

데이터마이닝 기법을 이용한 변압기 부하패턴 분석

신진호, 김영일, 이봉재, 송재주, 양일권
한전 전력연구원

Load Pattern Analysis of Distribution Transformer using Data Mining Techniques

Jin-Ho Shin, Young-Il Kim, Bong-Jae Yi, Jae-Ju Song, Il-Kwon Yang
Korea Electric Power Research Institute

Abstract - 시간 데이터마이닝은 기존 데이터마이닝에 시간 개념을 추가하여 시간 속성을 가진 데이터로부터 이전에 잘 알려지지 않는 패턴을 찾아내고 잠재적으로 유용한 시간 지식을 탐사하는 기술이다. 이 논문에서는 시간 속성을 가진 변압기 부하 패턴에 대해 시간의 변화에 따른 적용 시점이 명확한 지식 탐사가 가능하고, 향후 부하 예측에 있어 탐사된 규칙과 시간 지식을 이용함으로써 기존의 정적인 분류규칙을 적용한 방법보다 더 정확한 예측을 할 수 있는 새로운 시간 패턴 마이닝 기법을 제안한다.

1. 서 론

전력 부하 예측을 위한 정확한 분석 모델은 전력 운영과 계획에 필수적이며 전력 산업에 있어서의 부하 예측은 전력의 구입, 생산, 로드스위칭, 그리고 기반 시설의 업데이트 등에 대한 의사 결정을 함에 있어서 도움을 준다. 특히, 부하 예측은 공급자, ISO 회사 및 재무기관이 전기 에너지 생산, 전송, 분배, 및 관련 마케팅에 있어서 매우 중요하다.

전력산업에 있어서 효율적인 운용과 계획을 위해 정확한 부하 및 부하 패턴 예측 기술이 필요하며, 이를 위해 통계 및 데이터마이닝 등과 같은 수학적 방법들이 부하 분석 모델링을 위해 사용된다. 부하 패턴 예측 과정은 기존의 부하패턴을 식별하고 통계적 분석 및 데이터마이닝 기술을 적용하여 새로운 부하 예측 하는 것을 말한다. 특히, 전력시스템에서의 데이터마이닝 기술은 부하데이터로부터 규칙성을 인지하고 추출하는 가장 대표적인 기술이며 데이터마이닝 기법 적용 결과인 패턴 규칙 집합은 부하 데이터로부터 이전에 알려지지 않은 부하 패턴을 식별할 수 있게 한다. 또한 부하 패턴 규칙 집합은 현재의 부하패턴과 분석을 위해서 비교 되어 질 수 있다. 일반적으로, 데이터마이닝 기술을 적용한 부하패턴 예측은 관련된 정보로부터 부하 패턴 모델을 생성하고 이 모델을 적용하여 새로운 부하패턴을 예측한다.

이 논문에서는 시간 속성을 가진 변압기 부하 패턴에 대해 시간의 변화에 따른 적용 시점이 명확한 지식 탐사가 가능하고, 향후 부하 예측에 있어 탐사된 규칙과 시간 지식을 이용함으로써 기존의 정적인 분류규칙을 적용한 방법보다 더 정확한 예측을 할 수 있는 새로운 시간 패턴 마이닝 기법을 제안한다. 먼저 캘린더 스키마 및 캘린더 패턴의 시간 마이닝 개념을 정의하고, 기본 시간 간격에서의 3차원(시간-변압기-부하패턴) 패턴 마이닝 수행을 위한 3D 큐브 마이닝 수행하며, 3D 큐브 마이닝 수행 결과인 3차원 패턴에 대한 달력(calendar) 패턴을 주기성 표현 방식을 이용하여 개발결과를 설명한다.

2. 본 론

2.1 캘린더 패턴과 3차원 큐브 마이닝

캘린더 스키마는 달력의 개념 계층에 의해 결정되어지고 유효성 제약 조건을 갖는 관계형 스키마이다. 캘린더 스키마(CS : Calendar Schema), 캘린더 스키마는 달력 표현의 시간 단위와 그 단위에서의 가능한 도메인의 집합으로 정의되며, 그 형태는 다음과 같다.

$$CS = (G_n : D_n, G_{n-1} : D_{n-1}, \dots, G_1 : D_1)$$

$1 \leq i \leq n$ 에 대해, 속성 G_i 는 년, 월, 일 등과 같은 달력 개념에서의 시간 단위이고, 각 D_i 는 양의 정수의 유한 집합으로 속성 G_i 의 도메인 값의 집합을 나타낸다. 캘린더 패턴(CP : Calendar Pattern), 캘린더 패턴은 주어진 스키마 $CS = (G_n : D_n, \dots, G_1 : D_1)$ 의 인스턴스이며, $CP = (d_n, \dots, d_1)$ 으로 표현된다. 여기서 각 d_i 는 D_i 의 도메인 값이거나 문자 '*'이다. 만약 d_i 가 '*'이라면 그 의미는 도메인 D_i 의 모든 값을 나타내고 "every"로 해석한다.

캘린더 스키마 CS 에 대한 캘린더 패턴 CP 가 주어지면, CP 에 의해 포함되어지는 타임스탬프의 트랜잭션을 $D(CP)$ 로 나타낸다. 문법적으로 캘린더 연관규칙은 $\langle CAR, CP \rangle$ 형태로 표현되며, CAR 은 클래스

연관규칙이고 CP 는 CS 에 대한 캘린더 패턴이다. 또한 $\Phi(CP_0)$ 은 CP 의 모든 기본시간단위 집합을 나타낸다.

기존의 2D FCP(Frequent Closed Pattern) mining 알고리즘은 2차원 매트릭스 표현 데이터를 이용하여 빈발 항목집합을 발견하였다. 따라서 3D 부하패턴 데이터에서 시간의 속성을 고려한 빈발 항목집합을 발견하는데 적용할 수는 없다. 하지만, Representative Slice Mining framework는 기존의 2D FCP 알고리즘을 적용하여 3D 데이터에서 빈발 항목 집합을 발견하는 것이 가능하다. Representative Slice Mining framework는 3D 데이터 셋 $O = H \times R \times C$ 을 $O = H \times Slice_{R \times C}$ 로 표현하는 것은 기본으로 한다. 따라서 H 집합을 처음으로 계산하는 것과 같이 임의의 차원을 기준으로 Slice의 조합을 생성할 수 있다. 다음으로 $R \times C$ 와 같은 Slice 조합은 2D FCP 알고리즘을 적용할 수 있다. 다음으로 post-processing 단계에서 unclosed된 항목 집합을 제거한다.

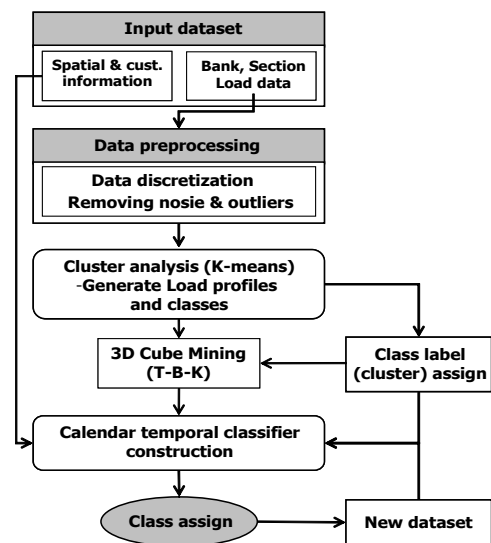
3차원 큐브 마이닝은 FCCs 마이닝을 직접 3D 데이터 집합 (Time-Bank-Cluster)에 적용하여 시간의 변화에 따른 부하 패턴 항목을 찾을 수 있는 새로운 방법이다. 이 방법은 전체 3D 데이터 집합에서 cutter로 불리는 Z의 집합을 이용하여 순환적으로 분리해 나가는 방법을 이용하여 Z를 순차적으로 사용하여 cube의 값이 1을 가질 때 까지 반복된다.

2.2 시간변화에 따른 부하패턴 예측 분류 규칙

시간 클래스 연관규칙 탐사로 생성된 규칙들은 신속하고 효율적인 분류규칙 생성을 위해 사용된다. 이 논문에서의 부하패턴 예측을 위한 시간 연관적 분류규칙 탐사의 전체 프레임워크는 그림 1과 같다.

시간 연관적 분류규칙 탐사를 위해서 먼저, 3D 큐브 마이닝 및 캘린더 패턴 클래스 연관규칙 탐사 알고리즘을 적용하여 모든 규칙들을 생성한다. 그 다음 생성된 분류규칙들은 테스트 데이터를 이용하여 정확성 여부를 검사한다. 만약 정확한 분류규칙들이 생성된다면 클래스 라벨이 없는 새로운 데이터를 시간 속성을 고려하여 분류하게 된다.

탐사된 시간 클래스 연관규칙들로부터의 분류규칙 생성을 위해서 모든 규칙들은 각 캘린더 패턴 CP_i ($CP_i \in \Phi(CP_0)$)에 의해 그룹핑 된다. 또한 각 그룹(같은 캘린더 패턴)의 규칙($CAR_k(CP_i)$)들은 기준에 의해 내림차순으로 정렬된다.



<그림 1> 시간 데이터마이닝 프레임워크

2.3 변압기 부하패턴 분석

무선 부하 감시 시스템에 의해 30분 단위로 측정하는 변압기의 대표 패턴을 저압고객 특성 및 변압기 공간 정보 등을 고려한 시간 데이터마이닝 기법적용을 수행한다. 변압기의 하루 단위 부하 패턴을 생성하기 위하여 무선 부하 감시 데이터에서 बैं크 ID와 일시 ID, 변압기 부하량 데이터를 추출한다.

〈표 1〉 부하 패턴 기본 데이터

속성명	데이터 타입	코드명
BANK_ID	nominal	뱅크 ID
DATE_TIME_ID	continuous	일시 ID
KW	continuous	변압기 부하량 (일별 30분 간격)

3D 큐브 마이닝에 변압기의 부하 패턴을 적용하기 위해서 추출된 데이터에서 부하 대표 패턴을 생성한다. 큐브 마이닝을 위해서 생성된 부하 대표 패턴을 बैं크 ID, 일시 ID, 부하패턴 속성으로 기본 데이터를 구성한다. 생성된 기본 데이터의 बैं크 ID, 일시 ID, 부하 패턴 군집 번호를 3차원 속성으로 지정한다. 지정된 속성을 순차 정렬하고, 이를 인덱스 형태로 전환하여 기본 데이터를 [뱅크 ID X 일시 ID X 부하 패턴 군집수] 크기의 0과 1로 구성된 순차적인 3차원 큐브형 데이터로 전환한다.

〈표 2〉 변압기 인덱스 형태 전환

INDEX	0	1	2	3	4	...	209	210
뱅크 ID	22	24	25	27	28	...	4844	4845
뱅크 매칭	r1	r2	r3	r4	r5	...	r210	r211

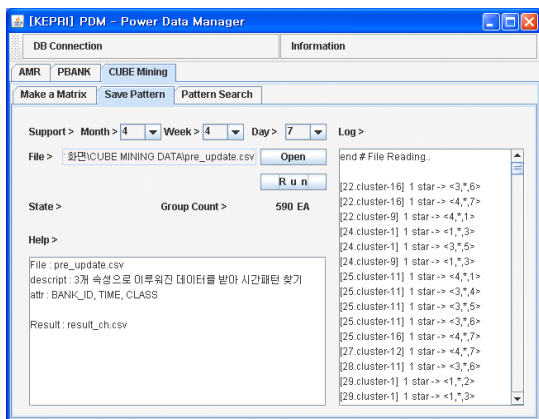
시간의 변화에 따른 빈발한 बैं크_ID 및 대표 패턴(cluster)을 찾기 위해서 생성된 매트릭스 데이터를 큐브 마이닝 알고리즘에 적용한다. 3D 큐브 마이닝 알고리즘을 실행하기 위해서는 알고리즘 실행하기 전에 입력 데이터의 길이(시간 ID 개수), 행의 개수(변압기 수량), 열의 개수(대표 패턴 수)를 입력한다. 또한 3가지 각 수량에 대한 기본 지지도를 입력하여 사용자가 원하는 빈발 패턴을 연도록 한다. 알고리즘 수행 결과에서 기본 시간 간격에서의 패턴을 찾기 위해서는 캘린더 패턴에 대한 분류규칙 형식으로 전환한다. 변압기 부하 패턴 예측에 사용된 캘린더 스키마는

$$CS = (G_n : D_m G_{n-1} : D_{n-1}, \dots, G_1 : D_1) \Rightarrow$$

$$CS = (Month : \{1, 2, 3, 4\}, Week : \{1, 2, 3, 4\}, Day : \{1, 2, 3, 4, 5, 6, 7\})$$

이다.

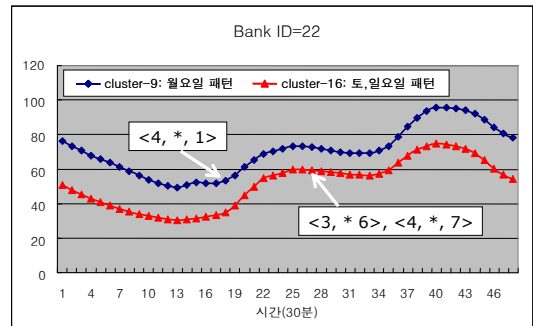
알고리즘 적용 후 일정한 달력 패턴을 추출하기 위해서 기본 시간 간격에서의 패턴을 갱신한다(그림 2). 기본 시간 간격 갱신은 3가지의 각 속성의 개수가 사용자에게 의해 정해진 지지도를 만족 할 경우 '*'로 표시하고 'EVERY'로 해석한다. 예를 들어, <MM, WEEK, DAY>의 갱신 결과가 <3, *, 6> 일 경우 '3월 달 매주 토요일'로 해석된다. 추출된 달력 패턴으로 변압기의 향후 발생하는 부하 패턴을 예측한다.



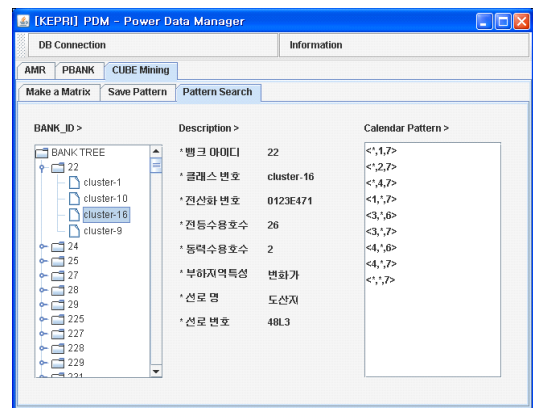
〈그림 2〉 기본 시간패턴 갱신 화면

그림 3,4는 시간 데이터마이닝 기법을 적용한 변압기 부하 패턴 예측에 대한 예이다. 그림에서 Bank ID=22인 Bank는 '3월 중 매주 토요일과 4월 중 매주 일요일' 대표 패턴(cluster 16)이 같으며, '4월 중 매주 월요일'

일' cluster 9번의 대표 패턴을 갖는다는 의미이다. 또한 이러한 시간 패턴을 가지는 22번 बैं크의 특징은 '전산화 번호가 0123E41이고 전등수용호수가 26, 동력수용호수 2, 공간적 특성으로 변화가인 부하지역특징 변화, 선로명이 도산지'인 특징을 가진다.



〈그림 3〉 변압기 시간패턴 및 부하패턴 예시



〈그림 4〉 변압기 부하패턴의 주기성 분석 화면

3. 결 론

공급과 수요의 변동, 기후의 변화, 에너지 가격의 변동 등 여러 요소에 의해서 부하 예측은 전력산업에 있어서 더욱 중요하게 되었다. 단기 예측 같은 경우에는 부하의 흐름을 예측하여 과부하의 발생을 방지할 수 있으며, 이와 같은 주기적인 작업들은 전체 네트워크의 안정성을 제고할 수 있고 장비의 고장 현상을 줄일 수 있다. 뿐만 아니라 계약물에 대한 평가, 그리고 전력 시장에서 가격 변동에 의한 다양하고 복잡한 상품들에 대한 평가를 함에 있어서도 중요하다.

이 논문에서는 데이터마이닝 기법을 이용하여 변압기 부하패턴을 분석하였다. 시간 패턴 마이닝은 기본 시간 간격에서의 3차원(시간-변압기-부하패턴) 패턴 마이닝 수행하고, 3D 큐브 마이닝 수행 결과인 3차원 패턴에 대해 캘린더 패턴을 적용하여 변압기 부하패턴의 주기성을 평가하는 알고리즘을 개발하였다. 향후에는 날씨민감도와 고객변동을 고려한 변압기 부하패턴 예측 연구를 수행하여 데이터의 신뢰성을 개선할 예정이다.

[참 고 문 헌]

- [1] Huang S. J., Shih K. R., "Short-term load forecasting via ARMA model identification including non-Gaussian process considerations,"IEEE Trans. Power System. Vol. 18, No. 2, pp. 673-679. 2003.
- [2] Pitt B. and Kirchen D., "applications of data mining techniques to load profiling." In Proc. IEEE PICA, pp. 131-136. 1999.
- [3] Li Y., Ning P., "Discovering Calendar-based Temporal Association Rules," In Proc. of the 8th Int'l Symposium on Temporal Representation and Reasoning.
- [4] Jain A. K., Dubes R. C., "Algorithms for Clustering Data," Englewood Cliffs, NJ: Prentice-Hall. 1998.
- [5] Chicco G., Napoli R., Postulache P., Scutariu M., and Toader C., "Customer characterization options for improving the tariff offer,"IEEE Tans. Power System, Vol. 18, pp. 381-387, 200