Boosting collaborative filtering with sentiment analysis

Xinyu He
University of Illinois, Urbana-Champaign
Illinois, USA
xhe34@illinois.edu

ABSTRACT

Sentiment analysis has been used to do rating prediction and achieved remarkable success. However, it requires user's review as input, which makes it impossible to predict ratings for items that user has never interacted with. Therefore, only using sentiment analysis for recommendation is impractical.

Given the success of sentiment analysis, researchers start to investigate how to integrate sentiment analysis with collaborative filtering. However, previous work ([1]), [3]) only focus on how to use the benefits of sentiment analysis to boost collaborative filtering, but ignore the feedback that collaborative filtering can contributes to sentiment analysis. Sentiment analysis can provide continuous approximation of ratings instead the discrete values; while collaborative filtering can provide extra user and item information for sentiment analysis. Therefore, it's important to maximize the mutual interest of sentiment analysis and collaborative filtering.

1 PROBLEM DEFINITION

Input: User-item interaction history with reviews: $\{(u, i, s, r)\}$, where the four entities in one history record are user, item, rating, review, respectively.

Output: Let $\mathcal{U}$ and $\mathcal{I}$ be the collection of users and item, for any user-item pair $(u, i) \in \mathcal{U} \times \mathbf{I}$, output probability of u interact with i in the future p_{ui}

2 PLANNED APPROACH

Maximizing Mutual Information is a popular technique in recent years that tries to maximize the agreement between two models. Although directly maximizing mutual information is intractable, InfoNCE Loss[4] is often adopted to maximize the lower bound of mutual information. Therefore, we can leverage both benefits of semantic analysis and collaborative filtering by maximizing the mutual information (adding InfoNCE Loss term) between two backbone model: semantic analysis backbone model $\mathcal{M}_s$ that output review embedding $\{e_r\}$; collaorative filtering backbone model $\mathcal{M}_c$ that output user embedding $\{e_u\}$ and item embedding $\{e_i\}$.

Specifically, for a record (u, i, s, r), we want to maximize the mutual information between $\{e_r\}$ and $f(Agg(e_u, e_i))$, where $Agg()$ is a aggregation function which could be concatenation; $f()$ is a projection function that project the aggegated embedding into the embedding space of e_r, and it can be implemented by simple neural network (e.g., fully-connected layer+activation function). Finally, the backbone model can be implemented by Transformer [5] and LightGCN [2]. Note that these backbone models are just selected for this project, it can be replaced by any other models that give desired embeddings.

3 DATASET

I'm planning to use the Amazon Digital Music Dataset[1].

4 EXPECTATION AND EVALUATION

I aim to compare the proposed method with two baseline methods: Transformer [5] and LightGCN [2] (which is also the selected backbone models). The proposed method is expected to outperform baselines by several evaluation metrics: NDCG@K, Recall@K and Precision@K (K=10, 20, etc.)

5 PROGRAMMING LANGUAGE

Python.

6 WORKLOAD

- Dataset Preprocessing: 3 hrs.
- Transformer Implementation: 6 hrs.
- LightGCN Implementation: 6 hrs.
- Proposed Method Implementation: 6 hrs.
- Evaluation: 3 hrs.
- Paper Work (Reports and Documentation): 4 hrs.

REFERENCES

[1] Miguel Á García-Cumbreras, Arturo Montejo-Ráez, and Manuel C Díaz-Galiano. 2013. Pessimists and optimists: Improving collaborative filtering through sentiment analysis. *Expert Systems with Applications* 40, 17 (2013), 6758–6765.

[2] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval.* 639–648.

[3] Cane WK Leung, Stephen CF Chan, and Fu-lai Chung. 2006. Integrating collaborative filtering and sentiment analysis: A rating inference approach. In *Proceedings of the ECAI 2006 workshop on recommender systems.* 62–66.

[4] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* (2018).

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[1] https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/