Report from Dagstuhl Perspectives Workshop 16152

Tensor Computing for Internet of Things

Edited by

Evrim Acar[1], Animashree Anandkumar[2], Lenore Mullin[3], Sebnem Rusitschka[4], and Volker Tresp[5]

1 **University of Copenhagen, DK,** evrim@life.ku.dk
2 **University of California – Irvine, US,** a.anandkumar@uci.edu
3 **University of Albany – SUNY, US,** lmullin@albany.edu
4 **Siemens AG – München, DE,** sebnem.rusitschka@siemens.com
5 **Siemens AG – München, DE,** volker.tresp@siemens.com

Abstract

This report documents the program and the outcomes of Dagstuhl Perspectives Workshop 16152 "Tensor Computing for Internet of Things". In an interactive three-day workshop industrial and academic researchers exchanged their multidisciplinary perspectives through impulse talks, panel discussions, and break-out sessions. Internet of Things (IoT) or Cyber-physical systems (CPS) bring out interesting new challenges to tensor computing, such as the need for real-time analytics and control in interconnected dynamic networks, e.g. electricity, transportation, manufacturing. On the other hand, IoT/CPS have characteristics that make tensor methods applicable to extract information very efficiently. During our discussions we identified an action plan to have a structured approach that will enable the multidisciplinary community of domain and control experts, data scientists, and distributed, embedded software developers to share knowledge and best practices, compare and exchange tensor models depending on data types and applications in distinct IoT/CPS scenarios.

Perspectives Workshop April 10–13, 2016 – http://www.dagstuhl.de/16152

1998 ACM Subject Classification D.4.7 Organization and Design, D.4.5 Reliability, I.2.6 Learning, I.1 Symbolic and Algebraic Manipulation, I.2.11 Distributed Artificial Intelligence

Keywords and phrases Tensor Methods, Multi-way Data Analysis, Multi-linear Algebra, Tensor Software, Distributed & Parallel Computing, Big Data Computing & Analytics, Cyber-physical Systems, Intelligent Autonomous Systems, Applications in Smart Grid, Mobility, Smart City

Digital Object Identifier 10.4230/DagRep.6.4.57

1 Executive Summary

Evrim Acar
Animashree Anandkumar
Lenore Mullin
Volker Tresp
Sebnem Rusitschka

In April 2016, Dagstuhl hosted a Perspectives Workshop on Tensor Computing for the Internet of Things. The prior year, industrial researchers had formulated the challenges of gaining insights from multi-dimensional sensory data coming from large-scale connected energy, transportation networks or manufacturing systems. The sheer amount of streaming multi-aspect data was prompting us to look for the most suitable techniques from the machine learning community: multi-way data analysis. Hence, we organized a three-day interactive workshop with two separate questions bringing two formerly distinct communities together: (i) How can we assure performance and reliability given the increasing complexity and data of an always-on connected world? (ii) Can we exploit the power of tensor algebra to solve high-dimensional large-scale machine learning problems that such a world poses?

The workshop focused on the Internet of Things (IoT), i.e. devices, which have the capability to sense, communicate, and more so, control their environments. These devices are increasingly becoming a part of complex, dynamic, and distributed systems of electricity or mobility networks, hence our daily lives. Various sensors enable these devices to capture multiple aspects of their surroundings in real-time. For example, phasor measurement units capture transient dynamics and evolving disturbances in the power system in high-resolution, in a synchronized manner, and in real-time. Another example is traffic networks, where a car today can deliver about 250 GB of data per hour from connected electronics such as weather sensors within the car, parking cameras and radars. Experts estimate that the IoT will consist of almost 50 billion objects by 2020 [2], which will trigger the Era of Exascale computing necessitating the management of heat and energy of computing in concert with more and more complex processor/network/memory hierarchies of sensors and embedded computers in distributed systems. Crucial for the extraction of relevant information is the format in which the raw data from such systems is represented. Crucial for viable efficiency of information extraction in IoT is which operations are used guaranteeing various attributes of resource use and management. Tensors can be viewed as data structures or as multilinear operators.

The goal of the workshop was to explore tensor representations and computing as the basis for machine learning solutions for the IoT. Tensors are algebraic objects which describe linear and multilinear relationships, and can be represented as multidimensional arrays. They often provide a natural and compact representation for multidimensional data. In the recent years, tensor and machine learning communities – mainly active in the data-rich domains such as neuroscience, social network analysis, chemometrics, knowledge graphs etc. – have provided a solid research infrastructure, reaching from the efficient routines for tensor calculus to methods of multi-way data analysis, i.e., tensor decompositions, to methods for consistent and efficient estimation of parameters of the probabilistic models.

Some tensor-based models have the intriguing characteristic that if there is a good match between the model and the underlying structure in the data, the models are much better interpretable than alternative techniques. Their interpretability is an essential feature for the machine learning techniques to gain acceptance in the rather engineering heavy fields of automation and control of cyber-physical systems. Many of these systems show intrinsically multilinear behavior, which is appropriately modeled by tensor methods and tools for controller design can use these models. The calibration of sensors delivering data and the higher resolution of measured data will have an additional impact on the interpretability of models.

Various presentations on tensor methods by established researchers from different application domains underscored that tensor methods are reaching a maturity tipping point. However, knowledge of usage characteristics of tensor models is scattered. Discussions of the currently independent perspectives on the usage of tensor methods showed convergence potential which we will detail in the Dagstuhl Manifesto. During our discussions based on the presentations of the IoT industrial researchers, it quickly became clear that we would need benchmark challenges for cyber-physical systems and benchmark data in order to be able to replicate the successes in machine learning for neuroscience, image processing or chemometrics, for example.

The tensor computing community will equally benefit from the new types of data, requirements and characteristics of IoT, which can lead to techniques that increase success rates of previous applications, as was the case with the challenges of social network data analysis leading to better tensor models/algorithms that can analyze data sets with missing entries, now used in many other fields in addition to social network analysis. Additionally, as opposed to standardized machine learning techniques, tensor computing currently lacks a common language and the homogeneity to flexibly exchange models. Hence, a hub platform bringing data and domain knowledge of cyber-physical systems together with a variety of practitioners of tensor computing would enhance increasing coherence of terms, best practices in data acquisition and structuring methods as well as model benchmarking, cataloging, and exchange of methods.

Furthermore, industrial researchers from IoT, automation and control domains highlighted their view that tensor computing methods are currently still inaccessible to the majority of the industrial practitioners even though there has been a considerable progress in developing tools for tensor computing. Matlab extensions to enable the use of tensor analysis are quite mature [1] [3] [4]. Matlab is widely used by control and automation practitioners. Python ecosystem for machine learning practitioners is very quickly adopting extensions for enabling tensor operations [5] [6]. However, both are mainly for prototyping and ultimately do not fulfill the need for a unified framework for industrial grade development and deployment of models in highly distributed cyber-physical systems. Interestingly, just five months prior to our workshop, Tensorflow [7], a numerical computation library aiming at capturing structures in multidimensional data as well as supporting both prototyping and production level algorithms was open sourced. Tensorflow can run on server clusters as well as embedded systems such as smart phones [8]. Another framework, unifying both batch and streaming data analysis, is Apache Spark [9]. Spark provides seamless scalability of software code to run on multiple machines. Recently there have been deployments of tensor methods on the Spark platform.

As a multidisciplinary community we believe that we will be able to formulate requirements and provide support in developing improvements for unifying frameworks. The required skill set is quite rare: we are in need of software developers that can create reliable high-performant code for both server-side distributed training on massive amounts of data and deployment of trained models in embedded distributed system. Heterogeneous processor architectures are predominant in cyber-physical systems. Either these software developers should be data scientists proficient in tensor computing and very good at communicating with domain experts or we need tooling such that data scientists and domain experts can collaboratively model data for cyber-physical systems. We will detail these discussions in the Manifesto: We believe that it is feasible to create such tooling that automates the generation of reliable, secure code, which accounts for the adaptive logic of devices interacting with their dynamic physical environment – but also through which there is a direct feedback between data scientist, domain or control expert, and the adaptive control device.

The Manifesto, which will be published on www.dagstuhl.de/16152/ will include a roadmap of how we as a newly formed multidisciplinary community want to start with a knowledge hub on tensors, and iterate through data grand challenges from IoT pilots, results dissemination, into what may one day become collaborative modeling hub for learning cyber-physical systems.

References

1 Tamara G. Kolda Brett W. Bader et al. Matlab tensor toolbox version 2.6. http://www.sandia.gov/~tgkolda/TensorToolbox/index-2.6.html

2 Statista. Internet of things (iot): number of connected devices worldwide from 2012 to 2020 (in billions). http://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

3 Ivan Oseledets. Tt (tensor train) toolbox version 2.2.2. https://github.com/oseledets/TT-Toolbox

4 Laurent Sorber Marc Van Barel Nico Vervliet, Otto Debals and Lieven De Lathauwer. Tensor lab user guide, release 3.0. http://www.tensorlab.net/userguide3.pdf

5 Ivan Oseledets. Python implementation of the tt-toolbox. https://github.com/oseledets/ttpy

6 Maximilian Nickel. scikit-tensor: Python library for multilinear algebra and tensor factorizations. https://github.com/mnick/scikit-tensor

7 Google. Tensorflow: An open source software library for numerical computation using data flow graphs. https://www.tensorflow.org/

8 Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek Gordon Murray, Benoit Steiner, Paul A. Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zhang. Tensorflow: A system for large-scale machine learning. *CoRR*, abs/1605.08695, 2016

9 Apache Spark. Apache spark: A fast and general engine for large-scale data processing. http://spark.apache.org/

2 Table of Contents

3 Overview of Talks

Impulse talks were grouped into two sessions: “IoT Applications and Computing Infrastructures” by industrial IoT/CPS researchers, who gave an overview and some deep dives into the IoT applications, multidimensional IoT data, what information we want to extract from it for what purpose, data sources and peculiarities, IoT/CPS computing infrastructures and peculiarities. “Challenges of Tensor Computing for IoT” consisted of impulse talks by academic researchers with vast experience in applying tensor decompositions, available tools, and emerging computing paradigms.

Impulse Talks I: IoT Applications and Computing Infrastructures

3.1 IoT and Applications such as Predictive Maintenance

Christine Preisach (SAP SE – Walldorf, DE)

In this presentation Predictive Maintenance, one major application in IoT, is described using an example use case and applied data mining methods. Moreover challenges encountered in the data used for Predictive Maintenance are stated.

3.2 Supervisory Control & Fault Diagnosis – Tensors in a Networking World

Gerwald Lichtenberg (HAW – Hamburg, DE)

In this impulse talk “Internet of Things” is approached from a systems and control theory perspective. Interesting challenges of IoT include highly connected control structures, in which now non-experts not only have access to sensors but also to actors in the system. Challenging scientific questions are briefly addressed such as performance and robustness in heterogeneous networks, finding optimal solutions for fault diagnosis and supervision, and deriving useful models with structural properties from continuous and discrete data [1]. Tensors could help in investigating these questions, by utilizing the inherent multilinearity of continuous and discrete dynamics, the model structure of networks and systems, among other characteristics [2].

References

1 Thorsten Müller, Kai Kruppa, Gerwald Lichtenberg, and Nicolas Réhault. Fault detection with qualitative models reduced by tensor decomposition methods. *IFAC-PapersOnLine*, 48(21):416–421, 2015

2 Georg Pangalos, Annika Eichler, and Gerwald Lichtenberg. Hybrid multilinear modeling and applications. In *Simulation and Modeling Methodologies, Technologies and Applications*, pages 71–85. Springer, 2015.

3.3 Heterogeneous Computing Infrastructures in Future Energy and Transportation Networks

Sebnem Rusitschka (Siemens AG – München, DE)

Our world becomes increasingly computerized, "we put little computers in our ears to hear better, we put our bodies into computers that drive," traditional sectors like energy, transportation, manufacturing increasingly rely on computerized automation [1].

The computing resources we'll find in these systems are heterogeneous, with highly differing capabilities and restrictions, some instructions will be time-critical, safety-critical. So the applications/algorithms, which will learn from the data, running in these heterogeneous computing environments will need end-to-end support to verifiably execute in the most optimal and safe way.

Can we facilitate this? How long will it take?

References

1 Sebnem Rusitschka and Edward Curry. Big data in the energy and transport sectors. In *New Horizons for a Data-Driven Economy*, pages 225–244. Springer, 2016

Impulse Talks II: Tensor Decompositions

3.4 Tensor analysis for handling huge amounts of -omics data

Rasmus Bro (University of Copenhagen, DK)

In this talk I will show how tensor methods are crucial in understanding the bigger and bigger datasets obtained on health, environmental and food research. With tensors methods we are able to convert huge and difficult chemical measurements into smaller data sets that are easier to handle, present the information in a more 'correct' manner and helps eliminating spurious findings which often arise from mining huge unfiltered data [1].

References

1 Age Smilde, Rasmus Bro, and Paul Geladi. Multi-way Analysis: Applications in the Chemical Sciences, West Sussex: Wiley, 2004.

3.5 Tensor Applications in Neuroscience

Morten Mørup (Technical University of Denmark – Lyngby, DK)

In this talk a range of applications of tensor decomposition in neuroscience will be outlined in particular for the modeling of electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) data. These spatio-temporal data of brain activity naturally form tensors when measured across multiple subjects, trials or conditions. Modeling approaches based on the CP (Canonical Decomposition [2]/Parallel Factor Analysis [3]) decomposition, shifted CP and convolutive CP decomposition will be discussed as well as approaches for the modeling of functional segregation and integration using tensor models with TUCKER2 and PARAFAC2 structure. The talk will finally outline potential areas of neuroscience where tensor modeling may become relevant as well as key challenges modeling neuroimaging data using tensor decomposition [1].

References

1 Morten Mørup. Applications of tensor (multiway array) factorizations and decompositions in data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):24–40, 2011

2 J. Douglas Carroll and J.-J. Chang. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35:283–319, 1970.

3 Richard A. Harshman. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA working papers in phonetics*, 16:1–84, 1970.

3.6 Data Fusion using Coupled Matrix and Tensor Factorizations

Evrim Acar (University of Copenhagen, DK)

Data fusion, i.e., extracting knowledge through the fusion of complementary data sets, is a topic of interest in many fields. For instance, in metabolomics, analytical platforms such as Liquid Chromatography-Mass Spectrometry and Nuclear Magnetic Resonance spectroscopy are used for chemical profiling of biological samples. Measurements from different platforms are capable of detecting different types of chemical compounds with different levels of sensitivity. Jointly analyzing those measurements can provide more accurate characterization and understanding of a physiological/pathological condition. Fusing data from multiple sources has proved useful in various disciplines including metabolomics, neuroscience, social network analysis and signal processing. However, data fusion remains a challenging task since there is a lack of data mining tools that can jointly analyze imperfect (i.e., with missing entries) heterogeneous (i.e., in the form of higher-order tensors and matrices) data sets, and capture underlying shared/unshared structures. In this talk, we discuss the formulation of data fusion as a coupled factorization problem. We give a brief overview of data fusion models based on coupled matrix and tensor factorizations (CMTF), demonstrate the use of various CMTF models with applications from different disciplines and discuss open research problems [1].

References

1 E. Acar, R. Bro, and A. K. Smilde. Data fusion in Metabolomics using Coupled Matrix and Tensor Factorizations. Proceedings of the IEEE, 103: 1603-1620, 2015.

3.7 Tensors for Representational Learning

Volker Tresp (Siemens AG – München, DE)

Embedding learning, a.k.a. representation learning, has been shown to be able to model large-scale semantic knowledge graphs. A key concept is a mapping of the knowledge graph to a tensor representation whose entries are predicted by models using latent representations of generalized entities. Latent variable models are well suited to deal with the high dimensionality and sparsity of typical knowledge graphs [1]. In recent publications the embedding models were extended to also consider temporal evolutions, temporal patterns and subsymbolic representations. We map embedding models, which were developed purely as solutions to technical problems for modelling temporal knowledge graphs, to various cognitive memory functions, in particular to semantic and concept memory, episodic memory, sensory memory, short-term memory, and working memory.

References

1 Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. *arXiv preprint arXiv:1503.00759*, 2015

Impulse Talks II: Tensor Computing and Beyond

3.8 Advances in the numerical computation of (coupled) tensor decompositions

Lieven De Lathauwer (KU Leuven, BE)

We give a short overview of recent advances in the numerical computation of tensor decompositions and coupled matrix/tensor factorizations. We pay special attention to large-scale aspects. We also highlight some features of Tensorlab, of which version 3.0 has been released in March [1].

References

1 Laurent Sorber Marc Van Barel Nico Vervliet, Otto Debals and Lieven De Lathauwer. Tensor lab user guide, release 3.0. http://www.tensorlab.net/userguide3.pdf

3.9 Tensors and IoT Computational Issues

Lenore Mullin (University of Albany – SUNY, US)

The high-performance computing (HPC) community widely recognizes that the current LINPACK-based benchmark used to estabish the TOP500 supercomputer ranking now focuses on the wrong things. It measures near-peak floating-point speed but tolerates relatively low communication performance, but what is now needed to predict application performance is just the opposite: we need to measure communication performance at all levels of the memory hierarchy, and assume the floating-point hardware is overprovisioned. The solution may reside in the observation that the kernel of LINPACK, matrix-matrix multiplication, is actually a special case of n-dimensional tensor operations. The broader category includes both the Fast Fourier Transform (FFT) and matrix-matrix multiplication that has been optimized by hierarchical blocking to match memory levels [1]. We first unify the generalized task of n-dimensional tensor operations and then show that we can simply dial a different point of the spectrum of such workloads to restore predictive value to the TOP500 ranking on real applications.

References

1 Lenore Mullin and James Raynolds. Scalable, portable, verifiable kronecker products on multi-scale computers. In *Constraint Programming and Decision Making*, pages 111–129. Springer, 2014

3.10 Guaranteed Learning Using Tensor Methods

Animashree Anandkumar (University of California – Irvine, US)

Unsupervised learning is the challenging problem of making automated discoveries without external supervision. It requires fitting unlabeled data to large-scale latent variable models. Traditional learning approaches such as expectation maximization or variational inference are slow to converge and get stuck in local optima due to non-convexity of the likelihood function. In contrast, we have developed a method of moments approach, based on decomposition of low order moment tensors, which is guaranteed to learn the correct model under mild conditions with (low order) polynomial sample and computational complexity [1].

In practice, tensor methods significantly outperform previous learning approaches, both in training time and model fitting, on a wide range of problems such as document categorization, social network analysis, discovering neuronal cell types, and learning sentence embeddings. Further, we have established that tensor methods are guaranteed to find the globally optimal solution to other challenging non-convex problems such as training multi-layer neural networks and reinforcement learning of partially observable Markov decision processes. These positive results demonstrate that many learning tasks, previously considered intractable, can be solved efficiently under mild and transparent conditions.

References

1 Majid Janzamin, Hanie Sedghi, and Anima Anandkumar. Beating the perils of non-convexity: Guaranteed training of neural networks using tensor methods. *CoRR abs/1506.08473*, 2015

3.11 Tensor techniques for the visualization of multidimensional data

Renato Pajarola (Universität Zürich, CH)

We will present the application of tensor approximation methods in the context of interactive visualization of large 3D volume data as well as the processing of visual data in the compressed domain [1].

References

1 Renato Pajarola, Susanne K Suter, and Roland Ruiters. Tensor approximation in visualization and computer graphics. 2013

4 Panel discussions

Panel discussions provided us with an additional method of knowledge and experience sharing and exchange of our different perspectives. We distributed panel discussions in two parts over the course of the workshop. In “Part 1: Challenges of Tensor Representations for IoT Data” industrial researchers shared their experience with real-world multidimensional sensor data. In “Part 2: Challenges of Tensor Computing for IoT” panelists shared their view on computational and process challenges with respect to wide spread applicability of tensors in IoT/CPS.

Part I: Challenges of Tensor Representations for IoT data

4.1 Distributed PARAFAC for Large Industrial Internet-scale Dense Tensors with Skewed Modes

Kareem Aggour (General Electric – Niskayuna, US)

Kareem’s work is driven by the need to analyze large Industrial Internet-scale datasets such as those generated by equipment sensors over time. GE is OEM of turbines among other industrial equipment. Typical analysis consists of data from 10-100 turbines, with 60-100 sensors, which deliver per second data. Sensor data is typically noisy. Peculiarities regarding equipment data is that they are both sparse and dense, and very skewed over the time mode. Multimodal concepts, and tensor decomposition lends itself to naturally find relations in this data. Kareem modified the classical PARAFAC algorithm to run in distributed computing environments to decompose large, dense, and skewed 3-way datasets. Distributed PARAFAC is implemented and extensively evaluated on two computing platforms HPC vs. Hadoop. Both scale well with large tensors.

It was mentioned in the audience that there are a few other tools like Splat that focus on sparse and dense data which also enable distributed scaling. Apache Spark also has an implementation for dense tensors.

4.2 Sensors in Power Networks – Representing Network Dynamics over Space and Time

Denis Krompaß (Siemens AG – München, DE)

Data quality is always challenging, especially in large-scale distributed CPS. There is a wide need for algorithms to handle the preprocessing of data. Use of tensor factorization as feature extractor on sensor data, is promising. A similar challenge is labeling of data. There is not much labeled sensor data – and no easy way to do so reliably – especially not without the constant involvement of a domain expert. Small questions like "how do you normalize the data – per device or per measurement type?" turn out to be bigger design issues. Important issue is that it is not clear to a non-expert where to pay attention in the streams of sensor data. Effective labeling of data is a dialog between domain expert and the data scientist.

Comments from the audience was that latent variables can be exposed in a unique way through tensor decompositions. CP is good for decomposition, good for learning representations, but has problems with numerical analysis. Maybe considering subspaces, as in Tucker, and working with residuals would give a decomposition similar to CP but without the problem of numerical analysis. Additionally, if the decomposed tensor respects the physical model of the system, e.g. power network, that would expose the subspace to focus attention on. Another way could be to guide experts to attach domain knowledge to the data, e.g. which are the relevant/irrelevant areas. Such domain checks could be realized as feedback.

4.3 Loose Semantic Coupling in IoT and the Role of Tensors

Edward Curry for Souleiman Hasan (National University of Ireland – Galway, IE)

Decoupling is a main principle for scalability. Can tensor models provide a principled data representation that is easy to exchange in a decoupled IoT environment? Smart City scenarios deal with data from a lot of distributed sensors, and applications therein are mainly concerned with event processing [1]. Hence, Souleiman and Edward investigated computational semantics by representing the data in vector space and analyzing collocation. The main challenge is how do you find the relevant data on the fly?

References

1 Souleiman Hasan and Edward Curry. Thingsonomy: Tackling variety in internet of things events. *IEEE Internet Computing*, 19(2):10–18, 2015

Part II: Challenges of Tensor Computing for IoT

4.4 Key Aspects in Tensor Computing in the Context of IoT

Benoit Meister (Reservoir Labs, Inc. – New York, US)

Benoit's working environment mainly applies tensors in cybersecurity applications. The models are shipped with appliances for handling massive throughput, e.g. high-end 100 Gbit/s or 20 Gbit/s. Solutions also include tools that allow the user to select areas for analysis. Security experts use visualization components of the tensor analysis tool. Modeling is done jointly with own domain experts and external security experts. CP is a preferred model because it is easier to communicate, clarify questions for research when structure is interpretable.

The application in IoT scenarios will bring interesting research challenges along, such as faster execution of tensor methods, streaming or low-rank updates for real-time analysis. Efficient distributed data and computation models are required for handling tensor computing in heterogeneous processor architectures with multi-core and embedded systems, operating at low power. Essentially algorithms need a power efficient design. Applying compressed sensing techniques on multidimensional sensor data to reduce computation costs, data movement, and eventually power utilization are promising areas [1].

References

1 Nicholas P Carter, Aditya Agrawal, Shekhar Borkar, Romain Cledat, Howard David, Dave Dunning, Joshua Fryman, Ivan Ganev, Roger A Golliver, Rob Knauerhase, et al. Runnemede: An architecture for ubiquitous high-performance computing. In *High Performance Computer Architecture (HPCA2013), 2013 IEEE 19th International Symposium on*, pages 198–209. IEEE, 2013

4.5 Modeling of Complex, Man-made Systems

Bülent Yener (Rensselaer Polytechnic Institute – Troy, US)

Interdisciplinarity is a main reason for huge, repetitive communication overhead in CPS projects. Tensors must be a part of a bigger computational machinery supporting the domain experts. It would be great if we eventually can qualify "it depends on the data", e.g. through recipes, best practices. For example in IoT/CPS data is always analyzed over time and space. In IoT, prediction (trending) is very important; e.g. anomalies in massive streams of IP traffic data coming from interconnected routers.

4.6 Future Research Directions on Tensor Computing for IoT

Vagelis Papalexakis (Carnegie Mellon University, US)

Tensor decompositions are very versatile and powerful tools, ubiquitous in data mining applications. They have been successfully integrating in a rich variety of real-world applications, and due to the fact that they can express and exploit higher order relations in the data, they tend to outperform approaches that ignore such structure. As a result, tensor modeling and analysis in Internet of Things applications, where the data are inherently multi-aspect and diverse, has a great potential [1].

The success that tensors have experienced in data mining during the last few years by no means indicates that all challenges and open problems have been addressed. Quite to the contrary, there exist challenges, such as scaling up to bigger data, modeling space and time, unsupervised model selection, or connections with heterogeneous information networks.

References

1. Evangelos E Papalexakis, Christos Faloutsos, and Nicholas D Sidiropoulos. Parcube: Sparse parallelizable tensor decompositions. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 521–536. Springer, 2012

5 Working groups

Break-out Sessions I: Value Proposition of Our Multidisciplinary Research

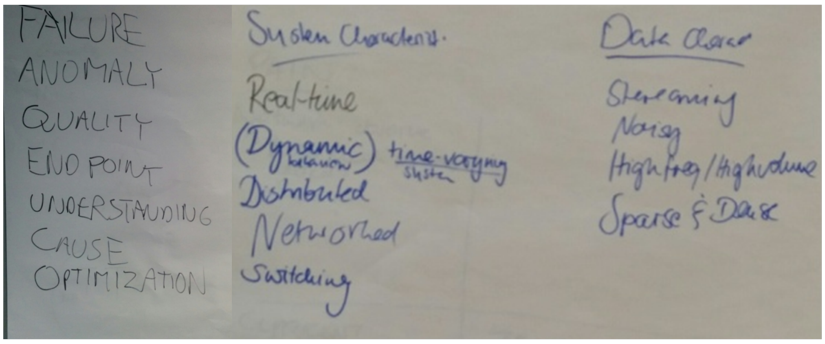
In the afternoon of the second day, after having shared our diverse perspectives, we took a "map" of our skills and research interest. Especially we were anxious to see how we are going to move from IoT applications of tensor decompositions to running software in decentralizing cyber-physical systems. Below is a snapshot:

Participants and their area of interest for determining break-out sessions.

Based on this snapshot, we divided into two break-out sessions: Applications and Frameworks.

5.1 Applications of Tensor Decomposition in IoT Scenarios

We discussed typical scenarios in CPS and IoT from domains such as energy, automation, transportation, etc., where operators/users are interested in taking actions based on the inferred knowledge about faults, anomalies or general events in the system. Especially operators, e.g. in traffic management, or grid management etc. are interested in understanding root cause and ability to optimize system over longer periods as well as in real-time.



IoT/CPS applications, system and data characteristics.

The systems in these domains are real-time, show dynamic behavior, are time-varying, distributed and networked. Usage and operations of these systems include switching, which requires both continuous and discrete variables in representations. The data coming from the sensors embedded into these CPS/IoT systems is streaming, noisy, both high-frequency and high-volume, both sparse and dense.

Many information of interest is conveyed from multiple aspects through different sensors capturing various characteristics of the system. In automation scenarios, when the control loop is closed, system modeling becomes too complicated to model. State space models, such as Kalman filters, are a very natural representation of such closed dynamic systems [1]. State space models are essentially tensorized representations of the physical system.

Novel techniques via tensors, which can handle both inference and modeling, could be applicable for increased interpretability. The above described application, system, data characteristics typically call for a multivariate statistical model of the multi-aspect streaming data. Computational methods in the language of tensors enables the design of provable algorithms, which are otherwise harder to get.

There are also multiple dimensions of challenges. Connectivity for example requires richer tensor models, e.g. hierarchical models. Time dimension requires dynamic processing, e.g. of incoming slices; or active learning of entries, by algorithms that interactively query the user or other data sources to obtain labels. Another approach is the explicit modeling of time or system evolution. The natural structure of tensors to be used both for representation and computationally does simplify the information extraction process. Later arriving data can be updated. For time and connectivity interaction, i.e. change over time that affects connectivity, the whole model can be updated, e.g. via coupled factorization.

On the other hand, the knowledge about the network structure from CPS can be fed into the model as apriori knowledge. Tensor decompositions are more efficient and accurate in representing graphs. An open research question is whether such IoT network graphs are representable through tensor graph networks, if the structure changes dynamically. The connectivity structure of IoT/CPS can help to establish associations between the various modalities capturing data about the system. Tensors can be used to combat curse of dimensionality: tensor methods excel when associations exist in the data. When a few signals dominate, then low rank, i.e. low parametric, representation becomes possible. De-noising of data could also benefit from exploiting the connectivity structure of IoT/CPS for tensor decompositions.

Today in large CPS such as power transmission systems or transportation, logistics systems etc. an important question is how many sensors are needed and where they should be placed. Optimal sensor placement studies need to become dynamic as the underlying systems get more dynamic [2]. Tensors, on the other hand, can uncover hidden structures in sparse data or from compressive resampled data. Monitoring aging devices with sampling only from time to time can already show a trend. Tensor representations are very efficient for such few parameters. Open research questions again arise from the dynamic nature of IoT/CPS in which dynamic states require that the models need to derive state from input/output data.

Many interesting questions arouse when we discussed the need to justify models. Model justification is crucial in CPS and IoT such that system users understand algorithmic inference. The coupling with semantics and links to Multi-agent Systems are a few of the interesting topics to investigate. Modeling of actions in these dynamic systems, especially actions taken based on the intelligence through the algorithmic inference is an open question.

Even in the face of evidence or interpretability of a model, trust into algorithmic inference is a big issue. Domain experts oftentimes underestimate how complex their systems are or will become through increasing digitalization and connectivity. Overreliance on simulations and rule-based systems give a false sense of complexity management. Tensor decompositions are not only interpretable but also well visualizable. New technology for visualization combined with effective visualization methods are open research directions towards creating better understandability of the models and creating trust.

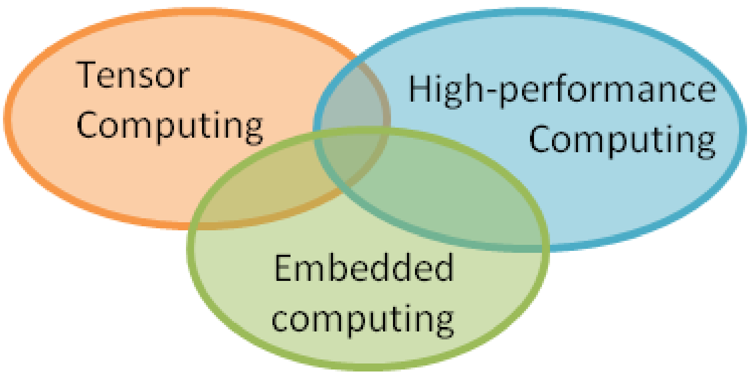
References

1 Mohammad Niknazar, Hanna Becker, Bertrand Rivet, Christian Jutten, and Pierre Comon. Robust 3-way tensor decomposition and extended state kalman filtering to extract fetal ecg. In *21st European Signal Processing Conference (EUSIPCO 2013)*, pages 1–5. IEEE, 2013

2 Farrokh Aminifar, Mahmud Fotuhi-Firuzabad, and Amir Safdarian. Optimal pmu placement based on probabilistic cost/benefit analysis. *IEEE Transactions on Power Systems*, 28(1):566–567, 2013

5.2 Frameworks for Tensor Computing in IoT

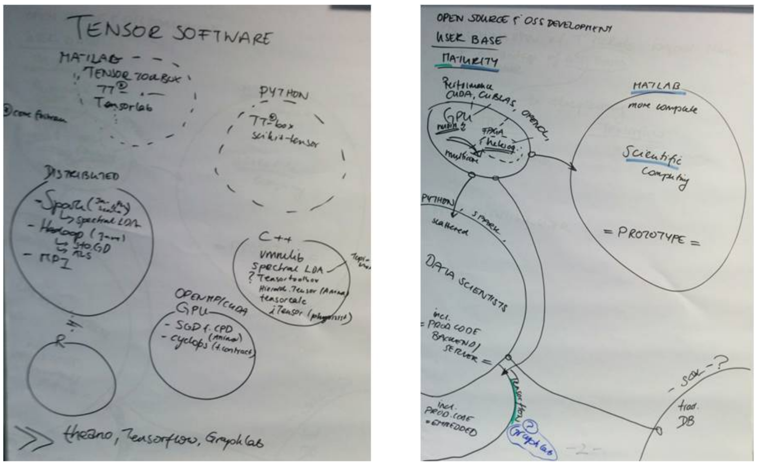
IoT applications will not only require a close collaboration between data scientists, control experts, and domain experts for the application of tensor computing, but also software developers who are proficient in embedded systems, distributed systems, and high performant code.



Rare overlap of skill set requires supporting SW frameworks.

Such a niche overlap as depicted above calls for frameworks abstracting away the complexities of each area to a degree that ultimately IoT application developers should be able to utilize these frameworks with a tolerable learning curve.

In this break-out session we started off by a depiction of currently available tensor software, libraries, and frameworks that are enabling tensor computing. Open source software (OSS) and OSS development is acknowledged as a driving force for increasing user bases and accelerated maturity. At the same time, there are currently two major user groups, data scientists and scientific community/control experts who mostly prefer Python or Matlab respectively. Matlab has a quite mature libraries, such as Tensor Toolbox [1], Tensor Train Toolbox [3], or TensorLab [4]. Python has the NumPy package for scientific computing which has an N-dimensional array object. However, tensor additions such as PyTensor or scikit-tensor [5] have not yet gained community support.

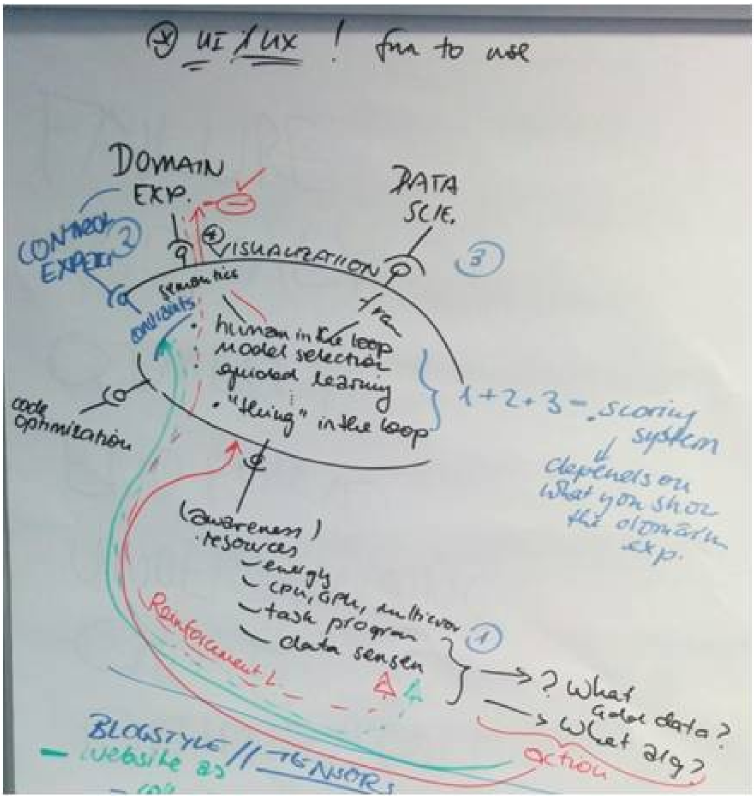


Available SW frameworks and applicability in productive IoT/CPS.

When it comes to production level code for use in IoT applications on massive amounts of streaming data, current frameworks, e.g. Apache Spark [8] or Tensorflow [6] are rather geared towards data scientists with Python as a front end language. Although Spark has SQL interface for traditional DBs [2], the group discussion concluded that traditional relational DBs are less likely to be used in information extraction from IoT. Tensorflow, additionally, enables seamless deployment of production code not only on distributed server clusters but also down to embedded devices.

One important aspect is that currently all these frameworks depend on the same libraries for performant execution of linear algebraic routines, e.g. BLAS, in vendor-specific or -neutral frameworks such as in CUDA or OpenCL, etc. CUDA fine-tunes operations on GPUs of NVIDIA, whereas OpenCL programs execute across heterogeneous platforms including CPUs, GPUs, DSPs, FPGAs or hardware accelerators.

During the course of the discussions we identified that such libraries can be enhanced by taking advantage of tensor algebra. Algebraic analysis can assist in statically solving the optimal partitioning of computations across available processor architectures in an IoT application. Additionally, the Tensorflow framework currently supports deployment on embedded devices [7], but not distributed embedded application, scoring, or update of a trained model. Currently distributed learning is being applied and improved for cluster computing to reduce training times. However, there is no research into distributed embedded application, e.g. retraining, of models.



Potential outline of a Tensors for IoT Framework.

References

1 Tamara G. Kolda Brett W. Bader et al. Matlab tensor toolbox version 2.6. http://www.sandia.gov/~tgkolda/TensorToolbox/index-2.6.html

2 Michael Armbrust, Reynold S Xin, Cheng Lian, Yin Huai, Davies Liu, Joseph K Bradley, Xiangrui Meng, Tomer Kaftan, Michael J Franklin, Ali Ghodsi, et al. Spark sql: Relational data processing in spark. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1383–1394. ACM, 2015

3 Ivan Oseledets. Tt (tensor train) toolbox version 2.2.2. https://github.com/oseledets/TT-Toolbox

4 Laurent Sorber Marc Van Barel Nico Vervliet, Otto Debals and Lieven De Lathauwer. Tensor lab user guide, release 3.0. http://www.tensorlab.net/userguide3.pdf

5 Maximilian Nickel. scikit-tensor: Python library for multilinear algebra and tensor factorizations. https://github.com/mnick/scikit-tensor

6 Google. Tensorflow: An open source software library for numerical computation using data flow graphs. https://www.tensorflow.org/

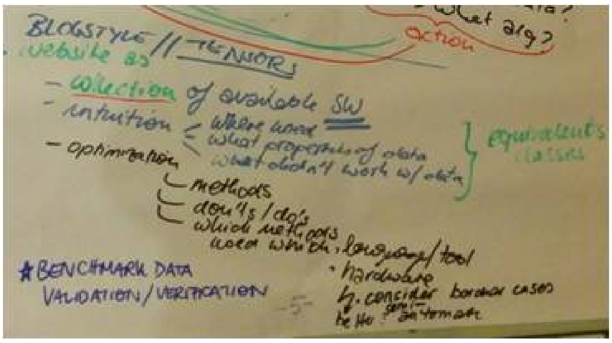
7 Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek Gordon Murray, Benoit Steiner, Paul A. Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zhang. Tensorflow: A system for large-scale machine learning. *CoRR*, abs/1605.08695, 2016

8 Apache Spark. Apache spark: A fast and general engine for large-scale data processing. http://spark.apache.org/

Break-out Sessions II: Call to Action

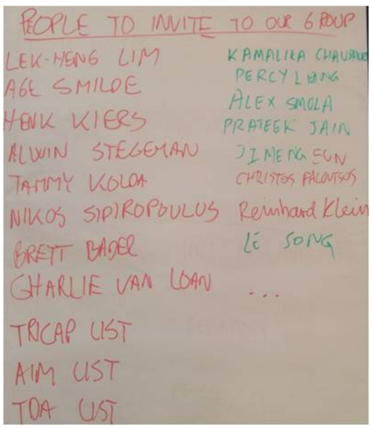
In this break-out session the group concentrated on establishing concrete steps to remove the identified inhibitors towards a multidisciplinary research on tensor decomposition applications in cyber-physical systems:

1. Improve accessibility of tensor methods and knowledge in the CPS/IoT domains:
   - Tensorize Wikipedia – Anima, Vagelis
   - IEEE Proceedings Special Issue for Manifesto – Volker
   - IEEE IoT Paper on Why Tensors – Edward
   - Paper on Multidisciplinary Perspective – Anima, Bülent, Vagelis
   - Paper on Common Notation/Language with 4–10 authors – Vagelis, Rasmus, Taylan, Lieven
2. Create IoT benchmark data repository for a data grand challenge – Sebnem, Edward
   - "30 days of traffic" city benchmark – Edward
   - Satellite/areal imagery – Renato
   - Open smart cities – Edward
   - Seismic data – Bülent, Taylan
3. Connect with tensor operation implementers (in Python, C++,…) – Ivan, Benoit, Sebnem
4. Create a website to link above efforts: community building incl. mailing list, wiki, repository, for defining common problems, data type etc. – Rasmus, Renato



Rough sketch of contents of the website.

5. Invite further experts and practitioners to the community – all



(Incomplete) list of experts to invite to the community.

Additionally, some concrete research interests were formulated based on synergies identified during this workshop

1. Incomplete CP for streaming – Evrim, Vagelis
2. Data Fusion based on moments – Evrim, Anima
3. "Teddy" Tensor decomposition dynamics – Lenore, Sebnem

6 Open Problems

Cyber-physical systems (CPS) or Internet of Things (IoT) describe the now visible trend of increased digitalization and interconnectivity that bridge the physical system and its virtual representation through data from multi-modal sensors embedded into such systems. Network characteristics and dynamics are manifested in the data. Tensor methods are considerably suitable for uncovering such hidden structures, latent variables, in the data if such associations exist. However, there exist challenges, such as scaling up to bigger data, modeling space and time, unsupervised model selection, or connections with heterogeneous information networks.

Model justification is crucial in CPS and IoT such that system users understand algorithmic inference. Modeling of actions in these dynamic systems, especially of how to model the actions taken based on the intelligence gained through the algorithmic inference is an open question. For a better understanding by users, it is of advantage that tensor decompositions are well visualizable. New technology for visualization combined with effective visualization methods are valuable research directions to create traction.

Regarding performance and power efficiency of models and algorithms, we identified that current libraries can be enhanced by taking full advantage of tensor algebra. Algebraic analysis can assist in statically solving the optimal partitioning of computations across available processor architectures in an IoT application. Current frameworks support distributed deployment of tensor computations across a cluster of servers and down to embedded devices. What is missing is distributed embedded application, scoring, or update of a trained model in a streaming real-time manner, which we deem necessary in IoT/CPS applications.

IoT/CPS are dynamic networks. Time and connectivity interactions, i.e. change over time that affects connectivity, require that the whole model can be updated online. One interesting aspect is that the knowledge about network structure from CPS can be fed into the model as apriori knowledge. Tensor decompositions are more efficient and accurate in representing graphs. An open research question is whether network graphs are representable through tensor graph networks, if the structure changes dynamically. Update mechanism and online tensor decompositions need to be researched in order to handle dynamic data. Whether connectivity associations can be taken advantage of in tensor decompositions for better de-noising of data is yet another open question.

7 Outlook

The Dagstuhl Perspectives Workshop was a very suitable format to bring in the multidisciplinary community to exchange knowledge, discuss synergies, and identify potential inhibitors of applications of tensor methods in IoT/CPS. From the many discussions we had, the concrete next steps we identified, one aspect stands out: We want to establish a better foundation for Tensor Computing in IoT through a knowledge hub website. Such a website has the potential to bring together domain experts and data scientists to exchange best practices. One important pillar of the knowledge hub will be the benchmark IoT data repository that this group wants to harvest.

We believe that IoT/CPS data grand challenges can be an appropriate format to bring domain and control experts together with data scientists effective in applying tensor methods. Over time the community will be able to catalogue and efficiently compare and exchange models, data types, and application types, which we started discussing in this workshop. The hub has the potential to move from knowledge exchange to guided preprocessing of IoT data and model selection. Further more, algorithms embedded in the CPS/IoT could utilize such a hub for active learning and reinforcement learning given sufficient quality content and appropriate interfaces.

In the meantime, we are committed to work on dissemination and seek academic and industrial feedback. Aforementioned activities such as creating and maintaining the knowledge hub, defining IoT data grand challenges to move tensor computing research towards industrial applied research in CPS/IoT will require national and international funding. The potential outcome of more efficient algorithms and better interpretable models for data-driven automation and control of large-scale digitalized systems are of utmost importance to users as well as operators of future electricity, transportation, and manufacturing systems.

Participants

- Evrim Acar
University of Copenhagen, DK
- Kareem Aggour
General Electric – Niskayuna, US
- Animashree Anandkumar
Univ. of California – Irvine, US
- Rasmus Bro
University of Copenhagen, DK
- Ali Taylan Cemgil
Bogaziçi Univ. – Istanbul, TR
- Edward Curry
National University of Ireland – Galway, IE
- Lieven De Lathauwer
KU Leuven, BE
- Hans Hagen
TU Kaiserslautern, DE
- Souleiman Hasan
National University of Ireland – Galway, IE
- Denis Krompaß
Siemens AG – München, DE
- Gerwald Lichtenberg
HAW – Hamburg, DE
- Benoit Meister
Reservoir Labs, Inc. – New York, US
- Lenore Mullin
Univ. of Albany – SUNY, US
- Morten Mørup
Technical Univ. of Denmark – Lyngby, DK
- Axel-Cyrille Ngonga-Ngomo
Universität Leipzig, DE
- Ivan Oseledets
Skoltech – Scolkovo, RU
- Renato Pajarola
Universität Zürich, CH
- Vagelis Papalexakis
Carnegie Mellon University, US
- Christine Preisach
SAP SE – Walldorf, DE
- Achim Rettinger
KIT – Karlsruher Institut für Technologie, DE
- Sebnem Rusitschka
Siemens AG – München, DE
- Volker Tresp
Siemens AG – München, DE
- Bülent Yener
Rensselaer Polytechnic Institute – Troy, US