ppmls

Principles and Practice of Multiple Learning Systems
A Workshop of ECML 2007
Sept 17-21, Warsaw, Poland

ECML 2007 PKDD

Background

Multiple Learning Systems, aka multiple classifier systems, ensembles, information fusion, have emerged as one of the strongest pattern recognition techniques of the last decade, influencing every area of Machine Learning and Data Mining. Techniques like Boosting or Bagging, that create multiple classifiers then combine them by voting or averaging are becoming a staple part of the ML toolbox. All techniques in this family rely on the general principle of combining information from compatible sources, whether they be classifiers, regressors, clusters, kernels, proximity measures, or procedures. In empirical evaluations, MLS consistently give outstanding performance compared to more complex state-of-the-art learning methods.

On a deeper level, the "combining" principle has recently been extended in several new directions. Cluster ensembles, semi-supervised learning, changing environments, kernel combining methods are all utilizing this principle. In particular, this workshop aims to highlight these advanced uses of the "combining" principle, right across the spectrum of machine learning and data mining.

This is the first ECML workshop to address this important principle, aiming to cover both the deep theoretical questions and practical applications of the idea.

Workshop Goals

The goal of this workshop is to bring together researchers that work with the combining principle in a very broad sense, from the theory of classifier ensembles through to meta-learning such as the integration of complementary learning methodologies. We aim to provide a platform for exchanging ideas between people from different sub-fields of machine learning or data mining. We expect participants that focus on various learning problems such as classification, clustering, regression, semi-supervised learning, multi-task learning, novelty detection, data organisation or information fusion and use either theory or efficient heuristics. We aim to discuss the state-of-the-art techniques, ideas originating from various subfields of computational intelligence and to be representative of the remaining open problems in the field.

Theory

- Methods of classifier combination
- Diversity measures
- Classifier selection vs classifier combination
- Weak and strong models vs overtraining
- Handling noise and outliers

Multiple Representations

- Kernel combining and kernel selection
- Learning from multiple data sets
- Combining proximity measures
- Information fusion and sensor fusion

Paradigm combination

- Hybrid ensemble methods
- Integration of statistical and structural techniques
- Combination of generative and discriminative models

Learning schemes

- Learning classifier systems
- Rule ensembles
- Classifier selection vs classifier combination
- Boosting and bagging methods
- Feature subspaces / feature projections

Data flexibility

- MLS in changing/dynamic environments
- Cluster ensembles
- Stream mining
- Learning from distributed data
- Semi-supervised learning

Applications

- Image classification/retrieval
- Data mining
- Text classification
- Computer vision
- Biometric identification

We particularly welcome novel ideas on combination of complementary methodologies, understood in a broad sense, such as combination of generative and discriminative models or probabilistic and information theoretic models for learning. We also encourage the researchers to share their both positive and negative experience with the combining paradigm.