First Insights into the Impact of Concept Drift on Process Complexity

Maxim Vidgof[1] [0000−0003−2394−2247]

Vienna University of Economics and Business
Welthandelsplatz 1, 1020 Vienna, Austria
maxim.vidgof@wu.ac.at
https://complex.wu.ac.at/nm/vidgof

Abstract. Process mining is concerned with extracting knowledge about business processes based on digital traces left during their execution. Most existing process mining methods are focused on stable processes. However, processes tend to change over time in response to changing regulations or market conditions but also due to internal dynamics. Concept drift is a phenomenon in process mining when a process changes during the timeframe when it is analyzed. Following the change in how a process is executed, process complexity also changes over time. Despite this apparent connection, the relationship between these concepts have not been explicitly studied. In this paper, some theoretical considerations on how concept drift can be reflected in process complexity are presented. These considerations are then evaluated using artificial and real-life event logs. The results support the considerations and invite a multitude of questions for future research.

Keywords: Process complexity · Concept drift · Process mining

1 Introduction

Business processes change over time. This change can be caused by a number of reasons, both internal to the process and external. New regulations or change in the market can cause the processes to add activities or variants. Process optimization and streamlining initiatives, on the other hand, can eliminate redundant activities and shrink the variability of the process. Even without any top-down incentive, processes can still change as process participants optimize their work, find workarounds, or change preferences from one path to the other. All these changes in the way a process is executed fall under the category of concept drift.

Another important aspect of any business process is its complexity. Complexity of business processes has multi-faceted influence on their execution and ultimately process performance [11]. As processes change over time, so does process complexity. However, the relationship between these two concepts have not yet been made explicit and studied rigorously. In this paper, the first considerations about the impact of concept drift are presented and evaluated with a simple artificial scenario as well as a real-life event log.

2 Background

2.1 Concept Drift

Concept drift is a known problem in data mining. It describes the case where the relationship between the input and the target variable changes over time [9]. In process mining, it refers to a situation where the process changes over time.

Drift detection techniques analyze event logs or streams of events in search of concept drift. The simplest techniques merely report whether a drift was detected or not, mode advanced ones can also report when the drift happened, what type of drift it was and which part of the process was affected. Most techniques apply statistical hypothesis testing, i.e. they compute some characteristics of the process in different time windows and compare them using statistical tests. If the difference is significant, a drift is reported. Other techniques rely on trace clustering and detect drift by detecting the changes in cluster composition over time. Change point detection methods convert event logs into multivariate time series and identify changes in their values [14].

There are several aspects of concept drift . First, one can distinguish between momentary and permanent drift. The former is often considered an outlier and is filtered out, while most techniques focus on the latter. Another aspect is the type of drift. One can distinguish between *sudden*, *gradual*, *incremental* and *recurring* drift [3]. As process mining deals with multiple perspectives, also concept drift can be related to one (or more) of the following perspectives: *control-flow*, *time*, *resource* and *data*. While there are now emerging approaches to study e.g. drifts in resource perspective [5], most approaches focus on the control-flow perspective. [12] defines the following patterns of concept drift in imperative process models: adding/deleting fragments, moving/replacing fragments, adding/removing levels, adding control dependencies, changing transition conditions.

2.2 Process Complexity

Complexity of business processes manifests itself in multiple ways [8]. There have been several approaches to quantitatively assess process complexity, however, they either relied on perceptual measures or on complexity of process models. Recently, event log complexity has received increased attention as it uses evidence of how process was really executed to numerically assess process complexity. It can be distinguished between four categories of event log complexity [1,11]. Complexity measures related to *size* capture the number of traces, events, event classes and lengths of traces. *Variation*-related metrics capture the number of variants of how process can be executed. *Distance*-related metrics try to numerically assess the degree to which these variants differ. Finally, *entropy*-based metrics try to unify these three aspects. Most complexity metrics in these streams relate to control flow, however some approaches to also incorporate data complexity have emerged [10]. Some studies have already delved into change of processes over time and connected change of complexity [11,7,13]. However, none of these papers has explicitly connected process complexity with concept drift. Also, [7] used random drifts in the process, while the observations of business processes claim that the real drift patterns are more limited.

3 Theoretical Considerations

Concept drift implies that there are two versions of the process: *before* and *after* the change point. The claim of this paper is that a drift in control-flow perspective is connected to change in at least one of the complexity metrics. The following observations outline the expected connection in more detail.

Observation 1. If activities are added or removed from the process, this will be captured by size-related complexity metrics, i.e. size-related complexity values will differ in two states.

Observation 2. If process variant is added or removed, this will be captured by variation-related complexity metrics.

Observation 3. If process variants are added or removed, or some process variants are being changed while others remain intact, this will be captured by distance-related complexity metrics.

Observation 4. During the drift, when both process versions coexist, variation-related metrics will see bursts with higher values than both before and after the drift, since they consider both variants from the old process as well as variants from the new process.

Thus, for instance, especially gradual drifts increase the number of variants observable in transition period, which will be captured by variation-related metrics. After the old version is replaced, however, complexity will drop again. In case of recurring drift, these dynamics will be observed multiple times if the windows are granular enough. Otherwise, with coarse-grained windows, it may appear that variant-related complexity is permanently high.

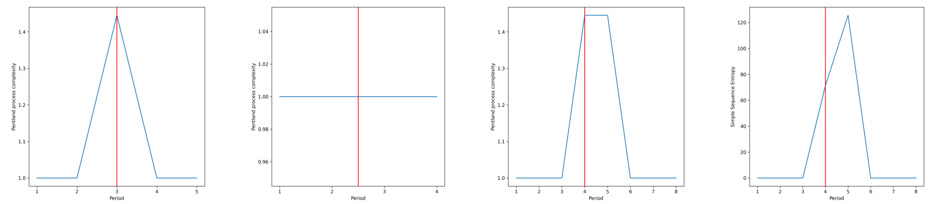
4 Evaluation

In order to empirically evaluate the connection between process complexity and concept drift, both artificial and real-life event logs were used. This section describes the evaluation scenario and the observed results.

4.1 Artificial Logs

As a starting point, this work uses the first pattern described in Section 2.1, namely deleting the activities. The initial process model is a simple order-to-cash process described in [2]. It is a sequential process with single variant having 6 activities. Three of these activities are removed during the drift. The simulations were performed using CDLG tool [4]. For the sake of brevity, only a brief excerpt of the results is presented, however, full results together with the process model are available on GitHub[1].

[1] `https://github.com/MaxVidgof/complexity-drift-data`



(a) Pentland process complexity (sliding windows, size 40)

(b) Pentland process complexity (sliding windows, size 50)

(c) Pentland process complexity (sliding windows, size 50, step 20)

(d) Sequence entropy (sliding windows, size 50, step 20)

Fig. 1: Complexity over multiple periods in presence of sudden drift.

As activities are removed, size-related complexity metrics report lower values after the drift. Bursts in variation-related complexity can indeed be seen. However, only if window size is such that some windows include both old and new version of the process. In case of gradual drift, the entire drift period has higher variation-related complexity. Entropy-based complexity can offer a more detailed view as is becomes more evident when and how many of the new versions are added and when old versions start to die out. In incremental drift, every drift point is associated with a burst in variation-related complexity.

4.2 Real-life Logs

In the Italian helpdesk log [2], there are 2 known drifts: on July 25th 2011, and on September 11th 2012 [6, p. 45]. For the evaluation, this event log was split into non-overlapping monthly windows by trace start. The change in complexity observed in this event log is multi-faceted, however some complexity metrics seem to correlate with drifts. For instance, magnitude and support (size-related metrics) both drop around the drift points. Average trace length drops at the first drift and recovers at the second, while staying low in between. Both drifts are also characterized by abrupt falls in Pentland task complexity and Lempel-Ziv complexity (variation-related metrics) around the drift points, signalling a reduction in the variance of the process around these timestamps. This is also mirrored by the lower number of distinct traces.

4.3 Discussion

Even the first step of the evaluation allowed to confirm some of the observations and to make further ones. As in the artificial log the drift was related to number of activities, size-related metrics are different before and after the drift, confirming Observation 1. Bursts in complexity predicted by Observation 4 can indeed be observed, however, under some additional conditions. These bursts are only

[2] `https://data.4tu.nl/articles/_/12675977/1`

seen when in some window both the old and the new versions of the process coexist. This, however, cannot be guaranteed by non-overlapping windows. Thus, in contrast to concept drift detection techniques that benefit from non-overlapping windows, observation of process complexity over time should be done with overlapping windows. This also provides the explanation to bursts in complexity observed in [7]: such bursts are present only for variation-related complexity and only if windowing strategy captures the coexistence of the process versions.

5 Conclusion

Concept drift is a phenomenon where a process changes while being analyzed. These changes have been classified, and algorithms have been developed for their detection. However, the relationship between concept drift and complexity of business processes has not been studied. In this paper, a first step has been made towards explaining this connection. Preliminary results support the expected relation, however, many open questions for further exploration remain.

5.1 Future Work

This paper presents several directionf for future work. First, it is planned to extend the evaluation on artificial logs in multiple directions. This includes testing multiple drift types described in [3], multiple scenarios described in [12], employing further windowing strategies, including cumulative complexity measurements similar to [10], as well as evaluating on further existing artificial logs [9].

Second, more thorough evaluation using real-life logs is planned. This includes not only employing more event logs with known drifts but also more thorough data preparation (e.g. noise filtering) and more detailed analysis of the results.

Third, it is worth explicitly incorporating the model complexity into this discussion. In presence of concept drift, the models discovered from different observation periods may not only differ in their structure and content but also in their complexity, which would add another interesting angle to this phenomenon.

Fourth, given the observed connection, it may be reasonable to also develop a drift detection technique based on process complexity or at least incorporate process complexity as one of the variables into drift detection techniques based on statistical tests or multivariate time series analysis.

Finally, the opposite direction of this connection can be studied, as, for instance, increased complexity of the process may lead to its intentional or unintentional simplification, i.e. concept drift.

References

1. Augusto, A., Mendling, J., Vidgof, M., Wurm, B.: The connection between process complexity of event sequences and models discovered by process mining. Inf. Sci. **598**, 196–215 (2022). `https://doi.org/10.1016/J.INS.2022.03.072`

2. Dumas, M., Rosa, M.L., Mendling, J., Reijers, H.A.: Fundamentals of Business Process Management, Second Edition. Springer (2018)
3. Gama, J., Zliobaite, I., Bifet, A., Pechenizkiy, M., Bouchachia, A.: A survey on concept drift adaptation. ACM Comput. Surv. **46**(4), 44:1–44:37 (2014). https://doi.org/10.1145/2523813
4. Grimm, J., Kraus, A., van der Aa, H.: CDLG: A tool for the generation of event logs with concept drifts. In: Janiesch, C., Francescomarino, C.D., Grisold, T., Kumar, A., Mendling, J., Pentland, B.T., Reijers, H.A., Weske, M., Winter, R. (eds.) Proceedings of the Best Dissertation Award, Doctoral Consortium, and Demonstration & Resources Track at BPM 2022 co-located with 20th International Conference on Business Process Management (BPM 2022), Münster, Germany, September 11th to 16th, 2022. CEUR Workshop Proceedings, vol. 3216, pp. 92–96. CEUR-WS.org (2022), https://ceur-ws.org/Vol-3216/paper_241.pdf
5. Klijn, E.L., Mannhardt, F., Fahland, D.: Multi-perspective concept drift detection: Including the actor perspective. In: International Conference on Advanced Information Systems Engineering. pp. 141–157. Springer (2024)
6. Ostovar, A., Leemans, S.J.J., Rosa, M.L.: Robust drift characterization from event streams of business processes. ACM Trans. Knowl. Discov. Data **14**(3), 30:1–30:57 (2020). https://doi.org/10.1145/3375398
7. Pentland, B.T., Liu, P., Kremser, W., Hærem, T.: The dynamics of drift in digitized processes. MIS Q. **44**(1) (2020), https://misq.org/the-dynamics-of-drift-in-digitized-processes.html
8. Revina, A., Aksu, Ü., Meister, V.G.: Method to address complexity in organizations based on a comprehensive overview. Inf. **12**(10), 423 (2021). https://doi.org/10.3390/info12100423
9. Sato, D.M.V., Freitas, S.C.D., Barddal, J.P., Scalabrin, E.E.: A survey on concept drift in process mining. ACM Comput. Surv. **54**(9), 189:1–189:38 (2022). https://doi.org/10.1145/3472752
10. Vidgof, M., Mendling, J.: Leveraging event data for measuring process complexity. In: Montali, M., Senderovich, A., Weidlich, M. (eds.) Process Mining Workshops - ICPM 2022 International Workshops, Bozen-Bolzano, Italy, October 23-28, 2022, Revised Selected Papers. Lecture Notes in Business Information Processing, vol. 468, pp. 84–95. Springer (2022). https://doi.org/10.1007/978-3-031-27815-0_7
11. Vidgof, M., Wurm, B., Mendling, J.: The impact of process complexity on process performance: A study using event log data. In: Francescomarino, C.D., Burattin, A., Janiesch, C., Sadiq, S. (eds.) Business Process Management - 21st International Conference, BPM 2023, Utrecht, The Netherlands, September 11-15, 2023, Proceedings. Lecture Notes in Computer Science, vol. 14159, pp. 413–429. Springer (2023). https://doi.org/10.1007/978-3-031-41620-0_24
12. Weber, B., Reichert, M., Rinderle-Ma, S.: Change patterns and change support features - enhancing flexibility in process-aware information systems. Data Knowl. Eng. **66**(3), 438–466 (2008). https://doi.org/10.1016/J.DATAK.2008.05.001
13. Wurm, B., Grisold, T., Mendling, J., vom Brocke, J.: Measuring fluctuations of complexity in organizational routines. In: Academy of Management Proceedings. vol. 2021, p. 13388. Academy of Management Briarcliff Manor, NY 10510 (2021)
14. Yeshchenko, A., Ciccio, C.D., Mendling, J., Polyvyanyy, A.: Visual drift detection for event sequence data of business processes. IEEE Trans. Vis. Comput. Graph. **28**(8), 3050–3068 (2022). https://doi.org/10.1109/TVCG.2021.3050071