Hybrid Improvement on Bayesian Rank Prediction and WR-MF

MEHDI VALINEJAD, *Student*

***Abstract*—To be completed**

***Index Terms*—Hybrid Recommendation, Bayes Theorem, Bayesian Prediction, Item Based Recommendation System, Content Based Recommendation System, Weighted Ranking, Matrix Factorization, AUC, ROC.**

I. Introduction

ITEM recommendation plays a pivotal role in delivering tailored content or products to users, anticipating their preferences and needs. The fundamental objective of this task is to predict a personalized ranking among a myriad of items, such as websites, movies, stocks, cryptocurrency or products. In this paper, I delve into a prevalent scenario where implicit feedback, including user actions like clicks, acquisition or purchases, guides the recommendation process.

Common approaches to item recommendation from implicit feedback, such as matrix factorization (MF) or adaptive k-nearest-neighbor (kNN), have been designed with the aim of personalized ranking. However, an intriguing observation surfaces: these methods, despite their intended purpose, are not directly optimized for the intricate task of ranking. In response to this gap, my research uses a set of optimization criterion, termed BPR-Opt [1] and WR-MF [2], specifically tailored for personalized ranking. This criterion arises as the maximum posterior estimator derived from a Bayesian analysis of the problem.

In an effort to enhance the efficacy of BPR and WR-MF approach, I propose the incorporation of Content-Based analysis and ranking. This strategic addition aims to elevate the probability and evaluation metrics of BPR and WR-MF, further refining the accuracy and relevance of personalized recommendations. In the realm of content-based analysis, I aim to elevate the metrics of BPR and WR-MF through enhancements utilizing Cosine and Jaccard Similarity.

II. Methodology

To improve the performance metrics of Bayesian Personalized Ranking (BPR) and Weighted Regularized Matrix Factorization (WR-MF), a comprehensive formulation has been devised. This formulation is centered on extracting non-quantitative features from the content of items and leveraging them to establish similarity between items. The process involves selecting items with prior ratings for a specific user, considering them as potential items of interest related to the user's preferences. These selected items are then added to the list of highly ranked items for each user, thereby enhancing the recommendation system by incorporating both content-based and collaborative filtering approaches.

Given $Sr(u_i)$ as the set of similar items for a particular user u_i with ranking higher than min_r, the process involves computing the collection of movies that are similar to the user's top-rated movies (those with ratings higher than min_r).

$$Sr(u_i) = \begin{cases} Sims(m_i) & \text{if } r_i \geq min_r; \\ & \text{if } r_i < 3.0. \end{cases}$$

Following the computation of the list of new ratings for each user, the next step involves appending these calculated ratings into the original rating array:

$$N = \begin{bmatrix} Sr(u_1) \\ Sr(u_2) \\ \vdots \\ Sr(u_n) \end{bmatrix}$$

Encompassing both the pre-existing ratings and the newly calculated ratings, N amalgamated array serves as a comprehensive repository of user preferences, capturing the dynamics of both historical and recently generated ratings. this will be used as input variable to BPR and WR-MF algorithms for further and extended processing.

Leveraging genres as a pivotal content medium, the extraction process involved meticulously curating a list of movies exhibiting a similarity threshold surpassing 0.9 for each user's set of rankings. Following this, a nuanced approach was implemented to append new ratings to these movies. The methodology for assigning these ratings was intricately tailored to align with the user's existing ranked values, drawing inspiration from the reference movie ranks.

This comprehensive procedure was systematically executed for a dataset comprising 100 movies, ensuring a diverse representation across genres. To validate the robustness of the approach, the examination was extended to 100 users randomly sampled for each movie. This iterative analysis provided a nuanced understanding of how user preferences and rankings were dynamically influenced by the intricate interplay of genres and similarity metrics within the dataset.

1. BPR

Implemented the Bayesian personalized ranking [1] algorithm to forecast rankings using both the pure and extended ratings data frames. Employed the AUC Score metric as a benchmark study, serving as a pivotal measure for evaluating the prediction accuracy.

III. Results

The findings demonstrate a minor improvement, revealing a 0.5% enhancement in the prediction AUC score for the training set and a noteworthy 0.7% improvement for the test set when employing the Bayesian personalized ranking algorithm (BPR). This evaluation encompassed a dataset comprising 100 movies, and the user ranks were systematically shuffled across 100 iterations.

Notably, these positive outcomes underscore the efficacy of the BPR algorithm in refining predictive performance. Moreover, the observed gains hint at the potential for even more remarkable results by augmenting the dataset with an increased number of novel ranks. As the number of ranks expands, there is a promising prospect of further refining the algorithm's ability to capture nuanced patterns and enhance its overall predictive accuracy.

IV. Dataset

Movie Lens 1M

The ML-1M dataset was utilized to extract detailed information, encompassing rankings, movies, user profiles, and genre classifications. This dataset stands as a valuable resource, facilitating a comprehensive exploration and analysis of the intricate relationships and patterns within the realm of movies.

References

[1] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "Bpr: Bayesian personalized ranking from implicit feedback," 2012.

[2] H. Zhang, I. Ganchev, N. S. Nikolov, Z. Ji, and M. O'Droma, "Weighted matrix factorization with bayesian personalized ranking," in *2017 Computing Conference*, 2017, pp. 307–311.

Mehdi Valinejad Received the B.S. degree in industrial engineering from the Azad University (South Tehran Branch) in 2012, and is currently working Master's. degree at the University of Bahcesehir at Istanbul.