



감정 인식 평점 재조정을 통한 추천 시스템 성능 향상

장지원¹, 최재준², 추상현^{1,*}

¹ 국립금오공과대학교 산업·빅데이터공학부, ² 경희대학교 산업경영공학과

Introduction

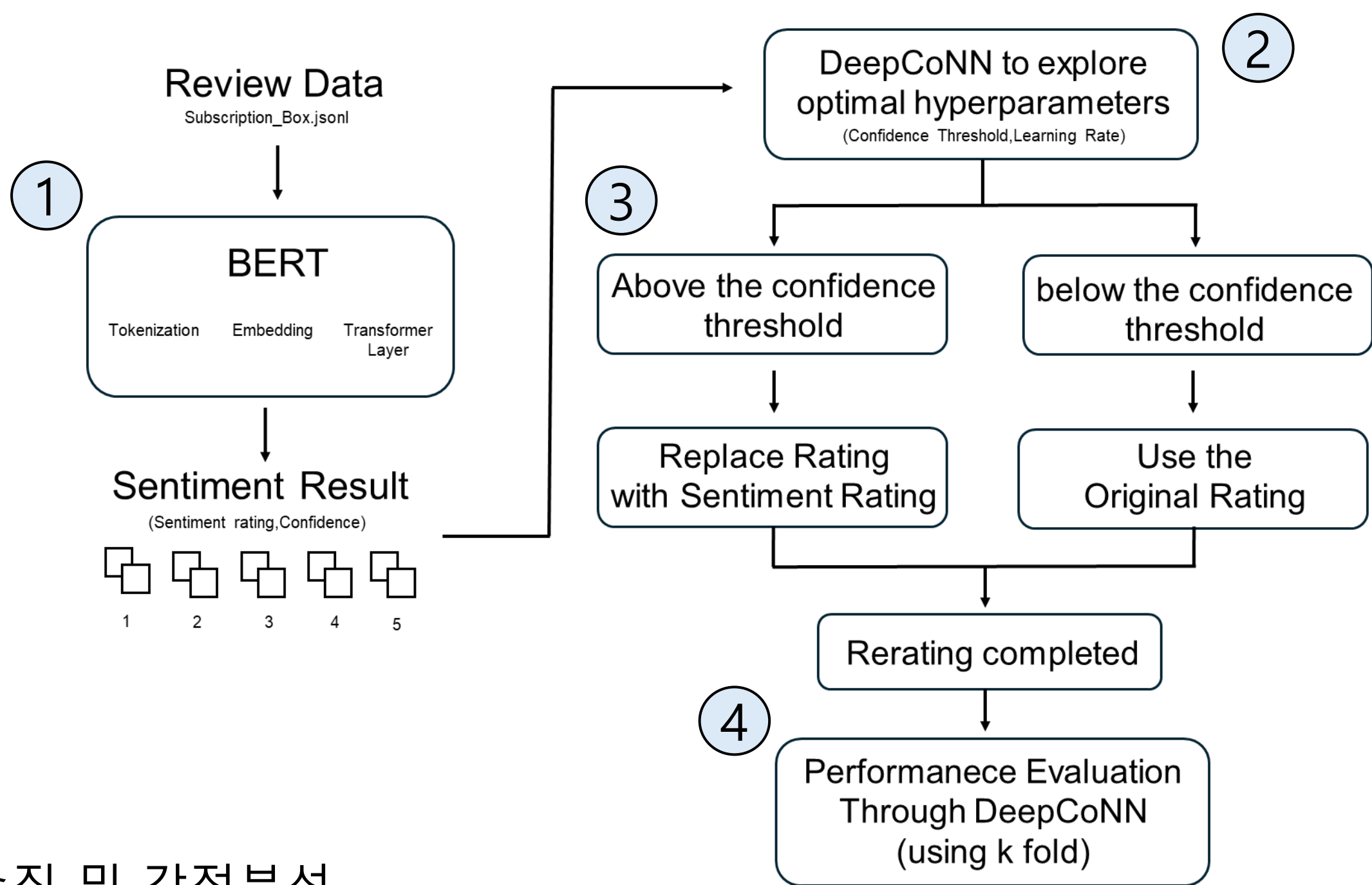
- 추천 시스템은 디지털 환경에서 사용자에게 맞춤형 콘텐츠를 제공하는 핵심 메커니즘임.
- 특히 추천 시스템은 사용자에게 최적화된 제품 및 서비스를 제안하는데 결정적인 역할을 하며, 리뷰 중 사용자 평점은 추천 시스템의 주요평가 지표임.
- 사용자 평점이 실제 텍스트에 담긴 감정적 맥락을 제대로 반영하지 못하는 경우가 종종 발생함.
- 예를 들어, 매우 긍정적인 내용의 리뷰임에도 낮은 평점을 주거나, 반대로 부정적인 리뷰임에도 높은 평점을 주는 경우가 발생함.
- 이러한 리뷰-평점 간 불일치 현상은 추천 시스템의 정확성을 저하시키고 추천 결과에 대한 신뢰성에도 부정적 영향을 끼침.
- 기존의 연구 [1] [2]에서는 감정 분석 결과(긍정/부정)를 평점 수식에 단순 반영하는 방식을 사용하였으나, 이러한 방법론은 감정 분석 결과의 신뢰도를 고려하지 않아 분석의 불확실성이 그대로 추천 결과에 반영될 우려가 존재함.
- 반면 본 연구에서는 감정 분석 결과의 신뢰도를 고려하여 임계값을 넘을 때만 평점을 대체함으로써, 추천 정확도를 향상하는 차별적 접근을 시도함. 이때 감정 분석 모델로는 BERT-multilingual-uncased[3]를, 최적의 신뢰도 임계값을 탐색하기 위해 DeepCoNN 모델[4]을 활용하였음.

Dataset

User_ID	Item_ID	Review Text	Rating
AEMJ2EG5OD ...	B07G584SHG	Absolutely useless nonsense and a ...	1
AEEJBFZKUB ...	B07QL1JRCN	With a couple of the items, I wasn't ...	2
...
AGSVZNZBTS ...	B07RBYJN37	Although I don't remember ...	5
AFDERNB6BI ...	B07KM6T8GV	I loved everything and could ...	5

- UCDS의 McAuley Lab에서 수집한 2023년 Amazon 리뷰(Subscription_Boxes 카테고리, 총 14,082개)[5]를 사용함.
- 각 리뷰는 User_ID, Item_ID, Review text, Rating으로 구성됨.
- 불용어 및 불필요한 구두점을 제거하는 전처리 과정을 거침.

Methods



① 데이터 수집 및 감정분석

- 대규모 리뷰 데이터를 수집한 후 BERT-base-multilingual-uncased 감정 분석 모델을 활용해 각 리뷰 텍스트에 대한 감정 평점(1~5점)과 신뢰도(0~1; 1에 가까울 수록 신뢰도가 높다)를 산출.

② 모델 학습 및 하이퍼파라미터 실험

- 텍스트 분석에 특화된 병렬 구조의 TextCNN기반 DeepCoNN 모델을 적용하여 신뢰도 임계값 (Confidence Threshold)과 학습률(Learning Rate)에 따른 하이퍼 파라미터 조합을 실험.
- 각 조합별 모델을 학습한 후 평균제곱오차(Mean Squared Error, MSE) 지표를 통해 성능 평가 후, 수치를 종합적으로 비교하여 최적의 하이퍼 파라미터를 결정.

③ 평점 재조정

- 최종적으로 결정된 신뢰도 임계값을 기준으로, 기존 평점을 감정 분석 평점으로 대체할지 여부를 판단.
- 신뢰도가 임계값 이상이며 동시에 감정 분석 평점이 기존 평점과 상반될시 평점 재조정.

④ 성능 평가

- BERT-base-multilingual-uncased 모델을 통해 각 텍스트에 대한 감정 평점과 신뢰도를 계산한 후, 최적의 하이퍼 파라미터 조합 탐색을 위해 K-fold 교차 검증을 활용해 DeepCoNN 모델을 반복적으로 학습시켜 다양한 하이퍼 파라미터 조합에 대한 성능 실험을 진행.

Results

하이퍼파라미터 조합에 따른 평점 재조정 DeepCoNN모델의 평균 MSE 성능

Confidence (Threshold)	Learning rate	Mean (Std)
0.5	0.001	2.69 (0.81)
	0.01	2.72 (0.40)
	0.1	9587.57 (20495.70)
0.6	0.001	2.63 (0.09)
	0.01	2.47 (0.12)
	0.1	285093.22 (637432.75)
0.7	0.001	2.62 (0.06)
	0.01	3.05 (0.64)
	0.1	185893.41 (398262.32)

- 신뢰도 (0.5,0.6,0.7)와 학습률(0.1,0.01,0.001)을 결합한 9가지 케이스를 학습 및 평가한 결과, 신뢰도 임계값 0.6, 학습률 0.01에서 MSE 2.47(Std: 0.12)로 가장 우수한 성능이 도출됨.

Baseline과 평점 재조정 모델의 성능 비교

Models	Baseline model	Re-rating model
Mean (Std)	2.56 (1.42)	2.47 (0.12)

- 본 연구에서 제시한 모델의 유효성을 확인하기 위해, Baseline 모델(DeepCoNN 원형)과 평점 재조정 모델을 동일 데이터셋에서 비교함.
- Baseline 모델은 기존 DeepCoNN의 하이퍼파라미터인 학습률 0.002를 그대로 적용하였으며, 재조정 모델은 앞서 구한 최적의 하이퍼파라미터인 신뢰도 0.6과 학습률 0.01을 적용함.
- 재조정 모델이 Baseline 대비 평균 MSE가 더 낮게 나타나, 추천 성능 향상에 기여함을 확인함.

Conclusions & Future work

- 본 연구에는 감정분석을 활용하여 사용자 평점을 재조정하는 방법을 제안하고, 이를 통해 추천 시스템 성능 향상 가능성을 실증적으로 확인함.
- 제안된 접근법은 리뷰 텍스트에 담긴 실제 감정 정보를 정량화하고, 해당 감정 분석의 신뢰도가 충분히 높을 때만 기존 평점을 대체함으로써 불일치를 효과적으로 완화함. 그 결과 추천 정확도가 개선되는 모습을 확인함.
- 하지만 추천 알고리즘에 대한 적용 가능성은 아직 입증되지 않았으므로, 본 방법론을 다양한 추천 알고리즘에 적용하거나 언어 자원이 더욱 다양한 환경에서의 적용 가능성을 검증함으로써 본 접근법의 일반화를 더욱 확장해 나갈 예정임.

References

- [1] 김상진, 김한주, "감성분석을 활용한 딥러닝 기반 개인화 추천 시스템," 한국정보기술학회 논문지, 제17권, 제5호, pp. 99–106, 2019.
- [2] J. T. Jo & S. H. Choi. Sentiment Analysis of movie review for predicting movie rating. Management & Information Systems Review, 34(3), 161-177. 2015
- [3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint arXiv:1810.04805, 2019.
- [4] L. Zheng, V. Noroozi, and P. S. Yu, "Joint Deep Modeling of Users and Items Using Reviews for Recommendation," Proc. 10th ACM Int. Conf. Web Search Data Mining (WSDM), pp. 1–10, 2017.
- [5] H. Luo, Y. Yu, Y. Zhu, and Q. Zhang, "Large Language Models are Zero-Shot Rankers for Recommendation," arXiv preprint, arXiv:2403.03952, 2024.