COMP90042

Workshop Week 05

☐ Homework 2

Due: Sunday, April 1

□ No workshops next week (Easter holiday break)

Syllabus

1	Introduction and Preprocessing	Text classification
2	Lexical semantics	Distributional semantics
3	Part of Speech Tagging	Hidden Markov Models
4	Unsupervised Hidden Markov Models	Context-Free Grammars
5	Probabilistic Parsing	Dependency parsing
	Easter hol	iday break
6	N-gram language modelling	Deep learning for language models
		and tagging
7	Information Extraction	Question Answering
8	Topic Models	ANZAC day holiday
9	Information Retrieval Boolean	Indexing and querying in the vector
	search and the vector space model	space model, evaluation
10	Index and vocabulary compression	Efficient query processing
11	The Web as a Graph: Page-rank & HITS	Machine Translation (word based)
12	Machine translation (phrase based)	Subject review
	and neural encoder-decoder	

Outline

- Parsing
 - ☐ The Chomsky Normal Form (CNF)
 - ☐ The CYK algorithm

- Consider the following simple context-free grammar
 - \square S -> NP VP
 - □ VP -> V NP | V NP PP
 - □ PP -> P NP
 - \square V -> "saw" | "walked"
 - \square NP -> "John" | "Bob" | Det N | Det N PP
 - □ Det -> "a" | "an" | "the" | "my"
 - □ N -> "man" | "cat" | "telescope" | "park"
 - □ P -> "on" | "by" | "with"
- ☐ Parse the following sentence
 - "an park by Bob walked an park with Bob"

A possible parsing tree using the rules

S											
NP VP											
Det	N	Р	Р	V	NP		Р	Р			
Det		Р	NP	V	Det	N	Р	NP			
an	park	by	Bob	walked	an	park	with	Bob			

- \square S -> NP VP
- □ VP -> V NP | V NP PP
- \square PP -> P NP
- V -> "saw" | "walked"
- \square NP -> "John" | "Bob" | Det N | Det N PP
- □ Det -> "a" | "an" | "the" | "my"
- □ N -> "man" | "cat" | "telescope" | "park"
- □ P -> "on" | "by" | "with"

The CYK parsing algorithm

- □ Convert grammar to Chomsky Normal Form (CNF)
- ☐ Fill in a parse table
- ☐ Use table to derive parse
- Convert result back to original grammar

Rewrite in Chomsky Normal Form (CNF)

- \square VP -> V X
- \square X -> NP PP
- \square NP -> Det Y
- □ Y -> N PP

- \square S -> NP VP
- □ VP -> V NP | V NP PP
- □ PP -> P NP
- \square V -> "saw" | "walked"
- □ NP -> "John" | "Bob" | Det N | Det N PP
- □ Det -> "a" | "an" | "the" | "my"
- □ N -> "man" | "cat" | "telescope" | "park"
- □ P -> "on" | "by" | "with"

In CNF, each rule consists of either:

a (single) non-terminal which re-writes as a single terminal, or

a (single) non-terminal which re-writes as exactly two non-terminals

Original rules

	S											
NP VP												
Det	N	Р	Р	\/	NP		PP					
Det		Р	NP	V	Det	N	Р	NP				
an	park	by	Bob	walked	an	park	with	Bob				

□ CNF (Binary tree)

	S										
	N	Р				VP					
	Υ					>	<				
Det	NI	Р	P V		N	Р	Р	Р			
	N	Р	NP		Det	N	Р	NP			
an	park	by	Bob	walked	an	park	with	Bob			

	an	park	by	Bob	walked	an	park	with	Bob	
	[0,1]	[0,2]	[0,3]	[0,4]	[0,5]	[0,6]	[0,7]	[0,8]	[0,9]	
		[1,2]	[1,3]	[1,4]	[1,5]	[1,6]	[1,7]	[1,8]	[1,9]	
			[2,3]	[2,4]	[2,5]	[2,6]	[2,7]	[2,8]	[2,9]	
				[3,4]	[3,5]	[3,6]	[3,7]	[3,8]	[3,9]	
					[4,5]	[4,6]	[4,7]	[4,8]	[4,9]	
	function CI	XY-PARSE(w	ords, gramm	ar) returns to	able	[5,6]	[5,7]	[5,8]	[5,9]	
	for all	om 1 to LENG $\{A \mid A \rightarrow w \in table[i-1, i]\}$	$ords[j] \in gr$	rammar			[6,7]	[6,8]	[6,9]	
$table[j-1,j] \leftarrow table[j-1,j] \cup A$ $for \ i \leftarrow from \ j-2 \ downto \ 0 \ do$ $for \ k \leftarrow i+1 \ to \ j-1 \ do$ $[7,8]$										
for all $\{A \mid A \to BC \in grammar \text{ and } B \in table[i,k] \text{ and } C \in table[k,j]\}$ $table[i,j] \leftarrow table[i,j] \cup A$ Figure 12.5 The CKY algorithm.										

an	park	by	Bob	walked	an	park	with	Bob		
[0,1]	[0,2]	[0,3]	[0,4]	[0,5]	[0,6]	[0,7]	[0,8]	[0,9]		
	[1,2]	[1,3]	[1,4]	[1,5]	[1,6]	[1,7]	[1,8]	[1,9]		
		[2,3]	[2,4]	[2,5]	[2,6]	[2,7]	[2,8]	[2,9]		
			[3,4]	[3,5]	[3,6]	[3,7]	[3,8]	[3,9]		
				[4,5]	[4,6]	[4,7]	[4,8]	[4,9]		
function C	KY-Parse(n	ords, gramm	ar) returns t	able	[5,6]	[5,7]	[5,8]	[5,9]		
for all	om 1 to LENG $ \begin{cases} A \mid A \rightarrow w \\ table[i-1, i] \end{cases} $	$ords[j] \in gr$	rammar}			[6,7]	[6,8]	[6,9]		
$table[j-1,j] \leftarrow table[j-1,j] \cup A$ $for \ i \leftarrow from \ j-2 \ downto \ 0 \ do$ $for \ k \leftarrow i+1 \ to \ j-1 \ do$ $[7,8]$										
for all $\{A \mid A \to BC \in grammar \text{ and } B \in table[i,k] \text{ and } C \in table[k,j]\}$ $table[i,j] \leftarrow table[i,j] \cup A$ Figure 12.5 The CKY algorithm.										

an	park	by	Bob	walked	an	park	with	Bob		
Det [0,1]	[0,2]	[0,3]	[0,4]	[0,5]	[0,6]	[0,7]	[0,8]	[0,9]		
	N [1,2]	[1,3]	[1,4]	[1,5]	[1,6]	[1,7]	[1,8]	[1,9]		
		P [2,3]	[2,4]	[2,5]	[2,6]	[2,7]	[2,8]	[2,9]		
			NP [3,4]	[3,6]	[3,7]	[3,8]	[3,9]			
				V [4,5]	[4,6]	[4,7]	[4,8]	[4,9]		
function (CKY-PARSE(и	vords, gramm	ar) returns t	able	Det [5,6]	[5,7]	[5,8]	[5,9]		
	for $j \leftarrow$ from 1 to LENGTH(words) do for all $\{A \mid A \rightarrow words[j] \in grammar\}$ [6.7]									
$table[j-1,j] \leftarrow table[j-1,j] \cup A$ $for \ i \leftarrow from \ j-2 \ downto \ 0 \ do$ $for \ k \leftarrow i+1 \ to \ j-1 \ do$ $[7,8]$										
for all $\{A \mid A \to BC \in grammar \text{ and } B \in table[i,k] \text{ and } C \in table[k,j]\}$ $table[i,j] \leftarrow table[i,j] \cup A$ Figure 12.5 The CKY algorithm.										

an	park	by	Bob	walked	an	park	with	Bob		
Det [0,1]	[0,2]	[0,3]	[0,4]	[0,5]	[0,6]	[0,7]	[0,	[0,9]		
	N [1,2]	[1,3]	[1,4]	[1,5]	[1,6]	[1,7]	[1,8]	[1,9]		
		P [2,3]	[2,4]	[2,5]	[2,6]	[2,7]	[2,8]	[2,9]		
	NP (3,4] [3,5] [3,6] [3,7] [3,8]									
				V [4,5]	[4,6]	[4,7]	[4,8]	[4,9]		
function C	KY-Parse(n	vords, gramm	ar) returns t	able	Det [5,6]	[5,7]	[5,8]	[5,9]		
for all	om 1 to LENG $\{A \mid A \rightarrow w \}$ table [i = 1 i	$[ords[j] \in gr$	rammar			N [6,7]	[6,8]	[6.9]		
$ \begin{array}{c} table[j-1,j] \leftarrow table[j-1,j] \cup A \\ $										
for all $\{A \mid A \to BC \in grammar \text{ and } B \in table[i,k] \text{ and } C \in table[k,j]\}$ $table[i,j] \leftarrow table[i,j] \cup A$ Figure 12.5 The CKY algorithm.										

an	park	by	Bob	walked	an	park	with	Bob
Det	NP		NP,X					
[0,1]	[0,2]	[0,3]	[0,4]	[0,5]	[0,6]	[0,7]	[0,8]	[0,9]
	N		Υ					
	[1,2]	[1,3]	[1,4]	[1,5]	[1,6]	[1,7]	[1,8]	[1,9]
		Р	PP					
		[2,7]	[2,8]	[2,9]				
			NP					
		[3,8]	[3,9]					
				V				
				[4,5]	[4,6]	[4,7]	[4,8]	[4,9]
					Det			
function	n CKY-PARSE(u	ords, gramn	nar) returns	table	[5,6]	[5,7]	[5,8]	[5,9]
	-from 1 to LENG					N		
for	f all $\{A \mid A \rightarrow w\}$	[0]	,			[6,7]	[6,8]	[6,9]
for	$table[j-1, j \\ i \leftarrow \mathbf{from} \ j-2 \ \mathbf{c}$		Р					
	for $k \leftarrow i + 1$ to		[7,8]	[7,9]				
	for all $\{A \mid A\}$	table[k,j]	_	NP				
	table			[8,9]				
Figure 12	.5 The CKY al			L / J				

an	park	by	Bob	walked	an	park	with	Bob	
Det	NP		NP,X						
[0,1]	[0,2]	[0,3]	[0,4]	[0,5]	[0,6]	[0,7]	[0,8]	[0,9]	
	N		Υ						
	[1,2]	[1,3]	[1,4]	[1,5]	[1,6]	[1,7]	[1,8]	[1,9]	
		Р	PP						
[2,3] [2,4] [2,5] [2,6] [2,7]								[2,9]	
	NP S								
	[3,4] [3,5] [3,6] [3,7]								
				V		VP			
				[4,5]	[4,6]	[4,7]	[4,8]	[4,9]	
					Det	NP			
function	n CKY-PARSE(и	ords, gramm	ar) returns t	able	[5,6]	[5,7]	[5,8]	[5,9]	
	-from 1 to LEN					N			
for	all $\{A \mid A \rightarrow w \}$ table $[j-1, j]$	[0]	,			[6,7]	[6,8]	[6,9]	
for	iable[j-1,j] $i \leftarrow \mathbf{from} \ j-2 \ \mathbf{c}$		Р						
	[7,8]	[7,9]							
		NP							
$table[i,j] \leftarrow table[i,j] \cup A$									
Figure 12	.5 The CKY al	gorithm.						[8,9]	

an	park	by	Bob	walked	an	park	with	Bob
Det	NP		NP,X			S		
[0,1]	[0,2]	[0,3]	[0,4]	[0,5]	[0,6]	[0,7]	[0,8]	[0,9]
	N		Υ					
	[1,2]	[1,3]	[1,4]	[1,5]	[1,6]	[1,7]	[1,8]	[1,9]
		Р	PP					
[2,3] [2,4] [2,5] [2,6] [2,7]								[2,9]
			NP			S		
		[3,7]	[3,8]	[3,9]				
				V		VP		VP*2
				[4,5]	[4,6]	[4,7]	[4,8]	[4,9]
					Det	NP		NP,X
function	n CKY-PARSE(w	ords, gramn	ar) returns	table	[5,6]	[5,7]	[5,8]	[5,9]
	-from 1 to Lend	,				N		Υ
for	$A = A \mid A \rightarrow W$	[0]	,			[6,7]	[6,8]	[6,9]
for	$table[j-1, j \\ i \leftarrow \mathbf{from} \ j-2 \ \mathbf{d}$		Р	PP				
10.	for $k \leftarrow i + 1$ to		[7,8]	[7,9]				
	for all $\{A \mid A\}$	table[k,j]		NP				
	table[[8,9]				
Figure 12	.5 The CKY al	gorithm.						L - / - J

an	park	by	Bob	walked	an	park	with	Bob
Det	NP		NP,X			S		S*2
[0,1]	[0,2]	[0,3]	[0,4]	[0,5]	[0,6]	[0,7]	[0,8]	[0,9]
	N		Υ					
	[1,2]	[1,3]	[1,4]	[1,5]	[1,6]	[1,7]	[1,8]	[1,9]
		Р	PP					
[2,3] [2,4] [2,5] [2,6] [2,7]								[2,9]
		S		S*2				
		[3,7]	[3,8]	[3,9]				
				V		VP		VP*2
				[4,5]	[4,6]	[4,7]	[4,8]	[4,9]
					Det	NP		NP,X
function	n CKY-PARSE(w	ords, gramn	ar) returns	table	[5,6]	[5,7]	[5,8]	[5,9]
$\mathbf{for}\ j \leftarrow$	-from 1 to Lend	GTH(words)	do			N		Υ
for	all $\{A \mid A \rightarrow w\}$	[0]	,			[6,7]	[6,8]	[6,9]
for	$table[j-1, j]$ $i \leftarrow \mathbf{from} \ j-2 \mathbf{d}$		P	PP				
101	for $k \leftarrow i + 1$ to		[7,8]	[7,9]				
	for all $\{A \mid A\}$	able[k,j]	[,,0]					
	table[2 10 20		NP				
Figure 12	.5 The CKY al			[8,9]				

Two possible trees in CNF

5									
NP				VP					
Det	Υ				NP				
		PP		V		Υ			
	N	Р	NP	V	Det	N	PP		
		r	INF				Р	NP	
an	park	by	Bob	walked	an	park	with	Bob	
S									
3									
NP				VP					
	Υ				X				
Det	N	PP		V	NP		PP		
		Р	NP		Det	N	Р	NP	
an	park	by	Bob	walked	an	park	with	Bob	

Convert result back to original grammar

S									
NP				VP					
Det	N	PP			NP				
		J P	NP	V	Det	N	PP		
		Γ					Р	NP	
an	park	by	Bob	walked	an	park	with	Bob	

S									
NP				VP					
Det	N	PP		\/	NP		PP		
		Р	NP	V	Det	N	Р	NP	
an	park	by	Bob	walked	an	park	with	Bob	

The CYK parsing algorithm

- □ Convert grammar to Chomsky Normal Form (CNF)
- ☐ Fill in a parse table
- Use table to derive parse
- Convert result back to original grammar