ANALYSING THE BIG DATA CHARACTERISTICS USING ABE TECHNIQUE

[1]M.V.Saranya, [2] R.Rajasekar
[1]Research Scholar, Department of Computer Science, Tagore Institute of Engineering and Technology, Deviyakurichi
[2]Assistant Professor, Department of Computer Science, Tagore Institute of Engineering and Technology, Deviyakurichi
[1]mvsaranyacse@gmail.com

Abstract: Big Data concern large-volume, complex, growing data sets with multiple, autonomous sources. With the fast development of networking, data storage, and the data collection capacity, Big Data are now rapidly expanding in all science and engineering domains, including physical, biological and biomedical sciences. This paper presents a HACE theorem that characterizes the features of the Big Data revolution, and proposes a Big Data processing model, from the data mining perspective. This data-driven model involves demand-driven aggregation of information sources, mining and analysis, user interest modeling, and security and privacy considerations. Analyze the challenging issues in the data-driven model and also in the Big Data revolution.

1. INTRODUCTION

To make the problem even more complicated, let us assume that 1) the elephant is growing rapidly and its pose changes constantly, and 2) each blind man may have his own (possible unreliable and inaccurate) information sources that tell him about biased knowledge about the elephant (e.g., one blind man may exchange his feeling about the elephant with another blind man, where the exchanged knowledge is inherently biased). Indeed, many data mining algorithm are designed for this type of problem settings. If only facing with huge amounts of structured data, users can solve the problem simply by purchasing more storage or improving storage efficiency.

Issues and significance. A systematic investigation on pattern matching, pattern mining with wildcards, and application problems as follows: Exploration of the NP-hard complexity of the matching and mining problems. Multiple patterns matching with wildcards. Approximate pattern matching and mining, application of our research onto ubiquitous. Personalized information processing and bioinformatics.

The problem of full search, but also for finding global models those traditional mining methods cannot find. Local pattern analysis of data processing can avoid putting different data sources together to carry out centralized computing

For data anonymization, the main objective is to inject randomness into the data to ensure a number of privacy goals. For example, the most common k-anonymity privacy measure is to ensure that each individual in the database must be indistinguishable from k - 1others. The domain and application knowledge can also help design achievable business objectives by using Big Data analytical techniques. For example, stock market data are a typical domain that constantly generates a large quantity of information, such as bids, buys, and puts, in every single second. The market continuously evolves and is impacted by different factors, such as domestic and international news, government reports, and natural disasters, and so on. Model mining and correlations are the key steps to ensure that models or patterns discovered from multiple information sources can be consolidated to meet the global mining objective.

2. RELATED WORK

Dynamic networks have recently being recognized as a powerful abstraction to model and represent the temporal changes and dynamic aspects of the data underlying many complex systems. Significant insights regarding the stable relational patterns among the entities can be gained by analyzing temporal evolution of the complex entity relations. This can help identify the transitions from one conserved state to the next and may provide evidence to the existence of external factors that are responsible for changing the stable relational patterns in these networks. This paper presents a new data mining method that analyzes the time-persistent relations or states between the entities of the dynamic networks and captures all maximal non-redundant evolution paths of the stable relational states. Experimental results based on multiple datasets from real world applications show that the method is efficient and scalable.

Behavioral economics tells us that emotions can profoundly affect individual behavior and decision-making. Does this also apply to societies at large, i.e. can societies experience mood states that affect their

collective decision making. By extension is the public mood correlated or even predictive of economic indicator. Here author investigate whether measurements of collective mood states derived from large-scale Twitter feeds are correlated to the value of the Dow Jones Industrial Average (DJIA) over time. Author analyze the text content of daily Twitter feeds by two mood tracking tools, namely Opinion Finder that measures positive vs. negative mood and Google-Profile of Mood States (GPOMS) that measures mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). Author cross- validates the resulting mood time series by comparing their ability to detect the public's response to the presidential election and Thanksgiving day in 2008. A Granger causality analysis and a Self-Organizing Fuzzy Neural Network are then used to investigate the hypothesis that public mood states, as measured by the Opinion Finder and GPOMS mood time series, are predictive of changes in DJIA closing values. Results indicate that the accuracy of DJIA predictions can be significantly improved by the inclusion of specific public mood dimensions but not others. Author fined an accuracy of 87.6% in predicting the daily up and down changes in the closing values of the DJIA and a reduction of the Mean Average Percentage Error by more than 6%.

Over the last decade, there has been an explosion of interest in network research across the physical and social sciences. For social scientists, the theory of networks has been a goldmine, yielding explanations for social phenomena in a wide variety of disciplines from psychology to economics. In this essay, author review the kinds of things that social scientists have tried to explain using social network analysis and provide a nutshell description of the basic assumptions, goals and explanatory mechanisms prevalent in the field. Author also give a brief history of network research in the social sciences and identify some historical criticisms and current challenges facing the field. Author hope to contribute to a dialogue among researchers from across the physical and social sciences who share a common interest in understanding the antecedents and consequences of network phenomena.

Author presents a collective approach to learning a Bayesian network from distributed heterogeneous data. Author fast learn a local Bayesian network at each site using the local data. Then each site edentates the observations that are most likely to be evidence of coupling between local and non-local variables and transmit a subset of these observations to a central site. Another Bayesian network is learnt at the central site using the data transmitted from the local site. The local and central Bayesian networks are combined to obtain a collective Bayesian network that models the entire data. Experimental results and theoretical rustication that demonstrate the feasibility of our approach are presented.

There is significant current interest in the problem of influence maximization: given a directed social network with influence weights on edges and a number k, find k seed nodes such that activating them leads to the maximum expected number of activated nodes, according to a propagation model. Kempeet al. Showed, among other things, that under the Linear Threshold model, the problem is NP-hard, and that a simple greedy algorithm guarantees the best possible approximation factor in PTIME. However, this algorithm suffers from various major performance drawbacks. In this paper, Author propose SIMPATH, an efficient and effective algorithm for influence maximization under the linear threshold model that addresses these drawbacks by incorporating several clever optimizations.

Through a comprehensive performance study on four real data sets, Author show that SIMPATH consistently outperforms the state of the art w.r.t. running time, memory consumption and the quality of the seed set chosen, measured in terms of expected influence spread achieved.

Authors are at the beginning of the multicore era. Computers will have increasingly many cores (processors), but there is still no good programming framework for these architectures, and thus no simple and unified way for machine learning to take advantage of the potential speed up. In this paper, Author develops a broadly applicable parallel programming method, one that is easily applied to many different learning algorithms. Our work is in distinct contrast to the tradition in machine learning of designing (often ingenious) ways to speed up a single algorithm at a time. Specifically, Author show that algorithms that fit the Statistical Query model can be written in a certain ―summation form,‖ which allows them to be easily parallelized on multicore computers. Author adapt Google's map-reduce

paradigm to emonstrate this parallel speed up technique on a variety of learning algorithms including locally weighted linear regression (LWLR), k-means, logistic regression (LR), naive Bayes (NB), SVM, ICA, PCA, gaussian discriminant analysis (GDA), EM, and back propagation (NN). Our experimental results show basically linear speedup with an increasing number of processors.

Data anonymization techniques have been the subject of intense investigation in recent years, for many kinds of structured data, including tabular, item set and graph data. They enable publication of detailed information, which permits ad hoc queries and analyses, while guaranteeing the privacy of sensitive information in the data against a variety of attacks. In this tutorial, Author aim to present a unified framework of data anonymization techniques, viewed through the lens of data uncertainty. Essentially, anonym zed data describes a set of possible worlds, one of which corresponds to the original data. Author show that anonymization approaches such as suppression, generalization, perturbation and permutation generate different working models of uncertain data, some of which have been well studied, while others open new directions for research. Author demonstrate that the privacy guarantees offered by methods such as k- anonymization and `-diversity can be naturally understood in terms of similarities and differences in the sets of possible worlds that correspond to the anonym zed data. Author describe how the body of work in query evaluation over uncertain databases can be used for answering ad hoc queries over anonym zed data in a principled manner. A key benefit of the unified approach is the identification of a rich set of new problems for both the Data Anonymization and the Uncertain Data communities.

Many modern enterprises are collecting data at the most detailed level possible, creating data repositories ranging from terabytes to petabytes in size. The ability to apply sophisticated statistical analysis methods to this data is becoming essential for marketplace competitiveness. This need to perform deep analysis over huge data repositories poses a significant challenge to existing statistical software and data management systems. On the one hand, statistical software provides rich functionality for data analysis and modeling, but can handle only limited amounts of data; e.g., popular packages like R and SPSS operate entirely in main memory. On the other hand, data management systems—such as Map Reduce-based systems—can scale to petabytes of data, but provide insufficient analytical functionality. Author report our experiences in building Ricardo,a scalable platform for deep analytics. Ricardo is part of the extreme Analytics Platform (XAP) project at the IBM Almaden Research Center, and rests on a decomposition of data-analysis algorithms into parts executed by the R statistical analysis system and parts handled by the Hadoop data management system. This decomposition attempts to minimize the transfer of data across system boundaries. Ricardo contrasts with previous approaches, which try to get along with only one type of system, and allows analysts to work on huge datasets from within a popular, well supported, and powerful analysis environment. Because our approach avoids the need to re-implement either statistical or data-management functionality, it can be used to solve complex problems right now.

The promise of data-driven decision-making is now being recognized broadly and there is growing enthusiasm for the notion of ―Big Data,‖ including the recent announcement from the White House about new funding initiatives across different agencies that target research for Big Data. While the promise of Big Data is real for example, it is estimated that Google alone contributed 54 billion dollars to the US economy in 2009 – there is no clear consensus on what is Big Data. In fact, there have been many controversial statements about Big Data, such as ―Size is the only thing that matters.‖In this panel Author will try to explore the controversies and debunk the myths surrounding Big Data.

In this paper Author address the issue of privacy preserving data mining. Specifically, Author consider a scenario in which two parties owning confidential databases wish to run a data mining algorithm on the union of their databases, without revealing any unnecessary information. Our work is motivated by the need to both protect privileged information and enable its use for research or other purposes. The above problem is a specific example of secure multi-party computation and as such, can be solved using known generic protocols. However, data mining algorithms are typically complex and, furthermore, the input usually consists of massive data sets. The generic protocols in such a case are of no practical

use and therefore more efficient protocols are required. Author focus on the problem of decision tree learning with the popular ID3 algorithm. Our protocol is considerably more efficient than generic solutions and demands both very few rounds of communication and reasonable bandwidth. Key.

Recent events have shown online service providers the perils of possessing private information about users. Encrypting data mitigates but does not eliminate this threat: the pattern of data accesses still reveals information. Thus, Author present Shroud, a general storage system that hides data access patterns from the servers running it, protecting user privacy. Shroud functions as a virtual disk with a new privacy guarantee: the user can look up a block without revealing the block's address. Such a virtual disk can be used for many purposes, including map lookup, micro blog search, and social networking. Shroud aggressively targets hiding accesses among hundreds of terabytes of data. Author achieves our goals by adapting oblivious RAM algorithms to enable large-scale parallelization. Specifically, Author show, via new techniques such as oblivious aggregation, how to securely use many inexpensive secure coprocessors acting in parallel to improve request latency. Our evaluation combines large-scale emulation with an implementation on secure coprocessors and suggests that these adaptations bring private data access closer to practicality.

This paper evaluates the suitability of the Map Reduce model for multi-core and multi-processor systems. Map Reduce was created by Google for application development on data-centers with thousands of servers. It allows programmers to write functional- style code that is automatically parallelized and scheduled in a distributed system. Author describe Phoenix, an implementation of Map Reduce for shared-memory systems that includes a programming API and an efficient runtime system. The Phoenix runtime automatically manages thread creation, dynamic task scheduling, data partitioning, and fault tolerance across processor nodes. Author study Phoenix with multi-core and symmetric multiprocessor systems and evaluate its performance potential and error recovery features. Author also compares Map Reduce code to code written in lower-level APIs such as P-threads. Overall, Author establishes that, given a careful implementation, Map Reduce is a promising model for scalable performance on shared-memory systems with simple parallel code.

Using Storage, users can remotely store their data and enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources, without the burden of local data storage and maintenance. However, the fact that users no longer have physical possession of the outsourced data makes the data integrity protection in a formidable task, especially for users with constrained computing resources. Moreover, users should be able to just use the storage as if it is local, without worrying about the need to verify its integrity. Thus, enabling public audit ability for storage is of critical importance so that users can resort to a third party auditor (TPA) to check the integrity of outsourced data and be worry-free. To securely introduce an effective TPA, the auditing process should bring in no new vulnerabilities towards user data privacy, and introduce no additional online burden to user. In this paper, Author proposes a secure storage system supporting privacy-preserving public auditing. Author further extend our result to enable the TPA to perform audits for multiple users simultaneously and efficiently. Extensive security and performance analysis show the proposed schemes are provably secure and highly efficient.

Real-world data mining deals with noisy information sources where data collection inaccuracy, device limitations, data transmission and discretization errors, or man-made perturbations frequently result in imprecise or vague data. Two common practices are to adopt either data cleansing approaches to enhance the data consistency or simply take noisy data as quality sources and feed them into the data mining algorithms. Either way may substantially sacrifice the mining performance. In this paper, Author consider an error-aware (EA) data mining design, which takes advantage of statistical error information (such as noise level and noise distribution) to improve data mining results. Author assume that such noise knowledge is available in advance, and Author propose a solution to incorporate it into the mining process. More specifically, Author use noise knowledge to restore original data distributions, which are further used to rectify the model built from noise corrupted data. Author materialize this concept by the proposed EA naive Bayes classification algorithm. Experimental comparisons on real-world datasets will demonstrate the effectiveness of this design.

Many large organizations have multiple data sources, such as different branches of an interstate company. While putting all data together from different sources might amass a huge database for centralized processing, mining association rules at different data sources and forwarding the rules (rather than the original raw data) to the centralized company headquarter provides a feasible way to deal with multiple data source problems. In the meanwhile, the association rules at each data source may be required for that data source in the first instance, so association analysis at each data source is also important and useful. However, the forwarded rules from different data sources may be too many for the centralized company headquarter to use. This paper presents a weighting model for synthesizing high- frequency association rules from different data sources. There are two reasons to focus on high- frequency rules. First, a centralized company headquarter is interested in high-frequency rules because they are supported by most of its branches for corporate profitability. Second, high-frequency rules have larger chances to become valid rules in the union of all data sources. In order to extract high- frequency rules efficiently, a procedure of rule selection is also constructed to enhance the weighting model by coping with low-frequency rules. Experimental results show that our proposed weighting model is efficient and effective.

Authors propose a new online feature selection framework for applications with streaming features where the knowledge of the full feature space is unknown in advance. Author define streaming features as features that flow in one by one over time whereas the number of training examples remains fixed. This is in contrast with traditional online learning methods that only deal with sequentially added observations, with little attention being paid to streaming features. The critical challenges for Online Streaming Feature Selection (OSFS) include 1) the continuous growth of feature volumes over time, 2) a large feature space, possibly of unknown or infinite size, and 3) the unavailability of the entire feature set before learning starts. In the paper, Author presents a novel Online Streaming Feature Selection method to select strongly relevant and non- redundant features on the fly. An efficient Fast-OSFS algorithm is proposed to improve feature selection performance. The proposed algorithms are evaluated extensively on high-dimensional datasets and also with a real- world case study on impact crater detection. Experimental results demonstrate that the algorithms achieve better compactness and higher prediction accuracy than existing streaming feature selection algorithms.

In this paper, Author propose a new research problem on active learning from data streams, where data volumes grow continuously, and labeling all data is considered expensive and impractical. The objective is to label a small portion of stream data from which a model is derived to predict future instances as accurately as possible. To tackle the technical challenges raised by the dynamic nature of the stream data, i.e., increasing data volumes and evolving decision concepts, Author propose a classifier ensemble- based active learning framework that selectively labels instances from data streams to build a classifier ensemble. Authors argue that a classifier ensemble's variance directly corresponds to its error rate, and reducing a classifier ensemble's variance is equivalent to improving its prediction accuracy. Because of this, one should label instances toward the minimization of the variance of the underlying classifier ensemble. Accordingly, Author introduces a minimum-variance (MV) principle to guide the instance labeling process for data streams. In addition, Author derives an optimal-Authority. Calculation method to determine the weight values for the classifier ensemble. The MV principle and the optimal weighting module are combined to build an active learning framework for data streams. Experimental results on synthetic and real-world data demonstrate the performance of the proposed work in comparison with other approaches.

3. SYSTEM MODEL

Big Data starts with large-volume, heterogeneous, autonomous sources with distributed and decentralized control, and seeks to explore complex and evolving relationships among data. These characteristics make it an extreme challenge for discovering useful knowledge from the Big Data. In a naïve sense can imagine that a number of blind men are trying to size up a giant Camel, which will be the Big Data in this context. The goal of each blind man is to draw a picture (or conclusion) of the Camel according to the part of information he collects during the process. Because each parson's view is limited to his local region, it is not surprising that the blind men will each conclude independently that the camel ―feels‖ like a

rope, a hose, or a wall, depending on the region each of them is limited to.

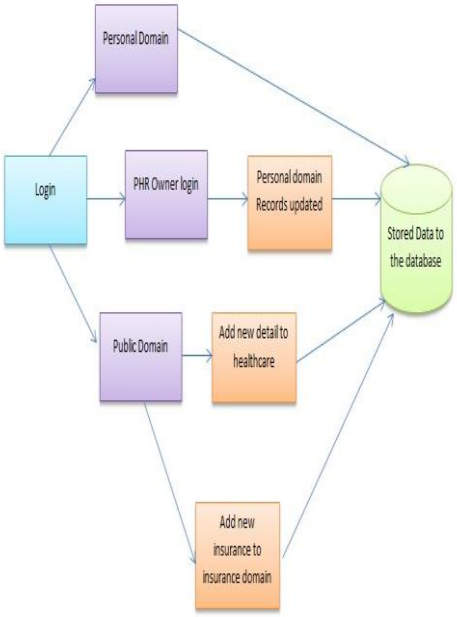

Figure 1: Architecture Diagram

To make the problem even more complicated, let us assume that the camel is growing rapidly and its pose changes constantly, and each blind man may have his own (possible unreliable and inaccurate) information sources that tell him about biased knowledge about the camel (e.g., one blind man may exchange his feeling about the camel with another blind man, where the exchanged knowledge is inherently biased).

System Model

- Registration
- Upload files
- ABE for Fine-grained Data Access Control
- Setup and Key Distribution
- Break-glass module

3.1Registration

In this module normal registration for the multiple users. There are multiple owners, multiple AAs, and multiple users. The attribute hierarchy of files – leaf nodes is atomic file categories while internal nodes are compound categories. Dark boxes are the categories that a PSD's data reader has access to. Two ABE systems are involved: for each PSD the revocable KP-ABE scheme is adopted for each PUD, our proposed revocable MA-ABE scheme.

- PUD - public domains
- PSD - personal domains
- AA - attribute authority
- MA-ABE multi-authority ABE
- KP-ABE - key policy ABE

3.2 Upload files

In this module users upload their files with secure key probabilities. The owners upload ABE-encrypted PHR files to the server. Each owner's PHR file encrypted both under a certain fine grained model.

3.3 ABE for Fine-grained Data Access Control

In this module ABE to realize fine-grained access control for outsourced data especially, there has been an increasing interest in applying ABE to secure electronic healthcare records (EHRs). An attribute-based infrastructure for EHR systems, where each patient's EHR files are encrypted using a broadcast variant of CP-ABE that allows direct revocation. However, the cipher text length grows linearly with the number of un revoked users.

In a variant of ABE that allows delegation of access rights is proposed for encrypted EHRs applied cipher text policy ABE (CP-ABE) to manage the sharing of PHRs, and introduced the concept of social/professional domains investigated using ABE to generate self-protecting EMRs, which can either be stored on servers or cell phones so that EMR could be accessed when the health provider is offline.

3.4 Setup and Key Distribution

In this module the system first defines a common universe of data attributes shared by every PSD, such as ―basic profile‖, ―medical history‖, ―allergies‖, and ―prescriptions‖. An emergency attribute is also defined for break-glass access. Each PHR owner's client application generates its corresponding public/master keys. The public keys can be published via user's profile in an online healthcare social-network (HSN).

There are two ways for distributing secret keys.

First, when first using the PHR service, a PHR owner can specify the access privilege of a data reader in her PSD, and let her application generate and distribute corresponding key to the latter, in a way

resembling invitations in Google Doc.

Second, a reader in PSD could obtain the secret key by sending a request (indicating which types of files she wants to access) to the PHR owner via HSN, and the owner will grant her a subset of requested data types. Based on that, the policy engine of the application automatically derives an access structure, and runs keygen of KP-ABE to generate the user secret key that embeds her access structure.

3.5Break-glass module:

In this module when an emergency happens, the regular access policies may no longer be applicable. To handle this situation, break-glass access is needed to access the victim's PHR. In our framework, each owner's PHR's access right is also delegated to an emergency department ED to prevent from abuse of break-glass option, the emergency staff needs to contact the ED to verify her identity and the emergency situation, and obtain temporary read keys. After the emergency is over, the patient can revoke the emergent access via the ED.

4. CONTROL OVER ACCESS TO RECORDS

To assure the control over access to records or files it is a promising method to encrypt the data before outsourcing .Here propose a novel data-centric framework and suite of mechanisms for data access control for information leverage attribute based encryption (ABE) technique to encrypt each data. Propose a novel data-centric framework and a suite of Mechanisms for data access control stored in semi-trusted servers. To achieve fine-grained and scalable data access control for files.

4.1 Reduce Key Management Complexity

In this proposed technique focus on the multiple data set scenarios. Then divide the data set in the database into multiple security domains. After this process is complete in this domain that greatly reduces the key management complexity.

5. CONCLUSION

Here focus on the multiple data set scenarios, and divide the data set in the database into multiple security domains that greatly reduces the key management complexity. Here propose a novel data-centric framework and suite of mechanisms for data access control for information, leverage attribute based encryption (ABE) technique to encrypt each data. Propose a novel data-centric framework and a suite of Mechanisms for data access control stored in semi-trusted servers. To achieve fine-grained and scalable data access control for files. To break through the limitations of traditional data mining methods, we have studied heterogeneous information discovery and mining in complex inline data, mining in data streams, multigranularity knowledge discovery from massive multisource data, distribution regularities of massive knowledge, quality fusion of massive knowledge.

REFERENCES

[1] D. Boyd and N. Ellison, ―Social network sites: Definition, history, an scholarship,‖ Journal of Computer-Mediated Communication, vol. 13,11 no. 11, October 2007.

[2] A. Chapanond, M. S. Krishnamurthy, and B. Yener, ―Graph theoretic and spectral analysis of enron email data,‖ Comput. Math. Organ. Theory, vol. 11, no. 3, pp. 265–281, 2005.

[3] X. Liu, J. Bollen, M. L. Nelson, and H. Van de Sample, ―Co-authorship networks in the digital library research community,‖ Information Processing and Management, vol. 41, no. 6, pp. 1462–1480,2005.

[4] Y. Koren, S. C. North, and C. Volinsky, ―Measuring and extracting proximity graphs in networks,‖ ACM Trans. Knowl. Discover. Data, vol 1, no. 3, p. 12, 2007.

[5] M. J. Zaki, ―Efficiently mining frequent trees in a forest,‖ in ACM KDD 02. 2002, pp. 71–80.

[6] A. Inokuchi, T. Washio, and H. Motoda, ―An apriori-based algorithm for mining frequent substructures from graph data,‖ in Proc.of the 4th European Conf. on Principles of Data Mining and Knowledge Discovery. Springer Verlag, 2000, pp. 13–23.

[7] J. Huan, W. Wang, and J. Prins, ―Efficient mining of frequent sub graph in the presence of isomorphism,‖ in Proc. of the 3rd IEEE Int. Conf. on Data Mining (ICDM), 2003, pp. 549–552.

[8] M. Kuramochi and G. Karypis, ―An efficient algorithm for discovering Frequent sub graphs,‖ IEEE TKDE, vol. 16, no. 9, pp. 1038–1051, 2004.

[9] A. Labrinidis and H. Jagadish, ―Challenges and Opportunities with Big Data,‖ Proc. VLDB Endowment, vol. 5, no. 12, 2032-2033,2012.

[10] Y. Lindell and B. Pinkas, ―Privacy Preserving Data Mining,‖J. Cryptology, vol. 15, no. 3, pp. 177- 206, 2002.

[11] J. Lorch, B. Parno, J. Mickens, M. Raykova, andJ. Schiffman,―Shoroud: Ensuring Private Access to Large-Scale Data in the Data Center,‖ Proc. 11th USENIX Conf. File and Storage Technologies (FAST '13), 2013.

[12] X. Wu, K. Yu, W. Ding, H. Wang, and X. Zhu,―Online Feature Selection with Streaming Features,‖ IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 35, no. 5, pp. 1178-1192, May2013.

[13] X. Zhu, P. Zhang, X. Lin, and Y. Shi, ―Active Learning From Stream Data Using Optimal Weight Classifier Ensemble,‖ IEEE Trans. Systems, Man, and Cybernetics, Part B, vol. 40, no. 6, pp. 1607- 1621, Dec. 2010.

[14] C. Ranger, R. Raghuraman, A. Pelmets, G. Bradski, and C. Kozyrakis, ―Evaluating Map Reduce for Multi-Core and Multiprocessor Systems,‖ Proc. IEEE 13th Int'l Symp. High Performance Computer Architecture (HPCA '07), pp. 13-24, 2007.

[15] C. Wang, S.S.M. Chow, Q. Wang, K. Ren, andW. Lou, ―Privacy-Preserving Public Auditing for Secure Storage‖ IEEE Trans.Computers, vol. 62, no. 2, pp. 362-375, Feb. 2013.