11th International Research/Expert Conference
"Trends in the Development of Machinery and Associated Technology"
TMT 2007, Hammamet, Tunisia, 05-09 September, 2007.

COLLABORATION KNOWLEDGE BASED SYSTEMS AND DATA MINING METHODS USING FUZZY CLUSTERING

Brano Markić, Ph.D.,
Faculty of Economics University of Mostar,
Matice hrvatske bb, Mostar
Bosnia and Herzegowina

Dražena Tomić, Ph.D.,
Faculty of Economics University of Mostar
Matice hrvatske bb, Mostar
Bosnia and Herzegovina

ABSTRACT
Knowledge based systems are computer programs which contain formalized knowledge in a narrow problem domain. Expert knowledge is extracted and formalize in various forms such as: production rules, semantic networks, triplet attribute-object-value, frames or first order predicate. But the knowledge can operate only on information getting from data stored in data bases or data warehouses. Extracting information from data is the main task of data mining tools. Therefore data mining methods and knowledge based systems are complement and their collaboration is naturally determined.
Different data mining methods help in finding patterns and relationships among objects representing by the set of attributes values. These methods cope with bulk of available data, and intend to discover hidden trends and relations among data in order to achieve faster, better and easier decision making. Cluster analysis is one of the ways that enables unsupervised classification, i.e. it is the process of grouping the data into classes so that the data objects (examples) are similar to one another within the same cluster and dissimilar to the objects in other clusters. This study uses the fuzzy c-means clustering method that was first developed by Dunn (1973) and that is based on the crisp k-means clustering method. The main goal of the paper is to investigate and present one possible integration of knowledge based systems with fuzzy c-means clustering. It was developed adequate software solution. Such integration has a great potential in different application fields, has a common importance, but in this paper we concentrate on application related to decision making in different business function.
Keywords: knowledge based systems, data mining**,** cluster analysis, data staging process.

1. INTRODUCTION
Knowledge Based System is a computer program that contains stored knowledge and solves problems in a specific field in much the same way that a human expert would. One of the main problems and most difficult tasks in building knowledge based systems is representing the knowledge discovered by data mining tools such as fuzzy c-means clustering.
The goal of paper is to investigate the possibility of knowledge extraction from database and building knowledge management system for applying and using knowledge in management at different organisation levels. The investigation therefore includes a few steps:

a) building data model
b) adjustment of data model to format that fuzzy c-means clustering algorithm expects as its input
c) representation the results of fuzzy c-means clustering as B-tree and production rules.
d) building and propose adequate software tools

Database management system provides an environment that manages and accesses the huge volume of data. Deriving knowledge from databases is impossible without adequate tools and programs, without computer assistance. Knowledge discovery in data (KDD) is such a "generic" approach to analyze and extract useful knowledge from databases. The process of knowledge discovery inherently consists of several steps. The first step is *to understand the application domain and to formulate the problem*. Our application domain is segmentation of customers according to appropriately attributes [3] . Customer relationship management faces the problem of customers clustering with the final goal to formulate adequate price discounts and payments as key element of the contracts between supplier and customer.
This step is clearly a prerequisite for extracting useful knowledge and for choosing appropriate data mining methods in the third step according to the application target and the nature of data.
The second step is *to collect and preprocess the data*, including the selection of the data sources, the removal of noise or outliers, the treatment of missing data, the transformation (discretization if necessary) and reduction of data, etc.

2. DATA

The first step in building knowledge management system is establishing data model and its physical implementation in any data base management system. Database stores data concerning various types of objects: customers, products, suppliers, accountants, orders, calculations etc. Databases contain objects that belong to different types and four types of relationships *1:1*, *1:n (or n:1)* and *n:m*.

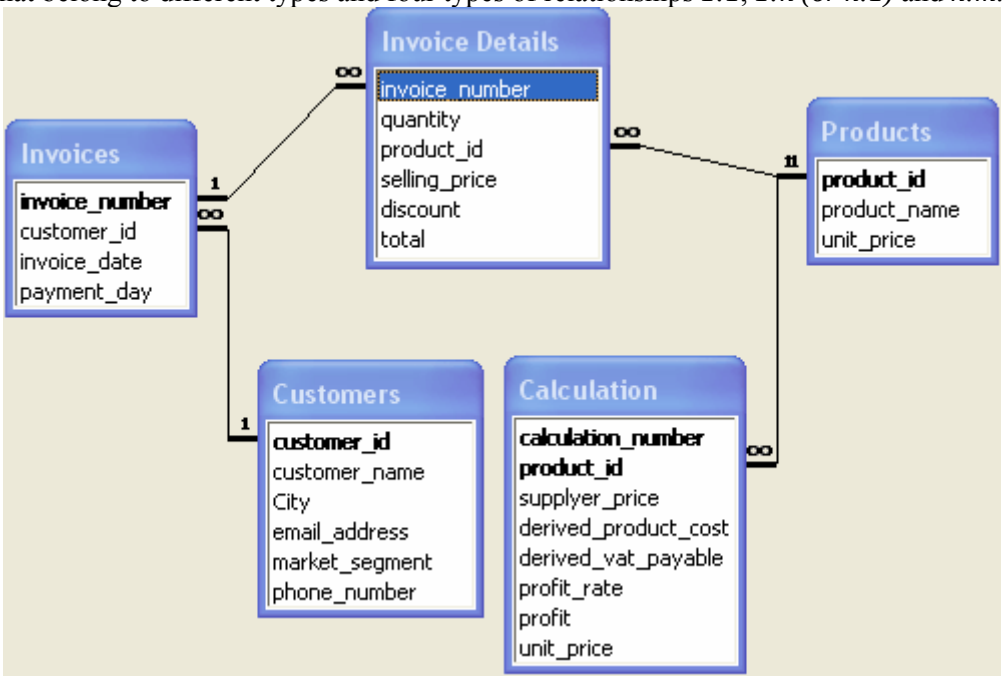

Figure 1. Relational data model

Relational data model relevant for building knowledge management system comprises six relational tables:
Invoce Details (invoce_number, product_id, quantity, selling_price, discount),
Invoices (invoce_number, customer_id, invoice_date, payment_day, total),
Customers (customer_id, customer_name, city, phone_number, market_segment, email_adress),
Products (Product_id, product_name, unit_pricer), Calculation(calculation_number, product_id, supplyer_price, derived_product_cost, derived_vat_payable, profit_rate, profit; unit_price).
Data warehouse is an attempt to integrate data from different and disintegrated operational, both from inside and outside of organization. As data is entered into the data warehouse, it is done in a way so that many inconsistencies at the application are undone. A data warehouse is a database management system that exists separate from the operations systems. Therefore our goal is to build data warehouse that meets the requirements of fuzzy c-means clustering as data mining algorithm.

2.1. A data set representation framework for database clustering

Now is necessary to adjust data model to format that fuzzy c-means clustering algorithm expects as its input.

An **object identification mechanism** [2] that defines what classes of objects will be clustered and how those objects will be uniquely identified. For creating the table adjusted for implementation fuzzy c-means algorithm is implemented adequate SQL statement.

Modular unit represents a particular perspective of the objects to be clustered. In the context of the relational data model modular units are defined as procedures that associate a bag of tuples with a given object. Using this framework, objects to be clustered are characterized by a set of bags of tuples, one bag for each modular unit [5]. It is obviously that data for customers clustering are stored in relational data warehouse that is temporarily loading from transactional data bases.

3. COLLABORATION KNOWLEDGE BASED SYSTEM AND THE FUZZY C-MEANS CLUSTERING IMPLEMENTATION RESULTS

Clustering is the process of grouping the data into classes (clusters) so that the data objects (examples) are similar to one another within the same cluster and dissimilar to the objects in other clusters. A *good clustering* method will produce high quality clusters with high *intra-class* similarity and low *inter-class* similarity. Hard k-means algorithm executes a sharp clustering, in which each object is either assigned to a cluster or not.[3]. Fuzzy c-means is an iterative algorithm. The aim of fuzzy c-means is to find cluster centers (centroids) that minimize a dissimilarity function.

The model of integration the results of c-means clustering and manager's experience and knowledge represents Figure 2. :

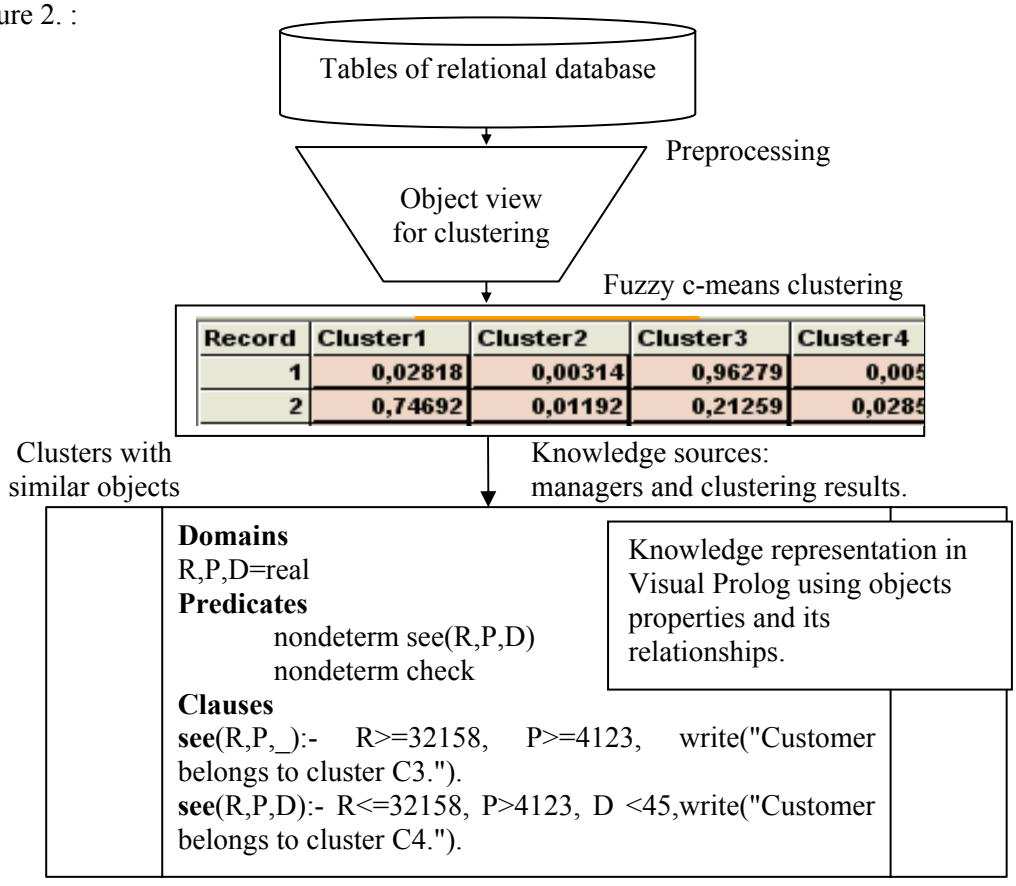

Figure 2. The conceptual model of collaboration knowledge based system and the fuzzy c-means clustering

Cluster centers (centroids) are initializing using randomly selecting records and fuzzy c-means does not ensure that it converges to an optimal solution. An knowledge based system is a computer program that contains stored knowledge and solves problems in a specific field in much the same way that a human expert would. The knowledge typically comes from a series of conversations between

the developer of the expert system and one or more experts. A peculiarity of knowledge based systems for interpretation the results of c-means clustering are that the knowledge comes from two sources. The first source is dimensions of centroids in clusters (dimensions are revenues, profit and days of payment) and the second source is manager.
The key components of knowledge based system for interpretation the results of customers clustering will be transformed into the statements (clauses) of Visual Prolog. The knowledge is represented by production rules: IF (condition) THEN (action).
Now is possible for each new customer, by consulting with expert system, find adequate cluster to which customer belongs. If the relationships among revenues, profit and days of payments are not acceptable for current market state then expert system will react and give managers adequate warning [4].
Now is clustering fully integrating with technology of expert systems whereas clustering algorithm (k-means) helps to determine the vector dimensions (centroids) which are the input information for building knowledge base of expert system. The output of clustering algorithms is input to expert system. This integration shows enormous application power.

4. CONLUSION

This paper clearly shows and realizes the collaboration among knowledge based systems and fuzzy c-means clustering. Fuzzy c-means clustering automatically clusters the database records into a number of groups and the results are the inputs into knowledge based system. Knowledge based system integrates managers knowledge and knowledge extracted by clustering. Clustering methodology is appropriate for the exploration of the interrelationships among samples and knowledge based system shows the strength and power for interpretation received results.

5. REFERENCES

[1] Dzeroski,S., Lavrac N.,eds., Telational Data ining, Springer, Berlin: Germany, 2001
[2] Han, J., Kamber,M., *Data Mining: Concepts and Techniques*, Morgan Kaufman, San Franciso,2000
[3] Markić, B., Tomić, D.; *Software solutions in marketing research for knowledge discovery in databases by fuzzy clustering*, Informatologija, Vol.39 No.4, str.240-244, Zagreb,2006 *Poslovna informatika*, Napredak Sarajevo, 2002.
[4] Markić, B., Tomić, D.; *Integrating cluster algorithms and expert systems,* The 8th International Conference " Modern Technologies in Manufacturing", Technical University of Cluj-Napoca, Romania, October 2005.
[5] Tae-Wan Ryu, Christoph F. Eick *A Database Clustering Methodology and Tool,* Department of Computer ScienceUniversity of Houston, *Information Science* in Spring 2005.
[6] Radovan, M., *Programiranje u Prologu*, Informator, Zagreb, 1991.
[7] Witten, I.H., Frank, E., Data Mining, Academic Press, 2000.