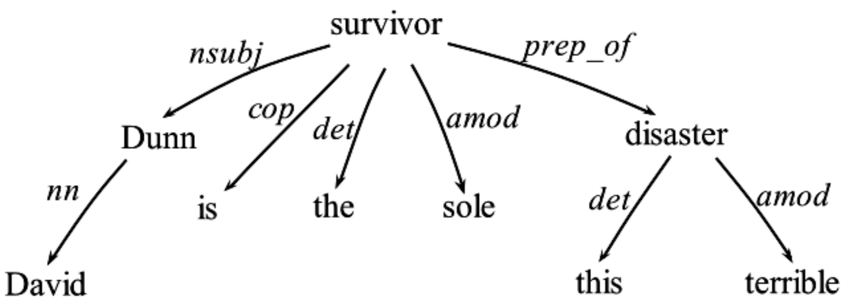
An Integer Linear Programming Framework for Mining Constraints from Data

Tao Meng, Kai-Wei Chang

University of California, Los Angeles

Constraints in Structured Output Predictions

- Many ML tasks involve structured labels that follow underlying constraints
- Our goal: Mine these constraints from input-label pairs



Structural: Tree

- Flower
- Tiger
- Car
- Plant
- Phone

Label Dependency

Constraint: At Least 1 Verb

An Integer Linear Programming (ILP) framework

$\boldsymbol{y}$: structured output labels

$$\min_{\boldsymbol{y} \in \mathbb{Z}^d} \boldsymbol{w} \cdot \boldsymbol{y} \qquad s.t. \;\; A\boldsymbol{y} \leq \boldsymbol{b}$$

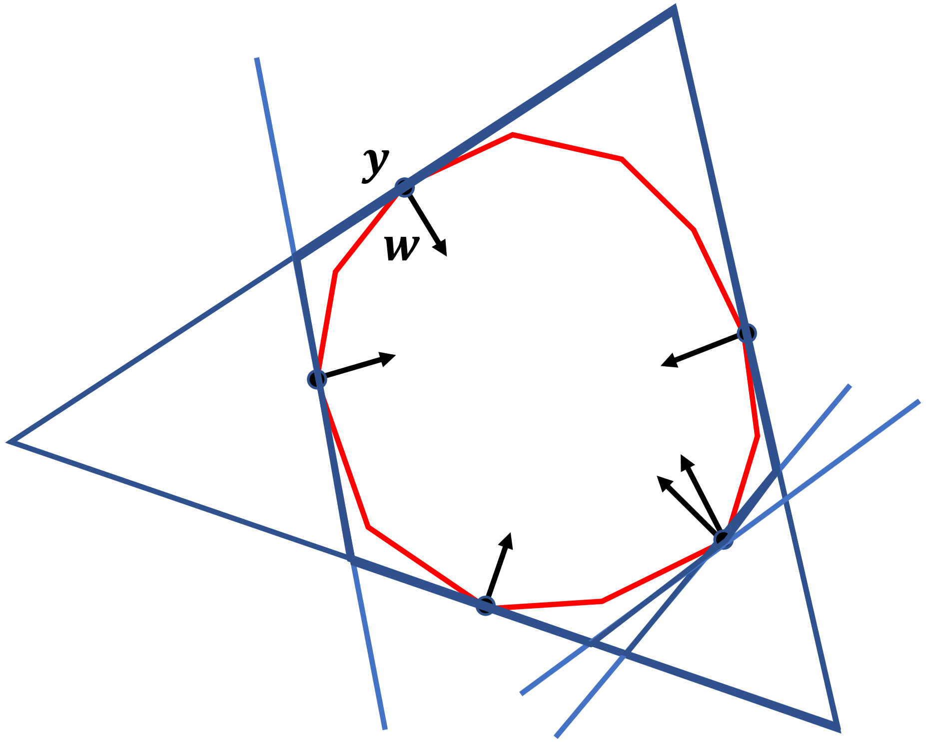
$\boldsymbol{w}$: weight of the objective function
(e.g., scores from local classifiers)

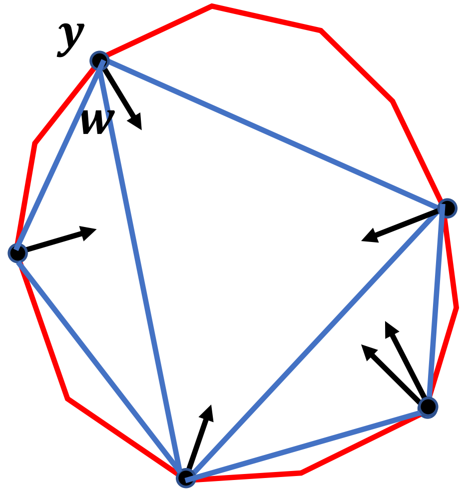
$A, \boldsymbol{b}$ defined constraints

- Inference in structured prediction tasks can be formulated as ILP
- Given $(\boldsymbol{w}, \boldsymbol{y})$ pairs, our goal is to identify A, b
  - *Example:* $\boldsymbol{w}$: adjacent matrix of a graph, $\boldsymbol{y}$: coressponding minimal spanning tree
    Identify A, b so that $A\boldsymbol{y} \leq \boldsymbol{b}$ representing a tree structure

Feasible Set Estimation

- Estimate feasible set defined by the constraints
- Overall idea: we find an outer polytope (left) and an inter polytope (right), and squeeze them towards feasible set.



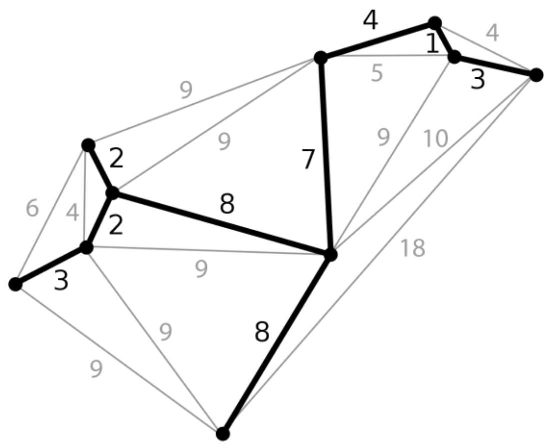


Equality Constraints Mining and Latent Variable

- Equality Constraints Mining
  - Equality constraints $W_{eq} \cdot \boldsymbol{y} = c$ can be found by solving the kernel of matrix $[\boldsymbol{y}, \mathbf{1}]$.
- Latent Variables
  - A set of variables that help formulate the constraints
- See details in paper.

Experiments: Synthetic Data

- Minimal Spanning Tree (MST)
  - Learn the complex tree structure
  - Neural models ~10% EM v.s. Ours ~90% EM
- Sudoku
  - Learn the rules of Sudoku
  - Neural models ~15% EM v.s. Ours 100% EM

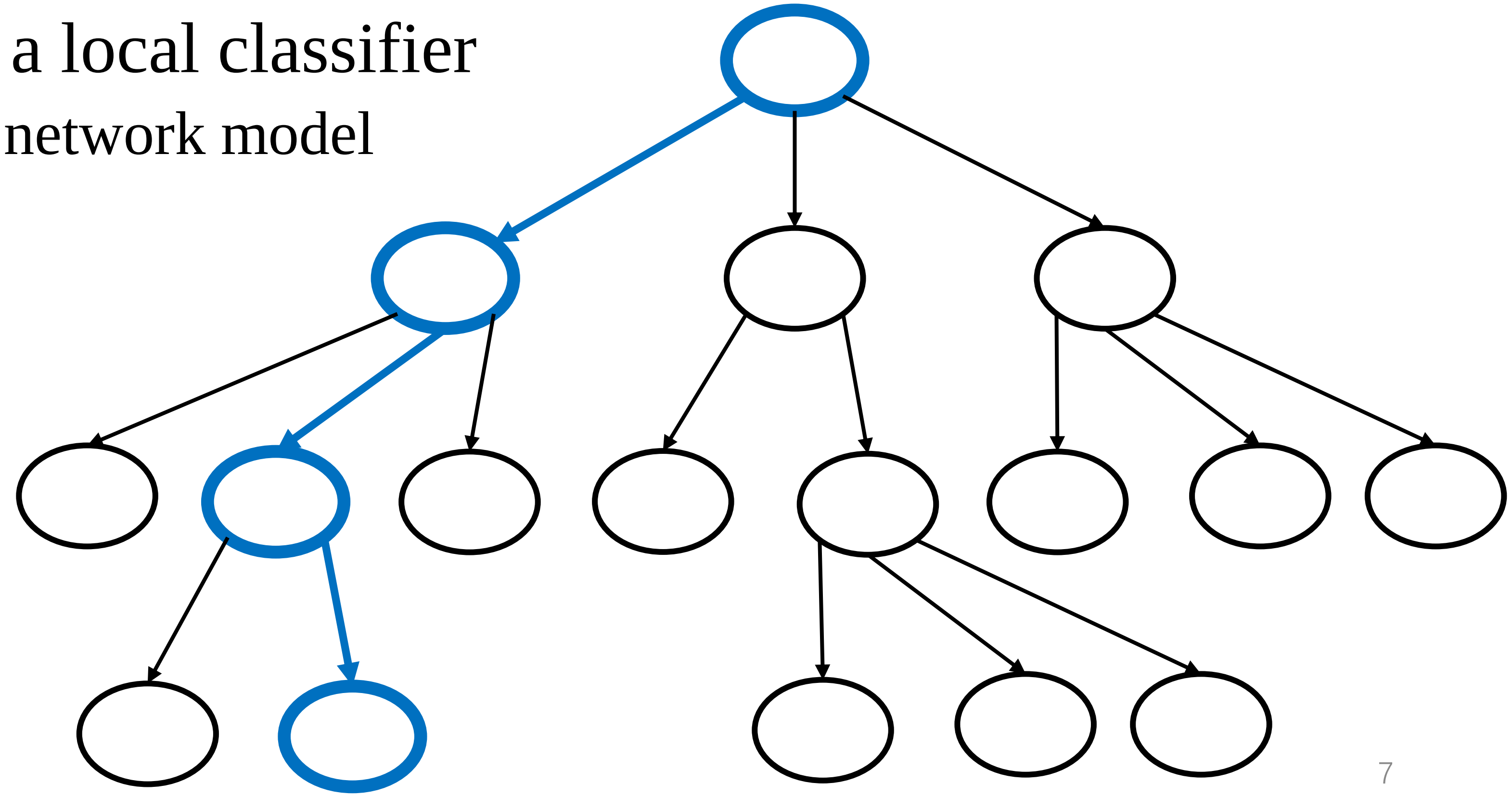


5	3			7				
6			1	9	5			
	9	8					6	
8				6				3
4			8		3			1
7				2				6
	6					2	8	
			4	1	9			5
				8			7	9

Metric: Exact Match (EM)

Experiments: Hierarchical Multiclass Classification (HMC)

- Label Dependency
  - Labels are organized as a tree structure, output is a path from root to a leaf
- Weights $\boldsymbol{w}$ is given by a local classifier
  - $\boldsymbol{w}$ is given by a neural network model



Experiments: Hierarchical Multiclass Classification (HMC)

- Baseline: conduct the inference without constraints
- Our method: conduct constrained inference w/ learned constraints
- Results:
  - Learned $\boldsymbol{w}$ is noisy
  - Able to identify constraints and improve performance

Paper and Code

Paper: https://arxiv.org/abs/2006.10836

Code: https://github.com/uclanlp/ILPLearning