Personalized tourism information push algorithm based on smart tourism big data analysis model

Hua Zhang
Party School of Shaanxi Provincial Party Committee (Shaanxi Administration College), Xi'an, Shaanxi, China

ABSTRACT

With the development of society and the improvement of people's living standards, tourism is becoming a new way of life. Anyone can take a trip as soon as they say, and the era of national tourism and mass tourism has arrived. Moreover, with the rapid popularization of information technology, when making travel plans, users will generally give priority to searching relevant travel information through search platforms. Personalized travel intelligent push service is mainly based on the comprehensive analysis of user basic information, behavior preference information and scene feature elements to build an interest model for each user, and then push the corresponding personalized service. This paper proposes a smart tourism information push system based on the cloud platform, and studies and solves some key issues involved in the system. In view of the massive data processing problem in the tourism system, the massive data of tourism information is processed through the powerful computing power of cloud computing.

Key words: Smart tourism; Big data analysis model; Personalized tourism; Information push

1. INTRODUCTION

The concept of "smart earth" was put forward by the company, and then "smart city" was derived from it, which was regarded as a brand-new concept and exploration path to promote the process of urban intelligence [1]. With the help of emerging information technology, we can perceive, capture, transmit and process information anytime and anywhere, which can not only manage the city in a refined, intelligent and comprehensive way [2]. Improve public service facilities and urban functions, create a safe, convenient, efficient and environmentally friendly environment, and enable people to work, study and experience more intelligently. Based on this, it is urgent for experts and scholars at home and abroad to analyze and solve how to scientifically obtain the information that users are really interested in from massive resources [3].

The recommendation system has been studied for a long time. Because of its good practicability and potential commercial value, it has been pursued by scientific researchers and business people. Therefore, its development has never stopped. From popular recommendation to personalized recommendation, it has been widely used [4]. During this period, a wide variety of recommendation technologies were produced. These technologies have their own characteristics and different effects. Each technology has its own advantages and disadvantages. Today's Internet information has undergone earth-shaking changes compared with the past, and the capacity of information is increasing exponentially. There are numerous websites about tourism on the Internet. Whenever users search for relevant tourism information through the search engine, their search results are a large number of advertisements and tourism consulting. It is difficult for tourists to find the information they care about in these information, which increases the difficulty of tourists' information query [5].

*2858573809@qq.com

Seventh International Conference on Mechatronics and Intelligent Robotics (ICMIR 2023), edited by Srikanta Patnaik, Tao Shen, Proc. of SPIE Vol. 12779, 1277919 · © 2023 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.2688887

At present, there are many smart travel search engine, such as where to go and Ctrip. These search engines are aimed at the vast number of users. When users use search engines to retrieve tourism information, if the search engines can't find out the important information that users need and put it in the front position, it will greatly affect users' satisfaction with the search engines [6]. The data accurately reflect where the tourist market of this city is and which products are concerned by tourists, which provides important data support for precision marketing and has a subversive effect on decision-making marketing. Although big data is a good thing and its application prospect in the tourism industry is very broad, there are still great obstacles in the application of big data in the whole industry [7]. These obstacles come from data collection, but also from data analysis and mining. When pushing the tourism information by using the regular rotation algorithm, the client regularly consults the server, and determines the tourism information to be pushed according to the historical information of the user in a period of time, ignoring that the user's interest in tourism will change, resulting in the push result not satisfying the user [8].

Therefore, this paper combines information recommendation, cloud computing, information push and other technologies, and proposes a framework of intelligent tourism information push service system based on the cloud platform based on the previous recommendation system based on a single machine, and carries out theoretical research on cloud computing, recommendation system, multiple recommendation algorithms, and information push technology.

2. THE BASIC THEORY OF TOURISM SERVICE RECOMMENDATION

2.1 Personalized recommendation system

The construction of user interest model is the key link of personalized information push application and the basic part of providing personalized service for users. The quality of user modeling directly determines the level of personalized push service [9]. Finally, a computer-readable model is designed and constructed, which can accurately express users' interest preferences and effectively obtain the changes of their interests. In all kinds of personalized recommendation service platforms, only by establishing an accurate user interest model can we carry out the next project matching, intelligent push and other related work [10]. The process of interest modeling is shown in Figure 1.

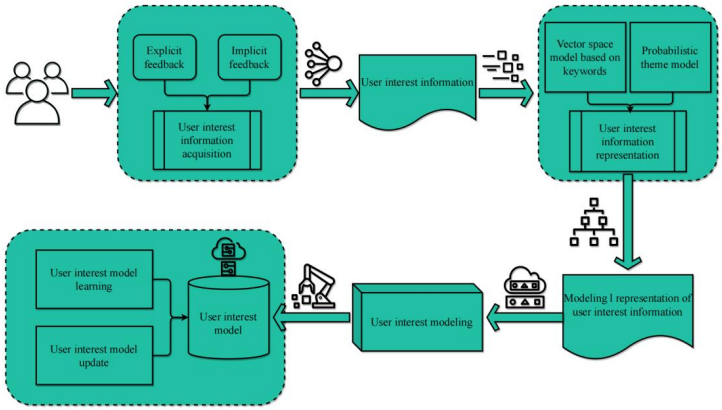

Figure 1. Basic flow chart of user interest modeling

In terms of extracting user feedback content, there are currently two different technical means to record various types of user feedback data. One is to obtain user preferences through direct evaluation of projects by users, which is called explicit feedback technology; The other is that the user's actual evaluation and active participation are not required in the whole process. Through the system's analysis and monitoring of the user's behavior process, this kind of technology is called implicit feedback technology.

In order to discover users' potential knowledge and interests, personalized recommendation system is highly practical. Facing massive data, it is difficult but valuable to help users find information with high inclination. Personalized recommendation makes the information stored in the website transparent to users, and users only need to judge the recommended information. For each user, it will provide targeted solutions or potentially needed solutions, unlike the

general travel service platform, which can only complete transactions online and only provide such popular services. Good recommendation results will greatly enhance the satisfaction of tourists. After purchasing infrastructure services, users can directly build their own platforms and applications on the infrastructure layer. Infrastructure layer provides users with virtual software and hardware resources, which can be dynamically allocated according to different needs of users. If users' software systems can run on the computing resources provided by infrastructure layer, users can transplant their software systems to services to reduce the hardware burden caused by users' software system operation and maintenance. In the cloud computing environment, tourism information self-organization push algorithm, as the core technology of the new generation of browsers, adopts corresponding technical guidelines and relevant protocols, and can self-organize and deliver the required tourism information to users on a regular basis in the Internet. For the implementation of adaptive self-organization push algorithm of tourism information, the most important thing is to actively push the latest and most relevant tourism information to users according to certain categories according to users' needs, so as to improve the efficiency of people's access to tourism information.

2.2 Definition of big data analysis

With the advent of the cloud era, big data has also attracted more and more attention. "Big data" needs a new processing mode to have more insight, decision-making power and massive and diverse information with high growth rate and process optimization ability. The significance of big data technology is not just to master huge data information, but to refine and specialize these data with various meanings. In other words, if big data is compared to an industry, the key to achieving profit in this industry is to improve the "reprocessing ability" of the original data, and realize the "value-added and efficiency" of the data through "repeated processing". In addition to the distributed file system, Hadoop is also accompanied by MapReduce architecture for large data set processing. According to the authoritative report, many enterprises have begun to use or evaluate Hadoop technology as the standard of their big data platform. The Hadoop ecosystem is shown in Figure 2.

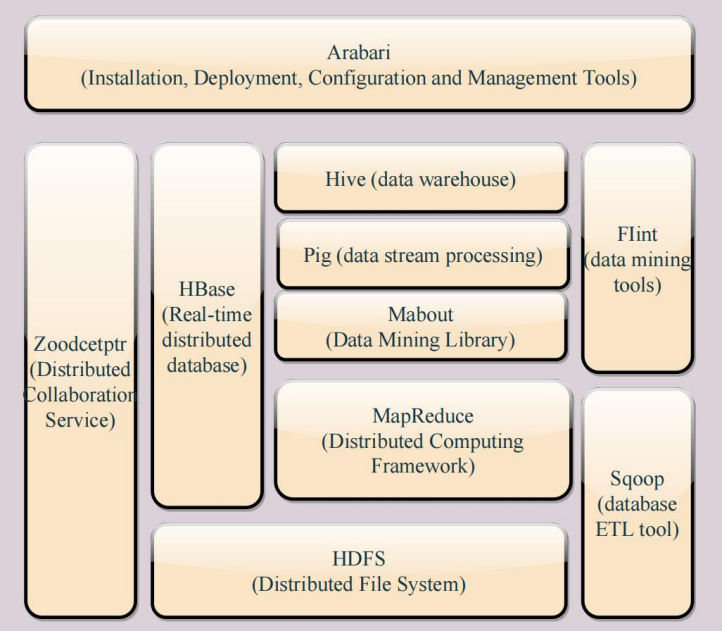

Figure 2. Hadoop ecosystem diagram

At present, Hadoop can be said to have become the de facto standard in the field of industrial big data, which is mainly represented by Yahoo, Facebook, eBay, IBM, etc. abroad; In China, Internet companies such as Baidu, Tencent and Ali are the main companies. Hadoop is an open source software, and the use and improvement iterations of these big companies further improve and promote the further development of Hadoop.

The remaining machines are DataNode. Namenode generally does not store data. It stores metadata and information about the storage and distribution of small data files. When computing needs to be performed, the master is responsible for scheduling all tasks that constitute a job and assigning these tasks to each slave. And the master monitors the execution of

each node. For the heartbeat test package sent back by the slave, the master judges whether the task is completed or not, and reassigns it to other nodes to execute the failed task.

3. ANALYSIS OF PERSONALIZED TOURISM INFORMATION PUSH ALGORITHM BASED ON SMART TOURISM

3.1 Algorithm model of personalized tourism information push

In the tourism information service platform, there are many kinds of tourism information, which can be roughly divided into tourist attractions information service, traffic information service, hotel reservation information service, air ticket reservation information service and so on. According to the opinions of the group with the same hobbies as the target user, the interested travel information is recommended to the target user, the similarity between the target user and other users is calculated, the neighbor set closest to the target user is determined according to the calculated similarity, and the favorite travel information of the closest neighbor is pushed to the target user. At the beginning of the system operation, users can add their own interesting content manually, and then initialize the model. In the process of use, users can also add their own interesting information, which is the explicit feedback of user information. Usually, the main way for users to browse the web is to browse at will, without searching for a specific goal. For example, in the selection of tourist attractions, some users prefer natural attractions such as Guilin mountains and waters, Qinghai Lake, Huangshan Mountain and other natural scenery, while some users prefer cultural landscapes such as the Forbidden City, the Terra Cotta Warriors and other historical sites; In the hotel reservation service, users with good economic conditions prefer to choose star hotels, while some users with ordinary economic conditions prefer to choose ordinary hotels. Therefore, users and tourism information services have their own context information. The user's interest in tourism information will change over time. Fuzzy sets can deal with the complicated relations such as time-varying and uncertainty based on the characteristics of self-adaptation and self-organization, so fuzzy sets theory is used to deal with the time-varying problems in the process of tourism information push. According to the collected user travel information, the corresponding evaluation index is constructed, the evaluation index is fuzzed, and then the weighted average anti-fuzzy processing is carried out, so that the user's weight output can be determined. The formula is described as follows:

$$f(x|\omega)=\frac{\sum_{i=1}^{n}\omega_i E_i(x)U_i a_i}{\sum_{i=1}^{n}\omega_i E_i(x)U_i} \tag{1}$$

In formula (2), ω_i is used to describe the rule weight; $E_i(x)$ is used to describe the ambiguity of the input value x; U_i is used to describe the volume of the set; a_i is used to describe the set weight.

$$k\times(k-1) \tag{2}$$

Its value is the ratio of the actual total number of links to the theoretical total number of links, reflecting the close degree of links between all nodes, and is given by the tourism network density formula (3).

$$D=(2\sum_{i=1}^{k}d_i(n_i)/(k*(k-1))) \tag{3}$$

Among

$$d_i(n_i)=\sum_{j=1}^{k}d_i(n,n_j) \tag{4}$$

In order to realize the parallel algorithm of the recommended algorithm, we must first find the elements that can be used for parallel computing. According to the standard process of parallel computing, it includes the disassembly of large tasks, the parallel input of input data, and the merger of operation results. There are two kinds of big tasks, one is that the tasks are completely independent of each other, so that the parallel algorithm can be used directly, and the other is that the small tasks are interdependent. Such tasks can be decomposed recursively into subproblems, and the tasks that can be calculated

in parallel can be designed in the subproblem so that they can be used in parallel mode, and the parts that cannot be parallel can be distributed to individual computers for serial execution. Through analysis, the attribute information in the original information is extracted, such as spatial location, scenic spots and hotels. The extracted attribute information contains a lot of related structured and unstructured data. Among them, structured data generally has a relatively clear meaning for describing the project, and its value often falls within a limited range, such as age, date, etc. Unstructured data is not clear and clear when describing the project, and cannot be used directly, but needs to be converted into structured data, After transformation, it can be used in the model. Finally, the system will list a list of items that users are likely to be interested in. This is achieved by performing the corresponding cosine similarity analysis between the user's prototype vector and the product vector.

3.2 Analysis of experimental results

The experiment was carried out on a tourism website, which contains a lot of tourism information. Because of the huge information scale of the website, cloud computing technology was adopted. 92 users who registered the website were randomly selected as the research objects, and according to the historical tourism information searched by users, self-organized travel information was pushed for them. In the cloud computing environment, all the tourism information that meets the requirements of users is queried, and the tourism information is actively pushed to users through the algorithm in this paper. According to users' preferences and behavior habits, the tourism information is divided and pushed to users, so as to enhance the push quality of tourism information. Tourism information mainly includes five indicators: tourism location, price, evaluation, type and time. The weight of each indicator can be given corresponding weight through the historical information of the research object. Table 1 describes the weight given to each indicator.

Table 1 Weights of indicators

Name	Position	Price	Evaluate	Type	Time
Weight	0.15	0.22	0.18	0.30	0.15

In order to verify the quality and timeliness of the algorithm in this paper, Bayesian algorithm and neural network algorithm are compared, and the hit rate of push, DOA standard and real-time performance of tourism information push are used as evaluation indicators for experimental testing. The spatiotemporal distribution pattern of the selected information is statistically analyzed. With the number of tourists as an indicator, the time distribution characteristics of the inbound tourist flow in Xi'an are studied, and the changes of the number of tourists with the month, the number of tourists and the time of stay, and the number of tourists and the number of accompanying tourists are separately counted. Run the recommended algorithm on a single machine and a cloud platform with virtual machines. The calculation items are: 10000 - 50000 pieces of data, and the settlement results are shown in Figure 3.

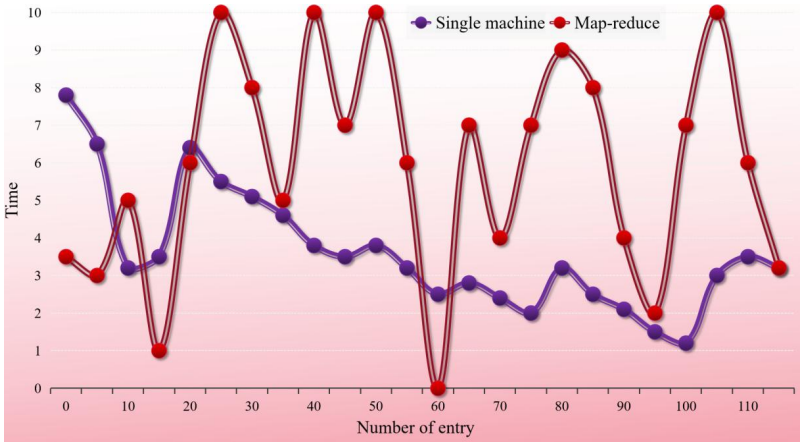

Figure 3. Comparison of efficiency of recommendation algorithms in stand-alone and cloud platforms

Run the recommendation algorithm on the cloud platform with 3-7 virtual machines respectively, and the running data is 20,000 projects, and the running time is as shown in Figure 4.

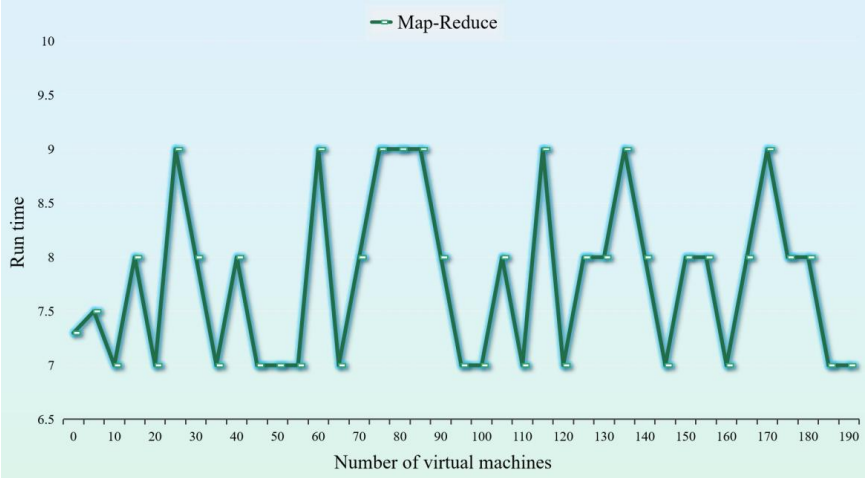

Figure 4. Comparison of operating efficiency of recommendation algorithms with different virtual machines

From the above algorithm operation efficiency analysis chart, it can be seen that the recommended algorithm running on the cloud platform is much more efficient than that running on a single machine.

Through the comparison of the above experimental results, it can be concluded that the search accuracy of the optimized SM-Page Rank algorithm is higher than that of the traditional Page Rank algorithm, and the secondary sorting based on the user interest model makes the search results more suitable for the needs of users. In general, the execution efficiency of the algorithm tends to be proportional to the number and scale of cluster nodes. The advantage of using cluster is that the larger the amount of data involved in the calculation, the more performance advantages can be reflected. In addition, due to the limited memory space of a single machine, recursive operations need to consume a lot of memory, and large-scale I/O operations will inevitably cause running waiting or blocking. However, MapReduce operation mode is adopted. Due to the design characteristics, data is allocated to multiple Datanodes by Namenodes for operation, jobs are allocated to each node, and the working pressure of the program is shared. In addition, Map/Reduce is two separate operation processes, In terms of I/O operations, MapReduce costs relatively small. By adopting MapReduce to parallelize the FP-Growth algorithm, the running time is reduced and the running performance is guaranteed.

4. CONCLUSION

The era of big data information has arrived, and people's lives are increasingly dependent on the Internet, so search engines have become one of the main ways for people to obtain tourism information. By comprehensively considering and analyzing users' multi-dimensional characteristic information, and combining users' basic characteristic dimensions and historical behavior dimensions, a multi-dimensional preference model of users in tourism scenarios is constructed. At the same time, the weight of each dimension is calculated by algorithm analysis, and the calculation method of the weight of each dimension is explained by an example. It is a relatively new research idea to implement tourism service recommendation on the cloud platform, which not only experiments the storage and calculation of big data, but also enhances travel information. In this paper, the traditional smart tourism information service based on a single machine is migrated to the cloud platform, which can effectively and rapidly process the increasingly complex tourism information big data through the powerful processing capability of cloud computing. The experiment shows that compared with the classical algorithm, the SM-PageRank algorithm and the secondary ranking based on the user interest model proposed in this paper make the information query more reliable and accurate, and can meet the user's search needs.

ACKNOWLEDGMENTS

This work is supported by Key Research and Development Project of Shaanxi Province(2022GY-326)

REFRENCES

[1] Jeong, M., Shin, H. H., Tourists' Experiences with Smart Tourism Technology at Smart Destinations and Their Behavior Intentions:. Journal of Travel Research, vol.59, no.8, pp.30 (2020).

[2] Koo, C., Cantoni, L., Information processing and management special issue on informatics/data analytics in smart tourism editorial information processing and management xxx (xxxx) xxx-xxx. Information Processing & Management, vol.57, no.1, pp.22 (2019).

[3] Torabi, Z. A., Shalbafian, A. A., Allam, Z., et al., Enhancing Memorable Experiences, Tourist Satisfaction, and Revisit Intention through Smart Tourism Technologies. Sustainability, vol.14, no.3, pp.26 (2022).

[4] Lima, M., Emerson CleisterDandolini, Gertrudes AparecidaBiz, Alexandre AugustoRibeiro, Alessandro Costa. Customer knowledge management and smart tourism destinations: a framework for the smart management of the tourist experience - SMARTUR. Journal of knowledge management, vol.25, no.5, pp.18 (2021).

[5] Jeong, M., Shin, H. H., Tourists' Experiences with Smart Tourism Technology at Smart Destinations and Their Behavior Intentions. Journal of Travel Research, vol.44, no.10, pp.37 (2019).

[6] Lee, P., Hunter, W. C., Chung, N., Smart Tourism City: Developments and Transformations. Sustainability, vol.58, no.9, pp.27 (2020).

[7] Pai, C., Kang, S., Liu, Y., et al., An Examination of Revisit Intention Based on Perceived Smart Tourism Technology Experience. Sustainability, vol.13, no.2, pp.50 (2021).

[8] Liu, S., Ma, X., How Social Networks Affect the Spatiotemporal Planning of Smart Tourism: Evidence from Shanghai. Sustainability, vol.48, no.12, pp.47 (2021).

[9] Xiang, Z., Stienmetz, J., Fesenmaier, D. R., Smart Tourism Design: Launching the annals of tourism research curated collection on designing tourism places. Annals of Tourism Research, vol.8, no.6, pp.36 (2021).

[10] Lan, F., Huang, Q., Zeng, L., et al., Tourism Experience and Construction of Personalized Smart Tourism Program Under Tourist Psychology.. Frontiers in psychology, vol.18, no.6, pp.21 (2021).