Massively collaborative machine learning
Rijn, J.N. van

Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/44814 holds various files of this Leiden University dissertation

Author: Rijn, Jan van
Title: Massively collaborative machine learning
Issue Date: 2016-12-19

Curriculum Vitae

Jan van Rijn was born on Wednesday 24 June 1987 in Katwijk. From 1999 until 2004 he was a student at Andreas College (location Pieter Groen) in Katwijk. In 2008 he obtained his Bachelor of Informatics at Leiden University of Applied Sciences, after which he obtained his master Computer Science in 2012 at Leiden University (thesis: 'Playing Games: The complexity of Klondike, Mahjong, Nonograms and Animal Chess' [120]).

During his studies he participated in various Algorithm Programming Contests, among which the North Western European Regional Contest in 2010 in Bremen, Germany. He was as student-assistant actively involved in various courses of the Informatics program at Leiden University. Furthermore, during his studies he was also active as Software Developer for a small business in Amsterdam.

From 2012 until 2016 he conducted his doctoral research under the supervision of Dr. Joaquin Vanschoren, Dr. Arno J. Knobbe and Prof. Dr. Joost N. Kok at Leiden Institute of Advanced Computer Science (LIACS). During this period he played a major role in the development and maintenance of OpenML. Furthermore, he was actively involved in teaching; he assisted at various Bachelor courses and gave several guest lectures about his research.

During his time as PhD student, he made several academic visits to foreign universities. He visited the University of Waikato three times for a period of several months, as invited by Prof. Dr. Bernhard Pfahringer and Prof. Dr. Geoffrey Holmes. He also visited the University of Porto for several weeks, as invited by Prof. Dr. Pavel Brazdil. Furthermore, Jan was also a member of the Institute Council, an advisory organ within the institute, and the Social Committee, for organizing social staff events.

Currently, Jan is working as researcher at the University of Freiburg. After obtaining his PhD he wants to pursue a career in academic.

Acknowledgements

This journey started in the spring of 2012, when I was in the process of finishing my Master's Thesis. My supervisors back then, Dr. Walter Kosters and Dr. Hendrik Jan Hoogeboom, encouraged me to have an informal chat with my current advisor, Prof. Dr. Joost Kok, about the open PhD positions. It was their guidance that made me enthusiastic about scientific research and led me towards my first academic position.

Many thanks towards the active OpenML community, that made the biannual workshops something to look forward to. In particular, I would like to thank the various people that I met at several locations around the world: Bernd Bischl, Paula Branco, Giuseppe Casalicchio, Matthias Feurer, Frank Hutter, Dominik Kirckhoff, Lars Kotthoff, Michel Lang, Rafael Mantovani, Luís Torgo and Joaquin Vanschoren.

I have been very lucky to work in the inspiring presence of many friends and colleagues from LIACS (and the Snellius building): Frans Birrer, Hendrik Blockeel, Benjamin van der Burgh, Ricardo Cachucho, André Deutz, Wouter Duivesteijn, Kleanthi Georgala, Vian Govers, Arno Knobbe, Rob Konijn, Michiel Kosters, Irene Martorelli, Marving Meeng, Annette Mense, Shengfa Miao, Siegfried Nijssen, Peter van der Putten, Kristian Rietveld, Jurriaan Rot, Claudio Sá, Marijn Schraagen, Hanna Schraffenberg, Marijn Swenne, Mima Stanojkovski, Anna Stawska, Frank Takes, Ugo Vespier, Jonathan Vis and Rudy van Vliet.

I would like to thank Prof. Dr. Bernhard Pfahringer and Prof. Dr. Geoffrey Holmes for hosting me multiple times at the University of Waikato. It has been a great pleasure to work together in such a prestigious and friendly environment. I also would like to thank several others from the Department of Computer Science, who made my visits both instructive and interesting: Christopher Beckham, Felipe Bravo-Marquez, Bob Durant, Dale Fletcher, Eibe Frank, Henry Gouk, Brian Hardyment, Simon Laing, Tim Leathart, Michael Mayo, Peter Reutemann, Sam Sarjant and Quan Sun. Our time in the lab was silver, but our time in the Tea Room was golden.

Furthermore, I would like to thank prof. dr. Pavel Brazdil and dr. Carlos Soares for inviting me to visit the University of Porto. It was a great experience to spend such cold months in the warm environment of LIAAD. Also many thanks to Salisu Abdulrahman, for the great collaborations prior, during and after my stay in Portugal.

Finally, I would like to mention my family and friends across the world. In particular, Nico, Leuntje, Niels, Annelies, Bas and Leoni van Rijn, for providing the warm and stable environment that I always could rely on. This thesis is dedicated to you.

Publication List

Below is a chronological list of publications by the author up to October 2016.

- S. M. Abdulrahman, P. Brazdil, J. N. van Rijn, and J. Vanschoren. Speeding up algorithm selection using average ranking and active testing by introducing runtime. *Machine Learning, Special Issue on Metalearning and Algorithm Selection*, forthcoming, 2016

- M. J. Post, P. van der Putten, and J. N. van Rijn. Does Feature Selection Improve Classification? A Large Scale Experiment in OpenML. In *Advances in Intelligent Data Analysis XV*, pages 158–170. Springer, 2016

- J. N. van Rijn, S. M. Abdulrahman, P. Brazdil, and J. Vanschoren. On the Evaluation of Algorithm Selection Problems. In *Proceedings of the 25th Belgian-Dutch Conference on Machine Learning (BeNeLearn 2016)*, 2 pages, 2016

- J. N. van Rijn, G. Holmes, B. Pfahringer, and J. Vanschoren. Having a Blast: Meta-Learning and Heterogeneous Ensembles for Data Streams. In *Data Mining (ICDM), 2015 IEEE International Conference on*, pages 1003–1008. IEEE, 2015

- J. N. van Rijn, F. W. Takes, and J. K. Vis. The complexity of rummikub problems. In *Proceedings of the 27th Benelux Conference on Artificial Intelligence (BNAIC 2015)*, 8 pages, 2015

- J. N. van Rijn, G. Holmes, B. Pfahringer, and J. Vanschoren. Case Study on Bagging Stable Classifiers for Data Streams. In *Proceedings of the 24th Belgian-Dutch Conference on Machine Learning (BeNeLearn 2015)*, 6 pages, 2015

- J. N. van Rijn and J. Vanschoren. Sharing RapidMiner Workflows and Experiments with OpenML. In J. Vanschoren, P. Brazdil, C. Giraud-Carrier, and L. Kotthoff, editors, *Proceedings of the 2015 International Workshop on Meta-Learning*

and Algorithm Selection (MetaSel), number 1455 in CEUR Workshop Proceedings, pages 93–103, Aachen, 2015

- S. M. Abdulrahman, P. Brazdil, J. N. van Rijn, and J. Vanschoren. Algorithm selection via meta-learning and sample-based active testing. In J. Vanschoren, P. Brazdil, C. Giraud-Carrier, and L. Kotthoff, editors, *Proceedings of the 2015 International Workshop on Meta-Learning and Algorithm Selection (MetaSel)*, number 1455 in CEUR Workshop Proceedings, pages 55–66, Aachen, 2015

- J. Vanschoren, J. N. van Rijn, B. Bischl, G. Casalicchio, M. Lang, and M. Feurer. OpenML: a Networked Science Platform for Machine Learning. In *ICML 2015 MLOSS Workshop*, 3 pages, 2015

- J. Vanschoren, J. N. van Rijn, and B. Bischl. Taking machine learning research online with OpenML. In *Proceedings of the 4th International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, pages 1–4. JLMR.org, 2015

- J. N. van Rijn, S. M. Abdulrahman, P. Brazdil, and J. Vanschoren. Fast Algorithm Selection using Learning Curves. In *Advances in Intelligent Data Analysis XIV*, pages 298–309. Springer, 2015

- J. Vanschoren, J. N. van Rijn, B. Bischl, and L. Torgo. OpenML: networked science in machine learning. *ACM SIGKDD Explorations Newsletter*, 15(2):49–60, 2014

- J. N. van Rijn and J. K. Vis. Endgame Analysis of Dou Shou Qi. *ICGA Journal*, 37(2):120–124, 2014

- J. N. van Rijn, G. Holmes, B. Pfahringer, and J. Vanschoren. Towards Meta-learning over Data Streams. In J. Vanschoren, P. Brazdil, C. Soares, and L. Kotthoff, editors, *Proceedings of the 2014 International Workshop on Meta-learning and Algorithm Selection (MetaSel)*, number 1201 in CEUR Workshop Proceedings, pages 37–38, Aachen, 2014

- J. N. van Rijn, G. Holmes, B. Pfahringer, and J. Vanschoren. Algorithm Selection on Data Streams. In *Discovery Science*, volume 8777 of *Lecture Notes in Computer Science*, pages 325–336. Springer, 2014

- H. J. Hoogeboom, W. A. Kosters, J. N. van Rijn, and J. K. Vis. Acyclic Constraint Logic and Games. *ICGA Journal*, 37(1):3–16, 2014

- J. N. van Rijn and J. K. Vis. Complexity and retrograde analysis of the game dou shou qi. In *BNAIC 2013: Proceedings of the 25th Benelux Conference on Artificial Intelligence*, 8 pages, 2013

- J. N. Van Rijn, V. Umaashankar, S. Fischer, B. Bischl, L. Torgo, B. Gao, P. Winter, B. Wiswedel, M. R. Berthold, and J. Vanschoren. A RapidMiner extension for Open Machine Learning. In *RapidMiner Community Meeting and Conference*, pages 59–70, 2013

- J. N. van Rijn, B. Bischl, L. Torgo, B. Gao, V. Umaashankar, S. Fischer, P. Winter, B. Wiswedel, M. R. Berthold, and J. Vanschoren. OpenML: A Collaborative Science Platform. In *Machine Learning and Knowledge Discovery in Databases*, pages 645–649. Springer, 2013

- J. N. Van Rijn and J. Vanschoren. OpenML: An Open Science Platform for Machine Learning. In *Proceedings of the 22th Belgian-Dutch Conference on Machine Learning (BeNeLearn 2013)*, 1 page, 2013

Titles in the IPA Dissertation Series since 2013

H. Beohar. *Refinement of Communication and States in Models of Embedded Systems*. Faculty of Mathematics and Computer Science, TU/e. 2013-01

G. Igna. *Performance Analysis of Real-Time Task Systems using Timed Automata*. Faculty of Science, Mathematics and Computer Science, RU. 2013-02

E. Zambon. *Abstract Graph Transformation – Theory and Practice*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2013-03

B. Lijnse. *TOP to the Rescue – Task-Oriented Programming for Incident Response Applications*. Faculty of Science, Mathematics and Computer Science, RU. 2013-04

G.T. de Koning Gans. *Outsmarting Smart Cards*. Faculty of Science, Mathematics and Computer Science, RU. 2013-05

M.S. Greiler. *Test Suite Comprehension for Modular and Dynamic Systems*. Faculty of Electrical Engineering, Mathematics, and Computer Science, TUD. 2013-06

L.E. Mamane. *Interactive mathematical documents: creation and presentation*. Faculty of Science, Mathematics and Computer Science, RU. 2013-07

M.M.H.P. van den Heuvel. *Composition and synchronization of real-time components upon one processor*. Faculty of Mathematics and Computer Science, TU/e. 2013-08

J. Businge. *Co-evolution of the Eclipse Framework and its Third-party Plug-ins*. Faculty of Mathematics and Computer Science, TU/e. 2013-09

S. van der Burg. *A Reference Architecture for Distributed Software Deployment*. Faculty of Electrical Engineering, Mathematics, and Computer Science, TUD. 2013-10

J.J.A. Keiren. *Advanced Reduction Techniques for Model Checking*. Faculty of

Mathematics and Computer Science, TU/e. 2013-11

D.H.P. Gerrits. *Pushing and Pulling: Computing push plans for disk-shaped robots, and dynamic labelings for moving points*. Faculty of Mathematics and Computer Science, TU/e. 2013-12

M. Timmer. *Efficient Modelling, Generation and Analysis of Markov Automata*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2013-13

M.J.M. Roeloffzen. *Kinetic Data Structures in the Black-Box Model*. Faculty of Mathematics and Computer Science, TU/e. 2013-14

L. Lensink. *Applying Formal Methods in Software Development*. Faculty of Science, Mathematics and Computer Science, RU. 2013-15

C. Tankink. *Documentation and Formal Mathematics — Web Technology meets Proof Assistants*. Faculty of Science, Mathematics and Computer Science, RU. 2013-16

C. de Gouw. *Combining Monitoring with Run-time Assertion Checking*. Faculty of Mathematics and Natural Sciences, UL. 2013-17

J. van den Bos. *Gathering Evidence: Model-Driven Software Engineering in Automated Digital Forensics*. Faculty of Science, UvA. 2014-01

D. Hadziosmanovic. *The Process Matters: Cyber Security in Industrial Control Systems*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2014-02

A.J.P. Jeckmans. *Cryptographically-Enhanced Privacy for Recommender Systems*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2014-03

C.-P. Bezemer. *Performance Optimization of Multi-Tenant Software Systems*. Faculty of Electrical Engineering, Mathematics, and Computer Science, TUD. 2014-04

T.M. Ngo. *Qualitative and Quantitative Information Flow Analysis for Multi-threaded Programs*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2014-05

A.W. Laarman. *Scalable Multi-Core Model Checking*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2014-06

J. Winter. *Coalgebraic Characterizations of Automata-Theoretic Classes*. Faculty of Science, Mathematics and Computer Science, RU. 2014-07

W. Meulemans. *Similarity Measures and Algorithms for Cartographic Schematization*. Faculty of Mathematics and Computer Science, TU/e. 2014-08

A.F.E. Belinfante. *JTorX: Exploring Model-Based Testing*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2014-09

A.P. van der Meer. *Domain Specific Languages and their Type Systems*. Faculty

of Mathematics and Computer Science, TU/e. 2014-10

B.N. Vasilescu. *Social Aspects of Collaboration in Online Software Communities*. Faculty of Mathematics and Computer Science, TU/e. 2014-11

F.D. Aarts. *Tomte: Bridging the Gap between Active Learning and Real-World Systems*. Faculty of Science, Mathematics and Computer Science, RU. 2014-12

N. Noroozi. *Improving Input-Output Conformance Testing Theories*. Faculty of Mathematics and Computer Science, TU/e. 2014-13

M. Helvensteijn. *Abstract Delta Modeling: Software Product Lines and Beyond*. Faculty of Mathematics and Natural Sciences, UL. 2014-14

P. Vullers. *Efficient Implementations of Attribute-based Credentials on Smart Cards*. Faculty of Science, Mathematics and Computer Science, RU. 2014-15

F.W. Takes. *Algorithms for Analyzing and Mining Real-World Graphs*. Faculty of Mathematics and Natural Sciences, UL. 2014-16

M.P. Schraagen. *Aspects of Record Linkage*. Faculty of Mathematics and Natural Sciences, UL. 2014-17

G. Alpár. *Attribute-Based Identity Management: Bridging the Cryptographic Design of ABCs with the Real World*. Faculty of Science, Mathematics and Computer Science, RU. 2015-01

A.J. van der Ploeg. *Efficient Abstractions for Visualization and Interaction*. Faculty of Science, UvA. 2015-02

R.J.M. Theunissen. *Supervisory Control in Health Care Systems*. Faculty of Mechanical Engineering, TU/e. 2015-03

T.V. Bui. *A Software Architecture for Body Area Sensor Networks: Flexibility and Trustworthiness*. Faculty of Mathematics and Computer Science, TU/e. 2015-04

A. Guzzi. *Supporting Developers' Teamwork from within the IDE*. Faculty of Electrical Engineering, Mathematics, and Computer Science, TUD. 2015-05

T. Espinha. *Web Service Growing Pains: Understanding Services and Their Clients*. Faculty of Electrical Engineering, Mathematics, and Computer Science, TUD. 2015-06

S. Dietzel. *Resilient In-network Aggregation for Vehicular Networks*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2015-07

E. Costante. *Privacy throughout the Data Cycle*. Faculty of Mathematics and Computer Science, TU/e. 2015-08

S. Cranen. *Getting the point — Obtaining and understanding fixpoints in model checking*. Faculty of Mathematics and Computer Science, TU/e. 2015-09

R. Verdult. *The (in)security of proprietary cryptography*. Faculty of Science, Mathematics and Computer Science, RU. 2015-10

J.E.J. de Ruiter. *Lessons learned in the analysis of the EMV and TLS security pro-*

tocols. Faculty of Science, Mathematics and Computer Science, RU. 2015-11

Y. Dajsuren. *On the Design of an Architecture Framework and Quality Evaluation for Automotive Software Systems*. Faculty of Mathematics and Computer Science, TU/e. 2015-12

J. Bransen. *On the Incremental Evaluation of Higher-Order Attribute Grammars*. Faculty of Science, UU. 2015-13

S. Picek. *Applications of Evolutionary Computation to Cryptology*. Faculty of Science, Mathematics and Computer Science, RU. 2015-14

C. Chen. *Automated Fault Localization for Service-Oriented Software Systems*. Faculty of Electrical Engineering, Mathematics, and Computer Science, TUD. 2015-15

S. te Brinke. *Developing Energy-Aware Software*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2015-16

R.W.J. Kersten. *Software Analysis Methods for Resource-Sensitive Systems*. Faculty of Science, Mathematics and Computer Science, RU. 2015-17

J.C. Rot. *Enhanced coinduction*. Faculty of Mathematics and Natural Sciences, UL. 2015-18

M. Stolikj. *Building Blocks for the Internet of Things*. Faculty of Mathematics and Computer Science, TU/e. 2015-19

D. Gebler. *Robust SOS Specifications of Probabilistic Processes*. Faculty of Sciences, Department of Computer Science, VUA. 2015-20

M. Zaharieva-Stojanovski. *Closer to Reliable Software: Verifying functional behaviour of concurrent programs*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2015-21

R.J. Krebbers. *The C standard formalized in Coq*. Faculty of Science, Mathematics and Computer Science, RU. 2015-22

R. van Vliet. *DNA Expressions – A Formal Notation for DNA*. Faculty of Mathematics and Natural Sciences, UL. 2015-23

S.-S.T.Q. Jongmans. *Automata-Theoretic Protocol Programming*. Faculty of Mathematics and Natural Sciences, UL. 2016-01

S.J.C. Joosten. *Verification of Interconnects*. Faculty of Mathematics and Computer Science, TU/e. 2016-02

M.W. Gazda. *Fixpoint Logic, Games, and Relations of Consequence*. Faculty of Mathematics and Computer Science, TU/e. 2016-03

S. Keshishzadeh. *Formal Analysis and Verification of Embedded Systems for Healthcare*. Faculty of Mathematics and Computer Science, TU/e. 2016-04

P.M. Heck. *Quality of Just-in-Time Requirements: Just-Enough and Just-in-Time*. Faculty of Electrical Engineering, Mathematics, and Computer Science, TUD. 2016-05

Y. Luo. *From Conceptual Models to Safety Assurance – Applying Model-Based Tech-*

niques to Support Safety Assurance. Faculty of Mathematics and Computer Science, TU/e. 2016-06

B. Ege. *Physical Security Analysis of Embedded Devices*. Faculty of Science, Mathematics and Computer Science, RU. 2016-07

A.I. van Goethem. *Algorithms for Curved Schematization*. Faculty of Mathematics and Computer Science, TU/e. 2016-08

T. van Dijk. *Sylvan: Multi-core Decision Diagrams*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2016-09

I. David. *Run-time resource management for component-based systems*. Fac-

ulty of Mathematics and Computer Science, TU/e. 2016-10

A.C. van Hulst. *Control Synthesis using Modal Logic and Partial Bisimilarity – A Treatise Supported by Computer Verified Proofs*. Faculty of Mechanical Engineering, TU/e. 2016-11

A. Zawedde. *Modeling the Dynamics of Requirements Process Improvement*. Faculty of Mathematics and Computer Science, TU/e. 2016-12

F.M.J. van den Broek. *Mobile Communication Security*. Faculty of Science, Mathematics and Computer Science, RU. 2016-13

J.N. van Rijn. *Massively Collaborative Machine Learning*. Faculty of Mathematics and Natural Sciences, UL. 2016-14