Data Science

Data science is an interdisciplinary research field that concerns methods, techniques, and tools for organizing and analyzing data, as well as extracting knowledge or insights in order to facilitate decision making. Research in this area may be found under several headings, including knowledge discovery, data and text mining, machine learning, big data analytics, predictive analytics, and intelligent data analysis. Data science can facilitate effective and efficient Artificial Intelligence (AI) by exploiting huge amounts of data sources and providing machines with algorithms and techniques for self-learning.

Analyzing large and complex data sets to support decision making

Our core research focus is on machine learning methods, both supervised and unsupervised, for facilitating decision making and efficient and effective Artificial Intelligence (AI) in the presence of complex data sources.

One of our core research themes involves the development of methods and techniques for sequential and temporal data mining, with emphasis on sequential and temporal data, such as time series, event sequences, time-interval sequences, and on building effective and efficient classification models, searching and indexing data structures, as well as temporal abstractions.

Moreover, we are actively working on interpretable and explainable machine learning models with emphasis on understanding the reasoning and causality behind the predictions.

Another active research theme is predictive data mining using ensemble methods, i.e., techniques for generating sets of models that collectively form predictions by voting, such as random forests, with emphasis on complex data sources, such as time-dependent variables and histogram variables.

Our research on clinical text mining focuses on efficient and resource-lean methods using language technology for very large text sets.

The main application area for our research is healthcare analytics, which aims for providing efficient and effective decision support for healthcare practitioners and policy makers. Both research groups in the area have been in close collaboration for several years with clinical practitioners, researchers in the medical, biomedical, healthcare, and epidemiological domain, as well as with computational chemists in the pharmaceutical industry. This has resulted in new techniques and tools for building predictive models from healthcare data registries, as well as from observed biological activities, e.g., the toxicity levels of chemical compounds, which are currently being used in the industry.

Another application area is modeling of component-wear in heavy trucks using data mining and to provide decision support for optimizing heavy truck fleet utilization.

Finally, our recent application theme is to explore the area of network traffic control and prediction using deep learning.

Contacts

Panagiotis Papapetrou (Professor)
Lars Asker (Assoc. professor)
Tony Lindgren (Assoc. professor)
Jaakko Hollmen (Senior lecturer)
Isak Karlsson (Postdoc. researcher)
Amin Azari (Postdoc. researcher)

For contact information, please consult https://dsv.su.se/en/research/research-areas/datascience

For information about language technology and clinical text mining, please consult Professor Hercules Dalianis and https://dsv.su.se/en/research/research-areas/language

Focus areas

Core research:

- Machine learning methods for Artificial Intelligence
- Sequential and temporal data mining
- Interpretable machine learning
- Ensemble learning
- Natural language processing

Applied research:

- Learning from electronic health records
- Integrated vehicle health management
- Network traffic control and prediction
- Clinical text mining

Ongoing Projects

- Temporal data mining for detecting adverse events in healthcare (Swedish Research Council)
- CODA: Predictive models for interpretability and concept drift analytics (Vinnova/Scania)
- EXTREME: Explainable and Ethical Machine Learning for Knowledge Discovery from Medical Data Sources (ICT-TNG)
- CorIL: Analyzing registry data for improving heart failure treatment (Stockholm County Council)
- TROPICAL: Network traffic control and prediction using artificial intelligence (Huawei Flagship)