

지도학습 오토인코더를 이용한 전문어의 범용어 공간 매핑 방법론

Domain-Specific Terminology Mapping Methodology Using Supervised Autoencoders

윤 병 호 (Byung Ho Yoon) 국민대학교 비즈니스IT 전문대학원 박사과정
김 준 우 (Junwoo Kim) 국민대학교 비즈니스IT 전문대학원 석사과정
김 남 규 (Namgyu Kim) 국민대학교 비즈니스IT 전문대학원 교수, 교신저자

요 약

최근 비정형 자료인 텍스트를 벡터로 변환하고 이를 통해 다양한 목적으로 방대한 양의 자연어를 분석하는 시도가 이루어지고 있다. 특히 코퍼스 규모가 제한적일 수밖에 없는 전문적인 도메인의 텍스트에 대해서도 분석 수요가 급증하면서, 해당 전문 분야의 문서를 범용 문서와 함께 분석하기 위한 연구가 활발하게 이루어지고 있다. 특정 전문어를 해당 전문어 코퍼스 외부의 일반적인 범용어와 함께 분석하기 위해서는, 전문어 임베딩 공간을 범용어 임베딩 공간과 일치시키는 것이 필요하다. 기존에는 변환 행렬 또는 매핑 함수 등을 통해 전문어 코퍼스로부터 얻은 전문어 임베딩 값을 범용어 임베딩 공간으로 변환, 일치시키려는 시도가 있었지만, 변환 행렬을 기반으로 하는 선형 변환은 국지적인 범위에서만 근사적인 변환 효과가 있다는 일반적인 선형 변환의 한계를 극복하지 못했다. 이러한 선형 변환의 한계를 극복하기 위해 최근에는 다양한 형태의 비선형적인 변환 방법이 제안되고 있으며, 본 연구에서는 오토인코더(Autoencoder)와 회귀 모델을 동시에 학습하는 종단형 학습을 통해 전문어 임베딩 공간을 범용어 임베딩 공간으로 변환하여 임베딩 공간을 일치시키는 모델을 제안한다. 실제 “보건의료” 분야의 R&D 문서에 대해 임베딩 변환 실험을 진행한 결과, 제안 방법론이 기존의 오토인코더를 활용한 방법 대비 변환 정확도 측면에서 우수한 성능을 보임을 확인하였다.

키워드 : 텍스트 분석, 워드 임베딩, 워드 임베딩 변환, 지도학습 오토인코더, 종단형 학습

I. 서 론

기술 발전과 인터넷 환경의 개선으로 인해 텍스트 데이터의 양이 가파르게 증가하고 있다. 동시에 컴퓨팅 기술 및 인공지능 알고리즘의 진보는 텍스트 데이터에 대한 정량적인 분석을 가능

케 했다. 특히 Word2vec(Mikolov *et al.*, 2013a)과 같은 워드 임베딩(Word Embedding) 방법이 소개 되면서, 비정형 자료인 텍스트를 벡터(Vector)로 변환하고 이를 통해 다양한 목적으로 방대한 양의 자연어를 분석하는 시도가 이루어지고 있다. 최근에는 제한적인 역사적 자료와 같이 텍스트의

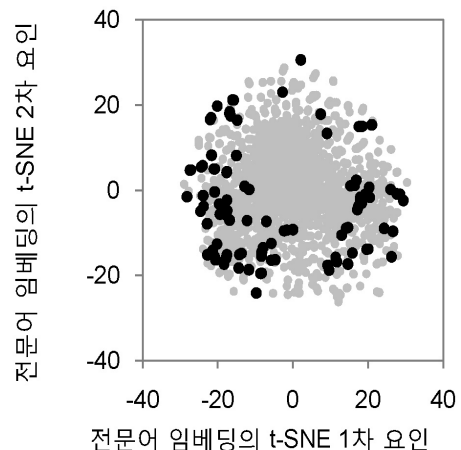
양이 한정적인 경우에 대한 자연어 분석(오준호, 2019)도 이루어지고 있다. 구체적으로 과거 사료 뿐 아니라 코퍼스 규모가 제한적일 수밖에 없는 전문적인 도메인의 텍스트에 대해서도 분석 수요가 급증하면서, 해당 전문 분야의 문서 분류, 전문어와 범용어의 유사도 분석, 또는 전문어 간의 의미 차이 분석 등을 목표로 한 연구가 활발하게 이루어지고 있다. 특히, 한의학(오준호, 2019), 법률(김성원, 박광렬, 2021; 할타르, 이말레, 2020), 그리고 IT(김민호 등, 2021) 등 전문어의 비중이 높고 또한 범용어와 동음이의어인 전문어를 다수 포함하고 있는 여러 분야에서 전문어 임베딩을 통해 전문 분야의 텍스트를 이해하려는 시도가 이어지고 있다.

특정 전문어를 해당 전문어 코퍼스 외부의 일반적인 범용어와 비교하기 위해서는, 전문어 임베딩 공간을 범용어 임베딩 공간과 일치시키는 것이 필요하다. 이는 각각의 코퍼스를 통해서 계산된 단어 임베딩은 해당 코퍼스 내 문서를 기초로 계산된 만큼, 동일한 단어라 할지라도 코퍼스별로 임베딩 값은 서로 다르게 산출되기 때문이다. 이러한 사유로 인해 코퍼스 간 임베딩 값의 단순 비교가 불가능할 뿐 아니라 코사인 유사도 등의 평가 기준이 의미를 갖기 어려우므로, 전문어 임베딩 공간과 범용어 공간의 일치하는 전문어와 범용어의 혼합 분석을 위해 반드시 필요한 과정이라고 할 수 있다.

기존 연구에서는 전문어와 범용어의 혼합 분석을 위해 두 코퍼스를 혼합하여 전문어와 범용어를 모두 포함한 하나의 임베딩 공간을 생성하고, 하나의 공간에서 각 용어의 임베딩 값을 계산하고 이를 비교하는 시도(김병태, 김남규, 2021)가 이루어졌다. 해당 연구에 따르면 전문어와 범용어 간의 공통 임베딩 공간을 통해 전문어의 확장된 의미 분석이 가능했지만, 범용어 코퍼스 대비 전문어 코퍼스의 규모가 작을 경우에는 확장 후 최종 임베딩 품질이 저하됨을 확인했다. 이러한 한계는 범용어 코퍼스 대비 전문어의 코퍼스

규모가 작은 전문어의 의미 분석 연구에서는 혼합 임베딩 방법론이 충분한 성능을 나타내기 어려움을 의미한다.

혼합 임베딩 방법 외에도 변환 행렬(Transformation Matrix) 또는 매핑 함수(Mapping Function) 등을 통해 전문어 코퍼스로부터 얻은 전문어 임베딩 값을 범용어 임베딩 공간으로 변환, 일치시키려는 시도(Artetxe et al., 2016; Grave et al., 2019; Mikolov et al., 2013b; Xing et al., 2015)가 있었다. 이들 연구는 우선 상이한 두 임베딩 공간을 일치시키기 위해 양 코퍼스에 공통적으로 존재하는 단어들을 선정한다. 다음으로 선정된 공통 단어를 기초로 두 임베딩 간 오차를 최소화하는 변환 행렬을 계산하고, 이를 임베딩 공간 변환기로 활용했다(Mikolov et al., 2013b). 공간의 변환은 이후 여러 연구(Artetxe et al., 2016; Grave et al., 2019; Xing et al., 2015)를 통해 직교 변환 행렬의 특성을 활용하여 임베딩 변환 전과 후 단어 간의 의미 관계가 보존되는 변환으로 발전하여 현재까지 활용되고 있다. 다만, 변환 행렬을 기반으로 하는 선형 변환은 국지적인 범위에서만 근사적인 변환 효과가 있다는 일반적인 선형 변환의 한계를 극복하지 못했으며, 이는 <그림 1>을 통해 설명될 수 있다.



<그림 1> 선형 변환의 한계

<그림 1>은 200차원의 R&D 전문어 임베딩을 t-SNE를 통해 2차원으로 도식화한 예로, 각 점은 전문어 임베딩 공간에서 각 전문어가 차지하는 위치를 2차원으로 보여준다. 이 중에서 범용어 공간으로 변환이 원활하게 이루어지는 전문어, 즉 전문어를 범용어 공간으로 선형 변환한 결과와 범용어 임베딩 공간 내 실제 임베딩 값과의 코사인 유사도가 높은 상위 100개의 전문어를 검은색으로 구별하여 표시했다. 만약, 전문어 변환이 전역적으로 양호하게 이루어진다면, 유사도 상위 100개 단어는 전체적으로 넓게 분포될 것으로 기대할 수 있다. 그러나 <그림 1>에서 확인할 수 있듯이, 변환 결과가 양호한 상위 100개 단어는 분포의 외곽부에 집중되어 있다. 이는 특정 전문어 집단이 다른 전문어들에 비해 변환 효과가 상대적으로 양호하게 나타난 것으로 해석할 수 있으며, 선형 방법론이 나타내는 국소적인 변환의 한계를 방증한다. 임베딩 공간 변환은 전체 단어에 대해 동일한 선형 변환을 통해 일괄적으로 이루어지기 때문에, 이러한 한계는 임베딩 결과의 신뢰성을 훼손하는 큰 문제로 지적되어 왔다. 특정 단어군에 대한 변환 성능은 양호하지만 그 외의 단어에 대한 변환 성능이 떨어지는 등 단어 간 변환 성능 격차가 클 경우, 임베딩 변환의 평균 성능도 낮게 나타날 뿐 아니라 임베딩 변환 자체에 대한 신뢰도 역시 저하되기 때문이다.

이러한 선형 변환의 한계를 극복하기 위해 최근에는 다양한 형태의 비선형적인 변환 방법이 제안되고 있으며, 이 가운데 특히 딥러닝 기반의 오토인코더를 활용한 비선형 변환 시도가 이루어지고 있다. 오토인코더는 신경망 구조 중 하나로서 입력변수와 동일한 출력변수를 사용하며, 그 사이에 차원을 압축한 잠재벡터(Latent Vector)를 포함함으로써 압축된 입력변수 정보를 제공한다는 특징이 있다(Baldi and Hornik, 2016). 이러한 오토인코더의 잠재변수를 매개로 전문어 임베딩을 범용어 임베딩 공간으로 변환할 수 있으며, 특히 오토인코더와 FFNN(Feed Forward Neural Network)을 조합할 경우 기존 방법 대비 전문어 변환 성능이 우

월함(김준우 등, 2021)을 보였다. 다만, 기존 연구는 오토인코더와 FFNN의 학습이 순차적, 독립적으로 이루어지기 때문에, 잠재변수의 학습 시 FFNN에서 이루어지는 하위 태스크(Task)의 성능이 충분히 고려되지 못했다는 한계가 있다.

따라서 전문어 임베딩 변환 시 오토인코더와 FFNN을 활용하되, 오토인코더의 복원 정확도뿐 아니라 FFNN의 하위 태스크 정확도까지 감안한 공동 손실 함수(Joint Loss Function)를 최소화하는 학습을 통해 추가 성능 향상이 기대된다. 이와 같이 오토인코더의 잠재변수에 FFNN이 추가된 오토인코더를 지도학습 오토인코더(Le et al., 2018)라고 한다. 이러한 지도학습 오토인코더는 종단형(End-to-End) 학습으로서, 전역적인 최적해 도출이 가능할 뿐 아니라 모델의 과적합 리스크를 줄여준다는 장점을 갖는다(Muller et al., 2005). 다만 다수의 모델로 구성되어 있거나 모델의 구성이 복잡한 경우에는 종단형 학습이 오히려 비효율적인 것으로 알려져 있다(Muller et al., 2005). 그러나, 본 연구에서 제안하고자 하는 전문어 임베딩 변환 모델은 오토인코더와 FFNN 각 1개로 구성된 단순한 모델인 만큼, 종단형 학습의 단점 대비 장점이 클 것으로 기대한다.

따라서 본 연구에서는 전문어의 의미 분석을 위한 전단계로서, 오토인코더와 임베딩 변환 태스크를 동시에 학습하는 종단형 학습을 통해 전문어 임베딩 공간을 범용어 임베딩 공간으로 변환하여 임베딩 공간을 일치시키는 모델을 제안한다. 구체적으로 본 연구에서는 오토인코더와 임베딩 변환 태스크를 순차적으로 진행했던 기존의 방법론과 다르게 종단형 학습을 제안함으로써, 모델의 성과와 간결성 측면에서 보다 개선된 연구 방법론을 제시하고자 한다. 전체적으로 제안 모델은 오토인코더의 잠재변수에 하위 태스크를 위한 분류 모델을 직접 연결하는 구조를 가지며, 구체적으로는 40차원의 잠재벡터를 중심으로 인코더와 디코더 각각 완전연결계층(Dense Layer) 5개와 2개의 스킵 연결로 구성된다. 또한 제안 방법론을 통해 전문어

임베딩 값을 범용어 공간으로 변환하고, 변환된 임베딩과 해당 전문어의 범용어 공간 내 실제 존재하는 임베딩의 코사인 유사도를 비교하여 제안 방법론의 성능을 평가하고자 한다.

본 논문의 이후 구성은 다음과 같다. 제Ⅱ장에서는 오토인코더 및 종단 학습에 대한 기존 연구를 소개한 뒤, 제Ⅲ장에서는 전문어 임베딩을 범용어 공간으로 변환하는 과정에서 지도학습 오토인코더를 활용하는 방법을 제안한다. 제Ⅳ장에서는 제안 방법론의 성능 평가를 위한 실험 과정 및 결과를 소개하고, 마지막으로 제Ⅴ장에서는 본 연구의 기여와 추후 연구 과제를 서술한다.

Ⅱ. 관련 연구

2.1 전문어 변환

전문어 임베딩 공간을 범용어 공간으로 변환하면, 범용어와의 직접 비교를 통해 전문어의 의미를 분석할 수 있다. 예를 들어 N 개의 단어에 대해 M 차원의 임베딩 값으로 구성된 $N \times M$ 차원의 전문어 임베딩 행렬이 있을 때, 이 행렬과 $M \times M$ 차원의 변환 행렬(Transformation Matrix) 또는 매핑 함수(Mapping Function)와의 행렬곱을 통해 원본 행렬을 $N \times M$ 차원의 새로운 임베딩 값으로 전환(Artetxe et al., 2016; Grave et al., 2019; Mikolov et al., 2013b; Xing et al., 2015)함으로써 임베딩 공간의 변환이 가능하다. 일반적으로 양쪽 코퍼스에 공통적으로 존재하는 단어를 기준으로, 전환된 임베딩과 실제 존재하는 임베딩 간 오차를 최소화하는 행렬을 임베딩 공간 변환기로 활용한다(Mikolov et al., 2013b). 이후, 변환 행렬이 직교 행렬이 되도록 제약조건을 추가하여, 전환 전의 단어 간 의미 관계를 보존할 수 있게 되었다(Artetxe et al., 2016; Grave et al., 2019; Mikolov et al., 2013b; Xing et al., 2015). 이는 선형 전환 전후의 기하학적인 정보를 유지하는 직교 행렬의 성질을 이용한

것으로, 이러한 변환 방법을 프로크루스테스(Procrustes) 변환이라 한다. 프로크루스테스 변환은 계산이 용이할 뿐 아니라 단어 간 기존 의미 관계를 보존할 수 있다는 장점 때문에 단어의 의미 변화 분석에 여전히 빈번하게 활용되고 있다.

다만 프로크루스테스 변환을 포함한 일반적인 선형적 변환 방법은 앞서 <그림 1>을 통해 확인한 바와 같이 변환 성능이 국지적으로 차이를 보인다는 한계가 있다(Nakashole and Flauger, 2018). 실제로 선형 변환을 이용한 기계 번역의 경우 유사한 언어 간 번역에서도 성능이 높지 않게 나오는 등 많은 한계를 보였다(Søgaard et al., 2018). 즉 선형 변환은 특정 군집 단어에 대해서는 변환 성능이 양호하지만 타 군집에 대해서는 변환 성능이 떨어지는 등 군집 별로 성능의 편차가 크기 때문에, 임베딩 변환 결과의 안정성 및 신뢰성을 보장할 수 없다.

이러한 선형 변환의 한계를 극복하기 위해 최근 비선형 변환에 대한 다양한 연구(Nakashole and Flauger, 2018)가 이루어지고 있다. 특히, 임베딩 변환이 자주 쓰이는 분야 중 하나인 이중 언어 사전 구축(Bilingual Lexicon Induction) 및 기계 번역(Machine Translation) 분야에서는 단어 번역을 위해 딥러닝 기반의 오토인코더가 널리 활용되고 있다. 구체적으로는 독일어와 영어 각각의 워드 임베딩에 대해 디노이징 오토인코더를 활용하여 단어 번역, 즉 임베딩 변환 성능을 개선한 연구(Kim et al., 2019)도 수행된 바 있다. 뿐만 아니라 오토인코더 기반의 GAN 모델을 활용하여 번역의 성능을 개선한 연구(Zhang et al., 2020), 오토인코더와 FFNN을 조합한 비선형 임베딩 변환을 통해 전문어의 의미를 분석한 연구(김준우 등, 2021) 등, 오토인코더를 활용하여 임베딩 변환의 성능을 향상시킨 사례가 다수 보고되고 있다.

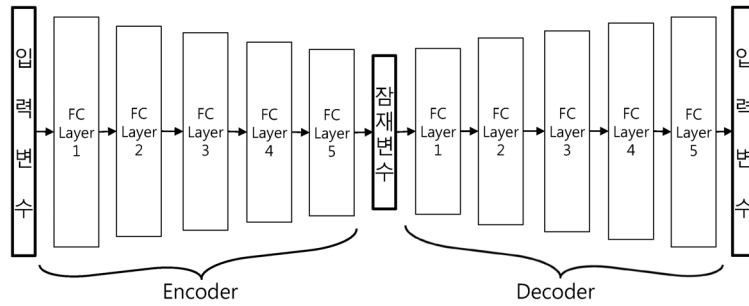
2.2 오토인코더

오토인코더란 신경망 기반 모형 중 하나로서,

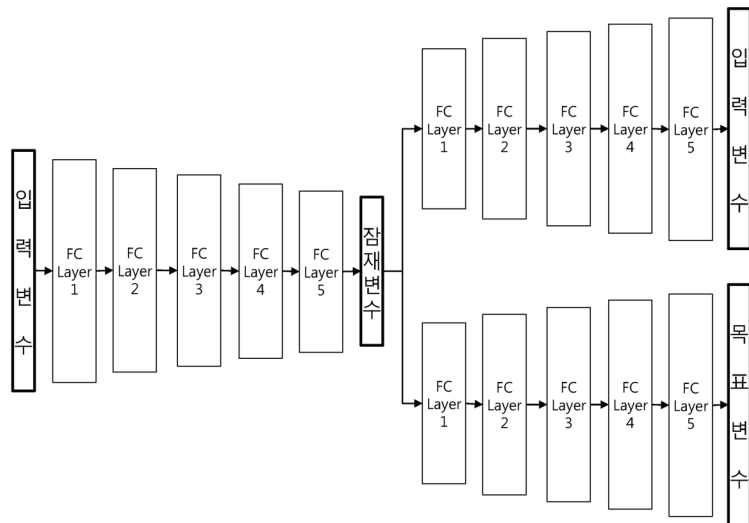
동일한 입력변수와 출력변수 사이에 병목층을 중심으로 좌우 대칭되는 은닉층 구조를 갖는다는 특징이 있다(<그림 2>). 오토인코더는 출력변수가 입력변수를 최대한 정확하게 복원할 수 있도록 은닉층의 가중치를 학습하게 된다. 이때, 병목층을 잠재변수라고 하며, 일반적으로 입력변수보다 낮은 차원으로 설정한다. 잠재변수에 입력변수 고유의 정보와 차원을 압축하여 저장하는 방식은 텍스트 분석뿐 아니라 이미지 분석, 디노이징 등 다양한 모델에서 활용되고 있다(Bank *et al.*, 2020).

오토인코더 내의 잠재변수는 입력변수의 정보를 압축하여 저장할 뿐 아니라, 잠재변수 자체가

분류 등 다양한 하위 태스크에 사용될 수 있다는 특징을 갖는다. 특히 오토인코더 내 잠재벡터에 하위 태스크를 연결하여 오토인코더와 하위 태스크를 동시에 학습하는 경우, 잠재변수는 입력변수를 잘 복원하기 위한 정보뿐 아니라 하위 태스크를 정확하게 수행하기 위해 필요한 정보까지 함께 압축하게 된다. 이러한 구조를 지도학습 오토인코더(Supervised Autoencoder)라고 하며, 그 구조는 <그림 3>과 같다. 실제로 오토인코더의 잠재변수에 하위 태스크 수행을 위해 FFNN을 추가한 오토인코더가 오토인코더와 FFNN을 순차적으로 학습하는 것에 비해 높은 성능을 보임을 MNIST 및



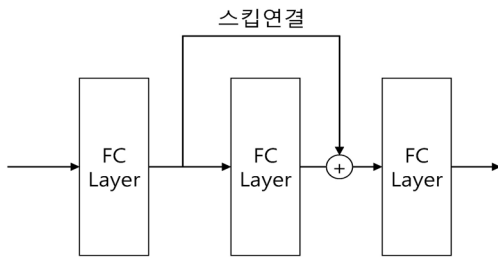
<그림 2> 일반적인 오토인코더 구조



<그림 3> 지도학습 오토인코더 구조

CIFAR-10 데이터를 통해 확인하였으며(Le *et al.*, 2018), 최근 지도학습 오토인코더는 분자 특성 분석에까지 활용(Oliveira *et al.*, 2022)되는 등 그 범위가 확대되고 있다.

다만, 일반적인 신경망 구조와 마찬가지로, 오토인코더 역시 은닉층이 깊어짐에 따라 기울기 소실(Vanishing Gradient)이 발생할 우려가 있다. 기울기 소실 문제를 해결하기 위해 이미지 분석 모델인 ResNet(He *et al.*, 2016)은 은닉층의 입력을 그대로 출력값에 더해주는 스킵연결(Skip Connections)을 제안했으며(그림 4), 이 방법은 결과적으로 기울기 소실을 극복하고 모델의 학습 효율을 향상시켰다. 최근에는 은닉층의 입력과 출력을 연결하는 것에 그치지 않고, 오토인코더의 인코더와 디코더를 직접 연결하여 이미지를 복원(Oliveira *et al.*, 2022)하는 등 다양한 변형 구조에 대한 연구가 이어지고 있다. 스킵 연결은 이미지 연구뿐 아니라, 하위 모델의 클러스터링 성능 개선(조인수 등, 2021), 트랜스포머(Transformer) 기반의 텍스트 분석(Kapoor *et al.*, 2012), 그리고 다중 레이블 분류 문제(김무성, 김남규, 2021) 등 다양한 형태의 텍스트 분석에도 널리 적용되고 있다.



〈그림 4〉 일반적인 스킵 연결 구조

2.3 종단형 학습

복잡한 문제 해결을 위해 점차 딥러닝 모델의 규모가 확대되면서, 파라미터를 조정하거나 변수를 가공(Feature Engineering)할 때 도메인 지식에 대한 의존이 심화하는 문제가 중요하게 다루어지

게 되었다. 이에 도메인 지식 없이도 대규모의 데이터와 모델을 일시에 학습함으로써 강건한 결과를 얻으려는 시도가 이루어지고 있다. 이러한 학습 방법을 종단형(End-to-End) 학습이라고 하며, 자율주행 중 장애물을 인식하는 태스크에 처음으로 적용(Muller *et al.*, 2005)되었다. 해당 연구는 CNN 모델을 기초로 주변 이미지와 운전자의 회피 기동을 각각 입력변수와 출력변수로 하는 종단형(End-to-End) 학습 모델을 제안했다. 또한 본 모델을 통해 종단형 학습이 장애물 인식 과정에서 모델의 과적합 리스크를 줄이고, 전역적인 최적해를 도출할 수 있음을 보였다.

장애물 인식 모델을 통해 종단형 학습의 효과를 확인한 이후, 다양한 분야에서 종단형 학습이 활용되었다. 우선, 음성 인식 분야에서는 음성 파일로부터 특성을 포착하는 모델, 음소를 추출하는 모델 등을 각각 학습하는 기존 접근에서 벗어나, 음성 파일로부터 바로 음성을 인식하는 신경망 모델(Lewis *et al.*, 2017)이 제안되었다. 또한 이미지 분석 분야에서도, 도메인 분류기와 조명 예측기를 동시에 학습함으로써 이미지의 색 항상성 모델의 성능을 향상시킨 연구(김준우, 김남규, 2021)도 수행된 바 있다. 뿐만 아니라, 기존 데이터가 충분치 않은 경우 데이터 증강 모델과 예측 모델의 동시 학습(김준우, 정호영, 2020)을 통해 성능 개선이 가능함을 보이는 등, 종단형 모델을 적용한 성능 개선 사례가 다수 보고되고 있다.

종단형 학습에 대한 긍정적인 연구 결과와 함께 종단형 학습의 한계도 최근 다루어지고 있다(Glasmachers, 2017). 특히, 소규모의 모델이 다수 포함된 종단형 학습 모델의 경우, 개별 모델의 학습이 충분히 이루어지지 못해서 전체 모델의 성능을 떨어뜨릴 수 있다. 그리고 모델의 일부분을 미세하게 수정하고자 하는 경우에도 전체 모델을 다시 학습해야 한다는 점은 학습 효율성 저하의 요인으로 지적되고 있다. 특히, 모델의 복잡도가 높고 규모가 큰 경우에는 종단형 학습 보다 단계적인 순차 학습의 결과가 더 좋을 수 있음이 알려져 있다.

III. 제안 방법론

3.1 방법론 개요

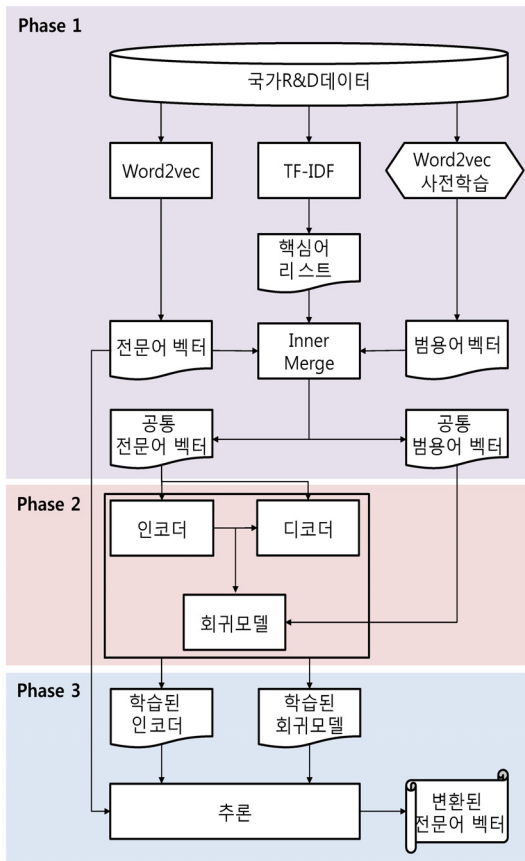
본 장에서는 전문어 임베딩 변환을 위해 <그림 5>의 방법을 제안한다. 우선 2절에서는 Phase 1을 통해 전문어 코퍼스 내에서 분석 대상이 될 전문어를 선정하고, 전문어 공간 임베딩과 범용어 공간 내 임베딩을 계산하는 방법을 소개한다. 이를 바탕으로, 3절에서는 오토인코더와 회귀 모델로 이루어진 지도학습 오토인코더 모델인 Phase 2의 학습 과정을 설명한다. 마지막으로 4절에서는 학습된 모델을 이용해 전문어를 범용어 임베딩 공간으로 변환하는 Phase 3을 구체화한다.

3.2 데이터 전처리

본 절에서는 <그림 5>의 Phase 1, 즉 데이터 전처리 과정에 대해 소개한다. 전문어 코퍼스 내 문서의 과제명과 연구내용을 결합하여, 이에 대해 Okt(Open Korea Text) 형태소 분석기를 활용하여 형태소 분석을 실시하였다. 또한 텍스트 정규화(Text Normalizaion)를 위해 코퍼스의 텍스트 내 외국어, 특수기호, 조사, 공백 등을 제거하는 전처리를 수행했다. 한편 전문어 코퍼스 내 전처리된 텍스트 데이터를 벡터로 변환하기 위해 Word2vec 모델을 활용하였으며, 이를 통해 각 단어에 대한 200차원의 임베딩을 도출하였다. 다만 범용어에 대한 임베딩은 별도로 도출하지 않고, 기존의 Word2vec 내에 사전학습된 200차원의 임베딩 결과를 그대로 활용했다.

이후의 분석은 전문어 코퍼스 내 문서별 TF-IDF(Term Frequency - Inverse Document Frequency) 상위 10개의 단어로 구성된 핵심어 집합에 대해서만 진행했다. 모든 전문어가 아닌 핵심어만을 대상으로 분석을 수행하는 이유는, 전문어 코퍼스 내 등장 빈도가 현저히 낮은 단어는 단어의 임베딩을 충분히 신뢰할 수 없고, 반대로 지나치게 자주 등장하는 전문어는 범용어의 특성을 갖기 때문이다. 이러한 과정을 통해 기존 연구(김준우 등, 2021)와 동일한 방법으로 전문어를 선정하고, 각 단어가 200차원의 전문어 공간에서 갖는 임베딩을 도출하였다.

다음으로, 앞서 선정된 핵심 전문어와 범용어에 대해 양 측에 공통으로 존재하는 단어(공통어)를 선별했다. 전문어와 범용어 간 공통어는 Word2vec을 통해 전문어 코퍼스로부터 직접 계산한 전문어 공간 내 임베딩과 Word2vec에 내재된 범용어 공간 내 임베딩을 동시에 갖는다. 따라서 공용어를 통해 전문어 임베딩을 범용어 공간으로 변환하는 변환기 학습이 가능하다.

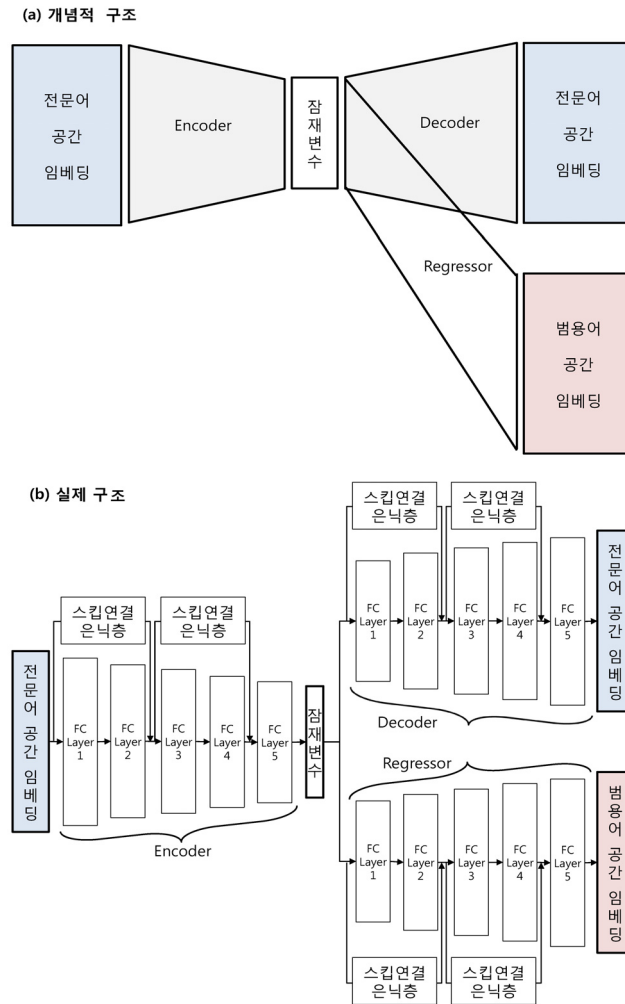


<그림 5> 제안 방법론 개요

3.3 지도학습 오토인코더를 통한 전문어 임베딩 변환

본 절에서는 <그림 5>의 Phase 2, 즉 지도학습 오토인코더의 구조 및 학습 과정을 소개한다. 전문어 임베딩 값을 범용어 임베딩 공간으로 변환하기 위한 지도학습 오토인코더의 구조는 <그림 6>과 같다. <그림 6(a)>는 오토인코더 모델과 회귀 모델로 이루어진 구조를 개념적으로 도식화하여 보인다. 오토인코더의 입력과 출력으로 전문어

공간 임베딩을 사용하고, 잠재변수에 연결된 회귀 모델의 출력변수는 범용어 공간 임베딩을 사용한다. 한편 <그림 6(b)>는 <그림 6(a)>를 구현하기 위한 구체적인 구조를 나타낸다. <그림 6(b)>에서 오토인코더 내 인코더와 디코더는 5개의 완전연결계층(Fully Connected Layer)으로 구성되어 있으며, 잠재변수는 40차원으로 설정하였다. 이렇게 구성된 오토인코더는 입력된 전문어의 임베딩을 정확하게 복원하는 작업을 수행하며, 이를 위해 복원 결과와 실제 임베딩 간 오차의 제곱합을 단



<그림 6> 제안 모델의 개념적 구조와 실제 구조

어별로 평균(MSE)한 손실 함수를 최소화하도록 오토인코더 가중치에 대한 학습이 진행된다.

전문어 임베딩을 범용어 임베딩으로 전환하기 위해 오토인코더 내 잠재변수에 5개의 완전연결 계층으로 구성된 FFNN을 연결했다. 본 FFNN은 40차원의 전문어 잠재변수를 200차원의 임베딩 값으로 변환하는 회귀 모델(Regressor)로, 입력변수인 전문어 잠재변수를 범용어 공간으로 변환한 결과와 실제 범용어 공간 내 임베딩 값과의 오차가 최소화되도록 학습한다. 결과적으로 지도학습 오토인코더는 <그림 6>과 같이 잠재변수를 중심으로 인코더, 디코더, 회귀 모델이 연결된 구조를 갖는다. 이로써 오토인코더의 손실 함수와 회귀 모델의 손실 함수의 가중합이 최소가 되도록 하는 종단형 학습이 가능해진다.

다만, 이와 같은 구조에서는 오토인코더가 다수의 은닉층으로 구성되어 있으므로 값의 전달 과정에서 기울기 소실이 발생할 우려가 있다, 특히 본 제안 모델은 오토인코더뿐 아니라 잠재변수에 연결된 회귀 모델까지 동시에 학습하는 만큼, 단순한 오토인코더에 비해 기울기 소실이 발생할 가능성이 더욱 크다. 따라서 본 연구에서는 모델 내 인코더와 디코더, 그리고 회귀 모델에 스킵연결을 추가하여 기울기 소실 가능성을 줄일 수 있는 구조를 제안한다. 구체적으로는 <그림 6>과 같이 인코더, 디코더, 회귀 모델 각각에 대해 동일하게 첫 번째 완전연결계층의 입력과 두 번째 완전연결계층 출력을 스킵연결하고, 세 번째 완전연결계층의 입력과 네 번째 완전연결계층의 출력을 스킵연결했다. 그리고 스킵연결의 입력과 출력 간의 차원을 맞추기 위해 스킵연결 은닉층(Skip-Connection Hidden Layer)을 추가했다.

3.4 추론

본 절에서는 학습된 지도학습 오토인코더를 사용하여, 전문어 임베딩을 범용어 임베딩 공간으로 변환하는 추론 과정에 대해 설명한다. 추론 시에

는 <그림 6>의 인코더와 회귀 모델만을 활용하며 디코더는 사용하지 않는다. 즉, 변환하고자 하는 전문어의 임베딩 값을 오토인코더 내 인코더에 입력변수로 제공하여, 해당 단어의 잠재변수를 계산한다. 이렇게 계산된 잠재변수는 회귀 모델의 입력으로 사용되며, 회귀 모델은 최종적으로 해당 전문어의 범용어 공간 내 임베딩 값을 반환한다.

IV. 실험

4.1 실험 개요

본 장에서는 제안 방법론을 활용하여 실제 데이터의 전문어 임베딩을 범용어 임베딩 공간으로 변환한 실험의 과정과 결과를 소개한다. 분석에 사용할 전문어 코퍼스는 2011년부터 2020년까지 국가 R&D 과제 내 “보건의료” 분야의 문서 총 77,578건의 제목과 요약문으로 선정했다. 이들 문서에 대해 특수 문자, 영어, 숫자 등을 제외하고 형태소 분석을 실시한 뒤, Word2vec 모델을 통해 200차원의 전문어 임베딩 벡터를 계산했다. 또한 각 문서별로 TF-IDF가 높은 단어 10개씩을 선정하고 중복을 제거하여 최종적으로 19,117개의 전문어를 분석 대상인 핵심어로 추출했으며, 범용어는 위키피디아의 대용량 말뭉치를 사전 학습한 뒤 공개되어 실무뿐 아니라 학술적으로 널리 쓰이는 범용 모델인 Word2vec 모델 내 30,185개의 단어와 해당 단어의 임베딩을 사용하였다.

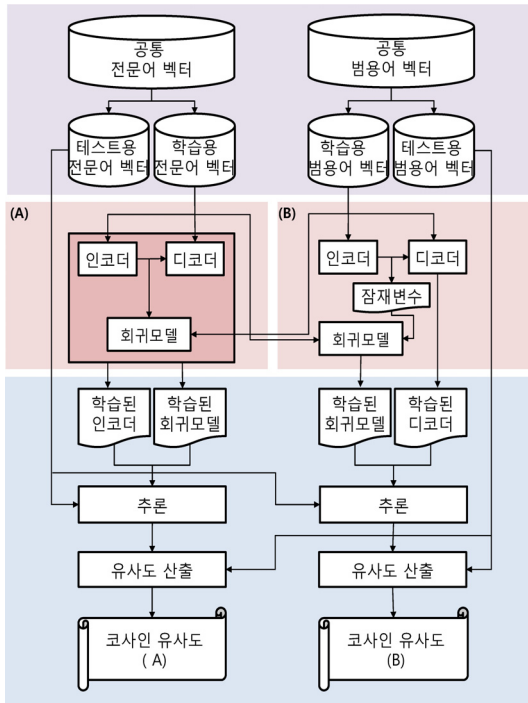
최종적으로 <표 1>과 같이 19,117개의 전문어와 30,185개의 범용어에 대해 공통으로 존재하는 공통어 7,317 단어를 선별하고, 이들 단어들이 전문어 공간과 범용어 공간 각각에서 갖는 임베딩을 모델의 학습 및 평가에 활용했다. 구체적으로 7,317개의 공통어 중 80%에 해당하는 5,854개는 학습용으로 사용하고, 나머지 1,463개는 학습에는 사용하지 않고 테스트에만 사용했다.

<그림 7>은 제안 모델인 지도학습 오토인코더 모델(A)과 비교 모델인 순차 오토인코더 모델(B)

과의 성능 비교를 위한 실험 과정이다. 지도학습 오토인코더 모델은 전문어 공간 임베딩을 오토인코더 내 인코더의 입력과 디코더의 출력으로 활용한다. 이와 동시에 전문어 공간 임베딩의 잠재변수를 입력으로 하고 범용어 공간 내 임베딩을 출력으로 하는 회귀 모델 학습을 진행한다.

<표 1> 분석 데이터

	전문어 데이터	범용어 데이터
학습 대상 문서	국가 R&D과제 중 “보건의료” 제목 및 요약문	Word2vec 사전학습 자료
문서	77,578건	2,203,153건
단어	19,117개 (핵심어)	30,185개
공통어	7,317개	
	학습용	테스트용
	5,854	1,463



<그림 7> 실험 전체 프로세스

반면, 비교 모델은 전문어 공간이 아닌 범용어 공간 내 임베딩 값을 입력 및 출력으로 활용하여 오토인코더를 1차적으로 학습한다. 제안 모델과 달리 오토인코더의 학습이 완료된 이후에 전문어 공간 내 임베딩을 입력변수로 하고 오토인코더에서 계산된 범용어 공간 기반의 잠재벡터를 출력변수로 하는 회귀 모델을 학습함으로써, 전문어 공간의 임베딩을 범용어 공간으로 변환하게 된다. 제안 모델과 비교 모델의 오토인코더는 각각 전문어 공간의 임베딩과 범용어 공간의 임베딩을 입출력 변수로 활용했다는 점에서 근본적인 차이가 있으나, 이를 제외한 나머지 오토인코더의 구조는 동일하다. 즉, 두 모델 모두 2개의 스킵연결과 5개의 완전연결계층을 갖는 인코더와 디코더, 그리고 40차원의 잠재벡터로 구성된다. 회귀 모델 역시 제안 모델과 비교 모델 모두 동일하게 2개의 스킵연결이 포함된 5개의 완전연결계층으로 구성된다. 또한 활성화 함수(Activation Function) 역시 두 모델 모두 동일하게 ‘ReLU’를 사용했다. 모델의 학습과 추론 등을 포함한 모든 실험은 Python을 사용하여 진행했으며, 구체적인 H/W 및 S/W 환경은 <표 2>와 같다.

<표 2> 실험 환경

HW	GPU	GeForce GTX 1060
	CPU	Intel(R) i7-8700 3.20GHz
	Memory	16GB
SW	Python	3.7.11
	Tensorflow-gpu	2.3.1
	Keras	2.4.3

학습이 완료된 이후의 추론 과정은 다음과 같다. 우선 제안 모델은 전문어 공간의 임베딩이 학습된 인코더와 회귀 모델을 거치면서 추론, 즉 범용어 공간으로의 임베딩으로 변환된다. 그리고 비교 모델은 회귀 모델과 디코더를 통해 전문어 공간의 임베딩이 범용어 공간으로 변환된다.

기존 모델 대비 제안 모델의 변환 성능을 평가

하기 위해 모델 별 변환 결과와 실제 범용어 공간 내 존재하는 임베딩 간의 코사인 유사도를 분석했다. 즉, 제안 모델의 추론 결과를 기초로 계산한 코사인 유사도 평균과 비교 모델의 추론 결과로 계산한 코사인 유사도의 평균을 비교하여, 이 값이 더 큰 모델이 상대적으로 우수한 성능을 보이는 것으로 판단했다. 이를 위해 전체 공통어의 20%에 해당하는 1,463개의 단어를 테스트용 데이터셋으로 활용하여, 각 모델의 임베딩 변환 성능을 비교했다.

4.2 전문어 변환 결과

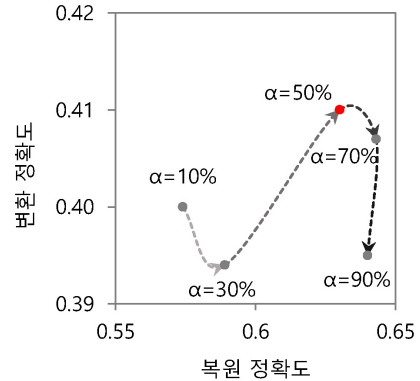
본 절에서는 실제 실험을 통해 나타난 제안 모델과 기존 모델의 성능을 비교한다. 이를 위해 테스트용 데이터인 공통어 1,463개에 대해 각각의 모델로 임베딩 변환을 진행한 후, 변환된 결과와 해당 단어의 실제 임베딩 간의 코사인 유사도를 산출하고 평균을 취하여 평가 지표로 삼는다. 이때, 오토인코더의 복원 손실과 회귀 모델의 변환 손실 간 가중치 구성의 변화에 따라 전체 모델의 성능이 영향을 받은은 당연하다. 따라서, 복원 손실 가중치(α)와 변환 손실의 가중치($1-\alpha$)를 변화시키며 동일한 실험을 반복 수행하였으며, 그 결과를 <표 3>에 요약하였다

<표 3> α 변화에 따른 모델 성능 변화

	복원 손실 가중치(α)				
	10%	30%	50%	70%	90%
변환 정확도	0.400	0.394	0.410	0.407	0.395
복원 정확도	0.574	0.589	0.630	0.643	0.640

<표 3>를 통해 가중치가 변화함에 따라 회귀 모델의 변환 정확도와 오토인코더의 복원 정확도가 변화하는 양상을 확인할 수 있다. 일반적으로 가중치(α)의 변화에 따라 변환 정확도와 복원 정확도는 서로 역의 상관 관계를 가지며 변화할 것으로 예상할 수 있다. 그러나 <표 3>에서 복원 정

확도와 변환 정확도 간의 강한 역의 상관 관계는 찾아볼 수 없으며, 이는 <그림 8>을 통해 시각적으로 더욱 명확하게 확인할 수 있다.



<그림 8> α 변화에 따른 모델 성능 변화 양상

구체적으로 살펴보면 복원 손실 가중치(α)가 증가함에 따라 복원 정확도는 전반적으로 향상되었으나, 가중치가 50% 이상일 때는 가중치(α)의 증가에 따른 복원 정확도의 개선이 미미하거나 오히려 감소하는 경우도 발생하였다. 한편 변환 손실 가중치($1-\alpha$)의 변화와 변환 정확도의 관계는 뚜렷이 나타나지 않았다. 본 실험을 통해 복원 손실과 변환 손실 간의 균형이 필요함을 알 수 있었으며, 최적 균형을 찾기 위해 다양한 가중치 조합에 대한 반복 실험이 필요함을 확인하였다. 이후 부분에서는 반복 실험 결과 변환 정확도가 가장 높을 뿐 아니라 복원 정확도 역시 양호한 성능을 보여준 모델, 즉 복원 손실 가중치와 변환 손실 가중치를 각각 50%로 가중합한 모델로 실험을 수행한 결과를 소개한다.

기존 모델과 제안 모델의 평균 코사인 유사도를 정리한 <표 4>에 따르면 제안 모델이 기존 모델 대비 코사인 유사도의 평균이 소폭 높음을 알 수 있다. 이는 제안 모델의 변환을 통해 실제 임베딩 값과 더 유사한 벡터를 도출할 수 있음을 의미하는 것으로, 제안 모델의 성능이 상대적으로 우수함을 나타낸다.

<표 4> 모델 간 성능 비교

	기존 모델	제안 모델
변환 정확도	0.394	0.410
복원 정확도	-	0.633

본 연구에서는 제안 모델의 특성을 보다 세밀하게 파악하기 위해, 평균 코사인 유사도뿐 아니라 개별 단어의 변환 결과를 분석하는 추가 실험을 수행하였다. <표 5>는 기존 모델 대비 제안 모델의 코사인 유사도 증가폭이 가장 크게 나타난 주요 5개 단어를 보인다. 예를 들어 단어 “항우울제”의 경우 기존 모델의 변환 결과는 코사인 유사도가 0.15 정도로 매우 낮았으나, 제안 모델을 적용할 경우, 코사인 유사도가 0.54로 대폭 개선됨을 알 수 있다

<표 5> 코사인 유사도 개선 상위 주요 단어

	기존 모델	제안 모델	성능 차이
치매	0.29	0.78	+0.49
항우울제	0.15	0.54	+0.39
폐암	0.29	0.61	+0.32
천식	0.49	0.76	+0.27
홍부	0.45	0.71	+0.26

반면 제안 모델을 통한 성과 개선이 미미했던 단어들은 <표 6>에 요약하였다. 이 표에 제시된 단어들은 기존 모델을 통한 변환에서 나타난 코사인 유사도와 제안 모델을 통한 변환에서 나타난 코사인 유사도가 서로 동일하게 나타난 단어들이다.

<표 6> 모델 간 코사인 유사도가 유사한 주요 단어

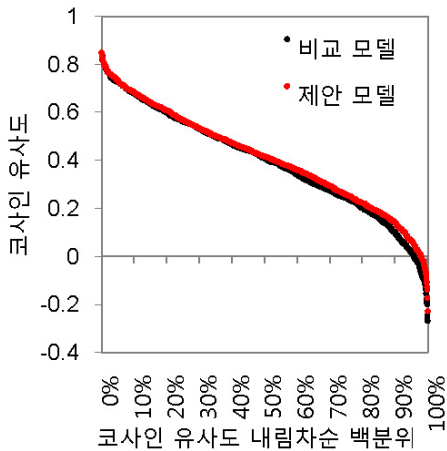
	기존 모델	제안 모델	성능 차이
컨트롤러	0.85	0.85	0.00
카트리지	0.76	0.76	0.00
휴대용	0.73	0.73	0.00
파이프	0.64	0.64	0.00
포인트	0.57	0.57	0.00

<표 5>와 <표 6>의 비교를 통해 흥미로운 사실을 발견할 수 있다. 제안 모델의 코사인 유사도는 두 표에서 크게 다르지 않게 나타난 반면, 기존 모델의 코사인 유사도는 <표 6>에 비해 <표 5>에서 현저하게 낮게 나타난다. 즉 기존 모델을 통한 변환 결과 코사인 유사도가 낮게 나타난 단어일수록 제안 모델을 통한 큰 폭의 성능 개선을 기대할 수 있음을 암시한다. 또한 <표 5>에 나타난 단어, 즉 기존 모델에서의 변환 성능이 저조하게 나타난 단어들의 경우 전문 코퍼스인 “보건의료” 문서에서 사용이 빈번한 단어라는 특징이 있고, <표 6>에 나타난 단어들은 <표 5>의 단어들에 비해 전문적이지 않은 범용 코퍼스에서도 자주 사용되는 단어들이라는 특징이 있다. 즉 전문성이 높은 단어일수록 기존 모델을 통한 변환의 성능은 낮게 나타나지만, 제안 모델의 경우 이러한 전문어에 대한 변환 성능도 우수하게 나타났음을 알 수 있다.

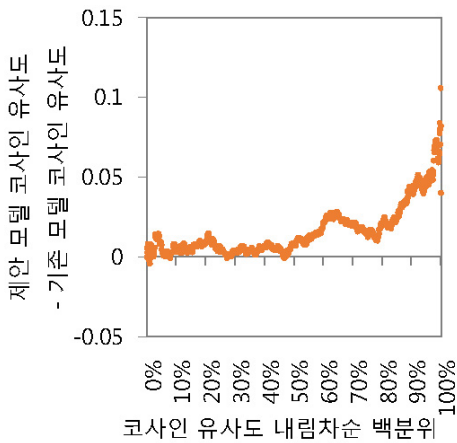
<그림 9>는 실제 임베딩과 각 모델의 변환 결과 간의 단어별 코사인 유사도를 정렬하여 도식화한 산점도이다. 그림에서 코사인 유사도가 높은 왼쪽 구간에서는 두 모델의 변환 정확도가 큰 차이를 보이지 않지만, 가로축의 오른쪽 부분, 즉 코사인 유사도가 낮은 구간에서는 제안 방법이 기존 모델 대비 변환 결과가 양호함을 확인할 수 있다. 이는 <표 5> 및 <표 6>에서 나타난 현상과 일맥상통하는 결과로, 변환 품질이 좋지 않은 구간에서 제안 방법이 기존 방법 대비 양호한 변환 결과를 제공함을 나타낸다.

한편 <그림 10>은 <그림 9>와 동일한 가로축에 대해 제안 모델과 기존 모델 간의 코사인 유사도의 차이를 세로축으로 도식화한 결과이다. <그림 10>의 대부분의 구간에서 차이 값이 양수(+)로 나타남을 확인하였으며, 이는 제안 방법이 전반적으로 기존 방법 대비 변환 성능이 높음을 나타낸다. 특히 우측 구간, 즉 변환 결과가 좋지 않은 구간의 경우 두 모델의 유사도 값의 차이가 더욱 크게 나타나며, 제안 방법이 기존 모델 대비 최대 0.1 가량 코사인 유사도가 높게 나타난 경우도 확인하였다.

이를 통해 제안 모델의 성능이 전반적으로 우수하며, 특히 변환이 쉽지 않은 단어들에 대해 제안 모델이 기존 모델 대비 뛰어난 변환 성능을 제공하는 것을 알 수 있다.



〈그림 9〉 백분위 순위에 따른 모델 간 코사인 유사도 추이



〈그림 10〉 백분위 순위에 따른 코사인 유사도 개선 정도

이처럼 변환 정확도가 낮은 구간에서 제안 모델이 기존 모델 대비 우수한 성능을 보인 이유는, 제안 모델이 오토인코더와 회귀 모델을 동시에 학

습함으로써 잠재변수의 품질을 개선하였기 때문인 것으로 판단한다. 즉, 종단형 학습을 거치면서 제안 모델의 잠재 변수는 기존 모델과 달리 임베딩의 변환에 기여할 수 있는 정보까지 학습할 수 있었으며, 그 결과 기존 모델로는 임베딩 변환이 어려웠던 구간에서 제안 모델이 큰 폭의 성능 개선을 가져온 것으로 판단한다.

V. 결 론

최근 다양한 텍스트 데이터 분석에 대한 수요가 늘어남에 따라, 뉴스와 같은 범용적인 코퍼스뿐 아니라 의학, 법률과 같은 전문 분야에서도 텍스트 분석에 대한 필요성이 제기되고 있다. 다만 전문 분야의 경우, 코퍼스의 규모가 상대적으로 작아 만족스러운 수준의 텍스트 분석이 이루어지기 어려웠다는 한계가 있다. 이에 본 연구는 기존 사전학습 언어 모델로 임베딩된 범용어 공간에 전문어를 추가하는 방법을 제안했다. 제안 모델은 선형 모형을 통한 임베딩 변환의 한계를 극복한 기존 오토인코더 활용 모델을 계승하되, 오토인코더와 회귀 모델을 동시에 학습하는 구조로 개선되었다. 또한 본 연구에서는 실험을 통해 지도학습 오토인코더 모델이 기존 오토인코더 모델 대비 개선된 변환 성능을 보였고, 특히, 변환 정확도가 낮은 구간에서 제안 모델의 성능이 우수하게 나타남을 확인했다.

본 연구는 최신 딥러닝 기술을 활용하여 단어 벡터의 비선형 변환 방법을 제안하였으며, 특히 기존의 임베딩 변환 모델과 달리 오토인코더와 회귀모델을 종단형으로 연결하여 성능을 향상시켰다는 점에서 학술적 기여를 인정받을 수 있다. 특히, 변환 정확도가 낮은 구간에서 기존 모델 대비 제안 모델의 성능이 개선되었다는 것이 특징적이며, 종단형 연결 구조의 경우 모델의 간결함으로 인해 추후 다른 모델과 연계하여 다양한 후속 연구에도 활용될 수 있을 것으로 기대한다. 또한 제안 방법을 통해 기존 범용 언어 모델에 전문어를

추가하게 되면, 추가 사전학습과 같은 컴퓨팅 부담이 없이 전문어에 대한 다양한 분석이 가능하다. 이는 소규모 코퍼스는 데이터가 충분치 않아 분석 결과의 신뢰성이 담보되기 어렵고, 코퍼스 내 단어가 충분하지 않아서 다양한 단어 간 비교를 통한 의미 파악이 쉽지 않다는 한계를 제안 방법론이 극복하였음을 의미한다. 본 연구는 선행 연구와의 비교를 위해 보건의료 분야를 대상으로 분석을 실시했으나, 제안 방법론은 특정 분야에 종속되지 않고 실무적으로 다양한 분야의 전문 코퍼스에 적용 가능할 것으로 기대된다. 이와 같이 소규모 코퍼스인 전문 코퍼스를 포함할 수 있게끔 텍스트 분석 범위를 확장한 점은 본 연구의 실무적 기여로 인정받을 수 있다.

본 연구에서는 제안 모델의 성능을 기존 모델과 비교하기 위해, 기존 연구가 수행된 환경과 동일한 실험 환경을 구성하여 코퍼스 내 문서별 TF-IDF 상위 10개의 단어로 핵심어 집합을 구성하여 실험 결과를 제시하였다. 하지만 제안 모델의 성능은 핵심어 집합의 규모 및 특성에 따라 상이하게 나타날 가능성이 충분히 있으므로, 향후 다양한 환경에서 반복적으로 실험을 수행하여 제안 모델의 특성을 더욱 세밀하게 확인할 필요가 있다. 이외에도, 제안 모델을 학습하기 위해 전문어 코퍼스와 범용어 코퍼스에 동시에 출현하는 공통어만 활용했다는 점은 본 연구의 한계로 지적될 수 있다. 소수의 공통어만으로는 충분한 학습이 이루어지지 않을 수 있기 때문이다. 본 제안 모델이 사용한 오토인코더는 공통어가 아닌 순수 전문어에 대한 학습도 가능하다는 특성을 활용하여, 향후 연구에서는 오토인코더 학습 시에는 전체 전문어를 활용하여 더욱 성능을 개선할 수 있을 것으로 기대한다. 즉, 준지도학습 오토인코더의 구조를 응용하여 오토인코더 학습 시에는 전문어 전체를, 그리고 회귀 모델 학습 시에는 공통어만 활용하는 학습을 구현함으로써, 본 연구에서 제안한 전문어 임베딩 변환 방법론의 성능을 개선하기 위한 시도가 이루어져야 한다.

참고 문헌

- [1] 김무성, 김남규, “다중 레이블 분류의 정확도 향상을 위한 스킵 연결 오토인코더 기반 레이블 임베딩 방법론”, *지능정보연구*, 제27권, 제3호, 2021, pp. 175-197.
- [2] 김민호, 윤호열, 최상욱, “인공신경망 기반 자연어처리를 적용한 연도별 정책내용 변화 분석에 관한 연구: 일본 IT신전략(2018-2020)을 대상으로”, *정보통신정책연구*, 제28권, 제1호, 2021, pp. 1-27.
- [3] 김병태, 김남규, “혼합 임베딩을 통한 전문 용어 의미 학습 방안”, *정보시스템연구*, 제30권, 제2호, 2021, pp. 57-78.
- [4] 김성원, 박광렬, “딥러닝을 이용한 법률 분야 한국어 의미 유사판단에 관한 연구”, *정보처리학회논문지: 소프트웨어 및 데이터 공학*, 제11권, 제2호, 2022, pp. 93-100.
- [5] 김준우, 김남규, “이질적 이미지의 딥러닝 분석을 위한 적대적 학습기반 이미지 보정 방법론”, *정보처리학회논문지: 소프트웨어 및 데이터 공학*, 제10권, 제11호, 2021, pp. 457-464.
- [6] 김준우, 윤병호, 김남규, “전문어의 범용 공간 매핑을 위한 비선형 벡터 정렬 방법론”, *지능정보연구*, 제28권, 제2호, 2022, pp. 127-146.
- [7] 김준우, 정호영, “제한된 학습 데이터를 사용하는 End-to-End 음성 인식 모델”, *말소리와 음성과학*, 제12권, 제4호, 2020, pp. 63-71.
- [8] 오준호, “한의학 고문헌 데이터 분석을 위한 단어 임베딩 기법 비교: 자연어처리 방법을 적용하여”, *대한한의학회지*, 제32권, 제1호, 2019, pp. 61-74.
- [9] 조인수, 강윤희, 최동빈, 박용범, “스킵연결이 적용된 오토인코더 모델의 클러스터링 성능 분석”, *정보처리학회논문지: 소프트웨어 및 데이터 공학*, 제9권, 제12호, 2020, pp. 403-410.
- [10] 할타르, 이말레, “인공지능 기법을 활용한 법률안 예측 모델 연구”, *한국디지털컨텐츠학회*

- 논문지, 21권, 제8호, 2020, pp. 1443-1452.
- [11] Artetxe, M., G. Labaka, and E. Agirre, “Learning principled bilingual mappings of word embeddings while preserving monolingual invariance”, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 2289-2294.
- [12] Baldi, P. and K. Hornik, “Neural networks and principal component analysis: Learning from examples without local minima”, *Neural Networks*, Vol. 2, No. 1, 2016, pp. 53-58.
- [13] Bank, D., N. Koenigstein, and R. Giryes, “Autoencoders”, Available at arXiv preprint arXiv:2003.05991, 2020.
- [14] Glasmachers, T., “Limits of end-to-end learning”, *Proceedings of the Ninth Asian Conference on Machine Learning*, Vol.77, 2017. pp. 17-32.
- [15] Grave, E., A. Joulin, and Q. Berthet, “Unsupervised alignment of embeddings with wasserstein procrustes”, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, Vol.89, 2019, pp. 1880-1890.
- [16] He, K., X. Zhang,, S. Ren, and J. Sun, “Deep residual learning for image recognition”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.
- [17] Kapoor, A., R. Viswanathan, and P. Jain, “Multilabel classification using bayesian compressed sensing”, *Advances in Neural Information Processing Systems*, Vol.25, 2012.
- [18] Kim, Y., J. Geng, and H. Ney, “Improving unsupervised word-by-word translation with language model and denoising autoencoder”, Available at arXiv preprint arXiv:1901.01590, 2019.
- [19] Le, L., A. Patterson, and M. White, “Supervised autoencoders: Improving generalization performance with unsupervised regularizers”, *Advances in Neural Information Processing Systems*, Vol.31, 2018.
- [20] Lewis, M., D. Yarats, Y. N. Dauphin, D. Parikh, and D. Batra, “Deal or no deal? end-to-end learning for negotiation dialogues”, Available at arXiv preprint arXiv:1706.05125, 2017.
- [21] Mikolov, T., K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space”, Available at arXiv preprint arXiv:1301.3781, 2013.
- [22] Mikolov, T., Q. V. Le, and I. Sutskever “Exploiting similarities among languages for machine translation”, Available at arXiv preprint arXiv:1309.4168, 2013.
- [23] Muller, U., J. Ben, E. Cosatto, B. Flepp, and Y. Cun, “Off-road obstacle avoidance through end-to-end learning”, *Advances in Neural Information Processing Systems*, Vol.18, 2005.
- [24] Nakashole, N. and R. Flauger, “Characterizing departures from linearity in word translation”, Available at arXiv preprint arXiv:1806.04508, 2018.
- [25] Oliveira, A. F., J. L. Da Silva, and M. G. Quiles, “Molecular Property Prediction and Molecular Design Using a Supervised Grammar Variational Autoencoder”, *Journal of Chemical Information and Modeling*, Vol.62, No.4, 2022, pp. 817-828.
- [26] Sogaard, A., S. Ruder, and I. Vulić, “On the limitations of unsupervised bilingual dictionary induction”, Available at arXiv preprint arXiv:1805.03620, 2018.
- [27] Xing, C., D. Wang, C. Liu, and Y. Lin, “Normalized word embedding and orthogonal transform for bilingual word translation”, In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational*

Linguistics: Human Language Technologies,
2015, pp. 1006-1011.
[28] Zhang, Y., Y. Li, Y. Zhu, and X. Hu, "Wasserstein

GAN based on Autoencoder with back-translation
for cross-lingual embedding mappings", *Pattern
Recognition Letters*, Vol. 129, 2020, pp. 311-316.

Information Systems Review

Volume 25 Number 1

February 2023

Domain-Specific Terminology Mapping Methodology Using Supervised Autoencoders

Byung Ho Yoon* · Junwoo Kim* · Namgyu Kim**

Abstract

Recently, attempts have been made to convert unstructured text into vectors and to analyze vast amounts of natural language for various purposes. In particular, the demand for analyzing texts in specialized domains is rapidly increasing. Therefore, studies are being conducted to analyze specialized and general-purpose documents simultaneously. To analyze specific terms with general terms, it is necessary to align the embedding space of the specific terms with the embedding space of the general terms. So far, attempts have been made to align the embedding of specific terms into the embedding space of general terms through a transformation matrix or mapping function. However, the linear transformation based on the transformation matrix showed a limitation in that it only works well in a local range. To overcome this limitation, various types of nonlinear vector alignment methods have been recently proposed. We propose a vector alignment model that matches the embedding space of specific terms to the embedding space of general terms through end-to-end learning that simultaneously learns the autoencoder and regression model. As a result of experiments with R&D documents in the “Healthcare” field, we confirmed the proposed methodology showed superior performance in terms of accuracy compared to the traditional model.

Keywords: *Text Analysis, Word Embedding, Terminology Mapping, Supervised Autoencoder, End-to-End Learning*

* Graduate School of Business IT, Kookmin University

** Corresponding Author, Professor, Graduate School of Business IT, Kookmin University

◎ 저자 소개 ◎



윤병호 (bhoyoon@kookmin.ac.kr)

현재 국민대학교 비즈니스IT전문대학원 박사과정에 재학 중이다. 연세대학교 경제학과에서 학사와 석사 학위를 취득하였고, 대우증권 리서치센터 등에서 국내 주식 및 선물 시장 내 이벤트 전략 및 차익거래 전략을 담당했으며 현재 관련 펀드를 운용하고 있다. 주요 관심분야는 텍스트 마이닝, 딥러닝 등이다.



김준우 (britz17@kookmin.ac.kr)

현재 국민대학교 비즈니스IT전문대학원 석사과정에 재학 중이며, 한국외국어대학교 언론정보전공으로 학사 학위를 취득하였다. 금융 빅데이터 활용 아이디어 공모전 대상, 데이콘 인공지능 비트 트레이더 경진대회, 자연어 기반 기후기술분류 등 AI 경진대회에서 다수 수상하였다. 주요 관심분야는 자연어 처리, 이미지 처리, 멀티모달, 딥러닝 등이다.



김남규 (ngkim@kookmin.ac.kr)

현재 국민대학교 비즈니스IT전문대학원 및 경영정보학부 교수로 재직 중이다. 서울대학교 컴퓨터공학과에서 학사 학위를 취득하고, KAIST 테크노경영대학원에서 Database와 MIS를 전공하여 경영공학 석사 및 박사학위를 취득하였다. 한국지능정보시스템학회 부회장, 한국정보기술응용학회 부회장, 한국경영학회 상임이사, 한국경영정보학회 이사, 한국인터넷정보학회 이사를 역임하였다. 주요 관심분야는 텍스트 마이닝, 딥러닝, 데이터 모델링 등이다.

논문접수일 : 2022년 10월 08일

게재확정일 : 2023년 01월 10일

1차 수정일 : 2022년 12월 04일