

RESEARCH ON BEHAVIORAL MOTIVATION INFERENCE METHOD FOR GEOGRAPHICAL SPATIOTEMPORAL LARGE DATA

Jingwen Li¹, Wenda Chen¹, Yuan Ma¹, Na Yu¹, Xu Li¹, Jianwu Jiang^{1*}

¹ Guangxi Key Laboratory of Spatial Information and Surveying, Guilin University of Technology, China

Commission VI, WG VI/4

KEY WORDS: Geographical Space-time Big Data, Geographic Context, Association Rule Algorithm, Data Mining

ABSTRACT:

Along with the rapid development of Internet technology, GNSS technology and mobile terminals, a large amount of information including geographical location and time attributes has been generated. Faced with large and complex Internet geospatial data, how to quickly and accurately extract valuable reference information becomes an urgent problem to be solved. And the user's demand for personalized information of recommendation information is getting stronger and stronger, and researching efficient and accurate personalized recommendation system has good application value. In this paper, based on the application requirements of personalized recommendation information, the GIS platform and related recommendation algorithms are used to fully exploit the user and location based on geographic space-time big data. It is divided into user explicit interest and user implicit interest, and then establishes a scientific and efficient user behavior motivation prediction model based on geographic situation. User interest information can be obtained from explicit interest information, implicit interest information and geographic situation interest information. Geographical environment, geographic location and other related context information. By introducing time factors, it is used to update and improve the user real-time interest model to achieve accurate prediction of user behavior motives under geographic spatio-temporal big data. Use Apriori algorithm to calculate the support and determine the current Frequent itemsets of user interest in geographic context, using frequent itemsets to generate strong association rules, and realizing the analysis of user behavior motives based on geography context. For geographic spatio-temporal big data, this paper proposes a personalized hybrid recommendation algorithm, which is based on users. Effective combination of collaborative filtering algorithms and association rules for geographic context-user behavioral interest adaptation.

* Jiang Jianwu - E-mail: fengbuxi@glut.edu.cn

1. INTRODUCTION

1.1 General Instructions

With the rapid development of Internet technology, we are in the era of information explosion. At the same time, with the rapid development of GNSS technology and mobile terminals, a large amount of data including geographical location information is also generated, such as location information, check-in area and time. Information, labels, and evaluations that support people's decisions. By mining these geo-tagged information, users can be provided with personalized user behavior prediction and location recommendation services.

Faced with a large amount of demand information that is frequently updated every day, research on user behavioral motivation prediction and personalized recommendation has emerged. The recommended technology brings together user behavior, data mining technology and human-computer interaction, and provides users with real-time and accurate personalized information services.

User behavior motivation inference is one of the current research hot issues, and a large number of domestic and foreign scholars have put forward their own views. Lee et al. proposed a hotel recommendation system based on geographic location. The rating of the target hotel is mainly based on the decision tree, and the recommendation service of tourism information under specific geographic scenarios is realized. Yap predicts the user's relevant interest information in real time, which improves the recommendation quality in the mobile context. Koren et al. creatively added time factors to collaborative filtering to provide users with personalized, real-time recommendations. Li and others explored the user's geographical location activity record and geographic situation to explore the relationship between users and geographical location, and better provide users with personalized location recommendations. At the same time, the HGSM framework was proposed (Li Q, et al., 2008). Tao TAO Yong-cai et al. designed a recommendation algorithm based on geographical location, which takes into account factors such as user location and reading preferences (TAO Yong-cai et al., 2016). Yuan et al. consider the time factor in terms of modeling, and the method used is to consider the time impact of recommending POI for users in LBSN. They divide the day into 24-hour periods and apply a memory or model-based collaborative filtering recommendation technique to infer the user's preference for location in each time slot (Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. 2013). Madkou et al. proposed a context-aware retrieval model that selects the required services based on functional context information and non-functional context information (Madkour M, et al., 2012). Shiraki et al. found that different context factors can affect the mobile recommendation system to varying degrees. Bao J et al., in the location-based service, divide users into travel experts and ordinary users, and calculate the score and heat of a certain project according to different weights according to the activity and credibility of different users (Bao J et al., 2012). In terms of model building of the recommendation system, Eksombatchai et al. proposed the pixie model, a scalable graphical-based real-time

recommendation system that improves the performance, scalability, and genericity of the algorithm (C. EKSOMBATCHAI, et al., 2017). Li N et al. consider the user's past activity records, demand preferences, and user's social attributes at the same time, and combine the construction models, and then recommend them to users according to the similarity weights (Li N and Chen G, 2009). Falcone D. and others at the University of Cambridge Computer Laboratory used blog posts published by Twitter users to consider the content of blog posts and time and space factors, and established a supervised learning model to determine the type of location of the user at the time (Falcone D, Mascolo C, Comito C, et al., 2019). Choi et al. introduced a bidirectional clustering recommendation algorithm that reduces space and improves recommendation quality with informational interpretability (S. CHOI, et al., 2017).

However, in many studies, there is a default phenomenon: the change of user interest depends only on historical behavior. However, in actual situations, the user's interest preference will be affected by many factors, such as the user's age and the rich experience. The influence of own factors, the geographical factors, the environment, the trend of the public and other situational factors, and the user's interest preferences will increase or decrease or even change over time, which requires a comprehensive consideration of the user's interests.

2. USER BEHAVIOR PREDICTION DATA ABSTRACT EXPRESSION METHOD

2.1 Data Preprocessing and Data Mining

In the process of user behavior prediction, complex heterogeneous data is filled. For example, user evaluation, browsing history, geographic situation, time stamp and other data are usually not directly used. This requires preprocessing the data to convert the unprocessed data into Suitable for analysis. Data mining originated in a variety of disciplines, the most important of which is statistics and machine learning. Statistics originated in mathematics, which emphasizes the accuracy of mathematics. Machine learning mainly originated from computer practice, which is more inclined to practice, and actively detects something to determine its manifestation. As an important branch of computational science, it is an important process of transforming massive data into knowledge. At the same time, as a carrier for storing data, the process of discovering knowledge from it is also an indispensable step. Data mining is the processing and mining of massive data to be processed. It not only solves the redundant data in the database, but more importantly, it can obtain valuable information hidden by the accumulated massive data.

Geospatial data contains geospatial information that is not available in general data. Therefore, geospatial data mining for geographic spatiotemporal data is different from the past and has more significant requirements. The requirements of spatial data mining are more extensive. The scale of space is added on the basis of ordinary data mining, in order to mine and analyze the data related to geographical context. Spatial data mining data often comes from geographic space-time big data, with huge

computational complexity and higher dimensions. The spatial data mining architecture is shown in Figure 1.

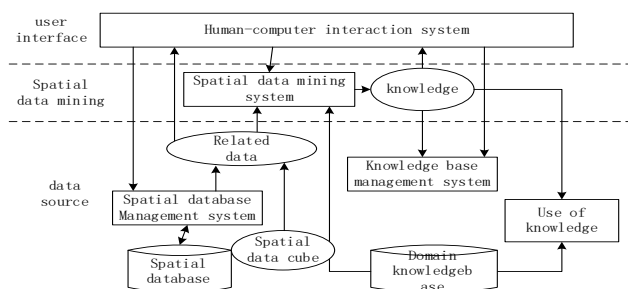


Figure 1. spatial data mining architecture

2.2 Data organization structure

There is a huge amount of geographic space-time big data on the Internet, and we need to organize and store it reasonably to support user behavior prediction and personalized recommendation. At the same time, in the recommendation process, the personalized recommendation system also performs data query on each function module, and the record information of the user in the recommendation process also needs to be stored, and the database is an essential component of the recommendation system. The design of the database takes into account the type and size of the data, and then designs from the conceptual, logical and physical levels. Finally, the implemented database needs to be evaluated and sent for inspection and maintained and updated from time to time. This will ensure the effective operation of the database. As shown in Figure 2.

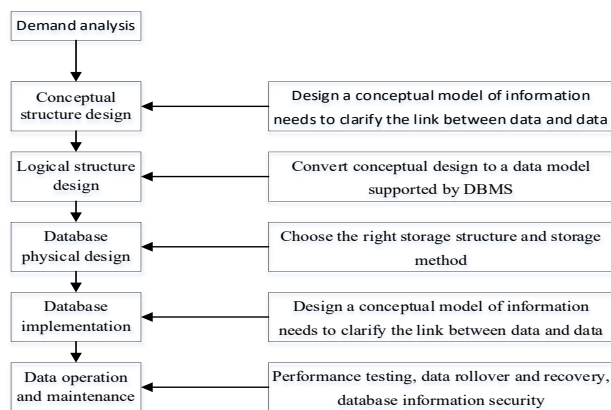


Figure 2. Database Design

3. CONSTRUCTION OF USER REAL-TIME BEHAVIOR MOTIVATION PREDICTION MODEL

User real-time interest modeling is a process of inferring and predicting information related to interests and needs through user history or historical data similar to the user to describe the preferences and preferences of the user pair. Whether the recommendation system can complete the personalized

recommendation that satisfies the user is whether the model can accurately predict the user behavior.

3.1 User interest information acquisition and representation

In the construction of user real-time interest model, the user's personal basic information and historical behavior are the most direct data sources, but such data and its scarcity, need to obtain relevant data in other aspects to support the establishment of the model. as the figure 3 shows:

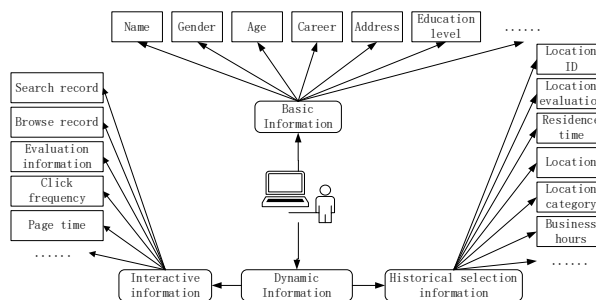


Figure 2. User interest information

For explicit and implicit feedback techniques, the same results can be divided into two categories. User explicit interest is called explicit interest, because the data source is the basic information of the user filled in the login account interface, and the submitted interest preference data, the extracted information has a clear meaning and explanation.

Compared with the user's explicit interest, as the name suggests, the user's implicit interest is because the information extracted through the mining has no clear meaning and explanation, which is due to the influence of the user's explicit interest, such as physical condition, mood change and other factors. Geographical factors such as geographical location and environment, these factors will also have different degrees of impact on users' interests.

3.2 Geographical Situation - User Behavior Interest Correlation Adaptation

First law of geography, All things are related, but nearby things are more related than distant things. Based on this law, when the distance between the user and the event location is large, the user's interest ratio is lowered. Therefore, in the motivation of user behavior motivation, not only user interest information is needed, but also we need to integrate various related heterogeneous information, including common user interest information such as user explicit interest information and user implicit interest information, geographical location information. Geographical spatiotemporal data such as time attribute information also needs to be considered to construct a recommendation algorithm based on multi-source heterogeneous information modeling. The specific operation is to establish an association rule between geographic context and user interest,

and use an algorithm to match geographical space-time location and user behavior.

The Apriori algorithm is a classic algorithm widely used in mining association rules between data. The core of the Apriori algorithm is iteration, which is a recursive algorithm based on frequent set search and frequency set rule generation process. The goal is to find the largest K item frequent set. The basic flow of the Apriori algorithm is: 1 By identifying the entire data set, all the frequent itemsets are calculated, and the frequent itemsets satisfying the minimum support are obtained. 2 Based on the rule that generates the frequent itemsets in the previous step, recalculate the support degree of each item. If the item is pruned. The final result is a frequent K-1 item set plus a non-empty K item set.

The confidence level is set to indicate the degree of matching of the user's behavioral interests in the current geographical context, and the probability of the simultaneous occurrence of the items A and B is abstracted. The description item contains the ratio of the number of items of A and B to the number of all items, such as Equation 1:

$$Support(A \Rightarrow B) = P(A \cup B) = \frac{Support_count(A \cup B)}{Total_count} \quad (1)$$

In the above formula, $Support_count(A \cup B)$ represents the number of items in the item set containing both A and B, $Total_count$ represents the number of items in the item set, and $P(A \cup B)$ represents the probability that both items A and B are included in the item set, and the association rule is established by using the confidence level, which can be expressed by the following Equation 2

$$Confidence(A \Rightarrow B) = P(B|A) = \frac{Support_count(A \cup B)}{Total_count(A)} \quad (2)$$

4. PERSONALIZED RECOMMENDATION ALGORITHM FOR USER INTEREST BASED ON GEOGRAPHIC CONTEXT

According to the user's geographical situation and user behavior, the association analysis is carried out to understand the situational features in the current geographical situation, and the adaptation association is based on this. Setting the geographic context GeoContext contains four types of scenes: A, B, C, and D. The user items include five types of user interests: e, f, g, h, and i. The user interests of the four types of scenes are: $A \rightarrow \{e, g, h\}$, $B \rightarrow \{f, g, i\}$, $C \rightarrow \{e, f, g, i\}$, $D \rightarrow \{f, i\}$. In this paper, the

Apriori algorithm is used to perform loop iteration and to find frequent k-item sets layer by layer.

Call the createC1 function to count the user interest information corresponding to the four types of scenes, call scanD to generate L1 to describe the number of occurrences of each interest, and delete the number of occurrences less than the given minSupport. The aprioriGen function is then called to generate the binomial C2. As shown in Figure 4.

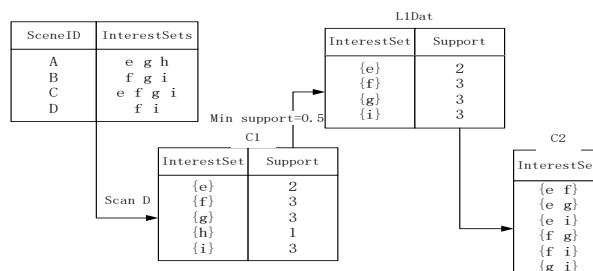


Figure4. Geographical Scene--User Behavior Mode

Compared with the collaborative filtering algorithm, the association rule analysis does not need to calculate the adjacency matrix of the two similarities. The correlation rule algorithm has relatively small computational complexity, and the association rules have the support degree, confidence and lifting degree to make the association rule analysis Strong interpretability, but the analysis of association rules has preconditions, and each data must have a certain relationship in order to carry out corresponding analysis. As shown in Figure 5.

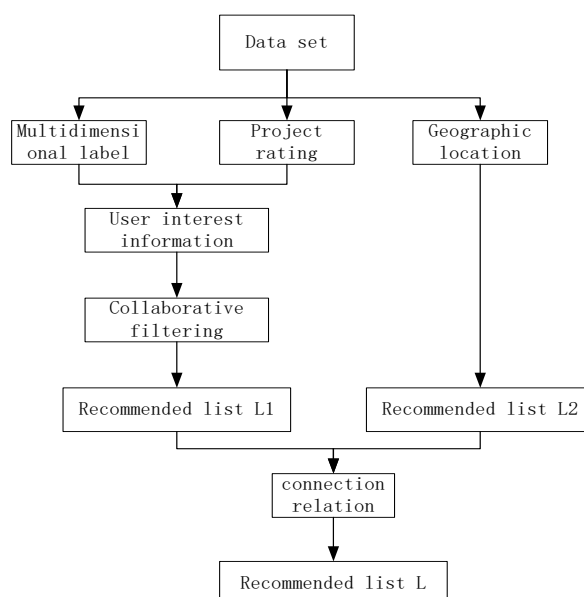


Figure5. Recommended algorithm flow

5. CONCLUSION

Personalized recommendation uses data mining and other related technologies to analyze the user's interest preferences and provide targeted personalized recommendations for users. Considering the fact that traditional personalized recommendation technology has certain defects due to less involvement of geographical context elements, the singularity of traditional collaborative filtering recommendation, and the flexibility of project resource flexibility, this paper designs a real-time interest of users who integrate geographic situations. Adaptation model, then a personalized hybrid recommendation algorithm based on user-based collaborative filtering algorithm and adaptive context association rules of geographic context-user behavior interest.

REFERENCES

- Bao J, Zheng Y, Mokbel M F. 2012: Location-based and preference-aware recommendation using sparse geo-social networking data International Conference on Advances in Geographic Information Systems. ACM, 199-208.
- C. EKSOMBATCHAI, P. JINDAL, J. Z. LIU, et al. 2017: Pixie: A System for Recommending 3+ Billion Items to 200+ Million Users in Real-Time.
- Falcone D, Mascolo C, Comito C, et al. 2015: What is this Place? Inferring Place Categories through User Patterns Identification in Geo-tagged Tweets 2014 6th International Conference on Mobile Computing, Applications and Services (MobiCASE). IEEE Computer Society, 10-19.
- Li N, Chen G, 2009: Multi-layered friendship modeling for location-based mobile social networks. Mobile and Ubiquitous Systems: Networking Services, MobiQuitous. MobiQuitous'09. 6th Annual International. IEEE, 1-10.
- Li Q, Zheng Y, Xie X, et al. 2008: Mining user similarity based on location history ACM Sigspatial International Conference on Advances in Geographic Information Systems. ACM, 1--10.
- Madkour M, et al. 2012: Context-aware service retrieval in uncertain context. In Proceedings of International Conference on Multimedia Computing and Systems, ICMCS. IEEE, 611-616.
- Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. 2013: Magnenat-Thalmann, "Time-aware point-of-interest recommendation," in Proc. 36th ACM Annu. Int. SIGIR Conf. Res. Develop. Inf. Retrieval, pp. 363–372.
- S. CHOI, H. HA, U. H. WANG, C. KIM, J. W. HA. 2017: Reinforcement Learning based Recommender System using Biclustering Technique. In Proceedings of Workshop on Multidimensional Information Fusion for User Modeling and Personalization (IFUP'18). ACM.
- TAO Yong-cai¹, LI Jun-yan¹, SHI Lei¹, WEI Lin². 2016: Hybrid Location-based Personalized News Recommendation. Journal of Chinese Computer Systems, 37(5): 943-947.