2015 IEEE International Conference on Data Science and Data Intensive Systems

From Data to Knowledge: a Cognitive Approach to Retail Business Intelligence

Atsushi Sato and Runhe Huang
Faculty of Computer and Information Sciences
Hosei University
Tokyo, Japan
Email: atsushi.sato.3r@stu.hosei.ac.jp, rhuang@hosei.ac.jp

***Abstract*— Consumer-oriented companies can no longer afford to make decisions or measure results based on gut feeling. They must be able to take advantage of all available data. Advanced analytics makes it possible to capture value and benefit from big data, however, this isn't a given. Companies must hire, develop, and retain skilled analysts, who can distinguish relevant from irrelevant data, draw the right assumptions, and translate information into insights. To lighten the burden on companies and support big data analytics, this paper presents a KID (Data-Information-Knowledge) model based on a cognitive approach which can accumulate experience and gain knowledge by continuously perceiving data, interpreting data into meaningful information, absorbing incoming information, and updating knowledge as humans do. This is a process of from data to knowledge and knowledge about correlations among attributes, making assumptions, and testing the assumptions with appropriate algorithms which are constantly updated and summarized in this data-information-knowledge cyclic process. This approach is applied to a retail business for understanding customer purchasing and product sale situations, so as to support provision of better service and timely adaptation of business strategy.**

Keywords- Retail business intelligence, cognitive analyst, KID model, big data, customer modeling, Business modeling

I. INTRODUCTION

Smart business, by definition, indicates the ability to achieve goals which are set according to the development tendency of business [1]. The key to successful implementation of the vision of smart business relies on a comprehensive understanding of the surrounded scenario in which a wide spectrum of elements are concerned. Instances simply include the vision of the company concerned, the global economic situation, current trends, the target market and consumers, etc. It is not difficult to come up with thousands of similar elements for consideration [2] [3]. However, an understanding of customers and products for consumer-oriented companies is the important element of data-driven insight [4] [5].

It is believed that big data and advanced analytics can deliver more useful insights than traditional tools. However, this isn't a given. [6] Companies must capture and manage mountains of data over several years. What is more, they must hire, develop, and retain skilled analysts who can distinguish relevant from irrelevant data, draw the right assumptions, and know what the appropriate tools or algorithms to use for translating information into insights are. The former takes time and the latter increases cost. Moreover, individuals with analytical talent as well as acute business acumen are in high demand and short supply.

To lighten the burden on companies and support big data analytics, this paper presents a KID model [7] based cognitive approach. Instead of big data, it continuously perceives incoming data piece-by-piece; interprets them into meaningful information; absorbs information into existing knowledge; and updates knowledge just as humans do. Prior knowledge about customers and products, and expert knowledge and skills in retail marketing can be pre-embedded into a knowledge repository in the KID model. Based on this prior knowledge, experience and new knowledge are continuously accumulated, summarized, and evaluated naturally in the data-information-knowledge cyclic process. To some extent, this knowledge is sufficient to turn data-driven insights into effective action on the front line. Meanwhile, it can support big data analytics in distinguishing relevant from irrelevant data, drawing the right assumptions, and translating information into insights.

This paper is organized as follows. Problems for this research to solve is raised in the next section. The KID model is briefly described in Sections III and IV. Section V explains the KID model based cognitive system. Pragmatic implementations for a retail business application are given in Section VI. Finally, conclusions are drawn and remarks are given in the last section.

II. THE PROBLEM DESCRIPTION

Retailers and consumer-packaged-goods producers normally capture information about every stock keeping unit sold to every customer at every store, every day. They have long had access to vast amounts of transaction data. However, it has always been a challenge to transform the available data into useful information and derive specific and timely knowledge about customers, products and markets which in turn can help boost profits, reduce costs and support better and more effective management. Conventionally, these businesses regularly use sophisticated marketing techniques to capture opportunities signaled in data. In

978-1-5090-0214-6/15 $31.00 © 2015 IEEE
DOI 10.1109/DSDIS.2015.106

recent years, adopting big data has become a trend. Can big data and advanced analytics truly deliver more useful insights than existing tools? The answer is yes, but this is costly in terms of the time it takes to collect data, requires skilled analysts and dynamic changing customer purchasing behavior.

Those who have big data experience must know it is not easy to make correct assumptions and find appropriate algorithms to realize the potential of this data unless one is a very skilled analyst or data scientist and has made a lot of trial and error experimentation. That is to say, setting right assumptions and deciding appropriate analytics are time consuming and requires manpower and skills which are beyond most ordinary people's capabilities. Moreover, hiring, developing, and retaining skilled analysts for making right assumptions and selecting appropriate analytic algorithms may be too costly for small-scale retail companies.

Instead of skilled analysts required for big data, it is suggested a cognitive model based system be used [8] [9]. By continuously processing incoming small sets of data rather than big data, a cognitive model based system can accumulate and aggregate experience and knowledge in each transformation process (from data to knowledge) cycle. As time goes by, when accumulated knowledge is sufficient to form useful insights, it is possible for a cognitive model based system to make reasonable assumptions and select appropriate algorithms, to some extent. In this paper, a KID model based cognitive approach is proposed to support the process of from data to knowledge and get insight from available data.

To counter the shortage of skilled analysts in big data, the KID cognitive model is an alternative or supplemental solution.

III. Why the KID Model?

Since the invention of programmable computers in the 1950s, AI researchers have increasingly written computer programs to simulate human mental processes. From the material view of the relationship between mind and body, the mind is like the actions of a machine [10]. Cognitive psychology focuses on the way human's process information, looking at how we process information that we receive and how this processing leads to behaviors. When looking at information-transmitting machines, cognitive scientists notice that human cognition has many of the properties of such machines [10]. For instance, humans take in information from the outside world, transform the information into neurological signals, and later base their behavior on stored information in certain associated structures. It can be claimed that any cognitive process consists of the transmission of information through a series of stages in which information is transformed [10].

The KID model is a cognitive model, where *K* refers to knowledge, *I* refers to information, and *D* refers to data since it is also a cognition process from data to knowledge. It adopts the results of psychologists' investigations, simulates human information processing and is built based on our argument that any cognitive model can be built with three transformation processes from data to knowledge, i.e., translating data to information by an interpretation function; absorbing information to existing knowledge by an assimilation function; and applying knowledge for effecting on the external by an instantiation function.

To sum up, processing retail business data and gaining useful insights from available data are human-like information processing. Piece by piece processing is more natural and closer to a human cognitive process. Therefore, the KID model is appropriate for the transformation process for data, to information, and to knowledge for retail business data. Besides this, the KID based cognitive approach is expected to have potential wide applications.

IV. Overview of the KID Model

As Fig. 1 shows, the KID model consists of three elements, *D*, *I*, *K*, and a knowledge repository, *K-store*. It provides three abstract functions, *interpretation()*, *assimilation()*, and *instantiation()* with some primitive functions, *stimulate()*, *map()*, *identify()*, *associate()*, *update()*, *copy()*, and *apply()*. Where, *D* refers to data which are symbols that represent observable properties of objects in the external world. *I* refers to information which is the result of data being interpreted by existing knowledge, that is to say, information is data endowed with meaning interpreted by knowledge. *K* refers to knowledge which is formed by assimilating information into existing knowledge or derived from updated knowledge. *D*, *I*, *K* are interrelated, their inter-relationships are defined by the three transformation functions.

A. The interpretation() function

If incoming data, *D* can be interpreted with meaning, it means that *K-store* can retrieve knowledge *K* to translate the data and transform it to information *I* which is meaningful to the existing knowledge. This transformation process can be expressed as

$$D \times K \rightarrow I. \tag{1}$$

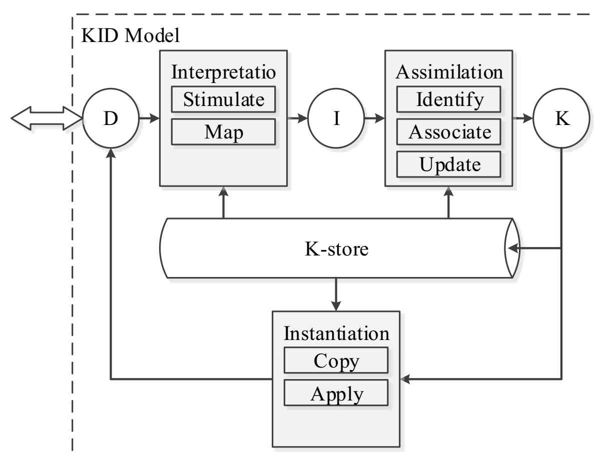

Fig. 1. The overview of the KID model.

The primitive functions, *stimulate()* and *map()* are pragmatic functions for implementing the transformation from data to information. In *stimulate()* function, data D is regarded as stimuli to activate an item of knowledge in the knowledge repository, *K-store* for translating the data. Therefore, its input is data D, and its output is K. While *map()* is to translate the original data to a meaningful representation, I which is specified in and understandable to this context or knowledge domain. We have,

$$stimulate(): D \rightarrow K,$$

$$map(): D \times K \rightarrow I.$$

B. The assimilation() function

The function, *assimilation()* is to link newly generated information in the interpretation process to relevant knowledge in *K-store*, whereby the existing knowledge is updated or may be enriched if new knowledge is derived due to the newly assimilated information. It can be expressed as

$$I \times K \rightarrow K. \tag{2}$$

The primitive functions, *identify()*, associate(), and update() are pragmatic functions to enable the transformation from information to knowledge. The function, *identify()* is to traverse *K-store* and identify those elements which are associated with the newly generated information, I and place the information in its appropriate location. The function, *associate()* is to build association links between the identified elements in *K-store* with the newly generated information, I. The function, *update()* is to updates existing knowledge after absorbing the newly incoming information by adding or deleting association links and modifying their association strength (weight values) of the information to the identified elements. We have,

$$identify(): I \times K \rightarrow K,$$

$$associate(): K \times K \rightarrow D,$$

$$update(): K \rightarrow K.$$

C. The instantiation() function

The function, *instantiation()* is to instantiate knowledge and apply it with knowledge about the current internal context or external request. Its effects can be on the internal system or on the external world. This function can be expressed as

$$K \times K \rightarrow D. \tag{3}$$

The pragmatic functions for implementing this function includes *copy()* and *apply()*. *Copy()* is to retrieve the relevant knowledge and get the copy of it. *Apply()* is to generate knowledge instance(s) for the current internal context or for an external request. We have

$$copy(): K \times K \rightarrow K,$$

$$apply(): K \times K \rightarrow D.$$

D. K-store

This is a knowledge repository in which items of knowledge are represented by certain representations and structures which can support the interpretation, assimilation, and instantiation processes. It is a cognitive model of the memory system. As depicted in Fig. 2, it includes two basic components, one (working memory) for current information processing, and another (long-term memory) for storing all the information acquired in the past and which might needed again in the future [11]. The memory system would be connected to other information processing systems, those for making sense of incoming information and planning behavior. Some are internal processes like the interpretation process, assimilation process, and instantiation process. Some may external processes, such as a customer modeling process or a retail business process. *K-store* is the core component for the KID model. How efficiently and effectively the KID model can process incoming data, turn them into knowledge and augment knowledge, depends on many factors. However, prior knowledge, knowledge representation and structure, and an inference mechanism are all critical factors.

V. The KID Model based Cognitive System

An array of cognitive models have been appearing in all fields of cognition at a rapidly increasing rate, ranging from perception to memory to problem solving and decision-making [10] [12] [13] [14]. A KID model based cognitive system is one of them. Having the KID model as its core, it includes some other modules like *D-store*, *I-Store*, customer model, retail business model, and retail business data in the external world as shown in Fig. 3. *D-store* is a warehouse which continuously stores all data that has passed through the element, D. As time goes by, it becomes retail business big data which can be shared by external entities. Similarly, *I-store* is a place for holding all information in a triple [original data: D, knowledge used for interpretation: K, resulting information: I]. Collected information in the triple could be as valuable as one's collected experience. The customer model module keeps each

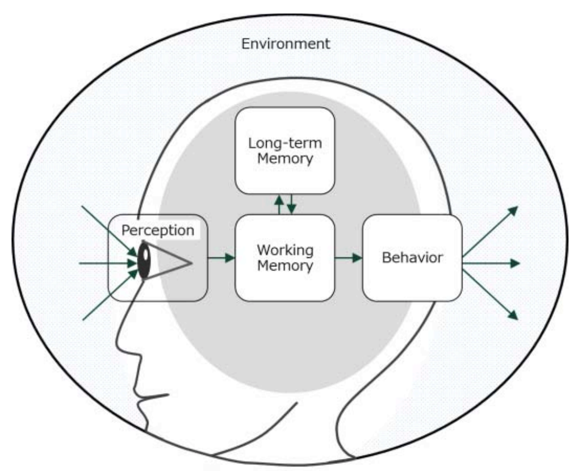

Fig. 2. A cognitive model of the memory system.

customer's digital description [15]. With more and more knowledge about customers contributed by the KID model, each customer model may approximate its real customer referring to their purchasing behavior aspect. The retail business model is a digital description of a retail business which grows with continuous contribution of knowledge about the business from the KID model. The retail business data is assumed to be the real-time received data from the front line. However, at the current stage, it is difficult. Instead, we use retail data collected piece by piece rather than the whole set of data to simulate the data from the front line.

In fact, our KID model based cognitive system can be used as a generic framework for data-driven and human-centric smart service systems [16] if we replace the customer model, retail business model, and retail business data with general-purpose module names such as human model, X-application model, and X-data. This is expected to have a variety of potential applications in the fields of smart service, business intelligence, smart city, and smart world. However, this paper is focused on its application to retail business intelligence and aimed at understanding customer purchasing and product sale situations from continuously perceiving and processing retail business data from the outside world. The results of processing in the KID model are experience and knowledge about customers, products, and retail marketing insights which are accumulated and aggregated in *K-store* as shown in Fig. 3. This experience and knowledge can be exploited in at least two ways. One is to apply knowledge to the front line of the retail business to provide better service to customers and to adjust retail strategy. Another is to apply knowledge to big data analytics with respect to distinguishing relevant from irrelevant data, drawing the right assumptions, and helping to select appropriate algorithms.

It is worthwhile to point out that the KID model is also a unified structure and data-driven process. It unifies data, information, and knowledge and gains data-driven knowledge. It is an abstract model and its formulation was given and explained

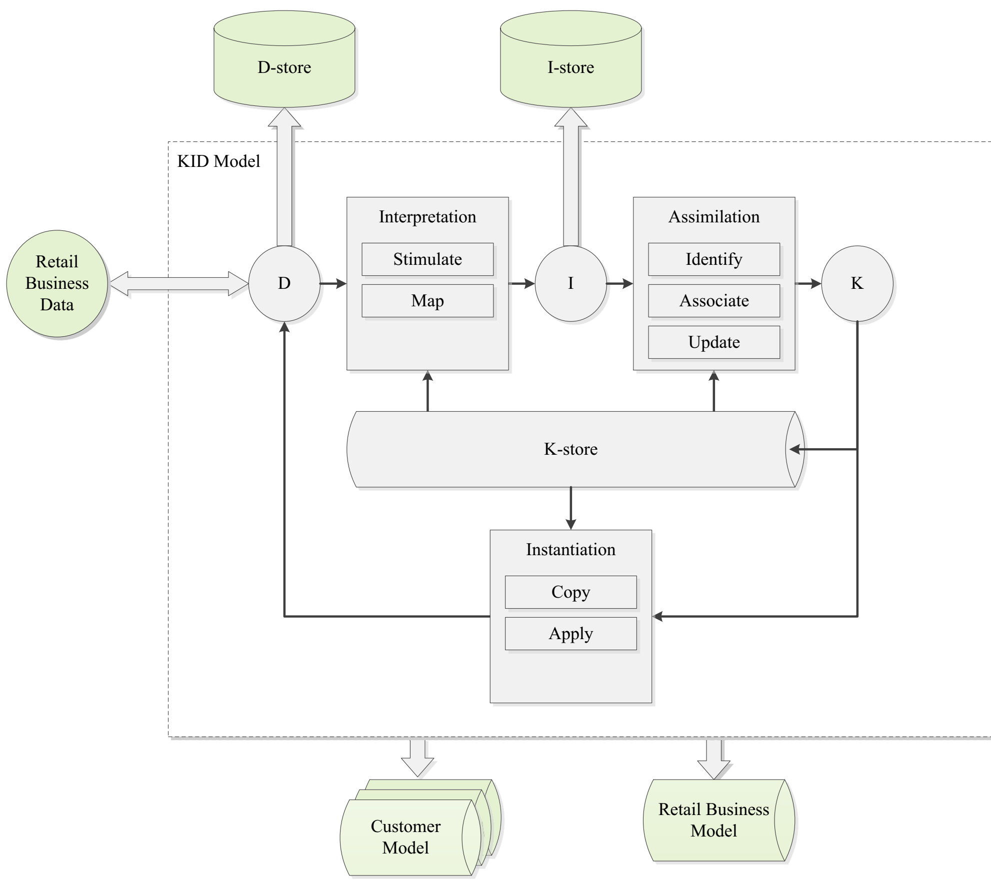

Fig. 3. A KID model encapsulated cognitive system.

in our previous research [7]. The following section is focused on describing how it is applied to retail business intelligence with pragmatic implementations of three abstract functions in the KID model.

VI. Pragmatic Implementation for a Retail Business Application

The KID model based cognitive approach for a retail business application is roughly divided into two parts: the external world and the internal KID model. This implementation is mainly focused on the internal KID model to demonstrate how it processes incoming retail data and generates useful insights from the data.

A. *Retail business data*

Although there are many different data which are related to a retail business company, the available data we have are two datasets. One is the customer profile dataset which has 22 attributes and belonging to which the number of registered customers is over 250,000. Another is the product history record dataset which has 38 attributes. The number of records is about 450,000 over one year, given in Fig. 4.

B. *Implementation of a KID model based system*

In the KID model based system, four classes, *World*, *Agent*, *KIDModel*, and *DataProvider* are defined. Where, *KIDModel* and *DataProvider* are two sub classes of *Agent*. In a "world" (instance of World class), there are three agents, "kid_model" (instance of *KIDModel* class), "customer_data_provider" (instance of *DataProvider*), and "product_data_provider" (instance of *DataProvider*). They are added to the "world" as objects and they interact in the world. The class diagram is given in Fig. 5.

Fig. 4. Retail data at a glance.

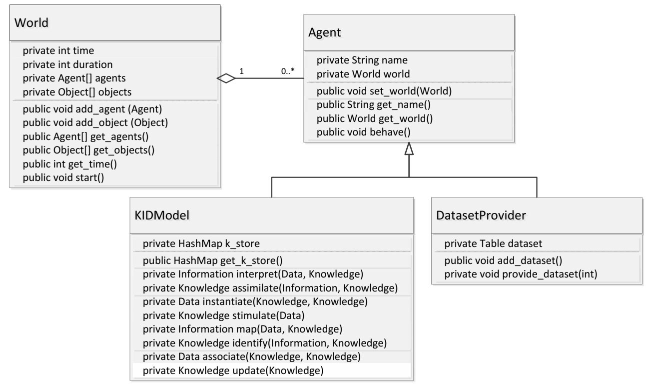

Fig. 5. The class diagram of the KID model system.

The system implementation diagram is shown in Fig. 6. *Initialize()* method generates three agents. *Start()* method sets the starting time and the duration of the system runs. Since this is a simulated front line data, it does not use real time but uses the number of processing cycles instead of "time". In *start()*, three threads start at the same time, *provide_dataset()* is to provide customer dataset and product dataset at a random time, respectively. "Objects" is an object store, once receiving an incoming dataset, it wakes up the KID model agent to process the dataset. The "time" variable post increments by 1 after each processing cycle. In the course of implementing the three functions, *interpretation()*, *assimilation()*, and *instantiation()* in the KID model, *K-store* is often interacted with for retrieving and updating knowledge.

This system is implemented in *R* language. Below, let us go through some cases with these two datasets to see how interpretation, assimilation, and instantiation functions in the KID

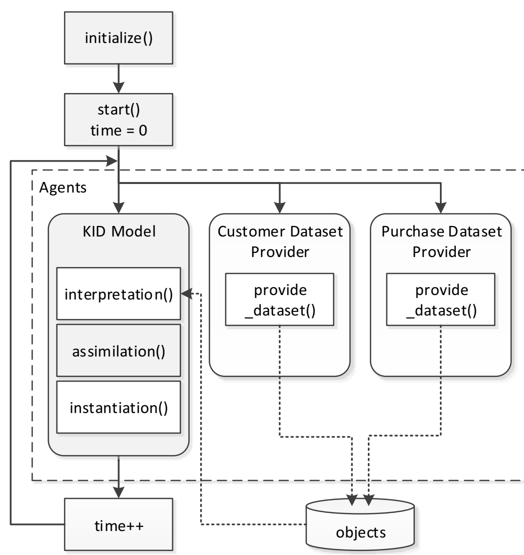

Fig. 6. The KID model system diagram.

model interact and work together with each other and see if we can obtain some useful insights with the two available datasets.

C. Data from the customer registration or purchasing front line

We assume that prior knowledge of customer profile data frame and product record data frame are already in *K-store*. Each data frame has knowledge about its attributes. Once a data or a dataset gets into the KID model, it stimulates a (or a set of) relevant knowledge in *K-store*. As Fig. 7 shows, *stimulate()* function takes data d_1 as input and returns interpreters which are stimulated with a nonzero value of intensity. *Map()* function takes data d_1 and the top one interpreter k_1 as inputs and maps it to d_2. In this implementation, d_1 is two instances of the customer's data as a customer dataset, k_1 is the knowledge about the customer data which is represented in a function, *$k.customer_dataset function(x)*. Finally, *interpret()* returns a triple [d_1, k_1, d_2] as the result, i.e. information of the interpretation process. Fig. 7 presents a practical case of the interpretation process and part of the implementation details.

D. How information i_1 is assimilated into K-store

We assume there has been some prior knowledge put in *K-store* and they are organized in a certain structure. Design of an efficient and effective human–like knowledge repository system for retail business is another challenging issue. At this stage of our research, we just use a simple nested list structure. A list structure in *R* returns a pair of reference and element. The reference can be an index which specifies the location of the element in the list. It can be also a specified key. For example,

```
mylist <- list (a = 1:5, b = "Hi There", c = function(x) x * sin(x))
mylist[2]
$b:
[1] "Hi There"
```

where, b in $b is the specified key for the element "Hi, There". A key can be used for identifying associated elements in *K-store* in the assimilation process for identifying a location where the information is to be associated and forming an association between the information and the identified elements in *K-store*. Finally, knowledge in *K-store* is updated accordingly after taking in the information.

As given in Fig. 8, information, i_1, as input of *identify()* function which finds knowledge k_2, as the identical element to i_1 in *K-store*. *Associate* () function takes d_1 contained in i_1 and k_2 output from *identify()* function, and returns the association of d_1 and d_1 and k_2. *Update()* function uses the result from *identify()* function to locate a place for d_1 to link with their association. In principle, d_1 and k_2 do not necessary to merge into the same physically same place, instead they are in different physical locations but connected with the association link. In this implementation, they are merged into a same place, i.e., in the customer data frame.

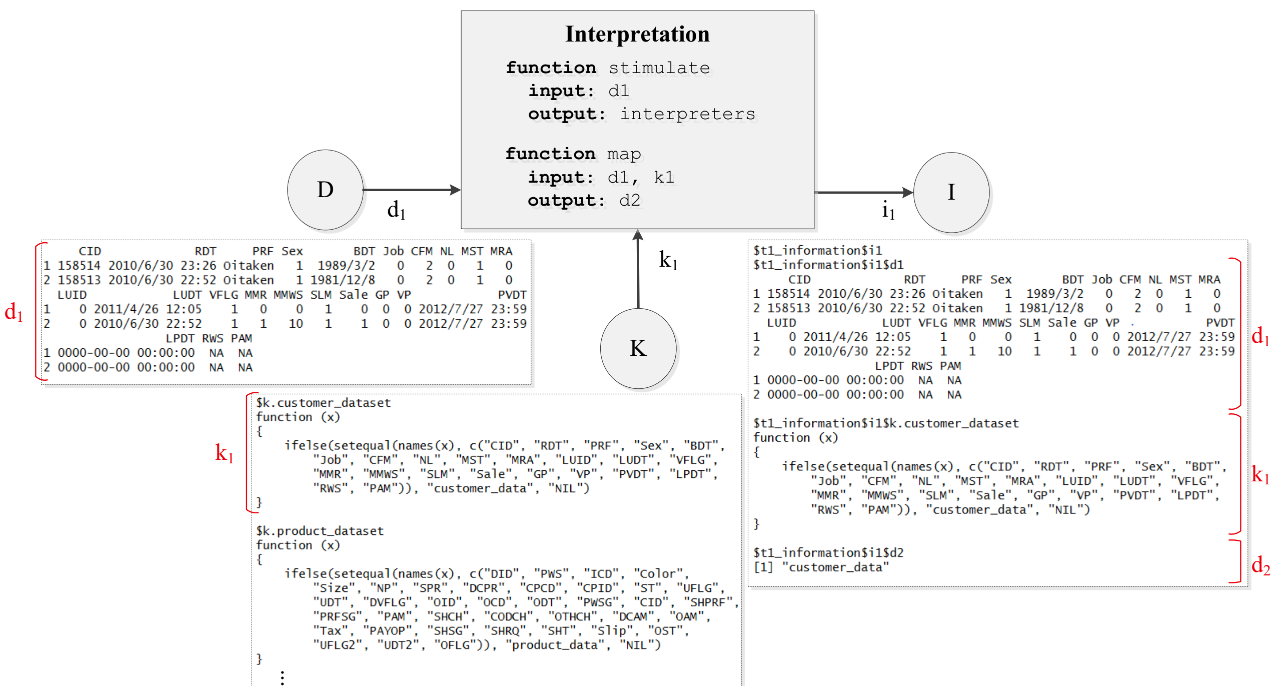

Fig. 7. The depiction of implementation of the interpretation process.

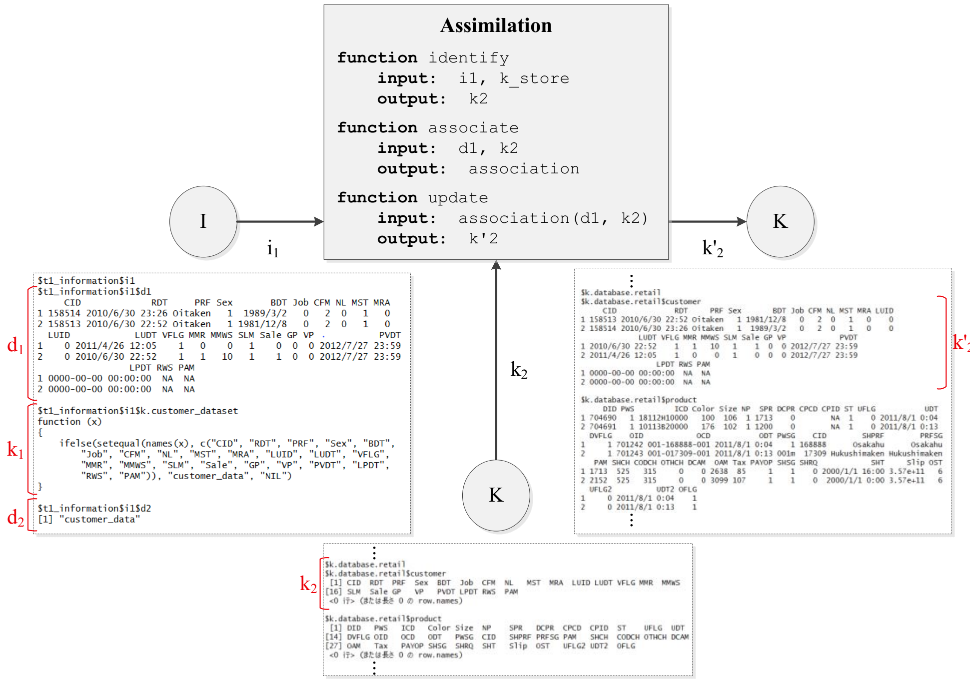

Fig. 8. The depiction of implementation of the assimilation process.

E. Enriching knowledge and find insights

In principle, it is possible to incorporate many existing *R* language methods, data mining algorithms, machine learning algorithms, and others in *K-store* as prior knowledge. With certain prior knowledge, knowledge in *K-store* may be updated with entailed new knowledge. The process is cyclically continued with incoming data from outside of the model, knowledge is enriched to be able to make a proposal or assumption. To these business datasets in this implementation, the correlation of a product's price and Tax is identified and also the two happy bag products are identified within the top 10 products. It is expected to find more valued insights from the incoming data and stream processing so as to make assumptions and proposals. However, the findings are limited due to the quality of datasets and insufficient prior knowledge of analysis algorithms. It may be improved with better quality of data in datasets and adding more data mining and analysis algorithms as prior knowledge. How prior knowledge related to the finding is another challenging problem for us to look at. This work is still in progress and will continue with more experimental results and obtaining more valid and better quality of data.

VII. Conclusion and Remarks

In this paper, the KID model is proposed as the core in a cognitive approach system. In order to demonstrate its feasibility and applicability, it is applied to a retail business intelligence application. This paper only demonstrates a part of the system's implementation and since some implementation work is still in progress. *R* language is a convenient programing language for processing data but it is not effective for knowledge representation. It is possible to use two implementation languages, for instance, *R* and *Java* which interface with each other, to improve system implementation.

The KID model is featured with a human-like cognitive process, simulating human information processing, an abstract model, a data-information-knowledge unified structure, and data-driven knowledge growth process cycle. With these features and its pragmatic implementation functions, the KID model based cognitive approach could have a variety of potential applications in the fields of smart things, intelligent systems, data mining and knowledge engineering, in particular, the future Internet of smart things [17]. We believe it is a promising but challenging approach. What we have done is the first step in a long march. We hope our work can serve as a trigger and draw more attention from researchers to get them involved toward its ultimate goal: formulating the KID model in mathematics and information algebra and proposing a unified theory of pragmatic mechanisms for a wide range of applications.

Our current work is just our research in progress and much work remains for the future. Of which three topics we would like to focus on for our future work are

(1) studying efficient and effective representation and structure of *K-store* for business intelligence applications.

(2) evaluating quality of information and knowledge (QoI and QoK) in the data transformation process with regards to achieving a goal.

(3) developing fusion mechanisms of data, information, and knowledge for efficiently and effectively deriving useful insights in the KID model.

ACKNOWLEDGMENT

The work is partially supported by the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (No. 25330270 and No. 26330350).

REFERENCES

[1] H. J. Watson, and B. H. Wixom, "The Current State of Business Intelligence," IEEE Computer, vol. 40, no. 9, pp. 96-99, 2007.

[2] W. F. Cody, J. T. Kreulen, V. Krishna, and W. S. Spangler, "The integration of business intelligence and knowledge management," IBM Systems Journal, vol. 41, no. 4, pp. 697-713, 2002.

[3] N. Sun, J. G. Morris, J. Xu, X. Zhu, and M. Xie, "iCARE: A framework for big data-based banking customer analytics," IBM Journal of Research and Development, vol. 58, no. 5/6, pp. 4:1-4:9, 2014.

[4] F. T. John, K. Peter, G. Carlos, and B. Ben. *The State of Data-Driven Decision-making in Retailing & CPG* [Online]. Available: http://www.baylor.edu/business/collaboratory/doc.php/230970.pdf [accessed: 2015/08/20]

[5] S. Ogden-Barnes. *Decision-Making in Retail Marketing: The End of Gut Feel?* [Online]. Available: http://www.pitneybowes.com.au/software/Decision-Making_in_Retail_Marketing_Thought_Leadership_Paper_Pitney_Bowes_Software_Australia.pdf

[6] P. Breuer, L. Forina, and J. Moulton, "Beyond the hype: Capturing value from big data and advanced analytics," pp. 4-9, in the Book: Perspectives on retail and consumer goods, no. 1, Spring 2013, McKinsey&Companies.

[7] A. Sato, and R. Huang, "A Generic Formulated KID Model for Pragmatic Processing of Data, Information and Knowledge," The 12th IEEE International Conference on Advanced and Trusted Computing (ATC2015), pp. 609-616, Beijing, China, August 10-14, 2015, IEEE publisher.

[8] J. R. Busemeyer, and A. Diederich, *Cognitive Modeling*. SAGE Publications, 2010.

[9] D. S. Modha, R. Ananthanarayanan, S. K. Esser, A. Ndirango, A. J. Sherbondy, and R. Singh, "Cognitive Computing," *Commun. ACM*, vol. 54, no. 8, pp. 62–71, 2011.

[10] R. K. Guenther, *Human Cognition*. Prentice Hall, ISBN-13: 978-0131402942, 1997.

[11] A. Sammons. *The cognitive approach: the basics* [Online]. Available: http://www.psychlotron.org.uk/newresources/approaches/as_aqb_approaches_cognitivebasics.pdf

[12] H. Reynolds, and S. Feldman. (2014, Jun. 27), *Cognitive Computing: Beyond the Hype* [Online]. Available: http://www.kmworld.com/Articles/News/News-Analysis/Cognitive-computing-Beyond-the-hype-97685.aspx

[13] R. High (2012, Dec. 12). *The Era of Cognitive Systems: An Inside Look at IBM Watson and How it Works* [Online]. Available: http://www.redbooks.ibm.com/redpapers/pdfs/redp4955.pdf

[14] J. E. Kelly, and S. Hamm, *Smart Machines: IBM's Watson and the Era of Cognitive Computing*. Columbia University Press, 2013.

[15] J. Ma, J. Wen, R. Huang, and B. Huang, "Cyber-Individual Meets Brain Informatics," IEEE Intelligent Systems, vol. 26, no. 5, pp. 30-37, Sep. 2011.

[16] R. Huang, A. Sato, T. Tamura, J. Ma, and N. Y. Yen, "Towards next-generation business intelligence: an integrated framework based on DME and KID fusion engine," Multimedia Tools and Application, pp. 1–22, Dec. 2014.

[17] P. Barnaghi, W. Wang, C. Henson, and K. Taylor, "Semantics for the Internet of Things: early progress and back to the future". International Journal on Semantic Web & Information Systems, vol. 8, no. 1, pp. 1-21, Jan. 2012.