Dual-Decomposed Learning with Factorwise Oracles for Structural SVMs of Large Output Domain

Ian E.H. Yen1, Xiangru Huang2, Kai Zhong2, Ruohan Zhang2, Pradeep Ravikumar2 and Inderjit S. Dhillon2.
1 Carnegie Mellon University. 2 University of Texas at Austin.

Abstract

- Many applications of machine learning involve structured outputs with large domains, such as Translation, Alignment, and Parsing.
- Learning of a structured predictor is prohibitive due to repetitive calls to an expensive inference oracle.
- We propose decomposing training of a structural SVM into factorwise multiclass SVMs connected with messages, replacing structured oracles with factorwise oracles.
- The proposed algorithm, Greedy Direction Method of Multiplier (GDMM), guarantees 𝜖-suboptimality in 𝑂(1/ε) iterations, and shows orders-of-magnitude speedup over state of the art on large-domain problems.

Structured Prediction of Large Output Domain

- We consider Structured Predictor of the form

\[ h(x; w) = \arg \max_{y \in Y(x)} \langle w, \phi(x, y) \rangle. \]

- For Structural SVM, we use the structured hinge loss

\[ L(w; x, y) = \max_{y \in Y(x)} \langle w, \phi(x, y) \rangle - \langle w, \phi(x, y) \rangle + \delta(y, \tilde{y}), \]

where the inner product allows factor decomposition of the form

\[ \langle w, \phi(x, y) \rangle = \sum_{f \in F} \sum_{t \in \mathcal{F}(x)} \langle w_f, \phi_f(x_t, y_t) \rangle, \]

and \( \delta(y, \tilde{y}) \) is a task-dependent error function (usually Hamming Error).

Convergence Analysis

Let \( d(\lambda) = \min_{\lambda \in [0,1]} \mathcal{L}(\alpha, \lambda) \) and \( \Delta^i_j := d(\lambda^i_j) - d(\lambda^i_j) \) be the dual and primal suboptimality respectively. The GDMM algorithm has

\[ E[\Delta^i_{j+1} + \Delta^i_{j}] \leq \epsilon \text{ for } t \geq \omega \log(\frac{1}{\epsilon}) \]

for some constant \( \omega > 0 \).

Experiments

- Sequence Labeling: POS (|Y| = 45), ChineseOCR (|Y| = 3039).
- Structural oracle uses Viterbi Algorithm.
- Multilabel with Pairwise Interaction: RCV1 (|Y| = 228).
- Structural oracle solves a Linear Program.