International Journal of Economics, Commerce and Management
United Kingdom Vol. IV, Issue 6, June 2016
http://ijecm.co.uk/ ISSN 2348 0386

MULTI-CRITERIA PRODUCTS CLASSIFICATION USING SQL SERVER

Yasser Al Hadad
Faculty of Cybernetics, Statistics and Economic Informatics, Department of Economic Informatics and Cybernetics, Bucharest University of Economic Studies, Romania
dukeyasser10@yahoo.com

Răzvan Daniel Zota
Faculty of Cybernetics, Statistics and Economic Informatics, Department of Economic Informatics and Cybernetics, Bucharest University of Economic Studies, Romania
zota@ase.ro

Abstract

Classification process is essential for a comprehensive understanding of any collection. Products classification process can be performed in many perspectives, especially when they are many properties to evaluate. Classification process cannot describe the dynamic of the products within the company based only on its attributes structure. Values of the attributes can determine better the chosen classification perspective. The collected values of every instance can play a big role by giving more weight for attributes that classify products better. There are many methods that can perform classification process. In this paper, we try to summarize and explore characteristics of products dataset belong to Romanian timber Export Company. Our dataset was analyzed and clustered in order to give a good classification that summarizes the essence of dataset and offers a brief view of timber export business branch at Romania for medium level companies. SQL server clustering algorithm was used to perform the classification process of our dataset.

Keywords: Classification Methods, Timber Products, Business Intelligence, Stock Management, Cluster Analysis, Learning Machine, Data Mining

INTRODUCTION

Companies have hundreds of products that it is difficult to manage and other factors that affect the performance of inventory control. A comprehensive understanding of products characteristics can increase stock management efficiency. Therefore, many solutions have been developed such as ABC analysis to solve inventory control problem. It separates the materials in several groups to ease the control of items (Partovi & Anandarajan, 2002). For inventory items, the classification has usually performed by single criterion. In real life, there are many factors should be considered in determining the inventory level of items such as sales volume, profit, lead time (Guvenir & Erel). Several solutions have been proposed for multi-criteria products classification. One of them is the analytic hierarchy process (AHP) (C.-W., 2008). The main idea of AHP is to evaluate the criteria by comparing each criteria to others criteria and then derives a single value of importance for each item. This method does not classify the items in groups. Other technique for classifying items is fuzzy C-means clustering method (FCM) (Gulsen & Coskun). FCM is a complex method that requires considerable effort calculation (Stancu, 2012), so it is better to avoid using the fuzzy systems where alternatives exist (Fuzzy logic toolbox, user's guide R2016a).

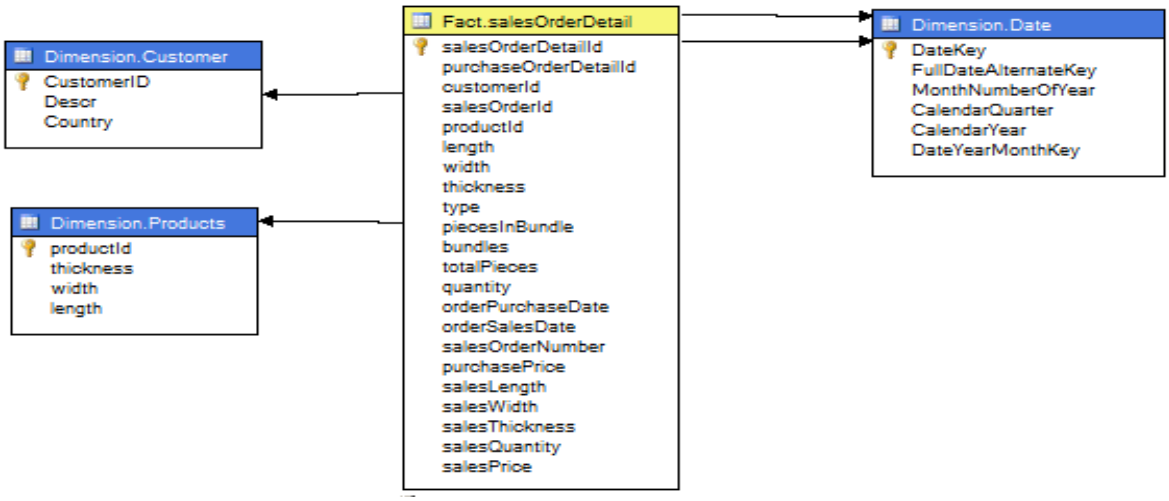
Multi-criteria classification problem can be solved by the means of unsupervised learning or clustering. The target of clustering is grouping the objects that are similar to each other in some way together in one group or cluster, then include the objects that seem or tend to be a part of this cluster. Many methods have been developed for unsupervised learning which are mainly based on the calculating the grade of similarity and divergence between the objects within a cluster such as hierarchical clustering, centroid-based clustering, distribution-based clustering, density-based clustering, etc (Tan, Steinbach, & Kumar, 2005). Many algorithms for distributed clustering have been created such as (Tasoulis & Vrahatis, 2004; Forman & Zhang, 2000) to adapt algorithms that rely on statistics. Hierarchical clustering was used by (VINTILĂ & GHERGHINA, 2014) for clustering companies belongs to European emerging countries. Core-sets have been used for distributed clustering as in (Zhang, Liu, & Wang, 2008).

In this paper, we try to explore characteristics of products dataset that belong to Romanian timber Export Company. At the field of timber export for medium company at Romania, there are no companies that utilize analytics systems. So there are not many researches that dedicate this business branch at local market. A case study of data set for one of Romanian timber Export Company is presented using SQL server to get the benefits from its powerful clustering algorithm implemented and its friendly user interface that is used for creating clustering analysis project.

Preparing Data for Clustering Analysis Project

The following section represents a case study for classifying products of timber Export Company at Romania. We use the business intelligence (BI) tools to perform our analytical task. In our case study, we try to get the benefit of using clustering method that is represent a part of methods implemented in any BI tool in order to classify our data set. All data related to the products are gathered from transactional database and archived data into data-warehouse using SSIS project (SQL server integration services). Then, a cube was modeled that includes all needed data for our proposed clustering model, see figure 1. It includes sales order data combined with purchase order at "Fact Sales Order Detail" table. The dimensions include date, customer and products. There are many properties or criteria that we can take in consideration for classification products. For this case, we choose three criteria that were considered the most important for classifying. The first criterion is total sales or turnover. It can be obtained by aggregating the total sales by products dimension. Also every instance at fact table is attached with its purchase order data. So for every row, we know both prices (purchase and sales). The profit of every instance can be obtained by getting the difference between both prices. The price is not only factor that determine the profit but also the dimensions of products. Every product is characterized by their dimensions (thickness, width and length). Real dimensions of timber products represent the measure noted at order plus an addition percent that represents the waste that can be lost later when the timbers get dry or processed. The company exports their products to many countries. Every destination has its standardized dimensions and its calculation method of waste percent. So these dimensions can be altered for some product in order to meet the required demands. So the final profit can be affected by these changed dimensions.

Figure 1: Sales Order Cube



Other criteria that we think it don't get enough concern is the storage period of products. In many companies, when their deposit is full, it easy to find any product type. The problem begin when the stock level begin to decrease and the most required products begin to end up first. So un adequate predictable method lead to that kind of inefficiency. Our goal is to engage that criteria in the classification process in order to a get a comprehensive understand of products involved in the classification.

Configuring Clustering Analysis Module

At this step, there are three ready attributes or criteria for clustering module (sales quantity, profit and storage period) which were obtained by the cube that was created in the previous section, see figure 1. First step at defining these criteria as input for clustering analysis module. Then normalizing step is followed to ensure the proper output and improve clustering results. It recommended especially for data that have a normal distribution. Normalizing process is aim to minimize the effect of input that might have many outliers and large magnitudes. In SQL Server 2008, configuration of clustering algorithm was changed to use z-score normalization by default. This feature can be disabled by add "NORMALIZATION" variable to parameter setting and set its value to 0.

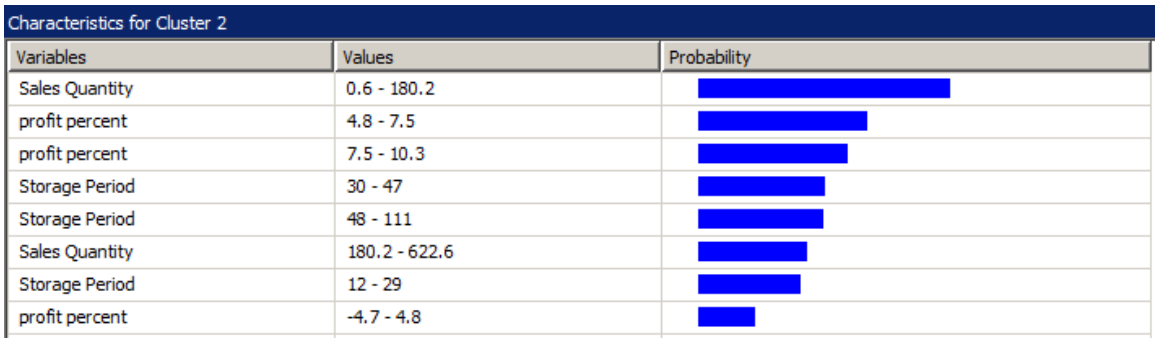
Microsoft clustering algorithm implements two clustering method k-means and EM(expectation maximization) and the second is the default one that is also used in our study case. Another variable that it can be setting is the number of clusters or numbers of groups that classify products, in our case; we set this variable to 3. Analysis model has grouped the products in three clusters. Figure 2, 3, 4 shows the most significant values that define the cluster combined with its relative probability presented by a bar that indicates how much in percent this range of values is a member of the current cluster. The unit used in sales quantity is cub meter, for storage day is days number and the value of profit criteria represents the profit percent.

Figure 2: Characteristics Diagram for Cluster 1

Characteristics for Cluster 1

Variables	Values	Probability
Sales Quantity	0.6 - 180.2	
Storage Period	12 - 29	
profit percent	-4.7 - 4.8	
profit percent	10.3 - 19.8	
Storage Period	30 - 47	
profit percent	4.8 - 7.5	
profit percent	7.5 - 10.3	
Storage Period	1 - 11	
Storage Period	48 - 111	

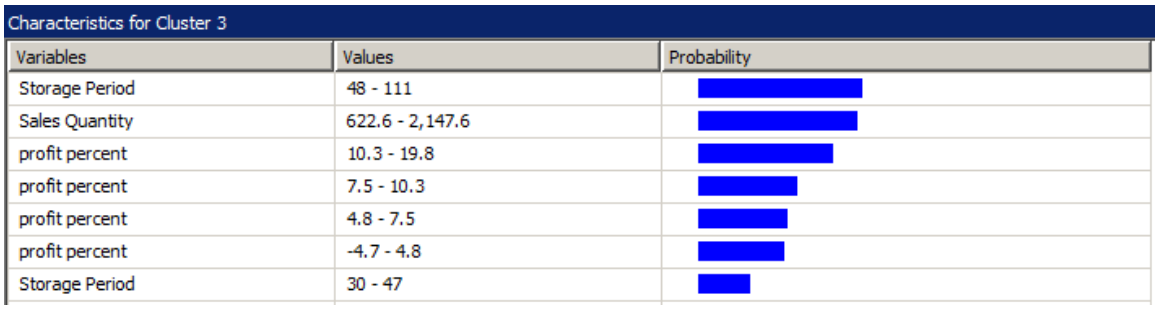
Figure 3: Characteristics Diagram for Cluster 2.



Characteristics for Cluster 2

Variables	Values	Probability
Sales Quantity	0.6 - 180.2	
profit percent	4.8 - 7.5	
profit percent	7.5 - 10.3	
Storage Period	30 - 47	
Storage Period	48 - 111	
Sales Quantity	180.2 - 622.6	
Storage Period	12 - 29	
profit percent	-4.7 - 4.8	

Figure 4: Characteristics Diagram for Cluster 3



Characteristics for Cluster 3

Variables	Values	Probability
Storage Period	48 - 111	
Sales Quantity	622.6 - 2,147.6	
profit percent	10.3 - 19.8	
profit percent	7.5 - 10.3	
profit percent	4.8 - 7.5	
profit percent	-4.7 - 4.8	
Storage Period	30 - 47	

DISCUSSION OF RESULT AND CONCLUSION

Analysis model has grouped the products in three clusters, see figure 2, 3, 4. The most weighted criteria that were chosen by clustering algorithm was sales quantity or turn over. We see in figure 4 that the most sold products were grouped in cluster 3. Cluster 1 includes products are less sold. Cluster 2 is near from cluster 1 but it includes also products that have an average sales between 180.2-622.6 mc(meter cube). Also storage period was chosen by analysis model to define clusters. We see in figure 4 that cluster 3 includes products that have a long storage period.

As a conclusion, we observe that the less important criteria was profit because there is no big differences between products in term of profit percent. Also we expect that storage period is long just for products that are not sold in large amount, but results shows that cluster 3 with best sold products has also the longest storage period. That means that the stock management is efficient or the process of preparing sales orders takes a long time until the entire order is collected from suppliers. The most important criteria is sales quantity as it appears in the top of each characteristics diagram. So we can conclude that sales quantity attribute has a normal distribution rather than an uniform distribution, so the system has rely on it mostly to define clusters.

ACKNOWLEDGMENT

Part of this work is done under the auspices of the doctoral studies within the Doctoral School of Economic Informatics, Bucharest University of Economic Studies.

REFERENCES

C.-W. C. (2008). Controlling inventory by combining ABC analysis and fuzzy classification. *Computers and Industrial Engineering , 55*, 841–851.

Forman, G., & Zhang, B. (2000). Distributed data clustering can be efficient and exact. *SIGKDD Explor. Newsl.* , 34-38.

Fuzzy logic toolbox, user's guide R2016a. The MathWorks.

G. A., & C. O. Multiple Criteria ABC Analysis with FCM Clustering. *Journal of Industrial Engineering, 2013*.

Guvenir, H. A., & Erel, E. Multicriteria inventory classification using a genetic algorithm. *European Journal of Operational , 105* (1), 29–37.

Partovi, F. Y., & Anandarajan, M. (2002). Classifying inventory using an artificial neural network approach. *Computers & Industrial Engineering* , 389-404.

Stancu, S. (2012). *Calcul neuronal - Teorie si aplicatii economice.* (e. ASE, Ed.) Bucharest.

Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining, (First Edition)* (Vol. 41). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.

Tasoulis, D. K., & Vrahatis, M. N. (2004). Unsupervised distributed clustering. *in Proceedings of the International Conference on Parallel and Distributed Computing and Networks.*

VINTILĂ, G., & GHERGHINA, Ş. C. (2014). Pattern recognition techniques to classify the european emerging markets companies from the valuation perspective. *ECECSR Journal , 48*.

Zhang, Q., Liu, J., & Wang, W. (2008). Approximate Clustering on Distributed Data Streams. *Proceeding ICDE '08 Proceedings of the 2008 IEEE 24th International Conference on Data Engineering* (pp. 1131-1139). Washington, DC, USA: IEEE Computer Society.