Learning Continuous Phrase Representations and Syntactic Parsing with Recursive Neural Networks

Richard Socher, Christopher D. Manning, Andrew Y. Ng
Department of Computer Science
Stanford University
richard@socher.org, manning@stanford.edu, ang@cs.stanford.edu

Marco Damonte - 1333293
• Proper syntactic representations are of importance to tasks such as relation extraction and semantic role labeling

• *Recursive Neural Networks* (RNNs) can provide us with vector space representations that can be exploited during parsing

• It is even possible to jointly learn representation and parse using the same deep network
Introduction
Vector Space Representation with Deep Learning

• *Deep networks* are commonly used to learn vector space representation of words $\rightarrow$ *Word Embeddings*

• RNNs generalize these embeddings for entire sentences
Methodology
Recursive Neural Networks

They are deep networks where we use the same set of weights recursively over a deep hierarchical structure.

\[
\text{score} = U^T p = \tanh(W[x; y] + b)
\]
Methodology
Recursive Neural Networks

They are deep networks where we use the same set of weights recursively over a deep hierarchical structure.

\[ score = U^T p \]
\[ p = \tanh (W[x; y] + b) \]
You might ask: *Why do we believe RNNs are good for this?*
You might ask: *Why do we believe RNNs are good for this?*
Methodology
Model 1: Greedy RNN

∀i, j

\[ \text{score}_{i,j} = U^T p_{i,j} \]

\[ p_{i,j} = \tanh(W[x_i; x_j] + b) \]
Methodology
Model 1: Greedy RNN

∀i, j
score_{i,j} = U^T p_{i,j}

p_{i,j} = \tanh(W[x_i ; x_j] + b)
Methodology

Model 1: Greedy RNN

\[ \forall i, j \]

\[ \text{score}_{i,j} = U^T p_{i,j} \]

\[ p_{i,j} = \tanh (W[x_i; x_j] + b) \]
Methodology

Model 1: Greedy RNN

∀i, j

score_{i,j} = U^T p_{i,j}

p_{i,j} = \tanh (W[x_i; x_j] + b)
Methodology

Model 2: Greedy Context-aware RNN

Context introduced in the first layer by allowing the representations of context words to modify the parsing decision

\[ score_{2,3} = U^T p_{2,3} \]
\[ p_{2,3} = \tanh(W[x_1; x_2; x_3; x_4] + b) \]
Methodology

Model 3: Greedy, Context-aware RNN and Category Classifier

Extension to greedy CRNN model: adding to each node a layer to predict class labels.

\[ P(y = c | x) = \text{softmax}(w_c^T x) = \frac{e^{w_c^T x}}{\sum_{c=1}^{C} e^{w_c^T x}} \]
Methodology

Model 4: Max-Margin Framework with Beam-Search

Instead of greedily collapse the best pairs:

- Formulate a global objective function that penalises choices far from the correct choice

\[ \sum_{i} s(x_i, y_i) - \max_{y} (s(x_i, y) + \Delta(y, y_i)) \]

- **Beam Search** instead of **Greedy Search** to find the best parse
Methodology

Training the RNN

We need to determine the set of weights $W$:

- Backpropagation

\[(f(g(x)))' = f'(g(x))g'(x)\]

- Gradient descent
Methodology
Training the RNN

We need to determine the set of weights $W$:

- Backpropagation

\[
(f(g(x)))' = f'(g(x))g'(x)
\]

- Gradient descent

Actually..

- Backpropagation Through Structure (BTS)
- Subgradient method
## Experiments

### Word embeddings

<table>
<thead>
<tr>
<th>Word</th>
<th>Initial Collobert &amp; Weston embedding</th>
<th>After RNN training</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>a, its, an, this, his, their, UNK</td>
<td>an, The, no, some, A, these, another</td>
</tr>
<tr>
<td>and</td>
<td>,, or, but, as, -, .., that, for, in</td>
<td>or, but, And, But, &amp;, least</td>
</tr>
<tr>
<td>in</td>
<td>at, on, from, for, over, after, and, as</td>
<td>In, from, into, since, for, For, like, with, what, who, if, this, some, which, If</td>
</tr>
<tr>
<td>that</td>
<td>which, but, ,, and, -, as, for, or, about, if</td>
<td>says, fell, added, did, rose, sold, reported</td>
</tr>
<tr>
<td>said</td>
<td>added, says, -, while, but, reported, on</td>
<td>they, I, we, you, It, she, it, He, We, They</td>
</tr>
<tr>
<td>he</td>
<td>she, it, they, which, also, now, who, we</td>
<td>increase, bank, income, industry, issue, state, sale, growth, unit, president</td>
</tr>
<tr>
<td>share</td>
<td>high, higher, business, market, current, stock, lower, increase, price, financial</td>
<td>where, how, which, during, including</td>
</tr>
</tbody>
</table>
Center Phrase and Nearest Neighbors

(A) Sales grew almost 2% to 222.2 million from 222.2 million.
  1. Sales surged 22% to 222.22 billion yen from 222.22 billion.
  2. Revenue fell 2% to 2.22 billion from 2.22 billion.
  3. Sales rose more than 2% to 22.2 million from 22.2 million.
  4. Volume was 222.2 million shares, more than triple recent levels.
## Experiments

### Parsing

<table>
<thead>
<tr>
<th>Method</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 (Greedy RNN)</td>
<td>76.55</td>
</tr>
<tr>
<td>Model 2 (Greedy, context-aware RNN)</td>
<td>83.36</td>
</tr>
<tr>
<td>Model 3 (Greedy, context-aware RNN + category classifier)</td>
<td>87.05</td>
</tr>
<tr>
<td>Model 4 (Beam, context-aware RNN + category classifier)</td>
<td>92.06</td>
</tr>
<tr>
<td>Left Corner PCFG, [MC97]</td>
<td>90.64</td>
</tr>
<tr>
<td>Current Implementation of the Stanford Parser, [KM03]</td>
<td><strong>93.98</strong></td>
</tr>
</tbody>
</table>
Conclusion

- Better word embeddings
- Sentence embeddings entailing syntactic and semantic information
- Almost state-of-the-art parsing