Randomized Greedy Search for Structured Prediction: Amortized Inference and Learning

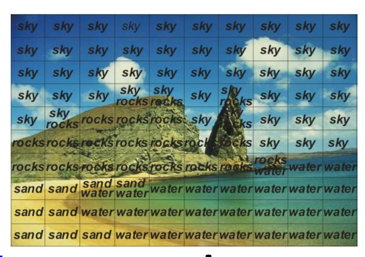
Chao Ma*, F A Rezaur Rahman Chowdhury*, Aryan Deshwal, Md Rakibul Islam, Jana Doppa and Dan Roth

Oregon State University
University of Pennsylvania

Structured Prediction

- **Given:** a set of structured input-output pairs of the form (x, y)
  - Handwriting recognition

    $x =$  $y =$ s t r u c t u r e d
  - Image labeling

    $x =$  $y =$ 
- **Learn:** a function $F: X \rightarrow Y$ to make predictions on new inputs
- **Evaluation:** against a loss function $L(x, y, F(x)) \in R^+$
  - Hamming loss, F1 score, B^3 score …

Key Challenge: "Argmax" inference

- **Popular Approaches: CRFs, Structured SVM …**
  - Learn parameters of linear models
  - $\phi(x, y)$ is n-dim feature vector over input-output pairs
  - w is n-dim parameter vector

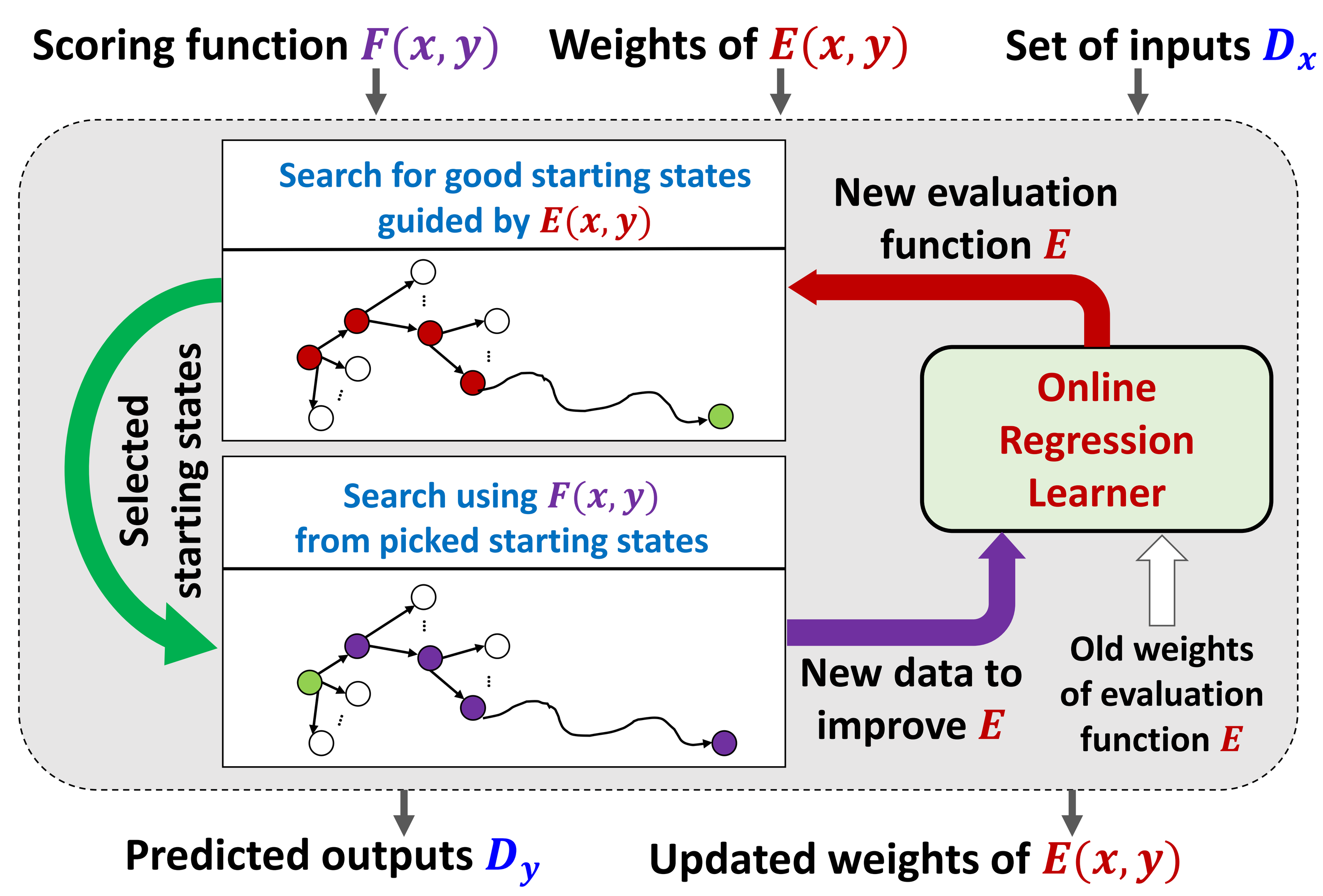
$$F(x) = arg \max_{y \in Y} w \cdot \phi(x, y)$$

$|Y|$ is generally exponentially large

- **"Argmax" inference:**
  - Find the max. scoring output in an exponentially large set of possible outputs. Computationally hard (NP-hard) for all but simplest dependency structure of features $\phi(x, y)$
- **Optimizing "non-decomposable" loss functions is hard**
  - Requires a loss-augmented inference solver, which is tractable only for decomposable loss functions
- **Training and Inference is expensive for complex models**
  - Typically, inference is performed *independently* for each structured input – No reuse of knowledge.
  - Repeated calls to inference solver during training of weights
- **Speedup Learning meets Structured Prediction**
  - Speedup learning methods improve the efficiency of problem-solving process via practice on training problems
  - If we abstract out inference solvers as computational search processes, we can learn search control knowledge to improve the efficiency of inference. Study this general idea in the context of randomized greedy search (RGS).
  - RGS (multiple iterations of greedy search from random outputs) requires many iterations to achieve high accuracy and doesn't work very well for large structured outputs.
  - Amortized RGS: Learn evaluation function to select good starting states to improve the accuracy of greedy search (high accuracy with a small number of RGS iterations)
  - Plug amortized RGS inference solver in the inner-loop of structured learning to improve the efficiency of training

Amortized RGS Inference and Learning

- **RGS(α) Inference Solver**
  - α fraction of the output variables are initialized with a learned IID classifier and remaining $1 - \alpha$ variable randomly. Prior work is a special case with $\alpha = 0$ [Zhang et al., 2014, 2015]
  - α controls the trade-off between the minimum depth at which target outputs can be located and diversity of starting outputs
  - Non-zero α will potentially help with large structured outputs
- **Learning for Amortized Inference**
  - Given a set of structured inputs D_x and scoring function $F(x, y)$ to score candidate outputs, we want to reduce the number of iterations of RGS(α) to uncover high-scoring structured outputs
  - Key Idea: Learn evaluation function $E(x, y)$ to select good starting states to improve the accuracy of greedy search guided by $F(x, y)$ [Boyan and Moore, 1997]



Experimental Results

- RGS with best α gives better accuracy than RGS(0) for tasks with large structured outputs. RGS(α) is competitive or better than many state-of-the-art methods.

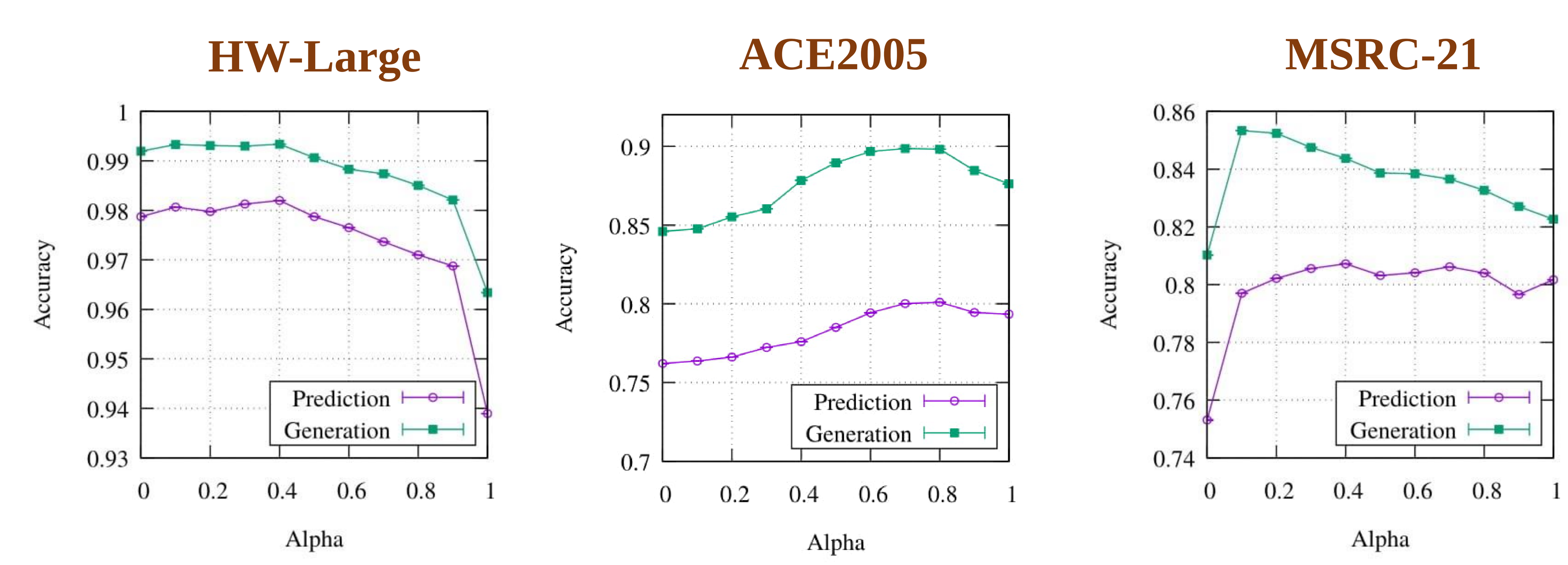
a. Coreference Resolution (ACE2005)

	MUC	BCube	CEAFe	CoNLL
Berkeley	81.41	74.70	72.93	76.35
RGS(0)	80.07	74.13	71.25	75.15
RGS(α)	**82.18**	**76.57**	**74.01**	**77.58**

b. Image Segmentation (MSRC21)

	Global	Average
ICCV2011	85	77
CRF-CNN	91.1	**90.5**
RGS(0)	81.27	73.14
RGS(α)	85.29	78.92
RGS(α)-CNN	**91.53**	90.28

- Prediction and generation accuracy of RGS(α) on development set



- Inference and training time improves significantly with amortized RGS

Testing Time (milli seconds)

	Bibtex	Bkmrks	HWLrg	MSRC	ACE05
SPEN	5791	63073	-	-	-
DVN	18086	211448	-	-	-
RGS	69890	288058	17323	9451	282864
A-RGS	20925	98921	4812	2294	55355

Training Time (minutes)

	Bibtex	Bkmrks	HWLrg	MSRC	ACE05
SPEN	114	237	-	-	-
DVN	20	204	-	-	-
DCD(RGS)	95	392	71	115	171
DCD(A-RGS)	32	319	44	27	39