Deep Learning for Vision & Language

Natural Language Processing I: Introduction

Natural Language Processing

The study of automatic reasoning over text / language

- Fundamental goal: *deep* understand of *broad* language
  - Not just string processing or keyword matching!
- End systems that we want to build:
  - Ambitious: speech recognition, machine translation, information extraction, dialog interfaces, question answering…
  - Modest: spelling correction, text categorization…

Slide by Dan Klein

Why is NLP Hard?

- Human Language is Ambiguous

Task: Pronoun Resolution

- Jack drank the wine on the table. ***It*** was red and round.
- Jack saw Sam at the party. ***He*** went back to the bar to get another drink.
- Jack saw Sam at the party. ***He*** clearly had drunk too much.

[Adapted from Wilks (1975)]

Slide by Yejin Choi

Why is NLP Hard?

- Human Language Requires World Knowledge

Task: Co-Reference Resolution

- The doctor hired a secretary because she needed help with new patients.
- The physician hired the secretary because he was highly recommended.

[From some of our group's work]

Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods
Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, Kai-Wei Chang.
North American Chapter of the Association for Computational Linguistics. **NAACL 2018**.

Adapted from a slide by Yejin Choi

Why is NLP Hard?

- Human Language is Ambiguous

  Learning mother tongue (native language)

  -- you might think it's easy, but…

  - compare 5 year old V.S. 10 year old V.S. 20 year old

- Learning foreign languages

  •– even harder

Slide by Yejin Choi

Word Segmentation

- Breaking a string of characters into a sequence of words.
- In some written languages (e.g. Japanese) words are not separated by spaces.
- Even in English, characters other than white-space can be used to separate words [e.g. , ; . - : ( )]
- Examples from English URLs:
  - jumptheshark.com ⇒ jump the shark .com
  - myspace.com/pluckerswingbar
    ⇒ myspace .com pluckers wing bar
    ⇒ myspace .com plucker swing bar

Slide from Ray Mooney

Morphological Analysis

- ***Morphology*** is the field of linguistics that studies the internal structure of words. (Wikipedia)
- A ***morpheme*** is the smallest linguistic unit that has semantic meaning (Wikipedia)
  - e.g. “carry”, “pre”, “ed”, “ly”, “s”
- Morphological analysis is the task of segmenting a word into its morphemes:
  - carried $\Rightarrow$ carry + ed (past tense)
  - independently $\Rightarrow$ in + (depend + ent) + ly
  - Googlers $\Rightarrow$ (Google + er) + s (plural)
  - unlockable $\Rightarrow$ un + (lock + able) ?

    $\Rightarrow$ (un + lock) + able ?

Slide from Ray Mooney

- ***German***

555 --> fünfhundertfünfundfünfzig

7254 → Siebentausendzweihundertvierundfünfzig

Part Of Speech (POS) Tagging

- Annotate each word in a sentence with a part-of-speech.

I ate the spaghetti with meatballs.

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

- Useful for subsequent syntactic parsing and word sense disambiguation.

Slide from Ray Mooney

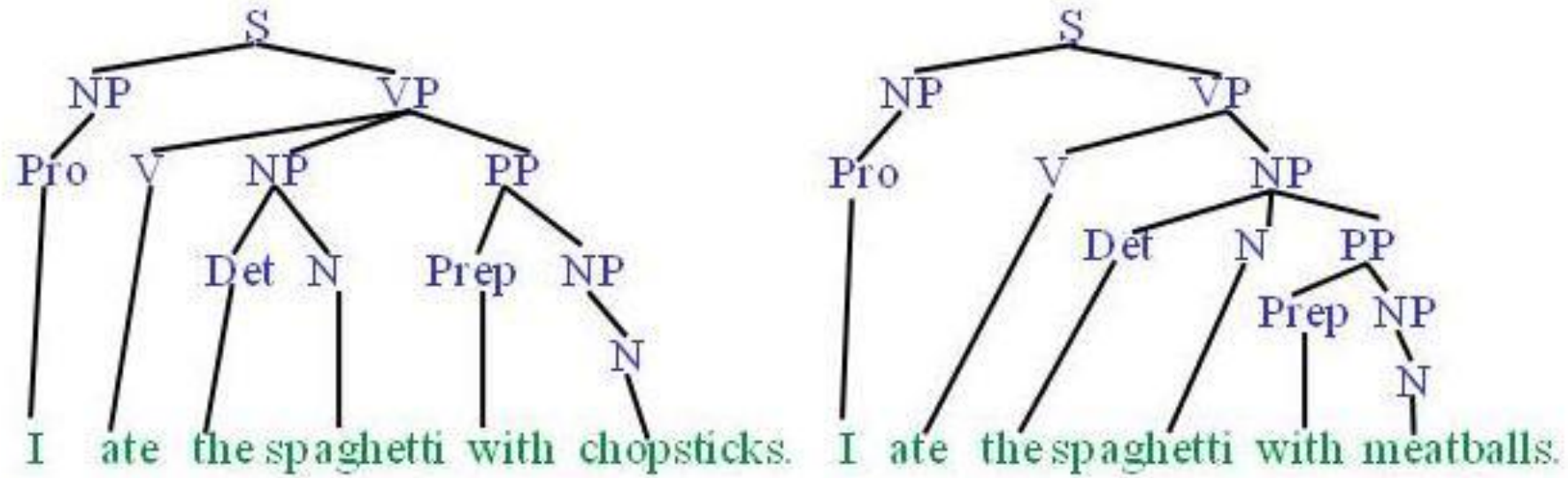
Phrase Chunking

- Find all noun phrases (NPs) and verb phrases (VPs) in a sentence.
  - [NP I] [VP ate] [NP the spaghetti] [PP with] [NP meatballs].
  - [NP He] [VP reckons] [NP the current account deficit] [VP will narrow] [PP to] [NP only # 1.8 billion] [PP in] [NP September]

Slide from Ray Mooney

Syntactic Parsing

- Produce the correct syntactic parse tree for a sentence.



Slide from Ray Mooney

Word Sense Disambiguation (WSD)

- Words in natural language usually have a fair number of different possible meanings.
  - Ellen has a strong interest in computational linguistics.
  - Ellen pays a large amount of interest on her credit card.
- For many tasks (question answering, translation), the proper sense of each ambiguous word in a sentence must be determined.

Slide from Ray Mooney

Other tasks more “advanced” tasks:

- Text classification
  - Simple: Is this text English or Spanish?
  - Harder: Is this text about Toys or Weapons?
  - Even harder: Is this text written by Shakespeare or not?
- Text Generation
  - Free form: Generate arbitrary human-like produced text
  - Conditional: Generate text that has certain style
- Entailment, can text statement A be deduced from text statement B?
- Question Answering (QA), Given a question text, give a text answer.
- Dialog – same as VQA but continuous back and forth

How to represent a word?

		one-hot encodings
dog	1	[1 0 0 0 0 0 0 0 0 0]
cat	2	[0 1 0 0 0 0 0 0 0 0]
person	3	[0 0 1 0 0 0 0 0 0 0]
holding	4	[0 0 0 1 0 0 0 0 0 0]
tree	5	[0 0 0 0 1 0 0 0 0 0]
computer	6	[0 0 0 0 0 1 0 0 0 0]
using	7	[0 0 0 0 0 0 1 0 0 0]

How to represent a word?

How to represent a phrase/sentence?

bag-of-words representation

person holding dog	{1, 3, 4}	[1 0 1 1 0 0 0 0 0 0]
person holding cat	{2, 3, 4}	[0 1 1 1 0 0 0 0 0 0]
person using computer	{3, 7, 6}	[0 0 1 0 0 1 1 0 0 0]

dog, cat, person, holding, tree, computer, using

person using computer person holding cat	{3, 3, 7, 6, 2}	[0 1 2 1 0 1 1 0 0 0]

What if vocabulary is very large?

Sparse Representation

bag-of-words representation

person holding dog	{1, 3, 4}	indices = [1, 3, 4]	values = [1, 1, 1]
person holding cat	{2, 3, 4}	indices = [2, 3, 4]	values = [1, 1, 1]
person using computer	{3, 7, 6}	indices = [3, 7, 6]	values = [1, 1, 1]

person using computer person holding cat	{3, 3, 7, 6, 2}	indices = [3, 7, 6, 2]	values = [2, 1, 1, 1]

Recap

- Bag-of-words encodings for text (e.g. sentences, paragraphs, captions, etc)

You can take a set of sentences/documents and classify them, cluster them, or compute distances between them using this representation.

Problem with this bag-of-words representation

my friend makes a nice meal

These would be the same using bag-of-words

my nice friend makes a meal

Bag of Bi-grams

indices = [10132, 21342, 43233, 53123, 64233]
values = [1, 1, 1, 1, 1]

my friend makes a nice meal

{my friend, friend makes, makes a,

a nice, nice meal}

indices = [10232, 43133, 21342, 43233, 54233]
values = [1, 1, 1, 1, 1]

my nice friend makes a meal

{my nice, nice friend, friend makes,

makes a, a meal}

A dense vector-representation would be very inefficient
Think about tri-grams and n-grams

Recommended reading: n-gram language models

Yejin Choi's course on Natural Language Processing

http://www3.cs.stonybrook.edu/~ychoi/cse628/lecture/02-ngram.pdf

Questions?