items which might be shown is a separate bandit, with a separate response rate.

As in the multi armed bandit problem, there is a trade off between exploration and exploitation - it is necessary to sometimes serve items other than the most popular in order to measure their response rate with sufficient precision to correctly identify which is the most popular. However, in this application there is a further complication - typically the rates of response to each item will vary over time, so continuous exploration is necessary in order to track this time variation, as old knowledge becomes out of date. An extreme example of this might be in choosing which news story to serve as the main story on a news page - interest in one story will decrease over time while interest in another will increase. In addition, the interest in several stories might vary in a similar, coherent way - for example a general increase in interest in sports stories at weekends, or in political stories near to an election. So there are typically two types of variation to consider - where response rates vary together, and where response rates vary completely independently.

Visual Object Classes Challenge

1 January - 30 June 2006

The Visual Object Classes Chellenges has the following objectives:

- To compile a standardised collection of object recognition databases
- To provide standardised ground truth object annotations across all databases
- To provide a common set of tools for accessing and managing the database annotations
- To run a challenge evaluating performance on object class recognition

Unsupervised Segmentation of Words into Morphemes Challenge

1 September 2005 - 12 April 2006

The objective of the Challenge is to design a statistical machine learning algorithm that segments words into the smallest meaning-bearing units of language, morphemes. 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

Second Recognising Textual Entailment Challenge

1 October 2005 - 10 April 2006

Textual Entailment Recognition has been proposed recently as a generic task that captures major semantic inference needs across many natural language processing applications, such as Question Answering (QA), Information Retrieval (IR), Information Extraction (IE), and (multi) document summarisation. This task requires to recognise, given two text fragments, whether the meaning of one text is entailed (can be inferred) from the other text.

By introducing a second challenge we hope to keep the momentum going, and to further promote the formation of a research community around the applied entailment task. As in the previous challenge, the main task is judging whether a hypothesis (H) is entailed by a text (T). One of the main goals for the RTE-2 dataset is to provide more "realistic" text-hypothesis examples, based mostly on outputs of actual systems. We focus on the four application settings mentioned above: QA, IR, IE and multi-document summarisation. Each portion of the dataset includes typical T-H examples that correspond to success and failure cases of such applications. The examples represent different levels of entailment reasoning, such as lexical, syntactic, morphological and logical.

XML Challenge

30 July 2005 - 1 April 2006

The objective of the challenge is to develop machine learning methods for structured data mining and to evaluate these methods for XML document mining tasks. The challenge is focused on classification and clustering for XML documents. Datasets coming from different XML collections and covering a variety of classification and clustering situations will be provided to the participants.

One goal of this track is to build a reference categorisation/ clustering corpora of XML documents. The organisers are open to any suggestion concerning the construction of such corpora.

Performance Prediction Challenge

1 October 2005 - 1 March 2006

This project is dedicated to stimulate research and reveal the state-of-the art in "model selection" by organising a competition followed by a workshop. Model selection is a problem in statistics, machine learning, and data mining. Given training data consisting of input-output pairs, a model is built to predict the output from the input, usually by fitting adjustable parameters. Many predictive models have been proposed to perform such tasks, including linear models, neural networks, trees, and kernel methods. Finding methods to optimally select models, which will perform best on new test data, is the object of this project. The competition will help identify accurate methods of model assessment, which may include variants of the well-known cross-validation methods and novel techniques based on learning theoretic performance bounds. Such methods are of great practical importance in pilot studies, for

which it is essential to know precisely how well desired specifications are met.

Inferring Relevance from Eye Movements Challenge

1 March - 1 September 2005

In the Challenge we have an experimental setup, where the test subject is first shown a question, followed by ten sentences. Five of the sentences are "relevant" to the question (they are of the same topic as the question) and five of the sentences are irrelevant (they have no relation to the topic of the question). One of the relevant sentences is the correct answer to the question. The experimental setup is designed to resemble a real-life information retrieval scenario as closely possible while at the same time retaining a controlled setup where the ground truth is known.

Thus, in the Challenge the meaning of "relevant" is defined in terms of this experimental setup. The objective of the Challenge is to find the best methods and features that can be used to predict the relevance from the eye movement measurements.

BCI Competition III Challenge

12 December 2004 - 22 May 2005

The goal of the "BCI Competition III" is to validate signal processing and classification methods for Brain-Computer Interfaces (BCIs). Compared to the past BCI Competitions, new challanging problems are addressed that are highly relevant for practical BCI systems, such as

session-to-session transfer (data set I)

small training sets, maybe to be solved by subject-to-subject transfer (data set IVa),

non-stationarity problems (data set IIIb, data set IVc),

multi-class problems (data set IIIa, data set V, data set II,),

classification of continuous EEG without trial structure (data set IVb, data set V).

Also this BCI Competition includes for the first time ECoG data (data set I) and one data set for which preprocessed features are provided (data set V) for competitors that like to focus on the classification task rather than to dive into the depth of EEG analysis.

PASCAL Ontology Learning Challenge

1 November 2004 - 30 April 2005

The aim of this challenge is to encourage work on automated construction and population of ontologies. For the purposes of this challenge, an ontology consists of a set of concepts and a set of instances. An instance can be assigned to one or more concepts. The concepts are connected into a hierarchy.

Several types of tasks are included in this challenge:

- Ontology construction: given a set of documents, construct an ontology with these documents as instances.
- Ontology extension: given a partial ontology and a set of instances, extend the ontology with new concepts using the given instances.
- Ontology population: given a partially populated hierarchy of concepts, develop a model that can assign new instances to concepts.
- Concept naming: given a set of instances and the assignment of instances to concepts, suggest user-friendly labels for the concepts.

Evaluation is based on comparing the results to a "golden standard" ontology prepared by human editors.

Recognising Textual Entailment Challenge

1 June 2004 - 10 April 2005

Recent years have seen a surge in research of text processing applications that perform semantic-oriented inference about concrete text meanings and their relationships. Even though many applications face similar underlying semantic problems, these problems are usually addressed in an application oriented manner. Consequently it is difficult to compare, under a generic evaluation framework, semantic methods that were developed within different applications. The PASCAL Challenge introduces textual entailment as a common task and evaluation framework for Natural Language Processing, Information Retrieval and Machine Learning researchers, covering a broad range of semantic-oriented inferences needed for practical applications. This task is therefore suitable for evaluating and comparing semantic-oriented models in a generic manner. Eventually, work on textual entailment may promote the development of generic semantic "engines", which will play an analogous role to that of generic syntactic analyzers across multiple applications.

101 Visual Object Classes Challenge

1 September 2004 - 31 March 2005

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 learning problem in that a training set of labelled images will be provided. The four object classes that have been selected are:

motorbikes
bicycles
people
cars

There will be two main competitions:

- For each of the 4 classes, predicting presence/absence of an example of that class in the test image.
- Predicting the bounding box and label of each object from the 4 target classes in the test image.

Contestants may enter either (or both) of these competitions, and can choose to tackle any (or all) of the four object classes. The challenge allows for two approaches to each of the competitions:

- Contestants may use systems built or trained using any methods or data excluding the provided test sets.
- Systems are to be built or trained using only the provided training data.

The intention in the first case is to establish just what level of success can currently be achieved on these problems and by