Part of speech tagging

COMP90042

Natural Language Processing

Lecture 5

Assignments

- 2 assignments (down from 3)
- 20% of subject (no change)
- 1st assignment will be released in week 4

Workshops

- **Online** workshops available till week 12
- Workshop slides by tutors:
  - Modules > Workshops > Workshop Slides

Correction on Lecture 3, Page 22

Backoff

- Absolute discounting redistributes the probability mass **equally** for all unseen n-grams
- Katz Backoff: redistributes the mass based on a **lower order** model (e.g. unigram)

$$P_{katz}(w_i|w_{i-1}) = \begin{cases} \frac{C(w_{i-1},w_i)-D}{C(w_{i-1})}, & \text{if } C(w_{i-1},w_i) > 0 \\ \alpha(w_{i-1})P(w_i), & \text{otherwise} \end{cases}$$

unigram probability for w_i

the amount of probability mass that has been discounted for context w_{i-1}

Correction on Lecture 3, Page 22

Backoff

- Absolute discounting redistributes the probability mass **equally** for all unseen n-grams
- Katz Backoff: redistributes the mass based on a **lower order** model (e.g. unigram)

$$P_{katz}(w_i|w_{i-1}) = \begin{cases} \frac{C(w_{i-1},w_i)-D}{C(w_{i-1})}, & \text{if } C(w_{i-1},w_i) > 0 \\ \alpha(w_{i-1}) \times \frac{P(w_i)}{\sum_{w_j:C(w_{i-1},w_j)=0} P(w_j)}, & \text{otherwise} \end{cases}$$

unigram probability for w_i

sum unigram probabilities for all words that do not co-occur with context w_{i-1}

the amount of probability mass that has been discounted for context w_{i-1}

?

What is Part-of-Speech (POS)?

- AKA word classes, morphological classes, syntactic categories
- Nouns, verbs, adjective, etc
- POS tells us quite a bit about a word and its neighbours:
  - nouns are often preceded by determiners the.
  - verbs preceded by nouns determiners
  - *content* as a **noun** pronounced as *CONtent*
  - *content* as a **adjective** pronounced as *conTENT*

Authorship Attribution Revisited

- Training data:
  - "The lawyer convinced the jury." → Sam
  - "Ruby travelled around Australia." → Sam
  - "The hospital was cleaned by the janitor." → Max
  - "Lunch was served at 12pm." → Max
- "The bookstore was opened by the manager." → ?
- Similar **structure** (passive voice).
  - Not captured by simple BOW representations.
- How to ensure a computer knows/learns this?

Information Extraction

- Given this:
  - "Brasilia, the Brazilian capital, was founded in 1960."
- Obtain this:
  - capital(Brazil, Brasilia)
  - founded(Brasilia, 1960)
- Many steps involved but first need to know **nouns** (Brasilia, capital), **adjectives** (Brazilian), **verbs** (founded) and **numbers** (1960).

Outline

Parts of speech, tagsets

Automatic tagging

POS Open Classes

Open vs **closed** classes: how readily do POS categories take on new words? Just a few open classes:

- Nouns
  - Proper (*Australia*) versus common (*wombat*)
  - Mass (*rice*) versus count (*bowls*)
- Verbs

  tenses
  - Rich inflection (*go/goes/going/gone/went*)
  - Auxiliary verbs (*be*, *have,* and *do* in English)
  - Transitivity (*wait* versus *hit* versus *give*) — number of arguments

POS Open Classes

- Adjectives
  - $\boxed{\text{Gradable}}$ (*happy*) versus non-gradable (*computational*)

    for sake of NLP.
- Adverbs
  - Manner (*slowly*)
  - Locative (*here*)
  - Degree (*really*)
  - Temporal (*yesterday*)

not universal
may not have adj. adv.

POS Closed Classes (English)

- Prepositions (*in*, *on*, *with*, *for*, *of*, *over*,…)
  - ***on*** *the table*
- Particles
  - brushed himself ***off***
- Determiners
  - Articles (*a*, *an*, *the*)
  - Demonstratives (*this*, *that*, *these*, *those*)
  - Quantifiers (*each*, *every, some, two,…*)
- Pronouns
  - Personal (*I*, *me*, *she,…*)
  - Possessive (*my*, *our,…*)
  - Interrogative or *Wh* (*who*, *what*, …)

POS Closed Classes (English)

- Conjunctions
  - Coordinating (*and*, *or*, *but*)
  - Subordinating (*if*, *although, that, …*)
- Modal verbs
  - Ability (*can, could)*
  - Permission (*can, may*)
  - Possibility (*may, might, could, will*)
  - Necessity *(must*)
- And some more…
  - negatives, politeness markers, etc

Ambiguity

- Many word types belong to multiple classes
- Compare:
  - *Time flies like an arrow*
  - *Fruit flies like a banana*

Time	flies	like	an	arrow
noun	verb	preposition	determiner	noun

verb.

Fruit	flies	like	a	banana
noun	noun	verb	determiner	noun

nouns

left party

POS Ambiguity in News Headlines

not sure what to do.

- British Left Waffles on Falkland Islands

verb.

- Juvenile Court to Try Shooting Defendant
- Teachers Strike Idle Kids
- Eye Drops Off Shelf

POS Ambiguity in News Headlines

- [British Left] [Waffles] [on] [Falkland Islands]
- Juvenile Court to Try Shooting Defendant
- Teachers Strike Idle Kids
- Eye Drops Off Shelf

POS Ambiguity in News Headlines

- [British Left] [Waffles] [on] [Falkland Islands]
- [Juvenile Court] [to] [Try] [Shooting Defendant]
- Teachers Strike Idle Kids
- Eye Drops Off Shelf

POS Ambiguity in News Headlines

- [British Left] [Waffles] [on] [Falkland Islands]
- [Juvenile Court] [to] [Try] [Shooting Defendant]
- [Teachers Strike] [Idle Kids]
- Eye Drops Off Shelf

POS Ambiguity in News Headlines

- [British Left] [Waffles] [on] [Falkland Islands]
- [Juvenile Court] [to] [Try] [Shooting Defendant]
- [Teachers Strike] [Idle Kids]
- [Eye Drops] [Off Shelf]

Tagsets

- A compact representation of POS information
  - Usually ≤ 4 capitalized characters
  - Often includes inflectional distinctions

sub category

- Major English tagsets
  - Brown (87 tags)
  - Penn Treebank (45 tags)
  - CLAWS/BNC (61 tags)
  - "Universal" (12 tags) cross lingual, general tags.
- At least one tagset for all major languages

NNPS
NNSP
NNS

Major Penn Treebank Tags

General N
November NNP

NN	noun	VB	verb
JJ	adjective	RB	adverb
DT	determiner the	CD	cardinal number 61
IN	preposition as,	PRP	personal pronoun
MD	modal should will	CC	coordinating conjunction
RP	particle	WH	wh-pronoun
TO	*to*		

Penn Treebank Derived Tags

sub category

NN: NNS (plural, *wombats*), NNP (proper, *Australia*), NNPS (proper plural, *Australians*)

VB: VB (infinitive, *eat*), VBP (1st /2nd person present, *eat*), VBZ (3rd person singular, *eats*), VBD (past tense, *ate*), VBG (gerund, *eating*), VBN (past participle, *eaten*)

JJ: JJR (comparative, *nicer*), JJS (superlative, *nicest*)

RB: RBR (comparative, *faster*), RBS (superlative, *fastest*)

PRP: PRP$ (possessive, *my*)

WH: WH$ (possessive, *whose*), WDT *(wh*-determiner, *who*)*,* WRB (*wh*-adverb, *where*)

Tagged Text Example

The/DT limits/NNS to/TO legal/JJ absurdity/NN stretched/VBD another/DT notch/NN this/DT week/NN
when/WRB the/DT Supreme/NNP Court/NNP refused/VBD to/TO hear/VB an/DT appeal/VB from/IN a/DT case/NN that/WDT says/VBZ corporate/JJ defendants/NNS must/MD pay/VB damages/NNS even/RB after/IN proving/VBG that/IN they/PRP could/MD not/RB possibly/RB have/VB caused/VBN the/DT harm/NN ./.

Why Automatically POS tag?

- Important for morphological analysis, e.g. lemmatisation
- For some applications, we want to focus on certain POS
  - E.g. nouns are important for information retrieval, adjectives for sentiment analysis
- Very useful features for certain classification tasks
  - E.g. genre classification
- POS tags can offer word sense disambiguation
  - E.g. *cross/NN* *cross/VB* *cross/JJ*
- Can use them to create larger structures (parsing)

Automatic Taggers

- Rule-based taggers
- Statistical taggers
  - Unigram tagger
  - Classifier-based taggers
  - Hidden Markov Model (HMM) taggers

Markov Chain

Rule-based tagging

- Typically starts with a list of possible tags for each word
  - From a lexical resource, or a corpus
- Often includes other lexical information, e.g. verb *subcategorisation* (its arguments) transitivity.
- Apply rules to narrow down to a single tag
  - E.g. If DT comes before word, then eliminate VB
  - Relies on some unambiguous contexts
- Large systems have 1000s of constraints

start with

Unigram tagger

- Assign most common tag to each word type
- Requires a corpus of tagged words
- “Model” is just a look-up table
- But actually quite good, ~90% accuracy
  - Correctly resolves about 75% of ambiguity
- Often considered the baseline for more complex approaches

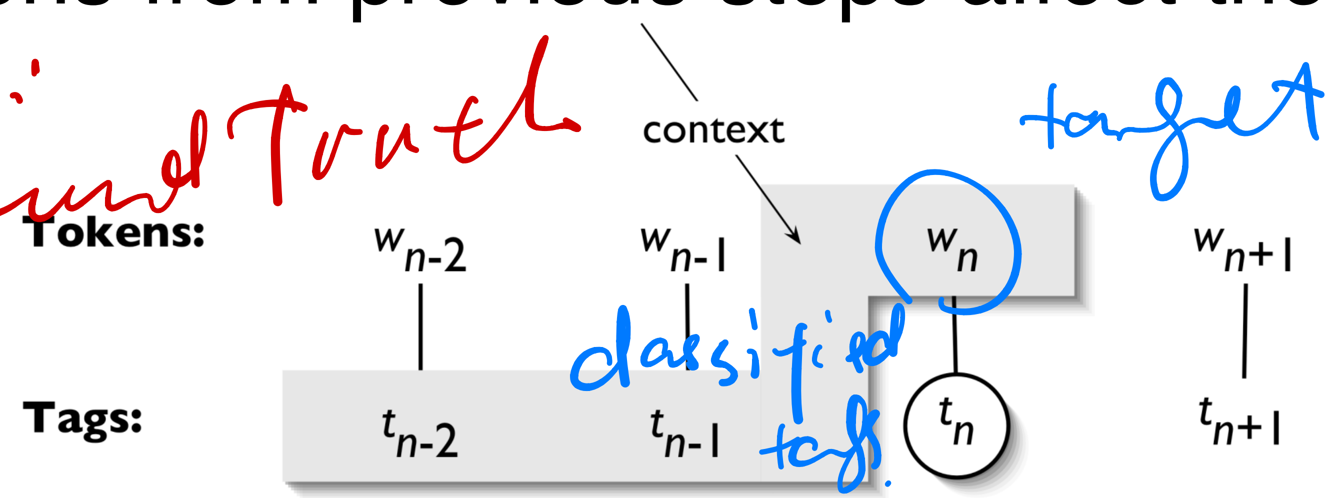
Classifier-Based Tagging

- Use a standard discriminative classifier (e.g. logistic regression, neural network), with features:
  - Target word
  - Lexical context around the word
  - Already classified tags in sentence
- Among the best sequential models
  - But can suffer from **error propagation**: wrong predictions from previous steps affect the next ones

noise propagation

Train time: always ground truth label

target word.



classified tags

Snowball the errors.

Hidden Markov Models

- A basic sequential (or structured) model
- Like sequential classifiers, use both previous tag and lexical evidence
- Unlike classifiers, treat previous tag(s) evidence and lexical evidence as independent from each other
  - Less sparsity
  - Fast algorithms for sequential prediction, i.e. finding the best tagging of entire word sequence

Unknown Words

- Huge problem in morphologically rich languages (e.g. Turkish)

both rare things

- Can use things we've seen only once (hapax legomena) to best guess for things we've never seen before
  - Tend to be nouns, followed by verbs
  - Unlikely to be determiners
- Can use sub-word representations to capture morphology (look for common affixes)

A Final Word

- Part of speech is a fundamental intersection between linguistics and automatic text analysis
- A fundamental task in NLP, provides useful information for many other applications
- Methods applied to it are typical of language tasks in general, e.g. probabilistic, sequential machine learning

Reading

- JM3 Ch. 8 8.1-8.3, 8.5.1

Reading

- JM3 Ch. 8 8.1-8.3, 8.5.1