Technical Review

Vaibhav Karanam

CS 410 (Fall 2022)

Introduction

This technology review is about the comparison between memory-based collaborative filtering versus model-based collaborative. Collaborative filtering was discussed in week six lectures as part of the recommender systems. While two sub-parts of memory-based collaborative filtering were discussed, model-based methods were not shown. With this technology review, I wanted to dive into model-based methods and how they differ from the memory-based methods shown in lectures. The review will further go into the comparison between the two and how they are used respectively.

Memory-Based

Memory-based is one of the sub-parts of collaborative filtering. In this approach, the goal is to find similar entities. Using this goal, this approach can be further divided into two categories – user filtering and item filtering. In user filtering, the goal is to find users who like the same attribute and form a group. This means we want to combine users with similar tastes and recommend items that other users in the group also like. Item filtering is based on a similar goal. However, instead of creating a group of similar users, we find users who like a particular item and then find other items they liked and recommend them to the users in the group.

Model-Based

The other sub-part of collaborative filtering is a model-based method. In this approach, machine learning is applied to predict users' preferences and what items users might like. This method is entirely different from the memory-based approach, where the decision is made on the similarity of the users and items. Here, most decisions are based on predicting the user's preference given the collected/presented data.

Workings

In memory-based methods, the similarity is usually calculated using some distance similarity like Cosine similarity, or if it is implemented using clustering, then K-nearest neighbors can be applied. Whereas in model-based methods, while clustering-based algorithms can usually be utilized, matrix factorization is popular as it offers an advantage to using multi-layer neural networks. Singular Value Decomposition (SVD) also offers an attractive solution for matrix factorization as the matrix is sparse.

Advantages of Memory-based Collaborative filtering

Memory-based methods work well on smaller-level data and give accurate quality predictions. They are also simple algorithms to apply as they usually use simple distance measures like cosine similarity or KNN for clustering methods. Since they use distance measures to predict similarity, the memory-based models account for every data point in the database when making decisions. This increases the accuracy of the prediction while increasing the odds that a strong prediction will be made.

Disadvantages of Memory-based Collaborative filtering

While using the entire database to make predictions offers strengths, it also disadvantages large data sets. The memory-based methods will be slow as the data points increase due to the calculation needed to find distance increases with the number of points. The performance also reduces when the data matrix is sparse. Another issue is that the predictions might overfit the data if there is a lot of noise or sparsity. This becomes worse when using clustering algorithms such as KNN since noise clusters can pull the prediction to deviate and not be accurate.

Advantages of Model-based Collaborative filtering

The model-based methods offer expandability to large datasets. Model-based methods allow better scalability to enormous datasets; this makes it much more attractive for working with real systems and big data. This is because while memory-based methods have to use the whole dataset to make predictions, model-based methods can use matrix factorization or neural networks to make decisions. This is also the reason why model-based offers faster prediction

speeds. Another reason why the model-based method is better on real large datasets is that it avoids overfitting to noise which is a disadvantage in memory-based methods.

Disadvantages of Model-based Collaborative filtering

One of the reasons why model-based methods can provide disadvantages is that it is harder to customize once the model is pushed to production. Adding data adds a challenge, as model-based methods do not rely on the whole dataset. The new data might not make influence the prediction-making. Another reason why the model-based method might not be best suited for all the datasets because the prediction quality might not be as good as the memory-based methods. Since the model-based method does not use the whole dataset, the prediction quality might suffer due to not considering the edge cases or rare results.

Conclusion

While in the beginning, the goal of this paper was to find which collaborative filtering is the best one out there. The review showed that each method offers its pros and cons and that each method's use depends on the dataset and the approach used. From the study, it might seem that the complex model-based method will be the best due to the implementation of matrix factorization and neural network expandability. However, as seen with the disadvantages, it might not be the best choice in some instances and might hurt the accuracy of the predictions.

References

Candillier, Laurent & Meyer, Frank & Boullé, Marc. (2007). *Comparing State-of-the-Art Collaborative Filtering Systems*. 4571. 548-562. 10.1007/978-3-540-73499-4_41.

Pocs, Mate. *How to Build a Model-Based Recommendation System Using Python Surprise.* https://towardsdatascience.com/how-to-build-a-model-based-recommendation-system-using-python-surprise-2df3b77ab3e5.

"Algorithm Performance." *Recommender Systems :: Results*, https://www.cs.carleton.edu/cs_comps/0607/recommend/recommender/results.html.

"Algorithms." *Algorithms :: Memory-Based Algorithms*, https://cs.carleton.edu/cs_comps/0607/recommend/recommender/memorybased.html.

"Algorithms." *Algorithms :: Model-Based Algorithms*, https://www.cs.carleton.edu/cs_comps/0607/recommend/recommender/modelbased.html.

Azhari, Gazza. "Model-Based Collaborative Filtering Systems with Machine Learning Algorithm." *Medium*, Medium, 13 Feb. 2019, https://medium.com/@gazzaazhari/model-based-collaborative-filtering-systems-with-machine-learning-algorithm-d5994ae0f53b.

Grover, Prince. "Various Implementations of Collaborative Filtering." *Various Implementations of Collaborative Filtering*, https://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0.

M, Darshan. "What Is Cosine Similarity and How Is It Used in Machine Learning?" *Analytics India Magazine*, 7 July 2022, https://analyticsindiamag.com/cosine-similarity-in-machine-learning/.

Maklin, Cory. "Memory Based Collaborative Filtering - User Based." *Medium*, Medium, 10 Aug. 2022, https://medium.com/@corymaklin/memory-based-collaborative-filtering-user-based-42b2679c6fb5.

Pocs, Mate. "How to Build a Memory-Based Recommendation System Using Python Surprise ..." *How to Build a Memory-Based Recommendation System Using Python Surprise*, https://towardsdatascience.com/how-to-build-a-memory-based-recommendation-system-using-python-surprise-55f3257b2cf4.