`MMRec`: Simplifying Multimodal Recommendation

Xin Zhou
Nanyang Technological University
`xin.zhou@ntu.edu.sg`

Abstract

This paper presents an open-source toolbox, `MMRec` for multimodal recommendation. `MMRec` simplifies and canonicalizes the process of implementing and comparing multimodal recommendation models. The objective of `MMRec` is to provide a unified and configurable arena that can minimize the effort in implementing and testing multimodal recommendation models. It enables multimodal models, ranging from traditional matrix factorization to modern graph-based algorithms, capable of fusing information from multiple modalities simultaneously. Our documentation, examples, and source code are available at https://github.com/enoche/MMRec.

1 Introduction

Multimodal recommendation models is a raising trend in current research community due to the following reasons:

- The prevalence of multimodal information (e.g. images, texts and videos);
- Leveraging multimodal information in recommendation can address the sparsity of interaction data [1, 2];
- The maturity of recent researches on multimodal learning in NLP and CV domains.

Compared with the conventional recommender systems [3–5] that solely leverage user-item interactions for recommendation, multimodal models involves the processes of preprocessing information from multiple modalities (e.g. k-coring user/item filtering, data splitting, vectorizing multimodal information, aligning items IDs with their multimodal information), fusing multimodal information etc. On the one hand, these tedious processes impede the progress of multimodal recommendation research. On the other hand, the wide variety of preprocessing methods make the models difficult to reproduce its performance and compare fairly with others.

To address these issues, we present `MMRec`, an open-source toolbox that simplifies the research on multimodal recommendation. `MMRec` provides a full stack toolbox that includes data preprocessing, multimodal recommendation models, multimodal information fusion, performance evaluation to minimize the cost of implementation and comparison of novel models or baselines. The toolbox is highly configurable and user-friendly.

2 Architecture and Key Features

Figure 1 presents the modules of `MMRec`. Noting the inputs of `MMRec` are raw data of multimodal and user-item interaction files.

Data Encapsulation. `MMRec` first preprocesses raw data and encapsulates user interactions and multimodal information into `DataLoader` of Pytorch. The format of raw data is consistent with Amazon Review Data [1]. `MMRec` performs k-core filtering to retain the users and items with at least k interactions, and aligns the multimodal information with the retained items. It then splits the whole
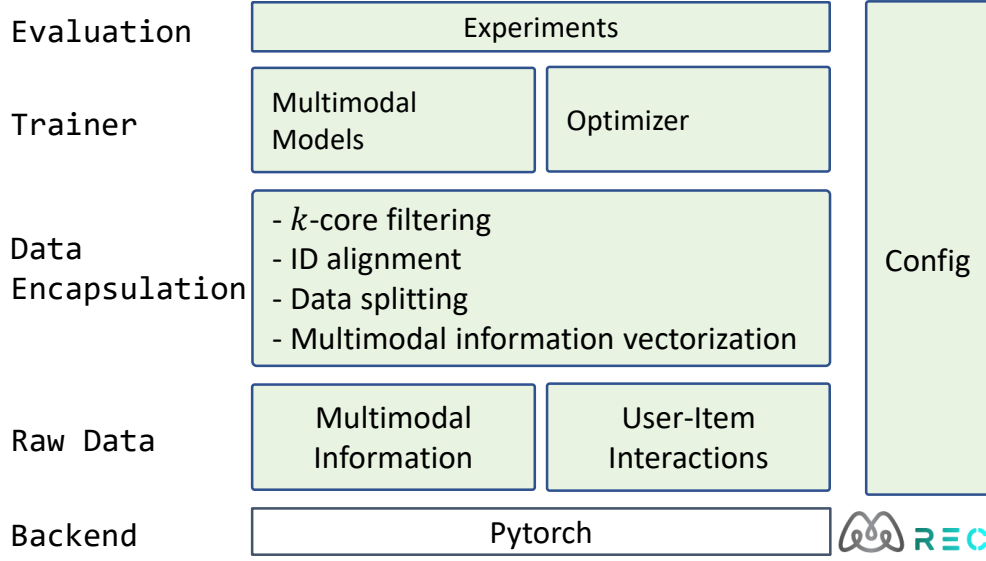
---

[1] http://jmcauley.ucsd.edu/data/amazon/links.html

Figure 1: The architecture of `MMRec`. `MMRec` consists of 4 modules ranging from raw data preprocessing to model performance evaluation. To ensure model reproducibility and performance consistency, `MMRec` consumes raw data as input.

interactions into Training/Validation/Test. The raw features of multimodal information are vectorized into numeric values leveraging pre-trained multimodal models, such as transformers.

Trainer. `MMRec` provides various optimizer to train the models. Both unimodal and multimodal models are supported in `MMRec`. In `MMRec`, information from each modality can be easily fused for multimodal recommendation. In its current version, `MMRec` supports four popular modalities: Text, Image, Audio, Video. `MMRec` unifies the training interface for all models. Customized models are merely required to implement two functions:

`calculate_loss:` The main part of the model, which defines how loss is generated from the model graph flow.

`full_sort_predict:` This function predicts the ranking of items for users.

Evaluation. This module features a wide set of commonly used metrics for recommender systems, such as Recall, NDCG, MAP etc.

It is worth noting that all modules can be customized and configured by modifying the configuration files. All changes will be reflected and loaded in **config** module. The **config** module also supports grid searching of models on hyperparameters. The results from all combinations of hyperparameters are summarized and reported to end-users after training. Reproducibility of models are secured by resetting the seed and raw data in `DataLoader` in each hyperparameter combination.

3 Comparison to Related Works

As far as we know, Cornac [6] is the only open-source library that supports multimodal information in recommendation. Although we observe 40+ algorithms are implemented in Cornac repository [2], most of them are general recommenders systems that do not utilize multimodal information. Furthermore, Cornac is limited by the fusion of modalities in its current stage. Specifically, Cornac cannot integrate multimodal information (i.e. text, image, audio, video) from more than one modality. The detailed comparisons between Cornac and `MMRec` are presented in Table 1.

---

[2]https://github.com/PreferredAI/cornac

Frameworks	Multimodal Algorithms				Multimodal Feature Fusion
	Text	Image	Audio	Video	
Cornac	✓	✓			
MMRec	✓	✓	✓	✓	✓

Table 1: Comparison of multimodal recommendation frameworks.

4 Conclusion

We presented `MMRec`, a multimodal recommendation library to facilitate researchers to implement their models and compare the results with the state-of-the-art baselines. Full documentation and source code are available at `https://github.com/enoche/MMRec`.

References

[1] Xin Zhou, Hongyu Zhou, Yong Liu, Zhiwei Zeng, Chunyan Miao, Pengwei Wang, Yuan You, and Feijun Jiang. Bootstrap latent representations for multi-modal recommendation. *arXiv preprint arXiv:2207.05969*, 2022. 1

[2] Xin Zhou. A tale of two graphs: Freezing and denoising graph structures for multimodal recommendation. *arXiv preprint arXiv:2211.06924*, 2022. 1

[3] Lingzi Zhang, Yong Liu, Xin Zhou, Chunyan Miao, Guoxin Wang, and Haihong Tang. Diffusion-based graph contrastive learning for recommendation with implicit feedback. In *International Conference on Database Systems for Advanced Applications*, pages 232–247. Springer, 2022. 1

[4] Xin Zhou, Donghui Lin, Yong Liu, and Chunyan Miao. Layer-refined graph convolutional networks for recommendation. *arXiv preprint arXiv:2207.11088*, 2022.

[5] Xin Zhou, Aixin Sun, Yong Liu, Jie Zhang, and Chunyan Miao. Selfcf: A simple framework for self-supervised collaborative filtering. *arXiv preprint arXiv:2107.03019*, 2021. 1

[6] Aghiles Salah, Quoc-Tuan Truong, and Hady W Lauw. Cornac: A comparative framework for multimodal recommender systems. *J. Mach. Learn. Res.*, 21(95):1–5, 2020. 2