Received December 21, 2020, accepted January 7, 2021, date of publication January 12, 2021, date of current version January 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3051051

Sport Location-Based User Clustering With Privacy-Preservation in Wireless IoT-Driven Healthcare

QIYUN ZHANG[1], YUAN ZHANG[1], CAIZHONG LI[1], CHAO YAN[2], YUCONG DUAN[3], (Senior Member, IEEE), AND HAO WANG[4], (Senior Member, IEEE)

[1]Weifang Key Laboratory of Blockchain on Agricultural Vegetables, Weifang University of Science and Technology, Weifang 262700, China
[2]School of Computer Science, Qufu Normal University, Rizhao 276826, China
[3]School of Computer Science and Cyberspace Security, Hainan University, Haikou 570228, China
[4]Department of Computer Science, Norwegian University of Science and Technology, 7491 Trondheim, Norway

Corresponding author: Chao Yan (yanchao@qfnu.edu.cn)

ABSTRACT The gradual prevalence of Internet of Things (IoT) and wireless communication technologies has enabled the wide adoption of various smart devices (e.g., smart watches) in provisioning the healthcare services to massive users. Besides monitoring the real-time health signals or conditions of users, smart devices can also record a series of sport-related user information such as user location information at a certain time point. The location sequence information is valuable to cluster the users who share the similar sport preferences or habits and therefore, is also playing a key role in providing wireless healthcare services to these users. However, the user location information is often sensitive to certain wireless users as they decline to reveal their daily sport behavior patterns to others. In this situation, a natural challenge is raised in securing the sensitive user location information while mining the users' daily sport behavior patterns and provisioning better healthcare services to the users. Considering this challenge, we take advantage of the well-known SimHash technique to protect users' location privacy while clustering the users who share similar sport preferences or habits for better healthcare services. At last, we validate the feasibility of the proposal through a set of simulated experiments conducted on a real-world dataset. Reported results demonstrate that our solution performs better than the other two competitive ones while securing user location information.

INDEX TERMS Sport location, user clustering, privacy, healthcare service, simhash, wireless network.

I. INTRODUCTION

The gradual prevalence of the Internet of Things (IoT) and the wireless communication technologies has successfully enabled the wide adoption of various smart devices (such as smart watches, smart rings and smart phones, and so on) in provisioning high-quality healthcare services to massive users [1]. Typically, a smart device can monitor and collect the current healthy conditions or signals (e.g., temperature, blood pressure, and so on) in real time and generate health reports to users. According to these reported healthy data, more scientific and rational healthcare decisions could be made, so as to improve the users' living conditions [2]. Currently, the wireless IoT-enabled real time healthy data collection and transmission are gaining more and more attentions in both industrial applications and academic research domains.

Besides monitoring the real-time health signals or conditions of users (e.g., temperature, blood pressure, and so on), smart devices can also record a series of sport-related user information such as user location information at a certain time point (also called "check-in" data in social network domain) [3]. The user "check-in" data or user location sequence information is specifically valuable to cluster the massive users into different groups as the users with similar location sequence information often share similar sport preferences or habits [4]. Moreover, the users in an identical group or cluster are prone to be with similar healthy conditions and would probably enjoy similar healthcare services.

However, the user location information is often sensitive to certain users as they probably decline to reveal their daily sport behavior patterns to others who are not dependable

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

enough [5], [6]. In this situation, a natural challenge is raised in securing the sensitive user location information while mining the users' daily sport behavior patterns and provisioning better healthcare services to the users. Considering this challenge, we take advantage of the well-known SimHash [7] technique to protect users' location privacy while clustering the users who share similar sport behaviors or habits for better healthcare services. At last, we validate the feasibility of the proposal (named SLUC: Sport Location-based User Clustering) through a set of simulated experiments conducted on a real-world dataset. Reported results demonstrate that our solution performs better than the other two competitive ones while securing user location information.

In summary, the major contributions of this research work are three-fold.

(1) We recognize the importance of sensitive user location information (e.g., sport location) in clustering massive users into different groups with distinct sport preferences and habits.

(2) We recruit the well-known SimHash technique to protect users' location privacy while clustering the users who share similar sport behaviors or habits for better healthcare services.

(3) We validate the feasibility of the proposal through a set of simulated experiments conducted on a real-world dataset. Experimental results prove the feasibility of the new algorithm brought forth in this research work.

The reminder of this work is structured as below. We review the current research literatures of the field in Section 2. A motivating example is presented in Section 3 for better understanding the motivation of our research. In Section 4, we formulate the sport location-based user clustering problem with privacy-preservation and introduce the concrete steps of the proposed SLUC method. In Section 5, a series of simulated experiments are enacted to show the advantages of our proposed solution. At last, we conclude the whole paper and discuss the possible research directions in the future work in Section 6.

II. RELATED WORK

In this paper, we only focus on the issue of sport location-based similar user finding and clustering with privacy-preservation. Therefore, to further emphasize the research significance or research background of the work in this paper, we summarize the current research status of the field through the following two profiles.

A. LOCATION-BASED SIMILARITY CALCULATION

User location or item location is employed in [8] to divide massive users or items into multiple categories. This way, only the similar users of a target user or the similar items of a target item are recruited to make similar user finding and similar item finding; thus, the similarity calculation complexity is reduced accordingly as the search space is decreased considerably. Through the above similar item finding and cluster process, an unavailable service can be replaced by another similar service; this way, service exceptions can be remedied and the system robustness can be enhanced accordingly. In [9], the authors investigate the correlation relationships between location context and service quality (QoS). In concrete, QoS is modeled as a function associated with the parameters or variants of location context. This way, location-aware missing quality data are enacted and performed by studying the location context of the missing QoS records.

Similar to [8], massive candidate users are divided into different categories in [10] according to the user location; likewise, massive candidate items are divided into different categories according to the item location. Thus, the users belonging to the same cluster and the items belonging to the same cluster are used to make a time-efficient and scientific prediction decision. The authors in [11] argue that the geographical coordinates of users are an important evaluation metric for close user finding. Thus, the absent QoS data experienced by a user (assumed X) could be predicted accurately by observing the existed QoS data experienced by another user (assumed Y) who is geographically close to user X, as geographically close users often share the same or similar QoS as QoS is often location-related.

Work [12] claims that a user's experienced QoS data is often related to his or her neighbors' experienced QoS. Motivated by this hypothesis, a novel matrix factorization-based service quality prediction model is suggested in [13] by introducing the correlation relationship between the QoS data experienced by neighbors. The moving trace of users is considered as a valuable reference in [14] to predict the next point of interests as well as the corresponding QoS data. According to the predicted point of interests and QoS data, a high-quality point of interest recommendation decision could be generated and the optimal solution is generated.

However, through the abovementioned analyses, we can find that existing literatures in the field of location-based service quality prediction methods often do not consider the sensitive location information of users or items as well as the corresponding protection issues.

B. PRIVACY-AWARE DATA PREDICTION

The authors of [15] insist that a service user can conceal most of his/her recorded sensitive QoS data over services by only revealing the optimal value of QoS performances to the third-party. This way, privacy protection goal can be achieved more or less. However, there is still a small volume of QoS data whose sensitive information is probably disclosed. The authors of [16] obfuscate the real QoS data with privacy by plus or minus a small value that is produced randomly. This way, the real data containing private information are secured successfully. However, with the obfuscated QoS data for subsequent QoS prediction, the prediction accuracy would be influenced and decreased accordingly. Distributed storage technology is recruited in [17] to divide the sensitive QoS data (assumed Z) into several pieces (assumed $z_1, z_2, \ldots, z_n$ satisfying $Z = z_1 + z_2 + \ldots + z_n$) and record them separately

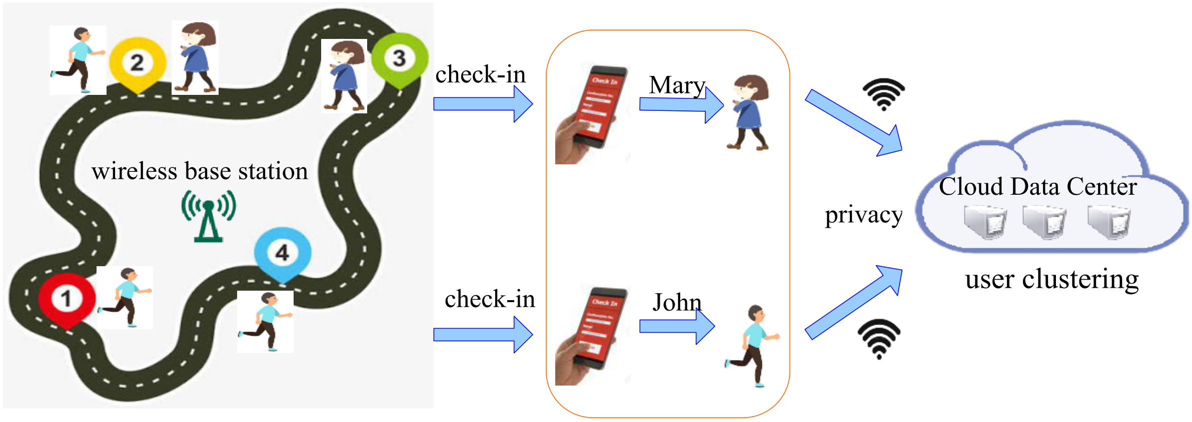


FIGURE 1. Sport location-based user clustering: privacy challenge.

by multiple parties. This way, each party only records a piece of sensitive QoS data (e.g., z_1), instead of the whole QoS data (i.e., Z); therefore, the disclosure of a party's recorded QoS data (e.g., z_1) would not lead to the danger of the entire QoS data (i.e., Z).

Hash technology is also employed to achieve the privacy protection goal, and there are currently a series of hash-based resolutions such as [18]–[20] using LSH and [21] using enhanced LSH. These hash-based variants generally use index or embedding to realize quick and accurate similarity calculation. As classic data protection ways, encryption [22], [23] and blockchain [24] are also popular in securing sensitive user data especially when data communications among different parties are necessary.

In summary, existing research works generally focus on either location-based similarity calculation or privacy-aware data prediction, while seldom integrates these two research challenges together. The authors in [25] consider item locations and LSH for efficient and secure recommendation decision-makings. However, they take user location as a fixed factor in recommendation process while neglect the more common scenarios where user locations are varied with time elapsing. Considering the drawbacks of existing literatures, a novel sport location-based user clustering method with privacy is proposed.

III. A MOTIVATING EXAMPLE

The motivating example in Fig.1 (here, the IDs {1, 2, 3, 4} are corresponding to four places) shows the research background and significance of our paper. In the exampled scenario, two users *Mary* and *John* both like to take exercises and are apt to check in with smart phones to record their respective ever-visited location information. In concrete, *Mary* checked in at $place_2$ and $place_3$, while *John* checked in at $place_1$, $place_2$ and $place_4$. In this situation, their historical check-in records provide a promising basis to evaluate their ever-visited location similarity so as to cluster them into different location groups.

However, for the central cloud platform which is responsible to cluster users, it is inevitable to read the historical check-in records of *Mary* and *John*. Such a process probably discloses the sensitive information of locations ever-visited by *Mary* and *John*, respectively. Therefore, a challenge is raised in terms of the conflict between the users' privacy disclosure concern and the cloud platform's centralized user clustering analyses. Motivated by this challenging issue, a sport location-based user clustering method considering privacy, i.e., SLUC is brought forth in the next section, which is basically based on the SimHash technique.

For more formalized description and discussions, we summarize the used symbols in our approach in Table 1. In this table, various symbols as well as their respective symbol descriptions are introduced in detail. These symbol descriptions have established the formal base for the subsequent algorithm specifications.

TABLE 1. Symbol descriptions.

Symbol	Description
$u_1, \ldots, u_m$	The set of users
$place_1, \ldots, place_n$	places ever-visited by users
$f_1, \ldots, f_M$	SimHash functions
$T_1, \ldots, T_N$	SimHash tables
$P_1, \ldots, P_h$	Check-in data storage platform
Len	Code length for each place
$Code(place_1), \ldots, Code(place_n)$	Codes of $place_1, \ldots, place_n$
$w_1, \ldots, w_n$	Weights of $place_1, \ldots, place_n$
$I_1, \ldots, I_n$	Indices of $place_1, \ldots, place_n$

IV. METHOD

The major idea of the proposed SLUC method is: first, we convert the ever-visited location information of users (containing some privacy) into a string of binary values without much privacy; second, we use the users' corresponding strings of binary values to look for the users who share the similar or same sport habits or behaviors, and then users with similar or same sport habits or behaviors are clustered into a single group. This way, we can realize privacy-free similar user finding and clustering based on the sensitive ever-visited location information by users.

In summary, we divide the SLUC method into the following three steps as presented in Fig.2. Here, user set is denoted

Step 1: Place encoding. For each place ever-visited by users, we assign a code constituted by a string of binary values. The encoding manner is mainly dependent on the size of location set *P*.

Step 2: User index generation. According to a user's ever-visited place set and the corresponding place codes derived in Step 1, we generate the user's index constituted by a string of binary values. Here, the index generation is based on SimHash.

Step 3: Similar user clustering. According to the generated user indices in Step 2 and the theoretical knowledge of SimHash, determine the users who share the same or similar sport habits or behaviors and cluster them together.

FIGURE 2. Concrete procedure of SLUC method.

by $U = \{u_1, \ldots, u_m\}$, users' ever-visited location set is denoted by $P = \{place_1, \ldots, place_n\}$. In addition, these three steps are of the sequential order; in other words, the output of Step 1 is the input of Step2 and the output of Step 2 is the input of Step 3.

A. STEP 1: PLACE ENCODING

As introduced in this section, users' ever-visited location set is denoted by $P = \{place_1, \ldots, place_n\}$. Next, for each place $place_j$ $(1 \leq j \leq n)$, we assign a new code constituted by a string of binary values according to SimHash. The concrete encoding manner is mainly dependent on the size of location set *P*. In more detail, if location set *P* has *n* elements, then the length of the string code for each place $place_j$ is $Len = \lceil \log_2^n \rceil$. For example, if there are ten places in set *P*, then the length of the string code for each place $place_j$ is $Len = \lceil \log_2^{10} \rceil = 4$. In concrete, the string codes of the ten places can be of the following forms:

Code ($place_1$): 0001

Code ($place_2$): 0010

Code ($place_3$): 0011

Code ($place_4$): 0100

Code ($place_5$): 0101

Code ($place_6$): 0110

Code ($place_7$): 0111

Code ($place_8$): 1000

Code ($place_9$): 1001

Code ($place_{10}$): 1010

B. STEP 2: USER INDEX GENERATION

According to a user's ever-visited place set and the corresponding place codes derived in Step 1, we can generate the user's index constituted by a string of binary values. The concrete process is as follows. For a user u_i $(1 \leq i \leq m)$, we only focus on his/her ever-visited places and use them to generate user index. For example, in Fig.1, *John* checked in at $place_1$, $place_2$ and $place_4$, then we only consider the string codes of $place_1$, $place_2$ and $place_4$. If we use the string codes presented in Step 1, then we get these three places as well as their string codes as below (Fig.3(a)).

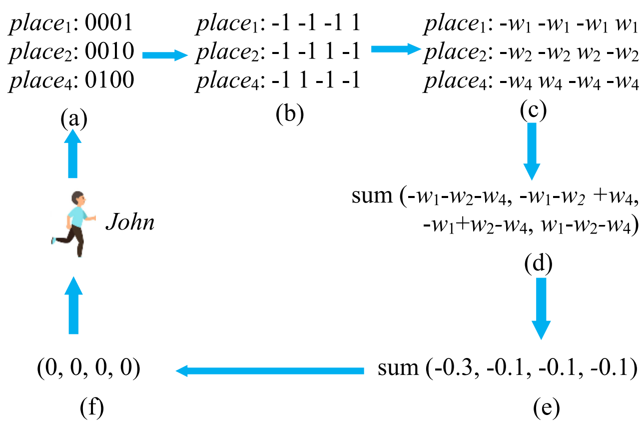


FIGURE 3. SimHash-based user index generation.

Next, we make a conversion for the string codes by replacing "0" with "−1" (Fig.3(b)). Afterwards, we introduce the weights for the *n* places in set *P* into the index generation process (assume the weights for the ten places in Step 1 are $w_1, \ldots, w_{10}$, respectively) by multiplying the weights with corresponding "1" or "−1" value (Fig.3(c)).

Then we count the sum of each column and get a sum vector as in Fig.3(d). Here, if we assume that $w_1 =, \ldots, = w_{10} = 0.1$, then the sum vector becomes (-0.3, -0.1, -0.1, -0.1) as presented in Fig.3(e). Finally, the positive entry in sum vector is replaced by "1" and the negative entry in sum vector is replaced by "0;" for example, the sum vector (-0.3, -0.1, -0.1, -0.1) is transformed into (0, 0, 0, 0) as in Fig.3(f). Then the vector (0, 0, 0, 0) is approximately regarded as the index for *John*. Thus, this step ends.

C. STEP 3: SIMILAR USER CLUSTERING

According to the procedure exampled in Fig.3, we can generate an index for each user in set *U*. Next, according to the user indices in Step 2 and the theoretical knowledge of SimHash, we can determine the users who share the same or similar sport habits or behaviors and cluster them together. Concretely, the similar user recognition process is as follows: considering two users u_x and u_y whose index values are I_x and I_y, respectively. If the equation in (1) holds, we can approximately regard u_x and u_y as similar users. In (1), "$*$" means the XOR operation between two vectors. At last, the similar users are clustered into an identical group. This way, we can finish the privacy-free similar user clustering process based on the sensitive sport location information generated by certain users in the historical visiting records. For more formalization, the pseudo code of the SLUC method

Algorithm 1 SLUC

Inputs:

(1) U = $\{u_1, \ldots, u_m\}$ // user set
(2) P = $\{place_1, \ldots, place_n\}$ // place set
(3) W = $\{w_1, \ldots, w_n\}$ // weight set for places
(4) R = $\{r_{i,j} | r_{i,j} = 1$ if u_i has visited $place_j$; otherwise, $r_{i,j} = 0$; $1 \le i \le$ m, $1 \le j \le$ n $\}$
// user-place visiting record set

Output: $G = \{group_1, group_2, \ldots\}$ // user clusters

$d = \lceil \log_2^n \rceil$;
For $j = 1$ to n **do**
assign a d-dimensional binary code c_j to $place_j$
End for
For $i = 1$ to m **do**
  M_i = Null // create a code matrix for user u_i
  $p_i = 0$ // record the row id of matrix M_i
  For $j = 1$ to n **do**
    If $r_{i,j} = 1$
      Then $p_i = p_i + 1$
        put c_j into M_i as its p_i-th row
        $w_{pi} = w_j$
      End If
    End For
    For $k = 1$ to p_i **do**
      For $l = 1$ to d **do**
        If $a_{k,l} = 0$
        Then $a_{k,l}$ = -w_k
        Else $a_{k,l} = w_k$
        End If
      End For
    End For
    For $l = 1$ to d **do**
      $sum_l = 0$
      For $k = 1$ to p_i **do**
        $sum_l = sum_l + a_{k,l}$
        If $sum_l > 0$
        Then $sum_l = 1$
        Else $sum_l = 0$
        End If
      End For
    End For
    $I_i = (sum_1, \ldots, sum_d)$
End For
For $i = 1$ to m **do**
  Cluster the users with the same index I_i into $group_i$
End For
FReturn $group_1, group_2, \ldots$

(i.e., the concrete procedure of Step 1 $\sim$ Step 3) is presented in Algorithm 1.

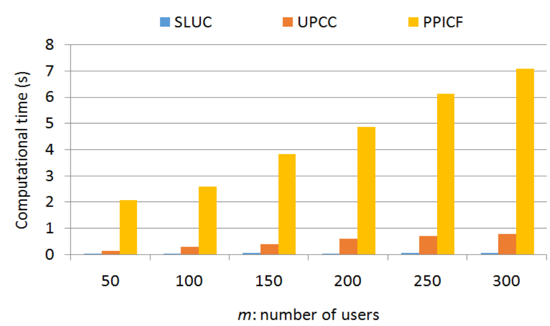
$$I_x * I_y \le 3 \tag{1}$$

V. EXPERIMENTS

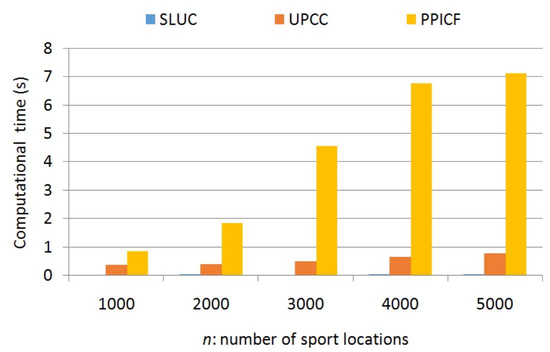
A set of experiments on real-world WS-DREAM [26] dataset are conducted in this section to prove the effectiveness and efficiency of the proposed SLUC method. The dataset is widely used in various user-service data related research domains. The simulated experiments are run in the OS of Windows 7 (64-bit) with 2.50 GHz processor and 16.0 GB RAM. We compare our work (i.e., SLUC) with two related solutions: UPCC [27] and PPICF [17]. Each test profile is executed 50 times.

A. COMPUTATIONAL TIME COMPARISONS WITH RELATED APPROACHES

Time efficiency (i.e., time complexity) is often a key criterion to evaluate the running performances of a recommender system. Motivated by this fact, we compare the time efficiency of different approaches and show the comparison results in Fig.4. Fig.4(a) measures the variation tendency of time efficiency of three solutions with respect to the number of users (i.e., m) and Fig.4(b) measures the variation tendency of time efficiency of three solutions with respect to the number of sport locations visited by users (i.e., n).



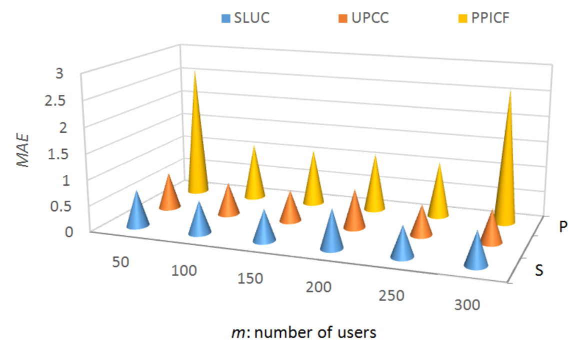
(a)



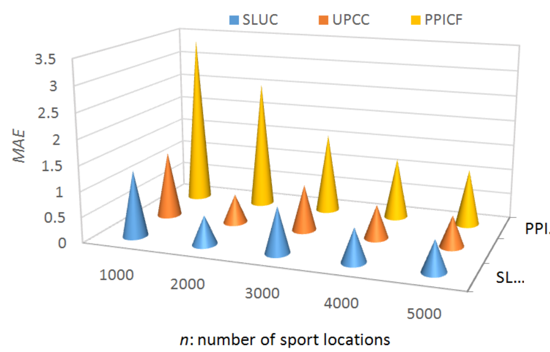
(b)

FIGURE 4. **Time efficiency comparisons.**

Through the comparisons in Fig.4, we can see that the time efficiency of the proposed SLUC method is superior to the other two related approaches, i.e., UPCC and PPICF (i.e., the consumed time of SLUC is the minimal). This is due to the fact that in SLUC, the user index items are always generated in an offline way (time complexity is approaching 0) and the online user query complexity based on user index items is approaching O(1). Therefore, as the results shown in Fig.4, the total computational time of SLUC solution is near to zero, which indicates that SLUC solution is very suitable for the big data environment.

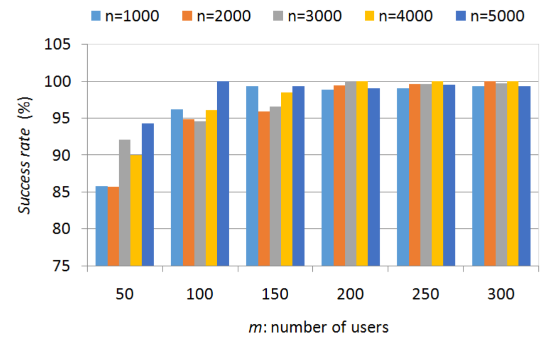


(a)

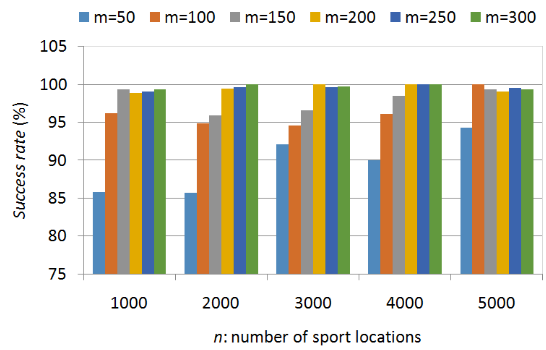


(b)

FIGURE 5. Prediction accuracy comparisons.



(a)



(b)

FIGURE 6. Success rate of SLUC solution.

B. PREDICTION ACCURACY WITH RELATED APPROACHES

Prediction accuracy (MAE, smaller is better) is another key metric to evaluate the performances of a recommender system as high prediction accuracy often brings high user satisfaction towards the recommended results. Motivated by this fact, we compare the prediction accuracy of different approaches and show the comparison results in Fig.5. Fig.5(a) measures the variation tendency of prediction accuracy of three solutions with respect to the number of users (i.e., m) and Fig.5(b) measures the variation tendency of prediction accuracy of three solutions with respect to the number of sport locations (i.e., n). Through the comparisons in Fig.5, we can conclude that the prediction accuracy of the proposed SLUC method is superior to the PPICF approach and approaches the baseline UPCC approach. This is due to the fact that in SLUC, the adopted SimHash strategy can achieve similarity-retaining, i.e., only those "most similar" users are clustered into the same group. Therefore, as shown in Fig.5, the prediction accuracy of SLUC is generally satisfactory in most experiments.

C. SUCCESS RATE OF SLUC WITH PARAMETERS m AND n

Failure is always inevitable for user clustering or recommender systems especially when the available data for clustering or prediction are very sparse. Motivated by this fact, we compare the success rate of SLUC with respect to parameters m and n, and show the concrete results in Fig.6. Fig.6(a) measures the success rate of SLUC with respect to the number of users (i.e., m) and Fig.6(b) measures the success rate of SLUC with respect to the number of sport locations (i.e., n). Through the reports in Fig.6, we can see that the success rate of SLUC method is increasing with the rising of m and n. This is due to the fact that in SLUC, more users or more sport locations often bring richer information for user clustering or similar user prediction. Therefore, as shown in Fig.6, the success rate of SLUC is approaching 100% when there are a big volume of users or sport locations.

D. PERFORMANCE CONVERGENCE

In this profile, we measure the performance (mainly MAE and Time cost) convergence of the suggested SLUC method with respect to the number of SimHash tables (denoted by m) used to cluster users. Concrete results are presented in Fig.7. In concrete, the performances of MAE and Time cost of SLUC method are becoming convergent when each group of experiments are repeated over 40 times. Therefore, it is appropriate to repeat 50 times and take their average measurement results in our experiment settings.

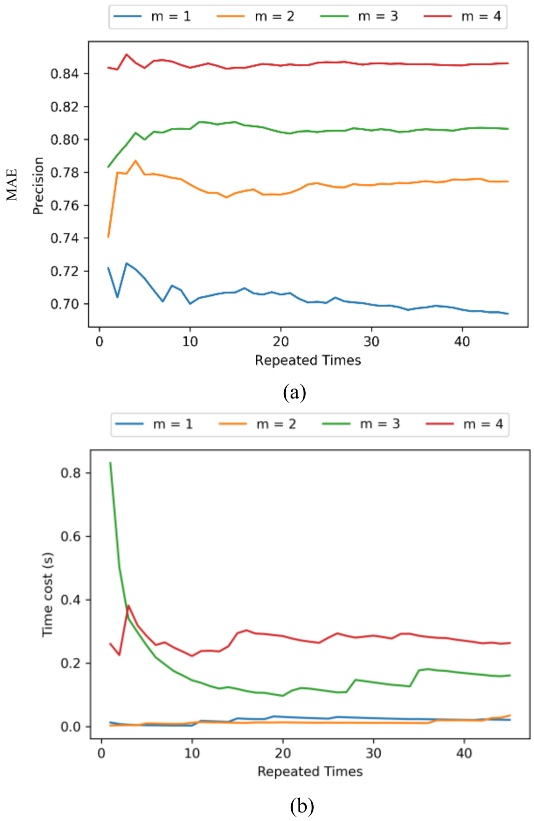


FIGURE 7. **Success rate of SLUC solution.**

VI. CONCLUSION

The gradual popularization of IoT and wireless communication technologies has enabled the wide adoption of various smart devices in provisioning the wireless healthcare services to massive users. Besides monitoring the real-time health signals or conditions of users, smart devices can also record user locations at a certain time point. Generally, the location sequence information is valuable to cluster the users who share the similar sport preferences or habits. However, the user location information is often sensitive to certain users. Considering this challenge, we take advantage of the well-known SimHash technique to protect users' location privacy while clustering the users who share similar sport preferences or habits for better healthcare services. At last, we validate the feasibility of the proposal through a set of simulated experiments conducted on a real-world dataset. Reported results demonstrate that our solution performs better than the other two competitive ones while securing user location information.

In the upcoming research, we will consider more context information or influencing factors [28]–[32] corresponding to users' location records. In addition, privacy protection capability measurement is also of practical and important significance for evaluating the recommendation performances [33]–[35]; therefore, how to quantify and measure the privacy protection performance of the hash-based solutions is also requiring challenging efforts in future study.

REFERENCES

[1] Z. Pan, X. Yi, and L. Chen, "Motion and disparity vectors early determination for texture video in 3D-HEVC," *Multimedia Tools Appl.*, vol. 79, nos. 7–8, pp. 4297–4314, Feb. 2020.

[2] 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, doi: 10.1109/JIOT.2020.2985082.

[3] U. Erra and N. Capece, "Engineering an advanced geo-location augmented reality framework for smart mobile devices," *J. Ambient Intell. Hum. Comput.*, vol. 10, no. 1, pp. 255–265, Jan. 2019.

[4] J. Kim, T. Guo, K. Feng, G. Cong, and A. Khan, "Densely connected user community and location cluster search in location-based social networks," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2020, pp. 2199–2209.

[5] Q. He, J. Yan, Y. Yang, R. Kowalczyk, and H. Jin, "Chord4S: A P2P-based decentralised service discovery approach," in *Proc. IEEE Int. Conf. Services Comput.*, Honolulu, HI, USA, Jul. 2008, pp. 221–228.

[6] Q. He, J. Yan, H. Jin, and Y. Yang, "ServiceTrust: Supporting Reputation-Oriented Service Selection," in *Proc. 7th Int. Conf. Service-Oriented Comput.*, Stockholm, Sweden, 2009, pp. 269–284.

[7] C. Sadowski and G. Levin, "Simhash: Hash-based similarity detection," Google, Menlo Park, CA, USA, Tech. Rep., 2007.

[8] G. Xu-Rui, W. Li, and W. Wei-Li, "Using multi-features to recommend friends on location-based social networks," *Peer Netw. Appl.*, vol. 10, no. 6, pp. 1323–1330, Nov. 2017.

[9] D. Ryu, K. Lee, and J. Baik, "Location-based Web service QoS prediction via preference propagation to address cold start problem," *IEEE Trans. Services Comput.*, early access, Apr. 2, 2018, doi: 10.1109/TSC.2018.2821686.

[10] Y. Yin, L. Chen, Y. Xu, and J. Wan, "Location-aware service recommendation with enhanced probabilistic matrix factorization," *IEEE Access*, vol. 6, pp. 62815–62825, 2018.

[11] S. Zhang, S. Zhang, N. Y. Yen, and G. Zhu, "The recommendation system of micro-blog topic based on user clustering," *Mobile Netw. Appl.*, vol. 22, no. 2, pp. 228–239, Apr. 2017.

[12] L. Kuang, L. Yu, L. Huang, Y. Wang, P. Ma, C. Li, and Y. Zhu, "A personalized QoS prediction approach for CPS service recommendation based on reputation and location-aware collaborative filtering," *Sensors*, vol. 18, no. 5, p. 1556, May 2018.

[13] D. Lian, K. Zheng, Y. Ge, L. Cao, E. Chen, and X. Xie, "GeoMF++: Scalable location recommendation via joint geographical modeling and matrix factorization," *ACM Trans. Inf. Syst.*, vol. 36, no. 3, pp. 1–29, Apr. 2018.

[14] Q. Li, T. Li, G. Ren, L. Cui, and W. He, "A new paradigm for personalized mashup recommendation based on dynamic contexts in mobile computing environments," *SCIENTIA SINICA Inf.*, vol. 46, no. 6, pp. 677–697, Jun. 2016.

[15] M. A. P. Chamikara, P. Bertok, D. Liu, S. Camtepe, and I. Khalil, "Efficient privacy preservation of big data for accurate data mining," *Inf. Sci.*, vol. 527, pp. 420–443, Jul. 2020.

[16] D. Yang, B. Qu, and P. Cudre-Mauroux, "Privacy-preserving social media data publishing for personalized ranking-based recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 3, pp. 507–520, Mar. 2019.

[17] 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.

[18] X. Chi, C. Yan, H. Wang, W. Rafique, and L. Qi, "Amplified locality-sensitive hashing-based recommender systems with privacy protection," *Concurrency Comput., Pract. Exper.*, vol. 1, Feb. 2020, Art. no. e5681, doi: 10.1002/CPE.5681.

[19] 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.

[20] X. Zhang, W. Dou, Q. He, R. Zhou, C. Leckie, R. Kotagiri, and Z. Salcic, "LSHiForest: A generic framework for fast tree isolation based ensemble anomaly analysis," in *Proc. 33rd IEEE Int. Conf. Data Eng. (ICDE)*, San Diego, CA, USA, 2017, pp. 983–994.

[21] 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, Jul. 27, 2020, doi: 10.1109/TNSE.2020.2969489.

[22] C.-Z. Gao, Q. Cheng, X. Li, and S.-B. Xia, "Cloud-assisted privacy-preserving profile-matching scheme under multiple keys in mobile social network," *Cluster Comput.*, vol. 22, no. S1, pp. 1655–1663, Jan. 2019.

[23] Q. Liu, Y. Tian, J. Wu, T. Peng, and G. Wang, "Enabling verifiable and dynamic ranked search over outsourced data," *IEEE Trans. Services Comput.*, early access, Jun. 11, 2019, doi: 10.1109/TSC.2019.2922177.

[24] 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. Services Comput.*, vol. 13, no. 2, pp. 289–300, May 2020.

[25] W. Lin, X. Zhang, L. Qi, W. Li, S. Li, V. S. Sheng, and S. Nepal, "Location-aware service recommendations with privacy-preservation in the Internet of Things," *IEEE Trans. Comput. Social Syst.*, early access, Feb. 4, 2020, doi: 10.1109/TCSS.2020.2965234.

[26] Accessed: Oct. 10, 2020. [Online]. Available: https://wsdream.github.io/

[27] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *Proc. 14th Conf. Uncertainty Artif. Intell.*, 1998, pp. 43–52.

[28] Z. Pan, "Adaptive fractional-pixel motion estimation skipped algorithm for efficient HEVC motion estimation," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 1, pp. 1–19, 2018.

[29] 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.

[30] 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.

[31] 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, 2020.

[32] X. Zhou, Y. Hu, W. Liang, J. Ma, and Q. Jin, "Variational LSTM enhanced anomaly detection for industrial big data," *IEEE Trans. Ind. Informat.*, early access, Sep. 11, 2020, doi: 10.1109/TII.2020.3022432.

[33] 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, Jul. 28, 2020, doi: 10.1109/TII.2020.3012157.

[34] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng, "Recent progress on generative adversarial networks (GANs): A survey," *IEEE Access*, vol. 7, pp. 36322–36333, 2019.

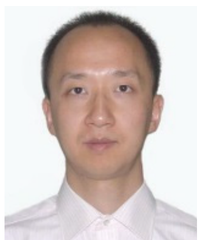
[35] L. Wang, X. Zhang, R. Wang, C. Yan, H. Kou, and L. Qi, "Diversified service recommendation with high accuracy and efficiency," *Knowl.-Based Syst.*, vol. 204, Sep. 2020, Art. no. 106196, doi: 10.1016/j.knosys.2020.106196.

QIYUN ZHANG received the bachelor's degree in physical education and training from the Anshan Normal College, in 1995, and the master's degree in physical education from Hebei Normal University. In 1995, he worked at the front line of physical education. In 2002, he began to coach the school track and field training team, and accumulated a wealth of teaching and training experience. Since 1999, he has been presiding over the school sports teaching and research office. He is currently working with the Weifang University of Science and Technology, China. Over the years, he has published many high-quality articles in various academic journals. His research interest includes sport exercise training.

YUAN ZHANG received the bachelor's degree in sport education from Linyi Normal University, China, in 2006. He is currently working with the Weifang University of Science and Technology, China, as a Professor. He has published several high-quality articles in various academic journals. His research interests include sport exercise and training.

CAIZHONG LI received the bachelor's degree in sports training from the Shandong Institute of Physical Education, China, in 2010. He is currently working with the Weifang University of Science and Technology, China. Since 2010, he has been working at the front line of physical education. During this period, he accumulated rich teaching experience and published many articles in journals. His research interests include physical education and sport training.

CHAO YAN received the master's degree from the Institute of Computing Technology, Chinese Academy of Sciences, China, in 2006. He is currently a Lecturer with the School of Information Science and Engineering, Qufu Normal University, China. His research interests include recommender systems and services computing.



YUCONG DUAN (Senior Member, IEEE) received the Ph.D. degree in software engineering from the Institute of Software, Chinese Academy of Sciences, China, in 2006. He is currently a Full Professor and the Vice Director of the Computer Science Department, Hainan University, China. His research interests include software engineering, service computing, cloud computing, and big data.

HAO WANG (Senior Member, IEEE) received the B.Eng. and Ph.D. degrees in computer science and engineering from the South China University of Technology, Guangzhou, China, in 2006. He is currently an Associate Professor with the Norwegian University of Science and Technology, Gjøvik, Norway. He has authored or co-authored more than 130 papers in reputable international journals and conferences papers. His current research interests include big data analytics, the industrial Internet of Things, high performance computing, and safety-critical systems. He is a member of the IES Technical Committee on Industrial Informatics. He served as the TPC Co-Chair for the IEEE DataCom 2015, the IEEE CIT 2017, and the ES 2017. He is a Reviewer for many journals, such as the IEEE Transactions on Knowledge and Data Engineering, the IEEE Transactions on Industrial Informatics, and the IEEE Transactions on Big Data.