Expert Systems with Applications 36 (2009) 3614–3622

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

An ontology-based business intelligence application in a financial knowledge management system

Hilary Cheng [a,*], Yi-Chuan Lu [b,1], Calvin Sheu [c,2]

[a] Department of Business Administration, Yuan Ze University, Chung-Li 320, Taiwan
[b] Department of Information Management, Yuan Ze University, Chung-Li 320, Taiwan
[c] Graduate School of Management, Yuan Ze University, Chung-Li 320, Taiwan

ARTICLE INFO

Keywords:
Business intelligence
Data mining
Knowledge management
Bond ratings

ABSTRACT

Business intelligence (BI) applications within an enterprise range over enterprise reporting, cube and ad hoc query analysis, statistical analysis, data mining, and proactive report delivery and alerting. The most sophisticated applications of BI are statistical analysis and data mining, which involve mathematical and statistical treatment of data for correlation analysis, trend analysis, hypothesis testing, and predictive analysis. They are used by relatively small groups of users consisting of information analysts and power users, for whom data and analysis are their primary jobs. We present an ontology-based approach for BI applications, specifically in statistical analysis and data mining. We implemented our approach in financial knowledge management system (FKMS), which is able to do: (i) data extraction, transformation and loading, (ii) data cubes creation and retrieval, (iii) statistical analysis and data mining, (iv) experiment metadata management, (v) experiment retrieval for new problem solving. The resulting knowledge from each experiment defined as a knowledge set consisting of strings of data, model, parameters, and reports are stored, shared, disseminated, and thus helpful to support decision making. We finally illustrate the above claims with a process of applying data mining techniques to support corporate bonds classification.

1. Introduction

Knowledge is power. Today's business environment has been tougher than ever. Enterprises experience global competitions. Customers demand more on product features and services. Corporate expenses are continuously increasing. To survive in the harsh environment, high-level management needs business intelligent information to efficiently manage corporate operations and support their making of decisions. Support-level staffs need knowledge information to provide better customer services for gaining satisfaction and retaining loyalty. Vast operating data is staggered into various corporate databases and needs consolidating. It has become more important than ever to access and generate valuable knowledge and share information among authorized users within a corporation and/or business partners. Thus, a system of integrating knowledge management and decision support processes is in great demand. As mentioned in (Bolloju, Khalifa, & Turban, 2002), a synergy can be created by the integration of decision support and knowledge management, since these two processes involve activities that complement each other. The knowledge retrieval, storage, and dissemination activities in knowledge management functionality enhance the dynamic creation and maintenance of decision support models, subsequently, enhancing the decision support process. From the system design's point of view, what we need is a new generation of knowledge-enabled system that provides enterprise an infrastructure to capture, cleanse, store, organize, leverage, and disseminate not only source data and information but also the knowledge or value-added information of the firm (Nemati, Steiger, Iyer, & Herschel, 2002).

We present the concept of financial knowledge management system (FKMS), which is a prototype of KM environment specifically for financial research purposes. What the environment generates is groups of knowledge set with strings of data, models, parameters, and reports. Ontology of knowledge management and knowledge sharing is presented. Finally, a realization of decision support and knowledge sharing processes to a corporate bond classification is illustrated. With FKMS, knowledge workers can freely extract sets of financial and economic data, analyze data with different decision support modules, rerun experiments with different sets of parameters, and finally disseminate value-added information (knowledge) through middleware or Internet to remote clients. Not to mention that the knowledge

* Corresponding author. Tel.: +886 3 4638800x2626; fax: +886 3 4630377.
E-mail addresses: hilary@saturn.yzu.edu.tw (H. Cheng), imylu@saturn.yzu.edu.tw (Y.-C. Lu), jwcalvin@saturn.yzu.tw (C. Sheu).
[1] Tel.: +886 3 4638800x2600; fax: +886 3 4352077.
[2] Tel.: +886 3 4638800x2693; fax: +886 3 4630377.

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.02.047

generated is being collected, classified, and shared with colleagues, and thus well archived into corporate business intelligence databank.

The remainder of this paper proceeds as follows. Section 2 reasons our motivation for developing FKMS. Section 3 introduces system architecture of FKMS. Section 4 presents the ontology of knowledge management and knowledge sharing, and demonstrates with a case of corporate bond classification problem. Section 5 concludes this paper.

2. Motivation of developing a financial knowledge management system

One of the biggest challenges that most security investment institutions experienced was the lack of an intelligent data mining system to support investment researches decisions. The problems that their system encountered including:

(i) lack of efficiency in managing vast financial data,
(ii) lack of communication and knowledge sharing among analysts,
(iii) lack of a mechanism to resolve synchronization problems when multiple users are accessing data,
(iv) lack of a mechanism to efficiently manage generated research experiments,
(v) lack of an automation to publish its reports to clients via its Web sites or email.

Though data for financial applications are simple data, the data typically includes time series information and the relationships among the financial instruments are complex. For example, consider a derivative security objects. The derivative security object often shares underlying securities with other derivatives. Underlying securities can come from many classes of instruments, from a simple currency to an interest rate swap to a hedge. As the securities become more complex, the problems of data management and knowledge discovery become more difficult. Consider a security portfolio, the portfolio construction is a process of quantitative analysis over massive amounts of data. The data cube and ad hoc analysis techniques are an invisible solution to support this process. A system that efficiently supports financial application thus would provide support for: (i) temporal objects, (ii) object management, which is the efficient storage and manipulation of complex data, (iii) knowledge discovery, the capability of extracting information as rules for decision making. The (i) and (ii) result in financial data modeling using object-oriented techniques, whereas the (iii) is merely data mining techniques.

3. System architecture

The architecture of the FKMS is a layered structure as shown in Fig. 1. The object-oriented design gives the system flexibility and expandability. FKMS consists of five layers: the resource layer, the data conversion layer, the data storage and management layer, the knowledge/trend/pattern layer, and finally the user process layer. Various sources of data are converted into the time series database according to predefined schema. With OLAP tool, users can easily define various periodical reports with report generator and generate sets of data cubes for analysis. The resulting data cubes are stored and managed in the FKMS that various valuation models, data mining techniques or statistical modules can be applied to. In addition, Web pages represented by the XML can be sent to major corporate clients (as a message), as well as posted on the enterprise information portal with Web-enabled modules and messaging tools. Finally, thousands of files generated by the analysts are well managed and monitored for knowledge sharing and as for internal performance evaluations.

The data cubes are stored in a traditional relational database management system (RDBMS); users can easily divert the data cubes via ODBC or JDBC for analytical applications at the knowledge/trend/pattern layer. The selected analytical applications are either designed or programmed by users, or the off-the-shell software such as Excel, Matlab, IMSL, SAS, SPSS, or S-Plus. A use case diagram in Fig. 2 depicted the function requirement for FKMS implementation.

3.1. Resource layer

Various resources of data are used by analysts when they write research reports or run valuation models. Typical examples of these data resources are financial databases from foreign data vendors such as Bloomberg, Data Stream, First Call, or from domestic data vendors like the TEJ and the SFI, as well as other reports from competitors, and some periodically published data on the Web sites. While some data are static, meaning that they are periodically released, some are dynamic, which means that they are not periodic. On the other hand, the format of data can be classified as structured, semi-structured, and unstructured. We focus on the management of structured data, as their format is clearly defined so that data operations or manipulations can be deployed.

3.2. Data conversion layer

A data warehouse should always provide its users with accurate, consistent, and real-time data. It should be flexible to support all corporate operations and changes. Corporations usually manage

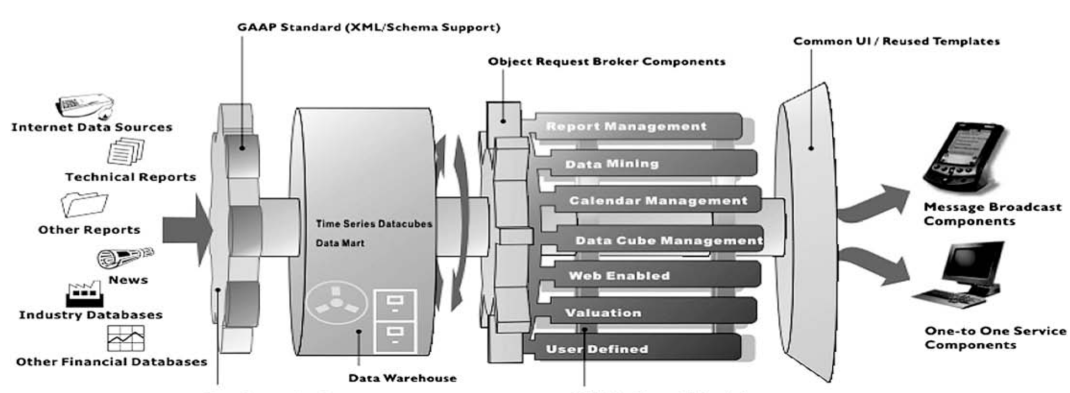

Fig. 1. The system architecture of financial knowledge management system.

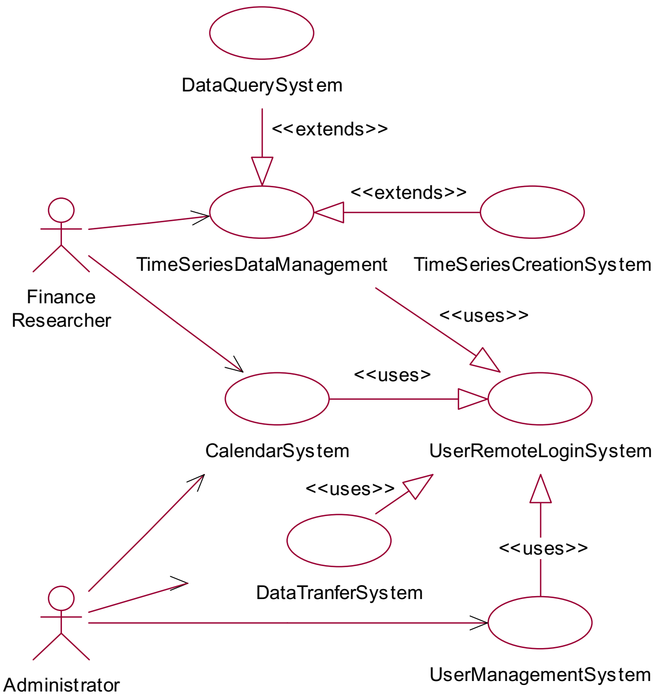

Fig. 2. Use case diagram of FKMS.

several different databases. To unite all these data sources into one data warehouse needs prudent consideration and technology. The data conversion layer of FKMS allows various data sources to be converted and loaded into its data warehouse for efficient storage and retrieval. Two important modules in the data conversion layer are shown in Fig. 3 and described as follows.

3.2.1. Enterprise metadata management (EMM) module

Metadata bridges the gap between data format and the physical data storage. In the aggregate planning of the metadata for financial domain, we take the following aspects for implementation into consideration: the attributes and the key of tables, the resources of tables, algorithms that may be used for data conversion, metadata update, etc.

The enterprise metadata management (EMM) module allows data administrator to define enterprise metadata to align with corporation structure and operations based on the business requirements. EMM is the brain of the data conversion layer. For the purpose of being expandable, scalable, and portable, the ontology of the knowledge abstraction was represented with XBRL schema. Also, the design of EMM is flexible to accommodate corporate structure and business changes if a corporate re-engineering is needed.

3.2.2. Extraction, transformation, and loading (ETL)

The extraction, transformation, and loading (ETL) module automates data conversion and loading into the data warehouse. It is integrated with EMM to map data in the various legacy databases to data warehouse. It supports and synchronizes with current corporate business needs. Should any data changes be required, the administrator can easily modify the mapping rules to accommodate the changes. It provides tremendous saving of time and money for the operations.

The next step after metadata definition is data conversion, which consists of four phases: creating the data schema, listed companies management, mapping rules for conversion, and finally the conversion process. The storage of data schema was managed by two tables: Schema_Field and Schema_NameSpace. Where there are data items with same names (such as price to book ratio; P/B) but from different databases (for example, Data Stream and Bloomberg), Schema_NameSpace gives a clear understanding of data sources. Besides, Schema_Field keeps good management of all data items in the data warehouse. Due to the increasing number of listed companies in the database, we provide a management function, which can automatically update when companies either are out of the market or change their identification codes. The rules mapping system not only allow the users to create relationship between the new data source and the predefined data items, but also create mapping rules between the new data item with the same value from the ElementSourceCode in the existing schema. The data conversion process ends with the management of values of all data items, company details and the data items.

Once the source data are transformed into the data warehouse, the knowledge discovery process begins with the creation of data cubes by knowledge workers for simple queries or running decision support application models.

3.3. Data storage and data management layer

The main functions in the data storage and data management layer include financial time series data management, user accounts management, data cube creation and management, and calendar management. The calendar management function is to allow an enterprise to manage and track financial markets' working calendar. Its friendly user interface allows users to easily define rules of a specific financial market working days, vacations, and holidays.

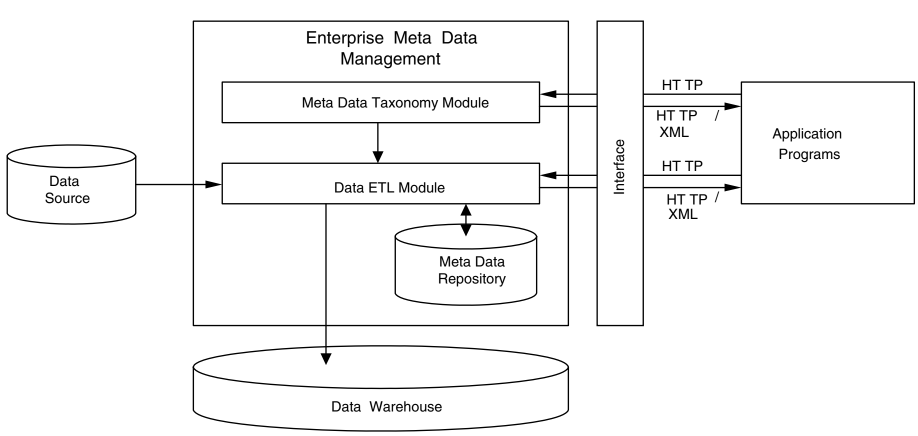

Fig. 3. EMM and ETL modules in the data conversion layer.

The module is also capable of checking exceptions to the defined rules. The calendar management is particularly important to financial time series management. The following queries for simple calculations help in explaining the need for calendar management:

A. To calculate the weekly, monthly, quarterly, and annually returns for each listed company in the past five years. The report must be updated by the last trading day of each week.
B. To summarize a company's performance by calculating the difference of the company's last two EBIT growths, and the projected P/E ratio.

From above examples, we see that each query defines a new cube containing time series with different time granularity (see Lu, Cheng, & Wang, 1999 for more discussion about calendar management). The creation of data cube takes five steps described as follows: select root data cube; select items; screening; add/delete qualified companies; final review. Knowledge workers that generate the new knowledge will have to define where the new knowledge comes from, and that is to define what the root data cube(s) is (are). Once the root data cube(s) is (are) defined, the knowledge workers will then select groups of items they are interested in for subsequent analysis. These items will be retrieved from the database once requested, and can be downloaded to either Word, Excel, simply to the Website for snapshot purpose, or data mining modules for decision support purpose (see (Lu, Cheng, & Sheu, 2002) for more details).

When conducting data mining technique to corporate bonds classification applications, knowledge workers will first choose a set of variables that he (she) believes that can sufficiently explain the default risk or solvency ability of those studied companies. The number of variables and the selected set of financial time series vary from test to test, so does the result of each experiment. Fig. 4 shows the metadata for a data cube described by a time series header:

CubeName = {backtesting}
Criteria = {ComID = '23xx', MarketCap > 100 mil., EPS > $0.5, MA(10)_close_daily < $30}
DateCreated = {03/12/1998}
TimeGranularity = {Monthly}
DataTimePeriod = {06/01/1993, 05/30/1995}
Priority = {6}

4. Integration of decision support and knowledge management process

As mentioned in Lu and Cheng (2003), a successful knowledge management system enhances the way people work together, enables knowledge workers and partners to share information easily so they can build on each other's ideas and work more effectively. The goal is to gather company proprietary knowledge in order to come up with the best decision-making, or to quickly seize the initiative with innovative ideas. To optimize the flow of information and knowledge-worker-to-knowledge-worker interaction so that company can always make better trading or investment decisions, the specific group of data and modeling results should be administered and properly shared. Future knowledge can be generated by capturing existing (shared) knowledge via filtering, storing, retrieving and disseminating explicit knowledge and by creating and testing new knowledge (Nemati et al., 2002).

According to the KDD process defined by Fayyad, Piatetsky-Shapiro, and Smyth (1996), as shown in Fig. 5, the flow of data and shared information goes as follows: users select data with different criterion to perform Quant analysis. The data retrieved will automatically be saved as data cubes. Each data cube records the time this file has been created, the screening rules, the data item names and types, and so on, so that colleagues can easily share and exchange their knowledge and value-added information with others. In the meantime, the company will keep good management of each individual worker's knowledge, and so be ready for good customer relationship management. In the following, we show how decision support and knowledge management is integrated.

4.1. Data mining application to corporate bond classification

Bond rating is a system whereby rating agencies analyze and assess the security of principal repayments and interest payments for individual bonds, and denote the level of issuer credit worthiness through easily understandable indicators such as AAA and AA. Rating agencies provide this investment information as a service to investors. The bond rating characterizes the default risk for the investor and affects the borrowing cost for the issuer. Rating agencies such as the Standard & Poor's and the Moody's determine the ratings from various financial data, which is not publicly disclosed. Furthermore, the agencies have publicly stated that the assessment of the ratings requires expert judgment in addition to statistical

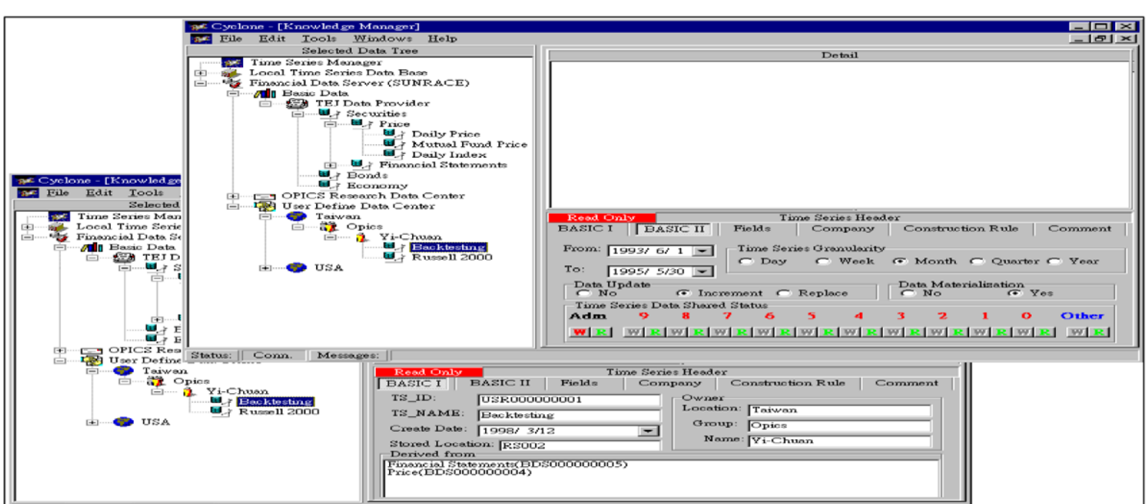

Fig. 4. A snapshot of GUI of financial knowledge management system (FKMS).

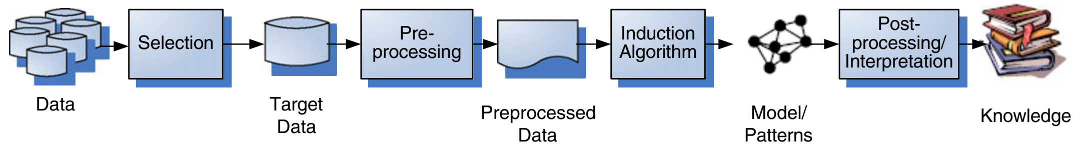

Fig. 5. The KDD process (adapted from (Fayyad et al., 1996).

analysis of financial data. It is thus generally believed that the ratings are to a certain extent judged on the basis of subjective factors that are not easily quantifiable and not directly related to a particular firm.

Due to these ambiguities in the bond rating process, the ratings cannot be reproduced 100% accurately. However, it is nonetheless helpful to be able to evaluate the default risk for the issuing company by estimating the ratings based on a model. The reasons for independent ratings estimates are in twofolds. First, not every firm is rated by the agencies. A financial institution may be interested in investing in a firm whose rating is not available. Second, ratings are reviewed only periodically. As a result, being able to predict rating changes before the announcement can be quite valuable. A predictive model that maps fundamental financial factors onto an estimated rating refined by the expert judgment can accomplish this job.

Traditional credit analysis for corporate bonds almost exclusively focused on the default risk, of which ratio analysis was the primary concern. Pointed out in Howe (1995, chapter 18), this kind of approach was deemed appropriate when interest rates were stable and investors purchased bonds with the purpose of holding them to maturity. However, as more and more investors trade bonds actively to make a profit on changes in interest rates or in absolute or relative credit quality, the analysis focused on the ratios and the trends of firms' profitability becomes important. Factors employed in common stock analysis models, such as return on equity (ROE), operating margins, and turnover ratios will serve as good indicators for changes of corporate credit.

Similar to common stock analysis, the bond ratings model we proposed is composed of three fundamental steps: industry analysis, company analysis, and financial analysis. The reasons to conduct both industry analysis and company analysis in addition to financial analysis follow from the facts that each industry has its unique characteristics and economic cycles. One company with a 10% growth in EPS may be deemed as good in comparison with the universe; nevertheless, it may turn out to be just fair when compared with its industry or segment growth. In addition to different industry structures and performances, an individual company may be underscored or overestimated if the company's unique characteristics are neglected in the evaluation process. For example, the company may have a negative or an ordinary EPS versus others in the same industry; however, it may incur a large amount of capital gain from disposing some of company's fixed assets such as land, which is often the case in Taiwan. Since each industry has its unique characteristics and economic cycles, it is thus believed to be critical to examine each industry as well as each company before any statistical or artificial intelligence techniques to be applied in the financial analysis. The financial data are then rated separately not only according to the industry, but also to a specific segment within an industry, such that the differences of heterogeneity can be minimized.

Most studies on bond ratings using either statistical analysis or techniques from artificial intelligence domains to perform ratio analysis. The applications of neural network to corporate bond ratings are presented in several literatures. Dutta and Shekhar (1988) developed multiple regression models and neural network models to detect AA bonds (as rated by the S&P) from a set of AA and non-AA bonds. Singleton and Surkan (1991) also studied the bond rating ability of neural networks. Utans and Moody (1991) developed a neural network bond rating predict or using the architecture selection methods through the prediction risk. When applying neural network techniques for bond rating prediction, the targets usually are the public ratings such as the ones from the Moody's, or the Standard & Poor's. In the absence of public ratings such as the case in Taiwan, the application of neural network by supervised learning becomes impractical, as there is no target to learn and train. Among various architectures and algorithms for artificial neural networks, the self-organizing feature map (SOFM) has a special property of effectively creating spatially organized internal representations of various features of input factors. We therefore apply this technique to the financial analysis in our model. The SOMF/LVQ technique and the model development are discussed as follows.

4.2. Overview of the hybrid SOFM/LVQ classifier for bond ratings

The key for clustering data is to determine their probability distributions. The self-organizing feature map (SOFM) Kohonen (1988,1990) is known as a nonparametric neural network approach for learning the probability density function of the input samples. The mechanism of map formation used in the SOFM is called the competitive learning in which only one output node, or one per local group of nodes at a time gives the active response to the current input signal. A good resulting map of the SOFM requires the minimization of the average distance with respect to the neuron field. For a mathematical formulation of this requirement, a quantity is defined

$$E(W) = \int \|X - W_i\|^2 P(X) \mathrm{d}X \tag{1}$$

where $E(W)$ is the mean distance with respect to the neuron field; $P(X)$ is the probability density describing the distribution of sample vector X; W_i is the weight of neuron i which corresponds to a cluster center vector.

Since there is a lack of knowledge of the function $P(X)$, the construction of cluster vectors relies on the accumulation of many individual steps. In each step, a cluster vector is moved by a small step size in response to the sample vector. A detailed mathematical analysis of Kohonen model can be found in Ritter, Schulten, and Kohonen (1988), and Ritter and Schulten (1989).

The purpose of using the SOFM is to group input vectors into one of the pre-specified clusters according to the features detected in the training data set. During training, cluster vectors (templates) that belong to the same region will move closer in response to the input sample thereby reducing the within-industry error. After learning is completed, each cluster vector represents a template of the training data.

The main focus of this paper is the problem of perceiving and identifying bond ratings. If the SOFM is to be used in the sense of classification, then the output nodes representing clusters are usually made to represent each bond rating. Their identity within the industrial class is no longer important. In particular, the criterion to evaluate the effectiveness of features must be a measure of the class overlap or separability, which depends not only on the within-industry criterion but also on the between-industry criterion. In order to seek the optimum cluster set, the resulting clusters from the SOFM need to be refined so that the heterogeneity among different ratings can be increased, and that heterogeneity can be done by introducing a supervised refining algorithm. We have chosen the supervised version of Kohonen's model known as the learning vector quantization (LVQ) (Kohonen, 1986) to refine selected features. Once the desirability of cluster analysis is clear, the practical problem becomes how to decide what the discriminant function will be. We have used a simple Euclidean distance criterion as the system discriminating function which results in a piecewise linear classifier. In the classification stage, each testing data sample is applied to the distance classifier sequentially. The minimum value selector outputs a class label as the one having the smallest distance to the input sample.

Several SOFM models reside in the SOFM block to learn the features detected in each industrial class of corporate bond rating training samples. This is to learn the class-conditional probability density function of each industry. After learning is completed, cluster vectors in each SOFM model represent templates of a given class training samples. We then assign a known target label to each template. The function of LVQ block is to group those labeled templates as an initial cluster vectors to learn classification boundaries. After LVQ training is done, the distance discriminant function is used to classify input data to the target class with the shortest distance.

4.3. Modeling of the hybrid SOFM/LVQ classifier for bond ratings

The construction of a clustering and classification model takes three basic stages: the collection of a set of measurement or feature vectors obtained from a set of companies that are the ratings of interest, the matching of the measurement with the bond rating templates, and finally, the classification. One of the known neural network techniques for the bond rating prediction is the multi-layer perceptron (e.g., back-propagation model). The multi-layer perceptron model learns the classifications on the basis of known bond ratings with a supervised learning. If the model had to decide by itself whether a group of bonds belong together or not, by examining the measurement or feature vectors, this becomes unsupervised learning problem. Due to the lack of the rating agencies for Taiwan bond market, the unsupervised learning seems to be a proper approach. In doing so, we consider the concept of clustering. Clusters are simply tools to interpret and to evaluate the measurements and features of bond ratings. If the clustering results in a grouping that corresponds to our understanding of the classification problem and our goal in recognition, then we have a good chance that the clusters have been selected appropriately. If the clustering result is unnatural and contradicts our intuition, then there is high possibility that we will not succeed in achieving our goals. Such "unnatural" clustering is a sign that we have to modify our interpretation of the problem.

To construct corporate bond clusters, one approach is to collect historic corporate bond feature vectors, which are then labeled individually by the expertise of bond analysts. The resulting feature vectors are then stored in a database as bond rating templates for further classification. The disadvantage of this approach is that a large number of reference bonds are required for the classifier. To avoid the memory overload while achieving a reasonable performance, one can use annually data instead of quarterly data. Another approach is to use a neural network based technique to automatically learn the representative bond clusters. The goal is to reduce the number of reference bond templates while maintaining a reasonably high correct classification rate. A hybrid SOFM/LVQ (self-organizing feature map/learning vector quantization) neural network model was developed and applied to the waveform target classification (Chang & Lu, 1994; Stewart, Lu, & Larson, 1994). The hybrid SOFM/LVQ system performs three basic functions: the learning of the within-industry homogeneity, the learning of the between-industry heterogeneity and the identification of unknown bond ratings.

Fig. 6 shows the SOFM/LVQ classifier for bond classification. We first use the SOFM to select a bank of neurons with each neuron tuned to a specific bond rating. The purpose of using the SOFM is to group input vectors into one of the pre-specified clusters according to the features detected in the training data set. During training, cluster vectors (templates) belonging to the same region will move closer in response to the input sample thereby reducing the within-industry error. After learning is completed, each cluster vector represents a template of the training data.

If the SOFM is to be used in the sense of classification, then the output nodes representing clusters are usually made to represent each bond rating, and their identity within the industrial class is no longer important. In particular, the criterion to evaluate the effectiveness of features must be a measure of the class overlap or separability, which depends not only on the within-industry criterion but also on the between-industry criterion. In order to seek the optimum cluster set, the resulting clusters from the SOFM need to be refined in order that the heterogeneity among different ratings can be increased, and that it can be done by introducing a supervised refining algorithm. Once the desirability of cluster analysis is clear, the practical problem becomes one of deciding what the discriminant function will be. We have used a simple Euclidean distance criterion as the system discriminant function, resulting in

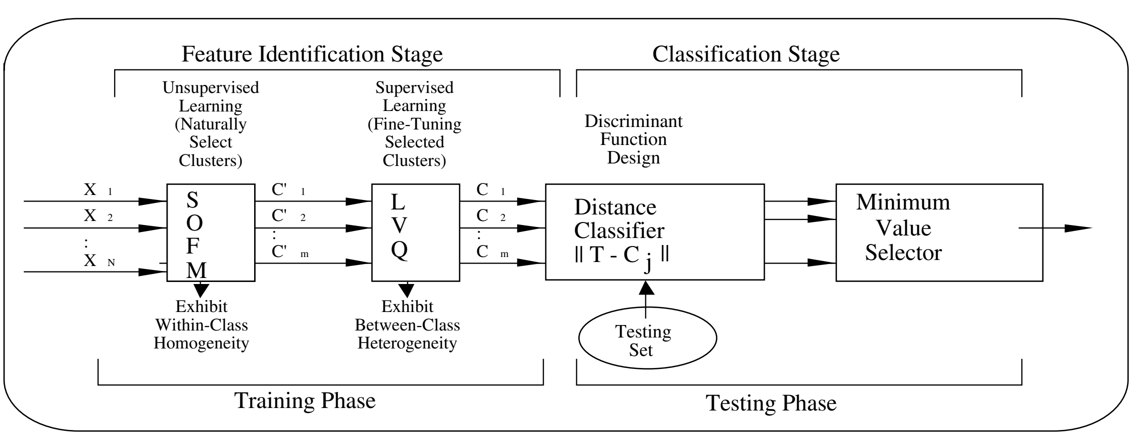

Fig. 6. A hybrid SOFM/LVQ classifier for corporate bond ratings.

a piecewise linear classifier. In the classification stage, each testing data sample is applied to the distance classifier sequentially. The minimum value selector outputs a class label as the one having the smallest distance to the input sample.

We then use the supervised version of Kohonen's model known as the learning vector quantization (LVQ) (Kohonen, 1986) to refine selected features and to learn the between-industry heterogeneity, which is done by collecting those selected neurons as the initial cluster centers for the LVQ to learn their class boundaries. This is to maximize the probability of correct classification. In the classification stage, the SOFM/LVQ classifier accepts input data of an unknown bond, computes distances of this data with those representative bond templates, and then classifies this data to the bond rating class with the shortest distance. The number of clusters for this experiment is set to be eight. The financial data are clustered separately not just according to the industry, but even to a specific segment within an industry, such that the differences of heterogeneity can be maximized.

In Fig. 6, several SOFM models reside in the SOFM block to learn the features detected in each industrial class of corporate bond rating training samples. This is to learn the class-conditional probability density function of each industry. After learning is completed, cluster vectors in each SOFM model represent templates of a given class training samples. We then assign a known target label to each template. The function of LVQ block is to group those labeled templates as an initial cluster vectors to learn classification boundaries. After LVQ training is done, the distance discriminant function is used to classify input data to the target class with the shortest distance.

4.4. Data selection

All corporate financial quarterly data in the cement and the electronics industry from March 1990 to June 2006 were selected for the study. The sample data were divided into two sets: the data in the period of March 1990 to December 2004 were used for training, and the rest were used for testing. Different industry data were rated separately.

As for the input financial factors, note that there are numerous financial ratios that can be considered as input factors on the default risk prediction model. In analyzing a particular industry or segment, however, factors selected may vary to reflect the specific or critical characteristics with that industry or segment. We use the twelve variables from Taiwan Corporate Credit Risk Index (TCRI) model by TEJ for comparison purpose. The TCRI was developed to predict the changes in the credit risk of companies. The variables used were ROE, ordinary income, operating income margin, ROA, quick ratio, interest expense ratio, debt ratio, receivables turnover, inventory turnover, operating income – also called earnings before interest and taxes (EBIT), total assets, and sales growth rate.

4.5. Test results

Fig. 7 shows the 3-D plot clustering result of cement companies for the period of year 1990–2006. The z-axis represents the clusters labeled from 1 to 8. Points in x–y plane are resulted from Sammon mapping (Sammon,1969), which maps 12-dimensional input vectors to 2-dimensional points on a plane whereby the distances between the image vectors tend to approximate to Euclidean distances of the input vectors. Similar plots for the clustering results for the electronics industry are shown in Fig. 8. While most companies maintain a pretty stable overall performance and therefore stay in the same (or neighboring) cluster from time to time, some companies have changed dramatically.

The training data from the SOFM model were initially clustered each quarter in terms of their stability, turnover rates, growth

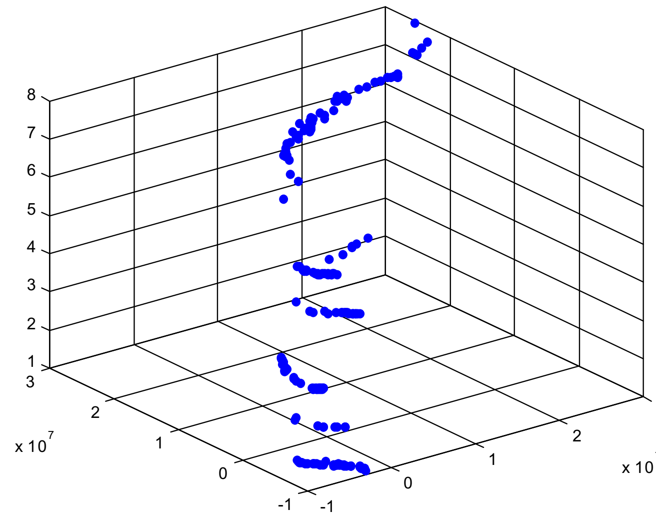

Fig. 7. A 3-D plot of cement companies for the year from 1990 to 2006.

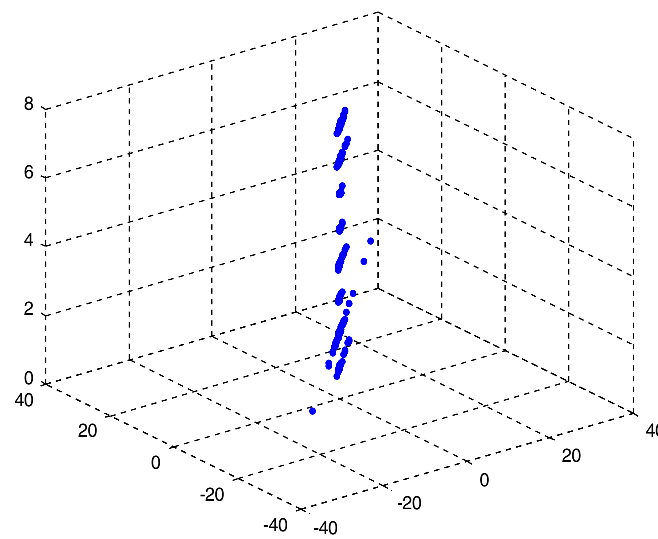

Fig. 8. A 3-D plot of company ratings in electronics industry from 1990 to 2006.

potential, and solvency, and so on. The bond analysts or managers then examine the initial clustering results utilizing the information from both industry analysis and company analysis and their expertise as well. The bond analysts play an important role in the modeling process as they will verify the correctness of the initial clustering results, identify any outliers that do not belong to a specific cluster, and eventually rank the refined clusters in the refinement/tuning process.

By looking at the quarterly ratings for a particular company, one gets better insight about the trend of performance for this company. The fine-tuned cluster data from the LVQ model for each industry can be stored as a template for predicting rating changes from time to time as long as new financial forecasts are on hand. Similarly, for newly issuing bonds, recent financial data can be tested using the same industry cluster template for predicting its credit.

4.6. Ontology of knowledge management

The knowledge for corporate bond classification can be discovered by applying various models as shown in past literature. The application of choosing SOFM/LVQ to manipulate data along with

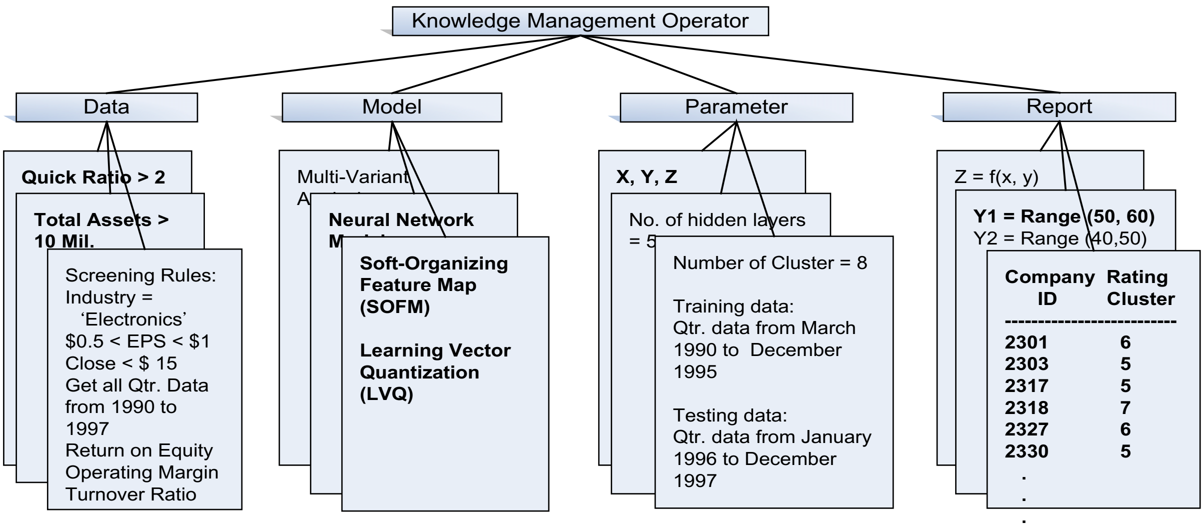

Fig. 9. Knowledge management ontology.

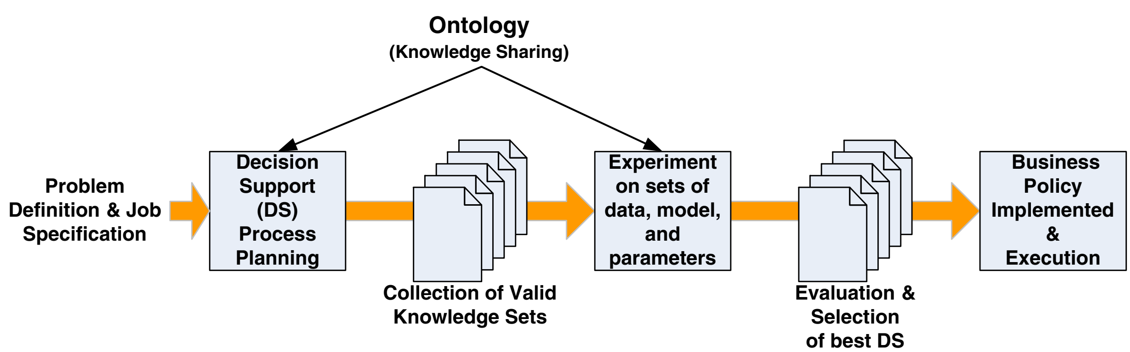

Fig. 10. Integration process of decision support and knowledge management by FKMS.

fine-tuning initial cluster results is merely one instance of knowledge sets within corporate business intelligence database. In this experiment setting, the knowledge set is described as: KnowledgeClass_BondClassification (data, model, parameter, report), where the attributes are detailed as follows:

BondClassification :=

Data = {ROE, ordinary income, operating income margin, ROA, quick ratio, interest expense ratio, debt ratio, receivables turnover, inventory turnover, EBIT, total assets, and sales growth rate}.
Model = {SOFM/LVQ}
Parameter = {8 clusters}
Report = {templates of fine-tuned cluster data}

The resulting ontology that manages the knowledge discovery process is depicted in Fig. 9.

4.7. Knowledge sharing

Knowledge sets stored in FKMS database can be classified and screened for further application. Fig. 10 shows the complete knowledge sharing process. Users searching for bond classification knowledge sets would simply do the following:

Select KnowledgeClass := {BondClassificaiton}
Select ModelType := {Back Propagation, SOFM/LVQ, and Regression}

Once the models were selected, the metadata associated with data cubes of each model can be analyzed; combination of data, model, and parameters can be manipulated; initial test results can be evaluated and fine-tuned based on experts' domain knowledge or the shared knowledge, and finally, the best models will be selected and implemented on new business policies for execution. In this case, the final learned clusters as well as the model parameters are being stored and classified, and can further be searched and applied to rating predictions of newly issuing bonds, or rating changes of existing bonds.

5. Conclusion

Domain experts need a flexible system environment where they can freely select data and models and run different settings of parameters for decision support purposes. Knowledge sets of each research experiment containing data, models, parameters, and results essentially provide great value to business intelligence generation. The integration of decision support and knowledge management processes is crucial for enterprises to create their niche business intelligence and to maintain global competitive advantages.

We present the concept of financial knowledge management system, FKMS, which is a prototype of KM environment specifically for financial research purposes. What FKMS environment generates are groups of knowledge sets containing strings of data, models, parameters, and reports for each analytical study. Ontology of knowledge management and knowledge sharing is presented. Finally, we demonstrate a business intelligence generation process, where corporate bonds are classified by domain experts using selected data mining approach, finally learned cluster data of bond features are being saved for prediction of rating changes or determining interest rates of newly issuing bonds.

As thousands of knowledge sets gradually piled up in the knowledge database, intelligent screening of ontology, spontaneous push-and-pull knowledge dissemination, and performance ranking will essentially and inevitably lead the way to more powerful knowledge generation.

References

Bolloju, N., Khalifa, M., & Turban, E. (2002). Integrating knowledge management into enterprise environments for the next generation decision support. *Decision Support Systems, 33*, 163–176.

Chang, K. C. & Lu, Y. C. (1994). Feedback learning: A hybrid SOFM/LVQ approach for radar target classification. In *International symposium on artificial neural networks.* Taiwan, ROC.

Dutta, S., & Shekhar, S. (1988). Bond rating: A non-conservative application of neural networks. *Proceedings of ICNN-88. (II)*, 443–450.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *CACM, 39*(11), 27–34.

Howe, J. T. 1995. Credit Analysis for Corporate Bonds. In *The Handbook of Fixed Income Securities* (4th ed.), Irwin.

Kohonen, T. (1986). Learning vector quantization. Helsinki University of Technology, Laboratory of Computer and Information Science, Report TKK-F-A-601.

Kohonen, T. (1988). *Self-organization and associative memory*. Berlin: Springer.

Kohonen, T. (1990). The Self-organizing Map. *Proceedings of the IEEE, 78*(9), 1464–1480.

Lu, Y. C., Cheng, H., & Wang, W. H. (1999). Technical report – A time series management system for financial decision support: Models, techniques, and implementations. Financial data mining lab., Yuan Ze University, Taiwan, ROC.

Lu, Y. C., Cheng, H., & Sheu, C. (2002). A knowledge management system with financial data mining embedded. In *Proceedings of the CIEF'2002 Conference*. NC: Research Triangle Park.

Lu, Y.-C., & Cheng, H. (2003). *Automated optimal equity portfolios discovery in a financial knowledge management system, computational intelligence in economics and finance*. Springer-Verlag. pp. 387–402.

Nemati, H. R., Steiger, D. M., Iyer, L. S., & Herschel, R. T. (2002). Knowledge warehouse: An architectural integration of knowledge management, decision support. *Artificial Intelligence and Data Warehousing, Decision Support Systems,* (33), 143–161.

Ritter, H., Schulten, K., & Kohonen, T. (1988). Self-Organizing Maps: Exploring Their Computational Capabilities. In *IEEE ICNN 88 Conference, I*. San Diego, pp. 109–116.

Ritter, H., & Schulten, K. (1989). Convergence properties of kohonen topology conserving maps: Fluctuation, Stability, and Dimension Selection. *Biological Cybernectics, 60*, 59–71.

Sammon, J. W. Jr., (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions Computing, C-18*(5), 401–409.

Singleton, J., & Surkan, A. (1991). Modeling the judgment of bond rating agencies: Artificial intelligence applied to finance. *Journal of the Midwest Finance Association, 20*, 72–80.

Stewart, C., Lu, Y. C., & Larson, V. (1994). A neural clustering approach for waveform classification. *Pattern Recognition, 27*(4), 503–513.

Utans, J., & Moody, J. (1991). Selecting neural network architecture via the prediction risk application to corporate bond rating prediction. In *Proceedings of the first international conference on artificial intelligence applications on wall street.* Los Alamitos, CA: IEEE Computer Society Press.