

에세이의 창의성 분류를 위한 어텐션과 역문서 빈도 기반의 자기부호화기 모델

정세진, 김덕기, 온병원*

군산대학교, 소프트웨어학과

luyttt4@gmail.com, thekey1220@kunsan.ac.kr, bwon@kunsan.ac.kr

An AutoEncoder Model based on Attention and Inverse Document Frequency for Classification of Creativity in Essay

Se-Jin Jeong, Deok-gi Kim, Byung-Won On*

Department of Software Science and Engineering, Kunsan National University

요 약

에세이의 창의성을 자동으로 분류하는 기존의 주요 연구는 말뭉치에서 빈번하게 등장하지 않는 단어에 초점을 맞추어 기계학습을 수행한다. 그러나 이러한 연구는 에세이의 주제와 상관없이 단순히 참신한 단어가 많아 창의적으로 분류되는 문제점이 발생한다. 본 논문에서는 어텐션(Attention)과 역문서 빈도(Inverse Document Frequency; IDF)를 이용하여 에세이 내용 전달에 있어 중요하면서 참신한 단어에 높은 가중치를 두는 문맥 벡터를 구하고, 자기부호화기(AutoEncoder) 모델을 사용하여 문맥 벡터들로부터 창의적인 에세이와 창의적이지 않은 에세이의 특징 벡터를 추출한다. 그리고 시험 단계에서 새로운 에세이의 특징 벡터와 비교하여 그 에세이가 창의적인지 아닌지 분류하는 딥러닝 모델을 제안한다. 실험 결과에 따르면 제안 방안은 기존 방안에 비해 높은 정확도를 보인다. 구체적으로 제안 방안의 평균 정확도는 92%였고 기존의 주요 방안보다 9%의 정확도 향상을 보였다.

주제어: 창의성, 에세이, 어텐션 메커니즘, 역문서 빈도

1. 서론

에세이의 창의성을 평가하기 위한 연구는 교육 관련 학과나 연구회에서 활발히 진행 중이다. 창의적인 에세이를 작성하는 것은 자기 생각을 다른 사람들과 공유하여 창의성을 기르기 위한 수단으로 많이 사용된다[1]. 이때, 에세이가 창의적인지 평가하기 위한 척도가 필요하다. 예를 들어, 미국에서는 학생들의 창의성을 기르기 위해 수업 외적으로 글쓰기 연습을 시킨다. 그리고 학생들이 쓴 글이 창의적인지 채점하기 위해 교사들이 모여 글쓰기의 창의성 평가 기준을 정한다. 또한, 사람마다 생각하는 창의성의 기준이 다르므로 에세이의 창의성 평가 척도 연구가 이루어지고 있다[2]. 이처럼 글쓰기에서 창의성을 평가할 기준을 정하는 것은 어렵다. 실제로 교육학이나 국문학에서는 학생들이 작성한 에세이가 창의적인지 평가하기 위한 에세이의 창의성 평가 지표 개발 연구가 끊임없이 이어지고 있다.

창의성을 평가하기 위한 기준을 마련하기 위해 국어교육과에서 연구한 다양한 종류의 글들의 창의성 평가 지표들을 참고한다. 일반적으로 글쓰기에서 창의성을 평가할 때 과정보다는 결과의 창의성만을 보고 평가한다.

[3]은 ‘내용의 창의성’, ‘표현의 창의성’의 평가 기준에 각각 ‘참신성’과 ‘효과성’이 있는데, 참신성과 효과성은 각각 어느 하나만 부족하더라도 텍스트가 창의적이라고 인정받기 어렵고, 효과성이라는 용어는 정확성이나 적절성, 그밖에 논리성 등 이른바 ‘좋은 글’의 요건들을 모두 포함하고 있다.

하지만 에세이를 평가할 때 창의성 평가 지표 영역의 항목들을 모두 만족할 수 없다. 왜냐하면, 창의성 평가 기준은 너무 광범위하기 때문이다. 예를 들어, 기존에 없던 새로운 단어를 사용하여 에세이를 작성하였지만, 에세이 내에서 단어의 사용이 부적절할 수가 있다. 이러면 주어진 에세이가 창의적인 에세이라고 보긴 어렵다. 또한, 대부분의 창의성 평가 항목들이 주관적으로 평가된다[2]. 주어진 에세이를 보고 A라는 사람은 에세이가 독창적이어서 창의적인 에세이라고 하지만 B라는 사람은 독특하긴 한데 엉뚱한 말을 써놓은 창의적이지 않은 에세이라고 평가할 수 있다. 즉, 사람마다 에세이의 창의성을 평가하는 기준이 다르다. 그래서 에세이의 창의성을 평가할 때 모든 창의성 평가 지표의 항목들을 고려하는 것은 불가능하다.

본 연구에서는 창의성 평가 지표의 모든 항목을 고려하여 자동화된 에세이 창의성 평가 모델을 구현하는 것은 불가능하므로, 에세이에서 핵심적이고 참신한 단어에 초점을 두어 에세이의 창의성을 평가한다. (여기서 참신한 단어는 말뭉치에서 빈번히 등장하지 않는 단어로 정의한다) 그러기 위해서 먼저 에세이를 문장에서 단어 단위로 나눈다. 에세이를 구성하는 각 단어는 에세이의 내용을 함축하는 역할을 한다[4]. 따라서 창의성 평가 기

- 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2022R1A2C1011404).

* 교신저자 : 군산대학교 소프트웨어학과 교수

준 항목들을 모두 포함하여 에세이의 창의성을 평가하지 않고, 에세이에서 핵심 단어이면서 참신한 단어를 가지고 에세이를 평가하는 방법을 제안한다.

이와 같은 목적을 가지고 본 논문에서는 주어진 에세이의 핵심적이고 참신한 단어를 찾아 에세이가 창의적인지 아닌지 분류하기 위해 어텐션(Attention) 기법과 역문서 빈도(Inverse Document Frequency; IDF), 그리고 오토인코더(AutoEncoder)를 이용한다. 어텐션 기법은 문장 안에서 각 단어의 중요도를 알려 주고, 역문서 빈도는 주어진 에세이의 단어 중에서 다른 에세이들에서 자주 등장하지 않는 단어에 그렇지 않은 단어들보다 좀 더 큰 값을 주는 역할을 하고, 오토인코더는 주어진 에세이와 상대적으로 다른 특징을 가지는 에세이를 비교한다. 본 연구에서는 어텐션과 역문서 빈도를 기반으로 주어진 에세이의 주요 내용을 전달하면서 참신한 표현을 가진 단어들을 오토인코더의 특징 벡터로 전달하여 에세이가 창의적인지 아닌지 분류한다.

이를 위해, 본 연구에서는 캐글(Kaggle) ‘The Hewlett Foundation: Automated Essay Scoring’에서 제공하는 학생들이 작성한 에세이[5]를 평가자들이 읽고 창의성을 평가하여 학습데이터로 생성한 후에, 창의적인 에세이와 그렇지 않은 에세이를 각각 400개씩 800개를 가지고 제안 모델의 입력 데이터로 사용한다. 이후 분류된 에세이를 구성하는 각 단어를 분석하여 두 종류의 에세이의 차이를 밝히고, 성능 평가를 위해 유사도를 이용하여 기존의 참신한 단어만을 가지고 창의적인 에세이를 분류했을 때와 비교 실험을 진행한다. 실험 결과, 기존의 방안보다 9% 높은 정확도를 보였다.

본 논문의 기여도는 다음과 같다.

- 기존의 주요 방안들과 비교했을 때 에세이의 단어 중에서 핵심적이면서 참신한 단어에 초점을 두어 주어진 에세이가 창의적인지 아닌지 분류하기 위한 딥러닝 모델을 최초로 제안한다.
- 본 연구의 실험을 위해 창의적인 에세이와 창의적이지 않은 에세이의 데이터 800개를 구축하고, 에세이의 창의성 분류 결과, 기존 방안보다 9%의 정확도가 향상되었다.

이 논문의 구성은 다음과 같다. 제2장에서는 제안 방안의 이해를 돕기 위해 관련 연구에 관해 기술한다. 제3장에서는 제안 방안에 관해 서술하고, 제4장에서는 제안 방안에 대한 실험 및 결과에 대해 분석한다. 마지막으로 제5장에서는 결론 및 향후 연구에 대해 다룬다.

2. 관련 연구

다양한 종류의 글에 대한 창의성 평가 지표 개발 연구가 국어 교육학에서 진행되어 왔다[2, 3, 6]. [2]에서는 서사문과 논설문의 창의성을 평가하기 위해 ‘아이디어/내용의 새로움’, ‘구성의 독창성’, ‘표현의 참신성’, ‘적절성’ 등 다양한 요소들을 다룬다. 하지만 이를 모두 만족하여 에세이의 창의성을 평가하는 것은 어렵다. [3]에서는 ‘참신성’, ‘효과성’을 가지고 글쓰기의 창의성을 평가한다. 여기서 ‘효과성’은 정확성이나

적절성, 논리성 등을 모두 포함하는 추상적이고 주관적인 요소이기 때문에 에세이의 창의성을 평가할 때 많은 사항을 고려해 주어야 하는 문제가 있다. 또한 [6]에서 글쓰기 평가자 개인에 의해 이루어진 창의성 평가 과정을 분석해보면, 평가자들의 평가 결과의 일치도는 매우 낮았다. 즉, 평가자 모두의 의견을 만족하여 에세이의 창의성 평가 지표를 일반화하는 것은 어려운 일이다. 따라서 에세이에 대한 모든 창의성 평가 지표를 가지고 딥러닝 모델을 구현하는 것은 불가능하므로 본 연구에서는 창의성 평가 지표의 모든 항목을 고려하기보다 에세이에서 핵심 내용을 전달하는 데 큰 영향을 끼치고 참신한 단어에 초점을 두어 에세이의 창의성을 평가한다.

그리고 국어 교육학에서와 달리 쓰기 영역의 창의성 분류에 관한 연구로 컴퓨터공학에서는 다양한 기계 학습 기법들을 이용하여 진행되어 왔다[7, 8, 9, 10]. [7]에서는 연구 논문의 참신함을 평가하기 위해 각 단어의 TF-IDF(Term Frequency-Inverse Document Frequency)를 구한다. 하지만 TF-IDF는 주로 문서에서 불용어를 제거하여 중요한 단어를 찾을 때 사용하기 때문에 참신한 단어를 찾기 위한 목적으로 사용하는 것은 부적합하다. 예를 들어, 에세이 내에서 상대적으로 중요한 단어에 큰 값을 부여하지만, 그 단어가 다른 에세이에서도 빈번하게 등장하면 낮은 값을 부여한다. 그러므로 참신하고 중요한 단어임에도 불구하고 모든 에세이에 등장하는 단어라면 TF-IDF 값은 0에 가까워진다. 또한, 단어의 순서를 고려하지 않으므로 문맥이 무시된다는 단점이 있다. 따라서 에세이의 참신성 평가에서 TF-IDF를 사용하는 것은 적합하지 않다. 그리고 [8, 9, 10]에서는 문서가 창의적인지 아닌지 분류하기 위해 합성곱 신경망(Convolutional Neural Networks; CNN)을 사용한다. 왜냐하면, CNN은 인접한 단어들을 하나의 벡터로 합쳐서 문장 내 문맥 정보를 보존해주기 때문이다. 하지만 에세이와 같이 문장의 길이가 자유로운 경우 고정 길이의 벡터로 차원을 축소하면 긴 문장의 핵심 정보가 유실될 수 있다.

본 논문에서는 문장 내에서 중요하고 참신한 단어에 초점을 맞추기 위해 어텐션과 IDF 기반의 모델을 제안하고, 입력 문장의 길이에 제한을 두지 않기 위해 오토인코더를 사용해 특징 벡터를 추출하여 유사도를 측정해 에세이가 창의적인지 아닌지 분류한다. 어텐션과 IDF를 활용함으로써 중요하면서 다수의 에세이에 등장하는 참신한 단어의 가중치가 낮아지는 문제를 해결하고, 오토인코더를 사용하여 길고 복잡한 문장을 가진 에세이도 창의적인지 아닌지 분류할 수 있다.

3. 제안 방안

그림 1은 제안 모델의 알고리즘이다. 제안 모델은 에세이에서 핵심적이고 참신한 표현을 가진 단어를 기반으로 에세이가 창의적인지 아닌지 분류한다. 그러기 위해 먼저, 3.1절 인코더(Encoder)에서 에세이를 문장으로 나누고, 문장을 다시 단어 단위로 나누어 연산할 수 있는 벡터 형태로 만들어 준다. 그리고 3.2절 디코더

(Decoder)에서 에세이를 구성하는 각 단어에 대한 ‘창의성 점수’를 구한다. (‘창의성 점수’는 3.2절에서 자세히 다룬다) 그런 다음, 3.3절에서 디코더에서 구한 단어의 창의성 점수를 오토인코더의 입력으로 넣고 하나의 특징 벡터로 함축하여 에세이의 창의성 점수를 구한 뒤, 유사도를 이용하여 에세이가 창의적인지 아닌지 분류한다.

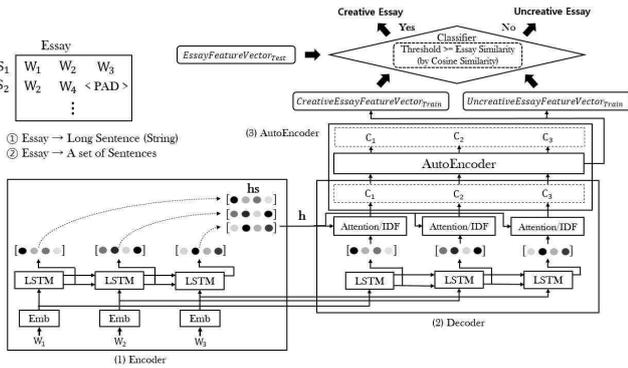


그림 1 제안 방안의 구성도

Fig. 1 System overview of the proposed deep learning model

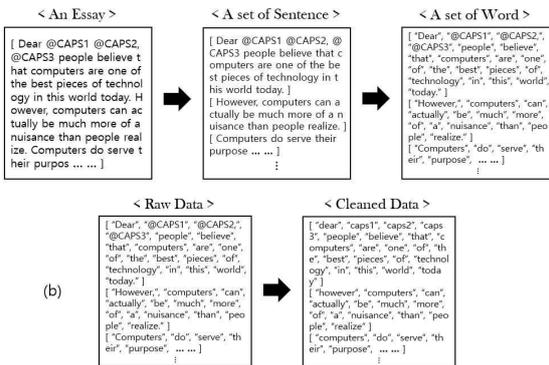


그림 2 에세이 전처리

Fig. 2 Essay preprocessing

3.1 인코더(Encoder)

에세이를 학습데이터로 사용하기 전에 계산할 수 있는 형태로 만들어 주어야 한다. 그러기 위해 에세이를 문장으로 나눈 뒤, 문장을 단어 단위로 토큰화 작업을 수행한다. (그림 2 (a)) 그리고 불용어를 제거하고, 대문자는 모두 소문자로 바꾸어 준다. (그림 2 (b))

그리고 나서, 문장의 최대 길이 벡터로 단어 임베딩 작업을 수행해 준다[11]. 하지만 임베딩만 되어 있는 단어 벡터는 문장의 문맥 정보를 반영하지 못한다. 그래서 그림 3과 같이 단어 사이의 정보를 반영해 주기 위해 LSTM 셀을 이용하여 입력된 모든 단어 벡터들을 하나의 벡터로 압축한다[12]. LSTM 셀은 긴 문장을 입력으로 하는 학습 모델에서 필요로 하는 학습 수행 능력을 갖추고 있다. 그리고 단순히 단어 하나에만 집중하여 문장을 표

현하는 것이 아니라 각 단어가 주변 단어에 대한 정보를 함께 가지게 하여 문장을 구성하는 단어 벡터들을 재조정할 수 있다.

그림 3에서 인코더의 LSTM 셀은 문장의 모든 단어를 입력받은 뒤에 마지막 시점의 은닉 상태 벡터(h)를 디코더 LSTM 셀로 넘겨주는데, 이를 컨텍스트 벡터(context vector)라고 한다[13]. 컨텍스트 벡터 h는 문장 전체에 대한 정보를 가지고 있다.

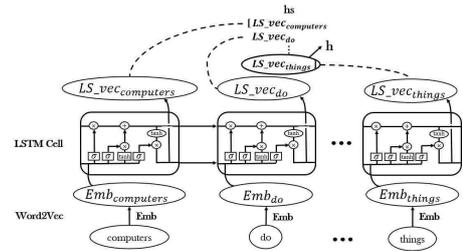


그림 3 인코더 모델의 구성도

Fig. 3 Network architecture of the encoder model

인코더에서 입력 문장들을 학습해 주기 위해 하이퍼 파라미터는 다음과 같이 설정한다. 전체 단어 종류에 대한 어휘의 수를 10,808개, 단어의 임베딩 차원 수는 100 차원, 은닉 상태 벡터의 차원 수는 64차원으로 정한 뒤, LSTM을 학습하고, Adam 옵티마이저[14]를 사용한다.

3.2 Attention과 IDF 기반의 디코더(Decoder)

에세이가 창의적인지 아닌지 분류하기에 앞서, 에세이의 창의성을 수치화하는 작업이 필요하다. 본 논문에서는 에세이의 문맥 벡터 생성 함수를 수식 1과 같이 정의한다.

$$C_i = (\alpha \cdot Attn_i + (1 - \alpha) \cdot IDF(w_i)) \quad (\alpha = 0.5)$$

수식 1 문맥 벡터 생성 함수

Eq. 1 Function for computing a context vector

수식 1은 에세이의 창의성 점수를 정의한 것이다. 에세이에서 참신한 단어는 역문서 빈도(Inverse document frequency; IDF)[15]를 이용하여 구하고, 문장에서 핵심적인 역할을 하는 단어는 어텐션(Attention) 기법[16]을 이용하여 구한다. 본 연구에서는 디코더의 i번째 단어에 해당하는 컨텍스트 벡터(C_i)를 α의 비율로 어텐션 가중치(Attn_i)와 IDF 점수(IDF(w_i))의 조합으로 설정하여 에세이의 창의성 점수를 구한다.

그림 4는 본 논문에서 정의한 ‘에세이 창의성 점수’를 가지고 에세이의 창의성을 평가하는 과정이다. 인코더와 같이 Word2Vec을 이용한 단어 임베딩과 LSTM 셀을 이용한 문장 임베딩이 이루어진다. 이때, 인코더와 달리, 각 단어가 입력되고 나서 출력된 모든 LSTM 셀들의

결과값들을 사용한다. 그리고 인코더의 마지막 시점의 은닉 상태 벡터(h)와 LSTM 셀들의 결과 벡터들을 내적하여 어텐션 가중치를 구한다. 어텐션 가중치를 구하면 문장에 대한 입력 단어의 중요성을 알 수 있다. 내적했을 때 값이 큰 단어는 문장에 영향력을 많이 주는 중요한 단어이고, 값이 작은 단어는 문장에 영향력이 크지 않은 중요하지 않은 단어로 구분할 수 있다. 어텐션 가중치와 함께 IDF 값을 구한 뒤, 각 단어에 해당하는 컨텍스트 벡터, 즉, 에세이의 창의성 점수를 구한 뒤, 오토인코더의 입력값으로 넘겨준다.

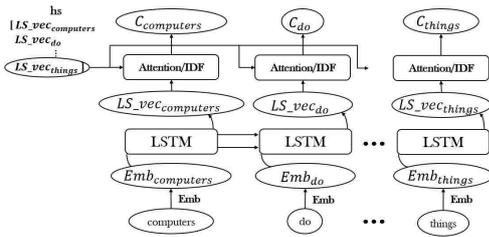


그림 4 어텐션과 역문서 빈도 기반의 디코더 모델의 구성도
Fig. 4 Network architecture of the decoder model using attention and IDF

디코더에서 입력 문장들을 학습해 주기 위해 하이퍼 파라미터는 다음과 같이 설정한다. 전체 단어 종류에 대한 어휘의 수를 10,808개, 단어의 임베딩 차원 수는 100 차원, 은닉 상태 벡터의 차원 수는 64차원으로 정한 뒤, LSTM을 학습하고, Adam 옵티마이저[14]를 사용한다.

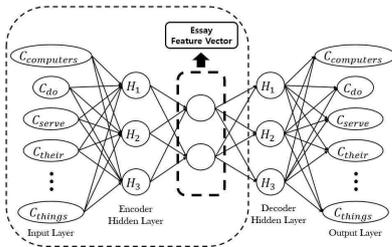


그림 5 자기부호화기 모델을 이용한 특징 벡터 추출
Fig. 5 Feature vector extraction using AutoEncoder

3.3 오토인코더(AutoEncoder)

디코더에서 에세이를 문장 단위로 쪼개어 각 문장에 대한 에세이 창의성 점수를 구하였기 때문에 이를 에세이 단위로 합쳐 주어야 한다. 이를 위해, 오토인코더(AutoEncoder) 모델을 사용한다.[17] 오토인코더 모델을 사용하면 그림 5와 같이, 입력층(Input Layer)에서 들어온 여러 개의 문장 컨텍스트 벡터들을 인코더 은닉층(Encoder Hidden Layer) 부분에서 하나의 벡터로 압축할 수 있다. 압축된 벡터는 입력층의 특징 벡터로 많이 사용하는데, 본 연구에서는 에세이의 문장들이 입력으로 들어왔기 때문에, 인코더 은닉층에서 압축된 벡터를 에

세이 특징 벡터(Essay Feature Vector; EFV)라고 한다.

그리고 나서 그림 6과 같이 오토인코더에서 구한 에세이 특징 벡터를 이용하여 에세이가 창의적인지 아닌지 분류한다. 이를 위해, 코사인 유사도를 이용하여 학습 에세이(Train Essay) 특징 벡터와 실험 에세이(Test Essay) 특징 벡터의 유사도를 구한다. 두 벡터의 유사도가 전체 에세이 특징 벡터의 평균보다 작으면 실험 에세이는 창의적인 에세이로 분류하고, 평균보다 크면 실험 에세이는 창의적이지 않은 에세이로 분류된다.

오토인코더에서 컨텍스트 벡터를 학습해 주기 위해 하이퍼 파라미터는 다음과 같이 설정한다. 입력 벡터의 크기를 문장의 최대 길이로 설정한다. 그리고 인코더 은닉층에 대한 벡터의 차원과 편향의 차원은 각각 100차원으로 설정해 오토인코더의 인코더를 학습한다. 오토인코더의 디코더도 인코더와 같다. 그런 다음, 파라미터들을 업데이트해 주기 위해 손실 함수로는 최소 제곱 오차를 사용하고, Adam 옵티마이저[14]를 사용한다.

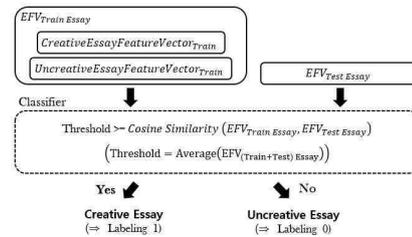


그림 6 창의적인 에세이 분류
Fig. 6 Creative essay classification

4. 실험

4.1 실험 환경

표 1은 실험에서 사용한 컴퓨터 사양이다. 구현을 위해 Python 3.6을 사용하고, 문장을 단어로 나누고 불용어를 제거하기 위해 NLTK 3.5[18]를 사용하였으며, 단어를 임베딩 하기 위해 Gensim 3.8.3[19]를 사용하였다. 또한, GPU를 이용한 딥러닝 모델 학습을 위해 Tensorflow-gpu 1.15[20]를 사용했다.

4.2 학습 데이터

본 논문에서는 캐글(Kaggle) ‘The Hewlett Foundation: Automated Essay Scoring’에서 제공하는 학생들이 작성한 영어 에세이[5]를 관련 전공자들이 읽고 창의적인 에세이와 창의적이지 않은 에세이로 분류하여 학습데이터로 사용한다. 에세이는 미국의 학년을 기준으로 7학년부터 10학년 학생들이 직접 작성하였다. 에세이의 주제는 ‘컴퓨터가 사람에게 미치는 영향에 대한 의견을 신문에 쓰는 편지 형식으로 작성하기’이다. 에세이를 구성하는 단어의 개수는 최소 150개부터 최대 550개이고, 개인 식별 정보를 제거하기 위해 인물, 기

관, 장소 등을 나타내는 단어는 익명화를 적용한다. 표 2는 학습데이터 세트의 교차검증 표이다. 창의적인 예

	Contents	Count	Size
OS	Ubuntu 18.04		
CPU	Intel Xeon Scalable 4414	2	
GPU	ZOTAC GAMING GEFORCE RTX 2080 Ti D6 11GB BLOWER	4	11GB
SSD	Samsung 970 PRO M.2 2280	1	1TB
HHD	16TB IronWolf PRO ST16000NE008	1	16TB

표 1 실험 사양

Table. 1 Experimental set-up

세이와 창의적이지 않은 에세이 각각 400개씩 800개를, 창의적인 에세이의 문장 8,633개, 창의적이지 않은 에세이의 문장 9,769개로 쪼개어 학습데이터로 사용한다.

	Creative Essays	Uncreative Essays	Sentences of Creative Essay	Sentences of Uncreative Essay
Fold 1	100	100	2,188	2,473
Fold 2	100	100	2,242	2,288
Fold 3	100	100	2,134	2,387
Fold 4	100	100	2,069	2,621
Total	400	400	8,633	9,769

표 2 학습데이터와 교차검증

Table. 2 Cross validation in the train set

4.3 실험 결과

표 3은 제안 모델에 따라 분류된 창의적인 에세이와 창의적이지 않은 에세이의 결과이다. 에세이가 잘 분류되었는지 알아보기 위해 각 단어의 창의성 점수(수식 1)를 구해 보았다. 각 단어의 창의성 점수는 0 이상 1 미만의 경우와 1 이상 2 미만의 경우, 그리고 2 이상의 경우로 나눌 수 있다. 이것을 3개의 등급으로 나누어 보면, 각각 전체 100% 중에서 56%, 34%, 10%로 나눌 수 있다. 에세이를 구성하는 단어의 개수가 최소 150개에서 최대 550개이기 때문에 각 단어에 해당하는 점수를 모두 나열하지 않고, 전체 에세이의 구성 단어들을 오름차순으로 정렬하였을 때 가장 점수가 높은 것을 기준으로 상위 1~10%에 해당하는 단어는 빨간색으로, 11~44%에 해당하는 단어는 노란색으로, 45 ~ 100%에 해당하는 단어는 초록색으로 표현하였다. 창의적인 에세이의 단어들은 대체로 빨간색이나 노란색이었지만, 창의적이지 않은 에세이의 단어들은 빨간색보다는 주로 초록색이나 노란색을 띠었다.

그림 7은 기존 연구와 제안 방안을 사용하여 에세이가 창의적인지 아닌지 분류한 정확도를 비교한 것이다. 기존 연구에서는 TF-IDF만 이용하여 단어의 창의성 점수를 구해서 에세이를 분류하였기 때문에 창의적이지 않고 새로운 단어로 구성된 에세이도 창의적인 에세이로 분류된다. 또한, TF-IDF를 사용하였을 경우, 에세이 내에서 상대적으로 중요한 단어의 창의성 점수는 높지만, 다른 에세이에서 같은 단어가 빈번하게 등장하면 창의성 점수는

낮아진다. 그래서 참신하면서 중요한 단어가 다수의 에세이에 등장하면 TF-IDF의 값은 매우 낮아진다.

Creativity	Essay Set	Labeling	
		Human	Model
Creativity (1)	Dear Local Newspaper, @CAPS1 is quite obvious that computers are in fact a benefit for the society, and they are convenient when keeping in touch, but in my opinion, computers are beginning to have a negative effect on people. They are affecting people mentally, physically, and they're also having an impact on people's safety. People need to unplug the hard drive and start living life unplugged. We've all heard the term before, a 'couch potatoe!' @CAPS2 pretty much the same for a computer except a 'desk potatoe'.	1	1
Uncreativity (4)	Dear Local Newspaper, I believe that computers have a negative effect on peoples lives. I believe this because who spend to much time on the computer don't get out as much as they should, don't spend enough time with their family, and the computer can't do everything. My first reason is I believe that people need to get out more. When they don't get out, they don't exercise and that is very unhealthy. Instead of watching the games or the scores they should get out and play the game.	0	0

Red Words : Top 1 ~ 10 % Yellow Words : Top 11 ~ 44 % Green Words : Top 45 ~ 100 %

표 3 제안 모델에 따라 분류된 창의적인 에세이와 창의적이지 않은 에세이의 결과

Table.3 Results of both creative and non-creative essays classified by the proposed model

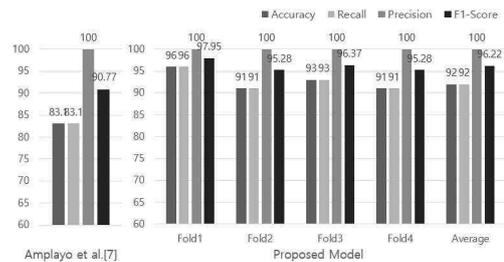


그림 7 기존 방안과 제안 방안의 정확도

Fig. 7 Results of the existing and proposed models

그에 반해, 제안 방안에서는 어텐션과 IDF를 이용하여 주어지 에세이의 문장에서 중요하고 참신한 단어에 초점을 맞추어 에세이의 창의성 점수를 구한다. 단어뿐만 아니라 문장의 문맥까지 고려하기 때문에 다른 에세이에 없는 새로운 단어로 구성된 에세이라도 주제에 맞지 않는 단어가 포함된 에세이는 낮은 창의성 점수를 받아 창의적이지 않은 에세이로 분류된다. 하지만 참신하고 문맥 간 자연스러운 표현으로 이루어진 에세이는 높은 창의성 점수를 받아 창의적인 에세이로 분류된다.

에세이가 창의적인지 아닌지 분류하기 위해 TF-IDF를 이용한 기존 방안의 정확도는 83.1%를 보였지만, 어텐션과 역문서 빈도를 이용한 제안 방안은 교차검증 결과 평균적으로 약 9% 증가하여 정확도가 92%를 보였다. 이는 에세이의 창의성을 평가할 때 새로운 단어에만 초점을 맞추어 창의성 점수를 주는 것보다 문장에서 핵심적이면서 다른 에세이에 자주 등장하지 않는 참신한 단어에 집중하여 에세이의 창의성 점수를 부여하는 방식이 에세이가 창의적인지 아닌지에 대한 분류를 더 정확하게 수행한다고 볼 수 있다.

5. 결론 및 향후 연구

본 논문에서는 사람이 쓴 에세이가 창의적인지 아닌지 분류하는 딥러닝 모델을 제안하였다. 기존 연구에서는 주어진 에세이에 포함된 단어들만 다른 에세이에 자주 등장하지 않으면 창의적인 에세이로 분류되었다. 하지만 주제와 연관성이 없는 단어들도 에세이에 많이 포함되면 창의적이지 않더라도 새로운 단어를 많이 포함하고 있으므로 창의적인 에세이로 분류되는 문제가 발생한다. 이러한 문제를 에세이 전처리 단계에서 사람이 직접 주제와 관련 없는 에세이를 골라낼 수 있지만, 에세이의 개수도 많고 평균적으로 에세이의 길이도 길어서 시간과 비용상 불가능하다. 본 논문에서는 에세이가 창의적인지 아닌지 분류하는 작업에서 기존 방안의 문제점을 해결하기 위하여 주어진 에세이의 문장에서 핵심적인 역할을 하면서 참신한 표현을 가진 단어를 기반으로 에세이의 창의성 점수를 구하고, 창의성 점수를 이용하여 다른 에세이와 유사도를 구해 에세이가 창의적인지 아닌지 분류하는 딥러닝 모델을 제안하였다. 주어진 에세이의 문장에서 단어가 중요하고 참신한지 알아내기 위해 어텐션과 역문서 빈도를 사용하였다. 기존의 TF-IDF만을 사용하였을 때보다 정확도가 약 9% 향상되었다.

향후 연구에서는 트랜스포머(Transformer) 모델을 사용하여 셀프 어텐션(Self-Attention)을 통해 에세이의 창의성 점수를 임베딩하는 것이 학습 모델을 개발할 것이다. 또한, 좀 더 다양한 평가 요소를 고려하여 에세이가 창의적인지 아닌지 분류할 계획이다.

참고문헌

- [1] Fox, Lynn H., and William G. Durden. "Educating verbally gifted youth." Fastback 176. Phi Delta Kappa, Eighth & Union, Box 789, Bloomington, IN 47402, 1982.
- [2] Han, Jung-Ran. "Development of standards for creativity assessment of narrative text and persuasive text," Graduate School of Korea National University of Education, 2015.
- [3] Park, Jong-Hoon. "A study on the concept of 'creativity' in Korean language education with regard to writing evaluation," Korean Language Education Research 20, 383-406, Seoul: Korean Language Education Association, 2004.
- [4] "Basic writing guide." Sogang Center for Writing. 2012.4.16. Modified. 2020.1.20. Content. https://writhing.sogang.ac.kr/front/guide/guide_01_02.html.
- [5] Hewlett Foundation [Online]. Available: <https://www.kaggle.com/c/asap-aes/data>. (downloaded 2019, Jan)
- [6] Oh, Sae-Young. "Study on the aspects of creativity scoring in the writing assessment of information text," Chinese Language Education 18, 59-93, Seoul: Korean Language and Literature Education Association, 2013.
- [7] Amplayo, Reinald Kim, Seung-won Hwang, and Min Song. "Evaluating research novelty detection: Counterfactual approaches." Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13). 2019.
- [8] Ghosal, Tirthankar, et al. "Novelty goes deep: a deep neural solution to document level novelty detection." Proceedings of the 27th International Conference on Computational Linguistics. 2018.
- [9] Ghosal, Tirthankar, et al. "To comprehend the new: On measuring the freshness of a document." 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019.
- [10] Nandi, Dipannyta, and Rohini Basak. "A Quest to detect novelty using deep neural nets." 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2020.
- [11] Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013).
- [12] Hochreiter, Sepp, and Jurgen Schmidhuber. "Long short-term memory." Neural Computation 9.8 (1997):1735-1780.
- [13] Sutskever. Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." arXiv preprint arXiv:1409.3215 (2014).
- [14] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).
- [15] Robertson, Stephen. "Understanding inverse document frequency: on theoretical arguments for IDF." Journal of Documentation (2004).
- [16] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473 (2014).
- [17] Vincent, Pascal, et al. "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion." Journal of Machine Learning Research 11.12 (2010).
- [18] Bird, Steven. "NLTK: the natural language toolkit." Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions. 2006.
- [19] GENSIM [Online]. Available: <https://radimrehurek.com/gensim/>
- [20] Tensorflow [Online]. Available: <https://www.tensorflow.org/?hl=ko>