

GIS-AMR 부하 분석 모델에서의 전력 부하 예측을 위한 시공간 패턴 마이닝¹⁾

이헌규*, 박명호**, 박진형**, 신진호***, 류근호**

*한국전자통신연구원 우정물류기술연구부

**충북대학교 데이터베이스 연구실

***한전 전력연구원 전력정보기술그룹

e-mail:hg_lee@etri.re.kr*, khryu@dblab.chungbuk.ac.kr**

Spatio-temporal Pattern Mining for Power Load Forecasting in GIS-AMR Load Analysis Model

Heon Gyu Lee*, Minghao Piao**, Jin Hyoung Park**, Jin-ho Shin***, and Keun Ho Ryu**

*Postal Technology Research Center, Electronics and Telecommunications Research Institute

**Database/Bioinformatics Lab., Chungbuk National University

***Power Information Technology Group, Electric Power Research Institute

요 약

변압기 무선부하감시 시스템에서 30분 간격으로 계측된 부하 데이터와 GIS-AMR 데이터웨어하우스로부터 변압기 속성 및 공간적 특징을 추출하여 정확한 변압기의 부하 패턴을 예측하기 위한 시공간 패턴 마이닝 기법을 적용하였다.

1. 서론

전력 부하 예측을 위한 정확한 분석 모델은 전력 운영과 계획에 필수적이며 전력 산업에 있어서의 부하 예측은 전력의 구입, 생산, 로드 스위칭, 그리고 기반 시설의 업데이트 등에 대한 의사 결정을 함에 있어서 도움을 준다. 특히, 부하 예측은 공급자, ISO 회사 및 재무기관이 전기에너지 생산, 전송, 분배의 관련 마케팅에 있어서 매우 중요하다. 따라서 전력산업에 있어서 효율적인 운영과 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, 데이터마이닝 기술은 부하 데이터로부터 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[1],[2].

이 논문에서는 변압기의 시간대별과 위치 정보를 고려

한 부하 패턴을 예측하기 위해서 변압기 무선부하감시 시스템으로부터 계측된 30분 시간 간격의 부하량과 기구측된 GIS-AMR[3] 데이터웨어하우스로부터 변압기의 속성 및 공간 정보를 추출하여 시공간 패턴 마이닝을 적용한다. 즉, 무선부하감시시스템에서 30분 단위로 계측된 부하데이터를 이용하여 변압기의 대표 부하패턴을 생성하고 계측되지 않은 변압기에 대해 부하를 예측을 수행하며, 상세 내용은 다음과 같다.

- 무선부하감시 시스템에서 30분 간격으로 계측된 부하량 데이터로 일별 변압기 부하패턴 생성
- GIS-AMR²⁾ 데이터웨어하우스로부터 변압기 속성 및 공간 속성 추출
- 일별 부하패턴의 군집화를 통한 대표 부하패턴 생성
- 주기적 시간 표현 및 공간 특징을 고려한 시공간 패턴 마이닝 수행을 통한 분류 모델 생성
- 미 계측된 변압기의 부하패턴 예측

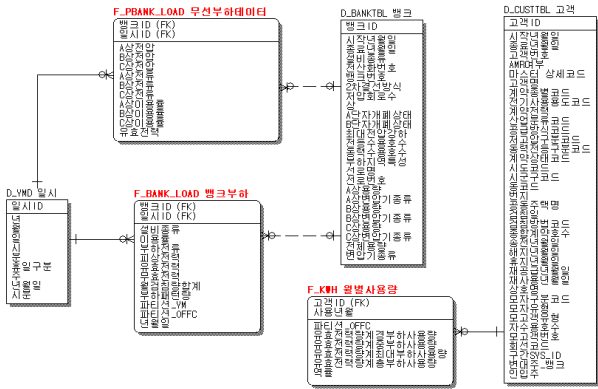
2. 변압기 부하 및 시공간 정보 추출

무선 부하 측정 뱅크에 대한 30분 단위의 부하 패턴을 생성하기 위해서 무선부하데이터(F_PBANK_LOAD) 테이블에서 유효전력과 일시 ID 속성을 추출한다. 무선 부하감시 시스템에 의해 30분 간격으로 측정된 뱅크 유효 전력의 시간 정보를 추출하기 위하여 일시(D_YMD) 테이블

1) 이 논문은 2009년도 정부(과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(R01-2007-000-10926-0).

2) GIS : 지리정보시스템 (Geographic Information System), AMR : 원격검침 (Automatic Meter Reading)

의 시간 속성을 이용한다. 그 다음, GIS-AMR 데이터웨어 하우스의 변압기 정보(D_BANKTBL) 테이블에서 변압기 구분을 위한 변압기 ID와 부하지역특성, 전체 용량 속성과 고객 정보(D_CUTTBL) 테이블에서 유효전력량계총부하사용량을 추출한다.



(그림 1) 변압기(BANK) 부하 분석 대상 테이블 (ERD)

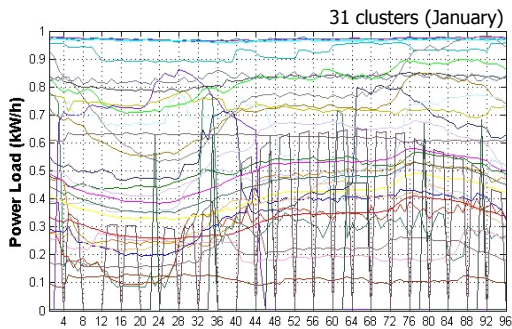
3. 변압기 대표 부하 프로파일의 생성

추출된 전기 사용량을 변압기별(bank), 일별 구분하여 벡터 표현으로 아래 식과 같이 재구성 하였다.

$$V^{(b)} = \{ V_0^{(b)}, \dots, V_h^{(b)}, \dots, V_H^{(b)} \} \quad (식 1)$$

(단, b=변압기, 0<h<2330, H=2330)

벡터로 표현된 모든 부하 패턴으로부터 K-means 알고리즘을 이용하여 [4]와 같은 방법으로 변압기의 일별 대표 부하패턴 생성을 위한 군집화를 수행한다.



(그림 2) 변압기 대표부하패턴 생성 예

이렇게 생성된 모든 변압기 대표부하패턴은 분류 모델 생성할 때의 클래스 라벨로 사용된다.

4. 시공간 패턴 마이닝

이 절에서는 변압기 부하패턴의 시간 및 공간적 특징을 추가한 분류 모델 생성 과정을 기술한다.

4.1. 분류 모델 생성을 위한 특징 추출

변압기의 부하패턴은 30분 간격의 측정치로 32 시간점에 대한 부하량이며, 이를 이용한 부하예측모델 생성은 너무 많은 특징 값들로 인해 성능 저하를 야기 시킨다. 더욱

이 시간 및 공간 특징들이 추가 된다면, 부하예측이 불가능할 정도로 많은 입력 벡터를 사용하게 된다. 따라서 이 논문에서는 특정 부하패턴을 대표하는 부하요소(load factor)를 추출하여 차원축소를 수행한다. 부하요소들은 전력계통의 전문가들에 의해 정의되며, 부하패턴에서의 주요한 특징들이다. 부하요소들은 다음 식들에 의해 정의되며, (그림 3)과 같은 영역을 포함한다.

Early Morning (0h~7h):

$$LF_1 = \frac{7 P_{avg. of early mor.}}{24 P_{avg of day}} \quad (식 2)$$

Morning (9h~12h):

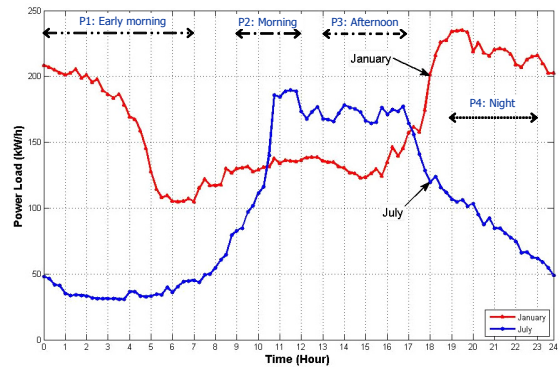
$$LF_2 = \frac{1 P_{avg. of morning}}{8 P_{avg of day}} \quad (식 3)$$

Afternoon (13h~17h):

$$LF_3 = \frac{1 P_{avg. of afternoon}}{6 P_{avg of day}} \quad (식 4)$$

Night (19h~23h):

$$LF_4 = \frac{1 P_{avg. of night}}{6 P_{avg of day}} \quad (식 5)$$



(그림 3) 부하요소에 대한 시간 영역

시간 및 공간적 특징과 추출된 모든 입력 특징은 <표 1>과 같고 연속형 실수값인 부하요소는 시공간 패턴 마이닝 적용을 위해 엔트로피 기반 이산화[5]를 수행하여 명목형(nominal)으로 변환한다. (그림 4)는 부하분석모델 생성을 위한 훈련데이터 집합의 예이다. 또한 공간적 특징 추출을 위해서 변압기 위치(location_id)로부터 부하지역 특징, 전등수용호수, 동력수용호수 등의 정보를 추출한다.

<표 1> 분류 모델 생성을 위한 특징 (속성)

속성	설명	
Bank_ID	변압기 ID	
Time_Pattern	시간 (월, 주, 일)	
Location_ID	변압기 위치 ID	
LF	LF ₁	부하요소 1
	LF ₂	부하요소 2
	LF ₃	부하요소 3
	LF ₄	부하요소 4
CLASS	대표부하패턴 (클래스 라벨)	

변압기 ID	기본시간패턴	변압기위치_ID	LF ₁	LF ₂	LF ₃	LF ₄	대표부하패턴 (클래스)
1	<1,1,1>	3125	a ₁	b ₃	c ₁	d ₃	Cluster3
2	<1,1,2>	3145	a ₄	b ₁₂	c ₂	d ₂	Cluster18
...
210	<4,4,6>	3547	a ₁₂	b ₅	c ₆	d ₆	Cluster3

변압기 위치값을 이용한 공간적 특징 추출

부하지역특성	선로명	전등수용호수	동력수용호수
변화가	3125	26	2
변화가	3145	37	4
...
210	3547	11	3

(그림 4) <표 1>에 대한 입력 데이터 예

4.2. 부하패턴의 주기 패턴 발견을 위한 시간 패턴식

<표 1>의 속성들에 해당하는 데이터로부터 시간의 주기적 표현이 가능한 시간 패턴식을 적용한다.

시간의 패턴 표현식³⁾은 시간 개념상의 단위의 집합으로 정의되며, 사용자 기반의 표현식이다. 예를들어 시간 단위 집합을 <월,주,일>로 가정하면, 시간 패턴 <1,1,1>의 의미는 "1월 첫째주 월요일"이며, <2,2,4>는 "2월 둘째주 목요일"을 의미한다. 또한 주기적 패턴(cyclic pattern)을 표현하기 위해서 문자 "*"를 정의하며, "*"의 의미는 해당 시간 단위의 모든 도메인값의 포함을 나타낸다. 예를들어 <1*,3>의 표현은 "1월 매주 수요일"을 의미하며, <*,*,1>는 "매월 매주 월요일"을 의미한다. 특히, "*"를 전혀 포함하지 않는 시간 패턴은 기본시간패턴이라고 정의한다.

4.3. 시공간 속성을 고려한 분류모델의 생성

(그림 4)의 훈련데이터 집합으로부터 분류모델 생성을 위한 규칙들을 추출한다. 규칙의 형태 R은 다음과 같다.

$$R: p_1 \wedge p_2 \wedge \dots \wedge p_n \Rightarrow C \quad (\text{식 6})$$

$p_1 \wedge p_2 \wedge \dots \wedge p_n$ 는 (그림 4)의 속성들에 대한 속성값이고 C는 클래스 라벨, 즉 대표부하패턴(cluster)이다.

전체 알고리즘 수행 단계는 다음과 같다.

- ① 모든 각각의 기본시간패턴에 대해서 규칙의 지지도 (support)를 만족하는 모든 패턴들을 추출한다.

$$support_i = \frac{Rule\ Count}{|D_i|} \quad (\text{식 7})$$

Rule Count는 해당 기본시간패턴 i에 포함된 데이터집합에서 $p_1 \wedge p_2 \wedge \dots \wedge p_n$ 와 해당 클래스, C가 동시에 발생하는 빈도를 의미한다.

- ② 추출된 모든 패턴으로부터 규칙의 신뢰도 만족시키는 정확성 높은 규칙들을 추출하며, 모든 규칙은 신뢰도를 기반으로 내림차순 정렬된다.

$$confidence_i = \frac{Rule\ Count}{|p_1 \wedge \dots \wedge p_n|} \quad (\text{식 8})$$

3) 달력 표현방식의 캘린더 패턴식이며, 자세한 정의 및 알고리즘은 [6] 논문 참조

- ③ 추출된 모든 규칙은 기본시간패턴을 가지므로 주기성⁴⁾ 발견을 위해서 "*"를 포함한 시간패턴으로 갱신한다. 갱신 알고리즘은 (그림 5)이다.

```

input Time pattern "*", a set of rules with
basic time pattern
output a set of rules with time pattern "*"

1  if time pattern "*" is first time updated then
2    Rk with "*" = Rk with basic time pattern;
3    update count = 1;
4  else
5    for each R ∈ Rk with basic time pattern - Rk
      with "*"
6      Rk.update count = 1;
7    end
8    for each R ∈ Rk with basic time pattern ∩
      Rk with "*"
9      R.update count++;
10   end
11   Rk = {R.update}
12  end if
    
```

(그림 5) 기본시간패턴의 갱신 알고리즘

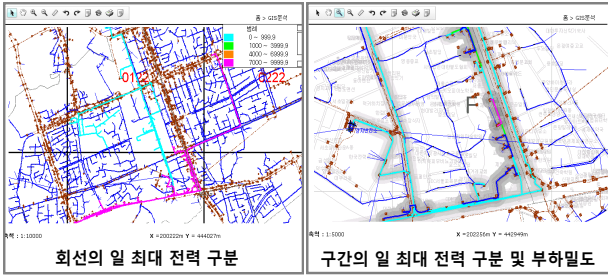
각 i 번째 기본시간패턴에 대한 규칙들은 각각의 해당되는 카운터를 가지고 있다. 만약 갱신 과정이 처음 이루어지는 단계(1~4라인)이거나 기본시간단위에서의 규칙이 처음 갱신 되어질 경우(5~7라인), 카운터를 1로 설정한다. 그렇지 않은 경우(8~10라인), 현재의 카운터를 1씩 증가시킨다. 해당 시간 단위에서의 모든 도메인값을 가진다면, 그 시간 단위의 패턴 표현은 "*"이 되고 주기성을 의미한다. 예를들어 어떤 규칙 R에 대해서 시간 단위 집합을 <월,주,일>로 지정했을 경우에, R: <2*,1>의 의미는 규칙 R이 기본시간패턴 {<2,1,1>, <2,2,1>, <2,3,1>, <2,4,1>}에서 모두 발생했다는 의미이다.

모든 시간 패턴 및 분류 규칙이 생성된 후에 대표부하패턴을 할당할 테스트 데이터셋에 대해서 먼저 테스트 데이터의 기본시간과 인스턴스값을 포함하는 모든 시간패턴을 가지는 규칙들이 선택되어지고 이러한 여러 규칙 중 가장 높은 신뢰도를 가지는 규칙의 클래스(대표부하패턴)를 할당한다.

5. 실험 및 성능 평가

제안된 시공간 패턴 마이닝의 실험 및 평가를 위해서 한국 전력 연구원에서 구축한 배전고압계통 GIS-AMR 부하분석 데이터웨어하우스에서 무선부하감시 시스템에 의해 수집된 부하 데이터를 적용하였다. 실험에 사용된 데이터는 수도권역 2,923개의 2007년 1월부터 4월까지의 데이터가 사용되었다.

4) 하나 이상의 "*"를 포함하도록 기본시간패턴을 갱신



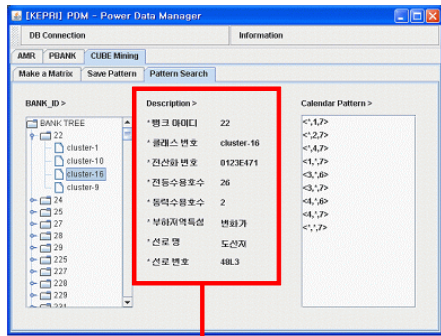
(그림 6) 한전 전력연구원 배전계통 GIS-AMR의 회선/구간/변압기 부하 현황과 부하 밀도[3]

<표 2>는 K-means 알고리즘을 이용한 4개월에 해당하는 변압기에 대한 대표부하패턴 생성 결과이다.

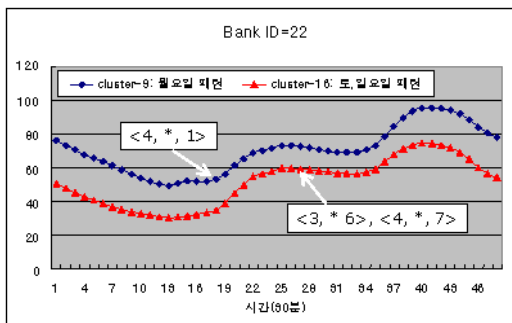
<표 2> 변압기 대표 패턴 생성 결과

월	대표부하패턴 수(# of cluster)
1	31
2	43
3	26
4	31
5	30

“*” 표현의 주기성을 갖는 시간 패턴 표현의 적용의 예는 (그림 7)이다. 그림에서 위의 ‘●’ 그래프는 9번 클러스터로 예측된 월요일 부하패턴이며 시간패턴 <4,* , 1>을 가지므로 그 해석은 “4월의 매주 월요일” 의미이다. ‘▲’ 그래프는 주말(토, 일요일)에 해당되는 부하패턴으로 두 가지, “3월 매주 토요일과 4월 매주 일요일 (<3,* 6>, <4,* ,7>)” 시간패턴을 가진다.



변압기 정보 표현과 시간패턴의 예



(그림 7) 시간패턴 분석결과 및 부하패턴 예측의 예

<표 3>은 기존의 정적인 분류모델과 시공간 패턴을 이용한 부하분석모델의 테스트 집합에 대한 결과이다. 실험의 2,923건의 데이터에 대해 10-교차검증 (CV-10) 방법을 이용하였다. 분류 결과 시공간 패턴 마이닝 적용의 결과 평균 168, 정적인 부하분석모델의 경우는 443.9개의 잘못된 대표부하패턴이 할당되었다.

<표 3> 변압기 부하분석모델의 정확도 평가

테스트 집합 (CV-10)	시공간 패턴 부하분석모델	정적 부하분석모델
	테스트 정확도	테스트 정확도
1	2806 / 2923	2484 / 2923
2	2807 / 2923	2488 / 2923
3	2800 / 2923	2806 / 2923
4	2807 / 2923	2501 / 2923
5	2625 / 2923	2491 / 2923
6	2776 / 2923	2463 / 2923
7	2683 / 2923	2464 / 2923
8	2655 / 2923	2463 / 2923
9	2815 / 2923	2484 / 2923
10	2776 / 2923	2483 / 2923
평균 정확도	94.3%	84.8%

6. 결론

이 논문에서는 30분 간격으로 예측되는 변압기의 정확한 부하패턴 예측을 위해서 시공간 속성을 고려한 부하분석모델 생성을 목적으로 시공간 패턴 마이닝 기법을 적용하였다. 제안된 시공간 패턴 마이닝의 적용결과 기존의 정적인 부하분석모델에 비해 약 10% 정도 향상된 예측 결과를 보였다.

참고문헌

- [1] Figueiredo V., et al., "An Electric Energy Consumer Characterization Framework Based on Data Mining Techniques," IEEE Trans. on Power Syst., Vol.20, No.2, pp.596-602, 2005.
- [2] Chicco G., et al., "Load Pattern-Based Classification of Electricity Customers," IEEE Trans. Power Syst., Vol.19, pp.1232-1239, 2004.
- [3] 한전 전력연구원, "인터넷 GIS 환경의 AMR 시스템 연계 모델 개발," R-2005-0-093, 2008.
- [4] 박진형, 이현규, 신진호, 류근호, 김희석, "기대치-최대화 군집 알고리즘과 출현 패턴 마이닝을 이용한 전력 소비 패턴 분석," 한국정보처리학회 춘계학술발표대회, 제15권, 제2호, pp.261-264, 2008.
- [5] Kohavi R., et al, "MLC++: A Machine Learning Library in C++," Tools with Artificial Intelligence, pp.740-743, 1994.
- [6] 이현규, 노기용, 서성보, 류근호, "캘린더 패턴 기반의 시간 연관적 분류 기법," 한국정보과학회 논문지 제32권, 제6호, 2005.