韓國IT서비스學會誌
第19卷 第5號
2020年 10月, pp.83-91

Journal of Information Technology Services
https://doi.org/10.9716/KITS.2020.19.5.083

사전 학습된 한국어 BERT의 전이학습을 통한 한국어 기계독해 성능개선에 관한 연구

이치훈* · 이연지** · 이동희***

A Study of Fine Tuning Pre-Trained Korean BERT for Question Answering Performance Development

Chi Hoon Lee* · Yeon Ji Lee** · Dong Hee Lee***

■ Abstract ■

Language Models such as BERT has been an important factor of deep learning-based natural language processing. Pre-training the transformer-based language models would be computationally expensive since they are consist of deep and broad architecture and layers using an attention mechanism and also require huge amount of data to train. Hence, it became mandatory to do fine-tuning large pre-trained language models which are trained by Google or some companies can afford the resources and cost.

There are various techniques for fine tuning the language models and this paper examines three techniques, which are data augmentation, tuning the hyper paramters and partly re-constructing the neural networks. For data augmentation, we use no-answer augmentation and back-translation method. Also, some useful combinations of hyper parameters are observed by conducting a number of experiments. Finally, we have GRU, LSTM networks to boost our model performance with adding those networks to BERT pre-trained model.

We do fine-tuning the pre-trained korean-based language model through the methods mentioned above and push the F1 score from baseline up to 89.66. Moreover, some failure attempts give us important lessons and tell us the further direction in a good way.

Keyword : Language Model, Masked Language Model, Question Answering, BERT, Fine-Tuning

Submitted : September 2, 2020 1st Revision : September 29, 2020 Accepted : October 18, 2020

* 티쓰리큐(주) 인공지능연구소장
** 티쓰리큐(주) 인공지능연구소 선임연구원
*** 국민대학교 경영학부 교수, 교신저자

1. 서 론

인공지능의 발전과 함께, 자연어처리에 관한 기술 연구도 통계적인 접근법을 넘어 딥러닝 기반으로 활발히 진행되고 있다. 특히, 자연어처리 분야 중에서 기계독해(Machine Reading Comprehension)는 주어진 본문과 질문을 컴퓨터가 이해하고 스스로 해답을 제시하는 인공지능 태스크로 질의응답, 챗봇, 계약서 및 약관리스크 분석, 그리고 금융권의 컴플라이언스 등 다양한 분야에서 응용될 수 있다.

2018년부터 현재까지 RNN 기반의 언어모델을 보완한 트랜스포머 언어모델인 BERT의 등장은 자연어처리 기술에 획기적인 발전을 가져왔다. 트랜스포머 기반 모델의 사전학습과 전이학습은 기계번역, 문장 분류, 기계 독해, 개체명 인식 등의 다양한 자연어 처리 분야에서 SOTA(State-Of-The-Arts)를 달성하는 업적을 이루었다. 특히, 기계 독해의 객관적인 지표로 활용되는 SQuAD 데이터셋의 리더보드에서는 BERT 기반의 사전학습 모델이 대부분을 차지하고 있을 정도로 BERT의 대외적인 성능과 활용도는 실질적으로 입증되고 있다.

하지만 BERT 기반 사전모델의 발전은 자연어처리의 특성상 특정 언어에 국한되는 경우가 많은데, 단적인 예로 해외에서 공개되는 다국어 사전모델은 위키백과 데이터로만 학습했기 때문에 데이터양이 상당히 부족하고 성능에 한계가 있을 수 밖에 없다. 이러한 문제점을 해소하기 위해 LG CNS, SK, ETRI, Twoblock AI 등의 한국 기업들이 딥러닝 기반 언어모델로의 변화에 적응하며 한국어에 특화된 대량의 데이터셋을 구축하여 배포하고, 학습데이터로 활용한 사전모델을 공개하고 있다.

BERT는 언어의 구조, 표현, 형태 등을 모두 학습할 수 있고, 이를 토대로 기계 독해뿐만 아니라 다양한 자연어처리 태스크에서 우수한 성능으로 두각을 나타내고 있다. 하지만, 여전히 언어학적으로 BERT의 연구 분야는 무궁무진하다. 이러한 특성은 사전학습보다 전이학습을 할 때, 더욱 강조될 수 있는데 그 이유는 진입장벽이 낮음과 동시에 확장성이 높기 때문이다. 특정 언어와 도메인에 특화된 사전학습을 하려면, 엄청난 시간, 자원, 비용과 방대한 데이터가 필요하다. 하지만 이에 반해 전이학습은 제한된 자원과 비교적 적은 데이터로 다양한 도메인과 태스크에 최적화하기 쉽다. 스탠포드 대학 자연어처리연구실의 2019년 수상 논문은 모두 BERT와 전이학습에 관련된 내용인데, 여기서 우리는 전이학습이 사전학습과 비교하면 접근성이 훨씬 뛰어나고 제한이 없어, 더 많은 연구가 가능하다는 것을 증명할 수 있다.

이렇게 전이학습의 중요성을 강조하며, 이 논문에서는 먼저 BERT 전이학습에 관한 내용을 전반적으로 다루면서 성능 개선을 위한 파인튜닝 기법을 중점으로 소개한다. 후반부에는 이 기법들을 활용한 실험 및 결과를 비교하여 설명한다.

2. 관련 연구

Wang(2019)의 논문에서는, BERT에 High-way LSTM과 DenseNet 네트워크를 각각 추가하여 앙상블 파인튜닝을 한 방법(stack-only)과 하이퍼파리미터 미세 조정(finetune-only) 두 가지 기법을 비교하였는데 앙상블 모델이 정확도와 F1 Score에서 모두 우세한 경향을 보였다.

같은 맥락에서, Ying(2019)과 Qin et al.(2019)은 BiDAF와 BERT를 앙상블하거나 BERT 임베딩과 다양한 알고리즘을 융합한 연구를 진행하였다. Dodge et al.(2020)의 연구에서는 전반적인 파인튜닝 기법에 대해 소개하였는데, 적은 데이터셋일 경우 Random Seed의 영향이 크며, 더 많이 할당할수록 성능이 개선된다는 점을 발견하였다. 또한, Correlation을 구해 성능이 좋지 못한 모델을 학습 초기에 구분하여 학습을 중지하는 등의 알고리즘을 제시하고 Early Stopping으로 4개의 데이터셋의 성능 개선을 증명하였다.

또한 이번 실험에서 활용한 No-Answer 증식 방법은 Semnani et al.(2019)의 연구에서 사용한 증식 방법으로 한국어에 적용해보았으나 방법이

정확하게 명시되어 있지 않아 직접 SQuAD 데이터셋을 분석하여 시행착오를 겪기도 하였다. 비슷하게 BERT가 아닌 Bi-LSTM RNN LM을 활용한 증식에 대한 연구로 Kobayashi(2018)는 긍정의 의미를 부정 단어들과 높은 확률로 연관 지어 추론하는 이슈를 해결하기 위해 conditional constraint 활용하고, 언어모델을 label-conditional LM으로 수정하여 문맥 기반의 증식을 성공시켰다.

기계독해 이외에도 다른 Downstream Task와 관련된 파인튜닝 기법이 연구된 사례도 많았다. Sun(2019)은 하이퍼 파라미터 튜닝, 레이어 수정, 긴 문장에 대한 해결 방법을 제시하며 문서 분류와 감성 분석 모델을 생성해 기존 연구들과 성능 비교를 제시했다.

3. 이론적 배경

3.1 Text Embedding

텍스트에 대한 분석을 하기 위해서, 자연어를 컴퓨터가 이해하기 위해서 수치화가 필요하다. 이 과정이 바로 벡터 공간에 텍스트를 적용하는 임베딩이다. 임베딩 방법은 단어의 의미가 어디서나 변하지 않는 Static Embedding과 문장 기반으로 문맥마다 동적으로 벡터를 생성하는 Contextual Embedding으로 나누어진다(Ethayarajh, 2019).

Word2vec, Glove 등의 단어 기반 임베딩 기법이 Static Embedding에 해당하는데, Static Embedding을 이용하여 임베딩을 하게 되면, 단어의 벡터 표현 값이 고정되어 있기 때문에 문서 텍스트의 어느 부분에서 등장하던지 그 벡터 값이 동일하다. Static Embedding은 단어 하나에 해당하는 여러 의미들도 오직 하나의 벡터 표현 값으로 임베딩하고, 단어의 순서를 고려하지 않은 채 학습한다. 그렇기 때문에, 문맥과 의도는 다르지만 모양이 같은 단어 집합들이 고정된 하나의 벡터 표현 값을 가지게 되면서 동음이의어가 구분되지 않는 단점이 존재한다.

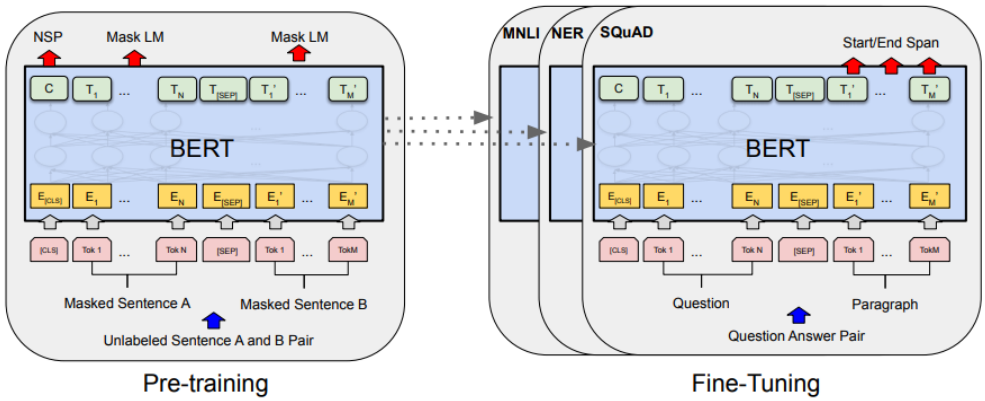
이러한 Static Embedding 기법의 단점을 해소하면서 문맥을 좀 더 깊이있게 표현할 수 있는 개념이 Contextual Embedding이다. ELMo, BERT 등의 언어모델의 문장 기반 임베딩 기법이 이에 해당하며 동음이의어가 각 문장별로 다른 벡터 표현 값을 가질 수 있도록 단어, 문장의 순서와 위치 정보, 주변 단어들을 학습한다. 또한 BERT의 경우, 사전학습 모델의 임베딩 레이어를 수정하여 적절한 전이학습을 통해 특정 도메인 태스크에 최적화된 벡터 표현 값을 가질 수 있다.

3.2 BERT

BERT는 여러 개의 어텐션 헤드를 가진 수십 개의 레이어로 이루어 진 트랜스포머 기반 언어모델로 사전학습, 전이학습을 통해 다양한 자연어 처리 태스크에 활용하는 취지에서 제일 처음 Google ai Research 팀에서 개발되었다. 사전학습의 경우, 레이블링 되지 않은 대량의 데이터에 대해 학습하는데 데이터의 일반적인 특성을 배우게 된다. 학습 시에는 학습할 데이터의 일부를 마스킹하여 모델이 마스킹한 부분을 맞추게 하는 기법인 Masked Language Modelling과 질의응답 태스크에 중요한 역할을 하는 문장별 관계를 학습하는 Next Sentence Prediction이 적용되었고, 양방향 인코더를 사용하여 모든 토큰에 양방향으로 어텐션 메커니즘이 동작하도록 설계되어 있다(Devlin et al., 2018). 하지만 사전학습의 경우 시간과 비용에 소요가 필요하기 때문에, 진입장벽이 상당히 높은 편이다.

BERT 기반의 사전학습은 국내외로 활발하게 공개되고 있으며, 이러한 사전모델을 파인튜닝하여 실제 태스크에 적용하는 것이 당연해질 정도로 자연어처리 분야에서 트랜스포머 언어모델은 필수적으로 자리 잡았다.

해외사례로는 구글이 영어 및 다국어로 사전 학습시킨 BERT-base, BERT-large 모델이 있으며 국내에서는 다양하고 대량의 한국어 데이터를 활용한 SKT의 Kobert, ETRI의 Korbert 등이 공



[그림 1] 사전 학습기반의 언어모델(BERT)

개되어 있다. 실제로 많은 공수를 들여 사전모델을 학습시키지 않아도, 여러 파인 튜닝 기법들을 적절히 조합하여 특정 태스크에서 만족할만한 성능을 충분히 보여주기 때문에 더욱 연구가 활발하게 진행되고 있다.

3.3 Fine Tuning 기법

파인 튜닝은 이미 특정 도메인에 사전 학습된 모델을 다른 태스크에 활용하기 위해 학습시키는 과정이다. 사전모델을 사용할 경우, 이미 기본적인 데이터의 특징을 학습했기 때문에 유사한 특징을 가진 데이터의 양이 적어도 파인튜닝을 하여 모델을 도메인에 최적화시키고 성능을 개선할 수 있다. 파인튜닝을 하는 방법은 데이터를 활용한 방법, 하이퍼 파라미터를 미세 조정해서 성능을 개선하거나 사전모델 구조에 추가적으로 네트워크를 구성하는 등의 다양한 방법이 존재하고 이 방법들을 동시에 적용할 수도 있다.

이 논문에서 적용한 파인튜닝 방법을 3가지 종류로 나누자면, 데이터 증식, 하이퍼 파라미터 튜닝, 파인튜닝 네트워크 추가로 각 방법에 대한 자세한 설명은 다음과 같다.

3.3.1 데이터 증식

자연어처리에서 딥러닝 기반 모델은 당연하게도 양질의 데이터가 보장되어야 한다. 너무 적거나 정제하지 않은 데이터는 오버피팅 등의 모델의 성능 저하를 초래하기도 한다(Kobayashi, 2018). 다른 분야에 비해 자연어처리 분야의 데이터 증식 방법은 많이 알려지지 않았으며, 나라마다 다양한 언어의 특성을 모두 똑같은 증식법을 적용할 수 없기 때문에 표준화에 난항을 겪기도 한다. 또한 증식을 할 때, 언어학적으로 뜻하는 의미가 유실되지 않기 위해 단어와 문장의 하위, 상위, 유사, 반의 관계 등을 모두 고려해야 하는 점도 상당히 어려운 부분이다. 이러한 텍스트 데이터 증식은 유사어를 치환하는 단순한 방법부터, 일부 노이즈를 사용하여 유사 데이터를 생성하는 Adversarial 증식까지 활발히 연구되고 있다.

3.3.2 Hyper Parameter Fine-Tuning

모델은 같은 하이퍼 파리미터 값을 가지고 있어도 결과가 다르게 나올 수 있다. 사전 언어 모델의 극히 일부에 해당하는 파라미터들을 조정해서 성능이 개선될 수 있는데, 그러기 위해서 여러 가지 조합을 실행해야 하고 자원이 어느 정도 지원되느냐에 따라 시간 비용이 증감할 수 있다. BERT를 포함한 사전모델에서는 Epoch 수를 조정하거나, Batch Size, Random Seed, Learning Rate 값을 조정하는 것이 여러 연구에서 활용한 공통적인 방법이다. 이 논문에서도 마찬가지로 Epoch, Batch

Size, Learning Rate, Random Seed, Max Sequence Length를 미세 조정하여 약 30개의 조합으로 실험을 수행하였다.

3.3.3 Recurrent Neural network

이전 단어를 기억하는데 제한점이 많은 전통적인 언어모델을 개선한 RNN(Recurrent Neural Networks)은 말뭉치의 모든 이전 단어들에 대해 학습을 할 수 있다. RNN은 히든 레이어의 노드가 방향을 가지고 연결되어 있어 순환 구조를 이루는 신경망으로, 이전 state 정보가 다음 state 정보를 예측하는데 사용된다. 시계열 데이터, 순서가 보장되어야 하는 텍스트 데이터 등에 효과적이다. 하지만 RNN은 Short Term memory의 약점을 가지고 있다. 데이터의 시퀀스가 길어질수록 정보를 가지고 현재 state까지 오는데 역전파 과정에서 기울기 소실이 발생한다. 가중치들이 너무 작아 업데이트가 제대로 되지 않으면 결국 학습이 멈추는 것이다. 이러한 단점을 해결하기 위한 네트워크가 바로 LSTM과 GRU인데, 내부적으로 '게이트'라는 개념이 추가되어 정보의 흐름을 조절하고 규제한다(Mohammadi et al., 2019).

1) LSTM

LSTM은 RNN 기반의 유사한 흐름을 가지고 있다. 하지만 RNN보다는 복잡하고, 셀상태와 연관된 게이트에 대한 개념이 중요하게 동작한다. 히든 레이어의 메모리 셀에 입력 게이트, 망각 게이트, 출력 게이트를 각각 추가하여 학습하는 동안 게이트들이 셀에 필요하지 않은 정보는 지워버리고, 중요한 정보로 판단되는 것들은 기억하도록 한다. 언급된 3개의 게이트에서는 공통적으로 출력 값이 0과 1사이의 값이 나오는 시그모이드 활성함수를 사용한다.

2) GRU

LSTM와 유사한 GRU는 RNN 기반의 네트워크 중 가장 최신의 개념이다. 셀상태를 사용하지 않으며 은닉 상태를 정보를 전달하는데 활용한다. 또한 LSTM과 달리 3개가 아닌 리셋과 업데이트 2개의 게이트를 가지고 있다. 리셋 게이트는 이전의 정보 중에 잊어버려도 되는 정보를 결정하고, 업데이트 게이트는 LSTM의 망각, 입력 게이트와 비슷한 역할로 다음 상태의 셀에 추가될 정보가 필요한지 판단한다. GRU와 LSTM은 성능 면에서 도메인 태스크마다 비슷하기 때문에, 데이터와 모델 최적화에 따라 여러 실험을 거쳐 적합한 알고리즘을 적용해야 한다는 의견이 분분하다.

4. 실험 및 결과

4.1 데이터셋

KorQuAD 1.0 데이터셋을 파인튜닝 학습에 적용하였다. KorQuAD 1.0는 LG CNS에서 한국어 위키백과 문서를 바탕으로 개발한 한국어 기계독해 데이터셋이다. 한국어 위키 백과 랜덤 탐색으로 총 1637건의 문서를 수집하고, 크라우드소싱을 통하여 수작업으로 70,000쌍 이상의 질문-답변 학습 데이터셋을 구축하였다. JSON 형식의 학습 데이터셋은 총 7 Depth로 구성되어 있으며, 한 개의 위키백과 타이틀에 해당하는 단문들이 여러 쌍의 질문-답변 리스트를 가진 형태이다(임승영 외, 2018). 특히, 좋은 품질의 데이터 셋 구축을 위해 위키백과의 알찬 글, 좋은글 등의 텍스트 데이터를 우선순위에 두고 제작되었다. KorQuad 1.0 데이터 셋에 대한 BERT 모델의 공식적인 baseline F1 Score는 82.20이다.

4.2 적용 모델

적용언어모델은 2020년 5월 인라이플과 LG CNS가 NLP 챌린저용으로 공개한 한국어 BERT소형 모델을 채택하였다. 구글 BERT의 경우 MLM기법으로 15% 랜덤 마스킹 또는 전체 Word 마스킹을 사용하지만, 해당 모델의 경우 n-gram 마스킹을 사용하였다. 또한, Next Sentence Prediction이 아닌 Sentence Order Prediction을 사용하여 좀 더 효율적으로 문맥을 해석하도록 구성하였다.

〈표 1〉 적용 모델 상세정보

Hidden	Layer	Attention-Head
256	12	8

Sub-word 레벨은 형태소 단위이며 모델 상세 정보는 <표 1>과 같다.

4.3 실험 방법

4.3.1 데이터 증식

데이터 증식 방법은 2가지를 사용하였다. 첫번째로 KorQuAD 1.0 학습 데이터셋을 SQuAD 2.0 데이터셋처럼 답이 없는 (No-Answer) 질의를 증식시켰다. 예를 들어, 위키백과의 한 주제에 해당하는 아티클 페이지를 보면, 문단이 여러개로 이루어져 있다. 각각 문단에는 해당하는 질문-정답 셋이 있는데, 그 질문-정답 셋을 정답이 있는 문단을 포함하여 모든 문단에 넣어 증식하는 방법이다(Semnani et al., 2019). 이렇게 되면, 정답이 아닐 경우 맞출 확률이 높아질 수 있고, 데이터의 양을 대폭 증가시킬 수 있다. 실제로, 이 방법을 사용했을 때 데이터가 260MB로, 기존 데이터셋에 비해 약 10배 정도 증가하였다. 두 번째로 Back Translation을 활용한 증식을 수행하였다. 이 증식법도 마찬가지로 질의 데이터에만 적용하였는데, 한글 데이터를 영어로 번역 후 다시 한글 데이터로 번역하는 방법이다. 번역시 데이터의 의미는 같으나 문체와 어휘가 달라지는 원리를 이용한다. <표 2>와 같이 증식 모델은 성능 개선효과는 미미하였다.

〈표 2〉 데이터증식모델과 소형모델의 성능비교표

	Epoch	F1 Score
Baseline	4	88.20
Aug-Model	4	88.35

4.3.2 파라미터 튜닝

하이퍼 파라미터 튜닝은 Batch size, Learning rate, Maximum sequence length, Epoch, random seed, Maximum query length 7개의 항목을 중심으로 약 30개의 조합을 수행하였으며, 최상위 모델 5개만 도출하였다. 하이퍼 파라메터 튜닝으로 오히려 성능이 감소하는 것으로 나타나 Baseline모델이 이미 최적화된 것으로 판단된다.

4.3.3 네트워크 설계

네트워크는 사전학습 모델에 Linear 레이어를 추가하여 Epoch과 Layer 개수에 변화를 주어 실험을 하였다. 이후에는 RNN 기반의 LSTM, GRU를 각각 기존 사전 모델과 앙상블 하여 파인 튜닝을 수행하였다. 마찬가지로 Layer수와 Epoch 수, 양방향 여부를 변경하며 실험을 진행하였다.

〈표 4〉 Linear 레이어 추가모델 성능비교표

	추가 레이어 수	Epoch	F1 Score
Case 1	2	4	87.85
Case 2	2	8	87.85
Case 3	1	4	87.12

〈표 3〉 하이퍼 파라미터 미세조정 모델 성능 비교표

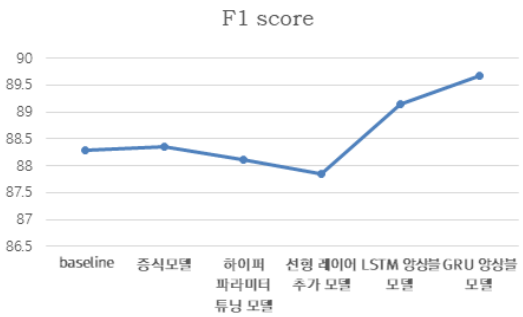
	Batch size	Learning rate	Max_seq_length	Epoch	Seed	Max_query_length	F1 Score
Baseline	16	5.00E-05	512	4	42	96	88.20
Case 1	16	5.00E-05	512	4	42	64	88.11
Case 2	16	5.00E-05	512	4	1000	96	87.84
Case 3	16	5.00E-05	512	8	1000	96	87.64
Case 4	8	3.00E-05	384	3	42	96	87.53
Case 5	32	2.00E-05	512	10	42	96	87.45

〈표 5〉 LSTM 네트워크 앙상블모델 성능비교표

	양방향	Epoch	레이어	F1 Score
Case 1	False	4	2	88.50
Case 2	True	4	2	88.90
Case 3	True	3	1	88.70
Case 4	**True**	**4**	**1**	**89.14**

〈표 6〉 GRU 네트워크 앙상블모델 성능비교표

	양방향	Epoch	레이어	F1 Score
Case 1	False	4	1	86.33
Case 2	True	4	1	89.36
Case 3	True	4	2	89.54
Case 4	**True**	**3**	**2**	**89.66**



[그림 1] 파인튜닝 모델별 F1 Score 추이그래프

파인튜닝을 수행한 3가지 방법 중 네트워크를 수정하거나 추가한 방법으로 가장 좋은 결과를 도출할 수 있었다. 기존 모델의 Baseline F1 Score는 88.20로, 하이퍼 파라미터 미세조정과 데이터 증식 방법을 통한 모델은 모두 Baseline 안팎으로 적은 범위의 미세한 성능 차이를 보였다. 그렇지만 LSTM과 GRU 네트워크를 앙상블한 모델은 Baseline보다 1.6% 정도의 F1 Score가 대폭 증가했다. 또한, 기계독해 태스크의 특성을 고려하여 양방향 학습을 수행한 결과 단방향일 때보다 크게는 3% 성능 개선의 효과가 있었다.

5. 결 론

BERT를 비롯한 최근 공개되는 언어모델은 대량의 말뭉치를 사전 학습된 상태로 제공된다. 그 이유는 첫째, 사전학습을 위해 엄청난 비용과 시간이 소요되며, 둘째, 사용자 입장에서는 소량의 특정 도메인 데이터만을 전이학습을 수행하여 비즈니스 요구를 충족하는 것이 가능하기 때문이다. 따라서 파인튜닝을 통한 성능개선의 접근방식이 점차적으로 중요해지고 있다.

이번 연구에서 네트워크를 추가한 앙상블 모델의 성능이 가장 좋았으나, 데이터 증식에 대한 정제가 제대로 수행되어 이 논문에서 실험한 모든 파인 튜닝 방법을 같이 적용한다면 더 획기적인 성능 개선의 효과를 볼 수 있을 것이라 기대한다.

특히, F1 Score 88.20%에서 89.66%로의 성능개선은 BERT의 사전모델을 보완하지 않고 순수한 파인튜닝만의 결과로 기계독해 분야 Task에 비용과 시간을 상당부분 줄일 수 있다고 생각한다.

이번 실험을 통해 여러 파인튜닝 기법을 적용해 볼 수 있었으나, 모델의 결과를 볼 때 그 원인을 분석하는 데 어려움이 있었다. 모델 결과에 사전학습 모델이 영향을 주는지 아니면 파인튜닝의 과정이 더 영향을 주는지 파악하기가 쉽지 않았지만, 더욱 다양한 실험과 연구를 통해 앞으로 파악해야 할 과제라고 생각한다. 또한, 한국어 기계독해 및 질의응답 등 한국어 데이터셋으로 학습된 한국어 언어모델을 최적화 하는 것뿐만 아니라, 실험을 진행하면서 한국어 자연어처리에 대한 많은 발전 가능성을 보았다. 특히, 한국어 자연어처리에 대한 연구가 부족한 상황에서 한국어에 최적화된 모델은 한국 연구자들에 의해 나와야 한다고 생각하며, 그런 의미에서 이번 논문은 공동 발전의 의미가 있다.

참고문헌

임승영, 김명지, 이주열, “KorQuAD : 기계독해를 위한 한국어 질의응답 데이터셋”, 한국정보과학회 학술발표논문집, 2018, 539-541.

Clark, K., U. Khandelwal, O. Levy, and C. Manning, “What Does BERT Look At? An Analysis of BERT’s Attention”, Stanford University,

Facebook AI Research, 2019. Available at https://arxiv.org/abs/1906.04341 (Accessed June 11, 2019).

Devlin, J., M.W. Chang, K. Lee, and K. Toutanova, "BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding, Google AI Language", 2018. Available at https://arxiv.org/pdf/1810.04805.pdf (Accessed May 24, 2019).

Dodge, J., G. Ilharco, R. Schwartz, A. Farhad, H. Hajishirzi, and N. Smith, "Fine-Tuning Pretrained Language Models : Weight Initializations, Data Orders, and Early Stopping", Cornell University, 2020. Available at https://arxiv.org/pdf/2002.06305.pdf (Accessed February 15, 2020).

Ethayarajh, K., "How Contextual Are Contextualised Word Representations? Comparing The Geometry of Bert, Elmo, And Gpt2", Stanford University, 2019. Available at https://arxiv.org/abs/1909.00512 (Accessed September 2, 2019).

Kobayashi, S., Contextual Augmentation : Data Augmentation By Words With Paradigmatic Relations, Preferred Networks, Inc., 2018. Available at https://arxiv.org/abs/1805.06201 (Accessed May 16, 2018).

Lalande, K.M., "CS224n Final Project : SQuAD 2.0 with BERT", 2019. Available at http://web.stanford.edu/class/cs224n/reports/default/15791990.pdf (Accessed September 5, 2020).

Marivate, V. and T. Sefara, *Improving short text classification through global augmentation methods*, CD-MAKE 2020 : Machine Learning and Knowledge Extraction, 2019, 385-399.

Mohammadi, M., R. Mundra, R. Socher, L. Wang, and A. Kamat, "Natural Language Processing With Deep Learning", Stanford University, 2019. Available at http://web.stanford.edu/class/cs224n/readings/cs224n-2019-notes03-neuralnets.pdf (Accessed June 10, 2020).

Qin, Z., W. Mao, and Z. Zhu, "Diverse Ensembling with Bert and its variations for Question Answering on SQuAD 2.0", 2019. Available at pdfs.semanticscholar.org/728e/855946e2683dd34fe8eb165f223059cb2961.pdf (Accessed October 10, 2020).

Semnani, J.S., R.K. Sadagopan, and F. Tlili, "BERT-A : Fine-tuning BERT with Adapters and Data Augmentation", Standford University, 2019. Available at https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1194/reports/default/15848417.pdf (Accessed June 10, 2020).

Sun, C., X. Qiu, Y. Xu, and X. Huang, How To Fine-Tune BERT For Text Classification?, Shanghai : Fudan University, 2020. Available at https://arxiv.org/pdf/1905.05583.pdf (Accessed June 10, 2020).

Wang, R., H. Su, C. Wang, K. Ji, and J. Ding, "To Tune or not tune? How about the best of both worlds?", Percent Group, AI Lab. 2019. Available at https://arxiv.org/pdf/1907.05338.pdf (Accessed Oct 10, 2020).

Yang, W., Y. Xie, L. Tan, K. Xiong, M. Li, and J. Lin, "Data Augmentation for BERT Fine-Tuning in Open-Domain Question Answering", 2019. Available at https://arxiv.org/pdf/1904.06652.pdf (Accessed April 14, 2019).

Ying, A., "Really Paying Attention : A BERT+ BiDAF Ensemble Model for Question-Answering", Standford University, 2019. Available at https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1194/reports/default/15792214.pdf (Accessed June 10, 2020).

About the Authors

이 치 훈 (chihoon.lee@t3q.com)

성균관대학교 전기・전자・컴퓨터공학 박사학위를 취득하였다. 삼성SDS에서 IT컨설팅 업무를 수행하였고, 현재는 티쓰리큐(주)에서 AI연구소장으로 AI서비스개발 및 연구과제 총괄을 맡고 있다. 주요관심분야는 딥러닝 기반 자연어 처리 및 영상인식기술, 유사이미지 검색 등이며, 현재 정보화진흥원 법률문서 리스크분석과제를 수행하고 있다.

이 연 지 (jennylee03@t3q.com)

Royal Holloway, University of London 컴퓨터과학(학사), 현재 티쓰리큐(주) AI연구소에서 선임연구원으로 재직 중이다. 주요 연구분야는 딥러닝 기반 자연어 처리 분야이고, 주요 관심분야는 텍스트 증식, 언어모델, 기계독해이다.

이 동 희 (donghl917@.naver.com)

국민대학교 경영학부 교수 (마케팅, 리더십), 국민대학교에서 경영혁신 전공으로 박사학위를 취득하였다. 삼성SDS에서 해외사업부장, e삼성 일본,인도 총괄, 마케팅홍보사업부장으로 근무하고 ㈜펜타크리드(SW개발) 대표이사를 역임하였다. 한국 창업교육협의회장, 한국블록체인경영학회장을 맡고 있으며, 과기부 국가디지털 전환사업심의위원, 서울경찰청 스마트치안자문단으로 활동 중이다.