Language-model based recommender algorithm

GROUP
fyp24108

Student: Xiong Hongruifeng

Supervisor: Prof Huang, Chao

Introduction

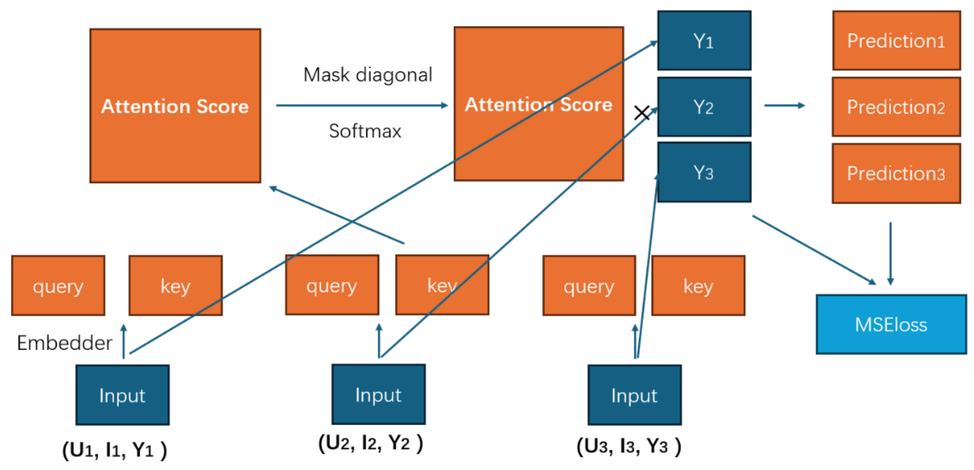
- **Problem**:
- Traditional ID-based recommendation algorithms heavily depend on collaborative data, which makes them vulnerable to the cold start problem. As a result, they struggle to offer effective recommendations for new users or items that lack sufficient interaction history.
- **Solution**:
- Combine language models with recommendation algorithms by utilizing text descriptions to generate embeddings through the language model. This approach helps to better capture the semantic information of items and users, enhancing the overall recommendation performance.
- **Previous work:**
- The paper titled "EasyRec: Simple yet Effective Language Models for Recommendation" used language model to embed descriptive language to users/items profiles, then do adversarial learning based on cosine-similarity.

$$\mathcal{L}_{\text{con}} = -\frac{1}{|\mathcal{B}|} \sum_{(u,i_{\text{pos}},i_{\text{neg}})} \log \frac{\exp(\cos(\mathbf{e}_u, \mathbf{e}_{i_{\text{pos}}})/\tau)}{\sum_{m \in \mathcal{N}_{\text{neg}}} \exp(\cos(\mathbf{e}_u, \mathbf{e}_m)/\tau)},$$

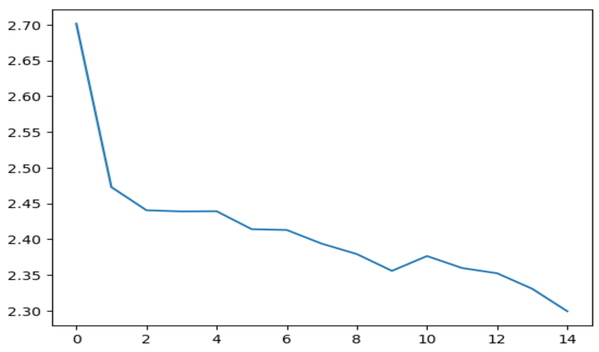
- **Key Innovation**:
- Propose an attention model to replace cosine-similarity based model
- Integrates language models for profile generation.

Methodology

- **Collaborative data part:**
- Firstly, use the following attention layer to make recommendation based on the collaborative data
- **Attention Layer Architecture:**
- Input: (User ID, Item ID, Interaction/Rating).
- Embedding: Concatenated user/item vectors (32-dim).
- Attention: Masked self-attention with learned Q/K matrices.
- Output: Interaction probability (MSE loss) or rating class (softmax).

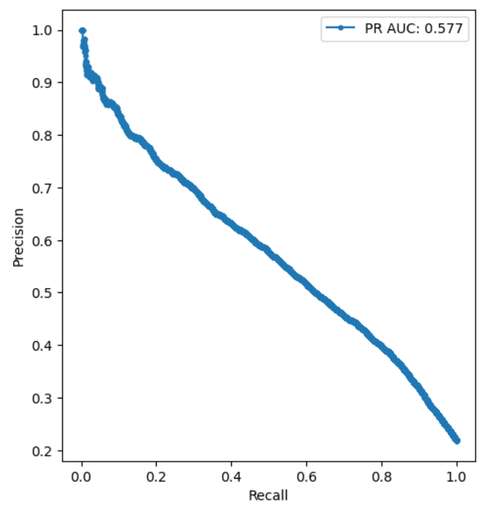


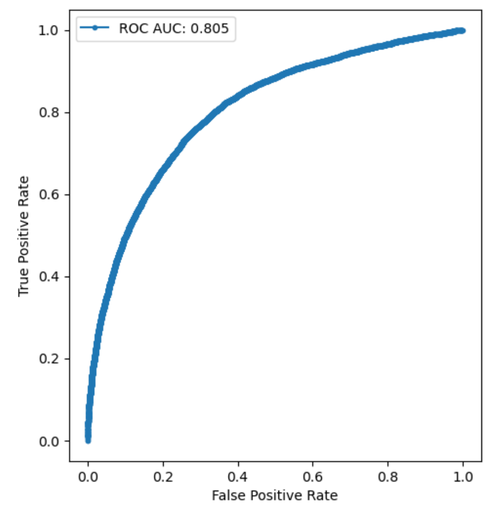
- **Language model part:**
- Secondly, use a sentence embedding model to embed items' text information to tensors. Specifically in this project, the movies name, genres and tags are used for embedding.
- Then, use the embedding from the language model to replace the weight of embedder of the previous model.
- Freeze the embedder and train the model.
- **Training loss:**



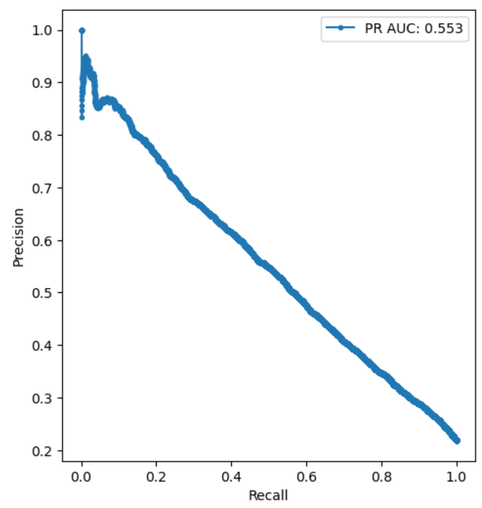
Evaluation

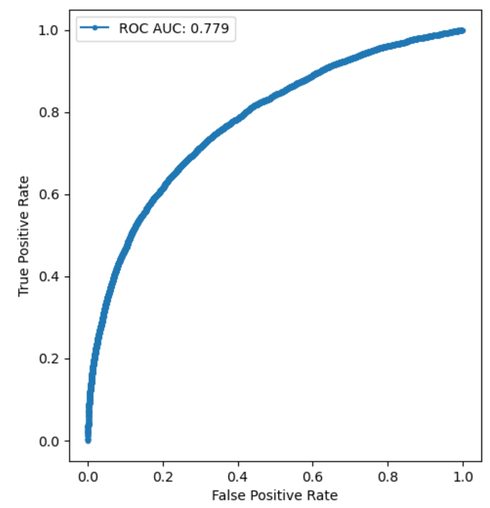
- **Matrix:**
- The model is evaluated on MovieLen dataset
- Positive samples rate: 21.97%
- PR-AUC and ROC-AUC is plotted.
- **Attention Layer Model only:**





- **Integration of Attention model and Language model:**





Summary

- In this project, a method that combines language models with recommendation algorithms is proposed, and its accuracy is validated on the MovieLens dataset. Additionally, a novel architecture is introduced, which performs predictions based on user-item pairs, replacing the traditional approach based on cosine similarity.

Reference

- [1] Xubin Ren, Chao Huang. "EasyRec: Simple yet Effective Language Models for Recommendation." arXiv preprint arXiv:2408.08821 (2024). https://doi.org/10.48550/arXiv.2408.08821.
- [2] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4: 19:1–19:19. https://doi.org/10.1145/2827872

Acknowledgement

- Thanks to Prof. Huang Chao for his foundational guidance