Statistical models and logic: handling inconsistencies (Project #1)

University of Oslo, Department of Mathematics

Four year PhD position (25% time dedicated to teaching and/or administrative tasks)

Description

Inconsistencies often occur in the analysis of data, in the form of contradictions between various parts of the data, between parts of the background knowledge which enters models, between predictions/estimates and what is known about these in advance, or with respect to a combination thereof. This project aims at bringing together the views and methods of statistics and logic on detecting and handling inconsistencies. The PhD student must expect to do exciting interdisciplinary work between these two scientific fields. The supervising team has expertise from both.

The project has two specific aims: **(i)** develop probabilistic models for partially observed preference data (e.g., but not limited to, user clicking data) for preference learning in the presence of inconsistencies in the input data; **(ii)** develop probabilistic approaches to the treatment of inconsistencies beyond prefernce learning, in more general settings. It will be useful to develop a suitable formal language to describe preference statements that enables detection and inference in the presence of inconsistencies and to handle inconsistencies as a piece of useful information.

Specific project requirements

- Master's degree in mathematics, statistics, computer science, data science or a related subject.
- Documented experience in statistics or logic will be an advantage.
- Contract beginning no later than 1 October 2024
- This PhD position has four instead than three years, because 25% of the time will be dedicated to teaching and/or administrative tasks.

Supervisors

Statistics/biostatistics:

- Associate Professor Ida Scheel, idasch@math.uio.no contact person for inquiries about the position
- Associate Professor Valeria Vitelli
- Professor Arnoldo Frigessi

Computer Science/Logic:

- Postdoctoral Fellow Anne-Marie George, annemage@ifi.uio.no contact person for inquiries about the position
- Professor Martin Giese
- Associate Professor Egor Kostylev
- Professor Arild Torolv Søetorp Waaler

Sparse models for machine learning from a Bayesian viewpoint (Project #2)

University of Oslo, Department of Mathematics

Three year PhD position

Description

Within deep neural networks, simplifying and pruning highly over-parameterized models have shown to improve accuracy, and results in sparse networks that also are more computationally efficient and sometimes easier to interpret. This is very important for next generation machine learning algorithms. In this project we aim at exploring Bayesian approaches to sparsification/pruning for neural networks. Focus will be on prior specifications, statistical properties and efficient algorithms. Exploration of the pruning approach in other machine/statistical learning settings will also be of interest. While sparse models usually are obtained by going from large to small, also procedures going from small to large will be considered.

Specific project requirements

- Master's degree in statistics, mathematics, computer science or a related quantitative subject with proven competence in statistics.
- Documented experience in scientific programming is an advantage.

Supervisors

- Professor Geir Olve Storvik, geirs@math.uio.no contact person for inquiries about the position
- Professor Fred Godtliebsen
- Associate Professor Ida Scheel
- Professor Sylvia Richardson

Developing novel information theoretic discrepancy measures (Project #3)

UiT The Arctic University of Norway, Department of Physics and Technology

Three year PhD position

Description

Machine learning increasingly needs to deal with new types of data, needs new ways to incorporate a priori knowledge for better generalization and explainability, and needs to discover causal relationships. A key research question for this project is: How do we develop novel measures of discrepancy or divergence to best capture information in data for efficient neural networks learning?

Specific project requirements

- Master's degree in machine learning, computer science, statistics, mathematics or related fields with a strong background in statistical learning theory, deep learning, pattern recognition.
- For this PhD position, the Master's thesis must have the grade B or better in the Norwegian educational system, when the degree is obtained.
- Strong programming experience and experience with training deep neural networks (e.g. PyTorch, Tensorflow).
- Documentation of English proficiency

Supervisors

- Professor Robert Jenssen, robert.jenssen@uit.no contact person for inquiries about the position
- Associate Professor Shujian Yu
- Associate Professor Johan Pensar
- Associate Professor Ali Ramezani-Kebrya
- Professor Sylvia Richardson

Admission to the PhD programme

For employment in the PhD position, you must be qualified for admission to the PhD programme at the Faculty of Science and Technology, and participate in organized doctoral studies within the employment period.

Admission normally requires:

- A bachelor's degree of 180 ECTS and a master's degree, or an integrated master's degree.

UiT normally accepts higher education from countries that are part of the Lisbon Recognition Convention.

In order to gain admission to the programme, the applicant must have a grade point average of C or better for the master's degree and for relevant subjects of the bachelor's degree. A more detailed description of admission requirements can be found here.

If you are employed in the position, you will be provisionally admitted to the PhD programme. Application for final admission must be submitted no later than two months after taking up the position.

General information

The appointment is made in accordance with State regulations and guidelines at UiT. At our website, you will find more information for applicants.

More practical information about working and living in Norway can be found here: https://uit.no/staffmobility

Exploration and control of the inner representation in generative AI models (Project #4)

UiT The Arctic University of Norway, Department of Physics and Technology

Three year PhD position

Description

There is a growing and urgent need for efficient control of generative neural networks. The goal of this project is to find new ways of controlling ML models directly using their inner or "latent" representation of the data input. The project will involve both experimental, methodological and theoretical work in order to discover general concepts for control and to better understand the inner mechanisms of these models.

Specific project requirements

- Master's degree in machine learning, computer science, mathematics, statistics or related field.
- For this PhD position, the Master's thesis must have the grade B or better in the Norwegian educational system, when the degree is obtained.
- Experience with deep neural networks is an advantage but is not necessary, as these can be learned.
- Strong programming skills is necessary (open-source activities, for example a Github account with open-source projects, will be an asset)
- Documentation of English proficiency

Supervisors

- Associate Professor Benjamin Ricaud, benjamin.ricaud@uit.no contact person for inquiries about the position
- Associate Professor Ali Ramezani-Kebrya
- Professor Arnoldo Frigessi

Admission to the PhD programme at UiT

For employment in the PhD position, you must be qualified for admission to the PhD programme at the Faculty of Science and Technology and participate in organized doctoral studies within the employment period.

Admission normally requires:

- A bachelor's degree of 180 ECTS and a master's degree, or an integrated master's degree.

UiT normally accepts higher education from countries that are part of the Lisbon Recognition Convention.

In order to gain admission to the programme, the applicant must have a grade point average of C or better for the master's degree and for relevant subjects of the bachelor's degree. A more detailed description of admission requirements can be found here.

If you are employed in the position, you will be provisionally admitted to the PhD programme. Application for final admission must be submitted no later than two months after taking up the position.

General information

The appointment is made in accordance with State regulations and guidelines at UiT. At our website, you will find more information for applicants.

More practical information about working and living in Norway can be found here: https://uit.no/staffmobility

Structure-generating models for transition metal complexes (Project #5)

University of Oslo, Department of Mathematics

Three year PhD position

Description

Current generative artificial intelligence approaches are typically based on Large Language Models (LLM). LLMs enable to estimate the probability of a sequence and to continue a started sequence. Graph-structured data is relevant in many fields, as they represent knowledge (e.g., in the form of taxonomies, ontologies, knowledge graphs) and disciplines such as chemistry (e.g., molecular graphs, chemical reaction pathways). Although LLMs can to some extent process graph structured data when it is simplified into sequences, this has drawbacks as the models then do not directly operate on graphs. Furthermore, these models typically do not take the semantics (e.g., axioms from an ontology) and the dependence structure of the data into account.

The goal of this project is to investigate the limitations of current approaches and to develop new methods to estimate the probability of a graph, to continue and extend graphs through LLMs. These approaches will then be applied to graphs that combine descriptions of the molecular structure with logical statements from an ontology related to chemistry. We will focus on Transition Metal Complexes (TMCs). TMCs are of great interest because of their potential uses in a wide variety of material technologies, including as metallodrugs for chemotherapy or catalysts for industrial chemical processes.

Specific project requirements

- Master degree in computer science, data science, mathematics, statistics, or other relevant field
- A solid background in statistics
- Good programming skills (e.g. Python) and ability to work with version control tools (e.g. Git)
- Experience with (graph) neural networks, graph-structured data, semantic technologies (RDF/OWL/SPARQL), and generative models is an advantage but is not necessary, as these can be learned.

Supervisors

- Researcher Basil Ell, basile@ifi.uio.no contact person for inquiries about the position
- Associate Professor Johan Pensar
- Associate Professor Riccardo De Bin

Benchmarking ethical reasoning in Large Language Models (Project #6)

University of Oslo, Department of Informatics

Three year PhD position
(An extension of the appointment by up to twelve additional months may be considered, which will be devoted to career-enhancing work duties, e.g. teaching or supervision. This will be dependent on the qualifications of the applicant and the specific teaching needs of the department.)

Description

This project will take a cross-disciplinary approach to the design and use of benchmarks for evaluating ethical and moral reasoning in Large Language Models (LLMs). Combining approaches from experimental philosophy and natural language processing, the project will address research questions like the following:

- How can we best design benchmarks for evaluating ethical reasoning in LLMs? What are the strengths and weaknesses of current approaches?
- How do the different stages in the LLM training pipeline or different sources of training data impact the knowledge learned about ethics?
- How robust and consistent are the moral judgments of a given LLM? This could be investigated with respect to, e.g., word choice or framing, and cross-lingual or cross-cultural effects.

This project forms part of a collaboration between researchers from several departments at the University of Oslo, including the Language Technology Group at the Department of Informatics, the Practical Philosophy Group at the Department of Philosophy, Classics, History of Art and Ideas, and the Department of Biostatistics at the Institute of Basic Medical Sciences.

Specific project requirements

- Master's degree in computer science with a specialization in Natural Language Processing, including background from neural Machine Learning.
- Demonstrable experience of/interest in philosophy and ethics. A formal qualification in philosophy is highly advantageous, but not a requirement.

Supervisors

- Professor Erik Velldal, erikve@ifi.uio.no contact person for inquiries about the position
- Professor Anna Smajdor
- Professor Manuela Zucknick

Discrepancy measures and accelerated likelihood-free inference for simulator-based models (Project #7)

University of Oslo, Oslo Centre for Biostatistics and Epidemiology, Institute of Basic Medical Sciences

Three year PhD position

Description

Multiple research domains ranging from physics to bio-medical research have started to incorporate simulators to answer scientific questions that have eluded us thus far due to lack of computational power and the modelling capabilities. Simulator-based inference is a computational tool that has emerged from the intersection of statistics and machine learning and that is rapidly growing. It enables to calibrate mechanistic models using computer simulations as the main tool. We will develop new efficient probabilistic surrogate models for the simulator-based inference problem. The surrogate modelling can be used in multiple ways in simulation-based inference, as it can be used to replace virtually any or all components of the stochastic simulation system. We will develop theoretical guarantees for our new methods. We will investigate whether our new methods for quantifying the difference between synthetic simulation data and observational real-world data will outperform workflow-based human-in-the-loop strategies.

Specific project requirements

- Master’s degree in statistics, computer science, data science, mathematics, or a related quantitative subject with proven competence in statistics and/or machine learning.
- Documented experience in scientific programming (especially Python) will be an advantage.

Supervisors

- Professor Jukka Corander, jukka.corander@medisin.uio.no contact person for inquiries about the position
- Researcher Henri Pesonen
- Professor Robert Jenssen

Knowledge and bias extraction from Large Language Models (Project #8)

University of Oslo, Department of Informatics

Three year PhD position
(An extension of the appointment by up to twelve additional months may be considered, which will be devoted to career-enhancing duties, e.g. teaching or supervision. This will be dependent on the qualifications of the applicant and the specific teaching needs of the department.)

Description

A major concern when dealing with complex machine learning models is how to determine what influences the outcome of such models, the well-known explainability challenge. There have been several approaches to extract interpretable abstractions from machine learning models, in particular Large Language Models (LLMs). These approaches greatly contribute to effectively detecting useful information from language models, such as harmful biases. However, they often tackle this issue from a purely practical point of view and therefore lack theoretical guarantees. This project aims to employ techniques based on computational learning theory to extract rules potentially expressing harmful biases from language models, while providing theoretical guarantees for the rules extracted.

Specific project requirements

- Master degree in computer science, machine learning, or other relevant quantitative fields.
- Experience with Natural Language Processing, in particular large language models (LLMs).
- Experience with working on propositional logic and logical reasoning is an advantage.

Supervisors

- Associate Professor Ana Ozaki, anaoz@ifi.uio.no contact person for inquiries about the position
- Professor Lilja Øvrelid
- Professor Erik Velldal

Embedded sufficient statistics (Project #9)

University of Oslo, Department of Informatics

Three year PhD position

Description

Popular Neural Networks have various types of layers such as feedforward, residual, multi-head attention, bias, and normalization layers. Each layer represents a statistical summary (embedding) of the input data. The goal of this project is to understand how different types of layers of such architectures summarize input data distribution for specific purposes. Merging together concepts, approaches and cultures of theoretical machine learning and statistics, we will develop new methods and theories to investigate the embedding of data as a way to compress and specialise information. The embedded sufficient statistics achieves the (almost) minimum description that optimally encodes the data distribution, which is a long-standing open problem in Machine Learning.

Specific project requirements

- Master degree in computer science, statistics, machine learning, data science, mathematics, physics, engineering, or other relevant field
- Programming experience with deep neural networks (e.g. PyTorch, Tensorflow) is an advantage but is not necessary, as it can be learned.

Supervisors

- Associate Professor Ali Ramezani-Kebrya, ali@ifi.uio.no contact person for inquiries about the position
- Professor Arnoldo Frigessi
- Professor Sylvia Richardson

Multi-modal data integration for personalized treatment recommendations (Project #10)

University of Oslo, Oslo Centre for Biostatistics and Epidemiology, Institute of Basic Medical Sciences

Four year PhD position

Start no later than 1 October 2024

Description

Many biomedical studies profile patients using multiple modalities (e.g. various omics and imaging technologies), but there is a lack of knowledge-driven approaches of how to best use these multimodal datasets together to improve patient management. Further, most real-world healthcare datasets contain high degrees of missing data, together with high-dimensional and correlated genomic and molecular features, which hinder the use of standard ML approaches in treatment response modelling and prediction. This project will develop new knowledge-driven ML methods to make the most of these incomplete datasets for personalized treatment recommendations using logic-based decision rules and decision systems, where biological and clinical knowledge guides the search space exploration. We will develop ML approaches that allow data integration even in the presence of data missing not at random, enabling prediction without imputation but by exploiting the modalities redundancy structure. Therefore, a key theme in this project is redundancy in the heterogeneous data sources.

Specific project requirements

- Master degree in machine learning, theoretical computer science, data science, (bio)statistics, bioinformatics, mathematics, or other relevant field
- Good programming skills (e.g. Python, R) and ability to work with version control tools (e.g. Git)
- Experience in bioinformatics data analysis tools for large-scale omics data is an advantage but is not necessary, as it can be learned
- Contract beginning no later than 1 October 2024
- This PhD position has four instead than three years, because 25% of the time will be dedicated to teaching and/or administrative tasks.

Supervisors

- Associate Professor Valeria Vitelli, valeria.vitelli@medisin.uio.no contact person for inquiries about the position
- Researcher Basil Ell
- Professor Tero Aittokallio

Learning differential models (and other logic rules) from data with uncertainty (Project #11)

University of Oslo, Oslo Centre for Biostatistics and Epidemiology, Institute of Basic Medical Sciences

Three year PhD position

Description

We want to discover the rules that govern relationship between factors, functions, or variables from data. This is useful for prediction, generalisation, and to understand processes and systems. We start with rules that can be written as differential equations for systems evolving in time. This belongs to physics-informed machine learning. The task can be for example to model the evolution of the volume of a tumour of a patient. In such a case, and in many other real-world situations, not much data is available, as measurements are clinically intensive. In this PhD project, we will develop new methods and algorithms in the situation when data is scarce. One approach will be to combine inference on the differential equation with a stochastic simulation of the system under study, so to be able to generate synthetic data which are then useful for inference. Symbolic regression (SR) is an interpretable approach to learn the time dynamics of the system under study, as it represents the differential equation as a (possibly sparse) tree. Because the data are always measured with error, and the simulation algorithm is stochastic, the estimated differential equation is uncertain and we will find new ways to quantify such uncertainty, for example in terms of distributions over trees of varying dimension. We will study how the uncertainty depends on data, and on the choice of basis functions used in SR; we will investigate the presence of phase transitions of the system under study, expressed by sudden changes in the closed form of the estimated differential equation.

This PhD project aims to develop new methods with theoretical understanding, but will also test the new approaches on real patient data, with the aim to discover for the first time the differential equations governing a breast tumour growth. We will also investigate the estimation of other types of rules, beyond differential equations, for example logic relations, from noisy data.

Specific project requirements

- Master’s degree in statistics, mathematics, machine learning, theoretical computer science or a related quantitative subject with proven competence in statistics and/or machine learning.
- Excellent and documented experience in scientific programming is necessary.

Supervisors

- Professor Arnoldo Frigessi, arnoldo.frigessi@medisin.uio.no contact person for inquiries about the position
- Associate Professor Benjamin Ricaud
- Senior Researcher Alvaro Köhn-Luque
- Professor Martin Giese

Scalable and knowledge-driven identification of anomalous structure in evolving data stream settings (Project #12)

Place of employment: Bosch Centre for Artificial Intelligence, Renningen, Germany
PhD programme: University of Oslo, Faculty of Mathematics and Natural Sciences, Department of Mathematics

Three year PhD position

About this position

This project is a collaboration between Integreat, the Bosch Centre for Artificial Intelligence (Germany) and the Department of Mathematics and Statistics, University of Lancaster (UK).

The selected applicant will be employed at Bosch Centre for Artificial Intelligence (Germany) and will be enrolled in the PhD programme of the Faculty of Mathematics and Natural Sciences at the University of Oslo, and must adhere to all requirements of this PhD programme. It is a requirement to spend minimum 18 months during your PhD at University of Oslo, the actual time for your stays will be agreed upon individually.

The applicants will be evaluated by all partners. As part of the recruitment and evaluation process, shortlisted candidates will be invited to visit Bosch before possible employment.

Description

Anomaly detection is the problem of detecting and locating points or regions - temporal intervals or subsets of features - in data that behave differently compared to some baseline behaviour. It is important to detect anomalies because they often indicate unexpected events of interest or behavioural patterns in complex systems such as manufacturing environments, finance and health. Automated, quick, customisable and reliable methods for anomaly detection are therefore in substantial and increasing demand. The PhD project will in particular focus on

- Online/real-time anomaly detection in data streams, including challenging scenarios when the baseline structure is evolving, potentially undergoing concept drift or being time-dependent.
- Online classification of anomalous-type, by exploring ways in which contemporary machine learning based classification methods like neural nets or autoencoders can be deployed to classify anomalous regions in real time.

We will develop new statistical and machine learning methodology and theory, together with efficient algorithms implemented and evaluated in challenging real-world settings at Bosch, with the possibility to be considered as industrial standard.

Project specific requirements

- Master’s degree in statistics, mathematics, computer science or a related quantitative subject with proven competence in statistics and/or machine learning.
- Genuine interest in methodological research
- Documented experience in scientific programming is necessary

Bosch offers:

- Competitive salary
- Access to travel budget for attending national and international conferences, schools, workshops, etc
- A unique research environment at Integreat and at Bosch with multiple opportunities to develop research themes at the forefront of modern science.
- A friendly professional and stimulating international working environment at Integreat and at Bosch .
- Access to a network of top-level national and international collaborators at Integreat and at Bosch .
- A vibrant international academic environment at Integreat and at Bosch.
- Career development programmes at Bosch and UiO and individual professional plan for the full duration of the doctoral research period.
- Funds through Integreat for shorter research mobility.
- Oslo’s and Stuttgart's family-friendly surroundings with their rich opportunities for culture and outdoor activities.

Supervisors

- Professor Ingrid K. Glad, UiO, glad@math.uio.no contact person for inquiries about this position
- Associate Professor Evgeny Kharlamov, UiO and Bosch
- Senior research scientist Martin Tveten, NR
- Professor Idris Eckley, University of Lancaster
- Associate Professor Michael Kampffmeyer, UiT

Uncertainty quantification in the presence of logical constraints (Project #13)

Place of employment: Bosch Centre for Artificial Intelligence, Renningen, Germany
PhD programme: University of Oslo, Faculty of Mathematics and Natural Sciences, Department of Informatics

Three year PhD position

About this position

This project is a collaboration between Integreat amd the Bosch Centre for Artificial Intelligence (Germany).

The selected applicant will be employed at Bosch Centre for Artificial Intelligence (Germany) and will be enrolled in the PhD programme of the Faculty of Mathematics and Natural Sciences at the University of Oslo, and must adhere to all requirements of this PhD programme. It is a requirement to spend minimum 18 months during your PhD at University of Oslo, the actual time for your stays will be agreed upon individually.

The applicants will be evaluated by all partners. As part of the recruitment and evaluation process, shortlisted candidates will be invited to visit Bosch before possible employment.

Description

One of the most exciting frameworks to quantify uncertainty of predictions is conformal prediction (CP). Under appropriate exchangeability conditions, it provides a confidence set (credibility set if prediction is Bayesian) for a multivariate estimate with statistical coverage guarantees. This PhD project will develop new CP methods for knowledge graphs (KGs), which are one of the most popular approaches for (semi-)structured data. There are many learning tasks that are based on KGs, such as KG completion, link prediction, and node and graph classification. Graph Neural Networks (GNNs) are very successful for learning on KGs and solving the mentioned tasks, but also have great potential for incorporating symbolic knowledge due to strong connections between GNNs and logics. We will develop methods for logic-aware CP on KGs using GNN for prediction and design new algorithms with theoretical guarantees. Then we will verify practical applicability and usefulness of these ideas and algorithms on benchmarks and on challenging real-world settings at BOSCH, with the possibility to develop industrial standard. This PhD project will be at the interface between statistics, logic and machine learning.

Requirements

- Master's degree in statistics, logic, mathematics, theoretical computer science or a related quantitative subject with proven competence in statistics and/or logic and/or machine learning.
- Genuine interest in methodological research.
- Documented experience in scientific programming is necessary

Bosch offers:

- Competitive salary
- Access to travel budget for attending national and international conferences, schools, workshops, etc
- A unique research environment at Integreat and at Bosch with multiple opportunities to develop research themes at the forefront of modern science.
- A friendly professional and stimulating international working environment at Integreat and at Bosch.
- Access to a network of top-level national and international collaborators at Integreat and at Bosch.

- A vibrant international academic environment at Integreat and at Bosch.
- Career development programmes at Bosch and UiO and individual professional plan for the full duration of the doctoral research period.
- Funds through Integreat for shorter research mobility.
- Oslo’s and Stuttgart's family-friendly surroundings with their rich opportunities for culture and outdoor activities.

Supervisors

- Professor Arnoldo Frigessi, UiO, frigessi@uio.no contact person for inquiries about the position
- Associate Professor Egor Kostylev, UiO
- Associate Professor Evgeny Kharlamov, UiO and Bosch