WASP | WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

September 24, 2018

Two PhD students in machine learning:

Probabilistic models and deep learning – bridging the gap

Fredrik Lindsten and Lennart Svensson

We are looking for two PhD students for a WASP collaboration project between Chalmers University of Technology and Linköping University. The project is entitled *Probabilistic models and deep learning – bridging the gap* and it involves fundamental research on the development of new machine learning models and computational algorithms as well as applied research to demonstrate the usefulness of the new methodologies. Section 1 presents the project organization and background and Section 2 contains a high-level presentation of the research that will be conducted within the project.

- **Linköping University:** Ref IDA-2018-00209. Application deadline: October 26, 2018. `https://liu.se/en/work-at-liu/vacancies`
- **Chalmers University of Technology:** Ref 20180540. Application deadline: November 2, 2018. `http://www.chalmers.se/en/about-chalmers/Working-at-Chalmers/Vacancies/`.
- **Contact:** Fredrik Lindsten `<fredrik@lindsten.net>`, Lennart Svensson `<lennart.svensson@chalmers.se>`

1 Background and organization

Wallenberg AI, Autonomous Systems and Software Program (WASP) is Sweden's largest ever individual research program, a major national initiative for strategically basic research, education and faculty recruitment. The program is initiated and generously funded by the Knut and Alice Wallenberg Foundation (KAW) with 2.6 billion SEK. In addition to this, the program receives support from collaborating industry and from participating universities to form a total budget of 3.5 billion SEK. Major goals are more than 50 new professors and more than 300 new PhDs within AI, Autonomous Systems and Software. The vision of WASP is excellent research and competence in artificial intelligence, autonomous systems and software for the benefit of Swedish industry. For more information about the research and other activities conducted within WASP please visit: `http://wasp-sweden.org/`

The present call is for two PhD students: one at the Division of Statistics and Machine Learning at Linköping University, and one at the Department of Electrical Engineering at Chalmers University of Technology, Göteborg. Together with two senior researchers—Fredrik Lindsten at LiU and Lennart Svensson at Chalmers—you will form a team that will jointly pursue the research presented in the next section. The collaboration within the team is of central importance for the project. We will therefore encourage regular research visits between the two participating research groups and offer joint supervision and PhD courses. Naturally, you will also be part of the local research group at your university, either Chalmers or LiU, with the possibility of collaboration, interaction, and discussions with other PhD students and senior researchers working on related projects.

The principal investigators have wide networks of international collaborators all around the world, for example at the University of Cambridge, University of Oxford, Columbia University in the City of New York, University of California, Berkeley, and University of British Columbia. We strive for all PhD students to get solid international experiences during their PhD studies. Furthermore, we have strong connections with Swedish industry and, indeed, the demonstrator applications that we will investigate in the present project have clear links with some of our company collaborators that may be involved in the research conducted in the project.

2 Research presentation

Probabilistic models—where unobserved variables are viewed as stochastic and dependencies between variables are encoded in joint probability distributions—are widely used in the areas of statistics and machine learning. It has been recognized that probabilistic models come with many desirable properties: they enable reasoning about the uncertainties inherent to most data by using well-developed probability theory; they can be constructed hierarchically to build complex models from simple parts; they provide a natural safeguard against overfitting; and they allow for fully coherent inferences over complex structures from data. Deep learning is a different branch of machine learning which has recently had remarkable success in a range of different applications related to computer vision, natural language processing and more. In most cases, the deep neural networks (DNNs) are trained discriminatively using supervised learning, with large sets of (annotated) training data. Unfortunately, the advancement of deep learning has come at a price—DNNs often lack many of the desirable properties of probabilistic models, such as uncertainty quantification and structure exploitation over well-defined probabilistic priors.

In this project, we will develop theory and methods related to the interplay between probabilistic models and deep learning. More specifically, we intend to develop both new models and new inference and learning algorithms for applications where unobserved variables are naturally characterized using, for instance, probabilistic graphical models or stochastic processes, whereas data is from some domain where deep learning has been successful (e.g., images). The family of problems that involve such interplay between deep learning and probabilistic models is general, and we expect to derive tools that are widely applicable. However, to make the research more concrete and to showcase the merits of the new methodology we will study three specific applications in more depth, each one of significant importance on its own. These are:

1. Weak annotations: The success of machine learning hinges on the availability of informative data. However, acquiring large amounts of high-quality data can be very costly and time consuming. In the *weak supervision* approach to machine learning, annotators (humans) are assumed to provide noisy annotations (labels) on a set of training data. Such weak annotations can often be obtained more cheaply, for instance via crowdsourcing, but at the potential cost of obtaining less reliable annotations. Probabilistic models appear naturally in the context of weak supervision. For instance, the difficulty of annotating a certain data point, along with the skill of a certain annotator, will influence the probability that the produced label is correct. The relationships between these variables can then be modelled using, e.g., a probabilistic graphical model. DNNs are also expected to be of great importance, not only as a final regression or classification model, but also during the annotation process since they can provide information about annotation quality. In this project, we will investigate how such DNNs can be combined with probabilistic models in order to produce reliable posterior distributions of the true annotations.

2. Dynamical systems: Dynamical systems are of great importance in modern society, for instance, due to their central role in robotics and autonomous vehicles. In that context, measurements often contain unstructured data from a camera or a radar sensor for which DNN-based models are suitable. We seek to estimate the states of both the agent (the robot or host vehicle) and its environment (e.g., nearby road users and stationary objects). In this project we will investigate the possibility of combining DNN-based measurement models with probabilistic state space models which allow us to exploit prior knowledge about the state representation and the structure of the dynamics. We will also consider how this approach can be used for "semi-supervised" learning, where for part of the data the state is known (e.g., initial learning using reference sensors) and for part it is unknown (e.g., refinements of the model on-the-fly).

3. Mitosis modelling: Detection and counting of cell mitosis (i.e., division of the cell nucleus) from microscopic tissue examinations is an important diagnostic tool, in particular when concerned with cancer. Deep learning has been successful in automating this task with the potential of making the process more cost efficient and accurate. However, existing approaches are mainly based on individual pixel-wise classification of the microscopy images. Probabilistic models have the potential of improving the performance and reliability of such methods by modelling spatial dependencies and *a priori* distributions of mitosis within a statistical framework. Moreover, this approach can be used for automated analysis at different scales, which is of great practical importance due to the extremely high resolution of histopathological images. By first computing a posterior belief regarding the mitosis intensity at a coarse scale, we can then use sequential decision making, based on the posterior uncertainty, to recursively decide where to "zoom in" and classify at a finer scale, etc.

All of the applications mentioned above involve a probabilistic description of some unobserved variables. Additionally, we have access to data which is naturally processed using a DNN model. The output of this DNN is in some sense related to the unobserved variables. Note that the DNN in this context no longer outputs a distribution over labels, as is customary, but provides one component which is integrated with the probabilistic model. A key research question in this project is to investigate how this combination of different types of models should be done in a principled, general, and statistically well-motivated way.

Furthermore, such model combinations give rise to the computational challenge of performing statistical inference over the unobserved variables and simultaneously learn the parameters of the model. Learning of the neural network parameters is typically done using stochastic gradient descent, whereas inference in probabilistic models relies on methods such as belief propagation, variational inference, and Markov chain Monte Carlo sampling. Another key research question in this project is therefore how to systematically combine such inference and learning techniques, thereby developing efficient computational algorithms for the type of probabilistic deep learning models described above.