

is based on explicit personalization, by exploiting the scientists' social networks, using gossip protocols that scale well. Relevance measures may be expressed based on similarities, users' confidence, document popularity, rates, etc., and combined to yield different recommendation criteria. With P2Prec, each user can identify data (documents, annotations, datasets, etc.) provided by others and send queries to them. For instance, one may want to know which scientists are expert in a topic and get documents highly rated by them. Or another may look for the best datasets used by others for some experiment. To efficiently disseminate information

among peers, we propose new semantic-based gossip protocols. Furthermore, P2Prec has the ability to get reasonable recall with acceptable query processing load and network traffic.

The second result deals with the efficient processing of scientific workflows that are computationally and data-intensive, thus requiring execution in large-scale parallel computers. We propose an algebraic approach (inspired by relational algebra) and a parallel execution model that enable automatic optimization of scientific workflows. With our algebra, data are uniformly represented

by relations and workflow activities are mapped to operators that have data aware semantics. Our execution model is based on the concept of activity activation, which enables transparent distribution and parallelization of activities. Using both a real oil exploitation application and synthetic data scenarios, our experiments demonstrate major performance improvements compared to an ad-hoc workflow implementation.

Please contact:
 Patrick Valduriez
 INRIA, France
 E-mail: Patrick.Valduriez@inria.fr

Performance Analysis of Healthcare Processes through Process Mining

by Diogo R. Ferreira

Process mining provides new ways to analyze the performance of clinical processes based on large amounts of event data recorded at run-time.

Hospitals and healthcare organizations around the world are collecting increasingly vast amounts of data about their patients and the clinical processes they go through. At the same time - especially in the case of public hospitals - there is growing pressure from governmental bodies to refactor clinical processes in order to improve efficiency and reduce costs. These two trends converge, prompting for the need to use run-time data in order to support the analysis of existing processes.

In the area of process mining, there are specialized techniques for the analysis of business processes according to a number of perspectives, including control-flow, social network, and performance. These techniques are based on the analysis of event data recorded in system logs. In general, any information system that is able to record the activities that are performed during process execution can provide valuable data for process analysis.

These event data become especially relevant for the analysis of clinical processes, which are highly complex, dynamic, multi-disciplinary, and ad-hoc in nature. Until recently one could only prescribe general guidelines for this kind

of process, and expect that medical staff comply. Now, with process mining techniques, it is possible to analyze the actual run-time behaviour of such processes and obtain precise information about their performance in near real-time.

Such an endeavour, however, is made difficult by the fact that reality is inherently complex, so direct application of process mining techniques may produce very large and confusing models, which are quite difficult to interpret and analyze - in the parlance of process mining, these are known as "spaghetti" models.

While the area of process mining is being led by Wil van der Aalst at the Technical University of Eindhoven in The Netherlands, here at the Technical University of Lisbon, in Portugal, we have been developing techniques to address the problem of how to extract information from event logs such that the output models are more amenable to interpretation and analysis. To this end, we have spent the last six years developing a number of clustering, partitioning, and preprocessing techniques. Such techniques have matured to the point that they can be systematically applied to real-world event logs,

according to a prescribed methodology, to produce understandable, useful, and often surprising results.

One of the latest developments in the field of process mining, introduced by Zhengxing Huang and others at Zhejiang University in China, concerns performance. Typically, a control-flow model must be extracted from the event log prior to performance analysis. However, to study the performance of healthcare processes, only a subset of the recorded activities is usually considered - these are the key activities that represent milestones in the process, and that are always present regardless of the actual path of the patient. The time span between these activities becomes a Key Performance Indicator (KPI).

The ability to measure this KPI directly from the event log is a major improvement with respect to previous performance analysis techniques which rely on a control-flow model that often includes too much behaviour. Here, we are interested in a predefined sequence of milestones and in retrieving the time span between any pair of milestones. Incidentally, this approach also provides the time span between the first and last activities, which can be used to

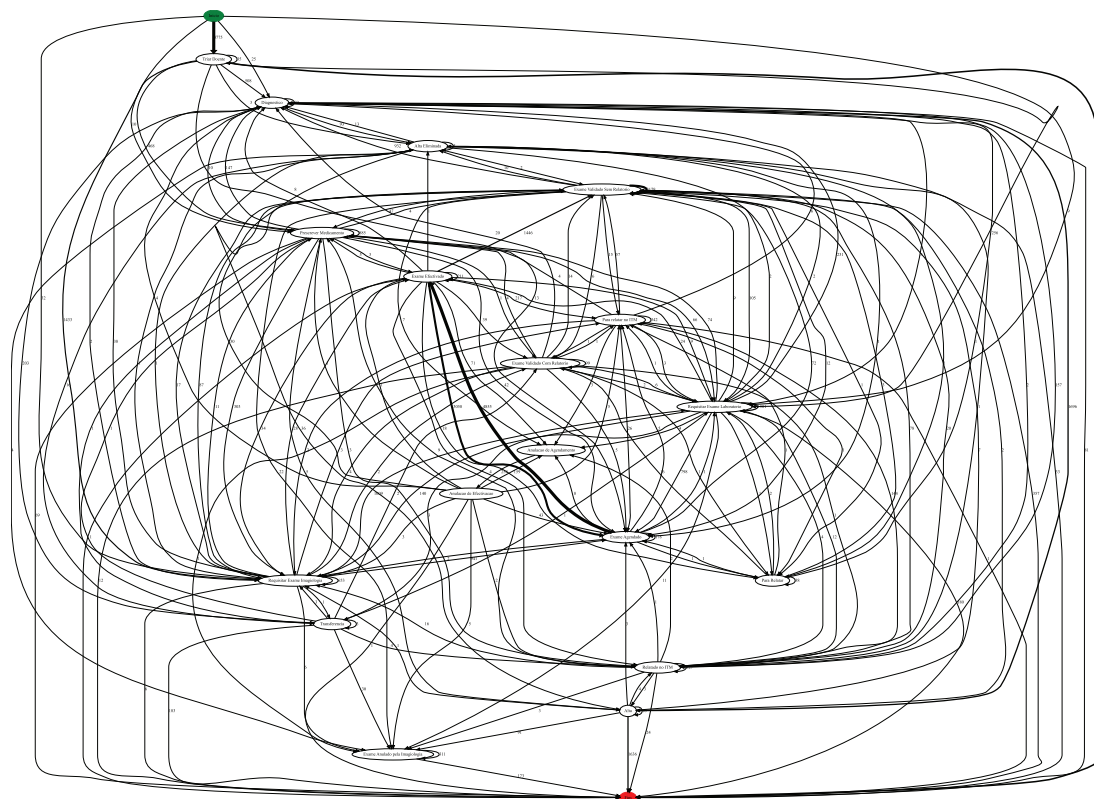


Figure 1:
Control-flow model

determine the length of stay (LOS) of the patient in the hospital, one of the most sought-after KPIs in healthcare processes.

Back in Portugal, we applied this approach in a case study carried out in the emergency department of a mid-

sized public hospital. The hospital has an Electronic Patient Record (EPR) system, which records the events that take place in several departments. The event log used in this experiment was collected over a period of 12 days. A total of 4851 patients entered the emergency department in that period,

resulting in over 30 000 recorded events, although there are only 18 distinct activities.

Figure 1 depicts a control-flow model for these activities, illustrating the reason why such diagrams are often called “spaghetti” models. In Figure 2, we present the results for some key activities. The first step – triage – determines the priority of the patient and takes place once the patient enters the hospital. For patients who require medical examination, it takes on average two hours to perform the first exam. About two hours and 30 minutes later, the patient receives the diagnosis, and then is quickly discharged, on average within three minutes. The resulting LOS amounts to an average of four hours and 30 minutes.

Figure 2 shows minimum, maximum, and average times, and standard deviations. While these results were gathered for all patients that entered the emergency department, similar analysis can be conducted for patients with certain conditions or with particular clinical paths.

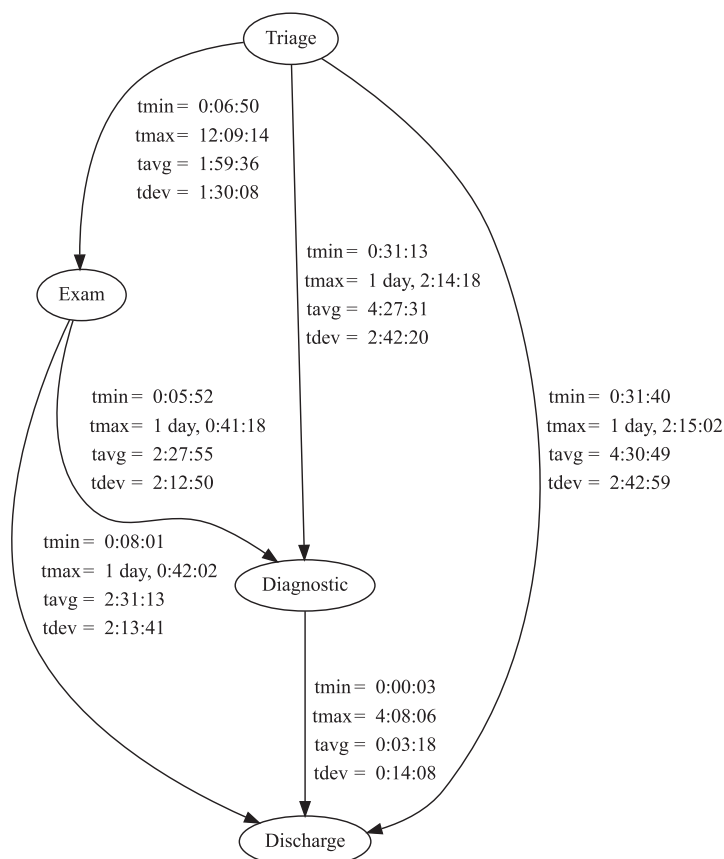


Figure 2: Performance analysis

Link: <http://www.processmining.org/>

Please contact:
Diogo R. Ferreira
IST – Technical University of Lisbon,
Portugal
Tel.: +351 21 423 35 52
E-mail: diogo.ferreira@ist.utl.pt