About Performance Evaluation of the Movie Recommendation Systems

Shreya Agrawal
(ME Student)
CSE
SVITS
Indore, MP, India

Pooja Jain
(Assistant Professor)
CSE
SVITS
Indore, MP, India

ABSTRACT
Movie recommendation systems are now becoming very popular both commercially and also in the research community, where many approaches have been proposed for providing recommendations. For more and more usage of any system, it is necessary to know about the efficiency of the system and for this reason performance evaluation of a Recommendation system is done. By doing the performance evaluation of a system, one can prove the potential of a recommendation system. The more high performance a system gives more is its worth as compared to others. And, on this basis we can get to know further research and improvement options for a system which gives rise to new advancements in the field. Indeed, movie recommendation systems have a number of properties that may affect user's experience, such as accuracy, quality, robustness, scalability, and so forth. In this paper, various important performance evaluation metrics are reviewed and discussed.

Keywords
Movie recommendation systems, performance evaluation, accuracy, scalability, quality

1. INTRODUCTION
Day by day as the people are getting easy access to the internet, there is an increase in the action on the Web. Whether it is a commercial usage or domestic usage, the Internet usage rates are rising towards the sky. Nowadays, users are having lots of data within their reach and also it is rising exponentially. In such an information overload condition, users wish to locate desired information quickly. So, a need arises for a system which can do this work for us easily.

Recommendation systems make recommendations of services or products that target users may like. A wide range of recommendation engines have been developed and also widely accepted like movie recommendation systems, music recommendation systems, job recommendation systems etc [2].

Collaborative filtering technique is the most used method in movie recommendation systems. It is based on the ratings provided to movies and make predictions on the basis of the interests of a user who has similar preferences to other like-minded users.

The system calculates similarities among users by using those ratings, and makes predictions about a user's liking for a movie which the user has not rated yet. For better recommendations, it is required that more and more users give ratings to more and more movies [2].

Earlier recommendation algorithms were mainly evaluated for accuracy [6]. In this paper, we have reviewed other aspects of performance evaluation also. There are many other parameters also on which the movie recommendation systems can be judged. These include prediction quality, scalability, computing time, memory requirements etc. Furthermore, prediction quality can be judged by four parameters like mean absolutes error, coverage, recall, precision.

Scalability means if the number of users or number of movies increases the system should cope up well and should be able to continue working. Computing time measures quickness of an algorithm which means that how quickly an algorithm provides good recommendations. Confidence metrics helps users to make more effectual decisions [6].

Novelty or Serendipity helps to ensure that a recommendation is not fake but real. Robustness deals with measuring the capability of an algorithm to make good predictions, whether data is sparse or noisy [6]. The coverage of a recommendation system is a measure of the domain of items over which the system can make recommendations. For movie recommendation systems, items will be movies.

Mean Absolute Error measures the average error in the predicted rating as compared to the true rating. This measure helps in checking system's accuracy. Recall is the measure of completeness that is the proportion of good recommendations that appear in top recommendations. Precision is the measure of exactness that is the proportion of recommendations that are good recommendations. Correctness measures how close are movie recommendations to a set of movie recommendations supposed to be right.

2. RELATED WORK
Iman Avazpour et al. [2014] have reviewed a span of evaluation metrics and measures as well as approaches which are used for evaluating recommender systems. They presented metrics which are being grouped under 16 different dimensions, for example novelty, correctness, coverage. They also reviewed these metrics according to their corresponding dimensions. An overview of approaches is presented using collections of recommender system dimensions and related metrics. They also provided advice for future research and practice way [1].

Francois Fouss et al. [2008] suggested measures evaluating other characteristics than accuracy of the recommendation algorithms. They investigated six collaborative recommendation methods. They showed that kernel-based algorithms provides the best results by analyzing the results on artificial data sets for robustness or on the real Movie Lens data set for novelty, accuracy, and computing time[6].

J. Herlocker et al. [2004] reviewed the conclusions required in evaluating a collaborative filtering recommendation system. They also presented empirical results by doing the analysis of various accuracy metrics. They strongly correlated the metrics within each equivalency class while uncorrelated the metrics from different equivalency classes [8].

3. PERFORMANCE EVALUATION METRICS FOR MOVIE RECOMMENDATION SYSTEM

According to many researches, we come to know that for evaluating performance of a movie recommendation system only accuracy is not the criteria [3] but we should also consider other important parameters. Recommendation quality is also equally significant metric.

Performance Evaluation Metrics can be divided into two parts [3]:

- Evaluation Metrics based on the recommendation algorithm

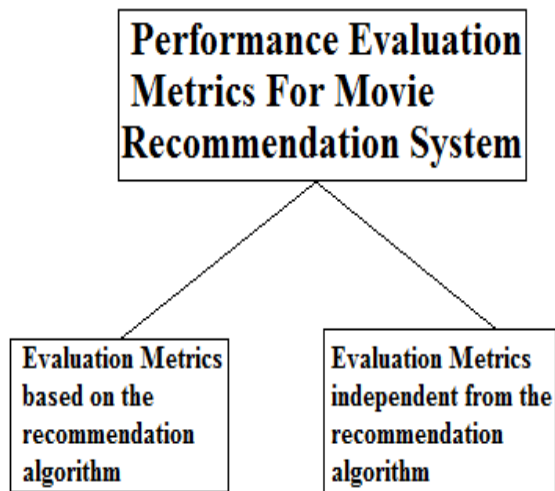- Evaluation Metrics independent from the recommendation algorithm

Figure 1: Performance Evaluation Metrics

3.1 Evaluation Metrics based on the recommendation algorithm

1. Diversity

Sometimes, it is of no use to users to know the similar values, since when there are a large number of items and it will take a longer time for exploring such a span of values. In these cases, diversity has its significance, since; diversity is the opposite of similarity [3].

2. Novelty

Novelty means new that means novelty in recommendations are those recommendations for movies that the user did not knew about before [3].

3. Serendipity

In content-based filtering recommendation system there is no serendipitous items, that is, serendipity is the capability of the system to give a product which appear surprisingly interesting to a customer, but not expected by the customer. For example, if a cosmetic of the same brand is recommended to the customer, the customer may already know regarding the cosmetic and, therefore, is not amazed. The recommendation would be serendipitous if the system is recommending another cosmetic of another company, and it seems unexpectedly interesting to the user. Two major properties of serendipity are attractive and surprising [3]. This means that more and more the recommendation is serendipitous, more and more the item is attractive, interesting as well as surprising [3].

4. Accuracy

Accuracy stands for the degree to which the result of a calculation, measurement, or specification fulfills to the precise value or a standard.

Accuracy can be classified among three different classes [3]:

- accuracy of ratings

- accuracy of usage

- accuracy of ranking of items

5. Coverage

The coverage of a movie recommendation system is a measure of the domain of movies over which the system can make recommendations. The term coverage was mainly associated with two concepts [4]:

1. Prediction Coverage: The percentage of the movies for which the system is able to produce a recommendation.

2. Catalogue Coverage: The percentage of the available movies which effectively are ever recommended to a user.

6. Mean Absolute Error (MAE)

Mean Absolute Error helps to measure the error that means it measures the average variation in the forecasted rating versus the true rating.

7. Recall

Recall is the measure of completeness, that is, the proportion of good movie recommendations that appear in top recommendations or the fraction of relevant movies retrieved out of all relevant movies.

8. Precision

Precision is the measure of exactness, that is, the proportion of movie recommendations that are good recommendations or the fraction of relevant movies retrieved out of all movies retrieved.
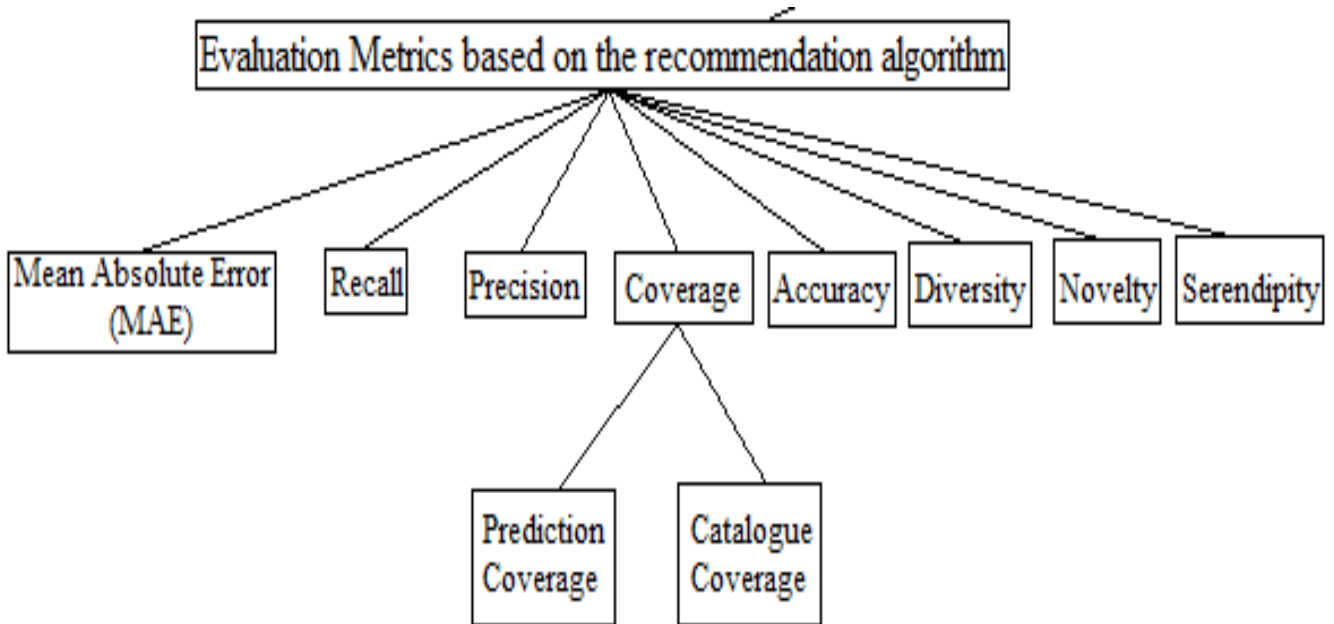
Figure 2: Evaluation Metrics based on the recommendation algorithm

3.2 Evaluation Metrics independent from the recommendation algorithm

3.2.1 From system's perspective

1. Scalability

The movie data and its users are increasing day by day on the web and algorithms for large, differential databases are weak that is scalability problem occurs. So to handle the problem of scalability, algorithm requires to be optimized.

2. Confidence [3]

Confidence refers to the recommendation system's faith in the predictions provided. The more the confidence the more is the quality of the recommendations.

3. Sparsity

Movie recommendation system works over the large datasets comprising of the set of users and the movies. It is assumed that data sparsity may cause smaller number of co-rated movies or no such movies between two users, resulting in untrue or unavailable similarity information as a result the concept of the nearest neighbor algorithms seems to be unable to predict any of the movies for any of the dynamic users, and further suffering from poor recommendation quality.

4. Adaptivity [3]

In the recommendation of news items like recommendations in the online newspapers, there are chances of changing the interest very rapidly means there is a fast shift in the taste of items. So, system must be adaptable to such swift changes in the trend in real environment.

5. Robustness [6]

Two features can be associated with the term robustness which is:

Accuracy: It refers to the system's accuracy that means does the system provide movie recommendations after an external deliberated or un-deliberated attempt to change the recommendations, movies that are actually valued.

Stability: It refers to the system's stability that means does the system provide different movie recommendations after an external deliberated or un-deliberated attempt to change the movie recommendations.

6. Computing Time [6]

Computing time is mainly classified into two components which are:

Training Phase: This phase calculates the time required by a particular algorithm to provide movie recommendations for the first time.

Run-Time Phase: This phase determines the time required to calculate new movie recommendations or to update old movie recommendations, when adding new movies, new users, or new links between them.

3.2.2 From user's perspective

1. User Preference [3]

User Preferences are the most important thing which should be taken care of while trying to improve a movie recommendation system. Until and unless a user is not satisfied with the system, he/she will not favor the system over the other. So, by getting to know the user preferences, we can satisfy user more and more by improving system in their direction.

2. Trust [3]

Trust denotes how much the user trusts the movie recommendation system. Trust of a user can be built up when the movie recommendation system is able to predict the user's interest accurately every time. Trust grows when the predicted movie recommendations maps to the real world user's movie preferences.

3. Privacy [3]

Privacy is the user's main concern nowadays, since day by day cyber crime is increasing. User gives ratings to several movies in the hope of getting beneficial movie recommendations. But, it is required for some users that their preferences should remain private.
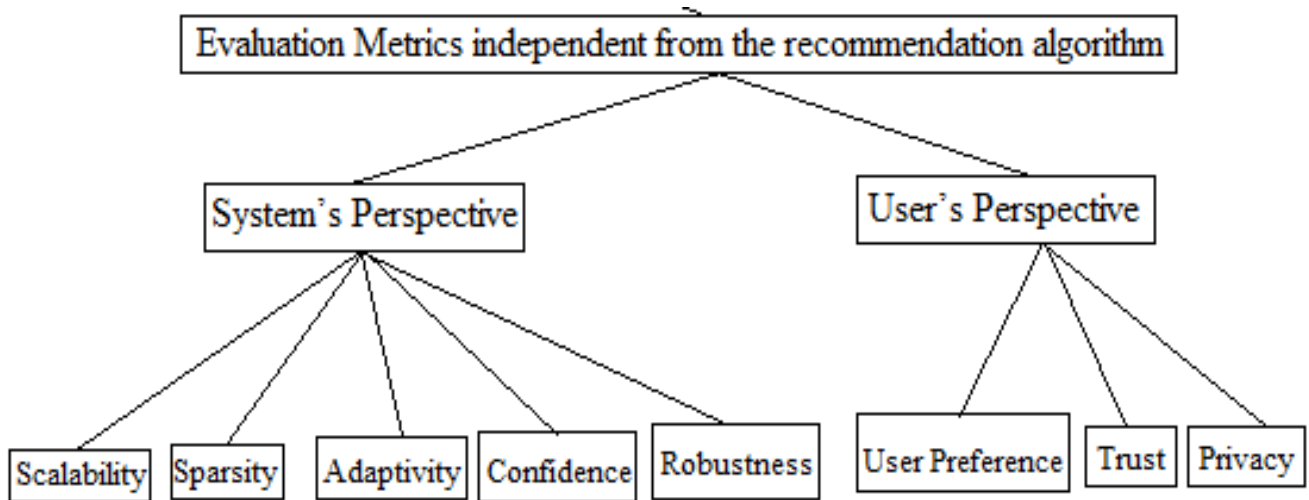
Figure 3: Evaluation Metrics independent from the recommendation algorithm

4. CONCLUSION

In this paper, we have reviewed and discussed the performance evaluation metrics required for movie recommendation systems. We have also surveyed related work in this field. We also hope that researchers will be benefitted from our description and it will assist them it constructing better movie recommendation systems in future. Basically, this paper will be helpful for researchers and practitioners in learning the performance evaluation criteria's for movie recommendation systems. This will aid in improving their choice while selecting a particular recommendation system for a project.

5. REFERENCES

[1] Iman Avazpour, Teerat Pitakrat, Lars Grunske and John Grundy (2014), "Dimensions and Metrics for Evaluating Recommendation Systems", Recommendation Systems in Software Engineering, Part II, pp 245-273, Springer Berlin Heidelberg.

[2] Woon-hae Jeong, Se-jun Kim, Doo-soon Park and Jin Kwak (2013), "Performance Improvement of a Movie Recommendation System based on Personal Propensity and Secure Collaborative Filtering", Journal of Information Processing Systems, Vol. 9, no. 1

[3] Wu, W., He, L., and Yang, J. (2012), "Evaluating Recommender Systems," 7th International Conference on Digital Information Management (ICDIM), pp.56–61

[4] Mouzhi Ge, Carla Delgado-Battenfeld, Dietmar Jannach (2010), "Beyond Accuracy: Evaluating Recommender Systems by Coverage and Serendipity", Proceedings of the fourth ACM conference on Recommender systems. ACM, 2010.

[5] Gunawardana, A. and Shani, G. (2009), "A Survey of Accuracy Evaluation Metrics of Recommendation Tasks", The Journal of Machine Learning Research, 10, p.2935-2962, 12/1/2009.

[6] Francois Fouss & Marco Saerens (2008), "Evaluating performance of recommender systems: An experimental comparison", IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.

[7] G. Adomavicius and A. Tuzhilin (2005), "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions", IEEE Transactions on Knowledge and Data Engineering, pages 734–749, 2005.

[8] J. Herlocker, J. Konstan, L. Terveen, and J. Riedl (2004), "Evaluating collaborative filtering recommender systems", ACM Transactions on Information Systems, 22(1):5–53, 2004.