Analysis and Prospect of Recommendation System on social media

Yilong Yuan *

University of California Irvine, Irvine, America

* Corresponding Author Email: z1034285762@gmail.com

Abstract. Recommendation system works surprisingly well on providing interesting and informative contents to individual user by matching suitable information with target users, and it tremendously reduces the time users spend for searching relative contents while it is also cutting back the workload of content providers. There are two popular recommendation techniques: content-based and collaborative based filtering. The abilities of each technique is compared and analyzed, and how well they match the need of users on the social media platform. In addition, a typical form of hybrid recommendation that has the advantages of multiple techniques is also taken in the comparison. From the perspective of users in social media, each specific technique achieves some degree of success for the overall user satisfaction. Integrating techniques is a valuable attempt. This paper presents a look the origin and the development of recommender system in the past. The evolution of recommendation provides insight regarding what the future of recommender will become. The architecture of recommender system is also reviewed for understanding the way system performs. The key component inside the recommender architecture including matching, filtering, and objects utilized by technique which are critical factors in determination of prediction are examined. Different techniques used in filtering are analyzed for recognizing the advantages and the disadvantages for the purpose of social media. After all, the effectiveness of the hybrid recommendation technique on dealing with limitations of popular filtering is demonstrated.

Keywords: Recommendation system, social media, integrating techniques.

1. Introduction

With the rapid development of world wide web and technology, the software development is evolving, and several turning points have been reached from straightforward email application to complex video streaming application with personalized recommendation feature built in. The need of people also evolved, and people wish to be interconnected with others in the vivid online world. Applications which are relative to social media are in the center stage of everyone's daily routine, and the recommendation algorithm is the key role in every social media application in order to guarantee the favorable user experience. People unconsciously expect the contents presented to them are highly personalized since the recommendation algorithm used by many mainstream social media applications made some considerable effort to keep users feeling interactive so that users are willing to spend more time and money in certain software [1]. The current recommendation algorithm seems to be very successful as more users were satisfied in general. However, today's recommendation algorithm is not in an ideal form. A major problem is that the recommendation algorithm in this time is still not highly flexible for each individual user, instead the current stage of recommendation mainly ensures the content generated for users would not go wrong on a large scale.

The value of recommendation is mutual, and it benefits both users and service providers. From users' perspective, personalized contents mean less noise in the desired information and less distraction when users are well served by interesting contents. From service providers' perspective, a good recommendation algorithm reduces the cost of providing content when information overload can be avoided and increases the profit by increasing users' dwell time. One major issue existed between seller and buyer or user and service provider: waste of resource and time. Many unpredictable and useless attempts and the time required for those attempts were wasted. Service providers were taking extra effort to provide the excess of information which may not be useful to users. Data feedback collected from users is also extremely difficult to handle and to convert it to

make proper decision. Users also spent extra time to browse over information to find some insights they did not know but needed.

With the assistance of modern recommendation algorithm, big data can be handled by recommendation algorithm, and users are more likely to see information which are more relevant to them. It is a modern way for service providers to effectively understand what users are expecting [2]. The waste of resource and time is greatly reduced but not eliminated. There are two limitations of current recommendation algorithm: inability to fully understand the input information from human beings and uncertainty in handling input information to produce a desired output decision. Human beings are not perfect, and some inaccurate data often produced to feed the algorithm. Under such a circumstance, recommendation algorithm could have a wrong assumption to user. On the other hand, the algorithm itself prefers stability over accuracy. Almost all users in short video platform will be recommended with the most popular topic in the period. Nevertheless, it is obviously not the best personalization experience for everyone. The preference of stability also reflects in other decisions produced by recommendation algorithm including recommending everything relative based on user's search query. If a user does a searching about latest high-end phone and purchase it, and algorithm keeps suggesting and pushing similar high-end phones to the user, it does not make a good sense in this case since most people only need one or two phones. Another common drawback is that the algorithm discreetly assumes and predicts what users could be aroused by. Male users may often be associated with super car and sport, and female users are often tied with makeup tutorial and dressing guide.

This paper reviews the timeline of the evolution of recommendation algorithm including its origin and the evolution, different types of recommendation system, and the performance of different recommendation algorithms. This paper also looks forward to the prospect of recommendation algorithm and some concerns worthy of the attention on the area of social media.

1.1. Timeline of recommendation algorithm

The happening of recommender system is not a coincidence. At different times, the system itself aims to solve specific issues, so it evolved.

1.2. Origin of the Recommendation System

The idea of recommender system is in an understandable form long ago. There are traces in our world indicating that specific patterns or orders exist in everywhere. The incentive for doing something and the result caused by the incentive are a perfect pair. It is a manner of doing things as responses on something else. People are facing lots of choices all the time. The right car to buy, the food to eat, and the book people want to read for the new year are all considered as options. The amount of information on the internet today could be overwhelming. A search query about keyboards can results in hundreds of pages of products. When it came to the beginning of the recommender system, the central idea is to help people find the most suitable and desirable things. In 1979, the first model of the recommender system is called the Grundy system which uses stereotypes as the main technique to build individual model for each user with limited amount of input data [3]. With this simple but rational logic, the Grundy system can recommend books to different users based on their behaviors [4]. The algorithm is not a perfect representation of a real-world scenario, but it is still the basic of recommendation. Later in 1992, soon after the invention of the stereotypes-based recommendation system, a more advanced recommender system named Tapestry presented a crucial concept: defining and understanding the similarity of users numerically with the collaboration of information from items themselves and reaction of other users [5]. Any user is able to retrieve some specific documents based on the content of the documents and feedback from various users. This concept is also the beginning of the rank filtering and collaborative filtering that is still used in many applications people used nowadays. In the same year, a recommender system, GroupLens [6], that can match users with potential valuable information is already possible. It is one of the first automated

recommendation system. For the first time, the algorithm itself can predict user preference by studying activities generated from users and utilizing machine learning technique.

1.3. Evolution of the Recommendation Algorithm

The amount of information people is dealing with is growing out of control. The number of new brands, new documents, and even the new domain sites is increasing faster than ever before. This is why the recommender system had come an independent research area and topic around 1970s [7]. Lots of work were done for the development of recommender system, and there are many popular social platforms are taking the advantages of recommender system to provide smooth service such as amazon, instagram, YouTube. Throughout the time, the recommendation system evolved to better serve people's need, and three tiers of recommender system is an indication of the evolution [8]. Since the origin of the recommender system, the majority of algorithm is relative to content based and collaborative based filtering. The decision made is determined based on information itself and users' perspectives, and this approach is considered as the first tier of recommendation system. The first tier has limitations and cannot fully predict the need of people since the information for the algorithm is also limited. Then, it comes to the second tier of recommendation system, which includes the social information as additional input to help algorithm to make decisions. The includes but not limited to user's information and everything the algorithm can learn from exterior conditions. A user portfolio is also created for individual user in the second-tier recommendation system, and the system is responsible of keep track of the information to update user portfolio. This makes discovering similar users much easier, and one real world scenario is about the potential known friends' people can find in the Instagram or Facebook. In the third-tier recommendation system, the system is made to understand more critical factors and some intangible factors like the geographical location, emotional change, and subtle change of user's perspective. In a short word, the recommender algorithm is powerful with proper and rich data input. From the origin to the system nowadays, the algorithm gets richer data and users activity history, the algorithm itself does not only generate few recommendations, but also creates extraordinary and diverse contents for users to browse and enjoy. The recommendation system becomes more powerful and results in excellent user satisfaction with the help of information retrieval and artificial intelligence [7]. Recommendation Algorithm is extremely flexible today with the option of many input data [8]. Depending on the thing people expect the algorithm to predict and make decisions, different input data can be chosen to satisfied the need.

2. Architecture of recommendation system

Even though plenty of recommendation systems are used in all different platform, the architecture of the recommendation system is roughly the same with some minor differences. Shopping websites like Tabao and Amazon make the use of recommendation system to recommend products to customers, and most of the revenue comes from the success of a good recommendation system since users are likely to make a purchase. The architecture of shopping platform recommendation system could be a representative example of all the recommenders. The recommendation system consists of multiple layers. As the figure 1 shows, the bottom layer of the recommendation is underlying basic data layer. This layer includes valuable data input such as user information, product information, history of user behavior. This is the most basic portion of a recommendation system which should be considered as a requirement. The information in this layer is foundation for further data process and analysis [9]. The next layer is processing layer, which is responsible for handling and pre-processing abundant data. During the process, the layer can recognize the focus of each kind of data. For instance, the gender of the customer is an important feature, and this feature may match some characteristic of the items. In term of matching, it is the work for the next layer named modeling layer. With the features gathered from the processing layer, the recommendation system can start the prediction. Three major modules work together to achieve the goal: matching, ranking, and strategy [9]. Multiple algorithms can run in the matching module, filtering technique can be implemented in this area, and

there are several of them, each has specific advantages and drawbacks. Then, the algorithm can match items with customers based on the history and features. The second module is ranking. Generally, each potential prediction will be given a score in a range. An example of the range can be between zero and one. The item with a higher score will be more suitable for the specific user. The third module is strategy. This is the last gate before pushing the prediction to users, and it is a stage to polish and improve the result from previous module. Some strategies like A/B testing can indicate what may be a better recommendation. The top layer is the service layer where the system can decide what recommendation to push. In this case, it could be an item, a person, or an advertisement. Obviously, this is not the only architecture of recommendation system. Some can differ in the organization of the layers, and others could have a different scope. This is mainly for getting a sense of what is inside the recommendation system, and it will be easier assume what layer positively or negatively affect the performance of the recommendation system later.

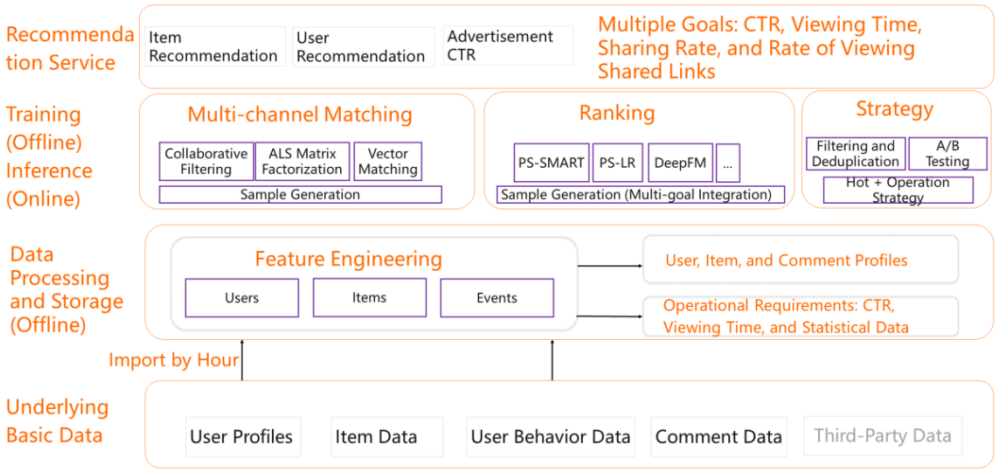


Fig. 1 Architecture of the recommendation

3. Type of recommendation filtering

The recommendation system uses filtering technique to provide high quality information from a large collection of data depending on the activity history of users, which is to be seem as a classification task [10]. The filtering system can handle tons of data by building a user model, unordered objects are classified into good type, bad type, and intermediate type. Good type represents the things that best matches the user preference. In the opposite, it is the bad type. Intermediate type can be seemed as something that is the second or third tier of the ranking result, and it can be considered to push to user as a recommendation in some cases. The performance of the recommendation system can vary dynamically with different filtering techniques, and it means to serve various purposes. There are two most popular filtering techniques: content-based filtering and collaborative-based filtering.

3.1. Content-Based Filtering

Content-based filtering technique applies features of objects or products to recommend similar items to what users like [11]. In the figure 2, the model knows what websites the user has visited. Each website is given an attribute. Features in those websites are used to find similar websites and continue recommend these websites to users. As long as users keep providing more data for the model by generating behaviors, the recommender system can recommend better predictions. On the other hand, the system will find it difficult to recommend anything without enough data for training. It could merely recommend some popular topics [12]. For all things in the figure to work properly, the model is built and trained first with the training data. The training data here is websites the user already read and websites the user added to their favorite list. Then, the content-based filtering system can determine what should be relevant or not with things it learns while user profile is frequently

compared with incoming items. Finally, the recommender system will make decisions by matching user preference and websites contents [10]. A positive circle is formed.

3.2. Collaborative-Based Filtering

Collaborative-based filtering somewhere in the opposite of the content-based filtering. It requires the data about user preference and user history interaction with objects instead of the feature of content. In other word, the system filtering uses the similarity between users and items for recommendation [13]. Users can get more diverse and serendipitous recommendations because one user who has lots of similarities with another user may be recommended with something another user likes. There are various strategies to get user interaction. The algorithm can either let users explicitly indicates what kind of items they like, or it can implicitly collect what items users interact with reference [13]. There are two kinds of collaborative-based filtering: User-based collaborative filtering and item-based collaborative filtering.

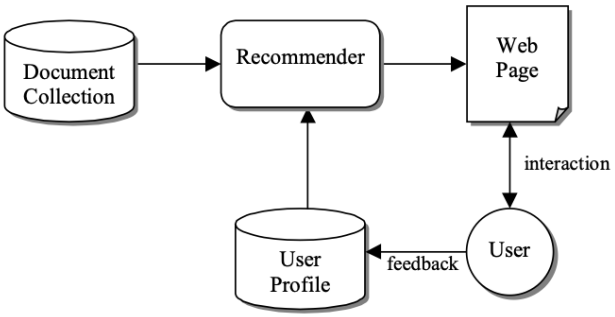


Fig. 2 Content-based logic

User-Based Filtering: User-based filtering rates items by first looking at the rating of the neighboring users [12]. The preference of neighboring users can greatly affect the decisions of recommendation system [14]. For instance, if a user liked a post about cat, and the system is about to recommend something else to this user, this filtering technique will first check some similar neighboring users and their preferences [15]. If there is another user who also liked the post about cat, and in addition, this user also like the post of a panda. The original user may also be recommend about panda in this case. Figure 3 is a great representation of how does user-based filtering work. In this fixed size example, the algorithm will find the top k most similar users, and makes use of their interaction history as the basic for recommendation, where the k is determined by the algorithm itself. In this example, user one liked all items: sun, thunder, moon, and cloud. The user two liked the thunder, and the user three liked the thunder and the moon. The algorithm will then consider user one and three as the most similar users. The recommender system finally decided to recommend something else liked by user one to user three: the sun and the cloud.

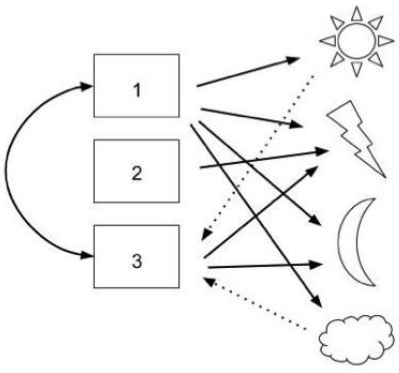


Fig. 3 User-based example

Item-Based Filtering: Item-based filtering is more popular in the shopping department of the social media, and it is first introduced by Amazon in 1998. Instead of looking at the similarity of the users,

the item-based filtering cares more about the similarity of the items. The rating of each item is then performed by taking users' rating on similar items. This filtering technique works great if the number of item is larger than the number of user. If a laptop was bought for ten times, when a phone and a desktop are bought for eight times and two times. The algorithm can conclude laptop and phone are similar items. Then, users who have not yet purchased a phone will be recommend with a phone. In the figure 4, the algorithm will pick several most similar items based on assumption. The moon was liked three times, the sun was liked two times, the cloud was liked one time, and nobody likes the thunder. The algorithm will pair the moon with the sun. The user will get the recommendation of the sun if they have not seem it.

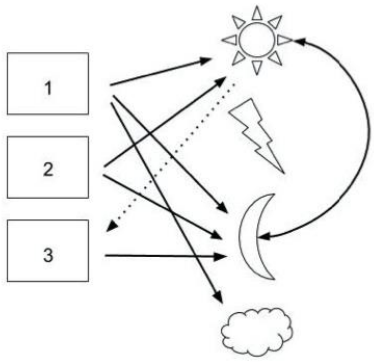


Fig. 4 Item-based example

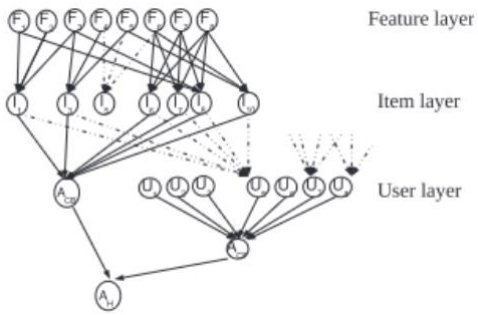


Fig. 5 Hybrid model

3.3. Hybrid Recommendation

Content-based and collaborative-based both works well in their ways, but limitations exist in both of them. Hybrid recommendation algorithm combines the content-based and collaborative-based filtering, and hybrid model with item instantiation outperforms the prediction result from content-based or collaborative-based filtering. In the figure 5, a top-down structure contains three layers. Feature layer has the feature of users and items as node. Item layer contains items node with corresponding value which indicates the relevance. User layer contains user nodes which are for rating of items. In content-based potion, user information is not in consideration so that there is no user node. The final prediction is pointed by two potions. The left one without user node is content-based part, and the right one is collaborative-based part. They both work with each other to get result. Finally, with everything together as a whole is the hybrid recommendation.

4. Algorithm performance

The standard evaluation of performance of recommendation algorithm can vary because there are many recommender systems that serve different purpose. In the field of social media, the purpose of the recommender system is to provide the most relevant contents and to predict accurate results based on users' interests, which are also the purpose of the most recommender system. There are four basic criteria to evaluate the performance of the recommendation system: coverage, diversity, novelty and

serendipity [8]. In the real world, a good recommendation system means good user satisfaction and considerable revenue. The four criteria precisely represent the algorithm performance in the real world. Users expect similar contents and something new but interesting, and more transaction created. Each filtering technique mentioned in the previous section does have advantage and disadvantage in some of the criteria.

In content-based filtering, the focus of the algorithm is about the current user. The algorithm can ignore other users which makes this technique extremely efficient so that this is scalable. One more important idea of the content-based filtering is that the algorithm can recommend unique recommendations that most of the other users may not know. This is relatively a good work in the aspect of novelty, and it is a fine work in coverage with limited percentage of possible recommendations. Each model of content-based filtering is only responsible for one user, and the model only looks for items that are similar to the interest of user. This is the reason it can provide recommendation for specific needs, but it also results in a serious issue: lack in diversity and serendipity. If a user never interacts with something new, the model cannot recommend it. This is a bad idea in the social media platform. Most users are curious about new things such as some new contents they never know, and the content-based filtering fails to expand the interest of users. Besides, for newly created user model, content-based filtering also fails to provide good user experience since there is no user interaction history when an account is created.

Hence, collaborative-based filtering may be a better option for recommender algorithm on social media. In term of expanding the interest of the user, this is one primary advantage of collaborative-based filtering by taking similar users into the consideration. This model can always help user find new interests, which is a match with serendipity. In a real-world scenario, one specific user can discover something new liked by friends or other users, and it seems like a must feature to let users connect with each other's. However, collaborative-based filtering has a major limitation on the scalability [14]. The process of getting user information and interests takes lots of time. This is especially noticeable in the user-based collaborative filtering. With the growing number of the user and the item, the calculation will also grow. Another serious issue is relative to the accuracy of the prediction: sparsity. In user-based collaborative filtering, it is required to look for neighboring users and then make a decision. The algorithm may fail to get neighboring users for the particular user which will then result in the loss in prediction accuracy.

Item-based collaborative filtering could be another option for the social media platform. Instead of focusing on the information of users, items themselves are taken into consideration. This means faster computation. Item-based collaborative filtering isolates the neighboring generation, and a model that cares about mostly static items instead of frequently changed users [15]. The computation can be pre-process while there is less computation but preciser prediction. One minor issue is that item-based filtering sometimes recommends boring items depending on the previous user experience. Furthermore, sites that has numerous users can operate in a large scale with the issue of scalability. Comparing with the user-based collaborative filtering, item-based is for suitable for real time calculation on the social media, and it is fast in responding and good enough for diverse content suggestion. Overall, item-based could be a better idea while there are always more items than users on social media.

Hybrid recommendation which integrates content-based and collaborative-based filtering could be an ultimate answer for the social media. With the advantage from both filtering, the algorithm can cover very unique recommendation as well as serendipitous contents. This is the way to ensure the coverage, diversity, novelty and serendipity in all cases.

5. Conclusion

Recommendation system is indispensable in the social media software. It provides the most personalized user experience. The recommendation filtering techniques inside the recommender system is the core element that can dynamically affect the performance of the algorithm. In the

environment of the social media platform, the criteria for evaluating the algorithm performance can be very complicated. Item-based collaborative filtering is better than user-based collaborative filtering for social media purposes. It has a guarantee on the availability of precision while there are not many neighboring users and the features. One major problem mentioned above is about the lack of personalization in the recommendation decisions. This is because most of the recommendation systems used in the mainstream social media platform prefer the collaborative-based filtering, which will lead to recommendations that are rich in diversity and serendipity but not unique for each specific user. The content-based filtering is the answer for extreme flexible recommendation, while it is not capable of providing diverse contents from similar users. It comes to a conclusion that hybrid recommendation that combines the advantages of multiple filtering is the most suitable filtering technique to cope with the complex need of users. The hybrid recommendation performs well, and there are still many rooms for it to evolve in many aspects. The hybrid recommendation system is not limited by the number of filtering it can contain. The future study should focus on how to integrate different filtering so that the advantage of each algorithm can be showed. The future study may also take the scalability as a topic. Expanding collaborative filtering to larger scale, and this technique will be operated in more important industrial project. Some alternative calculation methods are potentials for time-saving purposes. Different strategies for integrating filtering technique and what filtering to use in the hybrid recommendation system still need more researching to ensure user satisfaction in the field of social media, and better hybrid recommendation system will be happened.

References

[1] M. B Dias, D. Locher, L. Ming, "The value of personalised recommender systems to e-business," Proceedings of the 2008 ACM Conference on Recommender Systems. RecSys, Lausanne, Switzerland, pp. 23-25, October 2008.

[2] A. Rasquin, "History and Definition of the Rome Criteria," Humana Press, 2013.

[3] F. M. Harper, S. X. Li, C. Yan, "An Economic Model of User Rating in an Online Recommender System," International Conference on User Modeling. Springer-Verlag, June 2005.

[4] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Computer Science. Lomdon, May 2014.

[5] G. Adomavicius, A. Tuzhilin, "Towards the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," IEEE. London, vol. 45, pp. 121-129, March 2017.

[6] J. A. Konstan, J. Riedl, A. Borchers, "Recommender Systems: A GroupLens Perspective," IEEE. London, vol. 67, pp. 101-112, July 1998.

[7] R. Sharma, R. Singh, "Evolution of Recommender Systems from Ancient Times to Modern Era: A Survey," Indian Journal of Science and Technology. New York, vol. 9, pp. 601-614, May 2016.

[8] B. Sinha, R. Dhanalakshmi, "Evolution of recommender paradigm optimization over time," Journal of King Saud University - Computer and Information Sciences. London, vol. 34, pp. 1047-1059, April 2022.

[9] V. Bocci, P. Booth, M. Bozzo, "Basic concepts and architectural details of the Delphi trigger system," IEEE Transactions on Nuclear Science. London, vol. 42, pp. 837-843, March 1995.

[10] R. V. Meteren, "Using Content-Based Filtering for Recommendation," IEEE Transactions on Nuclear Science. London, vol. 72, pp. 817-823, March 2000.

[11] J. Kim, K. Kim, K. H. You, "An Approach for Music Recommendation Using Content-based Analysis and Collaborative Filtering," International Journal on Information. Annapolis, vol. 15, pp. 1985-1996, June 2012.

[12] J. Liu, J. Song, C. Li, "A Hybrid News Recommendation Algorithm Based On K-means Clustering and Collaborative Filtering," Journal of Physics: Conference Series. Alaska, vol. 3, pp. 032050, April 2021.

[13] Z. D. Zhao, M. S. Shang, "User-Based Collaborative-Filtering Recommendation Algorithms on Hadoop," Third International Conference on Knowledge Discovery and Data Mining, WKDD 2010, Phuket, Thailand, 9-10 January 2010.

[14] B. Sarwar, “Item-Based Collaborative Filtering Recommendation Algorithms,” Proc. the 10th International World Wide Web Conference (WWW10), Hong Kong, May 2001.

[15] L. Jiang, L. Liu, J. Yao, “A hybrid recommendation model in social media based on deep emotion analysis and multi-source view fusion,” Journal of Cloud Computing. New York, vol. 9, pp. 521-518, May 2020.