Aggregation of spatio-temporal and event log databases for stochastic characterization of process activities*

Rodrigo Gonçalves
Eindhoven University of Technology
Information Systems Group,
De Lismortel 2, 5612AR,
Eindhoven, The Netherlands
r.m.t.goncalves@tue.nl

Rui Jorge Almeida
Eindhoven University of Technology
Information Systems Group,
De Lismortel 2, 5612AR,
Eindhoven, The Netherlands
rjalmeida@tue.nl

João M. C. Sousa
Universidade de Lisboa,
Instituto Superior Técnico, IDMEC
Av. Rovisco Pais 1, 1049-001
Lisboa, Portugal
jmsousa@tecnico.ulisboa.pt

Process-aware information systems are typically used to log events (e.g. transaction logs, logistics, healthcare) describing the execution of such processes. The analysis of these logs can provide meaningful knowledge for organizations to improve the quality of their services as well as their efficiency [3]. In logistic domains, transportation planning and scheduling are made based on a-priori knowledge about activities. Thus, a proper identification and characterization of process activities is of extreme importance. Given the nature of such logistics business processes, the presence of spatio-temporal databases (form global positioning systems) is common.The analysis of spatio-temporal and event logs databases provide different information on the same business processes. Both databases are important for the correct stochastic characterization of process activities. Spatio-temporal data provide accurate information about location and time but uncertainty about which actions occurred. Event logs clearly identify which activities took place but when they are partially human generated, it is not assured that events are logged at the correct time and location. This leads to uncertainty related to the time at which events (recorded by means of user interaction) are logged. By aggregating both types of data, it is possible to reduce the activity recognition uncertainty.

Learning patterns of human behaviour from sensor data is extremely important for high-level activity inference [2]. To derive activity logs from low-level event logs, we define an activity as a finite sequence of events, over a finite period of time, where each event in the activity is an occurrence. The duration of an activity is given by the elapsed time between its first and last event. If the events are logged before (or after) the real occurrences, the duration of activities can not be correctly obtained. Using the spatio-temporal data: latitude, longitude and time-stamp, we defined time-windows based on the average travel speed. The portions of the trajectories where the truck was stopped are combined with the activities extracted from the event log to create an activity time-line. Despite the logging system being able to keep track of all activity types (e.g load, unload, rest, wait, etc.) dead-times between activities are possible to exist due human influence. Considering such premise, it is possible to formulate hypothesis to estimate activity durations based on the empty time on the neighbourhood of activities. Travel times can also be estimated by analysing the moving portions of trajectories. The framework enables the prediction of service times at specific costumers, by using load and unload activity durations for frequency analysis, as well as stochastic travel times by clustering trajectories [1]. Such knowledge, based on real-world data, can be used in vehicle routing applications for planning and optimization of business processes.

References

[1] J.-G. Lee, J. Han, and K.-Y. Whang. Trajectory clustering: a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 593–604. ACM, 2007.

[2] L. Liao. *Location-based activity recognition*. PhD thesis, University of Washington, 2006.

[3] W. Van der Aalst, T. Weijters, and L. Maruster. Workflow mining: Discovering process models from event logs. *Knowledge and Data Engineering, IEEE Transactions on*, 16(9):1128–1142, 2004.

*This work was performed as part of the DAIPEX project grand funded by Dinalog