

부하별 에너지 빅데이터 분석 소프트웨어 시스템

강정훈*, 유준재*, 최효섭*, 이태우*

*전자부품연구원

e-mail: budge@keti.re.kr

Energy big data analysis and classification software based on machine learning

Jeonghoon Kang*, June-Jae Yoo*, Hyoseop Choi*, Taewoo Lee*

*Korea Electronics Technology Institute

요 약

본 논문은 지속적으로 수집되는 전력량계 데이터를 자동으로 처리, 분석하기 위한 IoT 데이터 기반 자동분석 기법을 제시한다. 에너지 효율을 높이기 위해서는 대상 설비의 관리, 모니터링을 통해 운영을 최적화해야 한다. IoT 기술을 이용하여 에너지 설비 사용 효율을 확인하고, 관리 여부를 판단하는 진단기술을 구현하기 위해서는, IoT 전력량계를 통해 수집된 데이터를 다양한 머신러닝 알고리즘에 입력하여 관리에 필요한 결과 지표를 도출할 수 있어야 한다. 이런 기능을 제공하는 IoT 수집 시스템의 모니터링 및 자동 진단 시스템은 데이터 수집, 분석을 신속하게 수행할 수 있다. 데이터 수집과 고속, 대용량 데이터 저장에 적합한 분산 파일시스템과 고속 시계열 기능을 기반으로 의존도, 유사도 분석 실행을 제공하는 고속 전처리 시스템의 특징을 제안한다.

1. 서론

본 논문에서는 에너지 설비에 설치된 전력센서 데이터를 수집하여 데이터 패턴을 분석할 수 있는 기반 기술인 고속 검색과 분석 소프트웨어 기술에 대해 제안한다. 지금까지는 현장 에너지 설비상태 분석을 위해서 직접 방문하여 현장에서 측정 절차를 수행하고, 수동으로 에너지 소비 패턴 및 변수를 파악했다. 이를 통해 도출된 설정 방법을 사이트에 적용하였는데, 시간, 비용효율을 높이기 위해 근래에는 기계학습 소프트웨어를 적용하여 수작업으로 소요되는 시간을 줄일 수 있으며, 결과의 정확성도를 향상할 수 있다. 예전 수동 확인과 분석 방법은 비용효율이 낮아, 자동화된 분석 소프트웨어는 앞으로 다양하게 상용화가 시도될 것으로 예측된다.

에너지 분석 시스템이 적용된 경우, 12개월 데이터 저장 용량은 기가바이트 이상의 대용량이며, 여러 개의 사이트에서 수집된 데이터를 동시에 분석할 경우, 데이터 용량이 수십-테라바이트 규모이다. 이런 빅데이터를 관리자 또는 전문가들이 수동으로 분석, 정리하는 것은 한계가 있다. 자동화된 분석을 위해 제안한 시스템은 무선 메쉬 네트워크 기반의 전력량계 IoT 센서를 이용하여 데이터를 수집하고 데이터베이스 구축과 기반 시스템을 구현한다. 이를 기반으로 고속 시계열 데이터 접근 기술, 유사데이터의 분류, 분석을 담당하는 데이터마이닝 분석 부분으로 구성한다. 자동화된 분석을 위한 기계학습은 비쿼리방식(Query-Free Approach)의 분석 방법으로 유사도, 의존도

를 연산하고 관련도가 높은 데이터 패턴을 확인할 수 있다.

2. 전력량계 빅데이터 수집, 분석 시스템

무선 메쉬 기술은 에너지 측정 전력량계의 측정 정보 수집 단계에 사용하였다. 기존 유선 측정 시스템에 비해 설치와 유지관리 업무를 최소화할 수 있으며, 센서와 서버 구간은 저전력 무선 네트워크와 와이파이 메쉬 네트워크의 두 가지 네트워크 기술을 통합 제공한다.

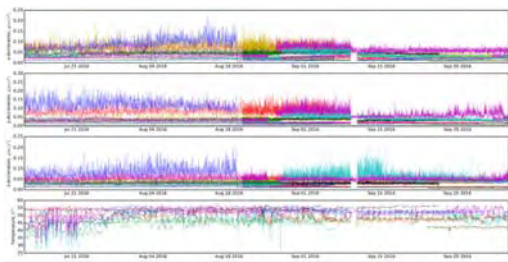
수집된 에너지 측정 정보를 저장하는 시스템은 분산 빅데이터 소프트웨어를 적용하여 구현하였다. 시스템의 저장 기능에 사용되는 HBases는 Hadoop 기반의 데이터베이스로써 Hadoop의 HDFS 위에 설치되며, Zookeeper는 Node 관리에 사용한다. RowKey를 기반으로 조회에 빠른 액세스 속도를 내지만, 쿼리문은 지원하지 않는다[1]. 대용량 데이터 Read/Write 기능과 고속스캔을 제공하고, Bigtable은 1,000개 이상의 클러스터 노드를 지원하여, 분산 컬럼 기반(distributed column oriented) NoSQL 데이터베이스 기능을 제공한다. 빅데이터 구축은 기계학습의 성능을 높이기 위해서 다양한 패턴의 데이터 특성을 도출할 수 있어야 한다. 빅데이터 기반의 기계학습을 위한 전처리, 데이터 학습 알고리즘 실행은 다양한 알고리즘을 적용해 보고, 중간 결과들을 확인하여 가장 적합한 방법을 선택하여 구현에 활용한다.



(그림 1) IoT 빅데이터 모니터링 시스템

빅데이터 시스템은 기계학습을 위해 저장된 데이터를 처리하여 주요 특징을 파악하고 데이터 패턴 학습을 수행한다. 데이터 처리 단계에는 데이터 요약, 압축과정과 같은 데이터 변환을 거쳐 정규 이벤트 (Regular Event Classification)와 비정규 이벤트 (Irregular Event Classification)를 탐색한다. 이후, 자동화된 affinity propagation을 사용하여 센서 데이터 사이의 유사성에 따라 클러스터링을 수행 한다.

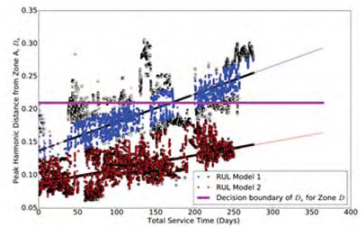
실시간 모니터링을 통해 에너지 피크를 실시간 모니터링 할 수 있으나, 측정값의 비용 및 시스템 의미를 분석하기 위해서는 6개월에서 2년 정도까지의 과거데이터를 동시에 처리하여, 현재 측정되는 데이터의 의미를 분석, 정리, 표출해야 한다. 제안한 시스템에서는 1년간의 전력 소비량 데이터에 대해서 학습을 수행하였다.



(그림 2) 데이터 히스토리 패턴

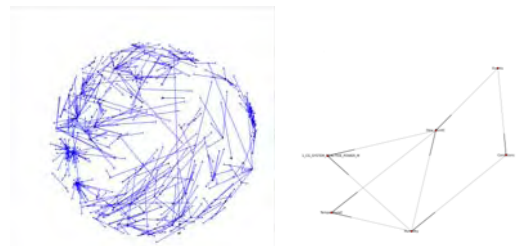
학습을 통해 기존 데이터의 패턴을 파악하면, 현재 입력 되는 데이터를 기반으로 에너지소비에 대한 전반적 판단이 가능해 진다. 사용자는 다양한 통계 연산 결과에 따라 현재 에너지 소비가 낭비 중인지, 현재 상태로 인해 전력 지출 비용이 크게 상승할 것인지에 대한 예측 결과를 실시간으로 확인할 수 있다. 이런 최종 결과의 근거로는 과거 데이터의 경험 수치가 중요하데, 기계학습 과정에서는 과거 패턴을 기억하여 현재 상황과 가장 비슷한 입력으로, 비슷한 입력 데이터가 적용된 경우에는 과거 어떤 결과를 갖게 되는지 저장된 수치를 이용하여, 예상되는 결과를 보여준다. 이런 학습방법과 과거 데이터를 기반으로 인공지능 기능을 구현하기 위해서는, 베이지안 네트워크나 라이클리후드 알고리즘을 사용한다.

이런 분석기능은 사용자의 에너지 사용 패턴을 분석하고 개선하는데 도움을 줄 수 있다. 피크분석을 예로 들면, 15분 전력 사용량 기준으로 현재 사용량과 예상 사용량을 과거 데이터에서 주기적으로 확인하는 것이 기본 기능이며, 이를 이용하여 유사한 결과를 도출하는 모델을 구현한다. 모델을 만들 때 수집한 빅데이터를 기반으로 세밀한 모델을 학습할 수 있다.

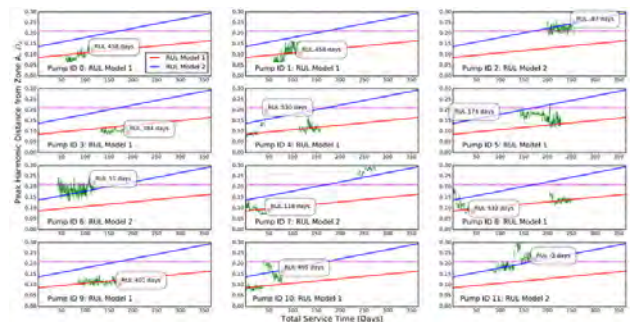


(그림 3) 원본 데이터와 학습결과의 패턴모델

시스템 표출은 전력 사용량의 히스토리 그래프와 그에 따른 예측 수치를 표시한다. 어제와 오늘 그리고 미래 전력 사용량 패턴을 확인할 수 피크 전력 예측에 도움이 되도록 하였다.



(그림 4) 데이터 유사도 연산 결과



(그림 5) 데이터 분석, 예측

3. 결론

제안한 시스템은 자동으로 에너지 소비 데이터를 학습하고 분석하여, 에너지 부하에 대한 특징을 학습하는 모델 소프트웨어이다. IoT 전력량계의 에너지 소비 형태를 분석하고 해당하는 부하를 자동으로 분류하는 것은, 에너지 효율성과를 검증하는데 필요한 기술로 해당 설비와 설치 실증 사이트가 많아질수록, 해당 시스템에 대한 요구가 늘어날 것이다. 현재는 500~1,500개의 전력량계를 대상으로 하였고, 지속적으로 추가 확장 가능하도록 구현 추진이 요구된다.

ACKNOWLEDGMENT

This work was supported by the Energy Efficiency & Resources of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government Ministry of Trade, Industry and Energy. (No.20152010103110)

참고문헌

- [1] openTSDB <http://opentsdb.net/>
- [2] HBase <http://hbase.apache.org/>
- [3] Hadoop <http://hadoop.apache.org/>