Part of Speech Tagging

Natural Language Processing: Jordan Boyd-Graber

University of Maryland

Hidden Markov Models

Adapted from material by Ray Mooney

Sequence Models

- The first big difference between ML and NLP
- Historically and pedagogically important
- Application: pos tagging
- What are models that do sequence modeling?

POS Tagging: Task Definition

- Annotate each word in a sentence with a part-of-speech marker.
- Lowest level of syntactic analysis.

John	saw	the	saw	and	decided	to	take	it	to	the	table

- Useful for subsequent syntactic parsing and word sense disambiguation.

POS Tagging: Task Definition

- Annotate each word in a sentence with a part-of-speech marker.
- Lowest level of syntactic analysis.

John	saw	the	saw	and	decided	to	take	it	to	the	table
NNP	VBD	DT	NN	CC	VBD	TO	VB	PRP	IN	DT	NN

- Useful for subsequent syntactic parsing and word sense disambiguation.

POS Tagging: Task Definition

- Annotate each word in a sentence with a part-of-speech marker.
- Lowest level of syntactic analysis.

John	saw	the	saw	and	decided	to	take	it	to	the	table
NNP	VBD	DT	NN	CC	VBD	TO	VB	PRP	IN	DT	NN

- Useful for subsequent syntactic parsing and word sense disambiguation.

POS Tagging: Task Definition

- Annotate each word in a sentence with a part-of-speech marker.
- Lowest level of syntactic analysis.

John	saw	the	saw	and	decided	to	take	it	to	the	table
NNP	VBD	DT	NN	CC	VBD	TO	VB	PRP	IN	DT	NN

- Useful for subsequent syntactic parsing and word sense disambiguation.

What are POS Tags?

- Original Brown corpus used a large set of 87 POS tags.
- Most common in NLP today is the Penn Treebank set of 45 tags. Tagset used in these slides for “real” examples. Reduced from the Brown set for use in the context of a parsed corpus (i.e. treebank).
- The C5 tagset used for the British National Corpus (BNC) has 61 tags.
- Universal Dependencies project tries to make a consistent tag set across languages

Open vs. Closed Class

- Closed class categories are composed of a small, fixed set of grammatical function words for a given language.
  - Pronouns, Prepositions, Modals, Determiners, Particles, Conjunctions
- Open class categories have large number of words and new ones are easily invented.
  - Nouns (Googler, textlish), Verbs (Google), Adjectives (geeky), Abverb (chompingly)

Open Class Tag Examples

- Noun (person, place or thing)
  - Singular (NN): boy, fork
  - Plural (NNS): boys, forks
  - Proper (NNP, NNPS): John, Springfields
- Verb (actions and processes)
  - Base, infinitive (VB): eat
  - Past tense (VBD): ate
  - Gerund (VBG): eating
  - Past participle (VBN): eaten
  - Non 3rd person singular present tense (VBP): eat
  - 3rd person singular present tense: (VBZ): eats
- Adjective (modify nouns)
  - Basic (JJ): red, tall
  - Comparative (JJR): redder, taller
- Adverb (modify verbs)
  - Basic (RB): quickly

Open Class Tag Examples

- Noun (person, place or thing)
  - Singular (NN): boy, fork
  - Plural (NNS): boys, forks
  - Proper (NNP, NNPS): John, Springfields
- Verb (actions and processes)
  - Base, infinitive (VB): eat
  - Past tense (VBD): ate
  - Gerund (VBG): eating
  - Past participle (VBN): eaten
  - Non 3rd person singular present tense (VBP): eat
  - 3rd person singular present tense: (VBZ): eats
- Adjective (modify nouns)
  - Basic (JJ): red, tall
  - Comparative (JJR): redder, taller
- Adverb (modify verbs)
  - Basic (RB): quickly

Open Class Tag Examples

- Noun (person, place or thing)
  - Singular (NN): boy, fork
  - Plural (NNS): boys, forks
  - Proper (NNP, NNPS): John, Springfields
- Verb (actions and processes)
  - Base, infinitive (VB): eat
  - Past tense (VBD): ate
  - Gerund (VBG): eating
  - Past participle (VBN): eaten
  - Non 3rd person singular present tense (VBP): eat
  - 3rd person singular present tense: (VBZ): eats
- Adjective (modify nouns)
  - Basic (JJ): red, tall
  - Comparative (JJR): redder, taller
- Adverb (modify verbs)
  - Basic (RB): quickly

Open Class Tag Examples

- Noun (person, place or thing)
  - Singular (NN): boy, fork
  - Plural (NNS): boys, forks
  - Proper (NNP, NNPS): John, Springfields
- Verb (actions and processes)
  - Base, infinitive (VB): eat
  - Past tense (VBD): ate
  - Gerund (VBG): eating
  - Past participle (VBN): eaten
  - Non 3rd person singular present tense (VBP): eat
  - 3rd person singular present tense: (VBZ): eats
- Adjective (modify nouns)
  - Basic (JJ): red, tall
  - Comparative (JJR): redder, taller
- Adverb (modify verbs)
  - Basic (RB): quickly

Open Class Tag Examples

- Noun (person, place or thing)
  - Singular (NN): boy, fork
  - Plural (NNS): boys, forks
  - Proper (NNP, NNPS): John, Springfields
- Verb (actions and processes)
  - Base, infinitive (VB): eat
  - Past tense (VBD): ate
  - Gerund (VBG): eating
  - Past participle (VBN): eaten
  - Non 3rd person singular present tense (VBP): eat
  - 3rd person singular present tense: (VBZ): eats
- Adjective (modify nouns)
  - Basic (JJ): red, tall
  - Comparative (JJR): redder, taller
- Adverb (modify verbs)
  - Basic (RB): quickly

Tag Examples (cont.)

- Personal pronoun (PRP): I, you, he, she, it
- Wh-pronoun (WP): who, what
- Preposition (IN): on, in, by, to, with
- Determiner:
  - Basic (DT) a, an, the
  - WH-determiner (WDT): which, that
- Coordinating Conjunction (CC): and, but, or,
- Particle (RP): off (took off), up (put up)

Tag Examples (cont.)

- Personal pronoun (PRP): I, you, he, she, it
- Wh-pronoun (WP): who, what
- Preposition (IN): on, in, by, to, with
- Determiner:
  - Basic (DT) a, an, the
  - WH-determiner (WDT): which, that
- Coordinating Conjunction (CC): and, but, or,
- Particle (RP): off (took off), up (put up)

Tag Examples (cont.)

- Personal pronoun (PRP): I, you, he, she, it
- Wh-pronoun (WP): who, what
- Preposition (IN): on, in, by, to, with
- Determiner:
  - Basic (DT) a, an, the
  - WH-determiner (WDT): which, that
- Coordinating Conjunction (CC): and, but, or,
- Particle (RP): off (took off), up (put up)

POS for QA

This Beligan artist painting a train emerging from a fireplace.

POS for QA

This Beligan artist painting a train emerging from a fireplace.

POS for QA

This Beligan artist painting a train emerging from a fireplace.

Lexical Answer Type

This/DET Belgian/ADJ artist/NN

POS for QA

This Beligan artist painting a train emerging from a fireplace.

Lexical Answer Type

This/DET Belgian/ADJ artist/NN

How hard is it?

- Usually assume a separate initial tokenization process that separates and/or disambiguates punctuation, including detecting sentence boundaries.
- Degree of ambiguity in English (based on Brown corpus)
  - 11.5% of word types are ambiguous.
  - 40% of word tokens are ambiguous.
- Average POS tagging disagreement amongst expert human judges for the Penn treebank was 3.5%

Ambiguity

“Like" can be a verb or a preposition

- I like/VBP candy.
- Time flies like/IN an arrow.

“Around” can be a preposition, particle, or adverb

- I bought it at the shop around/IN the corner.
- I never got around/RP to getting a car.
- A new Prius costs around/RB $25K.

What about classification / feature engineering?

- Let's view the context as input
- pos tag is the label
- How can we select better features?

Baseline

- Just predict the most frequent class
- 0.38 accuracy

Prefix and Suffixes

- Take what characters start a word (`un`, `re`, `in`)
- Take what characters end a word (`ly`, `ing`)
- Use as features

Prefix and Suffixes

- Take what characters start a word (`un`, `re`, `in`)
- Take what characters end a word (`ly`, `ing`)
- Use as features (Accuracy: 0.55)

Prefix and Suffixes

- Take what characters start a word (`un`, `re`, `in`)
- Take what characters end a word (`ly`, `ing`)
- Use as features (Accuracy: 0.55)
- What can you do to improve the set of features?

Error Analysis

- Look at predictions of the models
- Look for patterns in frequent errors

Errors from prefix / suffix model

said (372), back (189), get (153), then (147), know (144), Mr. (87), Mike (78)

Error Analysis

- Look at predictions of the models
- Look for patterns in frequent errors

Errors from prefix / suffix model

said (372), back (189), get (153), then (147), know (144), Mr. (87), Mike (78)

Confusion Matrix: Only Capitalization

	JJ	NN	NP	RB	VB
JJ	0	4119	235	0	0
NN	0	14673	713	0	0
NP	0	11	3330	0	0
RB	0	3760	531	0	0
VB	0	12291	338	0	0

Accuracy: 0.45

Incorporating Knowledge

- Use WordNet, an electronic dictionary in nltk
- (We'll talk more about it later)
- Now getting 0.82 accuracy

	JJ	NN	NP	RB	VB
JJ	3064	134	4	310	842
NN	554	13749	463	5	615
NP	90	204	3047	0	0
RB	744	420	314	2361	452
VB	83	1921	164	0	10461

Error Analysis

back	then	now	there	here	still	long	thought	want	even
223	145	140	116	115	100	99	88	79	67

A more fundamental problem . . .

- Each classification is independent . . .
- This isn't right!
- If you have a noun, it's more likely to be preceeded by an adjective
- Determiners are followed by either a noun or an adjective
- Determiners don't follow each other

这只丑狗是我的

Zhèzhǐ chǒu gǒu shè wǒde

这只丑狗是我的

Zhèzhǐ chǒu gǒu shè wǒde
PRP$

这只丑狗是我的

Zhèzhǐ	chǒu	gǒu	shè	wǒde
		NN		PRP$

这只丑狗是我的

Zhèzhǐ	chǒu	gǒu	shè	wǒde
		NN	VB	PRP$

这只丑狗是我的

Zhèzhǐ	chǒu	gǒu	shè	wǒde
DT	JJ	NN	VB	PRP$

Approaches

- Generative Statistical models $p(x, y)$: Hidden Markov Model (HMM)
- Discriminative Statistical models $p(y | x)$: Conditional Random Field (CRF), structured perceptron
- Neural sequence models: RNN / LSTM
- Transformers: BERT

Approaches

- Generative Statistical models $p(x, y)$: Hidden Markov Model (HMM)
- Discriminative Statistical models $p(y | x)$: Conditional Random Field (CRF), structured perceptron
- Neural sequence models: RNN / LSTM
- Transformers: BERT

Approaches

- Generative Statistical models $p(x, y)$: Hidden Markov Model (HMM)
- Discriminative Statistical models $p(y|x)$: Conditional Random Field (CRF), structured perceptron
- Neural sequence models: RNN / LSTM
- Transformers: BERT

Approaches

- Generative Statistical models $p(x, y)$: Hidden Markov Model (HMM)
- Discriminative Statistical models $p(y \mid x)$: Conditional Random Field (CRF), structured perceptron
- Neural sequence models: RNN / LSTM
- Transformers: BERT

Approaches

- Generative Statistical models $p(x, y)$: Hidden Markov Model (HMM)
- Discriminative Statistical models $p(y \mid x)$: Conditional Random Field (CRF), structured perceptron
- Neural sequence models: RNN / LSTM
- Transformers: BERT

What else can sequence models do?

- Find named entities: We went to New York City
- Find answers: The murderer was played by John Lithgow

What else can sequence models do?

- Find named entities: We/O went/O to/O New/B York/I City/I
- Find answers: The murderer was played by John Lithgow

What else can sequence models do?

- Find named entities: We/O went/O to/O New/B York/I City/I
- Find answers: The murderer was played by John Lithgow

What else can sequence models do?

- Find named entities: We/O went/O to/O New/B York/I City/I
- Find answers: The/O murderer/O was/O played/O by/O John/B Lithgow/I

What else can sequence models do?

- Find named entities: We/O went/O to/O New/B York/I City/I
- Find answers: The/O murderer/O was/O played/O by/O John/B Lithgow/I