CSC PhD projects LIACS 2021

Contents

CSC fully funded 4 years projects

1. Power and Limits of Adaptivity in Black Box Optimization

Keywords: Black-box optimization, one-shot optimization, artificial intelligence

Many, if not most, real-world optimization problems are not given as explicit mathematical formulae, but require expensive evaluations to assess and to compare the quality of different alternatives. Optimizing such ***black-box problems*** is typically done by heuristics which evaluate potential solutions and then adjust their sampling distribution according to the quality of the evaluated solutions. The question how to best adjust the sampling strategy is at the very heart of research in Artificial Intelligence. Unfortunately, some situations do not allow iterative sampling, e.g., when the evaluations are very costly or take too much time. In such cases, we need to resort to ***one-shot optimization***, where a set of solution candidates can be evaluated in parallel, but where a final decision has to be made after a single iteration. Interestingly, the performance gap between one-shot optimization and iterative optimization with the same budget can be surprisingly small for some applications, whereas it can be arbitrarily large in the worst case. The aim of this PhD project is to systematically analyze the power and limitations of adaptiveness in black-box optimization.

Depending on the candidate's background and interest, the project can be purely theoretical or fully experimental. The project will be carried out in collaboration with Carola Doerr, CNRS researcher at Sorbonne University in Paris, France.

Positions: 1 (4 years, fully funded by CSC)

Contact: Prof. Dr. Thomas Baeck, Head (t.h.w.baeck@liacs.leidenuniv.nl)
Natural Computing Group
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:

- an MSc degree in Computer Science;
- knowledge in Machine Learning, time series forecasting, and evolutionary computation is helpful;
- strong analytical and problem-solving skills;
- very good command of the English language (both written and spoken);
- excellent communication and presentation skills;
- excitement to work in a vibrant, dynamic, and exciting environment and to do cutting edge research;
- ability to work in an international research team.

2. Characterization of Discrete and Mixed-Integer Optimization Problems for Supervised Learning Approaches

Keywords: Black-box optimization, mixed-integer optimization, algorithm selection and configuration, machine learning.

Black-box optimization deals with the solution of problems that are not modeled as explicit functions, but which require simulations or experiments to assess the quality of potential solutions. A plethora of black-box optimization heuristics exists, with complementing strengths on different types of problems or instances. While this variety offers great flexibility, it leaves the user with the challenging task of selecting and configuring an appropriate algorithm for the problem at hand.

Until recently, algorithm selection and configuration have been solved more or less manually, relying entirely on the user's experience with similar optimization problems. Modern AI approaches aim at automating algorithm selection and configuration through data-driven performance extrapolation. A key ingredient for this automation are features which characterize the problem at hand. However, while feature extraction is very actively explored for some specific optimization problems (SAT solving, TSP problems, numerical black-box optimization), the important cases of discrete and mixed-integer black-box optimization have received much less attention, leading to unsatisfactory results.

The aim of this PhD project is to develop a meaningful set of features for discrete and for mixed-integer black-box optimization, to demonstrate its effectiveness, and to make a feature extraction module available to researchers and practitioners working with black-box optimization techniques. The module will be integrated into IOHprofiler, https://iohprofiler.github.io/, our benchmarking platform for black-box optimization algorithms.

The project will be carried out in collaboration with Carola Doerr, CNRS researcher at Sorbonne University in Paris, France.

Positions: 1 (4 years, fully funded by CSC)

Contact: Prof. Dr. Thomas Baeck, Head (t.h.w.baeck@liacs.leidenuniv.nl)
Natural Computing Group
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:

- an MSc degree in Computer Science;
- knowledge in Machine Learning, time series forecasting, and evolutionary computation is helpful;
- strong analytical and problem-solving skills;
- very good command of the English language (both written and spoken);
- excellent communication and presentation skills;
- excitement to work in a vibrant, dynamic, and exciting environment and to do cutting edge research;
- ability to work in an international research team.

3. Adaptive Variation Operators for Iterative Optimization Heuristics

Keywords: Evolutionary computation, adaptive variation operators, iterative optimization heuristics.

Iterative optimization heuristics (IOHs) are algorithms designed to solve a given optimization problem through a sequential process, which comprises the following steps:

1. select from the decision space a set of solution candidates,
2. evaluate these ***search points***,
3. adjust the strategy by which the next solution candidates are generated.

The third step is crucial for the algorithm's performance. The key challenge is to extract meaningful information from the search points evaluated so far. Many state-of-the-art heuristics use surprisingly simple rules, and only adjust the center of the search region, but not the intensity nor the direction of the search. This applies in particular to IOHs operating on discrete or on mixed-integer problems.

The aim of this PhD project is to develop variation operators that learn more actively from the optimization trajectory. In particular, we aim to develop techniques that learn to identify and to track which decision variables are the most relevant ones, how they interact, and by which intensity they should be perturbed to (a) maximize the information gain in early stages of the optimization process and to (b) focus the search on the relevant parts of the decision space in the later stages.

Depending on the candidate's background and interest, the project can be purely theoretical or fully experimental. The project will be carried out in collaboration with Carola Doerr, CNRS researcher at Sorbonne University in Paris, France.

Positions: 1 (4 years, fully funded by CSC)

Contact: Prof. Dr. Thomas Baeck, Head (t.h.w.baeck@liacs.leidenuniv.nl)
Natural Computing Group
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:

- an MSc degree in Computer Science;
- knowledge in Machine Learning, time series forecasting, and evolutionary computation is helpful;
- strong analytical and problem-solving skills;
- very good command of the English language (both written and spoken);
- excellent communication and presentation skills;
- excitement to work in a vibrant, dynamic, and exciting environment and to do cutting edge research;
- ability to work in an international research team.

4. Machine Learning for Evolutionary Strategies: The Self-Improving Optimization Algorithm

Keywords: Evolutionary algorithms, machine learning, exploratory landscape analysis, transfer learning, algorithm configuration.

As of today, evolutionary algorithms start from scratch on optimization problem they are solving, i.e., they do not have a memory of previously executed runs and optimization problems that they have been applied to, such that they could learn how to more efficiently solve a new problem.

One of my research visions is to change this by combining evolutionary algorithms with machine learning, such that the algorithm continuously learns. A first step towards this vision is based on the idea of problem classes in continuous optimization, i.e., subsets of optimization problems which are sufficiently similar such that some kind of transfer learning will be possible.

The aim of the PhD project is to develop a concept of similarity / distance measure between continuous objective functions, such that similarity of problem instances can be measured, e.g., through some kind of exploratory landscape analysis. Based on evolutionary strategies as a state-of-the-art algorithm for continuous optimization problems, we will then combine configurable evolutionary strategies, exploratory landscape analysis, and a suitable machine learning algorithm such as deep neural networks, for example, to facilitate the learning of a mapping between problem instances and algorithm configurations for evolutionary strategies, such that the best possible evolutionary strategy can be chosen automatically for optimizing a new problem instance. This approach will be extended to implement a lifelong learning algorithm which makes evolutionary strategies continuously more powerful with each new problem they have solved.

Positions: 1 (4 years, fully funded by CSC)

Contact: Prof. Dr. Thomas Baeck, Head (t.h.w.baeck@liacs.leidenuniv.nl)
Natural Computing Group
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:

- an MSc degree in Computer Science;
- knowledge in Machine Learning, time series forecasting, and evolutionary computation is helpful;
- strong analytical and problem-solving skills;
- very good command of the English language (both written and spoken);
- excellent communication and presentation skills;
- excitement to work in a vibrant, dynamic, and exciting environment and to do cutting edge research;
- ability to work in an international research team.

5. Automated Machine Learning for Multivariate Time Series Prediction

Keywords: Automated machine learning, multivariate time series prediction, hyperparameter optimization.

Multivariate time series prediction is an important field for real-world applications, such as industrial production process control and optimization, for example (smart industry). For deploying it in industry, however, the methods should be as automatic as possible, by training, validating, and updating the prediction models automatically.

Recent research in my group has shown that automated machine learning pipelines using feature engineering, feature selection, supervised machine learning and hyperparameter optimization approaches can provide excellent results when applied to time series classification problems in domains as divers as automotive and health. The aim of this PhD project is to extend such approaches for multivariate time series prediction.

The scientific challenges include, for example, the definition and computation of features that capture dependencies between time series (e.g., using information-theoretic approaches), the identification of time lags of such dependencies, and the utilization of a multi-output prediction approach such as deep neural networks, LSTMs, etc. In cooperation with our industrial partners, the methods will be tested on academic as well as real-world test data.

Positions: 1 (4 years, fully funded by CSC)

Contact: Prof. Dr. Thomas Baeck (t.h.w.baeck@liacs.leidenuniv.nl)
Natural Computing Group
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:

- an MSc degree in Computer Science;
- knowledge in Machine Learning, time series forecasting, and evolutionary computation is helpful;
- strong analytical and problem-solving skills;
- very good command of the English language (both written and spoken);
- excellent communication and presentation skills;
- excitement to work in a vibrant, dynamic, and exciting environment and to do cutting edge research;
- ability to work in an international research team.

6. Evolving pseudorandom generators with Genetic Programming

Keywords: pseudorandom generator; genetic programming; random number test

In the modern society, computing devices have truly become omnipresent. We use desktop computers, mobile phones, servers, sensor-based computing devices, etc. to automate and speed up a huge variety of tasks and constantly aspire for all such devices to become ever better. The majority of improvement of such devices comes from hardware. However, there is a clear potential for improvement in the algorithmic aspect of computation on these devices. One direction for algorithmic improvements of computing devices lies in the area of pseudorandom generators which are heavily used in various cryptographic and optimisation applications.

Modern computers, being as advanced as they are, are by design not capable of producing truly random numbers since, put simply, randomness is a lack of pattern and any computer code follows some 'pattern'. Thus, *true randomness* has to be replaced by *pseudorandomness* where a pseudorandom number is deterministically generated by a special algorithm - a pseudorandom number generator (PRNG) from a given 'seed'. The `pseudo' in pseudorandom stems from the fact that a sequence of numbers generated by a PRNG is completely determined by the seed.

Creating a good PRNG is not easy since it should be capable, within reasonable amount of time, of generating sequences of numbers that seem to behave as if they were generated randomly from a specified probability distribution. The quality criteria for PRNGs vary and depend on the intended application: fast statistically good versions for simulations and slower but unpredictable versions for cryptographic applications where data integrity is only guaranteed by a diminishingly low possibility of successful reverse engineering from the data.

This project aims at evolving new pseudorandom generators by means of Genetic Programming approach. It would involve:

- investigation of the state-of-the-art in the PRNGs creating and procedures for their benchmarking;
- formulation of a suitable objective function based on current benchmarking approaches for PRGNs;
- development Genetic Programming approach for PRGNs generation;
- investigation of the impact PRGNs of different quality have on applications such as optimisation, cryptography, etc.

Positions: 1 (4 years, fully funded by CSC)

Contact: Prof. Dr. Thomas Baeck (t.h.w.baeck@liacs.leidenuniv.nl)
Natural Computing Group
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:

- an MSc degree in Computer Science with the focus on Theoretical Computer Science or similar;
- strong analytical and problem-solving skills;
- excellent programming and debugging skills in C/C++, Java, and/or Python;
- basic knowledge of Unix;
- very good command of the English language (both written and spoken);
- excellent communication and presentation skills;
- ability to work in an international research team.

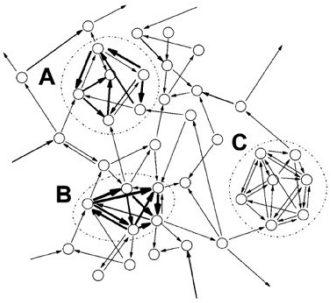
7. Analysis of evolving populations using network science

Keywords: network science; heuristic optimisation; simulated evolution; nature inspired optimisation; random graphs.

Many nature inspired optimisation algorithms are population-based. It is interaction of entities inside such simulated populations that drives the sought-after improvement of the objective function and explains the popularity of this class of algorithms. But what do we know about such interaction?

Surprisingly enough, not much. Thus, the proposed project concentrates around the question "Given volumes of evolution data from various nature inspired population-based algorithms, what can we learn about properties of interactions inside the evolving populations if we apply methods from network science?"

From this



to this

Positions: 1 (4 years, fully funded by CSC)

Contact: Prof. Dr. Thomas Baeck (t.h.w.baeck@liacs.leidenuniv.nl)
Natural Computing Group
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:

- an MSc degree in Computer Science or Mathematics;
- solid background in heuristic optimisation;
- knowledge of Graph Theory and/or Network Science;
- prior expertise in one or more of the following fields: random graphs, heuristic optimisation, nature inspired algorithms, population-based algorithms;
- strong analytical and problem-solving skills;
- excellent programming and debugging skills in C/C++, Java, and/or Python;
- basic knowledge of Unix;
- a very good command of the English language (both written and spoken);
- excellent communication and presentation skills;
- ability to work in an international research team.

8. Parameter setting in Differential Evolution

Keywords: differential evolution; parameter setting; algorithmic analysis; algorithm configuration.

Differential evolution (DE) and its variants are amongst the most successful algorithms for heuristic real-valued optimisation due to their simplicity and low number of parameters. However, surprisingly little research has been published regarding setting of these parameters. Thus, this project aims at investigating various strategies for setting parameters of basic and more advanced variants of DE and studying applicability of such setting. The project involves experimental and, where possible, theoretical approaches.

Positions: 1 (4 years, fully funded by CSC)

Contact: Prof. Dr. Thomas Baeck (t.h.w.baeck@liacs.leidenuniv.nl)
Natural Computing Group
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:

- an MSc degree in Computer Science or Mathematics;
- prior expertise in one or more of the following fields: heuristic optimisation, nature inspired algorithms, population-based algorithms;
- strong analytical and problem-solving skills;
- excellent programming and debugging skills in C/C++, Java, and/or Python;
- basic knowledge of Unix;
- a very good command of the English language (both written and spoken);
- excellent communication and presentation skills;
- ability to work in an international research team.

9. Abstract models for mobile application security

Keywords: Mobile security, abstract models, program analysis, malware detection, repackaging detection, automated testing

Mobile security, and most notably Android security, has become a very important issue in the past few years. Due to the proliferation of malicious and buggy applications, new techniques are being developed to identify malware, to find repackaged applications, and to automatically test third-party applications for crashes. These techniques frequently utilize abstract models of application code, like data flow graphs or control flow graphs. In this thesis project, the student will study these models to identify specific semantic representations that capture some desired properties, like "private data stealing" or "similar execution patterns". The main goal will be to develop a unified framework to express the abstract properties relevant for the task chosen by the analyst. An example of question that could be answered by this framework: "What abstract representation of an application code can be used to better capture the evolution of malware behavior within a single family?"

Positions: 1 (4 years, fully funded by CSC)

Supervisors: Dr. O. Gadyatskaya, Prof.dr. H.A.G. Wijshoff

Contact: Dr. O. Gadyatskaya (o.gadyatskaya@liacs.leidenuniv.nl)
High Performance Computing Group
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:

- Necessary: Interest in security, experience with programming (e.g., Python), MSc degree in Computer Science or a relevant domain.
- Preferably: experience with program analysis and large scale experiments, interest in formal methods. This project will entail a combination of theoretical and practical work, including producing abstract models for mobile apps, developing prototypes to extract these models from real apps, evaluating these prototypes experimentally on large application sets.

10. AutoML for Mobility Data

Keywords: Mobility modeling, spatio-temporal data modeling, urban computing, AutoML, open machine learning

Mobility datasets are a class of datasets generated as a result of everyday use of location-aware technology such as mobile phones, public transport cards, GPS enabled vehicles, etc. Such data is used in various application areas from smart mobility and urban planning, to designing context-aware mobile phone applications.

Using mobility data within machine learning algorithms is challenging as such data has properties that are distinctive from other data formats. A machine learning expert needs to design a solution for different phases of this process from pre-processing, feature-extraction, to algorithm selection and hyper-parameter setting. This process is often tedious, requires ad-hoc decision making, and often leads to non-optimal solutions that only work on one dataset, aiming at a specific application. All these complexities limit the accessibly of these models for application domain users who aim to use their mobility data as easily as possible.

AutoML as an emerging research field focuses on addressing this problem by developing methods and processes to automate all complicated tasks mentioned above. The goal of this PhD project is to research AutoML for mobility data. This involves proposing new representation learning algorithms and new efficient methods for algorithm selection and hyper-parameter optimization. We envision to demonstrate the results of this project by creating AutoML solutions for a set of well-known tasks in mobility data modeling.

Positions: 1 (4 years, fully funded by CSC)

Supervisors: Prof.dr. Holger Hoos, Dr. Mitra Baratchi

Contact: Dr. Mitra Baratchi (m.baratchi@liacs.leidenuniv.nl)
Prof.dr. Holger Hoos (h.h.hoos@liacs.leidenuniv.nl)
ADA Research Group
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:

- A master degree with major in Mathematics, Computer Science, Statistics.
- Interest in artificial intelligence, machine learning.
- Proficiency in a programming language
- Excellent English proficiency

11. RISC-V extensions for lightweight cryptography

Keywords: cryptographic coprocessors, instruction set extensions, RISC-V, lightweight cryptography

Summary: RISC-V is an instruction set architecture (ISA) that is free and open, i.e., no licenses or fees are required to use, extend or integrate RISC-V-based cores (https://riscv.org/). RISC-V was developed at UC Berkeley a decade ago and has been used worldwide by many universities and companies ever since. In order for RISC-V-based cores to be compliant with the low energy requirements of many Internet-of-Things (IoT) networks, ultra-low-power RISC-V cores have been proposed. An example is the Ibex core (https://ibex-core.readthedocs.io/en/latest/).

When personal or company-critical data are exchanged in IoT networks, it is important to foresee security mechanisms that have a minimal impact on the energy consumption and the cost of the individual IoT devices. Therefore, the National Institute of Standards and Technology (NIST) issued a competition for the standardization of lightweight cryptographic algorithms for encryption and authentication. This PhD project investigates how custom hardware modifications to RISC-V-based processors can be deployed to efficiently execute lightweight cryptographic algorithms. The resulting coprocessors and/or instruction set extensions will be evaluated in terms of energy consumption, silicon area and performance. Further, countermeasures against physical attacks will be designed, implemented and evaluated.

Positions: 1 (4 years, fully funded by CSC)

Supervisors: Prof. dr. ir. Nele Mentens

Contact: Prof. dr. ir. Nele Mentens (n.mentens@liacs.leidenuniv.nl)
Systems & Security Cluster
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:
computer architecture, hardware design (VHDL or Verilog)

12. Novel hardware security architectures for remote attestation and remote updates in IoT networks

Keywords: remote attestation, secure remote updates, hardware security

When electronic devices are deployed in the Internet of Things (IoT), it is important to make sure that these devices are not running malware. On the one hand, the status of the firmware should be regularly checked. This can be done through remote attestation, i.e., a challenge-response protocol between a verifier (e.g., a central server) and a prover (i.e., the IoT device). On the other hand, the process of updating the firmware of an IoT device (to heal the device after malware injection or to perform an upgrade) should be done in a secure way. For both remote attestation and remote updates, the highest security level can be achieved through hardware modifications to the processor system. Especially in large IoT networks, it is challenging to enable remote attestation and remote updates in an efficient way with respect to energy consumption, performance and cost. In this PhD project, we will work on novel protocols and hardware security architectures for remote attestation and remote updates.

Positions: 1 (4 years, fully funded by CSC)

Supervisors: Prof. dr. ir. Nele Mentens

Contact: Prof. dr. ir. Nele Mentens (n.mentens@liacs.leidenuniv.nl)
Systems & Security Cluster
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:
computer architecture, hardware design (VHDL or Verilog)

13. Hardware Acceleration for (Big) Data Pipelines

Keywords: big data, FPGA, acceleration, Spark, Hadoop, hardware

The increase in produced and analyzed data poses great problems to widely used computation platforms. This means that we cannot solely depend on faster CPUs to tame the demand of data processing. In big data processing workloads the CPU is already proven to be the major bottleneck. One of the possible solutions to this problem is to take advantage of the processing power of specialized hardware and accelerators, such as FPGAs, GPUs, ASICs, NVMe etc. For example, FPGAs are programmable devices which can be used to accelerate computation. They can be placed close to storage devices or even in network switches so that they alleviate part of the CPU-pressure by pre-filtering data, decompressing high-performance file-formats, or performing hash-table lookups. In this project we seek to investigate the role of such specialized hardware in big data processing pipelines: (1) where do such hardware fit the data processing ecosystem? (2) how can we redesign big data processing pipelines to accommodate such hardware? (3) how much performance can we gain by using these novel designs? We encourage applications from students with a keen interest in low-level systems and their inner workings, as well as analytical skills.

Positions: 1 (4 years, fully funded by CSC)

Supervisors: Dr. A. Uta, Prof.dr.ir. N. Mentens

Contact: Dr. A. Uta (a.uta@liacs.leidenuniv.nl)
High Performance Computing Group
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:
(At least 1-2 of these, or a keen interest in): low-level systems programming, some experience with distributed systems (for example Spark, Hadoop), GPU and/or FPGA programming, computer architecture, hardware design

14. Reproducible Performance Analysis in Large Scale Systems

Keywords: performance analysis, reproducibility, big data, machine learning, Spark, Tensorflow, large-scale infrastructure

In computer systems, performance evaluation involves assessing the time-to-solution, resource usage, scaling and efficiency of systems, as well as low-level hardware performance counters. Performance evaluation in modern computer systems (distributed systems, HPC, cloud computing) is inherently flawed by the lack of performance evaluation protocols and by the rapid dissemination pace of the publish-or-perish academic world. Coupling this with the highly variable cloud, performance evaluation becomes non-reproducible and produces possibly wrong performance conclusions. Wrong performance conclusions can possibly form a threat to users who might rely on guarantees that are not met in practice. In this project we aim to develop protocols and techniques for achieving reproducible performance evaluation. We aim to apply such techniques at large in assessing the performance of modern computing frameworks related to cloud computing (e.g., serverless, FaaS), big data (Spark, Hadoop, large-scale storage), machine learning (Tensorflow, pytorch), HPC (large-scale simulations). We encourage applications from students with a keen interest in low-level systems and their inner workings, as well as analytical skills.

Positions: 1 (4 years, fully funded by CSC)

Supervisors: Dr. A. Uta, Prof.dr. H.A.G. Wijshoff

Contact: Dr. A. Uta (a.uta@liacs.leidenuniv.nl)
High Performance Computing Group
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:
(At least 1-2 of these, or a keen interest in): low-level systems programming, some experience with distributed systems (for example Spark, Hadoop), some experience with HPC (MPI, openMP), math and statistics.

15. AI Applications in Business

Keywords: AI, Responsible AI, Explainable AI, Technology transfer, Business Applications, Python

This research focuses on identifying solutions to next generation business problems that require AI solutions, including AI assisted processing, hyper personalisation, intelligent chatbots, and other areas. It centres on applying existing fundamental research in the areas of Machine Learning, Deep Learning, Text Mining, and Rule-Based Solutions (and possibly combinations thereof) to these business issues. It involves developing prototype solutions, preferably in Python, that can be demonstrated to companies as a basis for technology transfer and further implementation in business environments.

Positions: 1 (4 years, fully funded by CSC)

Supervisors: Prof Dr Joost Visser, Dr Guus Ramackers

Contact: Prof.dr. Joost Visser (j.m.w.visser@liacs.leidenuniv.nl)
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:

- Some knowledge, or motivation to develop knowledge, of AI algorithms and libraries
- Some knowledge of, or motivation to develop skills in, programming in Python
- A hands-on approach to prototyping solutions
- Good level of spoken and written English

16. AI Applications in the System Development Process

Keywords: AI, Enterprise Architecture, ML Integration, Technology transfer, Model Driven Development, Python

This research focuses on the application of fundamental AI technology in the context of large scale Enterprise System Development. It addresses the issue that modelling tools are too complex for effectively involving business experts and end users in the requirements and conceptual modelling phases of system development. In particular, the generation of UML models from verbal and textual input in the context of a Model Driven Development approach in business environments is of interest. In addition, development and deployment of such models as part of an Enterprise Architecture that includes ML components is a focus. It involves developing prototype solutions, preferably in Python, that can be demonstrated to companies as a basis for technology transfer and further implementation in business environments.

Positions: 1 (4 years, fully funded by CSC)

Supervisors: Prof Dr Joost Visser, Dr Guus Ramackers

Contact: Prof.dr. Joost Visser (j.m.w.visser@liacs.leidenuniv.nl)
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:

- Some knowledge, or motivation to develop knowledge, of UML and BPMN modelling technology
- Some knowledge, or motivation to develop knowledge of a Model Driven approach to software development and code generation
- Some knowledge of, or motivation to develop skills in, programming in Python
- A hands-on approach to prototyping solutions
- Good level of spoken and written English

17. From science to innovation: An analysis of patent in text references

Keywords: Text mining, information retrieval, patent analysis, bibliometrics, strategy, innovation.

This PhD project aims to study how science feeds into technological innovation, using patent in-text references as an instrument for tracing the knowledge flow. The project can take several different directions ranging from a computer science focus to a strategy/innovation focus. One possible direction is to develop machine learning methods to extract reference strings in patent full texts and then matching them to scientific publications. Another possible direction is to analyze the text around references strings to understand the context of referencing. The third possible direction is to studying how different types of science leads to different types of innovation.

Positions: 1 (4 years, fully funded by CSC)

Supervisors: Prof.dr. Simcha Jong, Dr. Jian Wang

Contact: Dr. Jian Wang (j.wang@sbb.leidenuniv.nl)
Science Based Business
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:
Student should be enthusiastic about the topic, preferably with education background in one or more of the following fields: computer science, statistics, information science, economics, sociology, management.

18. Understanding funding allocation in science: The role of panel composition

Keywords: Funding policy, sociology of science, economics of science, bibliometrics, innovation.

Scientific funding is increasingly allocated through competitive project grants, and the review panel in funding agencies play an important role in funding decisions. This PhD project aims to answer the following questions: (1) who end up on funding review panels, (2) how panel composition affects funding decisions, and (3) how it is different across scientific disciplines.

Positions: 1 (4 years, fully funded by CSC)

Supervisors: Prof.dr. L. Waltman (CWTS), Prof. Dr. S. Jong, Dr. V. Traag (CWTS), Dr. Jian Wang

Contact: Dr. Jian Wang (j.wang@sbb.leidenuniv.nl)
Science Based Business
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:
Student should be enthusiastic about the topic, preferably with education background in one or more of the following fields: computer science, statistics, information science, economics, sociology, management, public policy.

19. New Programming Models for High Performance Computing

Keywords: High Performance Computing, Programming Models, Program Transformations, Compiler Technologies

Efficient system software (runtime systems, I/O, optimizing compilers) is essential for obtaining maximum performance on High Performance Computers. The development of such software is a difficult and time-consuming process involving intricate optimizations and requiring expert programmers. Underlying hardware platforms continue to advance, for instance resulting in a higher number of cores per socket (and thus per node), modified memory hierarchies and new computing paradigms such as GPUs and domain specific architectures.

Within this project the students will work on a novel programming model which enables system software efficiency at the same time as application programming. This programming paradigm will also lead to new architectural requirements for future high performance computing platforms. These requirements range from memory design, cache coherence implementations as well as optimized (parallel) I/O. This requires from students an overall in-depth knowledge on parallel applications as well as on modern computer architectures.

Positions: 1 (4 years, fully funded by CSC)

Supervisors: Prof.dr. H.A.G. Wijshoff, Dr. K.F.D. Rietveld

Contact: Prof.dr. H.A.G. Wijshoff (h.a.g.wijshoff@liacs.leidenuniv.nl)
High Performance Computing Group
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Requirements:

- Master degree in Computer Science.
- Must have taken courses in the area of computer systems and high performance computing, with excellent results.
- Excellent programming and debugging skills, with experience in C, C++ and Python on UNIX-based systems.
- Very good English proficiency.

Keywords for XJTU-Leiden joint Phd projects / own projects

We are looking for students already involved in a PhD process at XJTU and with relevant experience and publications in one or more themes as illustrated by the keywords below:

- **Keywords:** Decision making, Logistics, Transport planning
  Contact: Dr. Yingjie Fan (y.fan@sbb.leidenuniv.nl)
  Science Based Business
  Leiden Institute of Advanced Computer Science (LIACS), Leiden University

  Decision making for resilient global logistics systems
  The aim of this project is to develop methods and tools to provide decision support for transportation planning in order to improve the resilience and flexibility of global logistics networks. The successful candidate will work on data analytics in predicting potential disruptions of global logistics systems and developing resilient transportation plans (e.g., route plans, mode choices, vehicle schedules) in disruptive environments. Optimization, simulation, and data analytics will be used in this research project.

- **Keywords:** Analytics for threat modeling, threat intelligence analytics, security visualisations using threat models, security information sharing
  Contact: Dr. Olga Gadyatskaya (o.gadyatskaya@liacs.leidenuniv.nl)
  High Performance Computing Group
  Leiden Institute of Advanced Computer Science (LIACS), Leiden University

- **Keywords:** AI for IoT security, AI for embedded security, AI for hardware security, AI for network security
  Contact: Prof. Dr. Ir. Nele Mentens (n.mentens@liacs.leidenuniv.nl)
  Systems & Security Cluster
  Leiden Institute of Advanced Computer Science (LIACS), Leiden University