
EM 알고리즘기반의 공기 유량 및 전력 데이터 분류 분석

심재용* · 노영빈* · 정회경** · 김용철***

*(주)벤텍
**배재대학교
***용인대학교

EM Algorithm based Air Flow and Power Data classification Analysis

Jae-Ryong Shim* · Young-Bin Noh* · Hoe-kyung Jung** · Yong-Chul Kim***

*VETEC CO.,LTD, Daejeon 305-509, Korea

**Department of Computer Engineering, Paichai University, Daejeon 302-735, Korea

***Department of Logistics & Statistical Information, Yongin University, Yongin 449-714, Korea

E-mail : yckim@yongin.ac.kr

요 약

공기압축기는 공장 및 설비 가동에 사용되는 필수 장비로서 국내 산업용 전기의 20%이상을 소비하고 있어 실시간 센서 데이터 모니터링에 의한 소비전력 절감 분석은 중요하다. 이러한 모니터링 변수들 중 특히 유량과 압력은 소비 전력과 직접적인 상관관계가 있다.

본 논문은 EM 알고리즘을 이용한 유량과 전력의 이변량 분류 분석을 통하여 유량 센서의 측정치가 센서의 측정 한계에 의한 오류인지를 파악하는 방법을 제시하여 우측 한계 측정치가 존재하는 데이터에서 더욱 정확한 유량과 전력간의 상관관계를 통한 분석이 가능하도록 하였다.

ABSTRACT

Since air compressor, as an essential equipment used in the factory and plant operations, accounts for around 20% of the total domestic electricity consumption, a real time sensor data monitoring based analysis for electricity consumption reduction is important. In particular, flow rates and pressures of these monitored variables has a direct correlation with the power consumption. This paper proposes a method to identify if the measurement error of the flow rate sensor comes from the sensor measurement limit through bivariate classification analysis of the flow rate and power using the EM (Expectation and Maximization) Algorithm and show how to enable more accurate analysis by the correlation between the flow rate and power on the right-censored data.

키워드

공기 압축기, 분류 분석, 유량 센서, 한계 데이터, EM 알고리즘

I. 서 론

공기압축기는 공기를 압축 저장했다가 각종 산업 용 장비에 공급하는 설비로, 자동차나 중공업 등 각 산업계에서 필수 장비로 활용되고 있으며 국내 전력 사용량의 20%이상을 점유하는 주요 소비 에너지원이다.

일반적으로 압축 공기 생산은 VSD(Variable Speed Drive)와 FSD(Fixed Speed Drive) 타입의 공기 압축기를 복수로 설치하여 운영하는데, 본 논문에서 참조한 시스템은 VSD 타입의 공기압축기가 먼저 운전을 시작하고 압축공기 요구량이 사전에 설정된 유량을 초과하는 시점에, 다른 FSD 타입의 공기압축기들이 순차적으로 운전을 시작하는 방식이다. FSD 타입의 경우 완전부하, 무부하, 정지의 3가지 운전종류가 있으며, 무부하운전의 경우 전력소비는 발생하지만 압축공기는 생산하지 않는다. 따라서 FSD 공기압축기의 무부하 운전을 최소화하도록 VSD 공기압축기의 출력을 가변해 수요량에 최적화된 설비를 가동하고, FSD 타입 공기압축기의 무부하 운전을 최대한 회피하여 에너지 절감효과를 얻을 수 있다[1].

공기압축기의 소비전력의 절감 방안으로는 수요 관리 기술 적인 측면에서 실시간 센서 데이터 모니터링에 의한 소비전력 절감 분석은 중요하다. 이러한 모니터링 변수들 중 특히 유량과 압력은 소비전력과 직접적인 상관관계가 있다. 이러한 상관관계는 일정 유량을 공급하기 위한 필요한 전력 요구량을 추정할 수 있고 시스템의 요구사항에 최적화된 시스템 설계의 하나의 지표로 사용될 수 있다.

본 논문은 EM 알고리즘을 이용한 유량과 전력의 이변량 분류 분석을 통하여 유량 센서의 계측치가 센서의 측정 한계에 의한 오류인지를 파악하는 방법을 제시하여 우측 한계 측정치가 존재하는 데이터에서 더욱 정확한 유량과 전력간의 상관관계를 통한 분석이 가능하도록 하였다.

II. 본 론

EM 알고리즘은 보이지 않는 잠재 변수 (latent variable)에 의존하는 확률모델에서 모수(parameter)들의 최대우도 추정치 (maximum likelihood estimates of parameters)를 찾고자하는 알고리즘이다. 즉 Expectation(E) 단계에서는 잠재 변수의 기대치를 계산하고, Maximization (M) 단계에서는 주어진 데이터와 기대치가 부여된 잠재 변수를 이용하여 모수들의 최대우도 추정치를 계산한다 [2]. EM 알고리즘은 (E) 단계와 (M) 단계의 두 가지 단계로 구성된다. 각각의 단계에서는 한 쪽 변수를 고정한 채 나머지를 갱신한다. 이런 교대 갱신 (alternating update) 알고리즘은 한 번에 수렴하지 않기 때문에, EM 알고리즘은 (E) 단계와 (M) 단계를 알고리즘이 수렴할 때 까지 반복 한다.

본 논문에서는 정상적인 데이터와, 우측 한계 측

정치가 존재하는 비정상 데이터를 구별하는 분류 변수가 잠재변수이고 확률모형은 혼합 가우시안 으로 가정 한다.

III. 분석

사례 분석 데이터는 VSD (250마력) 공기 압축기의 유량계와 전력계로부터 수집하였다. 이 현장에는 VSD 공기 압축기가 기본 로드형으로 설정되어있고 부하/무부하 타입 FSD(200마력) 공기압축기가 여러 대 설치되어있으며, 압축공기 저장 탱크를 통과하여 사용처로 나가는 배관에 전체 압축공기량(유량)을 측정하는 전체 유량센서가 설치되어 있다. 그림 1은 현장 시스템 구성도 이며 FSD#6 와 FSD#7은 보조 FSD 로서 데이터 수집 기간에는 비활성 상태였다.

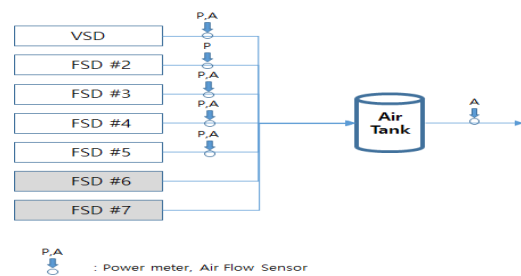


그림 1. 공기압축기 시스템

P,A는 전력계와 유량계, P는 전력계, 그리고 A는 유량계가 설치된 측정 지점을 의미한다. 본 논문에서는 전체 유량 센서 에서 우측 한계 측정치가 존재한 시점의 데이터를 이용하였다. 2015년 5월 1일부터 6월 12일까지 수집된 4,051개의 15분 데이터의 1초 평균 압축공기량 과 전력 소비량 데이터를 EM 알고리즘으로 분류하고 우측한계 측정치에 속하지 않는 그룹의 데이터만을 사용하여 설치 문제가 존재해도 현장의 압축공기량과 전력과의 상관관계를 밝힐 수 있음을 보여준다.

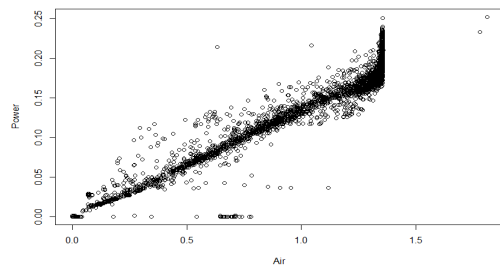


그림 2. 압축공기량 (Q (m3))과 전력(KWh)의 산점도

그림 2는 압축공기량과 전력의 산점도이며 산점도(Scatter Plot)를 통하여 압축공기 측정 센서가 우측 한계 측정치 (Right censored data)를 보이고 있음을 알 수 있다.

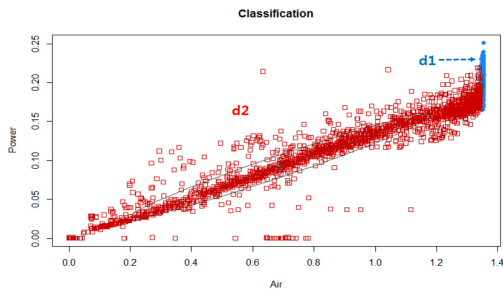


그림 3. EM 알고리즘 방법에 의한 압축공기량 (Q (m3)) 과 전력 (KWh)의 분류도

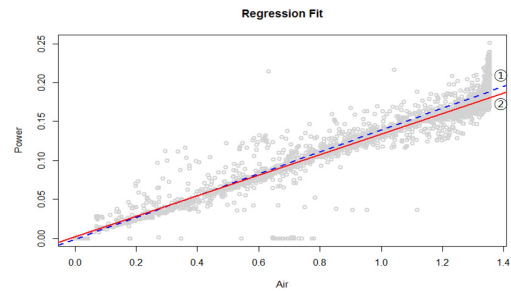


그림 4. 압축공기량 (Q (m3)) 과 전력 (KWh)의 회귀 분석

그림 3은 좌 상단의 두 개의 왜도 포인트를 제거한 후 압축공기량 (Q)와 전력의 이변량 데이터를 EM 알고리즘으로 분류한 결과를 보여준다. 분류 및 회귀 분석은 R 통계 패키지를 이용하였다[3].

그림 3의 d1은 한계측정치에 속할 수 있는 데이터의 분류를 보여주며 d2는 정상 데이터의 분류를 보여 준다. 통계적 분석으로는,

$$d1 \text{의 평균 과 공분산은 식 (13) 과 식 (14) 이며,} \\ \begin{bmatrix} air \\ pwr \end{bmatrix}_{d1} = \begin{bmatrix} 1.35 \\ 0.196 \end{bmatrix} \quad (13)$$

$$\begin{bmatrix} air \\ pwr \end{bmatrix}_{d1} = \begin{bmatrix} 0.06 & 0.085 \\ 0.085 & 2.174 \end{bmatrix} \times 10^{-4} \quad (14)$$

$$d2 \text{의 평균 과 공분산은 식 (15) 와 식 (16) 이다.} \\ \begin{bmatrix} air \\ pwr \end{bmatrix}_{d2} = \begin{bmatrix} 0.689 \\ 0.0927 \end{bmatrix} \quad (15)$$

$$\begin{bmatrix} air \\ pwr \end{bmatrix}_{d2} = \begin{bmatrix} 0.24 & 0.032 \\ 0.032 & 0.004 \end{bmatrix} \quad (16)$$

d1에 속하는 샘플 수는 1,185개, d2에 속하는 샘플 수는 2,864개이다.

d1에 약 30%에 해당하는 데이터 속해있으며 평균값(1.35) 이 최대 압축공기량인 (1.356)과 매우 근접해있고 동시에 매우 낮은 분산 값(6×10^{-6})을 보여주어 한계측정치가 존재함을 알 수 있다.

그림 4는 전체 샘플에 대하여 회귀 분석을 했을 경우 와 전체 샘플 중 EM 알고리즘에 의해 분류된 d2 영역에 속한 샘플만 이용하여 회귀 분석을 했을 경우를 보여준다.

전체 샘플을 적용했을 경우, ① ($pwr = 0.14air - 0.001$)의 표준 잔차(Residual standard error)가 0.0157이고 d2에 속한 샘플만 사용했을 경우 ② ($pwr = 0.132air + 0.002$)의 표준 잔차는 0.0146 을 보여 본 논문에서 제시하는 방법을 적용한 경우에 더욱 향상된 회귀 모형을 추정함을 알 수 있다.

IV.결 론

공기압축기는 공장 및 설비 가동에 사용되는 필수 장비로서 국내 산업용 전기의 20%이상을 소비하고 있어 실시간 센서 데이터 모니터링 에 의한 소비전력 절감 분석은 중요하다. 이러한 모니터링 변수들 중 특히 유량과 압력은 소비 전력과 직접적인 상관관계가 있다.

본 논문은 EM 알고리즘을 이용한 유량과 전력의 이변량 분류 분석을 통하여 분류의 혼합 확률, 평균, 분산 그리고 최대 측정값을 기반으로 유량 센서의 계측치가 센서 의 측정 한계에 의한 오류인지를 파악하는 방안을 제시하고 한계 측정치가 존재하는 데이터에서 더욱 정확한 유량과 전력간의 상관관계를 통한 분석이 가능하도록 하였다.

ACKNOWLEDGMENTS

This research was supported by the project 20142020104070 of KETEP (Korea Institute of Energy Technology Evaluation and Planning). We thank people from KETEP for invaluable contributions.

참고문헌

- [1] 심재용, 김용철, 노영빈, 정희경, “공기 변화량 분포를 이용한 효율적인 인버터타입 압축기 시스템” 한국정보통신학회논문지 Vol. 19, No.10, pp. 2396-2402, Oct, 2015
- [2] A.P. Dempster, N. M. Laird, D. B. Rubin “Maximum Likelihood from Incomplete Data via the EM Algorithm” Journal of the Royal Statistical Society. Series B (Methodological), Vol. 39, No.1, pp. 1-38, Jan., 1977
- [3] W. N. Venables, and D. M. Smith “An Introduction to R, 2013