Theodoros (Theodore) Vasiloudis

CONTACT INFORMATION

SICS Swedish ICT AB
Decisions, Networks and Analytics Lab
Swedish Institute of Computer Science
Royal Institute of Technology, KTH

Phone: +46 700477423

tvas@kth.se

RESEARCH SUMMARY

My research deals with large scale learning for graphs and decision trees, including the development of novel algorithms and contributions to data analytics frameworks.
The projects that I have been involved in included developing algorithms that make use of streaming frameworks such as Apache Flink and Apache Samoa, and contributing to the design and implementation of learning systems capable of handling massive datasets, such as the machine learning library for Apache Flink. As a committer for Apache Flink, I am responsible for coordinating the machine learning efforts for the project.

A research area that I am currently exploring is decision trees, and more specifically calculating the uncertainty in the predictions of random forests, with the aim of bringing the techniques to online learning. This work is currently under submission.

EDUCATION

Royal Institute of Technology, Stockholm, Sweden

Ph.D. Candidate **November 2014 – Currently**

- Academic Supervisor: Professor Anders Holst
- Academic Co-supervisors: Professor Henrik Boström, Professor Seif Haridi
- Industrial Supervisor: Daniel Gillblad

Royal Institute of Technology, Stockholm, Sweden

M.Sc. in Machine Learning **August 2012 – July 2014**

- Thesis: Extending recommendation algorithms by modelling user context.
- Performed at: Spotify AB
- Supervisor: Associate Professor Hedvig Kjellström
- Industrial Supervisor: Boxun Zhang

Aristotle University of Thessaloniki, Thessaloniki, Greece

B.Sc. in Computer Science **September 2004 – March 2011**

- Thesis: Implementation of two algorithms for the classification of multi-label data using Bayesian networks.
- Supervisor: Assistant Professor Grigorios Tsoumakas

PUBLICATIONS

Theodore Vasiloudis, Foteini Beligianni and Gianmarco De Francisci Morales. *BoostVHT: Boosting Distributed Streaming Decision Trees.*
ACM International on Conference on Information and Knowledge Management (CIKM) 2017, Singapore.

Theodore Vasiloudis, Puya Vahabi, Ross Kravitz, Valery Rashkov. *Predicting Session Length in Media Streaming.*
ACM International Conference on Research and Development in Information Retrieval (SIGIR) 2017, Tokyo, Japan.

Olof Görnerup, Daniel Gillblad, **Theodore Vasiloudis**. *Domain-Agnostic Discovery of Similarities and Concepts at Scale.*
Knowledge and Information Systems, Springer (2016).

Olof Görnerup, Daniel Gillblad, **Theodore Vasiloudis**. *Knowing an Object by the Company It Keeps: A Domain-Agnostic Scheme for Similarity Discovery.*
IEEE International Conference on Data Mining (ICDM) 2015, Atlantic City, USA.

Professional Experience

Swedish Institute of Computer Science, Stockholm, Sweden

Researcher **August 2014 – Currently**

I am currently employed as a researcher at the Swedish Institute of Computer Science (SICS) where I also have the opportunity to pursue a PhD title from the Royal Institute of Technology, KTH. My work as a researcher is aligned to my research interests in large scale machine learning algorithms and applications.

Amazon, Seattle, WA, USA

Scientist intern **June 2017 – September 2017**

During the summer of 2017 I worked as an applied scientist intern at Amazon with the Alexa Voice Shopping team. My project was aimed at aiding the modeling efforts across the organization by generating realistic queries using Generative Adversarial Networks (GAN) and Recurrent Neural Networks (RNN). Our results showed that text-oriented GANs are not yet able to match the performance of RNNs in query generation.

Pandora Media, Oakland, CA, USA

Scientist intern **June 2016 – August 2016**

During the summer of 2016 I was a scientist intern at Pandora, working on user modelling and more specifically predicting the length of user sessions on the service. Utilizing techniques from survival analysis, mixed effects modeling and gradient boosted trees we were able to significantly outperform previous approaches and produced new insights on users' use of the service.

Data Artisans, Berlin, Germany

Software Engineering Internship **April 2015 – July 2015**

During my time at Data Artisans I contributed to the design and implementation of the distributed Machine Learning library for Apache Flink. This included work on algorithms and tooling, such as developing a convex optimization framework, aiding in the design of machine learning pipelines, and developing an evaluation and cross-validation framework for machine learning models. I also assisted in expanding the library by reviewing and advising commiters and students from Google Summer of Code and TU Berlin.

Spotify, Stockholm, Sweden

Master Thesis **January 2014 – July 2014**

I performed my thesis at Spotify AB, where I studied the effects of including user context in the recommendation pipeline. We explored the effects of adding contextual information like the day of the week into a traditional recommender system, and showed improvements in performance vs. baselines like item-to-item similarity.

Github Profile

http://github.com/thvasilo

Talks

Uncovering Concepts in Graphs *Spotify NYC, Nov. 2015*
FilnkML: Machine Learning for Apache Flink, *FOSDEM HPC & Big Data Devroom, Feb. 2015*
A Field Fuide to the Machine Learning Zoo, *FOSDEM HPC & Big Data Devroom, Feb 2016*
Machine Learning in Production *Linköping University, Mar. 2017*

Service

SIGIR 2017, *Reviewer, Short paper track*
Workshop on Data Management for End-to-Machine Learning @ SIGMOD 2017, *Reviewer*
EuroPar 2018, *Reviewer*
Workshop on Data Management for End-to-Machine Learning @ SIGMOD 2018, *PC*