

A Data Analytics Framework for Health Record Management

R.Tamizhazhagi, Dr.Sankar

PG Scholar, Dept. of CSE, Sri Venkateswara of College of Engineering and Technology, Thiruvallur, India

Professor, Dept. of CSE, Sri Venkateswara of College of Engineering and Technology, Thiruvallur, India

ABSTRACT: Personal health record (PHR) is an emerging patient-centric model of health information exchange, which is often outsourced to be stored at a third party, such as cloud providers. However, there have been wide privacy concerns as personal health information could be exposed to those third party servers and to unauthorized parties. To assure the patients' control over access to their own PHRs, it is a promising method to encrypt the PHRs before outsourcing. Yet, issues such as risks of privacy exposure, scalability in key management, flexible access, and efficient user revocation, have remained the most important challenges toward achieving fine-grained, cryptographically enforced data access control. In this paper, we propose a novel patient-centric framework and a suite of mechanisms for data access control to PHRs stored in semitrusted servers. To achieve fine-grained and scalable data access control for PHRs, we leverage attribute-based encryption (ABE) techniques to encrypt each patient's PHR file. Different from previous works in secure data outsourcing, we focus on the multiple data owner scenario, and divide the users in the PHR system into multiple security domains that greatly reduces the key management complexity for owners and users. A high degree of patient privacy is guaranteed simultaneously by exploiting multiauthority ABE. Our scheme also enables dynamic modification of access policies or file attributes, supports efficient on-demand user/attribute revocation and break-glass access under emergency scenarios. Extensive analytical and experimental results are presented which show the security, scalability, and efficiency of our proposed scheme.

I. INTRODUCTION

1.1 DATA MINING

Dr. Yan Mo won the 2012 Nobel Prize in Literature. This is probably the most controversial Nobel Prize of this category. Searching on Google with "Yan Mo Nobel Prize," resulted in 1,050,000 web pointers on the Internet (as of 3 January 2013). "For all praises as well as criticisms," said Mo recently, "I am grateful." What types of praises and criticisms has Mo actually received over his 31-year writing career? As comments keep coming on the Internet and in various news media, can we summarize all types of opinions in different media in a real-time fashion, including updated, cross-referenced discussions by critics? This type of summarization program is an excellent example for Big Data processing, as the information comes from multiple, heterogeneous, autonomous sources with complex and evolving relationships, and keeps growing.

Along with the above example, the era of Big Data has arrived. Every day, 2.5 quintillion bytes of data are created and 90 percent of the data in the world today were produced within the past two years. Our capability for data generation has never been so powerful and enormous ever since the invention of the information technology in the early 19th century. As another example, on 4 October 2012, the first presidential debate between President Barack Obama and Governor Mitt Romney triggered more than 10 million tweets within 2 hour.

II. EXISTING SYSTEM

- Every day, 2.5 quintillion bytes of data are created and 90 percent of the data in the world today were produced within the past two years.
- Our capability for data for data generation has never been so powerful and enormous
- The unprecedented data volumes require an effective data analysis and prediction platform to achieve fast response and real-time classification for such Big Data.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

- In typical data mining systems, the mining procedures require computational intensive computing units for data analysis and comparisons.

DISADVANTAGE

- 1 Not Secure
- 2 Only stored in minimum values.
- 3 Un authorized persons are get the details by authorized person

III. PROPOSED SYSTEM

- In the proposed system, we provide a analysed big data set for that characterizes the future of symptoms. Under big data mining perspective, multisource, massive, heterogeneous data has analysed effectively using machine learning tool. All the heterogeneous has a big challenge is to transform quantity to quality to do so we provided systematic approaches such as classifier, clustering with effective result.

In our system we have analysed a mining methods with provides

- 1) The efficiency improvement of single-source.
- 2) Knowledge discovery methods.

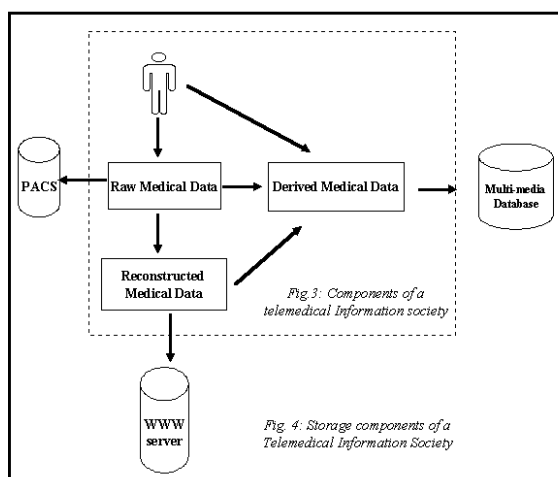
ADVANTAGE

secure

Accuracy

Un authorized persons are not hacked

IV. SYSTEM DESIGN



System architecture

4.1.1 Analyzing Heterogeneous And Big-Data Data

Our ability to collect data in various sensors, devices, formats, either from independent or connected applications, has significantly outpaced our capability to process, analyze, store and understand these datasets. Recently, the flourish of social networking applications, like Facebook, Twitter, that allow users to create contents freely, further amplifies the already huge web volume. Furthermore, with mobile phones becoming the sensory gateway to get real-time data on people from different aspects, the vast amount of data that mobile carrier can potentially process to improve our daily life has significantly outpaced our past CDR (call data record)-based processing for billing purposes only. People and devices (from home coffee machines to cars, to buses, railway stations and airports) are all loosely connected.

Trillions of such connected components will generate a huge data ocean, and valuable information must be discovered from the data to help improve quality of life and make our world a better place. For example, after we get up every morning, in order to optimize our commute time to work and complete the optimization before we arrive at office, the

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

system needs to process information from traffic, weather, construction, police activities to our calendar schedules, and perform deep optimization under the tight time constraints. In all these applications, we are facing significant challenges in leveraging the vast amount of data, including challenges in (1) system (2) algorithm (3) business model – application. This workshop aims to bring together people from both academia and industry to present their most recent work related to these issues, and exchange ideas and thoughts in order to advance this big challenge that has been considered as one of the most exciting opportunities in the past 10 years.

4.1.2 Data Ruling From Technical Details Using J-Rip Technique

We have proposed a technique that uses JRip classifier and association rule mining to select the most relevant features from a data set. JRip extracts the rules from a data set and then association rules mining technique is applied to rank the features. Twenty datasets are tested ranging from binary class problem to multi-class problem. Extensive experimentation is carried out and the proposed technique is evaluated against the performance of various familiar classifiers. Experimental results demonstrate that while employing less number of features the proposed method achieves higher classification accuracy as well as generates less number of rules.

They argue that holding back drug information makes it hard for patients and doctors to make informed decisions about treatment, and increases the chance that drug trials will be needlessly repeated, wasting money. Some drug companies are reluctant, citing concerns about commercial privacy.

4.1.3 Finding The Decision For Classification

Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with **decision nodes** and **leaf nodes**. A decision node (e.g., Outlook) has two or more branches (e.g., Sunny, Overcast and Rainy). Leaf node (e.g., Play) represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called **root node**. Decision trees can handle both categorical and numerical data.

Classification trees and regression trees predict responses to data. To predict a response, follow the decisions in the tree from the root (beginning) node down to a leaf node. The leaf node contains the response. Classification trees give responses that are nominal, such as 'true' or 'false'. Regression trees give numeric responses.

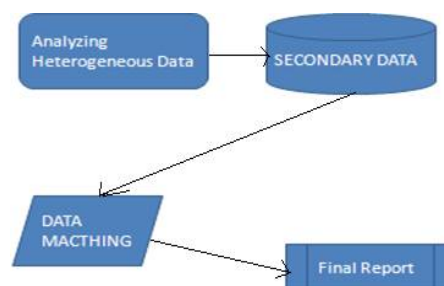
4.1.4 Hierarchical Clustering For The Optimized Result

In data mining, **hierarchical clustering** (also called **hierarchical cluster analysis** or **HCA**) is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types:

- **Agglomerative**: This is a "bottom up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.
- **Divisive**: This is a "top down" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

In general, the merges and splits are determined in a greedy manner. The results of hierarchical clustering are usually presented in a dendrogram.

4.1.5. Prescription Result



International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

V. CONCLUSION

We have analyzed several challenges at the data, model, and system levels. To support big data mining, high-performance computing platforms are required, which impose systematic designs to unleash the full power of the big data. At the data level, the autonomous information sources and the variety of the data collection environments, often result in data with complicated conditions, such as missing/uncertain values. In other situations, privacy concerns, noise, and errors can be introduced into the data, to produce altered data copies.

Developing a safe and sound information sharing protocol is a major challenge. At the model level, the key challenge is to generate global models by combining locally discovered patterns to form a unifying view. This requires carefully designed algorithms to analyze model correlations between distributed sites, and fuse decisions from multiple sources to gain a best model out of the big data. At the system level, the essential challenge is that a big data mining framework needs to consider complex relationships between samples, models, and data sources, along with their evolving changes with time and other possible factors.

VI. FUTURE ENHANCEMENT

A system needs to be carefully designed so that unstructured data can be linked through their complex relationships to form useful patterns, and the growth of data volumes and item relationships should help form legitimate patterns to predict the trend and future. big data technologies, we will hopefully be able to providemost relevant and most accurate social sensing feedback to better understand our society at realtime. We can further stimulate the participation of the public audiences in the data production circle for societal and economical events. the era of big data has arrived.

REFERENCES

1. Budhaditya Sahay, Dinh Phungy, Duc-Son Phamz , and Svetha Venkateshy z,"Clustering Patient Medical Records via Sparse Subspace Representation",vol vol. 315, no. 5814, pp. 972-976, 2007.
- 2.Fernando Berzal, Juan-Carlos Cubero, Nicolás Marín, José-María Serrano,"USABILITY ISSUES IN DATA MINING SYSTEMS",ISBN 0-385-26774-6.
- 3.DAVID DILTS, PhD, JOSEPH KHAMALAH, MASC, ANN PLOTKIN, OD, MSc,"Using Cluster Analysis for Medical Resource Decision Making",vol 1995;15:333-347.
- [4] C. Dong, G. Russello, and N. Dulay, "Shared and Searchable Encrypted Data for Untrusted Servers," *J. Computer Security*, vol. 19, pp. 367-397, 2010.
- [5] Y. Zheng, "Privacy-Preserving Personal Health Record System Using Attribute-Based Encryption," master's thesis, Worcester Polytechnic Inst., 2011
- [6] M. Pirretti, P. Traynor, P. McDaniel, and B. Waters, "Secure Attribute-Based Systems," *J. Computer Security*, vol. 18, no. 5, pp. 799-837, 2010.
- [7] S. Müller, S. Katzenbeisser, and C. Eckert, "Distributed Attribute-Based Encryption," *Proc. 11th Int'l Conf. Information Security and Cryptology (ICISC 08)*, pp. 20-36, 2009.
- [8]S.D.C.di Vimercati,
Charlet Annie2,"CLUSTERING DICHOTOMOUS DATA FOR HEALTH CARE"vol Vol.2, No.2, March 2012. Kenneth Revett, Sergio Tenreiro de Magalhaes, and Henrique M. D. Santos, "Data Mining a Prostate Cancer Dataset Using Rough Sets".vol12