Deep Learning in Lexical Analysis and Parsing

Wanxiang Che (HIT) and Yue Zhang (SUTD)
<table>
<thead>
<tr>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task Introduction</td>
</tr>
<tr>
<td>Deep Learning Background</td>
</tr>
<tr>
<td>Main Statistical Methods</td>
</tr>
<tr>
<td>Break</td>
</tr>
<tr>
<td>Neural Graph-based Methods</td>
</tr>
<tr>
<td>Neural Transition-based Methods</td>
</tr>
<tr>
<td>Applications</td>
</tr>
</tbody>
</table>
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**

- **Premier Li Keqiang**
- **study**
- **Shanghai Waigaoqiao**

**Total**

- **Premier Li Keqiang**
- **study**
- **Shanghai Waigaoqiao**

**Premier Li Keqiang study Shanghai Waigaoqiao**
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.
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

Example from (Yih & Toutanova, 2006)

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
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.

Combinatory Categorial Grammars (CCG)

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

- CCG Categories
  - Basic building block
  - Capture syntactic and semantic information jointly
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. \texttt{ORG}], easily topping forecasts on [Wall Street \texttt{LOC}], ...

• Alternative Output (Tagging)
  – Profits/\texttt{O} soared/\texttt{O} at/\texttt{O} Boeing/B-\texttt{ORG} Co./I-\texttt{ORG} ,/\texttt{O} easily/\texttt{O} topping/\texttt{O} forecasts/\texttt{O} on/\texttt{O} Wall/B-\texttt{LOC} Street/I-\texttt{LOC} ,/\texttt{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
  – \([\text{Yesterday}_{\text{time}}], [\text{Mary}_{\text{buyer}}] \text{ bought/pred } [\text{a shirt}_{\text{merchandise}}] \text{ from } [\text{Tom}_{\text{seller}}]\)

• 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

He goes on the road with his piano

A bitter conflict with global implications

<table>
<thead>
<tr>
<th>frequency cut-off</th>
<th># cat types</th>
<th># cat tokens in 2-21 not in cat set</th>
<th># sentences in 2-21 with missing cat</th>
<th># cat tokens in 00 not in cat set</th>
<th># sentences in 00 with missing cat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1225</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>409</td>
<td>1933 (0.2%)</td>
<td>1712 (4.3%)</td>
<td>79</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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