www.czasopisma.pan.pl www.journals.pan.pl

Management and Production Engineering Review
Volume 7 • Number 2 • June 2016 • pp. 29–33
DOI: 10.1515/mper-2016-0014

APPLICATION OF EXPECTATION MAXIMIZATION METHOD FOR PURCHASE DECISION-MAKING SUPPORT IN WELDING BRANCH

Agnieszka Kujawińska, Michał Rogalewicz, Magdalena Diering

Poznan University of Technology, Chair of Management and Production Engineering, Poland

Corresponding author:
Michał Rogalewicz
Poznan University of Technology
Chair of Management and Production Engineering
Piotrowo 3, 61-255 Poznań, Poland
phone: (+48) 61 665-27-98
e-mail: michal.rogalewicz@put.poznan.pl

Received: 30 March 2016
Accepted: 1 June 2016

Abstract
The article presents a study of applying the proposed method of cluster analysis to support purchasing decisions in the welding industry. The authors analyze the usefulness of the non-hierarchical method, Expectation Maximization (EM), in the selection of material (212 combinations of flux and wire melt) for the SAW (Submerged Arc Welding) method process. The proposed approach to cluster analysis is proved as useful in supporting purchase decisions.

Keywords
cluster analysis, welding process, expectation maximization method.

Introduction

In industrial companies, as well as in the metallurgical industry, decision-making processes in terms of production preparation are a fundamental and key element of the whole manufacturing process. Making a choice about a specific decision variant is preceded by analysis based mostly on the knowledge and intuition of the decision-maker. One of the processes related to production preparation is the purchase of materials necessary for its realization. The problem that responsible persons often encounter is the too large number of commercially available possibilities. That is why selecting the optimal material, considering costs and usability criteria is often difficult. In such a context, selecting the resources necessary for the manufacturing process can be considered a multi-criteria decision problem [1]. An example of such a problem is making decisions about the purchase of so-called additional materials (here: fluxes) for the Submerged Arc Welding (SAW) method.

A high competition and a diversified assortment on the fluxes market, as well as the various criteria of their selection (e.g. Boniszewski basic index value, weld mechanical properties, chemical composition and price) are reasons why an engineer planning the purchase of additional materials for welding processes should not use solely his intuition and experience. It may not lead to obtaining the most beneficial variant.

That is why in such a situation it is worth using decision-making support methods. One of the possible methods is cluster analysis, which belongs to the wide range of Data Mining tools (discovering knowledge from data) [2–5]. It realizes the task of grouping (clustering) objects into homogeneous groups based on their features [6]. The possibility of applying cluster analysis in the flux selection problem is investigated by the authors of this paper. Another group of methods worth mentioning in this contex are intelligent methods based on incomplete information processing (eg. [7]). Cluster analysis methods can be divided into hierarchical [8] and non-hierarchical ones [9]. The first group contains algorithms which can be applied to obtain so-called dendrograms, representing a structure of hierarchically ordered clusters. Within its scope, agglomerative and dividing methods can be distinguished. The first one build a tree of connections, starting from single objects, putting the most similar ones into larger clusters until one large cluster is obtained. The second one work in an opposite manner: they start from the full group of objects and divide them until reaching a level of single objects.

The second group of cluster analysis methods contains non-hierarchical methods. In this case as a result no relation clusters structure is obtained. In most cases that methods need to know a priori a number of clusters which will be obtained as result of the analysis. The most frequently used algorithm in the scope of this approach is the k-means algorithm. As a result of that algorithm, the observations (objects) are assigned to a set of clusters in such a way as to maximize the distances (minimalize similarities) between objects in a multidimensional space. Another approach to the non-hierarchical clustering is the Expectation Maximization (EM) method [10]. It is the usability of the EM method in decision-making support regarding the selection of additional materials for the SAW process that the authors of this paper decided to study.

The EM method assumes that a data set is described by a general distribution, which is a mixture of several probability distributions, separate for each cluster. The fundamental algorithm of this method approximates the distribution in such a way that maximizes the general probability for a given cluster's division. A distinctive feature of this approach is that no unequivocal belonging to a given cluster is determined. Instead, the probability of its membership to each cluster is defined.

Data acquisition

Acquisition of the data used for the analysis was conducted in several stages. In the first one, five producers offering fluxes on the Polish market (who agreed on using data from their commercial offers) were selected. In the next stage, the description and fluxes selection criteria were determined. The following criteria were used: the Boniszewski basic index (WZB), properties of the weld formed after the process (element contents – carbon, silicon, manganese, molybdenum, chromium, nickel; strength, yield point and elongation) as well as price. It is worth mentioning that the weld properties are affected not only by the flux, but also by the welding wire used in the welding process as well. It is because there is a strong relation between these two additional materials for the SAW method. Using the same flux with various wires results in obtaining welds of different mechanical and chemical properties. That is why in the last stage of data acquisition, the properties of the fluxes and welds were analyzed with distinguishing the welding wire used in the process. As a consequence, the data set (possible variants) contained 274 records. Next data cleaning was performed. It is very important in each Data Mining project [11], especially in cluster analysis. After studying the discriminating capability of the diagnostic features (all features were discovered to be important), data gaps and tensile strength (Rm) highly correlated ($r = 0.8$) to yield strength (Re) were removed. Then standardization of the features was performed. As a result of these actions, the final data set was obtained, it consists on 212 records (objects). A single object, as mentioned before, is described in a combination of flux and weld wire, as well as weld properties after the welding process and the 10 diagnostic features describing them.

Experimental work

The research was conducted in phases. In the first stage, a hierarchical clustering methods was deployed – the Ward method with Euclidean distance – to determine the most appropriate number of clusters for the analyzed problem. This made it possible to build a tree diagram – dendrogram (Fig. 1), showing a hierarchical structure of objects in the function of decreasing similarity (increasing distance) among them. Drawing a horizontal line in the diagram enables one to select ("cut off") the number of clusters which best describes a given dataset. Analysis of the diagram reveals that division into 3–5 groups will be the most appropriate. The authors decided to use four clusters ($k = 4$) and to apply this number as a parameter for further analysis with the se-

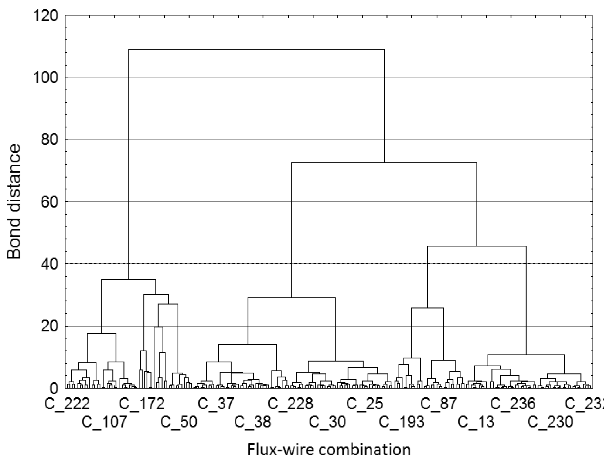

Fig. 1. Tree diagram of agglomeration method results using Ward method.

lected non-hierarchical method – Expectation Maximization. As emphasized before, this method is aimed at supporting decision-making by a SAW process engineer.

As part of building a model using the EM method, the authors decided to confirm the correctness of selecting 4 clusters. In order to do this, a 10 – fold cross validation test was used. Its results, illustrated in Fig. 2 using an earlier defined cost curve, also indicate that division into 4 clusters is an apt choice (it is clearly visible that the decrease in cost after changing from 4 to 5 clusters is no longer relevant). The next stage was analysis of the results obtained using the EM method.

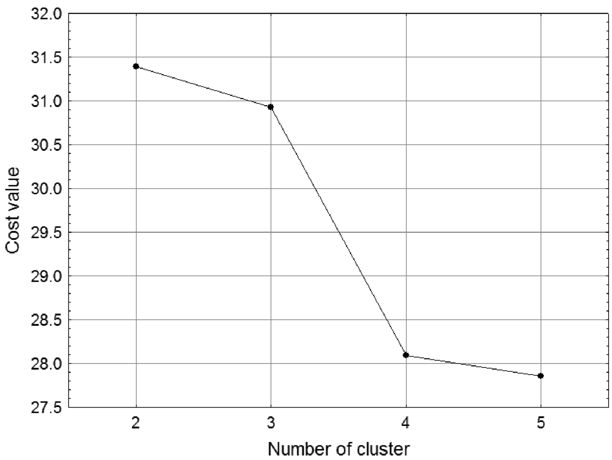

Fig. 2. Cost sequence diagram – result of 10 – fold cross validation

The results of the variance analysis conducted in the scope of the EM method (Table 1) indicate that each of the selected diagnostic features significantly differentiates among the obtained clusters (the p values for each of them are definitely lower than commonly accepted significance levels). It confirms the analysis results of the discriminating capability of the features that describe a flux – wire combination and ensures the reliability of their choice. In Table 2, the average values for each cluster obtained using the EM method are presented (the algorithm assigned a given object to a cluster with the highest affiliation probability).

Table 1
Results of variance analysis. Designations: SS_{inter} – SS (sum of squares) between clusters; df – degree of freedom, SS_{in} – SS inside clusters; F – test statistic, p – p-value

	SS_{inter}	df	SS_{in}	df	F	p
WZB	31.77	3	132.21	208	16.66	0.000
C [%]	0.04	3	0.27	208	11.20	0.000
Si [%]	0.75	3	9.90	208	5.29	0.002
Mn [%]	14.89	3	161.46	208	6.39	0.000
Mo / %	181.75	3	340.87	208	36.97	0.000
Cr [%]	17861.46	3	484.43	208	2556.40	0.000
Ni [%]	13771.18	3	14497.59	208	65.86	0.000
Re [N/mm^2]	209591.92	3	1345373.53	208	10.80	0.000
A5 [%]	3124.44	3	3471.45	208	62.40	0.000
Price [PLN/kg]	912.63	3	1525.33	208	41.48	0

Table 2
Average values for 4 clusters obtained by k-means method. Designations: WZB – Boniszewski basic index; C – carbon; Si – silicon; Mn – manganese; Cr – chromium; Ni – nickel; Re – yield point; A5 – elongation; N – number of cases; PN – percentage of cases.

Cluster no.	1	2	3	4
WZB	2.09	1.638	2.784	1.710
C [%]	0.032	0.043	0.078	0.063
Si/ %	0.29	0.439	0.296	0.419
Mn [%]	1.397	1.991	1.092	1.318
Mo [%]	1.729	2.6	0.487	0.16
Cr [%]	20.032	20.563	0.748	0.088
Ni [%]	10.055	22.378	1.137	0.083
Re [N/mm^2]	445.968	433.594	534.677	451.61
A5 [%]	29.258	34.906	22.935	25.025
Price [PLN/kg]	10.119	15.837	11.075	9.912
N	31	32	31	118
PN [%]	14.62	15.09	14.62	55.66

The diagram of averages, shown in Fig. 3, also confirms the conclusions from the variance analysis – for each diagnostic feature, the average value of at least one cluster significantly differs from the others (see Table 1). To make the average values for par-

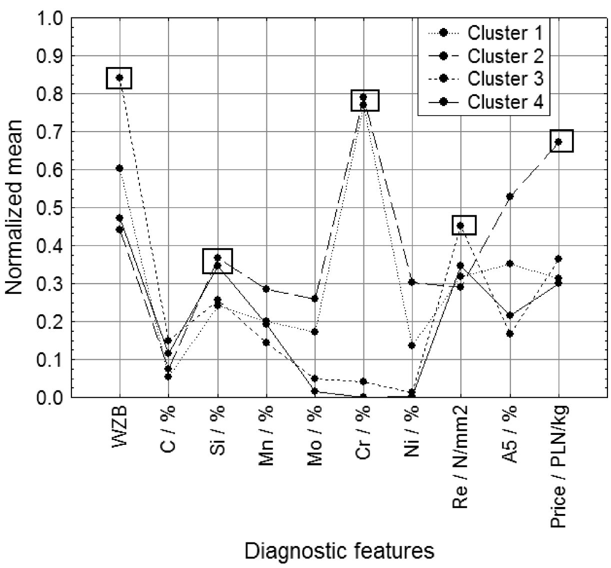

Fig. 3. Diagram of average values of particular features for constructed clusters

Table 3
Results of SAW flux clustering on basis of features.

Cluster No., elements	Cluster description
Cluster 1 101-105, 113-117, 119, 179-185, 193, 197-200, 202-203, 221-224, 226-227	Cluster 1 is a group of inert fluxes with a low melting temperature. This cluster is characterized by low contents of carbon in the weld, high contents of silicon, manganese and chromium. Their application enables one to obtain very good welding properties, but they leave more inclusions, which decrease the possible impact strength. It is the group with a relatively high price per kg.
Cluster 2 41-46, 48-53, 106-112, 118, 120, 172-173, 189-192, 194-196, 201, 225	The second cluster is a group of basic fluxes characterized by low silicon content, high contents of molybdenum, manganese, chromium and nickel. That is why they are recommended mostly for welding high alloyed steels. Cluster 2 fluxes are of low yield strength and high elongation. They are more expensive than the fluxes from the 3rd and 4th cluster, but cheaper than those from cluster 1.
Cluster 3 5-6, 10-12, 14-16, 27, 84, 86-88, 96, 187-188, 235-236, 239-240, 246-249, 252-253, 256-258, 263, 266	Cluster 3 is a group of high-basic clusters (WZB $>$ 2.0). Therefore, they have high melting temperatures. This group of fluxes, with proper welding procedures, enables one to obtain a high melt purity and high impact strength value at low temperatures. The fluxes from this cluster are also characterized by a high ductility value of the welded joint, the lowest elongation and relatively high amount of carbon. The contents of manganese, molybdenum, chromium and nickel are insignificant. Their price is on an average level – this group was ranked 3rd by this criterion.
Cluster 4 1-4, 7-9, 13, 17-26, 28-40, 70-83, 85, 89-95, 97-100, 166-171, 174-178, 186, 204-220, 228-234, 237-238, 241-245, 250-251, 254-255, 259-262, 264-265, 267-274	Cluster 4 is a group of acidic fluxes with a low Boniszewski index value. This cluster is characterized by high contents of carbon and silicon in the weld and high yield strength with low elongation. The weld formed after the process contains numerous inclusions lowering the impact strength. It is the cheapest group of fluxes.

ticular clusters comparable in the range of all the analyzed features, the Y axis was properly scaled, while the average values were shown using standardized values. Means with particularly different values were designated using rectangles.

In Table 3, the flux – wire combinations belonging to particular clusters are presented. Such a juxtaposition will make it possible for a process engineer to select an appropriate variant among significantly different clusters and then select a combination inside one cluster which will best suit his expectations, e.g. regarding price.

Conclusions

The paper presents usefulness of the Expectation Maximization method to support decisions made by the engineer responsible for the purchase of additional materials for the SAW process. On the basis of the obtained analysis results, one can select an appropriate combination of flux and welding wire among four isolated, different clusters. An unquestionable advantage of this method is assigning each of the 212 records a probability of affiliation to all the four clusters. That is why it is possible to select cases which are ambiguous or do not fully fit inside just a single cluster.

Further studies by the authors will be focused on comparing the analysis results obtained by means of the EM method with other approaches to cluster analysis.

The presented results derive from scientific statutory research conducted by Poznan University of Technology, Poland, supported by the Polish Ministry of Science and Higher Education from financial means in 2015–2016.

References

[1] Chai J., Liu J., Ngai E., *Application of decision-making techniques in supplier selection: A systematic review of literature*, Exp. Sys. with Appl., 40, 10, 3872–3885, 2013.

[2] Rogalewicz M., Kujawińska A., Piłacińska M., *Selection of data mining method for multidimensional evaluation of the manufacturing process state*, Man. and Prod. Eng. Review, 3, 2, 27–35, 2012.

[3] Cao L., *Domain-driven Data Mining: challenges and prospects*, Know. and Data Eng., IEEE Trans., 22, 6, 755–769, 2010.

[4] Janćikova Z., Roubićek V., Juchelkova D., *Application of artificial intelligence methods for prediction of steel mechanical properties*, Metalurgija, 47, 4, 339–342, 2008.

[5] Grudzień Ł., Hamrol A., *Information quality in design process documentation of quality management systems*, Int. J. of Inf. Man., 36, 4, 599–606, 2016.

[6] Popat S., Emmanuel M., *Review and comparative study of clustering techniques*, Int. J. of Comp. Sc. and Inf. Tech., 5, 1, 805–812, 2014.

[7] Stachowiak A., Żywica P., Dyczkowski K., Wójtowicz A., *An Interval-Valued Fuzzy Classi_er Based on an Uncertainty-Aware Similarity Measure*, Intelligent Systems'2014, Advances in Intelligent Systems and Computing, Springer, 32, 741–751, 2015.

[8] Murtagh F., Contreras P., *Algorithms for hierarchical clustering: an overview*, Wiley Int. Rev.: Data Mining and Know. Disc., 2, 1, 86–97, 2012.

[9] Jain K., Murty M., Flynn P., *Data clustering: a review*, ACM Comp. Surv. (CSUR), 31, 3, 264–323, 1999.

[10] Abbas A., *Comparisons between data clustering algorithms*, Int. Arab. J. of Inf. Tech., 5, 1, 320–325, 2008.

[11] Sika R., Ignaszak Z., *Implementation of the KMES Quality system for data acquisition and processing on the example of chosen foundry*, Arch. of Foundry, 8, 3, 97–102, 2008.