ISSN 2347-4289

Impulsion Of Mining Paradigm In ERP

Kowsalya.S, M.Phil Scholar

Department of Information Technology, AJK College of Arts & Science, Navakkarai, Coimbatore, Tamil Nadu, India - 641105.
Email: kowsalya.selvaraj@gmail.com

ABSTRACT: Traditionally, the construction industry has been faced with the problems of meeting project schedule, budget, and specifications set by the owner and architect/engineer. The proper utilization of internal and external resources is essential if construction companies are to make the best business decisions, maximize business goals, and survive in the competitive environment. Although the construction industry is one of the largest contributors to the economy, it is considered to be one of the most highly fragmented, inefficient, and geographically dispersed industries. To overcome this inefficiency, a number of solutions have long been offered including adaptation of information technology and information systems. Major construction companies embarked on the implementation of integrated IT solutions such as enterprise resource planning (ERP) systems to better integrate their various business functions. Each construction project is characterized by a unique set of site conditions, project team, and the temporary nature of relationships between project stakeholders. As a result, construction companies are required to have extensive customization of preintegrated business applications from the vendors of ERP systems. Unfortunately, such extensive customizations result in a greater challenge in implementing ERP systems. Therefore, finding the best ERP systems implementation strategy is needed to maximize the benefits of such integrated IT solutions for construction companies. Here I suggest an approach and implementation of mining paradigm in ERP system which provides the best solutions and better results. Since ERP deals with large volume of data and the challenge in that is its way of representation. By getting mining paradigm into its way there reside various approaches to bring the data in crisp and legitimate formats. My approach on this doesn't end with getting the mining paradigm into ERP's implementation; also I have developed a POC (Proof of Concept) on DBSCAN method of mining which depicts both functional and technical workings into act. To make this concept a proved one with real time workings, I have provided the facts and approaches that a construction company follows in real world and the data they dealt with. This paper also encloses how this DBSCAN method works over this business functionality in different dimensions to prove its working.

Keywords : DBSCAN, SOM, BOM.

1 INTRODUCTION

ERP is an application software and utilization paradigm by which an organization can integrate its diverse functions. ERP is structured such that, if one part of an organization's working impacts another part, the impacted departments information base gets altered automatically and ensuring that the departments function in tandem, and not in isolation as they usually do An ERP system is an integrated solution, sharing a centralized database, with all 'users', Human Resources / Payroll / Benefits, E-procurement, Accounting, Budgets, etc. being served by the same database through one point of entry. Data need only be entered or updated once, reducing errors, time and labor for reports, analysis, and planning and program management. Ultimately, time and resources are shifted to innovating, problem solving and direct service to customers rather than inputting, processing, organizing, verifying and related "busy work" that burns through time and money. Online ERP is the web-enabled front-end for the company wide ERP software. Users are provided with web-based information and a central and uniform data basis. A Workflow Management System ensures targeted and secure user guidance through the required business processes. The online ERP software, needs the permanent availability of user-specific Internet applications is ensured. The online Enterprise resource planning software forms the backbone of business systems in manufacturing firms and many other businesses. The online ERP solution for the users are a strategic and coherent approach to the management of organizations for providing facilities such as loans to the customers or giving credit to the suppliers or to reflect their financial status in the respective industrial environment. It is the most valuable asset for the people working there who are responsible both - individually and collectively for the achievement of the organization's goals. Different ERP systems have similar features to work effectively. Firstly, the systems are componentized so that diverse business functions are designed as different components. The components are usually integrated and seamless data flow between components makes them to collaborate as one function. Different ERP systems are flexible and expandable. As a result, they can work together harmoniously and any changes in business operations are easy to fulfill their needs.

2 ERP IN CONSTRUCTION INDUSTRY

ERP for Construction is a comprehensive ERP solution designed by taking utmost care with core functionality of the industry in mind. While developing the ERP for Construction Industry, core functional areas such as Project Monitoring and Control, Sales Management, Tendering, Bill of Quantities, Purchase and Supplier Management, Onsite Engineers Portal, Material Management, Labor and Contractors Management, Accounts Management and Human Resource Management have been given extra care and prominence. All activities right from pre-construction stage to post construction stage can be captured and monitored. Management can access and manage onsite activities from anywhere anytime. Independent Estimates can be generated for each task based on respective quantities. Estimating is facilitated by system prompted activities. All different activities can be defined as different assemblies. Associated with each assemblies are bill of quantities, labours, turnaround time for completion. Construction industry encloses the below mentioned modules.

1. Sales and Marketing
2. Project Management and Control
3. Bill of Quantity Management
4. Onsite Engineers Portal
5. Purchase Management
6. Inventory & Material Management
7. Contract and Contractors Management
8. Human Resource & Labor Management
9. Accounts and Finance Management

Now a days members of the Construction Industry are facing biggest challenges and need to meet today's market demands. To meet today's market demands, needs to manage information. In order to manage information/data, deliver high-quality information to the decision-makers at the right time and

to automate the process of data collection, collation and refinement, organizations have to make information technology to utilize it in the best way possible.

3 CHANLLENGES IN ERP

From the above listed modules, there remain large data sets to be managed, and not all the data sets are equally related and have impact all the time. Based on the time and work progress the impact and need of the data set varies. At the same time, one cannot ignore the data that got passed its need over a particular time; there may be a need of past data over a period of time for any reason. So it remains a challenging task to maintain the data sets and representing that data set based on its requirement. Not all the data can be represented in same format all the time and in other words, the same data shall be in need to represent in different format on various instances. For example, in the HR and Labor management module, the need for employee details may vary at the time of attendance verification or wages calculation or master detail preparation. The same way in CRM module, the customer details shall be required for different reasons as to know the status of payment and outstanding, receipt details, payment completed details and contact details. So this could be a challenging task to represent the same data in different format with relevant information. In this thesis I focused on this factor to get solve the need of data set with right format with right content without noise. Here comes the need of clustering concepts to achieve this goal. Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. ERP systems always store their data in a database which makes it a viable target for data mining activities, but many times companies use diverse systems together to form complete ERP functionality and their data is scattered across multiple databases. In these cases, it is sometimes necessary to gather the scattered data into a single database called a Data Warehouse (DW), before submitting it to data mining activity.

4 DATA MINING LITERATURE SURVEY

Computer scientists often refer to Moore's law, which states that computer processing speed doubles about every 18 months. It is less well known that computer storage capacity doubles about every nine months. Like an ideal gas, computer databases expand to fill available storage capacity. The resulting large amounts of data in databases represent an untapped resource. Like a gold mine, these data could be extracted into information. That information could then be converted to valuable knowledge with data mining techniques. It is difficult to convey the vast amount of unused data stored in very large databases at companies, universities, government facilities, and other institutions throughout the world and its current rate of increase. It could soon be the case that computer data storage will exceed human capability to use that data storage and the data it contains. A process for converting large amounts of data to knowledge will become invaluable. A process called Knowledge Discovery in Databases (KDD) has evolved over the past ten to fifteen years for this purpose. Data mining algorithms are included in the KDD process. A typical database user retrieves data from databases using an interface to standard technology such as SQL. A data mining system takes this process a step further, allowing users to discover new knowledge from the data. Data mining, from a computer scientist's point of view, is an interdisciplinary field. Data handling techniques such as neural networks, genetic algorithms, regression, statistical analysis, machine learning, and 15 cluster analysis are prevalent in the literature on data mining. Many researchers state that data mining is not yet a well-ordered discipline. The major opportunities for improvement in data mining technology are scalability and compatibility with database systems, as well as the usability and accuracy of data mining techniques. The conceptual success model of ERP has been shown below on the basis of KDD (before applying any pre-processing steps with respect to mining paradigm). In later part of this thesis the same success model will be depicted by applying mining paradigm of technology acceptance model (TAM)

5 CLUSTERING METHODOLOGY

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics. Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including values such as the distance function to use, a density threshold or the number of expected clusters) depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It will often be necessary to modify data preprocessing and model parameters until the result achieves the desired properties. Clustering can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. So, the goal of clustering is to determine the intrinsic grouping in a set of unlabeled data. But how to decide what constitutes a good clustering? It can be shown that there is no absolute "best" criterion which would be independent of the final aim of the clustering. Consequently, it is the user which must supply this criterion, in such a way that the result of the clustering will suit their needs.

5.1 Clustering Vs Classification

Clustering and classification are both fundamental tasks in Data Mining. Classification is used mostly as a supervised learning method, clustering for unsupervised learning (some clustering models are for both). The goal of clustering is descriptive, that of classification is predictive. Since the goal of clustering is to discover a new set of categories, the new groups are of interest in themselves, and their assessment is

intrinsic. In classification tasks, however, an important part of the assessment is extrinsic, since the groups must reflect some reference set of classes. Clustering groups data instances into subsets in such a manner that similar instances are grouped together, while different instances belong to different groups. The instances are thereby organized into an efficient representation that characterizes the population being sampled. Formally, the clustering structure is represented as a set of subsets C= C1, . . ., Ck of S, such that any instance in S belongs to exactly one and only one subset. Since clustering is the grouping of similar instances/objects, some sort of measure that can determine whether two objects are similar or dissimilar is required. There are two main types of measures used to estimate this relation, distance measures and similarity measures. In a database of customer records, where each record represents a customer's attributes. These can include identifiers such as name and address, demographic information such as gender and age, and financial attributes such as income and revenue spent. Clustering is an automated process to group related records together. Related records are grouped together on the basis of having similar values for attributes. This approach of segmenting the database via clustering analysis is often used as an exploratory technique because it is not necessary for the end-user/analyst to specify ahead of time how records should be related together. In fact, the objective of the analysis is often to discover segments or clusters, and then examine the attributes and values that define the clusters or segments. As such, interesting and surprising ways of grouping customers together can become apparent, and this in turn can be used to drive marketing and promotion strategies to target specific types of customers.

5.2 Clustering Algorithm Selection

This step refers to the choice of an algorithm those results in the definition of a good clustering scheme for a data set. A proximity measure and a clustering criterion mainly characterize a clustering algorithm as well as its efficiency to define a clustering scheme that fits the data set.Proximity measure is a measure that quantifies how “similar” two data points (i.e. feature vectors) are. In most of the cases we have to ensure that all selected features contribute equally to the computation of the proximity measure and there are no features that dominate others. In the clustering criterion step, we have to define the clustering criterion, which can be expressed via a cost function or some other type of rules. We should stress that we have to take into account the type of clusters that are expected to occur in the data set. Thus, we may define a “good” clustering criterion, leading to a partitioning that fits well the data set.

5.3 Validation of the Results

The correctness of clustering algorithm results is verified using appropriate criteria and techniques. Since clustering algorithms define clusters that are not known a priori, irrespective of the clustering methods, the final partition of data requires some kind of evaluation in most application.

5.4 Interpretation of the Results

In many cases, the experts in the application area have to integrate the clustering results with other experimental evidence and analysis in order to draw the right conclusion. In general terms, clustering may serve as a pre-processing step for other algorithms, such as classification, which would then operate on the detected clusters. Not all the algorithms will suite for all the defined problems, even though the basic process of clustering is same as described in previous sections, each one has its own thread variation between others in providing the solutions. On comparing all the available clustering techniques, I have decided to go with DBSCAN algorithm so as to meet the defined problems in the ERP based construction industry. Various factors has been taken into consideration while comparing the clustering techniques such as the type of data each model focus as input and complexity of results in the output, outliers left, input parameters and criteria that must get satisfied in each technique. Based on these factors I have depicted the reason why the DBSCAN technique is more suited for my problem

S.No	Name	Type of Data	Outliers	Input Parameters	Geometric Output	Results
1	K-Means	Numerical	No	No. of Clusters	Non Convex Shapes	Centre of cluster
2	K-Mode	Categorical	No	No. of Clusters	Non Convex Shapes	Modes of cluster
3	PAM	Numerical	No	No. of Clusters	Non Convex Shapes	Medoids of cluster
4	CLARA	Numerical	Partial	No. of Clusters	Non Convex Shapes	Medoids of cluster
5	CLARANS	Numerical	Partial	No. of Clusters Max no of neighbor	Non Convex Shapes	Medoids of cluster
6	FCM	Numerical	No	No. of Clusters	Non Convex Shapes	Centre of cluster beliefs
7	BIRCH	Numerical	Yes	Radius of Clusters	Non Convex Shapes	N' Number of points in clusters; 'F' linear sum of points in clusters; 'LS' the square sum of N data
8	ROCK	Numerical	Yes	No. of Clusters	Arbitrary Shapes	Assignment of data values to clusters
9	DBSCAN	Categorical	Yes	Cluster radius and minimum number of objects	Arbitrary Shapes	Assignment of data values to clusters

Table-1: *Characteristics comparison of clustering technique to decide the suitable model.*

The above table shows the comparative study made over different clustering techniques on various dimensions. As per the problem defined in this thesis for ERP implementation the data to be dealt with will be huge volume and in all type of formats. In construction industry, they might be handle text, image numeric calculations data related to respective modules, so the input parameters will also need to be accommodated for all the types. Modules that deals with cash flows will expect 100 percent correction in the results and the modules like estimation and production may have variations when compared to the actual and scheduled conventions where there remain outliers. So we cannot stick to a particular format for the resulting data. It has to be accommodated as per the type and nature of the work. For this flexibility the DBSCAN technique will suite close when compared to the others. This has been clearly presented in the above table. As all the other techniques deals with specific data types as this focus on categorical data. The input parameters of others are number of clusters while the DBSCAN deals with radius of cluster and number of objects, with this it is clear that we can customize the input

based on the modules as said earlier. The geometrical output from all the techniques is non convex shapes it means one cannot assure the fixed format for the results in the other techniques, but in DBSCAN the output will be arbitrary shape which suits the customer's needed format. Hence I prefer the DBSCAN technique for this ERP implementation in construction industry.

5.5 Algorithm Working

STEP-1: The algorithm initiates by randomizing the map's node weight vectors.

STEP-2: The second step is to grab an input vector $\mathbf{D(t)}$ from the defined data set.

STEP-3: This is followed by traversing each node in the map. The iteration in this step is

STEP-3.1: Using the Euclidean distance formula to find the similarity between the input vector and the map's node weight vector.

STEP-3.2: Track the node from which produces the smallest distance from the iteration (this node is then considered to be the best matching unit, BMU)

STEP-4: After iteration process, update the map's nodes in the neighborhood of the BMU (including the BMU) by pulling them closer to the input vector, hence arriving the equation as follows

$$\mathbf{Wv}(s + 1) = \mathbf{Wv}(s) + \Theta(u, v, s)\, \alpha(s)(\mathbf{D}(t) - \mathbf{Wv}(s))$$

STEP-5: The algorithm conclused by increasing and repeating from step 2 while $s < \lambda$.

5.6 Training SOM Algorithm

The training SOM algorithm relies initially on setting the weights and learning rate. The input vectors that are to be clustered shall be presented to the network. Based on the initial weights, the winner unit is then calculated either by sum of products method or Euclidean distance method. Depending on the winner unit selection, the weights are updated for the respective winner unit. This in turn an epoch is said to be completed once if all the input vectors are presented to the network. Hence by updating the learning rate, several epochs of training shall be performed. A specimen of two dimensional Kohonen Self Organizing Feature Map network with sample data set is shown in below Fig- 1

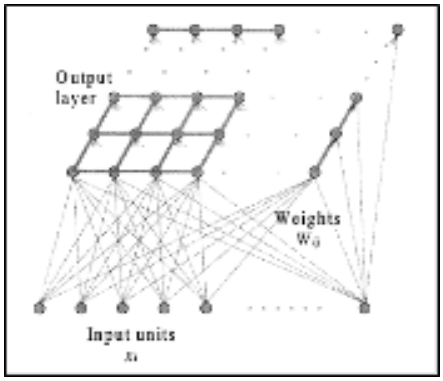

Fig-1: *The SOM Network*

6. Evaluation and Results

6.1 The Time Characteristics and Clustering Precisions

The superiority of our proposed algorithm is to improve the clustering results precision. Therefore, the F-measurement which is under supervision (F Measure) is used in experiments. In order to prove our proposed algorithm effectively, we make the contrast experiments with *I DBSCAN* and *DBSCAN* algorithms. To illustrate relevant problem comparatively, we suppose *MinP ts*=4 and *Eps*=*Eps*4 in two algorithm. And the comparison data is shown in below Table

DataSet	Dimension	Object Number	Algorithm	Eps	Min pts	Operation Time (ms)	Accuracy
DS1	2	159	I-DBSCAN	33.62	2	208	98.74
			DBSCAN	18.34	4	122	37.61
DS2	2	170	I-DBSCAN	18.46	4	245	90
			DBSCAN	10.81	4	165	87.64
Iris	2	150	I-DBSCAN	4.61	4	211	88.67
			DBSCAN	2.53	4	162	54
DS3	4	600	I-DBSCAN	3.74	5	680	83.83
			DBSCAN	1.92	4	488	28.5
DS4	5	700	I-DBSCAN	4.42	4	908	86.57
			DBSCAN	2.43	4	792	59.14

Table-2: *Comparison Data between I-DBSCAN and DBSCAN with parameter working results of different dimensions.*

From Table 5, we can see that I-DBSCAN algorithm spends more time than DBSCAN Algorithm. This phenomenon is a result of abundant computing involved in I-DBSCAN Algorithm. For every data object in experimental data sets, its corresponded DISTn_4 need to be calculated. So its time efficiency is comparatively low. But fortunately, these two algorithms' time spared doesn't have order of magnitude difference. Moreover, it appears that the accuracy is not satisfactory when MinP ts is 4 and Eps is Eps4 in DBSCAN algorithm. On the contrary, it has higher precision in I-DBSCAN algorithm. Especially, more irregular data objects behave, more evident I-DBSCAN algorithm's advantage is. In conclusion, it would rather determine the parameter MinP ts according to the chracteristics of data set than take a fixed value. DBSCAN is a classical density-based clustering algorithm, which can form clusters in different sizes and shapes, determine the cluster number automatically, and not be affected by the noises or outliers. However, two parameters (Eps and MinP ts) are required to be inputted manually.

6.2 Clustering of Uncertain Data

In this section, we present our error-aware density-based clustering approach. We start with a formal specification of the clustering problematic of uncertain data in problem specification which is followed by error aware extension of DBSCAN as first approach and clustering aggregation as second approach. The specific application scenario for the clustering is characterized by the following aspect: Each data object O is represented by a number of n different captured sensor values Si (1 <= i <= n) including some error or uncertain range. As a result, a data object is described not only by one single feature vector as an n dimensional data point but by an n dimensional region in which all points within this region equally likely represent the

object as shown in Figure-9(a). In the ongoing description, we denote the n dimensional region as data region. More formally, a data object O is described by the following vector:

O = {S1,……..,Sn, E1,…….,En},

Where n represents the number of captured sensor values. S1,……..,Sn are the measured sensor values and each Ei (1<=i <=n) represents a complex function as error quantification for each sensor value. These error functions depend on several factors. Moreover, these error functions may be independent of or dependent on several measured sensor values.

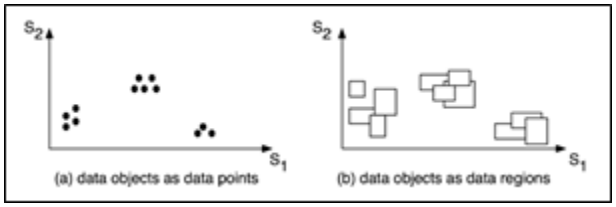

Figure-1: *Representation of Data Objects*

The above figure shows where the error functions are independent from each other, and therefore, the shapes of the data regions are hyper cubes. However, shapes of the data regions may be arbitrary. Within the data region, we assume that each possible data point represents the data object with the same probability.

7 Conclusion

More and more organizations are seeking to integrate the core functions of their business with technological advances. Enterprise systems facilitate this integration process through a single software architecture that links all aspects of business to function as one unit. Organizations continue to reap the benefits of enterprise systems, but they also encounter challenges. As a construction based organization face many obstacles in achieving their goals in terms of time, cost and customer satisfaction. Only on implementing the above said concepts it is clear the management can utilse the same available data in various dimensions which give different results. In this thesis I focused on DBSCAN methodologies that suits best in achieving the expected results. Since ERP itself is dealt with numerous raw data feed up from users and managing those data remains a challenge, the methods and algorithm suggested are best suit for handling the data and providing crisp information. According to the suggested algorithm, we need to provide input in different format, as partitioning method implies the split of data at each level under a common functionality. This best suits for estimation module in ERP where the total project cost can only been achieved from its base level by splitting the data from work group level to element level(BOM). As a result the project cost can be calculated at any level and at any set of combinations. On the other hand density linkage approach focuses on identifying the difference in the property of data. This well suits to the sales module where the data are scattered from flat allotment to receipt collection. First the data are populated as a pre-processing step as a input to the algorithm, on grouping them under a particular property the data will be shown such as flat availability, Payment outstanding and Payment completed (Receipt collection). While this is achieved through uniform kernel method where as in flexible beta method the uniqueness of data property is find out and allowing the user to make their input as customized one. With this the results will be formulated and represented in multiple dimensions ensuring the organization's progress and future assessment.

7.1 Operational Benefits

Implementing DBSCAN in ERP with the purpose of synchronizing the functions of different departments. An integrated system reduces the time used in processing documents, such as payrolls and other external documents. Information visibility and transparency within an organization is a benefit that facilitates the different operations carried out by the various departments. By using an ERP system with the described methodologies implemented, the companies improve the quality and efficiency of customer service, production and distribution. ERP helps the manufacturing process flow more smoothly, and it improves visibility of the order fulfillment process. This provides a real-time view of the company's overall performance which helps managers to make proper decisions in a timely manner.

7.2 Managerial Benefits

Managers find it less hectic to oversee operations and to ensure that key business objectives are achieved through enterprise systems. Because they are able to access information from a centralized server, managers find that the decision-making process becomes more informed and yields better results. On bringing the ERP functionality with DSCAN algorithm's technical assistance the management feels that the system allows organizations to be more flexible so that they can more easily adapt and capitalize on new business opportunities. This not only increases confidence in the accuracy of data and information but also ensures that all programs, whether it is sales related, purchase related or logistics related, are up-to-date. For any business to avoid wasting valuable time, it is crucial that while one operation is asleep, another can have total confidence in the shared information that it's using to run its own. This is as well as having one such system that implements the DBSCAN techniques with the ERP functionality that reflect both international and local business processes and compliance requirements.

Acknowledgment

I wish to thank my research guide Ms.Gomathi, Head of the Department, Department of Computer Science, AJK College of Arts & Science. Also I thank all the other staff members and scholars in my department who helped and provide support to perform such research workings with a better outcome. I owe the reviewers and authorities of IJTEEE for taking my research study to external viewers.

References

Books:

[1] Kaufman L. and Rousseeuw P. J (1990), "Finding Groups in Data: An Introduction to Cluster Analysis", John Wiley & Sons.

[2] Ankerst M., Markus M. B., Kriegel H., Sander J(1999), "OPTICS: Ordering Points To Identify the Clustering Structure", Proc.ACM SIGMOD'99 Int. Conf. On Management of Data, Philadelphia, PA, pp.49-60.

[3] Guha S, Rastogi R, Shim K (1998), “CURE: An efficient clustering algorithm for large databases”, In: SIGMOD Conference, pp.73~84.

[4] Ester M., Kriegel H., Sander J., Xiaowei Xu (1996), “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”, KDD’96, Portland, OR, pp.226-231.

[5] Wang W., Yang J., Muntz R(1997), “STING: A statistical information grid approach to spatial data mining”, In: Proc. of the 23rd VLDB Conf. Athens, pp.186~195.

[6] Raymond T. Ng and Jiawei Han (2002), “CLARANS: A Method for Clustering Objects for Spatial Data Mining”, IEEE Transactions on Knowledge and Data Engineering, Vol. 14, No. 5.

[7] Rakesh A., Johanners G., Dimitrios G., Prabhakar R(1999), “Automatic subspace clustering of high dimensional data for data mining applications”, In: Proc. of the ACM SIGMOD, pp.94~105.

[8] Efficient Map/Reduce-Based DBSCAN Algorithm with Optimized Data Partition Bi-RuDai - dept. of Comput. Sci. & Inf. Eng., Nat. Taiwan Univ. of Sci. & Technol Taipei, Taiwan.

[9] An Improved DBSCAN Algorithm for High Dimensional Datasets by Glory Shah Kaufman L. and Rousseeuw P. J (1990), “Finding Groups in Data: An Introduction to Cluster Analysis”, John Wiley & Sons.

[10] Ankerst M., Markus M. B., Kriegel H., Sander J(1999), “OPTICS: Ordering Points To Identify the Clustering Structure”, Proc.ACM SIGMOD‟99 Int. Conf. On Management of Data, Philadelphia, PA, pp.49-60.

[11] Guha S, Rastogi R, Shim K (1998), “CURE: An efficient clustering algorithm for large databases”, In: SIGMOD Conference, pp.73~84. Ester M., Kriegel H., Sander J., Xiaowei Xu (1996), “A Density-Based

[12] Algorithm for Discovering Clusters in Large Spatial Databases with Noise”, KDD‟96, Portland, OR, pp.226-231. Wang W., Yang J., Muntz R(1997), “STING: A statistical information grid approach to spatial data mining”, In: Proc. of the 23rd VLDB Conf. Athens, pp.186~195.

Thesis:

[13] http://genome.tugraz.at/MedicalInformatics2/SOM.pdf

[14] http://www.cs.bham.ac.uk/~jxb/NN/l17.pdf

[15] http://lib.tkk.fi/Diss/2002/isbn951226093X/article4.pdf

[16] http://www.cs.hmc.edu/~kpang/nn/som.html

[17] http://www2.tku.edu.tw/~tkjse/5-1/5-1-5.pdf

[18] http://users.ics.aalto.fi/mikkok/thesis/book/node12.html

[19] http://kanaya.naist.jp/SOM/