Journal of Network Computing and Applications (2019) 4: 21-25
Clausius Scientific Press, Canada

DOI: 10.23977/jnca.2019.41004
EISSN 2371-9214

Design and Implementation of Scientific Research Data Platform Based on Psychiatric Big Data

Feng Gao, Wei Zhong[a,*], Xinlei Chen and Guangzhong Yin

Suzhou GuangJi Hospital, Suzhou 215137, Jiangsu, China

[a]13862339400@163.com

*Corresponding author: 13862339400@163.com

Keywords: Psychiatry, Big data, Research platform

Abstract: Health care big data is an important basic strategic resource of the country. The application of medical big data is of great significance for clinical medical research, scientific management and the transition and development of medical service mode. Regarding the mining and research of big data in psychiatry, there is no corresponding research in China and abroad. Establish a clinical data center system: collect all the data of the hospital's clinical information system, including structured data and pathological reports in text format, scans of past medical records, etc. The main function of such system is to clean, store and refactor data for clinical and scientific analysis with the common data model (Common Data Mode1) as the core. In this article, in order to illustrating design and implementation of such system, architecture design, key technology and implementation steps are introduced in detail as well as the safety and reliability. Relied on the big data mining and analysis technology of electronic medical records, in the future, a clinical data center system would be applied to clinical medicine, scientific research data support, and guiding treatment programs.

1. Introduction

One of the national strategic guidelines is the construction of big data, in which national health big data is an important part of big data construction [1]. The document of the General Office of the State Council, “Guiding advice on Promoting and Regulating the Development of Big Data Applications for Health Care,” clearly states that health big data is a basic strategic resource to the Country [2]. The collection and application of health big data is of great significance for the development of clinical medicine, health management and medical services in the next 20 years [3-5]. Especially for psychiatry, it is more suitable for analyzing and researching with clinical big data, and achieving momentous research results.

2. Research status

The current state of foreign research: In recent years, the global development of big data is blooming, and it has been involved in all areas of industry [6]. With the popularization and application of various information systems, the medical industry has owned a large amount of

medical clinical data, which has met the general standard of big data. From the perspective of the development of the entire medical industry, the transformation and application of medical big data has become a trend of development in the industry [7-8]. In 2008, Science published an article on Big Data: Science in the Petabyte Era. The three basic theories in the 2011 IDC research report "Extracting Value from Chaos" constitute the theoretical basis of big data [9]. The American Public Health Association has launched Flu near you, a website that monitors the spread of the flu and uses big data processing techniques to analyze any future flu outbreaks [10]. However, the mining and research of big data in psychiatry is still in primary stage in developed countries.

The current state of domestic research: At present, the research and application of health big data in China is relatively late compared with other big data research. No scale effect has been generated yet, however great progress has been made in recent two years, which focuses on the development of big data analysis driven by medical application needs [11-12]. For example, the Beijing Health Commission has begun to analyze and alert epidemic outbreaks, public emergencies, and population movements through big data mining and analysis techniques, which would provide data evidence of management. Another instance, Central South University has also worked on a combination of big data, mobile medical treatment, and the Internet of Things. It takes advantages of the Internet of Things terminal to collect various physiological indicators of the human body, transmits it through mobile networks then, and eventually stores the collected healthy big data to the Xiangya Data Center [13]. Unfortunately, the mining and research of big data for psychiatry is still a blank in China nowadays.

3. Research content

Our team has had the rich construction experience of our hospital in clinical research data center and related applications [14]. Based on our experience and the analysis of clinical and scientific research demand, it is determined that the content of this research and development turns into the following aspects:

Construction of a clinical data center system: gather all data from the hospital's clinical information system, including the pathological report of structured data in text format, the scan of past medical records, etc. These data would be cleaned, stored and reconstructed with using the Common Data Model (Common Data Mode1). Clinical Big Data Center of GuangJi Hospital can be built on this solid system, for clinical research and scientific analysis [15].

Patient Indexing Service: a unique identity token is established on the Big Data platform, which is associated with the information such as medical card, medical treatment number, and hospitalization number of the local HIS system. This identity of the big data system is unique in order to ensure the availability and validity of big data.

Integrated User Management Service: Users are able to enroll in the various medical systems with a unique account password due to a unified platform, namely the association of other software sub-systems, which is convenient for authorization and user management. On the other hand, the visible range of data can be defined according to the user’s roles, responsibilities, and tasks in the research project of different users.

4. System design and implementation

(1) Architecture design

The entire psychiatric data center system will adopt centralized data management architecture to manage the electronic medical record data of each department, including extracting the data of each system in the hospital, transforming and integrating the extracted data. Then, the data would be classified and stored. The architecture of the data processing platform is approximately divided into

four layers, namely, a data layer, a component layer, a functional layer, and a service layer. The service layer provides support services for the data processing platform, including data extraction, data desensitization, data cleaning and other functions. The functional layer is the core of the platform, which is responsible for the core metadata management, metadata exchange, log management, node management and other functions. The primary role of the component layer is to support the functional layer to implement the basic components of the function, including database adapters, file adapters, data transfer conversion components.

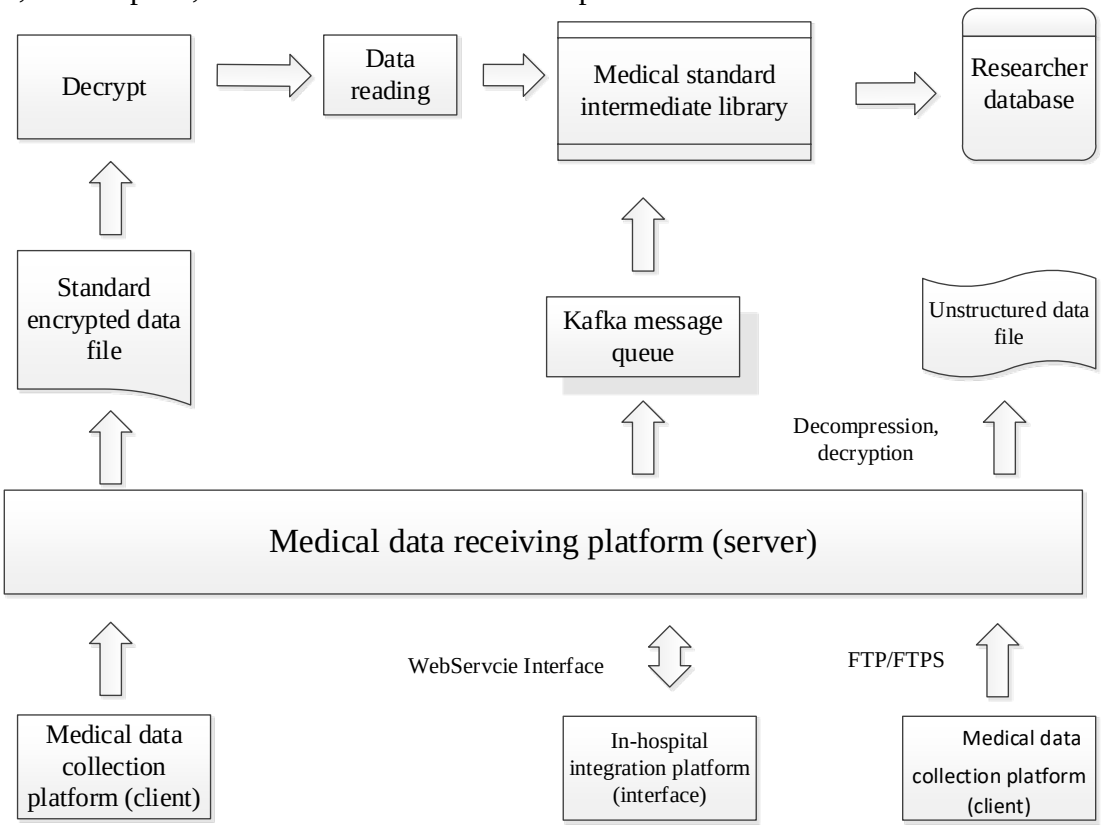

Figure. 1 Psychiatric data center system architecture

(2) Key technologies

Medical Vocabulary Module Design: Medical terminology dictionary is a medical term that standardizes and computerizes to reduce ambiguity. The medical knowledge base is a further standardized medical professional vocabulary. Through vocabulary standardization, it provides a standardized reference for the unification of complex, heterogeneous, and unstructured data in psychiatric clinical big data. This base construction improves data availability and standardization, laying the foundation for subsequent clinical scientific analysis. In the implementation phase, all words in the psychiatric terminology database are converted automatically to the standard terminology of the local knowledge base through natural language processing technology. After that, the vocabulary of special local concept would be added to the knowledge base, which aims to build a more complete vocabulary system. Researchers are able to search the knowledge base by keywords input, classify similar vocabulary concepts, create a concept set, and then define the disease queue using the concept set.

Text data natural language processing module design: The case data of the original HIS system and EMR system are most stored in the form of text. The common problem is that the clinical

description in text form is of ambiguity and non-standardized description. Therefore, the most important step in data processing is to transform the unstructured data into standardized, structured data. In combination with the characteristics of psychiatry, our hospital develops the following natural language processing module to actualize the structuration of data: (1) medical named entity recognition module (2) medical named entity automatic coding module.

(3) Implementation steps

The first stage: During the process of business analysis and data model establishment, the e-medical electronic medical record mining starts with the current mental illness treatment procedure in the hospital. The original structured data is directly extracted into the clinical resource center. Besides, the unstructured data would be cleaned, categorized, natural language processed, and then stored. Our goal is to complete the construction of a clinical big data platform with statistical analysis of information and corresponding data visualization. Gradually automate the extraction and processing of data in the psychiatric big data platform.

The second stage: Construct the model database for the recommendation system of the disease diagnosis and treatment plan based on the reference of the model of the clinical disease diagnosis system. This database would offer the data analysis evidence for the processing and analysis of clinical big data.

The third stage: Gradually improve the entities of various mental diseases and mental diseases in the database. The platform is able to generate new decision model parameters after new data was added; the establishment of grouping models of various diseases gradually completes the construction, transformation and data docking of the platform.

5. Safety and reliability design

Security: Adopt international advanced data desensitization technology to ensure that patient's individual privacy is not leaked and clinical trial data is not illegally infringed and modified. At the meantime, provide a variety of security inspection and audit tools to ensure the consistency of system data processing and data is not lost and damaged due to accidents. In terms of system security, the physical architecture including the security devices such as gatekeepers and firewalls are set up, and the internal and external networks in the hospital are isolated. Network user real name must be authenticated. At the network level, unlicensed, universal ports with security risks are closed. In terms of software, anti-virus software and policy prevention against viruses have recently appeared. Therefore, security settings and prevention are implemented in various aspects to prevent illegal users from intruding into the system, ensuring data security and data integrity.

Reliability: The platform has also made multiple preparations for the availability and the reliability. Firstly, it adopts a variety of high-reliability software architectures and high-availability technologies to ensure the stability of the normal operation of the system, especially for the uninterrupted operation of critical business and the operational protection of emergency. Back up the database in the clinical data center to ensure uninterrupted operation and rapid recovery of sudden failures. Ensure system reliability.

6. Summary

In the field of big data and precision medicine for mental illness, taking advantage of biomedical big data mining, a variety of life-study analysis and statistics technology, a unified and complete data view for all kinds of scientific research information applications can be created by the construction of a scientific research data platform based on psychiatry Big Data. Ultimately, the improvement of medical service quality, the reduction of medical errors, the improvement of clinical research level and the reduction of medical costs would be realized.

Acknowledgments

Funding: Science and Technology Demonstration Project of Suzhou Science and Technology Bureau, Big Data Analysis Processing and Clinical Application Research of Psychiatry (No. SS201706)

References

[1] Istephan S, Siadat M R. Extensible Query Framework for Unstructured Medical Data -- A Big Data Approach [C]// IEEE International Conference on Data Mining Workshop. IEEE, 2016.
[2] Zhang T, Chi H, Ouyang Z. Detecting Research Focus and Research Fronts in the Medical Big Data Field Using Co-word and Co-citation Analysis [C] // 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). IEEE, 2018.
[3] Mavroeidakos T, Tsolis N, Vergados D D. Centralized management of medical big data in Intensive Care Unit: A security analysis [C] // Smart Cloud Networks & Systems. IEEE, 2017.
[4] Shah F, Li J P, Shah F, et al. Quality Compression for Medical Big Data X-Ray Image using Biorthogonal 5.5 Wavelet[J]. International Journal of Engineering Research and Applications, 2016.
[5] Ishikawa K B. Medical Big Data for Research Use: Current Status and Related Issues [J]. Japan Medical Association Journal, 2016, 59 (2): 110-124.
[6] Windridge D, Bober M. A Kernel-Based Framework for Medical Big-Data Analytics [M] // Interactive Knowledge Discovery and Data Mining in Biomedical Informatics. Springer Berlin Heidelberg, 2014.
[7] Yao Q, Tian Y, Li P F, et al. Design and Development of a Medical Big Data Processing System Based on Hadoop [J]. Journal of Medical Systems, 2015, 39 (3): 23.
[8] Flynn F V. Medical Data Processing [J]. Proceedings of the Royal Society of Medicine, 1977, 70 (10): 750.
[9] Hiden H, Woodman S, Watson P. Prediction of workflow execution time using provenance traces: Practical applications in medical data processing [C] // 2016 IEEE 12th International Conference on e-Science (e-Science). IEEE, 2016.
[10] Csaba Horváth, Gábor Fodor, Ferenc Kovács, et al. A Proposed Scalable Environment for Medical Data Processing and Evaluation [J]. 2010.
[11] Popov A A, Yanenko V M, Zaitsev N G, et al. Automated system for medical data processing. [J]. International Journal of Bio-Medical Computing, 1970, 1 (3): 193-209.
[12] Xue J, Tian J, Dai Y K, et al. Processing Framework and the Fast Volume Rendering Algorithms for Out- of-Core Medical Data [J]. Journal of Software, 2009, 19 (12): 3237-3248.
[13] Asvestas P A, Matsopoulos G K, Nikita K S. Applications of fractal theory on medical data processing. [M]// Декабристы и их время /. Изд-во МГУ, 2000.
[14] Lim, Michael C L. METHOD AND SYSTEM FOR MEDICAL DATA PROCESSING [J]. 2011.
[15] Satoh H, Niki N, Takahashi E, et al. Teleradiology mobile internet system and home care medical system with a new information security solution [C] // SPIE Medical Imaging. 2015.