

<http://dx.doi.org/10.17703/JCCT.2024.10.6.703>

JCCT 2024-11-86

언어모델의 단어벡터를 이용한 영화 텍스트 분석 기법 연구

Analysis Method Study of Film Text using Word Vectors of Language Model

고광호* 백주련**

Kwangho Ko*, Juryeon Paik**

요약 언어모델을 구축하기 위한 딥러닝 기법인 LSTM의 경우 대형언어모델과 달리 컴퓨팅 자원이 작은 시스템에서도 수월하게 학습시킬 수 있다. 소규모 텍스트에 대해 LSTM 기반의 언어모델을 학습시키고, 텍스트를 구성하는 어휘의 단어벡터를 이용하여 해당 텍스트의 주요 주제어에 대해 객관적인 의미 및 관계 분석을 할 수 있는 융복합적인 기법을 제안하였다. 데이비드 로워리 감독의 2021년도 영화 '그린 나이트'의 영어 대본을 텍스트로 삼아 학습시킨 소규모 언어모델의 단어벡터를 이용하여 주요 주제어의 의미와 관계를 분석할 수 있는 기법을 제안하였다. 단어벡터의 유사도 연산을 통해 각 주제어들과 유사도가 높은 단어를 분석하여 그 의미와 상징성을 객관적으로 분석할 수 있고, 차원감소시킨 2차원 단어벡터를 도시하여 각 주제어들의 관계를 직관적으로 인식할 수 있었다. LSTM 방식의 소규모 언어모델을 이용하여 학습에 필요한 비용을 최소화하면서도 복잡한 텍스트를 분석할 수 있는 단어벡터 활용법을 제안하였다.

주요어 : 딥러닝, 단어벡터, 녹색기사, 유사도, 영화텍스트, 데이비드 로워리

Abstract LSTM, a deep learning technique for building language models, can be easily trained on systems with small computing resources, unlike large language models. In this paper, we propose a convergent technique to train LSTM-based language models on small-scale texts and perform objective semantic and relational analysis on the main topic words of the text using the word vectors of the vocabulary comprising the text. Using the word vectors of a small language model trained on the English script of the 2021 movie "Green Knight" directed by David Lowery as a text, we proposed a technique that can analyze the meaning and relationship of the main topic words. Through the similarity operation of the word vector, the meaning and symbolism of each theme word can be objectively analyzed with the similarity scores between the words. The relationship between each theme word can be intuitively recognized by displaying the dimensionality-reduced two-dimensional word vector. By using a small-scale language model of the LSTM method, we proposed a method to analyze complex texts using word vectors while minimizing the cost of learning.

Key words : Deep Learning, Wordvector, Green Knight, Similarity, Film Text, David Lowery

*정희원, 성균관대학교 응용AI융합학부 교수
*정희원, 평택대학교 데이터정보학과 부교수
접수일: 2024년 8월 22일, 수정완료일: 2024년 9월 25일
게재확정일: 2024년 11월 5일

Received: August 22, 2024 / Revised: September 25, 2024
Accepted: November 5, 2024
** Corresponding Author: jrpaik@ptu.ac.kr
Dept. of Data Information, Pyeongtaek Univ, Korea

I. Introduction

언어모델(LM : Language Model)은 주어진 학습용 말뭉치(Corpus)를 일정한 길이 단위로 입력받아 말뭉치를 구성하는 단어(토큰, Token)의 순서를 기억하는 딥러닝(DL : Deep Learning) 모델이다. 이러한 딥러닝 모델을 구성하는 층(Layer)은 보통 임베딩(Embedding) - RNN(Recurrent Neural Network) - DNN(Dense Neural Network) 으로 구성된다[1]. RNN에 대해서 전통적인 LSTM(Long Short Term Memory) 방식을 사용할 수도 있고, 트랜스포머(Transformer)와 같은 최신 알고리즘을 적용할 수 있다. 챗GPT와 같은 대형언어모델(LLM : Large Language Model)의 경우 트랜스포머 알고리즘이 적용된 언어모델이다. LSTM을 적용하는 경우에는 약 천만 개 수준의 매개변수를, 트랜스포머의 경우 약 천억 개 수준의 매개변수를 학습시켜야 한다[2].

LSTM 방식의 언어모델은 대형언어모델처럼 방대한 추론(Inference) 과정이 필요한 자연스러운 대화, 응대, 코딩, 작문 등의 작업은 어려우나, 비교적 작은 규모의 텍스트 자료에 대한 토픽 분류, 추천 시스템 등에는 효과적인 모델이다. GPU와 메모리 같은 가용 자원이 부족한 상황에서 LSTM 방식의 언어모델이 사용 가능한 최적의 모델이 될 수 있으므로 모델의 활용도가 높다고 할 수 있다[3].

본 연구에서 LSTM 방식의 언어모델을 이용한 융복합적인 텍스트 분석 기법을 제시하고자 한다. 이해하기 어려운 맥락이나 미묘한 뉘앙스의 차이에 대해 언어모델의 단어벡터(Word Vector)를 활용하는 방식이다. 다만 이러한 방식을 적용하기 위해서는 해당 작품에 대한 융복합적인 이해를 필요로 하는 단점이 있으므로 협동적인 연구 과정을 통해 극복 가능한 기초적인 기법을 제안하였다.

II. Related Works

본 연구에서 말뭉치로 사용한 텍스트는 영화 'Green Knight'의 영어 대본으로, 데이비드 로워리(David Lowery) 감독의 2021년도 작품이다. 상영시간이 2시간 정도인 영화로, 14세기 잉글랜드의 중세 로망스 작품인 '가웨인 경과 녹색 기사(Sir Gawain and the Green

Knight)' 라는 서사시(저자 미상)를 각색한 영화이다. 영화 텍스트에 대한 이해도를 바탕으로 후후 언어모델을 적용하기 때문에 원작 서사시와 함께 영화 텍스트에 대한 이해를 위해 몇 가지 주요 사항들을 살펴보았다.

일반적인 중세 로망스는 기사의 무용담을 주요 소재로 다루는데, 역경을 이겨내고 진정한 기사로 성장하는 과정에서 숙녀(Lady)에 대한 사랑을 지켜내는 것이 주요 스토리이다. 하지만 '가웨인 경과 녹색기사' 서사시에서는 전형적인 중세 로망스의 공식이 깨어진다. 서사시의 시대 배경은 4세기 아서(Arthur) 왕의 치하에 있던 잉글랜드이다. 일반적인 로망스와는 달리 아서 왕의 조카인 가웨인의 모험담은 간략히 소개되고, 가웨인이 녹색 기사와의 대결에서 녹색 허리띠의 마법을 이용하여 자신의 목숨을 부지하는 과격적인 결말로 이끌어간다. 비록 비겁하더라도 목숨을 이어가는 것이 중요하고 이를 초월적인 존재인 녹색 기사가 인정하는 것이 주요 내용이다[4].

서사시의 녹색 기사가 거주하는 워렐(Wyrale) 숲은 서사시를 집필했던 14세기 당시 잉글랜드의 지배자였던 리처드 2세의 고향인 체스터(Chester)를 상징하고, 당시 상당한 오지였던 체스터 특유의 야생의 기질을 녹색 기사가 상징한다. 리처드 2세의 지원을 받아 제 2의 도시로 성장해 가는 체스터를 시기했던 런던의 호사가들은 세련되지 못한 지방 문화를 공격했고 이에 대한 반론의 의미를 서사시가 담고 있다. 서사시에서 녹색 기사가 거주하는 대저택의 실내 장식이나 그곳에서 벌어지는 세련되고 사교적인 의례를 자세히 묘사하고 있기 때문이다. 런던에서 워렐 숲으로 찾아온 가웨인을 녹색 기사가 인정해주고 격려하는 내용을 가지게 된 연유에는 이러한 당시의 사정이 있었다[5].

영화에서 가웨인은 철없는 청년으로 묘사되는데, 영화에서 '당신은 기사인가?'라는 질문을 끊임없이 받게 된다. 그의 대답은 언제나 '아직은 아니다.' 이다. '기사'가 되기 위해 숭고한 모험담이 필요한 상황에서 녹색 기사와 맞서게 되었지만, 녹색 기사를 만나기 위해 녹색 예배당을 찾아가는 여정에서 주인공이 겪게 되는 다양한 시련을 기사답게 맞서지 못한다. 오히려 어리둥절한 표정과 당황스러운 태도로 이를 대한다. 이는 일반적인 소년/청년이 성장하는 과정에서 사회가 요구하는 어른스러움을 수용하려고 하지만 제대로 그것을 이해하지도 못하고 적절하게 대처하지도 못하는 모습을 그

리는 것으로 이해할 수 있다. 그렇게 어리숙한 주인공이 녹색 허리띠라는 마법적인 물건을 이용해 자신의 목숨을 부지하고 비루한 인생을 견뎌내지만 결국은 이미 목이 잘린 것과 동일하다는 깨달음을 얻는 지난한 과정을 그리고 있다. 따라서 영화는 원작 서사시와 유사한 스토리를 가지면서도 소년에서 어른으로 성장해가는 과정의 지난함, 그 과정의 부조리함, 그것을 깨달은 자가 자신의 운명을 담담히 받아들이는 태도와 같은 주제를 그리고 있다. 정리하자면 이 영화는 성장의 부조리를 깨달아가는 과정에 대한 은유라고 할 수 있다[6].

영화에서는 녹색 예배당으로 가는 여정에서 4가지 에피소드를 그리고 있는데, 강도 소년(Scavenger), 목없는 소녀(위니프레드), 어딘가로 줄지어 걸어가는 거인들과의 만남, 대저택에서 성주와의 기이한 교환 게임이 그것이다. 제대로 보상을 받지 못했다고 가웨인을 공격하는 강도 소년, 잘린 머리를 찾아주면 어떤 보상을 하겠냐고 묻는 가웨인에게 그것이 무슨 상관이나고 되묻는 소녀 유령 위니프레드, 어깨에 태워가 달라는 가웨인의 요청에 손길을 내미는 신화적인 거인, 사냥감과 저택 내에서 얻은 것을 교환하자는 성주의 제안은 녹색 기사와 함께 ‘교환’의 문제에 대해 이야기하고 있다. 적절한 보상(강도 소년), 무조건적인 원조(위니프레드), 도움을 거절하는 두려움(거인), 목숨을 걸고 지키는 약속(성주 혹은 녹색 기사)을 의미하는 각 에피소드를 통해 일반적인 교환의 문제가 얼마나 복잡한 상황을 내포하고 있는지 잘 표현하고 있다.

현대 자본주의 사회에서 자본/상품의 교환은 사회의 운영/구성에 있어 핵심 문제이다. 사회에서 생활한다는 것은 쉽지는 교환 행위를 무한히 반복하는 것이다. 무의식적 차원에서 일어나는 교환 행위에서 약탈적 보상의 문제, 무상의 원조에 대한 문제, 원조 수용의 공포까지 포함될 수 있다[7]. 그 모든 중요하고도 목숨을 걸어야 하는 교환의 행위에 대해 녹색 허리띠를 밟고 목을 내밀었다가 도망친 주인공이 결국에는 별것 아닌 왕좌에 앉아 잘린 머리가 떨어지게 되는 최후를 맞이하는 것이 인생이라는 깨달음. 이러한 관점에서 녹색 허리띠는 현대 자본주의 사회의 ‘자본’이다. 그 자본을 일격에 끊어내는 것이 녹색 기사의 도끼라면, 이 도끼의 일격은 도무지 벗어날 수 없을 것처럼 느껴지는 자본주의의 교환 논리를 초월하는 깨달음의 순간이다. 따라서 녹색 기사와의 마지막 대면에서 가웨인이 돌려받는 도끼의

일격은 ‘성장/인생의 부조리함에 대한 깨달음’이라고 영화는 말하고 있다.

III. Result and Analysis

1. Language Model

연구에 사용된 언어모델은 LSTM 방식이다. Embedding - LSTM1 - LSTM2 - Dense Layer 로 구성된 언어모델로 단어 임베딩에 필요한 단어벡터의 차원은 1,000개이고, 두 개의 LSTM 층에서 은닉층의 개수는 동일하게 1,000개이다. 각 LSTM 층에는 드롭아웃(Drop-out)을 60% 적용하여 과적합을 방지했다. Dense는 완전연결(Fully Connected) 혹은 밀집(Dense) 층이다. 학습을 위해 입력되는 시퀀스 길이는 10으로 고정하였다. 이 경우 전체 매개변수는 21,478,734개이다.

표 1. 언어모델의 구조

Table 1. The Structure of Language Model

Layer Structure		
Embedding	(vocab_size, wordvec_size)	
LSTM1	(hidden_size, dropout=0.6)	
LSTM2	(hidden_size, dropout=0.6)	
Dense	(vocab_size, activation='softmax')	
Model Summary		
Layer	Output Shape	Parameters
Embedding	(None, 10*100, 1000)	2,734,000
LSTM1	(None, 10*100, 1000)	8,004,000
LSTM2	(None, 1000)	8,004,000
Dense	(None, 2734)	2,736,734
(Total parameters : 21,478,734)		
Hyperparameters		
vocab_size	2,734	Number of vocabulary
wordvec_size	1,000	Dimension of word vectors
hidden_size	1,000	Dimension of hidden
time_size	10	Length of sequence
corpus_size	8,591	Number of corpus

파이썬 아나콘다 플랫폼에서 텐서플로 라이브러리를 사용하여 모델을 학습시켰다. 사용된 컴퓨터의 제원은 Windows 11, Intel Xenon 2.11GHz, RAM 64GB, NVIDIA RTX A2000 12GB (GPU) 이다. 100 에폭(Epoch) 수준의 학습에 5분 내외의 시간이 소요되어 반복적인 학습 및 평가에 적합하다. 에폭(Epoch)은 주어진 학습용 데이터를 모두 사용하는 횟수를 의미하기 때문에 100에폭이면 주어진 텍스트를 100번 반복하여 학습시켰음을 의미한다.

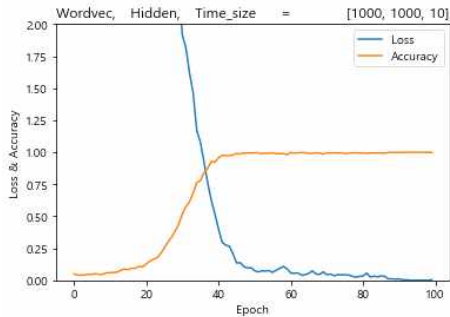


그림 1. 언어모델 학습 커브
Figure 1. Training Curve for LM

학습용 말뭉치로 영화 ‘Green Knight’의 영어로 된 대본을 사용하였다. 불용어(Stop-word) 처리를 하여 말뭉치 8,591개, 어휘(Vocabulary) 2,734개로 구성되어 있다. 일반적인 언어모델 학습을 위한 용도로는 부족한 크기이지만 앞에서 언급했듯이 개인 연구자가 모델 구성/학습/평가를 위한 용도로는 적절한 크기이다. 이러한 모델의 구성과 특성을 Table 1에 정리하였다.

이상의 말뭉치와 언어모델을 이용하여 학습한 결과 그래프가 Fig. 1이다. 약 100에폭 수준에서 적절하게 학습되는 양상을 파악할 수 있다. 이렇게 학습된 언어모델의 성능을 평가하기 위해 앞에서 제시한 주요 주제어들의 단어벡터를 활용한다[8]. 1,000 차원의 단어벡터는 말뭉치로 입력된 텍스트의 내용에 맞는 다양한 의미들을 표현하고 있다. 따라서 주요 주제어들의 단어벡터를 활용하여 그 벡터의 방향과 유사한 다른 단어들을 살펴볼 수도 있고, 차원 감소(Dimension Reduction) 연산을 통해 단어들을 2차원 평면에 도시하여 그 위치를 판단할 수도 있다.[9] 이러한 단어벡터를 활용하여 텍스트를 분석할 수 있다.

2. Word Vector - Similarity

1,000개 차원의 단어벡터를 활용하여 주요 주제어들과 벡터의 방향이 유사한 상위 5개 단어를 Table 2에 정리하였다. 일반적인 벡터의 코사인 유사도(Cosine Similarity) 계산법을 적용하여 단어 사이의 유사도를 구했다[10]. 다소 이해하기 어려운 단어들도 일부 포함되어 있으나 텍스트를 살펴보면 어느 정도 일관성이 있는 단어들이다.

녹색(green)의 경우 거역할 수 없는 자연의 힘, 부조리한 인생을 깨닫게 하는 영역이라고 영화에서 묘사되는데, ‘가까운(closer)’, ‘기사(knight)’, ‘용감한(brave)’, ‘당기다(pulls)’, ‘성내(courtyard)’ 등의 단어가 추천되었다. 거역할 수 없는 자연/죽음과 같은 ‘녹색’의 힘이 생각보다 인생에서 ‘가깝고’, ‘용감한’ ‘기사’들도 피해 가지 못하고 오히려 ‘끌어당길’ 수 있으며, 이미 안전한 ‘성내’에 도사리고 있다, 라고 이해할 수 있다.

목숨을 지켜주는 마법의 도구인 허리띠(girdle)에 대해서는 ‘땅(ground)’, ‘불(fire)’. ‘쥐다(grab)’, ‘부담(load)’, ‘새장(cage)’ 등의 단어가 추천되었다. 비록 허리띠를 매고 안전할 수는 있어도 그것은 영화에서 인상적으로 등장한 새장(cage)에 갇힌 시체와 같이 부담스러운(load) 것이고, 땅(ground)이나 불(fire)과 같은 마법적인 요소를 가웨인의 어머니(마녀)가 움켜쥐고(grab) 있는 것이라고 이해할 수 있다.

다소 억지스러울 수 있지만 유사도 추천을 통해 텍스트의 의미를 한번 더 깊게 감상할 수 있음을 알 수 있다. 주요 키워드의 상징과 의미를 다양한 방식으로 유추할 수 있어 언어모델의 단어벡터는 텍스트 분석에 대해 객관적이고 과학적인 기법을 제안한다고 할 수 있다[11].

표 2. 키워드의 유사도 연산 결과
Table 2. Similarity Result for the Keywords

scavenger company: 0.483 hunting: 0.4779 leans: 0.4523 king: 0.4353 blood: 0.431	green closer: 0.3806 knight: 0.3464 brave: 0.3215 pulls: 0.3162 courtyard: 0.316
honor hoof: 0.114 stands: 0.1048 trot: 0.0933 continuous: 0.092 wounded: 0.0914	girdle ground: 0.7148 fire: 0.7121 grabs: 0.6949 load: 0.6788 cage: 0.678
game girdle: 0.5974 winter: 0.5856 cut: 0.5811 load: 0.5767 fire: 0.5741	chapel winter: 0.5531 pass: 0.5406 moment: 0.54 slats: 0.5346 servants: 0.5325
mother moment: 0.265 friend: 0.2644 opportunity: 0.257 arm: 0.254 cold: 0.2471	axe castle: 0.5594 stumbles: 0.5018 trees: 0.4882 feel: 0.4657 two: 0.4589

3. Word Vector - Visualization

벡터의 차원 감소 연산을 통해 1,000차원의 단어벡터를 2차원의 단어벡터로 변환할 수 있다[12]. 2차원 벡터의 경우 2D 평면에 도시할 수 있는데, 영화 텍스트의 주요 키워드를 도시한 것이 Fig. 2 이다. 단어들의 군집 상태를 살펴보면 '기사/왕비/위니프레드'가 모여 있고, '에셀/게임', '어머니/키스', '가웨인/왕/강도소년', '녹색/거인/머리/명예'가 밀집되어 있다.

'기사'의 모험담을 강조하는 '왕비'와 기사의 도리를 다하라는 목없는 유령 소녀 '위니프레드'는 유사한 역할을 수행한다. 녹색 기사와의 '게임'에 대해 질문하는 가웨인의 연인 '에셀', 아직 유아적인 가웨인이 크리스마스 이브에 '어머니'에게 사랑의 '키스'를 전달하고, 게임의 중요성을 강조하는 '왕'과 충분한 보상을 요구하는 '강도소년'의 분위기는 비슷하다. 마법의 '허리띠'를 끊어내는 '도끼'는 충분한 거리를 두고 떨어져 있어 상반되는 의미를 잘 나타내고 있다.

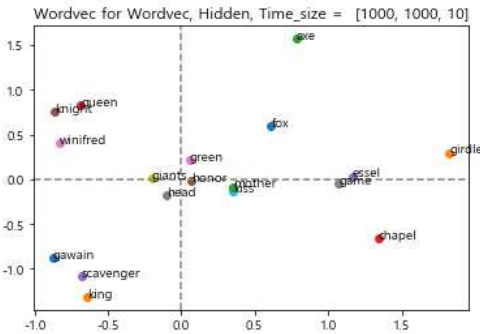


그림 2. 차원감소된 단어벡터 도시
 Figure 2. Display of the wordvectors in 2-dimension

이상과 같이 평면 도시를 통해 주요 키워드의 위치를 기준으로 텍스트의 상징과 의미를 분석할 수 있어 단어벡터를 활용할 수 있는 기법이 될 수 있다.

IV. Conclusion

언어모델의 단어벡터를 활용하여 텍스트의 의미를 분석할 수 있는 융복합적인 기법을 제안하였다. 단어벡터의 유사도 연산을 통해 주요 키워드 사이의 친연성을 분석하였고, 2차원 수준으로 차원감소한 단어벡터를 활용하여 키워드의 군집 상태를 도시하여 키워드 사이의

관계를 직관적으로 인식할 수 있었다.

이러한 유사도 및 평면 도시법은 모두 단어벡터를 이용하는 것인데, 언어모델 학습과정에서 단어벡터는 임베딩 과정을 통해 결정된다. 임베딩은 단어를 정해진 차원의 벡터로 초기화하는데, 이때 벡터는 무작위 숫자들로 구성된다[13]. 학습과정에서 벡터는 입력되는 텍스트 말뭉치의 맥락(context)을 확률적으로 예측/표현할 수 있는 숫자들로 변화하게 된다. 이렇게 결정된 단어벡터는 말뭉치를 구성하는 어휘(Vocabulary)에 해당하는 각 단어를 여러 개의 숫자로 표현하게 된다. 이 단어의 표현법(기표)은 벡터이고, 그 벡터를 구성하는 숫자들에 의해 그 의미(의기)가 결정되는 것이다.

이는 소쉬르(Saussure)의 기호이론과 유사한 측면이 있다. 소쉬르 기호학에 의하면 언어는 기호들의 관계망이며, 기호(Sign)를 구성하는 기표(Signifier)와 의기(Signified)는 서로 특별한 관계가 없이 전체 기호의 집합에서 차이만을 가지는 형식으로 기표가 결정된다. 이러한 기표들은 통시적(Diachronic) 혹은 공시적(Synchronic) 친연성에 의해 연쇄 혹은 맥락을 형성한다. 이러한 기호체계에서 시간/순서 상 적절하게 기표가 나열되기도 하고(통시성), 유사한 분위기/느낌/의미를 가지는 기표들로 나열되기도 하면서(공시성) 맥락을 구성한다. 이 맥락을 텍스트라고 부르고, 이러한 텍스트를 인간이 인식하면 다양한 이미지와 정동(Affection)을 맛보게 되는 것이다[14].

언어모델에 의해 결정된 단어벡터는 소쉬르의 기표를 정확하게 표현하고 있다. 단어벡터 역시 무작위 숫자들로 초기화되었다가 주어진 맥락을 구성하기 적합한 숫자들로 학습/변환 과정을 통해 결정되기 때문이다. 각 어휘에 해당하는 단어벡터는 서로 간의 차이가 중요하고, 입력된 맥락을 연쇄시키기 적합한 숫자들로 구성되는 것이다. 따라서 기의를 상징하는 벡터의 숫자들로 연산된 유사도 높은 단어를 살펴보고, 단어들의 군집 상태를 가시화하면 주어진 텍스트를 정확하고 객관적으로 분석하는 것이 가능하다.

본 연구에서는 영화 'Green Knight'의 영어 대본에 대해 단어벡터 분석법을 적용하여 주요 키워드의 의미를 살펴보았다. 유사도 연산 및 2D 가시화를 통해 키워드 사이의 관계와 의미를 객관적으로 분석할 수 있음을 알 수 있었다. 이를 토대로 영화 텍스트에 대한 분석기법을 제안하였다.

References

- [1] K. Hyungsuc, Y. Janghoon, "Analyzing Semantic Relations of Word Vectors trained by The Word2vec Model", *Journal of KIISE*, 46(10), pp. 1088-1093, 2019
- [2] L. Hickman, et al., "Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations", *Organizational Research Methods*, 25(1), pp.114-146, 2022
- [3] N. Fatima, et al., "A Systematic Literature Review on Text Generation Using Deep Neural Network Models", *IEEE Access*, 10, 53490 - 53503. 2022
- [4] D. Lee, "Shame and Guilt As Sanctions Controlling Gawain's Behavior", *Journal of British & American Studies*, 59, pp. 87-112, 2023
- [5] J. Kang, "The Geographical Imagination in Sir Gawain and the Green Knight: The Wilderness of Wirral", *Medieval and Early Modern English Studies*, 23(1), pp. 1-29, 2015
- [6] M. Rasmussen, "Tweaking the Tradition: Gawain as Perceval in David Lowery's *The Green Knight*", *Arthuriana*, 34(20), pp. 62-78, 2024
- [7] M. Eden, "Representing Sir Gawain and the Green Knight", *Arthuriana*, 34(2), pp. 16-61, 2024
- [8] Q. Yanan and T. Fuqiang, "Keyword Extraction for Film Reviews Based on Social Network Analysis and Natural Language Technology," *Journal of Technology and Information*, 2022, DOI:10.1051/e3sconf/202018903019
- [9] S. H. A. Latif, A. S. Alwan and A. M. Mohamed, "Principal Component Analysis as Tool for Data Reduction with an Application," *EUREKA: Physics and Engineering*, vol.5, pp.184-198, 2022, DOI:10.21303/2461-4262.2022.002577
- [10] F. Heimerl, M. Gleicher, "Interactive Analysis of Word Vector Embeddings", *Computer Graphics Forum*, 37(3), pp. 253-265, 2018
- [11] Z. Bingyu and N. Arefyev, "The Document Vectors Using Cosine Similarity Revisited," *Proceedings of the Third Workshop on Insights from Negative Results in NLP*, pp.129 - 133, Association for Computational Linguistics, 2022, DOI:10.18653/v1/2022.insights-1.17
- [12] K. Kwangho, et al., "Input Dimension Reduction based on Continuous Word Vector for Deep Neural Network Language Model," *Phonetics and Speech Sciences*, 7(4), pp. 3 - 8, 2015
- [13] K. Kusum and S. Panda, "Sentiment Analysis Using Global Vector and Long Short-term Memory," *International Journal of Electrical and Computer Engineering (IJECE)*, 2022, DOI:10.11591/ijeecs.v26.i1.pp414-422
- [14] Z. Wasik, "Uncovering the two conceptions of the linguistic sign in Saussure's lectures: An epistemological inquiry with comments on translational equivalence", *Sign Systems Studies*, 51(3/4), pp. 513-537, 2023