

IMPressED: Outcome-Oriented Interactive Multi-Interest Process Pattern Discovery Tool

Mozhgan Vazifehdoostirani^{1,*†}, Laura Genga^{1,†}, Xixi Lu^{2,†} and Remco Dijkman^{1,†}

¹Eindhoven University of Technology, Eindhoven, the Netherlands

²Utrecht University, Utrecht, the Netherlands

Abstract

Process pattern discovery methods (PPDMs) have been developed with the primary goal of identifying patterns of interest to users. Existing PPDM approaches are predominantly unsupervised and tend to focus on a single dimension of interest, such as discovering frequent patterns. We present IMPressed, an interactive tool for exploring process patterns leveraging a multi-dimensional notion of interest. IMPressed is designed to identify patterns that align with complex analytical objectives, such as deriving process patterns that affect the process outcome. Incorporating an iterative and interactive approach, this tool collaborates with domain experts to enhance pattern discovery.

Keywords

Process Pattern Discovery, Multi-interest Pattern Detection, Outcome-Oriented Process Patterns

1. Introduction

Process pattern discovery methods (PPDMs) aim to discover process patterns that are *of interest* for the human analyst. The interest of a pattern is usually computed according to one or more functions. Previous studies highlighted how these techniques often uncovered interesting behaviors that would otherwise remain hidden in start-to-end process models [1]. While various techniques have been proposed for discovering process patterns from event logs [1], most of them concentrate on a single interest dimension. This approach can lead to discovering numerous uninteresting patterns and missing valuable but infrequent ones [2]. Recent pattern mining research emphasizes patterns' multi-dimensional nature [3], relevant in process analysis due to the interaction of various factors [4]. A few PPDMs introduced a broader notion of interest by allowing users to define cut-off thresholds for various metrics [1] or using a composite metric during pattern generation [2]. However, these approaches have limitations in handling multi-dimensional pattern interest. Defining appropriate cut-off thresholds for conflicting metrics is a challenging decision that significantly affects results. Moreover, aggregating multiple dimensions into one obscures the interplay of different dimensions, which is especially critical in the presence of conflicting metrics. To address these challenges, we proposed a *multi-objective*


*Corresponding author.

†These authors contributed equally.

✉ m.vazifehdoostirani@tue.nl (M. Vazifehdoostirani); l.genga@tue.nl (L. Genga); x.lu@uu.nl (X. Lu); r.m.dijkman@tue.nl (R. Dijkman)



© 2023 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)

approach for process pattern detection.

Beyond the *multi-objective* challenge, most unsupervised PPDMs encounter pattern explosion in real-world event logs. Prior research suggests leveraging expert domain knowledge through an interactive setting, allowing the users to select and extend process patterns manually, can mitigate this issue [5]. However, this approach relies on frequency-based metrics and burdens users with manual tasks lacking sufficient guidance.

In this demo paper, we introduce the implementation of IMPresseD (**I**nteractive **M**ulti-interest **P**rocess **P**attern **D**iscovery) tool for multi-interest process pattern discovery. IMPresseD is designed to identify patterns fulfilling a complex, multi-dimensional notion of interest, thus supporting complex analytical objectives. In particular, in its current implementation, the tool discovers patterns affecting the process outcome, taking into account different interest functions to tackle the complex and multi-dimensional nature of the problem. To the best of our knowledge, most outcome-oriented pattern detection approaches do not support a multi-dimensional analysis. The main functionalities of the tool are described in the following section.

2. Innovation and functionality

IMPresseD, a Python-based tool implementing the framework introduced in [6], supports outcome-driven process pattern discovery through two distinct modes: interactive and fully automatic. There are variations in the output between the two modes, while both share a common set of core functionalities: multi-interest function analysis, pattern selection, and pattern extension.

- **Multi-interest Function:** Unlike prior outcome-oriented studies that mainly emphasized outcome correlation [7], IMPresseD argues for a broader perspective. First, we incorporate frequency alongside correlation-based interest. Ignoring the frequency measure may lead to identifying rare patterns that are often less interesting. In addition, frequent patterns that are not highly correlated may still be worth exploring. Moreover, it is well-known that potential confounding variables may play an important role in determining the outcome of a treatment process [8]. Consider a treatment pattern P_1 that negatively impacts outcomes. If P_1 is mainly applied to elderly patients, the age factor could actually be driving the results. To mitigate the effect of confounding variables, we consider the distance between cases with or without a specific pattern as the third interest dimension.
- **Pattern selection:** In practice, optimizing all objective functions simultaneously is often unattainable. Hence, our aim is to identify the *Pareto Front*, comprising patterns that are not dominated by any others concerning the multiple interest functions. This functionality empowers users to concentrate their efforts on these non-dominated patterns instead of exploring less interesting patterns.
- **Pattern extension:** The tool employs an iterative strategy for pattern construction, commencing with single-activity patterns and subsequently extending the most promising ones. This extension involves exploring the existing relationships between the selected patterns and other activities within the event log.

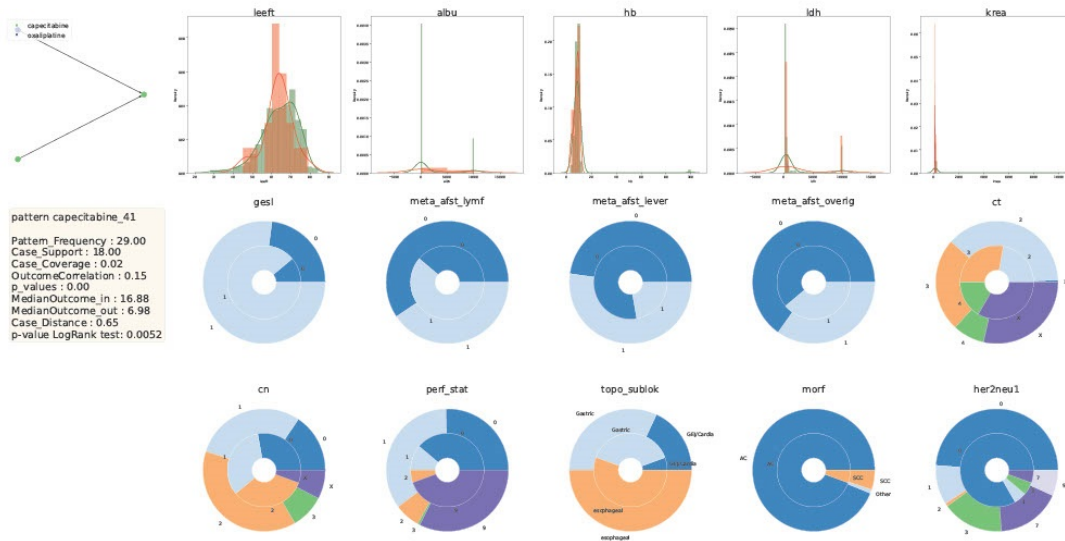


Figure 1: Dashboard visualization result example

2.1. Interactive mode

In addition to the functionalities mentioned earlier, there is a visualizations and user-interaction functionality specifically designed for the interactive mode. In this setting, selected patterns based on the Pareto Front are visualized, and a Python interface is designed to allow users to interact with the pattern discovery algorithm by selecting their desired pattern for extension in the next iteration. Figure 1 represents a visualization example provided by the tool for a healthcare case study in [6]. Moreover, Figure 2 demonstrates the interface designed for users to select their desired patterns from the Pareto Front. This interface allows users to make informed decisions, considering the values of each interest function, thus facilitating a more interactive and intuitive pattern exploration and extension process.

2.2. Automatic mode

In the automatic setting, all patterns chosen from the Pareto Front will automatically undergo extension in the subsequent iteration. In this mode, users specify the maximum number of iterations and initiate the pattern discovery process by clicking the designated button. This mode is optimized to uncover all relevant patterns for prediction purposes.

Note that in the automatic mode, the emphasis is on generating patterns and encoding them in the event log using frequency-based encoding. These encoded patterns serve as new features for each case, supporting the construction of outcome prediction models through machine learning algorithms. Thus, instead of visualizing the patterns, the tool outputs encoded patterns in training and testing sets and also individually in .json format.

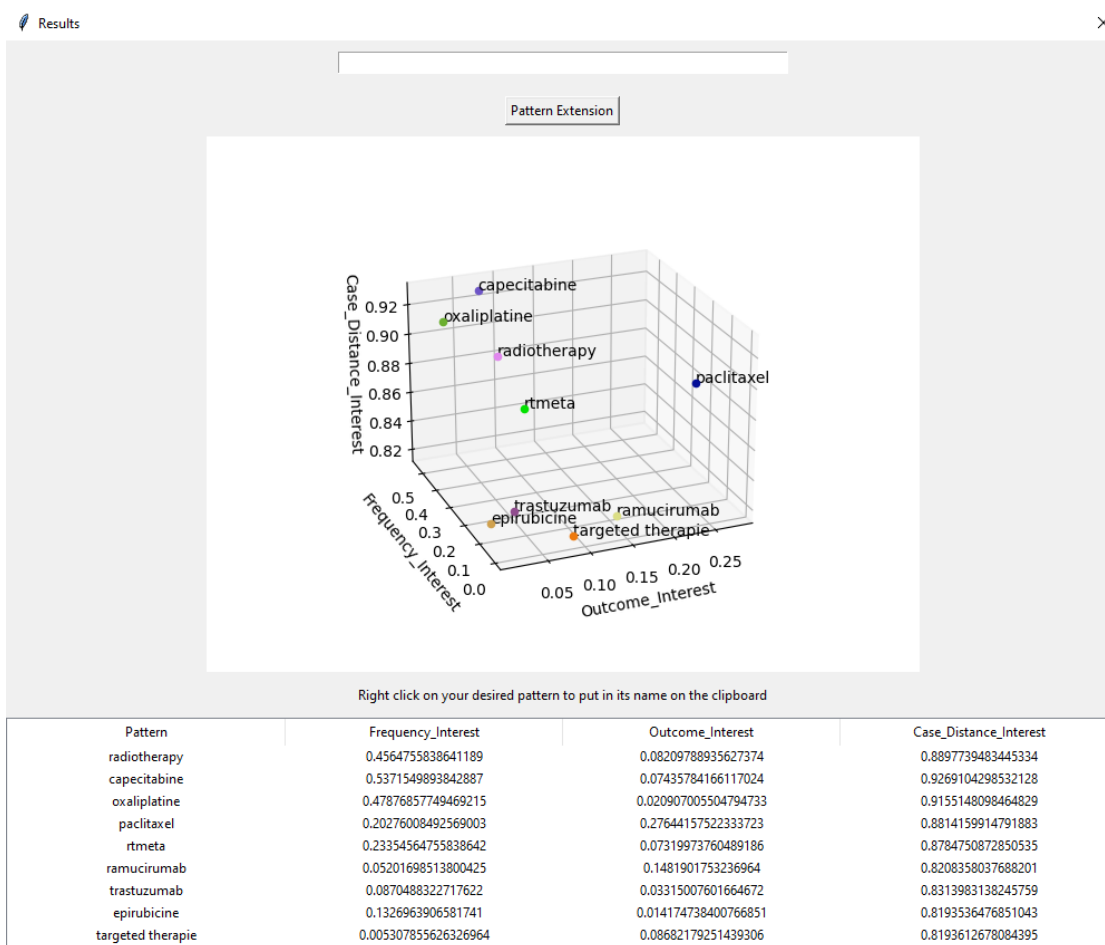


Figure 2: Pareto front visualization in tool user interface

3. Maturity and availability

The existing version of the tool is accessible online via GitHub¹. The tool has been used for analyzing private healthcare data provided by the Netherlands Cancer Registry (NCR) regarding the treatment process for patients with metastatic stomach or esophageal cancer. The tool has been evaluated in a real case study involving healthcare experts for the interactive mode. Furthermore, the automatic discovery version of the tool has been used for extracting outcome-oriented process patterns from publicly available datasets, namely *BPIC2012*, *BPIC2011*, and *Production*. These patterns have been encoded and leveraged for outcome prediction tasks. Our research demonstrates that utilizing these discovered patterns leads to prediction performance that is either comparable to or superior to using all potential patterns. Results of the above-mentioned case studies are reported in [6].

To get a firsthand look at how the tool operates in both interactive and automatic modes, we

¹<https://github.com/MozhganVD/InteractivePatternDetection>

have prepared a video demonstration that can be found online².

4. Conclusion and future work

This demo paper presents the implementation of the IMPresseD framework introduced in [6]. The tool supports the user to 1) interactively discover patterns affecting the process outcome and 2) discover all outcome-oriented patterns for prediction purposes. In future work, we plan to implement custom definitions of the interest functions. Furthermore, we intend to test the tool in other real-world case studies. Additionally, we intend to explore additional extension operators to discover more complex patterns.

References

- [1] N. Tax, N. Sidorova, R. Haakma, W. M. van der Aalst, Mining local process models, *Journal of Innovation in Digital Ecosystems* 3 (2016).
- [2] N. Tax, B. Dalmas, N. Sidorova, W. M. van der Aalst, S. Norre, Interest-driven discovery of local process models, *Information Systems* 77 (2018) 105–117.
- [3] W. Fang, Q. Zhang, J. Sun, X. Wu, Mining high quality patterns using multi-objective evolutionary algorithm, *IEEE Transactions on Knowledge and Data Engineering* 34 (2020) 3883–3898.
- [4] D. Fahland, Multi-dimensional process analysis, in: *Business Process Management: 20th International Conference, BPM 2022, Springer, 2022*, pp. 27–33.
- [5] X. Lu, D. Fahland, R. Andrews, S. Suriadi, M. T. Wynn, A. H. ter Hofstede, W. M. van der Aalst, Semi-supervised log pattern detection and exploration using event concurrence and contextual information, in: *OTM Confederated International Conferences” On the Move to Meaningful Internet Systems”*, 2017, pp. 154–174.
- [6] M. Vazifehdoostirani, L. Genga, X. Lu, R. Verhoeven, H. van Laarhoven, R. Dijkman, Interactive multi-interest process pattern discovery, in: C. Di Francescomarino, A. Burattin, C. Janiesch, S. Sadiq (Eds.), *Business Process Management, Springer Nature Switzerland, Cham, 2023*, pp. 303–319.
- [7] H. Nguyen, M. Dumas, M. La Rosa, F. M. Maggi, S. Suriadi, Mining business process deviance: a quest for accuracy, in: *OTM Confederated International Conferences” On the Move to Meaningful Internet Systems”*, 2014, pp. 436–445.
- [8] A. Terada, D. duVerle, K. Tsuda, Significant pattern mining with confounding variables, in: *Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2016*, pp. 277–289.

²<https://www.youtube.com/watch?v=Rrk5La8vwYU>