

Institution: Royal Holloway, University of London
Unit of Assessment: 11 Computer Science and Informatics
Title of case study: VC Learning Theory and Support Vector Machines
<p>1. Summary of the impact (indicative maximum 100 words)</p> <p>Results of research at Royal Holloway in Machine Learning – Support Vector Machine (SVM), kernel methods and conformal-prediction techniques – are at the source of the analytics and ‘Big Data’ revolution, whose impact is transforming the economy (and society), from data mining to machine vision, from Web search to spam detection.</p> <p>Although SVMs were not invented at Royal Holloway, our contributions include the original reference monograph (Vapnik 1998), basic underpinning theory (such as [3.2]), the first widely distributed open-source implementation [3.1], the first accessible textbook (Cristianini and Shawe-Taylor 2000), and multiple extensions (such as [3.3] and [3.4]).</p>
<p>2. Underpinning research (indicative maximum 500 words)</p> <p>The underpinning research was conducted at Royal Holloway from 1995-present by: academic staff – Chervonenkis (0.5), Gammerman, Shawe-Taylor, Shafer (0.5), Vapnik (0.5), Vovk, Watkins; RAs – Cristianini, Kandola, Vinokourov; and PhD students – Saunders, Stitson, Weston, Lodhi.</p> <p>We focus on three technologies: SVM, kernels, and conformal prediction.</p> <p>The Support Vector Machine is a general machine learning method that can efficiently learn complex prediction rules from large datasets. It is theoretically well justified, efficient and easy to apply, and it has excellent performance on many real-world problems. The method was originally proposed in 1995 at AT&T Bell Labs by Vapnik and collaborators. In the same year, Vapnik joined Royal Holloway and we started a research programme on SVMs and statistical learning theory. Together with AT&T Bell Labs, we wrote and distributed an open source SVM implementation, which was the first SVM software to be widely available [3.1].</p> <p>In 1998 Vapnik published <i>Statistical Learning Theory</i> (Wiley), a seminal monograph that gives the full, extended presentation of SVMs and rigorous exposition of the underlying learning theory, and proposes an extension of SVM to an important new setting (transductive SVM). In the postscript to the second edition (2006) of his 1982 book, Vapnik extended SVM in a different direction – to SVM+; [3.2] extended Vapnik’s theoretical analysis to include the case where the prediction rule complexity is chosen after looking at the data.</p> <p>Theoretical analysis of learning has been greatly extended by us since, through which it has acquired practical impact by underpinning insightful engineering design of learning algorithms. For example, two extensions of the original two-class SVM were made through which examples can be classified into several possible classes (e.g., to classify an image as one of the ten possible digits) [3.3], and a careful derivation of kernelised ridge regression was developed in [3.4].</p> <p>In 2000, Cristianini and Shawe-Taylor published <i>An Introduction to Support Vector Machines and other Kernel-Based Learning Methods</i> (CUP), which presented accessible theory and simple algorithms. This was the first book-length exposition accessible to general technical readers. It was influential, disseminating the research and stimulating widespread development of applications.</p> <p>SVMs are used with kernel functions, or ‘kernels’. A kernel is a user-defined measure of similarity between data examples required to satisfy a technical condition of positive definiteness. We invented two new classes of kernels:</p> <ul style="list-style-type: none"> • [3.5] introduced string kernels for comparing strings of different lengths. These stimulated much further development of applicable kernels for bio-sequences, parse trees, time series, molecules, and other discrete structures. • [3.6] constructed kernels that measure semantic similarity of documents based on a corpus of related, unlabelled documents. These kernels, applicable in information retrieval, led on to the development of cross-language document retrieval by correlational machine learning alone.

In contrast to Bayesian methods, SVMs did not provide estimates of the accuracy of each individual prediction. Motivated by this difficulty, we developed a radically new method of estimating the accuracy of each prediction as it is made, in a simple yet rigorously principled non-Bayesian way. This **conformal prediction** method can be applied to nearly any machine learning algorithm, not just SVMs. Royal Holloway has awarded 11 PhDs and published over 70 papers in Conformal Prediction, including the book *Algorithmic Learning in a Random World* (Springer 2005) by Vovk, Gammerman and Shafer.

3. References to the research (indicative maximum of six references)

- 3.1 Craig Saunders, Mark Stitson, Jason Weston, Léon Bottou, Bernhard Schölkopf, Alex Smola *Support Vector Machine Reference Manual* Technical Report: Department of Computer Science Royal Holloway CSD-TR-98-03, 1998
- 3.2 John Shawe-Taylor, Peter Bartlett, Robert Williamson, and Anthony Martin *Structural risk minimisation over data-dependent hierarchies*. IEEE Transactions on Information Theory, **5** pp1926-1940, 1998
- 3.3 Jason Weston and Chris Watkins. 1999. *Multi-class support vector machines*. In: Proceedings of the 6th European Symposium on Artificial Neural Networks (ESANN 99); John Platt, Nello Cristianini, John Shawe-Taylor *Large Margin DAGs for Multiclass Classification* NIPS 12 pp 547-553, 1999

Evidence of the quality of the under-pinning research

- 3.4 Craig Saunders, Alexander Gammerman, and Volodya Vovk *Ridge regression in dual variables* in Proceedings of the 15th International Conference on Machine Learning, pp 515-521, 1998.
- 3.5 Huma Lodhi, Craig Saunders, John Shawe-Taylor, Nello Cristianini, and Chris Watkins *Text classification using string kernels* Journal of Machine Learning Research (2002) **2** pp 419-444
- 3.6 Nello Cristianini, John Shawe-Taylor, Huma Lodhi *Latent Semantic Kernels*, pp 66-73 in Carla Brodley, Andrea Danyluk (Eds.): Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001), Williams College, Williamstown, MA, USA, June 28 - July 1, 2001. Morgan Kaufmann 2001 ISBN 1-55860-778-1

4. Details of the impact (indicative maximum 750 words)

The growing impact that the application of machine learning techniques on data – ‘Big Data’ – is having in society and the economy is profound, to such an extent that it is impossible to quantify. The key enablers of this explosion have been:

- The availability of reliable algorithms such as the various generations of SVMs: research at Royal Holloway [3.2-3.6] has been instrumental in the development of this technology, including string kernels and conformal prediction.
- The availability of open-source software, to which Royal Holloway contributed the first open-source SVM implementation to be widely available [3.1].
- Education: the first accessible textbooks on SVMs were produced at Royal Holloway, which are still at the source of course material that is being developed today all over the world – for example, a specialist course (<http://www.is.titech.ac.jp/~shimo/class/fukumizu/>) at the University of Tokyo in 2010. SVMs are a standard major topic in introductory machine learning courses such as the two best-known on-line machine learning courses in the world – “Learning from Data” (<https://www.edx.org/course/caltechx/cs1156x/learning-data/1120>) course CS1156x from CaltechX presented on edX fall 2013, and Andrew Ng’s Machine Learning Course at Coursera (<https://www.coursera.org/course/ml>).

Among SVM open-source implementations available during the REF period are those in R toolboxes, WEKA, and libsvm. Examples of SVM implementations that were commercially distributed within software suites during the REF period are SAS Enterprise Miner (2011), Oracle Data Miner (2008) and Avanade SQL Server Analysis Services (2012).

We now provide more concrete evidence that, during the REF period, SVMs have been widely used in data analytics and embedded applications, generating or enabling economic impact through improved products and services in ways that can be directly linked to our research.

Economic impact: improved products and services; creation of new business sectors

SVMs are widely used in big-data analytics, for example in: forecasting of demand; credit scoring; detection of fraudulent transactions; spam detection; and prediction of customer preferences. Two examples of successful companies that depend on SVMs for data analytics are:

- Idalab GmbH [5.1], which uses SVM as their key technology, for market research and business process risk analysis.
- TWIMPACT [5.2], whose clients include advertisement and public relations divisions of many high-tech companies worldwide.

SVMs are also used as embedded components of larger systems for classifying images, events or machine states from sensor data. These applications are typically proprietary and confidential. We therefore present specimen patent applications as indirect evidence of widespread impact because they show that companies are willing to spend money protecting rights to inventions that work. The following specimen examples cite (Cristianini and Shawe-Taylor 2000) and the preferred embodiments of the inventions use SVM:

- Patient state detection based on support vector machine based algorithm (Medtronic Inc 2010, WO2010126625)
- Diagnosis support system providing guidance to a user by automated retrieval of similar cancer images with user feedback (Sti Medical Systems Llc 2012, WO2012154216)
- Systems and methods for detecting obstructions in a camera field of view (Mobileye Technologies Ltd 2013, US8553088)
- Method, array, and use thereof [for diagnosis of pancreatic cancer] (Immunovia AB 2012, EP2490025)
- Anomaly detection for link-state routing protocols (Telecom Italia S.p.A. 2010, EP2235910)
- Process of checking the presence of a grid pattern in an X-ray image (Agfa Healthcare NV 2009, EP1696366)
- Method for producing peptide libraries and use thereof (Sanofi Aventis 2009, EP2137658)
- Radiation image processing method, apparatus, and program (Fujifilm Corp. 2008, WO2008084880)
- Support machine using iterative chunking and violators (Oracle Corp. 2011, US7937351)

Royal Holloway research on SVMs and **kernel methods** has also contributed to the rise of a new business sector – the SVM was modified to both learn from and to predict comparative preferences rather than absolute values, so that a machine learning system could learn to predict which retrieved items users would most like to see, and which possible advertisements users would be most likely to respond to. These techniques are used in data retrieval, web search, and recommendation systems, with global impact.

For example, the US patents [5.3] granted to Google Inc. are for a method of automatically generating additional query suggestions to give additional useful search results – for example, “car for hire” should be extended with the query “car for rent” in the US. These patents cite eleven non-patent sources; of the six that were chosen by the inventors, four are Royal Holloway publications: [3.5], [3.6], one paper by Kandola et al, and another by Vinokourov et al.

String kernels in particular have influenced many fields. Originally, SVM kernels were essentially smoothing operators for multivariate data. The impact of defining similarity by dynamic alignment of matching scores instead of Euclidean distance was a new methodology for developing application-specific kernels for many types of discrete data – for example, in bioinformatics, natural language processing, time-series and chemistry. A letter from the Head of Google Research, New York [5.4] describes the influence that [3.5] in particular had on research leading to deployment of AT&T’s US-wide automated phone helpline. Exemplary patent applications making reference to [3.5] are [5.5] and [5.6].

Finally, sophisticated users are already starting to apply **conformal prediction** – a supporting letter [5.7] describes its trial use for generating reliable warnings of plasma instability in experimental fusion reactors (the JET Tokamak at Culham, and the Stellarator at CIEMAT).

5. Sources to corroborate the impact (indicative maximum of 10 references)

The following sources corroborate the impact of our research in SVM [5.1, 5.2], kernels [5.3, 5.4, 5.5, 5.6], and conformal prediction [5.7].

- [5.1] Letter from a Managing Partner, idalab GmbH Germany.
- [5.2] Letter from the owner and CEO, twimpact UG, Germany.
- [5.3] US patent US7725485 (granted 2010) and its extensions US8015199 (granted 2011) and US8209347 (granted 2012) – *Generating query suggestions using contextual information*.
- [5.4] Letter from the Head of Google Research, New York.
- [5.5] US patent US7647534 (Xerox Corporation 2010) – *Method of avoiding repetition of user actions by using past users' experiences*.
- [5.6] US patent US8463718 ((Health Discovery Corporation 2013) – *Support vector machine-based method for analysis of spectral data*. Citation of [3.5] is in the body of the description of invention.
- [5.7] Letter from the Head of Data Acquisition Unit, CIEMAT, Spain.