Deep Learning in Lexical Analysis and Parsing

Wanxiang Che (HIT) and Yue Zhang (SUTD)

Outline

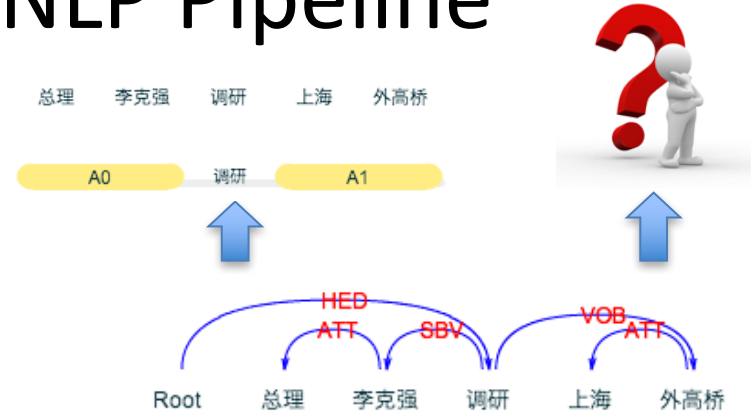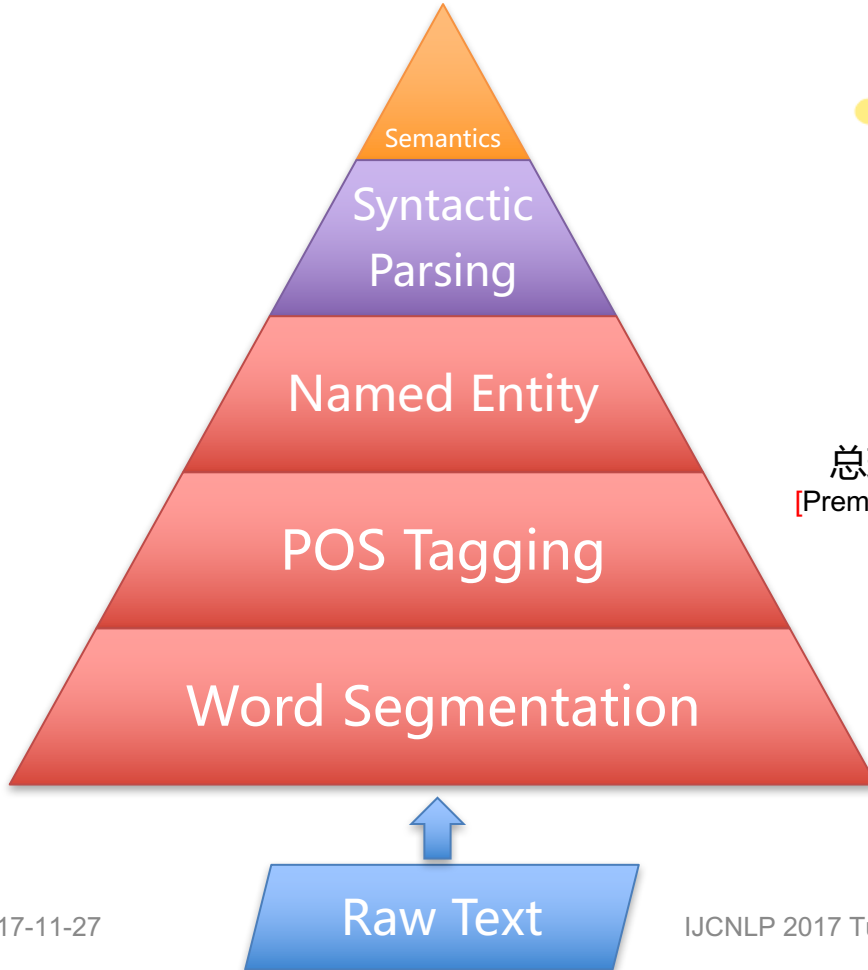| Content |
|---|
| Task Introduction |
| Deep Learning Background |
| Main Statistical Methods |
| Break |
| Neural Graph-based Methods |
| Neural Transition-based Methods |
| Applications |

Part 1: Tasks Introduction

Part 1.1: Lexical Analysis and Parsing

Fundamental NLP Pipeline

Semantics

Syntactic Parsing

Named Entity

POS Tagging

Word Segmentation

Raw Text

总理 李克强 调研 上海 外高桥

A0 | 调研 | A1

HED
ATT   SBV   VOB ATT

Root   总理   李克强   调研   上海   外高桥

总理/n [李克强 **人名**] 调研/v [上海 外高桥 **地名**]
[Premier]/n [Li Keqiang]/**Name** [study]/v [Shanghai Waigaoqiao]/**Location**

总理/n 李克强/nh 调研/v 上海/ns 外高桥/ns
[Premier]/n [Li Keqiang]/nh [study]/v [Shanghai]/ns [Waigaoqiao]/ns

总理 李克**强 调**研 上海 外高桥
Premier Li Keqiang sutdy Shanghai Waigaoqiao

总理李克强调研上海外高桥
Premier Li Keqiang study Shanghai Waigaoqiao

IJCNLP 2017 Tutorial

Word Segmentation

- Words are fundamental semantic units
- Chinese has no obvious word boundaries
- Word segmentation
  - Split Chinese character sequence into words
- Ambiguities in word segmentation
  - E.g. 严守一把手机关了
    - 严守一(name)/ 把(ba)/ 手机(mobile)/ 关(turn off)/ 了(le)
    - 严守(name)/ 一把手(first-leader)/ 机关(office)/ 了(le)
    - 严守(name)/ 一把(one time)/ 手机(mobile)/ 关(turn off)/ 了(le)
    - 严守一(name)/ 把手(handle)/ 机关(office)/ 了(le)
    - ……
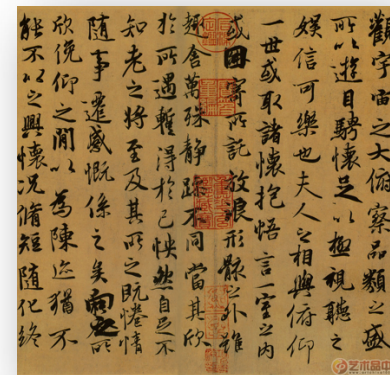
Part-of-speech (POS) Tagging

- A POS is a category of words which have similar grammatical properties
  - E.g. noun, verb, adjective
- POS tagging
  - Marking up a word in a text as a particular POS
  - based on both its definition and its context
- Ambiguities in POS Tagging
  - Time flies like an arrow.
  - 制服(subdue)了敌人 vs. 穿着制服(uniform)

Named Entity Recognition (NER)

- Named Entities
  - Persons, locations, organizations, expressions of times, quantities, monetary values, percentages, etc.
- Locating and classifying named entities in text into pre-defined categories
- Ambiguities in NER

Kerry to visit Jordan, Israel Palestinian peace on agenda.

Jordan

Person name(乔丹)？

Location name(约旦)？ ✓

Syntactic Parsing

- Analyzing a natural language string conforming to the rules of a formal grammar, emphasizing subject, predicate, object, etc.
  - Constituency and Dependency Parsing
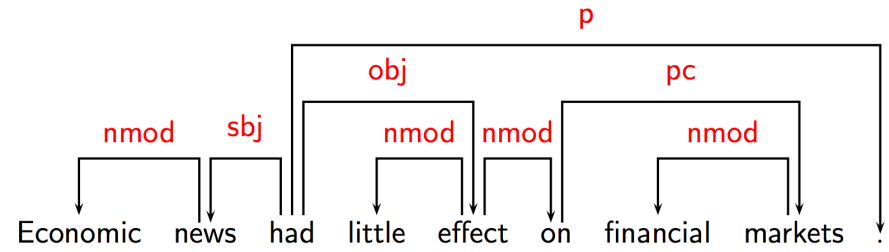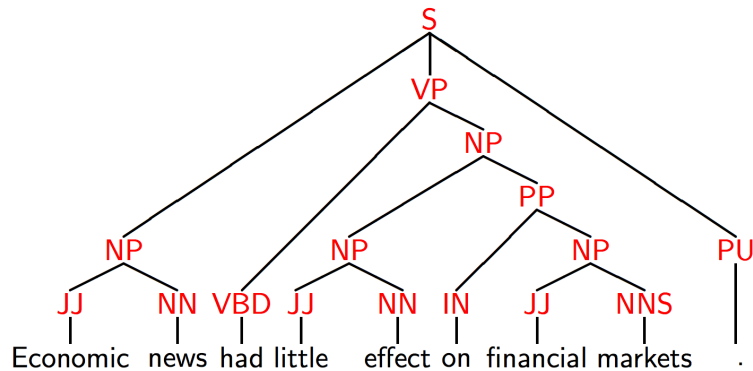
Semantic Role Labeling

- Recognizing predicates and corresponding arguments

TEMP   HITTER   THING HIT   INSTRUMENT

Yesterday, Kristina hit Scott with a baseball

Scott was hit by Kristina yesterday with a baseball

Yesterday, Scott was hit with a baseball by Kristina

With a baseball, Kristina hit Scott yesterday

Yesterday Scott was hit by Kristina with a baseball
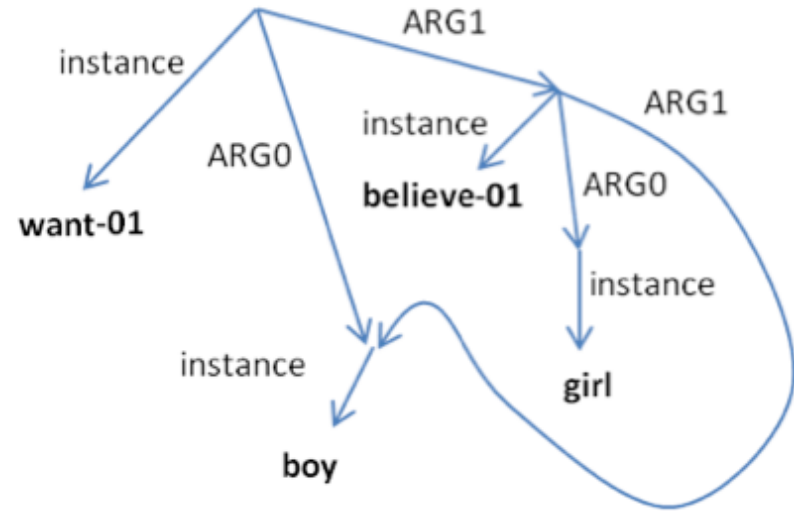
Kristina hit Scott with a baseball yesterday

Example from (Yih & Toutanova, 2006)

Semantic Role Labeling

- Answer "Who did what to whom when and where"
  - Question Answering
    - Yesterday $_{time}$, Mary $_{buyer}$ bought a shirt $_{bought\ thing}$ from Tom $_{seller}$
    - Whom $_{buyer}$ did Tom $_{seller}$ sell a shirt $_{bought\ thing}$ to, yesterday $_{time}$
  - Information Extraction
  - ……

Abstract Meaning Representation (AMR)

The boy wants the girl to believe him.
The boy wants to be believed by the girl.
The boy has a desire to be believed by the girl.
The boy's desire is for the girl to believe him.
*The boy is desirous of the girl believing him.*

http://www.isi.edu/~ulf/amr/help/amr-guidelines.pdf

Combinatory Categorial Grammars (CCG)

$$\frac{\begin{array}{ccc} \text{CCG} & \text{is} & \text{fun} \\ \hline \begin{array}{c} NP \\ CCG \end{array} & \begin{array}{c} S\backslash NP/ADJ \\ \lambda f.\lambda x.f(x) \end{array} & \begin{array}{c} ADJ \\ \lambda x.fun(x) \end{array} \end{array}}{\cfrac{S\backslash NP}{\cfrac{\lambda x.fun(x)}{\begin{array}{c} S \\ fun(CCG) \end{array}}<}>}$$

- CCG Lexical Entries
  - Pair words and phrases with meaning by a CCG category

  fun $\vdash ADJ : \lambda x.fun(x)$

  Natural Language | CCG Category

- CCG Categories
  - Basic building block
  - Capture syntactic and semantic information jointly

  Syntax $ADJ : \lambda x.fun(x)$ Semantics
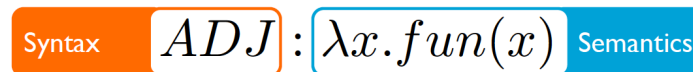
Part 1.2: Structured Prediction

Structured Prediction

- Predicting structured objects, rather than single value
- Output structures influence each other
- Categories
  - Sequence segmentation
  - Sequence labeling / Tagging
  - Trees
  - Graphs

Sequence Segmentation

- Break a sequence into contiguous parts
- For example: Word Segmentation
  - Input
    - 严守一把手机关了
  - Output
    - 严守一/ 把/ 手机/ 关/ 了/
- More examples:
  - Sentence segmentation
  - Paragraph segmentation
  - NER

Sequence Labeling/Tagging

- Given an input sequence, produce a label sequence of equal length
- Each label is drawn from a small finite set
- Label influence each other
- For example: POS tagging
  - Input
    - Profits soared at Boeing Co., easily topping forecasts on Wall Street, …
  - Output
    - Profits/N soared/V at/P Boeing/N Co./N ,/, easily/ADV …

Word Segmentation as Sequence Labeling

- Input
  - 严守一把手机关了
- Output
  - 严守一/ 把/ 手机/ 关/ 了/
- Alternative Output (Tagging)
  - 严/B 守/I 一/I 把/B 手/B 机/I 关/B 了/B
- Where

  - B: Begin of a word; I: Inside of a word

NER as Sequence Labeling

- Input
  - Profits soared at Boeing Co., easily topping forecasts on Wall Street, …
- Output
  - Profits soared at [Boeing Co. ORG], easily topping forecasts on [Wall Street LOC], …
- Alternative Output (Tagging)
  - Profits/O soared/O at/O Boeing/B-ORG Co./I-ORG ,/O easily/O topping/O forecasts/O on/O Wall/B-LOC Street/I-LOC ,/O …
- Where
  - B: Begin of entity XXX; I: Inside of entity XXX; O: Others

Semantic Role Labeling as Sequence Labeling

- Input
  - Yesterday, Mary bought a shirt from Tom
- Output
  - [Yesterday $_{time}$], [Mary $_{buyer}$] bought/pred [a shirt $_{merchandise}$] from [Tom $_{seller}$]
- Alternative Output (Tagging)
  - Yesterday/B-time ,/O Mary/B-buyer bought/pred a/B-merchandise shirt/I- merchandise from/O Tom/B-seller
- Where
  - B: Begin; I: Inside; O: Others

CCG Supertagging as Sequence Labeling

$$\frac{He}{NP} \quad \frac{goes}{(S[dcl]\backslash NP)/PP} \quad \frac{on}{PP/NP} \quad \frac{the}{NP/N} \quad \frac{road}{N} \quad \frac{with}{((S\backslash NP)\backslash(S\backslash NP))/NP} \quad \frac{his}{NP/N} \quad \frac{piano}{N}$$

$$\frac{A}{NP/N} \quad \frac{bitter}{N/N} \quad \frac{conflict}{N} \quad \frac{with}{(NP\backslash NP)/NP} \quad \frac{global}{N/N} \quad \frac{implications}{N}$$

frequency cut-off	# cat types	# cat tokens in 2-21 not in cat set		# sentences in 2-21 with missing cat		# cat tokens in 00 not in cat set		# sentences in 00 with missing cat	
1	1 225	0		0		12	(0.03%)	12	(0.6%)
10	409	1 933	(0.2%)	1 712	(4.3%)	79	(0.2%)	69	(3.6%)

Trees

- All kinds of algorithms converting sentences to tree or graph structures
  - Constituency and Dependency Parsing

Conclusion

- NLP Tasks
  - Word segmentation, POS tagging, named entity recognition
  - Constituent/dependency parsing
  - Semantic (graph) dependency parsing
  - AMR
- Structured Prediction
  - Sequence segmentation
  - Sequence labeling / Tagging
  - Trees