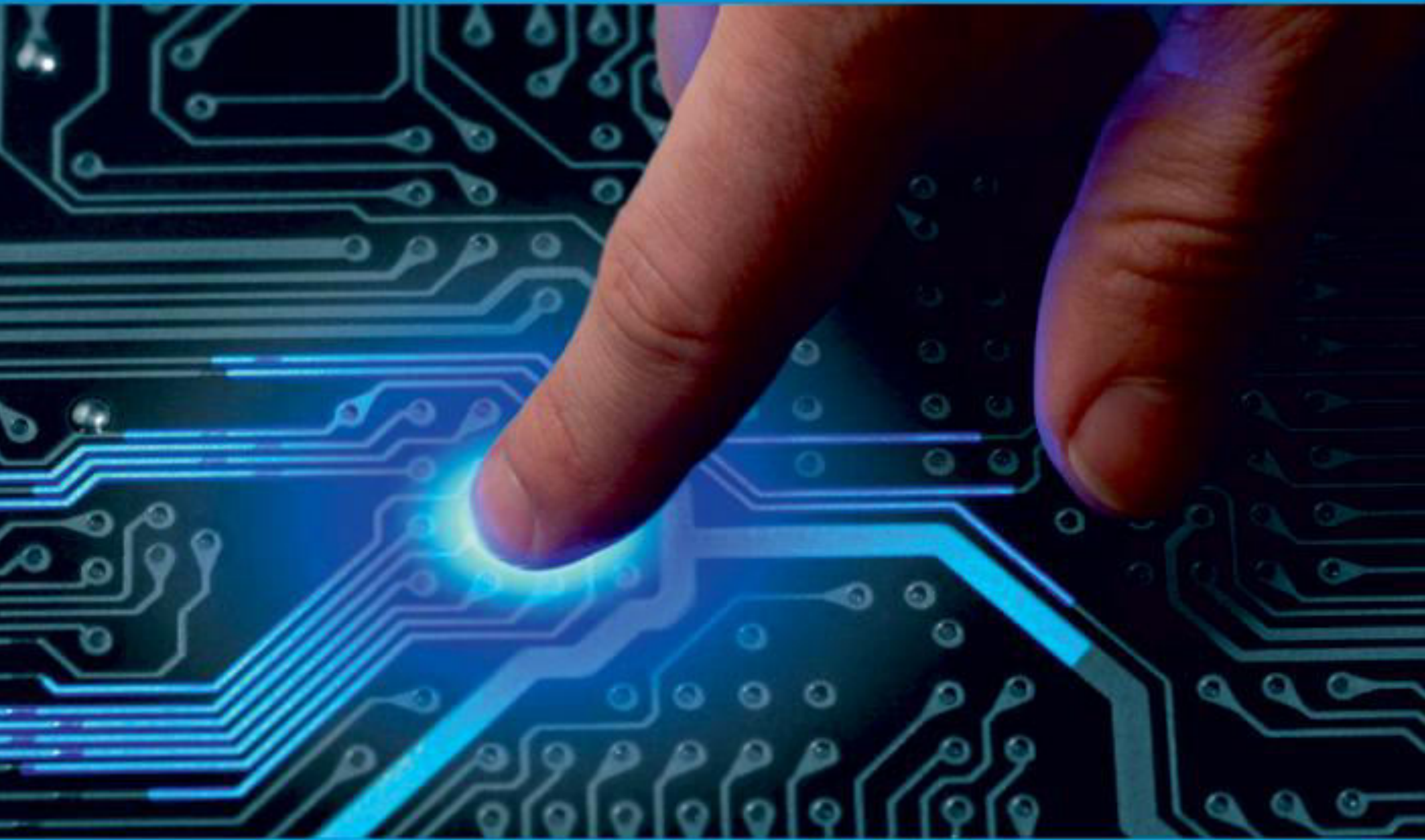
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

Online Social Voting with Collaborative Data Mining - A Survey

Namrata M. Dhandar, A. P. Kankale

PG Student, Dept. of CSE, RSCE, Buldana, Sant Gadge Baba Amravati University, Maharashtra, India

H.O.D., Dept. of CSE, RSCE, Buldana, Sant Gadge Baba Amravati University, Maharashtra, India

ABSTRACT: Social voting is a comparatively recent function in online social networks. It has its own set of challenges as well as the opportunity for referral. This overview describes several researches done on online social voting with the functions of collaborative data mining. This survey shows that, in recent years there has been a lot of development in the same area. Several techniques such as matrix factorization (MF), recommender systems (RSs), latent factor, etc. are taken into consideration for improvement.

KEYWORDS: Recommender system (RS), Collaborative Filtering (CF), latent features, matrix factorization, cold starts, social voting.

I. INTRODUCTION

Online social networks (OSN), such as Facebook and Twitter, make it simple for friends to share information. A user can not only exchange text, photo, and video updates with her immediate friends, but may also instantly spread such updates to a much broader audience of indirect friends, thanks to popular OSNs' extensive connection and global reach. Many OSNs now provide a social voting feature that allows users to give their opinions, such as likes or dislikes, on a variety of topics, such as user statuses, profile images, games played, products purchased, websites visited, and so on. Taking the like–dislike voting model a step further, certain OSNs, such as Sina Weibo, allow users to create their own voting campaigns on any issue of their choosing, with user-defined vote alternatives. A vote initiator's friends might join in the campaign or re-tweet it to their friends. Social voting provides several potential commercial benefits in addition to stimulating social relationships. Advertisers can use voting to promote specific brands. To perform market research, product managers might commence voting. Voting can be carefully launched by e-commerce operators to entice more online shoppers.

The growing popularity of social voting immediately raises the issue of "information overload": a user might quickly get overwhelmed by several votings launched, participated in, or re-tweeted by her direct and indirect friends. To improve user experience and enhance user involvement in social voting, it is vital and tough to present the "right voting's" to the "right users". Recommender systems (RSs) help consumers deal with information overload by proposing topics that might be of interest to them. Unlike conventional goods for recommendation, such as books and movies, social voting is spreading through social media. A user is more likely to be exposed to a voting if her friends started it, participated in it, or re-tweeted it. A user's voting visibility is closely associated with her social neighborhood's voting behaviors. Social impact becomes more significant as a result of social propagation: a user is more likely to vote if her friends have already voted. A user's voting habit is closely associated with her social friends due to social propagation and social impact.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907096 |

II. LITERATURE SURVEY

A. *Toward The Next Generation Of Recommender Systems: A Survey Of The State-Of The- Art And Possible Extensions* [2]

This paper provides an overview of the area of recommender systems and discusses the current generation of recommendation methods, which can be divided into three categories: content-based, collaborative, and hybrid approaches. The limits of current recommendation systems are also discussed in this paper and it also examines potential additions that could increase recommendation capabilities and make recommender systems more relevant to a wider range of applications. These enhancements include, among other things, a better understanding of users and objects, the introduction of contextual information into the recommendation process, support for multi-criteria ratings, and a more flexible ad platform. Over the last decade, several content-based, collaborative, and hybrid methods have been proposed, and several "industrial-strength" systems have been constructed, recommender systems have made substantial progress. Despite these advances, the current generation of recommender systems evaluated in this study still requires improvement in order for recommendation methods to be more successful in the real world.

This paper has examined the limitations of present recommendation algorithms and potential improvements that could improve recommendation capabilities. These enhancements include, among other things, improved user and item modeling, the incorporation of contextual information into the recommendation process, support for multi-criteria ratings, and a more flexible and less intrusive user interface.

B. *A Survey on Recommender Systems based on Collaborative Filtering Technique* [3].

Nowadays, product advertising and viewer preference are two crucial aspects of marketing. These two components combine to form a system known as the Recommender system. In internet technology, a recommender system is critical for data collection and ranking. Four types of filtering techniques are used in the Recommender System: demographic, content, collaborative, and hybrid. Collaborative filtering is the most common and commonly utilized technique. In this study, the previous three strategies are briefly discuss, but the primary focus is on collaborative filtering, its types, and important issues such as the cold start problem, sparse data, scalability, and accuracy.

They looked at a number of research publications, including some from the top, and discovered that Collaborative Filtering is the most often used filtering approach, but it also has certain drawbacks such as sparseness, accuracy, and scalability. Numerous studies have been conducted, and many writers have presented their findings. Scalability, Cold Start, Sparseness, and Accuracy are all priorities. However, there hasn't been much study done on the sparsity issue. Since today's internet data is growing at an exponential rate, sparsity is also expanding as new recordings, items, things, music, data, and so on are added and loaded on a daily basis. We will investigate the sparsity issue in future work because it is a significant obstacle that recommender systems confront today and in the future.

C. *Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model* [4]

Recommender systems give users with individualized product or service recommendations. These systems frequently employ Collaborating Filtering (CF), which analyses previous transactions to build links between users and products. The two most successful approaches to CF are latent factor models, which directly profile both users and products, and neighborhood models, which assess both people and items. This work adds some new elements to both techniques. The factor and neighborhood models can now be seamlessly blended, resulting in a more precise combined model. Extending the models to take advantage of both explicit and implicit inputs from users improves accuracy even more. The methods are put to the test using Netflix data. The findings are superior to those previously published on the same dataset. In addition, the work has proposed a novel assessment metric based on the performance of algorithms on a top-K recommendation job, which shows the disparities between them.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907096 |

This research offered enhancements to two of the most widely used Collaborative Filtering techniques. First, a new neighborhood-based approach based on formally maximizing a global cost function, unlike earlier neighborhood techniques. This improves prediction accuracy while keeping the benefits of the neighborhood technique, such as predictability and capacity to forecast. Second, modifications to SVD-based latent factor models that incorporate implicit feedback into the model allow for enhanced accuracy. One of the models also has features that are typically associated with neighborhood models, such as the capacity to explain recommendations and easily manage new users. Furthermore, the novel neighborhood model allows us to develop an integrated model that integrates the neighborhood and latent factor models for the first time. Because the neighborhood and latent factor models address data at various levels and complement each other, this is useful for increasing system performance.

The accuracy, diversity, ability to surprise with unexpected suggestions, explainability, appropriate top-K recommendations, and computational efficiency are all factors that influence the quality of a recommender system. Some of the characteristics, such as accuracy and efficiency, are reasonably easy to quantify and were addressed in this study. Other factors are more mysterious and difficult to define. The study proposed a novel method for evaluating the success of a top-K recommender, which is common in systems where only a few products should be recommended to each user.

D. *Temporal Collaborative Filtering with Bayesian Probabilistic Tensor Factorization* [5]

Traditional collaborative filtering techniques often rely on the premise that real-world relational data is stationary. The authors offer a factor-based approach that can account for time in response to our sales forecast challenge. This problem is formalized as a tensor factorization with a particular constraint on the time dimension by introducing additional time elements. The authors also offer a fully Bayesian approach to avoid tuning parameters and achieve automatic model complexity control. They create an effective sampling process capable of evaluating large-scale data sets in order to learn the model. The Bayesian Probabilistic Tensor Factorization (BPTF) algorithm is being tested on a variety of real-world issues, including sales forecasting and movie recommendation. The advantage of our temporal model is demonstrated by empirical results. For modeling changing relational data, we introduce the Bayesian Probabilistic Tensor Factorization technique. BPTF learns the global development of latent features by introducing a set of new time characteristics to classic factor-based collaborative filtering algorithms and imposing a smoothness restriction on those factors.

Automatic model averaging is achieved via an efficient MCMC approach that significantly removes the need for parameter adjustment on large-scale data. To demonstrate the superiority of temporal models over static models, they offer substantial empirical results on many real-world data sets. Instead of Gaussian models, we may utilize other types of observational models in the future, such as exponential family distributions. A Poisson model, for example, would be better suited to our sales challenge. This, however, may result in more intricate posterior distributions for which Gibbs sampling is ineffective.

E. *A Guide to Singular Value Decomposition for Collaborative Filtering* [6]

As the electronic commerce sector expands at a breakneck pace, it's more critical than ever to deliver tailored recommendations for different types of customers. Collaborative filtering is a useful tool for modeling and analyzing client preferences and making appropriate recommendations. One of the most often used methods for collaborative filtering is Singular Value Decomposition (SVD). Direct application of traditional SVD algorithms to collaborative filtering, on the other hand, may result in poor performance. They highlighted challenges that novices may encounter in this report, as well as useful SVD versions for collaborative filtering.

For collaborative filtering, a judicious use of Singular Value Decomposition is effective. Batch learning, also known as incomplete incremental learning requires a lower learning rate and has a more unstable performance than complete incremental learning, according to their findings. For collaborative filtering with millions of training instances, we find that incremental learning, particularly totally incremental learning, which adjusts values after looking at a single training score, is the best option.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907096 |

F. *Probabilistic Matrix Factorization* [7]

Many present approaches to collaborative filtering are incapable of dealing with huge datasets or users with few ratings. The Probabilistic Matrix Factorization (PMF) model is presented in this study, which grows linearly with the number of observations and, more crucially, performs well on the enormous, sparse, and highly imbalanced Netflix dataset. The authors expand the PMF model by including an adaptive prior on the model parameters and demonstrating how the model capacity can be automatically managed. Finally, they present a limited version of the PMF model based on the assumption that users who have rated similar sets of films have comparable preferences. For users with little ratings, the resulting model generalizes far better. They attain an error rate of 0.8861 when the forecasts of several PMF models are linearly merged with the predictions of Restricted Boltzmann Machines models, which is about 7% better than Netflix's own approach.

The probabilistic matrix factorization (PMF) and its two variations, PMF with a learnable prior and constrained PMF, were presented in this study. The researchers also showed that these models can be efficiently trained and deployed to a large dataset of over 100 million movie ratings. We would use hyper priors over the hyper parameters and MCMC methods to accomplish inference if we took a completely Bayesian approach. While a fully Bayesian treatment of the proposed PMF models would be computationally more expensive, preliminary results strongly imply that it would result in a large gain in predicted accuracy.

G. *Multi-Domain Collaborative Filtering* [1]

Collaborative filtering is a powerful recommendation method that predicts a user's choice for an item based on the preferences of other users with similar interests. The data sparsity problem, which generally arises because each user normally only rates a few items and so the rating matrix is highly sparse, is a significant obstacle when employing collaborative filtering methods. This problem is addressed in this study by addressing numerous collaborative filtering tasks in various domains at the same time and exploiting domain linkages. A multi-domain collaborative filtering (MCF) problem is what we call it. To solve the MCF problem, we present a probabilistic framework that models the rating problem in each domain using probabilistic matrix factorization and allows information to be adaptively transferred across domains by learning the association between domains automatically. We also provide the connection function to correct biases in distinct domains. When compared to certain exemplary methodologies, experiments done on various real-world applications indicate the efficiency of our solutions.

The multi-domain collaborative filtering issue, in which numerous rating prediction problems are collaboratively learned, is treated in this study. When all rating data is combined, it presents a probabilistic model that takes into account the association between different areas. Experiments on a variety of recommendation datasets show that our methods are effective. Active learning is another technique to deal with the data sparsity problem in CF. Unlike many traditional machine learning algorithms, which wait passively for labeled data to be delivered before beginning the learning process, active learning takes a more proactive approach by selecting unlabeled data points and querying an oracle or domain expert to lower the cost of labeling. Incorporating active learning into the future work is something the researchers would like to do to boost the learning performance.

H. *Top-N recommendations on Unpopular Items with Contextual Knowledge* [10]

Traditional recommender systems make item recommendations to users; more recently, some of them have taken into account the context of forecasts. The authors present a strategy in this research that uses traditional recommendation algorithms and post-filters recommendations based on the contextual information accessible to them. The most important relationships between context and item properties are identified using association rules. To give contextualized suggestions, the mined rules are used to filter the predictions made by typical recommender systems. Their experiments suggest that the proposed methodology can improve the output of traditional algorithms proposed in the literature, particularly for unpopular goods.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907096 |

They demonstrated how contextual rules, which express regular links between context information and the features of evaluated items, can improve the recall of state-of-the-art collaborative recommender systems in this paper. Both customized and non-personalized recommender systems can benefit from their approach. They plan to work on the cross-domain challenge in the future. They're specifically looking into the possibility of applying contextual rules gathered from one dataset to another.

I. *Making Recommendations from Multiple Domains* [8]

This research proposed a generalized cross-domain collaborative filtering paradigm that smoothly blends social network data with cross-domain data. This is accomplished through the use of tensor factorization in conjunction with topic-based social regularization. By identifying shared implicit cluster-level tensors from disparate domains, this system can transfer high-dimensional data without the requirement for decomposition. Extensive experiments on real-world datasets show that the proposed framework outperforms state-of-the-art algorithms in the areas of item, user, and tag recommendation.

They describe a unique collaborative filtering approach for merging social and cross domain networks in a unified framework using latent feature sharing and cluster-level tensor sharing in this paper. This framework makes use of data from many domains and allows useful knowledge to be transferred from an auxiliary domain to the target domain. Extensive studies on real-world datasets have revealed that our unified framework beats state-of-the-art approaches in all three recommendation tasks. They've also proved the algorithm's scalability by implementing it on a map-reduce architecture.

J. *Dynamic Matrix Factorization with Priors on Unknown Values* [9]

Based on explicit feedback, advanced and effective collaborative filtering algorithms presume that unknown ratings do not follow the same model as observed ratings (not missing at random). They build on this assumption in this paper and present a novel dynamic matrix factorization framework that permits an explicit prior to be established on unknown values. We can update the factorization in real time whenever new ratings, users, or objects join the system, regardless of the amount of the data (number of users, items and ratings). As a result, we can swiftly propose things to even new users. They put their methods to the test on three big datasets, two of which are extremely sparse, in both static and dynamic conditions.

They provided a new, simple, and effective technique to incorporate a prior on unknown ratings in various regularly used matrix factorization loss functions in this paper. They used an experiment to show how important it is to include such a prior in order to solve the collaborative ranking problem. They also addressed the issue of factorization updating when new users, items, and ratings are added to the system. They feel that this issue is crucial to real-world applications of recommendation systems, because new users frequently enter those systems, and the factorization must be maintained up to current in order to provide recommendations to them after just a few encounters with the platform. They provide an update technique whose complexity is independent of data size, making it a viable option for enormous datasets.

III. Conclusion

In this survey paper, we have discussed few of the existing work on the same topic. A vigorous research is done in this area and we can see that there has been a huge progress. Recommender systems are increasingly being utilized to assist filter out extraneous material and propose relevant information to users. Traditional recommendation systems are based on a single domain. Recent research has looked at the relationships between domains and developed models that leverage user preferences in one area to predict user preferences in another. There might occur an information overload. We may say that each system's performance differs depending on the protocol applied. And none is perfectly appropriate according to the recent scenario about the increasing popularity of social media. From that we can conclude that each work has a scope for improvement. The current scenario about

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907096 |

the increasing use of social media demands more work to be done. This paper shows that most of these systems need to be updated according to the recent use of social votings.

REFERENCES

[1] Thanh-Nam Doan, Shaghayegh Sahebi , Review-Based Cross-Domain Collaborative Filtering: A Neural Framework ,University at Albany

[2] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Trans. Knowl. Data Eng., vol. 17, no. 6, pp. 734–749, Jun. 2005. ,

[3] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," Adv. Artif. Intell., vol. 2009, Aug. 2009, Art. no. 421425, doi: 10.1155/2009/421425.,

[4] Y. Koren, "Factorization meets the neighborhood: A multifaceted collaborative filtering model," in Proc. ACM KDD, 2008, pp. 426–434. ,

[5] Y. Koren, "Collaborative filtering with temporal dynamics," in Proc. KDD, Paris, France, 2009, pp. 447–456. ,

[6] A. Paterek, "Improving regularized singular value decomposition for collaborative filtering," in Proc. KDDCup, 2007, pp. 39–42. ,

[7] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in Proc. NIPS, vol. 20. 2008, pp. 1257–1264. ,

[8] (2.9) Vivek kumara Krishna Mohan, PD Shrivastva, Shailendra Singh , Senior member, Cross Domain Recommendation Using Semantic Similarity And Tensor Decomposition, International Conference on Computational Modeling and Security (CMS 2016)

[9] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from noisy entries," J. Mach. Learn. Res., vol. 11, pp. 2057–2078, Jul. 2010. ,

[10] P. Cremonesi, Y. Koren, and R. Turrin, "Performance of recommender algorithms on top-N recommendation tasks," in Proc. ACM RecSys, 2010, pp. 39–46.