

유전자 알고리즘에 기반한 수산업 전력 수요 예측에 관한 연구

김형수, 이성근*
주식회사 글로비트

Forecasting of Electricity Demand for Fishing Industry Based on Genetic Algorithm approach

Heung-Soe Kim, Sung-Geun Lee*
Globit Co., Ltd.

요약 전력은 모든 나라에서 사회 발전과 경제 성장에 가장 기본적인 자원이다. 산업이 고도화 되고 경제의 규모가 발전하면서 전력의 소비량은 점점 증가하고 있다. 전력을 공급하는 쪽에서는 전력을 생산할 때 자원의 낭비를 줄이기 위해 전력 사용량을 예측하는 것은 중요한 일이다. 또한 전력 수요 예측을 통해 여름과 겨울의 피크 타임에서의 전력 수요를 분산하는 것이 가능하다. 그리고 소비 전력의 예측은 국내에서 수요자원 거래시장(Negawatt market)이 본격화되면서 더욱 중요하게 되었다. 더구나 전력 소비량 예측은 소비자가 전력 시장에 직간접적으로 참여하는 수요관리 방법을 제공해준다. 본 연구에서는 1999년부터 2011년까지의 국내총생산, 1인당 국민총소득, 부가세, 국내전력소비량을 이용하여 제주도의 어업 전력 사용량을 예측하는데 유전자 알고리즘을 사용하고 있다. 유전자 알고리즘은 다양한 조합 최적화 분야에서 최적해를 찾는 데 유용하게 사용되는 알고리즘이다. 본 논문에서 유전자 알고리즘에서 최적의 동작을 위한 파라미터들을 찾는다. 그리고 실제 전력 소비량 예측을 위해 사용되는 계수(coefficient)들의 최적값을 찾아 예측값과 실제 전력 소비량의 오차를 최소화하는데 목적이 있다.

• **주제어** : 전력 소비량 예측, 유전자 알고리즘, 네가와트, 수산업, 경제지표

Abstract Energy is a vital resource for the economic growth and the social development for any country. As the industry becomes more sophisticated and the economy more grows, the electricity demand is increasing. So forecasting electricity demand is an important for electricity suppliers. Forecasting electricity demand makes it possible to distribute electricity demand. As the market for Negawatt market began to grow in Korea from 2014, the prediction of electricity consumption demand becomes more important. Moreover, power consumption forecasting provides a way for demand management to be directly or indirectly participated by consumers in the electricity market. We use Genetic Algorithms to predict the energy demand of the fishing industry in Jeju Island by using GDP, per capita gross national income, value add, and domestic electricity consumption from 1999 to 2011. Genetic Algorithm is useful for finding optimal solutions in various fields. In this paper, genetic algorithm finds optimal parameters. The objective is to find the optimal value of the coefficients used to predict the electricity demand and to minimize the error rate between the predicted value and the actual power consumption values.

• **Key Words** : forecasting energy consumption, Genetic algorithm, Negawatt, fishing industry, economic indicator

*Corresponding Author : 이성근 (sglee@gbit.kr)

Received November 3, 2016
Accepted January 20, 2017

Revised December 8, 2016
Published January 28, 2017

1. 서론

전력 수요의 증가에 따라 전력 수요를 예측하는 기능이 중요한 기능이 되고 있다. 특히, 2014년에 개설된 에너지 효율화와 보존을 통한 소비감량을 의미하는 네가와트(Negawatt) 시장에서는 소비전력을 예측하는 일이 핵심이 되고 있다. 네가와트라는 용어는 전력단위인 메가와트(MegaWatt)와 네거티브(Negative)가 합쳐진 용어이다. 네가와트 시장에서 전력 예측이 중요한 이유는 에너지 정책이 공급위주에서 수요관리 중심으로 변하면서 수요자원의 중요성이 부각되고 있고, 2035년까지 수요관리를 통한 전력수요를 15% 감축을 목표로 하는 등의 수요 예측에 기반 한 전력 소비 감축이 중요한 기본이기 때문이다. 전력 수요의 예측에는 다양한 환경적 요인을 고려 [1,2]하기도 하고, 국가의 경제적 지표들(economic indicators)을 사용하기도 한다. 관련 연구로는 터키의 전력 수요 예측을 위해 유전자 알고리즘을 사용한 연구[3,4]와 기계학습을 사용하여 특정 건물의 전력 수요 예측 연구[5] 등이 존재 한다. 본 논문에서는 전력 수요 예측을 위해 경제적 지표를 이용하기 때문에 선형 수식(linear equation)에 비해 비선형 수식(non-linear equation)에 의한 예측이 좀 더 효과적인 것으로 보인다. 본 논문에서는 전력 수요 예측에 유전자 알고리즘(Genetic algorithm)을 사용한다. 유전자 알고리즘은 다양한 분야에서 최적해를 찾는데 유용하게 사용되는 알고리즘이다. 본 논문의 목적은 여러 산업 분야 중 제주 지역의 어업에 관련하여 전력 소비량을 예측하는 것을 목적으로 한다. 유전자 알고리즘을 이용하여 liner-logarithm model의 변수를 설정하여 실제 데이터와 예측 데이터의 차이를 최소가 되게 하고 결과 값을 기계학습(machine learning)을 이용한 결과와 비교한다[6].

향후 논문의 구성은 다음과 같다.

2장에서는 유전자 알고리즘을 설명과 문제를 정의하며, 3장에서는 실험 결과에 대해 정리와 향후 연구 방향을 기술한다.

2. 유전자 알고리즘

유전자 알고리즘은 다양한 최적화 문제에 우수한 성능을 발휘하는 하는 알고리즘으로 널리 알려져 있다. 이로부터 알 수 있듯이 자연생태계의 진화 모델을 모방한 것으로 1989년 Goldberg에 의해 고안되었다.[7] 기존의 최

적화문제를 위한 알고리즘으로 타부서치(tabu search), 시물레이티드 어닐링(simulated annealing) 등이 하나의 해를 사용하여 이웃해를 탐색하는 방식이므로 지역 최적해(local minima)에 수렴할 확률이 높았으나, 이들에 비해 유전자 알고리즘은 복수 개의 잠재적인 해의 집단(population)을 이용하여 탐색을 하고 우수한 해들을 다음 세대(generation)로 전달하기 때문에 전역해에 수렴할 확률이 높다.

유전자 알고리즘의 이론은 단순하고 해의 탐색이 우수하여 많은 최적화문제에 다양하게 적용 가능하며, 특히 복잡한 해 공간의 탐색이 가능하여 변수와 제약 조건이 많은 문제에 적용가능한 것이 장점이다. 특히, 유전자 알고리즘의 가장 큰 이점은 다음 탐색 공간으로 이동하기 위해 유전자에 유용한 정보나 최적해에 가까운 유전자 정보의 일부를 계속 사용하거나 발전시켜 나갈 수 있다는 것이다[3]. 유전자 알고리즘은 크게 4 개의 부분으로 구성된다. 첫 번째는 문제의 탐색 공간에서 불능해를 포함하여 해가 될 가능성이 있는 부분을 유전자적 표현(genetic representation)으로 나타내는 개체 표현 방법(encoding scheme)이고, 두 번째는 유전자들 중에서 최적해에 가까운 해들을 높게 평가하여 다음 세대에 유전될 확률을 높여 주는 평가함수(fitness function), 세 번째는 해의 탐색 공간을 넓히고 다양한 유전자적 표현을 만들어 내기 위한 유전자 연산(genetic operation)이며, 마지막으로 유전자 알고리즘 자체의 성능 향상을 위해 개체 집단의 크기, 유전자 연산자의 적용 확률 등을 나타내는 제어 파라미터(control parameter)로 구성된다. 유전자 알고리즘에서 세대(population)는 각 가능해(feasible solution)를 나타내는 염색체(chromosome)로 이루어지며, 염색체는 비트로 구성된 스트링으로 구성된다. 유전자 알고리즘의 특징 중 하나는 하나의 해(염색체)를 탐색 공간으로 한정하는 것이 아니라 많은 염색체로 구성된 세대(population)라는 넓은 탐색 공간을 가짐으로써 지역 해에 빠질 확률이 줄어든다. 세대(population)를 구성하는 염색체의 수는 최적화 문제에 따라 다양하게 구성될 수 있으며, 각 염색체의 평가는 목적함수(fitness function)에 의해 결정된다. 유전자 알고리즘이 수행되면서 각 세대는 이전 세대에서 목적함수에 의해 높은 평가를 받은 염색체 위주로 다양한 유전자 연산들을 수행하여 좀 더 좋은 해집단을 구성하게 된다. 다음은 유전자 알고리즘에서 사용하는 다양한 유전자 연산에 대해 설명한다[8].

2.1 유전자 연산

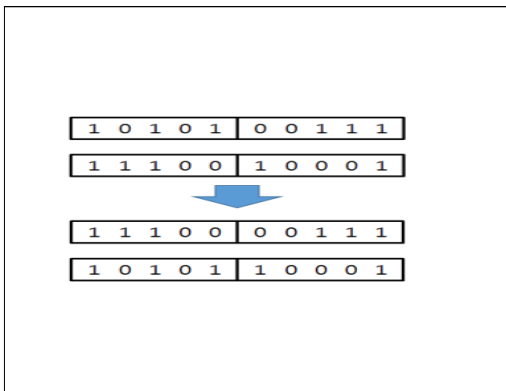
2.1.1 선택 연산(selection operator)

한 세대에서 다음 세대로 전달되는 해(chromosome) 들을 세대에서 선택하는 과정으로 문제에 따라 목적함수 (적합도 함수, fitness function)를 만들고 함수의 계산에 의해 만들어진 값(적합도)에 따라 선택한다. 선택 연산의 종류에 따라 최적해에 수렴하는 속도와 정도에 차이가 발생하며, 잘못된 선택 연산은 지역해에 빠질 가능성이 높아진다. 세대에서 염색체 선택 시에 적합도가 높은 유전자는 높은 확률로 선택되지만, 적합도가 낮은 유전자도 낮은 확률로 선택될 확률은 존재한다. 선택 연산의 종류는 룰렛휠 선택, 순위 선택, 토너먼트 선택 등 존재한다.

2.1.1 교차 연산(crossover operator)

자연환경에서는 교배를 통해 부모의 형질을 물려받아 다음 세대를 생성하게 되는데, 유전자알고리즘에서도 한 세대 내의 선택된 두 염색체들의 교배연산을 통하여 다음 세대에 좀 더 다양한 염색체를 구성하게 된다. 각각 부모 염색체에서 교차연산을 통해 서로 겹치지 않는 위치의 유전자를 이용하여 새로운 유전자로 구성된 자식 세대를 생성한다 이 때 교차 연산의 방법에 따라 일점교차, 이점교차, 균등교차, 순서교차, PMX(Partially Matched Crossover), 휴리스틱 교차 등이 존재한다[9].

다음 [Fig. 1]은 일점교차연산을 표현한 것이다. 일점 교차 연산을 통해 두 개의 부모 염색체 1010111과 1110011에서 두 개의 자식 염색체 1110011과 1010111이 생성된 것을 보여준다. 일점교차연산자는 그림과 같이 부모 염색체의 하나의 특정 지점에서 다른 부모 염색체와 교차가 발생하는 것을 알 수 있다.

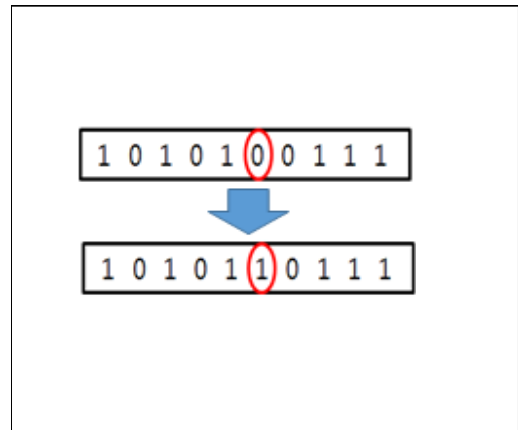


[Fig. 1] One point Crossover

2.2 돌연변이 연산(mutation operator)

기존 세대에서 얻은 염색체(chromosome)와 전혀 다른 해를 얻기 위해 유전자 알고리즘에서 사용하는 연산자이다. 이 연산을 통해 유전자 알고리즘이 지역해(local optima)에 수렴할 확률을 줄이고 다양한 해 공간을 탐색할 수 있는 방법을 제공한다. 교차 연산 후에 염색체 단위로 수행되며 일정한 확률 이하의 빈도로 염색체의 임의의 위치의 값을 교체하는 연산이다.

아래 그림의 경우 염색체1010100111에서 여섯 번째 비트가 0에서 1로 변경되어 1010110111로 변경된 것이다.



[Fig. 2] Mutation

2.3 적합함수(Fitness function)

각 세대에서 만들어낸 염색체들은 모델에 사용된다. 이 때 모델에서 예측된 값과 실제 데이터 값 사이의 차이가 최소가 되도록 적합함수가 정의된다. 그러므로 실제 전력사용량을 E_{actual} 라 하고, 예측된 전력사용량을 $E_{estimated}$ 라 하면, 평가함수는 다음과 같이 정의 된다[10,11].

$$\min f = 1/n \sum (E_{actual} - E_{estimated}) / E_{actual}$$

여기서 n은 population의 크기를 의미한다.

본 논문에서는 전력사용량을 예측을 위해 아래와 같은 선형로그함수모델(Linear logarithmic model)을 사용한다[12,13].

$$\ln y = a + b \times \ln x_1 + c \times \ln x_2 + d \times \ln x_3 + d \times \ln x_4$$

사용되는 데이터는 1999년부터 2015년까지의 국내총생산, 1인당국민소득, 부가세, 국내전력소비량으로써 두 개의 부분으로 나누어서 사용하게 된다. 전반부의 13개

데이터는 최적의 파라미터 값을 찾고 함수의 계수의 최적값을 찾기 위해 사용되고 후반부 4개의 데이터는 목적 함수의 성능을 평가한다[14]. 본 실험에서는 예측에 의한 값과 실제 예측값의 차이가 최소가 되게 하는 것이 목적이다.

3. 실험결과 및 향후 연구 방향

실험을 통해 최적의 적합도를 얻기 위해 유전자알고리즘의 적합도는 <Table 1>과 같다.

<Table 1> Parameters Value

| | |
|--------------------------|------|
| Population size | 100 |
| The number of iterations | 250 |
| Mutation rate | 0.02 |
| Crossover rate | 86% |

실험에 사용한 데이터는 1999년부터 2015년까지 17년간의 데이터로서 국내총생산, 1인당 국민총소득, 부가세, 국내전력소비량 그리고 결과값으로는 제주도 수산업 전력사용량을 사용하였다[15].

<Table 2> Results for Genetic Algorithm

| Year | Actual Data | Estimated Data | Error Rate(%) |
|--------------------|-------------|----------------|---------------|
| 2012 | 741 | 750 | 1.2 |
| 2013 | 790 | 810 | 2.5 |
| 2014 | 832 | 844 | 1.4 |
| 2015 | 829 | 840 | 1.3 |
| Average error rate | | | 1.6 |

<Table 2>는 유전자 알고리즘에 의해 수행된 실험의 결과이다. Error Rate는 (Estimated Data - Actual Data)/Actual Data의 값을 백분율로 나타낸 것으로 4년간의 평균 에러율은 1.6을 기록하였다.

<Table 3> Results for Machine Learning

| Year | Actual Data | Estimated Data | Error Rate(%) |
|--------------------|-------------|----------------|---------------|
| 2012 | 741 | 810 | 9.3 |
| 2013 | 790 | 840 | 6.3 |
| 2014 | 832 | 932 | 12.0 |
| 2015 | 829 | 940 | 13.4 |
| Average error rate | | | 10.25 |

반면 <Table 3>은 기계학습에 의한 실험 결과로써 Error Rate를 구하는 방식은 유전자 알고리즘과 동일하게 적용하였을 때, 평균 에러율이 10.25를 기록하였다.

결과적으로 기계학습에 의한 예측보다는 유전자 알고리즘에 의한 실험이 우수한 성능을 보이고 있으며, 이는 향후에 다양한 형태의 예측 시스템과 실시간 대용량 데이터를 다루는 빅데이터에서도 유전자 알고리즘을 이용한 방식이 더욱 효과적일 수 있다는 사실을 보여준다.

향후 연구에서는 전력사용량에 영향을 미치는 요소를 좀 더 광범위하게 파악할 필요성이 있으며, 빅데이터와 결합하여 실시간으로 환경적 요인을 수산업 전력량 예측에 활용한다면 좀 더 정확한 예측결과가 가능할 것으로 판단된다.

ACKNOWLEDGMENTS

본 논문은 2016년 미래창조과학부의 재원으로 SW융합기술고도화 사업의 지원을 받아 수행된 연구임 (S0177-16-1057)

REFERENCES

- [1] S. Y. Kim, H. W. Jung, J. D. Park, S. M. Baek, W. S. KIM, K. H. Chon and K. B. Song, "Weekly Maximum Electric Load Forecasting for 104 Weeks by Seasonal ARIMA Model", Journal of the Korean Institute of Illuminating and Electrical Installation Engineers, Vol. 28, No. 1, pp. 50-56, 2014.
- [2] H. W. Jung and K. B. Song. "Daily Maximum Electric Load Forecasting for the Next 4 Weeks for Power System Maintenance and Operation", The Transactions of the Korean Institute of Electrical Engineers, Vol. 63, No. 11, pp. 1497-1502, 2014.
- [3] H. Ceylan, H. Ozturk, "Estimation energy demand of Turkey based on economic indicators using genetic algorithm approach", Energy Conversion and Management, pp. 2525-2537, 2004.
- [4] H. Ozturk, H. Ceylan, O. E. Canyurt, A. Hepbasli, "Electricity estimation using genetic approach : a case study of Turkey", Energy 30, pp. 1003-1012, 2003.

[5] Elena Mocanu, Phuong H. Nguyen, Madeleine Gibescu, Wil L. Kling, "Comparison of Machine Learning Methods for Estimating Energy Consumption in Building", PMAPS, 2014.

[6] S. Wong, K. K. Wan and T. N. Lam, "Artificial neural networks for energy analysis of office buildings with daylighting", Applied Energy, Vol. 87, No. 2, pp. 5510557, 2010.

[7] Goldberg DE, Genetic algorithm in search, optimization and machine learning, Adison wesley, 1989.

[8] T. J. VanderNoot, I. Abrahams, "The use of genetic algorithms in non-linear regression of emittance data", Journal of Electro Analytical Chemistry, Vol. 448, Issue 1, pp. 17-23, 1998.

[9] S. N. Sivanandam, S. N. Deepa, "Introduction to Genetic Algorithm", pp. 80-92, 2007.

[10] D. Bunning, M. Sun, "Genetic algorithm for constrained global optimization in continuous variables", Applied Mathematics & Computation, pp. 604-636, 2005.

[11] H. Hasheminia, S. T. Akhavan Niaki, "A genetic algorithm approach to fit the best regression/econometric model among the candidates", Applied Mathematics & Computation, 2006.

[12] A. Sadeghi, "Economics faculty, Electricity consumption forecasting using econometrics methods", 2003.

[13] M. Zamani, economics faculty, Teheran University, Electricity demand function estimation in different economic sectors, 1998.

[14] A. Tang, C. Quek and G. Ng, "GA-TSKfnn : Parameters tuning of fuzzy neural network using genetic algorithms", Expert Systems with Applications, pp. 769-781, 2005.

[15] Yu SH, Choi JY, "The causal relationship between energy and GNP : an international comparison", Journal of Energy, pp. 249-272, 1985.

저자소개

김형수(Heung-Soe Kim)

[정회원]



- 2000년 2월 : 경희대학교 전자계산공학과 (학사)
- 2005년 6월 ~ 2010년 9월 : (주)인투비 솔루션개발부 이사
- 2010년 10월 ~ 현재 : (주)글로비트 대표이사

<관심분야> : 빅데이터, 수산업, 기계학습 알고리즘

이성근(Sung-geun Lee)

[정회원]



- 2004년 8월 : 경희대학교 컴퓨터공학과 (박사)
- 2004년 9월 ~ 2014년 2월 : (주)아이컨택트 연구소장
- 2014년 3월 ~ 현재 : (주)글로비트 수석연구원

<관심분야> : 최적화 알고리즘, 기계학습 알고리즘