

Simon Rawles

Please email me for the version of this CV with my address and phone number.

Email jobs@rawles.net
Web https://www.rawles.net/

Dated May 14, 2018

A research engineer with scientific and commercial R&D experience, including at project lead level, aiming to create and develop novel machine learning and data mining technologies to solve complex problems.

Programming: Java SE (20 years), Clojure, Perl, Prolog, Python (NumPy, SciPy), JavaScript.
Data analysis: R, scikit-learn, Map/Reduce, Cascalog, Hadoop, MySQL, PostgreSQL, PostGIS.
Machine learning: Learning from structured data, learning with ontologies/taxonomies, learning with logics, propositionalisation, feature construction, feature selection, computational linguistics, domain modelling, kernel partial least squares, time series analysis.

Experience

September 2016 – March 2017 Neodemographics Lab, University of Nottingham, UK
Postdoctoral Research Associate Principal Investigator: Dr. James Goulding
Investigated methods for flood detection and prediction in Tanzania using only mobile telephone metadata.

- Used a range of time series analysis methods: changepoint analysis in ARIMA models, outlier effect detection, intervention analysis, Poisson processes, Granger causality.
- Implemented a pre-processing pipeline to transform, clean and index the transaction log for mining.
- Developed a feature construction method, where a structured taxonomy of time series features are constructed from user behaviour criteria. These features defined cohorts of users.
- Detected changepoints were shown to correlate well with flood incidence for particular cell towers. Learning with linear combinations of cohort features demonstrated promising results.

February 2016 – July 2016 Dept. of Computer Science, University of Nottingham, UK
Postdoctoral Research Associate Principal Investigator: Prof. Thomas Gärtner
Investigated methods for predicting binding selectivity and activity of ligands for data-driven drug design. Selective ligands bind to a very limited set of receptors or receptor types. Of interest were learning approaches which select and rank candidate molecules according based on selectivity and not individual receptor activity.

- Collaborating with a computational chemistry domain expert, explored and analysed the activity data of 140 molecules. Visualised and clustered the data via knowledge-based embeddings.
- Investigated the performance of machine learning tools for QSAR regression, comparing established descriptor features with a graph kernel. Used partial least squares and kernel-based regression.
- Investigated adding an orthogonalisation term to the objective function of a kernel-based regularised least squares regression problem with multiple outputs. On synthetic and real-world data, demonstrated a 25%–45% drop in RMSE in selectivity prediction.
- Assisted with the preparation of a grant application.

October 2015 – June 2016 (since then I act as an advisor to the R&D team) Gluru Ltd.,
Soho, London, UK
Research Engineer
Research-led development of machine learning algorithms and software for an intelligent assistant app.

- Developed a context-aware recommender system algorithm which identifies relevant and related files, users, and other resources. The algorithm adapts to user behaviour, ratings and preferences.
- Developed and implemented a model to model and predict user intention. Given an incoming email, the model predicts with high precision and recall whether the user will choose to reply.
- Co-authored a patent application describing the company's core technology.

January 2013 – June 2015

Software Engineer (Cloud and Data, Data Science, Analytics)

SwiftKey/TouchType Ltd. (now Microsoft Research),

Southwark, London, UK

Working in and leading engineering and R&D teams developing novel tools and techniques for terabyte-scale data analysis. Applications include decision support, user profiling and language modelling.

- Performed explanatory and predictive modelling (including time-series) to detect subgroups of users in language and telemetry data. Analysed product performance and factors in customer retention.
- Applied machine learning techniques to computational linguistics problems, including topic modelling, phrase trending, named-entity recognition, neologism and profanity detection.
- Established and developed data mining infrastructure, including processing pipelines and efficient aggregate computation for language and telemetry data.
- Developed a web graph crawler/miner collecting linguistic source data guided by a learned heuristic. The crawler automatically explores relevant web pages and extracts relational knowledge from them.
- Performed data cleaning and integration, e.g. deduplication, smart match/merge and defect analysis.
- Constructed a user data triplestore for iterative extraction of domain-specific background knowledge.
- Led the first eight months of the company's data science department (four engineers). Set objectives, reviewed and planned projects and prepared proposals and grant applications.

September 2007 – December 2012

Development Manager (Intelligent Cache)

Multicom Products Ltd.,

Bristol, UK

I led a team of three senior research engineers developing OMNISCACHE, a rapid general-purpose intelligent cache, from the ground up, relying on machine learning technology for its accuracy and efficiency.

- Over three years, led the project from inception, to demo, to development for a FTSE-250 customer.
- Designed and implemented core algorithms, e.g. learning of cache eviction and filling policies.
- Coordinated review/planning, directed scientific discussion, and oversaw the technical direction.
- Prepared bids, requirements and specifications, collaborated with other teams and companies.
- Applying and integrating research from the fields of information retrieval, query refinement in description logics, and semantic web-style ontologies, co-developed OMNISEARCH, a commercial holiday package search engine. OMNISEARCH combines semantic reasoning, datacubes and free-text search to deliver sub-second results from many millions of records.

October 2000 – January 2002

Research Assistant

Department of Computer Science,

University of Warwick, Coventry, UK

- Investigated the application of software modelling methods to business process reengineering (BPR). Implemented a collaborative web tool to facilitate BPR.
- Conducted a comprehensive pedagogical evaluation of computer-assisted learning software as part of a large-scale research consortium, leading to an ACM journal article.

January 1998 – September 1999

Software Engineer (Research and Development)

GEC-Marconi Aerospace Systems (now BAe Systems),

Rochester, UK

- Researched, designed and implemented key components of a C++ avionics middleware. Coordinated an OO working group and management committees. Promoted COTS/OO approaches within GEC.
- Developed and tested display systems modules for the Eurofighter Typhoon fighter aircraft in Ada.

Qualifications

October 2002 – April 2007

Dept. of Computer Science, University of Bristol, UK

Ph.D.: *Object-oriented Data Mining.*

Supervisor: Prof. Peter A. Flach

- Developed and implemented COSINUS, a data mining system adopting the object-oriented data model to allow highly domain-centric mining of structured, heterogenous data objects belonging naturally in class taxonomies. COSINUS employs an object-specific refinement operator which performs domain-constrained feature construction, using these features in propositionalisation.
- Developed REFER, an efficient algorithm for eliminating logical redundancy in general Boolean multi-class feature sets without compromising the discovery of a classification rule. REFER therefore reduces complexity, especially in propositionalised data.
- Successfully applied object-oriented data mining to a real-world natural language processing task, producing a learner which reliably assigns grammatical/semantic roles to words and phrases, a key task in natural language querying and automatic translation.

October 1999 – September 2000

Dept. of Computer Science, University of Bristol, UK

M.Sc. with Commendation in Machine Learning. Grade 72%. Tutor: Prof. Christophe Giraud-Carrier

- Courses: Artificial Intelligence: 91%; Neural Networks: 82%; Evolutionary Computing: 77%; Advanced Topics in Machine Learning: 74%; Propositional Inductive Learning: 73%; Higher-order Inductive Learning: 72%; Computational Intelligence: 68%; Computational Learning Theory: 63%.
- Dissertation (individual): *Symbolic Processing in Neural Networks and its Application to Region Classification*. I devised a hybrid neural/symbolic learning system based on existing techniques for symbolic reasoning in neural networks and applied it to a traffic recognition task.

October 1994 – July 1997

University of Warwick, UK

B.Sc. Mathematics and Theory of Computation. Class 2:2.

Studied pure mathematics, theoretical computer science and German. Final year (individual) project compared object-oriented and declarative bases for modelling in simulation.

Publications

Tim Kovacs, Simon Rawles, Larry Bull, Masaya Nakata, and Keiki Takadama. XCS-DH: Minimal default hierarchies in XCS. In *IEEE Congress on Evolutionary Computation (IEEE CEC 2016) Special Session on New Directions in Evolutionary Machine Learning (to appear)*, 2016.

Simon Rawles and Peter A. Flach. Neighbourhoods of Examples for Detecting Logical Redundancy. In *Late-breaking Papers, ILP 2005*, pages 47–52, 2005.

Annalisa Appice, Michelangelo Ceci, Simon Rawles, and Peter A. Flach. Redundant Feature Elimination for Multi-Class Problems. In *Proceedings of ICML 2004*, pages 33–40, 2004.

Simon Price, Simon Rawles, and Peter Flach. Estimating whether partial FOAF descriptions describe the same individual. In *First Workshop on Friend of a Friend, Social Networking and the Semantic Web*, September 2004.

Mark-A. Krogel, Simon Rawles, Filip Železný, Peter A. Flach, Nada Lavrač, and Stefan Wrobel. Comparative Evaluation of Approaches to Propositionalization. In *Proceedings of ILP 2003*, pages 197–214, 2003.

Peter A. Flach, Hendrik Blockeel, Thomas Gärtner, Marko Grobelnik, Branko Kavšek, Martin Kejkula, Darek Krzywania, Nada Lavrač, Peter Ljubnič, Dunja Mladenić, Steve Moyle, Stefan Raeymaekers, Jan Rauch, Simon Rawles, Rita Ribeiro, Gert Sclep, Jan Struyf, Ljupčo Todorovski, Luis Torgo, Dietrich Wettschereck, and Shaomin Wu. *On the road to knowledge: mining 21 years of UK traffic accident reports*, chapter 11, pages 143–155. Kluwer Academic Publishers, 2003.

Lenka Nováková, Jiří Kléma, Michal Jakob, Simon Rawles, and Olga Štěpánková. Trend analysis and risk identification. In *Discovery Challenge Workshop, ECML 2003/PKDD 2003*, pages 95–107, 2003.

Mike Joy, Boris Muzykantskii, Simon Rawles, and Michael Evans. An Infrastructure for Web-Based Computer-Assisted Learning. *Journal on Educational Resources in Computing (JERIC)*, 2(4):1–19, 2002.

Simon Rawles, Mike Joy, and Michael Evans. Computer-Assisted Assessment in Computer Science: Issues and Software. Technical report, Department of Computer Science, University of Warwick, UK, 2002.

(232 citations, h -index = 5)

Conferences

- ICML 2005: 22nd International Conference on Machine Learning, Bonn, Germany. Attended.
- ILP 2005: 15th International Conference on Inductive Logic Programming, Bonn, Germany. Presented a late-breaking paper on recent developments in logical feature elimination.
- ICML 2004: 21st International Conference on Machine Learning, Banff, Canada. Presented a conference paper on the REFER local feature elimination technique.
- ILP 2003: 13th International Conference on Inductive Logic Programming, Szeged, Hungary. Co-presented a conference paper on propositionalisation during the first year of my Ph.D. studies.

Programme Committees

- WCCI/CEC 2016: IEEE World Congress on Computational Intelligence, Vancouver, Canada. Special Session on New Directions in Evolutionary Machine Learning.

Research visits

- Takadama Laboratory, Department of Informatics, University of Electro-Communications, Tokyo, Japan. Summer 2015. Collaborating with Tim Kovacs (Bristol), Masaya Nakata (Yokohama), Keiki Takadama (UEC Tokyo) and his laboratory, we focussed on techniques for symbolic machine learning incorporating the generality ordering and how they might be used in a more open-ended approach to machine learning and data mining. This work is still in progress and has led to one workshop paper.
- Gerstner Laboratory, Faculty of Cybernetics, Czech Technical University, Prague, Czech Republic. Summer 2003. Concentrating on the data transformation phase of the multi-relational data mining process, I collaborated with staff and Ph.D. students at the CTU during an EU-funded visit to contribute to an approach to the discovery challenge for the ECML/PKDD'03 conference, concerning a longitudinal study on atherosclerosis risk factors.

Patents

- US Patent application 15/368,263. *System and Method for Intelligent Resource Ranking and Retrieval*, filed December 2016.

Abstract of my PhD thesis, *Object-oriented Data Mining*

Attempts to overcome limitations in the attribute-value representation for machine learning have led to much interest in learning from structured data, concentrated in the research areas of inductive logic programming (ILP) and multi-relational data mining (MDRM). The expressiveness and encapsulation of the object-oriented data model has led to its widespread adoption in software and database design. The considerable congruence between this model and individual-centred models in inductive logic programming presents new opportunities for mining object data specific to its domain.

The thesis investigates the use of object-orientation in knowledge representation for multi-relational data mining. The thesis proposes a language for expressing object model metaknowledge and uses it to extend the reasoning mechanisms of an object-oriented logic. A refinement operator is then defined and used for feature search in a object-oriented propositionalisation-based ILP classifier. An algorithm is proposed for reducing the large number of redundant features typical in propositionalisation. A data mining system based on the refinement operator is implemented and demonstrated on a real-world computational linguistics task and compared with a conventional ILP system.