CONVOLUTION KERNELS FOR NATURAL LANGUAGE

Paper by: Michael Collins, Nigel Duffy Presented by: Ruinan (Renie) Lu

AYUSH BADOLA(B20MT012)

NATURAL LANGUAGE PROCESSING

- TRAINING - GIVEN A STRING OR SENTENCE FINDING THE HIDDEN STRUCTURE
- PARSING-TASKS INVOLVING TREE DATA STRUCTURE
- TAGGING-TASKS INVOLVING HIDDEN TEXT SEQUENCES

STRUCTURE IN NLP TASK

Parse tree: Lou Gerstner is chairman of IBM -- [S [NP Lou Gerstner] [VP is [NP chairman [PP of [NP IBM]]]]]

DEALING WITH AMBIGUITY

- Stochastic grammar: --
- PCFG (Probabilistic Context Free Grammar) for parsing
- HMM (Hidden Markov Model) for tagging
- Probabilities are attached to rules in the grammar.
- Rule probabilities are estimated using MLE (Maximum likelihood estimation).
- Probabilities are used to rank the competing analyses for the same sentence.

PCFGs AS PARSING METHOD

- Counts the relative number of occurrences of a given rule.
- Uses the count to represent its learned knowledge.
- Makes strong independence assumptions.
- Ignores substantial amounts of structural information (e.g. assume rules applied at level i in the parse tree are unrelated to those applied at level i+1).

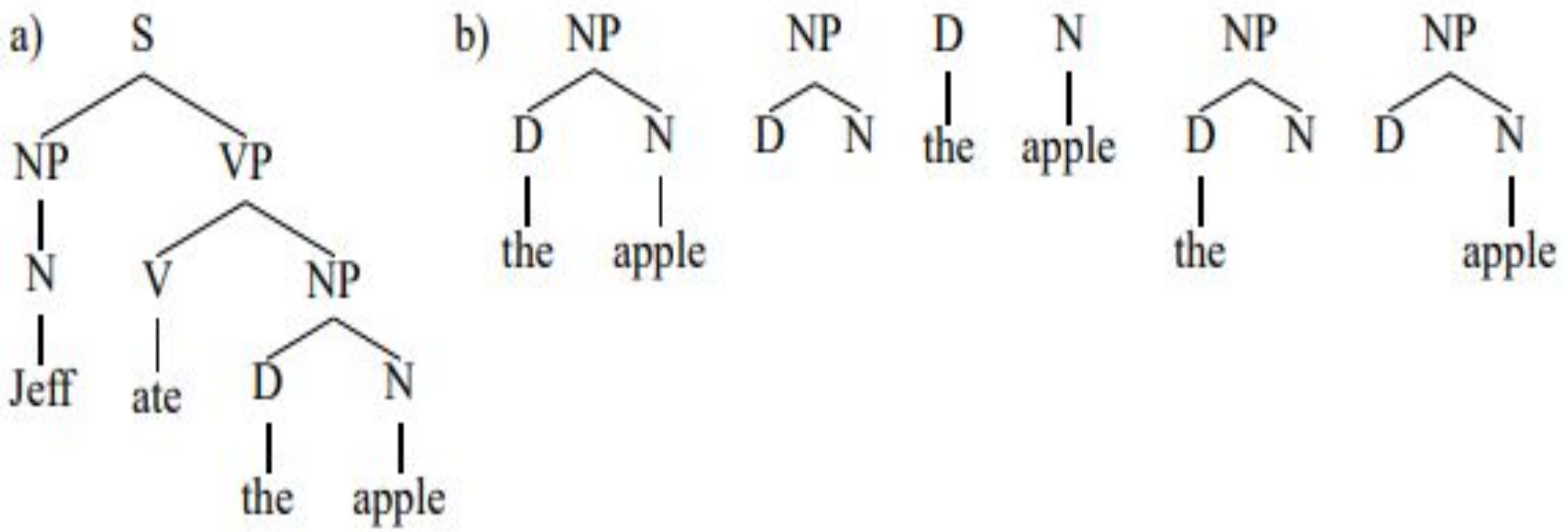
INTRODUCTION TO KERNELS

- Kernels basically takes the dot product between two feature vectors
- they try to find out similarities between objects
- reduces the complexity of computations

APPLYING KERNEL METHOD TO NLP PROBLEMS

- Input in NLP-Trees,sequences,etc.
- kernel used-tree kernel(convolution kernels)
- the kernel function counts the number of common subtrees in any two trees to find similarity between them.
- finds dependencies between structural information.

TREE AND SUBTREE



REPRESENTATION OF TREE KERNELS

- Enumerate (implicitly) all tree fragments:1,…,n.
- Represent each tree by an n-d vector:
- number of occurrences of the i’th tree fragment in tree T
- Tree T is represented as:
- $h(T)=(h1(T),h2(T),h3(T),......hn(T))$
- n is very large

DEFINING TREE KERNEL

- INNER PRODUCT BETWEEN TWO TREES T1 AND T2
- K(T1,T2)=h(T1).h(T2)
- INDICATOR FUNCTION(I)
- Ii(n)=1 if sub-tree is seen rooted at node ;

  and 0 otherwise

definition cont..

$$\mathbf{h}(T_1)\cdot\mathbf{h}(T_2)=\sum_i h_i(T_1)h_i(T_2)=\sum_{n_1\in N_1}\sum_{n_2\in N_2}\sum_i I_i(n_1)I_i(n_2)=\sum_{n_1\in N_1}\sum_{n_2\in N_2}C(n_1,n_2)$$

Recursive definition

- If the productions at n1 and n2 are different C(n1,n2)=0.
- If the productions at n1 and n2 are the same, and n1 and n2 are pre-terminals, then C(n1,n2)=1.
- Else if the productions at n1 and n2 are the same and n1 and n2 are not pre-terminals,

$$C(n_1,n_2) = \prod_{j=1}^{nc(n_1)} \left(1 + C(ch(n_1,j), ch(n_2,j))\right),$$

EXPERIMENT OUTCOMES

- applied tree kernel to the problem of parsing the Penn treebank ATIS corpus
- Applied Voted Perceptron using tree kernel to the test set.
- Applied PCFG on same
- voted perceptron using kernel performed better with better accuracy

PERFORMANCE OF PERCEPTRON WITH KERNEL(improvement is relative to PCFG)

Scale	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Score	77 ± 1	78 ± 1	79 ± 1	79 ± 1	79 ± 1	79 ± 1	79 ± 1	79 ± 1	78 ± 1
Imp.	11 ± 6	17 ± 5	20 ± 4	21 ± 3	21 ± 4	22 ± 4	21 ± 4	19 ± 4	17 ± 5