

Towards Zero-Defect Manufacturing: Machine Selection through Unsupervised Learning in the Printing Industry

Sotirios T. Spantideas¹, Anastasios E. Giannopoulos¹, Nikolaos C. Kapsalis², Angelos Angelopoulos¹, Stamatis Voliotis¹ and Panagiotis Trakadas¹

¹ National and Kapodistrian University of Athens, Psachna, Evia, 34400, Greece

² National Technical University of Athens, Patission Complex 42, Patission str, 10682 Athens, Greece

Abstract

Zero Defect Manufacturing (ZDM), being one of the main concepts of Industry 4.0, is especially critical in the offset printing industry, since it is associated with production enhancement and environmental footprint reduction. This work proposes a Machine Learning clustering-based approach to determine hidden order attributes that can be used to define a beneficial machine selection policy for the incoming orders in terms of fault occurrence reduction and production enhancement. Three clustering methods (k-means, agglomerative hierarchical clustering and density-based scanning) are modified in order to reveal the hidden order features that have a significant impact on the number of defected pieces. First, the ML framework of the clustering methods is presented, mainly including the fine-tuning of the learning parameters. Then, the trained ML models are compared in terms of their performance on unseen data to evaluate the machine selection process. The evaluation outcomes demonstrate the ability of the clustering ML framework to ensure proactive machine selection policy, reducing the printing defects.

Keywords

Machine learning, Industry 4.0, zero-defect-manufacturing, unsupervised learning, clustering.

1. Introduction

The primary target of Industry 4.0 is to ameliorate the current conventional production methods by combining innovative data technologies from both physical and digital contexts [1, 2]. This transformation will enable the manufacturing production to move from the state of a posteriori management to the state of timely prediction of optimal resource and process management, optimizing the quality of the product and the usage of raw materials, while also minimizing the production chain defects [3]. The concept of Zero-Defect Manufacturing (ZDM) has therefore been adapted by the majority of the stakeholders operating in the manufacturing domain, not only due to the effective cost reduction in their production chain, but also due to the reduction of their environmental footprint [4]. For these purposes, the whole industrial field is currently moving beyond reactive resource management towards proactive and predictive solutions, necessitating the establishment of Artificial Intelligence (AI)-assisted solutions. In the context of Industry 4.0, AI-based techniques and Machine Learning (ML) methods are used as the primary instigators to enable self-optimization and automation in the manufacturing process, as well as provide fault detection and real-time decision making functionalities towards ZDM [1–4].

By leveraging proactive and predictive measures in the industrial production chain, product quality can be effectively maximized and the cost associated with defects can be eliminated. AI/ML-assisted

Proceedings of the Workshop of I-ESA'22, March 23–24, 2022, Valencia, Spain

EMAIL: sospanti@uoa.gr (S. Spantideas); angianno@uoa.gr (A. Giannopoulos); ncapsalis@gmail.com (N. Kapsalis); a.angelopoulos@uoa.gr (A. Angelopoulos); svoliotis@uoa.gr (S. Voliotis); ptrakadas@uoa.gr (P. Trakadas)

ORCID: 0000-0002-1457-4131 (S. Spantideas); 0000-0002-8602-7401 (A. Giannopoulos); 0000-0002-8191-8010 (A. Angelopoulos); 0000-0001-5933-5129 (S. Voliotis); 0000-0002-5146-5954 (P. Trakadas)

© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)



solutions have therefore been developed for typical manufacturing applications such as fault detection, predictive maintenance, optimization of the manufacturing process and machine configuration parameters and enhancement of the energy savings [3, 5]. With the use of ML algorithms, prediction of defected products can be identified in advance and machine configuration parameters can be linked with the fault occurrence [6, 7].

Offset printing is one of the widely-used printing processes ever developed, accommodating many types of printing jobs, including newspapers, magazines, brochures, labels, books, and many others. The main identified issues that still remain unsolved are: (i) the number and the diversity of characteristics of the orders do not allow for an easy standardization of the processes; (ii) the deficiencies are typically observed during the quality control of the final product, thus leaving no space for corrective actions; (iii) existing rule-based optimization methods work in an a-posteriori management instead of timely prediction of optimal resource management; (iv) the printing industry exhibits a significant environmental footprint, since the manufacturing process involves extensive usage of raw materials (water, paper, ink, aluminum), where defected products contribute the largest part.

In the present exploratory work, we propose three modified Unsupervised Learning (UL) modeling algorithms in order to facilitate the standardization of ML-empowered methods in the offset printing industry. The present work proposes: (i) unsupervised learning-driven modeling of each machine with the goal of revealing the cluster of the minimum defects per order; (ii) investigation of the hidden attributes of machines to optimize the machine selection policy; (iii) construction of machine-specific clustering models using three well-established algorithms, namely k-means, agglomerative hierarchical clustering and density-based scanning (DBSCAN); (iv) a constraint-dependent clustering approach based on pre-defined functions; (v) quantitative validation of the developed models compared to the existing machine selection policy.

2. Methods

2.1. Dataset Description

The offset printing process consists of three phases, namely pre-press, press and post-press. In each one of these three phases, several (raw, organic, chemical and recycled) materials are used, including paper, water, ink, aluminum, alcohol solutions, having a direct impact not only on the environmental footprint but also on the economic growth. This paper exploits a subset of historical dataset obtained from the press stage of an offset printing company during the last two years. Data collection was performed at the single-order level for five operating printing machines, meaning that the features of a given order were recorded, along with the associated machine ID. Specifically, the dataset contains 10K entries per machine (a total of 50K data), corresponding to 10K historical orders.

The collected features for each order are: (i) *Quantity*: Number of pieces requested in a particular order. Indicative values range from 100 to 1000 pieces, depending on the order type; (ii) *Quality*: Paper quality requested in a particular order. *Quality* is a categorical variable that takes values ‘Velvet’, ‘Uncoated’ or ‘Illustration/Gloss’. Note that ‘Velvet’ is the most-frequently requested paper quality (57%), followed by ‘Uncoated’ (26%) and ‘Gloss’ (17%); (iii) *Color*: Color requirements of a particular order. Color is also a categorical variable that takes values ‘Color’ (typical 4-color printing, 88%), ‘Color+’ (4+1 color printing, 10%) or ‘B&W’ (grayscale printing, 2%); (iv) *Ink*: Ink level required for each piece of a particular order. Typical ink values vary between 0.1 to 1 gr; (v) *Type*: The requested outcome type of a particular order. Type is a categorical variable with values ‘Book’ (30%), ‘Poster’ (30%) or ‘Journal’ (40%); (vi) *Accuracy*: The ratio between the accurately printed pieces and the quantity of the order. *Accuracy* is a scalar variable ranging from 0 – 1 and reflects the percentage of defected pieces in the order (1 corresponds to zero defected pieces).

2.2. Machine Selection through Unsupervised Learning and Constraint Clustering

The modeling process follows the basic assumption that one or multiple order features are associated with enhanced accuracy levels. Given the variability in the number and shape of the clusters resulted by each algorithm, we determined an objective function to stabilize the algorithms' hyper-parameters (number of clusters for k-means and agglomerative clustering and minimum number of points in a ε radius for DBSCAN). The Accuracy Discrimination Score (ASD) is used as an objective function:

$$ASD_k = (\max_{k' \in K} [A_{k'}] + SS_k) / C_k, \quad (1)$$

where C_k is the number of clusters exceeding the accuracy threshold, A_k is the within-cluster accuracy score (90%) and SS_k is the silhouette score over the k clusters (ASD value is zero in case that $C_k = 0$). The ASD targets to jointly maximize the within-cluster accuracy score, while also minimizing the number of clusters that exhibit 90% accuracy levels, enabling the determination of constraint-clustering models.

3. Simulation Results

3.1. Hyperparameter Tuning

K-Means. The 10K dataset collected by each machine is provided to the algorithm using a varying number of clusters k (1 to 100). K-means iteratively assigns the data samples to k clusters, targeting to minimize the within cluster variance (6-dimensional squared distance between each sample and the cluster centroid). Figure 1 shows the ADS relative to k for each individual machine. Evidently, ADS is maximized for Machines 1-3 and 5 with relatively low number of total clusters k , while the dataset obtained by Machine 4 requires a significant $k=88$ in order to identify at least one cluster with accuracy level above 90%.

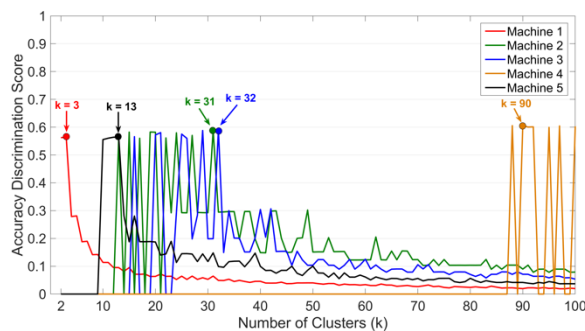


Figure 1: ADS relative to the number of clusters k for the five printing machines using k-means algorithm. Colored arrows designate the variable k exhibiting maximum ADS.

Agglomerative Hierarchical Clustering. The hierarchical clustering algorithm initially considers that each data sample forms its individual cluster. Then, depending on the distance between the data, adjacent samples in the 6-dimensional space are iteratively grouped together until the defined number of clusters k is reached. Similarly to the k-means algorithm, the ADS for varying k (1 to 100) is shown in Figure 2, along with the number of clusters k exhibiting the maximum ADS value for each machine.

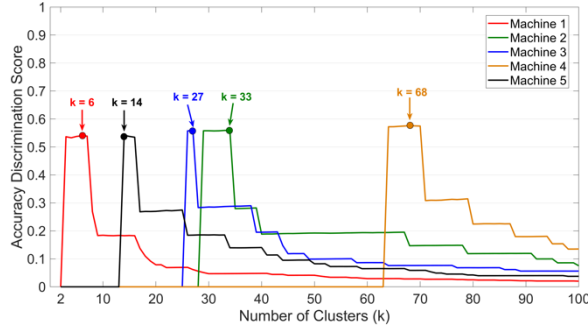


Figure 2: ADS relative to the number of clusters k using agglomerative hierarchical clustering. Colored arrows designate the variable k exhibiting maximum ADS.

DBSCAN. The density-based clustering algorithm is suitable for more complex clusters, e.g. when dense data areas are nested. The algorithm identifies core points in the data samples that are used to establish clusters depending on the minimum number of neighbouring data points N in radius ϵ . For this reason, the parameters N , ϵ are jointly varied in order to identify the optimal (N, ϵ) pair that maximizes the ADS. Figure 3 depicts the ADS as a function of (N, ϵ) as surface plot for each individual machine.

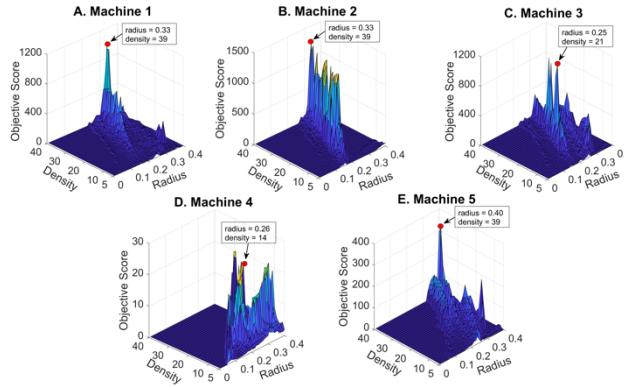


Figure 3: Machine-specific 2D-surface plot of Objective score (ADS) as a function of (N, ϵ) . Red dots indicate the (N, ϵ) points of objective maxima.

3.2. Validation Results and Machine Labeling

A validation dataset containing 100 unseen orders per machine, each one exhibiting accuracy levels above 90%, was used to verify the performance of the pre-trained models. The performance metric for each machine was calculated as the ratio between the number of validation samples grouped within the best-accuracy cluster and the total number of validation samples. For benchmarking purposes, Figure 4 depicts the performance of the three clusters along with a round-robin machine selection policy. All metrics are illustrated in relation to the ground-truth performance (Relative Performance Gain - RPG), resulted by the existing machine selection policy (rule-based approach, primarily exploiting specifications of the machines' manufacturers).

As evident from Figure 4, k -means outperforms the rest of the clustering models in Machine 1 and 3, implying that training datasets can be clustered following geometrically centroid-based criteria. On the contrary, datasets from Machines 2 and 5 formed density-based groups to isolate the best-accuracy clusters, thereby showing beneficial RPG for DBSCAN. Finally, Machine 4 did not reveal any excessive RPG score, concluding that there are no gains in using clustering methods for proactive machine selection. Note that, an RPG value of 1 denotes that a particular model performs equivalently with the currently used policy.

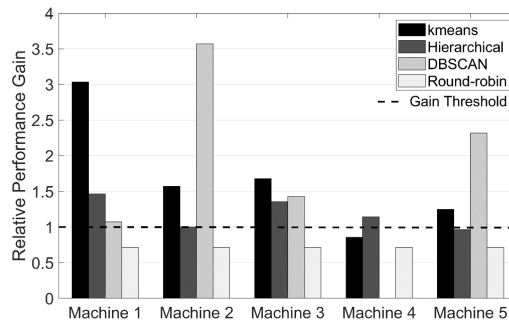


Figure 4: Relative Performance Gain of the ML clustering models and a round-robin scheme employing 100 validation data per machine.

The presented ML clustering methods can be used to further analyze the features of the data samples that form clusters with enhanced accuracy scores and determine the hidden order attributes for each machine. To this end, a machine/feature labeling can be established for beneficial machine selection policy (new orders are assigned to the printing machine showing suitable feature labels), which in turn will contribute to the enhancement of the production efficiency, the minimization of defected products and the reduction of the company's environmental footprint.

4. Acknowledgement

This work has been partially supported by the project Offspring, under the open call of ZDMP project, funded by the European Commission under Grant Agreement number 825631 through the Horizon 2020 program.

5. References

- [1] A. Angelopoulos, E. T. Michailidis, N. Nomikos, P. Trakadas, A. Hatziefremidis, S. Voliotis, T. Zahariadis, Tackling faults in the industry 4.0 era - a survey of machine-learning solutions and key aspects, *Sensors* 20 (2020) 109. doi: 10.3390/s20010109.
- [2] Y. S. Chuo, J. W. Lee, C. H. Mun, I. W. Noh, S. Rezvani, D. C. Kim, J. Lee, S. W. Lee, S. S. Park, Artificial intelligence enabled smart machining and machine tools, *Journal of Mechanical Science and Technology* 36 (2022) 1-23. doi: 10.1007/s12206-021-1201-0.
- [3] Z. M. Çınar, A. Abdussalam Nuhu, Q. Zeeshan, O. Korhan, M. Asmael, B. Safaei, Machine learning in predictive maintenance towards sustainable smart manufacturing in Industry 4.0, *Sustainability* 12 (2020) 8211. doi: 10.3390/su12198211.
- [4] P. Trakadas, P. Simoens, P. Gkonis, L. Sarakis, A. Angelopoulos, A. P. Ramallo-González, ... P. Karkazis, An artificial intelligence-based collaboration approach in industrial IoT manufacturing: Key concepts, architectural extensions and potential applications, *Sensors* 20 (2020) 5480. doi: 10.3390/s20195480.
- [5] A. Angelopoulos, A. E. Giannopoulos, N. C. Kapsalis, S. T. Spantideas, L. Sarakis, S. Voliotis, P. Trakadas, Impact of Classifiers to Drift Detection Method: A Comparison, in: *International Conference on Engineering Applications of Neural Networks*, Springer, Cham, 2021, pp. 399-410.
- [6] D. P. Penumuru, S. Muthuswamy, P. Karumbu, Identification and classification of materials using machine vision and machine learning in the context of Industry 4.0, *Journal of Intelligent Manufacturing* 31 (2020) 1229-1241. doi: 10.1007/s10845-019-01508-6.
- [7] O. Silvén, M. Niskanen, H. Kauppinen, Wood inspection with non-supervised clustering, *Machine Vision and Applications* 13 (2003) 275-285. doi: 10.1007/s00138-002-0084-z.