Structured Output Learning with A Random Sample of Spanning Trees

Hongyu Su

Helsinki Institute for Information Technology HIIT
Department of Information and Computer Science
Aalto University

November 18, 2014

Multilabel Classification

- Multilabel classification is an important research field in machine learning.
  - For example, a document can be classified as "science", "genomics", and "drug discovery".
  - Each input variable $\mathbf{x} \in \boldsymbol{\mathcal{X}}$ is simultaneously associated with multiple output variables $\mathbf{y} \in \boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{Y}} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_k$.
  - The goal is to find a mapping function that predicts the best values of an output given an input $f \in \mathcal{H} : \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{Y}}$.
- The central problems of multilabel classification:
  - The size of the output space $\boldsymbol{\mathcal{Y}}$ is exponential in the number of microlabels.
  - The dependency of microlabels needs to be exploited to improve the prediction performance.

Flat Multilabel Classification

- Multiple output variables are treated as a "flat" vector.
- It is difficult to take into consideration the correlation of labels.
- For example, ML-KNN, ADABOOST.MH, MTL, ...

Structured Output Learning

- There is an *output graph* connecting multiple labels.
  - A set of nodes corresponds to the multiple labels.
  - A set of edges represents the correlation between labels.
- Hierarchical classification:
  - The output graph is a rooted tree or a directed graph defining the different levels of granularities.
  - For example, SSVM, ...
- Graph labeling:
  - The output graph often takes a general form (e.g., a tree, a chain).
  - For example, $\mathrm{M}^3\mathrm{N}$, CRF, MMCRF, ...
- The output graph is assumed to be known *apriori*.

The Research Question

- The output graph is hidden in many applications.
  - For example, a surveillance photo can be tagged with "building", "road", "pedestrian", and "vehicle".
- We focus on the problem in structured output learning when the output graph is not observed.
- Our approach:
  - Assume the dependency can be modeled by a complete set of pairwise correlations.
  - Build a structured output learning model with a complete graph as the output graph.
  - Solve the optimization problem.

Contributions

- A structured output learning model which performs max-margin learning on a random sample of spanning tree.
- The model is not constrained to the availability of the output graph.
- The $\mathcal{NP}$-hard inference problem can be solved by a polynomial time algorithm with a condition guaranteeing the exact solution.
- The theoretical analysis and the empirical results verify the performance of the model.

Model

- The training examples are given in pair $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{m}$.
- Each example is mapped to a joint feature space $\phi(\mathbf{x}_i, \mathbf{y}_i)$.
- A compatibility score is defined as

$$F(\mathbf{x}_i, \mathbf{y}_i; \mathbf{w}) = \langle \mathbf{w}, \phi(\mathbf{x}_i, \mathbf{y}_i) \rangle$$

- $\mathbf{w}$ ensure an input $\mathbf{x}_i$ with a correct multilabel $\mathbf{y}_i$ achieves a higher score than with any incorrect multilabel $\mathbf{y} \in \mathcal{Y}$.
- The predicted output $\mathbf{y}_{\mathbf{w}}(\mathbf{x})$ for a given input is computed by

$$\mathbf{y}_{\mathbf{w}}(\mathbf{x}) = \underset{\mathbf{y} \in \mathcal{Y}}{\mathbf{argmax}}\, F(\mathbf{x}, \mathbf{y}; \mathbf{w}),$$

  also called *inference problem*.