



VORTRAG 19.11.2018 11.30 Uhr SE 15

**“Process Discovery: Extracting Business Processes
from Text Logs - A Mathematical Programming
Approach”**

Marcus Poggi, PUC-Rio, Professor Associado
Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro

Process Discovery aims at finding a process model from an event log of a business process. Petri nets are a choice for formalising a process model. One is interested in finding a Petri net which, when simulated, produces a log that corresponds to the original event log, or minimizes a metric between them. Given a set S of elements, each representing a process activity, an event log is a set $L = \{s_1; \dots; s_n\}$ where s_q represents a sequence of activities, called traces, that occur in the process. A Petri net is a bi-partite graph with directed arcs connecting transitions(activities) and places. van der Werf et al. 2008 presents an ILP that finds one place and its arcs to and from activities. This formulation has one constraint for each prefix of each trace. Practical instances have thousands of traces and dozens of activities. A Petri net solution for the Process.

Discovery problem is then obtained by solving a sequence of ILPs. Recently, Spyrides et al. (2018) proposes an ILP that finds all places and arcs of the Petri net. It allows considering global properties such as token balance and cohesion among places. The main drawback of this approach is how to scale to practical instances. Therefore, we devise a cut and column generation approach. The master problem considers the global properties of the Petri net, the column generation sub-problem, similar to the ILP proposed in van der Werf et al. 2008, obtains candidate places and arcs. Variants of then model are tested for solution quality on instances of the Process Discovery Contest @BPM2017. Further, since the quality of a process model is measured in terms of (replay) fitness, precision, simplicity and generalization, we extend the model above to consider only a high percentage of the traces in a log. We show that the resulting processes have almost the same fitness while increasing substantially the ability of interpretation provided by the model.

(Joint work with: Georges Spyrides, Beatriz Santiago, Helio Lopes)