IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125256

MULTI-SOURCE MEDICAL DATA INTEGRATION AND MINING FOR HEALTHCARE SERVICES

Sahil Ravindra Kadukar [1], Lowlesh N. Yadav[2], Neehal B. Jiwane[3]

B-Tech Final Year Student, Computer Science, and Engineering, Shri Sai College of Engineering and Technology, Bhadrawati, India[1]

Assistant Professor, Computer Science, and Engineering, Shri Sai College of Engineering and Technology, Bhadrawati, India[2]

Assistant Professor, Computer Science, and Engineering, Shri Sai College of Engineering and Technology, Bhadrawati, India[3]

Abstract: As the Internet of Health(IoH) period dawns, conventional medical or healthcare coffers are gradationally migrating to the web or the internet, performing in a massive influx of medical data relating to cases, physicians, medicinals, medical structure, and so on. This IoH data's good integration and analysis are ideal pointers for disaster opinion and medical care services. still, IoH is constantly divided into other departments and protects the druggies' sequestration. As a result, collecting or rooting critical IoH data, where stoner sequestration may be compromised, is constantly a delicate operation. To address the forenamed challenges, we concentrate on PDFM, multi-source medical data collecting and booby-trapping solution for bettered health care services(Data Fusion and Private Mining). Through PDFM, we can search for analogous medical records in a time-effective and sequestration-conserving manner, so as to offer cases with better medical and health services. A group of trials are legislated and enforced to demonstrate the feasibility of the offer in this work.

Index term: Service recommendations, Internet Health, site-sensitive hashing, user privacy, data integration.

I. INTRODUCTION

With the ever-adding fashionability of Information Technology and the gradational relinquishment of digital software in medical or healthy disciplines, colorful medical departments or agencies have accumulated a considerable quantum of literal data(e.g., cases ' medical records, healthy treatment results, and so on), which form a main source of big Internet of Health(IoT) data. The application degree of similar IoH data is a crucial criterion to estimate and quantify the information position of medical or healthy units or departments. Generally, the utmost of literal IoH data records contains precious information especially for the medical or healthy agencies, similar as the one complaint of a case at a time point.

Mining and assaying similar literal IoH data records can' scientific and reasonable opinion and treatment decision-material, as well as disaster trend vaticination and palladium. thus, it's of imperative necessity to collect, integrate, fuse, and dissect these multi-source IoH data records for high-quality healthcare services suitable for cases. 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. 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 privacy disclosure concerns.

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.

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.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

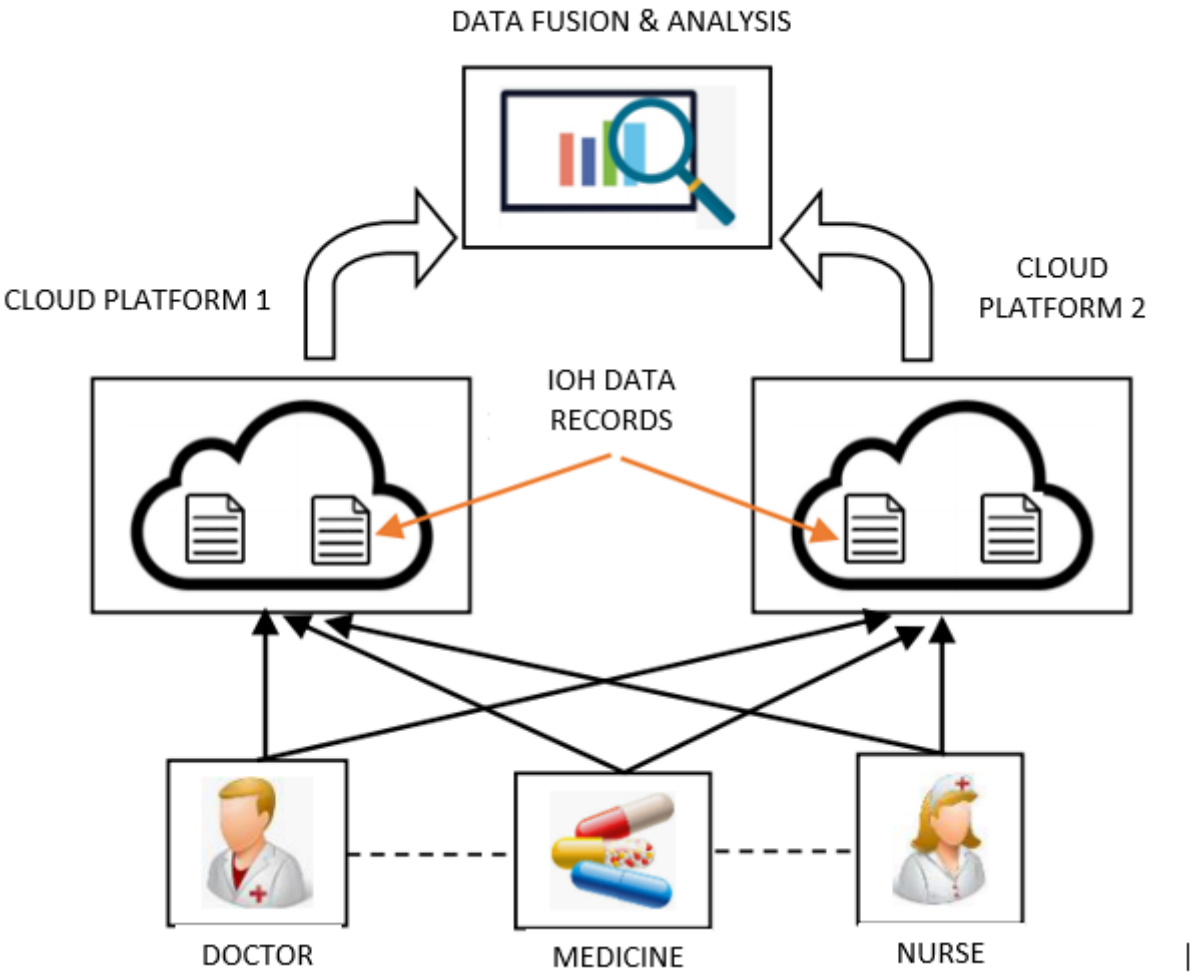
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125256

A. Encryption:- Encryption is a classic and effective way to secure sensitive user data, which has been investigated for a long time. 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.

B. Differentially privacy:- A differentially privacy-based improved collaborative filtering method named IPriCF was put forward in , to secure user privacy involved in the collaborative data integration process. By dividing user data and item data, IPriCF can effectively eliminate the disruption brought by noises incurred by differential privacy. This method can balance user data privacy and the accuracy of the recommended list. A stakeholder-feature-item matrix was built in to analyse 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 matrix factorization technique was used to convert sensitive user data into potential low-dimensional vectors; while a differential privacy technique was used to confuse the targeted object functions. However, when the number of dimensions grows, the prediction accuracy is reduced accordingly.

C. Anonymization:-Anonymization is an effective way for securing sensitive user data when making big data analyses and mining. 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 trade-off between data privacy and availability. The K-anonymity solution is adopted in to hide the key sensitive information involved in the data-driven decision-making process. A K-anonymity-based user location protection method is suggested in 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 data privacy and data utilization well as anonymized data would lose certain key information more or less.



International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125256

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 healthcare judgments to properly my 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.

CONCLUSION

Effective fusion and analyses of IoH data are of positive significance for scientific disaster diagnosis and medical care services. still, the IoH data produced by cases are frequently distributed across different departments and contain partial patient privacy. thus, it's frequently a grueling task to effectively integrate or booby-trap the sensitive IoH data without telling patient privacy. To attack this challenge, we bring forth a new multi-source medical data integration and mining result for better healthcare services, named PDFM.

Through PDFM, we can search for analogous medical records in a time-effective and privacy-conserving manner, so as to provide cases with better medical and health services. The trials on a real dataset prove the feasibility of PDFM. In the forthcoming exploration, we will modernize the suggested PDFM system by considering the possible diversity of data types(32) –(34) and data structure(35) –(38). In addition, how to fuse multiple sequestration results for better performances is still an open problem that requires ferocious and nonstop study.

REFERENCE

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

[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.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125256

[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