4th International Conference on Predictive Applications and APIs

Volume 82:

PAPIs 2017

Claire Hardgrove, Louis Dorard and Keiran Thompson

Preface

We introduce the proceedings of advanced talks presented at the 4th International Conference on Predictive Applications and APIs (PAPIs '17), in Boston, United States on October 24-25, 2017.

The conference featured 25 talks discussing various aspects of the integration of Machine Learning in real-world applications, processes and businesses. Speakers presented techniques, tools and lessons learned to an audience of researchers and industry practitioners with a wide range of experience levels. These proceedings contain papers from five of the talks that presented the most advanced and novel content.

The first paper demonstrates how to generate large training datasets by extreme augmentation of a small collection of coloured cel-style cartoon images, without overfitting, and with applications to automatic colouring of cartoon images in art, design and animation.

The second paper introduces MMLSpark, an open source library that combines the deep learning library Microsoft Cognitive Toolkit (CNTK) with Apache Spark, adding several memory and performance enhancing optimizations. The paper introduces java bindings for CNTK which enable deep learning on a large class of new devices running on the Java Virtual Machine. These bindings enable a SparkML transformer that distributes, loads, and executes arbitrary CNTK computation graphs inside SparkML pipelines. OpenCV is integrated as a SparkML transformer, and an automated wrapper generation system is introduced which generates Python wrappers for PySpark.

The third paper describes an architecture for fitting many related but independent predictive models, where models are similar enough to warrant investment in shared infrastructure, but each model requires specific customizations and insight. The authors focus is on Repeated Domain-Specific Modeling systems for specific business problems, and the paper provides a set of abstractions for thinking about this type of model, and an implementation in R.

The fourth paper presents an open-source platform, Marvin, designed to help data scientists manage the lifecycle of an artificial intelligence project. Marvin can be used either on-premise or in the cloud with any algorithm that can be implemented in general-purpose languages and provides more comprehensive tools than other comparable solutions. It provides tools for data exploration, versioning, tools for sampling development data, a unit test framework, skeleton project generation, artifact versioning, parallel and distributed computing integration, training pipeline interface, a feedback server to allow models to receive external signals, and a predictor server to maintain trained models in persistent memory storage.

The fifth paper demonstrates how to predict airline flight delays using a Variational Long Short-Term Memory model. It differs from previous work using Multi-Layer Perceptron and General Regression Neural Net, traditional machine-learning techniques, and a domain-specific technique in using Monte Carlo Dropout to provide robust uncertainty metrics required for a risk management setting.

Acknowledgments

Program Committee and Session Chairs:
Keiran Thompson (Stanford), Ikaro Silva (Philips Research), Paula Fadul (Telefonica), Yan Zhang (Microsoft), Sabrina Feder (Comma Soft), Vincent Van Steenbergen, Erran Li (Uber), Jeremy Achin (DataRobot), Sudarshan Raghunathan (Microsoft), Eli Amesefe (Scribd), Ruben Glatt (USP), Matt Higgs (University of Glasgow) and Sunanda Koduvayur (Wayfair)

PAPIs Team:
Nikola Mojic, Iñigo Martinez, Brice Blanloeil, Ruben Glatt and Leonardo Noleto

Sponsors:
Microsoft, Wayfair, DataRobot, ElementAI, Unity, BigML, Dataiku, VersaMe, Nexosis, Adobe, Oracle

Finally, we wish to thank the authors for their time and effort in creating such a fine program.

October 2017

The Editorial Team:

Claire Hardgrove, PAPIs '17 Publications Chair
IBM Research Australia
chardgro@au1.ibm.com

Keiran Thompson, PAPIs '17 Program Co-chair
Stanford University
keiran@stanford.edu

Louis Dorard, PAPIs.io General Chair
UCL School of Management
l.dorard@ucl.ac.uk

Table of Contents