Stochastic Models and Data Analysis for Processes

Sally McClean

School of Computing, Ulster University, Belfast, Northern Ireland.
e-mail: si.mcclean@ulster.ac.uk

Models, such as Markov chains or Petri nets, have often been used to describe processes where simulation tools can use the models to imitate important aspects of the behaviour of the system and allow experimentation without having to disrupt the real-life set-up. However, with the ever-increasing capability of computer systems to collect detailed information about all aspects of such processes, such as different tasks, events, alarms, intrusions or failures, has come increased ability to analyse process behaviours and outcomes leading to better understanding and improved efficiency. Such developments are leading to a tighter coupling of Stochastic Models and Data Analysis with obvious benefits.

Historically, much work has been carried out to develop models of individuals moving through states starting with simple Markov models which assume that probabilities of leaving a state do not depend on duration in that state and generalise to semi-Markov models where the durations in a state can have a more general duration distribution. As for modern business processes and other application domains, covariates may also be available relating to the whole process or specific states within the process. These have been modelled in different ways, such as stratification in the simplest case or, alternatively, the state durations distributions can be modelled as explicit functions of the covariates, for example, as in Cox regression. Another suitable modelling paradigm is the hidden Markov and semi-Markov model where we do not have data on the states but extract information about them through the covariates. Thus, the stochastic models of state occupancies can be unified with data modelling, through the associated state-specific covariates.

We here present a number of process modelling and mining problems, such as detecting anomalous sequences of tasks, predicting faults, non-compliance, or timely completion of processes, as well as analysing and forecasting likely occurrence of absorption. Such problems frequently occur in a real-time environment so predictions are often made as part of a monitoring process, alongside suggesting possible actions to operate the process in an efficient and profitable manner that can also manage expectations and avoid undesirable outcomes. Examples from healthcare, smart homes, business, and telecoms will illustrate how stochastic models can be combined with data analysis for better use of such complex processes.