Topological Data Analysis for Industrial Manufacturing – A Mini Review

Martin Uray[1], Barbara Giunti[2], Stefan Huber[1], and Michael Kerber[2]

[1] Salzburg University of Applied Sciences, Salzburg, Austria
`{martin.uray, stefan.huber}@fh-salzburg.ac.at`
[2] Graz University of Technology, Graz, Austria
`{bgiunti, kerber}@tugraz.at`

Abstract. Topological Data Analysis (TDA) is a branch of mathematics using topological tools for the analysis of complex, multi-dimensional data. It has been successfully applied in several fields such as medicine, material science, and biology. This mini-review summarizes the state-of-the-art of TDA in yet another area: industrial manufacturing in the context of *Industry 4.0*. It highlights its key benefits and challenges, and provides directions for future research.

Keywords: Industry · Manufacturing · Production · Topological Data Analysis · Persistent Homology · Mapper Algorithm · Review

1 Introduction

Industrial manufacturing is ongoing a deep transition in the course of the fourth industrial revolution, also known as *Industry 4.0*. A core aspect is the extensive digitalization and the convergence of Information Technology (IT) and Operational Technology (OT) to enable modern methods of machine learning and data science for the purpose of optimization, prediction, analytics and continuously increasing the levels of autonomy and flexibility for the industrial manufacturing processes.

Due to the natural characteristics of industrial production, topological and multi-dimensional data can be found all along the production processes, like shapes, textures of products, or trajectories. With data of this kind, traditional methods are often facing issues in effectively analyzing and using it.

Topology is a branch of mathematics that study the properties of geometric shapes that do not change after continuous deformations such as stretching. While studied for more than a century, the last 25 years have witnessed a steep increase in applications of topology in data analysis. This has led to the research field of Topological Data Analysis (TDA), which is extremely successful with over 400 applications [5] in many areas such as medicine (78 applications), material science (12), biology (37), cosmology (9), and finance (5). TDA captures information at different scales, and Persistent Homology (PH), one of the main tools in TDA, can be easily automated without a lot of parameters to be tuned. These properties makes TDA particularly attractive for changing specifications in flexible *Industry 4.0* settings, i.e., mass customization and lot-size one production.

The number of applications of TDA has become so numerous and diverse that no individual researcher can possibly keep track of all developments anymore. This poses the question of how the interface of TDA to application domains can be structured to become accessible from both sides. One recent initiative is DONUT [5], a search engine that allows for a lookup of applications of TDA in a simple way. A complementary effort are scientific survey articles that summarize and compare different approaches of how TDA has been linked to applied setups. There is a substantial body of such surveys and textbooks. Their focus is usually on explaining the theory and presenting a few sample applications, demonstrating the diversity of areas that can profit from TDA. This approach is reasonable since a comprehensive survey of all applications of TDA would result in a document of unmanageable size.

We suggest a different kind of survey, tailored towards domain experts in industrial applications. While this field is not recognized as the most successful application domain of TDA, we demonstrate in this article that there are various research articles scattered in the literature that point towards a rich interplay between TDA and industrial manufacturing, with a lot of untapped potential. We provide a first overview of which methods of TDA have been used in what step of the manufacturing processes.

In the following, section 2 defines the fundamental industrial and mathematical terminology used within this work. Section 3 gives a detailed description on the method used for the literature review, with the results shown in section 4. Our findings are then discussed and concluded in section 5.

2 Theoretical Background

The manufacturing of a product, in the traditional sense, involves several process steps. At the beginning of the process, the definition of the product need has to be identified, followed by a Conceptional Design and Evaluation of the said. Based on that, a prototype is created, enabling the creation of schematics for standardized reproduction. These schematics, in combination with the requirements of the product, define the specification for the selection of material, processes, and production equipment. The production itself is then accompanied and finished by an inspection and quality assurance before the products are packed. The terms *Manufacturing* and *Production* are used in literature for the process of creating products. Depending on the domain, e.g. Semiconductors, also the term *Fabrication* may be found [11].

Even when there is a semantic difference between the terms manufacturing, production, and fabrication, in this work, the term *Production* and *Manufacturing* are interchangeably used as an umbrella term for all three.

Manufacturing- or Production Engineering describes the branch of engineering working on the full process of manufacturing. Among others, the planning and optimization of the production processes are subject of interest to this discipline [16].

The field of TDA can be roughly clustered into two main approaches: *mapper* [19] and Persistent Homology (PH) [4,10]. Common to both of them is that the data at hand is first converted into a computer-interpretable geometric representation whose topological properties are analyzed. Importantly, this representation is not just one isolated shape, but rather a family of shapes that reveals properties of the data set at various scales. A key point of TDA is that most useful information on a data set becomes apparent by investigating how topological properties evolve across different scales.

Mapper is the conceptually simpler approach since the only topological property considered is connectivity. In essence, it is a topologically guided graph of clusters of an object set V in $\mathbb{R}^n$: First, V is mapped by f, the lens function, to a low-dimensional space $\mathbb{R}^d$, say, using PCA or autoencoders. Then $f(V)$ is split into (overlapping) sets $V_1, \ldots, V_k$. Each V_i is pulled back into $\mathbb{R}^n$ as $f^{-1}(V_i)$ and clustered using a clustering method of choice. All clusters of all $f^{-1}(V_i)$ form the vertices of the *mapper graph* G and if two such clusters intersect we add an edge to G. Note that the clustering happens in the original space of the point set, but it is guided by the filter function and the covering.

The mapper graph is used for explorative data analysis: usually, one looks for *flares* in the graph, that is, subpopulations of objects that are connected across several scales (intervals) and distinguished from the remaining objects on these scales. One then analyzes these subpopulations (potentially with traditional data analysis methods) to find a reason for their distinctiveness. This framework is described by Lum et al. [14], applying it to breast cancer data, the voting behavior of congressmen, and the analysis of basketball players in the NBA.

In practice, the major obstacle is the splitting of V into $V_1, \ldots, V_n$: the interpretability of the outcome entirely depends on its choice. While a few standard choices are known, it usually requires the prior knowledge of a domain expert to get meaningful insights out of the mapper pipeline. Since various other parameters can be chosen in the framework, mapper should be considered a powerful, versatile interactive tool that can reveal hidden connectivity in data sets.

Homology is a fundamental concept from algebraic topology, allowing us to identify shapes that cannot be continuously transferred into each other. An extensive treatment is beyond our scope; informally, homology reveals the number of k-dimensional holes of a shape for every integer k. For $k = 0, 1, 2$, this corresponds to the number of connected components, tunnels, and cavities in the shape. Crucially, given a continuous map between two shapes, for instance, an inclusion from X into Y, there is a well-defined map between these holes. For instance, X being a Swiss cheese in the shape of a torus, we get many cavities and one tunnel. If we imagine offsetting X by a small radius to obtain Y, we fill out all small cavities in its interior, but the overall torus shape remains; we can therefore label the tunnel in X as a *persistent feature* and the cavities as *spurious* because they can be removed via a small perturbation. The usual interpretation is that persistent features in data carry more meaning than spurious ones[3]

In the pipeline of PH, we build a sequence of increasing shapes X_r for every scale parameter $r \geq 0$, called a filtration, and observe how holes appear and disappear when we consider the increase of X_r as a continuous process. In the example above, X_r would be offsetting X with radius r, which is in fact the most common example in applications. The resulting evolution of topological features can be represented as a *barcode*, a collection of intervals (bars) that represent the lifetime of a hole within the filtration.

One advantage of this framework is that there is often a natural choice for picking a filtration, so the pipeline allows for easier automation than the mapper algorithm. There is also a rich theory for how to compare two data sets by comparing their barcodes, and how to integrate PH into machine learning methods, i.e., kernel-based methods or neural nets. The well-founded theory and the interpretability of the obtained features have contributed to the success of PH in practice.

3 Methodology

The method used for this review is a semi-exhaustive literature review, based on Randolph [18], and inspired by Tschuchnig et al. [20]. The problem defined for this work is the evaluation of methods from TDA on the application of industrial production processes.

For the definition of meaningful search queries, two categories are defined: *Method* and *Domain*. The keywords of the categories *Method* describe the methods we are interested to find applications in the keywords of the *Domain*. The keywords defined for the *Method* are *Topological Data Analysis*, *Persistent Homology*, and *Mapper Algorithm*. For the *Domain*, the keywords *Manufacturing*, *Production*, and *Fabrication* were chosen. The resulting search queries are formed by each combination of keywords from *Method* and *Domain*, so 9 in total.

For data collection, *Google Scholar* is defined as the search engine. Each defined search query was used and the first three pages of results are taken. For further processing, a set of criteria was defined for the inclusion of references into the review: the removal of duplicates, definition of a time period, filtering on relevant types of publication, the availability of a full text, and filtering on the context.

Some search queries returned overlapping results. These duplicates were removed. The timeframe to be considered for inclusion is defined to include all papers published before January 2023, with no restriction on a minimal year. In order to deliver a qualitative review, only publications from Conference Proceedings and Journals were taken into account.[4] References, like preprints, presentations, or reports are excluded. To be able to further extract information from the

[3] Cheese experts might disagree with that interpretation in this (toy) example.

[4] Note, here we are aware that not all Conference Proceedings and Journals published may be peer-reviewed.

publications, an english full text has to be available. So, non-english results and results without a full-text available (e.g. conference abstracts) were excluded. The remaining results are then analyzed, whether one keyword of the *Domain* and the *Method* categories are present within the title, abstract, or introduction. As a last step, all remaining publications were filtered, whether they are within the context of our research. Next, the results were manually clustered regarding their application. For further discussion, the following information was extracted from the full texts: *Task*, *Aim*, and *Applied Method*.

4 Results

The semi-exhaustive literature review resulted in 12 paper. In the following, these publications are listed, and grouped into three clusters, as identified during the manual grouping procedure. The results are clustered according to their stage of the production process. Each of the clusters is displayed in a table, stating the publication, the *Feature*, *Process*, or Task of *Quality Control*, including the *Aim* and the TDA *Tool* applied.

The first identified cluster is about detecting features of products, without any explicit task stated, but can be used at any stage during the production process. The results can be seen in table 1. Here, the publications are working on recognizing mechanical features [8], proposing a method for shape segmentation [22], or analyzing the shape of surfaces [23].

Table 1. Table showing cluster *Product Features* results by the review.

#	**Feature**	**Aim**	**Tool**
[8]	Feature	Recognition of Mechanical Features	PH
[22]	Shape	Proposal of a new method for Shape Segmentation, employing Graph Convolutional Networks	PH
[23]	Surface Texture	Analysis of shape of surfaces	PH

The second cluster contains publications with proposed methods or applications of TDA for *Manufacturing Engineering*. The results are listed in table 2. This cluster's result contains three different applications. To benchmark applications on optimizing material flow, Dassisti et al. [3] propose a method, evaluating the methods using PH. Given process variables, [6,7] use Mapper to forecast the quality of a product. The task of optimizing a configuration is shown by [15], on the application of energy-bound cages.

The third cluster shows publications explicitly stating methods that shall increase quality and detect, or even prevent, faulty products. Table 3 lists the results. For an explicit application towards quality control of products, 6 publications were identified. The included publications have applications during different phases of the product life-cycle. An application during the design phase is

Table 2. Table showing cluster *Manufacturing Engineering* results by the review.

#	Process	Aim	Tool
[3]	Material Flow Optimization	Visual benchmark on selecting the best material-flow path from depot to production line in a multi-vehicle routing problem	PH
[6], [7]	Manufactoring Productivity	Prediction of Product Quality, based on key process variables	Mapper
[15]	Configuration Optimization	Synthesizing planar energy-bound cages by identifying optimal configuration	PH

posed by [17], offering a simulation of physical behavior. Time-series of machinery signals is analyzed [12,13], so that signal artifacts do not lead to defective products. Hsu et al. [9] apply clustering on features, generated from a Vision Transformer, based on data from the production of a wafer. Defects on products can be detected using [1,21], as a final quality check. (In [1], Betti numbers and the Euler characteristics is used, i.e., plain homology rather than PH.)

Table 3. Table showing cluster *Quality Control* results by the review.

#	Task	Aim	Tool
[1]	Discrepancy Classification	Classification of Topological Discrepancies in Additive Manufacturing	H
[9]	Quality Clustering	Clustering on Defect Patterns for Wafer Production, based on Features extracted by a Vision Transformer	Mapper
[12]	Artifact Detection	Detection of True Steps in piecewise constant signals	PH
[13]	Artifact Identification and Detection	Identification and Detection of Chatter on Signals	PH
[17]	Design Simulation	Dynamic State Analysis of a Pendulum	PH
[21]	Fault Detection	Detection of Eccentricity of Electric Motors	PH

5 Discussion and Conclusion

The results of this study show that the interest in TDA methods for industrial applications increased over the past few years. But still, comparing the results with other application domains (c.f. [5]), TDA has still a low value in the application of industrial manufacturing. This work exposed 13 references, whereas other application domains show substantially more (78 in medicine, for example).

This study reveals a potential application direction of TDA methods. One advantage of methods on TDA exploited in other domains is that these can be fully automated. Where other methods need engineering parameters, these methods can be fully integrated into production systems, without the necessity for cherry-picking on parameters. Additionally, within production processes, data is often in a topological format, i.e. properties of products, like shapes, or textures. This kind of data is hard to exploit using traditional methods. Thus, TDA offers a range of methods, to take advantage of this data to be analyzed.

During the discussion of the results, and a complimentary check on DONUT, we observed that at least one relevant publication is not listed. The paper [2] was not even found with a reverse search using the occurring keywords. Still, for consistency within our method, this paper has not been manually included within the results. A potential reason could be our used method or the publisher's indexing strategy. A more exhaustive search is planned as future work.

In this paper, we provide an overview of the methods of TDA applied to industrial manufacturing. The numerous publications in the field imply that this also applies to the industrial setting. But here, compared to other reviews, only a limited number of publications were found. TDA seems to be an appropriate tool to be used in this setting, due to the natural characteristics of industrial production, where a lot of topological features can be found. In addition, since PH, one of the main tools in TDA, can be fully automated, it is especially applicable to the changing requirements posed by the demands of *Industry 4.0*.

Acknowledegments M. Uray and S. Huber are supported by the Christian Doppler Research Association (JRC ISIA). B. Giunti and M. Kerber are partially or fully supported by the Austrian Science Fund (FWF) P 33765-N.

References

1. Behandish, M., Mirzendehdel, A.M., Nelaturi, S.: A Classification of Topological Discrepancies in Additive Manufacturing. Computer-Aided Design **115**, 206–217 (2019)
2. Casolo, S.: Severe slugging flow identification from topological indicators. Digital Chemical Engineering **4**, 100045 (2022)
3. Dassisti, M., Eslami, Y., Mohaghegh, M.: Raw material flow optimization as a capacitated vehicle routing problem: A visual benchmarking approach for sustainable manufacturing. In: 2017 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI). pp. 168–174 (2017)
4. Edelsbrunner, H., Harer, J.: Computational Topology: An Introduction. American Mathematical Society (2010)
5. Giunti, B., Lazovskis, J., Rieck, B.: DONUT: Database of Original & Non-Theoretical Uses of Topology. `https://donut.topology.rocks` (2022), (Accessed: 2023-01-31)
6. Guo, W., Banerjee, A.G.: Toward automated prediction of manufacturing productivity based on feature selection using topological data analysis. In: 2016 IEEE International Symposium on Assembly and Manufacturing (ISAM). pp. 31–36 (2016)
7. Guo, W., Banerjee, A.G.: Identification of key features using topological data analysis for accurate prediction of manufacturing system outputs. Journal of Manufacturing Systems **43**, 225–234 (2017)

8. Harik, R., Shi, Y., Baek, S.: Shape Terra: Mechanical feature recognition based on a persistent heat signature. Computer-Aided Design and Applications **14**(2), 206–218 (2017)
9. Hsu, Y.M., Jia, X., Li, W., Lee, J.: A Novel Quality Clustering Methodology on Fab-Wide Wafer Map Images in Semiconductor Manufacturing. In: Proceedings of the ASME 2022 17th International Manufacturing Science and Engineering Conference. International Manufacturing Science and Engineering Conference, vol. 2: Manufacturing Processes; Manufacturing Systems. American Society of Mechanical Engineers (June 2022)
10. Huber, S.: Persistent homology in data science. In: Data Science Analytics and Applications, pp. 81–88. Springer Fachmedien Wiesbaden (2021)
11. Kalpakjian, S., Schmid, S.R.: Manufacturing Engineering and Technology. Pearson, seventh edition edn. (2014)
12. Khasawneh, F.A., Munch, E.: Topological data analysis for true step detection in periodic piecewise constant signals. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **474**(2218), 20180027 (2018)
13. Khasawneh, F.A., Munch, E., Perea, J.A.: Chatter Classification in Turning using Machine Learning and Topological Data Analysis. IFAC-PapersOnLine **51**(14), 195–200 (2018), 14th IFAC Workshop on Time Delay Systems TDS 2018
14. Lum, P., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J., Carlsson, G.: Extracting insights from the shape of complex data using topology. Nature Scientific Reports **3**(1236) (2013)
15. Mahler, J., Pokorny, F.T., Niyaz, S., Goldberg, K.: Synthesis of Energy-Bounded Planar Caging Grasps Using Persistent Homology. IEEE Transactions on Automation Science and Engineering **15**(3), 908–918 (2018)
16. Matisoff, B.S.: Manufacturing Engineering: Definition and Purpose, pp. 1–4. Springer Netherlands (1986)
17. Myers, A., Khasawneh, F.A.: Dynamic State Analysis of a Driven Magnetic Pendulum Using Ordinal Partition Networks and Topological Data Analysis. In: Proceedings of the ASME 2020 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, vol. 7: 32nd Conference on Mechanical Vibration and Noise. American Society of Mechanical Engineers (2020)
18. Randolph, J.: A Guide to Writing the Dissertation Literature Review. Practical Assessment, Research, and Evaluation **14**(13) (2009)
19. Singh, G., Memoli, F., Carlsson, G.: Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. Eurographics Symposium on Point-Based Graphics p. 10 pages (2007)
20. Tschuchnig, M.E., Gadermayr, M.: Anomaly Detection in Medical Imaging - A Mini Review. In: Data Science Analytics and Applications, pp. 33–38. Springer Fachmedien Wiesbaden (2022)
21. Wang, B., Lin, C., Inoue, H., Kanemaru, M.: Topological Data Analysis for Electric Motor Eccentricity Fault Detection. In: IECON 2022 48th Annual Conference of the IEEE Industrial Electronics Society. pp. 1–6 (2022)
22. Wong, C.C., Vong, C.M.: Persistent Homology Based Graph Convolution Network for Fine-Grained 3D Shape Segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 7098–7107 (October 2021)
23. Yesilli, M.C., Khasawneh, F.A.: Data-driven and Automatic Surface Texture Analysis Using Persistent Homology. In: 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA). pp. 1350–1356 (2021)