k-Surrounding Neighbors: Incorporating Serendipity in Collaborative Recommendations

Huidi Lu(✉)

Department of Marketing, NEOMA Business School, Mont-Saint-Aignan 76130, France
huidi.lu@neoma-bs.fr

Abstract. Social recommender systems have become ubiquitous in our online environment, but there are growing concerns that they narrow our horizons and polarize our opinions. This paper proposes a new recommendation algorithm, *k*-Surrounding Neighbors, based on the theory of weak ties, to increase the diversity and novelty of recommendations. The proposed method discards some nearest neighbors based on their similarity, giving more weight to less similar others, which can provide fresh information and new experiences. Validation tests using several metrics show that it significantly improves recommendation diversity and novelty at a minor cost in precision.

Keywords: Collaborative filtering · k-Nearest Neighbors · Serendipity · Recommender system

1 Introduction

The early days of the Internet promised a mind-expanding utopia, where we could freely exchange new ideas and contemplate other points of view. However, revelations involving cyber-bullying, troll factories, and alternative facts raise concerns about the use of the online space. We witness an alarming trend where people resist proper interactions with the outside, and their opinions polarize [2, 7, 9]. This problem could be due, in part, to the way we create and utilize our recommender systems.

Web platforms typically develop recommendation systems to predict consumer preferences accurately and suggest the most relevant items. Social networks are harnessed to check out the range of choices for us, reducing the options we need to review. The rationale for this strategy is to reduce information overload, assuming that consumers want to view movies, travel to places, and read books that others like them view, travel to, and read. However, this convenient function shields us from serendipitous new experiences or random discoveries, and our opinions and tastes become more solidified. Some attempts towards recommendation novelty or diversity have proven successful at increasing consumer satisfaction, [6] making it necessary to come up with potentially better solutions.

Previous approaches tackled this problem with various methods, including graph-based, [5, 10, 16] content-based [14] and model-based ones [1, 8]. Graph-based models are both effective and intuitive, as they directly rely on similar users to provide new yet

K. Hemachandran et al. (Eds.): ICAID 2023, AHIS 9, pp. 115–122, 2023.
https://doi.org/10.2991/978-94-6463-222-4_11

very relevant information [5]. Our proposed method builds upon this idea, suggesting a simple modification to the original *k*-NN voting algorithm, which discards a certain number of nearest neighbors based on the weak-tie theory. Many valuable contributions come from people with whom we have limited connections, providing us with fresh information that tends to be omitted if we only interact with similar people [4]. Therefore, the recommendation system should put more emphasis on less similar others to increase the possibility of discovering new information.

In the following analysis, we describe our proposed method in detail and validate its performance using several metrics, demonstrating that our model greatly extends the diversity and novelty of the recommendation list, though at a minor cost in its precision. This gap is negligible given the significant improvement in novelty, making this method a viable strategy for replacing the traditional algorithm.

2 Related Work

This study relates closely to literature on the collaborative filtering algorithm [11]. The basic type of collaborative filtering uses a memory-based approach which directly utilizes the ratings data to compute similarity and make recommendations. A typical example of this approach averages the ratings of the most similar individuals to predict how much the focal person will like that item. Another type of collaborative filtering is based on latent models, such as singular value decomposition or Bayesian networks. Memory based models are easier to implement but is less scalable than model-based approaches. Since the goal of this study is to examine the effectiveness of weak-tie information, we choose memory-based models as our starting point.

To date, there has been limited research that has focused on distinguishing between strong and weak ties in recommendation settings, with only a few exceptions alluding to this topic [5, 12, 13]. One example is to cluster users into overlapping groups based on similarity and then assign a description of their tastes to each group [5]. Recommendations are then produced based on neighboring groups with overlaps, rather than the user's own group. In contrast, our proposed method alleviates the need to explicitly segment users. Instead, we directly select the user's group of 'weak ties' and apply the collaborative filtering mechanism. We show that this simple method can provide novel recommendations while maintaining a reasonable level of accuracy. Some other researchers extended the Bayesian Personalized Ranking model to incorporate the distinction between strong and weak ties [12, 13]. Their method assumes that users have preferences for both items and tie strength. While their method is model-based, our modification to a memory-based approach can achieve the same objective with lower requirement on computation.

This idea also relates broadly to the design of recommendation systems. Although at present various recommendation systems have achieved success in many domains, there is still a long way to go towards user satisfaction. Researchers have started to focus on other aspects of a reliable recommendation algorithm, such as novelty, diversity, and serendipity [1, 9]. These concepts are closely related and suggest that avoiding narrow recommendations is generally preferable for user satisfaction. Users tend to prefer recommendation lists that are broader and more diverse. Our model builds on this

idea by proposing a new, simple recommendation method that achieves greater diversity and novelty in recommendations.

Lastly, this study is also related to the theory of weak ties. A number of studies have found that while close friends can provide support, those who are not as close are more effective in providing novel information [4]. This idea has been applied in various fields, including information dissemination, job search, and innovation. In the context of recommendation systems, the theory of weak ties suggests that recommendations from weakly connected individuals, such as acquaintances or strangers, could lead to more diverse and novel recommendations. Our study builds upon this idea and proposes a novel application of the theory of weak ties in the recommendation systems domain.

3 Proposed Method

Standard user-based collaborative filtering technique (i.e., with *k*-NN) averages the preferences of the most similar individuals (concept illustrated in Fig. 1). Recommendations are produced by sorting the inferred preferences for the focal user based on other similar users. However, it causes two problems. First, given the high similarity among these nearest neighbors, the system will constantly reinforce the focal user's current interest. Second, it is highly likely that the user is already aware of some missing items in the system. Since the user is very similar to his closest neighbors, their knowledge should be similar as well. The reason that we do not observe his/her rating for a specific item is not that he/she is not aware, but probably he/she does not like it. Consequently, it is a bad idea to still recommending these items to users.

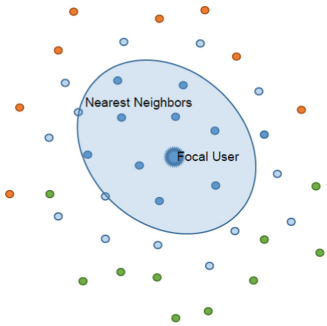

Fig. 1. Traditional Method

One can easily see that a main cause of the above issues is that nearest neighbors are too similar to the focal individual. Their salience renders the system susceptible to several biases. Our proposed method directly tackles this issue. Realizing that 'weak-tie' friends are more effective in communicating new information, we modify the original model so that we put more emphasis on individuals who are moderately close to our target. Specifically, we discard certain number of nearest neighbors before averaging, while include more people who were omitted in the traditional model because they were far away. The method is depicted in Fig. 2. The exact number of discarded individuals and

window length will be determined with training data, by balancing predictive accuracy with the ability to recommend novel items.

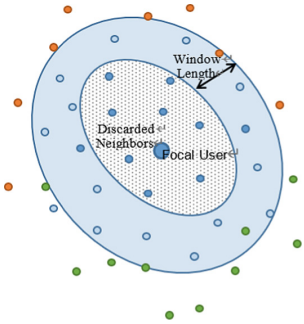

Fig. 2. Proposed Method

4 Evaluation on the MovieLens Dataset

To validate our model, we follow the previous literature and test it on the MovieLens Dataset. This dataset contains ratings from the MovieLens web site (movielens.org). It is one of the stable benchmark datasets, which records one million ratings from 6000 users on 4000 movies until Feb. 2003. Besides its wide publicity as a standard research dataset, another advantage of using movie related data is that this is a good setting for our project, since users not only want to watch relevant movies but also desire to explore more beyond what they are currently aware. Before evaluating the model, we randomly picked 30% of the data as an independent testing sample.

4.1 Prediction Accuracy

We conducted experiments by varying the number of discarded close neighbors and moderately close individuals to analyze the performance of our proposed method in different scenarios. The specification with 0 discards corresponds to the original k-NN method. The results on prediction accuracy are shown in Fig. 3. It indicates that discarding some nearest neighbors does have a slight negative impact on prediction accuracy compared with unmodified k-NN, but this is not significant. Even after excluding the top 180 neighbors, the accuracy level only decreased by about 1%. Furthermore, the accuracy steadily increases as we evaluate more than 600 individuals, which is a robust finding that remains unchanged when shifting the window away from the center of similarity.

4.2 Intra-list Diversity

Diversity metrics measures the variety of items in the recommendation set. We define the intra-list diversity for an individual u's top-N recommendations list $R_N(u)$ as follows:

$$ILD@N = \frac{1}{|U| \cdot N(N-1)} \sum_u \sum_{i \neq j \in R_N(u)} \left(\frac{1}{sim(i,j)} \right) \tag{1}$$

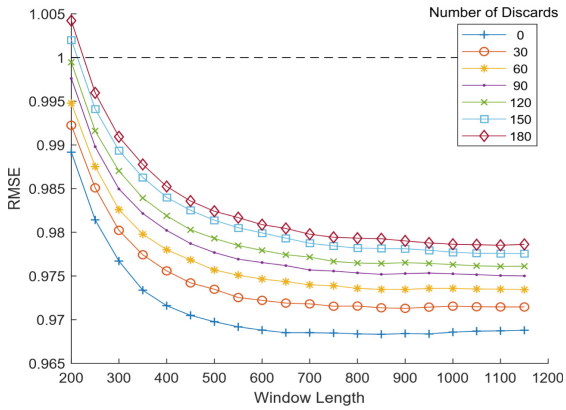

Fig. 3. Accuracy with Varying Window and Discarded Neighbors

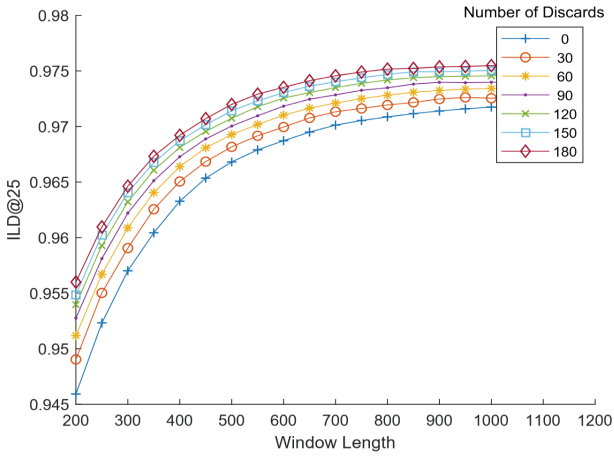

Fig. 4. ILD@25 with Varying Window and Discarded Neighbors

which is the inverse of all pairwise similarities of the N items recommended averaged across $|U|$ personal lists [15, 17]. Throughout the analysis, we set $N = 25$ to reflect the usual practice of real recommendation systems.

From the comparison in Fig. 4 we see that intra-list diversity tends to increase with window length, though at a decreasing rate. Shifting the window outwards moderately diversifies the recommended list. we see that intra-list diversity increased by approximately 1% after excluding about 100 nearest neighbors from our analysis.

4.3 Novelty

Recommender systems sometimes suggest popular items that are too obvious to be of much help. For instance, recommending a well-known movie like *Titanic* or *Star Wars* may be accurate, but it adds little value since the user is likely already aware of it. To

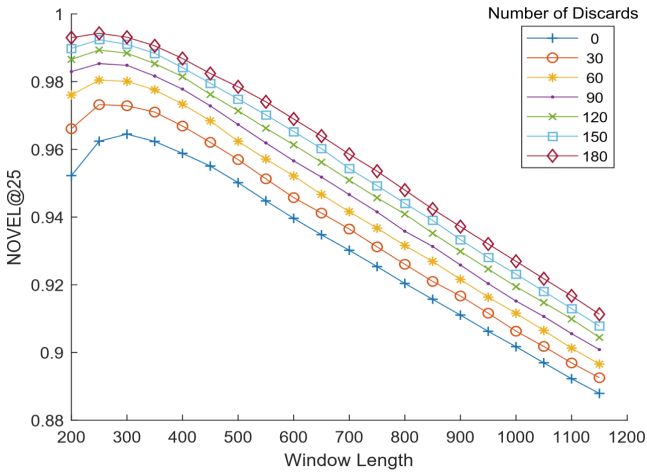

Fig. 5. NOVEL@25 with Varying Window and Discarded Neighbors

measure the system's ability to offer truly novel recommendations, we use the following formula [3]:

$$NOVEL@N = \frac{1}{|U|} \sum_{u}^{U} \frac{R_N(u) \backslash PM}{N}. \tag{2}$$

Here, *PM* represents a list of items recommended by a basic model assumed to have low unexpectedness. We utilize the top 500 items with the highest number of ratings to form the *PM* recommendation list. An item is considered novel if it does not belong to *PM*. A higher *NOVEL* score indicates that the method is more successful in offering fresh and new suggestions.

In Fig. 5, the results show a significant increase in novelty when discarding a certain number of nearest neighbors. Excluding the top 90 neighbors results in a more than 2% increase in this metric compared to the baseline (unmodified k-NN). Additionally, the plots demonstrate a concave shape, with the optimal window length at around 300, regardless of neighbor exclusion. Including too many individuals introduces strong popularity bias, as evidenced by the obvious decline when the window length is higher than 400. These findings are noteworthy and suggest that the proposed method can effectively increase novelty while avoiding popularity bias.

5 Conclusion

In this study, we propose a new recommendation method named 'k-Surrounding Neighbors' to address some of the limitations of traditional collaborative filtering techniques. The method discards some of the nearest neighbors to increase the diversity of recommendations, based on the theory of weak ties. The experiment results showed that this method improved recommendation variety and novelty, with a minimal decrease in predictive accuracy. However, further research is needed to compare its performance against

model-based methods and to improve its scalability. Overall, the proposed method is a promising solution to current recommendation system challenges.

References

1. Adamopoulos, P., Tuzhilin, A.: On unexpectedness in recommender systems: Or how to better expect the unexpected. ACM Transactions on Intelligent Systems and Technology (TIST) 5(4), 54 (2015)
2. Chaney, A.J., Stewart, B.M., Engelhardt, B.E.: How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In: Proceedings of the 12th ACM Conference on Recommender Systems. pp. 224–232. ACM (2018)
3. Ge, M., Delgado-Battenfeld, C., Jannach, D.: Beyond accuracy: evaluating recommender systems by coverage and serendipity. In: Proceedings of the Fourth ACM Conference on Recommender Systems. pp. 257–260. ACM (2010)
4. Granovetter, M.S.: The strength of weak ties. American Journal of Sociology 78(6), 1360–1380 (1973)
5. Kamahara, J., Asakawa, T., Shimojo, S., Miyahara, H.: A community-based recommendation system to reveal unexpected interests. In: Proceedings of the 11th International Multimedia Modelling Conference. pp. 433–438. IEEE Computer Society (2005)
6. Kotkov, D., Konstan, J.A., Zhao, Q., Veijalainen, J.: Investigating serendipity in recommender systems based on real user feedback. In: Proceedings of the 33rd Annual ACM Symposium on Applied Computing. pp. 1341–1350. ACM (2018)
7. Lazer, D.: The rise of the social algorithm. Science 348(6239), 1090–1091 (2015)
8. Lu, Q., Chen, T., Zhang, W., Yang, D., Yu, Y.: Serendipitous personalized ranking for top-n recommendation. In: 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT). vol. 1, pp. 258–265. IEEE (2012)
9. McNee, S.M., Riedl, J., Konstan, J.A.: Being accurate is not enough: how accuracy metrics have hurt recommender systems. In: CHI'06 Extended Abstracts on Human Factors in Computing Systems. pp. 1097–1101. ACM (2006)
10. Onuma, K., Tong, H., Faloutsos, C.: Tangent: a novel 'surprise me' recommendation algorithm. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 657–666. ACM (2009)
11. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. Advances in Artificial Intelligence 2009 (2009)
12. Wang, X., Hoi, S.C., Ester, M., Bu, J., Chen, C.: Learning personalized preference of strong and weak ties for social recommendation. In: Proceedings of the 26th International Conference on World Wide Web. pp. 1601–1610. International World Wide Web Conferences Steering Committee (2017)
13. Wang, X., Lu, W., Ester, M., Wang, C., Chen, C.: Social recommendation with strong and weak ties. In: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. pp. 5–14. ACM (2016)
14. Zhang, Y.C., Séaghdha, D.Ó., Quercia, D., Jambor, T.: Auralist: introducing serendipity into music recommendation. In: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining (WSDM). pp. 13–22. ACM (2012)
15. Zheng, Q., Chan, C.K., Ip, H.H.: An unexpectedness-augmented utility model for making serendipitous recommendation. In: Industrial Conference on Data Mining. pp. 216–230. Springer (2015)

16. Zhou, T., Kuscsik, Z., Liu, J.G., Medo, M., Wakeling, J.R., Zhang, Y.C.: Solving the apparent diversity-accuracy dilemma of recommender systems. Proceedings of the National Academy of Sciences 107(10), 4511–4515 (2010)
17. Ziegler, C.N., McNee, S.M., Konstan, J.A., Lausen, G.: Improving recommendation lists through topic diversification. In: Proceedings of the 14th International Conference on World Wide Web. pp. 22–32. ACM (2005)