## Lecture 23 - scribbles - L2S & submodularity

Friday, April 7, 2017 15:01

beam search:



15jues:

a) variciele langth output? -> end of sequence special character

b) long input sequence x needs to be summarized in fixed brath context vector

-> altertion mechanism

Learning to search (L2S)  
hw: 
$$k \rightarrow S$$
  
hw(x) = argman  $S(x,y;w)$ .  
yey  
learning to search's = split y is ordered # of decisions  
(y\_1,..., yr)  
learn a policy  $\pi(genessing) = y_{\mathcal{E}}$   
policy which takes an history to make next decision  
L3S: from  $\{z^{(i)}, y^{(i)}\}_{io}^{\mathcal{H}}$  and  $l(y, y')$   
learn a good policy  $\pi_{io}$  s.t.  $\hat{y}_{\mathcal{E}} = \pi_{io}(\hat{y}_{ii\in I}, z)$ 

$$V = \mathbb{E}_{1,r}, d\mathbb{E} \text{ is ground set}^{V}$$

$$Concrete examples Ising model E(y) = \mathbb{E}_{0}^{V} \mathbb{E}_{1,r} - \mathbb{E}_{0}^{V} \mathbb{E}_{0}^{V} \mathbb{E}_{1,r} \mathbb{E}_{$$

\* F(A) = g(IAI) if g is <u>cancare</u> randinality then F is submodular

\* embed sets as corrers of Mei hyperauble 
$$V(A) =$$

Sources extension flextends 
$$F(A)$$
 from corners to whole hypotrade  
 $f(\omega) = F(A)$  when  $\omega = V(A)$   
 $F$  submodular  $\Longrightarrow$  brass extension 5 is conver

## pointers:

- encoder-decoder model (seq2seq):
  - see <u>chapter 10.4</u> of deep learning book
  - $\circ~$  attention mechanism:  $\underline{12.4.5.1}$  of deep learning book
  - gated RNNs (LSTM or GRU) see <u>10.10</u>
- learning to search
  - see great ICML 2015 tutorial
  - LOLS paper: Kai-Wei Chang, Akshay Krishnamurthy, Alekh Agarwal, Hal Daume, John Langford, "Learning to Search Better than Your Teacher", <u>ICML 2015</u>
- submodularity:
  - website with tutorials and pointers: <u>http://submodularity.org/</u>
  - detailed monograph by Francis Bach: F. Bach. "Learning with Submodular Functions: A Convex Optimization Perspective." Foundations and Trends in Machine Learning, 6(2-3):145-373, 2013 | slides

aside: mentioned the counter-example by Nesterov that showed that Frank-Wolfe can sometimes not converge if the function is not differentiable;

see Example 1 in: Yurii Nesterov, "<u>Complexity bounds for primal-dual methods minimizing the model of objective function</u>", 2016