

# Visual Analytics in Process Mining for Supporting Business Process Improvement

Antonia KAOUNI<sup>a</sup>, Georgia THEODOROPOULOU<sup>a,1</sup>, Alexandros BOUSDEKIS<sup>a</sup>,  
Athanasios VOULODIMOS<sup>a</sup>, and Georgios MIAOULIS<sup>a</sup>

<sup>a</sup> Department of Informatics and Computer Engineering, University of West Attica,  
Athens, Greece

**Abstract.** The increasing amounts of data have affected conceptual modeling as a research field. In this context, process mining involves a set of techniques aimed at extracting a process schema from an event log generated during process execution. While automatic algorithms for process mining and analysis are needed to filter out irrelevant data and to produce preliminary results, visual inspection, domain knowledge, human judgment and creativity are needed for proper interpretation of the results. Moreover, a process discovery on an event log usually results in complicated process models not easily comprehensible by the business user. To this end, visual analytics has the potential to enhance process mining towards the direction of explainability, interpretability and trustworthiness in order to better support human decisions. In this paper we propose an approach for identifying bottlenecks in business processes by analyzing event logs and visualizing the results. In this way, we exploit visual analytics in the process mining context in order to provide explainable and interpretable analytics results for business processes without exposing to the user complex process models that are not easily comprehensible. The proposed approach was applied to a manufacturing business process and the results show that visual analytics in the context of process mining is capable of identifying bottlenecks and other performance-related issues and exposing them to the business user in an intuitive and non-intrusive way.

**Keywords.** Visual analytics, process analytics, process intelligence, visualization, data analytics, business process management

## 1. Introduction

The amount of data recorded in various domains has been growing exponentially [1]. This offers opportunities for algorithmic techniques, but also creates new challenges. The availability of data has extended conceptual modeling as a research field of manually created models with automatic techniques for generating models from data [2]. Process mining is one of these recent extensions. Process mining involves a set of techniques aimed at extracting a process schema from an event log generated during process execution [3][4]

Combining automatic analysis of event logs and visualization methods face several challenges [3]. While automatic algorithms for process mining and analysis are certainly needed to filter out irrelevant data and to produce preliminary results, visual inspection, domain knowledge, human judgment and creativity are needed for proper

---

<sup>1</sup> Corresponding Author, Georgia Theodoropoulou, University of West Attica, 12243 Ag. Spyridonos Athens, Greece; E-mail: gtheodoropoulou@uniwa.gr.

interpretation of the results [1]. Moreover, a process discovery on an event log usually results in complicated and/or huge process models. Most of the current visualizations display the conformance results in their entirety. From a user perspective, displaying the conformance results on a big and complicated process model with many nodes does not provide significant added value [5]. Such techniques turned out to have several problems when addressing processes with little or no structure, where the flow of the activities mostly depends on the decisions taken by the involved people, thus leading to the so-called “spaghetti-like” models, almost incomprehensible for a human analyst [6].

On the other hand, visual analytics has the potential to provide an answer to these problems by proposing a tight integration between automatic techniques and visualization. In the past ten years, a variety of visual analytics methods have been proposed to make machine learning more explainable, trustworthy, and reliable [7][8][9][10]. Similarly to the way that machine learning has coupled with visualization methods, process mining needs to follow the same direction. In this paper we propose an approach for identifying bottlenecks in business processes by analyzing event logs and visualizing the results aiming at supporting human decisions. In this way, we exploit visual analytics in the process mining context in order to provide explainable and interpretable analytics results for business processes without exposing to the user complex process models that are not easily comprehensible.

The rest of the paper is organized as follows: Section 2 describes the theoretical background upon which the proposed approach is based. Section 3 presents the proposed approach for identifying bottlenecks in business processes with the use of visual analytics. Section 4 outlines the implementation of the proposed approach and demonstrates the results from its application to a manufacturing business process. Section 5 concludes the paper and outlines our plans for future work.

## 2. Theoretical Background

The term visual analytics was coined by Jim Thomas to mean “the science of analytical reasoning facilitated by visual interactive interfaces” [11]. Over time, the scope of visual analytics broadened. Now the term refers to a multidisciplinary field that combines elements from human-computer interaction, geo-spatial and temporal data processing, data analysis and statistics [1] [12].

During the last years, machine learning approaches in process mining have gathered an increased research and business research by automating data analytics algorithms for business processes deriving detection and prediction outcomes. However, such approaches require the business analyst to configure and train the models in order to make the most out of them. On the other hand, visualization approaches utilize the power of human perception to simultaneously process large amounts of data [13], thus allowing computer and human working in optimal complementarity with non-intrusive decision support.

In this context, visual analytics exploits both the computational power of the computer and the human’s perception system to facilitate insights and enable knowledge discovery in large and complex bodies of data [13]. Visual analytics in process mining has been identified as a domain with high potential [1] and a key factor for the wide adoption of process mining methods, techniques and software tools [15], still facing several challenges [13].

Traditionally, visual analytics combined visualization with data mining techniques. However, given the maturity of process mining techniques, visual analytics based on process mining has been considered as a significant research direction [3]. Within process mining, performance analysis is used to explore issues such as identifying any bottlenecks in the process and the steps needed to optimize the process [5]. It is typically focused on performance indicators of the time dimension, such as the lead-, service- and waiting time and, as the name implies, is based on a process model [14]. Many commercial and free process mining tools allow doing such analysis; however, the commonly used model notations are not designed to project the time dimension on the model, i.e. changes over time cannot be represented in a comprehensible way [14]. On the other hand, although conformance results enable compliance and performance analysis, there are still some open challenges which could make the analysis difficult [5].

Since visual analytics aims at supporting human reasoning and gaining new insights, the quality of visual analytics approaches is hard to quantify [13]. However, process models extracted from the process mining algorithms aim at showing individual event sequences, conformance of individual event sequences, and variations of individual cases [13]. Therefore, they usually result in “spaghetti-like” process models that are not easily comprehensible by the business user, while they do not usually extract results related to time.

### 3. The Proposed Approach

The proposed approach aims at identifying bottlenecks in business processes by taking advantage of the large amounts of event logs becoming available through various information systems. In order to tackle the complexity of the process models extracted by the event logs, we propose a visual analytics approach that facilitates interpretability and explainability of the results in order to support business decisions. In this way, the human expert is in the loop having the capability of taking informed decisions. The proposed approach incorporates an incremental and iterative way and provides interactive visualization capabilities in order to facilitate the interaction with the human. The proposed approach consists of the following steps:

- (A) **Event Log Extraction:** The starting point is to extract the event logs. An event log is a file including the activities along with a timestamp. Every time a process is executed, a process instance is created. Each process instance is called a case (or trace). Each trace describes the lifecycle of a particular case (i.e., a process instance) in terms of the activities executed [15]. Every case has several ordered activities. The event log required for the proposed approach needs to incorporate the case number, the activity name, the timestamp (end time) and the start time.
- (B) **Filtering event logs:** We need to filter activities, which can be concluded to have faults from event log inspection. Also we can find outliers and filter them out. Filtering is done for two main reasons: cleaning the data or narrowing down the analysis. Filtering may concern removing a process instance (case), adding events, removing events or modifying events [1].
- (C) **Start and End Activities:** In this step, the proposed approach provides the identified start and end activities. In this way, the human may detect variations

in the process execution. For example, in case there are more than one end activities, the business process may suffer from incomplete instances.

- (D) **Start dates of traces and activities:** In this step, the proposed approach provides visualization results about the duration of the various activities. From these results, the human may detect the most long-lasting activities and how their duration changed throughout the time period under examination.
- (E) **Duration of the activities:** In this step, the visualization results provide information about the traces of the shortest and the longest duration in order to enable focusing on the improvement of the potentially problematic traces.
- (F) **Long-lasting traces:** In this step, the visualization results detect the total time of completion of the traces in order to identify the most long-lasting traces as well as their patterns throughout the time period under examination. In addition to that, the human is exposed to a more detailed graph combining the start dates and the activities of each trace.
- (G) **Human cognition:** In the final step, the human aggregates the aforementioned visualization results in order to make decisions about the efficiency of the business processes as well as about actions for improvement. The human decisions are supported by the results of visual analytics making capable of taking advantage of the large amounts of event logs generated by information systems without requiring advanced data analytics and machine learning skills.

#### 4. Implementation of a Visual Analytics Web Application

The proposed approach was implemented as a web application with the Python programming language. For this implementation, we used 3 Python libraries:

- **Pandas** (<https://pandas.pydata.org/pandas-docs/stable/index.html>) for data structuring. In particular, we used Pandas DataFrame which is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns. Pandas DataFrame consists of three principal components, the data, rows, and columns.
- **PM4py** (<https://pm4py.fit.fraunhofer.de/>) for process mining. PM4py provides functionalities for handling and filtering event data, performing process discovery, managing Petri nets and conducting conformance checking. It also performs evaluation and simulation, and provides process statistics capabilities.
- **Plotly** (<https://plotly.com/>) for delivering machine learning and data science to business users. Plotly enables building and deploying analytic web apps using Python, R, and Julia without the use of JavaScript or DevOps.

#### 5. Application to a Manufacturing Business Process

The proposed approach was applied to a manufacturing business process dealing with the control of a plant to increase its overall performance. The plant produces parts made of metal such as spurs, fastener, ball nuts, discs, tubes, wheel shafts, or clamps. To build these parts there are 28 machines for lapping, milling, turning, sinking, wire cutting, turning and milling, laser marking, and round and flat grinding [17] [18]. The

wide adoption of IoT devices, sensors and actuators in manufacturing environments has fostered an increasing research interest on real-time data analytics; however, event logs contain information regarding the whole factory cycle, either they have sensors installed or not, thus, having the credentials to move towards providing an all-around view of manufacturing operations on the shopfloor [19]. In this way, they may provide valuable information for planning and resource allocation [18].

First, according to the steps (A) and (B) of the proposed approach, the event log is extracted and filtered in order to be structured in a way that can feed into the subsequent visual analysis. The dataset contains process data from a production process, including data on cases, activities, resources, timestamps, etc. Table 1 presents a data sample from the extracted event log. As already mentioned, the most important attributes for the proposed approach are: the *Case ID*, the *Start Timestamp*, the *Complete Timestamp*, and the *Activity*. The *Resource* provides additional information related to the Activity and can be used for further analysis.

**Table 1.** Data sample from the event log of the manufacturing business process.

Case ID	Start Timestamp	Complete Timestamp	Activity	Resource
Case 1	1/29/2012 23:24	1/30/2012 5:43	Turning & Milling - Machine 4	Machine 4 - Turning & Milling
Case 1	1/30/2012 5:44	1/30/2012 6:42	Turning & Milling - Machine 4	Machine 4 - Turning & Milling
Case 1	1/30/2012 6:59	1/30/2012 7:21	Turning & Milling - Machine 4	Machine 4 - Turning & Milling
Case 1	1/30/2012 7:21	1/30/2012 10:58	Turning & Milling - Machine 4	Machine 4 - Turning & Milling
Case 1	1/31/2012 13:20	1/31/2012 14:50	Turning & Milling Q.C.	Quality Check 1
Case 1	2/1/2012 8:18	2/1/2012 8:27	Laser Marking - Machine 7	Machine 7- Laser Marking
Case 1	2/14/2012 0:00	2/14/2012 1:15	Lapping - Machine 1	Machine 1 - Lapping
Case 1	2/14/2012 0:00	2/14/2012 1:15	Lapping - Machine 1	Machine 1 - Lapping
Case 1	2/14/2012 9:05	2/14/2012 10:20	Lapping - Machine 1	Machine 1 - Lapping
Case 1	2/14/2012 9:05	2/14/2012 9:38	Lapping - Machine 1	Machine 1 - Lapping

Figure 1 depicts part of the process model derived from the aforementioned event log. As it can be noticed, it suffers from a high complexity making it impossible for the business user to understand the process. However, explainability and transparency in manufacturing is essential in order to inspire confidence that the system is secure, robust, and controllable [20]. This can be achieved with a visual analytics approach.

According to the step (C), the proposed approach presents to the human the start and the end activities. In the scenario under examination, there are several different start and end activities, as shown in Figure 2 and in Figure 3. Four start activities have a high frequency (i.e. *Turning & Milling – Machine 4*, *Turning & Milling – Machine 8*, *Turning & Milling – Machine 5*, *Turning & Milling – Machine 6*). Their context indicates that there are parallel production lines. However, there are also several other start activities. On the other hand, there are mainly two end activities (*Packing*, *Final Inspection Q.C.*), although there is a wide distribution on other activities. These facts imply a complex business process with high variety, e.g. depending on the products,

the maintenance operations, the remanufacturing, etc. On the other hand, this relatively unstructured process may indicate that the process either is executed in an inefficient way or it should be redesigned.

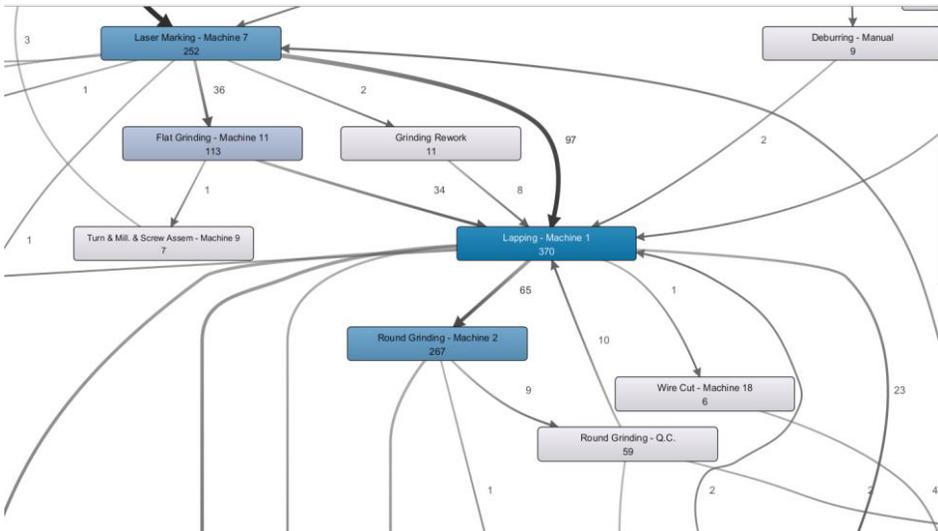


Figure 1. The process model of the manufacturing use case.

According to step (D), the proposed approach provides a Cleveland dot plot in order to result in an intuitive outcome regarding the activity durations through time. Cleveland dot plots are an alternative to bar graphs that reduce visual clutter and can be easier to read. Figure 4 depicts the duration of the various activities with respect to the start timestamp. Each data point has a different colored coding according to the activity it represents. In addition, Figure 5 depicts the mean duration of each activity enabling the detection of the most long-lasting activities.

Start Activities

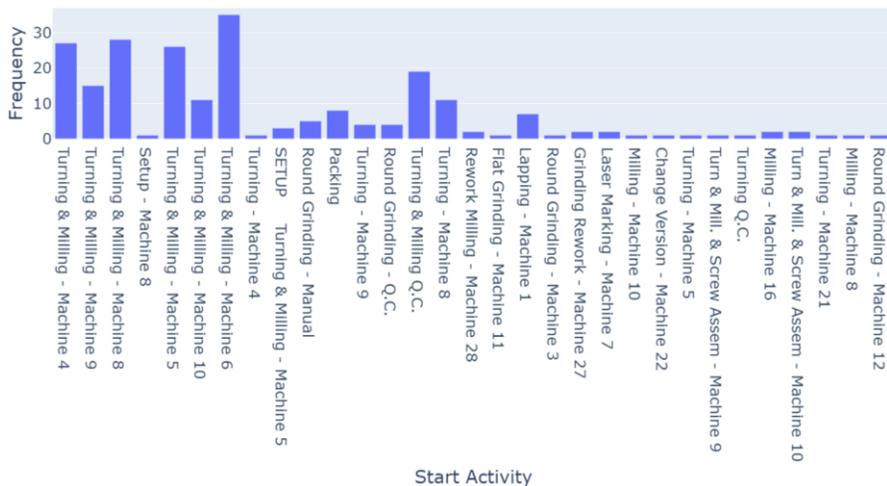
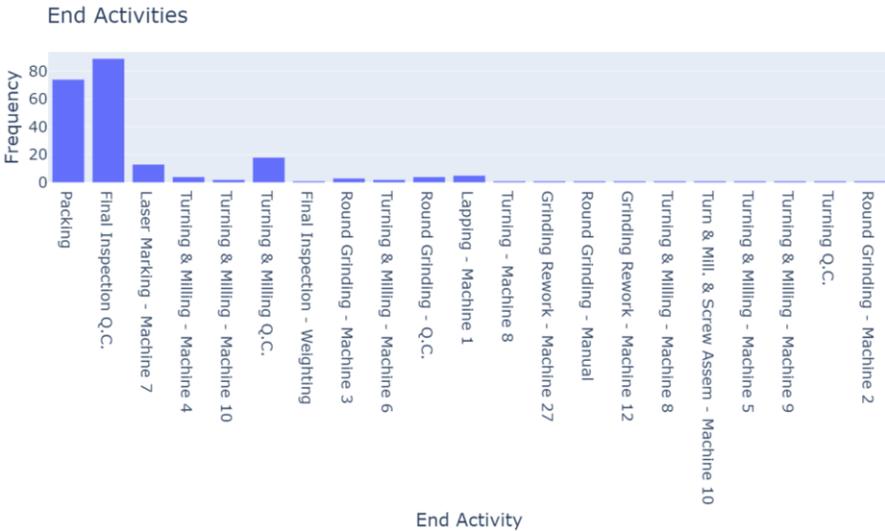


Figure 2. The start activities.



**Figure 3.** The end activities.

According to steps (E) and (F), Figure 6 depicts the total time for completion of the traces so that the business user is able to see the traces that required more time. For example, in the scenario under examination, one may notice that the traces tend to require much more time to be finished throughout time. Moreover, Figure 7 depicts a combination between the start timestamps of each trace and its activities.

According to step (G), the human decisions are supported by the aforementioned visual analytics on the manufacturing business process. These decisions may deal with both strategic and operational aspects of the business process. Therefore, for the scenario under examination, the business user may draw the following conclusions:

- The process under examination is subject to a high variety, since there are several different start and end activities. To some extent, this is reasonable due to the nature of the production network. However, the event log includes some start and end activities that seem to be outliers. Therefore, there are incomplete processes due to either the process execution or the process design itself.
- There are some variations with respect to the mean duration of similar activities. For example, the activities *Milling - Machine 8* and *Milling - Machine 16* have a significantly longer mean duration comparing to the activity *Milling - Machine 10*. Similarly, the activity *Grinding Rework - Machine 2* has a significantly longer mean duration comparing to the activity *Grinding Rework - Machine 12*.
- Some activities have a significantly longer duration comparing to others causing bottlenecks and leading to delays of the whole process.
- Some traces require more time to be completed comparing to others. Throughout time, the traces tend to have a longer duration revealing an increasing inefficiency and responsiveness in the process execution.

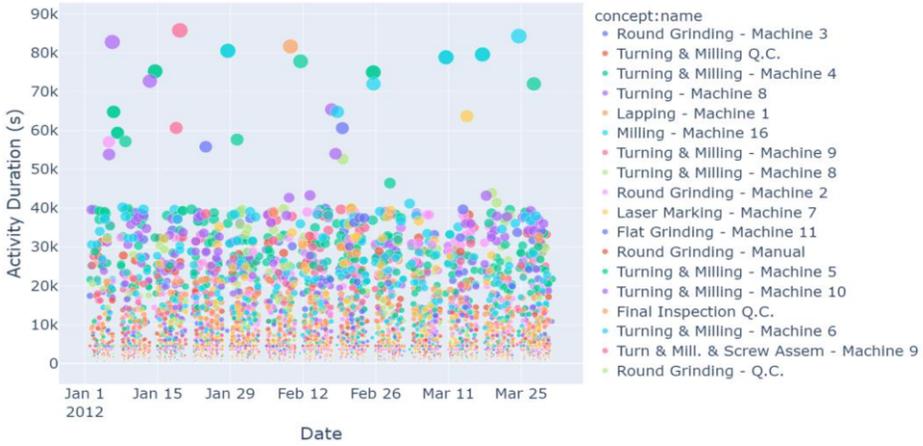


Figure 4. Activity duration through time (step D).

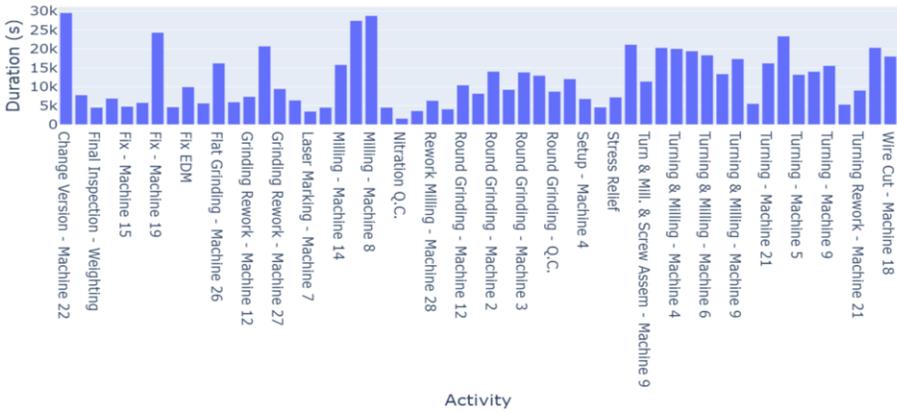


Figure 5. Mean duration of each activity (step D).

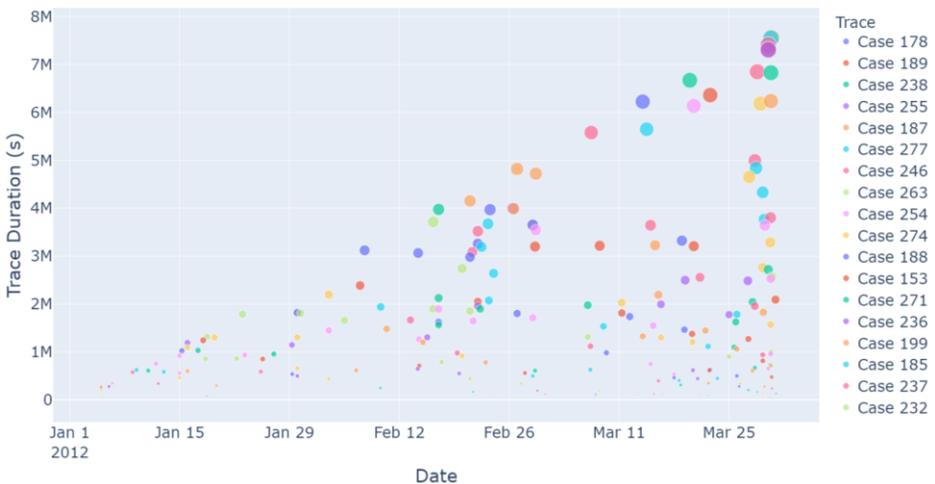


Figure 6. Trace durations through time (step E and F).

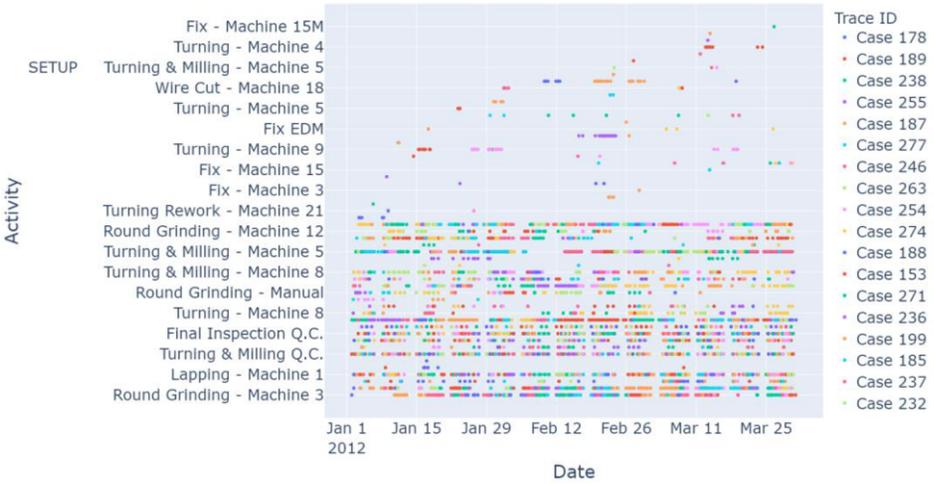


Figure 7. Activities and traces through time (step E and F).

## 6. Conclusions and Future Work

The increasing amounts of data have affected conceptual modeling as a research field by utilizing data analytics for generating process models. In this context, process mining involves a set of techniques aimed at extracting a process schema from an event log generated during process execution. While automatic algorithms for process mining and analysis are needed to filter out irrelevant data and to produce preliminary results, visual inspection, domain knowledge, human judgment and creativity are needed for proper interpretation of the results. Moreover, a process discovery on an event log usually results in complicated and/or huge process models not easily comprehensible by the business user. To this end, visual analytics has the potential to enhance process mining towards the direction of explainability, interpretability and trustworthiness in order to support human decisions.

In this paper we proposed an approach for identifying bottlenecks in business processes by analyzing event logs and visualizing the results. In this way, we exploit visual analytics in the process mining context in order to provide explainable and interpretable analytics results for business processes without exposing to the user complex process models that are not easily comprehensible. The proposed approach was applied to a manufacturing business process and the results show that visual analytics in the context of process mining is capable of identifying bottlenecks and other performance-related issues and exposing them to the business user in an intuitive and non-intrusive way.

Regarding our future work, we plan to implement advanced data analytics algorithms performing predictions on event logs. To do this, we will implement (deep) machine learning algorithms and we will focus on coupling them with visualization methods in order to achieve explainability and interpretability of complex algorithms and results.

## References

- [1] Van der Aalst, W. M., de Leoni, M., & ter Hofstede, A. H. (2011). Process mining and visual analytics: Breathing life into business process models. *BPM Center Report BPM-11-15*, *BPMcenter.org*, **17**, 699-730.
- [2] Yeshchenko, Anton & Di Ciccio, Claudio & Mendling, Jan & Polyvyanyy, Artem (2019), Comprehensive Process Drift Detection with Visual Analytics.
- [3] Van Der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., ... & Wynn, M. (2011b, August). Process mining manifesto. In International Conference on Business Process Management (pp. 169-194). Springer, Berlin, Heidelberg.
- [4] Van Der Aalst, W. (2016). *Data science in action*. In *Process mining* (pp. 3-23). Springer, Berlin, Heidelberg.
- [5] Buijs, J. C. A. M., & van der Aalst, W. M. P. (2017). Enabling interactive process analysis with process mining and visual analytics. *BIOSTEC 2017*, 573.
- [6] Diamantini, C., Genga, L., Potena, D., & van der Aalst, W. M. (2015). *Towards Process Instances Building for Spaghetti Processes*. In SEBD (pp. 256-263).
- [7] Cui, W. (2019). Visual analytics: A comprehensive overview. *IEEE Access*, **7**, 81555-81573.
- [8] Yuan, J., Chen, C., Yang, W., Liu, M., Xia, J., & Liu, S. (2021). A survey of visual analytics techniques for machine learning. *Computational Visual Media*, 1-34.
- [9] Ahn, Y., & Lin, Y. R. (2019). Fairsight: Visual analytics for fairness in decision making. *IEEE transactions on visualization and computer graphics*, **26**(1), 1086-1095.
- [10] Krak, I., Barmak, O., & Manziuk, E. (2020). Using visual analytics to develop human and machine - centric models: A review of approaches and proposed information technology. *Computational Intelligence*.
- [11] Thomas, J. J., & Cook, K. A. (2005). *Illuminating the path*. Los Alamitos, Calif, *IEEE Computer Society*.
- [12] Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., & Ziegler, H. (2008). Visual analytics: Scope and challenges, *Visual data mining* (pp. 76-90). Springer, Berlin, Heidelberg.
- [13] Gschwandner, T. (2015, December). Visual Analytics Meets Process Mining: Challenges and Opportunities, *International Symposium on Data-Driven Process Discovery and Analysis* (pp. 142-154). Springer, Cham.
- [14] Denisov, V., Belkina, E., Fahland, D., & van der Aalst, W. M. (2018, September). The Performance Spectrum Miner: Visual Analytics for Fine-Grained Performance Analysis of Processes. In *BPM (Dissertation/Demos/Industry)* (pp. 96-100).
- [15] Wong, P. C., Shen, H. W., Johnson, C. R., Chen, C., & Ross, R. B. (2012). The top 10 challenges in extreme-scale visual analytics. *IEEE computer graphics and applications*, **32** (4), 63-67.
- [16] Wil M. P. Aalst, 2015. "Extracting Event Data from Databases to Unleash Process Mining," *Management for Professionals*, in: Jan vom Brocke & Theresa Schmiedel (ed.), *BPM - Driving Innovation in a Digital World, edition 127*, pages 105-128, Springer.
- [17] D. Levy, "Production Analysis with Process Mining Technology. Dataset," 2014. [Online]. Available: [https://data.4tu.nl/repository/collection:event\\_logs](https://data.4tu.nl/repository/collection:event_logs)
- [18] Tello-Leal, E., Roa, J., Rubiolo, M., & Ramirez-Alcocer, U. M. (2018, November). Predicting activities in business processes with LSTM recurrent neural networks. In 2018 *ITU Kaleidoscope: Machine Learning for a 5G Future* (ITU K) (pp. 1-7). IEEE.
- [19] Lepenioti, Katerina & Pertselakis, Minas & Bousdekis, Alexandros & Louca, Andreas & Lampathaki, Fenareti & Apostolou, Dimitris & Mentzas, Gregoris & Anastasiou, Stathis. (2020). Machine Learning for Predictive and Prescriptive Analytics of Operational Data in Smart Manufacturing. 10.1007/978-3-030-49165-9\_1.
- [20] Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). A human cyber physical system framework for operator 4.0–artificial intelligence symbiosis. *Manufacturing Letters*, **25**, 10-15.