Research on Multi-Dimensional Dynamic Clustering Method of Big Data Alliance Users

Xiaoxiao Wang
School of Economics and Management, Harbin University of Science and Technology, Harbin, 150080, China

Lili Zhai
School of Economics and Management, Harbin University of Science and Technology, Harbin, 150080, China

Yanling Hu
School of Economics and Management, Harbin University of Science and Technology, Harbin, 150080, China

Shuchen Zhang
School of Economics and Management, Harbin University of Science and Technology, Harbin, 150080, China

Recommended Citation

Wang, Xiaoxiao; Zhai, Lili; Hu, Yanling; and Zhang, Shuchen, "Research on Multi-Dimensional Dynamic Clustering Method of Big Data Alliance Users" (2020). *WHICEB 2020 Proceedings*. 29.
https://aisel.aisnet.org/whiceb2020/29

Research on Multi-Dimensional Dynamic Clustering Method

of Big Data Alliance Users

Xiaoxiao Wang, Lili Zhai, Yanling Hu, Shuchen Zhang

School of Economics and Management, Harbin University of Science and Technology,

Harbin, 150080, China

Abstract: In order to improve the clustering accuracy of big data alliance users, this paper studies users' dynamic clustering based on their multi-dimensional attributes. First of all, the user profile of big data alliance is constructed from five dimensions of user basic attribute, domain attribute, preference attribute, social attribute and value attribute. And the K-means algorithm is used to cluster user profiles to complete the initial clustering. Then, based on the group user profile, combined with the user's recent dynamic behavior data, the FCM algorithm is used to achieve secondary clustering. Finally, the proposed user clustering method is tested by recommending data resources to the clustered user groups. The experimental results show that the user clustering method proposed in this paper has higher accuracy and lower error rate.

Keywords: big data alliance users, multi-dimensional attributes, dynamic clustering

1.   INTRODUCTION

With the development of cloud computing, Internet of things, artificial intelligence and other new generation of information technology, the huge value contained in the huge volume of data covering all walks of life has been gradually discovered and applied. In order to promote the sharing and application of data resources, and avoid the risks brought by information islands, the establishment of big data alliance has become an important means to promote the development of big data. Big data alliance provides users with corresponding data resources in time and accurately by connecting the upstream and downstream enterprises of big data industry for data fusion, which lays a solid foundation for the realization of alliance profits. However, the volume of users in big data alliance is huge, providing corresponding services based on users' real-time data will not significantly improve service efficiency, and will generate colossal resource consumption and high alliance operation costs. Therefore, in order to improve the efficiency of the alliance's management of users, it has become an inevitable trend to cluster users of the big data alliance. As we all know, the most commonly used user clustering method is the K-means algorithm, but users belong to only one specific group when using this method, which has some limitations. Existing researches on user clustering mainly focus on web user clustering, and get similar information needs from user behavior data. In addition, some scholars studied user clustering from such aspects as the time attributes of user preferences [1], various factors affecting user clustering, dynamic user clustering [2], and improvement of user clustering methods [3]. The existing research results of user clustering provide strong support for the research of big data alliance user clustering.

2.   MULTI-DIMENSIONAL DYNAMIC CLUSTERING OF USERS IN BIG DATA ALLIANCE

Big data alliance user clustering mainly consists of the following steps:

(1) User data collection and processing. The user data of big data alliance includes the basic information filled in by the user during registration and the behavior data left after the user operation. By integrating user data from alliance members and alliance platforms, the user registration data and behavior data are analyzed and processed to obtain user multi-dimensional attributes.

(2) Users are clustered for the first time. In order to fully describe the characteristics of users, we divide

the attributes of user behavior data based on the characteristics of user behavior data of big data alliance. When users browse and purchase data resources, they first pay attention to the domain of the data resource, and look for data resources that match their preferences in the corresponding domain. In order to understand the real situation of data resources, users will obtain detailed information by browsing the reviews or interacting with other users. After the corresponding data resource is determined, the big data alliance platform will evaluate the value of the user based on the user's visit and purchase behavior. Therefore, this paper divides user attributes into basic attribute, domain attribute, preference attribute, social attribute and value attribute. Based on the user's multi-dimensional attributes, a user profile of the big data alliance is constructed, and similar user profiles are aggregated by K-means algorithm to form a group user profile to complete the initial clustering.

(3) Second user clustering. The user profile of big data alliance is constructed without considering the timeliness of user data, and the user need changes dynamically with time, which result in the inability to provide accurate data resource services for the user groups only divided by the group user profile. Therefore, this paper uses the user's recent behavior data to mine the user's recent needs, and dynamically classifies the users in the same group profile according to the recent needs, so as to improve the accuracy of the clustering results. The user's recent behavioral data can be divided into two categories based on the explicitness and implicitness of user behavior. One is the explicit data left when users purchase and evaluate data resources, and the other is the implicit data left when users browse and search data resources. By analyzing and processing the user's explicit and implicit behavior data, users' recent needs are obtained. This paper uses the fuzzy c-means clustering method (FCM) to cluster users under the same group user profile according to the user's recent needs, and completes the second division of user groups.

3. CONCLUSIONS

The study of big data alliance users clustering is beneficial to improve the management efficiency of users in the alliance. However, the traditional user clustering methods do not consider the comprehensive information of users and the timeliness of user data, which result in unsatisfactory results. This paper integrates the user data shared by alliance members and the new user data after the alliance is formed, constructs user profiles from five dimensions of the user's basic attribute, domain attribute, preference attribute, preference attribute and value attribute, and aggregates similar individual user profiles to form group user profiles to complete the initial clustering. On the basis of group user profile, users are clustered again based on the user's recent explicit and implicit needs. Clustering users of the big data alliance through the above method can ensure the long-term preferences and recent needs of users in the same group are highly similar, improve the accuracy of clustering results and the efficiency of alliance management of users, clarify the use of alliance data resources and adjust the data resource layout in time to maximize data resources usage efficiency.

REFERENCES

[1] Wang Haiyan, Xiao Yikang. (2018).Dynamic group discovery method based on density peak clustering. Computer Research and Development, 55(02): 391-399.

[2] Liang Shangsong, Ren Zhaochun Zhao Yukun. (2017). Inferring dynamic user interests in streams of short texts for user clustering. ACM Transactions on Information Systems, 36(1):1-37.

[3] Irad B G, Shahar W, Gonen S. (2019).Clustering users by their mobility behavioral patterns. ACM Transactions on Knowledge Discovery from Data, 13(4): 1-28.