The Combination of Knowledge Management and Data mining with Knowledge Warehouse

Ming-Chang Lee
Department of Information Management, Fooyin University, Taiwan, ROC
Ming_li@mail2000.com.tw
doi: 10.4156/ijact.vol1.issue1.6

Abstract

Effective knowledge management (KM) enhances products, improves operational efficiency, speeds deployment, increases sales and profits, and creates customer satisfaction. Data warehousing provides an infrastructure that enables businesses to extract, and store vast amounts of corporate data. The purpose of data warehouse is to empower the knowledge workers with information that allows them to make decision based on a foundation of fact. The aim of this paper is to integrate a framework of knowledge management and data mining with knowledge warehouse. Therefore, first, it will brief review the existing of knowledge management, data mining, and decision support system. We then present a framework of knowledge management and data mining with knowledge warehouse. Second, knowledge, knowledge management and knowledge process is defined. Third, we introduce decision support, data mining and data warehouse support of knowledge management, and point out data mining in data warehouse environment. Four, the warehouse is defined. Using this definition, it can drive framework of knowledge warehouse. This framework contain 6 layers: knowledge input, knowledge activity, data store, application server, application system data base, and user client. In this paper some suggestions are made to get knowledge with data mining, which will provide the decision maker with an intelligent platform that enhances all phase of knowledge management and knowledge process.

Keywords

Knowledge Management, Data Warehouse, Data Mining, Decision Support, Knowledge Warehouse

1. Introduction

A knowledge warehouse (KW) is the component of an enterprise's knowledge management system, used to develop, store, organize, process, and disseminate knowledge. KW can be thought of as an "information repository" in which knowledge components are cataloged and stored for reuse. A knowledge warehouse enables a variety of different views of knowledge, useful in areas such as training or documentation. These views could be pre-set and organized by instructional designers or technical writers. Additionally, the knowledge warehouse could also support ad hoc queries, such as electronic performance support systems, intelligent help, or reference materials. Not incidentally, knowledge can be stored in several physical places, although that is not a requirement.

A data warehouse (DW) is a central repository for all or significant parts of the data that an enterprise's various. DW, an integral part of the process, provides an infrastructure that enables businesses to extract, cleanse, and store vast amount of corporate data from operational systems for efficient and accurate responses to user queries. DW empowers the knowledge workers with information that allows them to make decisions based on a solid foundation of fact [5]. Data mining is a decision-making functions (decision support tool). Data mining (DM) has as its dominant goal, the generation of no-obvious yet useful information for decision makers from very large data warehouse (DW). DM is the technique by which relationship and patterns in data are identified in large database [8]. In DW environment, DM techniques can be used to discover untapped pattern of data that enable the creation of new information. DM and DW are potentially critical technologies to enable the knowledge creation and management process [2]. The DW is to provide the decision-maker with an intelligent analysis platform that enhances all phase of the knowledge management process. Decision support system (DSS) or intelligent decision support system (IDSS) and DM can be used to enhance knowledge management and its three associated processes: i.e., tacit to explicit knowledge conversion, explicit knowledge leveraging, and explicit knowledge conversion [14].
DSS is a computer-based system that aids the process of decision-making [9]. DSS are interactive computer-based systems that help decision makers utilize data and models to solve unstructured problems.

DSS can also enhance the tacit to explicit knowledge conversion by eliciting one or more what-if cases (i.e., model instances) that the knowledge worker wants to explore. That is, as the knowledge worker changes one or more model coefficients or right hand side values to explore its effect on the modeled solution. That is, the knowledge worker is converting the tacit knowledge that can be shared with other workers and leveraged to enhance decision. DSSs which perform selected cognitive decision-making functions and are based on artificial intelligence or intelligent agent’s technologies are called Intelligent Decision Support Systems (IDSS) [10]. Dhar and Stein [6] use term to characterize the degree of intelligence provided by a decision support tool. It describes intelligence density as representing the amount of useful decision support information that a decision maker gets from using the output from some analytic system for a certain amount of time [6]. The goal of KW is to provide the decision maker with an intelligent platform that enhances all phase of knowledge management and knowledge process. Figure 1 is showed as a framework of knowledge management, data mining and IDSS with knowledge warehouse

2. Knowledge Management

Knowledge management is the process established to capture and use knowledge in an organization for the purpose of improving organization performance [16]. Knowledge management is emerging as the new discipline that provides the mechanisms for systematically managing the knowledge that evolves with enterprise. Most large organizations have been experimenting with knowledge management with a view to improving profits, being competitively innovative, or simply to survive ([4], [11], [13], [15]). Knowledge management systems refer to a class of information systems applied to managing organization knowledge, which is an IT-based system developed to support the organizational knowledge management behavior: acquisition, generation, codification, storage, transfer, retrieval [1]. There are two forms of knowledge explicit knowledge and tacit knowledge. Explicit knowledge is defined as knowledge that can be expressed formally and can be easily communicated or diffused throughout an organization [18]. Implicit (tacit) knowledge is knowledge that is unmodified and difficult to diffuse. The implicit knowledge is learned through extended periods of experiencing and doing a task, during which the individual develops a feel for and a capability to make intuitive judgments about the successful execution of the activity [3]. Nonaka and Takeuchi [12] view implicit knowledge and explicit knowledge as complementary entities. There contend that there are four modes (Socialization, Externalization, Combination, and Internalization) in which organizational knowledge is created through the interaction and conversion between implicit and explicit knowledge.
Common knowledge management practices include:
(1) Creating and improving explicit knowledge artifacts and repositories (developing better databases, representations, and visualizations, improving the real-time access to data, information, and knowledge; delivering the right knowledge to the right persons at the right time).  (2) Capturing and structuring tacit knowledge as explicit knowledge (creating knowledge communities and networks with electronic tools to capture knowledge and convert tacit knowledge to explicit knowledge).  (3) Improving knowledge creation and knowledge flows (developing and improving organizational learning mechanisms; facilitating innovation strategies and processes; facilitating and enhancing knowledge creating conversations/dialogues).  (4) Enhancing knowledge management culture and infrastructure (improving participation, motivation, recognition, and rewards to promote knowledge sharing and idea generation; developing knowledge management enabling tools and technologies).  (5) Managing knowledge as an asset (identifying, documenting, measuring and assessing intellectual assets; identifying, prioritizing, and evaluating knowledge development and knowledge management efforts; document and more effectively leveraging intellectual property).  (6) Improving competitive intelligence and data mining strategies and technologies.

3. Decision Support, Data mining and Data Warehouse support of Knowledge Management

3.1 Data mining vs. Data warehouse

Data mining includes tasks such as knowledge extraction, data archaeology, data exploration, data pattern processing, data dredging, and information harvesting. Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and subsequent knowledge from large databases [17]. DM uses well-established statistical and machine learning techniques to build models that predict customer behavior. Today, technology automates the mining process, integrates it with commercial data warehouses, and presents it in a relevant way for business users.

The data warehouse is a valuable and easily available data source for data mining operations. Data extractions the data mining tools work on come from the data warehouse. Figure 2 illustrates how data mining fits in the data warehouse environment. Notice how the data warehouse environment supports data mining.

3.2 Decision support system vs. Knowledge

In tacit to explicit knowledge conversion, the literature of knowledge acquisition in expert systems-based DSS is well established. DSS can enhance that the explicit knowledge conversion through the specification of mathematical models (e.g., linear programming models). The knowledge worker also explicitly specifies the model constrains in terms of the decision variables, and estimates both numerical coefficients of decision variables in each constraint and in the objective function (e.g. goal programming models).
The knowledge worker’s tacit knowledge is converted to explicit knowledge and stored in an appropriate form; it can be leveraged by making it available to others who need it. In adding, analyzing explicit knowledge to produce new knowledge can further it. Explicit knowledge stored in the form of instances of a mathematical model (what-if-cases) can be leveraged via deductive and/or inductive model analysis systems. Another form of explicit leveraging is found in case-based reasoning (CBR).

CBR is characterized by the knowledge worker making his or her interfaces and decision based directly on previous cases recalled from memory. That is, the knowledge worker tries to avoid, or reduce, the potential for failure by recalling previous similar failures and avoiding the associated pitfalls or changing key factors in those previous failures.

DSS or GSS can also provide valuable aids in internalizing explicit and new knowledge. One mode of internalizing explicit and / or new knowledge is through modifying the internal mental model that a knowledge worker uses to serve as a performance guide in specified situations.

3.3 Knowledge management system vs. Data warehouse

Knowledge management system (KMS) is a systematic process for capturing, integrating, organizing, and communicating knowledge accumulated by employees. It is a vehicle to share corporate knowledge so that the employees may be more effective and be productive in their work. Knowledge management system must store all such knowledge in knowledge repository, sometimes called a knowledge warehouse. If a data warehouse contains structured information, a knowledge warehouse holds unstructured information. Therefore, a knowledge framework must have tools for searching and retrieving unstructured information. Figure 3 is integration of KM and data warehouse.

4. Knowledge Warehouse

The DW is a type of database managed by a DBMS. Indeed, in its present form the DW is a database that uses a relational DBMS. The KW is a type of database managed by a Knowledge Base Management System (KBMS) and Artificial Knowledge Base (AKB). An AKB is the portion of an organization’s knowledge base expressed in the persistent storage and non-persistent memory of computers. Unlike a database which store records, an AKB stores a network of objects and components, and these encapsulate data and methods. KBMS is a computer application for managing (creating, enhancing, and maintaining) the AKB, just as a DBMS is a computer application for managing a database. KW may be viewed as subject-oriented, integrated, time-variant, and support of management’s decision processes. But unlike the DW, it is a combination of volatile and nonvolatile objects and components, and, of course, it stores not only data, but also information and knowledge.

In the DW, data about a subject is stored in a set of tables. As we have designed the KW, the storage structure is referred to as a knowledge base and is constructed as a tree with objects at the nodes. The KW can be accessed by executing stored knowledge. It can serve as the repository to document a company’s business knowledge. Since the KW contains a great deal of information encapsulated in attributes and methods, it can be dynamically queried again.

The KW can be through as six-layer processes of Knowledge input, Knowledge activity, data store; application server, application system database, and user client (see Figure 4).

Layer 1: Knowledge input

According [7] future companies will be dominated
by knowledge workers. In the next society, information and knowledge will be acquired easily through intranets, extranets, and the Internet. Knowledge flow will converge toward the Enterprise Information Portal (EIP). Knowledge flow also may come from non-IT channels, such as tacit knowledge (knowledge that is implicit by or inferred from actions or statements) and explicit knowledge (knowledge that is fully and clearly expressed, leaving nothing implied). All knowledge will be integrated into knowledge repositories.

Layer 2: Knowledge activity
The data warehouse “Extraction, transformation, Migration and load” (ETML) process has a parallel in the knowledge warehouse. KW also has logical structure to store knowledge that is analogous to the system tables that implement data store in the DW. Knowledge is applied through a layered representation that shields code until the bottom layer.

The spiral of knowledge postulates four interaction processes – socialization, externalization, combination, and internalization—that transfer individual employee knowledge to company knowledge.

Layer 3: Data Stores
An operational data store (ODS) is a database designed to integrate data from multiple sources to make analysis and reporting easier. Because the data originates from multiple sources, the integration often involves cleaning, resolving redundancy and checking against business rules for integrity. An ODS is usually designed to contain low level or atomic (indivisible) data (such as transactions and prices) with limited history that is captured "real time" or "near real time" as opposed to the much greater volumes of data stored in the data warehouse generally on a less frequent basis.

Data mart (DM) is a subset of an organizational data store, usually oriented to a specific purpose or major data subject that may be distributed to support business needs. DM is analytical data stores designed to focus on specific business functions for a specific community within an organization. Data marts are often derived from subsets of data in a data warehouse, though in the bottom-up data warehouse design methodology the data warehouse is created from the union of organizational data marts.

Layer 4: Application Server
This layer contains application server (or engine), such as BPE, KDD, MOLAP, and CTS. Middleware is computer software that connects software components or applications. The software consists of a set of services that allow multiple processes running on one or more machines to interact across a network. This technology evolved to provide for interoperability in support of the move to coherent distributed architectures, which are used most often to support and simplify complex, distributed applications. It includes web servers, application servers, and similar tools that support application development and delivery. Middleware is especially integral to modern information technology based on XML, SOAP, Web services, and service-oriented architecture.

Business Process Engine (BPE) is an orchestration tool that liberates data from process and process from application, allowing your business to adapt quickly to change. BPE can reduce transaction costs, accelerate order fulfillment and enhance customer satisfaction. It provides a flexible, robust and scalable system. Knowledge Discovery and Data Mining (KDD) is an interdisciplinary area focusing upon methodologies for extracting useful knowledge from data. The ongoing rapid growth of online data due to the Internet and the widespread use of databases have created an immense need for KDD methodologies. The challenge of extracting knowledge from data draws upon research in statistics, databases, pattern recognition, machine learning, data visualization, optimization, and high-performance computing, to deliver advanced business intelligence and web discovery solutions.

MOLAP is an alternative to the ROLAP (Relational OLAP) technology. While both ROLAP and MOLAP analytic tools are designed to allow analysis of data through the use of a multidimensional data model, MOLAP differs significantly in that it requires the pre-computation and storage of information in the cube - the operation known as processing. MOLAP stores this data in optimized multi-dimensional array storage, rather than in a relational database (i.e. in ROLAP).

A transaction server (CTS) is a software component that is used in implementing transactions. A transaction involves multiple steps which must be completed atomically. For example, when paying someone from your bank, the system must guarantee that the money is taken from your account and paid into the other persons account. It would simply be unacceptable for just one or the other action to take place; both must occur in order for the transaction to have taken place.

Layer 5: Application System
This layer contains application system, such as DSS, ES, and EIS. An Executive Information System (EIS) is a type of management information system intended to facilitate and support the information and decision-making needs of senior executives by providing easy access to both internal and external information relevant to meeting the strategic goals of
the organization. It is commonly considered as a specialized form of a Decision Support System (DSS). The emphasis of EIS is on graphical displays and easy-to-use user interfaces. They offer strong reporting and drill-down capabilities. In general, EIS are enterprise-wide DSS that help top-level executives analyze, compare, and highlight trends in important variables so that they can monitor performance and identify opportunities and problems. EIS and data warehousing technologies are converging in the marketplace.

An expert system (ES) is software that attempts to reproduce the performance of one or more human experts, most commonly in a specific problem domain, and is a traditional application and/or subfield of artificial intelligence. A wide variety of methods can be used to simulate the performance of the expert however common to most or all are (1) the creation of a so-called "knowledgebase" which uses some knowledge representation formalism to capture the Subject Matter Experts (SME) knowledge and (2) a process of gathering that knowledge from the SME and codifying it according to the formalism, which is called knowledge engineering. Expert systems may or may not have learning components but a third common element is that once the system is developed it is proven by being placed in the same real world problem solving situation as the human SME, typically as an aid to human workers or a supplement to some information system.

**Layer 6: User client**

A query is a form of questioning, in a line of inquiry.

A GUI is a graphical (rather than purely textual) user interface to a computer. As you read this, you are looking at the GUI or graphical user interface of your particular Web browser. Today's major operating systems provide a graphical user interface. Applications typically use the elements of the GUI that come with the operating system and add their own graphical user interface elements and ideas.

![Figure 4. Framework of knowledge warehouse](image-url)
Elements of a GUI include such things as: windows, pull-down menus, buttons, scroll bars, iconic images, wizards, the mouse, and no doubt many things that haven't been invented yet.

With the increasing use of multimedia as part of the GUI, sound, voice, motion video, and virtual reality interfaces seem likely to become part of the GUI for many applications.

5. Summary and Conclusion

The knowledge warehouse, as an extension of the data warehouse, provides a mechanism to capture, store, and access knowledge. Knowledge can be automated techniques such as are found in data mining. Knowledge can be stored in a tree structure with software objects placed at the tree nodes. Combined with tree search algorithms and tree manipulation methods, this provides a powerful built-in control structure at the system level. KW can provide a software object model of business process with embedded intelligence. In this paper, we have knowledge warehouse model, a knowledge warehouse architecture which is an IT-based system developed to support the organizational knowledge management behavior: acquisition, generation, codification, storage, transfer, retrieval. The framework of knowledge warehouse consists of an of knowledge input, knowledge activity, data store, application server, application system database, and user client.

6. Reference