IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2020.Doi Number

Multi-source Medical Data Integration and Mining for Healthcare Services

Qingguo Zhang[1, 2], Bizhen Lian[1,*], Ping Cao[2], Yong Sang[2], Wanli Huang[3], Lianyong Qi[3]

[1] China Basketball College, Beijing Sport University, Beijing, China

[2] Department of Physical Education, Qufu Normal University, Rizhao, China

[3] School of Information Science and Engineering, Qufu Normal University, Rizhao, China

Corresponding author: Bizhen Lian (e-mail: lian_bizhen@163.com).

This research is partially supported by the National Natural Science Foundation of China (No. 61872219).

ABSTRACT With the advent of Internet of Health (IoH) age, traditional medical or healthy services are gradually migrating to the Web or Internet and have been producing a considerable amount of medical data associated with patients, doctors, medicine, medical infrastructure and so on. Effective fusion and analyses of these IoH data are of positive significances for the scientific disaster diagnosis and medical care services. However, IoH data are often distributed across different departments and contain partial user privacy. Therefore, it is often a challenging task to effectively integrate or mine the sensitive IoH data, during which user privacy is not disclosed. To overcome the above difficulty, we put forward a novel multi-source medical data integration and mining solution for better healthcare services, named PDFM (Privacy-free Data Fusion and Mining). Through PDFM, we can search for similar medical records in a time-efficient and privacy-preserving manner, so as to offer patients with better medical and health services. A group of experiments are enacted and implemented to demonstrate the feasibility of the proposal in this work.

INDEX TERMS Service recommendation; Internet of Health; Locality-Sensitive Hashing; User privacy; Data integration

I. INTRODUCTION

With the ever-increasing popularity of Information Technology and the gradual adoption of digital software in medical or healthy domains, various medical departments or agencies have accumulated a considerable amount of historical data (e.g., patients' medical records, healthy treatment solutions and so on), which form a main source of big Internet of Health (IoH) data [1]. The utilization degree of such IoH data is a key criterion to evaluate and quantify the information level of medical or healthy units or departments [2].

Generally, most of historical IoH data records contain valuable information especially for the medical or healthy agencies, such as the past disease of a patient at a time point. Mining and analyzing such historical IoH data records can contribute much to doctors' scientific and reasonable diagnosis and treatment decision-makings, as well as disaster trend prediction and precaution [3]. Therefore, it is of emergent necessity to collect, integrate, fuse and analyze these multi-source IoH data records for high-quality healthcare services suitable for patients.

However, historical IoH data records from patients often contain sensitive patient privacy (e.g., blood pressure, temperature) as a patient is often not willing to let others know his or her historical disasters [4]. Therefore, the patients or the stakeholders of historical IoH data records dare not disclose their IoH data records to the public. In addition, they lack sufficient incentive for IoH data records sharing with others. The above two concerns significantly block the utilization of historical IoH data records. As a consequence, although many hospitals or other medical & healthy agencies have accumulated a considerable amount of historical IoH data records, they seldom release the data to the outside due to privacy concerns. Furthermore, the historical IoH data records are often distributed across different platforms or agencies, the integration and fusion of which further increases the privacy disclosure concerns.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3023332, IEEE Access

Considering the above challenge, we use hash techniques to realize private data protection when the multi-source IoH data are integrated together for subsequent IoH data mining and analyses. As hash techniques are single-directional data mapping ways, the goal of privacy protection can be achieved accordingly.

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

(1) We introduce the LSH (Locality-Sensitive Hashing) into multi-source IoH data fusion and integration so as to secure the sensitive information of patients hidden in the past IoH data.

(2) For the IoH data without patient privacy after LSH process, we bring forth a similar IoH data record search method for subsequent IoH data mining and analyses, so as to balance the IoH data availability and privacy.

(3) Based on a dataset collected by real-world users, we validate the advantages of the proposed work in this paper, through a set of pre-designed experiments.

The reminder of this paper is structured as below. We review the current research status of the field in Section II to further show the novelty of our work. In Section III, we use an intuitive example to motivate our research. The proposed multi-source medical data integration and mining method is specified in Section IV. Experiment comparisons are presented in Section V. At last, in Section VI, conclusions are drawn and the future research work is described in detail.

II. RELATED WORK

Many researchers have devoted themselves to the research of multi-source big data integration as well as the resulted sensitive data protection problems. In this section, we summarize the current research status as below.

A. Encryption

Encryption is a classic and effective way to secure sensitive user data, which has been investigated for a long time. In [5], Peng T. et al brought forth a multi-keywords sorting-based secure search method, which adopt the symmetric public key search encryption way to permit a user to make secure information retrieval in an encrypted dataset based on multiple keywords. The advantage is that it realizes secure service protection for cloud computing with limited resources. The disadvantage is that its computational efficiency is not high enough. Besides, the key disclosure risks are also present. In [6], Dai H. et al introduced a kind of oval curve encryption method to realize secure data use and proved that the oval curve encryption-based method is superior to the traditional FP-based method. The advantage is that it has a relatively high data security performance. However, it only considered the simple Boolean value-based keyword search, which narrows the method's application scope to some extent. In [7], Phuong T. V. X., et al employed vector space model and homomorphic encryption technique to realize encrypted data ranking, as well as multi-keyword data retrieval and file retrieval. The advantage is that it can guarantee high-level quality data protection. The disadvantage is that it brings additional computational time and communication cost that are often very high. For sortable and multi-keyword data encryption problem, the authors in [8] put forward a homomorphic encryption-based data retrieval method to aid the data stakeholders as each data item ready to be searched is homomorphic encrypted during information retrieval process. The proposal can solve most of the secure data processing requirements, however, it cannot support the possible fuzzy retrieval.

B. Differentially Privacy

A differentially privacy-based improved collaborative filtering method named IPriCF was put forward in [9], to secure user privacy involved in collaborative data integration process. Through dividing user data and item data, IPriCF can effectively eliminate the disruption brought by noises incurred by differentially privacy. This method can balance user data privacy and accuracy of the recommended list. A stakeholder-feature-item matrix was built in [10] to analyze the sparse data and provide optimal services. The authors can guarantee the privacy-preservation of involved data while maintaining an acceptable prediction accuracy loss. A differentially private matrix factorization method named DPMF was brought forth in [11]: matrix factorization technique was used to convert sensitive user data into potential low-dimensional vectors; while differentially privacy technique was used to confuse the targeted object functions. However, when the number of dimensions grows, the prediction accuracy is reduced accordingly. The authors in [12] improved the TrustSVD model by introducing differentially privacy and further propose a new model named DPTrustSVD. The new model can effectively reach a tradeoff among data privacy, data sparsity and data availability. Other similar work includes [13] in which the authors combined Differentially Privacy and Huffman Coding to put forward a privacy-aware location segments publishing method, and [14] in which the authors combined Differentially Privacy, Bayes network and entropy theory to bring forth a protection method for high-dimensional data.

C. Anonymization

Anonymization is an effective way for securing the sensitive user data when making big data analyses and mining [15]. Through hiding certain sensitive information (e.g., name, identity card no.) contained in data, anonymization can publish the rest of data (i.e., data after anonymization) to the public so as to achieve the tradeoff between data privacy and availability [16]. K-anonymity solution is adopted in [17] to hide the key sensitive

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3023332, IEEE Access

information involved in data-driven decision-making process. A K-anonymity-based user location protection method is suggested in [18] to hide the real user location or position. Although the above solutions can help to hide sensitive user data when performing data-driven business analyses and applications, they cannot balance the data privacy and data utilization well as anonymized data would lose certain key information more or less.

With the above summarization, it is noticed that although many big data fusion and mining solutions (such as [19-24]) have been proposed, they cannot make a balancing tradeoff among multiple conflicting criteria such as data security, data availability and so on. Considering this drawback of existing literatures, we look for better multi-source data fusion methods through another kind of privacy-preservation technique, i.e., hash. The details of our suggested solution will be described step by step in the following sections.

III. Motivation

Figure 1 shows the motivation of our paper, in which the doctor-medicine-nurse medical records of patients are partially located in cloud platforms cp_1 and cp_2, respectively. To comprehensively mine the valuable information from the IoH data distributed across platforms cp_1 and cp_2, we need to integrate and fuse these multi-source data for uniform data analyses and make more scientific healthcare decisions.

However, in the above IoH data fusion and analyses process, some privacy concerns are often raised as the historical IoH data records often contain partial sensitive information of patients. To encourage platforms cp_1 and cp_2 to release their IoH data records and alleviate the patients' privacy disclosure concerns, it is necessary to develop a novel data fusion method without revealing privacy.

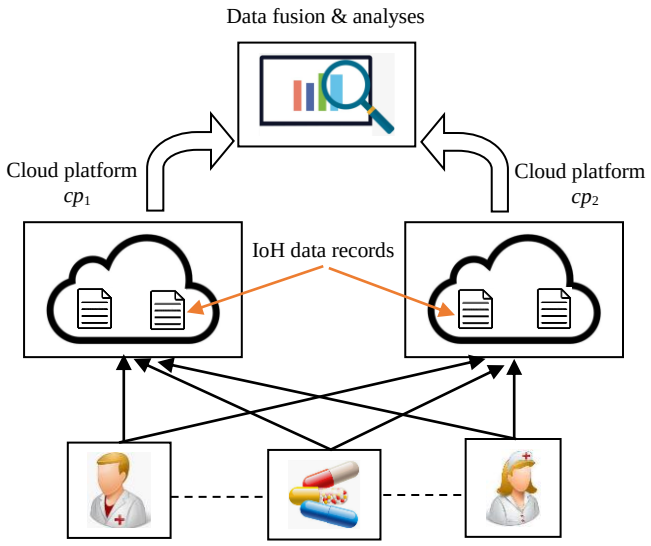


FIGURE 1. Multi-source IoH data fusion

Therefore, we explore a multi-source IoH data fusion method without privacy concerns in the next section. Furthermore, for better describe the details of our suggested data fusion method without revealing privacy information, we summarize the used symbols and their respective meanings in the method with TABLE 1.

TABLE 1. Symbol specifications

Symbols	Specification
$R_1, \ldots, R_n$	IoH data records
$q_1, \ldots, q_m$	Healthcare criteria
$f_1, \ldots, f_a$	Hash functions
$T_1, \ldots, T_b$	Hash tables
$cp_1, \ldots, cp_h$	Distributed cloud platform
$v_1, \ldots, v_m$	M dimensions of each hash function
$h_1(R_x), \ldots, h_a(R_x)$	Hash values of R_x based on $f_1, \ldots, f_a$
$H_1(R_x), \ldots, H_b(R_x)$	Indices of R_x in hash tables $T_1, \ldots, T_b$

IV. Approach

This section presents our proposed multi-source IoH data fusion and mining method, whose major procedure can be generalized with the following steps: First, the sensitive IoH data are projected based on LSH functions. Second, according to each IoH data record and its corresponding hash values derived after hash projection, we create a set of hash tables without patient privacy. Third, according to the derived hash tables, we make similar IoH data search and mining. In summary, the detailed three steps are listed in FIGURE 2.

Step-1: LSH-based IoH data projection. We randomly choose a group of LSH functions to project each piece of IoH data record. This way, for each piece of IoH data record, we obtain a group of hash values.

Step-2: Creation of hash tables without privacy. According to the "IoH data record-hash values" paris, we create a group of hash tables without much patient privacy. Each hash table includes the index of each IoH data record.

Step-3: Hash tables-based similar IoH data search and mining. According to the hash tables, we cluster the IoH data records into multiple groups and each group of records share the similar IoH data. Furthermore, we mine and analyze the IoH data records based on the derived multiple clusters.

FIGURE 2. Three steps of our proposal

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3023332, IEEE Access

(1) Step-1: LSH-based IoH data projection.

We use $R_i.q_j$ to denote the value of dimension q_j (j = 1, 2, …, m) of historical IoH data record R_i (i = 1, 2, …, n) from a patient. As $R_i.q_j$ is often sensitive to the patient, we need to secure the private information of $R_i.q_j$ when $R_i.q_j$ is published to the public. In this step, we use LSH strategy to achieve this goal.

For R_i (i = 1, 2, …, n), it has m criteria q_1, …, q_m. Therefore, the healthy information corresponding to R_i can be depicted by symbol $\overrightarrow{R_i}$ = ($R_i.q_1$, …, $R_i.q_m$). When releasing $\overrightarrow{R_i}$ to others, we need to make an LSH projection first. In concrete, we create a new vector V = (v_1, …, v_m) where v_j (j = 1, 2, …, m) is a randomly generated value from domain [-1, 1]. Thus, we create an LSH function f(.) as in equation (1).

$$\begin{aligned} f(\overrightarrow{R_i}) &= \overrightarrow{R_i} \bullet V \\ &= (R_i.q_1, \ldots, R_i.q_m) \bullet (v_1, \ldots, v_m) \\ &= R_i.q_1 * v_1 + \ldots + R_i.q_m * v_m \end{aligned} \quad (1)$$

Afterwards, we can get a $f(\overrightarrow{R_i})$ which may be either positive or negative. Next, we make the following mapping as in equation (2). Thus, through mapping, we convert $f(\overrightarrow{R_i})$ into $h(R_i)$ of a binary value: 0 or 1. Concrete procedure can be shown in Algorithm 1.

$$h(R_i) = \begin{cases} 1 & \text{if } f(\overrightarrow{R_i}) > 0 \\ 0 & \text{if } f(\overrightarrow{R_i}) \le 0 \end{cases} \quad (2)$$

Algorithm 1

Inputs:
(1) R_1, …, R_n: historical IoH data records;
(2) q_1, …, q_m: quality dimensions of IoH data.

Output:
(1) $h(R_i)$: Boolean value of $\overrightarrow{R_i}$ after mapping.

```
for j = 1, …, m do
    v_j = random [-1, 1]
end for
for i = 1, …, n do
    sum = 0
    for j = 1, …, m do
        sum + = R_i.q_j * v_j
    end for
    if sum > 0
        then f(R_i) = 1
    else f(R_i) = 0
    end if
    return f(R_i)
end for
```

(2) Step-2: Creation of hash tables without privacy.

The $f(\overrightarrow{R_i})$ derived in Step-1 can be regarded as a hash value of R_i through a projection process. However, one projection process is not enough for convert R_i into a privacy-free index. Considering this, we repeat Algorithm 1 for multiple times by projections of f_1, …, f_a, and after which we get an a-dimensional hash vector $H(R_i)$ as in equation (3). Thus the mappings of "$R_i \rightarrow H(R_i)$" (i = 1, 2, …, n) constitute a hash table, denoted by "T". In other words, through "T", we can query about the index value of R_i; on the contrary, given an index value R_i, we cannot infer the real value of R_i. This way, the privacy of patients contained in R_i is secured.

$$H(R_i) = (h_1(R_i), \ldots, h_a(R_i)) \quad (3)$$

However, one hash table "T" cannot always accurately reflect the real index of each IoH data record. Considering this limitation, we repeat the creation process of "T" several times and get b tables: T_1, …, T_b. This step can be formalized with the pseudo code of Algorithm 2.

Algorithm 2

Input:
(1) $h(R_1)$, …, $h(R_n)$: Boolean values of IoH data records;
(2) f_1, …, f_a: LSH functions.

Output:
(1) T_1, …, T_b: b hash tables.

```
for x = 1, …, a do
    repeat Algorithm 1 based on f_x
end for
for i = 1, …, n do
    H(R_i) = (h_1(R_i), …, h_a(R_i))
    put "R_i → H(R_i)" into T
end for
return T
repeat lines 1-8 b times
```

(3) Step-3: Hash tables-based similar IoH data search and mining.

In Step-2, b tables: T_1, …, T_b are generated. And in each table, there would be a set of corresponding "$R_i \rightarrow H(R_i)$" (i = 1, 2, …, n) pairs. Furthermore, $H(R_i)$ is approximately regarded as the index of R_i in the table. According to Locality-Sensitive Hashing theory, the IoH data records with the same index would be approximately similar. As a result, if two records R_1 and R_2 share the same index, then R_1 and R_2 are probably similar records. This way, we can mine the potential similar IoH data records through check their respective index values without much privacy.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3023332, IEEE Access

However, for two IoH data records R_1 and R_2, $H(R_1) = H(R_2)$ is a rather rigid constraint condition as each dimensional value of $H(R_1)$ should be exactly equal to that of $H(R_1)$. Such a rigid constraint condition is apt to produce an empty result of similar IoH data records search, which makes nonsense to privacy-free IoH data fusion and mining.

Considering this drawback, we relax the above rigid condition by generating more hash tables instead of only one. In concrete, considering the b tables created in Step 2, i.e., T_1, …, T_b, if $H(R_1) = H(R_2)$ holds in any T_y (y = 1, 2, …, b), then it is simply conclude that R_1 and R_2 are probably similar IoH data records. Thus, the similar IoH data records search condition is relaxed accordingly. Therefore, for a specific IoH data record R_x, we can look for its similar record set Sim_Set (R_x) through the above idea. Details of this step are presented in Algorithm 3. And finally, we return Sim_Set (R_x) as the final output of the proposal in this work.

Algorithm 3

Inputs:
(1) T_1, …, T_b: b hash tables;
(2) R_1, …, R_n: historical IoH data records;
(3) R_x: a target IoH data record whose similar records are required.

Output:
Sim_Set (R_x) : similar IoH data records of R_x

```
1: Sim_Set (Rx) = Φ
2: for y = 1 to b do
3:    for i = 1, …, n do
4:       if H(Ri) = H(Rx)
5:          then put Ri into Sim_Set (Rx)
6:       end if
6:    end for
7: end for
8: return Sim_Set (Rx)
```

IV. EXPERIMENTS

To validate the effectiveness of our solution in dealing with privacy-free data fusion and mining (abbreviated as PDFM), a group of experiments are designed and compared with existing approaches.

A. CONFIGURATION

We use the public data released by http://inpluslab.com/wsdream/ for simulation purpose. In the dataset, each user-item-qos pair is taken as a patient-criterion-value pair in multi-source IoH data fusion scenarios. 90% entries of the dataset are used to train the parameters of the data fusion and mining model, while the reminder of 10% entries are employed for test and validation.

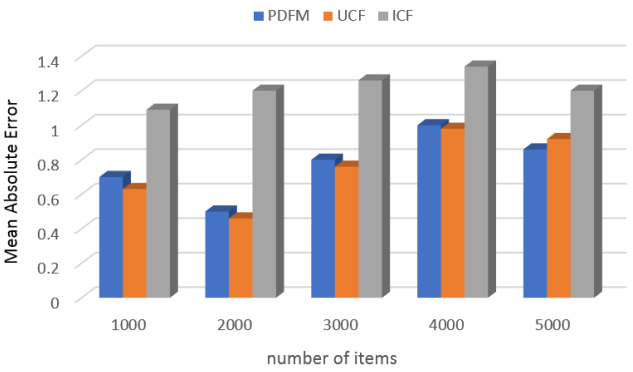
To show the competitive advantages of PDFM, UCF (baseline) and ICF methods are used to compare with PDFM. The compared metrics include missing data prediction accuracy (Mean Absolute Error) and computational time (s). The software and hardware configurations include: 2.80 GHz processor, 8.0 GB memory, Windows 7 operation system and JAVA 8. We run each experiment 50 times and report their average.

B. COMPARISONS

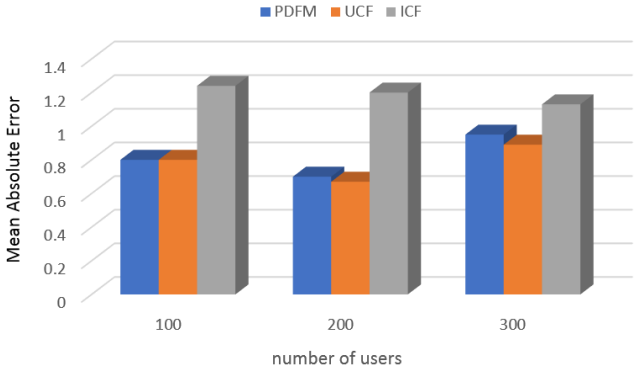
Parameters are of the following values: a = {2, 4, 6, 8, 10}, b = {2, 4, 6, 8, 10}. Concrete comparison results are presented in detail as follows.

(1) Mean Absolute Error comparison.

We measure and compare the Mean Absolute Error of three methods. The parameter settings are as follows: the user volume is 339, item volume is varied from 1000 to 5000, a = b = 10. Here, two sets of experiments are tested, respectively. First, we test the variation trend of Mean Absolute Error of three methods with the change of the number of items in the used dataset. Second, we test the variation trend of Mean Absolute Error of three methods with the change of the number of users in the dataset.



(a) number of users = 339



(b) number of items = 5825

FIGURE 3. Mean Absolute Error comparison

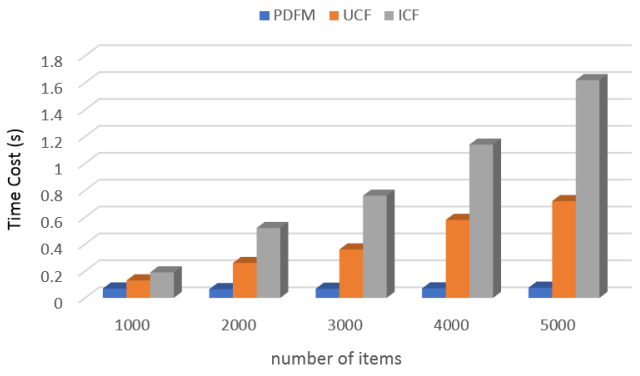
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3023332, IEEE Access

Comparison results are shown in Fig.3. In both Fig.3(a) and Fig.3(b), we can observe a clear advantage of PDFM and UCF compared to ICF, as UCF is a baseline method and PDFM is an approximate solution to UCF. Besides, PDFM achieves an approximate Mean Absolute Error of UCF as the LSH strategy adopted in PDFM can promise a good similarity-maintenance property. Moreover, PDFM has an advantage of privacy-preservation capability which is not owned by UCF.

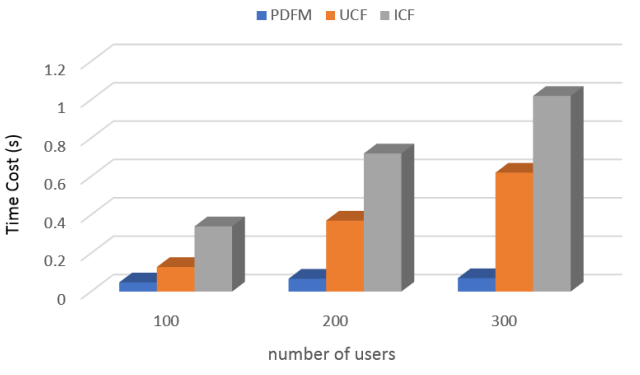
(2) Computational time comparison

We measure and compare the computational time of three methods. The parameter settings are as follows: the user volume is varied from 100 to 300, item volume is varied from 1000 to 5000, a = b = 10. Compared data are reported in Fig.4.

As can be seen from Fig.4, the consumed time of three methods approximately grows when the number of users or the number of items rises. Specifically, UCF and ICF consumes more time than PDFM as heavy-weight user similarity calculation or item similarity calculation is required in UCF and ICF, respectively. While in PDFM, the time cost can be divided into two parts: (1) hash table creation, which can be finished offline; as a consequence, the time complexity is of O(1); (2) similar IoH data record retrieval, which needs to be done online and its time complexity is O(1). As a result, PDFM can often return similar IoH data records within a small response time and hence, our method can be applied to the big IoH data environment.



(a) number of users = 339



(b) number of items = 5825

FIGURE 4. Computational time comparison

(3) Mean Absolute Error of PDFM

PDFM method is based on LSH strategy whose performances are often related to some key factors including parameters a and b. Considering this, we observe the performances of PDFM associated with a and b. The parameter settings are as follows: the user volume is 339, item volume is 5825, a = {2, 4, 6, 8, 10}, b = {2, 4, 6, 8, 10}. Compared data are reported in Fig.5.

As reported in Fig.5, the Mean Absolute Error of PDFM increases with the rise of parameter b and the decline of parameter a. This is due to the following reasons: (1) when there are more hash tables (i.e., b increases), the similar IoH data record retrieval condition becomes looser; as a result, more similar records are returned and correspondingly, the Mean Absolute Error is rising; (2) when there are more hash functions (i.e., a increases), the similar IoH data record retrieval condition becomes stricter; as a result, less similar records are returned and correspondingly, the Mean Absolute Error is decreased. Moreover, we can observe that more hash functions (i.e., a larger a) and less hash tables (i.e., a smaller b) will bring better prediction accuracy.

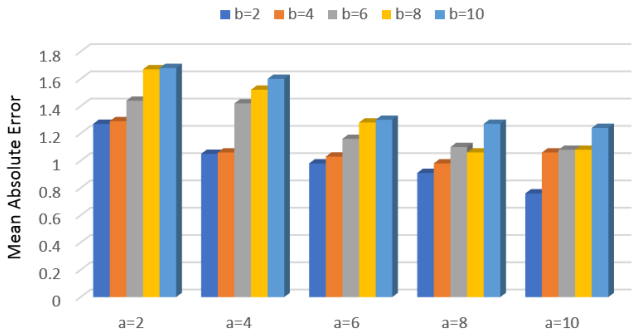


FIGURE 5. Mean Absolute Error of PDFM w.r.t. (a, b) pairs

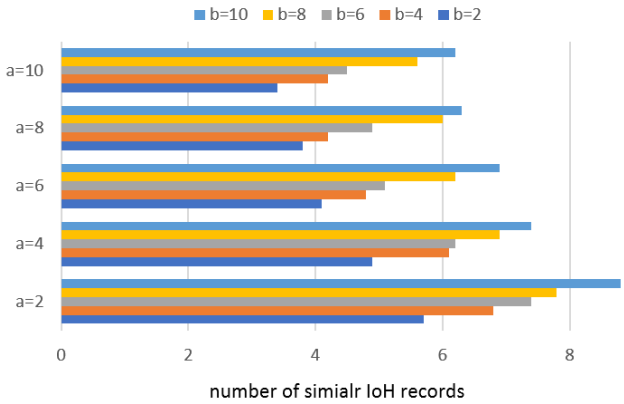


FIGURE 6. Number of returned results of PDFM w.r.t. (a, b) pairs

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3023332, IEEE Access

(4) Number of returned results of PDFM

As analyzed in the above analysis, PDFM method is based on LSH strategy whose returned result volume is often related to some key factors such as parameters a and b. Considering this, we observe the returned result volume of PDFM associated with a and b. The parameter settings are as follows: the user volume is 339, item volume is 5825, a = {2, 4, 6, 8, 10}, b = {2, 4, 6, 8, 10}. Compared data are reported in Fig.6.

As reported in Fig.6, the returned result volume of PDFM increases with the rising of parameter b and the dropping of parameter a. This is due to the following reasons: (1) when there are more hash tables (i.e., b increases), the similar IoH data record retrieval condition becomes looser; as a result, more similar records are returned; (2) when there are more hash functions (i.e., a increases), the similar IoH data record retrieval condition becomes stricter; as a result, less similar records are returned. Moreover, we can observe that more hash functions (i.e., a larger a) and less hash tables (i.e., a smaller b) will bring fewer returned results.

C. FURTHER DISCUSSIONS

The gradual popularity of big data technology has enabled a number of successful big IoH applications [25-27]. However, in our suggested IoH data fusion and mining method PDFM, there are several limitations.

(1) First of all, we only consider a simple IoH data of continuous type without considering the possible data type diversity (e.g., continuous data, discrete data, Boolean data) and data structure diversity.

(2) Second, how to measure the capability of securing sensitive patient information is still not introduced in PDFM.

(3) Third, how to further fuse different privacy protection solutions [28-31] for better performances is still an open problem that calls for future study.

(4) There is often an intrinsic tradeoff between data privacy and data availability. So it is inevitable to reduce the data availability when protecting the data privacy. As a result, our proposed LSH-based privacy-aware data fusion method cannot always guarantee 100% data availability. However, due to the inherent property of LSH, our proposal can guarantee 99.99% prediction accuracy if appropriate parameters are selected.

(5) At last, the reason that we choose LSH technique for privacy protection goal is that: LSH has a good property of similarity keeping. Concretely, if two points are neighboring points, then they would be projected into the same bucket after hash projection by LSH.

VI. CONCLUSION

Effective fusion and analyses of IoH data are of positive significances for scientific disaster diagnosis and medical care services. However, the IoH data produced by patients are often distributed across different departments and contain partial patient privacy. Therefore, it is often a challenging task to effectively integrate or mine the sensitive IoH data without disclosing patient privacy. To tackle this challenge, we bring forth a novel multi-source medical data integration and mining solution for better healthcare services, named PDFM. Through PDFM, we can search for similar medical records in a time-efficient and privacy-preserving manner, so as to provision patients with better medical and health services. The experiments on a real dataset prove the feasibility of PDFM.

In upcoming research, we will update the suggested PDFM method by considering the possible diversity of data types [32-34] and data structure [35-38]. In addition, how to fuse multiple existing privacy solution for better performances is still an open problem that requires intensive and continuous study.

REFERENCES

[1] S. Din, A. Paul. Smart health monitoring and management system: toward autonomous wearable sensing for Internet of Things using big data analytics. Future Generation Computer Systems, 111: 939-939, 2020.

[2] C. B. Natalie, C. V. Tiffany, J. S. Cynthia. Broadband Internet access is a social determinant of health! American Journal of Public Health, 110(8): 1123-1125, 2020.

[3] E. Sillence, J. Blythe, P. Briggs. A revised model of trust in internet-based health information and advice: cross-sectional questionnaire study. Journal of Medical Internet Research, 21(11), Article e11125, 2019.

[4] K. Szulc, M. Duplaga. The impact of Internet use on mental wellbeing and health behaviours among persons with disability. European Journal of Public Health, 29(4), 2019.

[5] T. Peng, Y. Lin, X. Yao, W. Zhang. An efficient ranked multi-keyword search for multiple data owners over encrypted cloud data. IEEE ACCESS, 6: 21924-21933, 2018.

[6] H. Dai, Y. Ji, G. Yang, H. Huang and X. Yi. A privacy-preserving multi-keyword ranked search over encrypted data in hybrid clouds. IEEE Access, 8: 4895-4907, 2020.

[7] T. Phuong, G. Yang, W. Susilo, et al. Sequence aware functional encryption and its application in searchable encryption. Journal of Information Security & Applications, 35: 106-118, 2017.

[8] Z. Xia, X. Wang, X. Sun, et al. A secure and dynamic multikeyword ranked search scheme over encrypted cloud data. IEEE Transactions on Parallel and Distributed Systems, 27(2): 340-352, 2016.

[9] M. He, M. Chang, X. Wu. A collaborative filtering recommendation method based on differential privacy. Journal of Computer Research and Development, 54(7): 1439-1451, 2017.

[10] T. Wang, S. He. An improved collaborative filtering recommendation algorithm with differentially privacy. Information Security and Technology, 7(4): 26-28, 2016.

[11] Z. Xian, Q. Li, X. Huang, J. Lu, L. Li. Differential privacy protection for collaborative filtering algorithms with explicit and implicit trust. Acta Electronica Sinica, 46(12): 3050-3059, 2018.

[12] C. Yin, L. Shi, R. Sun & J. Wang. Improved collaborative filtering recommendation algorithm based on differential privacy protection. The Journal of Supercomputing, 76: 5161-5174, 2020.

[13] Y. Xiao, L. Xiong, S. Zhang, Y. Cao. LocLok: location cloaking with differential privacy via hidden Markov model. Proceedings of the VLDB Endowment, 10(12): 1901-1904, 2017.

[14] X. Ren, J. Xu, X. Yang, et al. Bayesian network-based high-dimensional crowdsourced data publication with local differential

privacy (in Chinese). SCIENTIA SINICA Informationis, 49: 1586-1605, 2019.

[15] J. Wang, Z. Cai, Y. Li, D. Yang, J. Li, H. Gao. protecting query privacy with differentially private k-anonymity in location-based services. Personal and Ubiquitous Computing, 22: 453-469, 2018.

[16] S. Zhang, X. Li, Z. Tan, T. Peng, G. Wang. A caching and spatial k-anonymity driven privacy enhancement scheme in continuous location-based services. Future Generation Computer Systems, 94: 40-50, 2019.

[17] F. Casino, J. Domingo-Ferrer, C. Patsakis, D. Puig and A. Solanas. A k-anonymous approach to privacy preserving collaborative filtering. Journal of Computer and System Sciences, 81(6): 1000-1011, 2015.

[18] P. Zhao, J. Li, F. Zeng. ILLIA: enabling k-anonymity-based privacy preserving against location injection attacks in continuous LBS queries. IEEE Internet of Things Journal, 5(2): 1033-1042, 2018.

[19] S. Wang, Y. Zhang, Y. Li, et al. Single Slice based Detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization. Multimedia Tools and Applications. 77(9): 10393-10417, 2018.

[20] S. Wang, J. Sun, I. Mehmood, et al. Cerebral micro‐bleeding identification based on a nine‐layer convolutional neural network with stochastic pooling, Concurrency and Computation: Practice and Experience, 32(1): e5130, 2020.

[21] Y. Zhang, V. Govindaraj, C. Tang, et al. High performance multiple sclerosis classification by data augmentation and AlexNet transfer learning model, Journal of Medical Imaging and Health Informatics, 9(9): 2012-2021, 2019.

[22] Y. Zhang, S. Wang, Y. Sui, et al. Multivariate approach for Alzheimer's disease detection using stationary wavelet entropy and predator-prey particle swarm optimization, Journal of Alzheimer's Disease, 65(3): 855-869, 2018.

[23] C. Kang, X. Yu, S. Wang, et al. A heuristic neural network structure relying on fuzzy logic for images scoring, IEEE Transactions on Fuzzy Systems, 2020, doi: 10.1109/TFUZZ.2020.2966163.

[24] S. Wang, Y. Zhang, M. Yang, et al. Unilateral sensorineural hearing loss identification based on double-density dual-tree complex wavelet transform and multinomial logistic regression, Integrated Computer-Aided Engineering, 26: 411-426, 2019.

[25] S. Wang, J. Sun, P. Phillips, et al. Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units, Journal of Real-Time Image Processing, 15(3): 631-642, 2018.

[26] S. Wang, C. Tang, J. Sun, Y. Zhang. Cerebral micro-bleeding detection based on Densely connected neural network, Frontiers in Neuroscience, 13, Article ID: 422, 2019.

[27] S. Wang, S. Xie, X. Chen, et al. Alcoholism identification based on an AlexNet transfer learning model, Frontiers in Psychiatry, 10, Article ID: 205, 2019.

[28] L. Qi, X. Wang, X. Xu, W. Dou, S. Li. Privacy-Aware Cross-Platform Service Recommendation based on Enhanced Locality-Sensitive Hashing. IEEE Transactions on Network Science and Engineering, 2020. DOI: 10.1109/TNSE.2020.2969489.

[29] 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 Transactions on Services Computing, vol. 13, no. 2, pp: 289-300, 2020.

[30] L. Qi, C. Hu, X. Zhang, M. R. Khosravi, S. Sharma, S. Pang, T. Wang. Privacy-aware Data Fusion and Prediction with Spatial-Temporal Context for Smart City Industrial Environment. IEEE Transactions on Industrial Informatics. 2020. DOI: 10.1109/TII.2020.3012157.

[31] Y. Xu, C. Zhang, G Wang, Z. Qin, and Q. Zeng, "A Blockchain-enabled Deduplicatable Data Auditing Mechanism for Network Storage Services," IEEE Transactions on Emerging Topics in Computing, 2020. DOI: https://doi.org/10.1109/TETC.2020.3005610.

[32] W. Zhong, X. Yin, X. Zhang, S. Li, W. Dou, R. Wang, L. Qi. Multi-Dimensional Quality-Driven Service Recommendation with Privacy-Preservation in Mobile Edge Environment. Computer Communications, 157: 116-123, 2020.

[33] X. Chi, C. Yan, H. Wang, W. Rafique, L. Qi. Amplified LSH-based Recommender Systems with Privacy Protection. Concurrency and Computation: Practice and Experience, 2020. DOI: 10.1002/CPE.5681.

[34] C. Zhou, A. Li, A. Hou, Z. Zhang, Z. Zhang, F. Wang. Modeling Methodology for Early Warning of Chronic Heart Failure Based on Real Medical Big Data. Expert Systems with Applications, 2020, DOI: 10.1016/j.eswa.2020.113361.

[35] T. Cai, J. Li, A. S. Mian, R. Li, T. Sellis, J. X. Yu. Target-aware holistic influence maximization in spatial social networks. IEEE Transactions on Knowledge and Data Engineering, 2020. DOI: 10.1109/TKDE.2020.3003047.

[36] L. Qi, Q. He, F. Chen, X. Zhang, W. Dou, Q. Ni. Data-Driven Web APIs Recommendation for Building Web Applications. IEEE Transactions on Big Data, 2020. DOI: 10.1109/TBDATA.2020.2975587.

[37] H. Liu, H. Kou, C. Yan and L. Qi. Keywords-Driven and Popularity-Aware Paper Recommendation Based on Undirected Paper Citation Graph. Complexity, Volume 2020, Article ID 2085638, 15 pages, 2020.

[38] J. Li, T. Cai, K. Deng, X. Wang, T. Sellis, F. Xia. Community-diversified influence maximization in social networks. Information Systems, Vol. 92, pp. 1-12, 2020.