Transactions on Information and Communications Technologies vol 19 © 1998 WIT Press, www.witpress.com, ISSN 1743-3517

Learning From Data: Building, Evaluating and Understanding Models

Andreas S. Weigend
Stern School of Business, New York University
www.stern.nyu.edu/~aweigend

Abstract of the Invited Technical Conference

Analyzing tick-by-tick market data in real time, uncovering trading styles and understanding their profit and risk characteristics, managing portfolios of thousands of securities, and flagging potential fraud in millions of daily transactions are some of the challenges now sweeping the financial industries. Knowledge discovery and data mining approaches address these challenges and try to extract previously unknown, valid and comprehensible knowledge from large data sets. These approaches have emerged from several historically disjoint communities that include artificial intelligence, neural networks, evolutionary programming, reinforcement learning, signal processing, decision science, statistics, econometrics, probabilistic modeling and computational learning.

This talk provides a conceptual framework that unifies the principles and methods of knowledge discovery and data mining. It emphasizes the importance of formulating meaningful questions, highlights the assumptions of different model classes, stresses the critical evaluation and comparison with established methods, offers techniques for understanding and interpreting the results, and discusses various sources of noise and uncertainty in predictive models. Expressed as the seven steps of modeling, this framework clarifies the few fundamental principles that underlie most new methods being proposed today.

To gain insight into the practical advantages and disadvantages of these methods, several real-world applications are presented, including: predicting the probability distribution of returns, evaluating density forecasts, revealing hidden states in the market, understanding and managing the risk of a portfolio, and optimizing and evaluating trading strategies on risk-adjusted performance measures.

Andreas S. Weigend, Leonard N. Stern School of Business, New York University

Andreas Weigend (PhD Physics 1991, Stanford University; Diplom Physik, Philosophie 1986, Universitaet Bonn) is an Associate Professor of Information Systems at the Stern School of Business, New York University. He came to Stern from the University of Colorado at Boulder where he had founded the Time Series Analysis Group, after working at Xerox PARC (Palo Alto Research Center) on knowledge discovery. At the Santa Fe Institute, he co-directed the Time Series Prediction Competition that led to the volume Time Series Prediction: Forecasting the Future and Understanding the Past (1994, Addison Wesley)

His research focuses on basic methodologies for modeling and extracting knowledge from data and their application across different disciplines. He develops and integrates ideas and analytical tools from statistics and information theory with neural networks and other machine learning paradigms. His approach, *basic science on real problems*, emphasizes the importance of rigorous evaluations of new methods in data mining. His recent work uses computational intelligence to extract and understand hidden states in financial markets, and to exploit this information to improve density predictions. He has published about one hundred articles in scientific journals, books and conference proceedings. He co-edited four books including Decision Technologies for Financial Engineering (1998, World Scientific).

Prof. Weigend received a Research Initiation Award by the National Science Foundation (NSF), a major grant by the Air Force Office of Scientific Research (AFOSR), a Junior Faculty Development Award by the University of Colorado and a NYU Curricular Development Challenge Grant for his innovative course Data Mining in Finance. This course covers the foundations of knowledge discovery, data mining, prediction and nonlinear modeling, as well as specific techniques including neural networks, graphical models, evolutionary programming and clustering techniques. It develops solutions to current problems in finance and includes integrated in-depth projects with major Wall Street firms.

Prof. Weigend is organizing the sixth international conference, Computational Finance, taking place at Stern on January 6-8, 1999. This conference, previously called Neural Networks in the Capital Markets, is an annual research forum for advanced computational applications in finance. He has given tutorials and short executive courses on time series analysis, volatility prediction, nonlinear modeling and decision making under uncertainty and consulted for a broad spectrum of firms ranging from financial boutiques to Goldman Sachs, J. P. Morgan, Morgan Stanley and Nikko Securities.