Output Space Search for Structured Prediction
Janardhan Rao (Jana) Doppa, Alan Fern, and Prasad Tadepalli

Structured Prediction

- **Given**: a set of structured input-output pairs of the form \((x, y)\)
  - Handwriting recognition
  - Image labeling
- **Learn**: a function \(F: X \rightarrow Y\) to make predictions on new inputs
- **Evaluation**: against a loss function \(L(y, F(x)) \in [0, 1]\)
  - Hamming loss, F1 score, B3 score ...

Search Spaces over Complete Outputs

- Leverage powerful recurrent classifiers to automatically define search spaces over complete outputs
- **Recurrent Classifier**
  - **Key idea**: use previous label(s) as input features for predicting the current label
  - Training is performed via exact imitation
- **Limited Discrepancy Search** [Harvey and Ginsberg, 1995]
  - **Key idea**: correct the response of recurrent classifier at a small no. of critical errors to produce high-quality outputs

Output Space Search Framework

- **Key Elements**:
  - Search space over complete outputs
  - Cost function to evaluate complete input-output pairs
  - “Rank-based” search procedure
- **Main idea**: run search procedure for a specified time bound, return the least cost output that is uncovered
- **Illustration with greedy search**

Cost Function Learning

- Assumes a fixed “rank-based” search procedure
- **Key idea**: learning for search in the framework of imitation learning
  - Generate ranking examples to imitate searches conducted by loss function and learn a cost function via a rank learning algorithm
- **Illustration**: Ranking examples for greedy search

Experimental Results

- **Greedy search**: the most basic instantiation of our framework
  - Performance of LDS is superior to the state-of-the-art SP algorithms
  - Demonstrates the efficiency of LDS
- **Results improve with additional search**
- **LDS vs. Flip-bit**: LDS has significantly better performance and anytime behavior

Properties

- Anytime predictions
- Minimal restrictions on the complexity of cost function

Search in the space of discrepancy sets from small to large

Flip-bit Search Space

- Initialize with output of recurrent classifier and search in the space of single label changes (“bit flips”)
- Similar to search space underlying Gibbs sampling

Experimental Results

- **ALGORITHMS**
  - LDS-Greedy
  - Flip-Greedy
- **DATASETS**
  - Handwriting (HW)
  - Image Labeling (Image)
  - Scene Labeling (Scene)
- **Performance of LDS is superior to the state-of-the-art SP algorithms**
  - Demonstrates the efficiency of LDS
- **Results improve with additional search**
- **LDS vs. Flip-bit**: LDS has significantly better performance and anytime behavior

Performance improves with higher-order cost functions