RESEARCH IN LOGISTICS & PRODUCTION

THE USE OF CLUSTER ANALYSIS AND THE THEORY OF MATHEMATICAL RECORDS IN THE PROCESS OF PLANNING THE PRODUCTION-WAREHOUSE FLOW

Andrzej Bujak*, Katarzyna Topolska** and Mariusz Topolski***

WSB University in Wrocław, Fabryczna 29/31, 53-609 Wrocław, Poland
*Email: andrzej.bujak@wsb.wroclaw.pl
**Email: katarzyna.topolska@wsb.wroclaw.pl
***Email: mariusz.topolski@wsb.wroclaw.pl

Abstract The paper proposes a new approach to the agglomeration of data in cluster analysis. These approach faces the problem of data classification where under the same conditions different conclusions are draw. Such problems occur in many areas of daily life: medicine, technology, and economic and social sciences and economics. The new approach assumes that events like the harvest are attributed to the cumulative probability of their occurrence at the same time. Such approaches will not be found in probability. Thanks to the mathematical theory of records fairly accurate classification of the object can be provided. The paper presents a method of agglomeration of measurement data using the mathematical theory of records. This is the method which can be used in the cluster analysis by agglomeration.

Paper type: Research Paper

Published online: 16 July 2016
Vol. 6, No. 3, pp. 259-267
DOI: 10.21008/j.2083-4950.2016.6.3.6

ISSN 2083-4942 (Print)
ISSN 2083-4950 (Online)

Keywords: *cluster analysis, warehousing, logistics flows*

1. INTRODUCTION

A contemporary economy is based on knowledge. Each day new algorithms are created as well as methods aiming to support the solution of decision problems. Due to high complexity of the area of classified objects, the use is computer is necessary to make necessary calculations. One of the biggest problems which researchers face is the elaboration of methods which will precisely classify objects under uncertain conditions and in view of incomplete information. However, it is limited to estimation of probability of one incident and not a group of incidents happening at one time. Often in everyday life we deal with situations when various decisions have to be made on the basis of premises. This problem may be dealt with by the theory of mathematical records. Whereas the problem of classifying similar objects may be solved by the cluster analysis. The second point of this paper is devoted to this analysis. The research aims to look for a computer algorithm assigning an object to a certain class under uncertain conditions and uncertainty of measure information in particular with the same premises but different decisions.

2. CLUSTER ANALYSIS

One of the problem contemporary researchers face is organization and grouping data in sensible structures. The term cluster analysis was introduced by Tryon in 1939. The cluster analysis is also called as an exploratory data analysis aiming to assign data to specific structures of highest similarities inside (Kurzyński, Woźniak, & Żołnierek, 2007). At the same time combining objects with other objects should be the smallest. The above classification is very often used for classifying animals into specific species, transmitting sources of contamination, evaluating characteristic features of damages to mechanical elements etc. Apart from the cluster analysis used for evaluating the number of factors affecting the measurement of a specific feature, methods of factor analysis or multiple regression are used. These methods are predominantly used in the field of technical sciences.

We apply cluster analysis when we want to distinguish groups of similar objects if the objects are described by more than one feature.

Cluster analysis is a method distinguishing clusters of similar objects when new objects are described by more than one feature. The application of cluster analysis is very comprehensive. It is used by search engines for creating thematic groups, in psychology and medicine for aggregating psychological and somatic symptoms. So clustering or a group of a class shall be used to describe a set of objects where a similarity between a pair of objects is greater than a similarity of any object belonging to a class and an object not belonging to it (Gatnar & Walesiak, 2004). There are numerous methods describing distances between clusters i.e. the nearest neighbour, the furthest neighbourhood, a median, a group mean, a centre of gravity or Ward (Grabiński, 1988).

In cluster analysis methods of clustering are very important and are divided into two categories: hierarchical and non-hierarchical. The hierarchical method groups objects in a iterative way into smaller and smaller clusters. In non-hierarchical models you transfer objects between clusters looking for the best solution proceeding according to the established criteria. The most common cluster methods are an agglomeration method and K-mean method. Hierarchical models create a hierarchy of classifications for a set of objects starting with the set in which every object creates an independent cluster and ending with a division of objects making up a cluster. In cluster analysis the following agglomeration methods may be differentiated:

- single-linkage clustering,
- complete-linkage clustering,
- average linkage clustering,
- weighted average clustering,
- centroid linkage,
- medians,
- Ward's methods.

Distance measurements are as follows:

- Euclidean,
- city,
- Chebychev,
- power,
- percent disagreement,
- 1-r Pearson's.

3. ELEMENTS OF THE THEORY OF MATHEMATICAL RECORDS

The theory of mathematical records is also called the theory of belief functions and Dempster Shafer. The theory allows us to create models of uncertainty dealing with accepting numerous values of a particular attribute at one time. It allows us to treat plausibility in a subjective way.

When classifying an object on the basis of features $x \in \boldsymbol{X}$, it is classified to a certain class $j \in \Theta$, where $\Theta = \{1,2,..,M\}$, and M is the number of classes. Certain prior knowledge about classes is represented by a prior function of probability allocation (Stanisz, 2007).

DEFINITION 1. A prior function of probability allocation, determined on the basis of subjective judgment (e.g. a learning set) corresponds to prior probability of classes and can be written down as follows (Topolski, 2008):

$$\sum m(\Theta) = 1 \qquad (1)$$

$$m(\Theta) > 0$$

Where set Θ is a focal element for Bel belief function meeting condition $m(\Theta) > 0$. In the Beys classifier knowledge is represented by conditional and unconditional probability. However, in reasoning by means of the theory of mathematical records we assign sentences with *Bel* value being a belief degree. The probability allocation function (1) is a component of the *Bel* belief function according to the theory of Dempster-Shafer. The *Bel* belief function is the basis for reasoning and may be classified as follows:

DEFINITION 2. A belief function for a certain fuzzy set *Bel(Y)* is referred to the function which results from base distribution function of probability allocation a priori class $m(\Theta)$ with the function of probability allocation *m(Y*)* of the fuzzy set *Y** satisfying the dependency (Topolski, 2008):

$$Bel(Y) = \sum_{Y^{**} \in Y} m\,(Y^{**}) = \sum_{Y^{**} \in Y} m\,(\Theta) \oplus m(Y^{*}) \qquad (2)$$

$$\sum_{Y^{**} \in Y} m\,(\Theta) \oplus m(Y^{*}) = \sum_{Y^{**} \in Y} \frac{\sum_{\Theta \cap Y^{*} = Y^{**}} \quad m(\Theta) \cdot m(Y^{*})}{1 - \sum_{\Theta \cap Y^{*} = \emptyset} m\,(\Theta) \cdot m(Y^{*})} \qquad (3)$$

where *Y*, *Y**, *Y*** = Θ for Beys belief function. In definition 2 there is formula 3 which is called the rule of Dempster's combination. This function enables to make independent convictions and updates.

4. CASE STUDY OF A WAREHOUSING PROCESS – THE MODEL OF CLUSTER ANALYSIS MERGER AND THE THEORY OF MATHEMATICAL RECORDS

In warehousing we have to deal with four main processes such as receiving, storing, picking and releasing. When accepting, materials or goods ordered by an enterprise should be checked for the quantity and quality before recording them into stocks. By the acceptance we understand activities such as:

- accepting the shipment from a carrier or a supplier (unpacking, counting, weighing or measuring and comparing the quantity with shipping documents from the supplier),
- checking if the shipment is not damaged or if it does not have any defects caused during transportation or by the supplier,
- checking the quality of the shipment (Dudziński, 2011).

Goods and materials accepted to a warehouse must come from external deliveries or internal supplies of an enterprise. External deliveries include goods and materials purchased in a different enterprise, returned as unaccepted by a recipient or from external processing. Internal supplies of an enterprise comprise materials not used by production departments, returned, transferred between warehouses, deposits accepted for keeping or finished goods received by warehouses from production departments (Dudziński, 2011).

Regardless of the way of accepting materials and goods coming into a warehouse, the warehouse should be notified about kinds of shipments being delivered to it. It enables people to organize a place in a warehouse necessary for receiving the cargo. The following departments cooperate with the warehouse when accepting goods:

- procurement departments checking compliance of the shipment with the order,
- transport and loading teams,
- a quality team assessing the quality of delivered materials,
- a team of warehouse employees unpacking and putting goods into designated places (Dudziński, 2011).

The model proposed below may be applied in a task of detecting association rules for the purposes of establishing what products customers most often buy together. The model may be used to evaluate the sequence of occurring events e.g. the order of places visited by tourists. The method may also be applied to the analysis of customers' data and to infer about possibilities of doing shopping.

When classifying clusters on the basis of object features $\boldsymbol{x} \in \boldsymbol{X}$ an object is classified to a certain class $j \in \Theta$ (cluster) where $\Theta = \{1,2,..,M\}$, and M is the number of classes (clusters). Certain knowledge about probabilities a priori of classes is represented by a prior probability allocation function:

Using the method of the nearest neighbour according to the theory of cluster analysis as a distance (between two objects x_i *and* x_k) we are going to use the Euclidean distance expressed by the formulas:

$$d(x_i, x_k) = d_{ik} = \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{kj})^2} \qquad (4)$$

where x_{ij} – the value of object x_i in terms of a feature j, whereas p – is the number of these features.

Having a matrix n of objects and p of variables we design a matrix of distances between particular objects:

$$D = [d_{ik}],\ i,k=1,..n \tag{5}$$

Matrix D is used to calculate the total of all distances.

$$d = \sum_{d_{ik}} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{kj})^2} \tag{6}$$

The function of probability allocation for each distance is established on the basis of the formula:

$$m(\{d_{ik}\}) = \frac{d_{ik}}{d} \tag{7}$$

Meeting the condition $\sum m(\{d_{ik}\}) = 1$ and $m(\varnothing) = 0$.

Recognition of similar objects is an essential element of classification. When analysing object 6, we compare its probability with 5 other objects. We might have a situation that the classified object in its vicinity has 3 objects from cluster 1 and two from cluster 2. When the difference between clusters is 1, the classification is flawed. Thus the distance between the two points is modified to the set:

$$m(\{d_{ik1}, d_{ik2}, ..., d_{ikN}\}) = \frac{\sum_{ik} (d_{ik1}, d_{ik2}, ..., d_{ikN})}{d} \tag{8}$$

The next step modifies each element of matrix d using the rule:

Finally the affiliation to the cluster is established by maximizing a belief function with the formula:

$$Bel(Y) = \min\left[\sum_{Y \in d} m(\Theta) \oplus m(d)\right] = \min\left[\sum_{Y \in d} \frac{\sum_{\Theta \in d = Y} m(\Theta) \cdot m(d)}{1 - \sum_{\Theta \in d = \varnothing} m(\Theta) \cdot m(d)}\right] \tag{9}$$

Storing material goods in a warehouse is related to a basic function of the warehouse namely temporary storage. Storing is a set of activities preceded by receiving goods to the warehouse. Material goods, usually in the form of loadable units, are put in a systematic way in the warehouse. For the purpose of the most effective use of warehousing space special machines are used (pallet-trucks, fork lifts, cranes and other equipment adjusted to load units and ways of storing materials) (Fertsch, 2006).

Methods of storing martials depend on the type of a warehousing building, its surface and its layout as well as other factors. Goods may be for instance stored in a designated place, stacked in multiple layers depending on the specificity of a load unit. The most frequent method of storing materials is a high-bay warehouse (mainly closed warehouses). It reduces the amount of space occupied by a warehouse and is quite comfortable as each load unit has a location of its own.

Storage conditions are essential parameters characterizing each stored material. The parameters include a temperature, humidity, air cleanness, requirements for limited access or fire regulations. It is vital to put each material in a place in a warehouse to secure appropriate storing conditions (Fertsch, 2006).

The next activity involves picking which is one of the basic stages in a warehousing process compiling stored items according to external requirements made in line with customers' orders. Due to the fact that order picking involves making decisions about moving the assortment line from a supplier to a recipient, it generates the highest execution costs as it is threatened with a possibility of mistakes and faults. This situation makes us perceive picking as a critical element of entire warehousing and if optimized may improve effectiveness of the entire process – which may be achieved from the process as well infrastructural perspective. The other aspect refers to a technical side of the issue and highly automated technologies used for picking. Automation of the picking process is characteristic for the so called dynamic variant where stored assortment is moved from the storing area (by means of automated trucks, cranes, carriers) to special picking stands where it is picked up by the staff and put on the ordered load unit. Depending on the storage, the assortment is delivered to stands in single packages or entire units where it is stored. In the latter case a load unit returns to the storage place after the right number of materials is taken. Because of the way goods are being picked it is called goods-to-man picking system.

Finally, goods are released which is the last warehousing stage. It involves taking the right number of goods from their storage place, checking their weight and type requirements with an issuance order, picking and packaging and transferring them to an authorized person (a driver, a carrier or a customer's representative) (Dudziński, 2011).

The basis for releasing goods form the warehouse is constituted by an order issued by relevant organizational units in charge of the stock. These orders are in the form of goods issued notes (issued internally), limit cards (issued mainly by production companies applying the rule of limiting issued materials), goods dispatched note (releasing goods externally e.g. stales), “orders and invoices” (used by commercial companies mainly instead goods dispatched notes), warehouse transfer sheets (Dudziński, 2011).

Most released objects are load units which originate from picking. Others are load units which are released as they were delivered to the warehouse or as they were stored. When necessary, materials are packaged and formed into transport units. Then individual requirements are agreed with the customer or related to expected transport conditions. Materials ready for release must be checked. In case of external orders’ fulfilment, the result of the check-up leads to loading them into an external transport means. Similarly to goods receiving, an internal transport means is used and load handling devices. If goods are released internally then the internal transportations means is used (Fertsch, 2006).

A warehouse manager should remember about the principle of releasing goods from the warehouse. Goods which came first should first go (FIFO). He should receive an order for releasing goods in advance so that he had time for preparing the order. It shortens waiting for the goods and eliminates delays. When releasing the goods the warehouse manager should carefully check the number, the type and the quality of released goods. He cannot sell things or accept money (Dudziński, 2011).

Analysing the warehousing process in a real enterprise which operates on the basis of Total Quality Management System (TQM). TQM in an enterprise combines traditional values with effective planning and quality check in eight stages defined by PDCA techniques.

At present goods are stored according to two guidelines. The first one refers to the origin of goods. Does the component come from the European Union member state, Asia or Turkey? The second one refers to information about which factory goods got to. Each factory has its individual suppliers so stored goods cannot be mixed up with goods from other factories and each has a designated space. In this case a specific rack and lines.

Table 1 Advantages of using cluster analysis and the theory of mathematical records in a warehousing process; Source: Elaboration of one’s own

Warehousing process	Advantages of changing guidelines for storing goods
Receiving	Optimized process of receiving deliveries, Freedom in configuring accepted deliveries, Enhanced capacities of accepted pallets,
Picking	Faster order picking (especially in case of fast rotating goods being put in front racks),
Storing	Faster location of pallets in racks, Safe placement of goods in high bay warehouse,
Releasing	Visible distinction of goods in the release area, Faster location of particular goods in the release area.

Thanks to the use of cluster analysis and the theory of mathematical records in the process of planning warehouse flows, wastage may be eliminated and the flow of materials ordered throughout all warehouse activities. Additionally a periodic ABC analysis may be applied and fast rotating goods may be put in areas closest to the shipment. The table 1 presents advantages of using cluster analysis and the theory of mathematical records for the purposes of facilitating the entire warehousing process.

5. CONCLUSIONS

Very often researchers face the problem of classifying objects to specific groups. One of the method solving this problem is cluster analysis. The paper has presented a method of agglomerating measure data by means of the theory of mathematical records. This is the method which can be applied to the cluster analysis using the agglomeration method. Thanks to the application of the merger of two classifiers: kNN algorithm (k nearest neighbours) and a belief function, a model has been drawn up which seems to discriminate the space of objects. The proposed model may be used for classifying goods in a warehouse or for analysing similar spaces of the data.

REFERENCES

Dudziński Z., (2011), Vademecum organizacji gospodarki magazynowej, Wyd. ODDK, Gdańsk.

Gatnar E. & M. Walesiak, (2004), Metody statystycznej analizy wielowymiarowej w badaniach marketingowych, Wrocław, Wyd. Akademii Ekonomicznej.

Grabiński T., (1988), Metody taksonometrii, Kraków, Wyd. Akademii Ekonomicznej.

Kurzyński M., Woźniak M., & Żołnierek A., (2007), Analiza przydatności wybranych metod rozpoznawania sekwencyjnego dla problemu z zakresu diagnostyki medycznej, Inteligentne Wydobywanie Informacji, Technologie Informacyjne: Diagnostyka. PWNT Gdańsk.

Stanisz A., (2007), Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny, Tom 2. Modele liniowe i nieliniowe, Kraków.

Topolski M., (2008), Komputerowe algorytmy rozpoznawania sekwencyjnego łączące teorię zbio-rów rozmytych z teorią ewidencji matematycznej, Raport PRE 1/08, Politechnika Wrocławska.

Fertsch M. (ed.), Podstawy logistyki, Instytut Logistyki i Magazynowania, Poznań 2006.