

웹 서비스 사용자 경험(UX) 향상을 위한 BERT 기반 전이 학습 모델 적용 방안 연구

차주호

청운대학교 공과대학 멀티미디어학과

e-mail:jhcha@chungwoon.ac.kr

A Study on the Application of BERT-based Transfer Learning Model for Enhancing UX in Web Services

Jooho Cha

Dept. of Multimedia Science, Chungwoon University

요약

본 논문은 웹 서비스에서 사용자 경험(UX) 향상을 위한 베이스 모델로 BERT(Bidirectional Encoder Representations from Transformers)를 사용하는 전이 학습 모델을 제안한다. 다양한 자연어 처리(NLP) 문제에 최적화된 BERT 모델을 베이스 모델로 사용하여 사전 학습된 BERT 모델의 출력층을 변경함으로써 웹 서비스의 사용자 경험 데이터를 처리할 수 있는 구조로 변경할 수 있다. 전이 학습 모델을 위한 학습 데이터는 Kaggle에서 제공하는 데이터 셋을 사용하며, 전이 학습을 통해 사용자 선호도와 경험에 대한 향상에 기여하는 요소들을 분석하고 찾아낼 수 있다. 이를 통해 BERT 기반 전이 학습 모델이 변화하는 사용자의 요구에 더 지능적이고 더 사용자 중심적인 웹 서비스를 제공할 수 있다.

1. 서론

최근 웹 서비스에서 사용자 경험(UX)의 중요성은 지속적으로 높아지고 있다. 특히 전자상거래와 모바일 애플리케이션에서는 사용자에게 원활하고 매력적이며 개인화된 경험 제공이 웹 서비스의 성공 여부를 결정짓는다[1]. 따라서 본 논문은 Kaggle에서 제공하는 공개 데이터 셋인 Google Play Store Apps[2]와 Amazon Reviews Dataset[3]을 활용하여 웹 서비스에서 사용자 경험 향상을 제공하기 위한 BERT(Bidirectional Encoder Representations from Transformers)[4] 기반 전이 학습 모델이 성공적인 웹 서비스를 제공할 수 있는지 연구하고자 한다. 전이 학습 모델에서 학습에 사용할 두 데이터 셋은 각각 Google Play Store의 앱 정보 및 사용자 리뷰와 Amazon 제품에 대한 사용자 리뷰를 포함하고 있어서 사용자 선호도 및 경험에 대한 풍부한 정보를 제공한다.

2. BERT 베이스 모델

BERT는 google에서 발표한 자연어 처리(NLP)를 위한 사전 학습된 딥러닝 모델이다. 기존의 RNN이나 LSTM 등과 달리 Transformer 기반의 아키텍처를 사용하여 양방향으로 문맥을 처리한다. BERT 모델은 대용량 데이터 셋을 이용해 사전에 학

습되어 있기 때문에 일반적인 번역, 감정 분석, 개체명 인식 등 다양한 자연어 처리 작업에 사용할 수 있다. 따라서 본 연구에서는 사용자 경험 데이터에 특화된 전이 학습 모델을 구축하기 위하여 BERT를 베이스 모델로 선정하였다. 우리는 BERT 베이스 모델을 기반으로 웹 서비스에서 사용자 경험을 향상시킬 수 있는 전이 학습 기반의 새로운 딥러닝 모델을 생성할 수 있다.

3. 전이 학습 모델 구축

3.1 학습 데이터 셋

우리의 전이 학습 모델에서는 데이터 셋으로 Google Play Store Apps와 Amazon Reviews Dataset을 활용하여 학습을 수행한다. Google Play Store Apps는 Google Play Store에 있는 앱 정보와 사용자 리뷰 데이터를 포함하고 있어서 사용자들의 선호도와 리뷰를 분석하여 웹 서비스에서 사용자 경험 향상을 위한 데이터를 얻을 수 있다. 또한 Amazon Reviews Dataset은 아마존에서 제공하는 상품에 대한 사용자 리뷰 데이터가 포함되어 있어서 사용자들의 경험을 이해하고 개선할 수 있는 방향을 찾을 수 있다.

따라서 두 데이터 셋 모두 사용자 리뷰 데이터를 제공할 뿐만 아니라 웹 서비스 분야와 관련이 있어 BERT 기반 전이 학습 모델을 활용한 연구에 적합하다.

3.2 전이 학습 모델 구축

다음은 BERT 기반 전이 학습 모델을 구축하기 위한 절차이다.

1. BERT 베이스 모델: 미세 조정될 BERT 모델의 헤드를 제거하여 특징 추출 부분만 남긴다.
2. 도메인 특화된 미세조정: 사용자 경험과 관련된 데이터로 BERT 모델을 미세 조정한다. 이 과정에서 손실 함수를 사용하여 사용자 경험 데이터와 같은 도메인 특화된 정보가 모델에 통합되도록 한다.
3. 새로운 출력 층: BERT 베이스 모델의 헤드를 제거한 후, 사용자 경험 향상과 관련된 웹 서비스의 목적에 맞게 새로운 출력 층을 추가한다. 출력 층은 일반적으로 하나 이상의 밀집 층 활성화 함수 및 다른 구성요소로 이루어진다.
4. 추가 구성요소: 사용자 경험 데이터를 더 잘 반영하기 위해 새로운 출력 층에 여러 구성요소를 추가할 수 있다. 웹 서비스 사용 시간이나 페이지 이동 시간 등의 사용자 행동 데이터를 처리하는 Attention 메커니즘을 적용할 수 있다.
5. 멀티태스크 학습: 사용자 경험 데이터를 구성하는 속성 간 상관관계를 고려하여 멀티태스크 학습을 수행하는 구조를 갖도록 한다. 이를 통해 모델은 동시에 여러 가지 사용자 경험 요소를 학습하여 서로간의 관계를 잘 이해하고 예측할 수 있게 된다.

3.3 전이 학습 모델 구조

웹 서비스의 사용자 피드백 및 행동 데이터와 같은 도메인 특화 데이터를 활용하여 베이스 모델의 지식을 적용하고 확장하기 위해서는 사전 학습된 BERT 모델을 사용자 경험 데이터에 맞추어 미세조정을 해주어야 한다. 이를 위해 BERT 베이스 모델의 기존 출력 층(헤드)을 제거하고, 새로운 출력 층을 구성해야 한다. 새롭게 구성되는 전이 모델의 새로운 출력 층은 다음과 같은 구성 요소를 포함한다.

1. Classifier 층: 사용자 피드백 및 의견을 분류하기 위해 다양한 클래스를 고려한 분류기 층으로, 사용자 경험 데이터를 기반으로 텍스트 분류 및 해당 웹 서비스에서 사용자 만족도와 관련된 항목(예: 제품별 카테고리)을 분류하는데 도움이 된다.
2. Regression 층: 사용자 경험의 연속적인 가치(예: 별점, 페이지 유지 시간 등)를 예측하기 위한 회귀 계층으로, 이를 통해 웹 서비스 제공자는 사용자 만족도를 정량적으로 측정하고 개선할 수 있다.
3. 추가 Attention 메커니즘: 다양한 사용자 경험에 따른 특성들을 반영하기 위해 기존 BERT 모델의 Attention 메

커니즘에 추가로 적용할 수 있는 Attention 메커니즘을 고려한다. 이를 통해 모델은 작성자별, 제품별, 웹 서비스 특성별 등 다양한 상황에서의 텍스트로부터 중요한 정보를 검출할 수 있다.

4. 테스트 및 성능지표

BERT 기반 전이 모델의 입력 데이터는 Google Play Store Apps와 Amazon Reviews Dataset의 앱 및 제품에 대한 사용자 리뷰, 별점 및 의견 등의 정보를 포함한다. 이 데이터를 전처리하여 텍스트 데이터는 토큰화, 정제, 정규화되며 이후 BERT 모델의 입력 형태에 맞게 인코딩된다. 또한 훈련 데이터(80%)와 시험 데이터(20%)로 나누어 전이 학습 모델의 학습 및 평가를 진행한다. 본 연구의 전이 학습 모델의 성능 평가를 위해 정확도, 정밀도, 재현율을 사용한다. BERT 기반 전이 학습 모델을 활용한 사용자 경험 데이터 분석은 앱 사용 시간이 짧은 경우 사용자들이 정보를 찾기 어려운 것이 원인인지 또는 사용자들이 제품에 긍정적인 평가를 한 경우 그 원인이 풍부한 콘텐츠 때문인지 등을 확인하고 개선을 추진할 수 있다.

5. 결론

본 연구에서는 사용자 경험 향상을 목표로 BERT 기반 전이 학습 모델을 구축하고 웹 서비스 사용자 경험 데이터에 적용하여 그 성능을 평가하는 것을 목표로 하였다. 그러나 본 논문은 아직 논문에서 설계한 전이 모델에 대한 실질적인 테스트를 수행하지는 못하였다. 따라서 본 연구에서 사용할 입력 데이터에 대해 본 연구의 전이 학습 모델을 적용해 보고, 모델의 성능 및 확장성을 분석하는 것이 중요하다.

참고문헌

- [1] 차시호, “딥러닝을 이용한 프론트엔드 웹의 사용자 경험(UX) 개선”, 2023년 산학기술학회 춘계학술대회, 2023.5.
- [2] Kaggle, Google Play Store Apps, <https://www.kaggle.com/datasets/gauthamp10/google-playstore-apps>.
- [3] Kaggle, Amazon Reviews Dataset, <https://www.kaggle.com/datasets/anushabellam/amazon-reviews-dataset>.
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, <https://arxiv.org/pdf/1810.04805.pdf>.