

FAQ 분류 성능 향상을 위한 클래스 일치 여부 결합 학습 모델

양동일^o, 함진아, 이강욱, 이지연

삼성전자 삼성리서치

di87.yang@samsung.com, j.ham@samsung.com, kw.brian.lee@samsung.com, jyeon0.lee@samsung.com

Jointly learning class coincidence classification for FAQ classification

Dongil Yang^o, Jina Ham, Kangwook Lee, Jiyeon Lee
Samsung Electronics Samsung Research

요약

FAQ(Frequently Asked Questions) 질의 응답 시스템은 자주 묻는 질문과 답변을 정의하고, 사용자 질의에 대해 정의된 답변 중 가장 알맞는 답변을 추론하여 제공하는 시스템이다. 정의된 대표 질문 및 대응하는 답변을 클래스(Class)라고 했을 때, FAQ 질의 응답 시스템은 분류(Classification) 문제라고 할 수 있다. 종래의 FAQ 분류는 동일 클래스 내 동의 문장(Paraphrase)에서 나타나는 공통적인 특징을 통해 분류 문제를 학습하였으나, 이는 비슷한 단어 구성을 가지면서 한 두 개의 단어에 의해 의미가 다른 문장의 차이를 구분하지 못하며, 특히 서로 다른 클래스에 속한 학습 데이터 간에 비슷한 의미를 가지는 문장이 존재할 때 클래스 분류에 오류가 발생하기 쉬운 문제점을 가지고 있다. 본 논문에서는 이 문제점을 해결하고자 서로 다른 클래스 내의 학습 데이터 문장들이 상이한 클래스임을 구분할 수 있도록 클래스 일치 여부(Class coincidence classification) 문제를 결합 학습(Jointly learning)하는 기법을 제안한다. 동일 클래스 내 학습 문장의 무작위 쌍(Pair)을 생성 및 학습하여 해당 쌍이 같은 클래스에 속한다는 것을 학습하게 하면서, 동시에 서로 다른 클래스 간 학습 문장의 무작위 쌍을 생성 및 학습하여 해당 쌍은 상이한 클래스임을 구분해 내는 능력을 함께 학습하도록 유도하였다. 실험을 위해서는 최근 발표되어 자연어 처리 분야에서 가장 좋은 성능을 보이고 있는 BERT의 텍스트 분류 모델을 이용했으며, 제안한 기법을 적용한 모델과의 성능 비교를 위해 한국어 FAQ 데이터를 기반으로 실험을 진행했다. 실험 결과, 분류 문제만 단독으로 학습한 BERT 기본 모델보다 본 연구에서 제안한 클래스 일치 여부 결합 학습 모델이 유사한 문장들 간의 차이를 구분하며 유의미한 성능 향상을 보인다는 것을 확인할 수 있었다.

주제어: 자연어 처리, FAQ 분류, 결합 학습

1. 서론

최근 다양한 기업에서는 고객 상담 서비스를 자동화하고자 FAQ(Frequently Asked Questions) 질의 시스템을 도입하고 있다. FAQ 질의 응답 시스템은 특정 도메인(Domain)에서 자주 묻는 질문 및 이에 대한 답변을 정의하고, 사용자 질의에 대해 정의된 답변 중 가장 알맞는 답변을 추론하여 제공하는 시스템이다.

현재까지도 대부분의 FAQ 질의 응답 시스템은 단어 기반 검색을 바탕으로 질의에 대한 답변을 추론하고 있다. 그러나 최근에는 자연어 처리 분야에서 기계 학습이 활발하게 연구되면서 SVM(Support Vector Machine), CNN(Convolutional Neural Network), RNN(Recurrent Neural Network) 등과 같은 기계 학습 방법론을 활용한 연구 방법들도 제시되고 있다.

기계 학습을 활용한 FAQ 질의 응답 시스템에서는 대표 질의 및 그에 대응하는 답변을 클래스(Class)로

정의하고, 질의 응답 시스템 내 여러 클래스 중 주어진 질의와 가장 적합하게 대응되는 클래스를 추론하기 위해 다중 클래스(Multi-class)에 대한 텍스트 분류(Text classification) 기법을 활용하고 있다. 자연어 질의를 입력 받아 다중 클래스로 분류하는 문제에서는 보통 클래스 내 동의 문장(Paraphrase)에서 공통적으로 나타나는 특징을 학습하는데, 이러한 학습 방법은 비슷한 단어 구성을 가진 문장들이 동일한 클래스로 분류되도록 학습하는 데에 효과적이지만, 반대로 단어 구성이 비슷하더라도 단어 한 두 개에 의해 의미가 달라지는 문장들은 구분하여 학습하기 어려운 문장 분류 시 오류가 발생하는 문제점이 있다.

표 1은 FAQ 데이터를 기반으로 텍스트 분류 문제를 학습한 모델의 실험 결과를 나타낸 것이다. 표 1을 살펴보면, 앞서 언급한 문제점과 같이 ‘스마트홈 연결 안됨’과 ‘스마트씽스 연결 안됨’이라는 문장 내 글자 구성이 매우 유사한 두 클래스가 있을 때 ‘스마트홈이 동작을 안해’라는 입력에 대해 모델이 분류를 실패하는

것을 확인할 수 있다.

표 1. FAQ 텍스트 분류 결과

사용자 질의	스마트홈이 동작을 안해
정답 클래스	스마트홈 연결 안됨
예측 클래스	스마트스위치 연결 안됨
예측 점수	0.995524

종래의 텍스트 분류는 질문과 그에 대응하는 응답에 대해서만 학습이 이루어지므로 다른 클래스에 해당하는 학습 데이터에 비슷한 의미를 가지는 질문이 존재한다면 분류에 모호성이 생겨 성능이 저하될 수 있다. 본 연구에서는 이를 해결하고자, 텍스트 분류 문제를 학습할 때 서로 다른 클래스 내의 비슷한 의미를 가지는 문장들이 서로 상이한 클래스임을 구분할 수 있도록 클래스 일치 여부(Class coincidence classification) 문제를 결합 학습(Jointly learning)하는 기법을 제안한다. 클래스 일치 여부 문제에서는 아래 두 가지 측면의 학습을 포함한다.

- 1) 동일 클래스 내 데이터 쌍 대상 클래스 일치 학습
한 클래스 내 두 개의 동의 문장을 무작위로 선택하여 문장 쌍이 서로 같은 클래스임을 학습한다.
- 2) 다른 클래스 내 데이터 쌍 대상 클래스 상이 학습
서로 다른 두 개 클래스 내의 학습 문장을 각각 무작위로 선택하여 문장 쌍이 서로 다른 클래스임을 학습한다.

이러한 두 가지 측면의 학습을 통해 동일 클래스 내 문장들의 학습은 강화하면서, 동시에 서로 다른 클래스 내 학습 문장들의 차이를 구분하는 능력도 함께 학습하도록 유도하고자 하였다.

본 논문에서는 성능 비교 연구를 위해 FAQ 분류 기본 모델로는 최근 발표되어 자연어 처리 분야에서 가장 좋은 성능을 보이고 있는 BERT [1]의 텍스트 분류 모델을 활용하였고, 실험 데이터로는 자사 FAQ 데이터를 활용하였다.

비교 실험 결과, 다중 클래스 텍스트 분류 문제만 단독으로 학습한 모델보다 본 연구에서 제안한 클래스 일치 여부 결합 학습 모델이 유사한 문장들에 대한 차이를 구분하며 유의미한 성능 향상을 보인다는 것을 확인할 수 있었다.

2. 관련 연구

FAQ 분류 연구는 최근까지 활발하게 진행되고 있다. 기계 학습 방식을 도입하여 맞춤법 오류를 최소화 하는 FAQ 분류에 대한 연구[2]나 질의와 정의된 대표 문장들 간 유사도를 바탕으로 하는 FAQ 분류 연구[3] 등을 예로 들 수 있다.

FAQ 질의 응답 시스템의 기본이 되는 다중 클래스 텍스트 분류 문제를 해결하고자 하는 기계 학습 기법은 지속적으로 연구되어 왔다. 기계 학습 모델 중 이미지 처리 문제에 주로 활용되고 있던 CNN 을 문장 분류 문제

해결 기법으로 활용하는 연구 [4]에서는, 자연어를 이미지와 같은 다중 채널 벡터로 변환하고 Convolution 과 Pooling 을 거쳐 다중 채널 벡터 내 Local feature 를 찾아 학습하고자 하였다. 시퀀스 데이터(Sequence data) 처리에 특화된 RNN 기반의 연구 [5]도 있는데, 여기서는 입력 시퀀스 데이터 내 예측에 도움을 주는 부분을 참고하는 기법을 더한 Attention RNN 모델 기반 텍스트 분류 모델을 설명하고 있다.

최근 발표된 BERT 의 텍스트 분류 모델을 포함하여, 기존에 제안된 다중 클래스 텍스트 분류 모델들은 모두 동일 클래스 내 동의 문장에서 나타나는 공통적인 특징에 의존하여 학습하고, 다른 클래스의 학습 문장들과의 연관성을 학습하는 부분은 포함하고 있지 않다. 이에 따라 비슷한 의미 혹은 비슷한 단어 구성의 유사한 문장을 클래스 내 동의 문장으로 가지는 서로 다른 클래스에 대해 분류 오류가 발생할 수 있는 문제점을 가지고 있다. 본 연구에서는 유사한 문장 간의 차이를 학습할 수 있도록 텍스트 분류 문제에 클래스 일치 여부 문제를 함께 학습하여 비슷한 의미 혹은 비슷한 단어 구성의 유사 문장을 클래스 내 학습 데이터로 가지는 서로 다른 클래스를 혼동하지 않고 구분하는 성능을 높이고자 하였다.

본 실험에서는 문장 유사도 관련 연구 [6]를 참조하여, FAQ 데이터로부터 동일 클래스 내 동의 문장의 무작위 쌍과 다른 클래스 내 학습 문장의 무작위 쌍을 생성하고, 각각 클래스 일치 여부에 따라 클래스 일치 / 불일치로 라벨링(Labeling)하여 문장 간 클래스 일치 여부를 분류 문제와 함께 학습할 수 있도록 데이터를 구성하였다. 최종적으로 FAQ 질의 응답 시스템에서 텍스트 분류와 클래스 일치 여부 문제의 결합 학습을 통해 텍스트 분류만 학습한 모델과의 성능 비교 실험을 진행하였다.

3. FAQ 분류 모델

3.1 BERT 모델

자연어 처리 분야에서 딥러닝으로부터 좋은 성능을 기대하기 위해서는 다량의 데이터가 필요하지만, 라벨(Label)이 포함된 학습 데이터를 다량으로 얻기란 쉽지 않다. 목표하는 태스크(Task)의 학습 데이터가 적다면, 유사한 태스크의 학습 데이터가 다량으로 준비가 된 경우 전이 학습(Transfer learning)을 통해 성능을 개선할 수 있다. 전이 학습이란, 특정 목적으로 사전 학습(Pre-training)된 모델을 다른 목적으로 사용하기 위해 본 학습(Fine-tuning)을 진행 하는 것이며, 사전 학습 모델의 학습 데이터가 많을수록 보다 좋은 효과를 얻을 수 있는 것으로 알려져 있다.

BERT 는 트랜스포머(Transformer) [7] 구조로 되어 있으며, 전이 학습 방식을 따르고 있어 총 두 단계로 학습이 이뤄지는데 1) 많은 양의 말뭉치로 언어 모델을 생성하기 위한 학습이 진행되는 사전 학습 단계와 2) 특정 태스크를 학습하기 위한 본 학습 단계로 진행된다. 언어 모델은 많은 양의 말뭉치로 학습이 이루어지기 때문에, 본 학습에서 적은 양의 데이터만으로도 전이

학습을 적용하지 않은 모델에 비해 상대적으로 좋은 성능을 기대할 수 있다.

3.1.1 사전 학습(Pre-training)

BERT 는 두 가지의 비지도 학습(Unsupervised learning)을 통하여 언어 모델 생성을 위한 사전 학습을 진행하는데, 비지도 학습은 데이터에 라벨이 포함되지 않아도 학습이 가능하기 때문에 많은 양의 데이터를 보다 쉽게 확보할 수 있다는 장점이 있다.

두 가지의 비지도 학습은 결합된 형태로 학습을 진행하게 되는데, 1) 가려진 단어를 유추하는 MLM(Masked Language Model)과 2) 다음 문장을 예측하는 NSP(Next Sentence Prediction)를 통해 이루어진다.

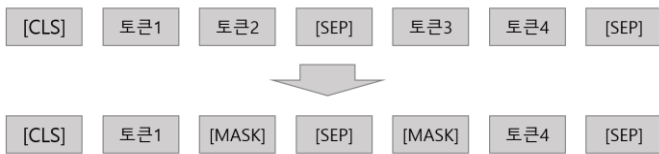


그림 1. 사전 학습 토큰 형태

MLM(Masked Language Model): 언어 모델 학습에서는 주어진 앞 뒤 단어의 정보를 바탕으로 문맥에서 찾고자 하는 목표 단어를 예측하는 CBOW 방식과 주어진 하나의 단어 주변에 문맥상 올 수 있는 단어를 예측하는 Skip-gram 방식이 주로 사용되어 왔다. 이 방식들은 이전 One-hot 인코딩(Encoding) 방식에 비해 토큰(Token)에 대한 많은 정보를 표현 할 수 있으며, 적은 차원으로 단어를 표현할 수 있다는 장점이 있다. 하지만 단어 단위의 임베딩(Embedding)만으로는 문장의 의미를 함축하는 분산 표상(Distributed Representation)을 만들기 어렵다는 문제가 있었다. BERT 는 문장의 전체 토큰 문맥을 반영하기 위하여, 입력 문장 전체의 15%에 해당하는 토큰들을 무작위로 [MASK] 토큰으로 변경하고 문장 전체 토큰을 활용하여 변경 전 토큰을 예측하는 방법으로 모델을 학습한다. 이 때, 셀프 어텐션(Self-attention) 기반인 트랜스포머 인코더(Encoder)를 사용하여 한 문장의 토큰들의 관계를 학습함으로써 문맥 정보를 반영하여 학습을 할 수 있다.

NSP(Next Sentence Prediction): NSP 에서는 질의 응답, 문장 유사도 예측 등 자연어 이해 분야에서 두 문장 관계를 파악하는 성능을 향상하고자 주어진 두 개의 문장이 문맥상 관계가 있는 문장인지 학습한다. 보다 구체적으로는, 모델의 마지막 층을 이진화 된 벡터 가중치로 구성하고 그 후 Softmax 활성화 함수(Activation function)를 사용하여 두 문장의 관계를 예측함으로써 모델을 학습한다. 두 문장의 관계 학습을 위하여 한 문단의 연속적인 두 개의 문장을 [SEP] 토큰으로 분리하거나 서로 다른 문단에서 임의로 선택된 두 개의 문장을 [SEP] 토큰으로 분리하여 학습 데이터를

구성한다. 이렇게 구성된 학습 데이터를 토대로 두 문장이 연관 관계에 있는 지를 학습한다.

제안된 모델은 BERT 와 같은 방식으로 사전 학습을 진행한다.

3.1.2 본 학습(Fine-tuning)

사전 학습을 바탕으로 생성된 언어 모델을 토대로 목표 태스크에 따른 본 학습을 진행한다. 본 연구에서 선택한 목표 태스크인 FAQ 분류 모델의 학습은 문장 분류 학습과 동일하게 진행되며, 문장 분류는 단일 문장과 그에 대응되는 클래스 라벨로 구성된 학습 데이터를 사용한다. 이 문장들은 그림 1 과 같이 사전 학습에서 사용되었던 토큰 구성과 동일하게 입력 데이터를 구성한다. [CLS]라는 특별한 토큰으로 문장 시작을 알 수 있도록 구성하며, 문장의 마지막에는 [SEP] 토큰을 삽입하여 [CLS]와 [SEP] 사이의 토큰들이 하나의 문장임을 알 수 있도록 한다. 사전 학습된 언어 모델을 기반으로 입력 토큰들을 이용해 본 학습을 진행하며, 트랜스포머 인코딩의 마지막 은닉 층에 있는 [CLS] 토큰 위치의 벡터 값을 이용해 입력에 대응되는 FAQ 클래스를 예측하는 방법으로 학습한다.

3.2 클래스 일치 여부 학습 결합 모델

CNN, RNN, BERT 등 많은 딥러닝 모델들은 FAQ 분류에 있어서 좋은 성능을 보여왔다. 하지만 서로 다른 클래스에 같은 단어가 반복적으로 중복되거나 문장 자체의 의미가 비슷한 경우 FAQ 성능이 저하 될 수 있다. 이는 각각의 문장들이 동일한 트랜스포머 인코더로 인코딩 되기 때문에 입력 문장들은 유사한 임베딩으로 변형될 것이며, 유사한 임베딩은 분류에 혼란을 야기할 수 있기 때문이다. 만약, 클래스 간의 경계를 명확하게 구분할 수 있도록 학습 문장들의 클래스 일치 여부를 고려한다면, 비록 비슷한 단어로 구성되어 있는 문장이지만 서로 다른 의미를 갖고 있는 경우에 대응할 수 있을 것이다. 본 논문에서는 이러한 가정을 기반으로 기존 FAQ 분류 학습 방법과 각 클래스 학습 데이터 간의 클래스 일치 여부 학습을 결합한 형태의 분류 모델을 제안한다.

3.2.1 모델 구조

본 논문에서 제시하고자 하는 모델은 BERT 모델을 기반으로 구성된다. 주어진 문장은 트랜스포머 인코더를 거쳐 인코딩 되며, 인코더의 마지막 은닉 계층(Hidden layer) 이후에 FAQ 분류 학습, 클래스 일치 여부 학습을 위한 두 개의 은닉 계층으로 나뉜다. 이 두 개의 은닉 계층은 하나의 입력 문장에 대해 병렬로 전방 전달(Feed forward) 및 손실 함수(Loss function)를 통해 각각의 손실을 계산하게 된다. 이후 각각의 손실은 더해지고 두

개의 은닉 계층이 연결된 형태로 학습이 진행된다. 즉, 하나의 입력에 대해 두 개의 태스크가 동시에 진행 된다.

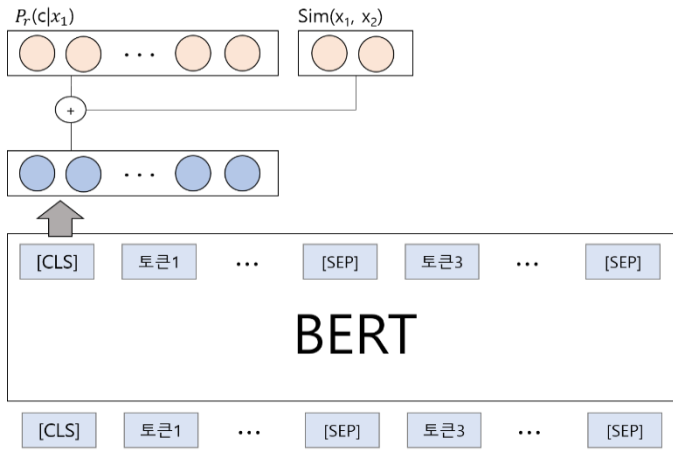


그림 2. 제안 모델의 구조

그림 2 는 제안한 BERT 기반 결합 학습 모델의 구조를 나타낸 그림이며, 각 은닉 계층 구조에 대해서는 아래에서 자세히 설명한다.

FAQ 분류 은닉 계층: 학습하고자 하는 FAQ 분류 클래스의 개수에 따라 은닉 계층의 크기가 결정되며, 트랜스포머 인코더의 [CLS] 토큰 위치의 벡터 가중치와 완전 결합(Fully connected)하게 된다. 각 클래스의 확률 값을 얻기 위하여 Sigmoid 활성화 함수 (1)를 거치게 되며, 활성화 함수에서 나온 결과값은 다시 수식 (2)와 같이 Sigmoid Cross Entropy 함수를 활용하여 손실을 계산하게 된다. 일반적인 문장 분류에서는 Softmax 활성화 함수를 사용하여 입력 질의가 특정 클래스에 분류되도록 학습하지만, FAQ 분류 특성상 예측 단계에서 입력 질의가 모든 클래스에 속하지 않는 경우 분류 배제(Rejection)가 일어나도록 학습하기 위하여 Sigmoid 활성화 함수를 사용하였다.

$$\sigma(s_i) = \frac{1}{1+e^{-s_i}} \text{ for } i = 1, \dots, K \quad (1)$$

$$CE = \sum_{i=1}^K -t_i * \log(\sigma(s_i)) - (1 - t_i) * \log(1 - \sigma(s_i)) \quad (2)$$

이 때, s_i 는 마지막 은닉 계층의 가중치 합(Weighted sum), t_i 는 라벨, K 는 클래스 개수를 뜻한다.

클래스 일치 여부 학습 은닉 계층: 클래스 일치 여부 학습은 일치, 불일치 두 개의 경우의 수만 존재하기 때문에 2 차원의 벡터 가중치로 이루어진 은닉 계층을 사용하게 된다. 이 은닉 계층 역시 FAQ 분류 은닉 계층과 마찬가지로 트랜스포머 인코더의 [CLS] 토큰 위치의 벡터 가중치와 완전 결합 후, LogSoftmax 활성화 함수 (3)를 거쳐 수식 (4)와 같이 손실을 계산하게 된다.

$$f(s_i) = \log\left(\frac{e^{s_i}}{\sum_{k=1}^2 e^{s_k}}\right) \text{ for } i = 1, 2 \quad (3)$$

$$Loss_{sim} = -Y * f(s) \quad (4)$$

여기서 Y 는 One-hot 형태의 2 차원 라벨 벡터이다.

학습된 모델로 FAQ 질의 예측 시, FAQ 분류 은닉 계층으로만 전방 전달을 통하여 예측이 진행되며 예측 단계의 손실 역시 FAQ 분류 예측 손실만 이용한다.

3.2.2 입력 토큰 구성

하나의 학습 데이터를 통해 두 개의 태스크를 학습하기 위해서는 특별한 토큰을 이용하게 되는데, 사전 학습과 같이 시작을 나타내는 [CLS] 토큰과 각 문장의 끝을 나타내는 [SEP] 토큰을 사용한다. 첫 번째 문장을 구성하는 토큰들은 [CLS] 토큰과 다음 첫 [SEP] 토큰 사이에 삽입되며, 두 번째 문장을 구성하는 토큰들은 첫 [SEP] 토큰과 두 번째 [SEP] 토큰 사이에 삽입되는 형태로 입력 데이터가 구성된다.

학습 시 위에서 언급한 것처럼 두 개의 문장이 입력되는데 첫 문장 토큰 자리에는 FAQ 분류를 위한 문장이 구성되고, 다음 문장 토큰 위치에는 주어진 첫 문장과 클래스 일치 여부 학습을 위한 다른 문장이 구성되어 입력으로 주어지게 된다. 비교를 위한 문장 선택은 일치, 불일치 두 개의 경우에 대해 편향되지 않도록 각각 50%의 해당하는 학습 문장을 이용하게 된다. 클래스 일치 여부 학습 데이터를 구성할 때에는 FAQ 학습 문장으로 선택된 클래스와 동일한 클래스에 속하는 학습 문장 중 하나를 사용하거나, 다른 임의의 클래스에 속하는 문장 중 하나를 무작위로 선택하여 사용한다.

4. 실험

4.1 실험 구성

본 실험에서는 FAQ 분류기 성능 개선을 위해 제안한 클래스 일치 여부 학습 결합 모델의 효과를 알아보기 위해, 1) BERT 의 기본 텍스트 분류 모델과 2) 제안한 구조를 더한 모델을 이용해 비교 실험을 진행했다.

실험 데이터로는 자사 FAQ 데이터를 활용했는데, 해당 데이터는 FAQ 클래스를 라벨로 가지며 각 클래스에 포함되는 동의 문장들의 집합으로 이루어져 있다. 데이터셋에 포함된 문장은 총 2734 개, 클래스는 총 166 개이고, 실험에서는 전체 데이터를 학습 데이터(2563 개)와 검증 데이터(171 개)로 나누어 학습 및 평가를 진행하였다.

두 가지 모델의 성능 비교를 위한 평가 지표로는 분류 모델에서 가장 보편적으로 활용되고 있는 F1 점수를 사용하였다. F1 점수는 정밀도(Precision)와 재현율(Recall)의 조화 평균으로, 모델이 예측한 클래스를 기반으로 계산하였다.

표 2. 실험 검증 데이터에서의 두 가지 모델의 F1 점수, 정밀도, 재현율 비교

임계값	BERT 기본			BERT 기본 + 클래스 일치 여부 학습		
	F1 점수	정밀도	재현율	F1 점수	정밀도	재현율
0.1	0.688	0.725	0.655	0.765	0.792	0.739
0.3	0.676	0.784	0.594	0.739	0.812	0.679
0.5	0.672	0.806	0.576	0.720	0.839	0.630
0.7	0.660	0.834	0.546	0.659	0.833	0.545
0.9	0.585	0.841	0.449	0.604	0.856	0.467

표 3. 실험 검증 데이터에서 제안한 모델이 효과를 보인 입력 데이터의 예시

입력 질의	정답 클래스	BERT 기본 예측 클래스	제안 모델 예측 클래스
냉동실인데 냉기가 별로 없어	냉동이 약합니다	냉동실이 동작되지 않습니다	냉동이 약합니다
냉장고에 정수기 떼어 버릴 수 있어?	지펠 냉장고 정수기 철거가 필요합니다	지펠 냉장고 정수기 설치가 필요합니다	지펠 냉장고 정수기 철거가 필요합니다
홈바가 안맞고 잠그려지지 않습니다	홈바가 안 닫힙니다	홈바가 빨리 열립니다	홈바가 안 닫힙니다

4.2 실험 결과

전체 데이터 학습 횟수인 에폭(Epoch)을 제외한 나머지 모든 학습 인자(Hyperparameter)는 BERT 에서 제안한 기본 학습 인자를 활용하였다. 에폭은 기본 모델의 본 실험 데이터 대상 성능이 가장 좋았던 50 으로 선정하였고, 마지막 은닉 계층에서 사용한 Sigmoid 값에 대해 정답 인정 여부를 결정하는 임계값은 0.1 부터 0.9 까지 0.2 단위로 수치를 달리하여 검증 데이터 171 개에 대한 두 모델의 F1 점수를 비교하였다.

표 2 를 보면, BERT 의 기본 텍스트 분류 모델과 비교하였을 때 본 논문에서 제안한 모델이 F1 점수에서 임계값 0.7 을 제외하고 다른 모든 임계값에서 낮게는 1.9%, 높게는 7.7%까지의 성능 향상을 보임을 확인할 수 있다. 임계값 0.7 의 경우, BERT 기본 모델의 성능이 더 높지만 0.1%의 차이로 사실상 두 모델 간 성능 차이가 미미하다고 볼 수 있다. 이에 따라, 클래스 일치 여부 문제를 결합한 모델은 정답 예측을 위한 임계값과 상관없이 대부분의 경우에서 성능이 향상됨을 확인할 수 있다. 특히, BERT 기본 모델의 성능이 가장 높게 측정된 에폭 50 을 기준으로 비교했음에도 불구하고 동일 실험 조건에서 제안한 모델의 성능이 전반적으로 우수하다는 것을 확인할 수 있다.

실험 데이터에서 제안한 모델이 효과를 보인 실제 입력의 예시를 표 3 에 나타내었다. 기본 모델에서는 ‘냉동’, ‘냉장고 정수기’, ‘홈바’ 와 같은 특정 표현이 나타나는 입력에 대해 동일한 표현을 가진 클래스 간 혼동이 일어났으나, 제안한 모델에서는 문장의 의미에 맞는 정답 클래스로 예측이 성공하였다. 이를 통해, 클래스 일치 여부 문제의 학습이 단어나 특정 표현으로 인한 분류 혼동을 효과적으로 해결할 수 있음을 확인하였다.

이러한 실험 결과를 바탕으로, 본 연구에서 제안한

텍스트 분류 및 클래스 일치 여부 결합 학습 기법이 FAQ 분류 모델 성능 향상에 효과적이라는 것을 확인할 수 있었다. 이는 FAQ 분류 학습 시 학습 문장들의 클래스 일치 / 불일치 여부를 추가로 학습하는 것이, 학습 문장과 그에 대응하는 라벨 데이터만을 이용하여 학습하는 것보다 클래스 간 문장들의 차이를 학습하는 데에 도움을 줄 수 있기 때문이라고 해석될 수 있다. 즉, 서로 비슷한 단어로 구성되어 있지만 다른 클래스에 속하는 문장들을 구분할 수 있도록 하는 것이 클래스들 간의 경계를 더욱 명확하게 학습할 수 있게 하여 분류 성능을 향상시킬 수 있다는 결론을 얻을 수 있었다.

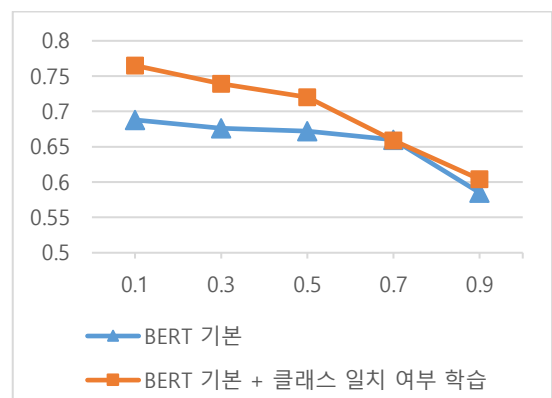


그림 3. 두 가지 모델의 F1 점수 비교 그래프

5. 결론

본 연구에서는 FAQ 분류 문제에서 비슷한 의미 혹은 비슷한 단어로 구성되어 있으나 서로 다른 클래스를 가지는 질문을 제대로 분류 할 수 없다는 문제를 해결하기 위하여 클래스 일치 여부 학습을 결합한 텍스트 분류 모델을 제안했다. 제안 모델의 효과를 알아보기 위해서 비교 실험을 수행했으며, 그 결과,

질문과 그에 대응하는 라벨 데이터만으로 학습한 분류 모델보다 쌍으로 주어진 질문들의 클래스 일치 여부를 함께 학습한 모델이 유의미한 수준으로 분류 성능을 향상시킬 수 있음을 확인할 수 있었다. 즉, 주어진 문장들이 동일한 클래스에 속하는 지 상이한 클래스에 속하는 지 판단할 수 있도록 결합 학습하는 것이, 비슷한 단어를 공유하는 질문들이 존재하는 상황에서의 FAQ 분류 성능 향상에 효과적인 방법이라는 결론을 내릴 수 있었다.

본 연구에서는 주어진 질문에 대응되는 정답이 반드시 존재하는 FAQ 상황만을 고려하였으나, 향후에는 질문에 대응되는 정답이 없는 상황에도 대응할 수 있도록 분류 배제를 고려한 FAQ 분류 연구를 진행할 계획이다.

참고문헌

- [1] J. Devlin, M. W. Chang, K. Lee and K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, In Proceedings of NAACL-HLT, pp. 4171-4186, 2019.
- [2] 장영진, 김학수, 강동호, 김세빈, 장현기, 자모 단위 합성곱 신경망 기반 맞춤법 오류가 포함된 자주 묻는 질문 자동 분류, 정보과학회논문지 vol.46 no.6, pp.564-569, 2019.
- [3] Wataru Sakata, Tomohide Shibata, Ribeka Tanaka, Sadao Kurohashi, FAQ Retrieval using Query-Question Similarity and BERT-Based Query-Answer Relevance, In Proceedings of SIGIR, pp. 1113-1116, 2019.
- [4] Yoon Kim, Convolutional Neural Networks for Sentence Classification, In Proceedings of EMNLP, pp. 1746-1751, 2014.
- [5] Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu, Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification, In Proceedings of ACL, pp. 207-212, 2016.
- [6] Daniel Cer, Mona Diab, Eneko Agirre, Inigo Lopez-Gazpio, and Lucia Speciad, SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Cross-lingual Focused Evaluation, In Proceedings of ACL, pp. 1-14, 2017.
- [7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin, Attention Is All You Need, In Proceedings of NIPS, pp. 6000-6010, 2017.