

Continual Learning을 이용한 한국어 기계독해

신중민^o, 조상현, 권혁철
부산대학교 정보융합공학과

tiswndals13@naver.com, delosycho@gmail.com, hckwon@pusan.ac.kr

Korean Machine Reading Comprehension using Continual Learning

JoongMin Shin^o, Sanghyun Cho, Jaehoon Choi, Hyuk-Chul Kwon
Dept. of Information Convergence Engineering Pusan National University

요 약

기계 독해는 주어진 지문 내에서 질문에 대한 답을 기계가 찾아 답하는 문제이다. 딥러닝에서는 여러 데이터셋을 학습시킬 때에 이전에 학습했던 데이터의 weight 값이 점차 사라지고 사라진 데이터에 대해 테스트 하였을 때 성능이 떨어진 결과를 보인다. 이를 과거에 학습시킨 데이터의 정보를 계속 가진 채로 새로운 데이터를 학습할 수 있는 Continual learning을 통해 해결할 수 있고, 본 논문에서는 이 방법을 MRC에 적용시켜 학습시킨 후 한국어 자연어처리 Task인 Korquad 1.0의 MRC dev set을 통해 성능을 측정하였다. 세 개의 데이터셋중에서 랜덤하게 5만개를 추출하여 10stage를 학습시킨 50K 모델에서 추가로 Continual Learning의 Learning without Forgetting를 사용하여 학습시킨 50K-LWF 모델이 F1 92.57, EM 80.14의 성능을 보였고, BERT 베이스라인 모델의 성능 F1 91.68, EM 79.92에 비교하였을 때 F1, EM 각 0.89, 0.22의 향상이 있었다.

주제어: 기계 독해, Continual Learning

1. 서론

MRC(Machine Reading Comprehension) 기계 독해는 주어진 지문 내에서 질문에 대한 답을 기계가 찾아 답하는 문제이다.

모델은 Training 과정에서 지문과 질문과 답이 포함되어 있는 MRC 데이터셋을 학습한 후에 Test과정에서 MRC 데이터셋으로 모델을 평가한다. 대표적인 한국어 MRC 데이터셋으로 KorQuad 1.0[1], KorQuad 2.0[2], AIHub 기계독해, AIHub 도서자료 기계독해, KLUE-MRC[3] 등이 있다. Training 데이터셋과 Test 데이터셋은 동일한 데이터셋일수도 다를수도 있는데, 본 논문에서는 학습 데이터셋으로써 KorQuad 1.0, AIHub 기계독해, AIHub 도서자료 기계독해를 사용하였고, 평가 데이터셋으로써 가장 처음 학습시킨 데이터셋 KorQuad 1.0 dev set을 사용하였다.

기존의 딥러닝의 Training과정에서 여러 데이터셋 학습 후 테스트시 이전에 학습했던 데이터에서 생성된 weight들이 점차 사라짐과 동시에 성능이 떨어지는 catastrophic forgetting[4] 현상이 나타난다. 이러한 문제를 Continual Learning 학습 방법을 통해 해결할 수 있다.

본 논문에서는 Continual learning을 MRC에 적용하여 문제를 해결해보고자 하였다. 세 개의 데이터셋들을 총 5개의 방법으로 BERT[5]에 학습시켜 모델을 생성한 뒤 3개의 데이터셋에서 랜덤하게 3만개를 추출해 학습시킨 model_30k, 5만개를 추출한 model_50k, model_50k에서

LWF방법으로 추가학습시킨 model_50k_LWF, epoch마다 각각 데이터셋 전체를 하나씩 학습시키는 model_epoch, 각 데이터셋들중 데이터 하나씩 순차적으로 계속 학습시키는 model_multitask, 총 5개 모델을 테스트하였다.

2. 관련 연구

기존의 딥러닝의 Training 과정에서 여러 데이터셋 학습 후 테스트시 이전에 학습했던 데이터에서 생성된 weight들이 점차 사라짐과 동시에 성능이 떨어지는 catastrophic forgetting 현상이 나타난다. 이러한 문제를 Continual Learning을 통해 해결할 수 있다. Continual Learning에 대한 연구는 [4]가 있다. Continual Learning은 새로운 데이터를 학습하면서도 동시에 기존에 학습된 데이터의 정보를 손실시키지 않고 가지고 있도록 하는 것이다. 대표적으로 Fine-tuning, EWC, LWF가 있다.

Fine-tuning은 가장 기본적인 방법으로 이전 stage에서 학습된 모델 파라미터를 사용해 다음 stage의 데이터를 tuning해 나가는 방법이다.

EWC(Elastic Weight Consolidation)[4]는 구글 딥마인드에서 제시한 방법으로 이전 stage에서 학습된 모델 파라미터들마다 중요도를 설정하고 그 수치의 크기만큼 다음 stage의 학습에서 변하지 않도록 하는 방법이다. 본 논문에서는 EWC는 사용하지 않았다.

LWF(Learning without Forgetting)[7]는 현재 stage의 데이터를 학습할 때 이전 stage에서 학습된 모델로 현재

stage의 데이터를 예측한 뒤, 이 예측값을 현재 학습의 label로써 포함해 사용하는 것이다.

3. Continual Learning 기반 질의응답

본 논문에서는 KorQuad 1.0, AIHub 기계독해, AIHub 도서자료 기계독해 곧 총 세 개의 태스크를 돌아가며 학습시키는 Continual Learning을 활용하여 모델을 학습시키고, KorQuad 1.0 dev 데이터셋에서 모델의 성능평가 테스트를 진행하였다.

Continual Learning을 통해 만들어진 모델은 총 네 개로, 30K, 50K, 1epoch, multitask가 있다. 30K는 각 데이터마다 데이터 삼만개씩 랜덤으로 추출하여 학습시켰고, 50K는 각 데이터마다 5만개씩 랜덤으로 추출하여 학습시켰다. 1epoch는 각 epoch마다 하나씩 학습시킨 것이고, multitask는 각 데이터셋을 1:1:1비율로 동일하게 학습시킨 것이다.

3.1 KorQuad 1.0

KorQuad 1.0은 한국어 MRC(Machine reading Comprehension)을 위해 만들어진 데이터셋으로, 1,560개의 Wikipedia article에서 10,645건의 문단과 66,181개 (Training set : 60,407, Dev set : 5,774)의 질의응답 쌍으로 구성되어 있다. 본 논문에서는 Continual learning 기반 학습을 위해 사용한 3개의 데이터셋 중 하나로 사용하였으며, 또한 evaluation 셋으로도 사용하였다.

표1. 학습 데이터 및 평가 데이터 통계

데이터	데이터양	사용 용도
KorQuad 1.0	질문과 답(6만)	Training, Evaluation
AIHub 기계독해	질문과 답(25만)	Training
AIHub 도서자료 기계독해	질문과 답(70만)	Training

3.2 AIHub 기계독해

AIHub 기계독해는 AIHub에서 제공하는 MRC 데이터셋으로, 45만개의 article에서 질문과 답을 포함한 25만개의 표준 데이터셋과 본문에서 답을 찾을 수 없는 질문을 포함한 10만개의 정답 없는 데이터셋, 그리고 질문과 답과 그 답을 선택한 단서를 포함한 10만개의 설명 가능 데이터셋으로 이루어져 있다.

본 논문에서는 질문과 답만을 포함한 25만개의 표준 데이터셋을 Continual learning 기반 학습을 위해 사용한 3개의 데이터셋 중 하나로 사용하였다.

3.3 AIHub 도서자료 기계독해

AIHub 도서자료 기계독해는 AIHub에서 제공하는 MRC 데이터셋으로, 100만개의 지문-질문-답변 데이터셋 중 정답이 있는 데이터셋 70만개와 정답이 없는 데이터셋 30만개로 이루어져 있다.

본 논문에서는 정답이 있는 데이터셋 70만개를 Continual learning 기반 학습을 위해 사용한 3개의 데이터셋 중 하나로 사용하였다.

3.4 BERT

BERT는 트랜스포머기반 사전언어훈련 모델이다. 현재에는 더욱 성능이 좋은 RoBERTa, DistilBERT, Electra 등의 BERT 개량형 모델이 있지만 본 논문에서는 Continual Learning의 MRC 적용 및 성능 향상이라는 목표를 확인하기 위해 가장 기본 BERT모델을 베이스라인 모델로써 사용하였다.

3.5 Continual Learning

기존의 딥러닝의 Training과정에서 여러 데이터셋 학습 후 테스트시 이전에 학습했던 데이터에서 생성된 weight들이 점차 사라짐과 동시에 성능이 떨어지는 catastrophic forgetting 현상이 나타난다. 이러한 문제를 Continual Learning을 통해 해결할 수 있다. Continual Learning은 새로운 데이터를 학습하면서도 동시에 기존에 학습된 데이터의 정보를 손실시키지 않고 가지고 있도록 하는 것이다.

본 논문에서는 이전 스테이지에서 학습된 모델의 예측값과 새로운 태스크에서의 라벨을 함께 학습하는 LWF 방법을 적용했다. 각각 다른 태스크로 학습된 이전 스테이지에서의 예측값과 새로운 스테이지에서의 예측값이 최소화하도록 학습하며, 이전 연구[7]에서 적용한 것과 같이 온도 값 T를 1보다 크게 설정하여 기존에 큰 값은 더 작게, 작았던 값은 더 크게 만든 로짓값을 이용하였다.

4. 실험 및 결과

본 논문에서는 Fine-tuning을 기반으로 여러 가지 학습 방법을 고안하여 사용한 4개의 모델(model_30k, model_50k, model_epoch, model_multitask)과 Fine-tuning 이후 Continual learning의 LWF 방법을 사용한 모델의 성능을 비교하였다.

model_30k는 Fine-tuning을 기반으로 하여 추가적인 방법으로 학습 후 생성된 모델 4개 중 하나이다. continual learning에서는 한 epoch을 스테이지로 정의하는데, 한 스테이지당 3개의 학습 데이터셋 중 랜덤하게 3만개를 뽑아 20 stage를 학습시켰다.

model_50k는 Fine-tuning을 기반으로 하여 추가적인 방법으로 학습 후 생성된 모델 4개 중 하나이다. 한 스

페이지당 3개의 학습 데이터셋중 랜덤하게 5만개를 뽑아 10 stage를 학습시켰다.

model_epoch는 Fine-tuning을 기반으로 하여 추가적인 방법으로 학습 후 생성된 모델 4개 중 하나이다. 본 모델은 한 epoch 즉 한 스테이지마다 각 데이터의 training셋 전체를 하나씩 학습시켜 생성한 모델이다. korquad 6만 -> AIHUB 기계독해 25만 -> AIHUB 도서자료 기계독해 70만의 과정으로 100 stage를 학습시켰다.

model_multitask는 Fine-tuning을 기반으로 하여 추가적인 방법으로 학습 후 생성된 모델 4개 중 하나이다. 본 모델은 stage단위로 학습시키지 않고 각 데이터셋들 중에서 데이터를 한 개씩 뽑아서 korquad -> AIHUB 기계독해 -> AIHUB 도서자료 기계독해 순으로 지속적으로 학습 시켰다.

model_LWF는 Fine-tuning 후 continual learning의 LWF 방법을 사용하여 생성된 모델이다. LWF는 Learning Without Forgetting의 약자이다. 총 세개의 데이터셋을 순서대로 학습시키며 현재 stage 데이터셋의 라벨과 이전 stage에서 학습된 모델로 현재 stage의 데이터를 예측한 확률값을 같이 라벨로써 사용해 학습시켰다. 본 논문에서는 model_50K에서 추가로 LWF 방법을 사용해 학습 시켰다.

표2. 각 모델별 성능비교

모델	F1	EM
BERT	91.68	79.92
30K	91.89	79.5
50K	92.18	80.25
50K-LWF	92.57	80.14
epoch	91.53	78.73
multitask	90.59	77.37

기본 베이스라인 BERT 모델에서 91.68(79.92)가 나왔고, Fine-tuning기반으로 추가적인 방법을 사용한 4개의 모델들 중에에선 50K가 F1 92.18, EM 80.25로 베이스라인 모델보다 F1 0.5, EM 0.38정도 가장 향상된 결과를 보였다.

50K에서 추가적으로 LWF방식을 사용하여 학습 및 생성된 모델이 F1 92.57, EM 80.14로 50k 모델보다 F1에서 0.39 향상되었고, EM은 0.11정도 감소했으나 기본 베이스라인인 BERT와 비교하였을 때 F1은 0.89, EM은 0.22정도 향상이 있었다.

5. 결론

본 연구에서는 기존의 딥러닝의 Training 과정에서 여러 데이터셋 학습 후 테스트시 이전에 학습했던 데이터에서 생성된 weight들이 점차 사라짐과 동시에 성능이 떨어지는 catastrophic forgetting 현상을 해결하기 위한 방법인 Continual Learning을 MRC 기계독해 task에 적용하여 성능을 측정하여 기본 베이스라인 모델과 Fine-tuning 단계에서 여러 가지 방법을 사용하여 학습시킨 모델들과 비교하여 보았다. 결과적으로

Fine-tuning 단계에서 랜덤으로 5만개를 추출하여 학습시킨 후 Continual Learning의 LWF 방식을 사용하여 추가학습을 시킨 model_50K_LWF 모델에서 성능이 향상된 것을 확인할 수 있었으며, 이를 통해 Continual Learning을 통해 이전 학습 데이터의 weight값이 감소하는 catastrophic forgetting 문제를 MRC task에서 또한 해결할 수 있음을 확인할 수가 있었다.

향후 연구로는 Continual Learning의 방법 중 하나인 EWC를 사용하여 모델을 학습시킨 후 LWF와 어느 것이 더 효율적인지 연구를 할 계획이다.

참고문헌

- [1] 임승영, 김명지, 이주열, KorQuAD: 기계독해를 위한 한국어 질의응답 데이터셋, 한국정보과학회 학술발표논문집, pp. 643-645, 2018
- [2] 김영민, 임승영, 이현정, KorQuAD 2.0: 웹문서 기계독해를 위한 한국어 질의응답 데이터셋, 한국정보과학회, pp. 577-586, 2020
- [3] Sungjoon Park, KLUE: Korean Language Understanding Evaluation, arXiv preprint arXiv:2105.09680, 2021
- [4] Kirkpatrick, J., Pascanu, R., et al., Overcoming catastrophic forgetting in neural networks, In: PNAS, 2017
- [5] 이동현, 박천음, 이창기, 박소윤, 임승영, 김명지, 이주열, BERT를 이용한 한국어 기계 독해. 한국정보과학회 학술발표논문집, pp. 557-559, 2019.
- [6] Hyo-Eun Kim, Seungwook Kim, Jaehwan Lee, Keep and Learn: Continual Learning by Constraining the Latent Space for Knowledge Preservation in Neural Networks, arXiv preprint arXiv:1805.10784, 2018
- [7] Geoffrey Hinton, Oriol Vinyals, and Jeffrey Dean. Distilling the knowledge in a neural network. In NIPS Deep Learning and Representation Learning Workshop, 2015.