한국정보전자통신기술학회논문지(jkiiect)'21-2, Vol.14 No.1
논문 21-14-01-14 http://dx.doi.org/10.17661/jkiiect.2021.14.1.14

기계학습을 이용한 유선 액세스 네트워크의 에너지 소모량 예측 모델

서유화*, 김은회**

Prediction Model of Energy Consumption of Wired Access Networks using Machine Learning

Yu-Hwa Suh*, Eun-Hoe Kim**

요 약 그린 네트워킹(Green networking)은 유선 데이터 네트워크(Wired data network)에서 통합적인 에너지 관리를 통해 에너지 낭비와 CO2 배출 감소를 유도하기 위해 주요 관심분야가 되었다. 그러나 액세스 네트워크(access networks)는 유선 데이터 네트워크 영역에서 사용자 단말을 제외하면 가장 많은 에너지를 소비하는 영역임에도 불구하고 그 범위가 매우 광대하여 통합적인 관리가 어렵고, 그 에너지 소모량과 에너지 절약 잠재성을 예측하기가 매우 어렵다. 본 논문에서는 기존의 다양한 수학적 예측 모델과 실험 및 실측 데이터를 이용하여 유선 액세스 네트워크의 에너지 소모량 데이터를 수집하고 머신러닝(Machine learning)의 지도학습을 이용한 다중 선형 회귀모델을 생성한다. 또한 생성한 모델로부터 다양한 실험을 통해 회귀모델의 성능을 최적화하여 유선 액세스 네트워크의 에너지 소모량을 예측하였고 생성한 회귀모델은 널리 알려진 평가 지표를 통해 성능을 평가하였다.

Abstract Green networking has become a issue to reduce energy wastes and CO2 emission by adding energy managing mechanism to wired data networks. Energy consumption of the overall wired data networks is driven by access networks, expect for end devices. However, on a global scale, it is more difficult to manage centrally energy, measure and model the real energy use and energy savings potential of the access networks. This paper presented the multiple linear regression model to predict energy consumption of wired access networks using supervised learning of machine learning with data collected by existing investigated materials, actual measured values and results of many models. In addition, this work optimized the performance of it by various experiments and predict energy consumption of wired access networks. The performance evaluation of the regression model was achieved by well-knowned evaluation metrics.

Key Words : Energy Consumption, Green Networking, Linear Regression, Machine Learning, Supervised Learning, Wired Access Network

1. 서론

오늘날 환경문제는 세계적인 주요 현안이 되었고, 생존의 문제로 인식됨에 따라, 인터넷에서의 에너지 소모와 CO_2 배출에 대한 문제 또한 인터넷 발전의 한계점으로 새롭게 등장한 요인이 되었다. 오늘날 사용되는 엄청난 수의 전자기기들은 대부분 네트워킹 기능을 내장하고 있으며, 네트워크의 유비쿼터스화는 인터넷의 에너지 소비량을 더욱 증가시킬 것으로 예상되고 있다.

그린 네트워킹(Green networking)은 에너지 절약과 CO_2 배출 감소를 목적으로 통합적인 에너지 관리를 연구하는 영역이다. 이 영역의 주요 이슈는 현재 많은 유선 데이터 네트워크가 전통적인 설계 원리에 따

*Baird University College, Soongsil University
**Corresponding Author : Department of Software Engineering, Seoil University (ehkim@seoil.ac.kr)
Received February 01, 2021 Revised February 01, 2021 Accepted February 04, 2021

라 설계됨에 따른 문제를 해결하는 것이다. 전통적인 유선 데이터 네트워크는 최대 네트워크 처리량과 최소 지연을 목표로 최대 로드에서 성능 저하를 견디도록 설계되기 때문에, 유휴시간에 조차 많은 양의 에너지를 소비하여 높은 에너지 비용을 가진다.

오늘날 유선 데이터 네트워크의 에너지 요구량은 사용자 단말 장비를 제외하면, 액세스 네트워크에 의해 지배된다. 전형적인 유선 데이터 네트워크 구조에서 트랜스포트(transport) 및 코어(core) 네트워크 장비는 전체 유선 데이터 네트워크 에너지 요구량의 약 30%, 액세스 네트워크 장비는 약 70%를 차지한다[1].

그린 네트워킹 기술은 특정한 종류의 개별적인 장비의 에너지 효율성 자체보다 해당 기술의 적용을 통한 통합적인 전체 네트워크의 에너지 절감이 큰 의미를 가진다. 그러나, 엑세스 네트워크는 그 규모로 인해 장비의 규모와 다양성 및 세부적인 사항들의 파악이 어려워, 에너지 소모량과 제안된 기술 적용 시 전체적인 성능을 파악하기가 어렵다.

기존의 액세스 네트워크의 에너지 소모량을 추정하기 위해 제안된 대표적인 모델들[2,3,4,5]은 주로 가입자 수와 몇몇 특정 모델의 장비 당 에너지 소모량의 통합에 따른 상향식(bottom-up) 접근방식을 따른다. 그러나 이러한 방식은 특정한 네트워크 장비와 네트워크 기술이나 개별 장비의 에너지 소모량과 장비 수에 의존하기 때문에 범용성과 확장성이 부족하고, 추정된 결과의 편차가 커 신뢰성이 부족하다.

이러한 문제를 개선하여 본 연구의 선행연구로써 대규모 유선 액세스 네트워크 환경에서 에너지 소모량과 안티그리닝 영향도 추정 모델[6]에서는 액세스 네트워크의 에너지 소모량 추정 모델을 제안한 바 있다. 이 모델은 하향식 접근법을 통해 액세스 네트워크가 위치한 영역 기반으로 에너지 소모량을 추정함으로써 확장성 있는 수학적 모델이며, 기존의 다른 모델링 방식들과 정확성과 신뢰성 측면에서 성능 비교 분석을 수행하여 우수성을 보였다. 그러나 선행연구의 제안 모델은 네트워크들 간의 상대적인 비교에서는 유용하게 사용될 수 있으나 실제 액세스 네트워크의 에너지 소모의 절대적 수치에 있어서 모델의 적용환경에 따라 상이한 결과가 나타날 수 있기 때문에 정확도 측면에서 더 발전된 연구가 필요하다.

오늘날 네트워크 영역에서는 AI 기술 적용에 대한 필요성이 대두되고 있고, 네트워크에서 발생하는 데이터를 기계학습 알고리즘의 학습 데이터로써 효율적으로 활용하여 지능화된 네트워크 운영 방안이 연구되고 있으나, 아직은 초기 단계에 있다. 특히 그린 네트워킹의 분야는 그 규모의 광범위한 특성 상 기존의 제안 모델이나 기술로써 전체적인 에너지 소모량에 대한 정확한 예측이 어려운 한계점이 있음으로 AI의 기술을 이용한다면 환경적 제약을 극복하여 에너지 소모에 대한 보다 정확한 예측이 가능할 것으로 보인다. 그러나 광범위한 영역에 위치한 액세스 네트워크는 학습 데이터 셋을 수집하는 것이 매우 어렵고 알려진 데이터 셋도 거의 부재하기 때문에 방대한 데이터가 필요한 딥러닝의 신경망 학습을 통한 예측방식으로는 적용에 어려움이 따른다. 따라서 본 연구에서는 선행연구에서 제안한 액세스 네트워크 에너지 소모량 추정 모델[6]의 변수를 기반으로 하여 데이터를 수집하고 AI의 기계학습 회귀모델 생성하여 정확성을 향상시킨 유선 액세스 네트워크의 에너지 소모량을 예측하고 생성한 회귀모델의 성능을 평가한다.

2. 관련연구

2.1 액세스 네트워크의 에너지 소모량 추정 모델[6]

본 연구는 액세스 네트워크의 에너지 소모량 추정 모델[6]에서 사용되는 변수들을 회귀 모델의 설명변수로 활용하고 있기 때문에 해당 모델의 주요 변수들을 소개한다.

액세스 네트워크의 에너지 소모량 추정 모델은 하향식 접근법을 통해 대규모 유선 액세스 네트워크의 에너지 소모량을 수학적 모델링을 통해 추정하였다. 대규모 액세스 네트워크의 에너지 소모량 추정에 있어, 기존의 제안 모델들에서는 개별적인 장비의 에너지 소모량의 합산을 통한 상향식 접근 방식을 취하고 있다. 이러한 방식은 사용자 수와 장비 수, 장비 당 에너지 소모량의 파악이 어렵고, 네트워크 장비의 중복성과 냉각 및 전력 공급 장치 등의 추가적인 오버헤드를 고려하

기 어려워 대규모 액세스 네트워크에서는 적용하기가 어려웠다. 이러한 한계점을 개선하여 액세스 네트워크의 에너지 소모량 추정 모델에서는 액세스 네트워크 장비들이 대부분 빌딩 영역에 위치하고 있다는 점을 고려하여 영역 기반의 에너지 소모량 추정 방식을 사용하며 하향식 접근 방식을 통해 가입자 수나 특정한 장비 및 네트워크 기술에 의존하지 않고 액세스 네트워크가 위치한 빌딩 영역을 기반으로 에너지 소모량과 CO_2 배출량을 추정한다.

액세스 네트워크의 에너지 소모량 추정 모델은 추정 대상이 되는 액세스 네트워크가 위치한 영역을 로컬 네트워크 영역이라 정의하며 이 영역은 해당 영역 전체 에너지 소모량이 추정 가능한 영역으로 한 개의 로컬 네트워크 영역에는 한 개의 액세스 네트워크를 포함한다. 상위 네트워크 영역은 로컬 네트워크 영역을 포함하는 더 큰 규모의 네트워크 영역으로 해당 영역의 전체 에너지 소모량이 추정 가능한 영역이다. 상위 네트워크의 에너지 소모량은 식 (1)과 같이 정의하며 식(2), (3), (4), (5)에 의해 유도된다. 각 식에서 사용된 변수들은 표 1과 같다.

$$A_{E_{ne}}(t) = \sum_{i=1}^{n} R_{ne_i}(t) + \sum_{i=1}^{n} E_{ne_i}(t) \quad (1)$$

$$R_{ne_i}(t) = E_i(t) \times B_i(t) \times r \cdot \frac{1}{PUE_{anet}} \times w_i \quad (t = t_0) \quad (2)$$

$$E_{ne_i}(t) = E_i(t) \times B_i(t) \times e \cdot \frac{1}{PUE_{anet}} \times w_i \quad (t = t_0) \quad (3)$$

$$R_{ne_i}(t) = R_{ne_i}(t_0) \cdot (1+\beta)^{\frac{t-t_0}{\tau}} \quad (t > t_0) \quad (4)$$

$$E_{ne_i}(t) = E_{ne_i}(t_0) \cdot (1+\beta)^{\frac{t-t_0}{\tau}} \quad (t > t_0) \quad (5)$$

본 연구의 회귀 분석 모델에서는 액세스 네트워크의 에너지 소모량 추정 모델에서 사용하고 있는 주요 변수들을 활용하여 머신러닝 지도학습을 위한 데이터 셋을 수집하였다.

2.2 선형회귀 분석[7]

회귀분석이란 어떤 자료에서 그 값에 영향을 주는 조건$(x_1, x_2, x_3 \cdots)$에 대해 각 조건에 대한 영향력 $(\beta_1, \beta_2, \beta_3 \cdots)$을 고려하여 해당 조건에서의 y를 예측하는 방법이다. 이때 y를 예측하는 함수 $h()$를 회귀모

표 1. 액세스 네트워크의 에너지 소모량 추정 모델의 변수 정의[6]

Table 1. Parameters of the estimation model of energy consumption in access networks[6]

Notation	Meaning
n	The number of local network areas to be estimated
t	A year to be estimated
t_0	A base year
$AE_{ne}(t)_{BAU}$	Energy consumption of the global access network in the year t
$R_{nei}(t)_{BAU}$	Energy consumption used by the local residential network i in the year t
$E_{nei}(t)_{BAU}$	Energy consumption used by the local enterprise network i in the year t
$E_i(t)$	Total energy consumption used by the local network area i in the year t
$B_i(t)$	The share of energy consumption used by buildings in the local network area i in the year t
PUE_{anet}	The share of the access network equipment of energy consumption used by the building
r	The share of the residential access network equipment of energy consumption used by the building
e	The share of the enterprise access network equipment of energy consumption used by the building
w_i	The level of development ans use of network equipment in the local access network i
β	the annual growth rate of the energy consumption of wired access network equipment

델이라고 하며 식 (6)과 같이 나타낼 수 있다.

$$y = h(x_1, x_2, x_3 \cdots x_k ; \beta_1, \beta_2, \beta_3, \cdots, \beta_k) + \varepsilon \quad (6)$$

여기서 조건이 되는 입력변수 x를 설명변수, 출력결과인 y를 목적변수, β를 회귀 계수(coefficient)라고 한다. 그리고 ε는 오차항으로 현실적인 한계로 인해 발생하는 불확실성으로 일종의 잡음(noise)이다. 회귀

모델에서는 예측치와 실측치의 차이인 오차를 최소화하는 방향으로 회귀계수를 추정한다. 이 때 오차는 손실함수(Loss function)로써 계산된다.

선형회귀(Linear Regression)는 회귀분석 중 가장 단순한 방법론이다. 이 모델은 설명변수와 목적변수 간의 관계를 설명하며 단순선형회귀 모델의 경우 하나의 설명변수는 하나의 목적변수와 관련되며 다중선형회귀 모델은 여러 개의 설명변수와 하나 또는 여러 개의 목적변수가 존재한다. 다중선형회귀 모델은 하나의 목적변수 y는 여러 개의 설명변수 $(x_1, x_2, x_3, \cdots)$에 대해 식 (7)과 같이 나타낼 수 있다.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + \varepsilon \quad (7)$$

다중선형회귀 모델에서는 식 (9)의 오차(e_i)를 최소화하는 방법으로 회귀계수의 추정하여 회귀식 (8)을 찾는 것을 목적으로 한다.

$$\hat{y} = \hat{\beta_0} + \hat{\beta_1} x_1 + \hat{\beta_2} x_2 + \hat{\beta_3} x_3 + \cdots + \hat{\beta_k} x_k \quad (8)$$

$$e_i = y_i - \hat{y_i} \quad (9)$$

3. 실험 및 평가

3.1 데이터 셋(Dataset)

본 연구에서는 액세스 네트워크의 에너지 소모량 추정 모델[6]의 변수를 이용하여 데이터 셋을 수집하였고 머신러닝의 다중선형회귀 모델을 구현하여 지도학습을 통해 유선 액세스 네트워크의 에너지 소모량을 예측하고 그 성능을 평가하였다. 본 연구의 데이터 셋은 다양한 규모의 유선 네트워크 환경에서의 액세스 네크워크에서의 에너지 소모량에 대한 실측데이터와 여러 모델로부터 추정된 약 2900건의 데이터를 사용하였다. 사용한 데이터는 액세스 네트워크의 에너지 소모량 추정 모델[6]에서 사용한 2007~2016년 국가별 전력소모량과 네트워크 장비의 보유 정도와 기술 수준에 대한 지표를 사용하였고, 실제 사례의 액세스 네트워크에 대한 조사 및 실측 데이터로써 LBNL(Lawrence Berkeley National Laboratory)[8]과 CEC(California Energy Commission)[9]의 연구를 통해 수집된 데이터를 사용하였다. 표 2는 수집한 데이터 셋에서 액세스 네트워크의 에너지 소모량을 예측하기 위한 회귀 모델 생성에 사용된 설명변수들의 상세한 내용을 나타낸다.

표 2. 회귀모델의 설명변수
Table 2. Explanatory variables of the regression model

Notation	Meaning
NAE	Access network area's the total energy consumption
FTS	Fixed-telephone subscriptions per 100 inhabitants
MCS	Mobile-cellular subscriptions per 100 inhabitants
IUB	Internet bandwidth Bit/s per Internet user
PHC	Percentage of households with computer
PHA	Percentage of households with Internet access
PII	Percentage of individuals using the Internet
FBS	Fixed-broadband subscriptions per 100 inhabitants
MBS	Mobile broadband subscriptions per 100 inhabitants
ICTS	ICT use skills
NDI	The level of development ans use of access network equipment in the network area
BT	The share of energy consumption used by buildings in the access network area
RT	The share of the access network equipment of energy consumption used by the building

본 연구의 회귀모델에서 사용된 데이터는 액세스 네트워크의 에너지 소모량 추정 모델[6]과 같이 레지던셜과 앤터프라이즈 네트워크를 포함하며 두 가지 네트워크 모두 빌딩 영역에 존재한다. 각각의 네트워크에 해당되지 않는 설명변수의 데이터는 0의 값을 가진다.

3.2 회귀모델 생성

실험에서는 먼저 설명변수들 간에 설명력이 높은 변수와 높지 않은 변수를 확인하기 위해 파이썬

statsmodels 모듈의 다중 회귀분석 모듈을 이용하여 목적변수에 영향을 미치는 것이 유의미하다고 할 수

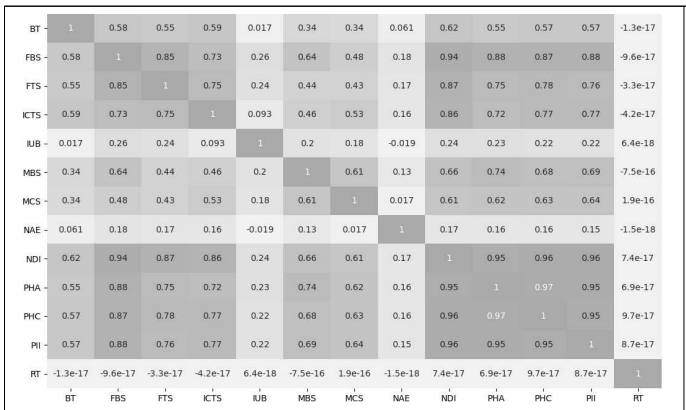

그림 1. (a) 설명변수들 간의 상관관계, (b) 회귀계수 및 p-value
Fig. 1. (a) Correlations of Explanatory variables, (b) Regression coefficients and p-values

설명변수들 간의 상관관계와 p-value를 그림 1과 같이 확인하였다. 그 결과 IUB를 제외한 다른 변수들 간에 높은 상관관계를 가지는 것을 볼 수 있다. 그림 1의 설명변수들 간의 상관관계는 하나의 변수의 값이 증가함에 따라 다른 변수의 값이 증가 또는 감소하는 선형관계를 나타내며 그 수치의 절대값이 1에 가까울 수록 변수들 간의 상관계가 높다고 할 수 있다. 이것은 목적변수에 대해 설명변수들 간에 중복된 설명력을 가짐을 의미하며 다중공선성 문제를 발생시킨다.

다중선형회귀 분석에서는 설명변수들 간의 설명력이 중복된 경우가 대부분이며 설명변수들 간에 강한 선형관계에 있을 때 목적변수를 설명하는 변동성이 겹치는 다중공선성이 있다고 한다. 높은 다중 공선성은 잘못된 변수의 해석을 야기시키고 예측의 정확도를 하락시킨다. 다중선형회귀 모델에서는 설명변수들끼리의 독립 관계일수록 데이터의 잡음이 없는 양질의 데이터를 이용한 학습이 가능하고 설명변수들 간에 낮은 상관계를 가질 수 록(독립관계일 수 록) 모형의 속도가 빠르고, 변수들 간의 중복된 설명력에 따른 다중공선성 문제를 완화시켜 모형이 높은 정확도를 가질 수 있다.

PHC, NDI, PII, IUB 변수들은 p-value도 매우 높은 값을 보이는데 이것은 낮은 유의성을 나타낸다. 일반적으로 p-value는 0.05보다 작은 경우 설명변수가

	VIF Factor	features		VIF Factor	features
0	146.078253	const	7	39.906126	PII
1	2.921624	NAE	8	25.426146	FBS
2	12.482221	FTS	9	2.955295	MBS
3	2.350128	MCS	10	21.863625	ICTS
4	1.217580	IUB	11	423.758332	NDI
5	33.167372	PHC	12	1.882623	BT
6	37.868122	PHA	13	1.001261	RT

그림 2. 설명변수들의 VIF
Fig. 2. VIF of Explanatory variables

있다. 회귀 분석은 목적변수의 변동성(분산)을 설명변수들이 얼마나 잘 설명하는가가 중요하며 다중 선형회귀에서는 각각의 설명변수가 목적 변수를 설명하는 변동성(분산)이 클수록 좋은 변수이며 설명변수의 변동성(분산)이 크면 변수들 간에 낮은 p-value를 가진다.

그림 2는 이들 설명변수들의 다중공선성을 진단하기 위한 VIF(Variance inflation factor)를 나타낸 결과이다. 그림 2의 결과 PHC와 PHA, PII, FBS, ICTS, NDI는 다중공선성이 높은 변수들로써 VIF 수치가 매우 높은 결과를 보인다. 그림 1과 그림 2에 따라 실험에서는 설명력이 높은 변수와 낮은 변수를 구별하고 높은 설명력을 가지는 변수만 선택하여 모델을 생성한다.

실험에서는 가정한 설명변수를 모두 포함하는 모델

을 model1, 설명변수를 추가해가며 성능지표를 비교

표 3. 회귀 모델별 성능지표
Table 3. Evaluation metrics of the regression models

	model 1	model 2	model 3	model 4
The number of Explanatory variables	13	7	6	7
R^2	0.698	0.697	0.697	0.697
MSE	0.3186	0.3196	0.3207	0.3196
RMSE	0.5644	0.5653	0.5663	0.5653
AIC	2271.86	2262.57	2262.92	2262.57
MAE	0.1774	0.1769	0.1754	0.1769

	coef	std err	t	P>|t|	[0.025	0.975]
NAE	0.0014	2.21e-05	63.378	0.000	0.001	0.001
BT	1.3811	0.105	13.116	0.000	1.175	1.588
ICTS	-0.0256	0.007	-3.652	0.000	-0.039	-0.012
MBS	0.0019	0.000	4.877	0.000	0.001	0.003
MCS	-0.0011	0.000	-4.109	0.000	-0.002	-0.001
FTS	0.0026	0.001	2.452	0.014	0.001	0.005
FBS	-0.0028	0.002	-1.529	0.126	-0.006	0.001

그림 3. model 4의 설명변수의 p-values
Fig. 3. p-values of model4

하는 방법인 전진 선택법(Feedforward Selection)으로 생성한 모델을 model2, 변수를 제거해가며 성능지표를 비교하는 방법인 후진소거법(Backward Elimination)에 의해 생성한 모델을 model3, 전진선택과 후진선택법의 혼합적인 방법으로 가장 유의한 변수를 추가하거나 유의하지 않은 변수를 제거하는 단계적 선택법(Stepwise)으로 생성한 모델을 model4로써 생성하여 모델의 성능을 비교하여 모델을 선택하였다. 데이터는 학습데이터와 검증데이터를 7:3으로 하여 검증데이터를 통해 성능지표를 산출하였고 표 3은 3가지 방법에 의해 생성한 모델의 성능 비교이다.

■ **성능지표**

- **설명변수의 수** : 같은 성능을 가지는 모델일 때 설명변수의 수가 작을 수 록 좋음
- **MSE(Mean Square Error)** : 실제 값과 예측한 값의 차이를 제곱한 평균, 작을 수 록 좋으나, 과도한 MSE 축소는 과적합의 가능성이 있음
- **RMSE(Root Mean Square Error)** : MSE가 제곱의 평균으로 실제 오류 평균보다 커지는 특성이 있어 MSE에 루트를 취한 값
- **R^2 (R-squared)** : 입력변수로 목적변수를 설명할 수 있는 변동성을 나타내는 지표, 1에 가까울수 록 좋음
- **AIC(Akaike information criterion)** : 변수의 수가 증가하면 R^2이 증가하기 때문에 MSE에 변수의 수만큼 penalty를 주는 지표, 낮을 수 록 좋음
- **MAE(Mean Absolute Error)** : 실제 값과 예측한 값의 차이의 절대값의 평균

표 3에 따르면 가정한 모든 설명변수를 사용한 model1이 다른 세 모델보다 R^2의 값이 1에 가깝고 MSE가 작지만 설명변수가 소거된 세 개의 모델들의 R^2와 MSE 차이는 유의미하게 큰 차이를 보이지는 않는다. 또한 변수의 수가 많아지면 R^2이 커지는 현상이 발생하고 작은 MSE의 값은 과적합을 의미할 수 도 있기 때문에 R^2과 MSE의 값은 선택의 기준에서 제외하였다. AIC와 MAE의 값은 설명변수를 소거한 모델이 더 작은 값을 가지며 유사한 성능을 가지는 모델에서는 설명변수의 수가 적은 모델을 선택하는 것이 성능면에서 타당함으로 변수를 소거한 모델을 선택하였고 이들 중 가장 좋은 성능을 보이는 model4 모델을 선택하여 유선 액세스 네트워크의 에너지 소모량 예측 모델로 선택하였다.

그림 3은 model4에서 선택한 유선 액세스 네트워크 에너지 소모량 예측에 있어 선택된 설명변수를 나타낸다. 그림 1의 (b)와 비교해 볼 때 설명변수 소거 후 변수들의 p-value가 0에 가까워진 것을 확인할 수 있으며 이것은 설명변수의 유의성이 높아졌음을 나타낸다.

3.3 회귀모델 평가

그림 4는 3.2에서 생성한 회귀모델을 진단하기 위한 지표로써 회귀모델의 잔차가 정규분포를 따르는지를 확인하기 위한 Nomral Q-Q 그래프를 나타낸다. Normal Q-Q그래프는 수집한 데이터의 잔차(Sample Quantiles)의 분포가 정규분포(Theoetical

Quantiles)의 분포와 같은지를 판단하기 위한 그래프로써 회귀모델의 진단하기 위한 주요 평가 지표이다. 그래프의 형태가 직선에 가까울 수 록 정규분포를 따른다고 할 수 있다. 생성모델은 양 끝단의 일부 데이터에 대해서는 직선의 형태에서 벗어나 있지만 대체적으로 정규분포에 가까운 형태를 띠고 있다고 할 수 있다.

그림 5는 회귀모델의 등분산성을 진단하기 위한 그래프로써 X축은 선형회귀로 예측된 Y값, Y축은 잔차를 보여준다. 선형 회귀에서 오차는 평균이 0이고 분산이 일정한 정규분포를 가정하였기 때문에 예측된 Y값과 무관하게 잔차의 평균은 0이고 분산을 일정해야 한다. 따라서 그래프의 기울기가 0에 가까울 수 록 이상적이라고 할 수 있다. 실험 모델의 결과 그래프의 기울기가 0이 되지는 않지만 0에 근접하고 있고 비교적 잔차가 일정한 분산을 가진다고 할 수 있다.

그림 6은 회귀모델의 관측치에 따른 잔차를 나타내는 그래프이다. 특정 구간의 이상치에서는 0의 값을 벗어나 있으나. 대체적으로 0에 가까운 오차를 보여주고 있다. 그림 4와 그림 5에 따르면 특정 구간 별로 나타나는 이상치를 제거하거나 다항 회귀분석을 수행하는 등의 회귀모델에 대한 성능 개선이 필요할 것으로 판단된다.

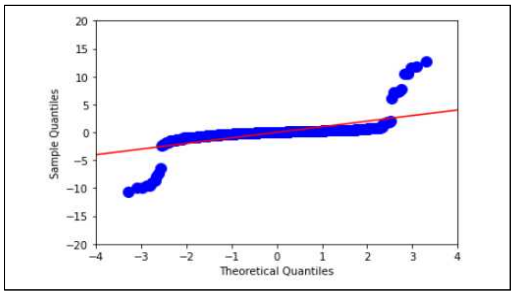

그림 4. 회귀 모델의 Normal Q-Q 그래프
Fig. 4. Normal Q-Q of the regression model

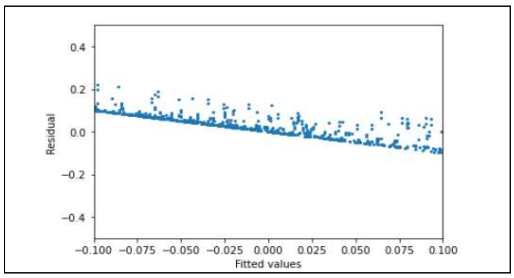

그림 5. 회귀 모델의 예측 값에 따른 잔차
Fig. 5. Residual vs fitted plot of the regression model

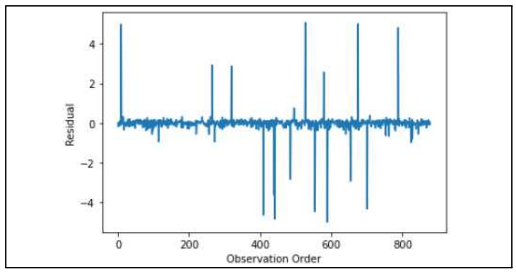

그림 6. 회귀 모델의 관측치에 따른 잔차
Fig. 6. Residual vs observation order of the regression model

4. 결론

본 논문에서는 선행연구인 대규모 액세스 네트워크의 에너지 소모량을 추정하는 수학적 모델[6]의 변수를 기반하여 데이터를 수집하고 이를 설명변수로 하는 AI의 기계학습 다중선형회귀 모델을 생성하였다. 그리고 생성된 모델로부터 전진 선택법, 후진소거법, 단계적 선택법을 이용하여 모델의 성능을 평가하고 개선하여 유선 액세스 네트워크 에너지 소모량을 예측하는 다중선형 회귀모델을 생성하였다. 그리고 생성한 회귀모델의 성능을 평가하기 위해 회귀모델의 일반적인 평가지표인 정규성과 등분산성의 확인하여 다중선형회귀 모델의 성능을 진단하였고, 대체적으로 회귀모델의 기본 가정을 만족하는 다중선형회귀 모델을 생성하였다.

향후에는 생성한 다중선형회귀 모델의 보다 이상적인 결과의 성능을 도출하기 위해 수집 데이터의 이상치에 따른 성능 저하를 개선하고 보다 신뢰성 높은 데이터의 수집과, 다항회귀분석 등을 시도하여 모델의 성능 비교 및 개선을 수행할 필요가 있다. 더불어 수집 데이터의 다중공선성을 제거하기 위한 다양한 형태의 모델 개선이 요구된다.

REFERENCES

[1] R. Bolla, R, Bruschi, F. Davoli, and F. Cucchietti, “Energy efficiency in the future Internet: A survey of existing approaches and trends in energy-aware fixed network infrastructures”, IEEE Commun. Surveys & Tuts., vol . 13, no. 2, pp. 223-244, 2011.

[2] J. Baliga, R. Ayre, W. V. Sorin, K. Hinton, and R. S. Tucker, “Energy consumption in optical IP networks,” J. Lightwave Technol., vol. 27, no. 13, pp. 2391-2403, 2009.

[3] R. Bolla, R. Bruschi, A. Carrega, F. Davoli, D. Suino, C. Vassilakis, and A. Zafeiropoulos, “Cutting the energy bills of Internet Service Providers and telecoms through power management: An impact analysis,” Computer Networks, vol. 56, pp. 2320-2342, 2012.

[4] S. Lambert, W. V. Heddeghem, W. Vereecken, B. Lannoo, D. Colle, and M. Pickavet, “Worldwide electricity consumption of communication networks,” Optics Express, vol. 20, no. 26, pp. B513-B524, 2012.

[5] V. C. Coroama, C. Schien, C. Preist, and L. M. Hilty, “The energy intensity of the internet: Home and access networks ICT innovations for sustainability,” Advances in Intell. Syst. Comput., vol. 310, pp. 137-155, 2015.

[6] Y. H. Suh, K. Y. Kim, “Estimation Modelling of Energy Consumption and Anti-greening I mpacts in Large-Scale Wired Access Networ ks”, J. Korean Institute of Communications and Information Sciences, Vol. 41 No. 08, p p. 928-941, 2016. 08.

[7] David A. Freedman (2009). Statistical Models: Theory and Practice. Cambridge University Press. p. 26.

[8] S. L.anzisera, B. Nordman, and R. Brown, “Data network equipment energy use and savings potential in buildings,” Energy Efficiency, vol. 5, no. 2, pp. 149-162, 2012.

[9] California Energy Commission, Small Network Equipment, California Energy Commission, Jul. 2013.

저자약력

서 유 화(Yu-Hwa Suh) **[정회원]**

- 2019년 3월~현재 : 숭실대학교 베어드교양대학 조교수
- 2016년 3월~2019년 2월 : 서일대학교 정보통신공학과 조교수
- 2016년 2월 : 숭실대학교 컴퓨터학과(공학박사)
- 2007년 11월~2009년 10월 : 정보통신산업진흥원 연구원
- 2005년 8월 : 숭실대학교 컴퓨터학과(공학석사)
- 2003년 2월 : 숭실대학교 컴퓨터학부(공학사)

관심분야 : 그린네트워킹, 유무선네트워크, 인공지능

김 은 희(Eun-Hoe Kim) **[정회원]**

- 2013년 3월~현재 : 서일대학교 소프트웨어공학과 조교수
- 2007년 9월~2013년 2월 : 숭실대학교 정보미디어기술연구소, 지능형로봇연구소 전임연구원
- 2006년 8월 : 숭실대학교 컴퓨터학과(공학박사)
- 1998년 8월 : 숭실대학교 컴퓨터학과(공학석사)
- 1993년 2월 : 숭실대학교 전자계산학과(공학사)

관심분야 : 빅데이터, 텍스트마이닝, IoT, 분산처리, 병렬처리