International Journal for Rapid Research in Engineering Technology & Applied Science
ISSN (Online): 2455-4723
Vol 2 Issue 2 March 2016

"An Improved Hybrid Filtering Approach for Recommender System"

[1]Nitika Kadam, [2]ShraddhaKumar
[1]M.Tech Scholar, [2]Assistant Professor
[1]Department of Computer Science &Engineering,SDBCTIndore,India
[2]Department of Computer science &Engineering,SDBCTIndore,India

ABSTRACT:

Recommender System is a subclass of information filtering system. It identifies similarity among users or items. It can be used as information filtering tool in online social network. Collaborative filtering recommendations are based on similarity of users or items, all data should be compared with each other in order to calculate this similarity. Due to large amount of data in dataset, too much time is required for this calculation, and in these systems, scalability problem is observed. It is better to group data, and each data should be compared with data in the same group. Content based filtering recommends items to users according to users history and items he liked in the past. To improve the working and quality of recommender system, a Hybrid approach by combining content based filtering and collaborative filtering, which includes Memory based (K-Nearest Neighbor) and Model based (clustering and rule based techniques),is presented in our proposed methodology. Using Hybrid approach, we get advantages from each other while drawbacks of both methods won't take into account.

Keywords: Recommender system, content based filtering, collaborative filtering, similarity function.

1. INTRODUCTION

Recommender system is a system which filters the informationfor recommending some items to its users; it filters the data and recommends the items. It is commonly used in movie lens, book-crossing, jester, wiki-lens that uses a collaborative filtering to present information on items and products that are likely to be of interest to the consumers. User's interest in the past is seen and analyzed for the recommendation of any items. While presenting the recommendations, the recommender system uses details of the account of the registered user's profile, behavior, preferences and habit of their whole group of users and comparison of the information to present the recommendations. It much relies on similarity calculation.

Types of Recommender System

Recommendation System can be classified into 3 different categories based on technique used; they are content based method, collaborative method, and hybrid method.

Content-Based Filtering: Content-based filtering recommends items to users that are almost alike to the ones that the user wished or desired in the past. It is done by first building relation between item and its properties in term of Matrix then select the most similar items to the target item by computing similarity based on the features associated with the compared items using various mathematic functions. The most common similarity function that used are Adjusted Cosine, Cosine or Pearson coefficient. Good similarity measures will result a high prediction quality.

Collaborative Filtering: Collaborative Filtering(CF) is a method of identifying the similar clients and recommending what the common clients prefer. This system recommends items to the active user or the target users with that of the other users with similar preferences in the past. The similarity in preferences of the two users is calculated/evaluated based on the similarity in the rating history of the different users. This is the reason why it's also known as "people-to-people correlation." The most common process of CF performs similarity computation on collection of user preferences that E-Commerce websites usually collect from rating made against products. The calculation used is the same technique in Content-Based Recommendation but focus on peer opinions. This filtering is considered to be the most successful, popular and widely implemented technique in Recommender Systems.

Types of Collaborative Filtering

Collaborative filtering is mainly based on 2 Types of techniques; they are Memory-Based or User Based Collaborative Filtering and Model- Based or Item Based Collaborative Filtering

Memory-Based or User Based Collaborative Filtering: The memory based collaborative filtering uses the entire user-item database to generate the recommendations for the users. K-Nearest Neighbour is an algorithm which is generally used. In neighborhood based algorithms, first step is that a subset of clients are picked focused around their closeness to the dynamic client, and a weighted combo of their appraisals is utilized to generate expectations of items for the dynamic client. This methodology is popular and well-known for its straightforwardness and its productivity and also has been very flourishing in past, but it has some difficulties like Scalability and Data Sparsity.

Model-Based or Item Based Collaborative Filtering: The model based collaborative filtering uses the ratings provided to the items to recommend them to the users. Ratings of the items are given preferences. Clustering and rule based techniques are some of the well known techniques which are used. To deal with the sparse data model based collaborative filtering such as clustering techniques provide more accurate predictions.

Hybrid Recommender Systems: This recommender system is based on the combination of the content based filtering system and the collaborative filtering system techniques. Hybrid approaches can be implemented in several ways: by making content-based and collaborative-based predictions separately and then combining them; by adding content-based capabilities to a collaborative-based approach (and vice versa); or by unifying the approaches into one model. Hybrid methods can provide more accurate recommendations than pure approaches. These methods can also be used to overcome some of the common problems in recommender systems such as cold start and the sparsity problem.

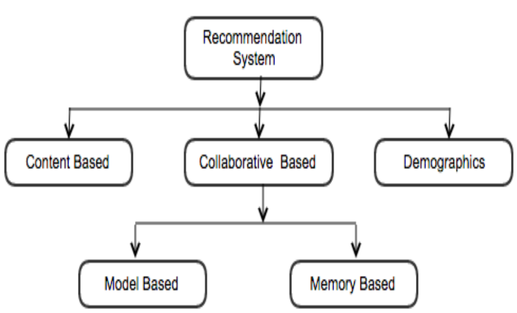

Figure 1.Type of Recommendation

2. LITERATURE REVIEW

HirdeshShivhareet. al. [2015] provides a combinatorial approach by combining fuzzy c means clustering technique and genetic algorithm based weighted similarity measure to develop a recommender system (RS). The proposed FCMGENSM recommender system provides better similarity metrics and quality than the ones provided by the existing GENSM recommender system but the computation time taken by the proposed RS is more than the existing RS.

Zan Wang et. al. [2014] proposed a hybrid model-based movie recommendation system which utilizes the improved K-means clustering coupled with genetic algorithms (GA) to partition transformed user space. In this way, the original user space becomes much denser and reliable, and used for neighborhood selection instead of searching in the whole user space. By this proposed method it will capable of generating effective estimation of movie ratings for new users via traditional movie recommendation systems.

Mohammad Yahya H. Al-Shamriet. al. [2014] proposed fuzzy weightings for the most common similarity measures for memory-based CRSs. Fuzzy weighting can be considered as a learning mechanism for capturing the preferences of users for ratings. Comparing with genetic algorithm learning, fuzzy weighting is fast, effective and does not require any more space. Experimental results show that fuzzy weighting improves the CRS performance irrespective of the fuzzy weighting variable where the fuzzy-weighted similarity measures outperform their traditional counterparts in terms of PCP, coverage, and mean absolute error.

F.Darvishi-mirshekarlouet. al. [2013] gave the review about the Recommender systems using collaborative filtering. It is the most popular and successful method that recommends the item to the target user. These users have the same preferences and are interested in it in the past. Scalability is the major challenge of collaborative filtering. With regard to increasing customers and products gradually, the time consumed for finding nearest neighbor of target user or item increases, and consequently more response time is required.

Forsati-Mohammadiet. al. [2013] introduces a new hybrid recommender system by exploiting a combination of collaborative filtering and content-based approaches in a way that resolves the drawbacks of each approach and makes a great improvement in the variety of recommendations in comparison to each individual approach. It introduce a new fuzzy clustering approach based on genetic algorithms and create a two-layer graph. After applying this clustering algorithm to both layers of the graph compute the similarity between web pages and users, and propose recommendations using the content-based, collaborative and hybrid approaches.

ShivaNadiet.al. [2011] proposed fuzzy recommender system based on collaborative behavior of ants (FARS). It works in two phases. First, user's behaviors are modeled offline and results are used in second phase for online recommendation.The performance is evaluated using log files. The results are promising and provides us with more functional and robust recommendations.

Subhash K. Shindeet.al.[2011]: proposed a novel modified fuzzy C- means clustering algorithm which is used for hybrid personalized recommender system. It works in two phases.In the first phase opinions from users are collected in form of user item rating matrix. In second phase recommendations are generated online for active users using similarity measures.

Kwoting Fanget.al.[2003]: proposes a recommender system with two approaches on user access behavior by fuzzy clustering method. It show similar preference to the given users and recommends what they have liked. The recall value was used to evaluate recommended accuracy with two algorithms (item-to-item) and (user-to-user).

3. PROBLEM DOMAIN

Collaborative Filtering systems cannot produce recommendations if there are no ratings available. They demonstrate poor accuracy when there is little data about user ratings. This problem is known as Cold-Start problem. Another problem is that Collaborative Filtering systems are not content aware meaning that information about items are not considered when they produce recommendations. Many of existing Collaborative Filtering systems work slowly on a huge amount of data. Several techniques such as clustering and parallelization were discovered to overcome the problem.

Content-based filtering system recommendations are limited in scope and require items and attributes must be machine-recognizable. It cannot filter items on some assessment of quality, style or viewpoint because of lack of consideration of other people's experience and also there is absence of personal recommendations.

In content-based filtering system there is no serendipitous items, i.e. Serendipity – is the ability of the system to give an item surprisingly interesting to a user, but not expected or possibly foreseen by the user. For example, if a book of the same author has been recommended, the user probably already knows about the book and, therefore, is not surprised. The recommendation would be serendipitous if the system recommends another book of another author and genre, but which seem unexpectedly interesting to the reader.

Content-based filtering system suffers from Synonymy. If there are two words spelled differently but having the same meaning – pure content-based filtering will recognize them as two independent words and will not find similarities among others.

Existing Method Problem: In Existing method the data is first clustered using the fuzzy c-means clustering technique and then this data is applied to the weighted similarity measure. Genetic algorithm is applied to find the optimal similarity measure and the similarity metrics to improve the quality of a RS. This method is based on pure collaborative model based approach which suffers from various problems of pure collaborative approach like cold start problems and increased computation time.

4. PROPOSED WORK

The proposed methodology is for improving the working and quality of the recommender system and also to remove the disadvantages of pure content based approach or pure collaborative filtering approach like limited scope of recommendations, cold start problems etc. We proposed Hybrid Approach which is combination of Contents and Collaborative Filtering, so that the approaches can be benefitted from each other.

To deal with the sparse data, model based collaborative filtering such as clustering techniques provide more accurate predictions. Proposed System also improves scalability.

The commercial recommender systems (CRS) employ memory- based methods, while model-based methods are usually associated with research recommender systems. To obtain the benefits of both memory based technique and model based techniques they both are also combined in the proposed method.

The existing methodology has been implemented on Movielens datasets only. Proposed System will be implemented on the Movielens, FilmAffinity and Netflix database.

To calculate the similarity between the different items in the given dataset in least time and efficiently and reduce computation time of the recommender system we can use cosine similarity. Both takes less execution time than other similarity measures like adjusted based similarity, correlation based similarity.

Proposed Algorithm divided into three algorithm:

Content-Based Filtering

For implementing a content-based filtering system following steps to be done:

- Terms can either be assigned automatically or manually. When terms are assigned automatically a method has to be chosen that can extract these terms from items.
- The terms have to be represented such that both the user profile and the items can be compared in a meaningful way.

A learning algorithm has to be chosen that is able to learn the user profile based on seen items and can make recommendations based on this user profile. Relevance feedback, genetic algorithms, neural networks, and the Bayesian classifier are among the learning techniques for learning a user profile.

Memory-Based or User Based Collaborative Filtering

For implementing Memory-Based or User Based Collaborative Filtering K-Nearest Neighbor Algorithm is used. The *k*-NN algorithm is used for estimating continuous variables. One such algorithm uses a weighted average of the *k* nearest neighbors, weighted by the inverse of their distance. This algorithm works as follows:

- Compute the Euclidean distance.
- Order the labeled examples by increasing distance.
- Find a heuristically optimal number *k* of nearest neighbors.
- Calculate an inverse distance weighted average with the *k*-nearest multivariate neighbors.

Model-Based or Item Based Collaborative Filtering

For implementing Model-Based or Item Based Collaborative Filtering Adjusted K-Means Algorithm is used. This algorithm works as follows:

- Input : The number of clusters k and items attribute features.
- Output: A set of k clusters that minimizes of the squared error criterion, and the probability of each item belonging to each cluster center are represented as a fuzzy set.

1. Arbitrarily choose k objects as the initial cluster centers;
2. Repeat (a) and (b) untilsm all change;
3. (Re) assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
4. Update the cluster means, ie. Calculated the mean value of the objects of each cluster.
5. Computing the possibility between objects and each cluster center.

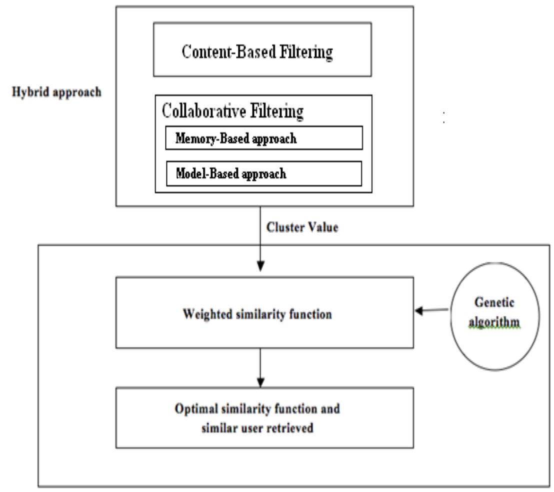

Figure 3: Hybrid Approach with Genetic Algorithm

6. RESULT:

We evaluate our proposed system on different parameters, which describe below:

- Recall
- Precision
- Accuracy
- Similarity
- Mean Absolute Error

Recall is the probability that a (randomly selected) relevant document is retrieved in a search. Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.

$$\text{recall} = \frac{|\{\text{relevant documentd}\} \cap \{\text{retrieved documentd}\}|}{|\{\text{relevant document}\}|} \ldots\ldots\ldots.(1)$$

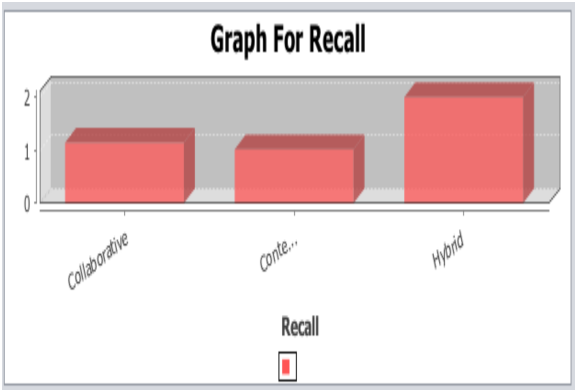

Figure 4: Recall graph for Collaborative, Content and Hybrid Approach

We calculate recall value for all algorithm collaborative, content base and hybrid. And results shown with help of diagram. We find that hybrid approach shows more true positive rate compare to collaborative and content base approaches.

precision

Precision is the probability that a (randomly selected) retrieved document is relevant. Precision is the fraction of retrieved documents that are relevant to the query:

$$precision = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

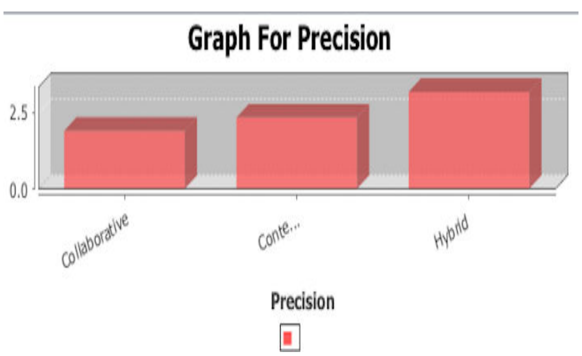

Figure 5: Precision graph for Collaborative, Content and Hybrid Approach

We calculate precision value for all algorithm collaborative ,content base and hybrid. And results shown with help of diagram. We find that hybrid approach shows more true positive rate compare to collaborative and content base approaches.

Accuracy

Accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined.

$$accuracy = \frac{number\ of\ true\ positives + number\ of\ true\ negatives}{number\ of\ true\ positives + false\ positives + false\ negatives + true\ negatives} \quad \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

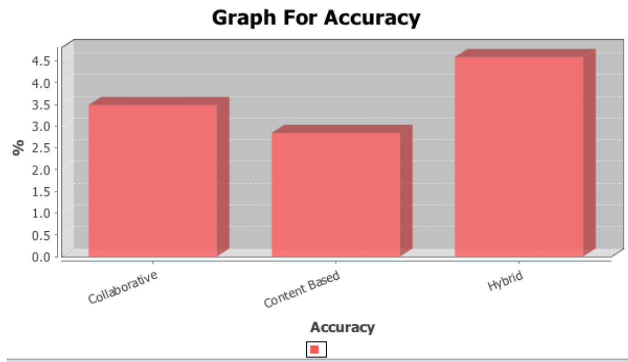

Figure 6: Graph of Accuracy for Collaborative, Content and Hybrid Approach

We calculate precision value for all algorithm collaborative, content base and hybrid. And results shown with help of diagram. We find that hybrid approach shows more true positive rate compare to collaborative and content base approaches.

Mean Absolute Error

The mean absolute error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. The mean absolute error is given by. As the name suggests, the mean absolute error is an average of the absolute errors, where is the prediction and the true value. The mean absolute error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. The mean absolute error is given by

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|fi - yi| = \frac{1}{n}\sum_{i=1}^{n}|ei|$$

As the name suggests, the mean absolute error is an average of the absolute errors $|e_i| = |f_i - y_i|$,

$$|ei| = |fi - yi|$$

Where

f_i is the prediction
y_i the true value.

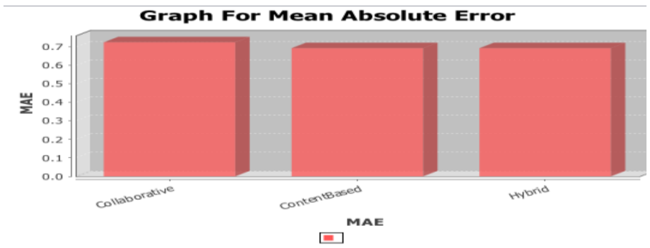

Figure 7: MAE graph for Collaborative, Content and Hybrid Approach

We calculate mean absolute value for all algorithm collaborative, content base and hybrid. And results shown with help of diagram. We find that hybrid approach shows lessvalue of meanabsolute error compare to collaborative and content base approaches.

Similarity

Cosine-based approach defines the cosine-similarity between two users x and y as:

$$\text{simil}(x,y) = \cos(\vec{x},\vec{y}) = \frac{\vec{x}\cdot\vec{y}}{||\vec{x}|| \times ||\vec{y}||} = \frac{\sum_{i\in I_{xy}} r_{x,i} r_{y,i}}{\sqrt{\sum_{i\in I_x} r_{x,i}^2}\sqrt{\sum_{i\in I_y} r_{y,i}^2}} \quad \text{……………..(8)}$$

The user based recommendation algorithm uses a similarity-based vector model to identify the k most similar users to an active user. After the k most similar users are found, their corresponding user-item matrices are aggregated to identify the set of items to be recommended.

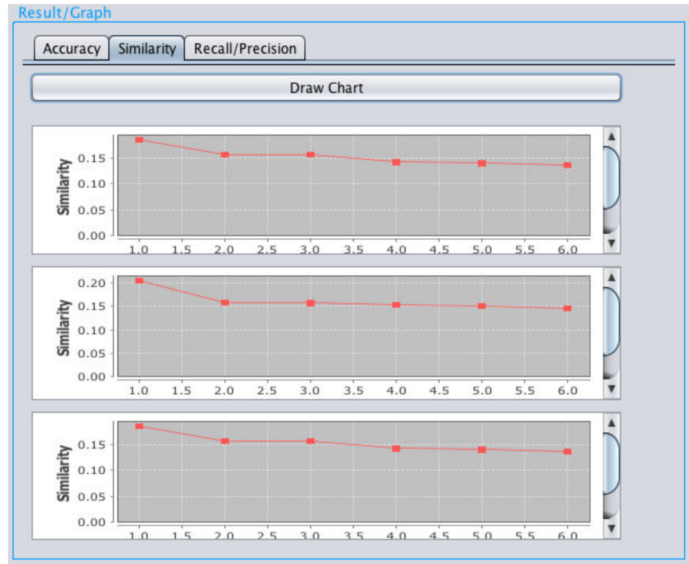

Figure 8: Cosine Similarity graph for Collaborative, Content and Hybrid Approach

We calculate cosine similarity between searched movie and recommended movies.

7. CONCLUSIONS:

Recommender systems are a powerful new technology for extracting additional value for a business from its user databases. These systems help users find items they want to buy from a business. Recommender systems benefit users by enabling them to find items they like. Conversely, they help the business by generating more sales. Recommender systems are rapidly becoming a crucial tool in E-commerce on the Web. Recommender systems are being stressed by the huge volume of user data in existing corporate databases, and will be stressed even more by the increasing volume of user data available on the Web. New technologies are needed that can dramatically improve the quality and scalability of recommender systems.

8. REFERENCES

[1] HirdeshShivhare, Anshul Gupta and Shalki Sharma (2015), “Recommender system using fuzzy c-means clustering and genetic algorithm based weighted similarity measure”, IEEE International Conference on Computer, Communication and Control.

[2]Antonio Hernando, Jesús Bobadilla, Fernando Ortega and Jorge Tejedor (2013), “Incorporating reliability measurements into the predictions of a recommender system”, International Journal of Information Sciences, vol. 218, pp. 1- 16.

[3] F. Ortega, J. Bobadilla, A. Hernando and A. Gutiérrez, (2013) “Incorporating group recommendations to recommender systems: Alternatives and performance”, International Journal of Information Processing and Management, vol. 49, issue 4, pp. 895- 901.

[4] Birtolo, C., Ronca, D., Armenise, R., &Ascione, M. (2011), “Personalized suggestions by means of collaborative filtering: A comparison of two different model-based techniques”, In NaBIC, IEEE (pp. 444–450).

[5] H Izakian, A Abraham , (2011) “Fuzzy Cmeans and fuzzy swarm for fuzzy clustering problem”, Expert Systems with Applications, Elsevier.

[6] Jesus Bobadilla, Fernando Ortega, Antonio Hernando and Javier Alcala (2011) “Improving collaborative filtering recommender system results and performance using genetic algorithms”, ELSEVIER Knowledge-Based Systems, Vol. 24, issue 8, pp. 1310–1316.

[7] Dr. P. Thambidurai and A. Kumar, (2010), “Collaborative Web Recommendation Systems -A Survey Approach”, in Global Journal of Computer Science and Technology Vol. 9 Issue 5 (Ver 2.0).

[8] Huang, C., & Yin, J. (2010). Effective association clusters filtering to cold-start recommendations. In 2010 Seventh int. conf. on fuzzy systems and knowledge discovery (FSKD), Vol. 5 (pp. 2461–2464).http://dx.doi.org/10.1109/ FSKD.2010.5569294.

[9] J. Bobadilla, F. Serradilla and J. Bernal (2010) “A new collaborative filtering metric that improves the behavior of recommender systems”, ELSEVIER Knowledge-Based Systems, Vol. 23, Issue 6, pp. 520-528.

[10] Jin-Min Yang, Kin Fun Li and Da-Fang Zhang (2009) “Recommendation based on rational inferences in collaborative filtering”, Knowledge-Based Systems, vol. 22, issue 1, pp. 105–114.

[11] J.L. Sanchez, F. Serradilla, E. Martinez, J. Bobadilla, (2008) “Choice of metrics used in collaborative filtering and their impact on recommender systems”, in: Proceedings of the IEEE International Conference on Digital Ecosystems and Technologies DEST, pp. 432–436.

[12] Kamal K. Bharadwaj and Mohammad Yahya H. Al-Shamri , (2008) “Fuzzy-genetic approach to recommender systems based on a novel hybrid user model”, Expert Systems with Applications.

[13] J. Ben Schafer, ShiladSen, Dan Frankowski, Jon Herlocker, (2007) “Collaborative filtering recommender systems” in In: The Adaptive Web, pp. 291–324. Springer Berlin / Heidelberg.

[14] E. Adomavicius, A. Tuzhilin, (2005) “Toward the next generation of recommender systems: a survey of the stateof- the-art and possible extensions”, IEEE Transactions on Knowledge and Data Engineering, vol. 17, issue 6, pp. 734–749.

[15] F. Kong, X. Sun, S. ye, (2005) “A comparison of several algorithms for collaborative filtering in startup stage”, in: Proceedings of the IEEE Networking, Sensing and Control, pp. 25–28.

[16] Joseph Konstan, George Karypis, BadrulSarwar, and John Riedl, (2001) “Item-Based Collaborative Filtering Recommendation Algorithms”, ACM.