CS 378 Lecture 9: Sequence Labeling, POS, HMMs

Announcements

- A2 due Thurs
- A1 back soon
- Bias response due Thurs

Today Set up the problem of sequence labeling

Example: part-of-speech tagging

Why sequence models?

Recap Course so far:

Input / language
- bag-of-words (tokenization, etc.)
- word embeddings

↓

Model
- LR
- FFNNs

↓

(Multiclass) predictions 2~20 classes

This lecture:

Sequential predictions

Part-of-speech tagging

Input: sentence $x_1, \ldots, x_n$

Output: POS tags $y_1, \ldots, y_n$
for each word

Why POS?

find the action verb

Teacher	strikes	idle	Kids
N	N	V	N
N	V	J	N
		adjective	

Text-to-speech: record

POS tags

Open-class: new words with these tags are always emerging

Closed-class: (function words) tags with a known fixed set of words

Open-class

(N) Nouns: Proper: Google
Common: cat, company
plural vs. singular

(V) Verbs: see, registered, (Google)
tense, subject agreement

(J) Adjectives: yellow

(RB) Adverbs: swiftly

Closed-class

(DT) Determiners: the, a (articles)

DT + N ⇒ NP

some, many

Cardinal: 1, 27, --

Conjunctions: and, or

Prepositions: up, on, in, to

Particles: made up

Auxiliary: had

Modal verbs: could / would / should

Fed raises interest rates 0.5 percent

Fed raises interest rates 0.5 percent

Fed
- VBD I fed the cat
- NNP
- VBN I had fed

raises
- NNS plural
- VBZ 3rd person present verb

interest
- NN
- VBP present "I interest you"
- VB infinitive "I want you to interest me"

rates
- NNS
- VBZ

alternate

0.5 CD

percent NN

Tagging with classifiers

Input: $\bar{x} = (x_1, \ldots, x_n)$

Position: i

Output: y_i (tag at i)

MCLR: $P(y_i = t \mid \bar{x})$ run for $i = 1 \ldots n$

$P(y_3 = N \mid \text{Fed raises interest} \ldots)$

Features: (1) bag-of-words ✗ DOES NOT work

$f(\bar{x}) = [\,0\;0\;1\;0\;0\;0\;1\;\ldots\,]$

raises interest

$f(\bar{x})$ doesn't depend on i, ignores order

$P(y_3 \mid \bar{x})$ vs. $P(y_2 \mid \bar{x})$ same? no!

(2) Features that depend on i

$f(\bar{x}, i)$ look at x_i and words around it

$f(\bar{x}, i=3) = [\ \ldots\ \underset{\text{interest}}{1}\ \ldots\ ]$

one-hot vector for x_3

$f(\bar{x}, i=3)$ {
- PrevWord = raises
- CurrWord = interest
- NextWord = rates

$f(\bar{x}, i=3) = [\ 0\ \ 0\ \ 1$

PrevWord = raises

"bag of positional words"

$0\ \ 0\ \ 0\ \ 1$

CurrWord = interest

$0\ \ \ldots\ ]$

"what score does CW = interest add to N"?

weights

$\bar{w}_N = [$ +1 $]$

CurrWord = interest

$i = 2$:

$f(\bar{x}, i=2) = [$ 1 0 0

CurrWord = raises

1 0 _ _ $]$

PrevW = Fed

Tagger:

Fed $\rightarrow f(\bar{x}, i=1) \rightarrow$ MCLR $\rightarrow y_1$

raises $\rightarrow f(\bar{x}, i=2) \rightarrow$ MCLR $\rightarrow y_2$

⋮

Problems with this

VBZ VBP

↑ ↑

Fed raises interest rates

Should not have 2 V tags adjacent

How to prohibit this?

① Incremental approach

$$y_1 \rightarrow y_2 | y_1 \rightarrow y_3 | y_1 y_2 \cdots$$

Problem: greedy

We want to model + predict the sequence

Hidden Markov Models

$P(\bar{y}, \bar{x})$ model whole sequence jointly