Formal methods for process knowledge extraction

Dmitry Morozov[1,2], Sergei O. Kuznetsov[3], Mario Lezoche[1,2], and Hervé Panetto[1,2]

[1] Université de Lorraine, CRAN, UMR 7039, Campus sciences B.P. 70239, Vandœuvre-lès-Nancy Cedex, 54506, France
[2]CNRS, CRAN, UMR 7039, France
{dmitrii.morozov, mario.lezoche, herve.panetto}@univ-lorraine.fr
[3]National Research University Higher School of Economics, Moscow, Russia
skuznetsov@hse.ru

Poster Summary

One approach to engineering of interoperating processes is based on different types and levels of abstraction or models. These models must express and formalize not only the structural aspect of the system components, but also their behaviour, the latter being constrained by domain-specific requirements of the process (business rules). The aim of this work is to study issues of cooperative processes engineering (engineered) models and actors (organizations, software systems, etc.) wishing to interoperate and to propose solutions to the highlighted problems.

We propose an application of Pattern Structures (Kuznetsov, 2001) to the field of Process Mining (van der Aalst, 2011). Process Mining refers to the extraction of process models from event logs. An event log simply presents a set of *sequences* of elementary events. These event sequences correspond to different executions of the process. The problem is that real-life processes tend to be less structured and more flexible. Therefore traditional Process Mining algorithms give spaghetti-like process models, which are hard to comprehend and use in practice. An approach to overcome this is to cluster an event log in such a way that each of the resulting clusters forms a coherent set of event sequences, which can be adequately represented by a distinct process model. In our poster we propose a trace clustering method based on Formal Concept Analysis framework, namely Pattern Structures.

A pattern structure on sequences is introduced by defining an operation of generalization of two sequences as a set of maximal common sub-sequences. With help of the generalization operation we define a subsumption relation and a pattern concept. An extent of a pattern concept is a set of objects, and their generalization (intent) is a set of their all-maximal common sub-sequences. We use a pattern concept to define a cluster of sequences. From a set of all concepts of a pattern structure we choose a small subset with the highest values of stability index, which gives us a complete or almost complete cover of the initial object set. Then the set of sub-sequences describing the objects of a stable concept is used as an input for the alpha-algorithm, which produces a petri net. All the petri-nets obtained in such a way give us a model of the *whole process* implicitly defined by the initial set of sequences.

Further research can be focused at improving the efficiency of the algorithm by finding a way to avoid construction of complete pattern structure with subsequent calculation of stability index.

References

W.M.P. van der Aalst. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer-Verlag, Berlin, 2011.

B. Ganter and S.O. Kuznetsov, Pattern Structures and Their Projections. In: Proc. G. Stumme and H. Delugach, Eds., ICCS 2001, LNAI (Springer), Vol. 2120, pp. 129-142, 2001.