Big Data and its Applications - A Survey

W. Melba Marries

M. Phil Computer Science, Noorul Islam University, Tamilnadu, India

Abstract— The Goal of this paper is to provide a comprehensive review of Applications of Big Data in Information Management. The characteristics of datasets in library conducted a review of the research work on the big data and then summarized the applications in this field. And also its apply to improve the health consciousness enhances. Nowadays the healthcare industry has become a hot spot in the society and some health monitoring systems emerge one after another in recent years. However, the most existing systems only focus on the logic, reasoning, but ignore the factor of the user's emotion, which is regarded as an important factor to impact human health. In this paper, We are going to discuss and analysis the life cycle of big data and its privacy applications which are applied in human healthcare also.

Keywords— Big data, cloud computing, big data storage, big data processing.

I. INTRODUCTION

In recent years, big data have become a hot research topic. The increasing amount of big data also increases the chance of breaching the privacy of individuals. For big data require high computational power and large storage, distributed systems are used. As multiple parties are involved in these systems, the risk of privacy violation is increased. There have been a number of privacy-preserving mechanisms developed for privacy protection at different stages (e.g., data generation, data storage, and data processing) of a big data life cycle [2].

Current analytical approaches in computational social science can be characterized by four dominant paradigms: text analysis (information extraction and classi_cation), social network analysis (graph theory), social complexity analysis (complex systems science), and social simulations (cellular automata and agent-based modeling). However, when it comes to organizational and societal units of analysis, there exists no approach to conceptualize, model, analyze, explain, and predict social media interactions as individuals' associations with ideas, values, identities, and so on. To address this limitation, based on the sociology of associations and the mathematics of set theory, this paper presents a new approach to big data analytics called social set analysis. Social set analysis consists of a generative framework for the philosophies of computational social science, theory of social data, conceptual and formal models of social data, and an analytical framework for combining big social data sets with organizational and societal data sets. Three empirical studies of big social data are presented to illustrate and demonstrate social set analysis in terms of fuzzy set-theoretical sentiment analysis, crisp set-theoretical interaction analysis, and event studies-oriented set-theoretical visualizations. Implications for big data analytics, current limitations of the set-theoretical approach, and future directions are outlined [3].

One of the biggest concerns of big data is privacy. However, the study on big data privacy is still at a very early stage. We believe the forthcoming solutions and

theories of big data privacy root from the in place research output of the privacy discipline. Motivated by these factors, we extensively survey the existing research outputs and achievements of the privacy yield in both applied and theoretical angles, aiming to pave a solid starting ground for interested readers to address the challenges in the big data case. We first present an overview of the battle ground by defining the roles and operations of privacy systems.

Second, we review the milestones of the current two major research categories of privacy: data clustering and privacy frameworks. Third, we discuss the effort of privacy study from the perspectives of different disciplines, respectively. Fourth, the mathematical description, measurement, and modeling on privacy are presented. We summarize the challenges and opportunities of this promising topic at the end of this paper, hoping to shed light on the exciting and almost uncharted land [4].

Mobile devices are increasingly becoming an indispensable part of people's daily life, facilitating to perform a variety of useful tasks. Mobile cloud computing integrates mobile and cloud computing to expand their capabilities and benefits and overcomes their limitations, such as limited memory, CPU power, and battery life. Big data analytics technologies enable extracting value from data having four Vs: volume, variety, velocity, and veracity. This paper discusses networked healthcare and the role of mobile cloud computing and big data analytics in its enablement the motivation and development of networked healthcare applications and systems is presented along with the adoption of cloud computing in healthcare. Cloudlet based mobile cloud-computing infrastructure to be used for healthcare big data applications is described. The techniques, tools, and applications of big data analytics are reviewed. Conclusions are drawn concerning the design of networked healthcare systems using big data and mobile cloud-computing technologies [9].

A micro-video recommendation system the recommendation algorithms are the core of this system. Traditional recommendation algorithms include content-based recommendation, collaboration recommendation algorithms, and so on. At the Bid Data times, the challenges what we meet are data scale, performance of computing, and other aspects. Thus, this paper improves the traditional recommendation algorithms, using the popular parallel computing framework to process the Big Data. Slope one recommendation algorithm is a parallel computing algorithm based on MapReduce and Hadoop framework which is a high performance parallel computing platform. The other aspect of this system is data visualization. Only an intuitive, accurate visualization interface, the viewers and producers can find what they need through the micro-video recommendation system [8].

Libraries play an important role at the intersections of government, universities, research institutes, and the public since they are storing and managing digital assets. The large amount of data and those data in a library need to be transformed into information or knowledge which then be used by researchers or users. Librarians might need to understand how to transform, analyze, and present data in order to facilitate knowledge creation. For example, they should know how to make big datasets more useful, visible and accessible. With new and powerful analytics of big data, such as information visualization tools, researchers/users can look at data in new ways and mine it for information they intend to have. In this work, we discussed the characteristics of datasets in the library, conducted a review of the research work on library big data and then summarized the applications in this field. The issues associated with it were also discussed and explored [8].

II. LIFE CYCLE OF BIG DATA

A. Data generation

Data can be generated from various distributed sources. The amount of data generated by humans and machines has exploded in the past few years. Usually, the data generated is large, diverse and complex. Therefore, it is hard for traditional systems to handle

them. The data generated are normally associated with a specific domain, such as business, Internet, research, [7] etc.

B. Data storage

This phase refers to storing and managing large-scale data sets. A data storage system consists of two parts, i.e., hardware infrastructure and data management. Hardware infrastructure refers to utilizing information and communications technology (ICT) resources for various tasks (such as distributed storage). Data management refers to the set of software deployed on top of a hardware infrastructure to manage and query large scale data sets. It should also provide several interfaces to interact with and analyze stored data [2].

C. Data processing

Data processing phase refers basically to the process of data collection, data transmission, pre-processing and extracting useful information. Data collection is needed because data may be coming from different diverse sources, i.e., sites that contain text, images and videos. Data collection phase, data is acquired from the specific data productions environment using dedicated data collection technology. In data transmission phase, after collecting raw data from a specific data production environment, we need a high speed transmission mechanism to transmit data into a proper storage for various types of analytic applications. Finally, the pre-processing phase aims at removing meaningless and redundant parts of the data so that more storage space could be saved. The excessive data and domain specific analytical methods are used by many applications to derive meaningful information. Although different yields in data analytics require different data characteristics, few of these yields may leverage similar underlying technology to inspect, transform and model data to extract value from it. Emerging data analytic research can be classified into the following six technical areas: structured data analytics, text analytics, multimedia analytics, web analytics, network analytics, and mobile analytics [2].

D. Data operation in privacy system

(a) Collecting: Data curators collect data from different data sources [4].

(b) Anonymizing: Data curators anonyms the collected data sets in order to release it to public [4].

(c) Communicating: Data users perform information retrieval on the released data sets [4].

III. Disciplines In Privacy

A. Cryptography

Cryptography can certainly be used in numerous fashions for privacy protection in the big data age. For example, a patient can use the public key of her doctor to encrypt her medical documents, and deposits the ciphertext into the doctor's online database for her treatment while her privacy is strictly preserved [4].

With the emergence of big data, clouds are built to serve many applications due to its economical nature and accessibility feature. The dilemma of encryption based privacy protection in big data is: on one hand, we need to offer sufficient privacy protection for users, at the same time; we have to make the encrypted data informative and meaningful for big data analysis and public usage [4].

One challenge is information retrieval on encrypted data. This research branch is also called *searchable encryption*, which boomed around year 2000 [46], [47]. The basic idea is as follows: a user indexes and encrypts her document collection, and sends the secure index together with the encrypted data to a server which may be malicious. To search for a given keyword, the user generates and submits a trapdoor for the keyword, which the server uses to run the search operation and recover pointers to the appropriate encrypted documents.

Another challenge here is operations on encrypted objects. This research branch is named as *homomorphic encryption* started in 1978 [4]. In this kind of encryptions, we expect carry out computations on ciphertext, and obtain an encrypted output. If we decrypt the output it

should match the result of operations performed on the original plaintext.

B. Biometric Privacy

Biometric is a powerful tool for security, which aims to identify individuals based on their physical, behavioral, and physiological attributes, such as face, iris, _ngerprint, voice, and gait. Biometrics has been widely used in access control, and the procedure includes two stages: enrollment and release. In the first stage, biometric features, such as fingerprints, are sampled, and the information is stored in a database either as a raw data or in a transformed form. In the second stage, the related biometric characteristics are sampled again on site, and compared with the stored one for authentication [4] [5].

C. Game Theory

Game theory is a rich set of mathematical models to deal with conflict and cooperation between intelligent and rationale decision-makers. It has been widely used in economics, political science, psychology, computer science, and so on. Given the nature of privacy protection, game theory is obviously a powerful tool to be hired to motivate privacy investment, settle argument among different participants, and so on. There is a long list of the applicable scenarios that game theory can contribute.

IV. BIG DATA TECHNOLOGY

Big data technology is an emerging technology which applies to data sets where data size is so large and common data related technology tools are hard to capture, manage, and operate under a multi Common data-related technology uses a consistent, single, traditional data management framework, which is only suitable for data with low diversity and volume. If coming from diverse data sets, data will not be correlated in space and time, and it is hard for those data to get correlated to a unified and generalized power system model. Therefore, traditional data-related technology will no longer guarantee reliable data management, operation, etc. In this way, big data technology provides integrated architectures for complex power grid, effective data analysis to aware the unfavorable situation in advance, and various data processing methods for different conditions [4][5].

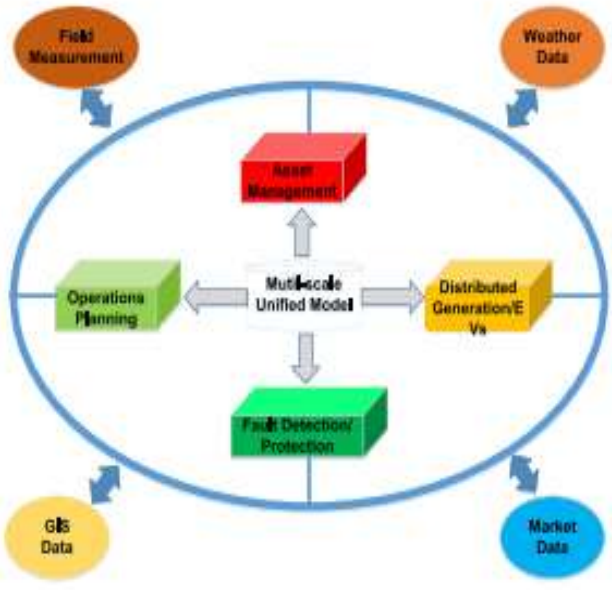

Figure: 1 – Big Data and its applications

V. CONCLUSION

This paper is show that the review of big data applications which is involved in the human health data management. The big data application is one of the most high valuable database management systems. I conclude this big-data and its applications are better research system of Information Management.

REFERENCES

[1] A Parallel Patient Treatment Time Prediction Algorithm and Its Applications in Hospital Queuing-Recommendation in a Big Data Environment, Kenli Li

(lkl@hnu.edu.cn), National Supercomputing Center in Changsha, Hunan, Changsha, 410082, China

[2] Protection of Big Data Privacy, Y. Xiang (yxiang@deakin.edu.au), School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798.

[3] Social Set Analysis: A Set Theoretical Approach to Big Data Analytics, R. R. Mukkamala (rrm.itm@cbs.dk) Computational Social Science Laboratory, Copenhagen Business School, Frederiksberg 2000, Denmark.

[4] Big Privacy: Challenges and Opportunities of Privacy Study in the Age of Big Data, SHUI YU, (Senior Member, IEEE) School of Information Technology, Deakin University, Waurn Ponds, VIC 3216, Australia (syu@deakin.edu.au)

[5] A MapReduce-Based Nearest Neighbor Approach for Big-Data-Driven Traffic Flow Prediction, Z. Zhang (zhangzl@swu.edu.cn), School of Information Technology, Deakin University, Geelong, VIC 3220, Australia.

[6] A TV Program Recommendation System Based on Big Data, Mengyi Zhang, Minyong shi, Zhiguo Hong, Songtao Shang and Menghan Yan, Department of Computer, Communication University of China, Beijing, China.

[7] A Micro-video Recommendation System Based on Big Data, Songtao Shang, School of computer science, Communication University of China, songtao.shang@foxmail.com.

[8] Exposing Library Data with Big Data Technology: A Review, Chunning Wang, National Geological Library of China, Beijing, China.

[9] Mobile Cloud Computing Model and Big Data Analysis for Healthcare Applications, L. A. Tawalbeh (latawalbeh@uqu.edu.sa), Department of Computer Engineering, Jordan University of science and Technology, Irbid 22110, Jordan

[10] System Design for Big Data Application in Emotion-Aware Healthcare, K. Lin (link@dlut.edu.cn), School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China.