# Re-Thinking Process Mining with Agents in Mind

Nils Erik Flick, Lawrence Cabac, Nicolas Denz and Daniel Moldt

Department of Informatics, University of Hamburg,
Vogt-Kölln-Straße 30, 22527 Hamburg, Germany
http://www.informatik.uni-hamburg.de

**Abstract.** Being able to keep business processes flexible and adapting to changing demands requires knowledge of the processes. The algorithmic generation of explicit models is a step towards making the relevant processes clearly comprehensible and exposing them for reflection.
We believe that processes which occur in practice, notably in organizational and inter-organizational contexts, possess a structure best described by systems of interacting agents.
We intend to carry out a mining of special reference nets which are structured in a multi-agent-like fashion. Process mining of reference nets might allow for structuring beyond mere model aggregation. One possible application of an algorithm providing such interpretations lies in spotting their weaknesses automatically with the aim to support of streamlining of organizational processes.

**Keywords:** agent technology, business process support, process mining, reference nets, workflow mining

## 1 Introduction

This paper is a position paper and a research agenda rather than a compendium of results. It serves to disseminate some ideas about a possible future of process mining at an early stage.

Process mining starts from the assumption that a log of activities contains interesting patterns. These patterns are deemed to be generated by a system whose structure must be recovered. In our opinion, this structure can be regarded as a multi-agent system.

Process mining should therefore be conducted from an explicitly agent-oriented perspective. In most cases, an agent-oriented view is either neglected or only implicitly taken in the current literature.

The remainder of the paper is structured as follows: In Section 2, we will position the present research within the emerging process mining landscape.

We will note the aspects we would like to concentrate on, because we perceive them as under-represented to date. These aspects include region theory and the organizational perspective of process mining.

We will work within the framework of reference nets, as defined in Kummer's monograph [1]. These higher-level Petri nets not only allow for the concise expression of concurrent processes (as do other Petri net formalisms) but lend themselves to formalizing multi-agent systems (MAS). Mining reference nets and mining agents are two sides of a medal, if the concept of MAS fits the target system: general reference nets might be too complex to be simply reconstructed from execution traces; agents must be formalized as nets when the goal is to offer an integrated net mining mechanism. Several essential implications of this multi-level, organizational view are examined in Section 3.

## 2 Relations to Prior Work

First, we note the existence of an algebraic approach for Petri net synthesis, the theory of regions, and its uses in process mining. Next, some aspects of the relationship of process mining to multi-agent systems are discussed and relevant publications noted.

### 2.1 Process Mining and Region Theory

Region theory has been put forward in its earliest form by Ehrenfeucht and Rozenberg [2]. Put simply, it allows the synthesis of Petri nets from their reachability graphs. This objective is closely related to what (part of) process mining strives to accomplish. Unsurprisingly, it has been turned into a promising avenue for traditional process mining, as judged by the growing number of publications.

The methods of van der Aalst et al. [3], van Dongen et al. [4], Carmona et al. [5], and Bergenthun et al. [6] all rely on converting an event log into a transition system in the first step. This automaton is an abstraction of the control flow represented in the log. In the second step, region theory is applied to turn the automaton into a more compact and possibly concurrent Petri net.

Region theory has to our knowledge only been applied to process mining in the control flow perspective. An interesting question is how region theory can contribute to other mining perspectives and which modifications and extensions of the original formalism might be necessary.

### 2.2 Agent Communities and Organizational Process Mining

Remarkably, the process mining community has, so far, treated the sequence and causality aspect more thoroughly than the semantic content of actions, maybe because the resulting structures are mostly limited to simple P/T nets.

In reality, *agents* espousing *roles* actually accomplish the actions in most cases. Arguably the agent metaphor applies in all cases, since even an organization can be modeled as an agent. Logs also commonly identify agents as originators of tasks.

Existing algorithms and implementations can mine the causal dependency between activities and also construct a social network. Minseok Song's work (especially in [7]) already shows the relevant data organized into a graph, a useful

decision which points in the direction of greater integration. Beyond that, the ProM framework [8] provides plugins for several algorithms aiming at organizational mining and putting the roles back into the processes (Bozkaya, Gabriels and Werf [9], Song and van der Aalst [10]), including role hierarchy mining.

Despite these possibilities, the results mined from different perspectives are seldom linked to provide better heuristics for structuring a log. Contributions like [11] and our own previous work [12] are a first step towards aggregating multiple perspectives in a common model. Another important property induced by the agent metaphor is often neglected as well: Current process mining techniques hardly pay attention to *locations* where actions take place. One exception is the work presented in [13] where physical places are considered.

Beyond that, our vision is a holistic approach that integrates these different perspectives in the system-theoretical viewpoint of multi-agent systems, based on the formalism of reference nets.

## 3 Where to Go from Here

Existing process mining methods have one shortcoming which is due to the concentration on the process-centered view *at the exclusion* of other relevant perspectives (see Cabac, Knaak, Moldt and Rölke [12] for one possible classification). There are at least three main points, all unified in a systems-theoretical view of multi-agent systems, where we realize the need for a new approach.

### 3.1 Agents as Coherent Entities

Certain processes are best expressed in terms of agents, not only because the agents execute the actions, but more pertinently because the concept of agents can be important for *structuring* purposes.

An agent, similar to an object in object-oriented programming, is a system characterized, among other properties, by a degree of coherence/cohesion and persistence of information attached to it. This correspondence can be used to uncover not directly observable dependencies by structuring actions according to (known or unknown) agents.

### 3.2 Agents as Situated Entities

Agents exist in an environment that is often defined in terms of distinct logical or physical locations with paths the agents can move on (e.g. [14]). Such topologies influence the behaviour of agents in several respects. Generally, an agent's behaviour might depend on its location. The ability to perceive and act might be restricted to a local radius (behavioural locality, e.g. [15]). Locations might serve as side conditions for the synchronization of agent behaviour as in rendez-vous synchronization (see e.g. [16]). The situatedness of agents is naturally represented in the reference net formalism.

As discussed above, locations and locality are seldom regarded in current process mining techniques. In our opinion, the focus on a nets-within-nets formalism can bring forward the handling of location-related information in process mining. On the one hand, locations and their properties might be reconstructed from a log based on hints of characteristic agent behaviour. On the other hand, available information about locations might provide heuristics to improve the reconstruction of process and organizational models.

### 3.3 Incorporation of Agents into Mining

Since actions are linked to agents, there is much to be gained by incorporating the agent side directly. If this information is ignored, one misses out on potentially crucial clues to understand the process. As an illustration, when a set of agents $\{A_i\}_{i \in I}$ for some $I \subseteq 2^A$ are involved sequentially in a process in certain roles, there must have been a connection between them.

When we observe that an action must happen in a certain sequence with other actions, this means that the agents participating are linked by a network of connections. Thus, the existence of dependencies between actions may establish, or betray, the fact that information must have been propagated.

### 3.4 Recovering Agents

The agent is 'defined' by its surroundings, at least in that it must reflect its relationships internally. Relationships also shape the ways in which it may interact. Recovering the agents could mean either of several things: (1) Building an operational model of the agent, (2) finding their relationships without trying to build such an agent, and (3) recovering a priori missing/hidden information (which will not be possible without further hints from the logs).

There is clearly a possible path for improving on methods which are blind to agents and only consider named activities, and on the existing approaches to role recovery. One could conceivably start from region theory and generalize regions further, as at present they ignore agents completely.

### 3.5 Recovering the Organization

Organization means that interdependencies and hierarchy exist. Some would argue that strongly linked subsystems form organizational units. It is unclear how well a 'strength of connections' to hierarchical clustering approach really helps in structuring the unknown domain of agents.

The intuition of clustering by separating weekly clusters can be misleading. Distance measures must be chosen with care with respect to the intended definition of 'similarity'. The agent metaphor provides several hints for similarity including similar behaviour, knowledge, frequency of communication, etc.

One can imagine a number simple examples: two secretaries from different companies who have a logical platform in common and frequently communicate

with each other, are closely linked but at the same time belong to separate organizations. System boundaries thus depend on semantic decisions.

The converse argument that such a naïve analysis can lead to a prejudice-free analysis of the situation is also valid. In a business setting, such a clustering result is useful as far as the goal is to examine the processes and the performance of the existing system on the technological/software side.

Opaque role annotations do not, per se, mean that one understands the relationships existing between the participants. Rather, mining this would be a step forward in doing *meaningful* process understanding (cf. roles as objects). Detecting the kind of relationship is possible because of the patterns generated.

Usually, the process perspective captures the dynamic behavior of the system, whereas the mining of more static, *structural* properties is also of value. There are prototypical kinds of relationship between such entities (see e.g. Jennings and Wooldridge [17]), which leave characteristic traces in the execution log.

## 4 Conclusions and Outlook

It must be stated that process mining is currently only at the beginning and many aspects have not yet been formally explored.

We have discussed the importance of considering agents as parts of processes in process mining. We argued in favor of integrating the mining of processes, the interrelated network of agents generating them, and the environment in which the agents are located. The agent metaphor naturally leads to the addressing of different perspectives beyond control flow mining as claimed in [18] and [19].

Reference nets might make the right kind of structuring possible. In our reference net-based MAS architecture Mulan [20], protocols, agents, and platforms are all nets, and as such first-order entities that can be reasoned about explicitly in process mining. They can represent other supplementary information as well, such as physical places, logical platforms and object flow.

Our prime objective at this point is to ascertain the possibility of recovering agents and mining static as well as dynamic relationships between them from sufficiently detailed logs. In case of success, further questions arise. As noted before in [21], dealing with noise, i.e. incorrectly logged information, may be a major issue and it is still not clear how to achieve robustness; this must certainly be addressed in further work. There is certainly a conflict between the uncompromising exactness of algebraic methods, the desire to detect even uncommon process variants and the quality of log data available to the analyst (many activities taking place outside of logs).

## References

1. Kummer, O.: Referenznetze. Logos Verlag, Berlin (2002)
2. Ehrenfeucht, A., Rozenberg, G.: Partial (set) 2-structures. Acta Informatica **27** (1990) 343–368

3. van der Aalst, W.M.P., Rubin, V., van Dongen, B.F., Kindler, E., Günther, C.W.: Process mining: A two-step approach using transition systems and regions. BPM Center Report BPM-06-30, BPMcenter. org (2006)
4. van Dongen, B.F., Busi, N., Pinna, G., van der Aalst, W.M.P.: An iterative algorithm for applying the theory of regions in process mining. In: Proceedings of the workshop on formal approaches to business processes and web services (FABPWS'07). Siedlce: Publishing House of University of Podlasie. (2007)
5. Carmona, J., Cortadella, J., Kishinevsky, M., Kondratyev, A., Lavagno, L., Yakovlev, A.: A symbolic algorithm for the synthesis of bounded Petri nets. Applications and Theory of Petri Nets (2008) 92–111
6. Bergenthum, R., Desel, J., Lorenz, R., Mauser, S.: Process mining based on regions of languages. Business Process Management (2007) 375–383
7. van der Aalst, W.M.P., Song, M.: Mining Social Networks: Uncovering interaction patterns in business processes. Business Process Management (2004) 244–260
8. van Dongen, B.F., de Medeiros, A.K.A., Verbeek, H.M.W., Weijters, A., van der Aalst, W.M.P.: The ProM framework: A new era in process mining tool support. Applications and Theory of Petri Nets 2005 (2005) 444–454
9. Bozkaya, M., Gabriels, J., Werf, J.M.: Process Diagnostics: A Method Based on Process Mining. In: International Conference on Information, Process, and Knowledge Management, 2009. eKNOW'09. (2009) 22–27
10. Song, M., van der Aalst, W.M.P.: Towards comprehensive support for organizational mining. Decision Support Systems **46** (2008) 300–317
11. Rozinat, A., Mans, R.S., Song, M., van der Aalst, W.M.P.: Discovering simulation models. Inf. Syst. **34** (2009) 305–327
12. Cabac, L., Knaak, N., Moldt, D., Rölke, H.: Analysis of multi-agent interactions with process mining techniques. In: Multiagent System Technologies. 4th German Conference, MATES 2006 Erfurt, Germany. Proceedings. Volume 4196 of Lecture Notes in Computer Science., Berlin Heidelberg, Springer-Verlag (2006) 12–23
13. Reijers, H., Song, M., Jeong, B.: On the performance of workflow processes with distributed actors: Does place matter? In Alonso, G., Dadam, P., Rosemann, M., eds.: Business Process Management. Volume 4714 of Lecture Notes in Computer Science. Springer Berlin/Heidelberg (2007) 32–47
14. Meyer, R.: Bewegungsgraphen – Ein Konzept für die räumliche Modellierung der Umgebung in der individuen- und agenten-basierten Simulation . In Wittman, J., Bernard, L., eds.: Simulation in Umwelt- und Geowissenschaften – Workshop Münster, Aachen, Shaker (2001) 47–60
15. Klügl, F.: Multiagentensimulation – Konzepte, Werkzeuge, Anwendung. Agententechnologie. Addison-Wesley, Munich (2001)
16. Jessen, E., Valk, R.: Rechensysteme: Grundlagen der Modellbildung. Studienreihe Informatik. Springer-Verlag, Berlin Heidelberg (1987)
17. Jennings, N.: Agent-oriented software engineering. Multi-Agent System Engineering (1999) 1–7
18. van der Aalst, W.M.P., Weijters, A.: Process mining: a research agenda. Computers in Industry **53** (2004) 231–244
19. Rembert, A.J., Ellis, C.S.: An initial approach to mining multiple perspectives of a business process. In: TAPIA '09: The Fifth Richard Tapia Celebration of Diversity in Computing Conference, New York, NY, USA, ACM (2009) 35–40
20. Rölke, H.: Modellierung von Agenten und Multiagentensystemen: Grundlagen und Anwendungen. PhD Thesis, Faculty of Informatics, University of Hamburg (2004)
21. van der Aalst, W.M.P.: Discovering coordination patterns using process mining. In: Workshop on Petri Nets and Coordination, Bologna, Italy, Citeseer (2004)