A novel framework for bringing smart big data to proactive decision making in healthcare

Shengyao Zhou
Runton Zhang
Beijing Jiaotong University, China

Donghua Chen
University of International Business and Economics, China

Xiaomin Zhu
Beijing Jiaotong University, China

Abstract

Big data have shown their great potential value to serve many aspects of human life. Due to complexity of the medical and healthcare big data in real life, traditional big data analysis methods are difficult to be dealt with. Therefore, a single method is unable to analyze and manage heterogeneous big data sources. To utilize data fully from the perspective of decision-making, we propose a novel framework which guides the healthcare big data to be smartly and proactively processed for decision-making without user interventions. The framework contains five stages, which are intelligent data cleaning, customized data fusion, analysis mapping, exploratory visualization analysis, and generation of decision-making reports. It also enables learning from the data and correlating them with the existing human knowledge. Subsequently, a smart big data-driven application exhibits innovative management in intelligent healthcare. The proposed framework provides the guidelines of the best practices of big data-driven analysis for intelligent healthcare according to our practical applications. The platform provides the appropriate reference for the big data-driven innovation of management in intelligent healthcare.

Keywords
data cleaning, data fusion, medical information systems, proactive decision making, smart big data

Corresponding author:
Runton Zhang, School of Economics and Management, Beijing Jiaotong University, No. 3, Shangyuancun, Haidian District, Beijing 100044, China.
Email: rtzhang@bjtu.edu.cn

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Introduction

A growing number of studies show that big data has demonstrated its tremendous value and has served every aspect of human life. Big data-driven algorithms and applications are capable of processing medical and health data in structured, semi-structured, and non-structured format. However, the traditional characteristics of big data, such as volume, velocity, variety, and even variability (4Vs), hinder the further use of big data. It is still challenging to turn medical and health data into insights to support intelligent healthcare. Although many researchers have proposed effective methods to utilize big data in critical studies, researchers are immersed in tedious data preprocessing work. The tedious and redundant work makes analysis lose its meanings of application practices, thereby hindering the development of high-level big data driven applications, such as establishing a clinical decision support system.

Currently, the analysis of medical and health big data results in several problems. First, common big data-related problems, including large data volume, diverse data formats and forms, and high data dimension, still exist. Second, a specific big data analysis method generally analyzes data for limited big data with specific format. Therefore, a single method is unable to analyze and manage heterogeneous big data sources. Third, the rules of machine-learning techniques are emphasized in the data analysis, while close correlations among heterogeneous data sources are ignored. Specifically, in the absence of an explicit application scenario, the big data in healthcare are almost worthless static data, which makes it difficult for big data to evolve into analysis models independently and to make possible proactive decisions of the innovation of management at various social and economic levels. The addressed problems are new challenges for various types of research and analysis-driven intelligent health-care programs. The arrival of the big data era has separated data-intensive science from the third paradigm, abandoned its dedication to cause-and-effect relationships, shifted its attention to related relationships, advocated data processing ahead of theoretical assumptions, and possibly arrived at previously unknown theories. Being part of the solutions of the aforementioned problems in utilizing medical and health big data, this study proposes a novel concept called “smart big data” that is based on existing big data theories to explore the best practices of the innovation of management in healthcare.

Smart big data

Due to complexity of the medical and healthcare big data in real life, traditional big data analysis methods are difficult to be dealt with. To utilize data fully from the perspective of decision-making, we propose a big data extended conceptual model. Thereafter, we explore the characteristics and its novel application models in intelligent healthcare.

Conceptual model

Similar with typical big data, smart big data exhibit the 4Vs and include Relationalization (R). Such 4V1R approach empowers a knowledge-enabled and self-evolving big data analytics. This model is a pervasive data-driven model for the analysis of big data in any field and not only for healthcare. On the basis of the practices of big data-driven analysis in intelligent healthcare, we introduce smart big data in healthcare solutions. Traditional big data are expressed abstractly by the R-based smart big data model where the novel consideration of various latent relationships within and among datasets in big data is considered. Equation (1) is an abstract definition of a smart big data model.
Relationalization (R) + Traditional Big Data (4Vs) = Smart Big Data (4V1R) \hfill (1)

The relationalization of smart big data in (1) refers to the utilization of various relationships in different contexts within big data. On the basis of the motivation, traditional big data models are extended to overcome the challenges of big data in analyzing complicated healthcare-related application cases. Knowledge sources in medical and health-care field in smart big data-driven analysis models in healthcare aim to promote integration, fusion, and management of medical and health big data. Accordingly, proactive decision-making models are established to promote the innovation of management in healthcare. Currently, few studies focus on the deep exploration of big data based on the relevance and regularity of knowledge and relationships formed from heterogeneous data sources. We believe that the future of big data should not focus on the analysis of static big data, but big data should be considered a form of smart big data based on the extracted complicated relationships of health-care data. Thus, big data after reconstruction using smart big data can achieve wide knowledge-based and self-evolving proactive decision-making models.10–12 To sum up, our original intention of proposing a smart big data driven model is suitable for healthcare.

Relationships within datasets in healthcare

Compared with traditional big data models for healthcare, a smart big data-driven model covers the relational transformation of healthcare data among different datasets.13 Utilizing various types of relationships within and among datasets in healthcare can profoundly integrate, manage, and fuse complicated big data and even realize the dynamics of original data.14 Therefore, the smart big data-driven analysis model is suitable for the innovation of management in intelligent healthcare.

Researchers believe that big data are intelligent in a decision-making process. In big data-driven management of healthcare, smart big data greatly depend on the latent relationships of healthcare found in the big data. An objective of introducing smart big data to the healthcare field is to discover relationships among and within different data sources automatically and precisely. The aforementioned relationships include various types of relationships from healthcare and non-healthcare fields. With the usage of the relationships, various data sets in the heterogeneous medical and health big data can be associated as much as possible with knowledge graphs that support further analysis and knowledge reasoning for specific problems in healthcare.15 Figure 1 shows the underlying relationships and reasoning capability of smart big data among datasets in healthcare.

In Figure 1, the \( i \)-th dataset \( d \) can be abstracted as follows:

\[
d_i = (x_1, x_2, \ldots, x_n); c, n \in N
\] \hfill (2)

where \( x_c \) represents the \( c \)-th column label of \( d \), \( c \) represents the total number of the columns of \( d \), and \( n \) represents the total number of data sets.

If \( r_{ij}^z \) is a medical and health semantic relationship between two datasets, \( d_i \) and \( d_j \), in (2), where \( r^0 \) represents a basic relationship of \( r \), \( r^1 \) denotes a first-level medical health semantic relationship, and \( r_{ij}^z \) denotes a semantic relationship in a big data set, then we have the following:

\[
r_{ij}^z = (k_1^z, k_2^z, \ldots, k_n^z), z \in Z
\] \hfill (3)

In (3), \( k^z \) denotes a concept of medical knowledge per \( z \)-level relationships. Relationships among concepts can be expressed as \( G(K) \), \( G \) represents a knowledge graph, and \( K = \{k\} \) represents a set of knowledge bases.
In medical health big data, a dataset based on the conceptual model of smart big data is $d_{i+1}$

$$d_{i+1} = rd_i \cup rd_{i-1} \cup \ldots \cup rd_1$$  \hspace{1cm} (4)

Equations (2)–(4) define data structure models in a smart big data-driven analysis model that enables a proactive decision-making process.

**Innovation of management in healthcare**

On the basis of the conceptual model and extraction of the relationship of smart big data, the application models of innovation management are explored. The models are extended to meet the requirement of 4V1R, which are suitable for various analysis needs in utilizing big data in healthcare. The traditional big analysis approaches generally follow proposing a hypothesis, selecting a specific method, and visualizing analysis results. By contrast, our smart big data-driven method allows the exploration of the meaningful use of big data in healthcare. Currently, research on the big data of medical field mainly covers knowledge discovery, health assessment, disease warning, service management, and service management practice.\(^{16}\) Therefore, smart big data-driven models for the innovation of management should be focused on these research directions.\(^{17}\)

A common application model of healthcare is to use big data in medical service organizations to analyze and generate reports for decision-making in health-care and insurance policies. The big data to be analyzed comes from different hospitals and other organizations who share their data for analysis. In this case, data managers define the standardized interfaces of storage such that these data are stored in a unified storage platform and they try to associate existing smart big data.
Through a data-cleaning process, data analysts set up a data-cleaning rule based on data extraction, transformation, and loading; and complete data integration, fusion, and management. Meanwhile, demand analysts for big data analysis, such as government agencies, establish general analysis models based on specific health-care application scenarios, such as disease spectrum analysis, medical service analysis, disease-based cost analysis, and decision support systems in health-care. As an example, big data analysis in healthcare will focus on big data and the perspectives of the innovation of management in actual problems of healthcare. Table 1 shows smart big data-driven application models for intelligent healthcare.

**Smart big data-driven analysis in healthcare**

On the basis of smart big data-driven model and our practices in healthcare, we present a novel framework for implementing proactive decision making for the innovation of management in healthcare. The framework provides the guidelines of best practices for smart big data-driven analysis in intelligent healthcare.

**Support of domain knowledge in healthcare**

The relationalization of big data in healthcare requires the involvement of domain knowledge in healthcare. Many useful knowledge bases can effectively support a complete understanding of big data in analyses. The exchanges of medical terminology standards play an important role in transferring medical and health data among different hospital information systems and in establishing a medical semantic relationship $S \rightarrow T$ between a source dataset $S$ and a target dataset $T$, thereby supporting knowledge discovery and reasoning in the medical context. Important terminology standards include International Classification of Diseases (ICD), Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT), and Unified Medical Language System (UMLS). Expert knowledge in knowledge bases can effectively utilize health-care health big data. Obviously, expert knowledge plays an important role in every aspect of smart big data-driven analysis model. However, the utilization of expert knowledge should be in an independent framework that is capable of the continuous expansion of knowledge in healthcare. Knowledge in medical and health data...
field $G(K)$ should be related to the discovery and evaluation of knowledge within big data. Here, we define a knowledge relationship model suitable for describing the relationships of knowledge as follows:

$$KB = G\{(S, R, T)\}$$  \hspace{1cm} (5)

where $S = \{p_s\}$ represents a source dataset, $T = \{p_t\}$ represents a target dataset, $R = \{p_r\}$ represents the set of mappings between $S$ and $T$, and $p_s$, $p_t$, and $p_r$ represent $S$, $T$, and $R$, respectively.

In (5), we use $R$ to bridge $S$ and $T$ from different knowledge domains. Here, $p$ denotes the attributes of $S$ and $T$. Thus, $KB$ in the model is stored in a form of graph $G$. The necessary elements to establish a knowledge base for healthcare include conceptual entity, knowledge hierarchy, entry attribute, entity relationship, and description sets. Relationships of which support the integration of expert knowledge into the smart big data-driven analysis models.

**Application mechanism for proactive decision making**

On the basis of the support of domain knowledge, a smart big data-driven mechanism is devised to implement proactive decision-making process by using medical and health big data. The framework processing big data in healthcare contains five stages, which are intelligent data cleaning, customized data fusion, analysis mapping, exploratory visualization, and intelligent report generation. This smart big data-driven framework in healthcare could effectively identify the user intentions of analysis and realize the actual needs of the innovation of management in big data-driven intelligent healthcare. Figure 2 presents an overall architecture of the framework in each stage.
Intelligent data cleaning

Medical information systems document massive medical and health data. However, due to differences in information standards, information entry, and other factors, the systems generate a large amount of dirty data that cannot be utilized. An objective of this stage aims to automatically extract, transform, and load the data features of big data for further analysis. In addition, this stage allows the framework to identify semantic relationships from the limited information of file directories to be associated with a huge generated semantic network, which represents relationships among different data sets. Meanwhile, these data are desensitized. Thereafter, different extraction rules for various types of data during the data-cleaning process are implemented (Table 2). This process embodies the heterogeneous data features of medical and health big data and promotes the initial fusion of raw big data in healthcare.

On the basis of extraction rules, we establish an intelligent data cleaning process. First, we explore potential semantic relationships in healthcare based on classification labels, data types, and topics of datasets. Second, we desensitize uploaded data and preserve the privacy of patient data. Sensitive records in the data will be removed, such as encrypting the patient’s medical insurance number and ID number through specific rules, deleting the patient’s address and telephone records, and keeping only the last name and replacing the first name with *. Third, by using information extraction techniques, we extract data features from the uploaded data. Subsequently, utilizing extracted data features, we implement unsupervised text classification techniques. Finally, we obtain the results of initial classification based on directory hierarchies. Therefore, the intelligent analysis of massive data should be applied to specific-application scenarios.

Customized data fusion

The meaningful use of existing big data with a relationship network generated in the data cleaning stage comes from a process of customized data fusion. The integration, fusion, and management of the datasets of big data provide the solutions of problems in healthcare. Multiple factors in customized data fusion, such as type of data set, format of data set, and actual scenes, should be considered. A process of data fusion, including data filtering and integration, is task-, application scenarios-, and user need-based. Users of the framework are required to provide a simple description of their user needs in healthcare.

With the extracted features of data, the process of data fusion will achieve the convergence of existing methods, such as union and intersection. Given more than one dataset, multiple datasets must be merged. Therefore, the directory structure information of big data should be involved, and extraction and recombination will be performed.

| Type of dataset          | Content of extraction      | Relationship of extraction                  |
|-------------------------|----------------------------|---------------------------------------------|
| Text file               | Keyword extraction         | 1. Extraction of category                   |
| Table file              | Column extraction          | 2. Extraction of meta data                  |
| Image file              | Image feature extraction   | 3. Association of meta data                 |
| Knowledge base          | Concept and knowledge      | 4. Discovery of latent semantic relationships|
|                         | relationship extraction    |                                             |
Methodology set mapping

Automatically mapping the user needs of analysis on the proper methodology sets of big data analysis could solve the problems in healthcare with human interventions. A key of this process is to create mapping relationships between the vague user needs of analysis and the definitive methodology sets of big data analysis. Accordingly, the framework identifies the needs to start big data analysis with proper methodologies suitable for specific health-care data. Supposing \( U \) is a set of the user needs of analysis, \( f \) is a set of rule-based mapping functions, \( m \) is a sub set of methodology, and \( M \) is a complete methodology set, we have the following:

\[
U \rightarrow m \in M
\]

Equation (6) represents a mapping process between \( U \) and \( M \) by using \( f \). The framework analyzes complex big data by using the selected methods \( (m) \). A complete methodology set includes predefined traditional analysis methods, such as clustering, classification, temporal retrieval, and anomaly discovery. The main difficulty in this stage focuses on resolving the intelligence of the dynamics of user needs and establishing mapping relationships between \( U \) and \( M \), the details of which are in Algorithm 1.

Algorithm 1. Methodology set mapping algorithm.

Input: a narrative healthcare related question and a set of data features of selected datasets, \( \text{User}(\text{question}, \{\text{features of dataset}\}) \).

Output: a list of mapping relationships between users' needs and a set of methodology, \( \text{Mapping}(\text{User}, \{(\text{term, methodology})\}) \).

Procedure:
1. Extract healthcare related terms from question of user into a term list
2. For each terms in the term list:
3. For each dataset in \( \{\text{features of dataset}\} \):
4. If term similar with any keyword from features of dataset:
5. If \( f(\text{terms}) \) in both \( U \) and \( M \):
6. Append \( (\text{term}, \{\text{methodology}\}) \) into \( m \)
7. Return \( \text{Mapping}(\text{User}, \{(\text{term, methodology})\}) \).

Exploratory visualization analysis

The smart big data-driven framework requires automatically exploring the potential values of the results of analysis in the framework. The results in previous stages sometimes turn out to be high dimensional data. To provide the users of the platform with a complete understanding of such results, we present an exploratory visualization analysis for results from the outputs of MapReduce-based models in Algorithm 2.

Algorithm 2. Exploratory visualization analysis algorithm for big data results.

Input: table files created by output of MapReduce based models, \( \text{R}(\text{table file}) \).

Output: exploratory charts and graphs, \( \text{G}(\{\text{chart or graph, (feature)}\}) \).

Procedure:
1. For each table in \( \text{R} \):
2. Extract features from the table into a feature list
3. Append the feature list into a complete feature list
4. For each group of features \( (\text{size} \geq 2) \) in the complete list:
5. If types of features in \( \{(\text{chart or graph, features})\} \):
6. Append (chart or graph, features) into \( \text{G} \).
7. Return \( \text{G} \).
Generally, the original outputs of big data analysis is a group of text-formatted table files. The algorithm explores potential correlations among the columns of table files. Meanwhile, domain knowledge in healthcare is involved. The algorithm generates two types of visualization results, charts and graphs, as outputs. The framework stores charts and graphs in relational and graph databases, such as Neo4J, respectively. To sum up, an exploratory visualization analysis for high-dimensional data in the framework provides the basis of the next stage, which is a decision making report based on results.

**Generation of decision-making reports**

The output of smart big data-driven analysis models in healthcare is the analysis report that is used for decision-making. By introducing natural language processing techniques, exploratory charts and graphs in the previous step are interpreted in human-readable reports. The reports contain necessary charts and narrative texts that explain the summarized meaningful points.\(^{26}\) In Algorithm 2 will transform into a list of tuples (chart or graph, narrative description) where the descriptions of charts and graphs are generated according to the comparison of the actual data of extracted features.

Decision-making reports generated by Bayesian prediction model and natural language processing technology in the framework can provide early warning and decision-making support in specific healthcare-related problems. This process makes a secondary analysis that provides the descriptive analysis in human-readable format.

**Smart big data-driven application in action**

The smart big data-driven mechanism enables healthcare analysis applications by using data from cooperative hospitals. Based on the intelligent big data analysis model proposed in this article and the five steps applied to the field of medical and health big data analysis, a medical and health big data analysis platform is established to support the intelligent mining and active decision-making of medical and health big data based on the relationship of medical data sets. The architecture of the platform is shown in Figure 3. Through the five stages, we apply smart big data-driven models for the decision support of various application scenarios in healthcare.

Figure 4 presents the several cases of big data-driven analysis in healthcare. Our best practices of the framework aim to solve the three types of healthcare-related problems, which are the analysis of abnormal behaviors in medical insurance, misdiagnosis, and big data-driven visualization. As in the cases in Figure 4, massive healthcare-related data are processed through our platform, and reports for decision-making are generated. The detailed process of utilizing such data is as follows. First, through the intelligent data cleaning process, data features are extracted from datasets. Data are then uploaded automatically to our Hadoop Distributed File System. Here, the distributed storage of the data to be analyzed and the semantic network of data features in healthcare are prepared. Second, on the basis of customized datasets, a methodology set mapping algorithm determines the set of actual big data-driven analysis methods for processing the uploaded data. Third, after the completion of big data analysis, an exploratory visualization analysis algorithm explores the potential value of big data results and provides an overview of visualized results to solve healthcare-related questions, details of which can be seen in Figure 4(a)–(d). Finally, reports based on the visualization analysis are generated for decision-making. In our practice, a novel concept of smart big data in healthcare is involved, thereby enhancing the ability of the automatic big data process in the framework.
Although our smart big data-driven framework is capable of solving most healthcare-related problems in a novel manner, it still has limitations. The smart big data model requires various types of relationships found within and among the datasets of analyzed data and such relationships in healthcare are sometimes difficult to determine. In addition, our big data practice in the field of healthcare mainly focuses on text-based data, and seldom analyzes medical images. Therefore, we need to consider how to apply this kind of medical image data in our framework, and it is necessary to conduct further research. Considering that our objective is to develop a
novel conceptual big data model to deal with complex problems in solving healthcare-related problems, the smart big data-driven framework still provides a high feasibility for big data analysis using novel healthcare relationships. To sum up, the framework provides five necessary stages with novel technologies and algorithms to solve the emerging challenges in big data-driven decision-making theory. However, further studies and practices are needed from the perspectives of big data in intelligent healthcare.

Figure 4. Results of several healthcare-related big data-driven analyses using our framework. Charts and graphs of visualization analyses in the different respects of healthcare are created automatically. With the use of smart big data, our framework provides big data-driven analyses from the perspectives of healthcare.
Conclusion

Traditional single big data analysis method can’t analyze and manage heterogeneous big data sources, and the close relationship between heterogeneous data sources is ignored in data analysis. Without clear application scenarios, it is difficult to develop medical big data into an analysis model independently. Therefore, this paper proposes a new concept model of intelligent big data analysis in medical field. This framework provides the capability for implementing proactive decision-making for management of healthcare. First, we analyzed the limitation of existing big data theories in health-related applications and introduce a novel conceptual model of smart big data, which utilizes relationships among the datasets of big data effectively. Second, we developed the smart big data-driven mechanism for proactive decision-making in healthcare. The proposed framework provides the guidelines of the best practices of big data-driven analysis for intelligent healthcare according to our practical applications. The platform provides the appropriate reference for the big data-driven innovation of management in intelligent healthcare. In our next work, we will consider the use of deep learning techniques in order to hopefully apply medical imaging data to our smart big data-driven framework and consider specific application practices.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by the Fundamental Research Funds for the Central Universities [grant number 2019YJS063], in part by the National Social Science Foundation of China (Major Project) [grant number 18ZDA086].

ORCID iD

Shengyao Zhou https://orcid.org/0000-0002-5676-7126

References

1. Allam Z and Dhunny ZA. On big data, artificial intelligence and smart cities. Cities 2019; 89: 80–91.
2. Janssen M, Luthra S, Mangla S, et al. Challenges for adopting and implementing IoT in smart cities an integrated MICMAC-ISM approach. Internet Res 2019; 29(6): 1589–1616.
3. Zeng DL, Tim YN, Yu JX, et al. Actualizing big data analytics for smart cities: a cascading affordance study. Int J Inform Manage 2020; 54: 102156.
4. Yang H and Kundakcioglu E. Healthcare intelligence: turning data into knowledge. IEEE Intell Syst 2018; 29(3): 54–68.
5. Gu DX, Li JJ, Li XG, et al. Visualizing the knowledge structure and evolution of big data research in healthcare informatics. Int J Med Inform 2017; 98: 22–32.
6. Monsted T. Achieving veracity: a study of the development and use of an information system for data analysis in preventive healthcare. Health Inform J 2019; 25(3): 491–499.
7. Urrea C and Venegas D. Automatized follow-up and alert system for patients with chronic hypertension. Health Inform J 2020; 26(4): 2625–2636.
8. Strang KD and Sun ZH. Hidden big data analytics issues in the healthcare industry. Health Inform J 2020; 26(2): 981–998.
9. Jones LK, Pulk R, Gionfriddo MR, et al. Utilizing big data to provide better health at lower cost. Am J Health Syst Pharm 2018; 75(7): 427–435.
10. Halper M, Gu HY, Perl Y, et al. Abstraction networks for terminologies: supporting management of big knowledge. *Artif Intell Med* 2015; 64(1): 1–16.

11. Nyantara W, Sarah KT, Elizabeth L, et al. Can electronic assessment tools improve the process of shared decision-making? A systematic review. *Health Inf Manag J*. Epub ahead of print 5 October 2020. DOI: 10.1177/1833358320954385.

12. Paez DG, Rodriguez MD, Sanz EP, et al. Healthy and wellbeing activities’ promotion using a Big Data approach. *Health Inform J* 2018; 24(2): 125–135.

13. Yang Y, Zheng XH, Guo WZ, et al. Privacy-preserving smart IoT-based healthcare big data storage and self-adaptive access control system. *Inform Sci* 2019; 479: 567–592.

14. Mirhaji P, Casscells SW, Allemang D, et al. Improving the public health information network through semantic modeling. *IEEE Intell Syst* 2007; 22(3): 13–17.

15. Singh G, Schulthess D, Hughes N, et al. Real world big data for clinical research and drug development. *Drug Discov Today* 2017; 23(3): 652–660.

16. Steinbrook R. Personally controlled online health data—the next big thing in medical care? *N Engl J Med* 2008; 358(16): 1653–1656.

17. Viceconti M, Hunter P and Hose R. Big data, big knowledge: big data for personalized healthcare. *IEEE J Biomed Health Inform* 2015; 19(4): 1209–1215.

18. Dehnavieh R, Haghdoost A, Khosravi A, et al. The District Health Information System (DHIS2): a literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries. *Health Inf Manag J* 2019; 48(2): 62–75.

19. Krishnan S. Application of analytics to big data in healthcare. In: 32nd Southern biomedical engineering conference, Shreveport, LA, 11–13 March 2016, pp.156–157. New York: IEEE.

20. Tian XM and Liu LB. Does big data mean big knowledge? Integration of big data analysis and conceptual model for social commerce research. *Electron Commer Res Appl* 2017; 17(1): 169–183.

21. Khan Z and Vorley T. Big data text analytics: an enabler of knowledge management. *J Knowl Manag* 2017; 21(1): 18–34.

22. CramerI KA, Maher L, Van DP, et al. Personal electronic healthcare records: what influences consumers to engage with their clinical data online? A literature review. *Health Inf Manag J*. Epub ahead of print 10 January 2020. DOI: 10.1177/1833358319895369.

23. Rotmensch M, Halpern Y, Tlimat A, et al. Learning a health knowledge graph from electronic medical records. *Sci Rep* 2017; 7(1): 5994.

24. Brimblecombe C, Crosbie D, Lim WK, et al. The Goals of Patient Care project: implementing a proactive approach to patient-centred decision-making. *Intern Med J* 2015; 44(10): 961–966.

25. Velampalli S and Jonnalagedda MV. Graph based knowledge discovery using MapReduce and SUBDUE algorithm. *Data Knowl Eng* 2017; 111: 103–113.

26. Malmir B, Amini M and Chang SI. A medical decision support system for disease diagnosis under uncertainty. *Expert Syst Appl* 2017; 88: 95–108.