

Big Data

Session co-chairs: Haiyong Xie, CAEIT and Tamas Sarlos, Google

Abstract

Big data is usually characterized by its uniquely complex combination of volume, variety of data sources, formats and use cases, velocity of changes and updates, and veracity of data quality. Many of today's most successful enterprises - in business and in science - are built on the collection and analysis of data. Science and practice of big data encompasses and extends the fields of statistics, machine learning, databases, distributed systems, algorithms, and applied mathematics. While it's surrounded by significant amount of publicity and sometimes unfounded promises or fears, it is steadily transforming most fields of science from data poor to data rich and data driven. In our session we will focus on two topics: how big data can positively impact our life and the tools and techniques required to develop such systems.

Increasing digitalization of all aspects of life enabled by the proliferation of sensors, rapid advances in computerized data acquisition, storage, and processing lets us analyze and solve challenges faced by human societies relying on far more data than it has been possible ever before. The first speaker of the session, Lakshminarayanan Subramanian, will discuss tackling big societal questions using big data. He will share us three stories on recognizing counterfeit goods, disease surveillance, and detecting events from news. Yanhua Li, our second speaker, will describe a framework for optimal placement of new electric vehicle charging stations that minimizes travel and waiting time by analyzing a large collection of electric taxi trajectories.

Handling large amount of data requires building sophisticated distributed software and hardware infrastructure. Our third speaker, Bin Zhou, will present the Big Search in Cyberspace project and discuss how to construct knowledge warehouses, understand the user's search intent expressed in natural language, match it to the knowledge warehouse, and finally ensure that the search process is secure. Quoc Le, our final speaker, will talk about the principles of large-scale deep learning, the common technique behind multiple recent breakthroughs in object recognition, speech recognition, and understanding natural language, and describe the Google Brain system.