Importance of Being Learnable
Colloquium 1998 at RHUL

1   Introduction

A series of one-day meetings were organised by the research groups in the Department.of Computer Science at Royal Holloway, University of London. The meetings were devoted to 30th anniversary of Computer Science Department that was established in 1968. The last meeting - the Colloquium *Importance of Being Learnable* – a paraphrase from Oscar Wilde play – was also to celebrate a foundation of a new Computer Learning Research Centre (CLRC) - estblished in February 1998.

The programme included invited lectures on key research topics in the rapidly developing field of *Machine Learning and Inductive Inference.*

Speakers included
Jorma Rissanen
Ray Solomonoff
Vladimir Vapnik
Chris Wallace
Alexei Chervonenkis.

2   Opening Address

Professor Alex Gammerman, Head of Department of Computer Science, Royal Holloway, University of London

'Today is a very special event, today is the 30th anniversary of our Department. What, perhaps, even more significant, we are celebrating today an establishment of new Centre of Machine Learning Research entitled *Computer Learning Research Centre - CLRC*. It was created in February this year and we would like to celebrate this in style. We have gathered together 5 great researchers, the very people who discovered new approaches, new paradigms in inductive inference and learning theory.

Back in the 1960's, exactly at the time when our Department was set up, R. Solomonoff published his papers on what is now known as algorithmic complexity. The idea was to provide a formal definition of simplicity used in Ockham

razor. Ray (and Kolmogorov and Chaitin) showed that simplicity can be measured by the length of the shortest program for a universal Turing machine. There are many possible "shortest" programs, but they differ in length by at most a constant. This is a beautiful result. The problem, however, is that it is impossible to compute "the length" - it is a non-computable function.

Fortunately, Chris Wallace and Jorma Rissanen developed approximate measures - MML and MDL principles, that can be used instead and have produced excellent results in practice. More importantly, perhaps, is that MML and MDL are new inductive principles, and these were very significant discoveries.

At the same time, in the 1960's, Vladimir Vapnik and Alexei Chervonenkis started the development of another induction principle - what is now known as Structured Risk Minimalisation (SRM) principle. This eventually resulted in the opening of new fields that include VC-theory, capacity control and Support Vector Machine (SVM).

We are very happy that the people who started this revolution in the theory of learning and inductive inference, back in the 60's, are here today, our speakers. Some of today's speakers are members of our Department, some of them are Visiting Professors in our Department, and all of them now are Fellows of CLRC. The Department and CLRC are recognised as one of the world leaders in the field of machine learning . We are privileged to work together and the simple fact that these distinguished speakers are here today is a reflection of our strength and puts us in a unique position among other Computer Science departments in this country and elsewhere. Some of the messages I have already received call this event an "historic" event.

I strongly believe that today's Colloquium will be remembered as a very important step in the development of modern induction theory, and the theory of learning. I hope you will enjoy the event."

3   Presentations

Professor Jorma Rissanen, Almaden Research Center, IBM Research Division, San Jose, CA, U.S.A.

MDL SELECTION OF REGRESSOR VARIABLES

Abstract: The idea with the MDL (for Minimum Description Length) principle is to calculate the shortest code length in a probabilistic sense of the observed data set, relative to a suggested class of probability models, archaically called 'hypotheses', which then can be used as a yardstick by which to judge the goodness of the class. Such a shortest code length, called stochastic complexity,

defines a universal model representing the entire class, which incorporates the optimal or maximum likelihood model.

This talk outlined the theory of MDL and stochastic complexity. As an application, the linear least squares regression problem was discussed. In particular, there was discussion on how to find the best, in the MDL sense, wavelet basis for data, as well as how to smoothen a data sequence by removing the MDL-defined noise from it.

Professor Vladimir Vapnik, AT&T Labs - Research, U.S.A. and Professor of Computer Science and Statistics, Royal Holloway, University of London

LEARNING THEORY AND PROBLEMS OF STATISTICS

Abstract: In the last three years, methods for constructing a new type of universal learning machine were developed based on results obtained in statistical learning theory. In contrast to traditional learning techniques, these novel methods do not depend explicitly on the dimensionality of input spaces and can control the generalization ability of machines that use different sets of decision functions in high dimensional spaces.

Along with a survey of the main ideas of statistical learning theory and a comparison of these ideas with the classical statistical approaches, the principles for constructing this type of machine were discussed.

Professor Alexei Chervonenkis, Institute of Control Sciences, Moscow

THE HISTORY OF THE SUPPORT VECTOR METHOD **Abstract**:

This talk reviewed developments over the last 30 years in the field of statistical learning theory. In particular, the following topics were addressed: "Generalized portrait" as a minimax vector that characterizes a class in pattern recognition cases. Converting the problem of "generalized portrait" search to a convex programming problem. Support vectors and their properties implied by the Kuhn-Tucker theorem. Evaluation of generalization performance using the number of support vectors.

Professor Chris Wallace, Univ of Monash, Victoria, Australia: and Visiting Professor in the Dept of Computer Science, Royal Holloway, University of London

BREVITY IS THE SOUL OF SCIENCE

Abstract: By "science" I mean the systematic study of natural phenomena, as carried out by a continuing society. That is, rather than looking at the actions or ideas of individuals, I am concerned with the overall growth and changes in a society's beliefs about natural phenomena, and what characterizes the theories and beliefs which gain long-term general acceptance.

There is ample evidence that a well-accepted scientific theory has great explanatory power. Much of what is observed is deducible from the theory, so the observed data may be reconstructed from a statement of the theory and a detailing of the undeducible residuum. Typically, the statement of theory and residuum is far shorter than a plain statement of the data.

Most accounts of "scientific method" emphasize predictions and the experimental checking of these. But predictive ability (although often justifying a society's investment in science) is not the touchstone of scientific enquiry. Quite accurate prediction of eclipses was achieved by several societies with little understanding of their cause, and as is shown by Solomonoff, the best possible prediction explicitly avoids commitment to any theory. In fact, however, human science seeks understanding, usually couched as a theory, and prediction is an economically-valuable by-product more often than the primary goal.

If we turn this descriptive account of science as finding concise, theory-based explanations of observed data into a prescription for how to handle data, we get various flavours of inductive inference techniques based on measures of information: Solomonoff prediction, MML and MDL theory selection, MML and BIC estimation. The differences reflect independent and near-simultaneous development, and are in practice far less important than the similarities. They use Shannon and Kolmogorov measures of "information". Formally, if not always philosophically, they are closer to Bayesian statistical inference than to classical methods. All rely on a fundamental trade-off between complexity of theory and fit to data encapsulated in the length of a message stating both theory and data.

By contrast, the Vapnik-Chervonenkis approach seems to come from a different world, in philosophy and form. It is in a sense an extension of classical statistical reasoning about "confidence" into a wider model space, bringing (like MML but by totally different means) both estimation and model selection under the same umbrella. Yet both these approaches demonstrably (and provably) work very well, having far more general application and giving consistently better results than previous methods of statistical and inductive inference. A fascinating open question is whether there is a deep relationship between the two approaches leading to their high degree of agreement in practical application.

Professor Ray Solomonoff, Oxbridge Research Inc, Cambridge, Mass, U.S.A:

HOW TO TEACH A MACHINE

Abstract: We now have many effective techniques for machine learning and they have been applied to a wide variety of problem areas. Efficient ways to use these techniques to bring an initially naive machine to a state of high intelligence was discussed.