

공공 데이터를 이용한 에너지 소비 예측 모델링에 관한 연구

박구락*, 정진영**, 안우영**, 정영석^o

*공주대학교 컴퓨터 공학과

**대전보건대학교 바이오 정보과

^o공주대학교 컴퓨터 공학과

e-mail: ecgrpark@kongju.ac.kr*, {jyjung,wyahn}@hit.ac.kr**, merope@kongju.ac.kr^o

A study on energy consumption predictive modeling using public data

Koo-Rack Park*, Jin-Young Jung**, Woo-Young Ahn**, Young-Suk Chung^o

*Dept. of Computer Science & Engineering, Kongju national University

**Dept. of Bio Information, Daejeon Health Sciences College

^oDept. of Computer Science & Engineering, Kongju national University

● 요약 ●

인터넷과 웹의 발전으로 수많은 정보가 발생하고 있으며, 공공공간도 많은 정보를 축적하고 있다. 이에 각 국에서는 공공공간이 보유하고 있는 데이터를 공개하고 있으며 우리나라도 통계청을 중심으로 다양한 데이터를 공개하고 있다. 그러나 공개된 자료의 활용도가 낮은 편이다. 본 논문에서는 공개된 공공데이터 중 에너지 소비 데이터를 활용하고자 한다. 에너지 소비 데이터를 미래 예측 연구에 많이 이용되고 있는 마코프 프로세스를 적용하여, 에너지 소비를 예측할 수 있는 모델링을 제안하고, 그 기대 효과에 대해 논의 한다.

키워드: 공공데이터(Public Data), 마코프 프로세스(Markov Process), 시뮬레이션(Simulation), 예측 모델링(Predictive modeling)

I. 서론

인터넷과 웹의 발전으로 현재 수많은 정보가 축적되고 있다. 정보의 총량은 증가하고 있고 2011년 현재 정보의 총량은 1ZB(제타바이트:10²¹바이트)를 넘어서는 것으로 추정된다. 그에 따라 공공공간에서 보유하고 있는 데이터도 늘어나고 있다. 공공 데이터는 정부 또는 공공공간에 보유하고 있는 데이터를 의미한다. 공공공간의 업무 진행 과정에서 얻어진 것으로 상대적으로 가치가 높은 데이터들이다. 미국, 영국 등 선진국에서는 공공데이터를 이용하여 정부의 대민 서비스에 공공데이터를 활용하고 있다[1,2]. 그 예로 구글의 경우 공공 데이터를 활용하여 데이터를 분석하여 보여주는 공공데이터 익스플로러를 운영하고 있고, 이것을 활용하여 독감과 관련된 단어의 등장하는 빈도수를 분석하여 독감의 확산을 예측하였다[3]. 그러나 우리나라의 경우 공공데이터를 활용하는 예는 많지 않다. 본 논문에서는 공공의 데이터 중 에너지 소비 데이터를 활용하여 에너지 소비를 예측하는 모델링을 제안한다. 에너지 소비를 예측하는 방법으로는 마코프 프로세스를 적용하였다. 마코프 프로세스는 미래를 예측하는 다양한 분야에 적용되고 있다. 본 논문의 구성은 다음과 같다. 2장 관련연구에서 공공의 데이터와 마코프 프로세스에 대해 논의하고 3장에서 공공데이터에 마

코프 프로세스를 적용한 미래 예측 모델링에 대해 제안한다. 마지막으로 결론 및 향후 연구과제에 대해 논의 한다.

II. 관련 연구

1. 공공데이터(Public Data)

공공데이터는 정부 내 산하 기관에서 보유한 데이터를 숫자로 된 정형화된 원시데이터로 볼 수 있다. 우리나라의 경우 통계청에서 국가통계포털을 통해 인구, 물가, 에너지 수급 등 다양한 통계자료 공개 하고 있으며, 서울시의 경우 “서울 열린 데이터 광장”을 통해 물가정보, 부동산 가격정보, 대기 환경 및 수질 정보, 수해 예방 상황 정보 등 다양한 정보를 공개 하고 있으며 이것을 스마트폰 등에 적용할 수 있도록 공개하고 있다[1].

2. 마코프 프로세스(Markov Process)

마코프 프로세스는 과거의 동적 특성을 분석하여 미래에 있을 변화를 예측하는 수학적인 기법을 말한다[4]. 마코프 프로세스는 상태간의 전이가 오로지 이전 n개의 상태에 의존하는 프로세스를 의미한다. 이때 이 모델을 n 차원 모델이라고 하고 n은 다음 상태

를 결정하는데 영향을 주는 상태들의 수로 정의된다. 마코프 프로세스를 $X(t)$ 라 하면 임의의 시간 $t_1 < t_2 < \dots < t_k < t_{k+1}$ 에 대해 $X(t)$ 가 이산 값이면

$$P[a < X(t_{k+1}) = x_{k+1} | X(t_k) = x_k, \dots, X(t_1) = x_1] \quad (1)$$

$$= P[X < X(t_{k+1}) = x_{k+1} | X(t_k) = x_k]$$

식(1)로 기술할 수 있고, $X(t)$ 가 연속 값이면

$$P[a < X(t_{k+1}) = b | X(t_k) = x_k, \dots, X(t_1) = x_1] \quad (2)$$

$$= P[a < X(t_{k+1}) = b | X(t_k) = x_k]$$

로 기술된다. 위의 식(1), 식(2)에서 t_k 는 현재, t_{k+1} 은 미래, t_1, \dots, t_{k-1} 은 과거의 시점이다[5]. 마코프 프로세스는 가능한 상태들의 집합인 상태 집합, 모델링의 초기화 확률인 초기 확률, 각 상태간의 전이 확률로 구성된 전이 행렬로 구성되어 있다[5,6].

III. 공공 데이터를 활용한 에너지 소비 예측 모델링

인터넷과 웹의 발전으로 다양한 데이터들이 늘어나고 있다. 공공기관도 예외는 아니어서 가지고 있던 다양한 분야의 데이터들을 공개하고 있다. 통계청은 월별로 각 에너지원(석탄, LNG, 수력, 원자력 등)의 소비치와 그 합계를 공개하고 있다. 본 논문에서는 공개된 데이터를 활용하여 다음 달에 발생할 에너지 소비를 예측하는 모델링을 제안한다. 제안된 모델링은 그림 1과 같다.

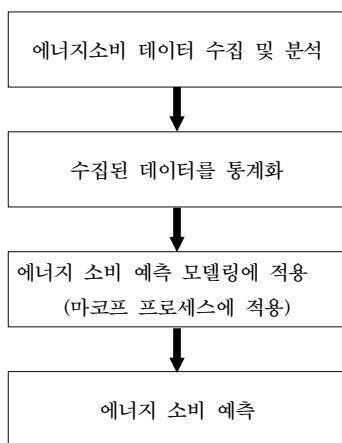


그림 1. 에너지 소비 예측 모델링

Fig. 1. Predictive modeling of energy consumption

각 모델링의 단계는 다음과 같다.

첫 번째 단계로 통계청에서 공개하는 월별 에너지 소비에 대한 데이터 중 예측을 원하는 에너지의 데이터를 수집한 후 그 데이터를 입력한다.

두 번째 단계로 예측을 원하는 에너지 데이터 중 최근 3년간의 자료를 바탕으로 임계값을 설정하고, 임계값을 바탕으로 상태를 정의한다. 그리고 정의된 상태를 예측을 원하는 에너지의 월별 에너지 데이터와 매칭 한다.

세 번째 단계로 최근에 발생한 예측을 원하는 에너지의 소비를 두 번째 단계에서 정의한 상태를 바탕으로 초기 확률을 산출하고, 전체 상태를 대상으로 각 상태간의 전이 확률 계산한다. 계산된 초기 확률과 전이확률을 마코프 프로세스에 적용하여 다음 달에 소비될 것으로 예측된 에너지 소비 확률을 생성 한다.

마지막 단계로 예측된 에너지 소비 확률과 최근 3년간의 에너지 소비 데이터 중 최대 값을 이용하여 다음 달에 발생할 에너지 소비를 예측한다.

IV. 결론

본 연구는 공공 데이터로 공개 되는 에너지 소비 데이터를 이용하여 에너지 소비를 예측모델링을 제안하였다. 예측에는 미래 예측 연구에 많이 활용되는 마코프 프로세스를 적용하였다. 본 논문에서 제안하는 에너지 소비 예측 모델링을 사용하면 다음 달에 발생할 에너지 소비를 예측하여 에너지 수급 정책 수립에 이용할 수 있다. 향후 연구 과제로 실제 에너지 소비 데이터를 바탕으로 에너지 소비 예측 값을 생성하는 것을 남긴다.

참고문헌

- [1] Manjai Lee, "Big Data and Utilization of Public Data", Internet and Information Security Vol.2, No.2, pp. 47~64, Nov. 2011.
- [2] 10 best practices global BigData, <http://www.itsa.or.kr/>
- [3] Google Flu Trends, <http://www.google.org/flutrends/intl/ko/>
- [4] Charles M. Grinstead, "Introduction to Probability: Second Revised Edition", American Mathematical Society, pp405-406, 1997.
- [5] Young-Gab Kim, Young-kyo Baek, Hoh Peter In, Doo-Kwon Baik, "A Probabilistic Model of Damage Propagation based on the Markov Process", Journal of KIISE, Vol133, No8, pp.524-535, 8, 2006.
- [6] Young-Suk Chung, Jin-Young Jung, "A Study of the Prediction of Incidence of Crime using Markov process" Vol.17, No.3, 3, 2012.