

Collaborative Filtering on Movie Recommendation Using Big Data

M. Manivannan¹, D. Sai Jahnvi^{2*}, A. Gowthami³, V. Jaswitha⁴, P. Vinay Yadav⁵, C. Jaya Prasad⁶

¹Associate Professor, Department of Computer Science and Engineering, Siddartha Institute of Science and Technology, Puttur, India

^{2,3,4,5,6}Student, Department of Computer Science and Engineering, Siddartha Institute of Science and Technology, Puttur, India

Abstract: These days, digitalization is increasing with rapid growth of personal and by home digital devices and the daily usage of internet, we generate a very large amount of amount of data, termed as "Big data". Movie recommendation systems can be enhanced to the needs of the users as individually. Collaborative filtering is a popular approach in big data domain to create recommendation systems. We describe a technique collaborative recommendation technique based on an algorithm specifically designed to mine association rules for this purpose. We use Alternating Least Squares. We use the Association rule mining approach to generate the rules to recommend movies to a user. We employ associations between users and association between the items. We build collaborative filtering in Apache spark. Apache Spark is the leading open- source unified analytics engine for big data processing. We use Euclidean distance similarity.

Keywords: Alternating Least Squares, Association rule, Big Data analysis, Collaborative filtering.

1. Introduction

Big Data is high-volume, high velocity, high variety information assets that demands the cost-effective, innovative forms of information processing for enhancing insight and decision making. Big data philosophy encompasses unstructured, semi-structured and structured data.

Collaborative Filtering is a technique used by recommender systems.

2. Existing System

In the existing system, hybrid methods have been implemented by bringing together content based and collaborative approaches of the system in making predictions. User will be recommended with items which he or she liked in the past. In this approach they will make use of the weighted technique in which the scores of each technique of collaborative filtering and content-based filtering are combined using a linear combination. Feature Weighted Linear Stacking approach is used for modeling by combining weights by making use of linear function for meta features. The output of one layer will be the input of the layer below it stacks. Similarity between the movies is computed using cosine similarity.

3. Proposed System

In this system, collaborative filtering is used to filter information by using the interaction and data collected by the system from other users. When we want to recommend something to a user, the best and most logical thing to do is to find people with similar interests so that we analyze their behavior, and make recommend our user the same items. Or we can look at the items which are similar to ones to which the user bought earlier and recommend movies which are like them. We describe a collaborative recommendation technique based on a algorithm using Alternating least Squares. Association Rule Mining is used to mine the association rules for this purpose. Rules are mined for a specific target user, which reduces the time required for big data processing. We make employ of the associations between the users as well as associations between items while making recommendations.

Historically, recommendation systems were devised with in mind without a regard for efficiency of computation. In our work, we will consider a computational efficiency. Using ARM. We build the project in Apache Spark. It makes use of the in-memory caching, and gives optimized query execution for fast analytic queries used against the data of any size. It is an open-source, it distributes the processing system that will be used for the big data workloads.

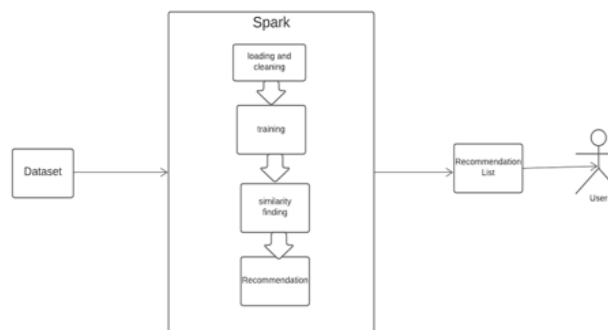


Fig. 1. System architecture

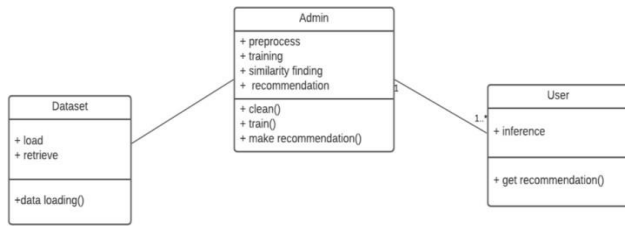


Fig. 2. Class diagram

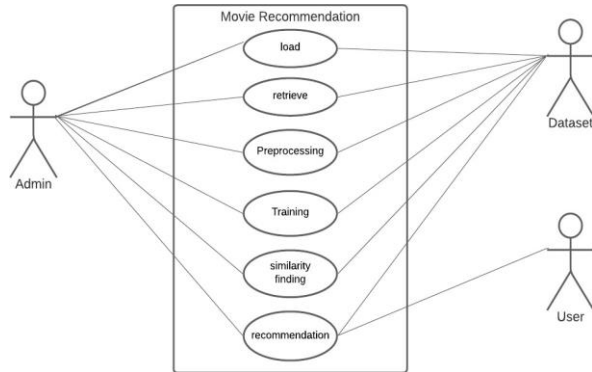


Fig. 3. Use case diagram

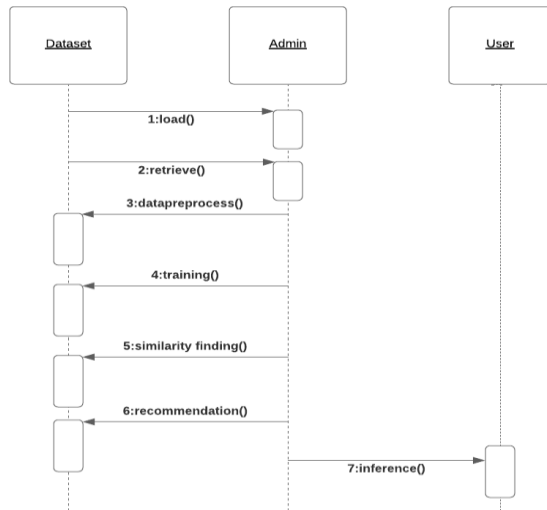


Fig. 4. Sequence diagram

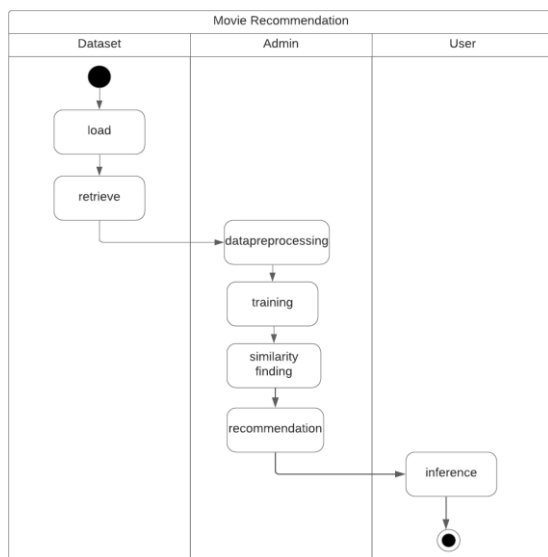


Fig. 5. Activity diagram

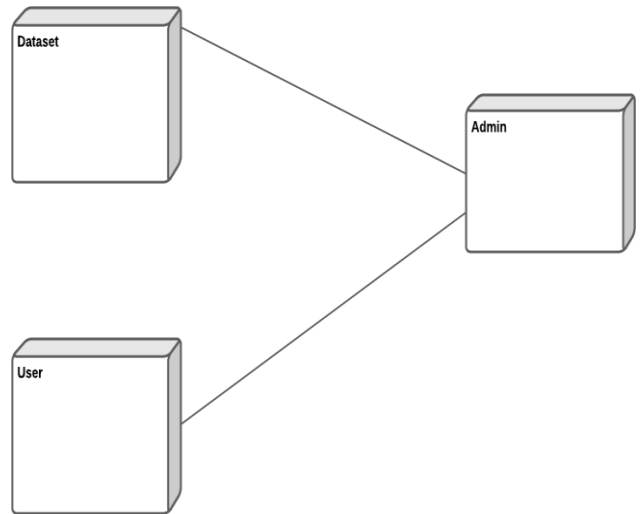


Fig. 5. Deployment diagram

4. Results

Recommended movies for the user are:

- The Million Dollar Hotel
- The 39 Steps
- License to Wed
- Terminator 3: Rise of the Machines
- 5 Card Stud

Recommendations for Iron Man:

- 1: Newton Boys, The (1998)
- 2: My Wife is an Actress (Ma Femme est une Actrice) (2001)
- 3: Glory Road (2006)
- 4: John Tucker Must Die (2006)
- 5: Strictly Sexual (2008)
- 6: Fired Up (2009)
- 7: Melancholia (2011)
- 8: Mirror Mirror (2012)
- 9: Best Exotic Marigold Hotel, The (2011)
- 10: Miss Peregrine's Home for Peculiar Children (2016)

Recommendations for Psycho:

- 'Big Fish',
- 'Star Wars',
- 'Titanic',
- 'Night on Earth',
- 'Rain Man',
- 'Sissi',
- 'Solaris',
- 'A Nightmare on Elm Street',
- 'The Hours',
- 'Monsoon Wedding']

5. Conclusion

We develop a movie recommendation system in Apache spark using collaboration filtering. We used Alternative Least Squares and Association Rule Mining methods. Our project analyzes the past data history of a user and predicts the future insights that are more likely to engage the user.

References

- [1] K. S and R. R. Badre, "Towards a Hybrid Recommendation System on Apache Spark," 2020 IEEE India Council International Subsections Conference (INDISCON), 2020, pp. 297-302.
- [2] A. Wijayanto and E. Winarko, "Implementation of multi-criteria collaborative filtering on cluster using Apache Spark," 2016 2nd International Conference on Science and Technology-Computer (ICST), 2016, pp. 177-181.
- [3] Sasmita Panigrahi, Rakesh Ku. Lenka, Ananya Stitipragyan, "A Hybrid Distributed Collaborative Filtering Recommender Engine Using Apache Spark," *Procedia Computer Science*, Volume 83, 2016, pp. 1000-1006.
- [4] Chong Yang, Xiaohui Yu, Yang Liu, Yanping Nie, and Yuanhong Wang. 2016. Collaborative filtering with weighted opinion aspects. *Neurocomput.* 210, C (October 2016), 185–196.
- [5] B. S. S. Govind, R. Tene and K. L. Saideep, "Novel Recommender Systems Using Personalized Sentiment Mining," 2018 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), 2018, pp. 1-5.
- [6] S. Hu, L. Chang and Y. Chen, "A Collaborative Filtering Recommendation Method based on Knowledge Points and Learning Ability Evaluation Model," 2020 IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI), 2020, pp. 264-267.
- [7] M. A. Lambay and S. Pakkir Mohideen, "Big Data Analytics for Healthcare Recommendation Systems," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 2020, pp. 1-6.
- [8] Chenjiao Feng, Jiye Liang, Peng Song, and Zhiqiang Wang. 2020. A fusion collaborative filtering method for sparse data in recommender systems. *Inf. Sci.* 521, C (Jun 2020), 365–379.
- [9] Y. Cai, H. -f. Leung, Q. Li, H. Min, J. Tang and J. Li, "Typicality-Based Collaborative Filtering Recommendation," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 3, pp. 766-779, March 2014.
- [10] J. Sun et al., "A Parallel Recommender System Using a Collaborative Filtering Algorithm with Correntropy for Social Networks," in *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 91-103, 1 Jan.-March 2020.