

# Enabling Real-Time Business Intelligence by Stream Mining

Simon Fong and Yang Hang  
*University of Macau,  
Macao SAR*

## 1. Introduction

Traditionally Business Intelligence (BI) is defined as “a set of mathematical models and analysis methodologies that exploit the available data to generate information and knowledge useful for complex decision making processes” (Vercellis, 2009). The real-time aspect of BI seems to be missing from the classical studies. BI systems technically combine data collection, data storage, and knowledge management with analytical tools to present complex and competitive information to business strategic planner and decision makers (Negash, 2003). This type of BI systems or architectures has served for business usage for past decades (Rao, 2000).

Nowadays businesses evolved to be more competitive and dynamic than the past, which demand for real-time BI and capability of making very quick decisions. With this new business market demand, recently published works (Yang & Fong, 2010; Sandu, 2008) advocated that BI should be specified in four dimensions: strategic, tactical, operational and real-time. Most of the existing decision-support systems however are strategic and tactical; BI is produced by data mining either in forms of regular reports or some actionable information in digital format within a certain frame time. Although the access to the BI database (sometimes called Knowledge base) and the decision generated from data-mining rules are instant; the underlying historical data used for analysis may not be most up-to-the-latest-minute or seconds.

Compared with the operational BI, real-time BI (rt-BI) shall analyze the data as soon it enters the organization. The latency (data latency, analysis latency, decision latency) shall be zero ideally. In order to establish such real-time BI systems, relevant technologies to guarantee low/zero latency are necessary. For example, operational / real-time BI data warehouse techniques are able to provide fresh data access and update. Thus operational BI can be viewed as rt-BI as long as it can provide analytics within a very short time for decision making. The main approach is: system response time shall stay under a threshold that is less than the action taking time; and the rate of data processing shall be faster than the rate of data producing. However, there are many real-time data mining algorithms in theoretical fields, but their applicability and suitability towards various real-time applications are still vague; so far no one has conducted an in-depth study for rt-BI with consideration of stream-mining. We take this as the research motivation and hence the contribution of this chapter.

The chapter is structured in the following way: Section 2 is an overview of rt-BI system; the high-level framework, system architecture and process are described. Section 3 is a

discussion of how rt-BI could be applied in several typical application scenarios. Section 4 details a set of experiments by simulating the different impacts of traditional data-mining and stream-mining in rt-BI architecture. A conclusion is drawn in the last section

## 2. rt-BI architecture

### 2.1 Overview of the rt-BI framework

rt-BI system relates to many technologies and tools evolved from strategic BI and tactical BI. Following previous research, a four-layer framework is proposed for rt-BI system in Figure 1. The main improvement is a real-time processing of whole knowledge discovery process.

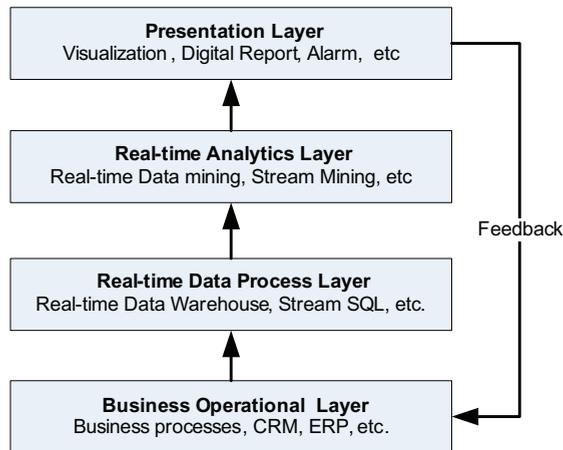


Fig. 1. Four-layer Framework

#### A. Business operational layer

This layer composes of two primary functions: business activity monitoring (BAM) and real-time process tuning (McCoy, 2002). Azvine (Azvine et al., 2006) presents the shortcoming of current BAM and process tuning technology for rt-BI: 1) current BAM can't make intelligent conclusion about the overall business process behavior; 2) and real business processes changes are carried throughout initiatives manually, that is expensive and time consuming. On the other hand, the level of automation is divided into two stages: semi- and fully-automatic. Our proposed framework tackles these problems with a fully-automated process. The system is built right on the top of business operations. It shall facilitate automated mapping of existing business operations within an organization, capture the knowledge to automate process tuning, optimization and re-engineering, and monitor people and systems for process conformance.

#### B. Real-time data process layer

This layer is responsible for providing qualified data to its upper layer – analytics layer. Data come from various resources in different formats. If the data contain too much noise, it will impair the business intelligence discovery. In this layer, the system is required to obtain the quality data within a time constraint. For this reason, preparation process should not take too long. Modern digital source is a kind of large volume and rapidly changing data.

Data stream technology (Botan, 2009) provides a good solution to build real-time data warehouse, with which increased refresh cycles to frequently update the data. This kind of data warehouse systems can achieve nearly real-time data updating, where the data latency typically is in the range from seconds to minutes.

### C. Real-time analytic layer

Traditionally, data analyzing follows “analyst-in-the-middle” approach where human expert analysts are required to drive or configure the information with BI software and tools. But such manual task will incur analysis latency. To this end, the analysis tools should provide a high degree of automation, which is relating to artificial intelligence technology. Data miners serve as the kernel to build models or extract patterns from large amounts of information (Hand & Mannila, 2001; Hand, 1999, Hoffmann et al., 2001). Analytics layer uses fast data mining method to interpret data to information. So far there are many real-time data mining algorithms and methodologies. The four popular types are: clustering, classification, frequency counting, and time series analysis. Stream processing engines are also used based on sliding windows technology (Dong, 2003).

### D. Presentation layer

This layer presents the BI to end-user in a highly interactive way in order to shorten action latency. The presentations vary in formats and designs. For examples, sophisticated time-series charts show a trend, and a KPI dashboard alarms off anomalies etc. Many companies provide these techniques as third party solutions, iNetSoft, SPSS, IBM, etc.

## 2.2 System architecture

Traditionally, the classic method to build model with data mining algorithm is training-then-testing approach. But the weakness is they may not suit large volume and high speed data.

A Mining Model Definition Language (MMDL) is used for stream mining system (Thakkar, 2008), but it has not illustrated how to design a stream mining system in a technical sense. A research (Stonebraker, 2005) proposed three real-time data stream processing architectures which can potentially be applied to solve high-volume low latency streaming problems but its both Rule Engine and Stream Process Engine architectures only rely on stream data querying (SQL). Mining data streams has been studied by many researchers. Gaber (Gaber, 2005) summarized the most cited data stream mining techniques with respect to different mining tasks, approaches and implementations. They proposed an adaptive resource-aware approach called Algorithm Output Granularity (AOG) (Gaber, 2004; Gaber, 2008).

The rt-BI system architecture described in this section is derived from the previous research in data mining and business intelligence. Different from the previous ones, the proposed architecture concentrates on constructing a system which is able to extract potential BI and return result to end-user in real-time.

Figure 2 shows a static view of rt-BI system architecture. *Firstly*, the rt-BI system collects a large amount of historical data from existing information system. *Secondly*, the system collects and monitors the new input data in real-time data process layer. If necessary it will transform the data into adequate forms. *Thirdly*, the system determines whether it relates to an established model in the real-time analytics layer. If so, the system matches it with the rules and returns BI result. Otherwise, the system runs on data-mining process in order to find new rules and BI. A newly found model is updated to the rule-based database. *Fourthly*, the discovered information is summarized as rt-BI result and presented in appropriate formats. During this process, any mis-prediction or incorrect-pattern will be updated to

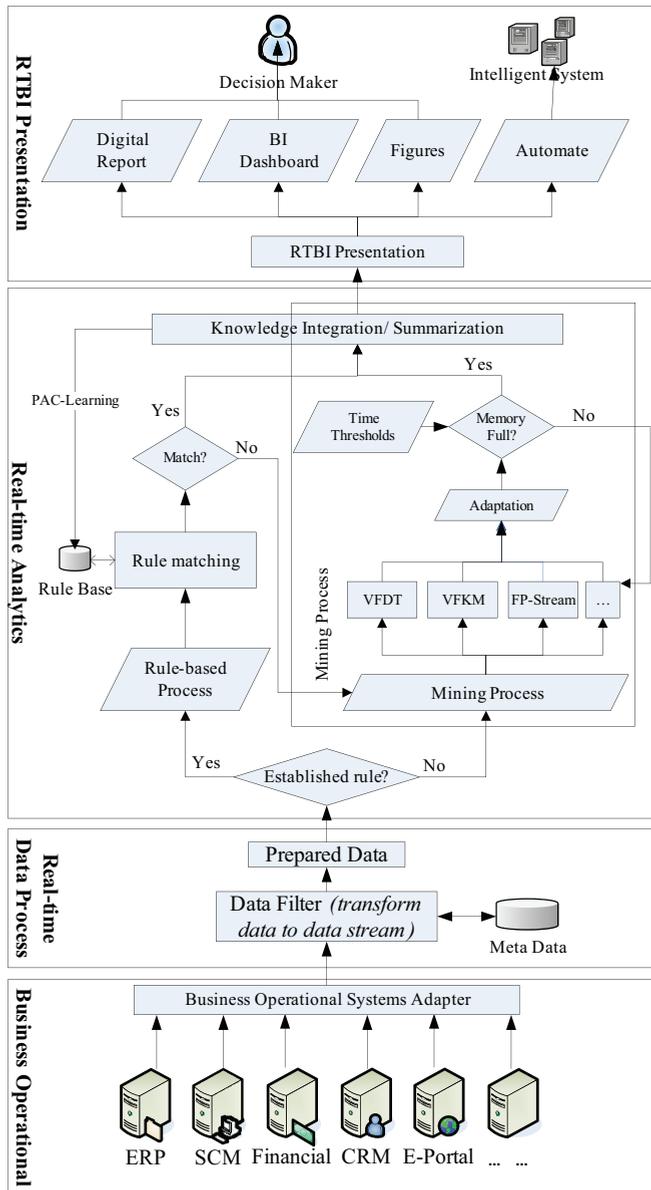


Fig. 2. rt-BI System Architecture

database in the case of next data mining happens. This process should be within a certain time threshold that the BI output is useful for decision making (to ensure no analytics latency). By this architecture, the system collects data and generates some prediction models in real-time. The data used to discover BI is not only dependent on historical but also the new incoming data.

### 2.3 rt-BI analyzing process

The analyzing kernel of an rt-BI system is the mining process. In this section, we show a dynamic diagram in Figure 3 to show how to implement the data mining process in RT-BI system.

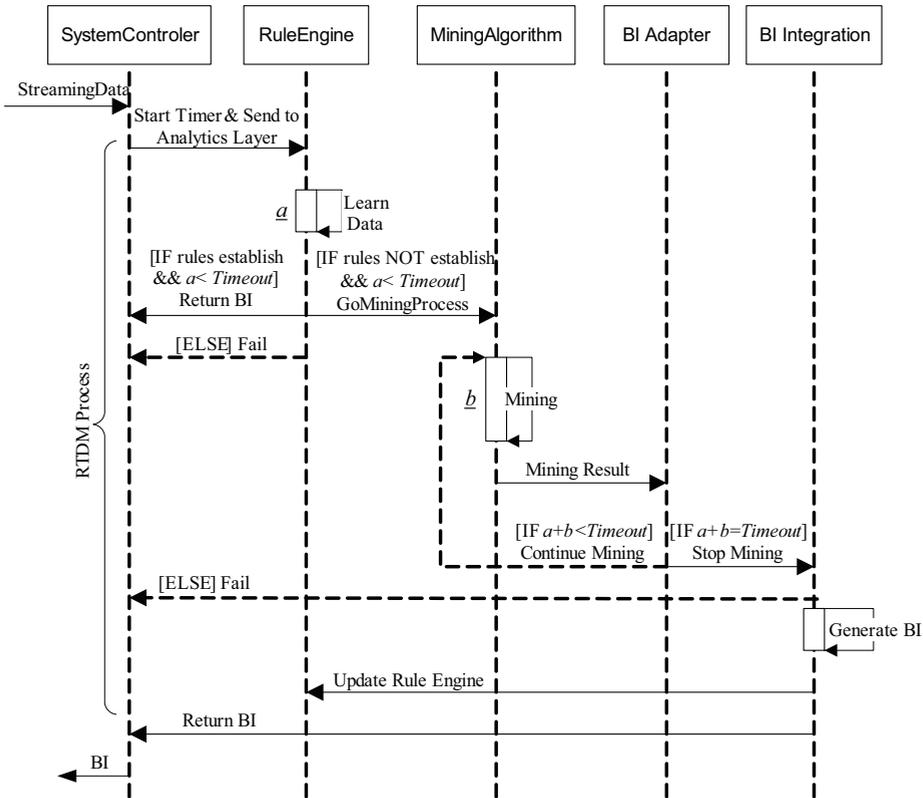


Fig. 3. rT-BI Generating Workflow

The process contains two segments: rule-based matching, and new BI mining. When new data comes, a timer is started to control the rt-BI running time so as to restrict analytics latency within an acceptable level. A timeout threshold is determined by the time required to make a decision, which restricts the rule-based matching time as well as the BI mining time. If new arrival data are correlating to the already established rules, the rule-based matching process activates and returns the BI within the time threshold. Otherwise, the new BI mining process will trigger. The determination should be within the time threshold. If it timeouts, the rt-BI system is deemed failed to discover new BI and returns the most recently updated information instead.

### 3. Applications of RT-BI system

The proposed system architecture can be applied in different applications. We illustrate four typical application domains. A more comprehensive comparison is presented in Table 1.

### *A. Anomaly detection and automated alerts*

Anomaly Detection refers to detecting patterns in a given data set that do not conform to an established normal behavior (Hodge, 2004). The detected patterns are called anomalies, which are also referred to as outliers, surprise, aberrant, deviation, peculiarity, etc. and often translated to critical and actionable information in several application domains. Many anomaly detection techniques have been specifically developed for certain application domains, while others are more generic. Its application domains mainly include: insurance fraud detection (Phua et al., 2004), network attack detection (Zhengbing et al., 2008), and credit card fraud detection (Quah & Sriganesh, 2008; Whitrow et al., 2009), etc. A survey (Chandola et al., 2009) tries to provide a structured and comprehensive overview of the research on anomaly detection, but it doesn't give a generic design for such kind of rT-BI system. This type of applications is not only required to find the anomaly pattern from a large amount of data in real-time, but to present the result to end-user reliably and take action efficiently.

### *B. Prediction and suggestions recommender*

Customer Relationship Management (CRM) systems apply data mining to analyze and predict the potential customer values. Although the analysis of available information for those customers who in the past have purchased product or services based on the historical data, and the comparisons with the characteristics of those who have not taken up the offer of the enterprise, it is possible to identify the segments with the highest potential. Commercial recommender systems use various data mining techniques to provide appropriate recommendations to users during real-time online sessions. E-business transactions usually take place over online networks. For analyzing e-Portal information, rT-BI system is recommends suitable suggestions to customers. A context-similarity based hotlinks assignment (Antoniou et al., 2009) analyzes the similarity of context between pages in order to suggest the placement of suitable hotlinks. Another real-time recommendation system based on experts' experiences is proposed in (Sun et al., 2008). It simplifies content-based filtering through computing similarity of the keywords and recommends common users the Web pages based on experts' search histories but not the whole Web pages. Online recommender systems often use the suggested purchase items, or the items in which costumer may be also interested, as the presentation of rT-BI. These techniques widely used in call centers to make investigation service in terms of the telephone call data stream.

### *C. Forecast and markets analysis*

Pricing network resources is a crucial component for proper resource management and the provision of quality of service guarantees in different markets. A model used data mining to forecast the stock market with time series (Dietmar et al., 2009). A competitive market intelligence system (Weiss & Verma, 2002) proposes to detect critical differences in the text written about a company versus the text for its competitor. However, the intelligence system is compelled to depend on empirical performance, which has to require human interaction to analyze the discovered patterns. As aforementioned, the latency is the primary constraint of operational BI and real-time BI. A business cannot respond to events as they happen if it cannot find out about these events for hours, days, or weeks. It also cannot immediately respond to events if the system that supplies the analyses of these events is down. If the problems of data latency and data availability are solved, then businesses react proactively to new information and knowledge rather than reactively.

Real-time business intelligence dashboards are used to bridge the gap between operational business intelligence and real-time business intelligence. It shall display not only historical information but also show the current status to support decision making.

#### *D. Optimization and supply chain management*

Supply Chain Management (SCM) is one of the hottest topics in e-Commerce. Online business transaction builds a dynamic pricing model that is integrated into a real-time supply chain management agent (Ku et al., 2008). Besides the pricing strategy, real-time supply chain management in a rapidly changing environment requires reactive and dynamic collaboration among participating entities. Radio Frequency Identification (RFID) is widely used in high-tech arena. It is described as a major enabling technology for automated contactless wireless data collection, and as an enabler for the real-time enterprise. Goods are supervised while they are embedded with RFID tags. The tags can send out electronic signal through its inside antenna. After capturing the data stream by sensors, RFID system is aware of the information of the goods, such as location and status. The real-time supervising and gaining visibility can achieve quick responsiveness and high efficiency in business flows, if RFID technology can be applied efficiently (Gonzalez et al., 2006).

The proposed architecture may address the challenge of processing high-volume, real-time data with requiring the use of custom code. rt-BI systems provide pattern discovery, trend detection, and visualization, controlling and improving the flow of materials and information, originating from the suppliers and reaching the end customers.

## **4. Experiments**

### **4.1 Simulation setup**

In our experiment, a simulation is programmed to verify the proposed framework. Simulated “real-time” environment runs through Massive Online Analysis (MOA), a framework for data stream mining. (Source: University of Waikato, [www.cs.waikoto.ac.nz](http://www.cs.waikoto.ac.nz)). MOA consists of a library of open source JAVA API extending from WEKA data mining. The experiment platform is a PC with 2.99 GHz CPU and 1 GB RAM. The main procedure flow is shown in Figure 4.

Experiments are performed on both synthetic data and real data. Synthetic data are generated by Random Tree Generator provided by MOA. Real data are collected from PKDD'99 conference. (Source: <http://lisp.vse.cz/pkdd99/Challenge/chall.htm>). We chose them because they came from real-world financial and banking applications. This procedure is simulating *Business Operational Layer* process.

The collected data are in different formats. Thus, a data filter simulates *Real-time Data Process Layer*, emerging different tables, eliminating noisy data, and converting different types of files to MOA readable ARFF format. The sensor is accountable in implementing the following tasks: partition datasets from data source with a pre-configured size; and transfer partitioned data to a mining engine in an interval time, simulating a continuous stream flow.

*Real-time Analytics Layer* uses Hoeffding Tree (Hoeffding, 1963) as the core fast mining algorithms. As a pre-configured threshold of maximum memory size (window size), remaining available time (time out threshold), a decision tree is generated for each window in real-time. With the time passing by, the tree structure changes simultaneously. After a decision tree is built, a set of test stream data are used to test the accuracy. With the tree structure changing, a simple chart is reported in *Presentation Layer*.

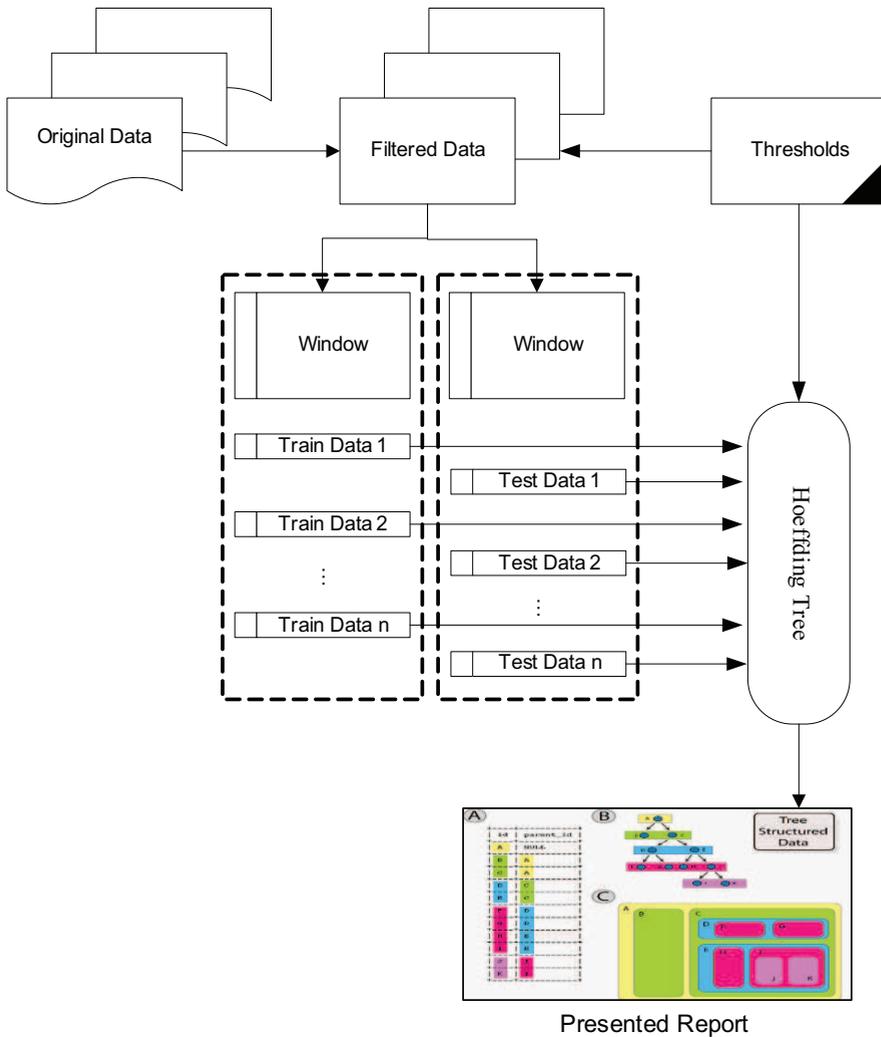


Fig. 4. Procedure Workflow

**4.2 Results and discussions**

Discussion of how rt-BI could be applied in several typical application scenarios. In the simulation, Hoeffding Tree algorithm is the main algorithm of very fast decision tree classification for real-time data mining. It is used with VFML10 (Bernhard et al., 2008) numeric estimator. The accuracy percentage is calculated by the Basic Classification Performance Evaluator provided in MOA, which refers to following formula:

$$Accuracy = \frac{CorrectObervationNumber}{TotalObervationNumber} \times 100\%$$

### A. Accuracy and window size

We extracted data segments of various sizes of windows from the same resource, respectively 1K, 5K, 10K, 100K, 250K, 500K, 750K and 1000K bytes. A streaming environment is simulated that data comes into rt-BI system continuously

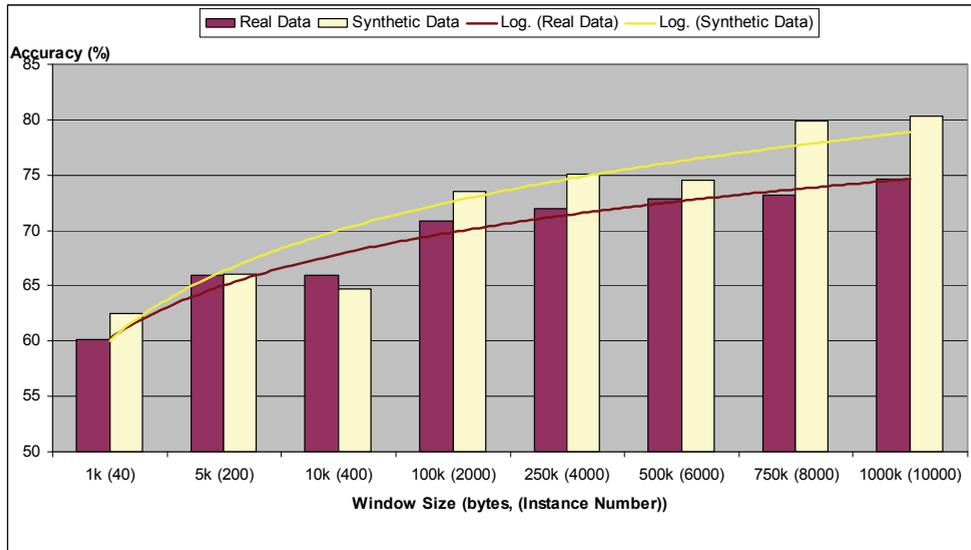


Fig. 5. Average Predication Accuracy Comparison

Generally, the synthetic data have a better performance than the real data result due to the controlled sparsity, but both the trends are correspondingly rising up. The accuracy increases with the window size growing (Figure 5). This experiment shows that: the larger the window size is, the higher accuracy. Thus, when it comes to designing a rt-BI system we have to consider what the optimal window size it should take. However, a large window size also yields certain delay.

### B. Window size and complexity

Data mining algorithm has different efficiency in terms of its complexity. There are some factors influencing the complexity. For instance in our experiment, Hoeffding Tree's complexity is effected by the serialized size. The bigger this size is, the higher the complexity. This size also influences the need of computing resource. In our simulation, when the size of window gets large, the Java virtual machine was allocated a generous amount of memory. It strongly indicates that the serialized size was the cause. We tested different windows with increasing size from 1 Kb to 1M bytes, and observed this trend.

Figure 6 shows that: (1) a significant growth of serialized size appears from 1Kb to 500 Kb; (2) and a relatively slowly growing appears from 500 K to 1 M bytes. In other words, the relatively high complexity is found in the bigger data size region; the complexity becomes steady when a threshold arrives, so the increasing trend slows down. When we design a system, we shall consider this threshold and required response time together. If the required response time (timeout) is less than this threshold, we will achieve zero analysis latency as a result.

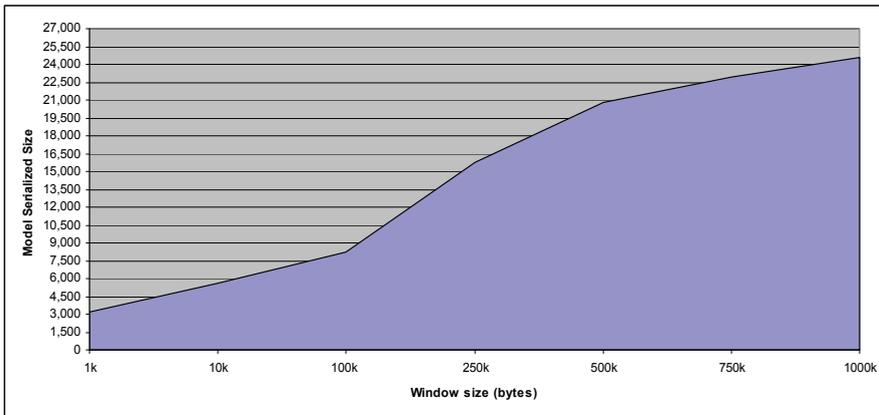


Fig. 6. Average Tree Complexity Comparison

C. Traditional and real-time learning methods

The biggest difference between traditional BI and rt-BI discovery is the data analyzing method. Due to the differences of the learning methods, rt-BI uses windowing technique, instead of training and testing the whole structured dataset as in traditional BI. The experimental data is the same as that in the previous experiments. From Figures 7 and 8, a classic classification method – Naïve Bayesian – is used in the traditional learning method. The ratio of training data size and testing data size is 1:1, 1:2, 2:1 respectively. rt-BI method based on Hoeffding Tree applies windowing technique. We can observe a significant trend that: traditional BI has a much better accuracy than rt-BI when the data size is relatively small; however, this advantage is not observed any more while the data size (window size) increases to a certain extent.

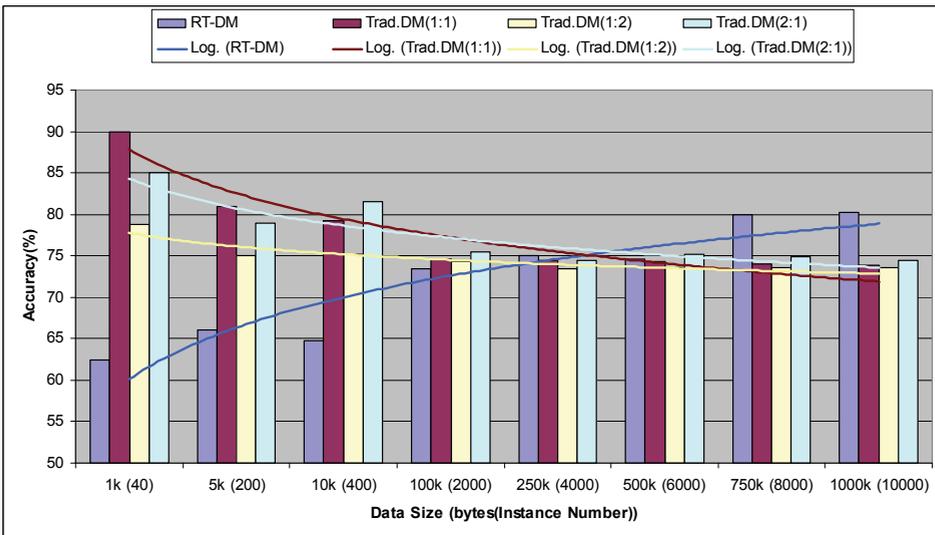


Fig. 7. Predication Accuracy Comparison between Bayesian and HTA

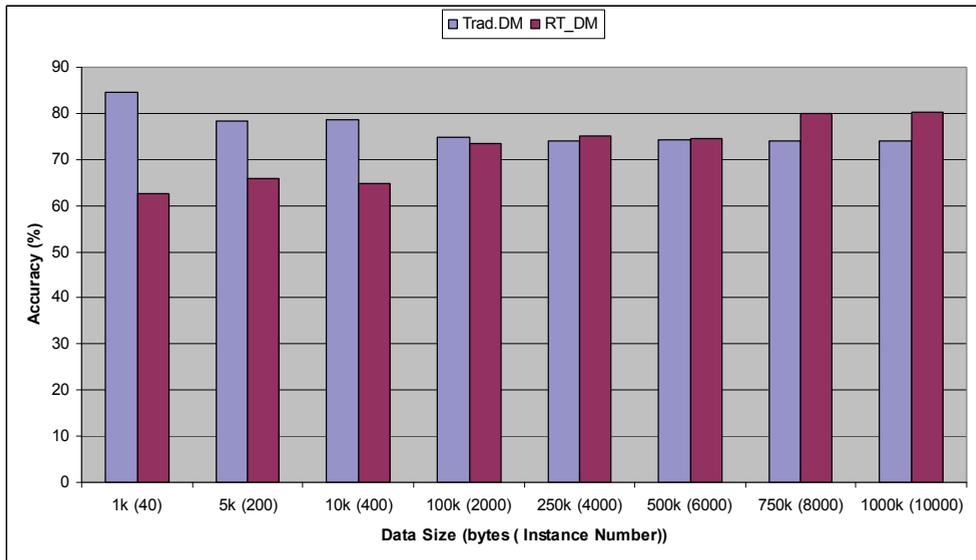


Fig. 8. Average Predication Accuracy Comparison between Bayesian and HTA

Figure 9 gives a comparison between a classic traditional decision tree J48 C4.5 and Hoeddfig Tree in our experiment. Synthetic data are used as input. We can see that both of them display an increasing accuracy while the data size is growing, C4.5 has a better performance than Hoeddfig tree but the former tree takes much more time than the latter one. For this reason, it may be unsuitable for rt-BI. Thus, we may make a short conclusion that: compared with traditional BI method (that is comprised of complete data training and testing), rt-BI equipped with stream data mining method obtains a better performance, and therefore it does suit environments characterized by huge and streaming data

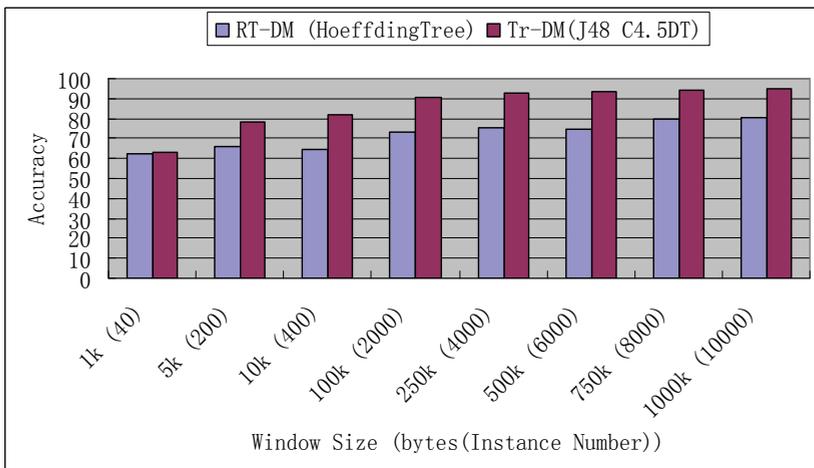


Fig. 9. Average Predication Accuracy Comparison by Window Sizes

## 5. Conclusion

Real-time business intelligence is a new concept in knowledge discovery. rt-BI requires exploring BI from a large volume and rapidly arriving data in business operations. rt-BI system aims to achieve very short time required in data process and analysis process for decision making. We proposed a generic framework architecture for rt-BI, followed by a discussion of rt-BI applications.

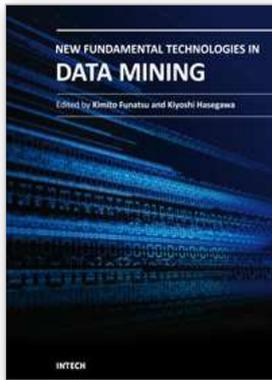
In addition, a simulation experiment is developed to validate the stream-mining performance. The results show that: 1) the window size is a key to determine the algorithm's accuracy in rt-BI system design; 2) the proposed framework is able to achieve nearly zero analysis latency within a threshold timeout. This shows using stream mining in rt-BI is desirable; 3) compared with traditional BI method, the rt-BI method has a better performance for a large volume of high speed streaming data.

## 6. References

- Antoniou, D.; Garofalakis, J.; Makris, C.; Panagis, Y.; Sakkopoulos, E. (2009). Context-similarity based hotlinks assignment: Model, metrics and algorithm, *Data & Knowledge Engineering, In Press, Corrected Proof, Available online*, 4 May 2009.
- Azvine, B.; Cui, Z.; Nauck, D. & Majeed, B. (2006). Real Time Business Intelligence for the Adaptive Enterprise. *Proceedings of the the 8th IEEE international Conference on E-Commerce Technology and the 3rd IEEE international Conference on Enterprise Computing, E-Commerce, and E-Services (CEC-EEE)*, pp. 29, San Francisco, CA, June 2006, IEEE Computer Society, Washington, DC
- Bernhard, P.; Geoffrey, H. & Richard, K. (2008). Handling Numeric Attributes in Hoeffding Trees, *Lecture Notes in Computer Science*, Springer Berlin/Heidelberg Volume 5012, 2008, pp. 296-307.
- Botan, I.; Cho, Y.; Derakhshan, R.; Dindar, N.; Haas, L., Kim, K. & Tatbul, N. (2009). Federated Stream Processing Support for Real-Time Business Intelligence Applications, *VLDB International Workshop on Enabling Real-Time for Business Intelligence (BIRTE'09)*, Lyon, France, August 2009
- Chandola, V.; Banerjee, A. & Kumar, V. (2009). Anomaly Detection: A Survey, *ACM Computing Surveys*, Vol. 41(3), Article 15, July 2009
- Dietmar, H.; Dorr, A. & Denton, M. (2009). Establishing relationships among patterns in stock market data, *Data & Knowledge Engineering*, Volume 68, Issue 3, March 2009, pp. 318-337
- Dong, G.; Han, J.; Lakshmanan, L.V.S.; Pei, J.; Wang, H. & Yu, P.S. (2003). Online mining of changes from data streams: Research problems and preliminary results, *Proceedings of the 2003 ACM SIGMOD Workshop on Management and Processing of Data Streams. In cooperation with the 2003 ACM-SIGMOD International Conference on Management of Data*, San Diego, CA, June 2003
- Gaber, M.; Zaslavsky, A. & Krishnaswamy, S. (2004). Towards an Adaptive Approach for Mining Data Streams in Resource Constrained Environments, *Proceedings of Sixth International Conference on Data Warehousing and Knowledge Discovery - Industry Track*, 2004

- Gaber, M.; Zaslavsky, A. & Krishnaswamy, S. (2005). Mining data streams: a review. *SIGMOD*, Rec. 34, 2, 2005, pp. 18-26
- Gaber, M. (2009). Data Stream Mining Using Granularity-Based Approach, *Studies in Computational Intelligence, Foundations of Computational*, Volume 6, 2009, pp. 47-66
- Gonzalez, H.; Han, J. & Li, X. (2006). Mining compressed commodity workflows from massive RFID data sets. *Proceedings of the 15th ACM international Conference on information and Knowledge Management. CIKM '06*. ACM, New York, NY, 2006, pp. 162-171.
- Hand, D. (1999). Statistics and Data Mining: Intersecting Disciplines, *ACM SIGKDD Explorations*, Vol.1, No.1, June 1999, pp. 16-19.
- Hand, D.; Mannila, H. & Smyth, P. (2001). Principles of data mining, *MIT Press*, 2001
- Hodge, V.J. & Austin, J. (2004). A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, Kluwer Academic Publishers, 2004, pp. 85-126.
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of American Statistic Association*, Volume 58, 1963, pp. 13-30
- Hoffmann, F.; Hand, D.J.; Adams, N.; Fisher, D. & Guimaraes, G. (2001). Advances in Intelligent Data Analysis. *Springer*, 2001
- Ku, T.; Zhu, Y. & Hu, K. (2008). A Novel Complex Event Mining Network for RFID-Enable Supply Chain Information Security. *Proceedings of the 2008 international Conference on Computational intelligence and Security*. Volume 1, CIS. IEEE Computer Society, Washington, DC, 2008, pp. 516-521.
- McCoy, D.W. (2002). Business Activity Monitoring: Calm Before the Storm, *Gartner Research*, ID: LE-15-9727
- Phua, C.; Alahakoon, D. & Lee, V. (2004). Minority report in fraud detection: classification of skewed data. *SIGKDD Explore News*. Vol. 6, No. 1, June 2004, pp. 50-59.
- Quah, J.T. & Sriganesh, M. (2008). Real-time credit card fraud detection using computational intelligence. *Expert Syst. Applications*. 35, 4, pp. 1721-1732.
- Sandu, D. (2008). Operational and real-time Business Intelligence, *Revista Informatica Economica*, Vol. 3, No. 47, pp. 33-36
- Stonebraker, M.; Çetintemel, U.; and Zdonik, S. (2005). The 8 requirements of real-time stream processing. *SIGMOD Rec*. 34, 4, 2005, pp. 42-47
- Sun, J.; Yu, X.; Wu, Z. & Li, X. (2008). Real Time Recommendation Utilizing Experts' Experiences. *Proceedings of the 2008 Fifth international Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. Volume 5, IEEE Computer Society, Washington, DC, 2008, pp. 379-383
- Whitrow, C.; Hand, D.J.; Juszczak, P.; Weston, D. & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge. Discovery*. Vol. 18, No. 1, 2009, pp. 30-55.
- Weiss, S. M. & Verma, N. K. (2002). A system for real-time competitive market intelligence.. *Proceedings of the Eighth ACM SIGKDD international Conference on Knowledge Discovery and Data Mining. KDD '02*. ACM, New York, NY, 2002, pp. 360-365.
- Yang, H. & Fong, S. (2010). Real-time Business Intelligence System Architecture with Stream Mining, *Proceedings of the 5th International Conference on Digital Information Management (ICDIM 2010)*, July 2010, Thunder Bay, Canada, IEEE.

Zhengbing, H.; Zhitang, L. & Junqi, W. (2008). A novel Network Intrusion Detection System (NIDS) based on signatures search of data mining. *Proceedings of the 1st international Conference on Forensic Applications and Techniques in Telecommunications, information, and Multimedia and Workshop*, Adelaide, Australia, January 21 - 23, 2008, ICST, Brussels, Belgium, pp. 1-7.



## **New Fundamental Technologies in Data Mining**

Edited by Prof. Kimito Funatsu

ISBN 978-953-307-547-1

Hard cover, 584 pages

**Publisher** InTech

**Published online** 21, January, 2011

**Published in print edition** January, 2011

The progress of data mining technology and large public popularity establish a need for a comprehensive text on the subject. The series of books entitled by "Data Mining" address the need by presenting in-depth description of novel mining algorithms and many useful applications. In addition to understanding each section deeply, the two books present useful hints and strategies to solving problems in the following chapters. The contributing authors have highlighted many future research directions that will foster multi-disciplinary collaborations and hence will lead to significant development in the field of data mining.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Simon Fong and Yang Hang (2011). Enabling Real-Time Business Intelligence by Stream Data Mining, New Fundamental Technologies in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-547-1, InTech, Available from: <http://www.intechopen.com/books/new-fundamental-technologies-in-data-mining/enabling-real-time-business-intelligence-by-stream-data-mining>

# **INTECH**

open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.