VECTORISING

Begriffsschrift

On the relevance of recent forays into the deep learning of word meanings to some traditional philosophical problems and vice versa

LAUC, Davor +,*; SKELAC, Ines +,**

+ University of Zagreb, Faculty of Humanities and Social Sciences, Unit of Logic
* TerraLogix.ai Group Ltd, Chief Data Scientist
** Algebra University College

HaPoC 2019
Agenda of the presentation

Introduction:
Motivation and our theses

NLP & Deep learning
Short review of the recent forays

Indeterminacy of translation
And deep neural model for machine translation

Frege’s contextuality principle
& dense vector representation of word meanings

Discussion
OUR THESES

01 Developments in NLP vindicates Frege’s context principle

02 Radical translation thought experiments restated in MT indicates that indeterminacy of reference is improbable

03 Transfer learning in MT indicates that Quine’s conjecture on holophrastic indeterminacy is too strong
AI and NATURAL LANGUAGE UNDERSTANDING
Deep Learning
and natural language understanding

Neural language models
2001

Word embeddings, (bi-)LSTMs
2013

Seq2seq models
2014

Pretrained language models, transfer learning
2018
GOOGLE BERT
ARCHITECTURE
The recent foray in NLU using Transformers

Google BERT, Facebook’s XLM, OpenAI GPT, RoBERTa, ....

<table>
<thead>
<tr>
<th>System</th>
<th>MNLI-(m/mm)</th>
<th>QQP</th>
<th>QNLI</th>
<th>SST-2</th>
<th>CoLA</th>
<th>STS-B</th>
<th>MRPC</th>
<th>RTE</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>392k</td>
<td>363k</td>
<td>108k</td>
<td>67k</td>
<td>8.5k</td>
<td>5.7k</td>
<td>3.5k</td>
<td>2.5k</td>
<td></td>
</tr>
<tr>
<td>Pre-OpenAI SOTA</td>
<td>80.6/80.1</td>
<td>66.1</td>
<td>82.3</td>
<td>93.2</td>
<td>35.0</td>
<td>81.0</td>
<td>86.0</td>
<td>61.7</td>
<td>74.0</td>
</tr>
<tr>
<td>BiLSTM+ELMo+Attn</td>
<td>76.4/76.1</td>
<td>64.8</td>
<td>79.8</td>
<td>90.4</td>
<td>36.0</td>
<td>73.3</td>
<td>84.9</td>
<td>56.8</td>
<td>71.0</td>
</tr>
<tr>
<td>OpenAI GPT</td>
<td>82.1/81.4</td>
<td>70.3</td>
<td>87.4</td>
<td>91.3</td>
<td>45.4</td>
<td>80.0</td>
<td>82.3</td>
<td>56.0</td>
<td>75.1</td>
</tr>
<tr>
<td>BERT&lt;sub&gt;BASE&lt;/sub&gt;</td>
<td>84.6/83.4</td>
<td>71.2</td>
<td>90.5</td>
<td>93.5</td>
<td>52.1</td>
<td>85.8</td>
<td>88.9</td>
<td>66.4</td>
<td>79.6</td>
</tr>
<tr>
<td>BERT&lt;sub&gt;LARGE&lt;/sub&gt;</td>
<td>86.7/85.9</td>
<td>72.1</td>
<td>92.7</td>
<td>94.9</td>
<td>60.5</td>
<td>86.5</td>
<td>89.3</td>
<td>70.1</td>
<td>82.1</td>
</tr>
</tbody>
</table>

Frege’s Context Principle

01 "never ... ask for the meaning of a word in isolation, but only in the context of a proposition"

02 sents = ['Rabbit is a mammal!', 'Rabbit is the same as Gavagai!']

03 [[[CLS], 'Rabbit', 'is', 'a', 'mam', '##al', '!', [SEP]],
[[CLS], 'Rabbit', 'is', 'the', 'same', 'as', 'Ga', '##va', '##gai', '!', [SEP]]]

('bert-base-multilingual-cased')

04 Vec('rabbit') in the 1. & 2. sentence cosine similarity 0.8167434
Frege’s Context Principle 2

01 "never ... ask for the meaning of a (sub)word (a word piece) in isolation, but only in the context of a proposition"

02 `sents = ['กระต่ายเป็นสัตว์เลี้ยงลูกด้วยนม', 'กระต่ายไม่เหมือนกับ Gavagai']`

03 `[CLS], ['ก', '##ร', '##ะ', '##ต', '##่า', '##เป็น', '##เป็น', '##ส', '##ำ', '##ต', '##ด', ... ('bert-base-multilingual-cased')]

04 Rabbit tokens (1-6) in the sentences 1-2 similarity 0.754323
THE FIRST THESIS

The improvements of NLU metrics when word meanings are represented by sentence-relative dense vectors indicate that Frege’s context principle is correct.
INDETERMINACY OF TRANSLATION

- indeterminacy of theories
- indeterminacy of reference
- holophrastic indeterminacy
I. of reference

Parts of the sentence may change in what they refer, but they will maintain the meaning of the sentence as a whole.

Holophrastic i.

A sentence may be correctly translated in multiple ways with different meanings.
<table>
<thead>
<tr>
<th>Radical Translation</th>
<th>Machine Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A target language is neither historically nor culturally linked to any known language</td>
<td>A target language is not closely linked to the language pair (i.e. another language family)</td>
</tr>
<tr>
<td>Translation of theoretical sentences is indeterminate</td>
<td>All sentences are theoretical; translation is based on MT methods</td>
</tr>
<tr>
<td>Input: behaviour, non-linguistic elements</td>
<td>Input: a large parallel text corpus</td>
</tr>
</tbody>
</table>
Deep learned Gavagai

01 Experiment measuring performance of the MT models trained on new (unseen) words

02 Experiment design:
Dataset 1: IWSLT-14 English-German MT corpus, 160,239 sentence pairs, 33 “rabbit” sentences
Dataset 2: Ted Talk Thai-English sentences, 304,245, 40 “กระต่าย” sentences
Library: Facebook FairSeq
DS1: Transformer wmt16.en-de Model (Ott et al., 2018)
DS2: vanilla wmt16
Deep learned Gavagai 2

03  MT Model learned on IWSLT-14 dataset with „rabbit / bunny” and „Hase / Kaninchen” sentences removed
• 100 epochs, loss 1.29
• %0 of test „rabbit” sentences translate correctly

04  MT Model learned on TED-talks DS without „กระต่าย” and „rabbit/bunny” sentences
• 100 epochs, loss 1.61
• %0 of test “กระต่าย” sentences translate correctly
Deep learned Gavagai 3

Model is incrementally trained on „rabbit” sentences - 23 sentences, 50 epochs each

% of correctly translated "rabbit" sentences

- % of "Hase/Kaninchen" (en->de)
- % of "rabbit/bunny" (th->en)
Radical translation thought experiments simulated with MT indicates that indeterminacy of reference is improbable.
<table>
<thead>
<tr>
<th>Translation manual</th>
<th>Deep NN model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different possibilities of translating the same sentence (different meaning); impossible to determine the right one</td>
<td>Different possibilities of translating the same sentence (different meaning); but with a high probability of correctness of one translation</td>
</tr>
<tr>
<td>Rules of translation</td>
<td>Rules, probability, transfer knowledge, vector representation</td>
</tr>
<tr>
<td>Context non-sensitive</td>
<td>Context-sensitive</td>
</tr>
</tbody>
</table>
Holophrastic indeterminacy & transfer learning

<table>
<thead>
<tr>
<th>Language Pair</th>
<th>Parent</th>
<th>Train Size</th>
<th>BLEU ↑</th>
<th>PPL ↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uzbek–English</td>
<td>None</td>
<td>1.8m</td>
<td>10.7</td>
<td>22.4</td>
</tr>
<tr>
<td></td>
<td>French–English</td>
<td>1.8m</td>
<td>15.0 (+4.3)</td>
<td>13.9</td>
</tr>
<tr>
<td>French’–English</td>
<td>None</td>
<td>1.8m</td>
<td>13.3</td>
<td>28.2</td>
</tr>
<tr>
<td></td>
<td>French–English</td>
<td>1.8m</td>
<td>20.0 (+6.7)</td>
<td>10.9</td>
</tr>
</tbody>
</table>

Transfer learning in MT indicates that Quine’s conjecture on holophrastic indeterminacy is too strong.
Thank You

Discussion time