Die Fakultät für Informatik der Universität Wien lädt Sie herzlich ein zum

CS-Colloquium

Big Data Analytics in Astrophysics

Prof. Dr. Katharina Morik
Technische Universität Dortmund

Wann?   19. März 2014, 15:00 Uhr

Wo?   Hörsaal 3
Währinger Straße 29
1090 Wien

Abstract

Big Data has become an increasingly popular term since 2011. Its facets, volume, variety, and velocity, have led to new storage and computation methods and hardware architectures. These developments may support but cannot replace the analysis of the data. Analytics is necessary in order to reap the data's benefit. In this talk, analysis methods for large volume data and for real-time velocity data are presented and illustrated by astrophysical experiments.

Astrophysical data can be so large, that they are faster to be shipped than to be broadcasted via satellite. For instance, the IceCube experiment at the South Pole produces 1 Terabyte of data per day. A satellite would need about 10 years to communicate the data of a year, but, stored on hard disk, they are shipped within 28 days from the South Pole to the University of Wisconsin. IceCube searches for neutrinos within a strongly biased distribution of signal and background events: 93,000 neutrinos versus 17.48 million background events. The challenge for Big Data Analytics is to detect the true neutrino events in the data and to carefully evaluate the learned model. Together with Dortmund colleagues from the IceCube collaboration, Wolfgang Rhode and Tim Ruhe, we could significantly increase the rate of detected true neutrinos and recognize a neutrino event at 1 Peta eV. The key to success was the preprocessing which automatically determined the right feature space.

Another physical experiment measures air showers by two Cherenkov telescopes, MAGIC and FACT at La Palma. It aims at detecting gamma-rays in order to better understand active galaxies. From gamma-rays, their source can be calculated, but they come along with several millions observations of other particles whose paths are influenced by intergalactic magnetic fields so that they do not allow to infer their source. Again, the challenge for Big Data Analytics is to find the needle in the hay stack and to justify the finding. An approach to process data streams in real-time and to combine this with the map-reduce approach is illustrated by inspecting the FACT data.

Bio

Katharina Morik is full professor for computer science at the TU Dortmund University, Germany. She earned her Ph.D. (1981) at the University of Hamburg and her habilitation (1988) at the TU Berlin. Starting with natural language processing, her interest moved to machine learning ranging from inductive logic programming to statistical learning, then to the analysis of very large data collections, high-dimensional data, and resource awareness.

Her aim to share scientific results strongly supports open source developments. For instance, RapidMiner started out at her lab, which continues to contribute to it. Since 2011 she is leading the collaborative research center SFB876 on resource-aware data analysis, an interdisciplinary center comprising 12 projects, 19 professors, and about 50 Ph D students or Postdocs.
She was one of those starting the IEEE International Conference on Data Mining together with Xindong Wu, and was chairing the program of this conference in 2004. She was the program chair of the European Conference on Machine Learning (ECML) in 1989 and one of the program chairs of ECML PKDD 2008. She is in the editorial boards of the international journals "Knowledge and Information Systems" and "Data Mining and Knowledge Discovery".

http://www-ai.cs.uni-dortmund.de
http://sfb876.tu-dortmund.de

Contact

Mag. Werner Schröttner, werner.schroettner@univie.ac.at, Tel. +43 1 4277 780 03
http://informatik.univie.ac.at