Neural Collaborative Filtering for Steam Dataset

Yueqiu Sun, Aodong Li

I. Introduction

IT is estimated that nearly 60 percent of Americans play video games, bringing in annual revenue of over $25 billion for PC gaming alone[1] The Steam digital distribution service is the largest digital stores for video games, thus providing a great place to start to look into users' behavior of purchasing video games. This paper aims to build a recommendation system for predicting user's preference towards video games based on the purchasing history obtained from Steam API using neural collaborative filtering. The paper also compare the performance between Generalized Matrix Factorization(GMF), one of the neural collaborative filtering model, and traditional Matrix Gactorization(MF).

II. Dataset

The data used in our analysis comes from the paper 'CONDENSING STEAM: DISTILLING THE DIVERSITY OF GAMER BEHAVIOR'[2]. The data is obtained by crawling through public profiles of 109 million gamers. Each gamer has a unique ID, for each ID, we can obtain his or her purchased games and the playing hours for each games. We will utilize the information of games owned and playing hours of each games for each gamers to build a collaborative filtering recommendation system.

III. Generalized matrix Factorization(GMF)

Matrix Factorization(MF) tries to capture user-item interaction using two matrixs which represent the latent space of the user and the item. The possible limitation of the matrix factorization is caused by the use of a simple and fixed inner product to estimate complex useritem interactions in the low-dimensional latent space.

The MF can be extended to incorporate more expressiveness and the complexity of the interactions by incorporating non-linearity setting to the model.

The Generalized matrix Factorization(GMF) can be formatted as follows. Let p_u denote the latent space of user u, and q_i denote the latent space of item i. The interaction between user u and item i can be written as follows,

$$y_{ui} = \sigma_{out}(c^T p_u \odot q_i)$$

where σ_{out} denote the activation function, and c is weights for the output of $p_u \odot q_i$. If σ_{out} is identity function, and matrix c is a vector of 1, we can recover MF.

[1]http://www.gamesindustry.biz/articles/2014-01-28-pc- gaming-market-to-exceed-USD25-billion-this-year-dfc/

[2]O'Neill, M., Vaziripour, E., Wu, J., Zappala, D. (2016). Condensing Steam (pp. 8195). Presented at the the 2016 ACM, New York, New York, USA: ACM Press. http://doi.org/10.1145/2987443.2987489

IV. Experiments

We conduct the experiment with the Steam data set to answer the following questions,

- Does the Generalized Matrix Factorization model outperforms the vanilla Matrix Factorization model?
- What does the the number of factors play in these too models?

The performance measure we use for a ranked list is the Hit Ratio (HR) and Nor- malized Discounted Cumulative Gain. In Particular, we will use HR@5 and NDCG@5 by truncating the ranked list to be top-5 list.

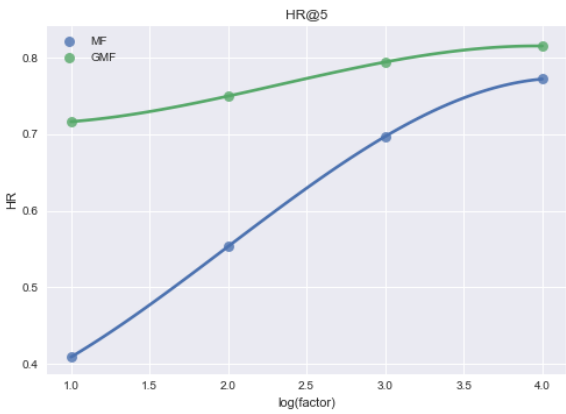

Fig. 1. HR@5 of MR and GMF methods w.r.t. the number of log(factors).

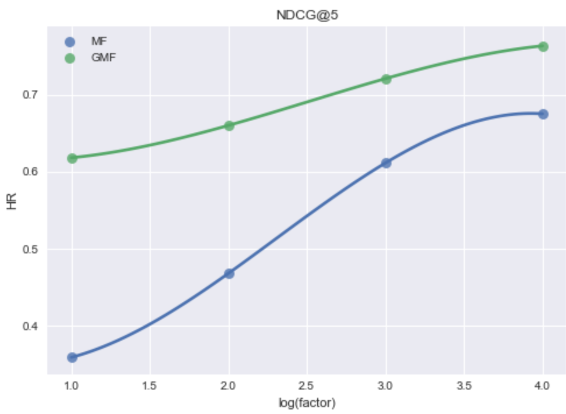

Fig. 2. NDCG@5 of MR and GMF methods w.r.t. the number of log(factors).

Figure 1 and 2 shows HR@5 and NDCG@5 for GMF and vanila MF. The GMF model outperforms the MF mdoel by a large margin. The more factors the model use, the more complex relationship between user and item the model can capture. More factors contirbute to the good performance of both model. The results also shows that the difference of performance for two models are getting smaller as the factors of the latent space increase. This is because the complexity of the neural networks is captured by the increase of the factors.

V. Conclusion

The paper examined the performance of Neural Collaborative Filtering on Steam Dataset, the largest game purchasing data set in public. We conclude that the generalized Matrix Factorization model outperforms the Matrix Factorization model. However, the difference in performance decrease as the factors of latent space increase.

References

[1] He, Xiangnan, et al. ”Neural collaborative filtering.” Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2017.

[2] O’Neill, M., Vaziripour, E., Wu, J., Zappala, D. (2016). Condensing Steam (pp. 8195). Presented at the the 2016 ACM, New York, New York, USA: ACM Press. http://doi.org/10.1145/2987443.2987489

[3] Zhang, Shuai, Lina Yao, and Aixin Sun. ”Deep learning based recommender system: A survey and new perspectives.” arXiv preprint arXiv:1707.07435 (2017).

[4] http://www.gamesindustry.biz/articles/2014-01-28-pc- gaming-market-to-exceed-USD25-billion-this-year-dfc/