Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 4, No. 3, 2022

Differential Privacy-based Personalized Recommendation Service for "Helping Farmers" Tourism in Guizhou Province

Zifang Xiao, Ting Li and Zhuo Liu

School of Mathematics and Big Data, Guizhou Education University, Guizhou 550000, China

Abstract: The revitalization of the countryside is an important strategy proposed in the report of the 19th Party Congress. Rural tourism in Guizhou Province is the key to the development of "big tourism" in the 14th Five-Year Plan of the province, and also the driving force of rural revitalization. Most of the current rural tourism personalized recommendation systems rely on user information, and the exposure of sensitive user information has become an obstacle to the development of smart tourism. Through the genetic matrix decomposition recommendation algorithm based on satisfying differential privacy protection and differential privacy recommendation algorithm for sparse data, privacy protection of similarity calculation, accuracy of recommendation and data security are realized. Actively develop secure and personalized rural tourism recommendations so that the majority of farmers can benefit from them.

Keywords: Rural tourism, Genetic matrix decomposition recommendation algorithm satisfying differential privacy preservation, Differential privacy recommendation algorithm for sparse data.

1. Introduction

Guizhou will usher in a new historical opportunity, according to the document "Opinions on supporting Guizhou to break new ground in the new era of western development" on. Tourism, as the characteristic advantageous industry of Guizhou and the new pillar industry that can pull regional affluence and achieve sustainable development, has been confirmed by more and more practice and recognized by the government, industry and farmers. Therefore, it is more important to seize the opportunity of western development and achieve a key breakthrough. Based on this paper, we study the personalized recommendation service of "helping farmers" tourism in Guizhou Province based on differential privacy, and provide "personalized and customized Guizhou tour" service for groups of all levels to form scale benefits.

Differential privacy[1] is a new secrecy system based on stochastic chaos techniques, which strictly applies mathematical definitions. It can quantify users' privacy and also defend against background knowledge attacks and synthetic attacks from attackers[2]. Differential privacy is increasingly used in data mining, data distribution, and location services, and has significant advantages over other privacy-preserving techniques such as random interference and data exchange for privacy protection of clustering analysis data[4]. Differential privacy has robust privacy-preserving capability, but it requires a third party that can be trusted to handle these data, while data leakage caused[7] by the third party is numerous, such as the current rural tourism personalized recommendation system mostly relies on user information, and the exposure of sensitive user information becomes an obstacle to tourism development, which not only causes the leakage of user privacy information, but also seriously damages the reputation of tourism software This not only causes the leakage of user privacy information[9], but also seriously damages the reputation of tourism software. On the premise of ensuring the availability of data, but also to meet the two algorithms. By perturbing the data at the user side, the de-third party protection of user privacy is achieved.

In addition, differential privacy techniques can still provide effective protection when an attacker knows all records except sensitive ones. The effect is that user records can be masked, guaranteeing privacy security while safeguarding the availability of data. It not only allows users to experience personalized travel recommendation service but also protects users' information security, so that tourists from all over the world can realize "easy travel", "safe travel" and "enjoyable travel" when they come to Guizhou. The service of "Easy Tour", "Safe Tour" and "Enjoy Tour" is available.

2. Data Sources and Research Methods

2.1. Data source

As shown in the figure 1, the big difference between the data of famous tourist attractions and non-famous tourist attractions in each city of Guizhou province. It can be seen that Guizhou province is very rich in tourism resources, especially karst landscapes. However, due to the lack of good promotion channels, these attention is not high. Through market research and analysis it is known that there are more than 1000 developable tourist attractions in the province. Here, there are towering wondrous peaks everywhere, strange rocks, splashing mountain springs, vast lakes and marshes, ancient caves, all showing the unique, ancient and mysterious charm of Guizhou plateau.

Therefore, it is more important to seize the opportunity of western development and achieve key breakthroughs. On the basis of this, we will study the personalized recommendation service of "helping farmers" tourism in Guizhou Province based on differential privacy, and provide "personalized and customized Guizhou tour" service for groups at all levels to form scale benefits. At the same time, based on the two algorithms, we will actively develop personalized tourism, safe tourism and rural tourism, so that all farmers can benefit from the development of rural tourism.

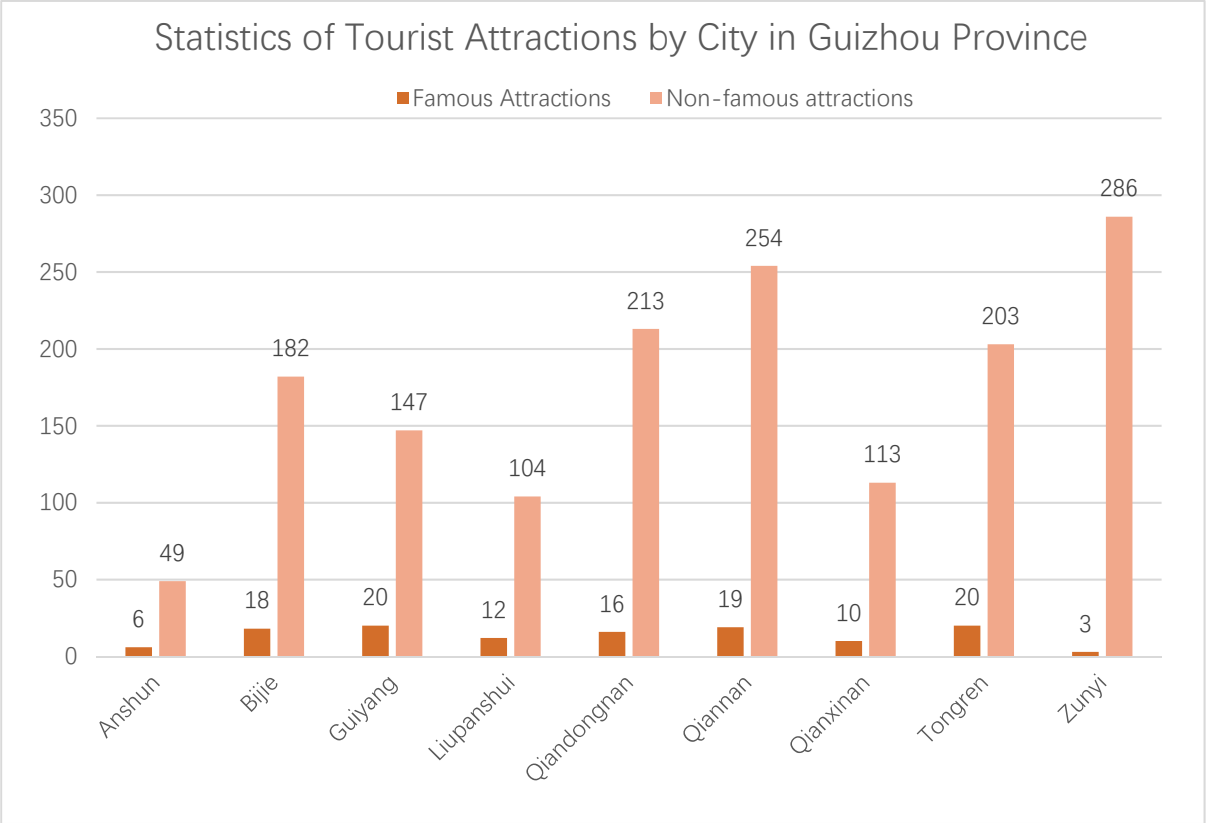

Figure 1. Statistics of Tourist Attractions by City in Guizhou Province

2.2. Research Methodology

At present, more than 90% of the travel platforms on the market, personalized travel need to obtain the user's personal information for data mining and data analysis, if you want to achieve intelligent and accurate analysis of personal recommendations, for the user's "attention", "collection", "search", "browsing preferences" and other network behavior data is the "cornerstone" of personalized recommendations. "search", "browsing preferences" and other network behavior data analysis, is the "cornerstone" of personalized recommendations, and at the same time also meet the need for a large number of users of big data as the algorithm of support for algorithms. However, if a large amount of data is obtained without encryption algorithm, it may lead to the risk of user privacy leakage.

The research involves personalized recommendation, tourism management, privacy information protection and other fields, and is a cross-disciplinary research combining tourism science and information security technology. The research starts from the perspective of "data" and "user", and constructs a personalized privacy protection requirement model based on two aspects: data value and user personalization. Secondly, based on the premise of two algorithms, another personalized recommendation model is constructed considering the privacy protection of users. Finally, game theory is used to reconcile the contradictory relationship between personalized recommendation quality and privacy protection. This project extends the connotation of personalized recommendation service and promotes the theoretical analysis of the interaction between "data value discovery - personalized recommendation service - privacy protection - tourism", and enriches and improves the theories and methods related to data analysis and decision making in the tourism industry.

The theoretical research results are applied to the tourism fields that need development such as data protection, route recommendation, personalized recommendation, information prediction and decision making in big tourism. The application results can not only provide users with personalized recommendations for tourism services, non-heritage cultural experience services, agricultural products purchase services and decision support, and help agricultural merchants understand users' travel preferences and shopping needs. Moreover, it can technically prevent the leakage and abuse of private sensitive information in the data application, so as to effectively play the value of data within a reasonable and legitimate scope, enhance the satisfaction of tourists, and create a satisfactory and safe "Guizhou tour".

3. Research Results and Analysis

An active recommendation system based on a situational awareness model, which focuses on different environmental elements and the interests[12] of the user. For the recommendation of travel itineraries, travel photos[12]of users on social networking sites can be analyzed. The system is based on a Bayesian learning model[14] and user interest models etc[13]. For travel recommendation, from the current research, as most of the recommendation systems rely on user's information, the identification of user's location and the exposure of sensitive information have become the biggest obstacles for the current project. To address this problem, the two algorithms based on the ability to achieve privacy protection of similarity calculation, the accuracy of recommendations and data security.

3.1. A study of content-based information retrieval and information filtering

Content-based recommendation is a technology based on information retrieval and information filtering, and it has become a mainstream personalized recommendation method. The content-based recommendation system is based on the content of text, and provides similar information to users by comparing the similarity of resources and resources, resources and users' interests for comparison.

3.2. Research on collaborative content-based recommendation

By filtering out information using similarity and then recommending users to those with the same preferences and parameter settings, it can also help users discover hidden and undiscovered interests by guessing what information they like.

4. Differential Privacy Algorithm

Based on the basic principle of differential privacy on the genetic matrix decomposition recommendation algorithm for differential privacy and the differential privacy non-negative matrix decomposition recommendation algorithm for sparse data are given to encrypt sensitive information and personalize the recommendation. A key part of the differential privacy algorithm[11] is to decompose the matrices of differential privacy. The main contributions of this paper are as follows: (1) The matrix decomposition is transformed into two alternating user hidden factor and item hidden factor optimization problems, which solve the problems of high dimensionality of the solution space and nonconvexity in the optimization in the solution process. (2) The genetic algorithm variation process is redesigned to improve the solution efficiency by considering the different degrees of user or project bias on the hidden factor; on this basis, an improved exponential enhancement mechanism is proposed, which reduces the influence of interference on the algorithm by using the enhanced exponential mechanism, and also can well balance the privacy protection and algorithm efficiency.

Definition 1: Set up two data sets D and D', with at most one record. Given a random function K, and this function provides ε-differential privacy protection, it holds for all A within the range of values of the function K, i.e., A Range(K). where r[] is the risk of disclosure of the event, controlled by the value of the random function. is the privacy protection budget parameter, the magnitude of the value of which directly determines the level of privacy protection

$$Pr[K(D) \in A] \leq Pr[K(D) \in A] \leq e^{\varepsilon} \times Pr[K(D') \in A] \quad (1)$$

Improved privacy genetic algorithm (APrivGene) input: to improve the rating prediction accuracy, the user rating matrix r is preprocessed, i.e., the boundary parameter is set to B. The ratings are transformed to the range [-B, B] to obtain a new user rating matrix R. Then, the matrix R is decomposed by a hidden factor, i.e.

$$P, Q = arg_{P,Q} min \sum_{(u,i)\in R} (R_{ui} - P_u^T Q_i)^2 \quad (2)$$

4.1. Privacy Protection Budget

From Definition 1, it is clear that the privacy-preserving budget is a key element in order for the inequality to hold. When the value of 0, the level of privacy protection is the highest level at this time, and when the becomes large, the exponential function exp() value monotonically increases and the level of privacy protection gradually decreases. When the smaller the value of the privacy protection level, the lower the availability of the original data. Thus, it seems that the the value of should be taken in context, only then, to ensure a balance between privacy and ease of use for users.

4.2. Noise mechanism

Differential privacy protection is a way to mask sensitive information by adding noise to the results of the query output. Noise mechanism is a method of adding noise to the calculation results. The noise mechanism is the addition of noise to the values. The amount of noise can have a direct impact on privacy protection, so we need a notion of sensitivity.

Definition 2: Sensitivity is a measure of the amount of noise added, for an arbitrary function f: D Rd, a set of data is input, and then a d-dimensional vector of real numbers is output. The data set D and D' differ by at most 1 record, with

$$\Delta f = \underset{D, D'}{Max} \| f(D) - f(D') \| \quad (3)$$

Δf is the global sensitivity of the function f.

5. Personalized Privacy Protection Requirements Model

We study how to quantitatively describe the personalized privacy protection needs of users. A personalized privacy protection model is constructed in terms of "user's associated information categories", "privacy content of information", and "user's privacy protection concern" and their interrelationships.

6. Personalized Recommendation Service

6.1. Genetic matrix decomposition recommendation algorithm

The first step of the genetic matrix decomposition recommendation algorithm for differential privacy preservation is to improve the rating prediction accuracy, and the second step is to decompose the initial problem into two attributes: a vector of hidden factors for solving users and a vector of hidden factors for solving items. The final objective function is listed as follows.

$$\int_P^i (D_i, Q_i) = -\sum_{u\in U_i} (R_{ui} - P_u^T Q_i)^2 = \sum_{t\in D_i} q(t, Q_i) \quad (4)$$

6.2. Differential privacy for sparse data

Based on this project, Jiagu model is selected as a tool for user keyword extraction and sentiment analysis and Spark platform to process large-scale data and perform unified analysis. Because Jiagu model is a domestic open source natural language processing tool, providing a variety of common natural language processing functions, rich API, and easy to operate, high stability, Spark features a unified platform; Spark platform can be used to extract key information to analyze user preferences to create a "personalized The Spark platform can be used to extract key information to analyze user preferences to create

"personalized" travel customization solutions.

Based on the above, our research objective is to provide "personalized" travel recommendation services to all visitors to Guizhou, with the goal of improving the timeliness of users and realizing the value of the data, while satisfying the two algorithms.

7. Research Findings and Future Prospects

7.1. Personalized travel recommendation based on differential privacy

1. According to the attractions or collections browsed by users and the data information of users' friends, collect, process, analyze and get the interest points of users.

2. According to the results of the analysis, it is necessary to recommend to the user tourist attractions that are similar to the user's browsing history in terms of content or other aspects.

3. can not only do a simple recommendation service, personalized recommendation is the most important thing is to pay attention to the user's privacy information leakage problem, need to take into account the user's privacy, in some aspects of the user wants to maintain privacy, the need to use differential privacy protection algorithm to first encrypt information before analysis.

7.2. The first Guizhou province "help farmers" tourism recommendation service

In accordance with the State Development Document No. 2, we will promote the construction of the countryside, and this project will not only help the construction of the countryside through rural tourism, but also provide employment opportunities for rural workers and villagers through this project, and drive villagers to get rich to help revitalize the countryside.

7.3. Research findings

Through the research, we found that, due to the prominent problem of information leakage in the process of using tourism software, this paper combines the "differential privacy non-negative matrix sorting recommendation algorithm for sparse data" and the "genetic matrix decomposition recommendation algorithm to meet differential privacy protection" with the Guizhou "help farmers" tourism personalized recommendation service. and the personalized recommendation service of Guizhou "help farmers", encrypting user information, and analyzing the encrypted user information with Jiagu model and Spark platform, so that users can get accurate travel recommendation contents and do not worry about the leakage of personal privacy when using personalized recommendation service. The personalized travel recommendation service provides users with accurate travel recommendation contents without worrying about the leakage of personal privacy information. The differential privacy personalized travel recommendation can achieve both personalized travel recommendation and protection of user information.

7.4. Future Outlook

In the next continue to strengthen (1) to ensure the correctness and accuracy of recommendations without revealing users' personal privacy and sensitive information; (2) to strengthen the modeling of personalized privacy protection needs of data models and user models, and construct new methods data perturbation methods to enhance the targeting of privacy protection; (3) to introduce perturbation noise for the characteristics of relevant links in the computation process of recommendation algorithms, taking into account privacy protection and recommendation accuracy.

It guarantees the availability of data while ensuring privacy and security. It not only allows users to experience personalized travel recommendation services but also protects users' information security, so that tourists from all over the world who come to Guizhou can realize "easy travel", "safe travel" and "enjoyable travel". The service of "Easy Tour", "Safe Tour" and "Enjoy Tour" is available.

Acknowledgment

The work described in this paper was supported by College Students' Innovative Entrepreneurial Training Plan Program of Guizhou Education University (No.202214223149), the Science and Technology Foundation Project of Guizhou Province (QianKeHeJiChu [2020] 1Y422, QianKeHeJiChu-ZK[2022]YiBan329, and QianKeHeJiChu-ZK[2022]YiBan331)

References

[1] DWORKC,ROTHA.The algorithmic foundations of difrential privacy[J]. Foundations and Trends in Theoretical Computer Science, 2013, 9(3/4): 211-407.

[2] DEWRIR, THURIMELLAR. Exploiting service similarity for privacy in location-based search queries[J]. IEEE Transactions on Parallel and Distributed Systems, 2014, 25 (2): 374-383.

[3] MEDKOVÁJ. Composition attack against social network data[J]. Computers & Security, 2018, 74: 115-129.

[4] Peng, H. L., Jin, K. Z., Fu, C. C., et al. A sequence lattice-based approach to privacy temporal pattern mining[J]. Journal of Electronics, 2020, 48(1): 153-163. PENG H L, JIN KZ, FU C C, et al. Private time series pattern mining with sequential lattice[J]. Acta Electronica Sinica, 2020, 48(1): 153-163. (in Chinese)

[5] Chen S, Fu A-M, Ke H-F, et al. MCDP: A neural network-based approach to multicluster distributed differential privacy data distribution[J]. Journal of Electronics, Acta Electronica Sinica, 2020, 48(12): 2297-2303. (in Chinese)

[6] XIAOXK, TAOYF, CHENMH. Optimal random-perturbation at multiple privacy levels[J]. Proceedings of the VLDB Endowment, 2009, 2(1): 814-825.

[7] KIFERD. on estimating the swapping rate for categorical data[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. new York, USA: ACM, 2015: 557-566.

[8] SUD, CAOJN, LINH, etal. Differentially private Kmeans clustering and a hybrid approach to private optimization[J]. ACM Transactions on Privacy and Security, 2017, 20(4): 1-33.

[9] NGUYENTD, GUPTAS, RANAST, etal. Privacy Aware K-Means Clustering with High Utility[C]//PacificAsia Conference on Knowledge Discovery and Data Min-. cham: Springer, 2016: 388-400.

[10] NGUYEN HH. Privacy-preserving mechanisms for kmodes clustering [J]. Computers & Security, 2018, 78: 60-75. [3]BRAUNHOFERM, RICCI F, LAMCHEB, etal. A context-aware model for proactive recommender systems in the tourism domain[C]// Proceedings of the 17th Intemational Conference

on Human-Computer Interaction with Mobile Devices and Services Adjunct.New York: ACM, 2015: 1070- 1075.

[11] ARASEY，XIEX, HARAT, etal. Mining people's trips from large scale geotagged photos [C]// Proceedings of the 18th ACM International Conference on Multimedia. New York: ACM, 2010.

[12] BRAUNHOFERM, RICCIF, LAMCHE B, etal. A context-aware model for proactive recommender systems in the tourism domain[C]// Proceedings of the 17th Intemational Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct.New York: ACM, 2015: 1070- 1075.

[13] CHENYY, CHENGA, HSUWH. Travel recommendation by mining people attributes and travel group types from community contributed photos [J]. IEEE Transactions on Multimedia, 2013, 15

[14] KURASHIMAT, IWATAT, IRIEG, etal. Travel route recommendation using geotagged photos [J]. Knowledge and Information Systems, 2013, 37(1): 37 -60.