Natural Language Processing
CSCI 4152/6509 — Lecture 14
POS Tagging and Hidden Markov Model

Instructors: Vlado Keselj

Time and date: 16:05 – 17:25, 28-Oct-2024

Location: Carleton Tupper Building Theatre C

Previous Lecture

- P0 discussion: P-22, P-25
- N-gram model as Markov chain
- Language model evaluation; Perplexity
- Text classification using language modeling
- N-gram Model Smoothing:
  - Add-one smoothing (Laplace smoothing)
  - Witten-Bell discounting

Bigrams and Higher-order N-grams

- Modified probability for seen bigrams

$$P(a|b) = \frac{\#(ba)}{\#(b) + r_b}$$

- Remaining probability mass for unseen events

$$\frac{r_b}{\#(b) + r_b}$$

- Estimate for unseen bigrams starting with b (N_b is the set of tokens that never follow b in training text):

$$P(a|b) = \frac{r_b}{\#(b) + r_b} \cdot P(a)/\Sigma_{x \in N_b} P(x)$$

The Next Model: HMM

- HMM — Hidden Markov Model
- Typically used to annotate sequences of tokens
- Most common annotation: Part-of-Speech Tags (POS Tags)
- First, we will make a review of parts of speech in English

Part-of-Speech Tags (POS Tags)

- Reading: Sections 5.1–5.2 (Ch. 8 in new edition)
- Word classes called **Part-of-Speech (POS) classes**
  - also known as **syntactic categories**, **grammatical categories**, or **lexical categories**
- Ambiguous example: Time flies like an arrow.

	Time	flies	like	an	arrow.
1.	N	V	P	D	N
2.	N	N	V	D	N
⋮					

- **POS tags:** labels to indicate POS class
- **POS tagging:** task of assigning POS tags

POS Tag Sets

- Traditionally based on Ancient Greece source: eight parts of speech:
  - nouns, verbs, pronouns, prepositions, adverbs, conjunctions, participle, and articles
- Computer processing introduced a need for a large set of categories
- Useful in NLP, e.g.: named entity recognition, information extraction
- Various POS tag sets (in NLP):
  Brown Corpus, Penn Treebank, CLAWS, C5, C7, ...
- We will use the Penn Treebank system of tags

WSJ Dataset

- WSJ — Wall Street Journal data set
- Most commonly used to train and test POS taggers
- Consists of 25 sections, about 1.2 million words
- Example:

```
Pierre NNP Vinken NNP , , 61 CD years NNS old JJ , ,
  will MD join VB the DT board NN as IN a DT
  nonexecutive JJ director NN Nov. NNP 29 CD . .
Mr. NNP Vinken NNP is VBZ chairman NN of IN
  Elsevier NNP N.V. NNP , , the DT Dutch NNP
  publishing VBG group NN . .
Rudolph NNP Agnew NNP , , 55 CD years NNS old JJ
  and CC former JJ chairman NN of IN Consolidated NNP
  Gold NNP Fields NNP PLC NNP , , was VBD named VBN
```

Open and Closed Categories

- Word POS categories are divided into two sets: *open* and *closed* categories:
- **open categories**
  - dynamic set
  - content words
  - larger set
  - e.g.: nouns, verbs, adjectives
- **closed categories** or **functional categories**:
  - fixed set
  - small set
  - frequent words
  - e.g.: articles, auxiliaries, prepositions

Open Word Categories

- nouns (NN, NNS, NNP, NNPS)
  - concepts, objects, people, and similar
- adjectives (JJ, JJR, JJS)
  - modify (describe) nouns
- verbs (VB, VBP, VBZ, VBG, VBD, VBN)
  - actions
- adverbs (RB, RBR, RBS)
  - modify verbs, but other words too

Nouns (NN, NNS, NNP, NNPS)

Nouns refer to people, animals, objects, concepts, and similar.

Features:

- number: singular, plural
- case: subject (nominative), object (accusative)
- Some languages have more cases, and more number values
- Some languages have grammatical gender

Noun Tags and Examples

NN for common singular nouns; e.g., company, year, market

NNS for common plural nouns; e.g., shares, years, sales, prices, companies

NNP for proper nouns (names); e.g., Bush, Japan, Federal, New York, Corp, Mr., Friday, James A. Talcott ("James NNP A. NNP Talcott NNP")

NNPS for proper plural nouns; e.g., Canadians, Americans, Securities, Systems, Soviets, Democrats

Adjectives (JJ, JJR, JJS)

- Adjectives describe properties of nouns
- For example: red rose, long journey
- Three inflective forms:

Form	Example	Tag
positive	rich	JJ
comparative	richer	JJR
superlative	richest	JJS

Periphrastic Adjective Forms

- Comparative and superlativ forms in English consist of several words for longer adjectives
- Example:
  intelligent — more intelligent — the most intelligent
- These are called **periphrastic forms**
- They are tagged as follows:
  `more JJR intelligent JJ`
  and
  `the DT most JJS intelligent JJ`

Verbs (VB, VBP, VBZ, VBG, VBD, VBN)

Verbs are used to describe:

- actions; e.g., throw the stone
- activities; e.g., walked along the river
- or states; e.g., have $50

Verb Tags

Verbs can have different forms and they are tagged accordingly:

Tag	Form name	Example
VB	base	eat, be, have, walk, do
VBD	past	ate, said, was, were, had
VBG	present participle	eating, including, according, being
VBN	past participle	eaten, been, expected
VBP	present non-3sg	eat, are, have, do, say, ’re, ’m
VBZ	present 3sg	eats, is, has, ’s, says

Gerund is a noun which has the same form as the present participle;

e.g., ‘Walking is fun.’

Verb Features

- number: singular, plural
- person: 1st, 2nd, 3rd
- tense: present, past, future
- aspect: progressive, perfect
- mood: possibility, subjunctive (e.g. 'They requested that he be banned from driving.')
- participles: present participle, past participle
- voice: active, passive:
  "He wrote a book." vs. "A book was written by him."

Verb Tenses

- present: I walk
- infinitive: to walk
- progressive: I am walking
- present perfect: I have walked
- past perfect: I had walked

Adverbs (RB, RBR, RBS)

- Adverbs modify verbs, but also other classes; e.g., adjectives and adverbs
- Some examples: *allegedly, quickly*
- **Qualifiers** or **degree adverbs** are closed adverbs; e.g., very, not
- Example of adverbs modifying verbs:
  She *often* travels to Las Vegas.
- Example of adverbs modifying verbs and adverbs:
  Unfortunately, John walked *home*
  extremely slowly yesterday.
- Example of adverbs modifying adjectives:
  a *very* unlikely event
  a *shockingly* frank exchange

Adverb Inflections

Adverbs can have three forms, similarly to adjectives;

Tag	Form	Examples
RB	positive	late, often, quickly
RBR	comparative	later, better, less
RBS	superlative	most, best

The superlative adverbs are tagged as RBT in the Brown corpus.

Adverbial Nouns

- Interesting example of blurred boundary between classes in some cases
- Adverbial nouns are nouns that also behave as adverbs
- Examples: ‘home’ and ‘tomorrow’

  I am going home.

  but not

  ** I am going room.*

- Tagged as nouns (NN), but in Brown corpus had a separate tag (NNR)

Closed Word Categories

- small, fixed, frequent, functional group
- typically no morphological transformations
- include:
  - determiners, pronouns, prepositions, particles, auxiliaries and modal verbs, qualifiers, conjunctions, numbers, interjections

Determiners (DT)

- articles: the, a, an
- demonstratives:
  - this, that, those; some, any; either,neither
- quantifiers: all, some

Interrogative Determiners (WDT)

- what, which, whatever, whichever

Predeterminers (PDT)

- Examples: both, quite, many, all such, half
- Examples in context:
  "half the debt", "all the negative campaign"
- Interesting classifications of determiners (Bond 2001)
  - by linear order: pre-determiners, central determiners, post-determiners
  - by meaning: quantifiers, possessives, determinatives

Pronouns (PRP, PRP$)

- PRP for personal pronouns
  - examples: I, you, he, she, it, we, you, they
- PRP tag for accusative case (diff. tag in Brown):
  - examples: me, him, her, us, them
- PRP tag for reflexive pronouns (diff. in Brown):
  - examples: myself, ourselves, ...
- PRP$ tag for possessive pronouns:
  - examples: your, my, her, his, our, their, its
- PRP for second possessives (diff. in Brown):
  - examples: ours, mine, yours, ...

Wh-pronouns (WP) and Wh-possessive (WP$)

- wh-pronouns (WP): who, what, whom, whoever, ...
- wh-possessive pronoun (WP$): whose

Prepositions (IN)

- Prepositions reflect spatial or time relationships.
- Examples: of, in, for, on, at, by, concerning, . . .

Particles (RP)

- frequently ambiguous and confused with prepositions
- used to create compound verbs
- examples: put off, take off, give in, take on, "went on for days", "put it off"

Possessive ending (POS)

- possessive clitic: ’s
- Example: John’s book
- tagged as: John NNP ’s POS book NN

Modal Verbs (MD)

- the examples of modal verbs: can, may, could, might, should, will
- and their abbreviations: 'd, 'll
- tag for modal verbs: MD
- negative forms are separated into a modal verb and an adverb 'not' (will be covered); e.g.: 'couldn't' is tagged as "could MD n't RB"
- *Auxiliary verbs* are: be, have, and do; and their different forms
- in Brown: each auxiliary verb has a separate tag
- in Penn Treebank: they are tagged in the same way as common verbs (we will see that later)

Infinitive word ‘to’ (TO)

- used to denote an infinitive: e.g., to call
- ‘na’ is marked as TO in ‘gonna’, ‘wanna’ and similar; e.g.: “gonna call” is tagged “gon VB na TO call VB”

Qualifiers (RB)

- qualifiers are closed adverbs, and they are tagged as adverbs (RB) (covered later)
- example: not, n’t, very
- postqualifiers: enough, indeed

Wh-adverbs (WRB)

Examples: how, when, where, whenever,. . .

Conjunctions (CC)

- words that connect phrases
- coordinate conjunctions (tag: CC) connect coordinate phrases:
- examples; and, or, but, yet, plus, versus, ...
- subordinate conjunctions connect phrases where one is subordinate to another
- examples: if, although, that, because, ...
- subordinate conjunctions are tagged as prepositions (IN) in Penn Treebank
- in Brown corpus, they used to be tagged CS

Numbers (CD)

Numbers behave in a similar way to adjectives: they also modify nouns.

There are two kinds of numbers:

- **cardinals** or **cardinal numbers**; for example: 1, 0, 100.34, hundred
- **ordinals** or **ordinal numbers**; for example: first, second, 3rd, 4th

Cardinal numbers are tagged as **CD**

Ordinal numbers have a separate tag in the Brown corpus—OD. In the Penn Treebank corpus, they are tagged as *adjectives: JJ*

Interjections (UH)

- Examples:
  yes, no, well, oh, quack, OK, please, indeed, hello, Congratulations, . . .

Remaining POS Classes

— **Existential 'there' (EX)** Belongs to closed word category; example: "There/EX are/VBP three/CD classes/NNS per/IN week/NN"
— **Foreign Words (FW)**
Examples: de (tour de France), perestroika, pro, des
— **List Items (LS)**
Examples: 1, 2, 3, 4, a., b., c., first, second, etc.
— **Punctuation**

Punctuation

Examples	Tag	Description
,	,	comma
; : ... - --	:	mid-sentence separator
. ! ?	.	sentence end
({ [<	(	open parenthesis
) }] >	)	closed parenthesis
‘ ‘‘ non-‘‘	‘‘	open quote
’ ’’	’’	closed quote
$ c HK$ CAN$	$	dollar sign
#	#	pound sign
- + & @ * ** ffr	**SYM**	everything else

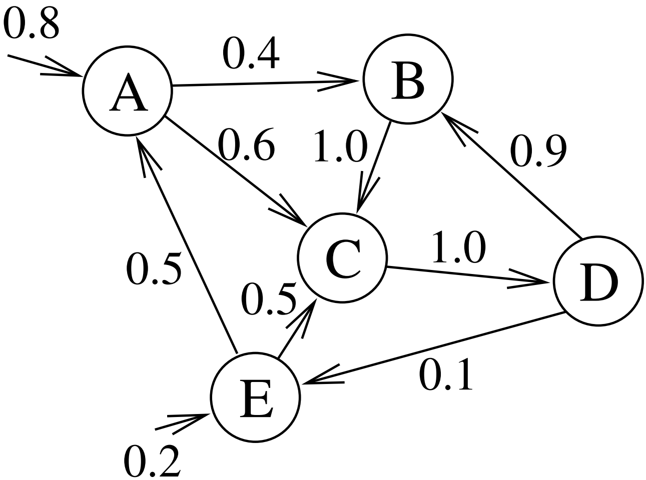
Some Tagged Examples

```
The/DT grand/JJ jury/NN commented/VBD on/IN
a/DT number/NN of/IN other/JJ topics/NNS ./.

Book/VB that/DT flight/NN ./.

Does/VBZ that/DT flight/NN serve/VB dinner/NN ?/.

It/PRP does/VBZ a/DT first-rate/JJ job/NN ./.

``/`` When/WRB the/DT sell/NN programs/NNS hit/VBP
,/, you/PRP can/MD hear/VB the/DT order/NN
printers/NNS start/VB to/TO go/VB ''/'' on/IN the/DT
Big/NNP Board/NNP trading/NN floor/NN ,/, says/VBZ
one/CD specialist/NN there/RB ./.

``/`` Do/VBP you/PRP make/VB sweatshirts/NNS or/CC
sparkplugs/NNS ?/.
```

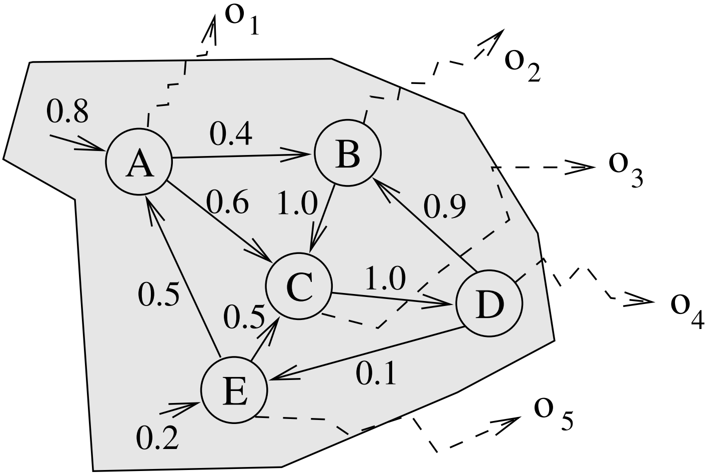
Hidden Markov Model (HMM)

- How do we apply Probabilistic Modelling to POS tagging?
- Idea: Model POS tag sequence as a Markov Chain
  - We can only observe words, which are generated from tags based on a probability distribution
- Model: a hidden Markov Chain with observable symbols emitted from hidden states based on a probability distribution
- This model is known as *Hidden Markov Model (HMM)*

Markov Chain Example



HMM Example



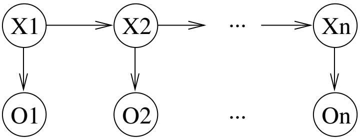
HMM Formal Definition

- Five-tuple: (Q, π, a, V, b) (there are other variations)

1. set of states $Q = \{q_1, q_2, \ldots, q_N\}$
2. initial distribution π: $\pi(q)$ for each state q
3. transition probabilities a: $a(q, s)$ for any two states q and s
4. output vocabulary $V = \{o_1, o_2, \ldots, o_m\}$
5. output probability b: $b(q, o)$ for each state q and observable o

HMM Assumption

- Another graphical representation



- HMM Assumption

$$P(X_1, O_1, \ldots, X_n, O_n) = P(X_1) \cdot P(O_1|X_1) \cdot$$
$$P(X_2|X_1) \cdot P(O_2|X_2) \cdot \ldots \cdot P(X_n|X_{n-1}) \cdot P(O_n|X_n)$$

HMM Application Areas

- Language Modelling
- Acoustic Modelling
- Part-of-Speech tagging (POS tagging)
- Many kinds of sequence tagging (e.g., extracting bio-medical terms)

HMM use in POS Tagging

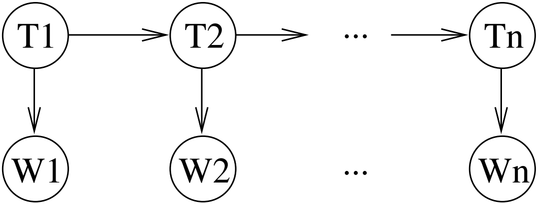
- Hidden states = POS Tags
- Observable variables = words
- In practice: higher-order HMM taggers are used, where the nodes keep a bit longer history (e.g., two previous tags)
- Described in [JM] Sec 5.5 (HMM POS Tagging)

Computational Tasks for HMM

- Evaluation: use HMM assumption formula
- Generation: generate in the order dictated by the "unrolled" graphical representation
- Inference:
  - marginalization, conditioning, completion
  - need for an efficient method (will discuss it)
- Learning: MLE if labeled examples are given

HMM POS Example

- Walk-through example to illustrate inference



- Conditional probability tables required:

$\mathrm{P}(T_1)$, $\mathrm{P}(T_{i+1}|T_i)$, and $\mathrm{P}(W_i|T_i)$

Learning HMM (Training)

- Let us Learn HMM from completely labeled data:

  `swat V flies N like P ants N`

  `time N flies V like P an D arrow N`
- We will use smoothing in word generation, by giving a 0.5 count to all unseen words

Generated Tables

T_1	$\mathrm{P}(T_1)$
N	0.5
V	0.5

T_{i-1}	T_i	$\mathrm{P}(T_i\|T_{i-1})$
D	N	1
N	P	0.5
N	V	0.5
P	D	0.5
P	N	0.5
V	N	0.5
V	P	0.5

T_i	W_i	$\mathrm{P}(W_i\|T_i)$
D	an	$2/3 \approx 0.666666667$
D	*	$1/3 \approx 0.333333333$
N	ants	$2/9 \approx 0.222222222$
N	arrow	$2/9 \approx 0.222222222$
N	flies	$2/9 \approx 0.222222222$
N	time	$2/9 \approx 0.222222222$
N	*	$1/9 \approx 0.111111111$
P	like	0.8
P	*	0.2
V	flies	0.4
V	swat	0.4
V	*	0.2