

시계열 예측 Application S/W 개발에 관한 연구

김치호 홍태화 김학배
연세대학교 전기컴퓨터공학과

Study on the Development of a Time-Series Prediction Application Software

Chiho Kim, Taehwa Hong and Hagbae Kim
Dept. of Electrical & Computer Engineering, Yonsei University

Abstract - 이 논문의 목적은 시계열 예측 엔진의 개발과 그 엔진을 Application S/W로 구현하는 것이다. 시계열 예측 엔진은 과거의 데이터를 분석하여 예측을 위한 식의 차수와 형태를 결정하고 이를 바탕으로 파라미터를 결정한 후 미래의 값을 예측하는 3가지 단계를 거친다. 여기에 쓰이는 기법들은 여러 가지가 있는데 본 논문에서는 ARMA(AutoRegressive Moving Average)를 기본으로 분석하였다. Application S/W는, 개발된 예측 엔진에서 분석될 과거 데이터를 입력받아 예측 엔진 구동에 사용되고 그 결과를 그래프로 나타내는 일련의 과정을 거친다. Application S/W 개발의 많은 Programming Language가 존재하지만 본 논문에서는 Visual C++을 사용하였다. 또한 이 논문에선, 특정 교차로를 통과하는 교통량 변화에 대한 데이터를 이용하여 예측을 수행하고, 그 결과를 Application S/W에 적용시켰다.

핵심어: 시계열 예측 엔진, Application S/W, ARMA

1. 서론

일반적인 경영 환경에서 발생하는 수요는 시간상의 변동에 대응하는 stochastic process로 예측되어질 수 있다. 물론 이를 위한 데이터 수집 방법론 및 파라미터화에 있어서 경제 사회적인 요소들이 중요한 역할을 하기 때문에 단순한 수리적 모델링이 용이하지는 않다. [1,3] 따라서 기존의 단순한 통계기법이나 인력에 의존한 합산 방식보다 유동적이고 통합적인 시스템적 접근방식을 통해 관련 전산시스템의 운영체계에 접목시키는 엔지니어링 측면의 접근방식과 최신 예측이론의 활용화가 선행되어야 한다.

기존의 예측모델은 시간에 대해 stochastic 시계열 함수(Time Series)로 표시되고, 이의 대표적인 형태는 다항함수 또는 saturation 곡선형이다. 그러나 이들은 수요의 극변동에 대한 고려가 용이하지 못하고, 장기적 변동을 감안한 계절적 효과 등이 구현될 수 없다는 제한점이 있다. 장기적 변동, 주기적 진동, 계절적 변화, 또는 불규칙적 변동들을 위해 계절적 조정기법이 고려되고 있으나 이 방법의 복잡한 구조와 수치적 사항의 현실화가 문제시되고 있다. 시간 변동 데이터를 직접적으로 고려하기 위해 AR(AutoRegressive)모델, ARIMA(AutoRegressive Integrated Moving Average)모델등의 stochastic 모델들이 검토되고 있으나 정확도 문제와 함께 경영의 제반 수요예측 적용에 필요한 가정이나 제약의 분석 결여로 보완과정을 거쳐야 한다. [1,2]

이러한 이유로 많은 예측 엔진들이 개발되었으나 어려운 수식과 실무자가 손쉽게 접근할 수 있는 Tool의 부재로 경영환경에 직접 쓰이지 못해온 것이 현실이다.

현재의 Application S/W들은 복잡하고 좋은 엔진보단도 User가 손쉽게 접근할 수 있는 쉬운 내용과 친근한 GUI로서 평가받고 있다.

Application S/W의 개발환경 역시 많은 변화를 겪고 있다. 하지만 많은 개발 language들이 있으나 대부분의 Application S/W들이 C/C++ base이다. 따라서 소스의 공유라는 점에서 많은 Developer들이 이해할 수 있는 C/C++ 환경에서의 개발이 권해지고 있는 현실이다.

따라서 본 논문에선 시계열 예측 엔진 개발과 그 엔진을 경영 실무자들이 친근하게 사용할 수 있는 Application S/W로 개발하는 일련의 과정을 담고있다. 2.1에선 시계열 예측 엔진 개발의 과정과 알고리즘을 소개한다. 2.2에선 개발된 예측 엔진을, Visual C++ 환경에서 구현된 S/W와 연결한 결과를 보여준다.

사회전반에 걸친 정보화 및 자동화 추세와 더불어 효율적 데이터 관리 및 운영이라는 측면에서 본 시계열 예측 S/W 개발에 관한 연구는 여타 분야의 관련 정보 및 경영기법에 큰 파급효과를 미치게 될 것이다.

2. 본론

2.1 시계열(Time Series) 예측 엔진

시계열 분석이란 일정기간동안 관찰할 수 있는 자료를 바탕으로 통계적인 예측을 하기 위한 일련의 작업이다. [4] 일반적으로 과거의 데이터를 분석하여 예측을 위한 식의 차수와 형태를 결정하고 이를 바탕으로 파라미터를 결정한 후 미래의 값을 예측하는 3가지의 단계를 거친다. [5] 여기에 쓰이는 기법들은 여러 가지가 있는데 본 논문에선 ARMA를 기본으로 분석하였다. [6,7,8]

통계적인 예측을 위해서는 (1)현재 가지고 있는 정보를 구체적으로 정리하고 (2)정리된 정보가, 과거의 통계를 이용하여 타당성이 있는가를 검증한 후 (3)통계적으로 검증된 정보를 이용하여 미래를 예측한다.

위에서 언급한 것과 같이 시계열의 분석에는 일반적으로 정형화된 3가지의 단계를 거친다. 각각의 단계를 보면 그림 1과 같다.

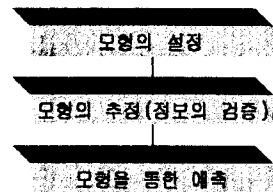


그림 1 시계열 분석의 단계

예측의 첫번째 단계는 예측을 위해 모형을 설정하는 것이다. 수요를 예측하는 데 관련되는 모든 정보를 잘

본 연구는 2000년도 두뇌한국21사업에 의하여 지원되었습니다.

조직화하여 아주 간단한 형태의 수식으로 표현하는 것을 말한다.

두번째 단계는 설정된 모형이 실제로 맞는지 틀린지, 혹은 맞다면 과거의 실적을 기준으로 그 맞는 정도가 어느정도 인지를 검증하는 것이다. 이 단계에서 중요한 점은 적절한 추정 기법(ARMA, ARIMA, SARIMA 등)을 사용하는 일이다. 기술적인 많은 방법들은 나름대로의 장점과 단점을 가지고 있다. 따라서 원하는 적절한 방법을 택하는 것은 매우 중요하다. 이러한 방법이 잘못 선택되거나 잘못 이용될 경우 추정의 결과가 사실과는 다르게 나타날 수 있기 때문이다.

마지막 단계는 추정된 모형으로 미래를 예측하는 일이다. 이 단계에서 중요한 점은 예측하고자 하는 변수에 영향을 주는 여러 여건들이 미래에 어떻게 변화할 것인가를 가정하는 일이다.

본 논문에선 ARMA(3,2)를 기본으로 하였다. 파라미터를 구하는 방법으로는 Newton-Gauss Method를 이용하였다.

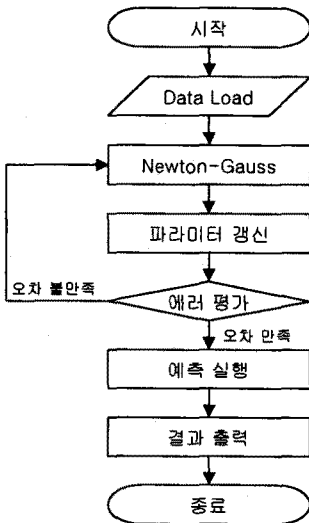


그림 2 시계열 예측 엔진 프로그램 순서도

원 그림 2의 순서도 에서 중요한 부분은 'Newton-Gauss'이다. ARMA(3,2)이므로 MA는 3개의 초기조건을 필요로 한다. 계산 시 초기조건은 '0'으로 가정하고 모든 파라미터들은 임의의 값(이 논문에선 0.1~0.3의 값을 random하게 발생시켰다.)으로 설정했다.

원 데이터의 그래프와, 계산 후 결정된 파라미터와 원 데이터 차수만큼의 계산 후에 나온 확률 오차 값들로 결정된 그래프를 비교해 보면 급격히 원 데이터의 이동에 접근해 간다. 데이터의 개수가 70개 정도인 모형은 대략 3번 정도의 Newton-Gauss 계산을 반복하면 일정한 에러가 나오며 그때 계산되는 파라미터들이 모델의 파라미터로 결정되어 진다.

그림 3은 특정 교차로를 통과하는 교통량 변화를 적용시켜 본 결과이다.

그림 3의 그래프에서 진한 선은 원 데이터를 나타낸다. 그리고 옅은 선은 파라미터에 의해 추정된 값이다. 추정된 데이터가 원 데이터를 잘 따라가고 있음을 알 수 있다. 이것으로 추정된 데이터에 대한 타당성을 검증했다.

그림 4는 5년 동안의 교통량 변화 데이터에 의해 미래의 6개월간 교통량을 예측한 결과이다.

2.2 Application S/W 구현

Application S/W는 Visual C++ 환경에서 구현했다.

크게 과거의 데이터를 입력받는 부분과 예측을 실행시

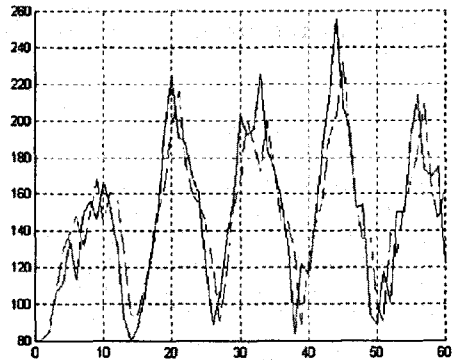


그림 3 교통량 변화

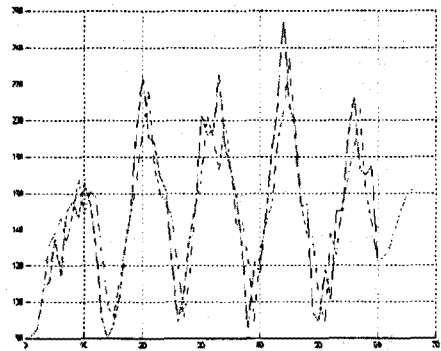


그림 4 교통량 예측

켜 미래의 데이터를 생성시키고, 추정 데이터와 미래의 데이터를 그래프로 나타내는 세 가지 step으로 진행된다.

우선 과거 데이터 입력은 그림 5의 Dialog Box에서 행해진다.

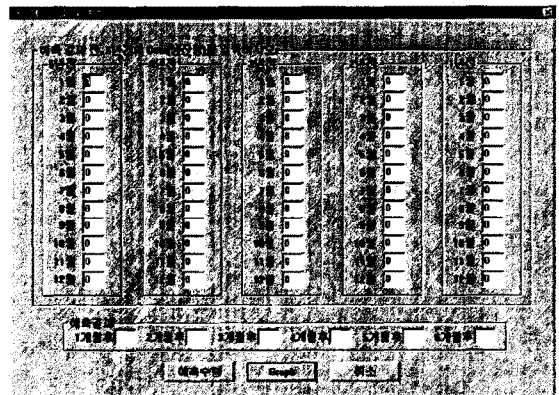


그림 5 '과거 데이터 입력과 예측' Dialog Box

그림 5의 Dialog Box는, 그림 6과 같이 메뉴의 '예측-'>'예측실행'을 선택하면 생성된다.



그림 6
Dialog Box
생성 메뉴

그림 5의 Dialog Box에 과거 데이터를 입력시킨 후 '예측수행' 버튼을 누르면 '예측결과' 그룹박스에 예측결과가 나타난다. 그 결과는 그림 7과 같다.

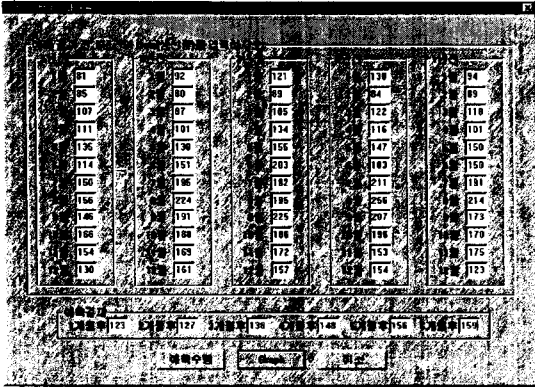


그림 7 예측결과

이 상태에서 'Graph' 버튼을 누르면 추정 데이터와 예측 데이터에 대한 그래프가 클라이언트 윈도우에 나타난다. 그 결과는 그림 8과 같다.

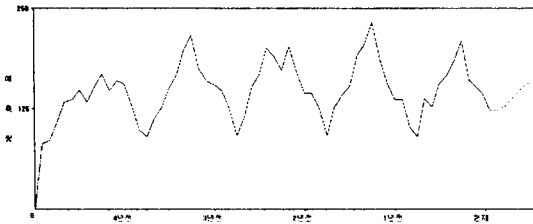


그림 8 추정 데이터와 예측 결과에 대한 그래프

그림 8에서 현재(진한 실선) 이후의 얇은 점선으로 표시된 부분이 예측결과에 해당한다. 이 S/W에서 구현된 그래프는 그림 4의 예측결과 그래프와 같다는 것을 알 수 있다.

3. 결 론

본 논문에서는 일반적인 경영 환경에서 발생하는 수요에 대한 예측을 Application S/W로 구현하여 현장에서 직접 사용할 수 있게 하였다. 우선 예측 엔진은 ARMA(3,2)를 기본으로 구현하였고, 이것은 직전 3개의 데이터로 다음 데이터를 추정한다. 추정한 데이터에 의해 미래의 값 예측이 가능하다. 이 논문은 원 데이터 60개(12개월×5년)를 추정하여 미래의 6개(6개월) 데이터를 예측하였다. 이렇게 생성된 예측엔진은, 수학적 마인드가 부족한 사람들이 적용하기엔 어려웠던 것이 현실이었다. 이에 Application S/W를 Visual C++상에서 구현하였다. 이 Application S/W는 통계학이나

컴퓨터 공학의 지식이 부족한 사람들도 편리하게 쓸 수 있도록 친숙하고 알기 쉬운 GUI로 구성되었다는 특징이 있다. 이 논문은 제안된 아이디어를 실제로 구현했기 때문에 통계적 예측이 필요한 환경에서 적극 활용될 수 있을 것이다.

(참 고 문 헌)

- [1] J. Hansen and R. Nelson, "Neural Networks and Traditional Time Series Methods: A Synergistic Combination in State Economic Forecasts", *IEEE Trans. on Neural Networks*, Vol. 8, No. 4, pp. 863-873, 1997
- [2] F. Lin and X. Yu, Shirley Gregor and Richard Irons, "Time Series Forecasting with Neural Networks", *Complexity International*, 1995
- [3] T. Weidner, "Hubbing in U.S. Air Transportation System: Economic Approach", *Transportation research record*, pp. 28-37
- [4] Nancy K. Groschwitz and George C. Polyzos, "A Time Series Model of Long-Term NSFNET Backbone Traffic", *IEEE*, pp. 1400-1404, 1994
- [5] A. Cavallini and G.C. Montanari, "A Parametric Approach to the Prediction of the Time-Behavior of Harmonic-Quantities in Electrical Networks", *IEEE*, pp. 2226-2232, 1995
- [6] Shiping Li and Bradley W. Dickinson, "An Efficient Method to Compute Consistent Estimates of the AR Parameters of an ARMA Model", *IEEE*, pp. 275-278, 1986
- [7] Gerard Alengrin and Josiane Zerubia, "A Method to Estimate the Parameters of an ARMA Model", *IEEE*, pp. 1113-1115, 1987
- [8] Shiping Li, Yao Zhu and Bradley W. Dickinson, "A Comparison of Two Linear Methods of Estimating the Parameters of ARMA Models", *IEEE*, pp. 915-917, 1989