

Received February 8, 2021, accepted February 22, 2021, date of publication February 26, 2021, date of current version March 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3062735

A Novel Similar Player Clustering Method With Privacy Preservation for Sport Performance Evaluation in Cloud

RUI MA¹, JIANQIANG LI², BAOHUI XING³, YUANYUAN ZHAO⁴, YUWEN LIU⁵,
CHAO YAN⁵, AND HANG YIN⁶

¹General Education Department, Shandong First Medical University (Shandong Academy of Medical Sciences), Tai'an 271000, China

²Nineteenth Middle School of Taian, Tai'an 271000, China

³First Primary School of Fengshui Town, Zibo, China

⁴Department of Basketball, Rizhao Sports School, Rizhao, China

⁵School of Computer Science, Qufu Normal University, Rizhao 276826, China

⁶School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China

Corresponding author: Hang Yin (myworkspace08@126.com)

This work was supported in part by the Natural Science Foundation of Shandong Province under Grant ZR2019MF001, and in part by the Open Project of State Key Laboratory for Novel Software Technology under Grant KFKT2020B08.

ABSTRACT With the ever-increasing popularity of sports and health ideas, people are paying more attentions to gaining high-quality healthy life through various taking various sport items or exercises. Through observing and analyzing the past sport exercise score records, we can cluster the players into different categories, each of which share the same or similar sport preferences or performances. However, the sport exercise score records are often massive and often stored in different cloud platforms, which raise a big difficulty for time-efficient player clustering. Furthermore, the sport exercise score records are a kind of privacy for most players; therefore, it is often not rational or legal to release these sensitive data to the public for similar player clustering purpose. Considering the above two issues, we use SimHash, a kind of privacy-aware approximate neighbor search technique, for similar player clustering by analyzing the sport exercise score records distributed across different cloud platforms. Thus, we can realize privacy-aware similar player clustering through SimHash. At last, we provide a set of experiments to validate the advantages of our proposed privacy-aware similar player clustering algorithm. Reported experimental results show the effectiveness of our proposal in remedying the big data volume and privacy concerns in player clustering based on sport exercise score records.

INDEX TERMS Similar player clustering, sport exercise score records, SimHash, privacy, big data, cloud platform.

I. INTRODUCTION

Along with the increasing development of society and economy, people's material living levels are continuously improved in the last decades. Therefore, people can enjoy more material conditions than ever before, typically in terms of foods, clothes, travelling and living. As a result, people can enjoy high-quality and comfortable living conditions, which bring people more happiness and satisfactions. However, on the other hand, people's living conditions are also influenced to some extent by some other factors as the improvement of material conditions is also double-edged.

The associate editor coordinating the review of this manuscript and approving it for publication was Qin Liu.

In concrete, the increasing improvements of material conditions have begun to bring several side effects to people's living levels. For example, people tend to become much fatter than ever before because of the rich living conditions. Consequently, obesity is becoming more and more popular and also brings several obesity-aware disasters such as hypertension, hyperlipemia and less exercises. In this situation, sports or exercises have become more and more popular as they have been proven an effective way to overcome the trouble brought by obesity and other health-related disasters.

For better sports or exercises, it is becoming a necessity to find out the users or players with same or similar sport preferences or habits [1], i.e., similar player clustering. However, similar player clustering is a non-trivial task due

to the absence of player clustering decision-making data. Fortunately, people's activity behavior data, especially the sport exercise score records from people (for example, students' taken sports and scores in universities), have offered a promising way to measure and evaluate the sport preferences or habits of individuals effectively and objectively [2]. However, the sport exercise score records (typically stored in different cloud platforms) are often massive, which raise a big difficulty for time-efficient similar player clustering. Furthermore, the sport exercise score records are a kind of privacy for most players (for example, the students with certain physical drawbacks); therefore, it is often not rational or legal to release these sensitive data to the public for similar player clustering purpose.

Considering the above two issues, we use SimHash, a kind of privacy-aware approximate neighbor search technique, to cluster the players with the same or similar sport preferences and habits by analyzing the sport exercise score records distributed across different cloud platforms. Furthermore, we put forward a SimHash-based similar player clustering (i.e., similar individual finding) method with privacy preservation, named SIF_{SimHash} (Similar Individual Finding based on SimHash).

In summary, the major contribution of this research work is three-fold.

(1) We observe the big data volume and data sensitivity in sport exercise score records-based similar player clustering process.

(2) We introduce SimHash technique into sport exercise score records-based similar player clustering process so as to cope with the big data volume and data sensitivity simultaneously.

(3) We design a set of experiments to validate the advantages of our proposed SIF_{SimHash} algorithm. The experimental reports show the effectiveness of SIF_{SimHash}.

The paper is organized as follows. Current research status of the field is studied and summarized in Section 2. In Section 3, privacy-aware player clustering problem is formalized and described intuitively. The suggested SIF_{SimHash} algorithm is clarified in detail in Section 4. A wide range of experiments are shown in Section 5 to prove the feasibility of SIF_{SimHash}. Conclusions are drawn in Section 6.

II. RELATED WORK

We summarize the state-of-the-art research work of the field from the following two aspects.

A. MISSING DATA PREDICTION AND SIMILAR ITEM CLUSTERING

The prediction of missing data has been a long-term research topic in big data-driven business applications. A considerable number of researchers have devoted themselves to this research topic and introduced various resolutions.

In work [3], a content-driven missing data prediction method is suggested. Typically, such a kind of content-based prediction is generally dependent on the contents that people

have browsed, read or rated in the past. For instance, if a user rated a 5-start score towards movie "Titanic", then the user will probably rate a 5-start score towards movie "Avatar" if his rating score towards "Avatar" is absent, as these two movies have been both conducted by director James Cameron. In literature [4], demography profile is employed for better prediction of missing data. Concretely, the demography information of a user is recruited to infer his possible preferences, e.g., user age, male/female, user income, affiliation and position, education background, and so on. In literature [5], knowledge-driven absent data prediction is developed, in which the prediction is mainly based on the knowledge associated with involved things. For instance, if today is rainy, a user would like to take an umbrella as there exists inherent knowledge that a person needs to play an umbrella in rainy days. Similar work is done in [6] where association rules are utilized to describe and quantify the correlations among different involved items or things. For instance, the beef price can be predicted by mining the Internet content such as historical user ratings, user feedback, user reviews, and so on, as there are hidden correlations between pork price and other information on the web.

Another category of traditional absent data prediction method is collaborative filtering. The general rationale of collaborative filtering is: if two users A and B share the same or similar preferences, then user A's rating on an item would be the same as user B's rating on the item. With the above analyses, missing data can be predicted accordingly. Concrete variants include user-based collaborative filtering [7], item-based collaborative filtering [8] and hybrid collaborative filtering [9].

B. PRIVACY-PRESERVATION

Privacy-preservation is a common concern in most big data-driven business application domains and has attracted considerable attentions and interests from both academy and industry.

If a user has multiple pieces of sensitive data, he/she can secure most of his/her privacy data by releasing only one piece of representative data stored in cloud to other people, without releasing all the pieces of data. This is the basic idea of literature [10] where only partial privacy information is disclosed while most privacy information is secured. Converting a piece of sensitive data into an encrypted text through a kind of encryption strategy is the major idea of encryption-based privacy-preservation in cloud environment [11]. For example, a symmetric public key encryption strategy is brought forth in [12] to achieve multiple keywords-driven and privacy-preserving information search.

Differential computing is a recently developed technique that guarantees user privacy during the execution of data-driven cloud business systems. The authors in [13] combine collaborative filtering with differentially privacy for high quality absent data prediction while maintaining a good privacy protection quality. However, both collaborative filtering and differentially privacy call for a heavy











						
Jack 	1	0	0	1	1	1
Alice 	0	1	1	1	0	0
Tom 	0	1	0	0	1	1
Judy 	1	0	0	1	1	1

FIGURE 1. Player-sport selection matrix.

computational cost as frequent calculation and model updates are necessary. In literature [14], differentially privacy and matrix factorization techniques are integrated to make full use of the advantages of these two techniques. Concretely, differentially privacy is responsible for protecting the sensitive user information, while matrix factorization makes accurate prediction of absent data. However, similar to literature [13], the computational time of [14] is often large. Literature [15] integrates Differentially Privacy and Huffman Coding for securing user location privacy. Similarly, Differentially Privacy is integrated with Bayes network in [16] for better prediction of absent high-dimensional data.

Hash is another category of effective resolutions when securing sensitive user data. The authors in [17] use Locality-Sensitive Hashing for privacy protection goal. Similar research is conducted in [18] where multiple pieces of quality data are secured, in [8] in which spatial-temporal context factors are employed and protected, in [7] where high prediction accuracy is achieved while protecting user privacy, and in [19] where privacy and diversity are considered simultaneously. Another hash variant for addressing privacy issues is Minhash. For example, the authors in [20] use Minhash to secure the private information hidden in the service intersection co-invoked by different users.

However, the abovementioned hash variants cannot protect the sensitive sport physique monitoring data that we focus in this paper very well. Motivating by this fact, we take advantage of the well-known SimHash technique to secure the sport exercise score data and meanwhile tackle the big data volume issues. The concrete details of the suggested $SIF_{SimHash}$ method will be described step by step in Section 4.

III. MOTIVATION

Next, we introduce the example in Figure 1 to describe the research focus and motivation of our work in this paper. In Fig.1, there are four players: Jack, Alice, Tom and Judy. Each player can select to take six sport items: football, volleyball, weight lifting, fencing, field hockey and boating. If a player chooses a sport item for sport exercise, then the corresponding entry is marked “1”; otherwise, if a player does not choose a sport item, then the corresponding entry is marked “0”.

To cluster the six players, it is necessary to analyze the player-sport selection matrix (typically stored in cloud platforms) shown in Fig.1. However, there are many sport exercise score records in the matrix, which make

Step-1: Sport item coding. For each sport item in the sport list, we assign it a specific code constituted by a 0-1 string.

Step-2: Player index creation. According to the sport item codes and the player-sport selection matrix, create an index for each player.

Step-3: Privacy-aware similar player clustering. According to player indices, cluster the similar players who take similar sport items.

FIGURE 2. Concrete steps of $SIF_{SimHash}$.

it time-consuming to cluster the six players. Moreover, the player-sport selection data in the matrix are often sensitive, as players are reluctant to reveal these records to the public. Thus, securing the player-sport selection records is a significant research topic when analyzing the player-sport selection records.

Thus, a challenge is raised when clustering the similar players according to the known player-sport selection records distributed across different cloud platforms while protecting the privacy of the involved players. Inspired by this challenge, we introduce a privacy-aware similar player clustering method, i.e., $SIF_{SimHash}$ based on SimHash. The concrete algorithm is specified in Section 4.

IV. APPROACH: $SIF_{SimHash}$

In summary, our suggested $SIF_{SimHash}$ method mainly includes the following three steps: (1) Sport item coding: each sport item is denoted by a 0-1 string; (2) Individual index creation: according to the sport item codes and the people-sport selection matrix, create an index for each individual; (3) Similar individual finding: according to individual indices, search for the similar individuals who take similar sport items.

A. STEP-1: SPORT ITEM CODING

The basic idea of SimHash [21] in securing user privacy information is that SimHash can convert a piece of sensitive user information into a privacy-free vector constituted by a Boolean string, e.g., 01100. To achieve this goal, it is necessary to assign an unique code to each sport item in the sport item list (see Fig.1). This procedure is named “sport item coding”.

For example, each sport item is assigned a Boolean string such as: football (100110), basketball (011010), weight lifting (101101), fencing (010011), field hockey (111101) and boating (110110). The coding strategy here is not fixed but varied. Generally, we can assign a distinct Boolean string for each sport item, as long as we can ensure that each sport item is assigned a concrete but distinctive Boolean string.

B. STEP-2: PLAYER INDEX CREATION

According to the pre-defined sport item codes (produced in Step-1) and the player-sport selection matrix (see Fig.1), we can generate an index for each player. The concrete procedure is described as follows.

For each player in the player-sport selection matrix, we only need to focus on his/her selected sport items, i.e., the sport items marked “1” (the sport items marked “0” mean that an item is not chosen by the player. Thus, they do not influence the final results of similar player clustering with high probably.

Considering the scenario in Fig.1, Jack selects four sport items: football, fencing, field hockey and boating. Therefore, we only need to consider the codes (generated in Step 1) corresponding to the four selected sport items. In concrete, the sport item selection vector presented in (1) is finally converted into a corresponding matrix, as presented in (2).

$$\text{Jack } (1, 0, 0, 1, 1, 1) \tag{1}$$

$$\text{Jack } \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \end{bmatrix} \tag{2}$$

Next, we project each entry in the matrix in (2). For example, each entry of “1” stays unchanged but each entry of “0” is changed to “-1”. Thus, we get a new matrix as shown in (3). To achieve personalized discrimination, we assign a concrete weight value to each sport item so as to quantify its significance in evaluating and clustering the player performances and preferences. Let’s consider the scenario in Fig.1, we assume the weights for the six sport items are w_1, w_2, \dots, w_6 , respectively. Then we multiply the weights with the matrix in (3). Thus, we derive a weighted matrix in (4).

$$\text{Jack } \begin{bmatrix} 1 & -1 & -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix} \tag{3}$$

$$\text{Jack } \begin{bmatrix} w_1 & -w_1 & -w_1 & w_1 & w_1 & -w_1 \\ -w_4 & w_4 & -w_4 & -w_4 & w_4 & w_4 \\ w_5 & w_5 & w_5 & w_5 & -w_5 & w_5 \\ w_6 & w_6 & -w_6 & w_6 & w_6 & -w_6 \end{bmatrix} \tag{4}$$

Next, for each column of matrix in (4), we compute the sum of all the element values in the identical column. In other words, a plus operation is imposed on each column. Afterwards, the matrix in (4) is converted into the vector in (5). Next, we project each entry in the vector in (5). In concrete, if the entry value is larger than 0, then we use “1” for replacement; otherwise, we use “0” for replacement. Afterwards, we get a vector formed by a Boolean string, e.g., we get a vector for Jack, i.e., (1, 1, 0, 1, 1, 0). Thus, the vector (1, 1, 0, 1, 1, 0) is taken as the index for Jack, denoted by $\text{Index}_{\text{Jack}}$. As (1, 1, 0, 1, 1, 0) does not contain much privacy of Jack, we can protect the privacy of Jack if we use index (1, 1, 0, 1, 1, 0) for the following similar player clustering. This is also the reason why we argue that our suggested SimHash-based similar player clustering solution, i.e., $\text{SIF}_{\text{SimHash}}$ is effective and efficient in securing the sensitive information of players.

$$\begin{aligned} &\text{Jack } (w_1 - w_4 + w_5 + w_5, -w_1 + w_4 + w_5 + w_6, \\ &\quad -w_1 - w_4 + w_5 - w_6, w_1 - w_4 + w_5 + w_6, \\ &\quad w_1 + w_4 - w_5 + w_6, -w_1 + w_4 + w_5 - w_6) \end{aligned} \tag{5}$$

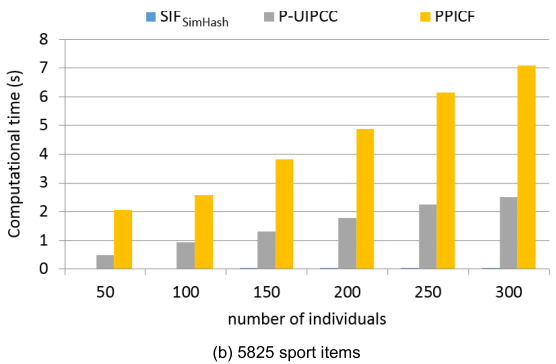
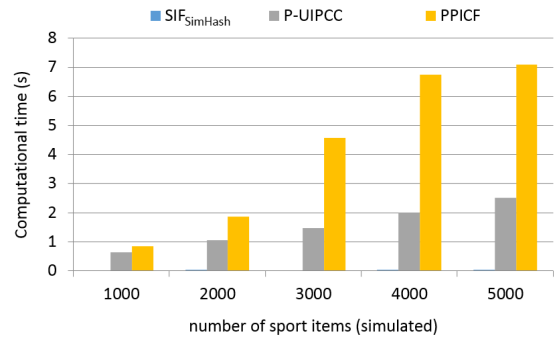


FIGURE 3. Running time of three solutions.

C. STEP-3: PRIVACY-AWARE SIMILAR PLAYER CLUSTERING

In Step-2, we have derived the index values of each player, i.e., $\text{Index}_{\text{Jack}}, \text{Index}_{\text{Alice}}, \text{Index}_{\text{Tom}}$ and $\text{Index}_{\text{Judy}}$. Next, we use the privacy-free index values to cluster the similar players among all the four players: Jack, Alice, Tom and Judy. The player clustering is based on a theoretical finding in SimHash. In concrete, if the Hamming Distance [22] between two vectors is not larger than 3, then we can approximately deem that these two vectors are close.

Let’s continue to consider the scenario in Fig.1, the index values of the four players are assumed to be:

$$\begin{aligned} \text{Index}_{\text{Jack}} &= (1, 1, 0, 1, 1, 0) \\ \text{Index}_{\text{Alice}} &= (0, 1, 1, 0, 1, 1) \\ \text{Index}_{\text{Tom}} &= (1, 1, 0, 1, 1, 1) \\ \text{Index}_{\text{Judy}} &= (1, 0, 0, 1, 1, 0) \end{aligned}$$

Next, we calculate the Hamming Distance (denoted by HD) of each player pair, whose results are:

$$\begin{aligned} \text{HD}(\text{Jack}, \text{Alice}) &= 4 \\ \text{HD}(\text{Jack}, \text{Tom}) &= 1 \\ \text{HD}(\text{Jack}, \text{Judy}) &= 1 \\ \text{HD}(\text{Alice}, \text{Tom}) &= 3 \\ \text{HD}(\text{Alice}, \text{Judy}) &= 5 \\ \text{HD}(\text{Tom}, \text{Judy}) &= 2 \end{aligned}$$

According to Hamming Distances, we can cluster similar players into four pairs, i.e., pair (Jack, Tom), pair (Jack, Judy),

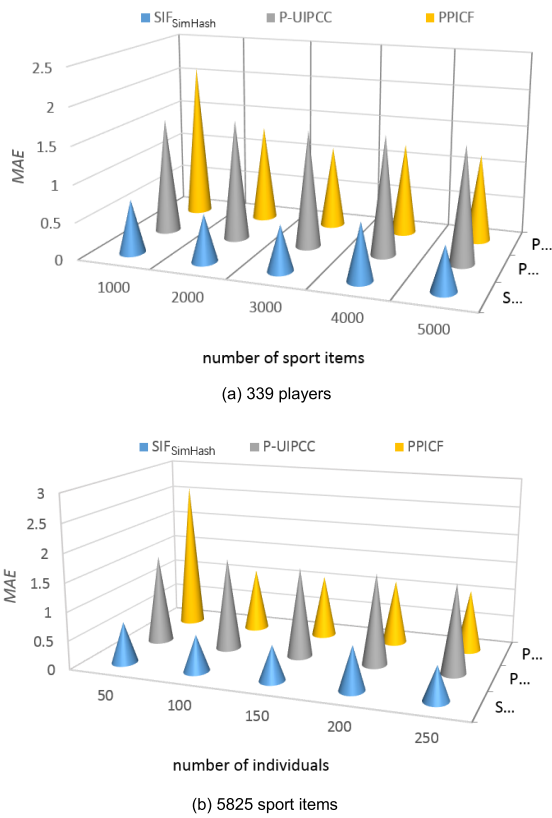


FIGURE 4. Prediction accuracy of three solutions.

pair (Alice, Tom), pair (Tom, Judy). As the clustering process is only based on privacy-free index values of the four players, we argue that the proposed SIF_{SimHash} solution is privacy-free. Next, we will test the effectiveness and efficiency of SIF_{SimHash} in performing similar player clustering in cloud environment in the next section.

V. EXPERIMENTS

A. CONFIGURATION

We use the WS-DREAM dataset for experiment evaluation in this section. WS-DREAM includes the historical service (totally 5825) invocation records by 339 users. Partial records are randomly removed for simulation purpose.

Our proposed SIF_{SimHash} solution is compared with two related methods, i.e., P-UIPCC [23] and PPICF [24]. Experimental configurations include 2.40 GHz CPU, 4.0 GB RAM, Windows 10 and JAVA 8. Each test is repeated 50 times and their average values are recorded.

B. RESULTS

The 5825 items in WS-DREAM is taken to simulate the candidate sport items and the 339 users are taken to represent the players who select the sport items.

1) RUNNING TIME

We test the running time of three solutions and the concrete experimental results are shown in Fig.3. As Fig.3 shows, for P-UIPCC and PPICF, the running time both increases with the

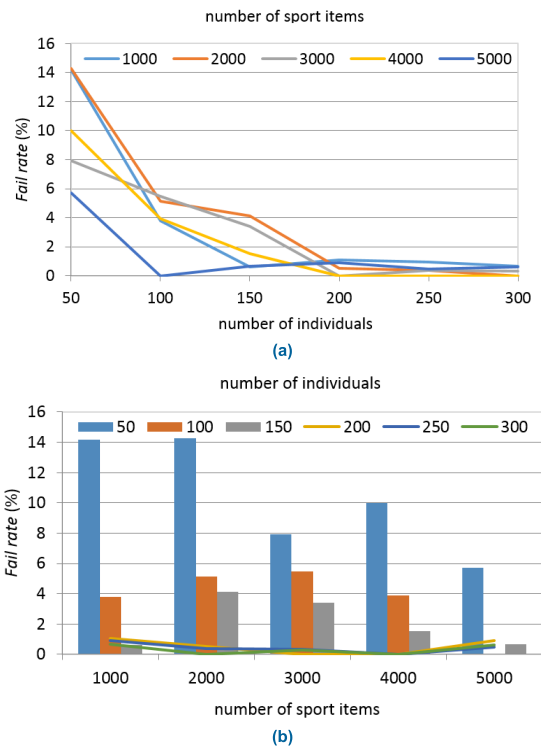


FIGURE 5. Fail rate of SIF_{SimHash}.

growth of the player volume and the sport item volume. This is because the similarity calculation operation in these two solutions involve all the player-sport item pairs. As a comparison, our SIF_{SimHash} solution performs the best; concretely, the running time of SIF_{SimHash} approximately approaches 0 as the player indices in Step-2 are generated in an offline manner. The low running time of SIF_{SimHash} also proves that SIF_{SimHash} can accommodate the big size of player-sport item selection matrix.

2) CLUSTERING ACCURACY

The final goal of three solutions (i.e., SIF_{SimHash}, P-UIPCC, PPICF) is the same, i.e., clustering the similar players. Therefore, the accuracy of the outputted similar player clusters by three solutions is also a critical metric to test the clustering performances. Motivated by this fact, we conduct a series of experiments to test the clustering accuracy of three solutions, whose test results are presented in Fig.4.

Fig.4 reports that the MAE performance of SIF_{SimHash} solution is smaller than P-UIPCC and PPICF, which shows a high prediction performance of SIF_{SimHash} solution. The reason is: SimHash technique used in SIF_{SimHash} solution is proved to be an effective clustering technique. Namely, SIF_{SimHash} solution can ensure to output only those “most similar” players among all ones. This is also the reason we choose SimHash for privacy-free similar player clustering.

3) FAIL RATE OF SIF_{SimHash}

SimHash has already been proven an effective neighbor search technique. Therefore, it is difficulty to ensure 100%

success when clustering similar players. Namely, SIF_{SimHash} solution failed to produce similar player clusters in some cases. Considering this shortcoming, we test the fail rate of SIF_{SimHash} solution and analyze its variation tendency, whose reports are presented in Fig.5.

As can be seen from Fig.5, the fail rate of SIF_{SimHash} generally drops with the growth of player volume and sport item volume. This can be explained as below: when player volume and sport item volume increase, more valuable information can be engaged in the similar player clustering process and therefore, fail rate of clustering is dropped accordingly.

VI. CONCLUSION

People are paying more attentions to gaining high-quality healthy life through various taking various sport items or exercises. Through observing and analyzing the past sport exercise score records, we can cluster the players into different categories, each of which share the same or similar sport preferences or performances. However, the sport exercise score records are often massive and often stored in different cloud platforms, which raise a big difficulty for time-efficient player clustering. Furthermore, the sport exercise score records are a kind of privacy for most players; therefore, it is often not rational or legal to release these sensitive data to the public for similar player clustering purpose. Considering the above two issues, we use SimHash, a kind of privacy-aware approximate neighbor search technique, for similar player clustering by analyzing the sport exercise score records distributed across different cloud platforms. Thus, we can realize privacy-aware similar player clustering through SimHash. At last, we provide a set of experiments to validate the advantages of our proposed privacy-aware similar player clustering algorithm. Reported experimental results show the effectiveness of our proposal in remedying the big data volume and privacy concerns in player clustering based on sport exercise score records.

In future study, we will continuously refine SIF_{SimHash} by considering more data types [25]–[29] and more optimization metrics [30]–[35]. Besides, how to integrate SIF_{SimHash} with other privacy protection techniques (e.g., in [36]–[45]) is another research focus.

REFERENCES

- [1] S. Din and A. Paul, "Smart health monitoring and management system: Toward autonomous wearable sensing for Internet of Things using big data analytics," *Future Gener. Comput. Syst.*, vol. 111, p. 939, Oct. 2020.
- [2] E. Silience, J. M. Blythe, P. Briggs, and M. Moss, "A revised model of trust in Internet-based health information and advice: Cross-sectional questionnaire study," *J. Med. Internet Res.*, vol. 21, no. 11, Nov. 2019, Art. no. e11125.
- [3] J. Li, Z. Xing, and A. Kabir, "Leveraging official content and social context to recommend software documentation," *IEEE Trans. Serv. Comput.*, early access, Mar. 6, 2018, doi: 10.1109/TSC.2018.2812729.
- [4] G. D. Poznik et al., "Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences," *Nature Genet.*, vol. 48, no. 6, pp. 593–599, Jun. 2016.
- [5] J. K. Tarus, Z. Niu, and G. Mustafa, "Knowledge-based recommendation: A review of ontology-based recommender systems for e-learning," *Artif. Intell. Rev.*, vol. 50, no. 1, pp. 21–48, Jun. 2018.
- [6] M. Y. Karim, H. Kagdi, and M. Di Penta, "Mining Android apps to recommend permissions," in *Proc. IEEE 23rd Int. Conf. Softw. Anal., Evol., Reeng. (SANER)*, Mar. 2016, pp. 427–437.
- [7] L. Qi, X. Wang, X. Xu, W. Dou, and S. Li, "Privacy-aware cross-platform service recommendation based on enhanced locality-sensitive hashing," *IEEE Trans. Netw. Sci. Eng.*, early access, Jan. 27, 2020, doi: 10.1109/TNSE.2020.2969489.
- [8] L. Qi, C. Hu, X. Zhang, M. R. Khosravi, S. Sharma, S. Pang, and T. Wang, "Privacy-aware data fusion and prediction with spatial-temporal context for smart city industrial environment," *IEEE Trans. Ind. Informat.*, early access, Jan. 28, 2020, doi: 10.1109/TII.2020.3012157.
- [9] L. Wang, X. Zhang, T. Wang, S. Wan, G. Srivastava, S. Pang, and L. Qi, "Diversified and scalable service recommendation with accuracy guarantee," *IEEE Trans. Comput. Social Syst.*, early access, Jul. 21, 2020, doi: 10.1109/TCSS.2020.3007812.
- [10] W. Dou, X. Zhang, J. Liu, and J. Chen, "HireSome-II: Towards privacy-aware cross-cloud service composition for big data applications," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 2, pp. 455–466, Feb. 2015.
- [11] Q. Liu, Y. Tian, J. Wu, T. Peng, and G. Wang, "Enabling verifiable and dynamic ranked search over outsourced data," *IEEE Trans. Serv. Comput.*, early access, Jun. 11, 2019, doi: 10.1109/TSC.2019.2922177.
- [12] J. Li, Y. Huang, Y. Wei, S. Lv, Z. Liu, C. Dong, and W. Lou, "Searchable symmetric encryption with forward search privacy," *IEEE Trans. Depend. Sec. Comput.*, vol. 18, no. 1, pp. 460–474, Jan. 2021, doi: 10.1109/TDSC.2019.2894411.
- [13] H. Ming, C. Mengmeng, and W. Xiaofei, "A collaborative filtering recommendation method based on differential privacy," *J. Comput. Res. Develop.*, vol. 54, no. 7, pp. 1439–1451, 2017.
- [14] W. Tong and H. Shubin, "An improved collaborative filtering recommendation algorithm with differentially privacy," *Inf. Secur. Technol.*, vol. 7, no. 4, pp. 26–28, 2016.
- [15] Y. Xiao, L. Xiong, S. Zhang, and Y. Cao, "LocLok: Location cloaking with differential privacy via hidden Markov model," *Proc. VLDB Endowment*, vol. 10, no. 12, pp. 1901–1904, Aug. 2017.
- [16] X. Ren, J. Xu, X. Yang, and S. Yang, "Bayesian network-based high-dimensional crowdsourced data publication with local differential privacy," (in Chinese), *Sci. Sinica Inf.*, vol. 49, no. 12, pp. 1586–1605, 2019.
- [17] X. Chi, C. Yan, H. Wang, W. Rafique, and L. Qi, "Amplified locality-sensitive hashing-based recommender systems with privacy protection," *Concurrency Comput., Pract. Exp.*, Feb. 2020, Art. no. e5681, doi: 10.1002/CPE.5681.
- [18] W. Zhong, X. Yin, X. Zhang, S. Li, W. Dou, R. Wang, and L. Qi, "Multi-dimensional quality-driven service recommendation with privacy-preservation in mobile edge environment," *Comput. Commun.*, vol. 157, pp. 116–123, May 2020.
- [19] L. Wang, X. Zhang, R. Wang, C. Yan, H. Kou, and L. Qi, "Diversified service recommendation with high accuracy and efficiency," *Knowl.-Based Syst.*, vol. 204, Sep. 2020, Art. no. 106196, doi: 10.1016/j.knsys.2020.106196.
- [20] V. Popic and S. Batzoglou, "A hybrid cloud read aligner based on MinHash and kmer voting that preserves privacy," *Nature Commun.*, vol. 8, no. 1, Aug. 2017, Art. no. 15311.
- [21] Z. Li and H. Yao, "Dynamic multi-keyword ranked search based on simhash in cloud computing," in *Proc. IEEE 21st Int. Conf. High Perform. Comput. Commun.*, Aug. 2019, pp. 591–598, doi: 10.1109/HPC/SmartCity/DSS.2019.00091.
- [22] R. Taheri, M. Ghahramani, R. Javidan, M. Shojafar, Z. Pooranian, and M. Conti, "Similarity-based Android malware detection using Hamming distance of static binary features," *Future Gener. Comput. Syst.*, vol. 105, pp. 230–247, Apr. 2020.
- [23] J. Zhu, P. He, Z. Zheng, and M. R. Lyu, "A privacy-preserving QoS prediction framework for Web service recommendation," in *Proc. IEEE Int. Conf. Web Services*, Jun. 2015, pp. 241–248.
- [24] D. Li, C. Chen, Q. Lv, L. Shang, Y. Zhao, T. Lu, and N. Gu, "An algorithm for efficient privacy-preserving item-based collaborative filtering," *Future Gener. Comput. Syst.*, vol. 55, pp. 311–320, Feb. 2016.
- [25] X. Zhou, Y. Li, and W. Liang, "CNN-RNN based intelligent recommendation for online medical pre-diagnosis support," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, May 14, 2020, doi: 10.1109/TCBB.2020.2994780.
- [26] Y. Wang, Y. Gao, Y. Li, and X. Tong, "A worker-selection incentive mechanism for optimizing platform-centric mobile crowdsourcing systems," *Comput. Netw.*, vol. 171, Apr. 2020, Art. no. 107144.
- [27] P. Lai, Q. He, M. Abdelrazek, F. Chen, J. Hosking, J. Grundy, and Y. Yang, "Optimal edge user allocation in edge computing with variable sized vector bin packing," in *Proc. 16th Int. Conf. Service-Oriented Comput.*, 2018, pp. 230–245.
- [28] X. Zhou, W. Liang, K. I.-K. Wang, H. Wang, L. T. Yang, and Q. Jin, "Deep-learning-enhanced human activity recognition for Internet of healthcare things," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6429–6438, Jul. 2020.

[29] J. Li, T. Cai, K. Deng, X. Wang, T. Sellis, and F. Xia, "Community-diversified influence maximization in social networks," *Inf. Syst.*, vol. 92, pp. 1–12, Sep. 2020.

[30] H. Kou, H. Liu, Y. Duan, W. Gong, Y. Xu, X. Xu, and L. Qi, "Building trust/distrust relationships on signed social service network through privacy-aware link prediction process," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106942, doi: [10.1016/j.asoc.2020.106942](https://doi.org/10.1016/j.asoc.2020.106942).

[31] Q. He, G. Cui, X. Zhang, F. Chen, S. Deng, H. Jin, Y. Li, and Y. Yang, "A game-theoretical approach for user allocation in edge computing environment," *IEEE Trans. Parallel Distrib. Syst.*, vol. 31, no. 3, pp. 515–529, Mar. 2020.

[32] X. Zhou, W. Liang, S. Huang, and M. Fu, "Social recommendation with large-scale group decision-making for cyber-enabled online service," *IEEE Trans. Comput. Social Syst.*, vol. 6, no. 5, pp. 1073–1082, Oct. 2019.

[33] T. Cai, J. Li, A. S. Mian, R. Li, T. Sellis, and J. X. Yu, "Target-aware holistic influence maximization in spatial social networks," *IEEE Trans. Knowl. Data Eng.*, early access, Jun. 17, 2020, doi: [10.1109/TKDE.2020.3003047](https://doi.org/10.1109/TKDE.2020.3003047).

[34] X. Xia, F. Chen, Q. He, J. C. Grundy, M. Abdelrazek, and H. Jin, "Cost-effective app data distribution in edge computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 32, no. 1, pp. 31–44, Jan. 2021, doi: [10.1109/TPDS.2020.3010521](https://doi.org/10.1109/TPDS.2020.3010521).

[35] X. Zhou, W. Liang, K. I.-K. Wang, R. Huang, and Q. Jin, "Academic influence aware and multidimensional network analysis for research collaboration navigation based on scholarly big data," *IEEE Trans. Emerg. Topics Comput.*, early access, Jul. 26, 2018, doi: [10.1109/TETC.2018.2860051](https://doi.org/10.1109/TETC.2018.2860051).

[36] Z. Cai, Z. He, X. Guan, and Y. Li, "Collective data-sanitization for preventing sensitive information inference attacks in social networks," *IEEE Trans. Depend. Sec. Comput.*, vol. 15, no. 4, pp. 577–590, Aug. 2018.

[37] Y. Xu, C. Zhang, G. Wang, Z. Qin, and Q. Zeng, "A blockchain-enabled deduplicatable data auditing mechanism for network storage services," *IEEE Trans. Emerg. Topics Comput.*, early access, Jun. 29, 2020, doi: [10.1109/TETC.2020.3005610](https://doi.org/10.1109/TETC.2020.3005610).

[38] Q. Liu, P. Hou, G. Wang, T. Peng, and S. Zhang, "Intelligent route planning on large road networks with efficiency and privacy," *J. Parallel Distrib. Comput.*, vol. 133, pp. 93–106, Nov. 2019.

[39] Z. Cai and X. Zheng, "A private and efficient mechanism for data uploading in smart cyber-physical systems," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 2, pp. 766–775, Apr. 2020.

[40] X. Zhou, Y. Hu, W. Liang, J. Ma, and Q. Jin, "Variational LSTM enhanced anomaly detection for industrial big data," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3469–3477, May 2021, doi: [10.1109/TII.2020.3022432](https://doi.org/10.1109/TII.2020.3022432).

[41] Q. Liu, G. Wang, F. Li, S. Yang, and J. Wu, "Preserving privacy with probabilistic indistinguishability in weighted social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 28, no. 5, pp. 1417–1429, May 2017.

[42] Y. Xu, J. Ren, Y. Zhang, C. Zhang, B. Shen, and Y. Zhang, "Blockchain empowered arbitrable data auditing scheme for network storage as a service," *IEEE Trans. Serv. Comput.*, vol. 13, no. 2, pp. 289–300, Apr. 2020.

[43] Z. Cai and Z. He, "Trading private range counting over big IoT data," in *Proc. IEEE 39th Int. Conf. Distrib. Comput. Syst.*, Jul. 2019, pp. 144–153, doi: [10.1109/ICDCS.2019.00023](https://doi.org/10.1109/ICDCS.2019.00023).

[44] S. Zhang, Q. Liu, and Y. Lin, "Anonymizing popularity in online social networks with full utility," *Future Gener. Comput. Syst.*, vol. 72, pp. 227–238, Jul. 2017.

[45] Z. Sun, Y. Wang, Z. Cai, T. Liu, X. Tong, and N. Jiang, "A two-stage privacy protection mechanism based on blockchain in mobile crowdsourcing," *Int. J. Intell. Syst.*, Jan. 2021, doi: [10.1002/int.22371](https://doi.org/10.1002/int.22371).



JIANQIANG LI received the bachelor's degree from the Shanghai University of Sport, China, in 2001. He is currently a First-Grade Teacher with the Nineteenth Middle School of Tai'an, Tai'an, China. His research interests include big data and sport training.



BAOHUI XING received the bachelor's and master's degrees from the Capital University of Physical Education and Sports, China, in 2011 and 2013, respectively. She is currently a First-Grade Teacher with the First Primary School of Fengshui Town, Zibo, China. Her research interest includes physical education.



YUANYUAN ZHAO received the master's degree in 2006. She is currently an Intermediate Teacher with Department of Basketball, Rizhao Sports School, Rizhao, China. Her research interest includes physical education and training.



YUWEN LIU received the bachelor's degree from the School of Computer Science, Qufu Normal University, China, in 2015. She is currently a Graduate Student with the School of Computer Science, Qufu Normal University. Her research interests include recommender systems and big data.



CHAO YAN received the master's degree from the Institute of Computing Technology, Chinese Academy of Sciences, China, in 2006. He is currently an Associate Professor with the School of Computer Science, Qufu Normal University, China. His research interests include recommender systems and big data.



RUI MA received the bachelor's and master's degrees from Qufu Normal University, China, in 2003 and 2011, respectively. He is currently a Lecturer with the General Education Department, Shandong First Medical University (Shandong Academy of Medical Sciences), Tai'an, China. His research interests include big data and medical healthcare.



HANG YIN received the M.S. degree in computer application from Liaoning Technical University, China, in 2007. He is currently a Lecturer with the School of Computer Science and Software Engineering, University of Science and Technology Liaoning. His research interests include data mining and big data management.

...