# A Framework of Business Intelligence-driven Data Mining for e-Business

Yang Hang, Simon Fong
Dept. of Computer and Information Science
University of Macau
Macau SAR
ma76562@umac.mo, ccfong@umac.mo

***Abstract*—This paper proposes a data mining methodology called Business Intelligence-driven Data Mining (BIdDM). It combines knowledge-driven data mining and method-driven data mining, and fills the gap between business intelligence knowledge and existent various data mining methods in e-Business. BIdDM contains two processes: a construction process of a four-layer framework and a data mining process. A methodology is established in setting up the four-layer framework, which is an important part in BIdDM. A case study of B2C e-Shop is provided to illustrate the use of BIdDM.**

***Keywords-Business intelligence; Data mining; BI-driven Data Mining***

## I. INTRODUCTION

e-Business has been rapidly evolving in the last two decades. Intelligent B2C recommender, smart online e-service, knowledge-driven customer-relation-management are some examples of emergent e-Services on Internet. Behind such e-Services, there are complex networked computing systems and tremendous amounts of data stored in databases. With the growth of demand, situational service that is comprised of two or more disparate e-Services are required, which have been combined to create a new integrated experience. As such, e-Service integration eventually becomes a mash-up system in terms of the e-Service construction changes.

Business Intelligence (BI) is a concept of applying a set of technologies to convert data into meaningful information [1]. BI methods include information retrieval, data mining, statistical analysis as well as data visualization. Large amounts of data originating in different formats and from different sources can be consolidated and converted to key business knowledge. Figure 1 presents a general view on how data are transformed to business intelligence. The process involves both business experts and technical experts. It converts a large scale of data to meaningful outcomes so as to provide decision-making support to end users.

On the other hand, DM is a core process to transform data to meaningful patterns. Rules can be extracted from DM models heuristically as output results. Traditionally, there are two types of data mining approaches: verification-driven and discovery-driven [2]. The former one is to purport some hypothetical association or pattern and then examine the data to find proof; while the latter one is relying on sophisticated manipulation of the data to discover associations, patterns, rules or functions. DM contains three components: data capturing, data mining and information representation. So far, however, most researches are focusing on either algorithm implementation in technical layer or knowledge representation in business layer. The two DM approaches however have been applied separately in many cases.

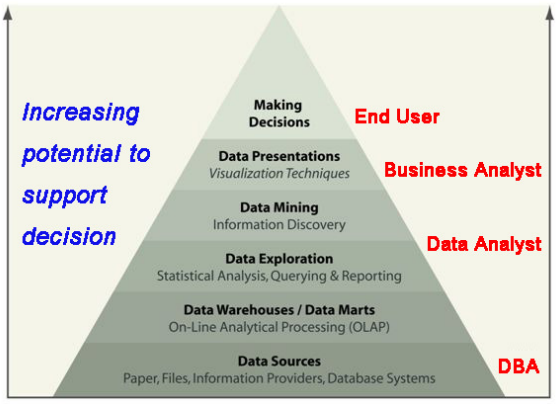


Figure 1. Business Intelligence Processing

Business Intelligence-driven Data Mining (BIdDM) is a new data mining methodology combining knowledge-driven data mining and method-driven data mining. It aims to propose a way to fill the gap between business intelligence knowledge in e-commerce and current various data mining methods. BIdDM pursues a flexible way to implement data mining in terms of business requirement. It contains two processes: four-layer framework construction and data mining. The four layers are knowledge layer, e-Service layer, method layer and data layer. The components in each layer are re-usable. The components can be flexibly added or dropped.

## II. BACKGROUND

### A. *Knowledge Discovery Proces*

The knowledge discovery process (KDP), is constantly seeking new knowledge in application domain. It is defined as a nontrivial process of identifying valid, novel, potentially meaningful, and ultimately understandable patterns in data [4]. It consists of many steps.-Each step attempts to complete

a particular discovery task and each accomplished by the application of a discovery method. KDP concerns about the following steps: how data are stored and accessed, how to use efficient algorithms to analyze massive datasets, how to interpret and visualize results, and how to model and support interactions between human and machine.

In general, BI can be generated directly from DM methods and the required data. SEMMA and CRISP-DM are two common methodologies to implement data mining. SEMMA focuses on the discovering and mining potential meaning from a data set; while CRISP-DM proposes a project implementation for data mining process. However, traditional methodologies like those requires substantial amount of human intervention.

### B. *Distributed E-Service Architecture*

Figure 2 shows a kind of new experience creation resulted from e-Service systems' frequent interaction. Such e-Services are distributed and their information transfer over the Internet. Service is on highest layer - application layer – in analogy to the ISO Seven-Layer Model [5]. Each e-Service has its own backend structure on downward layers, either distributed or centralized. Local database is providing a storage medium for system to place the relevant information, which is the data source of KDP.

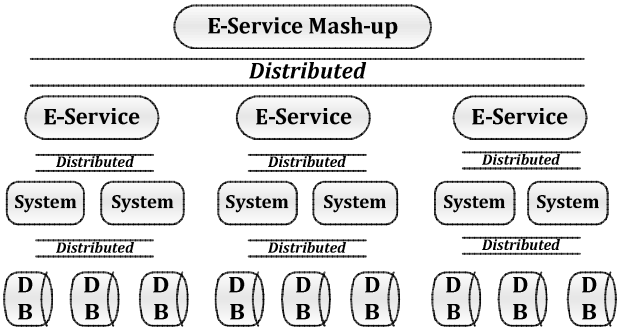


Figure 2. e-Service distributed architecture

### C. *Distributed Data Mining*

Distributed Data Mining (DDM) is the mining of distributed data sets, which are stored in different local databases. They are hosted by local servers and connected through a computer network. The local DM findings are composed to gain a global result [11]. DDM design follows a bottom-up approach, which relates to data fragmentation, replication, communication and integration tasks [6]. However DDM faces some challenges that include applying the discrete output and intelligent algorithms to obtain meaningful results without too much computation cost; ensuring it's capable of working with continual and categorical variable symmetrically; and the communication asynchronous problem in network. Under the e-Service mash-up's distributed architecture, DM over the e-Services becomes distributed as well. There have been many researches on DDM, for example, distributed classification and distributed clustering [6]. However, data are distributed in different formats, noisy data impede on DDM systems.

### D. *Method-driven and Knowledge-driven Data Mining*

Method-driven DM is a common method on data mining nowadays. It works on developing and applying algorithms to model and mine data. Until now, there have been a large n of studies established on the basics of mining algorithms, both for centralized and distributed data mining. For example, k-means clustering algorithms are developed to recover clusters that are hyper-spherical in shape [7]. Boosted regression tree [8] is used to model data fitting small regression trees to cases. As a result, many papers published in the research community discuss performance issues and merits of some particular algorithms.

Knowledge-driven DM advocates that the attention in data mining is directed from the type of knowledge that is deemed useful rather than concentrating on the algorithms and techniques [9]. It suggests that DM needs to make better use of knowledge to achieve a better outcome than it does now. It first determines what issues are analyzed, what knowledge is desired, then to associate with the pattern and trend discovery, etc.

Method-driven DM is implemented on a bottom-up design approach. It crunches on data firstly, and uses mining algorithms to discover patterns, then try to make sense out of them. Knowledge-driven DM is relying on the knowledge from experts in relevant fields during modeling and mining. One common method is ripple-down rules (RDR) [10]. Over time the classification of RDR and other experts converge where there will be a high layer of agreement between them. The gap between method-driven and knowledge-driven data mining brings a great challenge in BI applications.

## III. LITERATURE REVIEW

The gap between the fields of business and computing techniques poses an obstacle in business intelligence application. Insights into this challenge that users and developers face in enterprise wide business intelligence system are described in [14]. It identifies five potential influences that concerns with user empowerment, training, data interpretation, supporting for usage and negotiating authorship. It summaries each interwoven has its own advantages and disadvantages on either end user's or IS developer's views. Business metadata is proposed in paper [15]. It improves data interpretation by explaining the relevance and context of the data. This study targets the data in existent enterprise data warehouse (DWH). The metadata is derived from the use of model weaving [16]. However, some business intelligence comes from the data warehouse via data mining algorithm. DWH only provides an organized place for data storage. By DWH alone, it is impossible to generate BI. Paper [3] presents an ontology-based architecture to transforms data to comprehensible model in business model by using semantic middleware integration. It implements directly on data transaction of an individual system, and provides a view consumable by business users. But its four-layer architecture only involves query-and-mapping process between data and end users, which cannot be seen as real KDP. KDP concerns the entire knowledge abstract process from data to knowledge.

## IV. BI-DRIVEN DATA MINING

We proposed BIdDM as a flexible DM methodology, which can combine and re-use components from existent DM methodology to guide the creation of BI. It relieves human assistance in guiding the data mining process. Since an e-Service layer is incorporated in between knowledge and data mining method, knowledge-driven and method-driven data mining can be combined as an integrated process.

### A. *Overall Process*

BIdDM contains two processes: framework construction and data mining (Figure 3). Framework construction is the process to establish a four-layer framework, which is a top-down approach from BI to data. Data mining process is a bottom-up approach discovering the potential knowledge from existent data sources. But in here, once the user chooses a BI as prospective outcome, a set of appropriate data mining techniques and the process are selected automatically corresponding to the BI, even for novice data mining users.

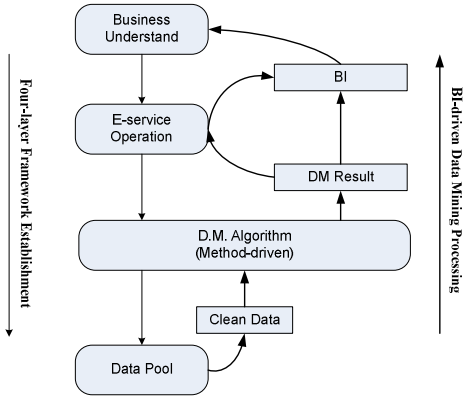


Figure 3. Overall Process Flowchart

### B. *Framework Construction Process*

Framework construction is the process of establishing the four-layer framework as shown in Figure 4. As a top-down process, the top level is to understand the business and which e-Service relates to the business, and what data source relates to the e-Service. Hence it proceeds to resembling a set of suitable data mining methods based on understanding. It loops until all BI relating to an e-Service have been found.

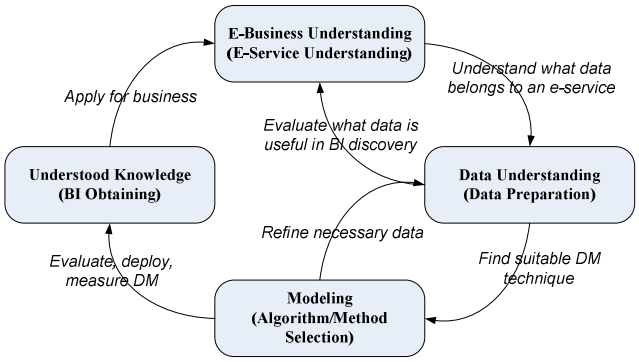


Figure 4. Infrastructure Construction Flowchart

### C. *Data Mining Process*

When the four-layer framework is established, one or several particular DM algorithms are tying to the selected BI. In this sense, the data mining methods are reusable as they can be tied to more than one BI. The data mining outcomes are labeled as BI Combo, which assembles to generate new business knowledge. The process is shown in Figure 5.

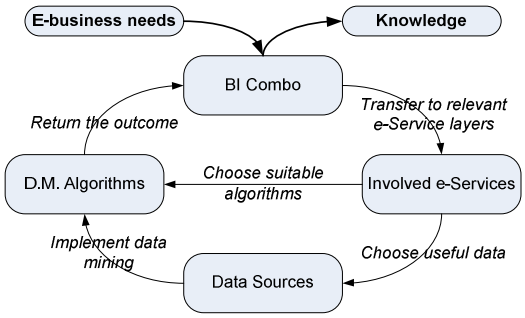


Figure 5. BI-driven Data Mining Process Flowchart

## V. FOUR-LAYER FRAMEWORK

BIdDM is a methodology built upon the four-layer framework which is shown in Figure 6. Our framework is extended from the four-layer architecture in [3]. It is applied for business intelligence derived from data mining methods, but only as a query-and-mapping data process.

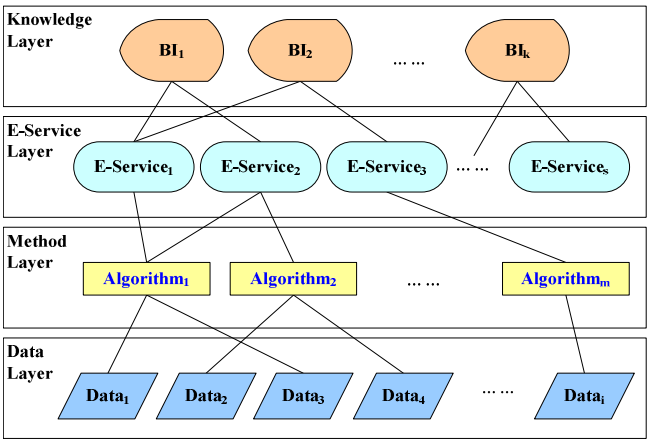


Figure 6. Four-layer BI-driven Data Mining

### A. *Knowledge Layer*

This layer is on top interfacing with end-users. It is referred as a knowledge or business decision-making support. BI combo reflects the outcome of knowledge discovery. Business managers can use BI findings to make strategic business decision. However, the end user does not have to know much about the technical details on the knowledge discovery system. Algorithms and data are selected as guided by the e-Service mash-up in BidDM. This layer harvests business intelligence from patterns that are derived from e-Business data.

This knowledge layer is adhering closely to the e-Business application that dictates the BI goals. It can be divided into several domains in e-Business, such as e-payment, e-logistic, e-procurement, e-shopping, e-community, or virtually anything that utilizes one or more e-Services. One BI may correspond to more than one e-Service. For example, if the expecting BI is about consumers' browsing and shopping behaviors, the records of customers' particulars, online requests and shopping behaviors are required from several e-Service components [12].

There are two types of e-Service systems in this layer. It is classified by the usage of BI. See from Figure 7, some e-Services, such as e-Service$_1$ and e-Service$_2$ are isolated to the others only belonging to BI$_1$; while e-Service$_2$ is a cross-BI e-Service belonging to BI$_2$ and BI$_3$. The similar relationships also appear in method layer and data layer.

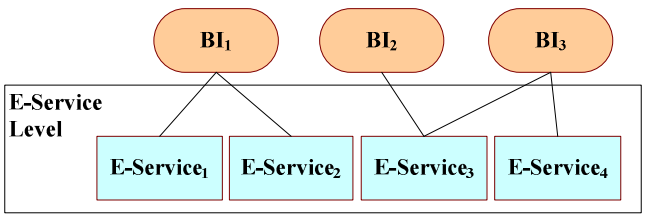


Figure 7. IES and CES

### B. Method Layer

This layer contains data mining algorithms, which is used to transform data into some meaningful expression. There are various algorithms that would be selected to use, one or several algorithms assembled to discover potential meaningful outcome. So far, a number of algorithms have been proposed as method-driven data mining. For example, mining the frequent patterns from web logs can discover the navigational behaviors of users [13].

Also, there are two types of data mining in this layer. It is classified by the usage of e-Service. Similar with the e-Service layer, some data mining algorithms are isolated to the others only belonging to one e-Service; while some are cross-service data mining belonging to different e-Services.

### C. Data Layer

This layer is the bottom one. It is responsible for providing data source of knowledge discovery. The data source has been preprocessed, which means transforming the raw data to cleaned data. The useful data come from raw data sources with the noise and inconsistencies filtered. Some BI may need cross-service data mining techniques. For this reason, e-Service mash-up system's data mining is usually distributed data mining. It is noted that distributed data mining implementation is more difficult than centralized data mining because of synchronous and asynchronous problems.

## VI. Definitions and Expressions

### A. Process Expression

The four-layer framework is implemented in a top-down approach. One certain knowledge in e-commerce is derived from a set of BI combo. Each BI is capitalized from some data mining methods (algorithms) used in e-Service. To implement the data mining algorithms, useful data is required as mining sources. This process is expressed in inference logic as in (1).

$$\begin{aligned} &KNOW(BI) \rightarrow KNOWLEDGE \\ &eSERVICE(METHOD) \rightarrow BI \\ &ALGORITHM(Data) \rightarrow METHOD \end{aligned} \tag{1}$$

### B. Component Constraint

The defined framework components has two constraints in Formula(2): ① every data mining algorithm in the project shall be applied in at least one e-Service to generate business intelligence;② and every data source in the project shall be applied in at least one data mining algorithm to generate business intelligence in e-Service.

$$\begin{aligned} &(\forall Method\ \exists EService\ BI(Method, EService)) \\ &\wedge (\forall Data\ \exists Method\ EService(Data, Method)) \end{aligned} \tag{2}$$

### C. Abbreviations

Here we explain the abbreviations used in four-layer framework components:

*Know(BIx): knows from BI x;*
*BI(ESx): BI comes from e-Service x ;*
*ES(Mx): e-Service relates to data mining algorithm x;*
*M(Dx): data mining algorithm relates to data set x;*
*ES: e-Service*
*CES: cross-BI e-Service;*
*IES: isolated e-Service for BI;*
*M: data mining algorithm*
*CM: cross-e-Service data mining algorithm;*
*IM: isolated data mining algorithm for e-Service;*
*D: data source;*
*CD: cross-algorithm data source;*
*ID: isolated data source for data mining algorithm*

### D. Construct Knowledge Layer

*Knowledge layer* is defined as logical expression in Formula(3). $Know(BI_x)$ is the intelligence known from $BI_x$. *BI* is the set of all business intelligence existing corresponding to a BI combo. If a set of BI is known to reflect one kind of business knowledge while there is no any other BI to reflect this knowledge, then the components in this BI set are business intelligence in knowledge layer.

$$\begin{aligned} &BI = \{BI_1, BI_2, ..., BI_n\} \\ &BI_i^n \in BI \wedge \forall \neg Know(BI_i^n) \notin Knowledge \\ &\rightarrow BI_i^n \in Knowledge \end{aligned} \tag{3}$$

BI outcome is derived from one or more e-Services. If the e-Service(s) is only corresponding to this BI, it is an IES. If the e-Service(s) belongs to one or more other BIs, it is a CES. One BI is derived from all of its IES and CES.

### E. Construct e-Service Layer

*e-Service layer* is defined as logical expression in Formula(4). $ES^n$ is a set of e-services. $CES^i$ is defined as a set of cross-BI e-service belonging to $BI_i$; while $IES^i$ is a set of isolated e-service belonging to $BI_i$. There is no intersection between IES and CES, while IES and CES are consisting of all e-Service components of this layer. Inheriting from Formula(3), the set $BI^n$ contains all business intelligence relating to one type of business knowledge. If every $BI_i^n$ belongs to $BI$, then all its $CES$ and $IES$ relate to the knowledge representation.

$$\left.\begin{array}{l} ES^n = \{ES_1, ES_2, ES_3, ..., ES_n\} \\ CES \cup IES = ES \\ CES \cap IES = \phi \\ \forall \neg BI(ES_i^n) \notin BI_i \wedge \exists ES_x^n \in BI_x \rightarrow \mathrm{ES}_\mathrm{x}^\mathrm{n} \in CES^i \\ \neg(\forall \neg BI(ES_i^n) \notin BI_i \wedge \exists ES_x^n \in BI_x) \rightarrow \mathrm{ES}_\mathrm{x}^\mathrm{n} \in IES^i \end{array}\right\}$$
$$\left.\begin{array}{l} \rightarrow BI(CES^i \cup IES^i) = BI_i \\ \forall BI_i \in BI_i^n \\ BI_i^n \in Knowledge \,//\, importFormula(3) \end{array}\right\} \quad (4)$$
$$\rightarrow \forall BI(CES^i \cup IES^i) \in Knowledge$$

### F. Construct Method Layer

*Method layer* is defined as logical expression in Formula(5). $M^n$ is a set of different data mining methods, such as clustering, classifying, etc. $CM$ is defined as a cross-e-Service method while $IM$ is an isolated method. There is no intersection between IM and CM, while IM and CM are consisting of all algorithm components of this layer. Inheriting the constraint of IES and CES from Formula(4), both of them inherit the characters belonging to ES. The set $ES^n$ contains all data mining algorithms relating to an e-Service. If every $ES^n$ belongs to a type of knowledge the corresponding the e-service reflects, then all the relevant algorithms of $ES^n$ lead to the knowledge representation.

$$\left.\begin{array}{l} M^n = \{M_1, M_2, M_3, ..., M_n\} \\ CM \cup IM = M \\ CM \cap IM = \phi \\ \forall \neg ES(M_i^n) \notin ES_i \wedge \exists M_x^n \in ES_x \rightarrow \mathrm{M}_\mathrm{x}^\mathrm{n} \in CM^i \\ \neg(\forall \neg ES(M_i^n) \notin ES_i \wedge \exists M_x^n \in ES_x) \rightarrow \mathrm{M}^\mathrm{n} \in IM^i \end{array}\right\}$$
$$\left.\begin{array}{l} \rightarrow ES(CM^i \cup IM^i) = ES_i \\ CES \cup IES = ES \,//\, importFormula(4) \\ CES \cap IES = \phi \,//\, importFormula(4) \end{array}\right\}$$
$$\left.\begin{array}{l} \rightarrow CES^i \cup IES^i = (CES(CM^i \cup IM^i)) \bigcup (IES(CM^i \cup IM^i)) \\ \forall ES_i \in ES^n \\ \forall BI(CES^i \cup IES^i) \in Knowledge \,//\, importFormula(4) \end{array}\right\}$$
$$\rightarrow \forall BI((CES(CM^i \cup IM^i)) \bigcup (IES(CM^i \cup IM^i))) \in Knowledge \quad (5)$$

### G. Construct Data Layer

*Data layer* is defined as logical expression in Formula(6). $D^n$ is a set of different clean data (D) in data sources. Isolated data (ID) and cross-method data (CD) exist in this layer. There is no intersection between ID and CD, while ID and CD are consisting as all data components of this layer. Inheriting the constraint of IM and CM from Formula(5), both of them inherit the characters belonging to M. The set $\mathrm{M}^n$ contains all data sets relating to a data mining algorithm. If every $M^n$ belongs to a type of knowledge the corresponding the algorithm reflects, then all the relevant data sets in $M^n$ lead to the knowledge representation.

$$\left.\begin{array}{l} D^n = \{D_1, D_2, D_3, ..., D_n\} \\ ID \cup CD = D \\ ID \cap CD = \phi \\ \forall \neg M(D_i^n) \notin M_i \wedge \exists D_x^n \in M_x \rightarrow \mathrm{D}_\mathrm{x}^\mathrm{n} \in CD^i \\ \neg(\forall \neg M(D_i^n) \notin M_i \wedge \exists D_x^n \in M_x) \rightarrow \mathrm{D}_\mathrm{x}^\mathrm{n} \in ID^i \end{array}\right\}$$
$$\left.\begin{array}{l} \rightarrow M(CD^i \cup ID^i) = M_i \\ CM \cup IM = M \,//\, importFormula(5) \\ CM \cap IM = \phi \,//\, importFormula(5) \end{array}\right\}$$
$$\left.\begin{array}{l} \rightarrow CM^i \cup IM^i = (CM(CD^i \cup ID^i) \cup IM(CD^i \cup ID^i)) \\ \forall M_i \in M^n \\ \forall BI((CES(CM^i \cup IM^i)) \bigcup (IES(CM^i \cup IM^i))) \in Knowledge \\ \,//\, importFormula(5) \end{array}\right\}$$
$$\rightarrow$$
$$\forall BI((CES((CM(CD^i \cup ID^i) \cup IM(CD^i \cup ID^i)))) \quad (6)$$
$$\bigcup (IES((CM(CD^i \cup ID^i) \cup IM(CD^i \cup ID^i))))) \in Knowledge$$

## VII. CASE STUDY

In this section, we give an example to demonstrate how to establish BIdDM framework in a B2C e-Shop.

### A. e-Service and Data Understanding

An example business model that we used to illustrate BIdDM is a B2C e-Shop. This e-Shop contains five primary e-Services in the whole business cycle as shown in Figure 8.

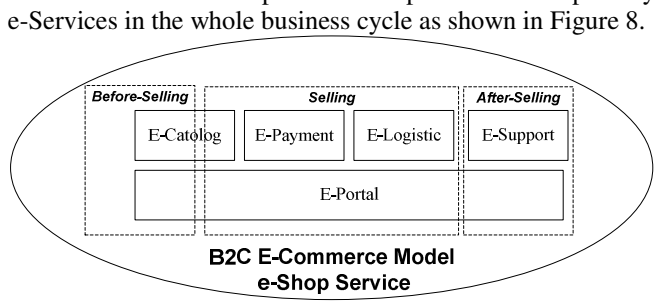


Figure 8. Sample e-Services of B2C e-Shop

*e-Portal:* a website that provides e-Commerce to customers. It will generate web log data on the web server. These data represent customers' navigating activities.

*e-Catalog:* a list of products, including products prices, introduction and other relevant information. By referencing

to this information, customers are able to decide which product to purchase.

*e-Payment:* an online transaction module to fulfill payment. The process usually relates to three entities: customers, e-Shop and payment gateway. In this B2C example, e-Shop's service only covers the transactions between customers and e-Shop.

*e-Logistic:* a fulfillment process of delivering goods to customer. This process is usually done by third-party logistic providers. Until the product is delivered to the customer, a selling process is not completed.

*e-Support:* a post-sales e-Service that provides relevant customer supports, such as sales records, complaints, goods-returned and servicing, etc.

### B. Methods Understanding

As examples in the case study, we choose three kinds of popular data mining algorithms: clustering, association rule and classifying.

*Partitioning around medoids (PAM) algorithm of clustering mining:* a representative object is chosen for every group. Once every medoids are chosen, other non-medoids will be thrown into a certain group according to the similarity, and the similarity is the Euclidean distance between any two objects [17]. It is a simple and efficient cluster algorithm to put the most similar data into the same group [18].

*FP-Tree of association rule mining:* it is a classic method to find association rules that satisfy the predefined minimum support and confidence from given data sources. FP-Tree is a frequent pattern mining. This algorithm can reduce the number of passes over the data source. It runs in two processes: constructing FT-Tree, and generate frequent patterns from FT-Tree [19, 20].

*Sequence mining algorithm of structure data mining:* Sequence mining is concerned with finding statistically relevant patterns between data examples where the values are delivered in a sequence. It is usually presumed that the values are discrete, and thus time series mining is closely related, but usually considered a different activity. It is a special case of structured data mining [21].

*Naive Bayes algorithm of statistical classification:* Bayesian classifiers assign the most likely class to a given example described by its feature vector. It has proven effective in much practical applications such as text classification, medical diagnosis etc [22].

### C. Prospective Business Intelligence

Through certain e-Services and data mining algorithms, some business intelligence result will be obtained from the e-Shop example. They are of importance in customer relationship management (CRM), providing potential insight for manager to make business strategies.

*Customer navigational behavior:* web log is the data source stored on e-Shop's WWW-server. It records all interactions between e-Portal and customers. Using methods of web usage mining, the most frequent visited pages of e-Portal are gained by FP-Tree association rule and Sequential mining algorithm from web logs [13].

*Customer online shopping habit:* e-Payment service provides some alternative choices for customers to pay the bills. Also, they are required to choose a logistic method for delivery. Using the Naive Bayes algorithm, online shoppers group themselves by their payment information and logistics preferences. By different online shopping habits, promotion packages can be prepared for distinct customer groups.

*Consumer's potential bought products*: e-Catalog contains the product's relevant introduction. According to the characters of the similar group's information, divide this data into two groups: the similar and dissimilar; then, use the method proposed in [18] will predict potential products that the current customer may probably buy.

### D. Four-layer Framework Construction

A four-layer framework of BI-driven Data Mining is constructed (in Figure 9) for e-Service of the B2C e-Shop business model. This is a top-down approach, which starts from choosing BI in knowledge layer, defines the relevant e-Service, and then identifies suitable data mining algorithms and useful data sources. Each line represents a dependent relationship between components across two layers.

In *knowledge-layer*, these three components compose one BI combo that will bring knowledge of customers shopping behavior to end-user to improve CRM strategies.

In *e-Service layer*, the isolated-BI e-services (IES) are e-Portal, e-Payment, e-Logistic and e-Support, while e-Catalog is cross-BI e-Service (CES) in this example. e-Portal is only relating to BI of customers' navigational patterns. So e-Portal is IES. e-Payment and e-Logistic are only relating to BI of e-Shopping habits, so they are IES too. e-Support is only relating to BI of potentially purchase products so it is also IES. e-Catalog is enabling BI of both online shopping habit and potentially purchase products, hence it is CES.

In *method layer*, the isolated-eService data mining algorithms (IM) are FP-Tree, sequence mining, while Naïve Bayes and PAM are cross-eService methods (CM). IM only relates to one e-Service, while CM relates to more than one e-Service in the four-layer framework architecture.

In *data layer*, the isolated-method data sources (ID) are payment options, delivery methods and sales records, while web log files and products information are cross-method data sources (CD). ID only relates to one data mining algorithm, while CD relates to more than one data mining.

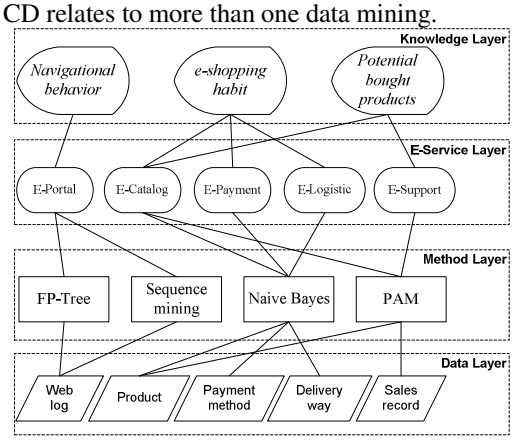


Figure 9. Four-Layer Framework of Sample e-Shop

## VIII. Conclusion and Future Work

BIdDM is proposed in this paper that is useful for e-Business to obtain business intelligence through some guided data mining methods by identifying the related e-Services. All the elements required under BIdDM for constructing a four-layer framework are shown in Figure 10.

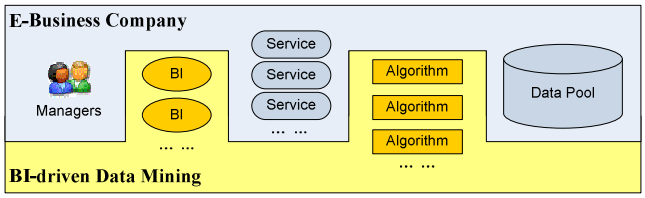


Figure 10. Elements under BIdDM

BIdDM represents a new data mining methodology combining knowledge-driven data mining and method-driven data mining. When the framework is established, business intelligence discovery will bring potential meaning to end-user in guided fashion because data mining methods are predefined and pre-identified to each e-Service which in turn contributes to certain types of BI. A B2C e-Shop case study is given for demonstrating how such a framework can be constructed. For future work, we would test BIdDM and construct more detailed examples from many other e-Business scenarios such as government services and B2B models.

## References


[1] H. P. Luhn. "A Business Intelligence System" (PDF). IBM Journal,October 1958.

[2] Simoudis, E., Reality check for data mining, IEEE Expert: Intelligent Systems and Their Applications , Volume 11 Issue 5, pp 26-33, Oct. 1996.

[3] Spahn, M., Kleb, J., Grimm, S., and Scheidl, S. Supporting business intelligence by providing ontology-based end-user information self-service. In *Proceedings of the First international Workshop on ontology-Supported Business intelligence* (OBI '08), vol. 308. ACM, New York, NY, pp. 1-12, 2008.

[4] K. J. Cios, W. Pedrycz, and R. M. Swiniarski. Data Mining: A Knowledge Discovery Approach., Chapter 2, pp. 9-24, Springer Press, 2007.

[5] Zimmermann, H., OSI Reference Model--The ISO Model of Architecture for Open Systems Interconnection, IEEE Transactions on Communications, Volume 28, Issue 4, pp 425- 432, 1980

[6] Fu Y., Distributed Data Mining: An overview. Newsletter of the IEEE Technical Committee on Distributed Processing, pp 5-9, Spring, 2001.

[7] Mu-Chun Su, M-C; and Chou, C-H. A Modified Version of the K-Means Algorithm with a Distance Based on Cluster Symmetry. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23, pp 674-680, 2001.

[8] Breiman, L. Random Forests. Machine Learning Journal,45, pp 5-32, 2001.

[9] Graco, W., Semenova, T., and Dubossarsky, E. Toward knowledge-driven data mining. In *Proceedings of the 2007 international Workshop on Domain Driven Data Mining*. DDDM '07. ACM, New York, NY, pp 49-54. 2007.

[10] S. Chakrabarti, "Data mining for hypertext: A tutorial survey." SIGKDD: SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery and Data Mining, ACM, Vol. 1, No. 2, pp. 1-11, 2000

[11] Tibor S., Iveta Z., Lenka L., "Distributed data mining and data warehouse", XXX.ASR 2005 Seminar, Instruments and Control,pp 417-420, April.2005

[12] Tian Chen; Zhi-geng Pan; Jian-ming Zheng, "EasyMall - An Interactive Virtual Shopping System," *Fuzzy Systems and Knowledge Discovery, 2008. FSKD '08. Fifth International Conference on* , vol.4, no., pp.669-673, 18-20 Oct. 2008

[13] Renata Ivancsy, Istvan Vajk, Frequent Pattern Mining in Web Log, 2006 Acta Polytechnica Hungarica, Vol 3, No.1, pp 77-90, 2006.

[14] B. Pinnington, B. L. B, and E. Ferneley. Too Much of a Good Thing? A Field Study of Challenges in Business Intelligence Enabled Enterprise System Environments. In Proceedings of the 15th European Conference on Information Systems, pages 1941--1952, 2007.

[15] V. Stefanov and B. List. Explaining Data Warehouse Data to Business Users - A Model-Based Approach to Business Metadata. In Proceedings of the 15th European Conference on Information Systems, pages 2062--2073, 2007.

[16] Breton E. and Bézivin J. Weaving Definition and Execution Aspects of Process Meta-models. In 35th Hawaii Int. Conf. on System Sciences, pp 290, 2002

[17] L. Kaufman and P. J. Rousseeuw, Finding Groups in Data: an Introduction to Cluster Analysis, John Wiley & Sons, 1990.

[18] Qingzhang Chen; Jianghong Han; Yuqing Chu; Xiaodong Ying, "Mining Consumers' Most Adaptive Products by Efficient Clustering Algorithm," *Artificial Reality and Telexistence--Workshops, 2006. ICAT '06. 16th International Conference on* , vol., no., pp.195-199, Nov. 29 2006-Dec. 1 2006.

[19] S. Kotsiantis, D. Kanellopoulos, Association Rules Mining: A Recent Overview, GESTS International Transactions on Computer Science and Engineering, Vol.32, Issue 1, pp. 71–82, 2006.

[20] Han, J. and Pei, J. Mining frequent patterns by pattern-growth: methodology and implications. ACM SIGKDD Explorations Newsletter 2, 2, pp.14-20, 2000.

[21] Pei, J., Han, J., Mortazavi-Asl, B., and Zhu, H. Mining Access Patterns Efficiently from Web Logs. In Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Current Issues and New Applications (April 18 - 20, 2000). T. Terano, H. Liu, and A. L. Chen, Eds. Lecture Notes In Computer Science, vol. 1805. Springer-Verlag, London, pp.396-407,2000.

[22] Domingos, P. and Pazzani, M. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. Mach. Learn. 29, 2-3, pp.103-130, 1997.