The Journal of Supercomputing (2020) 76:5161–5174
https://doi.org/10.1007/s11227-019-02751-7

Improved collaborative filtering recommendation algorithm based on differential privacy protection

Chunyong Yin[1] · Lingfeng Shi[1] · Ruxia Sun[1] · Jin Wang[2]

Published online: 8 January 2019

Abstract
In order to receive efficient personalized recommendation, users have to provide personal information to service providers. However, in this process, personal private data are in an extremely dangerous situation. Personalized recommendation technology based on privacy protection can enable users to enjoy personalized recommendations, while private data are also protected. In this paper, an efficient privacy-preserving collaborative filtering algorithm is proposed, which is based on differential privacy protection and time factor. The proposed method used the MovieLens data set in the experiment. Experimental results showed that the proposed method can effectively protect the private data, but the accuracy of recommendation is slightly inferior than the traditional collaborative filtering algorithm.

Keywords Collaborative filtering · Differential privacy · DiffGen · Time factor

1 Introduction

With the coming of the information age, the total amount of information disseminated on the Internet continues to increase. People have gradually moved from an information-deficient era to an information-overloaded era. In recent years, surveys show that the growth rate of global data has not slowed down, but it continues to accelerate. Faced with this critical data explosion, information users need to find truly valid information from vast amounts of information, while information producers must expend a great deal of effort to make the information widely available.

There are two main methods to deal with the problem of information overload: (1) the search. In order to express the information needs, users usually employ a

✉ Jin Wang
jinwang@csust.edu.cn

[1] School of Computer and Software, Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing 210044, China

[2] School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410004, China

few short words or phrases to form a query, which will be submitted to the search engine then. Finally, the query results of search engine are sent back to database in real time. (2) The recommendation. Users are often unaware of their specific needs, or cannot describe their requirements in detail. In this case, it is impossible to provide users with good services because of the exhaustive search requirements. For example, in e-commerce, the demands of users are often "potential" or "implicit." Consequently, how to accurately analyze the true intentions of users and make precise personalized recommendations is a problem every business must handle. The recommendation system is precisely a suitable solution to the above problem. The main work of the recommendation system is to establish the connection between users and information, which enables users to effectively obtain information they are interested in. It can also maximize the effective promotion of business information to achieve mutual benefit.

As we all know, personalized recommendation system exactly brings convenience to our lives and economic growth [1]. By contrast, our sensitive private information is leaked when the recommended system is operating. In order to obtain a recommendation accurately, the recommendation system builds an interest model of users, which is according to the behavior records of users and relevant data of nearby friends. It is inevitable to cause the leakage of private data when a large number of data are used for data mining. In the contemporary era when people pay more and more attention on their own privacy protection, only a little data or some fake data are provided to the data analysts. The recommended models are inevitably lack of representation, which are mined and constructed based on these false information. The recommended accuracy and efficiency of the recommendation system are both greatly deteriorated because of these bogus data. Therefore, it is imperative to ensure that users do not need to worry about their private data being leaked.

Applying the extremely strict privacy protection model to the personalized recommendation system can not only greatly avoid the leakage of private information of users, but also ensure the high efficiency of personalized recommendation system. So, privacy-preserving personalized recommendation system has great research value, which combines privacy protection and personalized recommendation technology.

2 Related work

In recent years, related researches and applications of personalized recommendation based on differential privacy protection are still in a budding and actively developing state at home and abroad. A complete personalized recommendation system contains two parts: data publishing and data analysis. The operations of these two parts will all leak the private information of users. The technology of differential privacy protection has attracted great attention of researchers, so the number of research results on data publishing and data mining of differential privacy protection continue to increase.

According to the differential privacy protection of classification methods in data mining, Blum et al. [2] proposed the method of Su LQ-based ID3. This method

combined the Laplacian noise mechanism and the information gain to select the corresponding segmentation attributes. The information gain corresponding to each attribute must be calculated, but it will cause a lot of privacy budgets to be wasted when the number of corresponding segmentation attributes is large. Friedman and Schuster [3] added index noise to the segmentation attributes in order to avoid the waste of private budgets, which can address the problem in Su LQ-based ID3. This method can only support a small amount of analytical queries, because it is based on an interactive query interface. DiffGen combines the advantages of both to find a better solution [4]. First of all, index mechanism and information gain are applied to this method to determine each segmentation attribute. Next, divide the data to leaf nodes through the top-down classification tree. Finally, complete privacy protection by adding Laplacian noise to the leaf nodes.

In order to achieve the purpose of active recommendation, the personalized recommendation system analyzes the interests and hobbies of users by collecting and learning information of users. Due to the sustained development of the Internet, the competition of the global market is increasingly fierce. In this case, personalized service is closely related to the development of business. The personalized recommendation system can fully improve the quality of service and the efficiency of accessing website. The current personalized recommendation systems can be mainly divided into the rule-based systems and information filtering systems according to different recommended technologies. Furthermore, the information filtering system can be further divided into the content-based filtering system and the collaborative filtering system.

Rule-based systems are simple and direct. These systems can find new points of interest, which is the most obvious advantage. However, these systems have a lower degree of personalization and cannot be dynamically updated. The rule extraction of these systems is also difficult and time-consuming. Moreover, as the number of rules increases, the system will become too difficult to manage. Moreno et al. [5] generated a series of association rules through a data mining algorithm. Then, a prediction model was established based on association rules, which can improve the efficiency and reduce the errors of the recommendation. Sun et al. [6] applied association rules to mine the correlation between projects, in order to reduce the sparseness of data in collaborative filtering algorithms. Chun et al. [7] proposed a hierarchical algorithm for association rules. The algorithm measured the relevance between a user and an association rule by comparing the attributes of users belonging to the same association rule. If the association rules are properly rated, the accuracy of filtering algorithms based on association rules can be greatly improved. Li et al. [8] reduced the number of rules that users were not interested in, by defining multiple templates of association rule. Because the recommendation algorithm based on association rules is simple and can be automatically processed, it is widely used in e-commerce recommendation systems, such as Amazon [9].

Content-based systems have intuitive recommendation results, which does not need any domain knowledge. However, it is difficult to distinguish the quality and style of the content of resources in these systems. These systems can only find resources similar to interests of users. In addition, the taste of users must be expressed in the form of content features, so it cannot explicitly get the judgment

of other users. The content-based recommendation technology is widely used in various types of machine learning algorithms, such as Bayesian classifier [10], clustering, decision trees and artificial neural networks [11]. Salem and Rauterberg [12] proposed a method of consolidating multiple user profiles in a context-based environment to deal with the problem of conflicts. Sparacino [13] introduced an approach based on Bayesian network, which was used to create a profile that provides personalized services to visitors in the museum.

The idea of applying differential privacy protection to recommendation systems was first proposed by McSherry and Mironov [14]. Firstly, the input of the recommendation system was interfered. Next, a project similarity covariance matrix was built in light of this input. Then, interference matrix was interfered by Laplacian noise. Finally, the conventional recommendation algorithm was carried out. Machanavajjhala et al. [15] applied differential privacy protection in a recommendation system of social network data.

Researches on the combination of privacy protection technology and personalized recommendation technology have attracted great attention of many scholars. Shen and Jin [16] proposed differential privacy data publishing method. The proposed model of privacy protection has lightweight computation and strong private guarantee. In order to achieve useful and feasible disturbances, they designed a new loose allowance mechanism, which can inject flexible instance based on noise. Jin and Jin [17] developed a built-in client system of personalized recommendation based on enhanced privacy, which uses data disturbance to protect privacy. Huang et al. provided an online anonymous method that deals with unknown and dynamic location-based service. Experiment showed that it is feasible to implement personalized recommendations while ensuring user privacy [18]. Ahmed et al. [19] proposed a method of group recommendation based on k-anonymity to protect the identity of unauthorized attackers.

3 Improved personalized recommendation based on privacy protection

3.1 Data publishing technology based on differential privacy protection

In this paper, a data publishing technology based on differential privacy protection is applied to protect the original data set. This method can effectively reduce the risk of leaking private data, while the required privacy protection services are received. This paper uses DiffGen algorithm to perform data publishing. The published data can be effectively used in data mining and have significant differences from the original data set.

3.1.1 Differential privacy protection

Differential privacy is a new privacy protection model proposed by Dwork [20]. This method can improve two shortcomings of the traditional privacy protection model:

1. It defines a rather strict attack model and does not care about how much background information an attacker possesses. Even if the attacker has mastered all the record information except a certain record, the privacy of the record cannot be disclosed.
2. The definition of privacy protection and the quantitative assessment method are given. Many advantages of differential privacy make it rapidly replace the traditional privacy protection model and become the research hot spot of privacy protection. Furthermore, differential privacy arouses the concern in many fields, such as theoretical computer science, database, data mining and machine learning.

The idea of differential privacy protection ensures that deletions or additions of records in one data set do not affect the results of the analysis by adding noise. Therefore, even if an attacker gets two data sets that differ only by one record, he cannot deduce the hidden information of that one record by analyzing their results [21].

Data sets D and D' have the same attribute structure, and the symmetry difference between the two is $D\Delta D'$. $|D\Delta D'|$ represents the difference number recorded in $D\Delta D'$. If $|D\Delta D'|=1$, they will be called the adjacent data set.

Differential Privacy: There is a random algorithm M. P_M is a set of all the possible outputs of M. For any two adjacent data sets D and D' and any subset S_M of P_M, algorithm M is called differential privacy protection, if the algorithm M satisfies

$$P_r\left(M(D) \in S_M\right) \leq \exp(\varepsilon) \times P_r\left(M\left(D^{'}\right) \in S_M\right) \tag{1}$$

Parameter ε is called private budget. The probability Pr [·] is controlled by the randomness of algorithm M and also indicates the risk of disclosing private data. The private budget ε denotes the degree of privacy protection. The smaller ε is, the higher the degree of privacy protection is [22].

The noise mechanism is the main technique to realize the differential privacy protection. Furthermore, the Laplace mechanism and index mechanism are the two basic mechanisms of differential privacy protection in differential privacy. The Laplace mechanism is suitable for the protection of numerical results, while the index mechanism applies to non-numeric results.

3.1.2 DiffGen

DiffGen is a privacy protection publishing algorithm that uses decision trees to publish data. The algorithm firstly generalizes the data set completely. Then, subdivide the data set according to the scoring strategy. Finally, Laplacian noise is added to the data to be released. The specific steps are shown in Fig. 1. The private budget in DiffGen algorithm is divided into two parts. The one is used to add Laplace noise to the equivalent classes, which is released at the end of the algorithm. The other is divided equally into exponential mechanisms in each iteration.

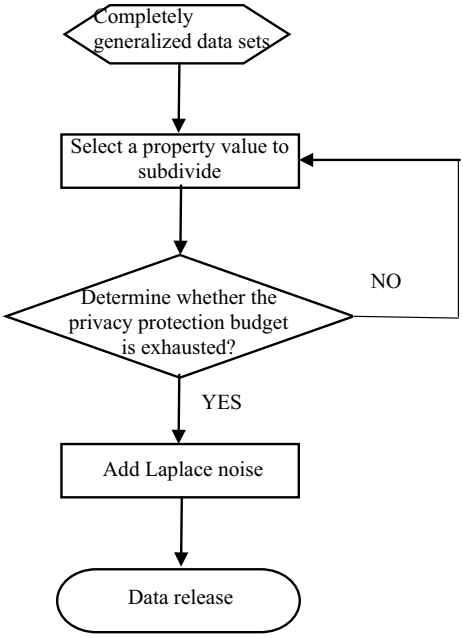

Fig. 1 The flowchart of DiffGen

3.2 Improved collaborative filtering algorithm based on time factor

This paper proposes an improved personalized recommendation method based on time factor. Firstly, the forgetting curve is used to track the change in interests with time. And then, a model of personalized recommendation based on dynamic time is built. Finally, the time forgetting factor is introduced in the collaborative filtering to improve the prediction effect of the interest-aware algorithm.

3.2.1 Collaborative filtering algorithm

The traditional collaborative filtering recommendation algorithms are mainly divided into two categories: memory-based collaborative filtering recommendation algorithm and model-based collaborative filtering recommendation algorithm [23]. Their respective common algorithms are shown in Fig. 2.

The memory-based collaborative filtering recommendation algorithm can be divided into two categories: user-based and project-based collaborative filtering recommendation algorithm. The user-based collaborative filtering recommendation algorithm mainly focuses on users. It searches the similar neighbors of the target users by calculating the similarities of users, and then the objects are recommended to the target users according to the found neighbors. The project-based collaborative filtering recommendation algorithm is mainly based on the project. It mainly calculates the similarity between items and searches items similar to the target item according to the similarities among the items, which is different from user-based

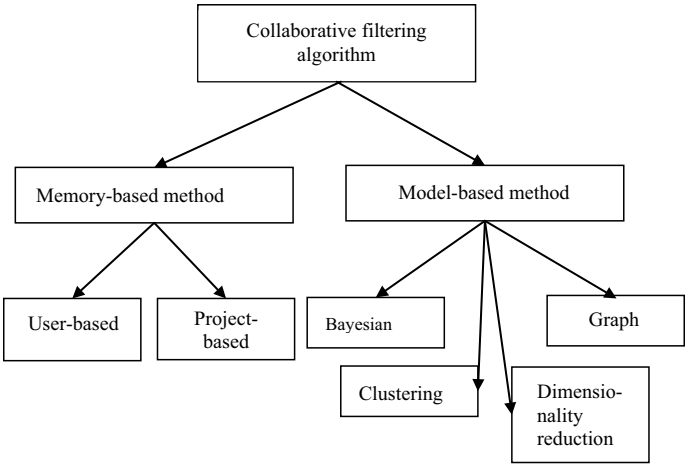

Fig. 2 Classification of collaborative filtering algorithms

collaborative filtering recommendation. Then, the rating of users on the target item is predicted according to the historical scores of the items.

The basic idea of the model-based collaborative filtering recommendation algorithm is to train a model based on the historical score data of the project, which will be used to predict the score. How to train an effective model is the key to this kind of method. Data mining, machine learning and other methods are often applied to train the model. However, this kind of method takes a long time to model training and assessing. This method relies on the complexity of training model method and data set size [24].

3.2.2 Improved collaborative filtering algorithm

As shown in Fig. 3, an improved personalized recommendation model based on time factor has been proposed. It can effectively improve the quality of recommendation and reduce the time complexity of calculation. The main method is to introduce the

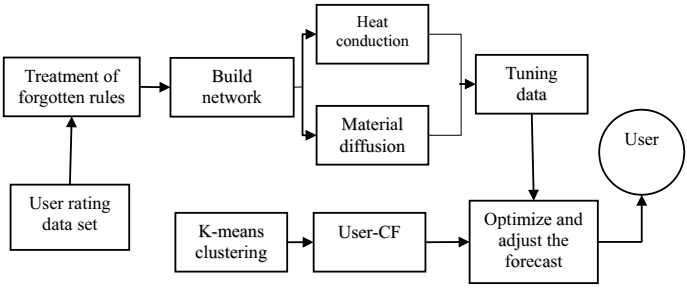

Fig. 3 The personalized recommendation model based on time factor

Table 1 The number of different user factors f corresponds to the RMSE value

Attribute	Property value								
Number of factors	1	2	3	4	5	6	7	8	9
RMSE	2.613	2.542	2.323	2.107	1.002	2.355	2.436	2.291	2.188

Fig. 4 RMSE curve corresponding to the number of different user factors f

forgetting curve based on the time factor to process the preference of users, cluster the set of users and optimize the predictive score.

As we all know, with the increase in age, personal interest will also change. Changes in education can also cause changes in the person's preferences, while personal experience can also affect one's interests. In order to improve the quality of recommendation, it is necessary to introduce a time factor, which is used to recommend a movie suitable for the user at that time. It requires that the current interest of the user can be accurately modeled.

This paper randomly selected information of different users from MovieLens. When taking the f value of different users, the RMSE (root-mean-square error [25]) value obtained by the corresponding Mysvd++ algorithm is shown in Table 1.

The changing curve of RMSE and factor f is shown in Fig. 4. It can be seen from Table 1 and Fig. 4 that the values of REMS vary greatly due to the different F values. The maximum value is 2.613, and the minimum is 1.002, which indicates that the recommended model is not stable.

Therefore, in order to fit the model better to adapt the scoring data, we use the sigmoid function to smooth it. This paper presents a method of simulating different users' interest with different models, which is shown in formula (2):

$$\hat{r}_{uit} = \bar{r} + r_t + b_u(t) + b_t + p_u^T \cdot q_i \tag{2}$$

In formula (2), b_u is the personal bias of user, and b_i is the bias of the movie itself. The implicit eigenvector p_u is the user's preference on each factor, and the element $p[u][k]$ in p_u indicates the greater the user's preference for factor k is, the more he likes best. The meaning of the implicit eigenvector q_i is the degree to which the movie i has various factors. The element $q[i][k]$ in q_i indicates that the movie i has the degree of factor k, and the larger the movie is, the more the factor k is. The number of implicit feature factors can be cross-validated to choose a better value. In other words, choose the number of different implicit feature factors f to obtain different corresponding feature factor numbers of the recommended results. Finally, select the minimum mean square error as the final number of eigenvectors.

Users have different requirements in different time periods. In the traditional recommendation system, it is difficult to obtain the current potential and accurate user preference information. Therefore, this paper introduces forgetting curve to adjust user preference information. As time goes on, changes in human memory are diminished by inhomogeneous memory, and the attenuation rate gradually decreases. Through a large number of experiments on the forgetting curve fitting calculation, we get formula (3):

$$r_{i,j} = r_0 + r_1 e^{-\lambda |t-t_0|} (t - t_0 < T) \tag{3}$$

t denotes the current time. $r_{i,j}$ denotes the preference value of the ith resource res_j by the ith user u_i at t [the default $r_{i,j} \equiv 5$ when r_i is calculated by formula (2)]. r_0 represents the minimum preference value of u_i at t. r_1 represents the preference value of u_i to res_j at t_0. T represents the valid time of the calculation formula.

The time factor is defined as λ, and $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_n]$. λ is determined by the user's real demand for resources and calculated by data mining algorithms, association rules and text analysis algorithms.

The minimum preference r_0 is used to adjust the user's preference for resources. For example, the system sets it to a larger value when the resource is just online or currently hot. Dynamic adjustment of r_0 can address the problem of resource cold start. This paper uses a one-dimensional vector to represent: $r_0 = [r_1, r_2, r_3, \ldots, r_n]$.

This paper presents an improved k-means algorithm to alleviate data sparsity and address high-dimensional problems [26]. Reduce data dimensions and mitigate data spars by reducing disruptive users.

The user set $U = \{u_1, u_2, \ldots, u_m\}$, where $u_i = (r_{i1}, r_{i2}, \ldots, r_{in})$, and k-means mainly decomposes the data set into k similar user clusters $(k < m)$ $\{S_1, S_2, \ldots, S_k\}$. S_k is the user cluster of the kth cluster [27].

K-Means calculation process:

Step 1: Select k users as cluster centers in U, which are denoted by $C = \{C_1, C_2, \ldots, C_k\}$, where $C_j = (c_{j1}, c_{j2}, \ldots, c_{jn})$.

Step 2: Traverse U and calculate the distance between the user and the center of the cluster by using the Euclidean distance, which is shown in formula (4):

$$\text{dis}_{u_i, c_j} = \sqrt{(r_{i1} - c_{j1})^2 + (r_{i2} - c_{j2})^2 + \cdots + (r_{ii} - c_{ji})^2} \tag{4}$$

Step 3: Calculate the mean point of the cluster and replace it with the new cluster center. For example, cluster $S_i = \{u_1, u_2, \ldots, u_a\}$, $u_i = (r_{i1}, r_{i2}, \ldots, r_{in})$ to get new cluster center C_j:

$$C_j = \begin{cases} c_{j1} = (r_{11} + r_{12} + \cdots + r_{1n})/n \\ c_{ja} = (r_{a1} + r_{a2} + \cdots + r_{an})/n \end{cases} \tag{5}$$

Step 4: Repeat steps 2 and 3 until the clustering result no longer changes. Determine whether the objective function no longer changes, which is as the end flag. The objective function is defined in formula (6):

$$\text{MIN} = \min\left(\sum_{i=1}^{k} \sum_{u \in S_i} \text{dis}(c_i, u)^2 \right) \tag{6}$$

Calculated by the *k*-means algorithm can be *k* user clusters, and each cluster represents a set of users with similar preferences. Formula (7) for calculating the maximum distance is as follows:

$$\mathrm{DIS}(j+1)=\prod_{i=1}^{j} \mathrm{dis}S_{u_i,u_{j+1}} \tag{7}$$

Compare the user points with the largest DIS $(j+1)$.

$$\mathrm{dis}S_{u_i,u_{j+1}} = \begin{cases} 1\ (\mathrm{dis}S_{u_i,u_{j+1}}=0) \\ \mathrm{dis}S_{u_i,u_{j+1}}\left(\mathrm{dis}S_{u_i,u_{j+1}}>0\right) \end{cases} \tag{8}$$

$$\mathrm{dis}S_{u_i,u_{j+1}} = \sqrt{(r_{i1}-r_{j+1,1})^2+(r_{i2}-r_{j+1,2})^2+\cdots+(r_{in}-r_{j+1,n})^2} \tag{9}$$

Formula (6) uses the product of the calculation method, which can effectively address the problem of center point set and other cluster centers adjacent to the interference. By using the additive method, it cannot be guaranteed that the cluster center is a single state under special circumstances [28].

The choice of learning rate can take a smaller constant or variables [29–31]. For example, the learning rate can be set as an inverse function on the number of iterations which avoid the optimal intersection point. Moreover, the convergence can be faster. Implementation steps of personalized recommendation algorithm based on time factor:

1. Training data preprocessing. Each user obtains a chronological ranking list, which is divided by the week and calculated by the average score of each shard. The same operation is applied to all users. Finally, one score corresponding to each user obtains a list, whose average points scored in chronological order. Similarly, the average score list for each time slice corresponding to each item can be obtained, which can be divided by time slice. The ratings of social groups on the movie are divided according to the time slice, and the average score *rt* corresponding to the film's average share in each time period is obtained.
2. Based on the improved algorithm, the user preference matrix is generated according to the existing user information, resource information and user history preference information in the system. Table 2 shows the basic code.
3. Prediction. According to the model obtained in (2), a predicted score can be obtained by inputting formula (2) for each test sample. *fu* and *fi* are the scores of user *u* and item *i* at time *t,* respectively, influenced by the score of *fu*, *fi* closest to time *t*. *f* is the number of potential factors.

Table 2 Improved personalized recommendation algorithm based on time factor

Algorithm: Improved personalized recommendation algorithm based on time factor
Data: train data
Result: all the variables of the Mysvd++ model
while have a user not been used in U do
Randomly initialize buto a smaller value;
Randomly initialize λuto a smaller value;
Randomly initialize w_{ut} $t \in 1, 2 \cdots f_u$ to a smaller value;
Randomly initialize p_{uj} $j \in 1, 2 \cdots f$ to a smaller value;
while have an movie not been used in I do
Randomly initialize b_i to a smaller value;
Randomly initialize λ_i to a smaller value;
Randomly initialize w_{it} $t \in 1, 2 \cdots f_u$ to a smaller value;
Randomly initialize q_{ij} $s_j \in 1, 2 \cdots f$ to a smaller value

4 Experimental analysis

4.1 Experimental data and framework

In this paper, we use the typical MovieLens data set in the recommended system. The specific experimental data are shown in Table 3. By comparing the RMSE indexes of the two methods under different neighbors, the advantages and disadvantages of the two algorithms can be demonstrated.

This paper designs a comparative experiment that compares the improved privacy-protection personalized recommendation algorithm with the original personalized recommendation algorithm. The DiffGen algorithm is used to perform data perturbation on the original data set, protect the original data set and publish the changed data set. Next, a personalized filtering recommendation algorithm based on

Table 3 Experimental data

Data set	Index	MovieLens data set
Training set	Number of score records	80,000
	Number of users	943
	Number of movies	1682
Test set	Number of score records	10,000
	Number of users	457
	Number of movies	1252
Verification set	Number of score records	10,000
	Number of users	457
	Number of movies	1256

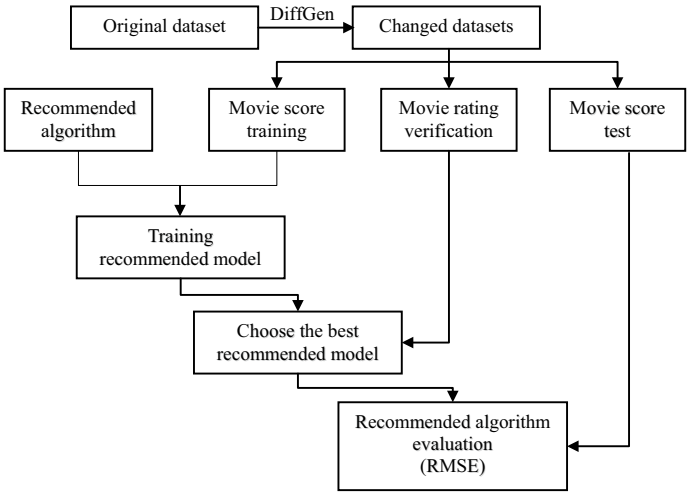

Fig. 5 Experimental framework

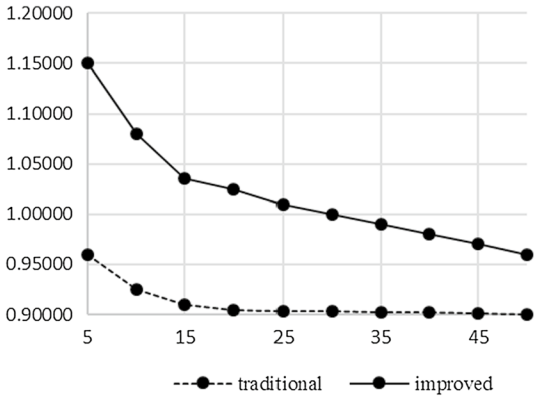

Fig. 6 Experimental comparison chart

preference time factor is used to personalize the published data set. The experiment flowchart is shown in Fig. 5.

4.2 Experimental result

The experimental results are shown in Fig. 6. In the figure, the vertical axis represents the RMSE, and the horizontal axis represents the number of neighbors k. By comparing the changes in RMSE indexes of two algorithms in different neighbors, we find that the improved method proposed in this paper provides effective privacy protection for the data, although it is not as good as the original algorithm in terms of recommendation.

5 Conclusion

This paper is devoted to effectively address the problem of privacy protection in personalized recommendation. The contradiction between the growing need for privacy protection and the harm of personalized recommendation technology for personal privacy data is still a daunting challenge for today's society. We propose an improved privacy-protection personalized recommendation algorithm to provide people with personalized recommendation services and protect the privacy of people's personal data at the same time.

In the following research, we hope to conduct research on the privacy protection technology in the privacy-protection personalized recommendation and design a more excellent privacy protection technology for personalized recommendation.

Acknowledgements This work was funded by the National Natural Science Foundation of China (61772282, 61772454, 61811530332). It was also supported by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX18_1032), Chongqing Research Program of Basic Research and Frontier Technology (No. cstc2015jcyjA40026, No. cstc2016jcyjA0568), Natural Science Foundation of Jiangsu Province (BK20150460) and Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology (CICAEET). It was also funded by the open research fund of Key Lab of Broadband Wireless Communication and Sensor Network Technology (Nanjing University of Posts and Telecommunications), Ministry of Education. Professor Jin Wang is the corresponding author.

References

1. Tang L, Jiang Y, Li L et al (2015) Personalized recommendation via parameter-free contextual bandits. In: International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, pp 323–332
2. Blum A, Dwork C, McSherry F et al (2005) Practical privacy: the Su LQ framework. In: Proceedings of the Twenty-Fourth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. ACM, pp 128–138
3. Friedman A, Schuster A (2010) Data mining with differential privacy. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp 493–502
4. Mohammed N, Chen R (2011) Differentially private data release for data mining. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp 493–501
5. Moreno MN, García FJ, Polo MJ et al (2004) Using association analysis of web data in recommender systems. In: International Conference on E-Commerce and Web Technology, pp 11–20
6. Sun X, Kong F, Chen H (2005) Using quantitative association rules in collaborative filtering. In: International Conference on Advances in Web-Age Information Management. Springer, pp 822–827
7. Chun J, Oh JY, Kwon S et al (2004) Simulating the effectiveness of using association rules for recommendation systems. In: Asian Simulation Conference on Systems Modeling and Simulation: Theory and Applications. Springer, Berlin, pp 306–314
8. Li J, Tang B, Cercone N (2004) Applying association rules for interesting recommendations using rule templates. In: Advances in Knowledge Discovery and Data Mining. Springer, Berlin, pp 166–170
9. Yin C, Feng L, Ma L (2016) An improved Hoeffding-ID data-stream classification algorithm. J Supercomput 72(7):2670–2681
10. Mooney RJ, Bennett PN, Roy L (2017) Book recommending using text categorization with extracted information. In: Recommender Systems Papers from Workshop 2017, pp 49–54

11. Pazzani M, Billsus D (1997) Learning and revising user profiles: the identification of interesting web sites. Mach Learn 27(3):313–331
12. Salem B, Rauterberg M (2004) Multiple user profile merging (MUPE): key challenges for environment awareness. In: European Symposium on Ambient Intelligence 2004, pp 196–206
13. Sparacino F (2003) Sto(ry)chastics: a Bayesian network architecture for user modeling and computational storytelling for interactive spaces. In: UBICOMP 2003: Ubiquitous Computing. Springer, Berlin, pp 54–72
14. McSherry F, Mironov I (2009) Differentially private recommender systems: building privacy into the net. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp 627–636
15. Machanavajjhala A, Korolova A, Sarma AD (2011) Personalized social recommendations: accurate or private. Proc VLDB Endow 4(7):440–450
16. Shen Y, Jin H (2014) Privacy-preserving personalized recommendation: an instance-based approach via differential privacy. In: IEEE International Conference on Data Mining 2014, pp 540–549
17. Jin H, Jin H (2016) EpicRec: towards practical differentially private framework for personalized recommendation. In: ACM Sigsac Conference on Computer and Communications Security. ACM, pp 180–191
18. Huang J, Qi J, Xu Y, Chen J (2015) A privacy-enhancing model for location-based personalized recommendations. Distrib Parallel Databases 33(2):253–276
19. Ahmed KW, Mouri IJ, Zaman R et al (2016) A privacy preserving personalized group recommendation framework. In: International Conference on Advanced Computing. IEEE, pp 594–598
20. Dwork C (2008) Differential privacy: a survey of results. In: International Conference on Theory and Applications of MODELS of Computation. Springer, pp 1–19
21. Clifton C, Tassa T (2013) On syntactic anonymity and differential privacy. In: 29th International Conference on. IEEE, pp 88–93
22. Yin C, Xi J, Sun R, Wang J (2018) Location privacy protection based on differential privacy strategy for big data in industrial internet-of-things. IEEE Trans Ind Inf 14(8):3628–3636
23. Kumari V, Chakravarthy S (2016) Cooperative privacy game: a novel strategy for preserving privacy in data publishing. Hum Centric Comput Inf Sci 6(1):12
24. Boutet A, Frey D, Guerraoui R, Jégou A, Kermarrec AM (2016) Privacy-preserving distributed collaborative filtering. Computing 98(8):827–846
25. Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)?–arguments against avoiding RMSE in the literature. Geosci Model Dev 7(3):1247–1250
26. Yin C, Zhang S (2017) Parallel implementing improved *k*-means applied for image retrieval and anomaly detection. Multimed Tools Appl 76(16):16911–16927
27. Yin C, Xia L, Zhang S, Sun R, Wang J (2017) Improved clustering algorithm based on high-speed network data stream. Soft Comput 22(13):4185–4195
28. Barzaiq OO, Loke SW (2016) Personal destination pattern analysis with applications to mobile advertising. Hum Centric Comput Inf Sci 6(1):17
29. Keegan N, Ji SY, Chaudhary A, Concolato C, Yu B, Jeong DH (2016) A survey of cloud-based network intrusion detection analysis. Hum Centric Comput Inf Sci 6(1):19
30. Yin C, Zhang S, Yin Z, Wang J (2017) Anomaly detection model based on data stream clustering. Cluster Comput 2017:1–10
31. Wang J, Cao J, Li B, Lee S, Sherratt S (2015) Bio-inspired ant colony optimization based clustering algorithm with mobile sinks for applications in consumer home automation networks. IEEE Trans Consum Electron 61(4):438–444