

Understanding the life of a knowledge worker

Tadej Štajner^{1,2}

¹ Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia

² Jožef Stefan International Postgraduate School (ICT, 1st year)

tadej.stajner@ijs.si

This poster presents an implementation of data mining techniques to obtain a probabilistic process model from the low-level event log of the activity of a knowledge worker. In these settings, knowledge workers are often involved in numerous projects, that require accessing different data sources, interacting with different individuals, exchanging messages, browsing the web, etc. An individual's role may vary from project to project, and the relevance of different portions of the document will vary according to the project and the individual reading the document. From such events, we obtain dynamic models and demonstrate the process on real-world data collected from knowledge workers in a large telecom company. The goal is to aid the user with improved information delivery when we are provided with better understanding of the knowledge process [1].

We describe the situation as following: given a database, describing events in a business setting, such as e-mail messages and visited URLs, executed in different contexts, produce a probabilistic temporal model that best describes the action patterns appearing in the event log. We obtain this by solving three tasks: context mining, action mining and process mining.

Context mining is the task where we want to discover the different contexts that the knowledge worker is involved in. It is obtained by performing semi-supervised clustering of events, where each cluster represents a distinct context in which the knowledge worker is working. In most work environments, contexts are most often seen as projects or clients, such as "*Process mining research project*" or "*Proposal for client X*". In general, we distinguish between them by different content keywords, resources, and people involved in the events.

Action Mining is the task where we wish to look at the events in a context-free manner and identify more general atomic actions. It is performed by clustering [2] of a context-free representation of events. When events are stripped of context-describing features and given additional metadata, we are left with clusters, which describe generalized representations of events, such as "*Send e-mail to group of co-workers*" or "*View intranet website*", which may occur in multiple contexts.

Process mining shows us the dynamics of the knowledge process of either a particular knowledge worker or an aggregate process model of an entire team. It gives us the probabilistic model of transitions between actions within a context. It is done by learning probabilistic deterministic finite automata [3] (PDFAs) and reduces the full models to only statistically significant frequent sequences of actions in the data. This gives us a pruned process model which resists noise by only using statistically grounded transitions and is easier to interpret.

Demonstration on real data shows that we can interpret several patterns using this model. For instance, the transitions between project-related actions are more common than transitions to administration-related actions. Also, web browsing events tend to have longer homogeneous action sequences than e-mails.

References:

- [1] J.M. Gomez-Perez, M. Grobelnik, C. Ruiz Moreno, M. Tilly, P. Warren. Using Task Context to Achieve Effective Information Delivery. *Workshop on Context, Information And Ontologies - CIAO at ESWC 2009*
- [2] B. Long, Z. Zhang, and P. S. Yu. A probabilistic framework for relational clustering. *In proceedings of the 13th ACM International Conference on Knowledge Discovery and Data Mining*, pages 470-479, 2007.
- [3] Jacquemont, S. and Jacquenet, F. and Sebban, M. Mining probabilistic automata: a statistical view of sequential pattern mining. *Machine Learning*, vol. 75, no. 1, pp. 91-127, Springer, 2009.