해서 유지로 판정하였다. 그리고 하강으로 판정된 패턴은 없었다. 이는

표 3. 테스트 데이터에 대한 인식률

패턴유형	판정	인식비율
패턴1'	상승 하강 유지	약 96%
패턴2'	상승 하강 유지	약 91%
패턴3'	상승 하강 유지	약 94%
패턴4'	상승 하강 유지	약 53% 0% 약 47%
패턴5'	상승 하강 유지	약 46% 약 21% 약 33%

주식 투자자가 판정할 때에도 유사하게 판정할 것으로 판단된다. 특히 하강으로 판정된 경우는 없었기 때문에 패턴 4'의 인식정확도도 매우 높은 것으로 판단된다.

상승, 유지, 하강의 테스트데이터 형태가 패턴 5'와 같이 상승, 유지, 하강이 혼재되어 있는 테스트데이터의 경우, 약 46%의 패턴에 대해서 상승, 약 21%의 패턴에 대해 하강, 약 33%의 패턴에 대해 유지를 판정하였다. 이 또한 주식 투자자가 판정할 때에도 유사하게 판정할 것으로 보여 패턴 4'의 인식정확도도 매우 높은 것으로 판단된다. 따라서 이 결과는 연속예측시스템과 주식투자시스템 개발의 핵심 기술로 사용할 수 있다.

6. 연속 예측 시스템 구현 결과

개발된 시스템의 연속 예측은 실제 주식 가격을 데이터베이스에 1분 단위로 들어가며 연속 예측을 하였다. 그림 11은 연속 예측 결과이다. 그림 11의 12:18분 사례에서 StreamSensor output 이후의 값들은 주식이가격 데이터의 선 처리 후 스트림센서를 통해 출력되는 값을 나타낸다. 이 값은 레벨1 노드의 입력이 된다. Result 뒤의 값은 ZetalTop노드의 출력으로서 최종 예측 결과이다. 그리고 값이 의미하는 추세를 Up, Down, Stable로서 표시해준다. 이러한 예측 결과는 주식거래 정보를 관리하는 모듈로 전달하여

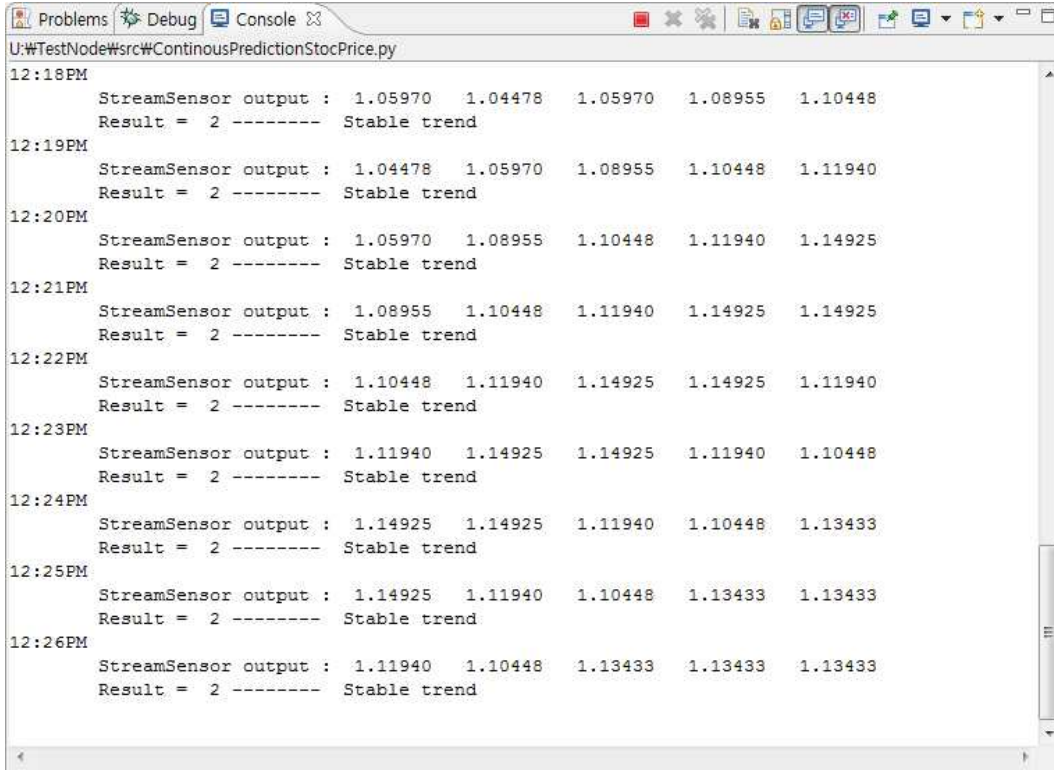


그림 11. 연속 예측 결과 값

사용자가 알아보기 쉽게 그래픽과 문자로서 보여 준다.

주식 거래 정보창에서는 데이터베이스로 부터 가격 데이터를 읽고 HTM 네트워크로 부터 인식 결과를 전달받는다. 그리고 화면에 현재가격 정보 및 전일 가격정보, 그리고 현재까지의 가격변동 그래프 및 예측 결과 등 사용자의 주식거래를 도와 줄 수 있는 기본 정보들을 보여준다.

그림 12는 13시 30분의 예측결과이다. 우측 최상단에는 현재 선택된 종목을 나타내고 좌측에는 금일 추가정보를 현재가, 시가, 최고가, 최저가, 전일대비 순으로 나타낸다. 금일 정보 아래 전일의 종가와 최고, 최저가를 보여준다. 우측에는 현재가를 기준으로 60분 이전까지의 값을 그래프로 보여주며 그래프 아래쪽으로는 현재 시간 및 현재 상태를 상승, 하강, 보합으로 보여준다.

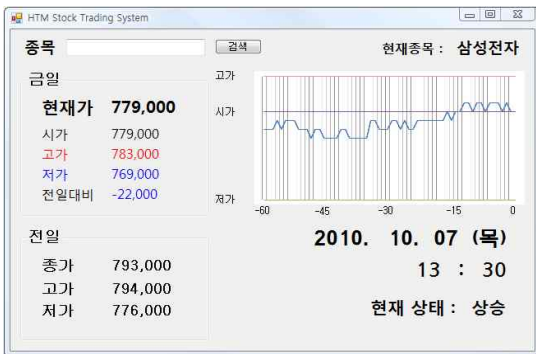


그림 12. 연속 예측 결과(13:30)

7. 결론 및 향후 과제

주식 가격 동향 예측에는 과거의 주식 가격 변동 추세를 바탕으로 미래 변화를 계속해서 예측하는 연속 예측이 필요하다. 이를 위해 연속적인 주식 가격 동향 예측을 도울 수 있는 시스템을 개발하였다. 본 논문에서는 HTM 기술을 기반으로 연속 예측에 적합한 새로운 HTM 네트워크 구성 방법을 제시하여 이를 이용한 주식 가격 동향 예측 시스템을 개발하였다. 실제 주식 가격 동향을 입력하여 실험한 결과

HTM에서의 추세 판단은 정확하였다.

본 시스템에서 나타나는 연속 예측의 결과들을 결합함으로써 주식 가격 동향 추세에 대해서도 보다 신뢰성있는 예측을 할 수 있는 결합 알고리즘에 대한 연구가 필요하다. 또한 이를 이용한 종합적인 자동 주식 거래 시스템의 개발 또한 필요하다. 그리고 주식 가격 데이터가 시계열 데이터의 한 종류이므로 이와 유사한 심전도, 심박수, 음파 등의 시계열 데이터에 대한 모니터링 및 예측에도 본 시스템의 활용이 가능하다.

참 고 문 헌

- [1] J. Hawkins, *On Intelligence*, New York, Henry Holt, 2005.
- [2] J. Hawkins and D. George, *Hierarchical Temporal Memory*, Numenta, Inc., 2007.
- [3] 하유민, 김상욱, 원정임, 박상현, 윤지희, “주가 예측을 위한 규칙 탐사 및 매칭,” 정보과학회논문지:데이터베이스, 제34권 제3호, pp. 179-192, 2007.
- [4] 이종우, 김유섭, 김성동, 이재원, 채진석, “패턴 매칭과 규칙 생성에 기반한 2단계 주식 트레이딩 시스템,” 정보처리학회논문지B, 제10-B권 제3호, pp. 257-264, 2003.
- [5] K. Lee and G. Jo, “Expert System for Predicting Stock Market Timing using a Candlestick Chart,” *Expert system With Applications*, 16, pp. 357-364, 1999.
- [6] 오유진, 김유섭, “2단계 하이브리드 주가 예측 모델 : 공적분 검정과 인공 신경망,” 정보처리학회논문지B, 제14-B권, 제7호, pp. 531-540, 2007.
- [7] J. Wana and S. Chanb, “Stock Market Trading Rule Discovery using Two-Layer Bias Decision Tree,” *Expert Systems with Applications*, 30, (4), pp. 605-611, 2006.
- [8] A. Fan and M. Palaniswami, “Sock Selection Using Support Vector Machine,” In *Proceeding International Joint Conference on Neural Networks*, pp. 1973-1983, 2001.
- [9] 황희수, “퍼지모델을 이용한 일별 주가 예측,” 정보처리학회논문지B, 제15-B 제6호, pp. 603-608, 2008.
- [10] J. Han and M. Kamber, “Data Mining: Concepts and Techniques,” Morgan Kaufman, 2001.
- [11] Y. Yang, X. Wu, and X. Zhu, “Proactive-Reactive Prediction for Data Streams,” Technical Report CS-05-03, Department of Computer Sciences, University of Vermont, USA, 2005.
- [12] 배선갑, 한창영, 서대호, 김성진, 배종민, 강현석, “HTM 기반의 소리 연식을 이용한 부품의 양·불량 판별 시스템,” 한국멀티미디어학회논문지, Vol.13, No.10, pp. 1494-1505, 2010
- [13] Haibao Nan, Bae Sun-Gap, Bae Jong-Min, and Kang Hyun-Syug, “A Recognition System for Multi-Form Korean Characters Based on Hierarchical Temporal Memory,” *한국멀티미디어학회논문지*, Vol.12, No.12, pp. 1718-1727, 2009
- [14] D. George and B. Jaros, *The HTM Learning Algorithms*, Numenta Inc., 2007.
- [15] D. George and B. Jaros., *Zatal Algorithms Reference*, Numenta, Inc., 2007.



서 대 호

2009년 경상대학교 컴퓨터과학과 졸업(학사)
2010년 경상대학교 대학원 컴퓨터과학과 졸업(석사)
관심 분야: XML, 기계학습, 데이터베이스 통합



강 현 석

1981년 동국대학교 전자계산학과 졸업(학사)
1983년 서울대학교 대학원 전산학과 졸업(석사)
1989년 서울대학교 대학원 전산학과 졸업(박사)

1981년~1984년 2월 한국전자통신연구원 연구원
1984년 3월~1993년 2월 전북대학교 전임강사, 부교수, 교수
1993년 3월~현재 경상대학교 컴퓨터과학과 교수
관심분야: 멀티미디어, 내장형 데이터베이스, 지능시스템



배 선 갑

1987년 경북대학교 물리학과 졸업(학사)
2004년 경상대학교 대학원 컴퓨터과학과 졸업(석사)
2010년 현재 경상대학교 대학원 컴퓨터과학과 졸업(박사)

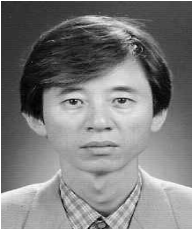
관심 분야: 패턴인식, 지능시스템, 내장형 데이터베이스



배 종 민

1980년 서울대학교 수학교육과 졸업(학사)
1983년 서울대학교 대학원 계산통계학과 졸업(석사)
1995년 서울대학교 대학원 계산통계학과 졸업(박사)

1982년~1984년 한국전자통신연구원 연구원
1997년~1998년 Virginia Tech. 객원연구원
1984년~현재 경상대학교 컴퓨터과학과 교수
관심분야: XML, 프로그래밍언어, 지능시스템



김 성 진

1979년 경북대학교 공과대학 전자공학과 졸업(공학사)
1981년 경북대학교 대학원 전자공학과 졸업(공학석사)
1995년 영남대학교 대학원 컴퓨터공학전공(공학박사)

1983년~1985년 삼성전자(주) 시스템개발부
1996년~1997년 University of California, Santa Cruz 연구교수
1985년 3월~현재 연암공업대학 컴퓨터공학과 교수
관심분야: 멀티미디어, 병렬처리, 운영체제