

Study of Temporal Data Mining for Transformer Load Pattern Analysis

辛 鎭 浩[†] · 李 倖 在^{*} · 金 榮 一^{**} · 李 憲 揆^{***} · 柳 根 鎬[§]
 (Jin-Ho Shin · Bong-Jae Yi · Young-Il Kim · Heon-Gyu Lee · Keun-Ho Ryu)

Abstract - This paper presents the temporal classification method based on data mining techniques for discovering knowledge from measured load patterns of distribution transformers. Since the power load patterns have time-varying characteristics and very different patterns according to the hour, time, day and week and so on, it gives rise to the uninformative results if only traditional data mining is used. Therefore, we propose a temporal classification rule for analyzing and forecasting transformer load patterns. The main tasks include the load pattern mining framework and the calendar-based expression using temporal association rule and 3-dimensional cube mining to discover load patterns in multiple time granularities.

Key Words : Transformer Load Pattern, Temporal Data Mining, Calendar Pattern, Association Rule, 3D Cube Mining

1. 서 론

전력산업에 있어서 효율적인 설비 운용과 계획을 위해 정확한 부하 및 부하패턴 예측 기술이 필요하며, 이를 위해 통계[1-2] 및 데이터마이닝[3]과 같은 수학적 방법들이 부하분석 모델링을 위해 사용된다. 특히, 전력시스템에서의 데이터마이닝 기술[4]은 부하데이터로부터 규칙성을 인지하고 추출하는 가장 대표적인 기술이며, 데이터마이닝 기법의 적용 결과인 패턴 규칙 집합[5]은 부하 데이터로부터 이전에 알려지지 않은 부하패턴을 식별할 수 있게 한다. 일반적으로 데이터마이닝 기술을 적용한 부하패턴 예측[6]은 관련된 정보로부터 부하패턴 모델을 생성하고 이 모델을 적용하여 새로운 부하패턴을 예측한다.

시간 데이터마이닝은 기존 데이터마이닝에 시간 개념을 추가하여 시간 속성을 가진 데이터로부터 이전에 잘 알려지지 않는 패턴을 탐색하는 기술이다. 시간 속성을 가지는 연관규칙에 대한 이전 연구들은 크게 주어진 시간 간격 동안 주기적으로 발생하는 현상, 즉 시간 간격에서의 완전한 주기성을 만족하는 연관규칙을 탐사하는 주기적 연관규칙 탐사[7]와 캘린더로 표현된 시간 패턴을 가지는 연관규칙을 탐사하는 캘린더 기반 연관규칙탐사[8-9]로 분류할 수 있다.

주기적 연관규칙이란 트랜잭션이 발생한 전체 시간을 사

용자에 기반한 시간 단위(년, 월, 일)에 따라 시간 간격의 집합으로 나누고, 시간 구간에서의 완전한 주기성을 갖는 모든 연관규칙들을 탐사한다. 그러나 실제로 주어진 시간 간격 동안에 완전하게 유지되는 규칙은 존재하지 않으며 대부분 불완전한 주기를 이루고 있다. 또한 주기적 연관규칙에서는 다양한 시간 단위를 표현하지 못하고 단 하나의 시간 단위만을 다룰 수 있다. 따라서 “매달 첫 번째 월요일”과 같은 실제 응용분야에서 적용되는 시간 표현은 불가능하다.

이전 연구[8-10]에서는 주기적 연관규칙을 확장하여 연관규칙을 탐사할 때 사용자 기반의 시간 패턴을 탐사하는 기법을 제안했다. 캘린더 연관규칙은 시간 패턴의 명시를 위해서 캘린더 대수(*calendar algebra*)를 사용한다. 캘린더는 시간 간격의 집합으로 정의되고, 캘린더 대수는 시간 연산자를 기반으로 캘린더 사이의 시간관계를 구할 수 있는 연산자를 제공한다. 이 기법은 주기적 연관규칙 보다 유용함과 실용적인 시간 연관규칙을 생성할 수 있으나 시간 패턴 탐사를 위해 사용자 기반의 캘린더 대수 표현이 요구되어지며, 이는 사용자 하위급 탐사될 시간 패턴에 대한 정확한 이전 지식을 필요로 하는 단점을 가진다. 또한, 시간 표현을 위한 방법으로 캘린더 시스템에 의존한 특정 시점이나 시간 간격의 데이터를 분석하는데 그쳤으며, 시간에 따른 다양한 상호관련성을 제시하는 측면에서는 한계성을 가진다.

따라서 이 논문에서는 시간 속성을 가진 변압기 부하패턴에 대해 시간의 변화에 따른 적용 시점이 명확한 지식 탐사가 가능하고, 향후 부하예측에 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 마이닝을 정의하고, 기본 시간 간격에서의 3차원(시간-변압기-부하패턴) 패턴 마이닝 수행을 위한 3D 큐브 마이닝 수행하며, 3D 큐브 마이닝 수행 결과인 3차원 패턴에 대한

† 교신저자, 正會員 : 韓國電力公社 電力研究員
 E-mail : jinho@kepco.co.kr

* 正會員 : 韓國電力公社 電力研究員

** 正會員 : 韓國電力公社 電力研究員

*** 非會員 : 忠北大學校

§ 非會員 : 忠北大學校

接受日字 : 2008年 9月 9日

最終完了 : 2008年 10月 20日

캘린더 패턴을 주기성 표현 방식을 이용하여 개발결과를 설명한다.

2. 캘린더 패턴과 3차원 큐브 마이닝

2.1 캘린더 스키마와 캘린더 패턴 마이닝

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

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

$1 \leq i \leq n$ 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 스키마가 $(G_n, G_{n-1}, \dots, G_1)$ 이고 $1 \leq i < n$ 일 때, 각 시간 단위 G_i 는 유일하게 G_{i+1} 에 포함된다. 캘린더 패턴 (CP: Calendar Pattern)은 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. $CS = (G_n, G_{n-1}, \dots, G_1)$, $CP = \{d_n, d_{n-1}, \dots, d_1\}$,

$CP' = \{d'_n, d'_{n-1}, \dots, d'_1\}$, $1 \leq i \leq n$ 인 각 i 에 대해, $d_i = '*'$ 이거나 $d_i = d'_i$ 일 때, CP 는 CP' 을 포함한다.

데이터베이스를 $D, I = \{i_1, i_2, \dots, i_n\}$ 을 모든 항목집합 그리고 $C = \{c_1, c_2, \dots, c_m\}$ 을 클래스 라벨 집합이라고 한다. 클래스 연관 규칙은 $X \rightarrow C_i$ 의 형태로 표현되고 $X \subset I, C_i \subset C$ 이다. 여기서 클래스 연관 규칙의 전제부를 ItemSet라하고 규칙 자체를 RuleItem라 한다. 클래스 연관 규칙의 지지도와 신뢰도 식 (2)와 같다.

$$support = \frac{RuleSupCnt}{|D|}, \quad confidence = \frac{RuleSupCnt}{ItemSupCnt} \quad (2)$$

RuleSupCnt는 RuleItem의 개수 값으로 ItemSet을 포함하면서 클래스 C_i 로 라벨된 D 안의 항목들의 개수이다.

ItemSupCnt는 ItemSet을 포함하는 D 안의 항목들의 개수이다.

시간 클래스 연관 규칙 탐사의 문제 정의를 위해서 먼저 모든 트랜잭션 데이터베이스 D 는 트랜잭션 시간이 주어진 타임스탬프와 관련된다고 가정한다. 캘린더 스키마 CS 에 대한 캘린더 패턴 CP 가 주어지면, CP 에 의해 포함되어지는 타임스탬프의 트랜잭션을 $D(CP)$ 로 나타낸다. 문법적으로 캘린더 연관 규칙은 $\langle CAR, CP \rangle$ 형태로 표현되며, CAR 은 클래스 연관 규칙이고 CP 는 CS 에 대한 캘린더 패턴이다. 또한 $\phi(CP_n)$ 은 CP 의 모든 기본시간단위 집합을 나타낸다.

시간 패턴 마이닝에서는 기존의 지지도, 신뢰도 외에도 유용성 측면에서 발생도(frequency)라는 새로운 임계값을 사용한다. 최소발생도(minimum frequency), 규칙, $\langle CAR, CP \rangle$ 가 $\phi(CP_n)$ 에 속한 기본시간단위의 $f\%$ 이상을 만족한다면, 규칙, CAR 은 캘린더 패턴 CP 를 만족한다. 이 때, $f\%$ 을 최소발생도, minFre이라고 정의한다. 즉, 규칙,

$\langle (A \wedge B) \Rightarrow (Class = 1) \rangle$ 은 매주 수요일에 발생된다는 의미이다. 최소발생도의 적용 이유는 실제 응용에서 주어진 시간 간격 동안에 모두 성립되는 규칙은 거의 없기 때문에 유연성을 주기 위함이다. 요약해 보면 시간 클래스 연관 규칙 탐사는 데이터베이스 D 로부터, 사용자가 미리 지정한 최소발생도 minFre에 대해서 최소지지도 minSup과 최소신뢰도 minConf를 만족하는 규칙들을 탐사하는 문제이며, 발생도를 고려함으로써 시간에 따라 변화하는 규칙을 탐사할 수 있다[9-10].

2.2 3차원 큐브 마이닝

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 항목집합을 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.

3차원큐브 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 데이터집합에서 cutter로 불리는 Z 의 집합을 이용하여 순환적으로 분리해 나가는 방법을 이용하여 Z 를 순차적으로 사용하여 cube의 값이 1을 가질 때 까지 반복된다.

2.3 캘린더 패턴의 갱신 알고리즘

임계값을 만족하는 큐브 연산 후의 모든 $CAR_k(CP_i)$ 들을 생성한 후, 각 i -star pattern에 대한 빈발한 $CAR_k(CP_i)$ 들을 계산한다. 이 단계에서는 i -star pattern으로 저장과 검색을 위해 해시트리를 이용한다.

- ① 루트 노드를 시작으로 하여 CP_i 안의 서로 다른 '*'의 분포로부터 루트의 자식노드로 차례로 매핑이 이루어진다. 예를 들어 만약, $n=3$ 이고 $i=2$ 이면 (n 은 캘린더 패턴의 시간 단위의 개수, i 는 '*'의 개수) 해시트리의 루트는 $\{*, *, d_1\}, \{*, d_2, *\}, \{d_3, *, *\}$ 으로 표현되는 3개의 자식노드를 갖는다.
- ② 루트의 한 자식노드를 시작으로 CP_i 가 저장될 위치를 결정한다. 여기서 자식노드는 해시백터를 이용하여 단말노드를 위치시킨다. 해시백터는 CP_i 안의 모든 양의정수로 구성된다. 예를 들어 $CP_i = (1, *, 2)$ 일 경우, 해시백터는 $\langle 1, 2 \rangle$ 가 된다.
- ③ 해시백터가 계산된 후, 캘린더 패턴 CP_i 그리고 CP_i 과 관련된 규칙인 $CAR_k(CP_i)$ 은 단말노드에 저장된다.

3. 시간변화에 따른 부하패턴 분류 규칙

시간 클래스 연관 규칙 탐사로 생성된 규칙들은 신속하고

효율적인 분류규칙 생성을 위해 사용된다. 이 논문에서의 부하패턴 예측을 위한 시간 연관적 분류규칙 탐사의 전체 프레임워크는 그림 1과 같다. 먼저 운영계 시스템에서 설비의 공간정보와 고객정보, 변압기의 집합인 बैं크와 구간정보, 현장 변압기에서 30분 단위로 계측되는 상별 전압과 전류의 부하 데이터를 연계하여 부하분석을 위한 데이터를 수집한다. 구축한 부하분석 데이터베이스에 포함된 오류 및 outlier 데이터는 군집분석의 성능을 크게 저하시킬 수 있기 때문에 데이터 정제를 위한 전처리 작업은 필수적이다. 계측된 원시 부하 데이터에서 하루 48개 데이터가 존재하지 않으면 분석 대상에서 제외하였다. 또한 outlier 처리를 위하여 데이터 정제 기법 중 SOM(Self-Organizing feature Map) 군집화 알고리즘을 적용하였다. 구성 매트릭스는 10 by 10 (100 군집)으로 한 클러스터에 포함된 데이터 객체가 1개 이하인 군집 결과는 outlier로 간주하여 제외하였다.

이때 원시 데이터를 그대로 사용하게 되면 군집화가 전력 사용량의 분포에 따라 형성된다. 따라서 변압기별 전류 최대값이 1이 되도록 정규화한 다음 군집분석을 수행해야 한다. 많은 군집화 알고리즘 중에서 본 논문에서는 대용량 데이터에서 빠른 군집 구성, 사용자 기반의 군집 수 결정방식, 군집의 적합성을 판단하는 재현율 적용 용이성 등을 고려하여 k-means 알고리즘을 적용하였다.

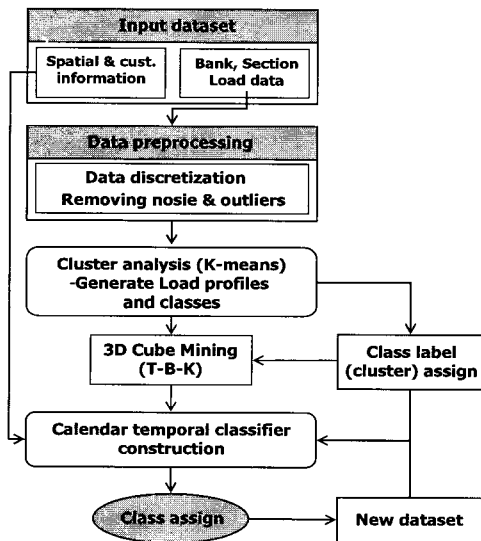


그림 1 시간 데이터마이닝 프레임워크
Fig. 1 Temporal data mining framework

시간 연관적 분류규칙 탐사를 위해서 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 ($CP_i \in \Phi(CP_i)$)에 의해 그룹핑된다. 또한 각 그룹(같은 캘린더 패턴)의 규칙($CAR_k(CP_i)$)들은 기준에 의해 내림차순으로 정렬된다.

두 규칙 $CAR(CP_i)_m, CAR(CP_i)_n$ 가 주어지면, 조건 (1), (2)에 의해 $CAR(CP_i)_m$ 은 $CAR(CP_i)_n$ 보다 높은 우선순위 ($CAR(CP_i)_m > CAR(CP_i)_n$)를 갖는다.

- (1) $Frequency(CAR(CP_i)_m) > Frequency(CAR(CP_i)_n)$ 또는,
 - (2) $Frequency(CAR(CP_i)_m) = Frequency(CAR(CP_i)_n) \wedge CAR(CP_i)_m$ 이 $CAR(CP_i)_n$ 의 Sub-Rule이다.
- ($\therefore CAR(CP_i)_m = (CAR(CP_i)_n)_{sub}$)

분류규칙 생성을 위한 규칙 선택시 *i-star pattern*에서 가장 발생도가 높은 규칙을 우선적으로 선택하는 것을 보장한다. 그러나 만약 두 규칙의 발생도가 같은 경우에는 일반화된 규칙을 선택하도록 한다. 캘린더 패턴에 대한 분류규칙의 형식 아래 식과 같다.

$$\langle CAR(CP_i)_1, CAR(CP_i)_2, \dots, CAR(CP_i)_n, deRule, \{CP\} \rangle \quad (3)$$

여기서, $CAR(CP_i)_j$ 는 탐사된 시간 클래스 연관규칙이고 *deRule*는 디폴트 규칙, $\{CP\}$ 는 그 규칙들이 유효한 캘린더 패턴 표현이다. 시간 클래스 연관규칙에 의한 연관적 분류규칙 생성의 기본접근법은 다음과 같다.

- ① 규칙들이 포함된 각 캘린더 패턴에 대해, 훈련 데이터의 기본시간단위가 캘린더 패턴에 포함되어지는 모든 데이터를 선택한다.
- ② 각각의 규칙 $CAR(CP_i)_j$ 가 정확하게 어떤 데이터를 포함(covering)할 경우, 규칙 $CAR(CP_i)_j$ 을 분류규칙 생성을 위한 규칙으로 선택한다.

데이터 $d(CP_0)$ 가 규칙 $CAR(CP_i)_j$ 의 ItemSet을 모두 포함한다면 $d(CP_0)$ 은 $CAR(CP_i)_j$ 에 의해 포함(covered)된다고 하고 또는, 규칙 $CAR(CP_i)_j$ 가 훈련 데이터 $d(CP_0)$ 를 포함(covering)한다고 말한다. 분류규칙의 생성 후에 클래스 라벨이 없는 새로운 데이터의 분류에 대해, 새로운 데이터는 그 데이터를 포함하는 첫 번째 규칙의 클래스 라벨을 할당 받는 것을 원칙으로 한다.

분류를 위한 모든 규칙들을 선택한 후에 새로운 데이터를 분류한다. 먼저 분류될 새로운 데이터의 기본시간단위를 포함하고 그 데이터를 정확하게 분류할 수 있는 모든 규칙들 $CAR(CP_i)$ 를 분류규칙으로 선택한다. 만약 모든 분류규칙들이 같은 클래스를 가지면 단순히 그 클래스를 새로운 데이터에 할당한다. 그러나 다른 클래스를 가질 경우, 규칙들을 각 클래스 별로 그룹핑한다. 그런 다음 각 그룹의 해당 클래스에 대한 강도를 *CMAR*[10-11]의 *weighted- χ^2* 상관분석을 적용하여 측정된 뒤 가장 강한 그룹의 클래스를 할당한다. *weighted- χ^2* 의 계산을 위해 먼저 각 클래스 별 그룹의 모든 분류규칙들에 대한 카이제곱(χ^2) 값의 계산이 필요하다. 카이제곱(χ^2)의 계산은 관찰된 값(O)과 기대 값(E)을 이용하며, 식(4)과 같다.

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad (4)$$

각 클래스별 그룹의 *weighted- χ^2* 의 계산은 먼저 각 규칙들의 *maximum- χ^2* 값을 구한다. 규칙 $R=A \rightarrow C$ 에 대한 *maximum- χ^2* 식은 다음과 같다.

$$maximum-\chi^2 = \left(\frac{support(A)support(C)}{N} \right)^2 \cdot N \cdot e \quad (5)$$

여기서 N 은 훈련데이터의 레코드 수이고 e 는 식 (6)이다.

$$e = \frac{1}{support(A)support(C)} + \frac{1}{support(A)(N-support(C))} + \frac{1}{(N-support(A))support(C)} + \frac{1}{(N-support(A))(N-support(C))} \quad (6)$$

따라서 χ^2 와 식 (5)의 $maximum \chi^2$ 을 이용하여 각 그룹의 $weighted \chi^2$ 은 식 (7)으로 정의되며, 가장 큰 값을 가진 그룹의 클래스를 분류된 데이터에 할당한다.

$$weighted-\chi^2 = \sum \frac{\chi^2 \chi^2}{maximum-\chi^2} \quad (7)$$

4. 변압기 부하패턴의 주기성 분석 구현 및 시험

서울시의 강남구와 서초구에 전력을 공급하는 한국전력공사 강남지점의 2007년 1월부터 10월까지의 471개의 변압기 무선부하감시 시스템을 대상으로 시험하였다. 변압기 무선부하감시 시스템은 현장의 변압기에서 30분 단위로 상변 전류와 전압을 측정하고 무선통신망을 이용하여 서버 시스템으로 전송하는 시스템이다. 변압기 부하패턴의 주기성 분석을 위해 변압기 무선부하감시 시스템에서 데이터를 연계하여 그림 2와 같이 부하분석 데이터베이스를 재구성 하였다. बैंक 기본정보와 일시정보를 분리하여 연계한 무선부하데이터와 전력을 계산한 बैंक부하 테이블을 분석이 용이하도록 구성하였다.

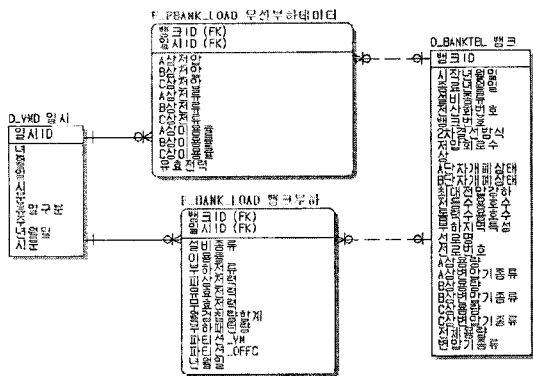


그림 2 변압기 부하 분석 대상 테이블 (ERD)

Fig. 2 Table for transformer load analysis (ERD)

변압기의 대표 패턴을 저압고객 특성 및 변압기 공간 정보 등을 고려한 시간 데이터마이닝 기법적용을 수행한다. 변압기의 하루 단위 부하패턴을 생성하기 위하여 무선부하 데이터에서 표 1의 (a)와 같이 बैंक ID와 일시 ID, 변압기 부하량 데이터를 추출한다.

표 1 기본 부하패턴 데이터

Table 1 Basic load pattern data

(a) 변압기 부하량 데이터

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

(b) 부하패턴 군집화 결과

속성명	데이터 타입	설명
BANK_ID	nominal	뱅크 고유 식별 코드
DATE_TIME_ID	continuous	일시 지정 코드
CLASS	nominal	부하 패턴 군집 번호

3D 큐브 마이닝에 변압기의 부하 패턴을 적용하기 위해서 추출된 데이터에서 부하 대표 패턴을 생성한다. 큐브 마이닝을 위해서 생성된 부하 대표 패턴을 बैंक ID, 일시 ID, 부하 패턴 속성으로 표 1의 (b)와 같이 군집화 결과 데이터를 구성한다. 생성된 기본 데이터의 बैंक 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 표 2와 같이 전환한다.

표 2 변압기 부하패턴 인덱스 형태 변환

Table 2 Index type conversion of transformer load pattern

BANK_ID	DATE_TIME_ID	CLASS		BANK_ID	DATE_TIME_ID	CLASS
22	36961	11	⇒	r1	h1	c11
24	36961	1		r2	h1	c1
25	36961	10		r3	h1	c10
27	36961	11		r4	h1	c11
28	36961	17		r5	h1	c17

시간의 변화에 따른 빈발한 बैंक_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 얻도록 한다. 알고리즘 수행 결과에서 기본 시간 간격에서의 패턴을 찾기 위해서는 캘린더 패턴에 대한 분류규칙 형식으로 전환한다. 변압기 부하 패턴 예측에 사용된 캘린더 스키마는 $CS = (G_n : D_n, G_{n-1} : D_{n-1}, \dots, G_1 : D_1) \Rightarrow CS = (Month : \{1,2,3,4\}, Week : \{1,2,3,4\}, Day : \{1,2,3,4,5,6,7\})$ 이다.

표 3 캘린더 패턴에 대한 의미 해석

Table 3 Interpretation about calendar pattern

속성	속성명	데이터 형식	데이터 해석	예제
MM	월	{1, 2, 3, 4}	각 달을 의미	{ 1 : 1월, 2 : 2월 ... }
EEK	주	{1, 2, 3, 4, 5, 6}	한달 중 주의 순서	{ 1 : 첫주, 2 : 둘째주, ..., 6 : 여섯번째주 }
DAY	요일	{1, 2, 3, ..., 7}	주 7일을 순서대로 매치	{ 1 : 월요일, 2 : 화요일, ..., 7 : 일요일 }

알고리즘 적용 후 일정한 캘린더 패턴을 추출하기 위해서 기본 표 4의 기본 시간 간격에서의 패턴을 그림 3과 같이 갱신한다.

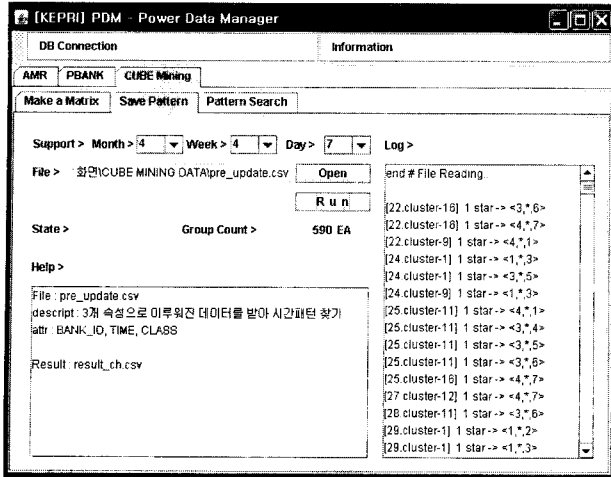


그림 3 기본 시간패턴 갱신 화면
 Fig. 3 Screen for basic temporal pattern update

기본 시간 간격 갱신은 3가지의 각 속성의 개수가 사용자에 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 캘린더 패턴으로 변압기의 향후 발생하는 부하 패턴을 예측한다.

표 4 기본 시간 간격과 캘린더 패턴
 Table 4 Basic time interval and calendar pattern
 (a) 기본 시간 간격

BANK_ID	CLASS	MM	WEEK	DAY
22	cluster-1	1	1	1
22	cluster-1	2	4	1
22	cluster-10	1	1	6
22	cluster-10	1	1	7
22	cluster-10	1	2	7

(b) 기본 시간 간격 갱신 후 캘린더 패턴

BANK_ID	CLASS	MM	WEEK	DAY
22	cluster-16	3	*	6
22	cluster-16	4	*	7
22	cluster-9	4	*	1
24	cluster-1	1	*	3
24	cluster-1	3	*	5
24	cluster-9	1	*	3
25	cluster-11	3	*	4
25	cluster-11	3	*	5
25	cluster-11	3	*	6
22	cluster-10	1	2	7

그림 4, 5는 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' cluster 9번의 대표 패턴을 갖는다는 의미이다. 또한 이러한 시간 패턴을 가지는 22번 은행의 특징은 '전산화 번호가 0123E41이고 전

동수용호수가 26, 동력수용호수 2, 공간적 특성으로 변화가 인 부하지역특징 변화가, 선로명이 도산지'인 특징을 가진다.

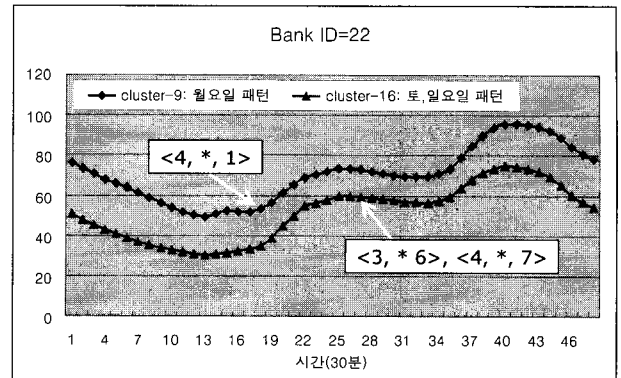


그림 4 변압기 캘린더 부하패턴 예시
 Fig. 4 Example of transformer calendar load pattern

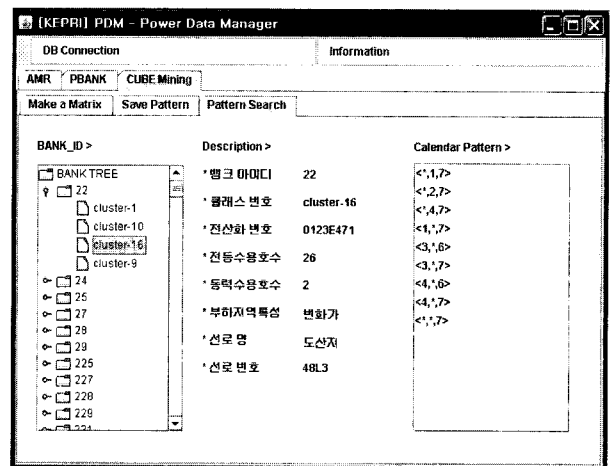


그림 5 변압기 부하패턴의 주기성 분석 화면
 Fig. 5 Frequency analysis screen of transformer load pattern

5. 결론

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. 뿐만이 아니라 계약들에 대한 평가, 그리고 전력 시장에서 가격 변동에 의한 다양한 복잡한 상품들에 대한 평가를 함에 있어서도 중요하다.

이 논문에서는 시간 데이터마이닝 기법을 이용하여 변압기 부하패턴을 분석하였다. 시간 속성을 가진 변압기의 부하 패턴에 대해 시간의 변화에 따른 적용 시점이 명확한 지식 탐사가 가능하고, 부하예측에 있어 탐사된 규칙과 시간 지식을 이용함으로써 기존의 정적인 분류규칙을 적용한 방법보다 더 정확한 예측을 할 수 있는 새로운 시간 패턴 마이닝 기법을 제안하였다. 이를 위해 캘린더 스키마 및 캘린더 패턴의 시간 마이닝의 문제를 정의하고, 기본 시간 간격에서의

3차원(시간 - 변압기 - 부하패턴) 패턴 마이닝 수행을 위한 3D 큐브 마이닝을 적용하고, 3차원 패턴에 대한 켈린더 기반의 시간 패턴으로 갱신 기법과 주기성 표현 알고리즘을 개발하였다. 향후에는 날씨민감도와 고객변동을 고려한 변압기 부하패턴 예측 연구를 수행하여 데이터의 신뢰성과 실용성을 개선할 예정이다.

감사의 글

본 연구는 지식경제부 전력산업기반기금의 지원으로 수행되었습니다.

참 고 문 헌

[1] S. J. Huang, K. Shih, "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] G. Chicco, R. Napoli, P. Postulache, M. Scutariu, C. Toader, "Customer characterization options for improving the tariff offer", IEEE Trans. Power System, Vol.18, pp.381-387, 2003.

[3] B. Pitt, D. Kirchen, "Applications of data mining techniques to load profiling", In Proc. IEEE PICA, pp.131-136, 1999.

[4] G. Chicco, R. Napoli, P. Postulache, M. Scutariu, C. Toader, "Load Pattern-Based Classification of Electricity Customers", IEEE Trans. Power System, Vol.19, pp.1232-1239, 2004.

[5] R. F. Chang, C. N. Lu, "Load profiling and its applications in power market", IEEE Power Engineering Society General Meeting, Vol.2, pp.974-978, 2003.

[6] N. Amjady, "Short-term hourly load forecasting using time-series modeling with peak load capability", IEEE Trans. Power System. Vol.16, No.3, pp.498-505, 2001.

[7] B. Liu, W. Hsu, Y. Ma, "Integrating classification and association rule mining", In Proc. of the 4th Int'l Conference Knowledge Discovery and Data Mining, 1998.

[8] Y. Li, P. Ning, "Discovering calendar-based temporal association rules", In Proc. of the 8th Int'l Symposium on Temporal Representation and Reasoning 2003.

[9] K. Verma, O. P. Vyas, "Efficient calendar based temporal association rule", SIGMOD Rec., Vol.34:3, pp.63-70, 2005.

[10] W. Li, J. Han, J. Pei, "CMAR: Accurate and Efficient Classification Based on Multiple Association Rules", In Proc. of 2001 Int'l Conference on Data Mining, 2001.

[11] www.csc.liv.ac.uk/~frans/KDD/Software/CMAR/cmar.html

저 자 소 개



신진호 (辛 鎭 浩)

1970년 7월 26일생
 1996년 한밭대 전자계산학과 졸업
 2004년 충북대 전자계산학과 졸업(석사)
 현재 한전 전력연구원 근무
 E-mail : jinho@kepco.co.kr



이봉재 (李 俸 在)

1959년 1월 13일생
 1982년 성균관대 수학과 졸업
 1985년 한국전력공사 입사
 현재 한전 전력연구원 근무
 E-mail : leeboja@kepco.co.kr



김영일 (金 榮 一)

1972년 11월 27일생
 1998년 충남대 컴퓨터공학과 졸업
 2000년 충남대 컴퓨터공학과 졸업(석사)
 현재 한전 전력연구원 근무
 E-mail : yikim@kepco.co.kr



이헌규 (李 憲 揆)

1975년 11월 7일생
 2002년 경대대 전자계산학과 졸업
 2004년 충북대 전자계산학과 졸업(석사)
 현재 충북대 전자계산학과 박사과정
 E-mail : hglee@dblab.chungbuk.ac.kr



류근호 (柳 根 鎬)

1953년 11월 11일생
 1976년 숭실대 전산학과 졸업
 1980년 연세대 전산전공(공학석사)
 1988년 연세대 전산전공(공학박사)
 현재 충북대학교 전기전자 컴퓨터공학부 교수
 E-mail : khryu@dblab.chungbuk.ac.kr