

데이터 마이닝을 이용한 양방향 전력거래상의 단기수요예측기법

김 형 중*, 이종수**, 신 명 철**, 최 상 열***
(주)오산에너지* 성균관대학교** 인덕대학***

Short-term demand forecasting method at both direction power exchange
which uses a data mining

Hyoung-Joong Kim*, Jong-Soo Lee**, Myong-Chul Shin**, Sang-Yeoul Choi***
Osan Energy Co. Ltd.*, Sungkyunkwan University**, Induk Institute of Technology***

Abstract - Demand estimates in electric power systems have traditionally consisted of time-series analyses over long time periods. The resulting database consisted of huge amounts of data that were then analyzed to create the various coefficients used to forecast power demand. In this research, we take advantage of universally used analysis techniques analysis, but we also use easily available data-mining techniques to analyze patterns of days and special days(holidays, etc.). We then present a new method for estimating and forecasting power flow using decision tree analysis. And because analyzing the relationship between the estimate and power system ceiling prices currently set by the Korea Power Exchange. We included power system ceiling prices in our estimate coefficients and estimate method.

1. 서 론

현재 진행되고 있는 전력구조개편에 따르면 이후 많은 민영 발전사업자들이 양방향 입찰 방식에 의해 전력거래가 이루어질 것이라 판단되어진다. 현재 한전에서 분리되어진 5개 발전회사에서는 이미 수요예측을 하고 있지만 소규모 발전사업자들은 아직 수요예측을 하지 못하고 있는 것이 현실이므로 좀 더 쉽게 전력부하를 예측하여 입찰의사 결정을 할 수 있는 과정이 필요하고 앞으로 많은 수요가 있을 것으로 보인다. 분산전원의 개발로 중소 규모 발전사업자들의 입찰과 이후 배전사업의 민영화이 이루어지면 배전사업자 역시 자신의 사업 범위에서의 수요를 예측하여 입찰을 해야 하므로 좀 더 쉬운 과정이 필요하다고 판단되어지기 때문에 비전문가도 수요를 예측할 수 있는 방법이 필요하다.

이전까지 전력수요예측에 사용되고 있는 방법은 안정적이고 범용적으로 사용되고 있는 시계열 방법중 시계열을 차분해서 자기회기이동평균모형(ARMA : Autoregressive Moving Average)이 되는 모형인ARIMA(Autoregressive Integrated Moving Average)[1]를 적용하고 있고 현재 KPX(Korea Power Exchange)에서 사용되고 있는 방식은 시계열의 계절성에 중점을 두고 계절마다의 패턴을 분석하여 예측하는 KULF(KPX-SNU Load Forecaster)를 사용하고 있다[2]. 현재 연구되어진 지수평활화 모형(Exponential Smoothing)은 단기예측에는 안정적이고 비전문가도 충분히 사용할 수 있는 기법이지만 기온과 같은 특수하고 불규칙한 변수가 있을 때는 오차 범위가 커지는 한계가 있고, 신경 회로망(Neural Network), 지식기반모형(Knowledge

based)[3]은 상당히 복잡한 과정을 통해서 예측하므로 비전문가가 구현하기는 한계가 있는 단점이 있어 본 연구에서는 데이터베이스의 기술의 발전과 Internet의 발달로 이제 방대한 양의 데이터 수집과 관리가 용이해졌기에 데이터 마이닝의 연관기법과 의사결정트리를 이용한 예측방법과 양방향전력거래에서 생성되는 계통한계 가격과 기온 그리고 전력량을 예측요소로 하여 수요예측 방법을 제시하고자 한다.

2. 본 론

2.1 데이터 마이닝 기법

데이터마이닝이란 과연 무엇일까? 커다란 데이터베이스를 기반으로 많은 양의 데이터 중 의사결정이나 실험에 필요한 데이터를 찾아내고 그 데이터를 여러 알고리즘을 통해 필요한 지식을 얻어내는 기법을 말한다.

2.1.1 데이터마이닝 지식발견절차

데이터마이닝 지식발견절차는 다음과 같은 과정들이 원하는 지식을 획득하기까지 반복되어진다.[4]

- (1) 데이터 정제(Data Cleaning) : 불필요하거나 일치하지 않는 데이터의 제거
- (2) 데이터 통합(Data Integration) : 다수의 데이터 소스들의 결합
- (3) 데이터 선택(Data Selection) : 필요한 데이터들을 데이터베이스로부터 검색한다.
- (4) 데이터 변환(Data Transformation) : 요약이나 집계 등과 같은 연산을 수행함으로써, 마이닝을 위해 적합한 형태로 데이터를 가공처리 한다..
- (5) 데이터 마이닝(Data Mining) : 데이터 패턴을 추출하기 위해 지능적 방법들이 적용되는 필수 과정
- (6) 패턴 평가(Data Evaluation) : 몇 가지 흥미로운 척도를 기초로 지식이 나타나는 패턴을 구별한다.
- (7) 지식 표현(Data Presentation) : 사용자에게 발견된 지식을 보여주기 위해 시각화와 지식 표현기법을 사용한다.

2.1.2 연관규칙

연관규칙 마이닝이란? 광범위한 데이터 중 필요한 데이터를 찾아내고 그 데이터들이 가지고 있는 공통된 특성이나 연관되어지는 관련성을 찾아내는 것을 말한다.[4] 다시 말해 집합에 교집합으로 표현되어질 수 있다. 대규모 데이터베이스에서 연관규칙을 발견하는 방법은 다음의 두 단계의 절차로 표현 할 수 있다.

- (1) 모든 빈발 항목집합 찾기 : 정의에 따라 모든 항목 집합은 빈번하게 발생하게 된다.

(2) 빈발 항목집합에서 강한 연관규칙 발견 : 의도에 따라 이 규칙들은 조건에 만족해야 한다[4]

2.1.3 의사결정트리(Decision tree)

의사결정트리란? 흐름도(Flow chart)와 유사한 트리 구조로 중간노드(Intermediate node)에는 속성에 대한 검사를 표시하고, 가지는 검사의 결과를 나타내며 잎(Leaf 또는 단말)노드는 클래스나 클래스의 분포를 나타낸다.

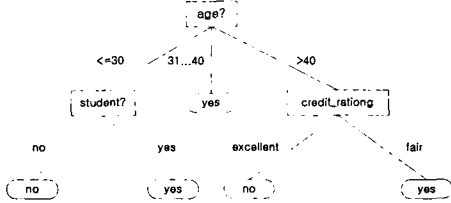


그림1. 의사결정 트리의 예

위의 예는 AllElectronics에서 고객이 컴퓨터를 구입할 것인지 아닌지를 예측한 의사결정 트리로 중간노드는 사각형, 잎노드는 타원형으로 나타났다.[4]

2.2 데이터 마이닝의 연관기법 적용

2.2.1 일별 부하 패턴 비교

우리나라 수요 패턴을 크게 분리하자면 평일과 토요일, 휴일, 특수일로 크게 구분할 수 있다. 먼저 한주간의 전력수요패턴을 비교하여 데이터를 분류하고자 한다. 따라서 2001년 1월 셋째주의 데이터를 가지고 한주동안의 수요를 비교하면 그림2와 같다.

2001년 1월 셋째주 전력 생산량 비교

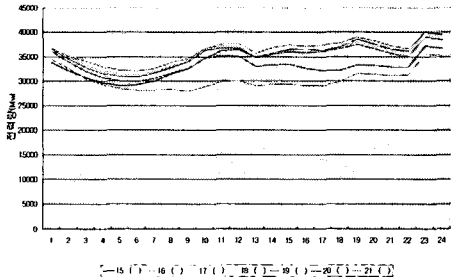


그림2. 2001년 1월 셋째주 전력 생산량 비교

그림2.에서 보듯이 평일의 수요패턴은 거의 흡사하지만 토요일과 일요일의 수요패턴이 다르고 토요일과 일요일의 수요 패턴도 다른 것을 볼 수 있다. 따라서 평일의 수요예측에 관한 데이터를 정제하거나 통합, 선택 할 때는 평일, 토요일, 일요일로 분류하는 것이 필요하다.

2.2.2 계절별 수요패턴 비교

우리나라는 4계절의 기후를 가지고 있기 때문에 4계절마다의 부하패턴이 다르다. 따라서 4계절을 나누는 기준은 3월~5월까지를 봄, 6월~8월까지를 여름, 9월~11월까지 가을, 12월~다음해 2월까지를 겨울로 보는 것이 통상적이다. 그러나 본 논문에서는 여름을 6월~9월까지 그리고 가을을 10월~11월까지 분류하고자 하는데 그 이유는 9월의 부하패턴을 분석하면 가을의 패턴보다는 여름패턴에 흡사하기 때문이다.

그림3은 4월, 7월, 10월, 12월로 계절을 대표하여 분석한 것으로 00시부터 08까지의 패턴은 비슷하지만 08시 이후 그리고 23시부터 00시까지는 계절별로 다른

수요패턴을 볼 수 있는데 이중 4월과 10월 즉 봄과 가을의 수요패턴은 생산량의 차이가 있을 뿐 수요패턴은

2001년 계절별 전력 생산량 비교

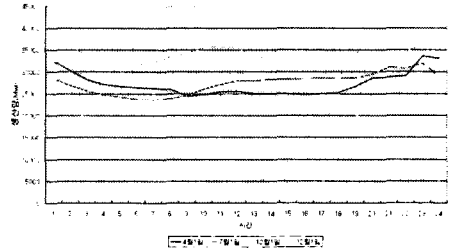


그림3. 2001년 계절별 전력 생산량 비교

흡사한 것을 볼 수 있다. 따라서 봄과 가을을 하나로 구분할 수 있음을 알 수 있다.

2.2.3 주말 및 휴일(특수일) 수요패턴 비교

앞에서 2.3.1에서 논한 것 같이 토요일과 일요일의 패턴이 다르다는 것을 알 수 있다. 따라서 토요일과 일요일의 수요패턴의 비교는 생략하고 일요일과 특수일 그리고 설·추석연휴의 수요패턴을 비교하고자 한다.

2001년도 특수일 전력 생산량 비교

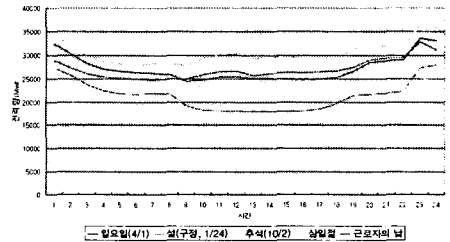


그림4. 2001년 특수일 전력 생산량 비교

그림4를 분석해보면 일요일과 일반 공휴일의 패턴은 흡사하고 설과 추석의 패턴이 비슷한 것 볼 수 있다. 단 계절에 따른 변화는 무시를 했다. 따라서 휴일의 패턴은 일요일과 휴일 그리고 설과 추석 즉 특수일로 분류할 수 있다.

2.2.4 기온에 따른 수요변동 패턴 비교

수요예측에서 가장 큰 영향을 주는 것은 바로 기후이다. 그중 기온의 변화에 따라 부하가 변동이 심하다는 것은 쉽게 알 수 있다. 물론 그날의 날씨가 흐린지, 비가 내리는지, 맑은 날씨등인지에 따라서도 영향을 줄 수 있지만 여기서 날씨를 맑다고 가정한다.

특히 4계절이 뚜렷한 우리나라에서는 기온의 영향을 많이 받게 되므로 날씨보다는 기온의 요소가 좀 더 중요하게 여겨지는데, 이는 기준에 많은 논문을 보더라도 알 수 있다.

2.2.5 가격과 전력량의 관계 비교

현재 우리나라는 전력구조개편을 진행하고 있는 상태이므로 가격과 전력량의 관계도 무시할 수 없는 요소가 되겠지만 아직 완전한 전력거래가 이루어지지 않고 있어서 시장경제원리에 의한 가격형성이 아닌 전력거래소에서 시간당 계통한계가격을 부가하는 방식의 비정상적인 방법을 사용하고 있는 과도기라 할 수 있다.

전력거래상에서의 가격형성은 발전회사에서 이익과 밀접한 관계를 갖게 되므로 가격도 충분히 예측 요소로서 중요하게 인식될 것이다.

2.2.6 데이터 마이닝의 의사결정트리 적용

그림5와 같이 데이터 마이닝을 하는 과정에서는 의사결정트리를 이용하고 그 이후 연관관계에서는 평활화법을 이용하여 수요치를 예측하고자 한다.

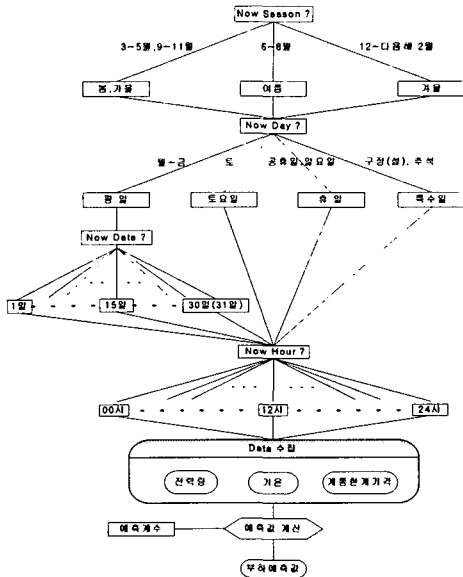


그림5. 의사결정트리를 이용한 수요예측 Flow chart 각 단계를 설명하면,

- ① Now Season : 계절에 대한 연관분석을 통해 봄·가을(3월~5월, 9월~11월), 여름(6월~8월), 겨울(12~다음해 2월)의 3부분으로 분류하였고 해당되는 결과를 선택한다.
- ② Now Day : 평일, 토요일, 휴일, 특수일의 4부분으로 구분하고 해당되는 결과를 선택한다. 단 평일은 월요일~금요일, 휴일은 일요일과 특수일을 제외한 법정휴일, 그리고 특수일은 설(구정), 추석연휴로 구분하였다.
- ③ Date : 평일의 경우 1월~30일(31일)중 해당되는 날짜를 선택한다. 그 외의 휴일과 특수일의 경우 날짜보다는 Day에 따라 형성되므로 적용을 생략한다.
- ④ Now Hour : 각 시간으로 현재 시간을 선택한다.
- ⑤ Data 수집 : 전력량과 기온, 계통한계가격의 요소에 해당되는 Data를 수집한다.
- ⑥ 예측값 계산 : 수집된 Data를 가공하고 시계열의 평활법과 예측계수를 이용하여 연산한다.
- ⑦ 예측계수 : 최근 3년간의 평균 전력수요 증감비를 적용한다.
- ⑧ 예측값 : 예측값 계산을 통한 단기수요전력 예측값을 출력한다.

2.3 사례적용

2001년 ~ 2004년까지의 3년간의 우리나라 발전량을 이용하여 2004년 1월의 평일에 대한 전력수요를 예측해 보았다.

2.3.1 2004년 1월 14일(평일) 수요예측

시간	수요치	마이닝 기법		KPX	
		예측치	오차(%)	예측치	오차(%)
1	44638	41091	7.9	41414	7.2
2	42823	35732	16.6	39609	7.5
3	41734	36526	12.5	38224	8.4
4	40292	36241	10.1	36943	8.3
5	38364	35299	10.4	36215	8.0
6	39200	36490	6.9	36090	7.9
7	39865	36800	7.7	36803	7.7
8	41031	37920	7.6	37983	7.4
9	42099	41473	1.5	39302	6.6
10	44413	41386	6.8	41266	7.1
11	44804	42719	4.7	41707	6.9
12	44946	42851	4.7	41705	7.2
13	41661	41367	0.7	38522	7.5
14	42944	41393	3.6	39639	7.7
15	43750	42191	3.6	40483	7.5
16	43559	42011	3.6	40193	7.7
17	43779	42527	2.9	40440	7.6
18	44603	42100	5.6	41423	7.1
19	45673	42727	6.4	42667	6.6
20	44443	42358	4.7	41601	6.4
21	44199	42215	4.5	41373	6.4
22	43626	43032	1.4	40684	6.7
23	45716	43438	5.0	42660	6.7
24	45915	43305	5.7	42641	7.1
평균값	43128	40650	6.0	39893	7.3

표1. 2004년 1월 14일(평일) 수요예측 결과 비교

2004년 1월 14일(평일)수요예측 결과 비교

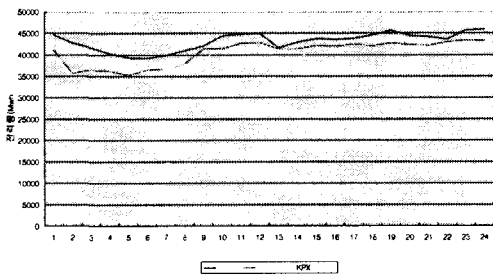


그림6. 2004년 1월 14일(평일) 수요예측 결과 비교

3. 결 론

본 논문에서는 평일의 데이터를 기준으로 데이터마이닝을 이용하여 새로운 예측방법을 제시하였고, 실험결과 현재 한국전력거래소에서 사용되고 있는 KULF(KPX-SNU Load Forecaster)방법과 비교하여 크게 뒤지지 않으며 오히려 주간과 오후의 경우는 전력거래소에서 예측한 결과보다 우수함을 볼 수 있었다. 그러나 데이터의 신뢰도에 따라 오차범위가 커지는 한계점을 보였는데 이는 데이터 정제과정과 데이터 신뢰성의 중요성을 알 수 있게 되었으므로 향후 이러한 부분에 있어서도 지속적인 연구가 필요하며 좀 더 다양한 예측요소의 개발로 오차범위를 줄일 수 있음을 증명하였다.

[참 고 문 헌]

- [1] 조신섭, 황선영, 이궁희, "시계열분석", 한국방송통신대학교 출판사, 2001년
- [2] "KPX-SNU Load Forecaster", 한국전력거래소·서울대학교, 2002년 12월
- [3] 황갑주, 김광호, 김성학 "주간수요예측 전문가 시스템 개발", 대한전기학회논문지, 48A 4호, 365~370, 1999년 4월
- [4] 박우창, 송현우, 용환승, 최기현 역, "데이터 마이닝 개념 및 기법", 자유아카데미, 2003년 9월 1일