Solving Cold Start Problem For Recommendation System Using Hybrid Model

Neha Chopra

Key Features

- **Problem Statement**
- **Importance of Solving Cold Start Problem**
- **Cold Start Problem**
- **Addressing the Cold Start Problem using Two Phase Recommender Strategy**
- **Baseline and Augmented Dataset Features**
- **Feature Engineering on Baseline Dataset to generate Augmented features**
- **Results and Conclusion**
- **Future Recommendation**
- **Reference**

Problem Statement

Problem: common cold start problem encountered by recommendation engines in new domains, where historical data is lacking

Solution: Addressing the Cold Start Problem using Two Phase Recommender Strategy backed by LLM

Importance of Solving Cold Start Problem

- Critical importance of providing precise and relevant recommendations in real-time to enhance user satisfaction and improve platform performance, especially in the highly competitive e-commerce industry.
- Increasing need to accurately predict the popularity of new products and suggest products to new users. These predictions have significant implications for various stakeholders, including sellers, marketers, and platform operators, within e-commerce platforms. Sellers can optimize inventory management, devise effective marketing strategies, and make informed pricing decisions based on popularity predictions. Platform operators can use these predictions to enhance the overall user experience and allocate resources more efficiently.

Cold Start Problem

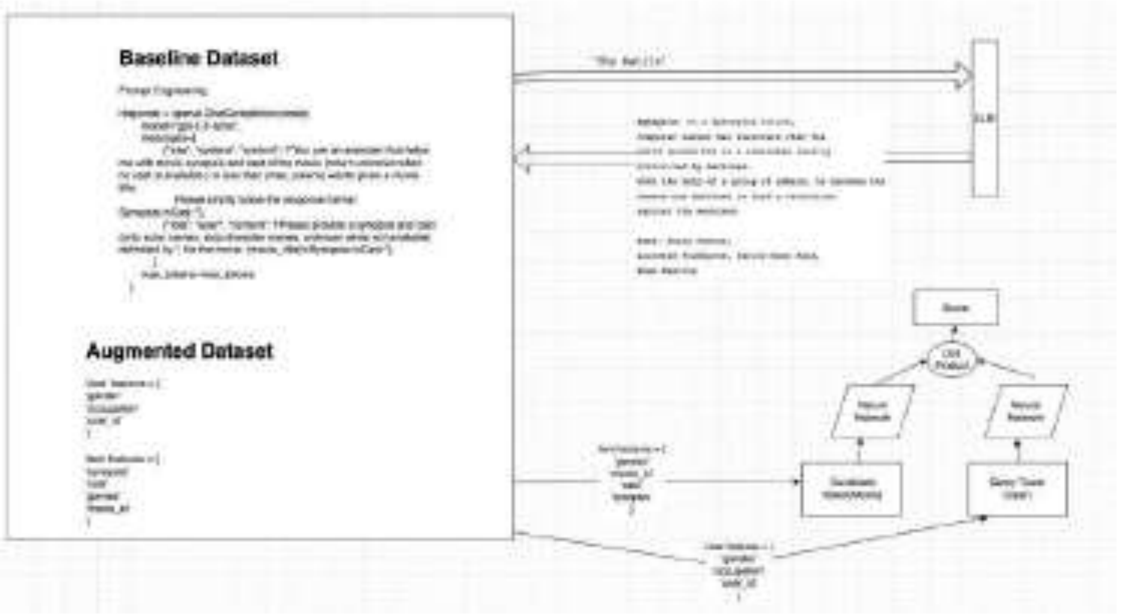
- The Cold Start Problem concerns the issue where RS cannot draw the inference for users or items for which it has not yet gathered sufficient information

- New items
    - Ex : a newly created movie
- New User
    - Ex: a new user who just signed into the system

Addressing the Cold Start Problem using Two Phase Recommender Strategy

- **Candidate/movie Augmentation :**
  Large Language Models (LLMs)
  to enhance the movie dataset by generating
  informative synopses and
  engaging cast descriptions for each movie.

- **Training Two Towel model on augment dataset :**
  The augmented dataset
  is instrumental in training the Two-Tower model,
  with a focus on the candidate tower.
  This tower undergoes training on various tasks
  using zero-shot learning,
  enabling the model to handle
  unseen tasks and predict outcomes for movies not
  present in the initial training data.



Baseline and Augmented Dataset Features

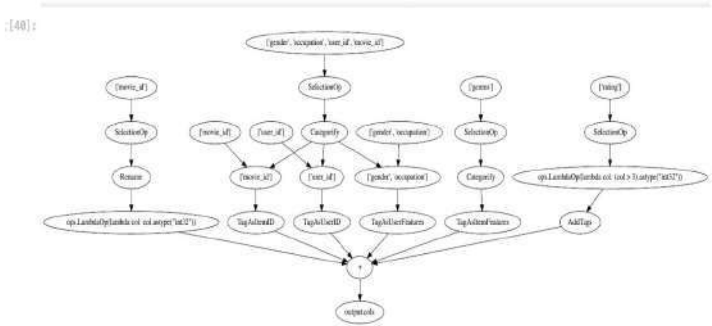


Figure 4.4.3 baseline dataset categorification

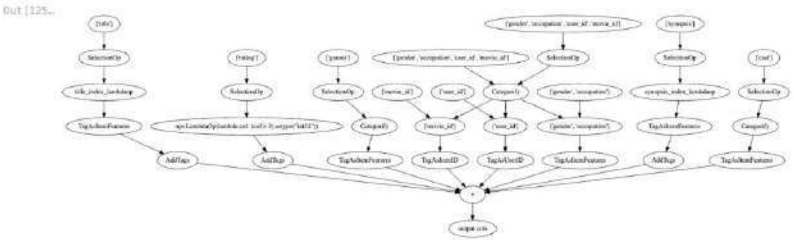


Figure 4.4.5 augmented dataset categorification

Comparison between Baseline and Augmented NDR models

Although both models exhibit comparable results at a top-k recommendation level of 100, the augmented model significantly outperforms the baseline model at a top-k recommendation level of 10. This suggests that the augmented model is particularly effective at providing top-10 recommendations, indicating its superiority in generating more relevant and contextually driven suggestions, especially for users or items with limited historical data (cold start scenarios).

Setting	Method	MovieLens 100K			
		HR@10	NDCG@10	HR@100	NDCG@100
	Baseline	0.0067	0.0024	0.1845	0.0342
NDR (Neural Deep Retrieval)	LLM Augmented	0.0158	0.0068	0.2039	0.0412

Table 5.2.1 Comparison between Baseline and NDR models

Result and Conclusion

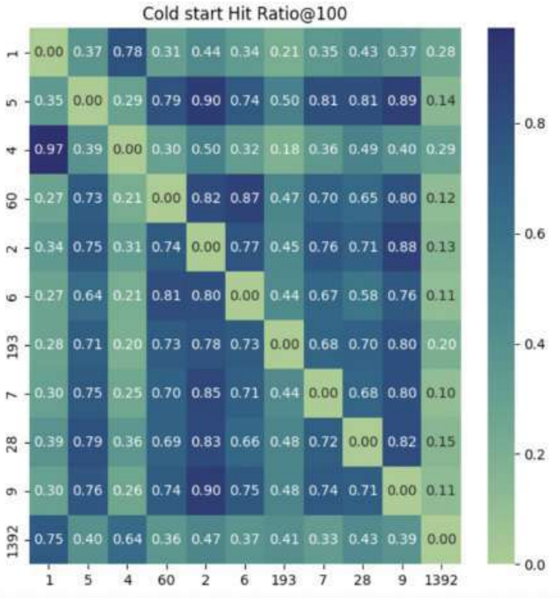


Fig 5.3.3 Hit Ratio@100 for baseline model for cold start

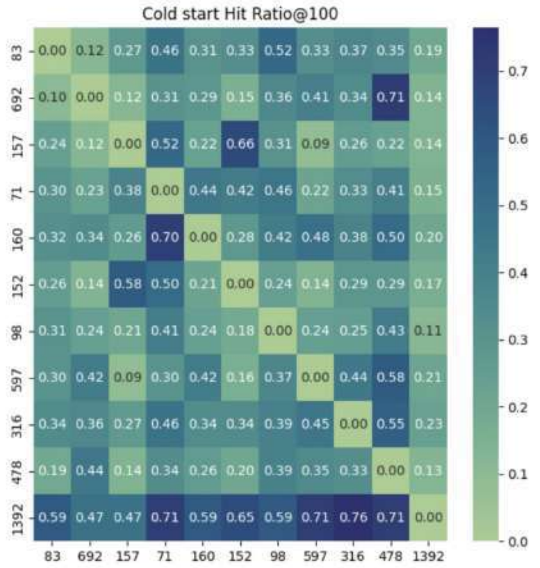


Fig 5.3.4.2 Hit Ratio@100 for augmented model for cold start

The baseline model tends to generate less accurate recommendations and exhibits a bias towards popular items, often failing to provide contextually relevant suggestions. In contrast, the augmented model showcases a higher level of precision in its recommendations and is less affected by biases driven by item popularity. This suggests that the augmented model's recommendations are more contextually relevant and excel at precisely recommending cold start items with similar contextual items., especially when compared to the baseline model

Future Recommendation

A potential future recommendation is to enhance the recommendation system by introducing an additional step in the pipeline where user preferences are actively taken into account. This step could help address the cold start problem from the user's perspective. By augmenting user preferences with additional item features generated from their interactions and choices, the system can become more attuned to individual user preferences and provide more personalized recommendations, even in scenarios with limited historical data. This user-centric approach could significantly improve the recommendation experience and adapt more effectively to each user's unique tastes and preferences.

Reference

Veres, C. (2022). Large Language Models are Not Models of Natural Language: They are Corpus Models. *IEEE Access*, *10*, 61970–61979. https://doi.org/10.1109/ACCESS.2022.3182505

Li, H., & Han, D. (2021). A Novel Time-Aware Hybrid Recommendation Scheme

Combining User Feedback and Collaborative Filtering. *IEEE Systems Journal*, *15*(4), 5301–5312. https://doi.org/10.1109/JSYST.2020.3030035

Haque, S., Eberhart, Z., Bansal, A., & McMillan, C. (2022). Semantic Similarity Metrics for Evaluating Source Code Summarization. *IEEE International Conference on Program Comprehension*, *2022-March*, 36–47. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Zhu, Y., Lin, J., He, S., Wang, B., Guan, Z., Liu, H., & Cai, D. (2020). Addressing the Item Cold-Start Problem by Attribute-Driven Active Learning. *IEEE Transactions on Knowledge and Data Engineering*, *32*(4), 631–644. https://doi.org/10.1109/TKDE.2019.2891530

Xin, S., Li, Z., Zou, P., Long, C., Zhang, J., Bu, J., & Zhou, J. (2021). ATNN: Adversarial two-tower neural network for new item's popularity prediction in e-commerce. *Proceedings - International Conference on Data Engineering*, *2021-April*, 2499–2510. https://doi.org/10.1109/ICDE51399.2021.00282

Lu, Z., Kazi, R. H., Wei, L. Y., Dontcheva, M., & Karahalios, K. (2021). StreamSketch: Exploring Multi-Modal Interactions in Creative Live Streams. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1). https://doi.org/10.1145/1122445.1122456

Chia, E. J., & Najafabadi, M. K. (2022). Solving Cold Start Problem for Recommendation System Using Content-Based Filtering. *Proceedings - 2022 International Conference on Computer Technologies, ICCTech 2022*, 38–42. https://doi.org/10.1109/ICCTech55650.2022.00015

Yanli Guo, nd. (n.d.). *Movie Big Data Intelligent Recommendation System Based on Knowledge Graph*. https://doi.org/10.1109/ISPA-BDCloud-SocialCom-SustainCom52081.2021.00080

Emon, M. I., Shahiduzzaman, M., Rakib, M. R. H., Shathee, M. S. A., Saha, S., Kamran, M. N., & Fahim, J. H. (2021). Profile Based Course Recommendation System Using Association Rule Mining and Collaborative Filtering. *2021 International Conference on Science and Contemporary Technologies, ICSCT 2021*. https://doi.org/10.1109/ICSCT53883.2021.9642633

Wibowo, F. W., & Institute of Electrical and Electronics Engineers. (n.d.). *2018 International Conference on Information and Communications Technology (ICOIACT) : 6-7 March 2018.*

Deng, Y. (2022). Recommender Systems Based on Graph Embedding Techniques: A Review. In *IEEE Access* (Vol. 10, pp. 51587–51633). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2022.3174197

Dou, Y., Yang, H., & Deng, X. (2016). *A Survey of Collaborative Filtering Algorithms for Social Recommender Systems*. https://doi.org/10.1109/SKG.2016.23

Yan, L., Qin, Z., Zhuang, H., Wang, X., Bendersky, M., & Najork, M. (2022). Revisiting Two-tower Models for Unbiased Learning to Rank. *SIGIR 2022 - Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2410–2414. https://doi.org/10.1145/3477495.3531837

Guan, R., Pang, H., Giunchiglia, F., Liang, Y., & Feng, X. (2022). Cross-Domain

Meta-Learner for Cold-Start Recommendation. *IEEE Transactions on Knowledge and Data Engineering*. https://doi.org/10.1109/TKDE.2022.3208005

Thank you !!!