SOM-based clustering and optimization of production

Primož Potočnik, Tomaž Berlec, Marko Starbek, Edvard Govekar

Faculty of Mechanical Engineering,
University of Ljubljana, Slovenia,
{primoz.potocnik, tomaz.berlec, marko.starbek, edvard.govekar}@fs.uni-lj.si

Abstract. An application of clustering methods for production planning is proposed. Hierarchical clustering, *k*-means and SOM clustering are applied to production data from the company KGL in Slovenia. A database of 252 products manufactured in the company is clustered according to the required operations and product features. Clustering results are evaluated with an average silhouette width for a total data set and the best result is obtained by SOM clustering. In order to make clustering results applicable to industrial production planning, a percentile measure for the interpretation of SOM clusters into the production cells is proposed. The results obtained can be considered as a recommendation for production floor planning that will optimize the production resources and minimize the work and material flow transfer between the production cells.

Keywords: production optimization, clustering, SOM neural network, *k*-means clustering, hierarchical clustering.

1 Introduction

Modern production strives toward the optimization of production costs by methods such as lean production, [1] which involves defining the internal value system and consequently analyzing and optimizing the internal production flow. Related methods known as "group technology" (GT) [2] have been proposed to implement cellular manufacturing systems (CMS) that group machines into production cells. A number of methods have been proposed for cell formation. [3-6] A unified approach that combines assigning parts to individual machines and forming machines into cells is proposed in [7]. A clustering model is introduced in [8], exploiting similarities between products. The ART1 neural network-based cell formation in GT has been proposed by [9] and group technology based clustering is applied by [10].

In this paper, we develop a framework for production optimization for small and medium enterprises (SMEs) that are characterized by individual and small batch production with many different products in their production range. This paper's findings are based on data obtained from a Slovenian manufacturing company. We propose a clustering approach for the segmentation of various groups of products that can be organized into small production cells. Based on the representatives of each production cell, an efficient floor plan can be designed that will support lean production objectives.

2 Data

The data considered in this paper are derived from the company KGL d.o.o., located in Slovenia. The company has been operating since 1985 in close relation to the automobile industry. The production line comprises mechanical services on CNC lathes and machining centres, pressing sheet metal, fabrication of cylinders for gasoline engines and assembling parts manufactured in blanks.

2.1 Collecting data

Data were collected in the company in 2009 and comprise 252 products with descriptions of properties and operations required to manufacture each product. Table 1 presents 39 operations applied during production. Beside the operations, various features are attached to each product, such as: material, form, weight, volume, shape, number of assembly parts, dimensional accuracy, appearance of the product, number of possible variants, request for examination, the need of parts protection and the value of the product.

Table 1. List of operations applied during the production.

No	Operation	No	Operation	No	Operation
1	Band cutting	14	Thermal cutting	27	Cutting on shears
2	Service-casting	15	Service-forging	28	Deploy profile
3	Broaching	16	Drilling	29	Single-spindle thread.
4	Multi-spindle thread.	17	CNC turning	30	3-axis CNC machin.
5	Brushing	18	Service-blasting	31	Service-deburring
6	Honing	19	Powder Coating	32	Dip galvanizing
7	Galvanic coating	20	Artificial aging	33	Carbonitrirating
8	Service-hardening	21	Washing	34	Precision polishing
9	Before assembly	22	Assembly 1	35	Hand deburring
10	Viewing area	23	Testing	36	Packaging
11	Chamfering machine	24	Tumble deburring	37	Compression-cutting
12	Remodeling	25	Compression	38	Progressive compress.
13	Progr. deep drawing	26	Deep Draw-transfer	39	Welding

2.2 Data preprocessing

The data about required operations and features of the products need some pre-processing before being applied to various clustering algorithms. The following rules were applied to prepare the data:

1. Operations are encoded with a single value ('operation is not required' = -1, 'operation is required' = 1).
2. Materials are encoded with a single descriptor ('material not used' = -1, 'material used' = 1) for each possible material (Al/CuZn, SL, Fe).

3. Form is encoded with a single descriptor (-1,1) for each form applied (cast/forged, rod, platinum, band, profile/tube).
4. Shape is expressed as 'simple' = -1, 'combined' = 1.
5. Dimensional accuracy is originally expressed as ['0.1', '0.01', '0.001'] and we encode this property as '0.1' = -1, '0.01' = 0, '0.001' = 1 to keep proper ordering.
6. Appearance of the product is encoded as ordered variable: 'unimportant' = -1, 'important' = 0, and 'very important' = 1.
7. Request for examination is expressed as a single value: 'No request' = -1, 'Functional examination' = 1.
8. The need of parts protection is encoded with (-1,1) for each category (no protection, mechanical, anticorrosion).
9. Weight, Volume, Nr. assembly parts, Nr. of operations, Nr. of possible variants, and Value are expressed as real values and are therefore not encoded but only scaled into [-1,1] intervals.

Finally, constant operations (packaging and progressive deep drawing) were eliminated. Without prior knowledge about the relative importance of various operations, we assumed equal importance for all operations. This assumption was encoded into data by using the same scaling in interval [-1,1] for all the attributes. The preprocessed data comprises 58 attributes: 37 operations, 3 materials, 5 forms, 1 shape, 1 dimensional accuracy, 1 appearance, 1 request for examination, 3 protections, and 6 for real valued categories.

3 Clustering algorithms

Several clustering algorithms were applied to the task of products clustering, as follows: hierarchical clustering, *k*-means clustering, and SOM neural network. An important ability of these algorithms is to also accept real-valued attributes, which is not possible with structures such as the ART1 neural network.

3.1 Hierarchical clustering

Hierarchical clustering algorithms produce a nested series of partitions based on a criterion for merging or splitting clusters based on similarity [11]. In this paper, we apply agglomerative hierarchical clustering, with a Euclidean distance metric and several linkage methods: s*ingle linkage, complete linkage, average linkage,* and *Ward linkage*.

3.2 *K*-means

The *k*-means is the simplest and most commonly used algorithm employing a squared error criterion [11]. It starts with a random initial partition and keeps reassigning the patterns to clusters based on the similarity between the pattern and the cluster centres until a convergence criterion is met.

3.3 SOM

A self-organizing map (SOM) was proposed by Kohonen [12]. In this paper, we examine a SOM algorithm with distances defined by the Euclidean distance metric, and two variants of two-dimensional grid (rectangular and hexagonal). We apply a two-dimensional grid with six clusters, arranged by two rows each with three elements. Such a topology seems to be well suited to production floor planning.

3.4 Evaluation of clustering results

Clustering validity measures fall broadly into three classes [13]:

a) *internal validation* (based on properties of the resulting clusters),
b) *relative validation* (running the algorithm with different parameters),
c) *external validation* (comparison with a given partition of the data).

In our case study, there is no possibility of evaluating the correct clustering based on external validation measures; therefore, we have to rely in internal and relative validation that makes clustering essentially a subjective visualization tool. Following the recommendation proposed in [13], we evaluate the clustering results based on silhouette values that seem to be a good internal validation measure and also provide good graphical representation of clustering quality. The silhouettes validation technique [14] calculates the *silhouette width* for each sample, the *average silhouette width* for each cluster and the overall *average silhouette width* for a total data set. In our study, we evaluate the clustering quality based on the average silhouette width for a total data set.

4 Clustering results

In this section, we present and compare various clustering results, obtained by the proposed clustering algorithms. Results are presented for hierarchical clustering, *k*-means, and SOM clustering. Within each clustering method, various parameters are optimized in order to obtain the best results.

From the perspective of production planning, the company would prefer a small number of clusters, i.e. condensed production cells sharing the necessary tools and operation within confined space. The number of clusters should be around $K \approx 5$ but this is only a recommendation and not a strictly defined condition. Consequently, in our study we fix the number of clusters to be $K = 6$, which supports the application of a two-dimensional clustering architecture (3×2).

4.1 Hierarchical clustering

Hierarchical clustering results based on Euclidean distance and several linkage methods are shown in Figure 1. Single linkage obviously yields the worst result, with an average silhouette $S = 0.1937$. Average and complete linkage give better clustering results, while the best result, $S = 0.4698$, is obtained by applying the Ward distance.

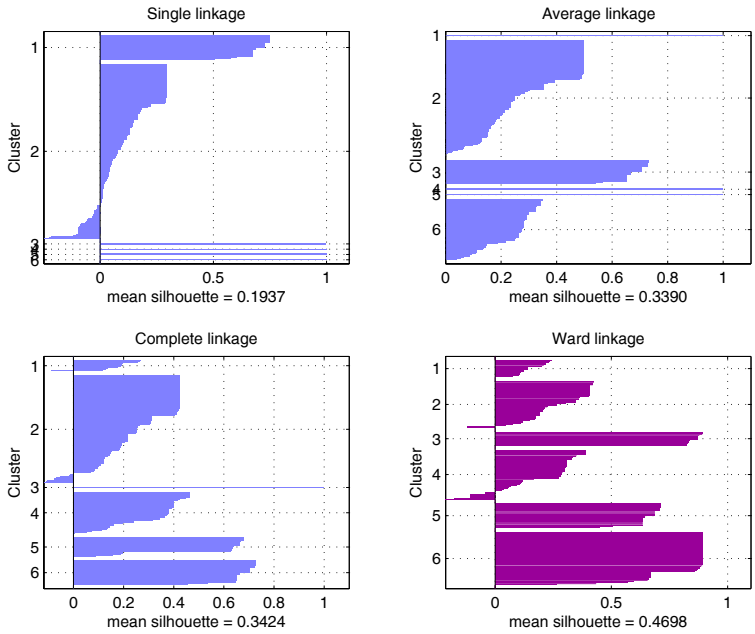

Fig. 1. Hierarchical clustering results compared by various linkage measures. Ward linkage results in best average silhouette value $S = 0.4698$.

4.2 *K*-means

Due to the sensitivity of the *k*-means algorithm to become trapped in a local minimum, the algorithm was restarted 100 times from various random initial positions. This effectively converged into a unique solution, as presented in Figure 2. The obtained average silhouette value amounts to $S = 0.4742$, which slightly exceeds the result obtained by hierarchical clustering.

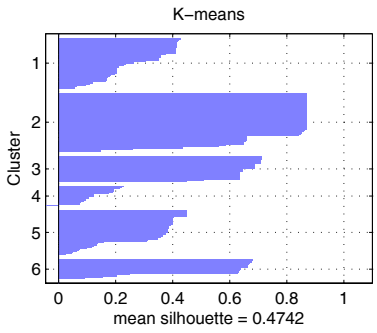

Fig. 2. *K*-means clustering, average silhouette value $S = 0.4742$.

4.3 SOM

Two variants of SOM grid organization are explored: rectangular and hexagonal grid. For the distance metrics and for the neighbourhood distance functions, the Euclidean distance metric is applied. Figure 3 displays the clustering results for both topologies (rectangular and hexagonal). The best result is obtained by the hexagonal topology and yields an average silhouette of $S = 0.4743$. Figure 4 shows a hexagonal SOM topology with class labels and number of samples in each class.

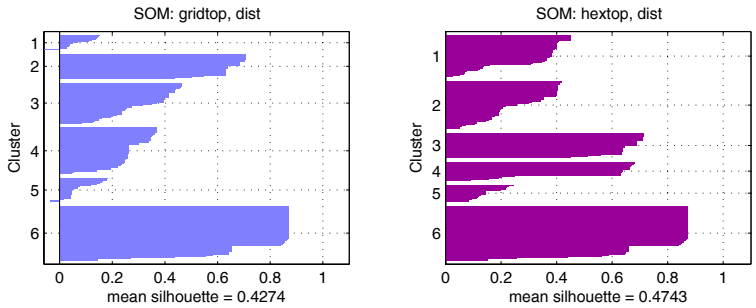

Fig. 3. SOM clustering results for rectangular and hexagonal 2-dimensional topology

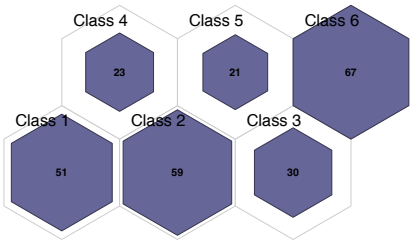

Fig. 4. Hexagonal SOM topology with class labels and number of samples in each class

4.4 Comparison of clustering results

The random initialization of the *k*-means and SOM clustering algorithm causes various final clustering arrangements in which class labels depend on initial conditions. In order to compare the silhouette plots of various clustering algorithms, the class labels should be aligned properly. We applied the method of class matching so that outcomes of various clustering methods were rearranged in a way that guarantees the highest cluster matching. The best results of hierarchical clustering, *k*-means, and SOM clustering are shown in Figure 5 and presented in Table 2. Table 2 also presents the similarity measure between various clustering approaches. The methods are compared to the SOM result with respect to the number of equally classified samples. After rearrangement, all three methods exhibit very similar results, with *k*-means and SOM being almost identical and hierarchical clustering diverging below 5%.

Table 2. Comparison of clustering results by average silhouette values.

Method	*Average silhouette S*	*Similarity to SOM*
Hierarchical clustering	0.4698	95.2%
K-means	0.4742	99.6%
SOM	0.4743	100%

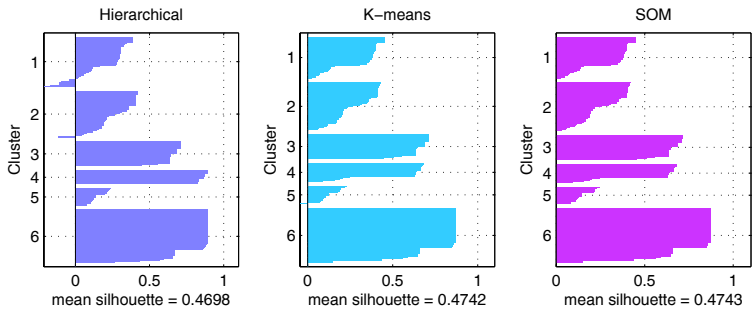

Fig. 5. Comparison of clustering results (hierarchical clustering, *k*-means, SOM).

The overall best result is considered to be obtained by SOM. Its average silhouette, $S = 0.4743$, is the highest score obtained in this study. The result is almost the same as *k*-means and only slightly exceeds hierarchical clustering, but there are two more advantages to support the selection of SOM clustering:

a) SOM result has no negative silhouettes, which means there are no products that are classified in a wrong cluster,

b) SOM topology shown in Figure 4 can be directly interpreted as a production floor plan.

SOM clusters maintain neighbourhood properties, which can be very helpful when designing a production floor plan. Operation clusters that are close to each other will probably share more equal operations and thus more material exchange than the clusters that are far apart. Therefore, we conclude that the SOM clustering method seems to be the most suitable approach for the task considered in this paper. The homogeneity of results obtained by various clustering methods only further supports the assumption that the obtained clustering result is meaningful and therefore applicable to production planning.

5 Application of clustering results

After obtaining a meaningful clustering result, the next step is to apply this result to the production environment. In this section, we propose an interpretation of clustering results that may yield an applicable industrial solution.

According to the SOM clustering result, six operation cells are arranged in a two-dimensional hexagonal grid. This architecture (shown in Figure 4) can already be

interpreted as an initial production floor layout. The next question is about which operations should be contained in the arranged production cells. This leads to the interpretation problem of clustering results as described below.

5.1 Interpretation of clustering results

The interpretation problem can be formulated as how to define a mapping from obtained SOM clusters into the real world production cells:

SOM clusters → Production cells

As SOM clusters are represented by prototypes, an initial estimation could be to directly translate SOM prototypes into production cells, but this turns out to be inappropriate. SOM prototypes have continuous values in the interval [-1,1], which is acceptable for the product features (weight, volume, shape, etc.) but not for the operations that should be either included or not included in the production cell. Therefore, some kind of discretization of SOM prototypes should be performed for logical descriptors (such as operations, materials, form, shape, etc.).

We propose a percentile measure for the interpretation of SOM clusters into the production cells. For each operation in a particular cluster, we can provide a percentage of samples in this cluster, which should contain this operation (or property) in order for this operation (or property) to be included into the production cluster. Various percentile margins can be defined, such as: p = {50th, 75th, 100th}. For each percentile margin, the particular operation should be included into the production floor planning cell if at least (100-p)% of samples in a particular SOM cluster require this operation. The 100th percentile should be interpreted as a limit value: if at least one sample requires an operation, it should be included in the production cell.

Figure 6 presents SOM interpretation results based on various percentile margins p = {50th, 75th, 100th}. The result for p = 50th seems to be under-populated as there are a significant number of operations missing in a complete production scheme because they are simply not frequently required by the production process. On the other side, the result for p = 100th is probably over-populated. An optimal arrangement can be expected somewhere between the 50th and 100th percentile margins; therefore, a 75th percentile margin can be taken as a guideline to successfully interpret SOM clusters into the production cells.

5.2 Finalization of production clusters

If SOM interpretation based on the 75th percentile is applied to cell arrangement, some operations are still not assigned to any production cell. This is a consequence of their rare application, but in practice the production process can not be completed unless all the required operations are available. Therefore, as a final stage, we propose assigning each missing operation into the cell that has the highest requirement for this operation among all cells. The final result is shown in Figure 7, where filled missing discrete operations are displayed in blue.

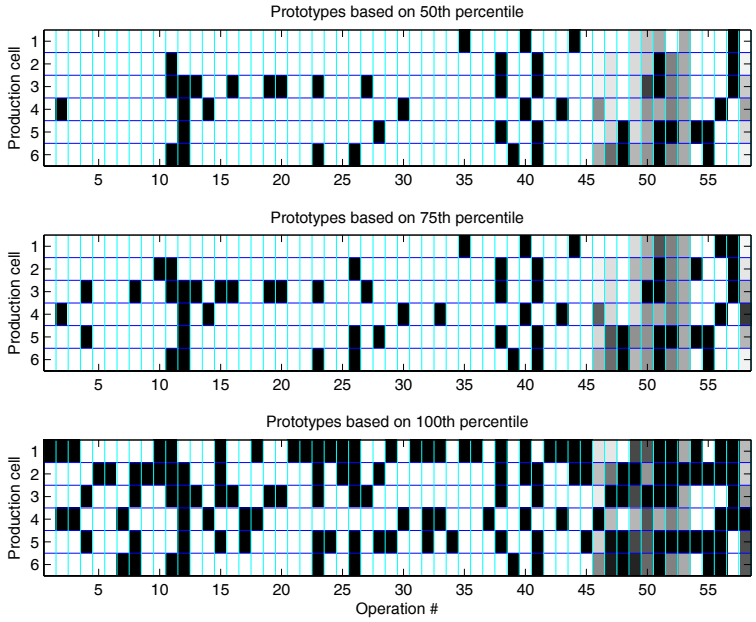

Fig. 6. Interpretation of SOM clusters into production cells based on percentile margins

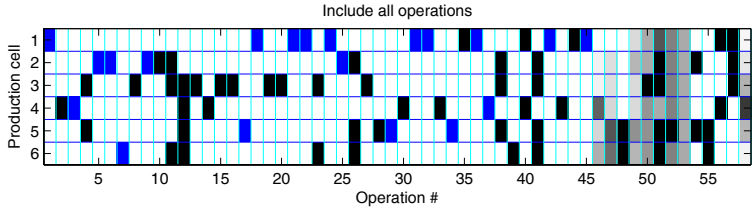

Fig. 7. Final interpretation of SOM clusters into production cells based on the 75th percentile margin and filling the empty operations

6 Conclusions

This paper presents an application of clustering methods to production planning. Various clustering methods, including hierarchical clustering, *k*-means and SOM clustering are applied to production data from the company KGL in Slovenia. The best results are obtained with SOM clustering, although the results are shown to be very consistent in comparison with other clustering methods. An interpretation method is proposed to translate the SOM clustering result into the production cells. The following two properties support the assumption that the resulting production cell arrangement is suitable for production planning:

a) Organization of production cells supports production of similar products in closed production units, which optimizes material and work flow, and reduces production costs. Clustering evaluation based on silhouette widths confirms good clustering quality, which means that the proposed clustering is meaningful.
b) Production cells are arranged according to SOM topology. This guarantees that neighbourhood properties of clusters are maintained and consequently, it can be expected that this will lead to the minimization of material and work piece exchange between production cells that are not close to each other.

The results reported can be considered as a recommendation to the production planning managers. We hope the proposed results will be a useful guidance to production planning in the company.

References

1. Hines P., Tayor D.: Going Lean, Lean Enterprise Research Centre, Cardiff Business School, Cardiff, UK, (2000)
2. Kusiak, A.: The generalized group technology concept. International Journal of Production Research, 25, 561–569, (1987)
3. Crama, Y., Oosten, M.: Models for machine-part grouping in cellular manufacturing. International Journal of Production Research, 34, 1693–1713, (1996)
4. Shanker, K., Agrawal, A.K.: Models and solution methodologies for the generalized grouping problem in cellular manufacturing. International Journal of Production Research, 35, 513–538, (1997)
5. Adenso-Díaza, B., Lozano, S., Racerob, J., Guerrerob, F.: Machine cell formation in generalized group technology. Computers & Industrial Engineering, 41, 227–240, (2001)
6. Fan, Z.P., Chen, Y., Mab, J., Zhu, Y.: Decision support for proposal grouping: a hybrid approach using knowledge rules and genetic algorithms. Expert Systems with Applications, 36, 1004–13, (2009)
7. Foulds, L.R., Neumann, K.: A network flow model of group technology. Mathematical and Computer Modelling, 38, 623–635, (2003)
8. Andrés, C., Albarracín, J.M., Tormo, G., Vicens, E., García-Sabater, J.P.: Group technology in a hybrid flowshop environment: A case study. European Journal of Operational Research, 167, 272–281, (2005)
9. Yang, M.-S., Yanga, J-H.: Machine-part cell formation in group technology using a modified ART1 method. European Journal of Operational Research, 188, 140–152, (2008)
10. Morača S., Hadžistević, M., Drstvenšek, I., Radaković, N.: Application of Group Technology in Complex Cluster Type Organizational Systems. Journal of Mechanical Engineering, 56, 663–675, (2010)
11. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Computing Survey, 31, 264–323, (1999)
12. Kohonen, T., Self-Organizing Maps, Second Edition, Berlin: Springer-Verlag, (1997)
13. Brun, M., *et al.*: Model-based evaluation of clustering validation measures. Pattern Recognition, 40, 807–824, (2007)
14. Rousseeuw, P.J.: Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. Computational and Applied Mathematics, 20, 53–65, (1987)
15. Rousseeuw, P.J.: Finding groups in data: An introduction to cluster analysis, New York: Wiley, (1990)