CS535 – Data Mining – Recommender System

Madhan Thangavel
Binghamton University
Vestal, NY 13902, USA
mthanga1@binghamton.edu

1. How my solution works

Here, in this project I have used Centered cosine similarity and user based similarity algorithm. By measuring the angle between the vectors, we can get a good idea of their similarity, and to make things even easier, by taking the Cosine of this angle.

Initially, the given dataset is stored into a matrix and cosine average is calculated then the matrix is adjusted based the cosine average, followed by rating values are predicted by iterating the user – item values. Prediction are made by rating for user 1 and item based on all the user who have rated the same item. This similarity is then used to predict potential user rankings for user-item pairs present in the dataset. We are going to predict all missing ratings based on the existing data. In order to do that, we need to compare the user-item ratings with differences of averages with matrix is calculated

Finally the rating values are adjusted in the range [1, 5] and retrieved in the resultant output.txt.

2. Strategies for cold start issue

Approaches:

At first, users rate different items in the system. Next, the algorithm calculates the similarities. After that, the system makes predictions for user-item ratings, which the user hasn't rated yet.

- Representative based: Representatives can be users whose linear combinations of preferences accurately approximate other users. For example, a famous representative based method, Representative Based Matrix Factorization (RBMF) is an extension of MF methods with an additional constraint that 'm' items should be represented by a linear combination of 'k' items.
- Deep learning: recent methods that tries to solve some of the issues tackled above but using a black box.
    - DropoutNet: This approach is that, in training deep learning based recommendation systems (e.g. collaborative filtering with many

layers), we can make it robust against the cold items by randomly dropping ratings of items and users. The key here is that, as opposed to standard dropout in neural net training, they drop the features, not the nodes.

  - Session-based RNN: This approach tries to utilize each session of users by feeding it into an RNN. Specifically, they trained a variant of Gated Recurrent Unit (GRU), where the input is the current state of the session and the output is the item of the next event in the session [1].

3. Examples of using a solutions of matrix completion problem

- Collaborative filtering:

  Collaborative filtering is the task of making automatic predictions about the interests of a user by collecting taste information from many users. Companies like Apple, Amazon, Barnes and Noble, and Netflix are trying to predict their user preferences from partial knowledge. In these kinds of matrix completion problem, the unknown full matrix is often considered low rank because only a few factors typically contribute to an individual's tastes or preference.

- System identification:

  In control, one would like to fit a discrete-time linear time-invariant state-space model to a sequence of inputs and outputs. The vector is the state of the system at time and is the order of the system model. From the input/output pair, one would like to recover the matrices and the initial state. This problem can also be viewed as a low-rank matrix completion problem.

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned}$$

4. Mini survey for the papers of recommended systems.

- Exploring Hybrid Recommender Systems for Personalized Travel Applications [2]

  This paper explains the research problems in the e-Tourism applications and presents the possible solution to achieve better personalized recommendations. For this, the authors have developed a Personalized Context-Aware Hybrid Travel Recommender System (PCAHTRS) by incorporating user’s contextual information. The proposed PCAHTRS is evaluated on the real-time large-scale datasets of Yelp and TripAdvisor. The experimental results depict the improved performance of the proposed approach over traditional approaches. The paper concludes with future work guidelines to help researchers to achieve fruitful solutions for recommendation problems.

- Research paper recommender system based on public contextual metadata [3]

  This paper proposes an alternative approach for proactively recommending scholarly papers to individual researchers using public contextual metadata for an independent framework that customizes scholarly papers, regardless of the research field and user expertise. This eliminates the drawbacks such as difficulty in recommending for new users as they are dependent on user profiles and limited scope of recommendation.

- Cultivation time recommender system based on climatic conditions for newly reclaimed lands on Egypt [4]

  This research proposes cultivation-time recommender system for predicting the best sowing dates for winter cereal crops in the newly reclaimed lands in Farafra Oasis, The Egyptian Western Desert. The main goal of the proposed system is to support the best utilization of farm resources. In this research, predicting the best sowing dates for the aimed crops is based on weather conditions prediction along with calculating the seasonal accumulative growing degree days (GDD) fulfillment duration for each crop. Various Machine Learning (ML) regression algorithms have been used for predicting the daily minimum and maximum air temperature based on historical weather conditions data for twenty-five growing seasons.

5. Other issues with recommender systems

- Synonymy

  Synonymy arises when an item is represented with two or more different names or entries having similar meanings. In such cases, the recommender cannot identify whether the terms represent different items or the same item. For example, a memory-based CF approach will treat "comedy movie" and "comedy film" differently. The variation in using descriptive terms is greater than commonly thought and the excessive usage of synonym words decreases the performance of CF recommenders. Since item contents are thoroughly ignored, therefore, the recommender does not consider the latent association between items. This is the reason why new items are not recommended as long as these are rated by the users. To alleviate the problems of synonymy, different techniques including ontologies, the Single Value Decomposition (SVD) techniques, and Latent Semantic Indexing (LSI) could be used [5].

- Shilling Attacks
  Malicious attacks can break the trust on the recommender system as well as decrease the performance and quality of recommenders. This threat is of more concern in CF techniques but lesser threat to the item-based CF technique. These types of attacks can be categorized by dimensions like intent of attacking, size of attack, and the required knowledge to start the attack.

- Privacy
  Feeding personal information to the recommender systems results in better recommendation services but may lead to issues of data privacy and security. Users are reluctant to feed data into recommender systems that suffer from data privacy issues.

- Grey Sheep
  Grey sheep occurs in pure CF systems where opinions of a user do not match with any group and therefore, is unable to get benefit of recommendations. Pure CB filtering can resolve this issue where items are suggested by exploiting user personal profile and contents of items being recommended. Similarly, sparse rating and first rater in CF filtering can also be resolved by CB filtering. Integrating CB with CF techniques may also yield more serendipitous and novel suggestions.

6. References

[1] Kojin Oshiba, At Robust Intelligence, https://kojinoshiba.com/recsys-cold-start/.

[2] Logesh R., Subramaniyaswamy V. (2019) Exploring Hybrid Recommender Systems for Personalized Travel Applications. In: Mallick P., Balas V., Bhoi A., Zobaa A. (eds) Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing, vol 768. Springer, Singapore. https://doi.org/10.1007/978-981-13-0617-4_52

[3] Haruna, K., Ismail, M.A., Qazi, A. et al. Research paper recommender system based on public contextual metadata. Scientometrics 125, 101–114 (2020). https://doi.org/10.1007/s11192-020-03642-y

[4] Nashwa El-Bendary, Esraa Elhariri, Maryam Hazman, Samir Mahmoud Saleh, Aboul Ella Hassanien, Cultivation-time Recommender System Based on Climatic Conditions for Newly Reclaimed Lands in Egypt, Procedia Computer Science, Volume 96, 2016, Pages 110-119, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2016.08.109. (http://www.sciencedirect.com/science/article/pii/S187705091631910X)

[5] Recommender Systems: Issues, Challenges, and Research Opportunities Shah Khusro, Zafar Ali and Irfan Ullah, Springer Science+Business Media Singapore 2016, DOI: 10.1007/978-981-10-0557-2_112.

[6] G. Guo, J. Zhang, Z. Sun and N. Yorke-Smith, LibRec: A Java Library for Recommender Systems, in Posters, Demos, Late-breaking Results and Workshop Proceedings of the 23rd Conference on User Modelling, Adaptation and Personalization (UMAP), 2015.

[7] J. Breese, D.. Heckerman, and C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering, Proc. Conf. Uncertainty in Artificial Intelligence, (UAI98) 1998

[8] J.L. Herlocker, J.A. Konstan, J.R.A. Borchers, and J. Riedl, An algorithmic framework for performing collaborative filtering, Proc. International on ACM SIGIR Research and Development in Information Retrieval, (SIGIR98) 1998

[9] P. Resnick and H.R. Varian, Recommender Systems, Special Issue of Communications of the ACM, 40(3), 1997