PRESENTATION OUTLINE: A Parallel Recommender System Using a Collaborative Filtering Algorithm for Movie Recommender System

Projna Saha
School of Computer Science
Carleton University
Ottawa, Canada K1S 5B6
projnasaha@cmail.carleton.ca

November 21, 2021

1 Introduction

- Title of the presentation and self-introduction

2 Background of Recommendation Systems

- What are Recommendation Systems? [1]
- Types of Recommendation Systems
- Collaborative Filtering for Movie Recommendation

3 Types of Recommendation Systems

- Collaborative Filtering [2]
- Content-Based Filtering [3]

4 Literature Review

- Similarity Computation –

i Cosine Vector (CV) Similarity [4]

ii Pearson Correlation (PC) Similarity [5] [6]

iii JacRA Similarity [7]

iv Spearman Correlation (SC) [8]

- Rating Prediction –

i Weighted Average (WA) [9]

ii Mean-Centering (MC) [10] [11]

iii Z-Score (ZS) [12]

5 Introducing Spark Framework

- What is Apache Spark? [13]
- Components of Apache Spark [14]
- Hadoop vs Apache [15]
- The task scheduling procedure in Spark [16]

6 Methodology

- Algorithms that we have used –

i k-nearest neighbors (KNN) Algorithm [17]

ii Alternating Least Square (ALS) [18]

iii Linear Regression Analysis [19]

- Measuring Accuracy of The Model – MAE, MSE, RMSE [20]
- Comparison between our dataset vs the paper we followed [16]
- Methodology for MovieLens-100k Dataset
- Removing Noise From ML-100k Dataset
- Methodology for Netflix-6.2M Dataset
- Applying Gaussian Distribution on Netflix for Selecting More Effective Dataset

7 Experimental Results

- MovieLens-100k & Netflix-6.2M Data Statistics
- Movie Recommender System using KNN Algorithm :: MovieLens-100k
- Movie Recommender System using ALS Algorithm :: Netflix-6.2M
- Comparison Between Netflix-6.2M & ML-100K Datasets ALS Method
- The Space and Time Complexity of User-Based and Item-Based Methods
- Executor Summary of Two Datasets
- Measuring performance of two datasets

8 Ganglia Cluster Report

- System Information for Movielens Cluster
- System Information for Netflix Cluster
- Ganglia Cluster Report :: MovieLens-100k
- Ganglia Cluster Report :: Netflix-6.2M

9 Conclusion

- Limitations & future work
- Discussion

References

[1] F. Ricci, L. Rokach, and B. Shapira, "Recommender systems: Introduction and challenges," *Recommender Systems Handbook*, p. 1–34, 2015.

[2] R. Ji, Y. Tian, and M. Ma, "Collaborative filtering recommendation algorithm based on user characteristics," *2020 5th International Conference on Control, Robotics and Cybernetics (CRC)*, p. 56–60, 2020.

[3] A. Pal, P. Parhi, and M. Aggarwal, "An improved content based collaborative filtering algorithm for movie recommendations," *2017 Tenth International Conference on Contemporary Computing (IC3)*, p. 1–3, 2017.

[4] K. B. Fard, M. Nilashi, and N. Salim, "Recommender system based on semantic similarity," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 3, no. 6, 2013.

[5] M. Deshpande and G. Karypis, "Item-based top- n recommendation algorithms," *ACM Transactions on Information Systems*, vol. 22, no. 1, p. 143–177, 2004.

[6] P. Ahlgren, B. Jarneving, and R. Rousseau, "Requirements for a cocitation similarity measure, with special reference to pearsons correlation coefficient," *Journal of the American Society for Information Science and Technology*, vol. 54, no. 6, p. 550–560, Apr 2003.

[7] X. Wu, Y. Huang, and S. Wang, "A new similarity computation method in collaborative filtering based recommendation system," *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, 2017.

[8] J. Bobadilla, A. Hernando, F. Ortega, and A. Gutiérrez, "Collaborative filtering based on significances," *Information Sciences*, vol. 185, no. 1, p. 1–17, 2012.

[9] A. Bonfietti and M. Lombardi, "The weighted average constraint," vol. 7514, 10 2012, pp. 191–206.

[10] M. Hofer, "Mean centering," *The International Encyclopedia of Communication Research Methods*, p. 1–3, 2017.

[11] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," Jan 1998. [Online]. Available: https://arxiv.org/abs/1301.7363

[12] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR 99*, 1999.

[13] J. L. Reyes-Ortiz, L. Oneto, and D. Anguita, "Big data analytics in the cloud: Spark on hadoop vs mpi/openmp on beowulf," *Procedia Computer Science*, vol. 53, p. 121–130, 2015.

[14] M. Zaharia, R. S. Xin, P. Wendell, T. Das, M. Armbrust, A. Dave, X. Meng, J. Rosen, S. Venkataraman, M. J. Franklin, and et al., "Apache spark," *Communications of the ACM*, vol. 59, no. 11, p. 56–65, 2016.

[15] A. Wakde, P. Shende, S. Waydande, S. Uttarwar, and G. Deshmukh, "Comparative analysis of hadoop tools and spark technology," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018, pp. 1–4.

[16] J. Sun, Z. Wang, X. Luo, P. Shi, W. Wang, L. Wang, J.-H. Wang, and W. Zhao, "A parallel recommender system using a collaborative filtering algorithm with correntropy for social networks," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, p. 91–103, 2020.

[17] P. Cunningham and S. Delany, "k-nearest neighbour classifiers," *Mult Classif Syst*, vol. 54, 04 2007.

[18] S. Ghosh, N. Nahar, M. Wahab, M. Biswas, M. Hossain, and K. Andersson, *Recommendation System for E-commerce Using Alternating Least Squares (ALS) on Apache Spark*, 02 2021, pp. 880–893.

[19] T. Jhalani, V. Kant, and P. Dwivedi, "A linear regression approach to multi-criteria recommender system," vol. 9714, 06 2016, pp. 235–243.

[20] W. Wang and Y. Lu, "Analysis of the mean absolute error (mae) and the root mean square error (rmse) in assessing rounding model," *IOP Conference Series: Materials Science and Engineering*, vol. 324, p. 012049, 03 2018.