Curriculum Vitae of Ralf Herbrich

Date of birth:	May 11, 1974
Nationality:	German
Marital Status:	Married / 2 children
Address:	6 Brierley Walk, CB4 3NH Cambridge, United Kingdom
Phone:	+44 1223 479803
Email:	rherb@microsoft.com

Education

Primary School Juri-Gagarin POS, Lunow (1980 – 1990)
Primary school leaving certificate (Mittlere Reife), grade: "excellent".

High School Einstein Gymnasium, Angermünde (1990 – 1992)
High school leaving certificate (Abitur), grade: "excellent".

Computer Science at Technical University of Berlin (1992 – 1993, 1994 – 1997)
Studies in computer science and business. Intermediate examination certificate (Vordiplom), grade: "good".

Computer Science Diploma (November 1997)
Master project: "Segmentation with Gabor filters for the induction of structural classifiers on images". Final grade: "excellent" (best possible final grade).

PhD in Computer Science at the Technical University of Berlin (1997 – 2000)
Machine Learning (Learning Theory) and Statistics with Prof. U. Kockelkorn and Prof. K. Obermayer at the Technical University of Berlin.

Doctor Rerum Naturalis (September 2000)
PhD thesis: "Learning Linear Classifiers: Theory and Algorithms". Final grade: "excellent" (best possible final grade). The thesis is published as a book in the Adaptive Computing and Machine Learning Series of The MIT Press.

Visiting Fellow at the University of Bristol (February 1999 – April 1999)
Collaborative work Dr. C. Campbell and Prof. J. Shawe-Taylor on *Bayes point machines* and *PAC-Bayesian learning theory.*

Visiting Fellow at the Australian National University (February 2000 – April 2000)
Collaborative work Dr. R. C. Williamson on *PAC-Bayesian margin bounds* and *online-learning.*

Research Fellow at Darwin College Cambridge (October 2000 – October 2004)
Microsoft Research Fellow, interested in the philosophy of Science.

Professional Experience

Civil Service at the Hospital Angermünde (1993 – 1994)
Paramedic in the emergency room.

Undergraduate Research Assistant at the Technical University of Berlin (1996 – 1997)
Non-teaching position in the project "Modern Methods of Machine Learning". Implemented a fully functional logical learning system (≈30 000 LOC).

Teaching Assistant at the Technical University of Berlin (1997 – 2000)
Teaching of Statistics (tutorials), theory of Machine Learning (seminars) and business mathematics (lecture). Supervised Diploma thesis: "Yang Gang: A comparison of learning algorithms on automobile recording data. 1999".

Postdoc Researcher at Microsoft Research Cambridge (2000 – 2001)
Frequentist guarantees for Bayesian algorithms, Algorithmic luckiness, Usefulness of unlabelled data.

Researcher at Microsoft Research Cambridge (2001 – 2006)
Bayesian inference, Reinforcement learning, Application of Bayesian methods in computer games.

Co-leader of the Applied Games Group at Microsoft Research Cambridge (since 2006)
Application of Bayesian methods in computer games and decision games.

Invited Technical Talks

Bayesian Ranking (2006)
Learning Workshop, Snowbird, Utah, USA.

Bayesian Ranking: Who is Really the Best at Halo 2 (2005)
Microsoft TechFest, Redmond, USA.

Machine Learning Meets Computer Games (2004)
Microsoft TechFest, Redmond, USA.

Directions in Statistical Learning Theory (2004)
University College London, London, UK.

Learning Kernel Classifiers (2002)
Gatsby Computational Neuroscience Unit, London, UK.

Algorithmic Luckiness (2001)
Dagstuhl Workshop on Inference and Model Selection, Wadern, Germany.

Bayes Point Machines (2001)
Gatsby Computational Neuroscience Unit, London, UK.

Learning Theory and Kernel Practice (2000)
NIPS*2000 Workshop on New Perspectives in Kernel Methods, Breckenridge, Colorado, USA.

Bayes Point Machines — Estimating The Bayes Point in Kernel Space (1999)
Microsoft Research Cambridge, Cambridge, United Kingdom.

Classification on Proximities with LP and QP Machines (1999)
Royal Holloway University of London, Egham, United Kingdom.

Statistical Learning Theory and Its Relation To Parametric Statistics (1999)
University of Augsburg, Augsburg, Germany.

A Survey of Statistical Learning Theory with Applications in Support Vector Machines (1998)
University of Potsdam, Potsdam, Germany.

Segmentation of Images using Model Selection Techniques (1997)
Daimler–Chrysler Research, Berlin, Germany.

Teaching Experience

Machine Learning and Statistics Seminar (Winter 1997)
Decision tree learning, neural network learning, statistical decision theory, support vector machines, induction principles.

Business Mathematics I Lecture (Summer 1998)
Set theory, series and sequences, financial mathematics, analysis, optimisation theory.

Statistical Learning Theory Seminar (Summer 1998)
Neural networks, Decision tree learning, VC and PAC theory, support vector machines.

Linear Models Tutorials (Winter 1998)
Linear algebra, correlations, one– and multidimensional regression, diagnosis.

Advanced Methods in Statistics Seminar (Winter 1998)
Statistical visualisation, theory of maximum likelihood, censored data, robust statistics, bootstrap estimators.

Theory of Machine Learning Seminar (Summer 1999)
VC, PAC and luckiness theory, boosting, functional analysis, on-line learning.

Advanced Methods in Statistics Seminar (Summer 1999)
Statistical visualisation, theory of maximum likelihood, censored data, robust statistics, bootstrap estimators.

Business Mathematics I Lecture (Winter 1999)
Set theory, series and sequences, financial mathematics, analysis, optimisation theory.

Statistical Foundations of Machine Learning (Winter 1999)
Probability theory, uniform convergence in function spaces, PAC/VC-theory, PAC-Bayesian analysis, transduction.

Business Mathematics II Lecture (Summer 2000)
Linear algebra, flow control, Markov chains, linear programming.

Reviewing

Journal of Machine Learning Research
Special issue on "Feature selection"

Machine Learning
Special issues on "Kernel methods" and "Model selection"

NIPS 1999 - 2006
Advances in Neural Information Processing Systems

ICML 2002-2006
International Conference on Machine Learning

COLT 2000 - 2006
International Conference on Computational Learning Theory

ESANN 1999, 2004
European Symposium on Artificial Neural Networks

IJCAI 1999
International Joint Conference on Artifical Intelligence

ALT 2001
International Conference on Algorithmic Learning Theory

CVPR 2003
International Conference on Vision and Pattern Recognition

IEEE AC
IEEE Transactions on Automation and Control

IEEE PAMI
IEEE Transactions on Pattern Analysis and Machine Intelligence

IEEE NN
IEEE Transactions on Neural Networks

IEEE SCM
IEEE Systems, Man and Cybernetics

Advances in Large Margin Classifiers 2000
Proceedings of the NIPS*99 workshop "Advances in Large Margin Classifiers"

Annals of the Institute of Statistical Mathematics
Special Issue on "Mathematics of Learning"

Neural Computation

Neurocomputing

Services

NIPS workshop "Learning on Relational Data Representations" (1998)
Co-organisation with K. Obermayer and T. Graepel.

NIPS workshop "Using Unlabelled Data For Supervised Learning" (1999)
Co-organisation with K. Obermayer and T. Graepel.

NIPS workshop "Learning To Rank" (2005)
Co-organisation with S. Agarwal and C. Cortes.

Program committee member COLT 2002, 2005, 2006
Fifteenth Annual Conference on Computational Learning Theory, Sydney.

Action Editor, Journal of Machine Learning Research
Section: Kernel Methods

Track Co-chair ICANN/ICONIP 2003
Special Track on "Support Vector Machines and Kernel Methods".

References

[1] S. Agarwal, T. Graepel, R. Herbrich, S. Har-Peled, and D. Roth. Generalization error bounds for the area under the ROC curve. *Journal of Machine Learning Research*, 6:393–425, 2005.

[2] S. Agarwal, T. Graepel, R. Herbrich, and D. Roth. A large deviation bound for the area under the ROC curve. In *Advances in Neural Information Processing Systems 17*, pages 9–16, 2005.

[3] T. Graepel, M. Goutrie, M. Krüger, and R. Herbrich. Learning on graphs in the game of Go. In *Proceedings of the Ninth International Conference on Artificial Neural Networks*, pages 347–352, 2001.

[4] T. Graepel and R. Herbrich. The kernel Gibbs sampler. In *Advances in Neural Information Processing Systems 13*, pages 514–520, 2001.

[5] T. Graepel and R. Herbrich. A pac-bayesian margin distribution bound for kernel classifiers (extended abstract), 2001. Presented at "New Directions in Kernel Based Learning Methods" NIPS*2001 workshop.

[6] T. Graepel and R. Herbrich. Invariant pattern recognition by semidefinite programming machines. In *Advances in Neural Information Processing Systems 16*, pages 33–40, 2004.

[7] T. Graepel, R. Herbrich, P. Bollmann-Sdorra, and K. Obermayer. Classification on pairwise proximity data. In *Advances in Neural Information Processing Systems 11*, pages 438–444, 1999.

[8] T. Graepel, R. Herbrich, and J. Gold. Learning to fight. In *Proceedings of the International Conference on Computer Games: Artificial Intelligence, Design and Education*, 2004. Winner of Best Paper Prize.

[9] T. Graepel, R. Herbrich, A. Kharechko, and J. Shawe-Taylor. Semidefinite programming by perceptron learning. In *Advances in Neural Information Processing Systems 16*, pages 457–464, 2004.

[10] T. Graepel, R. Herbrich, and K. Obermayer. Bayesian transductive classification by maximizing volume in version space. In *Proceedings of Learning 1999 Conference*, 1999.

[11] T. Graepel, R. Herbrich, and K. Obermayer. Bayesian Transduction. In *Advances in Neural Information Processing Systems 12*, pages 456–462, 2000.

[12] T. Graepel, R. Herbrich, B. Schölkopf, A. Smola, P. Bartlett, K. Robert-Müller, K. Obermayer, and R. Williamson. Classification on proximity data with LP–machines. In *Proceedings of the Ninth International Conference on Artificial Neural Networks*, pages 304–309, 1999.

[13] T. Graepel, R. Herbrich, and J. Shawe-Taylor. Generalisation error bounds for sparse linear classifiers. In *Proceedings of the Thirteenth Annual Conference on Computational Learning Theory*, pages 298–303, 2000.

[14] T. Graepel, R. Herbrich, and J. Shawe-Taylor. PAC-Bayesian compression bounds on the prediction error of learning algorithms for classification. *Machine Learning*, 59:55–76, 2005.

[15] T. Graepel, R. Herbrich, and R. C. Williamson. From margin to sparsity. In *Advances in Neural Information Processing Systems 13*, pages 210–216, 2001.

[16] A. Gretton, A. Doucet, R. Herbrich, P. J. W. Rayner, and B. Schölkopf. Support Vector regression for black-box system identification. In *11th IEEE Workshop on Statistical Signal Processing*, 2001.

[17] A. Gretton, R. Herbrich, O. Chapelle, B. Schölkopf, and P. J. W. Rayner. Leave-one-out error estimates for ν-svms (extended abstract), 2001. Presented at "New Directions in Kernel Based Learning Methods" NIPS*2001 workshop.

[18] A. Gretton, R. Herbrich, and A. Smola. The kernel mutual information. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 2003.

[19] A. Gretton, R. Herbrich, A. Smola, O. Bousquet, and B. Schölkopf. Kernel methods for measuring independence. *Journal of Machine Learning Research*, 6:2075–2129, 2005.

[20] A. Gretton, A. Smola, O. Bousquet, R. Herbrich, A. Belitski, M. Augath, Y. Murayama, J. Pauls, B. Schölkopf, and N. Logothetis. Kernel constrained covariance for dependence measurement. In *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, pages 112–119, 2005.

[21] E. Harrington, R. Herbrich, J. Kivinen, J. C. Platt, and R. C. Williamson. Online bayes point machines. In *Proceedings of the Seventh Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 241–252, 2003.

[22] R. Herbrich. Segmentierung mit Gaborfiltern zur Induktion struktureller Klassifikatoren auf Bilddaten. Master's thesis, Technical University of Berlin, 1997. in german.

[23] R. Herbrich. Exact tail bounds for binomial distributed variables, 1999.

[24] R. Herbrich. *Learning Linear Classifiers — Theory and Algorithms*. PhD thesis, Technische Universität Berlin, 2000.

[25] R. Herbrich. *Learning Kernel Classifiers: Theory and Algorithms*. MIT Press, 2002.

[26] R. Herbrich and T. Graepel. Large scale Bayes point machines. In *Advances in Neural Information Processing Systems 13*, pages 528–534, 2001.

[27] R. Herbrich and T. Graepel. A PAC-Bayesian margin bound for linear classifiers: Why SVMs work. In *Advances in Neural Information Processing Systems 13*, pages 224–230, 2001.

[28] R. Herbrich and T. Graepel. A PAC-Bayesian margin bound for linear classifiers. *IEEE Transactions on Information Theory*, 48(12):3140–3150, 2002.

[29] R. Herbrich and T. Graepel. Introduction to the special issue on learning theory. *Journal of Machine Learning Research*, 4:755–757, 2003.

[30] R. Herbrich, T. Graepel, P. Bollmann-Sdorra, and K. Obermayer. Learning a preference relation in IR. In *Proceedings Workshop Text Categorization and Machine Learning, International Conference on Machine Learning 1998*, pages 80–84, 1998.

[31] R. Herbrich, T. Graepel, P. Bollmann-Sdorra, and K. Obermayer. Supervised learning of preference relations. In *Proceedings Fachgruppentreffen Maschinelles Lernen*, pages 43–47, 1998.

[32] R. Herbrich, T. Graepel, and C. Campbell. Bayes point machines: Estimating the Bayes point in kernel space. In *Proceedings of IJCAI Workshop Support Vector Machines*, pages 23–27, 1999.

[33] R. Herbrich, T. Graepel, and C. Campbell. Bayesian learning in reproducing kernel Hilbert spaces. Technical report, Technical University of Berlin, 1999.

[34] R. Herbrich, T. Graepel, and C. Campbell. Robust Bayes Point Machines. In *Proceedings of ESANN 2000*, pages 49–54, 2000.

[35] R. Herbrich, T. Graepel, and C. Campbell. Bayes point machines. *Journal of Machine Learning Research*, 1:245–279, 2001.

[36] R. Herbrich, T. Graepel, and K. Obermayer. Regression models for ordinal data: A machine learning approach. Technical report, Technical University of Berlin, 1999.

[37] R. Herbrich, T. Graepel, and K. Obermayer. Support Vector learning for ordinal regression. In *Proceedings of the Ninth International Conference on Artificial Neural Networks*, pages 97–102, 1999.

[38] R. Herbrich, T. Graepel, and K. Obermayer. Large margin rank boundaries for ordinal regression. *Advances in Large Margin Classifiers*, pages 115–132, 2000.

[39] R. Herbrich, T. Graepel, and J. Shawe-Taylor. Sparsity vs. large margins for linear classifiers. In *Proceedings of the Thirteenth Annual Conference on Computational Learning Theory*, pages 304–308, 2000.

[40] R. Herbrich and E. Heymann. Animation von flexiblen Objekte. Technical report, Technical University of Berlin, 1997. in german.

[41] R. Herbrich, M. Keilbach, T. Graepel, P. Bollmann-Sdorra, and K. Obermayer. Neural networks in economics: Background, applications, and new developments. *Advances in Computational Economics*, 11:169–196, 1999.

[42] R. Herbrich and T. Minka. A multiple testing problem by jaynes, 2003. This note gives the solution to a crucial exercise in T. Jaynes book "Probability Theory: The Logic of Science".

[43] R. Herbrich and T. Scheffer. Generation of task specific segmentation procedures as a model selection task. In *Proceedings of the Visual Information Processing Workshop*, pages 11–24, 1997.

[44] R. Herbrich and J. Weston. Adaptive margin Support Vector Machines for classification learning. In *Proceedings of the Ninth International Conference on Artificial Neural Networks*, pages 880–885, 1999.

[45] R. Herbrich and R. C. Williamson. Algorithmic luckiness. *Journal of Machine Learning Research*, 3:175–212, 2002.

[46] R. Herbrich and R. C. Williamson. Algorithmic luckiness. In *Advances in Neural Information Processing Systems 14*, pages 391–397, 2002.

[47] R. Herbrich and R. C. Williamson. Learning and generalization: Theoretical bounds. *Handbook of Brain Theory and Neural Networks (2nd edition)*, pages 619–623, 2002.

[48] S. Hill, H. Zaragoza, R. Herbrich, and P. J. Rayner. Average precision and the problem of generalisation. In *Proceedings of the ACM SIGIR Workshop on Mathematical and Formal Methods in Information Retrieval*, 2002.

[49] N. Lawrence, M. Seeger, and R. Herbrich. Fast sparse gaussian process methods: The informative vector machine. In *Advances in Neural Information Processing Systems 15*, pages 625–632, 2003.

[50] N. D. Lawrence and R. Herbrich. A sparse bayesian compression scheme - the informative vector machine (extended abstract), 2001. Presented at "New Directions in Kernel Based Learning Methods" NIPS*2001 workshop.

[51] Y. Li, H. Zaragoza, R. Herbrich, J. Shawe-Taylor, and J. Kandola. The perceptron algorithm with uneven margins. In *Proceedings of the International Conference of Machine Learning*, pages 379–386, 2002.

[52] S. Rajaram, T. Graepel, and R. Herbrich. Poisson-Networks: A model for structured point processes. In *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, pages 277–284, 2005.

[53] S. E. Robertson, S. Walker, H. Zaragoza, and R. Herbrich. Microsoft cambridge at trec 2002: Filtering task. In *Proceedings of the Ninth Text Retrieval Conference (TREC-9)*, pages 361–368, 2002.

[54] T. Scheffer and R. Herbrich. Unbiased assessment of learning algorithms. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 798–803, 1997.

[55] T. Scheffer, R. Herbrich, and F. Wysotzki. Efficient theta-subsumption based on graph algorithms. *Lecture Notes in Artifical Intelligence*, 1314:212–228, 1996.

[56] T. Scheffer, R. Herbrich, and F. Wysotzki. Graph based subsumption algorithms for machine learning. In *Proceedings zum Fachgruppentreffen Maschinelles Lernen*, pages 101–105, 1996.

[57] B. Schölkopf, R. Herbrich, and A. J. Smola. A generalized representer theorem. In *Proceedings of the Fourteenth Annual Conference on Computational Learning Theory*, pages 416–426, 2001.

[58] B. Schölkopf, R. Herbrich, A. J. Smola, and R. C. Williamson. A generalized representer theorem. Technical report, Royal Holloway, University of London, 2000.

[59] D. Stern, R. Herbrich, and T. Graepel. Bayesian pattern ranking for move prediction in the game of Go. In *Proceedings of the International Conference of Machine Learning*, 2006.

[60] J. Weston and R. Herbrich. Adaptive margin support vector machines. *Advances in Large Margin Classifiers*, pages 281–296, 2000.

[61] H. Zaragoza and R. Herbrich. The perceptron meets reuters (extended abstract), 2001. Presented at "Machine Learning for Text and Images" NIPS*2001 workshop.