

# Recent Breakthroughs in Minimum Description Length Learning

ICML/UAI/COLT Workshop

Helsinki, 9 July 2008

Location: University of Helsinki, Main building, Fabianinkatu 33 – Room S7, 3rd floor  
Inquiries: Tim.van.Erven@cwi.nl

## Motivation

During the last few years (2004–2007), there have been several breakthroughs in the area of Minimum Description Length (MDL) modeling, learning and prediction. These breakthroughs concern the efficient computation and proper formulation of MDL in parametric problems based on the “normalized maximum likelihood”, as well as altogether new, and better, coding schemes for nonparametric problems. This essentially solves the so-called AIC-BIC dilemma, which has been a central problem in statistical model selection for more than 20 years now. The goal of this workshop is to introduce these exciting new developments to the ML and UAI communities, and to foster new collaborations between interested researchers.

Most new developments that are the focus of this workshop concern efficient (in many cases, linear-time) algorithms for theoretically optimal inference procedures that were previously thought not to be efficiently solvable. It is therefore hoped that the workshop will inspire original practical applications of MDL in machine learning domains. Development of such applications recently became a lot easier, because of the new (2007) book on MDL by Peter Grünwald, which provides the first comprehensive overview of the field, as well as in-depth discussions of how it relates to other approaches such as Bayesian inference. Remarkably, the originator of MDL, Jorma Rissanen, also published a new monograph in 2007; and a Festschrift in Honor of Rissanen’s 75th birthday was presented to him in May 2008.

## Program

09:00 – 10:10	P. Grünwald	MDL tutorial
10:10 – 10:30		Questions and discussion
10:30 – 11:00		Coffee break
11:00 – 12:30	P. Myllymäki S. de Rooij J. Ojanen	Fast computation of NML for Bayesian networks Nonparametric density estimation by switching Extensions to MDL denoising
12:30 – 14:30		Lunch break
14:30 – 16:00	T. Silander T. Zhang I. Täbuş	Sequential and factorized NML models Generalization Theory of Two-part Code MDL Estimator Normalized maximum likelihood models in genomics
16:00 – 16:30		Coffee break
16:30 – 17:00	M. Seeger	Information Consistency of Nonparametric Gaussian Process Methods
17:00 – 17:30		Panel discussion (with possible extension to 18:00)
18:00 – 20:00		Reception by the University of Helsinki

Workshop organizers: Tim van Erven, Peter Grünwald, Petri Myllymäki, Teemu Roos and Ioan Tabus



This workshop is being supported by the IST Programme of the European Community, under the PASCAL-2 Network of Excellence. This publication only reflects the authors’ views.