Simon Fischer, Ingo Mierswa (Eds.)

Proceedings of the 3rd RapidMiner Community Meeting and Conference (RCOMM 2012)

Shaker Verlag
Aachen 2012

Preface

Welcome to the RapidMiner Community Meeting and Conference 2012 in Budapest, Hungary. Being the third meeting in this series, it is becoming a more and more established event, and we, the organizers, hope that it will be as much fun as it was in the last years.

Again, we have selected a set of high-quality papers for presentation during the conference. Most of these cover new work, but some continue the successful work that has been started earlier and has been introduced already in previous RCOMMs.

Topics covered this year include various new extensions for supervised and unsupervised learning, building recommender systems, operating on images, text, and data streams, and combining RapidMiner with scripting languages. It is particularly remarkable that we have some very nice presentations coming from industry research as well as papers covering the application of RapidMiner in real-world use-cases. In addition to sixteen original research papers, Matthew North will launch his new data mining book, "Data Mining for the Masses", which uses RapidMiner as the primary software tool.

Due to the great demand in 2011, we decided to offer the training courses held on Tuesday and Friday again, in addition to the technical programme on Wednesday and Thursday. These include the certification exam on Friday which gives the participants the opportunity to become a certified Rapid-I Analyst. We also offer a special developer workshop enabling participants to create their own RapidMiner Extension. This workshop will also be a great opportunity to exchange and generate new ideas.

Last but not least, we will have one more edition of "Who wants to be a data miner", the rapid process design duel. We hope for many participants, challenging Rapid-I data miners.

Due to RCOMM taking place in beautiful Budapest we expect the dinner and concert as well as the trip to the city to become unforgettable events. The dinner will take place on board the A38 Ship, which was recently ranked 1st on Lonely Planet's list of 100 great bars of the world. We hope that you enjoy the programme and the presentations and utilize the time between the talks for exchanging ideas and establishing new collaborations.

June 2012

Simon Fischer
Head of Research and Development
Rapid-I GmbH

Contents

Part I – Data Analysis 1

1 Nearest-Neighbor and Clustering Based Anomaly Detection Algorithms for RapidMiner
Mennatallah Amer, Markus Goldstein **1**

2 LifeStyle Targeting on Big Data using RapidMiner
Maksim Drobyshev **13**

Part II – Extensions 19

3 Incorporating R Plots into RapidMiner Reports
Peter Jeszenszky **19**

4 An Octave Extension for RapidMiner
Sylvain Marié, Yaoyu Zhang **27**

5 NeurophRM: Integration of the Neuroph Framework into RapidMiner
M. Jovanović, J. Stojanović, M. Vukićević, V. Stojanović, B. Delibašić, M. Suknović **39**

6 Robust GPGPU Plugin Development for RapidMiner
Andor Kovács, Zoltán Prekopcsák **51**

Part III – Recommender Systems 63

7 Extending RapidMiner with Recommender Systems Algorithms
Matej Mihelčić, Nino Antulov-Fantulin, Matko Bošnjak, Tomislav Šmuc **63**

8 Design of User Based Collaborative Filtering in RapidMiner From Core Operators
Sérgio Morais, Carlos Soares **75**

Part IV – Unstructured Data **85**

9 Processing Data Streams with the RapidMiner Streams Plugin
Christian Bockermann, Hendrik Blom **85**

10 Image Mining Extension – Year After
Radim Burget, Václav Uher, Jan Mašek **99**

11 Automated Creation of Corpora for the Needs of Sentiment Analysis
Peter Koncz, Jan Paralic **107**

Part V – Applications **115**

12 Application of RapidMiner in Steel Industry Research and Development
Bengt-Henning Maas, Hakan Koç, Martin Bretschneider **115**

13 Comparison of Data-driven Models for Forecasting River Flows
Milan Cisty, Juraj Bezak **125**

14 Portfolio Optimization Using Local Linear Regression Ensembles in RapidMiner
Gábor Nagy, Gergő Barta, Tamás Henk **137**

Part VI – Making Your Life Easier **151**

15 Data Mining for the Masses: A New Textbook on Data Mining for Everyone
Matthew North **151**

16 Meta-Modeling Execution Times of RapidMiner Operators
Matija Piškorec, Matko Bošnjak, Tomislav Šmuc **159**

17 Get Some Coffee for Free – Writing Operators with RapidMiner Beans
Christian Bockermann, Hendrik Blom **169**