LINKÖPING UNIVERSITY

Linköping, February 15, 2019

Up to two PhD students in Statistical Machine Learning (Ref IDA-2019-00026)

Fredrik Lindsten
Division of Statistics and Machine Learning
Department of Computer and Information Science
Linköping University, Linköping, Sweden
Contact: `fredrik.lindsten@liu.se`

The Division of Statistics and Machine Learning (STIMA) at Linköping University is expanding. In this call we are looking for up to two new PhD students who want to join us in our effort of ramping up the research in statistical machine learning at the division. This document briefly presents the research at the division and the background of the PI (Section 1). This is followed by a few concrete suggestions for research topics for the PhD positions (Section 2). Some administrative details are:

- **Application deadline:** April 23, 2019
- **Application procedure:** All applications are to be submitted via Linköping University's online application system, available via this link:
  `https://liu.se/en/work-at-liu/vacancies?rmpage=job&rmjob=10575&rmlang=UK` (English)
  `https://liu.se/jobba-pa-liu/lediga-jobb?rmpage=job&rmjob=10473&rmlang=SE` (Swedish)
- **Questions:** If you have any questions do not hesitate to contact Fredrik via e-mail.

1 The Division of Statistics and Machine Learning

STIMA is a division of Statistics and Machine Learning that belongs to a department of computer science. This fact makes us unique in Sweden, and we like to view ourselves as Sweden's most modern division of statistics with a clear focus on state-of-the-art data analysis, prediction and decision making in complex systems. We are engaged in basic methodological research motivated by a wide range of problems in areas that span from journalism and psychology to genetics and climatology.

Fredrik Lindsten, who will be the main supervisor for the PhD students recruited in this call, is an Associate Professor at STIMA. His research background is in computational statistics, machine learning, and signal processing. In particular, his research has focused on probabilistic models and the development of computational algorithms for learning and reasoning about these models. Using probability theory, probabilistic models are able to systematically represent and cope with the uncertainty that is inherent to most data. This is of central importance in many applications of machine learning.

The group at STIMA has a wide network of strong international collaborators all around the world, for example at the University of Cambridge, University of Oxford, University of California, Berkeley, University of British Columbia (Vancouver), and University of New South Wales (Sydney). We strive for all PhD students to get a solid international experience during their PhD studies.

2 Potential Research Topics

The research projects for the advertised positions will be in the areas of statistical machine learning/computational statistics. A few examples of potential research topics are briefly outlined below. As an applicant you are not required to specify a specific research topic in your application, but you are of course welcome to do so if you want. The topics below are provided mainly to make the advertised positions more concrete. We do welcome own initiatives and the precise research topic of each PhD student will be decided in a dialog between the student and the supervisor after a successful appointment.

Spatio-temporal models: Processes that evolve over both time and space are ubiquitous in science and technology. Examples include the spread of a disease in a region, or the evolution of various climate variables pertaining to the risk of extreme weather events. However, many spatio-temporal processes encountered in practice involve complicated (nonlinear) dependencies both over time and space, non-Gaussian stochastic errors, partially missing observations, and other difficulties. As a consequence, constructing accurate models of such processes that can be used to predict their future values poses several challenges. A possible PhD project is to investigate how powerful machine learning technologies, such as convolutional neural networks, can be used for addressing these challenges. We will also investigate how this approach can be combined with conventional methods for spatio-temporal processes, including the use of so called probabilistic graphical models for encoding structure and enabling accurate reasoning about uncertainties in the learnt models and their predictions.

Combining simulation-based models and machine learning: Machine learning is predominantly data-driven, in the sense that generic model structures are used, which are then adapted to the application-specific data. Needless to say, this has proven to be a very successful approach for modeling the complex data dependencies that we often encounter in practice—very few assumptions are made and the data is "allowed to speak for itself". However, another very common approach to modeling and prediction-making is to use simulation-based models based on physical insights or first principles. Examples are found in chemical process engineering, climate science, and autonomous systems, to mention a few. This approach to modeling benefits from prior knowledge and can often leverage decades of research devoted to better understanding a specific problem area. This can be of paramount importance, in particular when data is scarce. In this project we will develop new ways of combining such simulation-based models with state-of-the-art machine learning technologies to, in some sense, reap the benefits of both approaches. Possible starting points for accomplishing this goal is to use simulators as priors for probabilistic models, or to use the output of these simulators as a type of pseudo-data for machine learning.

Uncertainty-aware deep learning: Deep learning is one of the leading approaches to many applications of machine learning, in particular those involving high-dimensional and unstructured data, such as images and text documents. In many of these applications, however, it is of paramount importance to be able to accurately reason about the uncertainties associated with the predictions made by these models. This includes medical diagnosis support, autonomous driving, and other applications relying on robust decision making. A third possible PhD project is to theoretically analyze various deep learning technologies to better understand how they affect the ability of the model to reliably quantify uncertainty. The insights gained from this analysis can then be used for developing new methods for improving the uncertainty-awareness of deep learning.

Probabilistic programming inference: A very flexible modeling approach are so called probabilistic programming languages (see `http://probabilistic-programming.org`), where probabilistic models are constructed by overloading standard programming operations to have probabilistic meanings. Hence, a software engineer can define a probabilistic model by writing more or less standard code, simply defining certain variables to be unknown and random. This type of generic modeling frameworks have a huge potential for making probabilistic models accessible to a wide range of (non-specialist) users in various application areas. A challenge, however, is to provide efficient and automated inference methods when very little is known about the model beforehand. Indeed, the model can be viewed as a black-box that can be simulated by executing the code, but otherwise its properties are unknown. A fourth possible PhD project is to develop new computational inference methods that can operate under these challenging conditions. The new methods can therefore be used as backbone inference engines for probabilistic programs, with a huge number of potential applications.