韓國컴퓨터情報學會 論文誌
第16卷 第11號, 2011. 11.

2011-16-11-5-4

풍력 데이터를 이용한 발전 패턴 예측

서 동 혁, 김 규 익*, 김 광 득**, 류 근 호*

Predicting Power Generation Patterns Using the Wind Power Data

DongHyok Suh*, Kyu Ik Kim*, Kwang Deuk Kim**, Keun Ho Ryu*

요 약

화석 연료의 무분별한 사용으로 환경이 심각하게 오염되고, 화석 연료의 고갈에 대한 문제가 대두됨에 따라서 화석 연료에 대한 문제를 해결 할 수 있는 대체 에너지원에 대해 관심이 집중되기 시작하였다. 현재 신재생 에너지 중에서 가장 각광을 받고 있는 에너지는 중에 하나가 풍력에너지이다. 풍력에너지 발전단지와 기존의 전력 발전소는 소비되는 전력에 대한 생산의 균형을 맞춰야하며, 풍력에너지단지에서 균형적인 생산을 하기 위해서는 풍력에너지에 대한 분석 및 예측이 필요하다. 이를 위해서 데이터마이닝 분야의 예측 기법이 활용 될 수 있다. 본 논문에서는 풍력 데이터를 이용하여 발전 패턴을 예측하기 위해 SOM(Self-Organizing Feature Map) Clustering 기법과 의사결정나무(decision tree)를 이용한 연구를 진행하였다. 즉, 1) 풍력 데이터의 누락된 데이터와 이상치 데이터를 처리하기 위하여, 전처리 과정을 수행하였고, 이 과정에서 특징 벡터를 추출하였다. 2) 전처리 단계를 거쳐 정제되고 정규화된 데이터 집합을 MIA(Mean Index Adequacy) 척도와 SOM Clustering 기법에 적용하여 대표 발전 패턴을 찾아내고 각각의 데이터에 해당하는 대표 패턴을 클래스 레이블로 할당하도록 하였다. 3) 의사결정나무 기반의 분류 기법에 데이터 집합을 적용시켜 새로운 풍력에너지에 대한 분석 및 예측 모델을 생성하였다. 실험 결과, 의사결정나무를 통한 풍력에너지 발전 패턴을 예측하기 위한 모델을 구축하였다.

▸ Keyword : 신재생에너지, 풍력에너지, 자기조직화형상지도, 패턴예측

Abstract

Due to the imprudent spending of the fossil fuels, the environment was contaminated seriously and the exhaustion problems of the fossil fuels loomed large. Therefore people become taking a

• 제1저자 : 서동혁 • 교신저자 : 류근호
• 투고일 : 2011. 08. 24, 심사일 : 2011. 09. 08, 게재확정일 : 2011. 10. 14.
* 충북대학교 컴퓨터과학과(Dept. of Computer Science, Chungbuk National University, KOREA)
** 한국에너지기술연구원 (Korea Institute of Energy Research)
※ 이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국 연구 재단의 지원(No. 한국연구 2011-0001044) 및 국토해양부 첨단도시기술개발사업 - 지능형국토정보기술혁신사업과제의 연구비지원(과제번호:06국토정보B01)과 한국에너지기술연구원의 지원으로 수행되었습니다.

www.kci.go.kr

great interest in alternative energy resources which can solve problems of fossil fuels. The wind power energy is one of the most interested energy in the new and renewable energy. However, the plants of wind power energy and the traditional power plants should be balanced between the power generation and the power consumption. Therefore, we need analysis and prediction to generate power efficiently using wind energy. In this paper, we have performed a research to predict power generation patterns using the wind power data. Prediction approaches of datamining area can be used for building a prediction model. The research steps are as follows: 1) we performed preprocessing to handle the missing values and anomalous data. And we extracted the characteristic vector data. 2) The representative patterns were found by the MIA(Mean Index Adequacy) measure and the SOM(Self-Organizing Feature Map) clustering approach using the normalized dataset. We assigned the class labels to each data. 3) We built a new predicting model about the wind power generation with classification approach. In this experiment, we built a forecasting model to predict wind power generation patterns using the decision tree.

▸ Keyword : New&Renewable Energy, Wind Power Energy, Self-Organizing Map, Predicting Patterns

I. 서 론

최근 환경오염이나 지구 온난화 등의 문제는 전 세계적으로 이슈가 되고 있는 문제점이다 [1]. 특히, 석유나 천연가스와 같은 화석 연료의 지나친 사용으로 인하여 심각한 오염과 자원의 고갈에 대한 문제가 대두되고 있다. 또한 매년 온실가스 배출량이 증가하고 있는 한국은 2005년 2월 16일에 발효된 교토의정서에 따라 2차 공약기간(2010년~2017년)까지 온실가스 배출량을 감축해야 하는 실정에 놓여있다. 그리하여, 애초에 이러한 문제점을 예측한 미국, 유럽, 일본 등과 같은 몇몇 선진국에서는 화석 연료를 대체할 새로운 에너지 자원에 대한 연구를 활발히 진행해왔다. 신재생 에너지 중 재생에너지 분야에는 태양열, 태양광발전, 풍력, 소수력, 바이오매스, 지열, 해양에너지 그리고 폐기물에너지로 분류된다[2]. 그 중에서 이슈가 되고 있는 신재생 에너지는 풍력과 태양광 분야이다. 본 논문에서는 투자자본 대비 발전 효율이 좋은 풍력 에너지에 대한 데이터를 이용한다. 이러한 자연계 에너지가 가지고 있는 속성은 실시간적이고, 무한량이며, 친환경적이라는 것이다. 그렇기 때문에 화석 연료가 가지고 있는 환경오염과 자원 고갈이라는 문제점들을 해결하기 위해 가장 적절한 에너지 자원이라고 할 수 있다.

신재생 에너지를 효율적으로 이용하기 위해서는 에너지 자원이 풍부한 지역을 찾아서 에너지단지를 조성해야 할 필요가 있는데, 국내에서는 이미 자원 지도에 대한 연구가 이루어진 상태이다. 또한, 풍력에너지는 다른 신재생 에너지 자원에 투자한 비용과 비교하여 전력 생산 효율이 높기 때문에 각광받고 있다. 하지만 이러한 신재생 에너지의 발전량 예측과 패턴에 대한 연구는 아직 미비한 실정이다. 앞으로 시행될 예정인 신재생 에너지의 의무할당제 등으로 인해 신재생 에너지를 이용하여 발전된 전력에 대한 공급을 점차 확대할 계획이다. 이처럼 신재생 에너지의 사용 비율이 증가하고, 소비자들이 고품질의 전력을 사용하기 위해서는 더 많은 연구와 분석이 필요하다. 특히, 신재생 에너지 발전단지를 조성하는데 있어서, 그 발전단지의 지역적인 특성과 발전 패턴을 분석하는 것은 소비자에게 고품질의 전력을 공급하는데 있어서도 중요한 요인이다. 그리하여, 최근에는 신재생 에너지의 발전량 예측에 대한 연구도 이루어지고 있다[3].

본 논문의 연구 내용은 다음과 같다.

첫째, 먼저 한국에너지기술연구원[4]으로부터 풍력 발전 터빈에서 1년 동안 수집된 데이터로부터 패턴을 찾아내기에 앞서, 누락된 데이터와 이상치 데이터를 처리하기 위한 전처리 과정을 수행한다.

둘째, 전력 부하 패턴의 특징 벡터 계산 방법을 이용하여, 풍력 발전 터빈의 패턴에 대한 특징 벡터를 계산한다[5, 6, 7, 8, 9].

셋째, 특징 벡터의 데이터 집합에서 대표 프로파일을 형성하기 위하여 자기조직화지도 (SOM: Self-Organizing Map) [10, 11] 군집 기법을 사용한다. 여기서 자기조직화지도는 Map의 사이즈를 지정해주어야 하는데, 최적의 군집 개수를 지정하기 위해 MIA 척도를 사용한다.

마지막으로, 풍력 발전기의 전력 생산 패턴의 분류 모델이

www.kci.go.kr

구축된다. 만들어진 분류 모델은 새로운 데이터 집합의 패턴을 예측하는 모델로 사용이 된다.

위의 절차대로 풍력에너지 발전량 패턴 예측에 대한 연구를 진행하였고, 단기간 동안 수집된 풍력 데이터를 이용하여 풍력 발전 터빈의 생산 패턴을 예측할 수 있었다.

논문의 구성은 다음과 같다. 2장에서는 연구를 진행하기 위해 필요한 기존 연구들에 대해서 소개를 한다. 그리고 3장에서는 이상치와 누락 데이터에 대한 처리와 특징 벡터를 추출하기 위한 전처리 과정과 대표 프로파일 생성을 위한 자기조직화지도 알고리즘 및 최적의 군집 개수 결정을 위해 사용한 MIA 척도에 대해 기술을 한다. 또한, 생성된 대표 프로파일을 이용하여 생산 패턴을 분류하는 기법에 대해 설명한다. 4장에서는 실험 평가를 설명한다. 그리고 5장에서 결론을 맺는다.

II. 관련 연구

기존의 연구들은 전기를 사용하는 소비자들의 전력 부하 패턴을 분석하는 연구가 많이 진행이 되어왔다. 대표적으로 국내에서는 시공간적인 요소를 고려하여 소비자의 전력 부하 패턴 예측의 정확도를 향상하는 연구[5]가 진행되었고, 국외에서는 소비자의 전력 부하 패턴을 찾아내어 최적의 전력 공급업체를 찾아내는 연구[6]가 진행된바 있다. 또한 우리는 기존에 이루어진 연구를 토대로 그 연구들에서 사용된 프레임워크와 군집 및 분류 기법들을 이용한다.

1. 특징 벡터 추출

논문에서 사용하는 데이터는 대용량의 고차원 데이터이다. 고차원의 데이터는 실험을 하는데 있어서 데이터로부터 규칙을 찾거나 알고리즘의 계산을 하는데 어려움을 겪는다. 이러한 문제를 해결하기 위해 데이터의 특징 벡터를 추출하는 연구가 제안되었다. 이 방법을 사용함으로써, 알고리즘의 계산 비용이 절감되고, 객체들의 대표 패턴을 찾기 위한 기반 데이터를 만들어낸다.

기존의 연구에서 특징 벡터는 부하 데이터로부터 추출을 한다. 부하 데이터는 하루 단위의 벡터 데이터로 만들어진다. Vector l(m)={l1(m),…,lh(m),…,lH(m)}형태로 벡터가 만들어지는데, 여기서 Vector l(m)은 m번째 소비자의 일간 평균 부하 데이터를 저장하는 역할을 한다.

2. MIA (Mean Index Adequacy)

객체들의 대표 패턴을 찾기 위해서는 군집 기법을 사용한다. 여기서 SOM Clustering 기법을 사용하는데, 이 기법은 클러스터의 수를 미리 정해주어야 하는 알고리즘이다. 이 때, 최적의 클러스터의 수를 정해주는 다양한 척도가 존재하는데, MIA 척도는 SOM Clustering에 좋은 성능을 나타낸다[12]. MIA 척도의 공식은 다음과 같다.

$$MIA = \sqrt{\frac{1}{K}\sum_{k=1}^{K} d^2\left(r^{(k)}, L^{(k)}\right)} \quad \text{.....(1)}$$

이 척도는 군집이 한 개 일 때의 MIA 척도 값을 측정하기 시작하여, 최대로 가질 수 있는 군집의 개수가 될 때까지 계산을 반복한다. 그리하여 가장 최적화된 클러스터의 개수를 선택할 수 있게 해준다. MIA 척도의 공식에서 K는 클러스터의 개수를 의미한다. d2는 유클리드 거리의 제곱을 말한다. 그리고 r(k)는 군집의 중심, L(k)는 군집에 포함되어 있는 값들을 의미한다.

3. 자기조직화 형상지도 (SOM : Self-Organizing feature Map)

자기 조직화 형상지도 (SOM) 알고리즘은 핀란드의 전기공학자인 코호넨(Kohonen)이 제안한 알고리즘이다[13]. 이 알고리즘은 2차원으로 구성된 경쟁층과 데이터의 입력을 위한 입력층이 있는데, 입력층을 통해 입력된 특징 벡터 데이터들은 경쟁층에 연결이 되어있으며, 모든 연결들은 입력층에서 경쟁층 방향으로 완전 연결(fully connected) 되어 있다. 그리고 경쟁층에서는 모든 뉴런들이 경쟁을 벌이며 스스로 유사한 값들끼리 조직화되며 지도를 생성한다.

3.1 자기 조직화 형상지도 학습단계

자기 조직화 형상지도는 다음과 같이 세가지 학습단계로 이루어져 있다.

첫번째 단계는 경쟁 단계 (Competitive Step)이다. 초기 모든 연결강도는 0에서 1사이의 값으로 정규화된 임의의 값이 부여된다. 그 다음 입력된 특징 벡터 데이터와 연결된 경쟁층의 뉴런들 중에서 연결 강도와 가중치를 계산하여 거리가 가장 가까운 뉴런이 경쟁에서 승리한다. 자기 조직화 형상지도의 철학이 '승자 독점 (Winner take all)'이기 때문에, 승리한 뉴런만이 출력값을 가진다. 그림 1은 자기 조직화 형상지도의 구성을 나타낸다.

www.kci.go.kr

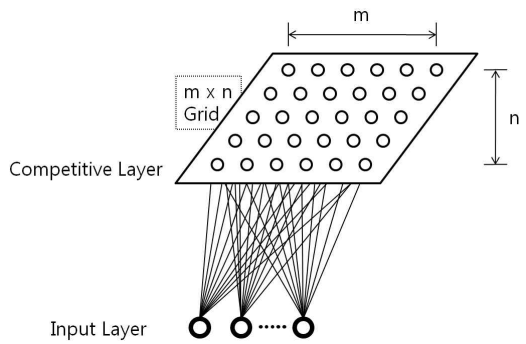

그림 1. SOM의 구조
Fig. 1. Structure of SOM

두번째 단계는 협동 단계이다. 경쟁 단계에서 승리한 승자 뉴런과 그 이웃 뉴런들만이 입력된 특징 벡터 데이터에 대한 학습이 가능하다. 즉, 이 단계에서는 유사한 특징 벡터 데이터에 더 잘 적응이 될 수 있는 지도를 형성할 수 있도록 승자 뉴런과 그 이웃 뉴런들의 연결 강도를 강화한다. 그림 2는 자기 조직화 형상지도의 승자뉴런과 그 이웃 뉴런들을 나타내는데, 다음과 같은 형태로 정의할 수 있다.

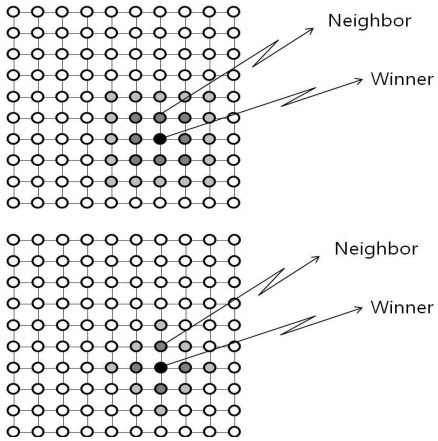

그림 2. SOM의 승자뉴런과 이웃뉴런
Fig. 2. Winner Neuron and Neighbor Neuron in SOM

세번째 단계는 적응단계이다. 앞의 두 단계를 거치면서 학습된 결과로 뉴런의 연결강도를 갱신한다. 이 단계를 통해 특정한 특징 벡터 데이터에 잘 적응하도록 뉴런들을 연결한 연결강도의 갱신이 이루어진다. 그림 3은 적응단계를 거친 지도의 모습을 나타낸다.

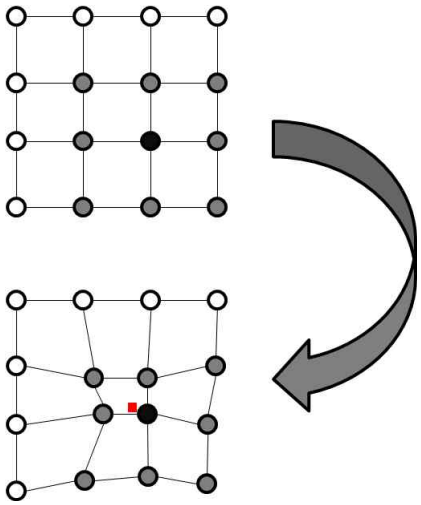

그림 3. SOM의 적응 단계
Fig. 3. The adaptation step of SOM

3.2 자기 조직화 형상지도 알고리즘

연결강도 초기화
- 연결강도는 임의의 값으로 초기화 됨

특징 벡터 입력
- 하나의 특징 벡터 데이터를 입력함
- 다시 반복될 때는 그 다음 특징 벡터가 입력됨

입력된 특징 벡터와 뉴런들의 거리 계산
- 특징 벡터 데이터와 뉴런 j사이의 거리 dj계산
- Xi(t) : 시각 t에서 i번째 입력된 특징 벡터
- Wij(t) : 시각 t에서 i번째 입력된 특징 벡터와 j번째 출력 뉴런 간의 연결강도

$$d_j = \sum_{i=0}^{N-1} \left(X_i(t) - W_{ij}(t) \right) \quad \text{.....(2)}$$

승자 뉴런 선택

승자 뉴런과 그 이웃들의 연결 강도 강화
- Wij(t+1) : 새로 갱신되는 연결 강도
- α : 시간이 경과함에 따라 작아지는 이득항

$$W_{ij}(t+1) = W_{ij}(t) + \alpha \left(X_i(t) - W_{ij}(t) \right) \quad \text{....(3)}$$

2번 단계부터 모든 데이터가 적용될 때까지 반복

www.kci.go.kr

III. 실험 및 평가

그림 4는 풍력 발전 패턴을 예측하기 위한 전체 프레임워크를 보여준다. 각각의 상세한 과정은 다음과 같다.

제공 받은 풍력 발전 데이터는 10분 간격으로 1년 동안 수집된 데이터이다. 하지만 설치된 장비의 안정화 작업이 지연됨으로 인해 데이터 수집에 어려움이 있었다. 이러한 환경적인 조건으로 인하여 하루에 144개의 레코드가 수집된 정상적인 데이터는 23일치에 불과하다. 이러한 제한점들로 인하여 계절적인 요인을 고려할 수 있을 만큼의 데이터가 확보되지 않았으며, 제공된 데이터는 제주도 한 지역에 설치된 하나의 터빈이라는 공간적인 제약도 존재하게 된다. 이러한 환경으로부터 수집된 데이터를 실험에 사용하기에는 데이터의 량이 턱없이 부족하기 때문에, 일정 수준만큼의 누락 데이터를 허용한다. 그리하면 124개 이상의 레코드가 수집된 199일치에 대한 데이터를 얻을 수 있다. 풍력 터빈으로부터 수집된 에너지 데이터에는 이상치 데이터와 누락 데이터가 존재한다.

전처리 과정을 통해 데이터 집합은 총 136일치에 해당하는 일별 대표 벡터 데이터만이 남겨진다. 이 데이터에 Clustering 기법이 적용되면 대표 발전 패턴이 생성된다. 이 과정에서는 MIA 척도와 SOM Clustering 기법을 사용한다. 그 뒤 Classification 기법을 이용하여 예측 모델을 생성한다.

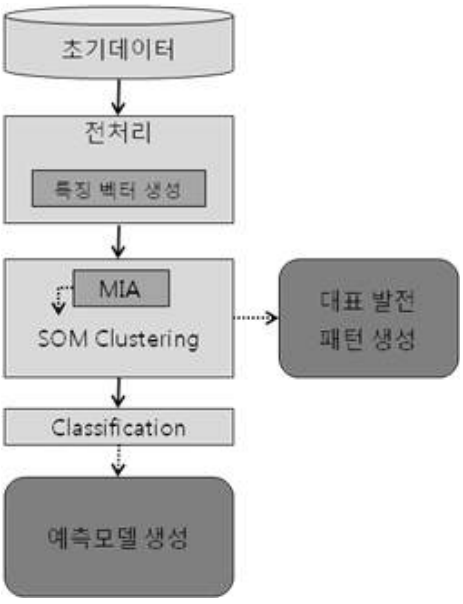

그림 4. 풍력 발전 패턴 예측을 위한 프레임워크
Fig. 4. The framework for predicting wind power generation patterns

1. 데이터 전처리

1.1 누락 데이터 처리

누락 데이터가 존재하는 경우는 그림 5와 같이 나타난다. 누락 데이터는 해당 데이터를 대체하는 기법을 사용하여 처리하는데, 우리는 SPSS의 결측값 대체 방법 중 <결측점에서 선형추세> 방법을 사용하여 누락 데이터를 처리한다. 이 과정에서 1023개의 데이터가 대체된다.

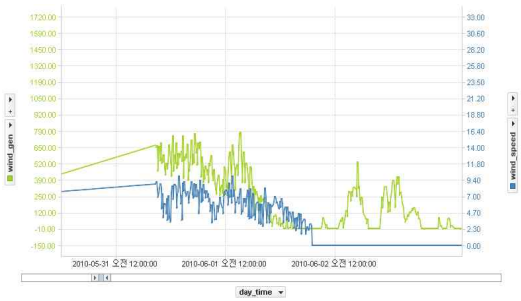

그림 5. 누락 데이터
Fig. 5. Missing Value

1.2 이상치 데이터 처리

본문 내용이상치 데이터에 대한 기준은 다음과 같이 세 가지 경우로 정리된다.

- 풍속은 유동적이나, 발전량이 0보다 작거나 같은 경우
- 풍속이 0보다 작거나 같으나, 발전량이 유동적인 경우
- 인위적인 판단으로 문제가 있어 보이는 경우

우리는 위에 해당하는 경우는 이상치 데이터가 존재한다고 판단할 수 있다. 몇 가지 경우는 DBMS에서 간단하게 처리가 가능하다. 여기에서 발전량이나 풍속이 0보다 작은 값을 포함하고 있는 레코드들이 100개 이상 존재하면, 이상치 데이터로 간주하여 해당 날짜의 데이터를 삭제한다. 이러한 경우가 첫번째와 두번째 경우에 해당한다. 하지만 그 이외의 경우는 인위적인 작업이 필요하다. 우리는 시각화 툴인 TIBCO Spotfire[14]를 이용하여 세번째 경우를 처리한다. 그림 6은 첫번째 경우부터 세번째 경우에 해당하는 데이터를 시각화 한 모습이다.

www.kci.go.kr

그림 6. 이상치 데이터
Fig. 6. Outlier Data

1.3 특징 벡터 추출

배전 계통의 전문가에 의해 제안된 특징 벡터를 계산하는 방법을 이용하여 풍력 발전 터빈의 전력 생산 패턴에 대한 특징 벡터를 계산한다[9].

표 1. 특징 벡터 정의
Table 1. Definition of Characteristic Vectors

Factor	Definition
F1 : Generating Electronic (24H)	$F1 = \frac{Avg.day}{Max.day}$
F2 : Night Generation (8H : 11pm~07am)	$F2 = \frac{1}{8} \times \frac{Avg.night}{Avg.day}$
F3 : Lunch Generating (3H : 12pm~03pm)	$F3 = \frac{1}{3} \times \frac{Avg.lunch}{Avg.day}$
F4 : Midnight Generating (7H : 00am ~ 07am)	$F4 = \frac{7}{24} \times \frac{Avg.midNight}{Avg.day}$
F5 : Morning Generating (3H : 09am~12pm)	$F5 = \frac{1}{3} \times \frac{Avg.morning}{Avg.day}$
F6 : Afternoon Generating (4H : 1pm~5pm)	$F6 = \frac{1}{4} \times \frac{Avg.afternoon}{Avg.day}$
F7 : Evening Generating (4H : 7pm ~ 11pm)	$F7 = \frac{1}{4} \times \frac{Avg.evening}{Avg.day}$
AVG : Average Power Generation in a Day	$AVG = Avg.day$
MAX : Maximum Power Generation in a Day	$MAX = Max.day$
AVG_WIND : Average of wind speed	$AVG\_WIND = Avg.day.wind$

[표 1]에 정의되어 있는 특징 벡터 계산 방법으로 우리는 하루에 해당하는 144개의 풍력 발전 레코드를 단 1개의 레코드로 요약하여 추출할 수 있다.

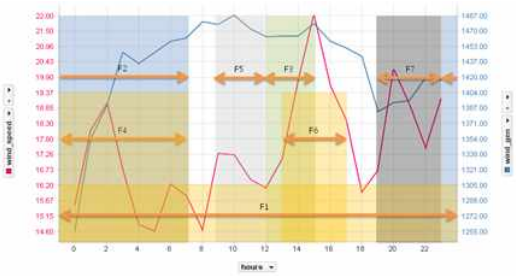

그림 7. 특징 벡터 추출을 위한 시간 범위

그림 7은 특징 벡터를 추출하기 위해 고려한 발전 시간 범위를 보여준다.

2. 특징 벡터에 대한 대표 프로파일 생성

2.1 군집 수 결정을 위한 척도

SOM Clustering 기법을 사용하기 위해서 생성될 군집의 개수를 정해줄 필요가 있다. 군집의 개수는 사용자가 임의로 정해 주어야 한다. 여기서 최적의 군집 개수를 찾기 위해 사용되는 척도는 여러 가지가 존재한다.

여기서 우리는 최적의 군집 개수를 찾아내기 위해 MIA 라고 불리는 척도를 사용한다.

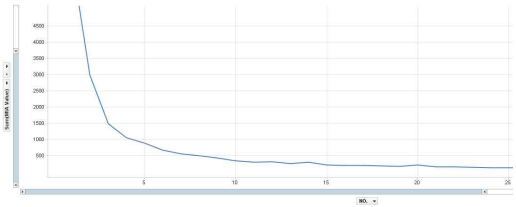

그림 8. 군집 개수별 MIA 척도의 값
Fig. 8. The Value of MIA measure each cluster

그림 8는 군집 개수별 MIA 척도의 값을 나타낸다. 우리는 MIA 척도 측정 실험을 토대로 계산된 값들의 분석을 토대로 값이 안정화되는 부분을 최적의 클러스터 개수로 지정하였다. 이 실험의 결과로 클러스터의 개수는 10개가 가장 적절하다고 분석하였다.

2.2 SOM Clustering 알고리즘

SOM Clustering 기법은 WEKA (Waikato Environment for Knowledge Analysis) Tools[15]의 plug-in으로 제공하는 알고리즘을 사용하였다. 전 단계에서 MIA 척도

www.kci.go.kr

를 통해 최적의 클러스터 개수를 10개로 정하였다. 이에 따라 SOM 클러스터 알고리즘의 MAP 사이즈도 5 by 2로 결정하였다. 그림 9는 풍력 생산 대표 패턴에 대한 각각의 클래스 레이블을 할당한 것이다. 수치는 각각의 대표 패턴별 빈도를 나타낸 것이다.

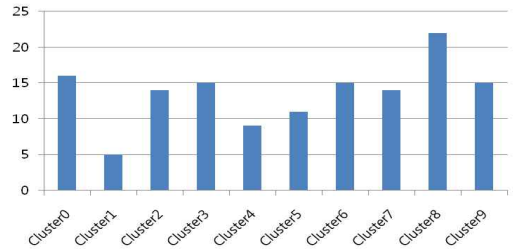

그림 9. 대표 패턴별 발생 빈도
Fig. 9. The Frequency each Representative Patterns

SOM Clustering을 통해 총 136개의 풍력 발전 벡터 데이터는 총 10개의 대표 프로파일을 갖는다. 이 데이터 집합은 예측 모델을 생성하는데 중요한 역할을 한다.

3. 생산 패턴 분류 기법

풍력 발전 패턴을 예측하기 위해서 분류 기법을 이용한 예측 모델을 생성한다. 분류 기법은 WEKA의 C4.5 알고리즘을 사용한다. 이 알고리즘의 의사결정규칙(decision rule)으로 예측 및 분류를 수행한다. C4.5 알고리즘[16]의 특징은 엔트로피 기반의 이득 비율이라는 분할 기준을 이용하여 최적화된 트리를 구축한다. 이 분류 기법을 통해 우리는 풍력 발전 패턴을 예측하기 위한 규칙과 모델을 얻을 수 있다.

4. 실험 결과

우리는 한국에너지기술연구원으로부터 1년 동안 수집된 풍력 발전 데이터 집합을 제공받아 실험에 사용하였다. 이 데이터는 전처리 과정에서 이상치 및 결측치가 처리되었으며, 총 136일치의 데이터 집합을 얻을 수 있었다. 그 다음, 특징을 추출하여 데이터 샘플을 수를 줄였으며, 추출된 특징값을 SOM Clustering에 적용하여 10개의 대표 패턴을 정의할 수 있었다. 각각의 풍력 발전 대표 패턴들은 그림 10과 같은 특징을 나타낸다.

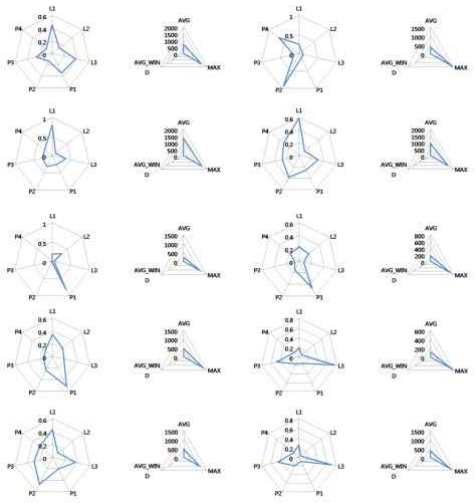

그림 10. 풍력 발전 대표 패턴별 속성값
Fig. 10. Attribute Values of Representative Wind Power Generation Patterns

10개의 대표 패턴을 가지고 있는 풍력 발전 데이터 집합을 통해 분류 모델을 구축한다. 여기서 만들어진 분류 모델을 이용하여 발전 패턴을 예측한다. 실험은 WEKA 툴의 C4.5 알고리즘으로 진행하고, 10 Folds Cross-Validation 기법을 사용하여 모델의 정확도를 평가하였다. 이 실험을 통해 우리는 82.3%의 정확도를 가진 예측 모델을 구축하였다. 그림 11 구축된 모델의 의사결정나무를 나타낸다. 이 모델에서 알 수 있듯이 예측에 큰 영향을 미치는 속성이 몇가지로 추려진 것을 알 수 있다.

```
AVG <= 634.53142
|   F7 <= 0.41394
|   |   F2 <= 0.139033
|   |   |   AVG <= 196.033392: cluster3 (16.0/1.0)
|   |   |   AVG > 196.033392
|   |   |   |   F3 <= 0.577468
|   |   |   |   |   MAX <= 1331: cluster4 (6.0)
|   |   |   |   |   MAX > 1331: cluster6 (3.0)
|   |   |   |   F3 > 0.577468: cluster5 (11.0)
|   |   F2 > 0.139033
|   |   |   F2 <= 0.244579
|   |   |   |   AVG <= 242.17875: cluster1 (5.0)
|   |   |   |   AVG > 242.17875
|   |   |   |   |   MAX <= 1528: cluster2 (12.0)
|   |   |   |   |   MAX > 1528: cluster6 (3.0/1.0)
|   |   |   F2 > 0.244579: cluster0 (17.0/1.0)
|   F7 > 0.41394: cluster7 (15.0/1.0)
AVG > 634.53142
|   F7 <= 0.15419: cluster6 (10.0)
|   F7 > 0.15419
|   |   AVG <= 1069.897222: cluster9 (17.0/2.0)
|   |   AVG > 1069.897222: cluster8 (21.0)
```

그림 11. 예측 모델의 의사결정나무
Fig. 11. The Decision Tree of Prediction Model

www.kci.go.kr

최종적으로 본 실험을 통해 우리는 기존에 수집된 풍력 발전량에 대한 데이터와 풍속과 같은 요소를 이용하여 풍력 발전 터빈의 발전 패턴을 예측 가능한 모델을 구축하였다. 이 모델에 데이터를 적용시켜 해당 데이터의 발전 패턴을 예측할 수 있으며, 추가적으로 패턴 예측에 중대한 영향을 끼치는 속성들을 알 수 있다.

IV. 결론

화석 연료의 많은 사용으로 인하여 심각한 환경오염 및 고갈 문제가 대두되었고, 이에 따라 화석 에너지를 대체 할 수 있는 신재생에너지가 각광을 받기 시작하였다. 이 에너지를 이용하여 발전을 하고, 소비자들에게 공급을 하기 위해서는 에너지에 대한 분석이 필요하다. 본 논문에서는 신재생 에너지 중에서도 발전 효율이 높은 풍력에너지의 예측 모델을 구축하는 실험을 진행하였고, 이전에 이루어진 소비자의 부하 패턴 예측에 관한 연구들을 통해 소비자의 생활패턴을 고려한 예측 프레임워크를 본 논문의 실험에 적용하였다. 그 결과 80% 이상의 정확도를 가진 예측 모델과 그에 따른 의사결정 규칙들을 얻을 수 있었다. 이 예측모델과 프레임워크는 앞으로 기존의 전력발전체계와 풍력에너지를 통해 전력을 균형적으로 공급할 수 있도록 활용될 수 있을 것이다.

향후에는 제주도에 설치되어있는 특정 발전 터빈이라는 제한성과 계절을 고려하지 않은 문제를 해결 할 수 있도록 하고, 또한 실험에 사용 가능한 데이터 집합을 충분히 확보하여 예측에 대한 정확도를 높일 수 있는 연구를 진행 할 것이다.

참고문헌

[1] S.K. Park, L. Wang, Y.K. Lee, D.J. Chai, C.H. Heo, K.H Ryu, K.D. Kim, “Design of Monitoring System Based on Sensor Network for Managing New&Renewable Energy Resources”, In Proc. Of KIPS, Vol.16 No.1, pp. 377-378, 2009

[2] What is the new & renewable energy? (2011, August, 15). Korea Institute of Energy Research Home Page. Retrieved July 15, 2011 from http://www.kier.re.kr/open_content/energy/new_energy/view.jsp

[3] G. H. Ryu, K. S. Kim, J. C. Kim, K. B. Song, "A Study on Estimation of Wind Power Generation using Weather Data in Jeju Island", The Transactions of The Korean Institute of Electrical Engineers, Vol. 58 No.12, pp. 2349-2353, 2009

[4] Korea Institute of Energy Research, http://www.kier.re.kr/

[5] H.G. Lee, Y.H Choi, J.H Shin, "Spatio-Temporal Mining for Power Load Forecasting in GIS-AMR Load Analysis Model", IJIPM: International Journal of Information Processing and Management, Vol. 2, No. 1, pp. 57-66, 2011

[6] V. Figueiredo , F. Rodrigues , Z. Vale and J. B. Gouveia "An electric energy consumer characterization framework based on data mining techniques", IEEE Trans. Power Syst., vol. 20, pp. 596-602, 2005.

[7] M.H. PIAO, J.H. Park, H.G. Lee, K.H. Ryu, “Classification Methods for Automated Prediction of Power Load Patterns”, KCC., Vol. 35, pp. 26-30, 2008.

[8] H. G. Lee, H. J. Shin, K.. H. Ryu, “Forecasting Electric Power Load Patterns using Unsupervised and Supervised Methods from Load Demand Data”, The First International Workshop on Frontiers of Information Technology, Applications and Tools, Vol.1, pp. 2-6, 2008

[9] J. H. Park, H. G. Lee, J. H. Shin, K. H. Ryu, H. S. Kim, “Power Consumption Patterns Analysis Using Expectation-Maximization Clustering Algorithm and Emerging Pattern Mining”, In Proc. Of KIPS, Vol.15 No.2, pp. 261-264, 2008

[10] S.V. Verdú, M.O. Garcı´a, C. Senabre, A.G. Marı´n and F.J.G. Franco, “Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps”, IEEE Transactions on Power Systems., Vol. 21, pp. 1672 -1682, 2006

[11] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer", IEEE Trans. Power System, vol. 18, pp. 381 - 387, Feb. 2003.

[12] G. Chicco, R. Napoli, P. Postulache, M.Scutariu, C.

www.kci.go.kr

Toader, "Load Pattern-Based Classification of Electricity Customers.", IEEE Trans. Power System, vol.19, pp. 1232-1239, 2004

[13] T. Kohonen, Slef-Organization and Associative Memory, 3rd Edition, NewYork: Springer-Verlag, 1989.

[14] Introduction of the TIBCO Spotfire Tools, http://spotfire.tibco.com/

[15] Weka Machine Learning Data Mining Tools, http://www.cs.waikato.ac.nz/ml/weka/

[16] J.R. Quinlan, "C4.5: Programs for Machine Learning, Morgan Kauffman.", 1993.

저 자 소 개

서 동 혁

1992~1997 : 한화통신 기술부 차장
1999~2007 : 벧엘네트워크 개발팀장
2005~2008 : 나사렛대학교 정보통신학과 겸임교수
2008~2009 : 나사렛대학교 멀티미디어학과 대우교수
2006~현재 : 충북대학교 전자계산학과 박사 수료
2010~현재 : 극동대학교 멀티미디어학과 겸임교수
관심분야 : WSN, 데이터퓨전
Email : hanhwaco@naver.com

김 규 익

2010 : 나사렛대학교 정보통신학과 공학사,
현 재 : 충북대학교 대학원 컴퓨터과학과 석사과정
관심분야 : 데이터베이스, 데이터 마이닝, 웹서비스
Email : kikim@dblab.chungbuk.ac.kr

김 광 득

1987 : 대전산업대학교 전자계산학과 공학사.
1999 : 전북대학교 대학원 전산통계학과 이학석사.
2010 : 충북대학교 대학원 전자계산학과 이학박사,
1981년~현재 : 한국에너지기술연구원 책임기술원
관심분야 : 컴퓨터 보안, 시공간 데이터베이스, 데이터 마이닝, 네트워크 관리
Email : kdkim@kier.re.kr

류 근 호

1976 : 숭실대학교 전산학과 이학사.
1980 : 연세대학교 공학대학원 전산전공 공학석사.
1998 : 연세대학교 대학원 전산전공 공학박사,
1976~1986 : 육군군수 지원사 전산실(ROTC 장교), 한국전자통신연구원(연구원), 한국방송통신대 전산학과(조교수) 근무, 1989~1991: Univ. of Arizona R esear ch Staff (Tem plS 연구원, Tempor al DB)
1986년~현 재 : 충북대학교 전기전자 컴퓨터공학부 교수
관심분야 : 시간 데이터베이스, 시공간 데이터베이스, Temporal GIS, 지식기반 정보검색 시스템, 유비쿼터스 컴퓨팅 및 스트림 데이터 처리, 데이터 마이닝, 보안, 데이터베이스, 바이오 인포메틱스
Email : khryu@dblab.chungbuk.ac.kr

www.kci.go.kr