CS378 Lecture 9: Sequence Labeling, POS, HMMs

Announcements

- A2 due Tues — Bias in embs response
- A1 back soon

Survey

- Lower Zoom volume
- Exercises / pace
- Ling / language
- More ~~state-of-the-art~~

Recap

Bias: sim(she, receptionist) > sim(he, receptionist)

Multilingual: dictionaries can help us align embeddings across languages

Today Sequence labeling
POS: definitions
Why sequence models for this?
HMM intro (if time)

Where we are

Classification: $\underset{y}{\text{argmax}}\ \bar{w}^T y\, \underbrace{f(\bar{x})}_{\text{NN or bag of words}}$

$\bar{x}$

~ sent

or


doc

→ y label binary/multi

Tagging: $y_1\ y_2\ y_3\ y_4$ label for each word

sent

Part-of-speech tagging

Input: sentence $x_1, \ldots, x_n$

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

Why POS?

Teacher	strikes	idle	kids
N	N	V	
N	V	ADJ	

disambiguate meaning

Text-to-speech: record

POS Tags

Open-class tags: new words with these tags are always coming out

Closed-class tags: more like function words in English

Open-class

(N) Nouns:
- Proper (NNP): Google
- Common (NN, NNS): cat, company

(V) Verbs: see, registered

Adjectives: yellow

Adverbs: swiftly

Closed-class

Determiners: the, a (articles), some, many

DT + N ⇒ NP

Conjunctions: and, or

Prepositions: up, on, in, to, ...

Particles: made up

Auxiliary verbs: had [V]

Modal verbs: could / would / should

Fed raises interest rates 0.5 percent

Fed — NNP proper noun
VBD past tense verb (I fed)
VBN participial (I had fed)

raises — NNS plural noun
VBZ 3rd person present verb

interest — NN
VBP I interest you
VB infinitive: to interest (I want you to interest me)

rates — NNS
VBZ

0.5 — CD cardinal

percent — NN

correct

alt

Tagging with classifiers

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

Output: ith position
label y_i at that position

(MC)
LR: $P(y_i = y \mid \bar{x})$, run for $i = 1 \ldots n$

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

① BoW features ✗ DOESN'T WORK

$f(\bar{x}) = [0\ 0\ \underset{\text{raises}}{1}\ 0\ 0\ \underset{\text{Fed}}{1}\ \ldots]$

$P(y_3 = N \mid \bar{x})$ same as $P(y_2 = N \mid \bar{x})$

same $f(\bar{x})$, doesn't know about i

Fed raises interest rates...
1 2 3 4

(2) Features that depend on i

$f(\bar{x}, i)$

$f(x_i)$ ← looks at one word, but we want word in context

Look at words "around" position i

$f(\bar{x}, i=3) = \begin{cases} \text{PrevWord} = \text{raises} \\ \text{CurrWord} = \text{interest} \\ \text{NextWord} = \text{rates} \end{cases}$

$f(\bar{x}, i=4)$...

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

"bag of positional words"

PrevWord=raises

CurrWord=interest

CurrWord=raises

$3 \cdot |V|$ feats

start-of-sentence

$f(\bar{x}, i=1)$ { PrevWord = <S>
--
--

y_1 (N) y_2 y_3

MC LR MC LR MC LR

$f(\bar{x}, i=1)$ $f(\bar{x}, i=2)$ $f(\bar{x}, i=3)$ ...

Fed raises interest

Problems with this

VBZ VBP ✗ wrong

Fed raises interest rates

We know this is wrong. VBZ–VBP will rarely happen

We want to prohibit this... how?

We want to model sequences

$P(\bar{y} \mid \bar{x})$ (MC LR: $\prod_{i=1}^{n} P(y_i = y \mid \bar{x})$)

$\bar{y}$: seq of labels

Structured prediction

predicting a sequence, tree, graph, ...

Hidden Markov Model: $P(\bar{x}, \bar{y})$

$\bar{x}$: words, $\bar{y}$: tags