Recent Advances in Parsing

Wanxiang Che
Harbin Institute of Technology
2016-10
An Example

山寨发布会阳淼
@️才看到。昨天手机打字，把“您转的这篇文章很无知”打成了“您转这篇文章很无知”，少了一个的字。抱歉。

山寨发布会阳淼
主语是那篇文章很无知。
Syntactic Parsing

• Analyzing a natural language string conforming to the rules of a formal grammar, emphasizing subject, predicate, object, etc.

• A traditional and core NLP task

http://www.ltp-cloud.com/demo/
Elements of Parsing

Algorithm

Grammar

Data
Constituency vs. Dependency
Syntactic vs. Stanford Dependency

Semantic Dependency Graph

ROOT 现在 她 脸色 难看 Exp
time Poss Exp mPunc mMod mTone
ROOT now she face terrible-looking seem disease already

http://www.ltp-cloud.com/intro/#sdp_how
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)

\[
\begin{align*}
\text{CCG} & \quad \text{is} \quad \text{fun} \\
NP & \quad S \setminus NP/ADJ \quad ADJ \\
CCG & \quad \lambda f . \lambda x . f(x) \quad \lambda x . fun(x) \\
& \quad \frac{S \setminus NP}{\lambda x . fun(x)} \quad \frac{S}{fun(CCG)} \\
\end{align*}
\]

http://groups.inf.ed.ac.uk/ccg/
Data

- Monolingual Single-domain
- Multilingual
- Multi-domain
- Universal Treebank

Rich-resource to Low-resource
Multilingual

• CoNLL 2007 Shared Tasks
  – Multi-lingual Dependency Parsing
  – 12 languages
  – http://ilk.uvt.nl/conll/

• CoNLL 2009 Shared Tasks
  – Syntactic and Semantic Dependencies in Multiple Languages
  – 7 languages
Multi-domain

• Syntactic Analysis of Non-Canonical Language (SANCL) 2012 Shared Task
  – https://sites.google.com/site/sancl2012/
  – Organized by Google

• Corpus
  – Google Web Treebank
  – Three web domains: CQA, Newsgroup, Online Review
Universal Treebank

- [http://universaldependencies.org/](http://universaldependencies.org/)
  - 40+ languages with **universal** dependencies and POS tags
- For multi-lingual and cross-lingual research
  - “Many Languages, One Parser” – CMU
- CoNLL 2017 Shared Task
  - Multilingual Parsing from Raw Text to Universal Dependencies
Transition-based Dependency Parsing

- Greedily predict a transition sequence from an initial parser state to some terminal states
- State (configuration)
  - $= \text{Stack} + \text{Buffer} + \text{Dependency Arcs}$

![Diagram showing configuration of Stack, Buffer, and Arcs with example sentence and transitions]

CCL 2016 Technical Review
Traditional Features

Feature Vector:
- Binary
- Sparse
- High-dimensional

Feature templates: a combination of elements from the configuration.
- For example: (Zhang and Nivre, 2011): 72 feature templates

Table 1: Baseline feature templates.

\[ \text{distance} \]
\[ w \text{ - word; } p \text{ - POS-tag.} \]

\[ \text{valency} \]
\[ \text{unigrams} \]
\[ \text{third-order} \]
\[ \text{label set} \]
Chen and Manning NN Parser

Softmax layer:
\[ p = \text{softmax}(W_2h) \]

Hidden layer:
\[ h = (W^w_x^w + W^t_x^t + W^l_x^l + b)_3 \]

Input layer: \([x^w, x^t, x^l]\)

Configuration

Stack
ROOT has_VBZ good_JJ
He_PRP

POS tags

Buffer
control_NN ...

Stack LSTM Parser


CCL 2016 Technical Review
Sentence-level Log Likelihood

\[ p(y_i \mid x, \theta) = \frac{e^{f(x, \theta)_i}}{\sum_{y_j \in \text{GEN}(x)} e^{f(x, \theta)_j}} \]

\[ f(x, \theta)_i = \sum_{a_k \in y_i} o(x, y_i, k, a_k) \]

SyntaxNet: Google

Training with Beam Search:

Sum scores of all decisions across entire history...

Incorrect parse

Correct parse

Update: maximize P(correct parse) relative to the set of alternatives

Globally Normalized SyntaxNet Architecture (Overview)

Changes of Performance

Test on PTB with Stanford Dependency


- Zhang & McDonald (2014)
- Chen & Manning (2014)
- Dyer et al. (2015)
- Zhou et al. (2015)
- Andor et al. (2016)

CCL 2016 Technical Review 22
Encoder-decoder Parsing


But it was the Quotron problems that...

Encoding:  But it was the Quotron problems that ...

Decoding:  But it was @L SBJ @L DEP the Quotron problems @L NMOD @L NMOD that ...

Some Open Questions

• Is Parsing Necessary?

<table>
<thead>
<tr>
<th></th>
<th>Bi-LSTM</th>
<th>Tree-LSTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stanford Sentiment TreeBank</td>
<td>49.8 / 50.7 (Segment)</td>
<td>50.4</td>
</tr>
<tr>
<td>Binary Sentiment Classification</td>
<td>79.0</td>
<td>77.4</td>
</tr>
<tr>
<td>Question-Answer Matching</td>
<td>56.4</td>
<td>55.8</td>
</tr>
<tr>
<td><strong>Semantic Relationship</strong></td>
<td><strong>75.2</strong></td>
<td><strong>76.7</strong></td>
</tr>
<tr>
<td><strong>Classification</strong></td>
<td><strong>75.2</strong></td>
<td><strong>76.7</strong></td>
</tr>
<tr>
<td>Discourse Parsing</td>
<td>57.5</td>
<td>56.4</td>
</tr>
</tbody>
</table>

More Questions

• Grammar
  – Is richer semantic better?
More Questions

• Grammar
  – Is richer semantic better?

• Data
  – How to obtain large (pseudo) annotation data?

• Algorithm
  – Can we utilize other supervision?
Future Trends

• Grammar
  – Trade-off between rich information and performance

• Data
  – Exploiting multiple treebanks

• Algorithm
  – Deep reinforcement learning
Thanks!