Applying Probabilistic Programming to an Intelligent Parking System

Yousuf Amanuel

Master’s thesis

Abstract

This work aims to identify the characteristics of Probabilistic Programming (PP) and its usage possibilities. We focus on Bayesian inference methods, which are either sample-based (MCMC) or optimization-based (VI). These methods are implemented by different PP languages, which facilitate creating statistical models that represent the output of a random event through a probability distribution called the posterior. The latter distribution is inferred using prior knowledge about the model and observations collected. The most critical point about PP is the requirement for high domain knowledge. Thus, the analyst can select appropriate inference algorithms and prior distributions for the application case. Criteria for making such selections are introduced in this work.

We applied PP to a parking management system and simulated the latter intending to predict the occupation of parking vacancies. Here, we assumed that our data was normally distributed. The predictions were used to calculate the availability of a parking place in further time points. The results of all selected inference algorithms (HMC, NUTS, and SVI) were satisfying in terms of a small and average loaded system, although the amount of data was small. For a high-loaded (crowded) system, the measured metrics delivered poor results. That happened because good domain knowledge was missing. The latter simulation case can be improved either by extending out domain knowledge or obtaining more data and repeating the whole process with another ML method, e.g., time series.

Overall, we can conclude that choosing the inference algorithm is based on a trade-off between the model's accuracy and the computational speed. MCMC methods should be chosen because of their high computational effort when the accuracy of the results is important. VI methods are faster and can be selected when an approximation of the model is enough. Furthermore, we consider PP as an appropriate method for creating Bayesian models when high domain expertise is available and obtaining much data is not easy. In contrast, traditional ML approaches are still a good choice when enough data is available, and the domain is not known well.