

Synergy of multi-label hierarchical ensembles, datafusion, and cost-sensitive methods for gene functional inference

Nicolo Cesa-Bianchi, Matteo Re and Giorgio Valentini

Mach. Learn.. In press

Functional genomics

- Multilabel classification.
- Structured functional classes.
- Multiple sources of data.
- Unbalanced classes.

Contributions

- Embed hierarchical constraints in multilabel prediction.
- Data fusion.
- Cost-sensitive techniques.
- Integrate them.

Experimental set-up

- Six different sources of yeast biomolecular data.
- Multilabel hierarchical vs. Flat methods.
- Cost-sensitive vs. Basic.
- Negative examples selection strategies.

Data fusion

- Kernel fusion.
- Ensemble weighted voting.

Hierarchical ensembles

- Hierarchical Top-Down ensemble (HTD).
- Hierarchical Bayesian ensemble (HBAYES).
- Hierarchical True Path Rule (TPR).

Cost-sensitive techniques

- HTD-CS, HBAYES-CS, TPR-W (Weighted).

Integration

1. Train a set of classifiers trained with multiple sources of data by kernel fusion methods.
2. Combine the predictions at each node to obtain the multilabel predictions using hierarchical multilabel methods (basic or CS.)