International Journal of Early Childhood Special Education (INT-JECSE)
DOI:10.9756/INTJECSE/V14I5.67 ISSN: 1308-5581 Vol 14, Issue 05 2022

MULTI-SOURCE MEDICAL DATA INTEGRATION AND MINING FOR HEALTHCARE SERVICES

Dr.C.K. Gomathy
Assistant Professor, Dept. Of CSE, SCSVMV (Deemed to be University), Kanchipuram, TamilNadu, India

K.Nishanth Reddy
Student, Dept. Of CSE, SCSVMV (Deemed to be University), Kanchipuram, TamilNadu, India

K.Sai Abhishek,
Student, Dept. Of CSE, SCSVMV (Deemed to be University), Kanchipuram, TamilNadu, India

Dr.V Geetha
Assistant Professor, Dept. Of CSE, SCSVMV (Deemed to be University), Kanchipuram, TamilNadu, India

ABSTRACT: As the Internet of Health (IoH) era dawns, conventional medical or healthcare resources are gradually migrating to the web or the internet, resulting in a massive influx of medical data relating to patients, physicians, pharmaceuticals, medical infrastructure, and so on. This IoH data's good integration and analysis are ideal indicators for disaster diagnosis and medical care services. However, IoH is frequently divided into other departments and protects the users' privacy. As a result, compiling or extracting critical IoH data, where user privacy may be compromised, is frequently a difficult operation. To address the aforementioned challenges, we focus on PDFM, a multi-source medical data collecting and mining solution for improved health care services (Data Fusion and Private Mining). We can search for similar medical data in a time-saving and private manner with PDFM, allowing us to deliver better medical and healthcare services to patients. To show the viability of a plan for this work, a test team is formed and employed.

INDEX TERMS: Service recommendations, Internet Health, site-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 healthcare domains, various medical departments or agencies have accumulated a large amount of historical data (e.g., patient medical records, healthy medical solutions, etc.), creating a huge source of data. Internet Health (IoH) [1]. The level of use of that IoH data is a key condition for evaluating and measuring the knowledge level of medical or health units or departments [2]. Most historical IoH data records offer useful information, particularly for medical and health authorities, such as a patient's prior sickness at a certain time point. Mining and analyzing historical IoH data records can help clinicians make more scientific and fair diagnoses and treatment decisions, as well as identify catastrophic trends and prepare for them [3]. As a result, collecting, integrating, fusing, and analyzing these multi-source IoH data sets for high-quality healthcare services acceptable for patients is a pressing need.

Individuals' past IoH data records, on the other hand, frequently contain sensitive patient privacy information (e.g., blood pressure, temperature), as patients are typically reluctant to tell others about their previous tragedies [4]. As a result, previous IoH data tracks patients or stakeholders. As a result, even though many hospitals and other medical and health institutions have amassed large amounts of historical IoH data records, they seldom share these with the public owing to privacy concerns. Furthermore, previous IoH data sets are frequently dispersed across several platforms or organizations, complicating the integration and fusion process and raising privacy problems.

In light of the aforementioned problem, we employ hash algorithms to ensure private data privacy when multi-source IoH and data are combined for further IoH data mining investigations.

In conclusion, the work in this study makes three key contributions.

(1) We use LSH (Locality-Critical Hashing) in multi-source IoH data fusion and integration to protect sensitive patient information buried in previous IoH data.

(2) We propose a comparable IoH data record search strategy for further IoH data mining and analysis for IoH data without patient privacy following the LSH procedure, in order to balance IoH data availability and privacy.

(3) We validate the benefits of the suggested work in this study using a dataset acquired from real-world users and a set of pre-designed trials.

This paper's reminder is organized as follows. In Section II, we evaluate the present state of research on the subject to demonstrate the uniqueness of our approach. We give an intuitive example to motivate our study in Section III. Section IV details the suggested multi-source medical data integration and mining approach. Section V compares the results of the experiments. Finally, in Section VI, conclusions are formed and a detailed description of future study efforts is provided.

II. RELATED WORK

Many studies have focused on multi-source big data integration as well as the sensitive data protection issues that have arisen as a result. The current research state is summarised in this section.

A. *ENCRYPTION*

Encryption is a tried-and-true method of securing sensitive user data that has been studied for a long time. Peng T. et al. proposed a multi-keywords sorting-based safe search approach in [5], which uses symmetric public key search encryption to allow a user to securely retrieve information from an encrypted dataset using multiple keywords. The benefit is that it provides secure service protection for cloud computing on a budget. Its computing efficiency is insufficient, which is a downside. Additionally, the main disclosure risks are present. Dai H. et al. [6] presented a type of oval curve encryption technology to provide safe data use and showed that the oval curve encryption-based approach outperforms the classic FP-based method. It has the benefit of having a pretty excellent data security performance. However, it only took into account simple Boolean value-based keyword searches, which limits the method's applicability. To achieve encrypted data ranking, as well as multi-keyword data retrieval and file retrieval, Phuong T. V. X., et al used a vector space model and homomorphic encryption approach [7]. It has the benefit of providing high-quality data protection. The drawback is that it adds to the amount of time it takes to compute and the amount of money it costs to communicate. The authors in [8] propose a homomorphic encryption-based data retrieval approach to help data stakeholders with sortable and multi-keyword data encryption problems since each data item ready to be searched is homomorphic encrypted throughout the information retrieval process. The idea can address the majority of secure data processing needs, however, it does not address the possibility of fuzzy retrieval.

B. *DIFFERENTIALLY PRIVACY*

To protect user privacy during the collaborative data integration process, [9] proposed a differentially privacy-based enhanced collaborative filtering approach called IPriCF. IPriCF can effectively prevent the interruption caused by sounds incurred by differential privacy by splitting user data and item data. This strategy can strike a balance between the privacy of user data and the accuracy of the recommended list. To assess the scarce data and deliver optimal services, [10] created a stakeholder-feature-item matrix. The authors can ensure that the privacy of the data involved is protected while retaining a reasonable prediction accuracy loss. In [11], the differentially private matrix factorization method DPMF was proposed: matrix factorization was utilized to turn sensitive user data into possible low-dimensional vectors, while differentially privacy was employed to mislead the targeted object functions. However, as the number of dimensions increases, the prediction accuracy decreases. [12] proposes a novel model called DPTrustSVD that improves the TrustSVD model by providing differentially privacy. The novel approach successfully achieves a compromise between data privacy, data sparsity, and data availability. Other related work includes [13], in which the authors combined Differentially Privacy and Huffman Coding to present a privacy-aware location segments publishing method, and [14], in which the authors combined Differentially Privacy, Bayes network, and entropy theory to present a high-dimensional data protection method.

C. ANONYMIZATION

When performing large data analysis and mining, anonymization is an excellent technique to protect sensitive user data [15]. Anonymization can achieve the tradeoff between data privacy and availability by masking specific sensitive information (e.g., name, identification card number) contained in data and publishing the remainder of the data (i.e., data after anonymization) to the public [16]. [17] uses the K-anonymity solution to conceal the main sensitive information involved in data-driven decision-making. proposes a K-anonymity-based user location protection solution for concealing the true user's location or position.

Although the aforementioned methods can assist in the concealment of sensitive user data when executing data-driven business analyses and applications, they are unable to strike a good balance between data privacy and data use since anonymized data would lose some critical information.

With the above summary, it is clear that, even though many big data fusion and mining solutions have been presented, they are unable to strike a balance between many competing requirements such as data security, data availability, and so on. Given this limitation of existing research, we want to improve multi-source data fusion approaches using a different type of privacy-preserving technology, namely hashing. The following sections will walk you through the specifics of our recommended remedy step by step.

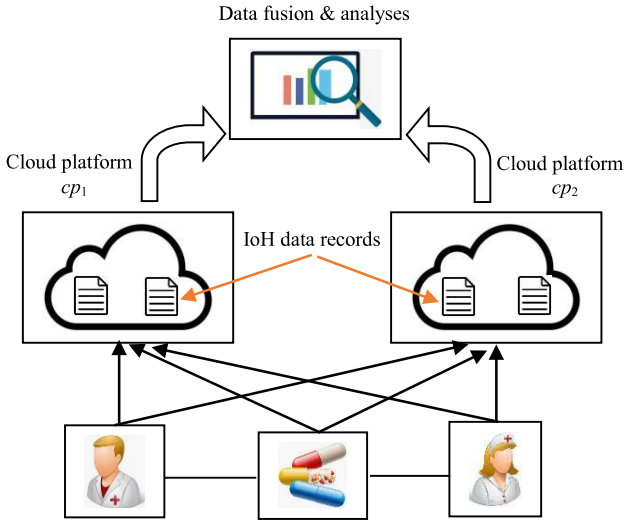


FIGURE 1. Multi-source IoH data fusion.

III. MOTIVATION

In Figure 1, medical records of doctor-nurse patients are partially available on cloud platforms cp1 and cp2, respectively, which served as motivation for our paper. We need to assemble and integrate this multi-source data to study the same data and make more scientific health care judgments to properly mine crucial information from IoH data spread across cp1 and cp2 domains.

However, further privacy problems are frequently highlighted throughout the aforementioned process of merging IoH data with analysis, as IoH historical data sets frequently contain incomplete patient-sensitive information.

It is vital to establish a new approach to gathering data without compromising confidentiality to persuade cp1 and cp2 forums to provide IoH data records and soothe worries about patient confidentiality.

As a result, in the following part, we'll look at how to integrate multi-source IoH data without sacrificing privacy. In addition, we outline the symbols used and their meanings in TABLE 1 to better explains the intricacies of our suggested data gathering approach without compromising sensitive information.

This section outlines our proposed multi-source IoH data fusion and mining approach, which may be summarised using the stages below: First, LSH functions are used to protect sensitive IoH data. Second, each IoH data record and its related hash values are examined.

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 $T_1, \ldots, T_b$

We generate a collection of hash tables without patient privacy using the results of hash projection. Third, we do analogous IoH data search and mining using the generated hash tables. In summary, FIGURE 2 shows the three phases in detail

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.

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

The value of dimension q_j (j = 1, 2, ..., m) of a history IoH data record R_i (i = 1, 2, ..., n) from a patient is denoted by $R_i.q_j$. We need to secure the private information of $R_i.q_j$ when $R_i.q_j$ is

disclosed to the public since $R_i.q_j$ is typically sensitive to the patient. To accomplish this purpose, we will apply the LSH technique in this stage.

For R_i (i = 1, 2, ..., n), it has m criteria q_1, ..., q_m.

As a result, the sign R_i may be used to represent the healthy information associated with R_i= ($R_i.q_1$, ..., $R_i.q_m$).

We must first build an LSH projection before releasing Ri to others. We make a new vector V = (v_1, ..., v_m), where v_j (j = 1, 2, ..., m) is a randomly generated value from domain [1, 1]. As a result, as shown in equation(1), we generate an LSH function f.

$$\begin{aligned} f(R_i) &= R_i \cdot V \\ &= (R_i.q_1,...,R_i.q_m) \cdot (v_1,...,v_m) \\ &= R_i.q_1 * v_1 + \ldots + R_i.qm * vm \end{aligned} \qquad (1)$$

Following that, we may generate an $f(R_i)$ that can be positive or negative. Then, as seen in the equation, we create the following mapping (2). As a result, we map $f(R_i)$ to $h(R_i)$, which is a binary value of 0 or 1. Algorithm 1 demonstrates a concrete technique.

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

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 R_i after mapping.

1: **for** j = 1, ..., m **do**
2: v_j = random[−1,1]
3: **end for**
4: **for** i = 1, ..., n **do**
5: sum = 0
6: **for** j = 1, ..., m **do**
7: sum + = $R_i.q_j * v_j$
8: **end for**
9: **if** sum > 0

10: **then** $f(R_i) = 1$
11: **else** $f(R_i) = 0$ 12: **end if**
$\overrightarrow{\quad}$
13: **return** $f(R_i)$
14: **end for**

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

The $f(R_i)$ 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 to convert R_i into a privacy-free index. Considering this, we repeat Algorithm 1 multiple times by projections of f_1, ..., f_a, after which we get an adimensional hash vector $H(R_i)$ as in equation (3). Thus the mappings of ‘‘R_i $H(R_i)$’’ (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 of R_i, we cannot infer the real value of R_i. This way, the privacy of patients contained in R_i is secured.

Through a projection procedure, the $f(R_i)$ generated in Step-1 may be thought of as the hash value of R_i. One projection procedure, however, is insufficient to transform R_i into a privacy-free index. In light of this, we repeatedly run Algorithm 1 using the projections of f_1, ..., f_a, and the result is the adimensional hash vector $H(R_i)$, as shown in equation (3). A hash table, designated by the symbol "T," is made up of the mappings of " R_i $H(R_i)$’’ (i = 1, 2, ..., n). In other words, we may inquire about the index value of R_i through "T," but we cannot determine the actual value of R_i from an index value. R_i protects the confidentiality of its patients in this way.

$$H(R_i) = (h_1(R_i),...,h_a(R_i)) \qquad (3)$$

One hash table "T" may not always reflect the actual index of each IoH data entry. Given this restriction, we create "T" several times, resulting in b tables: T_1, ..., T_b. With the help of Algorithm 2's pseudocode, this phase may be formalized.

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.
1: **for** x = 1, ..., a **do**
2: repeat Algorithm 1 based on f_x
3: **end for**
4: **for** i = 1, ..., n **do** 5: $H(R_i) = (h_1(R_i), ..., h_a(R_i))$
6: put ‘‘$R_i \rightarrow H(R_i)$’’ into T
7: **end for**
8: **return** T
9: **repeat** lines 1-8 b times
.

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

B tables T_1, ..., T_b is formed in Step 2. Additionally, each table would have a set of corresponding " R_i $H(R_i)$ " pairings I = 1, 2,..., n). Additionally, $H(R_i)$ is roughly thought of as the index of R_i in the table. The IoH data records with the same index should be somewhat comparable, according to the notion of location-sensitive hashing. As a consequence, records R_1 and R_2 are most likely related records if they have the same index.

In this approach, we may mine the prospective related IoH data records by verifying their respective index values without much privacy. However, for two IoH data records R_1 and R_2, $H(R_1) = H(R_2)$ is a relatively tight constraint requirement as each dimensional value of $H(R_1)$ should be precisely equivalent to that of $H(R_1)$. A tight constraint requirement like this is likely to return an empty result when searching for similar IoH data records, which is absurd for privacy. -free data mining and fusion for IoH

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 concluded 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.

Considering this shortcoming, we reduce the above strict criterion by producing additional hash tables instead of only one. In particular, if $H(R_1) = H(R_2)$ holds in any Ty (y = 1, 2,..., b), then it may be deduced that R_1 and R_2 are likely comparable IoH data records. This is done by taking into consideration the b tables produced in Step 2, i.e., T_1, ..., T_b. The search criteria for similar IoH data records are consequently loosened. Therefore, for a single IoH data record Rx, we may seek for its analogous record set Sim Set (Rx) utilizing the aforesaid approach. Details of this stage are described in Algorithm 3. And lastly, we return Sim Set (Rx) as the final result of the proposal in this effort.

V. EXPERIMENTS

A series of tests are planned and contrasted with current methods to verify the efficiency of our technique in addressing privacy-free data fusion and mining (abbreviated as PDFM).

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 (R_x) = *8*
2: **for** y = 1 to b **do**
3: **for** i = 1, ..., n **do**
4: **if** $H(R_i) = H(R_x)$

5: **then** put R_i into Sim_Set (R_x)
6: **end if**
6: **end for**
7: **end for**
8: **return** Sim_Set (R_x)

A. *CONFIGURATION*

We use the public data released by http://inpluslab.com/ wsdream/ for simulation purposes. In multi-source IoH data fusion situations, each user-item-QoS pair in the dataset is treated as a patient-criterion-value pair. The dataset's remaining 10% of entries are utilized for test and validation, while the remaining 90% are used to train the parameters of the data fusion and mining model.

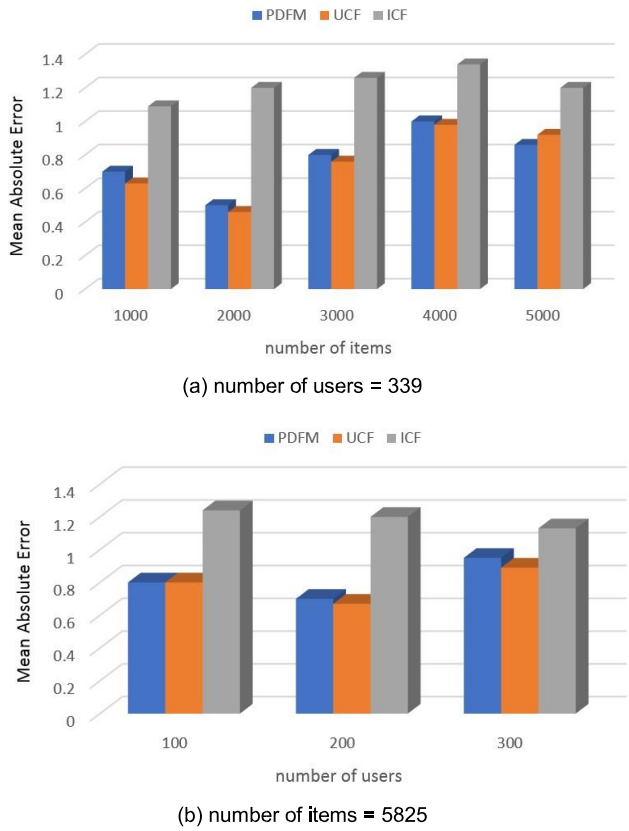
UCF (baseline) and ICF approaches are compared with PDFM to demonstrate the competitive benefits of PDFM. The compared parameters include computing time and the accuracy of missing data prediction (Mean Absolute Error) (s). 2.80 GHz processor, 8.0 GB of memory, Windows 7 operating system, and Java 8 are among the hardware and software options. Each experiment is conducted 50 times, and the average results are reported.

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 calculate three approaches' Mean Absolute Errors and contrast them. The settings for the parameters are as follows: The number of users is 339, the number of items is between 1000 and 5000, and a = b = 10. Here, two different experiment sets are put to the test. First, we examine how the amount of items in the dataset being utilized affects the variation trend of the Mean Absolute Error for three different techniques. Second, we examine how the number of users in the dataset affects the variation trend of the Mean Absolute Error for three different techniques. Figure 3 displays the comparison findings. As UCF is a baseline approach and PDFM is an approximation of UCF, we can see in Figs. 3(a) and 3(b) that PDFM and UCF have an obvious advantage over ICF. Additionally, because of the LSH approach used in PDFM, which can guarantee an excellent similarity maintenance characteristic, PDFM achieves an estimated Mean Absolute Error of UCF. Additionally, PDFM offers the advantage of privacy-preservation capabilities that UCF does not hold.



.

FIGURE 3. Mean absolute error comparison.

2) COMPUTATIONAL TIME COMPARISON

Three different counting techniques are measured and compared. User volume ranges from 100 to 300, object volume ranges from 1000 to 5000, and a = b = 10 are the parameter settings. Figure 4 displays comparative statistics.

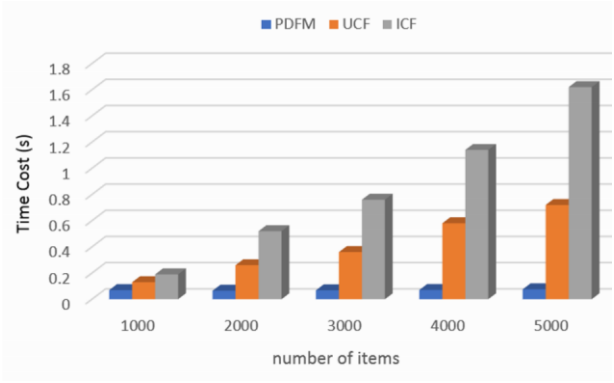
The time spent on the three techniques is probably going to rise when the number of users or the number of items increases, as shown in Fig. 4. Particularly, UCF and ICF take longer than PDFM because they demand more intensive user computation or object comparisons.

The cost of time in PDFM may be split into two categories: (2) The same retrieval of IoH data record, which has to be done online and the complexity of your time is O (3). (1) hash, which can be performed offline; as a consequence, the complex time is O (1). (1). The same IoH data records may now be retrieved using PDFM with a faster response time, allowing us to use our approach on a bigger IoH data region.

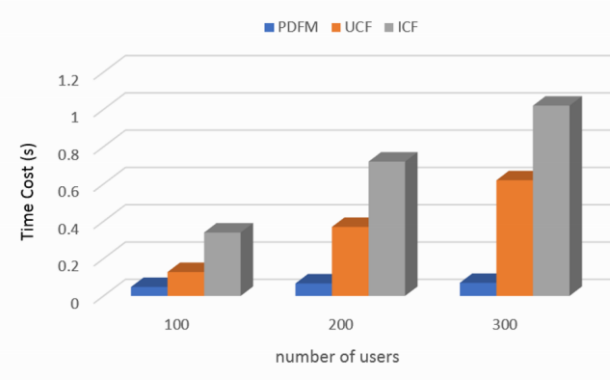
.

3) MEAN ABSOLUTE ERROR OF PDFM

The PDFM approach is based on the LSH strategy, whose performances are frequently correlated with some crucial variables, such as parameters a and b. In light of this, we examine the PDFM's associated with a and b performances. The parameters are as follows: the user volume is 339, the item volume is 5825, and the values for a and b are 2, 4, 6, 8, and 10, respectively. Data comparisons are shown in Fig. 5.



(a) number of users = 339



(b) number of items = 5825

FIGURE 4. Computational time comparison.

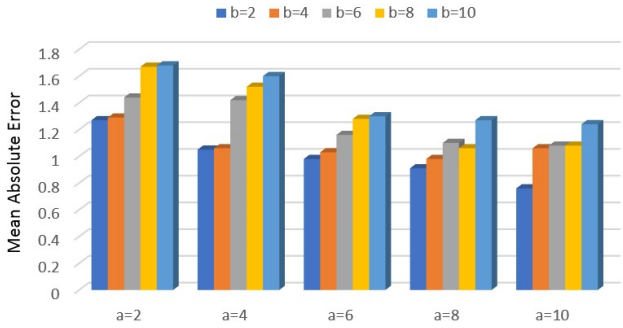


FIGURE 5. Mean absolute error of PDFM w.r.t. (a, b) pairs.

According to Fig. 5, the Mean Absolute Error of PDFM rises as parameter b rises and falls as parameter a. This is a result of the following factors: (1) The similar IoH data record retrieval condition becomes more lenient as the number of hash tables (i.e., b increases) increases; as a result, more similar records are returned and the Mean Absolute Error is rising; (2) As the number of hash functions (i.e., increases), the similar IoH data record retrieval condition becomes more stringent; as a result, fewer similar records are returned and the Mean Absolute Error is decreasing. Additionally, we can see that having more hash functions (bigger a) and fewer hash tables (smaller b) will result in more accurate predictions.

4) NUMBER OF RETURNED RESULTS OF PDFM

According to the study in the paragraph above, the PDFM approach is based on the LSH strategy, whose returned result volume is often correlated with some crucial variables such as parameters a and b. In light of this, we look at the PDFM returned result volume connected to a and b. The parameters are as follows: the user volume is 339, the item volume is 5825, and the values for a and b are 2, 4, 6, 8, and 10, respectively. In Fig.6, compared data are presented

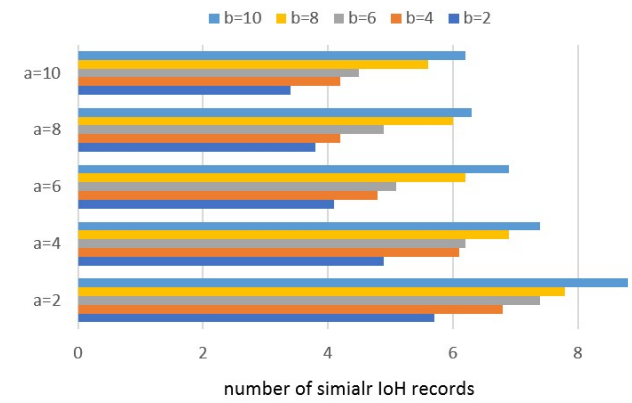


FIGURE 6. Several returned results of PDFM w.r.t. (a, b) pairs.

According to Fig. 6, the returned result volume of the PDFM grows when parameter b rises and parameter a falls. This is a result of the following factors: (1) The similar IoH data record retrieval condition becomes looser as the number of hash tables (i.e., b increases) rises, leading to the return of more similar records; (2) the condition becomes stricter as the number of hash functions (i.e., increases), leading to the return of fewer similar records. Furthermore, we can see that fewer hash tables (i.e., a smaller b) and more hash functions (i.e., a bigger a) would result in fewer returned results.

C. FURTHER DISCUSSIONS

Numerous successful big IoH applications have been made possible by the growing popularity of big data technologies. However, there are several restrictions in our proposed IoH data fusion and mining approach PDFM.

(1) To start, we just take into account straightforward IoH continuous data without taking into account potential data type and data structure variety (such as continuous, discrete, and boolean data).

(2) The second issue is that PDFM has not yet addressed how to assess the security of sensitive patient data.

(3) The third point is that there is still a need for research into how to better combine various privacy protection technologies for improved performance.

(4) There is frequently an inherent compromise between data availability and data privacy. To safeguard data privacy, it is unavoidable that data availability will be reduced. As a result, we cannot always ensure complete data availability using our suggested LSH-based privacy-aware data fusion strategy. However, if the right parameters are chosen, our solution can ensure 99.99 percent prediction accuracy owing to an intrinsic quality of LSH.

(5) Finally, LSH has an excellent similarity-keeping feature, which is why we selected it for privacy protection. In more concrete terms, following LSH's hash projection, two points that are close neighbors would be projected into the same bucket.

VI. CONCLUSION

Scientific disaster diagnostics and medical care services benefit from the effective integration and analysis of IoH data. The IoH data generated by patients, however, are frequently dispersed across many departments and only partially protect patient privacy. As a result, it might be difficult to successfully integrate or mine the sensitive IoH data without revealing patient privacy. We present PDFM, a revolutionary multi-source medical data integration and mining system for improved healthcare services, to address this problem. In order to give patients better medical and health services, we may quickly and privately search for similar medical data using PDFM. The practicality of PDFM has been demonstrated through trials on real data.

By taking into account the potential diversity of data types and data structures, we will upgrade the proposed PDFM approach in further study. Additionally, there is still much to learn about how to combine different privacy solutions that already exist for improved performance.

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, Feb. 2020.

[2] N. C. Benda, T. C. Veinot, C. J. Sieck, and J. S. Ancker, ‘‘Broadband Internet access is a social determinant of health!’’ *Amer. J. Public Health*, vol. 110, no. 8, pp. 1123–1125, Aug. 2020.

[3] E. Sillence, 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.

[4] K. Szulc and M. Duplaga, ‘‘The impact of Internet use on mental wellbeing and health behaviors among persons with disability,’’ *Eur. J. Public Health*, vol. 29, no. 4, pp. 185–425, Nov. 2019.

[5] T. Peng, Y. Lin, X. Yao, and W. Zhang, ‘‘An efficient ranked multi-keyword search for multiple data owners over encrypted cloud data,’’ *IEEE Access*, vol. 6, pp. 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*, vol. 8, pp. 4895–4907, 2020.

[7] T. V. Xuan Phuong, G. Yang, W. Susilo, F. Guo, and Q. Huang, ‘‘Sequence aware functional encryption and its application in searchable encryption,’’ *J. Inf. Secure. Appl.*, vol. 35, pp. 106–118, Aug. 2017.

[8] Z. Xia, X. Wang, X. Sun, and Q. Wang, ‘‘A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data,’’ *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 2, pp. 340–352, Feb. 2016.

[9] M. He, M. Chang, and X. Wu, ‘‘A collaborative filtering recommendation method based on differential privacy,’’ *J. Comput. Res. Develop.*, vol. 54, no. 7, pp. 1439–1451, 2017.

[10] Transformation of Data from RDBMS to HDFS by using Load Atomizer
IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 3 Issue 10, October 2016

[11] C.K.Gomathy.(2010), Cloud Computing: Business Management for Effective Service Oriented Architecture, International Journal of Power Control Signal and Computation (IJPCSC), Volume 1, Issue IV, Oct-Dec 2010, P.No:22-27, ISSN: 0976-268X.

[12] C.K.Gomathy and Dr.S.Rajalakshmi.(2011), Business Process Development In Service Oriented Architecture, International Journal of Research in Computer Application and Management (IJRCM) ,Volume 1,Issue IV, August 2011,P.No:50-53,ISSN : 2231-1009.

[13]C.K.Gomathy and Dr.S.Rajalakshmi.(2014), Software Pattern Quality Comportment In Service-Oriented Architectures, European Scientific Journal (ESJ) volume-10, No- 9, Issue-March 2014, P.No-412-423, ISSN-1857-7881.

[14]C.K.Gomathy and Dr.S.Rajalakshmi.(2014), A Business Intelligence Network Design for Service Oriented Architecture, International Journal of Engineering Trends and Technology (IJETT), Volume IX, Issue III, March 2014, P.No:151-154, ISSN:2231-5381.

[15]C.K.Gomathy and Dr.S.Rajalakshmi.(2014), A Software Design Pattern for BankService Oriented Architecture, International Journal of Advanced Research in Computer Engineering and Technology(IJARCET), Volume 3, Issue IV, April2014,P.No:1302-1306, ISSN:2278-1323