IoT Week 2018 Workshop
Making the Data Revolution Happen

Georg Wittenburg, Peter Wittenburg

This workshop raised three related topics on Big Data which are of great relevance and need to be addressed to overcome emerging bottlenecks:

- Trust and Technology - can it go together?
- Computational Power and Efficiency - will it help overcoming barriers?
- Machine Learning and Big Data - will it help transforming data into knowledge?

In the first session, **Dirk Helbing** described the comprehensiveness of information in the Internet about everything, in particular about citizens, and the inherent danger even for our democratic societies if we do not carefully design and take steps. Basically, he called for informational self-determination and made clear that the core of the actions to take is not technological, but social. Decentralised self-organizing systems are a way out and technology can assist. **Peter Wittenburg** described a new view on data that may help to construct a binding element between social and technical aspects. The concept of "Digital Objects" with clearly defined and identified data entities assigned by persistent and globally resolvable identifiers can act as the constitutional element of a rich data society which will remain an utterly dynamic universe in particular due to the volumes and complexity added through billions of smart IoT devices. **Visa Vallivara** pointed out that Blockchain technology combined with the notion of Digital Objects could be a key technology to build trust in the nascent data economy. Data identities can be coupled with Blockchain traces capturing contracts and events in a way that can be checked at all times and thus defining the basis for trust.
Summarising, we can state that there seem to be ways with Digital Objects and Blockchain technology to establish the basis for increased trust, but it will only be a small step towards trustworthy management and use of data. Societal norms are urgently needed to improve trust.

In the second session, **Thomas Lippert** explained how understanding the human brain -- as part of the Human Brain Project -- requires exascale computing, as well as Artificial Intelligence algorithms running on modular supercomputer architectures. It is in the nature of the large simulation models which will require centralised facilities. The simulation tasks are so extreme that four European HPC centres are now collaborating to easily shift computational jobs between their machines. **Dimitris Koureas** lined out that large computational facilities are one dimension, but that the huge fragmentation from data organisation to semantics needs to be overcome to make data work more efficient. As an example, he referred to the huge biodiversity collections available worldwide which cannot be exploited to its full potential to the benefit of environmental stability. **Milan Petkovic** listed challenges in the domain of healthcare research. In addition to the technical fragmentation, the legal fragmentation in Europe is hampering fast progress. New platforms under development to integrate health data and AI respecting the rules for privacy for example at Philips will facilitate big data approaches in health.
Summarising, we can indeed state that advances in providing large computational facilities to work on the huge volumes of IoT generated data with AI methods will be needed, but that we should not overestimate the use of edge computing for AI. Computational power alone however will not be sufficient to master the challenges. We need considerable investments to overcome the fragmentation at the various dimensions.

In the third session **Matthias Scheffler** gave an impression of big data applications in material science dealing with vectors of billions of elements. Advanced machine learning and other statistical methods running on high performance computers are used to come to new categorisations of

compound materials which may help finally to find answers to new challenges much faster, for example in the domain of IoT devices. **Wojciech Samek** opened the black box of machine learning and suggested methods to better understand how neural networks reason. We need to prevent situations where machine learning which will be deployed everywhere in future leads to erroneous decisions due to artefacts in the training material for example which the creators were not aware of. Finally, **Grazziela Figueredo** used examples from vehicular data and incidents to line out how to setup and run a process to get from abundant data to actual and actionable insights. Smart devices create so much data that proper selection, summarisation and presentation of data are essential to turn data into mind changing messages.

Summarising we need to reemphasise that we have a bunch of extremely helpful statistical algorithms and machines to execute complex calculations, but that the success is critically dependent on the availability of large data sets requiring appropriate regulatory mechanisms, that we need to be careful in relying on these statistical machines and that in many cases proper selection and presentation methods need to be applied to transmit messages.

This workshop was the second one on Big Data within the realm of the IoT Week. In 2017 we put the topics of "identification and identity resolution" (*of data being generated by IoT devices*) in the centre of the presentations and discussions with a clear message that identity of data objects needs to be taken seriously in science and industry. In 2018 we put three key areas into the focus which need to be addressed urgently to come to scalable methods.

- **Technology such as identified Digital Objects combined with Blockchain technology can assist in creating a more trustworthy data domain, but it will the combination of technological and social approaches (including science, laws, new paradigms, norms, ethics, and Zeitgeist) that will be able to fix our problems**
- **Large central computational facilities will be needed to tackle the many large problems and smaller facilities can be thought of facilitating the solution of the many small problems, but overcoming the huge fragmentation is urgently required to scale up.**
- **With machine learning and other statistical methods we have a fantastic set of methods to tackle even complex problems, but we need to better understand what the classifiers are actually doing and how to turn results into understandable messages.**
- **Additionally, it became clear in this workshop that AI methods and a higher degree of automation is needed to enable new scientific and business methods, e.g., to automatically generate approximate hypotheses to drive understanding of multi-dimensional problems.**

At the IoT Week in Aarhus in 2019, we will again organise a workshop on Big Data in IoT and suggestions for the focus are welcome.