The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)
Vol. 20, No. 2, pp.13-20, Apr. 30, 2020. pISSN 2289-0238, eISSN 2289-0246

https://doi.org/10.7236/JIIBC.2020.20.2.13
JIIBC 2020-2-2

KNIME 분석 플랫폼 기반 스마트 미터 빅 데이터 클러스터링

Clustering of Smart Meter Big Data Based on KNIME Analytic Platform

김용길*, 문경일**

Yong-Gil Kim*, Kyung-Il Moon**

요 약 빅 데이터 관련 주요 논제 중의 하나는 방대한 시간 기반 또는 원격 측정 데이터의 가용성에 관한 문제이다. 현재 저비용 획득 및 저장 장치의 등장은 더 세밀한 분석에 사용될 상세한 시간 데이터를 얻을 수 있어서 배후 시스템에 대해 여러 가지 지식을 갖거나 미래의 이벤트를 더 정확히 예측할 수 있다. 특히, 스마트 미터가 설치된 수많은 가정 및 기업 등을 대상으로 전기 사용에 관한 고객 맞춤형 계약을 정의하는 것은 다른 무엇보다도 중요한 문제이다. 수많은 스마트 미터 데이터를 바탕으로 공통적인 전력 소비 형태를 몇 가지 그룹으로 구분할 필요가 있다. 이에 본 연구에서는 스마트 미터 측정 관련 공개 데이터와 자바 기반 공개 소스인 KNIME 플랫폼을 사용하여 스마트 미터 관련 빅 데이터 변환과 클러스터링을 나타낸다. 빅 데이터 구성 요소는 공개 소스는 아니지만, 시험판으로 사용할 수 있다. 스마트 미터 빅 데이터를 가져오고, 정리하고, 변환한 후 전력 사용량 행위와 관련된 각 미터 ID의 해석과 클러스터링에 적합한 DTW 접근 방식을 통해 전력 사용 행위에 관한 스마트 계약을 정의할 수 있다.

Abstract One of the major issues surrounding big data is the availability of massive time-based or telemetry data. Now, the appearance of low cost capture and storage devices has become possible to get very detailed time data to be used for further analysis. Thus, we can use these time data to get more knowledge about the underlying system or to predict future events with higher accuracy. In particular, it is very important to define custom tailored contract offers for many households and businesses having smart meter records and predict the future electricity usage to protect the electricity companies from power shortage or power surplus. It is required to identify a few groups with common electricity behavior to make it worth the creation of customized contract offers. This study suggests big data transformation as a side effect and clustering technique to understand the electricity usage pattern by using the open data related to smart meter and KNIME which is an open source platform for data analytics, providing a user-friendly graphical workbench for the entire analysis process. While the big data components are not open source, they are also available for a trial if required. After importing, cleaning and transforming the smart meter big data, it is possible to interpret each meter data in terms of electricity usage behavior through a dynamic time warping method.

Key Words : Big data, Dynamic Time Warping, KNIME, Machine learning, Smart meter

*정회원, 조선이공대학교 컴퓨터보안과
**정회원, 호남대학교 컴퓨터공학과(교신저자)
접수일자 2020년 3월 1일, 수정완료 2020년 3월 30일
게재확정일자 2020년 4월 3일

Received: 1 March, 2020 / Revised: 30 March, 2020 / Accepted: 3 April, 2020
*Corresponding Author: kimoon@honam.ac.kr
Department of computer engineering, Honam university. korea.

Ⅰ. 서 론

빅 데이터는 기존의 데이터 처리 응용 프로그램으로 처리할 수 없는 방대하고 복잡한 데이터 집합 분석, 체계적인 정보 추출 및 처리 방법을 다루는 분야이다. 행 데이터는 대부분 통계에 의해 처리될 수 있지만, 열 또는 속성과 같은 복잡도가 높은 데이터는 잘못된 탐지를 유발할 가능성이 크다[1]. 빅 데이터 개념은 볼륨, 다양성 및 속도와 관련이 있다. 빅 데이터 처리는 표본 추출 대신에 단순히 어떤 일이 발생하는지 관찰하고 추적하기 때문에, 종종 전통적인 프로그램이 수용할 수 있는 시간과 값 내에서 처리할 수 있는 용량을 초과하는 크기의 데이터가 사용된다. 빅 데이터의 현재 적용은 예측 분석, 사용자 행동 분석 또는 데이터에서 가치를 추출하기 위한 어떤 다른 고급 데이터 분석을 사용하는 추세이다. 현재 이용 가능한 데이터의 양이 실제로 큰 것은 사실이지만, 이것이 새로운 데이터 생태계의 일반적인 특성인 것은 아니다[2]. 빅 데이터의 주요 구성 요소와 생태계는 첫 번째, A/B 테스트, 기계 학습 및 자연어 처리와 같은 데이터 분석 기법, 두 번째로 비즈니스 지능, 클라우드 컴퓨팅 및 데이터베이스와 같은 빅 데이터 기술, 세 번째로 도표, 그래프 및 기타 데이터 전시와 같은 시각화로 요약할 수 있다[3].

전력망에 사용되는 스마트 장치는 실시간으로 전력 소비를 측정하고 통신 네트워크를 통해 데이터를 전송하여 저장 및 분석할 수 있는 스마트 전력계이다. 이 기술의 흥미로운 점은 에너지 절약을 위해 전력 소비에 대한 실시간 피드백을 사용하는 데에 있다. 과거 에너지 사용자는 주로 소비하는 에너지의 양과 시기를 알고자 하지 않았는데, 주된 이유는 정보 수집의 번거로움과 저에너지 소비 활동 및 고에너지 활동의 영향에 대한 과소평가에 있었다. 이제 스마트 미터의 도입과 함께 미터기가 생성하는 방대한 데이터를 저장하는 데 도움이 되는 기술의 개발 덕분에 이제 전력 사용자는 실질적인 에너지 경제 활동이 가능해졌다. 특히, 에너지 공급 측면에서 스마트 미터에서 생성된 실시간 데이터 소비를 분석하여 저렴하고 생태적인 방식의 생산이 가능하게 되었다. 문제는 스마트 미터 데이터가 매우 규모가 큰 빅 데이터라는 점에서 데이터 관리가 매우 복잡하다는 것이다. 따라서 스마트 그리드 관련 비즈니스에서 빅 데이터 처리 및 분석 관련 적절한 플랫폼이 요구되고 있다. KNIME(Konstanz Information Miner)는 개방형 소스의 데이터 분석 통합 플랫폼으로 모듈형 데이터 파이프라인 개념을 통해 기계 학습 및 데이터 마이닝을 위한 다양한 구성 요소들을 통합하는 시스템이다[4]. GUI 및 JDBC의 사용은 최소한의 프로그래밍을 통해 데이터 모델링, 분석 및 가시화를 위한 전처리를 (추출, 변환 및 로딩과 같은) 포함하여, 서로 다른 데이터 소스를 혼합하는 노드 조합을 허용한다. KNIME는 데이터 흐름 또는 파이프라인을 시각적으로 만들고 일부 또는 모든 분석 단계를 선택적으로 실행한 후에 결과, 모델 및 대화형 보기를 검사할 수 있다. KNIME은 Java로 작성되었기 때문에, 확장 방식을 사용하여 추가 기능을 지원할 수 있다. KNIME의 주요 특징은 코어 아키텍처를 통해 대용량 데이터 볼륨을 처리할 수 있는데, 예를 들어, 3억 개의 고객 주소, 2천만 셀 이미지 및 1천만 분자 구조를 분석할 수 있다. 추가 플러그인을 사용하면 다양한 인공지능 기법들을 통합할 수 있다. KNIME는 Spark, Weka, H20, R 및 LIBSVM의 기계 학습 알고리즘과 같은 다양한 기타 개방형 소스 프로젝트를 통합한다. 자바로 구현되지만, Python, Perl 및 기타 코드 부분을 실행할 수 있는 노드를 제공하며, 다른 코드를 호출하는 래퍼도 허용한다. 현재 에너지 산업은 현재 상당한 변화를 겪고 있는 산업으로 매우 복잡한 네트워크, 초기 독점 시장의 경쟁자 개방, 에너지 전달 및 소비자 가격 규제 강화, 안전한 에너지에 대한 요구 사항이 늘고 있다. 긍정적인 추세 중 하나는 스마트 미터의 도입이다. 스마트 계량기는 에너지 회사가 고객이 에너지 사용을 이해하고 관리하도록 도와주는 장치이다. 또한, 에너지 회사가 최소 미터까지, 더 나아가서 기기 또는 장치 수준까지 에너지 요구 사항을 더 잘 예측할 수 있는 장치이다. 이에 본 연구에서는 에너지 산업의 지능화에 중점을 두고 공개적으로 사용 가능한 스마트 미터 빅 데이터와 공개 소스의 데이터 분석 플랫폼이라 할 수 있는 KNIME를 사용하여 빅 데이터 처리에 관한 통찰력을 높이는 데에 있다. 본 연구의 2장에서는 스마트 그리드의 빅 데이터 시스템 모델에 관한 관련 연구와 모델 구성체제 및 KNIME 분석 플랫폼의 작업 흐름도를 소개한다. 3장에서는 스마트 미터 시계열 데이터의 클러스터링과 관련하여 기존 k-Means 방식은 상당 부분 취약하다. 따라서, 미터 데이터의 동적인 시간 왜곡 문제를 처리하기 위한 접근 방식과 KNIME 작업 흐름도 구성을 나타낸다.

Ⅱ. 스마트 그리드 빅 데이터 시스템

스마트 그리드에 관해 최근 빅 데이터 관련 다양한 연구들이 진행되고 있다. 이 장에서는 스마트 그리드에 관

한 빅 데이터 개념 연구 및 관련 기술들에 관한 근래 연구를 요약하고, 이를 바탕으로 하는 스마트 그리드 빅 데이터 체제의 전반적인 구조와 데이터 분석 시스템 모델을 소개하고, KNIME 분석 플랫폼을 나타낸다.

1. 관련 연구

스마트 그리드의 빅 데이터 개념과 관련해서 Ye 등은 수요 반응 및 전역적인 모니터링 문제와 관련하여 스마트 그리드의 빅 데이터 분석 및 클라우드 컴퓨팅을 소개하고 있으며[5], Zhou 등은 에너지 빅 데이터의 잠재력을 실현하고 스마트 에너지 관리를 위한 통찰력을 얻기 위한 빅 데이터 중심 스마트 에너지 관리의 포괄적 연구를 진행하였다[6]. Vasilakos 등은 스마트 그리드 빅 데이터 분석 및 보안 문제에 관한 개발에 대한 조사와 논의를 제공하였으며[7], Munshi와 Mohamed는 스마트 그리드의 혁신과 관련하여 안전한 클라우드 기반 플랫폼에서 스마트 그리드를 구현할 수 있는 빅 데이터 틀을 제안하였다[8]. Liang 등은 다양한 데이터 가격 책정 모델, 그룹화 모델 등을 통해 데이터 거래의 수명 주기에 대한 포괄적인 조사를 제시하였다[9]. Hou 등은 대규모 스마트 그리드를 위한 새로운 시간적, 기능적, 공간적 빅 데이터 컴퓨팅 체제를 제안했다[11]. 스마트 그리드 빅 데이터의 취약한 부분에 대한 포괄적인 조사는 Chin 등을 참조할 수 있다[10]. 스마트 그리드의 빅 데이터 기술과 관련해서 Wen 등은 압축 기술을 제안하고 있고[11], Jiang 등은 신뢰할 수 없는 이기종 분산 시스템 환경에 저장된 암호화된 다차원 미터 데이터를 효율적으로 질의하는 방식으로 다차원 미터 암호화 빅 데이터에 대한 고성능 및 개인 정보 보호 질의 체계를 제안했다[12][13]. Chang과 Huang은 컴퓨팅 전력 및 대역폭이 제한되는 상황에서 네트워크 혼잡 발생을 방지하거나 신속하게 대응할 수 있는 빅 데이터 트래픽에 대해 경량급의 통신 접근법을 소개했다[14]. Liu 등은 현명한 에너지 관리를 위한 의사 결정 지원을 목표로 이상 소비 패턴을 탐지하는 람다 시스템을 제안했다[15]. 빈번한 패턴 마이닝을 통해 스마트 미터의 에너지 소비 빅 데이터에 적용되는 자율형 점진적 증가의 데이터 마이닝 방식은 Singh와 Yassine의 연구를 참조할 수 있다[16]. 스마트 그리드 빅 데이터에서 대량의 가격 데이터를 처리하는 새로운 전력 가격 예측 모델은 Wang 등의 연구를 참조할 수 있다[17]. Sheng 등은 빅 데이터 개념을 바탕으로 마이크로 그리드 최적의 스케줄링 및 부하 축소를 위한 새로운 모델을 제안하였으며[18], 빅 데이터를 기반으로 하는 트랜스포머 상태 파라미터에 대한 데이터 중심의 연관 규칙 마이닝 방법은 사전 확률적 그래픽 모델을 결합하는 형태로 소개되고 있다[19].

2. 스마트 그리드 빅 데이터 분석 시스템 모델

그림 1은 일반적인 스마트 그리드 빅 데이터 분석 시스템 모델의 구조이다. 시스템은 데이터 생성기, 데이터베이스 및 데이터 분석의 3가지 블록으로 나눌 수 있다. 데이터 생성기 블록과 관련하여, 스마트 그리드에서 빅 데이터 기술의 성능 평가를 위해서는 스마트 그리드 데이터 분석을 위해 페타바이트 규모의 데이터를 처리해야 한다. 현실적인 데이터를 사용하는 것이 훨씬 좋지만 개인 정보 보호 문제와 진실한 데이터 소스가 없으므로, 수백만 고객 레코드의 데이터 집합을 얻는 것은 힘든 일이다. 가상 데이터 집합 생성과 관련해서 Spark의 스마트 미터 데이터 생성기를 사용할 수 있다[20]. 데이터 생성기는 작은 실제 데이터 집합을 바탕으로 데이터를 생성할 수 있는 지도학습 형태의 기계 학습 알고리즘을 사용한다. 전형적인 데이터 생성기는 사전 프로세스 모듈, Spark 모듈 및 HDFS 파일 시스템 모듈과 같은 다른 블록을 가지고 있다. 전처리 모듈은 먼저 시드 데이터의 각 필드에서 형식과 값을 확인한 다음에 고객을 클러스터링하여 개인화된 에너지 효율 서비스를 제공한다[21]. 또한, Spark 모듈로 전송할 시계열 데이터를 준비하며, 데이터 생성기 엔진에는 메모리 기반 분산 컴퓨팅 체제인 Spark를 사용하여 구현되는 학습 모듈과 데이터 생성 모듈의 두 가지가 있다.

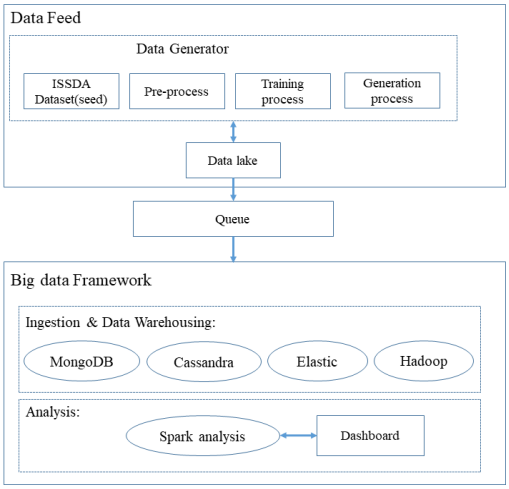

그림 1. 그리드 빅 데이터 분석 모델
Fig. 1. Grid Big Data Analytics Model

첫 번째 단계로 임시 데이터를 저장하기 위해 RDD (Resilient Distributed Dataset)가 RAM에 내장되어 있다. 시드 데이터는 주기적 변동을 평탄화하고 조정된 시계열을 만들기 위해 이동 평균 시계열 모델을 사용하는데 계절별로 정리되지는 않는다. 피크 주기를 평활화하여 이동 평균은 실제 시계열의 주기 변화를 줄이고 납작한 시계열로 훈련된 자동 회귀 모델을 사용하여 미터 판독 값을 예측하는데, 결국은 주기적인 변형이 다시 추가된다. 훈련 과정은 그룹화된 시드 데이터를 입력으로 하여 모델을 생성하고 출력을 HDFS에 기록한다. 생성된 데이터는 stream 지향적 특성을 갖기 때문에 Kafka의 경우는 모델에서 빅 데이터 체제의 메시지 큐로 사용된다. 출력은 Kafka 대기열을 통해 NoSQL 데이터베이스에 삽입된다. 생성된 데이터에는 미터 ID, 시간, 전력 소비 및 환경 온도를 포함한 4개의 필드가 있다. 미터 ID는 1000~5001000의 범위를 가지며, 시간은 5자리 숫자 필드로 저장되는데, 처음 3자리 숫자는 요일을 나타내고, 다음 2자리는 1~48 사이의 값으로 설정된 시계를 나타낸다. 시작 0의 시간은 자정에 해당한다. 전력 소비 필드는 수치 데이터로 모든 데이터는 열 기반으로 구성되며 변경 범위가 완전히 지정되어 있다[23].

3. KNIME 분석 플랫폼

KNIME는 데이터 과학 응용 프로그램 및 서비스 생성을 위한 개방형 소스 소프트웨어로 직관적이고 개방적이며 데이터 과학 작업 흐름도를 사용하여 데이터를 이해할 수 있도록 한다(그림 2 참조). KNIME 분석 플랫폼은 코드 작성이 거의 요구되지 않으며, 직관적인 drag & drop 스타일의 그래픽 인터페이스로 시각적 작업 흐름도를 구축할 수 있다. 2,000개가 넘는 모듈과 광범위한 통합 도구 및 다양한 고급 알고리즘을 제공한다. 분석 플랫폼에서 개별 작업은 그림 2와 같이 모두 노드로 표시되며 상호 연결된 노드 모음이 작업 흐름도이다. 이러한 작업 흐름도는 일반적으로 특정 데이터 분석 프로젝트의 일부 또는 전부를 나타낸다. 각 노드에는 구성 대화 상자를 통해 조정할 수 있는데, 이러한 조정을 하게 되면, 각 노드 아래에 신호등으로 노드 상태가 표시된다. 신호등이 빨강이면 구성되지 않은 상태이고, 구성된 경우에는 노란색, 실행 가능한 경우에는 초록색으로 표시되며, 에러가 있는 경우에는 빨간색의 x 표시가 나타난다. 노드는 파일 읽기/쓰기, 데이터 변환, 모델 훈련, 시각화 생성 등의 작업을 할 수 있다.

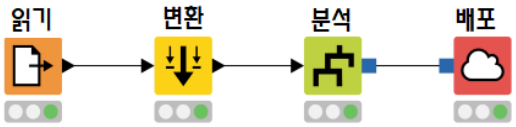

그림 2. KNIME 분석 플랫폼의 일반적인 작업 흐름도
Fig. 2. General workflow of KNIME analysis platform

KNIME의 작업 흐름도와 관련하여 작업대의 구성 요소는 다음과 같이 요약될 수 있다. 먼저, Explorer는 작업 공간에서 사용 가능한 작업 흐름도 및 작업 흐름도 그룹의 탐색에 사용된다(로컬 작업 공간, KNIME 서버와 개인 KNIME 허브 공간 등). 두 번째로 작업 흐름도 Coach는 KNIME 사용자 커뮤니티에서 구축한 작업 흐름도를 바탕으로 하는 노드 권장 사항들을 나열한다. 세 번째로 노드 Repository는 분석 플랫폼 및 확장에서 이용 가능한 모든 노드를 나열하는데 노드는 범주별로 구성되어 있다. 노드 저장소 맨 위 검색 상자에서 노드를 찾을 수도 있다. 네 번째의 작업 흐름도 편집기는 현재 작업 흐름도를 편집하기 위해 사용되는 일종의 캔버스이다. 다섯 번째로 Description은 현재 작업 흐름도 또는 선택 노드에 대한 설명을 나타낸다. 여섯 번째의 Outline은 현재 작업 흐름도에 대한 개요를 나타내며, 마지막으로 Console은 진행 중인 작업을 나타내는 실행 메시지를 표시한다.

Ⅲ. KNIME 기반 빅 데이터 분석

1. DTW 기반 스마트 미터 클러스터링

시계열 색인화와 관련된 알고리즘 대부분은 유클리드 거리 또는 그 변형을 사용하지만, 유클리드 거리는 매우 취약한 거리 측정값이다. DTW(Dynamic Time Warping)는 시계열에 대해 훨씬 더 강력한 거리 측정 방식으로, 시간 축에서 위상이 다른 경우에도 유사한 모양을 일치시킬 수 있는 유연성을 갖는다[22]. 변압기 스테이션 레벨 또는 네트워크의 다른 상위 레벨에서 여러 소비자를 분석하여 일반적인 일일 로드 프로파일의 모양을 식별할 수 있는데, 예를 들어, 15분 간격으로 일일 로드 프로파일을 작성하는 경우에 몇 개의 그룹으로 많은 일일 로드 프로파일을 그룹화하려면 거의 시간 간격이 비슷한 로드 프로파일을 그룹화해야 한다. 일일 로드 프로파일이 다른 peak 로드보다 15분 늦게 가질 수 있다. 즉, 동적 시간 왜곡이 발생하게 된다. 스마트 미터로부터 같은 시간 주

기에 걸쳐 추정된 두 가지 로드 곡선 $x_p = \{x_{p1}, ..., x_{pn}\}$와 $x_q = \{x_{q1}, ..., x_{qm}\}$을 고려하자. 여기서, n와 m는 각각 스마트 미터 p와 q에 의한 데이터 점들의 개수이다. 로드 곡선 x_p와 x_q의 데이터 점 j와 k의 로컬 거리 측정값은 다음과 같이 정의될 수 있다: $d(x_{pj}, x_{qk}) = \|x_{pj} - x_{qk}\|^2$. 크기 $n \times m$의 행렬식 으로 나타내면 거리 행렬 $D(x_p, x_q)$는 데이터 점들의 각 쌍의 로컬 거리로 구성된다. 거리 행렬 D의 a_l번째 행과 b_l번째 열에 대응하는 셀을 w_l라 할 때, 워핑 경로로서 $w = \{w_1, ..., w_L\}$을 정의할 수 있는 데, 타당한 워핑 경로가 되기 위해서는 첫 번째로 거리 행렬의 대각선 첫 번째 원소에서 시작하여 대각선 마지막 원소로 종료되어야 한다. 즉, $w_1 = (1,1)$이고 $w_L = (n,m)$. 두 번째는 워핑 경로 인접성 및 단조성으로 경로는 인접 셀들로만 국한되며, 단조 증가이어야 한다. 즉, $w_r = (c,d), w_{r+1} = (e,f)$에 관해 $0 \le e - c \le 1, 0 \le f - d \le 1$. 이제 워핑 경로의 총 거리는 다음과 같다.

$$d_w = \sum_{l=1}^{L} d(x_{pa_l}, x_{qb_l}) \quad (1)$$

x_p와 x_q 간에 DTW 거리는 모든 가능한 워핑 경로들 가운데 최소 거리로 이것은 동적 계획 알고리즘을 사용하여 구할 수 있다. 이러한 워핑 경로를 갖는 DTW를 유사성 측도로 사용하면, 시간 축의 신축 및 수축에 관계가 없이 유사한 모양의 로드 곡선 클러스터링을 할 수 있다. 특히, 초기 시드를 신중하게 선택하고 클러스터링 과정을 주시함으로써 기존 클러스터링 알고리즘의 한계를 극복하고 더 나은 클러스터링 결과를 얻을 수 있다. 클러스터링 성능은 최소 거리를 찾는 동적 계획법에서 순환 방식으로 오차 제곱 합을 구하는 방식으로 얻을 수 있다.

2. KNIME 작업 흐름도 구성

KNIME 분석 플랫폼을 기반으로 하는 스마트 미터 데이터 클러스터링을 위한 단계들은 4가지로 구분할 수 있다. 먼저, 각 미터 ID의 시계열 데이터를 정리한 후에 매일 또는 매시간 단위로 집계한다. 두 번째 단계는 가져온 시계열 데이터상에서 각 미터 ID에 관한 전력 사용량의 일부 측정값을 정의하고 계산한다. 세 번째 단계는 시계열 측정값들을 바탕으로 모든 미터 ID들에 대한 DTW 클러스터링을 실행하여, 분석 공간을 축소한다. 이는 추후 각 클러스터에 대해 과거 값을 바탕으로 추후 에너지의 사용량을 추정하기 위한 예측 모델을 생성하는 데에 있다. 마지막 단계는 빅 데이터 플랫폼을 사용하여 첫 번째와 두 번째 단계에서 행해진 작업을 다시 수행한다.

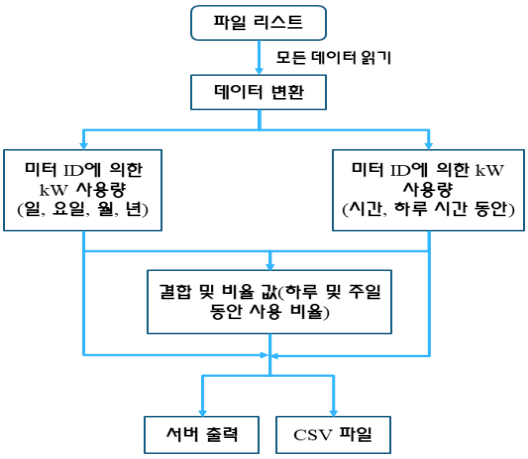

그림 3. 데이터 준비 작업 흐름도 구성
Fig. 3. Configure data preparation workflow

그림 3은 위 첫 번째와 두 번째 단계를 수행하기 위한 작업 흐름도이다. 즉, 데이터를 가져오고, 정리하고, 변환하고, 미터 ID당 전기 사용량에 대한 설명적인 측정값을 계산한다. 스마트 미터 데이터와 같은 빅 데이터의 처리는 하드 디스크를 비롯한 컴퓨팅 자원의 소비가 상당하므로, KNIME 서버와 같은 공유 플랫폼에서 데이터를 보유하는 것이 바람직하다.

로컬 파일 시스템이 아닌 특정 서버에서 데이터 접근을 하기 위해서 그림 3의 파일 리스트 작업은 그림 4와 같이 세분될 수 있다. 데이터베이스 커넥터 노드는 JDBC 드라이버를 통해 데이터베이스에 연결한다. 연결하려면 일반적으로 JDBC 드라이버 파일, 서버 URL, 액세스 포트, 데이터베이스 이름 및 데이터베이스에 액세스하기 위

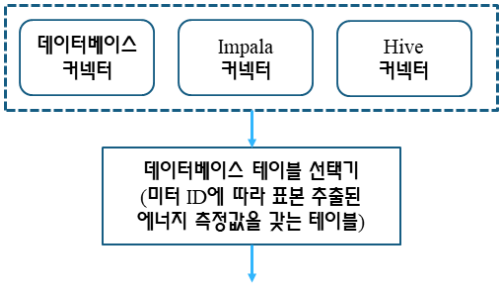

그림 4. 빅 데이터 플랫폼 연결
Fig. 4. Big data platform connection

한 자격 증명이 요구된다. 기본적으로 자격 증명은 암호화되지만 기본 설정 페이지에서 마스터키를 설정하여 사용자 이름 및 암호를 명시적으로 나타내야 한다. KNIME 플랫폼의 경우에 탐색기 패널에서 작업 흐름을 마우스 우측 버튼으로 클릭하여 작성할 수 있다. 노드 저장소 패널의 데이터베이스 범주에는 일반 데이터베이스 커넥터 노드가 있는데, 이 노드는 적절한 JDBC 드라이버 파일의 경로를 포함한 정보가 제공되면 가능한 모든 데이터베이스에 연결된다. 빅 데이터 플랫폼의 경우 사용하기 쉬운 전용 커넥터 노드가 많이 있다. 실제로 일반 데이터베이스 커넥터 노드를 사용하는 경우 빅 데이터 플랫폼에 연결하는 것은 복잡한 작업이다. 일부 설정은 이미 노드에 하드 코딩되어 있으므로 전용 노드를 사용하면 연결 설정을 더 쉽고 빠르게 수행할 수 있다. 그림 5는 유사한 전력 소비 행위의 그룹화를 위한 DTW 스마트 미터 ID의 클러스터링 작업 흐름도이다. 기본적인 소비 데이터는 5가지로 분류될 수 있는데, 첫 번째는 월요일부터 일요일까지 각 요일에 사용된 에너지의 평균 및 백분율, 두 번째는 이른 아침(7~9시), 아침(9~13시), 이른 오후(13~17시), 늦은 오후(17~21시) 및 야간(21~7시) 동안 사용된 에너지의 평균 및 백분율, 세 번째는 테스트 시간 동안 사용된 총 에너지(kW), 네 번째는 연간, 월간, 주별, 일별 및 시간별 사용된 에너지(kW) 평균, 다섯 번째는 주말과 평일에 사용된 평균 에너지이다. 클러스터링 작업과 관련하여 특정 변수가 다른 변수에 미치는 영향을 가능한 배제하기 위해 가능한 평균값은 사용하지 않는 것이 바람직하다. 클러스터링 알고리즘은 거리를 바탕으로 하므로 비교 가능한 범위인 0과 1 사이의 값으로 표준화시켜야 한다. 또한, 결측값 처리가 필요한데, 결측값 처리에는 다소간 통계적 기법이 요구되는데, 일단은 0으로 처리하여 에너지 소비가 없는 것으로 한다. 클러스터 개수는 30개 정도로 할 수 있지만, 현실적으로 20여 개 이상의 소비자 맞춤형 계약을 관리하는 데 어려움이 있다.

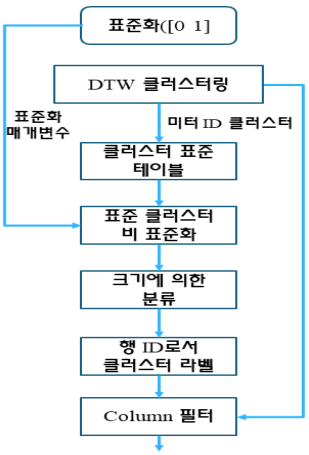

그림 5. DTW 기반 미터 ID 클러스터링
Fig. 5. DTW-based meter ID clustering

Ⅳ. 시뮬레이션

본 연구의 시뮬레이션에서는 KNIME에서 제공하는 아일랜드 스마트 에너지 프로젝트의 에너지 데이터를 사용한다(대략 6,000개의 가정 및 기업의 전기 사용량을 스마트 미터 ID를 통해 시간에 따라 모니터링). 먼저, 그림 6은 KNIME 분석 플랫폼에서 제안된 절차에 의한 스마트 미터 클러스터링의 전체 작업 흐름도이다. 그림 7은 그림 6의 File Reader와 관련된 설정 화면이고, 그림 8은 그림 6의 빅 데이터 환경 설정 관련 화면을 나타낸다.

그림 9는 일일 소비 데이터에 관한 클러스터링 결과 출력 화면의 일부를 나타낸다. 유사하게, 시간, 요일, 월별 클러스터링 결과를 출력할 수 있다. 일일 소비 데이터를 바탕으로 하는 클러스터링 결과 몇 가지 흥미로운 그룹들이 등장하는데, 주로 심야에만 전기를 소비하는 집단, 일일 전기 소비가 균등한 집단, 직장 퇴근 후에 저녁에 전기 소비가 많은 집단, 일일 전기 소비가 무분별한 집단 등으로 해석할 수 있는 다양한 소비 형태가 있음이 확인되었다.

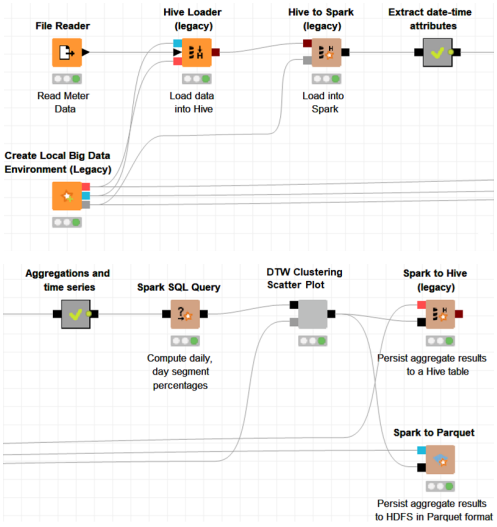

그림 6. 미터 클러스터링 작업 흐름도
Fig. 6. Meter clustering workflow

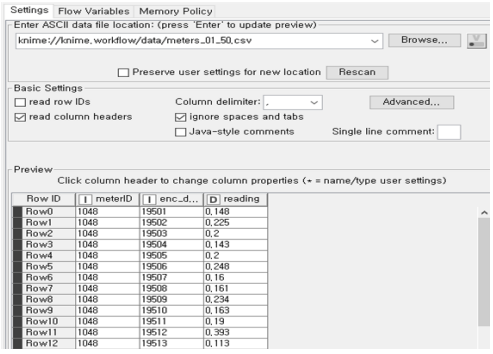

그림 7. 미터 데이터 설정 화면
Fig. 7. Meter data setting screen

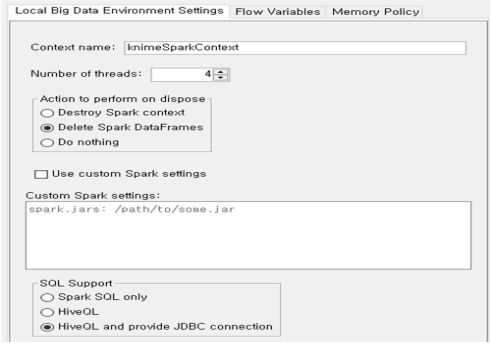

그림 8. 빅 데이터 환경 설정 화면
Fig. 8. Big Data Preferences Screen

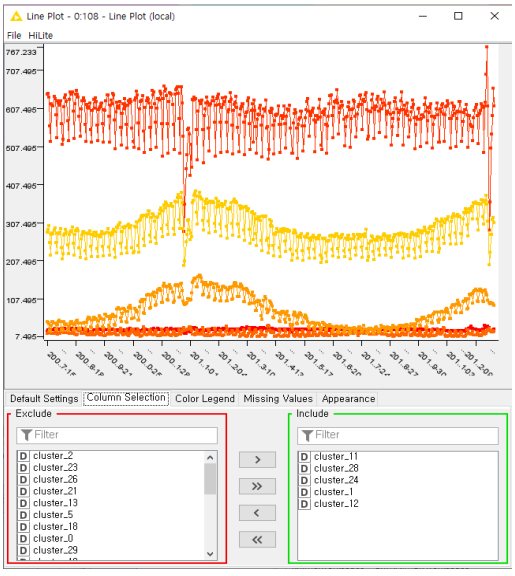

그림 9. 일일 소비에 의한 클러스터링
Fig. 9. Clustering by Daily Consumption

V. 결 론

스마트 미터 관련 빅 데이터를 정리하고 변환한 후에 유사한 스마트 미터 ID를 그룹화하는 작업은 추후 그리드 계획, 수요 반응 또는 에너지 효율 프로그램에서 소비자의 상세한 프로파일링, 확률적 로드 흐름 분석의 수행 등에 사용될 수 있다. 따라서 유사한 미터 ID를 그룹화하기 위한 클러스터링 알고리즘이 필요하다. 통상적으로, 유사한 에너지 소비를 기술하는데 사용된 정량적 특징과 클러스터링 절차에서 채택된 거리 측정이 기본이다. 여기서 잘 확립된 클러스터링 알고리즘 적용과 거리 측정은 가능하지만, 입력 패턴의 정량적 설명은 어려운 부분이다. 전력 사용의 시간에 따른 추세가 중요하다면, 하루 중 다른 요일과 시간에 소비된 에너지의 양을 측정할 필요가 있고, 적절한 정량적 측정을 통해 각 미터 ID 시계열을 분석하고, 관리 가능한 소수의 군집으로 그룹화시킬 필요가 있다. 이는 빅 데이터와 관련된 대부분 플랫폼에서 비교적 새로운 분석 영역이다. 특히, 시계열 분석에 적용하는 경우에 미래 데이터 예측과 관련하여 특정 데이터 변환이 필요한데, 이는 빅 데이터 분석과 관련하여 적합한 주제라 할 수 있다.

References

[1] T. Breur. “Statistical Power Analysis and the contemporary crisis in social sciences”. Journal of Marketing Analytics. Vol.4, No.2-3, pp.61-65, 2016. DOI: https://doi.org/10.1057/s41270-016-0001-3

[2] D. boyd, K. Crawford. “Six Provocations for Big Data. Social Science Research Network”. A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society. 2011. DOI: https://doi:10.2139/ssrn.1926431.

[3] J. M. Manyika, J. Bughin, B. Brown, R. Dobbs, C. B. Roxburgh, A. H. Byers. “Big Data: The next frontier for innovation, competition, and productivity”. McKinsey Global Institute. Retrieved 16 January 2016.

[4] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, K. Thiel, B. Wiswedel. “KNIME - the Konstanz information miner”. ACM SIGKDD Explorations Newsletter. Vol.11, No.1. 26, 2009. DOI: https://doi.org/10.1145/1656274.1656280

[5] F. Ye, Y. Qian, R. Q. Hu. “Big data analytics and cloud computing in the smart grid”. Smart Grid Commun. Infrastruct. pp.171-185, 2018. DOI: https://doi.org/10.1002/9781119240136.ch8

[6] K. Zhou, C. Fu, S. Yang. “Big data driven smart energy

management: from big data to big insights". Renew Sustain Energy Rev. 56: pp.215-225, 2016.
DOI: https://doi.org/10.1016/j.rser.2015.11.050

[7] A. Vasilakos, J. Hu. "Energy big data analytics and security: challenges and opportunities". IEEE Trans Smart Grid. pp.99:1, 2016.
DOI: https://doi.org/10.1109/TSG.2016.2563461

[8] A. A. Munshi, Y. A. Mohamed. "Big data framework for analytics in smart grids". Electric Power Syst Res. 151:3, pp.69-80, 2017.
DOI: https://doi.org/10.1016/j.epsr.2017.06.006

[9] F. Liang F, W. Yu, D. An, Q. Yang, X. Fu, W. Zhao. "A survey on big data market: pricing trading and protection". IEEE Access. 6:15, pp.132-154, 2018.
DOI: https://doi.org/10.1109/ACCESS.2018.2806881

[10] W. Chin, W. Li, H. Chen. "Energy Big Data Security Threats in IoT-Based Smart Grid". Communications (October), pp.70-75, 2017.
DOI: https://doi.org/10.1109/MCOM.2017.1700154

[11] W. Hou, Z. Ning, L. Guo, X. Zhang. "Temporal, functional and spatial big data computing framework for large-scale smart grid". IEEE Trans. Emerg. Topics Comput. 6750(1):1-1, 2017.
DOI: https://doi.org/10.1109/TETC.2017.2681113

[12] L. Wen, K. Zhou. S. Yang, L. Li. "Compression of smart meter big data: a survey". Renew Sustain Energy Rev. 91(April): pp.59-69, 2018.
DOI: https://doi.org/10.1016/j.rser.2018.03.088

[13] Y.G. Kim, K.I. Moom, "Noisy Weighted Data Aggregation for Smart Meter Privacy System,"The Journal of The Institute of Internet, Broadcasting and communication(JIIBC),Vol.18, No.3,PP.49-59. 2018
DOI: https://doi.org/10.7236/JIIBC.2018.18.3.49

[14] R. Jiang, R. Lu, K. K. R. Choo. "Achieving high performance and privacy-preserving query over encrypted multidimensional big metering data". Future Gener. Comput. Syst. 78: pp.392-401, 2018.
DOI: https://doi.org/10.1016/j.future.2016.05.005

[15] Y. C. Chang, T. C. Huang. "An interactive smart grid communication approach for big data traffic". Comput. Electr. Eng. 67: pp.170-181, 2018.
DOI: https://doi.org/10.1016/j.compeleceng.2018.03.024

[16] X. Liu, P. S. Nielsen. "Scalable prediction-based online anomaly detection for smart meter data". Inf. Syst. 77:pp.34-47, 2018.
DOI: https://doi.org/10.1016/j.is.2018.05.007

[17] S. Singh, A. Yassine. "Mining energy consumption behavior patterns for households in smart grid". IEEE Trans. Emerg. Topics Comput. 6750(c):1-1, 2017.
DOI: https://doi.org/10.1109/TETC.2017.2692098

[18] K. Wang, C. Xu, Y. Zhang, S. Guo, A. Zomaya. "Robust big data analytics for electricity price forecasting in the smart grid". IEEE Trans. Big Data 7790(c):1-1, 2017.
DOI: https://doi.org/10.1109/TBDATA.2017.2723563

[19] G. Sheng, H. Hou, X. Jiang, Y. Chen. "A novel association rule mining method of big data for power transformers state parameters based on probabilistic graph model". IEEE Trans. Smart Grid. 9(2): pp.695-702, 2018.
DOI: https://doi.org/10.1109/TSG.2016.2562123

[20] N. Iftikhar, X. Liu, S. Danalachi, F. E. Nordbjerg, J. H. Vollesen. "A scalable smart meter data generator using spark". In: OTM Confederated International Conferences On the move to meaningful internet systems. Cham: Springer; pp.21-36, 2017.
DOI: https://doi.org/10.1007/978-3-319-69462-7_2

[21] N. Iftikhar, X. Liu, F. E. Nordbjerg, S. Danalachi. "A Prediction-Based Smart Meter Data Generator". In: NBiS 2016 - 19th international conference on network-based information systems. pp.173-180, 2016.
DOI: https://doi.org/ 10.1109/NBiS.2016.15

[22] K. Eamonn, A. R. Chotirat. "Exact indexing of dynamic time warping". Springer-Verlag London Ltd. Knowledge and Information Systems. 2004.
DOI:https://doi.org/:10.1007/s10115-004-0154-9.

[23] J. Walkenbach. Kafka: "The definitive guide" 2010.
DOI: https://doi.org/10.1017/CBO9781107415324.004.

저 자 소 개

김 용 길(정회원)

- 1990년 호남대학교 전산통계학과 졸업(이학사)
- 1992년 광주대학교 대학원 컴퓨터학과 졸업(공학석사)
- 2014년 ~ 현재 : 조선이공대학교 컴퓨터보안과 부교수
- 관심분야 : 네트워크보안, 통신시스템, 정보보호

문 경 일(정회원)

- 1982년 서울대학교 계산통계학과 졸업(이학사)
- 1980년 서울대학교 대학원 계산통계학과 졸업(이학석사)
- 1991년 서울대학교 대학원 계산통계학과 졸업(이학박사)
- 1987년 ~ 현재 : 호남대학교 컴퓨터공학과 교수
- 관심분야 : 지능 시스템, 복잡성 과학