GAČR EXPRO
NEUREM³
Studying Representations

Ondřej Bojar

Sept 16, Pec pod Sněžkou
Deep NNs for Image Classification

It's deep if it has more than one stage of non-linear feature transformation.
Consider word2vec “comprehensive” test set (Mikolov et al., 2013):

- 8.8k “semantic” and 10.6k “syntactic” questions,
- w2v “accuracy is quite good” (eyeballing)
  - The authors do mention that exact-match is “only about 60%”).

Kocmi and Bojar (2016) carefully examined the test set:

- “Semantic” questions cover only 3 question types:
  - country→city, country→currency, masculine family member→feminine
  - Vylomova et al. (2016) test many other relations, e.g. walk-run, dog-puppy, bark-dog, cook-eat.

- “Syntactic” questions constructed by combinations:
  - starting from only 313 distinct word pairs,
  - (leading to only 35 different pairs per question on average),
  - And of the 313 pairs, 286 are formed regularly.
<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy on “Synt Qs”</th>
<th>Test Set by</th>
</tr>
</thead>
<tbody>
<tr>
<td>word2vec as released</td>
<td>62.5%</td>
<td>43.5%</td>
</tr>
<tr>
<td>word2vec on our data</td>
<td>42.5%</td>
<td>9.7%</td>
</tr>
<tr>
<td>SubGram on our data</td>
<td>42.3%</td>
<td>22.4%</td>
</tr>
</tbody>
</table>

Caveat on Evaluation (2/2)
Caveat on Evaluation (2/2)

<table>
<thead>
<tr>
<th>Accuracy on &quot;Synt Qs&quot;</th>
<th>Test Set by</th>
<th>Mikolov et al.</th>
<th>Kocmi et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>word2vec as released</td>
<td>62.5%</td>
<td>43.5%</td>
<td></td>
</tr>
<tr>
<td>word2vec on our data</td>
<td>42.5%</td>
<td>9.7%</td>
<td></td>
</tr>
<tr>
<td>SubGram on our data</td>
<td>42.3%</td>
<td>22.4%</td>
<td></td>
</tr>
<tr>
<td><strong>Nine rules</strong></td>
<td><strong>71.9%</strong></td>
<td><strong>66.4%</strong></td>
<td></td>
</tr>
</tbody>
</table>
Caveat on Ultimate Evaluation

Kocmi and Bojar (2016):
• submitted to TSD on March 22, 2016.
• appeared in TSD in September 2016.
... cited by 4.

Bojanowski et al. (2017):
• submitted to arxiv on July 15, 2016.
• appeared in TACL 2017.
... cited by 1024.
Caveat on Ultimate Evaluation

Kocmi and Bojar (2016):
• submitted to TSD on March 22, 2016.
• appeared in TSD in September 2016.
  ... cited by 4.
• No code released, no fast code implemented at all.

Bojanowski et al. (2017):
• submitted to arxiv on July 15, 2016.
• appeared in TACL 2017.
  ... cited by 1024.
• This is the FastText paper.
• Ondřej Bojar
• Pavel Pecina
• Jindra Helcl (non-autoregressive MT, i.a.)
• Ivana Kvapilíková (unsupervised MT)
• Michal Auersperger (document representations)
• (Jindřich Libovický) (MT with images, i.a.)
• (Petra Galuščáková) (something with video?)
Expected Outcomes of NEUREM³

- **Insight** into what the representations look like (for ASR and NMT).
- **Tools** for diagnosing:
  - Which tasks are learned implicitly with the main one.
  - Why is the network making some particular types of errors.
  - Which **generalizations** has the network learned and which not.
- **Methods** for:
  - semi-supervised and unsupervised learning.
  - pre-training, reuse of model parts, combining larger models, model interfacing,
  - successful multi-task training,
    all esp. in the areas of ASR and NMT.
Expected Outcomes of NEUREM

- **Insight** into what the representations look like (for ASR and NMT).
- **Tools** for diagnosing:
  - Which tasks are learned implicitly with the main one.
  - Why is the network making some particular types of errors.
  - Which generalizations has the network learned and which not.
- **Methods** for:
  - semi-supervised and unsupervised learning.
  - pre-training, reuse of model parts, combining larger models, model interfacing,
  - successful multi-task training,
    all esp. in the areas of ASR and NMT.
- **Good papers, good papers, good papers…**

