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 Types of Recommendation Systems

- Collaborative Filtering [1]
- Content-Based Filtering [2]

3 Literature Review

- Similarity Computation –

 i Cosine Vector (CV) Similarity [3]

 ii Pearson Correlation (PC) Similarity [4] [5]

 iii JacRA Similarity [6]

 iv Spearman Correlation (SC) [7]

- Rating Prediction –

 i Weighted Average (WA) [8]

 ii Mean-Centering (MC) [9] [10]

 iii Z-Score (ZS) [11]

4 Spark Framework

- Hadoop vs Apache [12]

- The task scheduling procedure in Spark [13]

5 Methodology

- Algorithms that we have used –

  i k-nearest neighbors (KNN) Algorithm [14]

  ii Alternating Least Square (ALS) [15]

  iii Linear Regression Analysis [16]

- Comparison between our dataset vs the paper we followed [13]

- 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

6 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

- Executor Summary of Two Datasets

7 Ganglia Cluster Report

- Ganglia Cluster Report :: MovieLens-100k

- Ganglia Cluster Report :: Netflix-6.2M

- MovieLens and Netflix Clusters Configurations - usage Overall

8 Conclusion

- Limitations & future work

- Discussion

References

[1] 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.

[2] 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.

[3] 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.

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

[5] 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.

[6] 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.

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

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

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

[10] 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

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

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