2000년도 대한전기학회 하계학술대회 논문집 2000.7.17-20

Levinson-Durbin 알고리듬과 Newton-Raphson Method를 이용한 개방형 시계열 데이터 예측엔진 구현에 관한 연구

구 진 모 홍 태 화 김 학 배
연세대학교 전기컴퓨터공학과

Implementation of an Open Prediction Engine for Time-Series Data Using Levinson-Durbin Algorithm and Newton-Raphson Method

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

Abstract - 시계열(time series)이란 한 사상 또는 여러 사상에 대하여 시간의 흐름에 따라 일정한 간격으로 이들을 관측하여 기록한 자료를 말한다. 이러한 시계열은 어떠한 경제현상이나 자연현상에 관한 시간적 변화를 나타내는 역사적 계열(historical series)이므로 어느 한 시점에서 관측된 시계열자료는 그 이전까지의 자료들에 주로 의존하게 된다. 따라서 시계열분석을 통한 예측에서는 과거의 자료들을 분석하여 법칙성을 발견해서 이를 모형화하여 추정하고, 이 추정된 모형을 사용하여 미래에 관측될 값들을 예측하게 된다. 본 연구에서는 ARMA (p,q)모형(autoregressive moving-average model)을 이용하여 시계열 데이터를 분석하며 계수의 추정에는 Levinson-Durbin 알고리듬과 Newton-Raphson Method를 이용한다.

핵심어: 시계열, 자료분석, Levinson-Durbin 알고리즘, Newton-Raphson Method, 계수추정

1. 서 론

ARMA(p,q)모형(AutoRegressive Moving-Average model)이란 관측된 시계열 데이터들이 정상적이고 관찰치가 1시차 이전의 데이터와 연관되어 있고, 과거시점의 확률오차에 관련되어 있을 때 많이 사용되는 모형이다. 정상적이라 함은 일정한 계열수준이나 변동을 중심으로 변하는 형태를 띠는 계열을 말한다. 시계열 분석에서 가장 중요한 것은 분석하고자 하는 계열에 적합한 ARMA모형의 차수를 가장 적절하면서도 최소의 차수를 결정하는 것과 이의 계수들을 얼마만큼 효율적이며 정확하게 구해내는 것이다. 적절한 모형과 계산되어진 계수를 이용하면 미래의 값을 예측할 수 있다는 시계열 분석의 특징 때문에 여러 분야에서 적용되고 있으며 시계열 분석의 효율적 계산방법이 연구중이며 발표되고 있다[1, 2, 3]. 일반적인 방법으로 계수를 구할 때는 p(AR부분의 차수)와 q(MA부분의 차수)중 하나의 차수만 변경되더라도 전부 새로 계산해야 하는 문제가 있다. 본 연구의 목적은 개방적인 시계열 데이터 예측엔진을 개발에 있으므로 기존 방법은 적절하지 않다. 즉, 계산초기에 차수를 고정하여 계산을 수행하는 방식은 다양한 차수에 대한 프로그램 구현에 많은 어려움이 따른다.

본 논문에서는 계산초기에 결정된 차수의 계수만을 얻을 수 있는 계산방식을 배제하고 한번의 계산과정에서 다양한 차수의 계수의 계산이 가능한 개방형 시계열 예측엔진에 적합한 알고리즘을 보인다. 이 알고리즘을 이용하면 한가지의 알고리즘만으로 다양한 모형과 차수에 적용이 가능하므로 프로그램 구현의 편리함과 코드자체의 용량을 줄일 수 있다.

본 연구는 2000년도 두뇌한국21사업에 의하여 지원되었습니다.

2. 본 론

2.1 AR(autoregressive) 부분의 계수 계산

2.1.1 순수한 AR(p)모형의 계수 계산

AR(p)모형은 계열이 정상적이고 계열 X_t의 각 관찰치가 1시점 이전의 과거 계열치와 연관되어 있을 때 그 관계를 나타낸 모형이다. 이 모형을 수식으로 표현하면 (1)식과 같다.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + e_t \quad (1)$$

AR(p)모형의 경우에는 AR(p)계수들은 Yule-Walker방정식을 만족하고 이 방정식은 Levinson-Durvin알고리즘으로 풀 수 있다. 이 Levinson-Durvin알고리즘은 시계열 분석에서 가장 기초적이고 중요한 알고리즘 중의 하나이다[4]. 이 알고리즘은 AR(p)모형의 계수들을 구하는데 유용한 알고리즘 이지만 Choi가 제안한 알고리즘[5]을 이용하면 ARMA(p,q)의 AR 계수들을 구하는데 이용할 수 있다. Levinson-Durvin알고리즘은 다음과 같다.

$$\begin{aligned}
&For\ k = 0,\ \ let \\
&\quad \phi_{1,1} = \rho_1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (2) \\
&\quad \lambda(1) = 1 - \phi^2_{1,1} \\
&\quad\quad For\ k = 1,\ 2,\ \cdots, \\
&\quad\quad\quad \theta(k) = \rho_{k+1} - \phi_{k,1}\rho_k - \cdots - \phi_{k,k}\rho_1 \\
&\quad\quad\quad \phi_{k+1,k+1} = \frac{\theta(k)}{\lambda(k)} \qquad\qquad\qquad\qquad\qquad (3) \\
&\quad\quad\quad \lambda(k+1) = \lambda(k)\{1-\phi^2_{k+1,k+1}\} \\
&\quad\quad\quad\quad For\ \ j = 1,\ 2,\ \cdots k, \\
&\quad\quad\quad\quad\quad \phi_{k+1,j} = \phi_{k,j} - \phi_{k+1,k+1}\phi_{k,k+1-j} \qquad (4)
\end{aligned}$$

여기서 ρ_k는 k시차의 자기상관 함수이다.

위의 식은 MA(q)를 구려하지 않은 AR(p)의 계수이다. 만약 구하려는 모델이 MA를 포함하지 않는다면 위의 식까지만 계산을 수행하면 적절한 값을 얻을 수 있다. 만약 p의 차수가 1이라면 k = 0인 (2)식까지만 계산이 필요하며 더 높은 k = 1, 2, ⋯, 에 따라 (3)식을 계산하면 요구되는 차수까지 계수를 구할 수 있다. (4)식은 k까지 계산을 수행했다면 $k+1$의 계산의 과정에서 필요한 원소들을 계산하는 과정이다. 이 식들을 이용하면 계산된 차수 이상의 계수들을 손쉽게 구할 수 있다.

2.1.2 ARMA(p,q)모형의 AR(p)

Levinson-Durvin알고리즘을 이용하여 순수한 AR(p)까지의 계수를 구한뒤에 다음의 Choi알고리즘을

이용하면 ARMA(p,q)모형의 차수에 맞는 계수들을 MA의 차수인 q만큼의 반복계산으로 비교적 쉽게 구할 수 있다. Choi알고리즘은 다음과 같다.

우선 Levinson-Durvin알고리즘을 사용하여 계산된 값을 $i=0$라고 표현하면 (5)식과 같이 나타낼 수 있다.

$$\{\phi_{k,j}^{(0)} \mid k = 1, 2, \cdots, \ j = 1, \cdots, k\} \qquad (5)$$

$For\ i = 1, 2, \cdots, q$

$\quad For\ j = 1, 2, \cdots,$

$$\phi_{k,0}^{(i)} = -1 \qquad (6)$$

$\qquad For\ j = 1, \cdots, k,$

$$\phi_{k,j}^{(i)} = \phi_{k+1,j}^{(i-1)} - \frac{\phi_{k+1,k+1}^{(i-1)}}{\phi_{k,k}^{(i-1)}}\phi_{k,j-1}^{(i-1)} \qquad (7)$$

위의 식을 살펴보면 q만큼의 반복계산이 필요하다는 것을 알 수 있다. (7)식을 살펴보면 $\phi_{k,j}^{(i)}$ 의 계산에는 $\phi_{k+1,k+1}^{(i-1)}$의 값이 필요하므로 요구되는 q만큼의 반복계산을 위해서는 Levinson-Durvin알고리즘에서 AR의 차수인 p의 횟수보다 좀더 많은 계산이 필요함을 알 수 있다. p의 횟수만큼의 계산과정에는 $q, q-1, \cdots, 1$에 해당되는 MA차수에 포함되는 ARMA(p,q)의 AR(p)계수를 같이 얻을 수 있으므로 모형식의 변경시 간단히 계수들을 얻을 수 있다. 계산된 결과에서 원하는 계수는 다음과 같이 결정할 수 있다.

$$\phi_{p,1}^{(q)} = \phi_1, \ \phi_{p,2}^{(q)} = \phi_2, \cdots, \ \phi_{p,p}^{(q)} = \phi_p \qquad (8)$$

2.2 MA(moving average)부분의 계수 계산

MA(q)의 계수들을 구하는 알고리즘에는 Wilson(1969)이 제시한 알고리즘이 있다. Choi(198 7)는 Wilson의 알고리즘을 확장하여 ARMA모형에서 MA계수들을 구하는 알고리즘을 제시했다[5]. Choi알고리즘은 근본적으로 다음의 방정식을 Newton-Raphson방법으로 푸는 것이다.

$$\begin{aligned} &\phi_0\sigma(j) + \phi_1\sigma(j-1) + \cdots + \phi_p\sigma(j-p) \\ &= (\psi_0\theta_j + \psi_1\theta_{j-1} + \cdots + \psi_{q-j}\theta_q)\sigma_\varepsilon^2 \\ &\qquad (j = 0, 1, \cdots, q) \end{aligned} \qquad (9)$$

이 알고리즘을 설명하기 위해서는 몇 가지 기호를 정의해야 한다. $\Sigma(q+1, p+1;0)$과 Ψ는 $(q+1)$ X(p+1) 행렬들로서 각 행렬의 (i,j)원소는 다음과 같다.

$$(\Sigma(q+1, p+1;0))_{i,j} = \sigma(i-j) \qquad (10)$$

$$(\Psi)_{i,j} = \begin{cases} 0 & (i<j) \\ \psi_{i-j} & (i \ge j) \end{cases} \qquad (11)$$

또한, Ψ_q는 $(q+1)$ X(q+1)행렬이고 이 행렬의 (i,j)원소는 다음과 같다.

$$(\Psi_q)_{i,j} = \begin{cases} 0 & (i>j) \\ \psi_{i-j} & (i \le j) \end{cases} \qquad (12)$$

이 기호들을 사용하면 (9)식을 다음과 같이 쓸 수 있다.

$$\Sigma(q+1, p+1;0)\phi_* = \sigma^2\Psi_q\Psi\phi_* \qquad (13)$$

Newton-Raphson방법을 이용하여 계수를 구하기 위해서는 기호의 정의가 더 필요하다.

$$c_0 = \sigma, \ c_1 = \sigma\psi_1, \ c_2 = \sigma\psi_2, \cdots, c_q = \sigma\psi_q \qquad (14)$$

$$C_q = \sigma\Psi_q \qquad (15)$$

$$C = \sigma\Psi \qquad (16)$$

$$c = (c_0, c_1, \cdots, c_q)' \qquad (17)$$

$$\phi_* = (\phi_0, \phi_1, \cdots, \phi_p)' \qquad (18)$$

$$(T_j)_{r,s} = \begin{cases} c_{j+r-s} & (1 \le s \le j+r, \ 1 \le r \le q-j+1) \\ 0 & otherwise, \end{cases} \qquad (19)$$

$$(s_j)_{r,s} = \begin{cases} c_{j-r+s} & (1 \le r \le q+1, \ \max\{1, r-j\} \le s \le q-j+1) \\ 0 & otherwise, \end{cases} \qquad (20)$$

$$U_j = T_j + S_j \qquad (j = 0, 1, \cdots, q) \qquad (21)$$

$$W = (U_0\phi_*, U_1\phi_*, \cdots, U_q\phi_*) \qquad (22)$$

위의 정의를 사용하면 (13)식을 다음과 같이 표현할 수 있다.

$$C_qC\phi_* = \Sigma(q+1, p+1;0)\phi_* \qquad (23)$$

본 논문의 목적은 위의 식을 Newton-Raphson방법을 이용하여 $c_0, c_1, \cdots, c_q$를 계산하는 것이다. 계산된 값들을 가지고 $\sigma, \psi_1, \psi_2, \cdots, \psi_q$를 구한 다음 2.1.2절에서 계산된 $\phi_1, \phi_2, \cdots, \phi_p$를 사용해서 $\theta_1, \theta_2, \cdots, \theta_q$를 계산한다. Newton-Raphson의 해답은 다음의 반복식에 의해서 결정된다.

$$c^{(n+1)} = \frac{1}{2}c^{(n)} + (W^{(n)})^{-1}\Sigma(q+1, p+1;0)\phi_* \qquad (24)$$

(23)식의 초기값으로 다음 값들을 사용한다.

$$\begin{aligned} c_0^{(0)} &= \{\sigma(0)\}^{1/2} \\ c_j^{(0)} &= \phi_1 c_{j-1}^{(0)} + \cdots + \phi_j c_0^{(0)} \\ &\qquad (j = 1, \cdots, q) \end{aligned} \qquad (25)$$

위의 알고리즘을 사용하면 대략 10번의 반복횟수로 일정한 값에 수렴하게 된다. 수렴된 값은 (26)식으로 MA(q)의 계수들을 계산하게 된다.

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 &= -\theta_1 + \phi_1 \\ \psi_2 &= -\theta_2 + \phi_2 + \phi_1\psi_1 \\ &\vdots \\ \psi_j &= -\theta_j + \phi_p\psi_{j-p} + \cdots + \phi_1\psi_{j-1} \\ &\text{단, } \psi_j = 0 \quad (j = -1, -2, \cdots) \\ &\qquad \theta_j = 0 \quad (j = q+1, q+2, \cdots) \end{aligned} \qquad (26)$$

2.3 수치 예제

앞의 알고리즘을 이용하여 Matlab으로 구현해 보았다. 앞에 설명된 알고리즘으로 계산을 하기 위해서는 ACF (autocorrelation function)과 PACF(partial autocorrelation function)를 먼저 계산해야 한다. 사용된 시계열 데이터는 그림 1과 같다. 그림 1을 보면 평균이 일정한 안정된 시계열 데이터임을 알 수 있다. ACF와 PACF의 분석결과 ARMA(3,2)의 모형임을 알 수 있었다.

그림 2는 앞에서 설명된 알고리즘을 가지고 계산된 계수를 이용하여 모형의 적합성을 검증한 그래프이다. 관찰의 편리를 위해 시차 30에서 106까지의 그래프를 보였다. 그림 2.에서 실선은 샘플데이터의 원본 그래프이며 점선은 계산된 계수를 이용한 계산치의 그래프이다. 계산치가 샘플데이터의 특성과 수치를 잘 나타내고 있으므로 만족할 만한 결과를 보인다고 할 수 있다. 101시

차부터 106시차까지의 값은 계산된 계수의 값을 바탕으로 미래의 데이터를 예측한 값이다. 예측값이 전체적인 데이터의 흐름에 일치되어 있다.

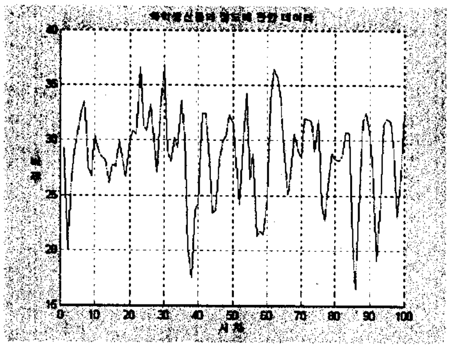

그림 1. 화학생산품 점도에 관한 데이터 그래프

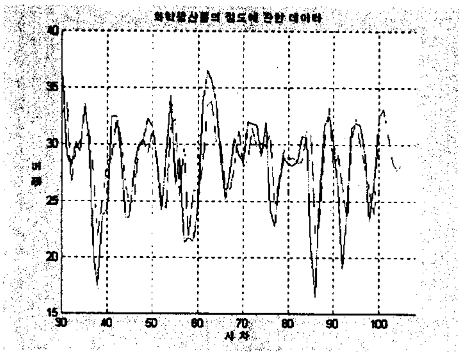

그림 2. 계산된 계수를 이용한 그래프와 예측값

3. 결 론

본 연구는 시계열 데이터 예측에서 가장먼저 선행되어야 할 계수의 분석알고리즘에 대하여 진행되었다. 본 연구에서 보인 알고리즘을 사용하면 ARMA(p,q)모형의 차수가 변경되더라도 예측엔진의 알고리즘을 변경시키지 않고 간단히 계수의 계산이 가능하다. 이 알고리즘은 적절한 차분이 이루어진 ARIMA(p,d,q)모형에도 적용이 가능하다. ARMA는 단기적인 예측에 적합한 모형이다. 따라서 장기적인 예측에 관련된 알고리즘의 연구가 더욱 진행되어야 할 것이다.

[참 고 문 헌]

[1] Shiping Li, Yao Zhu, and Bradley W. Dickinson, "A Comparison of Two Linear Methods of Estimating the Parameters of ARMA model", IEEE Transactions on Automatic Control, Vol. 34, No. 8, pp. 915-917, August 1989

[2] Gerard Alengrin and Josiane Zerubia, "A Method to Estimate the Parameters of an ARMA Model", IEEE Transactions on Automatic Control, Vol. AC-32, No. 12, pp. 1113-1115, December 1987.

[3] Shiping Li and Bradley W. Dickinson, "An Efficient Method to Compute Consistent Estimates of the AR Parameters of an ARMA Model", IEEE Transactions on Automatic Control, Vol. AC-31, No. 3, pp. 275-278, March 1986

[4] Brockwell, P. J. & R. A. Davis, "Time Series : Theory and Methods(Second Edition)", Springer-Verlag, pp241-245, 1990

[5] Choi, B. S.,"ARMA Model Identification", Springer- Verlag, chap. 1, 1992