

<http://dx.doi.org/10.17703/JCCT.2021.7.3.591>

JCCT 2021-8-71

딥러닝을 이용한 기형도 시의 핵심 이미지 분석

Deep Learning Application for Core Image Analysis of the Poems by Ki Hyung-Do

고광호*

Kwang-Ho Ko*

요약 전후방 단어들의 인접 여부 혹은 후방 단어들의 순서를 학습할 수 있는 통계 기법인 SVD, 딥러닝 기법인 CBOW, LSTM으로 단어벡터를 구할 수 있다. 이렇게 학습된 단어벡터를 기형도의 시에 적용하여 핵심 이미지를 대표하는 단어들과 유사도 높은 단어를 구해서 분석해 보았다. 시적 이미지와 어울리지 않는 단어들이 연산되기도 하지만 그 단어가 사용된 시적 맥락에서는 기준 단어와 유사한 이미지를 표현하고 있음을 알 수 있었다. 이러한 단어벡터를 활용하면 핵심 이미지를 대표하는 단어들의 관계와 유사한 관계의 다른 단어들도 유추할 수 있다. 따라서 통계 기법인 SVD 및 딥러닝 기법인 CBOW와 LSTM으로 구한 단어벡터의 유사도 및 유추 연산을 통해 대상 시를 다양하고 심도 깊게 분석할 수 있다.

주요어 : 딥러닝, 단어벡터, 기형도, 유사도, 유추, 이미지

Abstract It's possible to get the word-vector by the statistical SVD or deep-learning CBOW and LSTM methods and these ones learn the contexts of forward/backward words or the sequence of following words. It's used to analyze the poems by Ki Hyung-do with similar words recommended by the word-vector showing the core images of the poetry. It seems at first sight that the words don't go well with the images but they express the similar style described by the reference words once you look close the contexts of the specific poems. The word-vector can analogize the words having the same relations with the ones between the representative words for the core images of the poems. Therefore you can analyze the poems in depth and in variety with the similarity and analogy operations by the word-vector estimated with the statistical SVD or deep-learning CBOW and LSTM methods.

Key words : Deep-learning, Word-vector, Ki Hyung-do, Similarity, Analogy, Image

1. 서론

딥러닝을 이용하여 다양한 언어모델을 구축할 수 있다. 일반적으로 말뭉치를 구성하는 단어의 순서나 인접 여부를 기억하는 방식으로 언어모델을 구축하는데, 학습이

종료되면 일종의 부산물로서 각 단어의 분산표현(단어 벡터)을 얻을 수 있다[1]. 단어벡터를 구할 수 있는 기법으로는, 통계 기반의 SVD, 딥러닝 기반의 CBOW 및 LSTM 기법 등이 있다. 이러한 기법으로 구한 단어벡터를 사용하면 특정 단어와 유사도가 높은 단어를 얻을

*정희원, 평택대 스마트자동차학과 부교수 (제1저자)
접수일: 2021년 8월 2일, 수정완료일: 2021년 8월 7일
게재확정일: 2021년 8월 9일

Received: August 2, 2021 / Revised: August 7, 2021
Accepted: August 9, 2021
*Corresponding Author: kwangho@ptu.ac.kr
Dept. of Smart Automobile, Pyeongtaek Univ, Korea

수 있고, 또한 특정 단어 사이의 관계를 기준으로 원하는 단어와 동일한 관계를 형성하는 다른 단어를 유추할 수도 있다[2].

본 연구에서는 잘 알려진 문학 작품을 말뭉치로 사용하여 단어벡터를 구한 후 작품의 핵심 이미지와 유사한 단어를 단어벡터로 찾아 그 타당성을 평가해 보았다. 또한 작품의 핵심 단어 사이의 관계를 정의하고 그 관계와 유사한 다른 단어를 단어벡터로 구했다. 역시 이 단어들이 작품에서 드러난 독특한 관계와 어울리는지 평가해 보았다.

이러한 분석을 위한 문학 작품으로 기형도의 시를 선정했는데, 기형도 시인은 약 90편의 시가 실린 유고 시집 한 권만 남기고 30세의 나이로 1989년에 요절했다 [3]. 따라서 기형도의 시는 분량이 그리 많지 않고 그 핵심 이미지가 많은 선행 연구에 의해 잘 정리되어 있어 본 연구에 적절한 대상 작품이라고 판단했다. 또한 SVD, CBOW 및 LSTM 기법으로 구한 단어벡터를 기형도의 시에 적용하여 그 이미지와 유사도가 높은 단어 및 동일한 관계로 유추한 단어를 찾아 그 타당성을 각각 평가하였다. 이러한 분석을 통해 단순한 인접 여부 및 단어 순서를 기준으로 계산되는 단어벡터를 문학 작품에 적용하는 기법을 제안해 보았다.

II. 본 론

1. 단어벡터 : SVD, CBOW, LSTM

말뭉치에 대한 대표적인 딥러닝 기법으로는 CBOW (Continuous Bag of Words)와 LSTM(Long Short-Term Memory) 등이 있다. CBOW의 경우 단어의 전후방에 인접한 단어들을 추정할 수 있도록 학습한다. LSTM의 경우에는 단어의 후방에 이어지는 단어들의 순서를 학습한다. 즉, CBOW의 경우 단어의 순서와 관계없이 전후방의 인접 여부만 학습하고, LSTM에서는 후방에 이어지는 단어의 순서를 기억하는 것이다.

통계 기반의 SVD(Singular Value Decomposition) 기법도 있는데, 이는 전후방 단어의 인접 여부를 표현한 동시발생행렬(Co-occurrence Matrix)을 이용한다. SVD는 동시발생행렬에 확률 개념을 도입하여 계산된 PPMI(Positive Point-wise Mutual Information)를 직교/대각 행렬로 특잇값 분해한 후 차원 감소 연산을 적용하여 얻어낸 단어벡터를 이용한다[4]. CBOW와 LSTM

에서 얻은 단어벡터와 유사한 방식으로 SVD의 단어벡터 역시 유사도 높은 단어 및 동일한 관계의 단어를 유추하는데 사용할 수 있다. 이상의 대표적인 언어모델 구축 기법인 SVD, CBOW 및 LSTM의 단어벡터를 간략하게 살펴보면 다음과 같다.

우선 SVD의 경우, 단어 x 와 y 가 인접할(동시 발생할) 확률 개념을 적용하여 얻을 수 있는 PPMI에 특잇값 분해 및 차원 감소 연산을 적용하여 얻은 U_2 을 단어벡터로 사용한다.

$$PMI(x,y) = \log_2 \frac{P(x,y)}{P(x)P(y)}$$

$$PPMI(x,y) = \max(0, PMI(x,y))$$

$$PPMI = U \cdot S \cdot V^T$$

$$U_2 = U[:, :n]$$

위 식에서 P 는 확률을, S 는 대각행렬을 의미한다. 또한 직교행렬 U 에서 n 번째 열의 값만 취하여 단어벡터로 사용한다. 동시발생행렬의 경우 대부분의 원소의 값이 제로이기 때문에 직교행렬 U 의 앞쪽 열의 값이 크고 후방으로 갈수록 작아지므로 이를 차원 감소시킨 U_2 를 단어벡터로 사용하는 것이 효율적이기 때문이다.

두 번째로 CBOW의 경우 그림 1과 같은 딥러닝 계층을 통해 학습시켜서 단어벡터를 구한다. 말뭉치를 구성하는 단어의 앞뒤에 인접하는 단어에 대해서는 라벨 1을, 그 외의 단어에 대해서는 라벨 0를 부여하여 이진 분류(binary classification)를 적용한다. 즉 라벨에 따라 교차엔트로피오차(CEE, Cross Entropy Error)가 최소가 되는 매개변수(W_{in} , W_{out})를 경사하강법으로 계산한다. 학습 종료 후 입력 맥락(context)에 대한 임베딩(embedding) 가중치 W_{in} 을 단어벡터로 사용한다. 은닉층 결과(Hidden), 딥러닝 점수(Score) 및 손실함수(Loss)는 다음의 수식으로 계산한다.

$$Hidden = Average(Contexts \cdot W_{in})$$

$$Score = Hidden \cdot (Target \cdot W_{out})$$

$$Loss = CEE(Label, Sigmoid(Score))$$

마지막으로 LSTM의 경우 그림 2와 같은 딥러닝 계층을 구성하여 학습시키고, 역시 입력 단계의 임베딩

가중치를 단어벡터로 사용한다. LSTM에서는 시간 단위 (time size)에 해당하는 연속된 단어들을 순차적으로 임베딩한 결과에 LSTM 연산을 수행하여 내부적으로 사용되는 메모리셀 값과 다음 계층으로 전달되는 은닉층 값을 연산한다. 그 연산식은 그림 3에 정리하였다.

특히 LSTM 연산에 드롭아웃(Dropout)을 적용하여 과적합(overfitting) 문제를 예방하였고, 최종적으로는 어파인(Affine) 연산의 결과에 소프트맥스(Softmax)를 적용하여 확률을 계산하고, 이 확률값과 라벨을 비교하여 계산된 교차엔트로피오차를 손실함수로 이용한다. 이 손실함수의 값이 최소화되도록 경사하강법을 적용하여 매개변수를 최적화시킨다. 이상의 단어벡터를 연산하는 각 기법들의 특징을 표 1에 정리하였다.

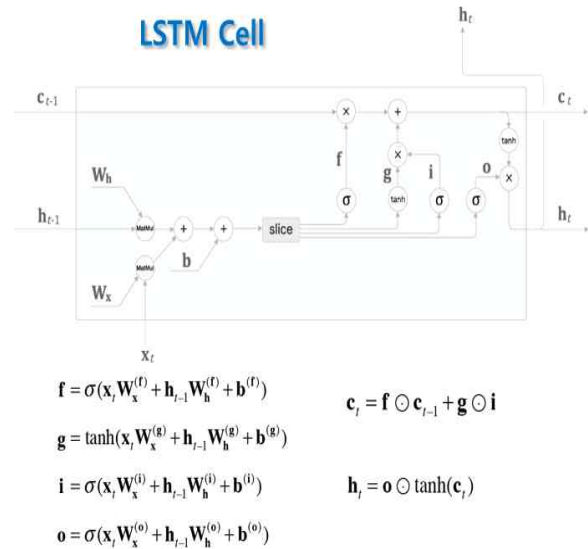


그림 3. LSTM 셀의 연산
 Figure 3. Calculations of LSTM Cell

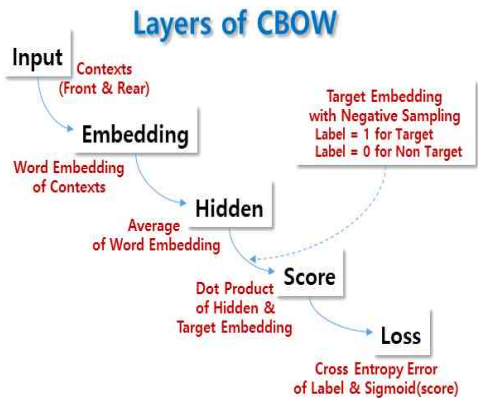


그림 1. CBOW 모델 계층
 Figure 1. Layers of CBOW Model

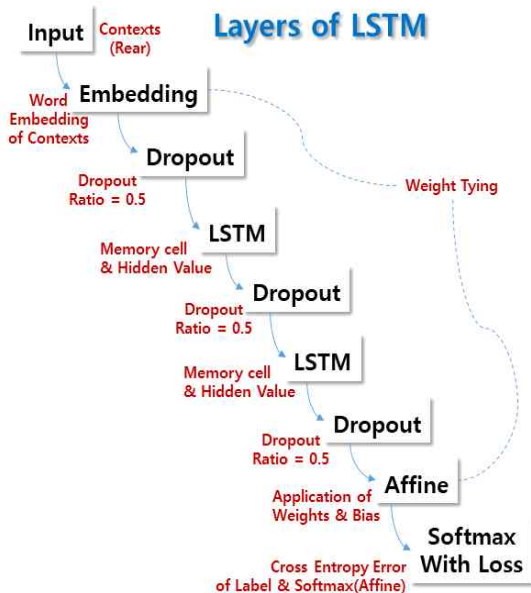


그림 2. LSTM 모델 계층
 Figure 2. Layers of LSTM Model

표 1. 단어벡터 기법의 특징

Table 1. The Characteristics of Word -vector Methods

Method	SVD	CBOW	LSTM)
Window size	5	5	10 (time size)
Wordvec size	40	150	150
Founda-tion	Statistics	Deep Learning	Deep Learning
Charac-teristics	PPMI Dimension reduction	Word embedding Binary classification	Word embedding 2 LSTM layers Dropout rate = 0.5 Weights tying

앞에서 살펴본 SVD, CBOW, LSTM을 적용하기 위해 기형도의 유작 시집(『입 속의 검은 잎』 (기형도, 문학과지성사, 1989년))을 말뭉치로 사용하였다. 응집확률 (cohesion probability) 기법을 적용하여 체언과 조사를 분리하고, 각종 구두점, 줄바꿈 기호 등을 삭제하여 말뭉치로 사용하였는데, 말뭉치를 구성하는 단어의 수가 11,346개이고, 어휘 수는 4,447개 이다. 이 말뭉치에 대해 단어벡터는 각 어휘(4,447개)에 해당하는 벡터 값을 의미하며, 이 벡터를 이용하여 코사인 유사도(cosine similarity) 및 벡터 연산을 적용할 수 있다. 원하는 단어와 유사도가 높은 단어를 구하거나, 원하는 관계와

동일한 관계에 있는 단어를 연산할 수 있는 것이다.

2. 기형도 시의 핵심 이미지

기형도는 「위험한 가게 1969」, 「삼촌의 죽음」, 「폭풍의 언덕」, 「엄마 생각」 등의 작품에서 유년 시절 경제적인 궁핍함과 가족들의 죽음(아버지, 삼촌, 작은 누나의 죽음)을 묘사하고 있다. 이러한 시들을 통해 기형도는 80-90년대의 가난, 우울함, 죽음 등을 상징하는 시인으로 유명세를 탔으며, 특히 1989년 그는 현실에서 죽음을 맞이하였고, 그 죽음이 그의 시적 정서와 맞물리며 90년대 초반 기형도는 그로테스크한 죽음의 미학을 표현했다는 평을 받고 있다[5].

기존의 기형도 시에 대한 연구를 살펴보면, 도정(道程)의 이미지가 강하다는 연구[6]; 안개나 강, 빗물처럼 흐르는 유체와 벽처럼 고정된 고체의 대조적인 이미지 형성에 대한 연구[7]; 소외된 주체가 길을 나서는 유목민적 이미지를 들뢰즈의 탈영토화 개념으로 접근한 연구[8]; 개인적인 내면을 위주로 전개된 시와 사회참여적인 내용의 시로 구분하여 두 가지 성격의 주체가 공존함을 밝힌 연구[9]; 기형도 시에 드러난 비극적인 구성 요소에 대한 분석[10]; 기형도 시를 심리치료에 활용하는 방안에 대한 연구[11] 등이 있다.

이러한 기존의 연구 결과를 기반으로 우선 기형도가 유년시절을 보낸 광명시 소하리에 대해 검토해보면, 기형도의 시의 핵심적인 이미지에 대해 쉽게 접근할 수 있다. 소하리는 전형적인 변두리 시골에서 공장 건립과 함께 제조업 중심의 공업도시로 변해간 지역이라고 할 수 있다. 서울을 중심으로 주변부에 형성되는 위성도시 혹은 배후도시의 대부분 이러한 공업화 과정을 거치면서 농촌에서 도시로 변모해가는 양상을 일반적으로 보인다. 따라서 서울 주변부의 도시들은 소규모 공동체 마을 중심의 촌락에서 중소 제조업 중심의 도시로 변모해가면서 거의 특색이 없는 무명(無名)의 도시화를 겪었다고도 할 수 있을 것이다. 안양천이 소하리를 관통하는데, 기형도는 ‘안개’라는 시에서 소하리 주변 도시화의 스산함을 묘사하였다. (본문에 등장하는 모든 시는 『입 속의 검은 잎』(기형도, 문학과지성사, 1989년)에서 인용하였다.)

아침저녁으로 셋강에 자욱이 안개가 낀다.
이 읍에 처음 와본 사람은 누구나

거대한 안개의 강을 거처야 한다.

앞서간 일행들이 천천히 지워질 때까지
쓸쓸한 가족들처럼 그들은

그 긴 방죽 위에 서 있어야 한다.

문득 저 홀로 안개의 빈 구멍 속에 갇혀 있음을 느끼고 경악할 때까지.

(……)

몇 가지 사소한 사건도 있었다.

한밤중에 여직공 하나가 겁탈당했다.

(……)

지난 겨울엔 방죽 위에서 취객 하나가 얼어 죽었다. 바로 곁을 지난 삼륜차는 그것이 쓰레기 더미인 줄 알았다고 했다.

그러나 그것은 개인적인 불행일 뿐, 안개의 탓은 아니다.

(……)

안개는 그 읍의 명물이다.

누구나 조금씩은 안개의 주식을 갖고 있다.

여공들의 얼굴은 희고 아름다우며

아이들은 무럭무럭 자라 모두들 공장으로 간다.

-「안개」

안양천 주변에 아침저녁으로 끼는 안개 속을 지나가는 여공들과 취객들의 모습을 애잔하고 무심하게 그리면서도, ‘안개의 주식’을 가진 ‘여공들’의 ‘아이들은 공장으로’ 가게 되는 자본주의의 무정함도 잘 표현하고 있다. 따라서 기형도의 시는 자본주의의 황폐함, 도시 빈민들의 애잔함에 어울리는 정서를 잘 표현했다고 할 수다.

또한 이렇게 배후도시로 밀려난 사람들이 자본주의와 관료주의의 사회에서 생활하며 다양한 우울감을 겪게 되는데 이러한 인간형은 카프카적이다. 카프카적인 인간형은 어떤 목표를 상실했지만 높은 공감능력 때문에 아름답고도 어리석은 인간이라고 벤야민이 묘사한 인간형이다[12]. 카프카적인 인간을 기형도는 애잔하고도 무심하게 표현했다.

눈은 퍼부었고 거리는 캄캄했다

움직이지 못하는 건물들은 눈을 뒤집어쓰고

희고 거대한 서류뭉치로 변해갔다

무슨 관공서였는데 희미한 불빛이 새어나왔다

유리창 너머 한 사내가 보였다
그 좁고 큰 방에서 서기는 혼자 울고 있었다!
(.....)
그리고 나는 우연히 지금 그를 떠올리게 되었다
밤은 깊고 텅 빈 사무실 창밖으로 눈이 퍼붓는다
나는 그 사내를 어리석은 자라고 생각하지 않는다

-「기억할 만한 지나침」

눈이 내리는 어두운 밤 관공서의 사무실에서 홀로
앉아 울고 있는 남자를 지켜보았던 나는 어느 날 그 남
자를 떠올리고 그를 ‘어리석은 자라고 생각하지 않는다’
고 말한다. 관료주의적 자본주의 체제하에서 고통 받는
남자의 처지가 자신의 처지와도 비슷하다는 연민과 공
감의 감정을 카프카적인 주체인 화자가 느끼는 것이다.
도시민들은 고향을 오래전에 떠나온 실향민에 가까
운 정서를 갖고 있다. 핵가족화 되고 마을 공동체가 사
라지는 환경에서 어릴 적 고향과 가족에 대한 향수를
간직하기도 할 것이다. 이러한 가족에 대한 그리움도
기형도는 잘 표현하고 있다.

열무 삼십 단을 이고
시장에 간 우리엄마
안 오시네, 해는 시든 지 오래
나는 찬밥처럼 방에 담겨
아무리 천천히 숙제를 해도
엄마 안 오시네, 배추잎 같은 밭소리 타박타박
안 들리네, 어둡고 무서워
금간 창 틈으로 고요히 빗소리
빈방에 혼자 엎드려 훌쩍거리던
아주 먼 옛날
지금도 내 눈시울을 뜨겁게 하는 그 시절, 내 유년의
윽목

-「엄마 생각」

열무, 배추, 시든 해의 이미지로 고된 어머니의 삶을,
찬밥과 혼자 훌쩍이는 아이의 이미지로 불우한 유년시
절을 형상화하고 있다. 숙제를 하며 일 나간 엄마를 기
다리는 소년의 안타까움은 도시에서 유년 시절을 보낸
사람들이라면 쉽게 공감할 수 있을 것이다.

이렇게 기형도의 시는 공업화되어 가는 배후도시를

이미지화 하고, 관료체계 하의 카프카적인 인간의 괴로
움을 울음으로 표현하기도 하며, 유년 시절의 고달픔과
그리움을 애잔하게 표현하기도 했다. 정리하면, 배후도
시, 공감능력, 유년시절에 대한 향수 등이 기형도 시의
핵심 이미지라고 할 수 있다.

3. 단어 백터 적용 : 유사도 및 유추

기형도 시의 핵심 이미지로 파악되는 ‘공장화 및 도
시화’에 대해서는 ‘도시’와 ‘공장’을 대표적인 시어로 선
정하였다. ‘카프카적 인간의 우울감’을 대표하는 시어로
는 ‘관공서’, ‘서류’를 선정하였고, ‘유년에 대한 향수’를
대표하는 시어로는 ‘엄마’와 ‘누이’를 각각 선정하였다.
시의 핵심 이미지를 대표하는 시어는 직관적으로 선정
하였다. 이 단어들과 유사도가 높은 단어를 단어백터로
찾아보면 표 2 와 같은 결과를 얻을 수 있다.

결과를 살펴보면 다소 미묘한 뉘앙스의 시어들이 추
천됨을 알 수 있다. 즉 직관적으로 찾은 키워드에 대해
직접적으로 파악하기 어려운 단어들의 유사도가 높게
연산된 것이다. ‘도시’라는 시어와 유사도가 높은 단어
로는 ‘빠지다’, ‘당황하다’, ‘내리다’, ‘아프다’, ‘노인들’,
‘천국’ 등이 백터 연산되었다. ‘천국’이나 ‘노인들’이라는
단어는 ‘도시’라는 단어와는 다소 무관하게 느껴지나 이
단어들이 사용된 기형도의 시를 살펴보면 어느 정도 그
타당성을 확인할 수 있다. 우선 ‘천국’이라는 단어는 다
음의 시에 등장한다. 지면상 줄바꿈 등을 생략하였다.

그대, 아직 내게 무슨 헤어질 여력이 남아 있어 붙들
겠는가. 그대여, 엑스자로 단단히 구두끈을 조이는
양복 소매 끈에서 무수한 달의 지느러미가 떨어진다.
떠날 사람은 떠난 사람, 그대는 천국으로 떠난다고
장기 두는 식으로 용감히 떠난다고 짧게 말하였다.
하늘나라의 달. (...) 여차피 내가 떠나기 전에 이미
나는 혼자 였다. 그런데 너는 왜 천국이라고 말하였
는지. 네가 떠나는 내부의 유배지는 언제나 푸르고
깊었다. 불더미 속에서 무겁게 터지는 공명의 방 그
리하여 도시, 불빛의 사이렌에 썰물처럼 골목을 우회
하면 고무줄처럼 먼저 튕겨 나와 도망치는 그림자를
보면서도 나는 두려움으로 몸을 떨었다. (...) 가끔은
시간을 앞질러 골목을 비어져 나오면 아, 온통 체온
계를 입에 물고 가는 술한 사람들 어디로 가조 꿈을
생포하려 예 누가요 꿈 따위는 없어 모두 어디로, 천국

으로 세상은 온통 크레졸 냄새로 자리 잡는다. 누가 떠나는 죽든 우리는 모두가 위대한 혼자였다. 살아 있으라, 누구든 살아 있으라. (...)

-「비가 붉은 달」

위 시에서 화자는 도시의 병동에서 앓고 있는 병자들에 대한 애잔한 마음을 잘 묘사하고 있다. 이 도시의 병동에서 앓다가 죽어가는 사람들이 가는 곳이 바로 ‘천국’인 것이다. 그 천국에서 즐겁고 행복하게 살아갈 것이라는 낙관보다는 ‘꿈 따위는 없’는 자포자기의 감정을

표 2. 단어벡터로 구한 핵심 이미지와 유사도 높은 단어
Table 2. Similar words to the core image recommended by word vectors

이미지	키워드	연관어 (Similarity)		
		SVD	CBOW	LSTM
공장화 및 도시화	도시	빠지다 민다 천국 가방 꽃히다 아프다	빠지다 당황하다 나가다 민다 잇다 성탄목	세우다 내리다 지치다 쉬다 꺼내다 노인들
	공장	빨리다 걸음 찬다 앞서다 숨구멍 식물	빨리다 찬다 앞서다 숨구멍 가다	수은주 실패 물 변화 빠르다 높다
관료적 (카프카적) 인간의 우울과 그에 대한 공감	관공서	변하다 희미하다 희다 새어나오다 들키다 송진탑	변하다 새어나오다 희미하다 희다 덩어리 멀다	자격 예배당 삼촌 입김 안심하다 거울
	서류	봉투 눈길 허리 처량하다 헤매다 황혼	봉투 눈길 허리 당황하다 헤매다 스팀	사각 석탄 보행 칼끝 펌프 풍병
유년에 대한 향수	엄마	오시다 시장 혜 숙제 시들다 배추	오시다 시장 일무 걱정 배추 삼십	가능하다 김장 뜸 헤매다 타박타박 놓다
	누이	몸 약값 비비다 뽀뽀린 뽀 마늘	약값 몸 마늘 뽀 뽀뽀린 구부러지다	햇살 망토 휴식 눈보라 스위치 담장

천국에 투사한 것으로 파악된다. 이렇게 ‘도시’라는 단어와는 다소 멀게 느껴지는 단어인 ‘천국’이 기형도 시의 어느 맥락에서는 ‘도시’와 비슷한 감정을 불러일으키는 것으로 볼 수 있다. 단어벡터를 활용하면 놓치기 쉬운 미묘한 시어의 용법에 대해서도 다시 한 번 숙고할 수 있을 것으로 판단된다.

또한 표 2에서, SVD와 CBOW의 결과는 거의 비슷하지만 LSTM의 결과는 상이함을 알 수 있다. 카프카적인 인간의 우울감과 공감에 대한 대표 시어로 선정된 ‘관공서’와 유사도가 높은 단어로 SVD와 CBOW는 ‘변하다’, ‘희미하다’, ‘새어나오다’ 등의 단어를 공통적으로 추천했지만, 이와 달리 LSTM의 경우 ‘자격’, ‘예배당’, ‘삼촌’ 등의 단어를 추천했기 때문이다. 이는 각 기법별로 단어벡터 연산의 기준이 상이하기 때문인 것으로 판단된다. 즉, SVD와 CBOW는 전후방 단어의 인접 여부로, LSTM은 후방 단어의 순서를 기준으로 단어벡터를 연산하기 때문이다. 단순한 인접 여부만 따지는 SVD와 CBOW의 결과가 유사하고, 이와는 다르게 후방 단어의 순서를 기준으로 단어벡터를 연산하는 LSTM의 결과는

표 3. 대표적인 시어의 관계에 해당하는 유추 단어
Table 3. Corresponding analogue words to the relations of the representative poetry words

이미지 / 키워드	유추어 (Analogy)		
	SVD	CBOW	LSTM
(공장 및 도시화) 공장 - 여직공 밭 - ?	지나다 생각하다 묵다 처음 사람들 감자	사건 한밤중 사소하다 개 작년 아이	성냥불 독서 기도 상점 망토 골목
(관료적 인간의 우울) 사무실 - 서류 집 - ?	아버지 어머니 여자 누이 소리 하늘	아버지 어둠 어머니 떠오르다 위로 액체	소리 여자 아이들 우리 기억 풀잎
(유년에 대한 향수) 누이 - 꽃 행인들 - ?	소리 검다 저녁 어머니 물 세상	검다 종소리 교회 지나가다 명랑하다 어머니	더럽다 빵 폐허 죽다 장마통 무수하다
(남녀 주체의 차이) 사내 - 눈물 여자 - ?	추억 아버지 시간 친구 돌아가다 다시	맑다 박수 감동 냄새 다스리다 울먹이다	발자국 유리 공기 빛깔 어둠 육체

다른 것이다. 이러한 특징에 유의하면서 유사도를 파악한다면 좀더 유의미한 분석이 될 것으로 판단된다.

표 3에는 핵심 이미지를 대표하는 시어들의 관계에서 유추된 단어들을 단어벡터로 연산해 보았다. ‘유년에 대한 향수’를 대표하는 시어로 ‘누이’와 ‘꽃’을 선정하고 이 두 단어 사이의 관계와 유사한 관계를 형성하는 ‘행인들’에 해당하는 단어를 연산한 것이다. 기형도의 시에서는 어린 나이에 죽은 ‘누이’를 ‘꽃’에 비유하는 경우가 있는데, 이와는 무관하게 느껴지는 ‘행인들’과 ‘누이-꽃’의 관계를 형성하는 단어들로, ‘소리’, ‘어머니’, ‘검다’, ‘폐허’, ‘뺨’, ‘지나가다’ 등의 단어가 추천되었다. ‘행인들’이라는 단어는 ‘흔해빠진 독서’에 등장하는데 그 분위기와 의미를 원문의 시에서 살펴보면 다음과 같다.

휴일의 대부분은 죽은 자들에 대한 추억에 바쳐진다. 죽은 자들은 모두가 겸손하며, 그 생애는 이해하기 쉽다. (...) 수북한 턱수염이 매력적인 이 두꺼운 책의 저자는 의심할 여지없이 불행한 생을 보냈다. 위대한 작가들이란 대부분 비슷한 삶을 살다 갔다. 그들이 선택할 삶은 이제 없다. (...) 그럴 때마다 내 손가락들은 까닭 없이 성급해지는 것이다. 휴일이 지나가면 그뿐, 그 누가 나를 빌려가겠는가 나는 분명 감동적인 충고를 늘어놓을 저 자를 눕혀두고 여느 때와 다를 바 없는 저녁의 거리로 나간다. 휴일의 행인들은 하나같이 곧 울음을 터뜨릴 것만 같다. 그러면 종종 묻고 싶어진다. 내 무시무시한 생애는 더 많이 대부분 쓸모없는 죽은 자들을 당신이 좀 덜어가 달라고

—「흔해빠진 독서」

이 시에서 화자는 휴일에 맑스의 책을 보던 심정을 묘사하고 있다. ‘까닭 없이 성급해지는’ 마음으로 저녁의 거리에 나가보자 ‘울음을 터뜨릴 것’ 같은 ‘행인들’이 보인다. 이 행인들은 무심하거나 방관하는 타자들이 아니라 화자와 공감하는, 혹은 화자가 자신의 감정 상태를 이입시킨 대상으로 파악된다. ‘누이-꽃’의 관계에서 유추된 ‘행인들-뺨’에서 ‘뺨’이라는 단어는 이 비유적인 관계와는 다소 무관한 단어로 보이는데, ‘뺨’이 등장하는 다음의 시에서 그 의미를 살펴보자.

김은 블라인드를 내린다. 무엇인가 생각해야 한다. 나는 침묵이 두렵다 침묵은 그러나 얼마나 믿음직한

수표인가 내 나이를 지나간 사람들이 내게 그걸 가르쳤다. 김은 주저앉는다. 어쩔 수 없이 이곳에 한번 꽃히면 어떤 건물도 도시를 빠져나가지 못했다. 김은 중얼거린다. 이곳에는 죽음도 살지 못한다. (...) 그러나 물을 끝없이 갈아주어도 저 꽃은 죽고 말 것이다. 뺨 꺾이기처럼 김은 상체를 구부린다. 뺨 부스러기처럼 내겐 얼마나 사건이 많았던가, 콘크리트처럼 나는 잘 참아왔다. 그러나 경험 따위는 자랑하지 말게 그가 텅텅 울린다. 여보게 놀라지 말게, 아까부터 줄곧 자네 뒤통에 앉아 있었네 김은 약간 몸을 부스러기된다. 이와, 우린 언제나 서류몽치처럼 속에 나란히 붙어 있네, 김은 어깨를 으쓱해 보인다. 아주 암전히 명함이나 타이프 용지처럼 햇빛 한 장이 들어온다. 김은 블라인드 쪽으로 다가간다 그러나 가볍게 건드려도 모두 무너진다. (...) 또다시 어리석은 시간이 온다. 김은 갑자기 눈을 뜬다. 갑자기 그가 울음을 터뜨린다. 갑자기 모든 것이 엉망이다. 예정된 모든 무너짐은 얼마나 질서 정연한가 김은 얼굴이 이그러진다.

—「오후 4시의 희망」

사무실에서 ‘김’과 화자가 나누는 대화나 상황을 묘사한 시로 파악된다. 여기에서 ‘김’이라는 인물은 사무실에 적응하지 못하고 우울해 하는 인물로, ‘한 번 꽃히면’ ‘빠져나가지 못하는’ 도시에 절망하고 있다. 이 ‘김’이라는 인물의 태도와 마음 상태를 구부러진 ‘뺨 꺾이기’나 ‘뺨 부스러기’로 묘사하고 있다. 앞에서 ‘행인들’을 화자의 절망적인 마음을 투사한 대상으로 보았는데, 여기서 ‘뺨’을 이용해 우울한 주체의 마음상태를 비유적으로 묘사하고 있는 것으로 볼 수 있다. 죽은 ‘누이’를 비유적으로 묘사하는 ‘꽃’과 비슷한 관계를 이룬다고 할 것이다. 즉, ‘누이-꽃’이 죽어서 상실된 대상에 대한 비유적 묘사의 관계를 의미한다면, ‘행인들-뺨’은 화자의 마음 상태를 투사시킨 대상과 그 대상의 상태를 비유적으로 묘사한 사물의 관계로 볼 수 있다. 이러한 관계에 대한 유추 단어를 통해서 시의 핵심 이미지를 더욱 깊고 다양하게 이해할 수 있는 것으로 판단된다.

III. 결 론

전후방에 인접한 단어 혹은 후방 단어의 순서를 학

습하는 CBOW 및 LSTM과 같은 딥러닝 기반의 기법들과 동시발생행렬을 이용하는 통계학 기반의 SVD 기법들을 활용하여 단어벡터를 구할 수 있다. 이 단어벡터를 이용하면 유사도가 높은 단어나 유사한 관계를 형성하는 단어를 찾을 수 있다. 이러한 단어벡터를 기형도의 시에 적용하여 시의 핵심 이미지에 해당하는 시어와 유사도가 높은 단어를 찾고, 핵심 시어들의 관계와 유사한 관계를 형성하는 다른 단어들을 유추해보았다.

직관적으로 선정한 시적 핵심 이미지를 표현한 대표 시어와 유사도가 높은 단어들을 단어벡터로 찾아보면 일반적인 이미지가 다소 기준 시어와 거리가 먼 것처럼 보이는 단어들이 있을 수 있다. 하지만 그 단어가 사용된 시의 맥락에서는 기준 시어와 유사한 이미지를 표현한다는 것을 알 수 있었다.

SVD 및 CBOW의 추천 단어와 LSTM의 추천 단어가 다소 상이한 특징을 보이는데, 이는 각 기법들의 연산 기준이 다르기 때문이다. SVD 및 CBOW는 전후방 단어의 인접 여부를 기준으로, LSTM은 후방단어의 순서를 기준으로 단어벡터를 연산하기 때문이다.

또한 시의 핵심 이미지를 나타내는 단어 두 개를 선정한 후 이 단어들 사이의 관계와 유사한 관계를 형성하는 다른 단어를 단어벡터로 유추해보았다. 이 추천된 단어들 중 다소 무관한 단어들도 있으나 그 단어들이 사용된 시를 살펴보면 비슷한 관계를 형성하는 것을 알 수 있었다.

이상의 분석을 통해 딥러닝 혹은 통계 기반의 단어벡터를 활용하면 유사도가 높은 단어 혹은 유사한 관계의 단어들을 쉽게 찾을 수 있다. 이렇게 구한 단어들 중에는 거리가 멀어 보이는 단어들도 있지만 그 단어들을 다시 살펴보면 그 기준에 부합하는 의미나 관계를 형성하는 경우가 많음을 알 수 있었다. 또한 이러한 과정을 통해, 시에 대한 좀더 깊고 다양한 분석과 해석을 시도할 수 있는 하나의 방법으로서 통계 혹은 딥러닝 기반의 단어벡터를 활용할 수 있을 것으로 사료된다.

References

- [1] Hyungsuc Kang, Janghoon Yang, "Analyzing Semantic Relations of Word Vectors trained by The Word2vec Model", Journal of KIISE, 46(10), pp. 1088-1093, 2019, DOI : 10.5626/JOK.2019.46.1.0.1088
- [2] Heimerl, F., Gleicher, M., "Interactive Analysis of Word Vector Embeddings", Computer Graphics Forum, 37(3), pp. 253-265, 2018, DOI : 10.1111/cgf.13417
- [3] Song, Ji-seon, "A Study on the Topophobia of Stuck 'Empty Place' in Modernity of Ki Hyungdo's Poetry", The Korean Language and Literature, vol. 89, pp. 177-207, 2021.
- [4] Basirat, A. Nivre, "Real-valued Syntactic Word Vectors", Journal of Experimental & Theoretical Artificial Intelligence, 32(4), pp. 557-579, 2020, DOI : 10.1080/0952813X.2019.1653385
- [5] Kim Hyeun, "Experiencing an Empty Room Closed Forever", 『Black Leaf in the mouth』, Moonji Publishing, pp. 135, 1989.
- [6] Lee Sekyung, "Space Consciousness of the Road Expressed in the Poems by Ki Hyungdo", Korean Culture Technology, 3(0), pp. 73-86, 2007.
- [7] Lee Seungchoel, "Cognitive Meanings Study on Fog of Hyeong-do Ki", Korean Language & Literature, 59(0), pp. 255-280, 2015.
- [8] Shin Jinsuk, "The Book/Cinema of the Refused World - Studies in 『Black Leaf in the Mouth』 by Ki Hyungdo", Cross-Culture Studies, 7(0), pp. 187-208, 2004.
- [9] Kwon Hyeockwoong, "The Subject in Gi, Hyoung-do's poems", Journal of Korean Literary Criticism, 34(0), pp. 65-87, 2011.
- [10] Lee Jongwon, "A Study on the Method of Reading Poetry Using the Concept of Descriptive Genre", Korean Literature Education Research, 51(0), pp. 177-208, 2016.
- [11] Park Jeonghye, "A Study on the Simsang-Poetry Therapy Utilizing Gi Hyeong Do's Poetry", Journal of Arts Psychotherapy, 9(3), pp. 39-66, 2013.
- [12] Oh Hyungyup, "An Inquiry of Walter Benjamins Literary Art Theory Laying stress on irony and allegory", Literary Criticism, 33, pp. 295-320, 2009.