

<http://dx.doi.org/10.7236/JIIBC.2013.13.1.101>

JIIBC 2013-1-15

정보기준과 효율적 자료길이를 활용한 시계열자료 운동패턴 예측 연구

A Study on Prediction the Movement Pattern of Time Series Data using Information Criterion and Effective Data Length

전진호*, 김민수**

Jin-Ho Jeon, Min-Soo Kim

요약 현실세계에서는 광범위한 업무영역에서 대용량의 시계열자료들이 실시간으로 발생되고 있다. 하지만 동적인 특징으로 표현되는 시계열자료들의 이해와 설명을 위한 최적의 모형을 결정하는 일은 쉽지가 않다. 이러한 시계열자료들의 특징을 잘 설명할 수 있는 모형을 추정하기 위하여 본 연구에서는 시계열데이터의 모형추정에 적합한 은닉마코프모형을 통해 시계열자료의 장, 단기 예측모형을 추정하였고 이를 통해 미래의 운동패턴예측을 확인하였다. 실제 주식시장의 여러 자료들을 통해 최적의 모형추정을 위한 정보기준과 가장 효율적인 자료길이를 통해 모형의 상태수를 정확하게 추정하는지를 확인하였다. 실험결과 유효한 상태의 수 추정과 단기의 예측이 장기예측보다 유사운동패턴 예측률이 더욱 유사함을 확인하였다.

Abstract Is generated in real time in the real world, a large amount of time series data from a wide range of business areas. But it is not easy to determine the optimal model for the description and understanding of the time series data is represented as a dynamic feature. In this study, through the HMM suitable for estimating the short and long-term forecasting model of time-series data to estimate a model that can explain the characteristics of these time series data, it was estimated to predict future patterns of movement. The actual stock market through various materials, information criterion and optimal model estimation for the length of the most efficient data was found to accurately estimate the state of the model. Similar movement patterns predictive than the long-term prediction is more similar to the short-term prediction of the experimental result were found to be.

Key Words : Information Criterion, Effective Data Length, Time Series Data, Prediction

1. 서 론

우리가 살아가는 실세계에서 생산되는 자료의 형태는 정보통신기술의 발전에 따라 다양한 업무 영역에서 시계열자료의 형태로 광범위하게 발생되고 있다. 이렇게 발

생되어지는 시계열자료들이 어떤 경향이나 패턴에 의해 움직인다고 가정하였을 때, 시계열자료들의 동적인 변화를 잘 설명할 수 있는 모형이 결정되어진다면 이를 적용하여 미래의 동적인 움직임에 대하여 효과적인 예측이 가능할 것이다. 하지만 현실세계를 나타내는 시계열자료

*정회원, 관동대학교 경영학과

**정회원, 관동대학교 무역학과(교신저자)

접수일자 : 2013년 1월 4일, 수정완료 : 2013년 2월 4일

게재확정일자 : 2013년 2월 8일

Received: 4 January 2013 / Revised: 4 February 2013 /

Accepted: 8 February 2013

*Corresponding Author: mskim@kd.ac.kr

Dept. of International Trade, Kwandong University, Korea

에 대한 최적의 분석모델을 결정하는 과정은 매우 어려운 작업이다. 특히, 경제시스템에서 발생되어지는 시계열 자료의 형태들에 대한 분석모형의 형성과정은 더욱 더 단순하지 않다. 이유는 경제시스템에 영향을 미치는 변수가 다양하고 예측 불가능한 예외적인 상황과 같은 불확실성의 요소들이 시시각각으로 발생되어지기 때문이라고 할 수 있다. 하지만 Occam's Razor의 원리처럼 아무리 복잡한 현상을 나타내는 시계열자료라도 최소의 자료로서 설명되어질 수 있을 것이다.

본 연구는 Occam's Razor처럼 동적인 특징을 갖는 시계열자료들에 대한 가장 적합한 분석모형을 결정하는데 필요한 적절한 효율적 자료길이를 정보기준을 통하여 추정하였으며 실험에 필요한 자료는 실제의 시계열자료, 즉 주식데이터를 적용하였다.

본 연구는 제1장 서론에 이어 제2장 관련연구에서는 시계열자료의 분석, 예측을 위한 모형 결정 방법론들을 살펴본다. 제3장에서 시계열자료의 모형 결정에 적합한 은닉마야코프모형(HMM)과 모델의 결정을 위한 정보기준과 효율적 자료 길이에 대한 방법론을 제시한다. 제4장에서 실제의 주식자료를 통한 모형에 필요한 상태 수와 자료의 길이를 추정과 예측을 통해 제시한 방법론의 유효성을 살펴본다. 제5장에서는 연구의 결과와 향후 연구 방향을 제시한다.

II. 관련 연구

동적인 특징으로 표현되는 시계열자료 분석은 기본적 분석(fundamental analysis)과 기술적 분석(technical analysis)의 연구영역으로 구분된다^[1].

기본적 분석이란 금융자산의 경제적 가치를 평가하는 것이다. 많은 연구에서 자산 가격이 랜덤윌을 따른다고 믿음으로서 기술적 분석은 소외되어 왔으나 최근에 일부 금융자산 가격이 랜덤윌을 따르지 않는다는 사실을 입증했다^[1].

기술적 분석방법은 자산가격의 과거 움직임이 미래의 자산 가격 움직임을 예측하는 데 의미 있는 정보를 제공한다는 가정에 바탕을 두고 있다. 실무상 기술적 분석방법이 보편적인 것과는 대조적으로 기술적 분석방법의 유용성을 제시한 연구는 많지 않다. 그러므로 본 연구에서는 모델기반의 예측을 통해 기술적 분석방법의 유효성을

평가하고자 한다. 기술적 분석은 모델기반 방법론이다.

자기회귀모델과 이동평균모델을 비롯하여 신경망^[2], ARMA(ARIMA)^[3], 마야코프모델, 은닉마야코프모델^[4,7], 상호정보^[5]를 이용한 방법들이다.

자기회귀모델과 이동평균모델은 자료들의 시점사이의 관계를 추정하는 방법이다. 자기회귀모델에서 종속변수는 현재 값이고 독립변수는 자기회귀의 차수로 불리는 N 개의 이전 값들로 이루어진다. 기본 개념은 시계열자료의 한 점에서의 값은 그 시점 이전과 이후의 값과 밀접한 관계를 갖는다는 사실에 근거를 둔다. 이 방법은 자연계나 사회적 현상의 다양한 형태를 모델링하고 예측하는데 자주 사용된다^[6]. 신경망은 많은 연구에서 시계열현상을 예측하는 영역에 적용되어 왔으나 모델의 구조가 알려져 있다는 것과 모델의 해석이 어렵다는 점들도 인하여 일반적인 시계열데이터의 모형결정에는 적합하지 않다^[2]. ARMA 모델은 계절성, 비정상성 등 다른 요소에 의해 예측에 한정된 역할을 하게 되는 단점이 있다^[3]. 상호정보를 포함한 다른 통계적 방법도 항상 고도의 기술적 제한 조건 및 적용 가능한 환경에 대한 조건을 제시하지 않을 경우 모델링은 매우 제한적이다^[5]. 마야코프모델은 간단하여 계산이 효율적이며 방향성 그래프로 표현되어 이해와 해석이 쉬우며 이산 적으로 표현되는 자료의 모델링에 적합하지만 연속적인 값을 갖는 시계열자료의 특징을 묘사하는 데 적합하지 않다는 제약점을 가지고 있다^[4]. 은닉마야코프모델은 관측되는 관측 자료에 대한 상태의 확률함수로 이루어져 연속적인 시계열자료의 표현 모델링에 적합하다. 이유는 각 상태에서 특징들에 대한 적합한 확률함수를 사용하여 연속적인 값을 갖는 시간적 특징을 갖는 데이터를 쉽게 처리하며 데이터의 묘사가 쉽기 때문이다^[4,7].

이에 따라 본 연구에서는 시계열자료의 운동패턴 예측모형의 추정과 예측에 은닉마야코프모델을 적용한다.

III. 은닉마야코프모형(HMM)

1. 은닉마야코프모형(Hidden Markov Model)

은닉마야코프모델은 은닉상태를 예측하기 위하여 실제적인 관측을 통해서 변화되는 통계적인 특징들을 확률적으로 모델링하는 방법이다. 은닉마야코프모델은 M 개의 상태들을 갖으며, 상태별로 K 특징들로 묘사되며 모

형의 구성은 초기상태확률, 상태전이확률, 방출확률의 3개의 확률 집합으로 이루어진다^[7,8].

은닉마야코프모델은 주어진 시계열자료에 대한 확률 추정 문제, 최적상태 순서 결정문제, 모델의 확률을 최대로 하는 매개변수의 추정문제가 있다. 확률추정의 문제는 식(1),(2)처럼 전향, 후향의 절차를 통해 구할 수 있다.

$$\alpha_t(i) = \left(\sum_{j=1}^N \alpha_{t-1}(j) a_{ij} \right) \cdot P_i(O_t) \quad (1)$$

$$\beta_t(i) = \sum_{j=1}^M a_{ij} \cdot P_i(O_t) \cdot \beta_{t+1}(j) \quad (2)$$

최적상태 순서 결정문제는 주어진 관측 열에 대응하는 최적의 상태 열을 찾는 즉, 가장 큰 확률을 갖는 상태 경로를 찾는 방법이다.

모델의 확률을 최대로 하는 매개변수의 추정문제는 새로 구성된 모델과 이전 모델이 관측 열을 생성하는 확률의 차이가 특정 값 이상이 될 때까지 모델의 매개변수를 변경하면서 새로운 모델을 형성해 나가는 바움-웰치 알고리즘을 적용하여 수행된다. 모델의 매개변수를 재추정 하는데 모델 매개변수 갱신은 다음의 식(3),식(4),식(5)과 같다.

$$a'_{ij} = \frac{\sum_{t=1}^L \alpha_t(i) a_{ij} P_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^L \sum_{j=1}^M \alpha_t(i) a_{ij} P_j(O_{t+1}) \beta_{t+1}(j)} \quad (3)$$

$$\mu'_{ik} = \frac{\sum_{t=1}^L \alpha_t(i) \cdot \beta_t(i) \cdot O_t^k}{\sum_{t=1}^L \alpha_t(i) \beta_t(i)} \quad (4)$$

$$\sigma'_{ik} = \sqrt{\frac{\sum_{t=1}^L \alpha_t(i) \cdot \beta_t(i) \cdot (O_t^k - \mu'_{ik})^2}{\sum_{t=1}^L \alpha_t(i) \beta_t(i)}} \quad (5)$$

2 모형 상태 수 결정위한 정보기준

주어진 시계열자료가 완전한 경우에는 한계우도의 계산이 매우 간단하지만 불완전한 경우에는 정확한 해를 얻는 것은 복잡하다. 그러므로 일반적으로 근사기법들이 사용된다. 근사기법으로 몬테카를로, 라플라스 근사법 등이 있다. 이 방법들은 정확한 결과 값에 수렴하지만 계산이 복잡하여 계산비용이 많이 드는 단점이 있다. 이 방법

들 보다 정확도는 떨어지지만 계산 복잡도를 줄여 효율성을 준 정보기준(Information Criterion)이 있다^[9]. 정보 기준은 다음 식(6),(7)으로 표현된다.

$$\log P(X|M) \approx \log P(X|\hat{\theta}, M) + \log P(\hat{\theta}|M) + \frac{d}{2} \log(2\pi) - \frac{1}{2} \log |A| \quad (6)$$

$$\log P(M|X) \approx \log P(X|M, \hat{\theta}) - \frac{d}{2} \log N \quad (7)$$

식(6),(7)에서 d 는 모형구성의 매개변수 수를 나타낸다. N 은 자료들의 수를 $\hat{\theta}$ 는 모형에 대한 한계우도의 매개변수이다. 식(7)에서 첫 번째 항은 시계열자료를 잘 설명할 수 있는 상세한 모델을 선호하는 우도 항과 두 번째 항은 매개변수의 수를 작게 하여 모형을 일반화시키는 항으로 구성되며 이 두 항이 조화를 통해 효율적인 모형이 결정되도록 유도된다. 즉, 주어진 정보기준을 통해 하나의 상태에서부터 시작하여 하나씩 증가해 가며 가장 높은 기준 값을 갖는 상태 수가 유효한 상태수를 나타내는 것으로 고려한다.

IV. 실험

시계열자료 운동패턴의 예측모형 결정을 위한 자료로 2009년 1월2일부터 12월25일, 2010년 1월2일부터 12월 25일, 2011년 1월 2일부터 12월25일까지의 3개년의 코스피(KOSPI) 지수, 반도체와 반도체장비 업종에서 대, 중, 소형 개별종목인 삼성전자, SK하이닉스 그리고 동부하이텍의 시계열자료를 선정하였다. 개별종목에서 대, 중, 소형으로 선정한 이유는 우리나라의 대표적인 우량종목으로서 삼성전자는 주가의 형성에 있어 외부의 시장 환경에 덜 민감할 것으로 생각되어 안정적인 예측모델이 가능할 것으로 생각되어 선택하였으며 중, 소형주는 예측률이 대형주와 차이가 있는지 확인함으로써 모델의 안정성을 확인하기 위하여 선정하였다.

실험에서 예측기간은 단기(10일)와 장기(30일)예측을 하였다.

단기(10일)예측 모형결정에 필요한 효율적인 자료 길이를 추정해보기 위해 4가지 각 지수별로 2011년1월2일부터 10주(50일)의 자료의 길이를 10, 20, 30, 40, 50일 단위로 모형을 결정해보고 예측하였다. 그림 1은 각 지수들

중 KOSPI 시계열자료의 운동패턴을 보여주고 있다.

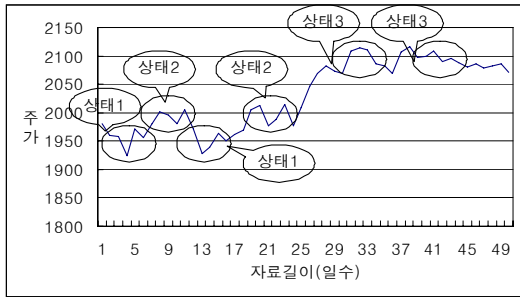


그림 1. 2011년 자료길이50 - KOSPI
Fig. 1. 2011 Data Length50 - KOSPI

정보기준의 유효성은 인공적인 실험데이터를 통한 이전의 연구에서 확인하였다⁹⁾. 본 연구에서는 실제의 주식 자료에 대하여서도 정보기준을 통해 정확한 상태 수를 추정하는지 확인한다.

그림 1에서 KOSPI 자료의 길이 10, 20일 까지는 2개의 상태를 30, 40, 50일 에서는 3개의 상태를 갖는 것으로 나타난다. 삼성전자의 경우, 자료에서 10일까지는 2개의 상태, 20일 까지는 3개의 상태, 30, 40, 50일에서는 4개의 상태 수를 갖는다. 각 지수별 자료의 길이별로 존재하는 상태 수와 정보기준을 통하여 추정된 상태 수는 표 1에 정리하였다. 표 1의 결과를 살펴보면, 10일과 20일에서는 모두 잘못된 상태 수 추정결과를 보여주고 있으며 30일, 40일, 50일에서는 모두 올바른 상태 수 추정 결과를 보여 주고 있다. 이유는 기간이 짧을수록 운동패턴에 대한 규칙성의 횡수가 적기 때문이다. 즉, 짧은 기간의 매 변곡점마다 상태로 보게 되면 상태 수가 많게 추정된다.

표 1. 정보기준에 의한 자료 상태 수 예측
Table 1. Data State Number Prediction by IC

종목 \ 길이	2주 (10일)	4주 (20일)	6주 (30일)	8주 (40일)	10주 (50일)
KOSPI	2(4)	2(3)	<u>3(3)</u>	<u>3(3)</u>	<u>3(3)</u>
삼성전자	2(3)	3(2)	<u>4(4)</u>	<u>4(4)</u>	<u>4(4)</u>
하이닉스	2(4)	2(3)	<u>3(3)</u>	<u>3(3)</u>	<u>3(3)</u>
동부하이텍	2(3)	2(3)	<u>3(3)</u>	<u>4(4)</u>	<u>4(4)</u>

위의 결과에 따라 그림 2는 상태 수를 정확히 추정된 KOSPI 자료길이 30일, 40일, 50일을 통해 생성된 예측모

형으로 각각 향 후 10일의 예측결과를 실제의 운동패턴 과 비교하여 나타냈다.

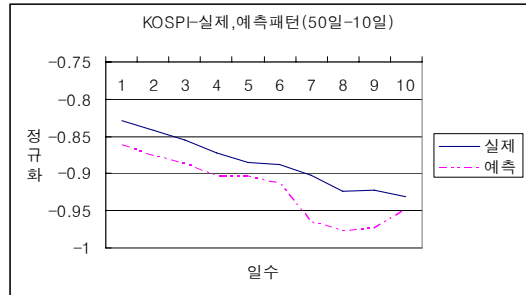
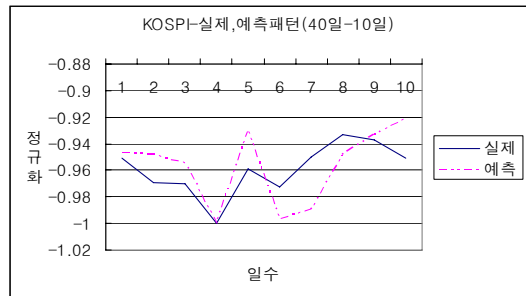
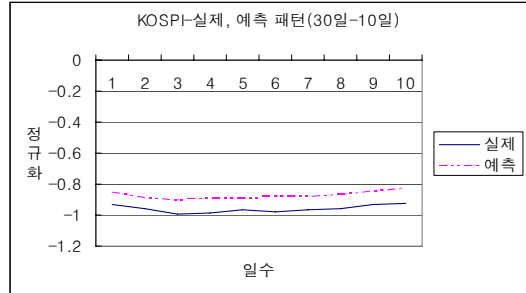


그림 2. 자료길이에 따른 예측패턴
Fig. 2. Prediction Pattern by Data Length

각 자료길이 모두 실제의 패턴과 유사한 운동패턴을 보여주고 있다. 30일의 자료로 모형을 생성 후 향 후 10일을 예측한 그래프는 3일 하락 후 점차 완만하게 상승하는 매우 유사한 운동패턴을 보여주고 있다. 40일과 50일의 자료로 모형 생성 후 향 후 10일을 예측한 그래프는 모두 8일까지는 매우 유사한 운동패턴을 9, 10일에서는 반대의 운동패턴을 보여주고 있다.

표 1과 그림 2의 결과를 종합해 보면 상태 수 결정과 유사 운동패턴 예측 모두 30일, 40일, 50일의 자료길이에서 유효한 결과를 보여준다. 결과적으로 10일의 단기예측을 고려한다면 계산비용 측면에서 30일의 자료길이가

가장 효율적인 자료의 길이라고 할 수 있다. 표 1에 따라 KOSPI, 삼성전자, SK하이닉스, 동부하이텍 각 지수별로 30일, 40일, 50일의 자료길이를 통해 10일 예측모형을 추정하였으며 이를 통해 생성된 모형을 기준으로 랜덤하게 각 지수들의 모형별로 운동패턴 예측을 30회씩 시행하였으며 유사패턴 예측률은 표 2와 같다.

표 2. 자료길이에 따른 유사패턴 예측률(단기)
Table 2. Prediction Rate the Similar Pattern by Data Length(Short-term)

종목 \ 길이	6주 (30일)	8주 (40일)	10주 (50일)
KOSPI	30(27)=0.90	30(26)=0.86	30(25)=0.83
삼성전자	30(28)=0.93	30(28)=0.83	30(27)=0.90
하이닉스	30(26)=0.86	30(25)=0.83	30(25)=0.83
동부하이텍	30(25)=0.83	30(21)=0.70	30(18)=0.60

예측패턴과 실제패턴의 유사성을 확인하는 과정은 예측과 실제 사이의 차이인 예측오차의 크기에 따라 평가하는 사후평가로 평균제곱오차(Mean Square Error)를 적용하였다. 평균제곱오차는 실제와 예측 모두 정규화 후에 일별변화의 차이를 고려한다. 두 운동패턴 사이의 평균치 0.5를 기준으로 평균치 임계점보다 낮을수록 유사함을 높을수록 비유사성을 나타낸다.

모든 지수들 모두 종합적으로 30일의 자료길이에 있어서 가장 높은 유사 패턴율을 보여주고 있으며 자료의 길이가 길어질수록 유사 패턴율이 점차 감소하는 것으로 나타났다. 이러한 이유는 자료의 길이가 길어질수록 주식 시장에서 발생하는 많은 불규칙적인 외적변수들이 반영되어 운동패턴에 내재되는 것으로 고려된다. 한 가지 특이한 점은 삼성전자와 SK하이닉스처럼 중, 대형주들은 업종지수를 이끌어가는 종목이라 동부하이텍처럼 소형주보다는 외부환경에 영향을 받는 것이 적은 것으로 나타난다. 동부하이텍은 30일 자료길이의 경우에서는 다른 종목들과 큰 차이는 없지만 자료길이가 길어질수록 급격하게 유사 패턴율이 감소하는 결과를 보여주고 있다.

단기(10일) 예측과 마찬가지로 장기(30일) 예측을 위한 효율적인 자료길이를 추정해보기 위해 4가지 각 지수별로 2011년1월2일부터 자료의 길이를 90, 120, 150, 180, 210일 단위로 모형을 결정하고 예측하였다.

표 3. 정보기준에 의한 자료 상태 수 예측
Table 3. Data State Number Prediction by IC

종목 \ 길이	90일	120일	150일	180일	210일
KOSPI	4(3)	4(3)	5(4)	<u>5(5)</u>	<u>6(6)</u>
삼성전자	4(3)	4(3)	<u>4(4)</u>	<u>5(5)</u>	<u>6(6)</u>
하이닉스	4(3)	4(3)	<u>5(5)</u>	<u>6(6)</u>	<u>6(6)</u>
동부하이텍	5(4)	6(5)	7(6)	<u>7(7)</u>	<u>8(8)</u>

먼저 시계열자료의 길이에 따른 상태 수 추정결과는 표 3과 같다. 코스피와 동부하이텍은 180일 이상에서 삼성전자와 SK하이닉스는 150일에서 올바른 상태 수를 추정하는 것을 확인하였다. 삼성전자와 SK하이닉스처럼 업종을 이끌어가는 대형 종목들은 시장 환경으로부터 받는 영향이 소형종목(동부하이텍)보다 변동성이 적기 때문에 짧은 시계열자료 길이를 통해서도 충분히 상태 수를 반영하는 것으로 생각된다. 장기(30일) 예측을 위해서는 모든 자료에서 최소 자료의 길이가 180일 이상이면 정확한 상태 수 추정을 통한 예측모형의 결정이 가능할 것이며 대형종목들은 150일에서도 정확한 상태 수 추정이 가능하였다.

표 4. 자료길이에 따른 유사패턴 예측률(장기)
Table 4. Prediction Rate the Similar Pattern by Data Length(Long-term)

종목 \ 길이	150일	180일	210일
KOSPI	—	30(23)=0.77	30(24)=0.80
삼성전자	30(25)=0.83	30(25)=0.83	30(26)=0.86
하이닉스	30(24)=0.80	30(25)=0.83	30(25)=0.83
동부하이텍	—	30(22)=0.73	30(21)=0.70

각 자료의 길이별로 30일 예측모형을 추정하였으며 이를 통해 생성된 모형을 기준으로 랜덤하게 각 자료별로 운동패턴 예측을 30회씩 시행했으며 결과는 표 4와 같다. 결과 장기(30일) 예측이 단기(10일) 예측보다 유사패턴 예측률이 낮은 점을 확인할 수 있다. 이유는 자료의 길이가 길어질수록 외부의 시장 환경요소에 많은 영향을 받기 때문에 예측률이 낮아진다고 고려된다. 표 5는 위의 방식으로 2009년에서 2011년까지의 3개년도의 자료를 통한 실험결과를 나타낸다.

실험결과 모형 추정과 예측에 적용된 시계열자료의 길이가 장기(30일) 보다는 단기(10일)의 예측모형이 유사 운동패턴을 잘 예측하는 것으로 확인되었다. 위의 결과를 통해 모형 추정과 예측에 적용된 시계열자료의 길이가 길어질 때보다 짧은 기간의 시계열자료에서 예측되어진 운동패턴들이 실제의 운동양태와 좀 더 유사한 운동패턴을 나타내는 것은 시계열자료의 길이(일수)가 길어질수록 시장에서 발생하는 수많은 불규칙적인 외적변수가 운동패턴에 반영되는 것으로 고려된다.

표 5. 실험데이터의 길이와 유사패턴의 비율
Table 5. Data length and rate of similar pattern

	30일-10일	180일-30일
유사패턴 예측비율 (3개년 자료별로 30회씩 예측한 비율)	30회 중 26.5회	30회 중 23.75회
	88%	79%

V. 결 론

본 연구는 다양한 영역에서 발생하는 시계열자료의 이해와 설명 그리고 미래의 예측문제에 적용 가능한 예측모형의 결정에 관한 방법을 제시하였다. 시계열자료의 모형에 적합한 은닉마야코프모델을 통해 모형의 추정과 예측에 적용하였다.

실험을 통하여 시계열자료 내에 내재하는 상태 수 추정의 문제와 이를 통해 장, 단기 시계열자료의 운동패턴 예측문제에 적용하였다. 실험결과 예측모형 결정의 상태 수를 추정하는 부분에서는 시간과 비용을 절약과 모형의 확장성을 고려하여 정비기준을 적용한 결과 실제의 자료 내에 내재하는 의미 있는 상태 수 추정결과를 나타내는 것을 확인하였다. 또한 이를 통한 모형의 생성을 통해 미래의 운동양태를 예측한 결과 예측의 운동패턴들이 실제의 운동패턴들과 매우 유사한 운동패턴을 예측하는 것으로 확인할 수 있었다.

본 연구에서 제시된 예측모형 결정 방법론을 적용한다면 생산 공정, 주가와 환율 예측, 고객 맞춤 마케팅 등 기업들의 다양한 업무영역에서 미래의 정확한 예측 문제들에 적용을 통해 기업의 목표를 극대화할 수 있을 것이다.

향 후 더 다양한 부분의 일반화된 시계열자료에 적용 가능한 연구를 통하여 일반적이고 유연한 분석, 예측모형을 세울 수 있는 부분으로 연구영역의 확대가 필요할

것이다.

참 고 문 헌

- [1] S. Shin, "Evaluation exchange rate of artificial neural network and moving average method", Finance Research, Vol. 9, no. 1, pp103-135, 1995.
- [2] J. Jeon, "A study on determining prediction models using model-based clustering of time series data", Dankook Univ Ph. D, 2007.
- [3] Y. Cho and G. Lee, "A Study on Improving Prediction Accuracy by Modeling Multiple Similar Time Series", The Institute of Webcasting, Internet and Telecommunication, Vol. 10, no. 6, pp137-143, 2010.
- [4] L. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition," Proc. of IEEE77, pp.257-286, 1989.
- [5] A. Sorjamaa, et al., "Methodology for long-term prediction of time series," Neurocomputing, pp. 178-186. Elsevier, 2007.
- [6] Y. Cho and G. Lee, "Prediction on Clusters by using Information Criterion and Multiple Seeds", The Institute of Webcasting, Internet and Telecommunication, Vol. 10, no. 6, pp145-152, 2010.
- [7] L. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition," Proc. of IEEE77, pp.257-286, 1989.
- [8] M. Siddiqi, J. Gordon and W. Moore, "Fast State Discovery for HMM Model Selection and Learning," In Proc. Int'l Conference on Artificial Intelligence and Statistics, 2007.
- [9] J. Jeon and m. Kim, "A study of criterion for efficient clustering estimation of temporal data", The Institute of Webcasting, Internet and Telecommunication, Vol. 11, no. 5, pp 139-144, 2011.

저자 소개

전 진 호(정회원)



- 1994년 : 관동대학교 경영학과 경영학사
- 1998년 : 명지대학교 경영정보학과 경영학석사
- 2007년 : 단국대학교 컴퓨터과학 이학박사
- 2009년 9월 ~ 현재 : 관동대학교 경영학과 조교수

<주관심분야 : 지능형시스템, 데이터마이닝>

김 민 수(정회원)



- 1997년 : 관동대학교 무역학과 경영학사
- 1999년 : 명지대학교 무역학과 경영학석사
- 2004년 : 명지대학교 무역학과 경영학박사
- 2009년 9월 ~ 현재 : 관동대학교 무역학과 조교수

<주관심분야 : 국제경영>