

SNS 빅데이터 분석을 통한 재생에너지 동향 및 관계구조★

김 종 민*

요 약

본 연구는 재생에너지와 관련된 에너지 분야의 동향과 관계구조를 분석하는 것이다. 이를 위해 본 연구에서는 SNS Data를 포함한 Big Data를 중심으로 분석하였다. SNS는 Instagram 플랫폼을 활용하여 재생에너지 해시태그들을 수집하였으며, 빅데이터 분석, 소셜네트워크 분석을 위한 워드임베딩 방법으로 사용하였고, 본 연구에서 도출된 결과를 토대로 재생에너지 산업의 발전에 활용할 수 있을 것으로 기대된다.

Renewable energy trends and relationship structure by SNS big data analysis

Jong-Min Kim*

ABSTRACT

This study is to analyze trends and relational structures in the energy sector related to renewable energy. For this reason, in this study, we focused on big data including SNS data. SNS utilizes the Instagram platform to collect renewable energy hash tags and use them as a word embedding method for big data analysis and social network analysis, and based on the results derived from this research, it will be used for the development of the renewable energy industry. It is expected that it can be utilized.

Key-words: SNS, Word2Vec, Renewable energy, Big Data

접수일(2022년 2월 07일), 수정일(2022년 3월 19일), 게재
확정일(2022년 3월 25일)

* 동신대학교 정보보안학과 교수

★ 본 과제(결과물) 2021년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과입니다.(1345341782)

1. 서론

전 세계적으로 온실가스 배출량의 증가로 인해 이상 기후변화가 발생하고 있으며, 이를 해결하기 위한 방안으로 재생에너지로의 전환이 시급하다.

2022년 현재의 시점으로 전 세계 인구의 25%, 10억 명 이상의 사람들이 재생에너지를 사용하고 있으며, 탄소 에너지를 줄이고 재생에너지로 전환하는 정책들을 도입을 하고 있다.

재생에너지의 글로벌 주요 정책은 에너지 및 기후 목표에 대해 달성하기 위해서 정부와 도시내에서 재생 가능 에너지를 생산하고, 건물, 차량 등의 재생 에너지들의 이용을 확대하여, 도시 전체에 재생 에너지를 사용할 수 있도록 하는 정책을 시행하고 있다.

이처럼 재생에너지 정책은 화석연료의 의존하는 에너지 소비를 대체하여 환경적인 문제를 해결할 수 있는 점에서 재생에너지로의 확대 및 전환에 대해 중요시 되고 있다. 이러한 의미에서 SNS상에서 재생에너지의 연관단어들을 분석하여 최근 동향에 대한 진단적 연구가 필요하다. 또한 이를 통해 재생에너지로의 전환에 중요한 전략적 방향성을 제시할 수 있을 것이다.

2. 관련연구

2.1 재생에너지

세계 각국은 전력 부문에서의 재생에너지 보급 목표를 수립하고 있으며, 보급 확대를 위한 정책적 지원이 증가하는 추세이다. 전 세계적으로 신규 재생에너지 설비 비중은 계속해서 증가하여 2019년에는 신규 발전설비의 75%를 기록한 바 있다. 재생에너지 신규 보급의 경우, 선진국은 주로 태양광과 풍력 설비를 추가하고 있으며, 개발도상국은 태양광과 풍력의 신규 보급 비중이 선진국보다 작지만, 그 수치가 빠르게 증가하고 있다. 2015년을 시작으로 중국의 재생에너지에 대한 투자가 급증하였고 현재는 개발도상국의 투자 비용이 선진국을 추월한 상태이며 인도, 브라질, 베트남을 포함하여 다른 개발도상국의 투자도 증가하고 있다[1].

2.2 SNS 분석

최근 스마트 스마트폰, 태블릿 PC와 같은 디바이스의 발달로 인해 SNS에 접속할 수 있는 환경이 조성되었고, 이로 인해 SNS 시장은 급속한 성장을 맞이하게 되었다. 특히 스마트폰을 기반으로 다양한 의견을 SNS에 실시간으로 표현할 수 있게 되었으며, 개인의 반응을 실시간으로 수집, 분석이 가능해졌다. 이에 따라 SNS 데이터를 기반으로 사용자들의 생각을 분석하고, 의미 있는 정보를 유추해내거나 특정 분야의 흐름을 예상할 수 있는 분석 기술들의 중요성이 커지고 있는 실정이며, 실제로 최근 SNS 데이터를 활용한 연구가 활발하게 이루어지고 있다[2][3][4].

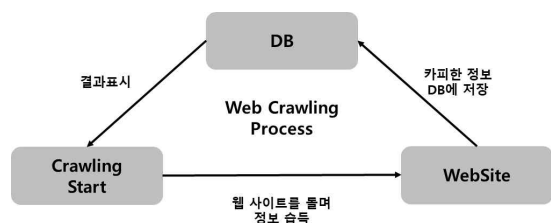
<표 1> SNS 분석 관련연구[5][6][7][8]

저자	내용
Y. D. Yun (2016)	인공지능에 대한 사람들의 의견 데이터를 연도별로 비교 분석
A. R. Lee (2015)	광고 데이터 수집 후 실시간으로 광고 효과 분석 시스템 설계 및 구축
Y. C. Hwang (2013)	악성 집단 댓글 분석에 의한 SNS 여론 소셜 데이터분석
S. I. Son (2013)	SNS의 다양한 요소를 추출하고, 기계학습을 통해 데이터 분석 에이전트 설계

3. 연구방법

3.1 데이터 추출(Web Crawling)

(그림 1)은 웹 크롤링의 흐름도이다.



(그림 1) Web Crawling Process

먼저, 크롤링을 시작하게 되면, 정보를 수집할 사이트에 접속을 하게 되고, 접속 후 가져올 정보에 대해 저장하게 된다. 정보를 가져오는 과정에서 모든 정보를 가져올 때까지 반복하여 수행을 하게 되는데 반복이 끝나게 되면 크롤링이 종료되고 정보가 DB에 저장하게 된다.

3.2 주요 연관단어 도출(Word2Vec)

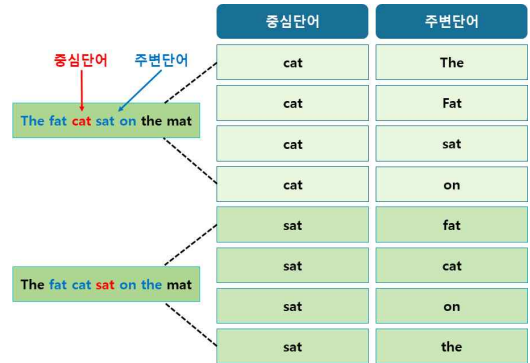
Word2Vec에는 CBOW(continuous bag of words)와 Skip-gram이라는 2개의 세부 학습모델이 있으며, 본 연구에서는 중심 단어를 통해 주변 단어를 예측하는 Skip-gram 모델을 사용하였다[9].

Word2Vec의 Skip-gram 모델을 활용하여 연관 단어들에 대한 유사도 분석 시에는 파라미터를 설정해야 하는데, <표 2>와 같이 파라미터에는 vector size, alpha, window size, min count, sg가 포함된다[10].

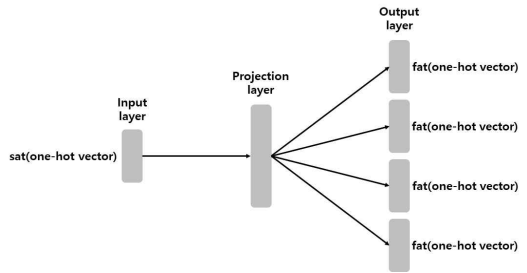
<표 2> Word2vec 분석 파라미터 설명

파라미터	설명
Vector size	벡터 차원의 크기
alpha	learning rate
window size	모델 학습할 때 앞뒤로 보는 단어의 수
min count	학습에 사용할 최소 단어 빈도 수
sg	skip-gram (default)

Skip-gram은 중심 단어에서 주변 단어를 예측하는 기법으로 Windows size가 2의 값을 가질 때, 데이터 셋은 (그림 2)와 같이 구성되며, 아키텍처는 (그림 3)과 같다[11].



(그림 2) Skip-gram 데이터셋 구성



(그림 3) Skip-gram Architecture

4. 분석 및 결과

크롤링을 활용한 데이터 수집 및 Word2vec을 사용하여 재생에너지와 관련된 연관단어를 추출하기 위한 분석환경은 <표 3>와 같다.

<표 3> 분석환경

환경	SPEC
운영체제(OS)	Windows10 Pro
CPU	Intel(R) Core(TM) i5-7500 CPU @ 3.40GHz 3.40 GHz
메모리(RAM)	20GB
분석 데이터	https://www.instagram.com/explore/tags/재생에너지

4.1 데이터 추출

연관단어 분석을 위해 재생에너지에 대한 데이터를 수집하였으며, <표 4>와 같이 데이터 수집을 위해 크롤링 기법을 적용하여 수집하였다.

<표 4> 데이터 수집 Code

```
import requests
from bs4 import BeautifulSoup
from selenium import webdriver as wd
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
import time
import re
import pandas as pd
import numpy as np

# 해시태그 검색어
keyword = "재생에너지"

# 사진 정보 정의
username = '인스타ID'
userpw = '인스타P/W'
time.sleep(3)

# 해시태그 url 값
url = "https://www.instagram.com/explore/tags/{}/".format(keyword)

... (중략)

# 결과값 저장
test = pd.DataFrame.from_dict(instadict)
time.sleep(20)# 추가
newtag = pd.DataFrame()# 추가
tag = test['hashtag']# 추가
test = test.drop(columns=['hashtag'], axis=1)# 추가

for k in range(0,mtag):# 추가
    newtag[str(k)]=tag.str.split(',').str[k]# 추가

newdata = pd.concat([test,newtag],axis=1)# 추가

newdata.to_csv('C:\jmtest\crawling\word2vec.csv', encoding='utf-8-sig')# 추가
# 크롬드라이버 종료
driver.close()
```

데이터 수집은 인스타그램에서 재생에너지와 관련된 해시태그들을 자동 추출하여, 지정된 경로에 .csv 형태로 저장되어지며, 수집된 데이터의 형태는 <표 5>와 같다.

<표 5> 수집 데이터 속성

none	id	date	hashtag
0	ktv_kr	2021-12-14 00:00:00	#문재인,#대통령,#정와대,#시드니,#코로나19,#COVID19,#경제회복,#기후변화,#한국기업,#양국협력,#신재생에너지,#에너지#전기차,#수소차,#대한민국,#KTV국민방송
1	dragon_t	2021-07-27 00:00:00	#녹색성장
2	ilovek_hnp	2020-11-13 00:00:00	#한국수력원자력,#한수원,#KHNP,#신재생에너지,#생활SOC,#연료전지
:(생략)			

4.2 연관단어 도출

Word2vec의 Skip-gram 모델을 통해 재생에너지 관련 주요 연관단어를 도출하였으며, 도출 시 사용된 파라미터의 값은 vector_size=150, alpha=0.025, window=2, min_count=50, sg=1 로 설정하여 분석을 진행하였다.

<표 6> Word2vec 분석 Code

```
import pandas , nltk
from gensim.models import Word2Vec
from nltk.tokenize import RegexpTokenizer

filename = ('C:\jmtest\crawling\word2vec.csv')
df = pandas.read_csv(filename, sep=',')

def nltk_tokenizer(_wd):
    return RegexpTokenizer(r'\w+').tokenize(_wd.lower())

df['Token_hashtag'] = df['hashtag'].apply(nltk_tokenizer)
df.head()

common_texts = df['Token_hashtag'].tolist()

model = Word2Vec(
    common_texts,
    vector_size=150,
    alpha=0.025,
    window=2,
    min_count=50,
    sg=1)

model.train(common_texts, total_examples=len(common_texts), epochs=10)
```

상위 10위의 재생에너지의 주요 연관단어에 대한 빈도분석 결과 <표 7>과 같이 에너지, 신재생에너지, 태양광발전, 친환경, 탄소중립, 환경산업,

친환경플라스틱, ecorn, 환경보호, 재생가능에너지 순으로 빈도가 높게 나타났다. 이들 주요 연관단어들의 주요 특성을 살펴보게 되면 환경과 관련된 내용들이 주를 이루고 있다.

<표 7> 재생에너지 연관단어 빈도

순위	해시태그	빈도
1	에너지	0.964
2	신재생에너지	0.960
3	태양광발전	0.956
4	친환경	0.953
5	탄소중립	0.951
6	환경산업	0.948
7	친환경플라스틱	0.947
8	ecorn	0.944
9	환경보호	0.942
10	재생가능에너지	0.940

<표 7>과 같이 재생에너지의 연관단어 빈도가 주로 환경과 관련된 내용의 이유는 대부분의 에너지생산이 탄소에너지 이용하고 있고, 이로 인해 이산화탄소 배출량의 증가로 지구온난화로 인한 이상기후 변화가 발생을 하고 있으며, 세계적으로 지구온난화의 주요원인인 온실가스 배출을 줄이기 위해 재생에너지의 전환을 정책적으로 유도함으로써, 전환 기업들에게는 감세 및 재정적인 지원을 하기 때문에 재생에너지의 연관단어에서 환경과 관련된 내용의 빈도가 높은 것으로 분석된다.

5. 결론

본 연구는 재생에너지 정책에 대한 연구로서 재생에너지의 주요 연관단어의 구조를 분석을 하였다. 이를 위해 대표 SNS인 인스타그램 해시태그를 토대로 빅데이터를 수집하여, Word2Vec을 활용하여 분석을 진행하였다.

분석 후 도출된 상위 10개의 주요 연관단어(에너지, 신재생에너지, 태양광발전, 친환경, 탄소중립, 환경산업, 친환경플라스틱, ecorn, 환경보호, 재생가능에너지)들의 특징으로는 환경과 관련된 내용이 주를 이루고 있었으며, 이는 전 세계적으로 지

구 온난화로 인한 이상기후변화에 대응하기 위해 탄소중립을 선언함으로써, 우리나라 정부도 탄소에너지에서 재생에너지로의 전환을 유도하는 정책을 시행함으로써 재생에너지의 주요 연관단어들이 환경과 관련된 단어들이 도출되었다.

도출된 결과를 토대로 현재의 재생에너지가 기업적인 사업수단이 아닌 환경과 관련하여 탄소중립 사회로의 전환으로 변화함을 알 수 있었으며, 향후 이 결과에 기초하여 재생에너지에 대한 정책적 추가 연구가 필요하다.

참고문헌

- [1] D. M. Lee, S. Y. Jung and J. S. Lim, "Analysis of Trends of New and Renewable Energy Businesses in Global Oil and Gas Companies", Journal of the Korean Society of Mineral and Energy Resources Engineers, Vol. 58, No. 4, pp. 241-352, 2021.
- [2] J. Y. Seo and C. Koh, "Big Data Analysis by Sensitivity Analysis.", Journal of the Society of Convergence Knowledge, Vol. 2, No. 1, pp. 15-21, 2014.
- [3] J. S. Sohn and S. W. Cho and K. L. Kwon and I. J. Chung, "Improved Social Network Analysis Method in SNS.", Journal of Intelligence and Information Systems, Vol. 18, No. 4, pp. 117-127, 2012.
- [4] J. Y. Chang, "Convergence of Education and Information & Communication Technology : A Study on the Communication Characteristics of SNS Affecting Relationship Development between Professor and Student.", Journal of the Korea Convergence Society, Vol. 6, No. 6, pp. 213-219, 2015.
- [5] Y. D. Yun, Y. W. Yang and H. S. Lim, "A SNS Data-driven Comparative Analysis on Changes of Attitudes toward Artificial Intelligence.", Journal of Digital Convergence, Vol. 14, No. 12, pp. 173-182, 2016.

- [6] A. R. Lee and J. S. Bang and Y. H. Kim, "A Design of a TV Advertisement Effectiveness Analysis System Using SNS Big-data.", KIISE Transactions on Computing Practices, KTCP, Vol. 21, No. 9, pp. 579-586, 2015.
- [7] Y. C. Hwang and C. Koh, "Analysis of Opinion Social Data on the SNS (Social Network Service) by Analyzing of Collective Damage Reply.", Journal of Digital Convergence, Vol. 11, No. 5, pp. 41-51, 2013.
- [8] S. I. Son and C. G. Park, "A Study on Design of Agent for SNS Big Data Analysis.", Journal of Industrial science re-searches, Vol. 31, No. 1, pp. 209-214, 2013.
- [9] Y. S. Sim and H. B. Kim, "The Trend and Structure of Domestic Travel related to Travel Week Using SNS Big Data: Focused on Hashtag of Instagram", Journal of Tourism Sciences, Vol. 42, No. 8, pp. 165-189, 2018.
- [10] I. B. Heo and H. J. Yun, "Clustering and Centrality of Keywords related to Solo Travel at Social Media : Using Word2Vec and Social Network Analysis", Journal of Tourism Management Research, Vol. 24, No. 2, pp. 767-787, 2020.
- [10] S. M. Shin and K. C. Kim, "Addressing the New User Problem of Recommender Systems Based on Word Embedding Learning and Skip-gram Modelling", Journal of The Korea Society of Computer and Information, Vol. 21, No. 7, pp. 9-16, 2016.

[저 자 소 개]



김 종 민 (Jong-Min Kim)
2015년 산업보안학박사
현 재 동신대학교 정보보안학과
교수

email : dyuo1004@gmail.com