

IoT 환경에서 예측 정확도 향상을 위한 계절성 비선형 시계열 알고리즘 설계

강정구*, 박석천**, 김종현***

*가천대학교 일반대학원 모바일소프트웨어학과

**가천대학교 컴퓨터공학과 정교수(교신저자)

***위세아이텍 대표이사

e-mail:webclub09@gmail.com

Design of Seasonal Nonlinear Time Series Algorithm for Improving Forecast Accuracy in IoT Environment

Jung-Ku Kang*, Seok-Cheon Park**, Jong-Hyun Kim***

*Dept. of Mobile Software, Gachon University

**Dept. of Computer Engineering, Gachon University(Corresponding Author)

***Representative Director, WISEITECH co., ltd

요 약

ICT 시대를 맞아 하루가 다르게 새로운 기술이 등장하고 있으며, 최근에는 사물인터넷 시대까지 도래하였다. 하지만 현재 사물인터넷에서 폭발적으로 발생하는 시계열 데이터를 분석하는 연구는 미비한 상태이다. 따라서 본 논문에서는 사물인터넷에서 발생하는 시계열 데이터의 예측 정확도 향상을 위해 사계절이 뚜렷한 우리나라의 계절성 특성을 고려한 SARIMA알고리즘과 비선형 특성 예측 알고리즘인 SVM을 결합한 하이브리드 SARIMA-SVM알고리즘을 제안 한다.

1. 서론

ICT(Information and Communications Techlogies)시대를 맞아 하루가 다르게 새로운 기술이 등장하며 최근에는 사물인터넷(IoT:Internet of Things) 시대까지 도래하였다.

사물인터넷이란 고유하게 식별 가능한 사물이 만들어낸 정보를 인터넷을 통해 공유하는 환경을 의미한다. 사물인터넷의 3대 핵심 기술로 사물에서 정보를 얻을 수 있는 센싱 기술, 확보한 데이터를 네트워크에 연결하여 전공하는 유무선 통신 및 네트워크 인프라 기술, 이러한 정보를 바탕으로 특정 기능을 수행하는 사물인터넷 서비스 기술을 들고 있다[1]. 사물인터넷 핵심 기술인 사물인터넷 서비스 기술 중 사물인터넷에서 발생된 센싱 데이터의 분석에 많은 관심이 모아지고 있다. 사물인터넷에서는 대규모의 시계열 데이터가 폭발적으로 발생하고 있으며, 이렇게 발생된 시계열 데이터의 분석에 대한 연구는 현재 미비한 상태이다.

또한 우리나라는 사계절이 뚜렷해 시계열 데이터 분석시 계절성 특성을 고려해야 하며, 계절성 시계열 데이터의 경우 선형과 비선형의 특성이 섞여 있는 형태 이므로 하나의 예측 기법으로 설명하거나 식별하는데 어려움을 가지고 있다.

이에 본 논문에서는 사물인터넷 환경에서 발생하는 시계열 데이터의 예측 정확도 향상을 위한 계절성 시계열 분석 기법인 SARIMA(Seasonal Autoregressive Integrated

Moving Average)알고리즘과, 비선형 시계열 알고리즘인 SVM(Support Vector Machine) 알고리즘을 결합한 하이브리드 SARIMA-SVM 제안 하고자 한다.

2. 관련연구

2.1 사물인터넷

최근 인터넷 기술은 언제, 어디서나, 모든 정보를 검색하고, 활용하여 새로운 정보로 융·복합하는 서비스들이 요구 되고 있다. 이러한 환경을 바탕으로 사람과 사물뿐 아니라 사물과 사물(machine to machine), 현실 세계와 가상 세계를 연결하는 등의 다양한 서비스들에 대한 연구가 진행 중이다[2].

사물인터넷(IoT)란 사물들이 네트워크로 연결되어 서로 데이터를 주고받으며 소통이 가능한 지능적인 서비스 인프라이다. 모바일, 클라우드 컴퓨팅, 빅 데이터 분석 기술과의 융·복합을 통해 초 연결사회의 핵심으로 부상 중이다.

사물인터넷의 주요 구성요소인 사물은 네트워크의 사용자(end device)뿐만 아니라 사람, 차량, IT기기, 자연, 물류 등 환경을 구성하는 모든 물리적 사물이 포함된다. 또한 현실 세계와 가상 세계의 모든 정보와 상호 작용하는 개념 역시 포함된다[3].

2.2 시계열 모형

두 변수 간에 성립되는 인과관계 또는 함수관계를 규명

하는데 주된 목적이 있는 회귀분석 방안은 어떤 경제현상을 특수 함수관계로 파악하고 또 성공적으로 그 관계를 추정할 경우 그 추정 결과에 근거하여 장래에 대한 예측을 시행한다. 반면, 시계열 분석방법론은 인과관계에 관한 어떠한 이론적 배경이 없이도 예측을 수행할 수 있는 방법론으로서, 한 변수의 미래 값을 단지 자신의 과거 관측치 값에 근거하여 파악할 수 있는 방법이며, 미래 예측을 주 관심사로 볼 때 오히려 회귀분석방법보다 쉽게 활용할 수 있다.

시계열 분석은 기본적으로 한 변수의 변동 내용을 장기적 추세변동(Secular Trend)과 주기적 변동(Cyclical Variation) 및 계절적 변동(Seasonal Variation) 그리고 불규칙 변동 등으로 구성되어 있는 것으로 보고 이를 분해하여 추정하는 방법을 사용한다.

대표적인 시계열 분석방법론에는 자기회귀(AR : Auto Regressive)모형과 이동평균(MA : Moving Average)모형 그리고 자기회귀 이동평균(ARMA : Auto Regressive Moving Average)모형과 ARIMA(Auto Regressive Integrated Moving Average)모형 그리고 SARIMA(Seasonal Auto Regressive Integrated Moving Average)모형 등이 있다[4].

2.3 SARIMA모형

일반적인 적분된 자기회귀 이동평균(Autoregressive Integrated Moving Average, ARIMA) 모형은 (식 1)과 같이 정의된다.

$$\phi_p(L)(1-L)^d = \theta_q(L)u_t \quad (1)$$

(식 1)은 ARIMA(p, d, q)로 정의되며 p는 자기 회귀(Autoregressive, AR)항의 차수, q는 이동평균(Moving Average, MA)항의 차수, d는 단위근의 개수를 의미한다. ARIMA모형은 일반적으로 불안정적인 시계열을 대상으로 하는 시계열 모형으로 단위근을 감안하는 모형이다. $\phi_p(L)$ 과 $\theta_q(L)$ 은 자기회귀항(AR term) 및 이동평균항(MA term)의 래그 다항식을 의미한다. 대표적인 사례로서 ARIMA(1, 1, 1) 모형은 p, d, q의 차수가 모두 1인 경우이다. ARIMA(1, 1, 1)의 모형은 (식 2)와 같다.

$$(1 - \phi_1 L)(1 - L)y_t = (1 + \theta_1 L)u_t \quad (2)$$

ARIMA 모형은 특정 연도의 연속적인 월별 자료 간의 특성을 고려하는 모형이며 SARIMA 모형은 이외에도 연속 연도의 동일 시점(월) 자료의 특성을 모두 고려하는 모형이다. SARIMA 모형은 크게 두 부분으로 정의되며 첫 번째 부분은 ARIMA 부분으로 (식 1)과 같다. 예를 들어, 첫 번째 부분은 2013년 12월 오존농도는 2013년 11월 자료와 관계가 있음을 의미한다. 계절성을 고려한 부분은 예를 들면 2013년 5월 오존 농도는 2013년 5월 오존농도와

관련 있음을 의미하며 계절성이 부여된 ARIMA 모형은 (식 3)과 같이 정의된다.

$$\Phi_P(L)(1-L^s)^D = \Theta_Q(L)u_t \quad (3)$$

이때 s는 만약 분기별 자료라면 4가 될 것이고, 월별 자료라면 12가 될 것이다. 또한 D는 계절 단위근의 차수, P는 계절 자기회귀항의 차수, Q는 계절 이동평균항의 차수이다. 또한 $\Phi_P(L)$ 과 $\Theta_Q(L)$ 은 각각 계절 자기회귀항(SAR term)과 계절 이동평균항(SMA term)의 다항식을 의미한다.

최종적으로 첫 번째 부분과 두 번째 부분을 모두 함께 고려하면 (식 4)와 같이 정의된다.

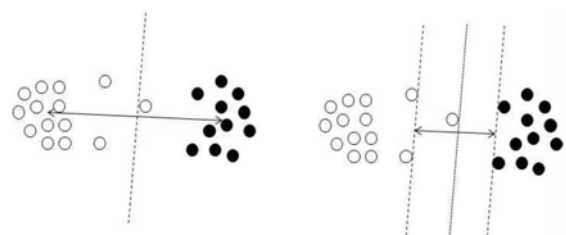
$$\phi_p(L)\Phi_P(L)(1-L)^d(1-L^s)^D y_t = \theta_q(L)\Theta_Q(L)u_t \quad (4)$$

(식 4)는 SARIMA 모형으로 차수는 (p, d, q)×(P, D, Q)로 정의된다[5].

2.4 SVM (Support Vector Machine)

패턴인식은 여러 그룹 혹은 두 그룹을 나누는 방법으로써 식별기준에 따라 각 그룹을 식별하는 인지과학 및 인공지능 분야에 속하는 방법론으로써, 최근인지과학에서 많이 사용되고 있다.

Support Vector Machine은 다양한 패턴인식 방법 중 우수한 해결능력을 보여주는 판별법으로써 두 그룹을 식별하는 식별방법이다.



(a) 일반적인 판별분석방법 (b) SVM method

(그림 1) 일반적인 판별 분석방법과 SVM방법

SVM 방법은 최근 일반적인 판별분석방법과 비교되고 있는데, 이러한 비교 분석을 (그림 1)에 나타내었다. 일반적인 판별분석방법으로 흰색 그룹과 검은색 그룹의 데이터가 주어졌을 경우, 두 그룹의 중심을 구한 후 그 가운데의 최적의 초평면(Hyperplane)을 구함으로써 흰색과 검은색 그룹을 나누는 방법이다. 하지만 오른쪽의 SVM 방법에서는 각 그룹의 중심이 아닌 두 그룹 사이의 경계에 있는 데이터에 초점을 맞추어 선을 그은 다음 그 가운데에 새로운 선을 그어 최적의 초평면을 정하게 된다. 이렇게 함으로써 새롭게 등장한 흰색 데이터는 왼쪽의 일반적인

방법에 의해서는 검은색으로 잘못 분류되지만 SVM 방법에서는 흰색으로 올바르게 예측된다는 것을 알 수 있다.

이렇게 SVM방법은 최적의 초평면을 각 데이터의 중심을 통해서 구하는 것이 아닌 경계를 통해 찾음으로써 흰색그룹과 검은색그룹을 구분하는 오차가 최소가 된다. SVM방법을 이용하여 경계를 나눌 시에 나누는 경계면을 (식 5)와 같이 나타낼 수 있다.

$$F = W^T \underline{X} + \underline{b}, W \in R^n \quad (5)$$

여기서 X는 각 축을 의미하고, W는 X의 방향을 나타내는 벡터이고, b는 축의 절편을 의미하고, R은 실수집합을 의미하게 된다.

(식 5)는 두 그룹의 마진이 최대가 되는 평면을 나타내게 된다. 최적의 초평면에서 각 그룹간의 거리는 $1/\|W\|$ 로 나타낼 수 있는데, 두 그룹의 에러율을 최소로 하기 위해서는 (마진을 최대화하기 위해서는) $2/\|W\|$ 를 최대화 시키는 것과 같다. 위의 설명을 (식 6)과 같은 목적함수로 표현 할 수 있다.

$$\min \frac{1}{2} W^T W \quad (6)$$

$$\text{제약조건, } y_i(W^T \underline{X} + \underline{b}) \geq 1, i = 1, 2, \dots, n$$

(식 6)을 라그랑지 함수로 표현하면 (식 7)을 유도할 수 있으며, (식 7)을 풀게 되면 최적 분류 초평면을 구할 수 있게 된다.

$$L(W, b, \alpha) = \frac{1}{2} \|W\|^2 - \sum_{i=1}^n \alpha_i [y_i(W^T \underline{X} + \underline{b}) - 1] \quad (7)$$

(식 7)을 최대화하기 위하여 (식 8)을 유도할 수 있으며, (식 8)을 풀게 되면 최적 분류 초평면을 구할 수 있게 된다.

$$\max L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (8)$$

$$\text{제약조건 } \sum_{i=1}^n \alpha_i y_i = 0, \alpha_i \geq 0$$

또한 SVM에서는 kernel 함수를 사용하여 저차원에서 발생한 문제를 고차원의 공간으로 사상(mapping)을 실시하게 되는데, 이러한 사상은 비선형문제를 효율적으로 해석하는데 많은 도움이 된다. 사상은 데이터들이 실제로 존재하는 입력 공간(input space, 저차원의 공간)에서는

잘 나누어지지 않는 문제를 feature space라는 고차원의 공간으로 이동시켜서 판별문제를 풀게 된다[6].

3. 계절성 비선형 시계열 알고리즘 설계

3.1 계절성 비선형 시계열 알고리즘 개요

본 논문에서는 계절성과 비선형을 고려한 시계열 알고리즘을 제안하고자 한다. 본 논문에서 제안하고자 하는 계절성, 비선형 시계열 알고리즘은 2단계로 설계 하였다.

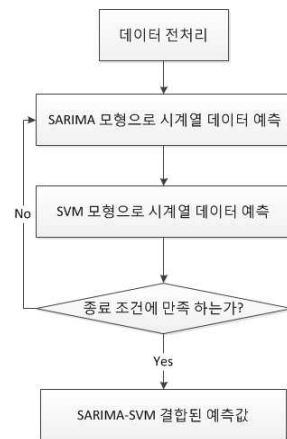
첫 번째 단계에서는 계절성 시계열 분석 알고리즘인 SARIMA모형을 통해 시계열 데이터의 계절성 특성을 분석하는데 사용되며, 계절성 특성에 대한 예측이 진행된다. 하지만 SARIMA모형은 데이터의 비선형적 특성을 예측하는데 미흡한 부분이 있어 비선형적 특성을 분석 하는 기법이 필요 하다.

두 번째 단계에서는 SARIMA모형의 단점인 시계열 데이터의 비선형적 특성을 분석하기 위해 비선형적 특성 분석, 예측 알고리즘인 SVM모형을 사용하였다.

따라서 본 논문에서는 계절성 특성을 고려한 SARIMA모형과 비선형 특성 예측 알고리즘인 SVM모형을 결합한 하이브리드 SARIMA-SVM알고리즘을 제안 한다.

3.2 하이브리드 SARIMA-SVM 알고리즘 설계

본 논문에서 제안하는 하이브리드 SARIMA-SVM알고리즘은 (그림 2)와 같다. 먼저, 사물인터넷에서 발생된 시계열 데이터를 활용하여 데이터 전처리를 진행한다. 다음으로 SARIMA 모형으로 계절성 특성에 대한 예측 분석이 진행된다. 분석이 진행된 데이터를 통해 SVM모형을 활용하여 비선형 특성의 분석이 진행되며, 결과 값에 만족할 경우 계절성 비선형 시계열 데이터의 예측이 종료 된다.



(그림 2) 하이브리드 SARIMA-SVM 알고리즘

4. 결론

ICT 시대를 맞아 하루가 다르게 새로운 기술이 등장하고 있으며, 그 중 사물인터넷은 최근 가장 큰 화두가 되고 있다. 사물인터넷 핵심 기술인 사물인터넷 서비스 기술

분야에서 센싱 데이터의 분석에 많은 관심이 모아지고 있다. 사물인터넷에서 발생하는 시계열 데이터를 활용하기 위한 방법론이 제기 되고 있다.

따라서 본 논문에서는 사물인터넷 환경에서 발생하는 시계열 데이터의 예측 정확도 향상을 위한 계절성 시계열 분석 기법인 SARIMA(Seasonal Autoregressive Integrated Moving Average)알고리즘과, 비선형 시계열 알고리즘인 SVM(Support Vector Machine) 알고리즘을 결합한 하이브리드 SARIMA-SVM 알고리즘을 제안 하였다.

향후 연구 과제로는 본 논문에서 제안한 하이브리드 SARIMA-SVM알고리즘을 통해 사물인터넷 환경에서 발생하는 시계열 데이터를 활용하여 하이브리드 SARIMA-SVM 모형을 구현할 예정이다.

사사의 글

본 연구는 미래창조과학부의 2015년 고용계약형 SW석사과정 지원 사업(과제번호:H0116-15-1003)으로부터 지원 받아 수행한 결과입니다.

참고문헌

- [1] 권준호, 최원익, “IoT와 소프트웨어 특집호를 내면서”, 2014
- [2] 표철식, 강호용, 김내수, 방효찬, “IoT(M2M) 기술동향 및 발전전망”, 한국통신학회지 제 30권 제8호, 2013.07, pp. 3-10
- [3] 문지현, “IoT환경 스마트 홈에서 대기전력 저감 방안에 대한 연구”, 2014
- [4] 김윤식, “SARIMA모형과 VAR모형을 이용한 철도 여객의 수송수요예측”, 2014
- [5] 이재민, 권용재, “계절성을 감안한 ARIMA모형을 이용한 교통수요 동태적 변화 연구”, 2011
- [6] 신준현, 이윤성, 김성열, 김진오, “SVM방법을 이용한 풍력발전기 고장 예측 및 발전수익 평가”, 2014