



추천 시스템 향상을 위한 감정-평점 분석 기반 최적 데이터 선별

김상민¹, 최재준², 추상현^{1,*}

¹ 국립금오공과대학교 산업·빅데이터공학부, ² 경희대학교 산업경영공학과

Introduction

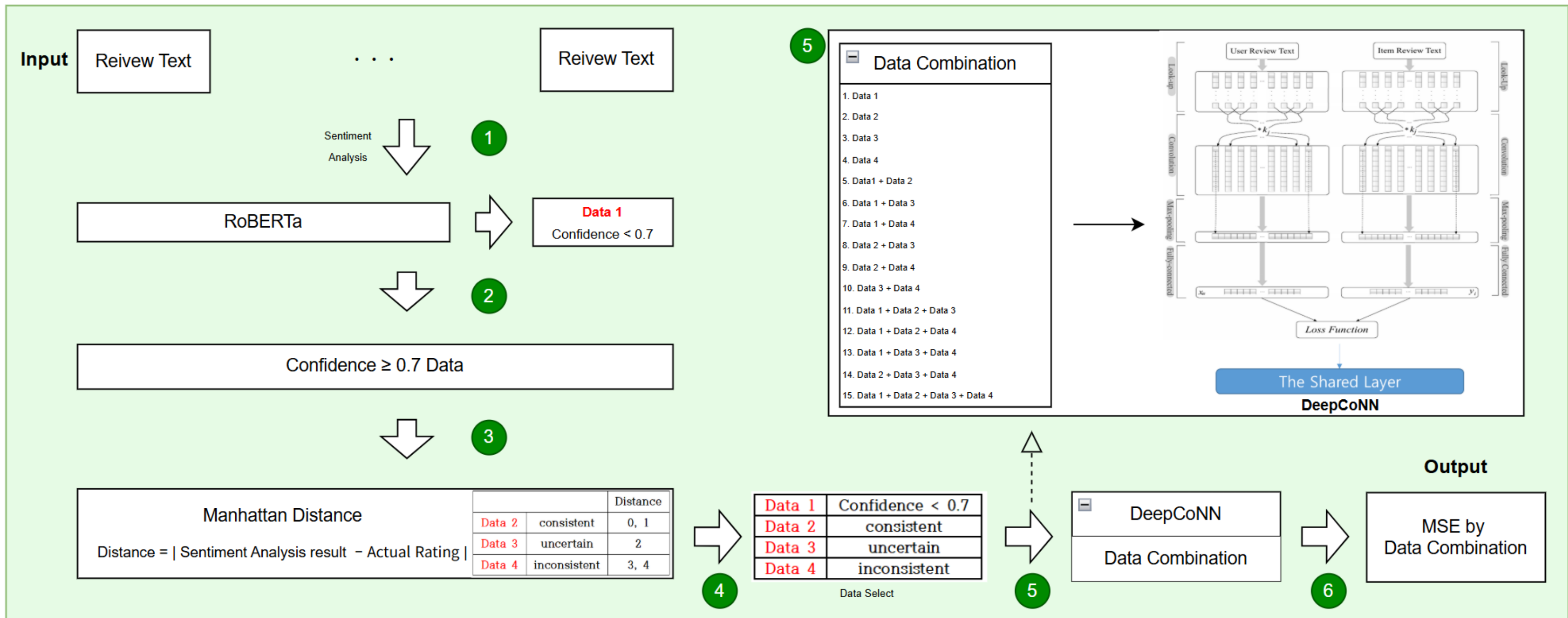
- 개인화 추천 시스템은 리뷰 텍스트와 평점을 기반으로 정밀한 사용자 경험을 제공하며, 이커머스 및 콘텐츠 플랫폼에서 핵심 기술로 자리잡고 있음.
- 하지만 리뷰 텍스트에 담긴 감정과 사용자 평점 간의 불일치 현상은 학습 데이터에 노이즈로 작용하여 모델 성능 저하의 원인이 될 수 있음.
- 기존 감성 기반 연구들[1]은 주로 감정 분석 결과를 보조 정보로 활용하거나 전체 데이터를 동일하게 처리함으로써, 감정-평점 간 정합성 문제를 충분히 반영하지 못했다는 한계가 있음.
- 본 연구는 감정 분석 신뢰도와 맨해튼 거리를 활용해 리뷰-평점 쌍을 non-confidence, consistent, uncertain, inconsistent로 분류하는 감성 기반 데이터 선택 방법을 제안함.
- 데이터 정제를 통해 추천 모델 학습 성능 향상을 도모하며, 이중 K-Fold 교차 검증으로 일반화 가능성을 검토함.

Dataset

| | user_id | asin | text | rating |
|---|------------------------------|------------|---|--------|
| 0 | AEMJ2EG5ODOCYUTI54NBXZHDJGSQ | B07G584SHG | Absolutely useless nonsense and a complete was... | 1 |

- Amazon의 Subscription Boxes 카테고리에서 14,082개의 리뷰 데이터를 수집하여 user_id, asin, text, rating을 주요 분석 변수로 사용함.
- 이 중 text는 감정 분석의 입력으로 사용되며, rating은 평점 예측을 위한 정답 레이블이자 감정-평점 간 불일치 판단 기준으로 활용됨.
- 본 연구는 리뷰 텍스트를 바탕으로 평점을 정밀하게 예측하며, 데이터 전처리 및 분류 과정은 모델의 학습 성능을 높이기 위한 기반 역할을 수행함.

Methods

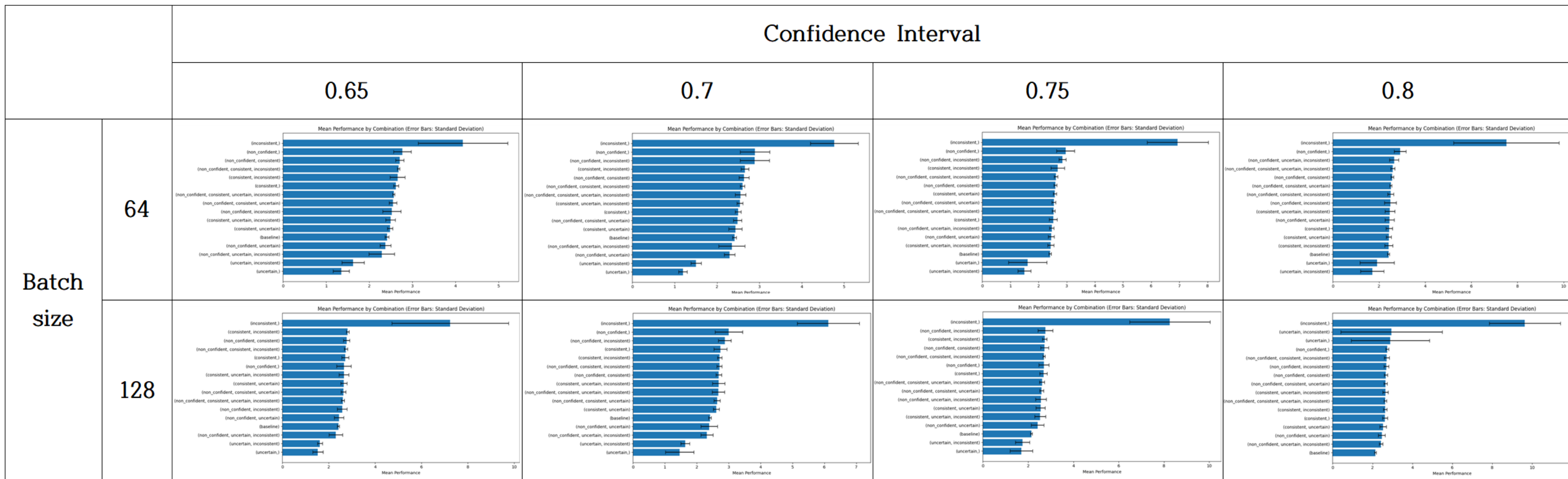


- Review Text에 RoBERTa 모델[2]을 적용해 감정 분석 시행
- Confidence 값이 0.7 미만인 데이터는 non-confidence로 분류, 0.7 이상인 경우 다음 단계 진행
- 감정 결과와 평점 간 맨해튼 거리를 기준으로 데이터를 세 그룹(consistent, uncertain, inconsistent)으로 분류
- 총 4개 그룹(non-confidence 포함) 구성
- 이 4개 그룹을 다양한 방식(총 15가지)으로 조합하여 DeepCoNN[3]모델에 적용하고 성능 비교 실험 수행
- 각 조합별 평균제곱오차(MSE)를 비교하여 최적의 데이터 구성 도출 및 추천 성능 평가

- DeepCoNN[3]은 사용자와 아이템 리뷰를 각각 처리한 후 결합해 평점을 예측하는 모델임. 감정-평점 정합성이 추천 정확도에 미치는 정확도에 미치는 영향을 정량적으로 분석함.
- 하이퍼파라미터 설정 : 학습률 0.001을 기준으로, 배치 크기(64, 128) 및 감정 분석 신뢰도 기준(0.65, 0.70, 0.75, 0.80)을 조합하여 실험을 수행함.
- 모델 평가 : 5-Fold 교차 검증과 Early Stopping(patience=5)을 적용하여 모델의 일반화 성능을 안정적으로 평가함.

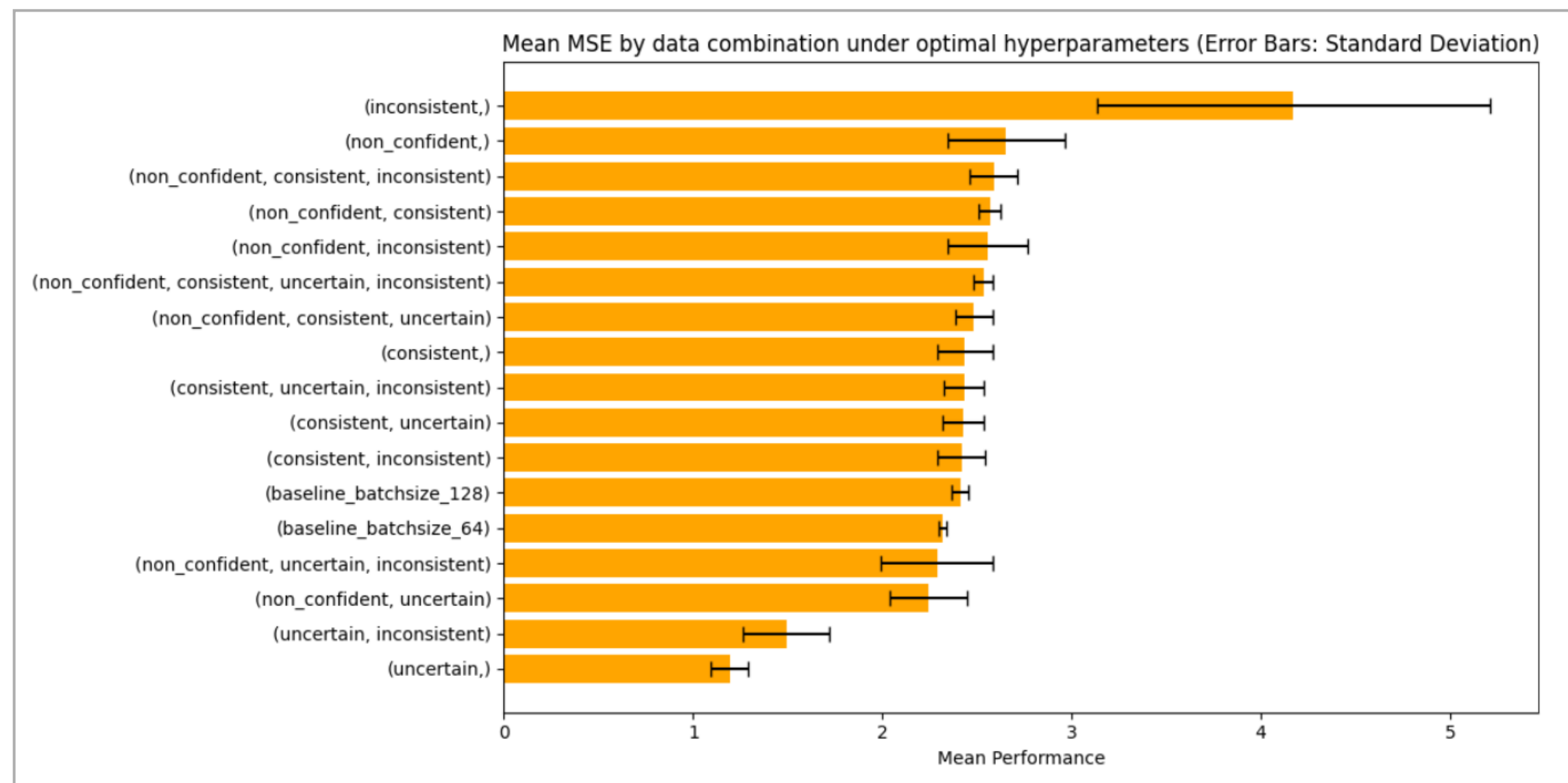
Results

Batch Size 및 Confidence에 따른 조합별 평균 MSE 비교



- 가장 우수한 조합은 학습률 0.001, 배치 크기 64, 신뢰도 0.7 조건에서 uncertain 데이터만 사용한 경우로, 평균 MSE 1.192 (Std: 0.100)를 기록함.
- 이는 감정과 평점 간 불일치가 일정 수준 존재하는 데이터(uncertain 그룹)가 오히려 정보 다양성을 제공하며, 추천 모델의 일반화 성능 향상에 기여함을 시사함.

최적 하이퍼파라미터에 따른 데이터 조합별 평균 MSE



- 감정 기반 데이터 조합별 MSE를 높은 순으로 정리한 위의 그래프는 데이터 선택 전략이 추천 성능에 미치는 영향을 시각적으로 보여줌.
- 특히 uncertain 데이터 단독 학습 시 이중 교차 검증 결과 평균 MSE가 1.165로 낮아져, 해당 그룹이 일반화 성능 향상에 효과적임을 확인함.

Conclusion & Future work

- 본 연구는 감정 분석과 최신 딥러닝 기반 추천 모델을 결합하여, 리뷰-평점 간 정합성을 정량적으로 평가하고 이를 바탕으로 학습 데이터의 신뢰도를 높이는 방법을 제안함.
- 실험 결과, 감정과 평점 간 불일치가 있는 uncertain 데이터만을 선별해 학습에 활용했을 때 가장 낮은 MSE를 기록하며, 모델의 예측 정확도와 일반화 성능 모두 향상되는 것을 확인함.
- 이중 교차 검증을 통해 해당 결과의 재현성과 안정성이 입증되었으며, 다양한 데이터 조합 및 하이퍼파라미터 실험에서도 일관된 성능 우위를 보임.
- 본 연구는 데이터 기반 접근법을 통해 추천 시스템의 정밀도와 신뢰도를 동시에 확보하며, 향후 실제 플랫폼 적용 및 도메인 확장을 통한 상용화 가능성을 제시함.

References

- [1] S. Kim, S. Kim, and J. Kim, "Sentiment Analysis Based Recommendation System Using Large Language Models for Review Data," Journal of Korean Institute of Intelligent Systems, vol. 34, no. 5, pp. 407–413, 2024. doi:10.5391/JKIS.2024.34.5.407.
- [2] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.
- [3] 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.
- [4] H. Luo, Y. Yu, Y. Zhu, and Q. Zhang, "Large Language Models are Zero-Shot Rankers for Recommendation," arXiv preprint, arXiv:2403.03952, 2024.