

## **GREAT08 Challenge**

1 March - 30 December 2008

The GRavitational IEnsing Accuracy Testing 2008 (GREAT08) Challenge focuses on a problem that is of crucial importance for future observations in cosmology. The shapes of distant galaxies can be used to determine the properties of dark energy and the nature of gravity, because light from those galaxies is bent by gravity from the intervening dark matter. The observed galaxy images appear distorted, although only slightly, and their shapes must be precisely disentangled from the effects of pixelisation, convolution and noise.

## **Causality Challenge #1: Causation and Prediction Challenge**

15 December 2007 - 30 April 2008

The focus of this challenge is on predicting the results of actions performed by an external agent. Examples of that problem are found, for instance, in the medical domain, where one needs to predict the effect of a drug prior to administering it, or in econometrics, where one needs to predict the effect of a new policy prior to issuing it. We focus on a given target variable to be predicted (e.g. health status of a patient) from a number of candidate predictive variables (e.g. risk factors in the medical domain). Under the actions of an external agent, variable predictive power and causality are tied together. For instance, both smoking and coughing may be predictive of lung cancer (the target) in the absence of external intervention; however, prohibiting smoking (a possible cause) may prevent lung cancer, but administering a cough medicine to stop coughing (a possible consequence) would not.

## **Human-machine comparisons of consonant recognition in noise challenge**

1 December 2007 - 31 March 2008

Listeners outperform automatic speech recognition systems at every level of speech recognition, including the very basic level of consonant recognition. What is not clear is where the human advantage originates. Does the fault lie in the acoustic representations of speech or in the recogniser architecture, or in a lack of compatibility between the two? There have been relatively few studies comparing human and automatic speech recognition on the same task, and, of these, overall identification performance is the dominant metric. However, there are many insights which might be gained by carrying out a far more detailed comparison.

The purpose of this challenge is to promote focused human-computer comparisons on a task involving consonant identification in noise, with all participants using the same training and test data. Training and test data and native listener and baseline recogniser results will be provided by the organisers, but participants are encouraged to also contribute listener responses.

## **Graph Labelling Challenge: Application to Web Spam Detection Challenge**

1 January 2007 - 1 January 2008

The goal of the Web Spam Challenge series is to identify and compare Machine Learning (ML) methods for automatically labeling structured data represented as graphs. More precisely, we focus on the problem of labeling all nodes of a graph from a partial labeling of them. The application we study is Web Spam Detection, where we want to detect deliberate actions of deception aimed at the ranking functions used by search engines.

## **Visual Object Classes Challenge 2007**

1 January - 31 October 2007

The goal of this challenge is to recognise objects from a number of visual object classes in realistic scenes (i.e. not pre-segmented objects). It is fundamentally a supervised learning problem in that a training set of labelled images is provided. The twenty object classes that have been selected are:

Person: person

Animal: bird, cat, cow, dog, horse, sheep

Vehicle: aeroplane, bicycle, boat, bus, car, motorbike, train

Indoor: bottle, chair, dining table, potted plant, sofa, tv/monitor

## **Unsupervised Morpheme Analysis -- Morpho Challenge 2007**

1 October 2006 - 21 September 2007

The objective of the Challenge is to design a statistical machine learning algorithm that discovers which morphemes (smallest individually meaningful units of language) words consist of. Ideally, these are basic vocabulary units suitable for different tasks, such as text understanding, machine translation, information retrieval, and statistical language modeling.

The scientific goals are:

To learn of the phenomena underlying word construction in natural languages

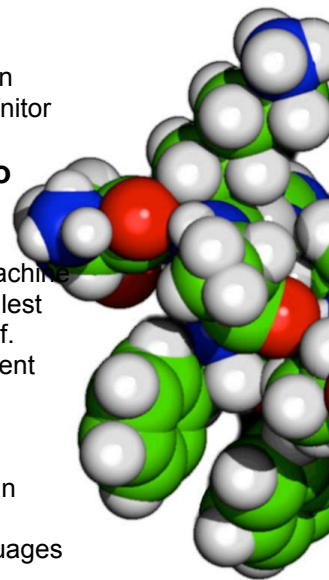
To discover approaches suitable for a wide range of languages

To advance machine learning methodology

## **Agnostic Learning vs. Prior Knowledge Challenge**

1 October 2006 - 1 August 2007

"When everything fails, ask for additional domain knowledge" is the current motto of machine learning. Therefore, assessing the real added value of prior/domain knowledge is a both deep and practical question. Most commercial data mining programs accept data pre-formatted as a table, each example being encoded as a fixed set of features. Is it worth spending time engineering elaborate features incorporating domain knowledge and/or designing ad hoc algorithms? Or else, can off-the-shelf programs working on simple features encoding the raw data without much domain knowledge put out-of-business skilled data analysts?



In this challenge, the participants are allowed to compete in two tracks:  
 The “prior knowledge” track, for which they will have access to the original raw data representation and as much knowledge as possible about the data.  
 The “agnostic learning” track for which they will be forced to use a data representation encoding the raw data with dummy features.

### Third Recognising Textual Entailment Challenge

1 December 2006 - 1 June 2007

RTE 3 follows the same basic structure of the previous campaign, in order to facilitate the participation of newcomers and to allow "veterans" to assess the improvements of their systems in a comparable test exercise. Nevertheless, the following innovations are introduced to make the challenge more stimulating and, at the same time, to encourage collaboration between system developers:  
 a limited number of longer texts, i.e. up to a paragraph- in order to move toward more comprehensive scenarios which incorporate the need for discourse analysis. However, the majority of examples will remain similar to those in the previous challenges, providing pairs with relatively short texts.  
 an RTE Resource Pool has been created where contributors have the possibility to share the resources they use.  
 an optional task, "Extending the Evaluation of Inferences from Texts", which explores two other tasks closely related to textual entailment: differentiating unknown from false/contradicts and providing justifications for answers.

### Learning when Test and Training Inputs have Different Distributions Challenge

1 June 2005 - 30 April 2007

The goal of this challenge is to attract the attention of the Machine Learning community towards the problem where the input distributions,  $p(x)$ , are different for test and training inputs. A number of regression and classification tasks are proposed, where the test inputs follow a different distribution than the training inputs. Training data (input-output pairs) are given, and the contestants are asked to predict the outputs associated to a set of validation and test inputs. Probabilistic predictions are strongly encouraged, though non-probabilistic "point" predictions are also accepted. The performance of the competing algorithms will be evaluated both with traditional losses that only take into account "point predictions" and with losses that evaluate the quality of the probabilistic predictions.

### Computer-Assisted Stematology Challenge

6 October 2006 - 14 April 2007

Stematology (a.k.a. stemmatics) studies relations among different variants of a document that have been gradually built from an original by copying and modifying earlier versions. The aim of such study is to reconstruct the family tree of the variants.

We invite applications of established and, in particular, novel approaches, including but of course not restricted to hierarchical clustering, graphical modeling, link analysis, phylogenetics, string-matching, etc.

The objective of the challenge is to evaluate the performance of various approaches. Several sets of variants for different texts are provided, and the participants should attempt to reconstruct the relationships of the variants in each data-set. This enables the comparison of methods usually applied in unsupervised scenarios.

### Type I and Type II Errors for Multiple Simultaneous Hypothesis Testing Challenge

1 January 2006 - 1 February 2007

Multiple Simultaneous Hypothesis Testing is a main issue in many areas of information extraction:

- rule extraction,
- validation of genes influence,
- validation of spatio-temporal patterns extraction (e.g. in brain imaging),
- other forms of spatial or temporal data (e.g. spatial collocation rule).
- other multiple hypothesis testing,

In all above frameworks, the goal is to extract patterns such that some quantity of interest is significantly greater than some given threshold.

### Letter-to-Phoneme Conversion Challenge

1 February 2006 - 31 January 2007

Letter-to-phoneme conversion is a classic problem in machine learning (ML), as it is both hard (at least for languages like English and French) and important. For non-linguists, a 'phoneme' is an abstract unit corresponding to the equivalence class of physical sounds that 'represent' the same speech sound. That is, members of the equivalence class are perceived by a speaker of the language as the 'same' phonemes: the word 'cat' consists of three phonemes, two of which are shared with the word 'bat'. A phoneme is defined by its role in distinguishing word pairs like 'bat' and 'cat'. Thus, /b/ and /k/ are different phonemes. But the /b/ in 'bat' and the /b/ in 'tab' are the same phoneme, in spite of their different acoustic realisations, because the difference between them is never used (in English) to signal a difference between minimally-distinctive word-pairs.

Although we intend to give most prominence to letter-to-phoneme conversion, the community is challenged to develop and submit innovative solutions to these related problems.

### Exploration vs. Exploitation Challenge

19 January - 15 November 2006

Touch Clarity ([www.touchclarity.com](http://www.touchclarity.com)) provides real time optimisation of websites. Touch Clarity chooses, from a number of options, the most popular content to display on a page. This decision is made by tracking how many visitors respond to each of the options, by clicking on them. This is a direct commercial application of the multi armed bandit problem - each of the