Project Proposal

1 Introduction

With the growth of digital music industry and popularity of music software and platform like Spotify, Google Play, etc., automatic music recommendation has become a hot research topic in the field of machine learning/information retrivial.

Two broad directions in music recommendation are *collaborative filtering* vs. *content-based recommendation*. While the collaborative filtering relies on the usage patterns (items that a specific user has consumed or rated which provide information about the user's preference more directly), content-based recommendation uses item content (audio signal) and metadata (artist, year of release, etc.). According to [Sla11], collaborative filtering generally outperforms content-based methods but it cannot handle cold-start problem (no usage data for new songs).

In this project, we aim to use a hybrid method of both content-based recommendation and collaborative filtering using mainly users' listening history as implicit feedback with deep learning methods to predict customs' preference. The dataset we will be using is from KKBOX. For more information about the dataset, please see here.

2 Related Work

Previous pioneering work presented by [HKV08] uses collaborative filtering with implicit datasets (no direct rating/preference information from users but only purchase histories) for TV shows recommendation. They transform the implicit feedback into preference-confidence paradigm and provide a latent vector algorithm to handle it. Inspired by their work, [VDS13] adopts their *weighted matrix factorization* (WMF) and *alternating least squares* (ALS) optimization methods for content-based music recommendation. However, different from other content-based methods such as Mcfee et al [MBL12], who use metric learning to learn a similarity metric for comparing bag-of-words representations of audio signals, [VDS13] adopts a deep convolutional neural network for predicting. They achieve a better performance and also handles the cold start problem so that new and unpopular songs can be recommended. Later on, Wang et al [WW14] develops a novel model which can unify the extraction and prediction into an automated simultaneous process, which further improves the performance of hybrid and content-based music recommendation systems.

3 Evaluation Criteria

3.1 Area under receiver operating characteristic curve

Area under receiver operating characteristic curve (short for ROC-AUC or simply AUC) is the area under the receiver operating characteristic curve (ROC). The latter one works by plotting the true positive rate (TPR = TP/ (TP+FN)) against the false positive rate (FPR = FP / (FP + TN)) at various threshold settings.

Since the result is a binary classification(whether a particular user will listen to a particular song). We can predict the probability of this event and evaluate this probability and the observed target under different threshold.

3.2 Song cluster prediction

Accuracy metrics by themselves do not provide enough insight into whether the recommendations are sound. To provide a more intuitive way. For each song, we searched for similar songs by measuring the cosine similarity between the predicted user patterns and compare this result with the ground truth. The result is presented into clusters of songs so that we can visualize it in an intuitive way.

4 Method

The task can be divided into three major parts: embedding the given data to an appropriate feature space, choosing an effective neural network model, and tuning the parameters by cross validation.

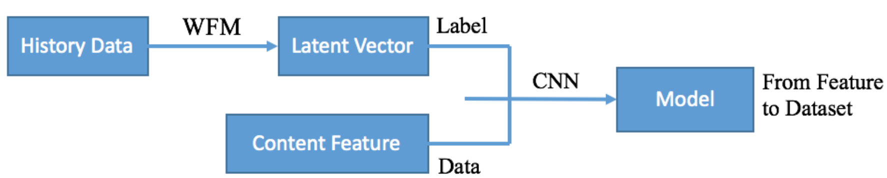
4.1 Data Processing & Tuning

As the names of the songs are available, we may try to augment the data (e.g. comments, reviews, or even audios) with online data by searching with its name. Process natural language data (lyricist, or augmented data as mentioned above) using techniques as TF-IDF and word2vec. Use one hot to embed categorical features. Use PCA, Random Forest or relevant techniques to rule out unimportant data and lower the dimension of the feature space.

Depending on the training time for each model, we may use k-fold validation or just hold-out validation to tune the parameters or network structures.

4.2 Models & Algorithms

We will mainly use the method suggested by [VDS13]. This method first uses weighted matrix factorization to learn latent user vectors and latent item vectors, then train a CNN with learned latent vectors as ground truth. Finally, use the trained CNN to predict the result. We may further improve its performance by adopting methods from [Che+16; WW14].



We may also try multilayer perceptron based methods such as [He+17; Lia+17; Che+16; Guo+17], or autoencoder-based methods such as item-based AutoRec [Sed+15] and its variations.

Except experimenting with different models, we may try to combine different models by either stacking, voting or directly embedding different network structures into one network.

References

[Che+16] Heng-Tze Cheng et al. "Wide & deep learning for recommender systems". In: *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*. ACM. 2016, pp. 7–10.

[Guo+17] Huifeng Guo et al. "DeepFM: A Factorization-Machine based Neural Network for CTR Prediction". In: *arXiv preprint arXiv:1703.04247* (2017).

[He+17] Xiangnan He et al. "Neural collaborative filtering". In: *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee. 2017, pp. 173–182.

[HKV08] Yifan Hu, Yehuda Koren, and Chris Volinsky. "Collaborative filtering for implicit feedback datasets". In: *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. Ieee. 2008, pp. 263–272.

[Lia+17] Jianxun Lian et al. "CCCFNet: A Content-Boosted Collaborative Filtering Neural Network for Cross Domain Recommender Systems". In: *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee. 2017, pp. 817–818.

[MBL12] Brian McFee, Luke Barrington, and Gert Lanckriet. "Learning content similarity for music recommendation". In: *IEEE transactions on audio, speech, and language processing* 20.8 (2012), pp. 2207–2218.

[Sed+15] Suvash Sedhain et al. "Autorec: Autoencoders meet collaborative filtering". In: *Proceedings of the 24th International Conference on World Wide Web*. ACM. 2015, pp. 111–112.

[Sla11] Malcolm Slaney. "Web-scale multimedia analysis: Does content matter?" In: *IEEE MultiMedia* 18.2 (2011), pp. 12–15.

[VDS13] Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen. "Deep content-based music recommendation". In: *Advances in neural information processing systems*. 2013, pp. 2643–2651.

[WW14] Xinxi Wang and Ye Wang. "Improving content-based and hybrid music recommendation using deep learning". In: *Proceedings of the 22nd ACM international conference on Multimedia*. ACM. 2014, pp. 627–636.