Masterarbeit/Bachelorarbeit

Karlsruher Institut für Technologie

Institut für Programmstrukturen und Datenorganisation (IPD)
Lehrstuhl für Systeme der Informationsverwaltung, Prof. Böhm

Masterarbeit/Bachelorarbeit

Change Detection in High-Dimensional Data Streams

In real-world applications such as manufacturing, environmental analyses, and e-commerce, data streams are high dimensional and evolve over time: Weather stations observe multivariate data which shows natural seasonality. Online retailers collect customer behavior through different indicators which may change with their interests. Cars or production machines monitor their internal state by analyzing multivariate sensor data. The detection of change in data streams, i.e., concept drift, is a key component in designing reliable and adaptive systems, as it enables them to react in the first place.

Although there exist numerous techniques for concept drift detection, most of them only handle low dimensional data. They are thus unable to deal with the following two key challenges: First, data streams are generally high-dimensional and concept drift might only be visible in selected subspaces. Second, different types of drift may occur simultaneously at various time scales and in arbitrary subspaces. To the best of our knowledge, there so far exists no approach which addresses both of these challenges.

The focus of this thesis is the development of a framework which identifies simultaneous concept drifts in multivariate data streams. This results in the following tasks:

- **In-depth literature review of concept drift detection in multivariate streams and related fields.**
- **Development of a framework for the detection of concept drift in multivariate data streams.**
- **Evaluation of the developed approach against baselines and state-of-the-art algorithms.**

We offer:

- Continuous and thorough mentoring
- A highly motivated and fun team, constructive teamwork
- Access to the required computing infrastructure

We expect:

- Basic knowledge in Python programming and Data Mining
- Ability to plan and work independently
- Very good knowledge of German or English
- High level of motivation, enthusiasm and curiosity

Throughout this work, the student will acquire state-of-the-art knowledge and practical experience in the domain of data stream monitoring and related fields. The thesis takes place as joint project between IPD Böhm and FZI, i.e., the student will benefit from supervision by both institutions. We are looking forward to your application to Marco Heyden (heyden@fzi.de) and Edouard Fouché (edouard.fouche@kit.edu). Please provide us with a Transcript of Records and Curriculum Vitae.

Responsible institute at KIT for WIWI faculty: AIFB | Prof. Dr. York Sure-Vetter
Responsible institute at KIT for Computer Science: IPD | Prof. Dr.-Ing. Klemens Böhm

Ansprechpartner

Edouard Fouché, M. Sc. edouard.fouche@kit.edu +49 721 608-47337 Raum: 342

Am Fasanengarten 5 76131 Karlsruhe Gebäude: 50.34