Evaluating Compositional Distributed Semantic Models with RELPRON

Stephen Clark

University of Toronto Department of Computer Science

31 March 2016

University of Cambridge Computer Laboratory
Compositional + Distributional ?

The cat sat on the mat.
Vector Space Semantics

- furry
- stroke
- pet
- cat
- dog
From Words to Sentences

\[ s_1 \]

man killed dog

\[ s_2 \]

man murdered cat

\[ s_3 \]

man killed by dog
Vector-Based Models of Sentences

Grefenstette et. al, New Directions in Vector Space Models of Meaning (ACL, 2014)
Roadmap

- Compositional Models
  - elementwise operators
  - neural networks
  - type-driven tensor-based models
- Evaluation
  - a new dataset based on relative clauses (RELPRON)
  - results using neural networks
Collaborators
Distributional Semantics

... examples of felines are the big cats – the lion, tiger, leopard, ... Calico cat is an American name for cats with three colors. Out ... Siamese is a breed of cat from Thailand. It is a well-known cat. ... Cats is a musical. It is based on Old Possums Book of Practical ...

<table>
<thead>
<tr>
<th></th>
<th>feline</th>
<th>lion</th>
<th>...</th>
<th>musical</th>
<th>color</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>10</td>
<td>20</td>
<td></td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>tiger</td>
<td>2</td>
<td>4</td>
<td></td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

\[
tTest(w_i, c_j) = \frac{p(w_i, c_j) - p(w_i)p(c_j)}{\sqrt{p(w_i)p(c_j)}}\]
### Dimensionality Reduction

#### CS+Norm+SVD Topics (K=20)

<table>
<thead>
<tr>
<th>Topic ID 1</th>
<th>Topic ID 2</th>
<th>Topic ID 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>animal</strong></td>
<td><strong>human</strong></td>
<td><strong>animals</strong></td>
</tr>
<tr>
<td><strong>food</strong></td>
<td><strong>food</strong></td>
<td><strong>with</strong></td>
</tr>
<tr>
<td><strong>buildings</strong></td>
<td><strong>plants</strong></td>
<td><strong>public</strong></td>
</tr>
<tr>
<td><strong>transport</strong></td>
<td></td>
<td><strong>wings</strong></td>
</tr>
</tbody>
</table>

- **Topic ID 1**: flower 0.4219, yellow 0.420, fruit 0.4180, eat 0.4033, salad 0.3918, foliage 0.3867, blue 0.3838, bird 0.3811, insect 0.3702, colour 0.3655
- **Topic ID 2**: salad 0.5368, meat 0.4934, dessert 0.4570, potato 0.4444, carrot 0.4441, eat 0.4255, noodle 0.4254, chicken 0.4037, tomato 0.3941, soup 0.3864
- **Topic ID 3**: bird 0.4874, mammal 0.4420, reptile 0.4099, parrot 0.3609, hummingbird 0.3517, pelican 0.3527, insect 0.3429, dragonfly 0.3424, gull 0.3311, stork 0.3303
- **Topic ID 4**: building 0.2918, wall 0.2978, floor 0.2855, ceiling 0.2650, brick 0.2624, roof 0.2616, bird 0.2607, mammal 0.2504, patio 0.2484, tower 0.2465
- **Topic ID 5**: flower 0.4482, foliage 0.4092, petal 0.3548, yellow 0.3488, leaf 0.3283, purple 0.3208, bright 0.3040, tree 0.2779, plant 0.2765, shade 0.268
- **Topic ID 6**: line 0.3395, rail 0.3111, road 0.3107, railway 0.2986, train 0.2977, station 0.2887, traffic 0.2517, subway 0.2474, bus 0.2460, downtown 0.2392
Neural Representations

CBOW

Skip-gram
Roadmap

- Compositional Models
  - elementwise operators
  - neural networks
  - type-driven tensor-based models
- Evaluation
  - a new dataset based on relative clauses (RELPRON)
  - results using neural networks
Elementwise Operators

\[
\begin{array}{cccc}
\text{black} & 0.34 & 0.64 & ... & -0.06 & ...
\end{array}
\]

\[
+\n\begin{array}{cccc}
\text{cat} & 0.15 & 0.29 & ... & -0.03 & ...
\end{array}
\]

\[
=\n\begin{array}{cccc}
\text{black + cat} & 0.49 & 0.93 & ... & -0.09 & ...
\end{array}
\]

\[
\begin{array}{cccc}
\text{black} & 0.34 & 0.64 & ... & -0.06 & ...
\end{array}
\]

\[
=\n\begin{array}{cccc}
\text{cat} & 0.15 & 0.29 & ... & 0.03 & ...
\end{array}
\]

\[
\begin{array}{cccc}
\text{black o cat} & 0.05 & 0.19 & ... & -0.002 & ...
\end{array}
\]
### Mitchell & Lapata 2010 Dataset:

**AN**: national government  
  - new information  
  - cold air  
  - further evidence  

**NN**: environment secretary  
  - party leader  
  - telephone number  
  - future development  

**VO**: offer support  
  - fight war  
  - provide help  
  - win battle
Composition in Neural Models

Deep Learning for NLP (Socher et al., 2013)
Composition in Neural Models

Recursive Matrix-Vector Model

Socher et al. (EMNLP 2013)
Syntactic Types to Tensors

\[
\text{cat} \quad \text{chases} \quad \text{dog}
\]

\[
\begin{array}{c}
\text{NP} \\
\text{S} \setminus \text{NP} / \text{NP} \\
\text{NP}
\end{array}
\]

\[
\begin{array}{c}
\text{N} \\
\text{S} \otimes \text{N} \otimes \text{N} \\
\text{N}
\end{array}
\]
Type and Tensor Reductions

Function application = taking inner products
Type and Tensor Reductions

\[
\begin{align*}
\text{cat} & \quad \text{chases} & \quad \text{dog} \\
\text{NP} & \quad (S\setminus\text{NP})/\text{NP} & \quad \text{NP} \\
S & \quad \otimes & \quad N & \quad \otimes & \quad N & \quad N \\
S & \quad \otimes & \quad N & \quad \otimes & \quad N & \quad N \\
S' & \quad \setminus & \quad \text{NP} \\
S & \quad \otimes & \quad N \\
S & \quad \otimes & \quad N \\
S & \quad \otimes & \quad N
\end{align*}
\]
Relative Pronoun Neural Network

person that use telescope

person that use telescope

use telescope

use telescope

person

that

use

telescope
Relative Pronoun Neural Network
Roadmap

• Compositional Models
  • elementwise operators
  • neural networks
  • type-driven tensor-based models

• Evaluation
  • a new dataset based on relative clauses (RELPRON)
  • results using neural networks
## Existing Datasets

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>national government</td>
<td>1</td>
</tr>
<tr>
<td>cold air</td>
<td></td>
</tr>
<tr>
<td>new information</td>
<td>6</td>
</tr>
<tr>
<td>further evidence</td>
<td></td>
</tr>
<tr>
<td>offer support</td>
<td>7</td>
</tr>
<tr>
<td>provide help</td>
<td></td>
</tr>
<tr>
<td>fight war</td>
<td>5</td>
</tr>
<tr>
<td>win battle</td>
<td></td>
</tr>
</tbody>
</table>

This hiker does a dance of joy on top of the mountain. 2.5
A man stands in a strange position on a rock high above some trees.
A brown and white dog trots across shallow water with his mouth open. 3.7
A white and brown dog is walking through the water.
RELPRON Dataset

- **wisdom**: quality that experience teaches
- **bowler**: player that dominates batsmen
- **theatre**: building that hosts premieres
Motivation

• Relative clauses have an interesting compositional structure

• Intermediate level of syntactic complexity with a function word

<table>
<thead>
<tr>
<th>quality</th>
<th>that</th>
<th>experience</th>
<th>teaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>(NP\NP)/(S/NP)</td>
<td>NP</td>
<td>(S\NP)/NP</td>
</tr>
<tr>
<td>N</td>
<td>N⊗N⊗S⊗N</td>
<td>N</td>
<td>S⊗N⊗N</td>
</tr>
</tbody>
</table>
The Ranking Task

navy
organisation that fleet destroy
organisation that sailor join
person that make journey
organisation that soldier join
device that seal opening
organisation that defeat fleet
organisation that establish blockade
mammal that fleet hunt
Terminology

telescope: device that detects planets

RELATIVE CLAUSE

property
Selection of Head Nouns and Terms

Head Nouns

activity, building, device, document, mammal, material, organization, person, phenomenon, player, quality, room, scientist, vehicle, woman

Terms

person: traveler, intellectual, follower, survivor, ...
quality: accuracy, balance, wisdom, rhythm, ...
Extraction of SVO Triples

- It allow traveler
- Website allow traveler
- Inn serve traveler
- Business serve traveler
- She become traveler
- Student become traveler
- Traveler make decision
- Traveler make journey
- Traveler take road
- Traveler take advantage
- Traveler have option

traveler: person that
business serves

traveler: person that
takes advantage
Manual Annotation of Properties

✖ traveler: person that it allows
✖ traveler: person that website allows
✔ traveler: person that inn serves
✖ traveler: person that business serves
✖ traveler: person that she becomes
✖ traveler: person that student becomes

✖ traveler: person that makes decision
✖ traveler: person that makes journey
✔ traveler: person that takes road
✖ traveler: person that takes advantage
✖ traveler: person that has options
✔ traveler: person that has baggage
Full List of Properties for navy

OBJ  navy: organization that sailor join
OBJ  navy: organization that fleet destroy
OBJ  navy: organization that vessel serve
OBJ  navy: organization that battleship fight

SBJ  navy: organization that use submarine
SBJ  navy: organization that maintain blockade
SBJ  navy: organization that defeat fleet
SBJ  navy: organization that protect waters
SBJ  navy: organization that establish blockade
SBJ  navy: organization that blockade port
## Data Set Statistics

<table>
<thead>
<tr>
<th>Head noun</th>
<th>Term</th>
<th>Prop</th>
<th>Head noun</th>
<th>Term</th>
<th>Prop</th>
</tr>
</thead>
<tbody>
<tr>
<td>phenomenon</td>
<td>10</td>
<td>80</td>
<td>quality</td>
<td>10</td>
<td>81</td>
</tr>
<tr>
<td>activity</td>
<td>10</td>
<td>83</td>
<td>organization</td>
<td>10</td>
<td>99</td>
</tr>
<tr>
<td>material</td>
<td>10</td>
<td>82</td>
<td>device</td>
<td>10</td>
<td>76</td>
</tr>
<tr>
<td>room</td>
<td>10</td>
<td>80</td>
<td>building</td>
<td>10</td>
<td>69</td>
</tr>
<tr>
<td>vehicle</td>
<td>10</td>
<td>79</td>
<td>document</td>
<td>10</td>
<td>69</td>
</tr>
<tr>
<td>mammal</td>
<td>10</td>
<td>70</td>
<td>person</td>
<td>10</td>
<td>86</td>
</tr>
<tr>
<td>woman</td>
<td>5</td>
<td>37</td>
<td>player</td>
<td>5</td>
<td>38</td>
</tr>
<tr>
<td>scientist</td>
<td>8</td>
<td>58</td>
<td>TOTAL</td>
<td>65</td>
<td>518</td>
</tr>
<tr>
<td>TOTAL</td>
<td>73</td>
<td>569</td>
<td>TOTAL</td>
<td>65</td>
<td>518</td>
</tr>
</tbody>
</table>
Evaluation

navy

organisation that fleet destroy

organisation that sailor join

person that make journey

organisation that soldier join

device that seal opening

organisation that defeat fleet

organisation that establish blockade

mammal that fleet hunt

\[
\text{Mean Average Precision (MAP)}
\]

\[
= \frac{1}{N} \sum_{t=1}^{N} \frac{1}{p_t} \sum_{k=1}^{M} \text{Prec}(k) \times \text{rel}(k)
\]
Roadmap

• Compositional Models
  • elementwise operators
  • neural networks
  • type-driven tensor-based models

• Evaluation
  • a new dataset based on relative clauses (RELPRON)
  • results using neural networks
Relative Pronoun Neural Network

person that use telescope

person

that

use
telescope
Training the Neural Networks

person that use telescope

person that use telescope
Anarchism is a political philosophy which considers the state undesirable, unnecessary, and harmful, and instead promotes a stateless society, or anarchy. It seeks to diminish or even abolish authority in the conduct of human relations. Anarchists may widely disagree on what additional criterion be required in anarchism.

Word2Vec (Mikolov et al.)
- promote society
- abolish authority
- include strain
- support economy
- become ruler
Anarchism is a political philosophy which consider the state undesirable, unnecessary and harmful, and instead promote a stateless society, or anarchy. It seek to diminish or even abolish authority in the conduct of human relation. Anarchist may widely disagree on what additional criterion be require in anarchism.
Holistic Relative Clauses?

person that use telescope

person

that

use telescope

use

telescope
Relative Clauses as Adjectives

“using-telescope” person

person that use telescope

use
Anarchism is a political philosophy which consider the state undesirable, unnecessary and harmful, and instead promote a stateless society, or anarchy. It seek to diminish or even abolish authority in the conduct of human relation. Anarchist may widely disagree on what additional criterion be require in anarchism.

Word2Vec (Mikolov et al.)

- political philosophy
- stateless society
- human relation
- additional criterion
- red cortina
Implementation Details

- tensorflow
- 300 dim. word2vec embeddings
- hidden layers: 1000, 600, tanh
- MSQE loss, Adagrad, minibatch
- 831M words Wikipedia dump
- 50 freq cutoff for holistic vectors
- 75,106 adjective nouns
- 54,825 verb objects
- 39,571 subject verbs
## Development Results

<table>
<thead>
<tr>
<th>Model</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Addition</strong></td>
<td>53.8</td>
</tr>
<tr>
<td><strong>Neural Net (all holistic)</strong></td>
<td>58.2</td>
</tr>
</tbody>
</table>
**Dictionary Data**

projection: a forecast or prognosis obtained by extrapolation
projection: a belief or assumption that others have similar thoughts and experiences as oneself
projection: the image that a translucent object casts onto another object
projection: an image of an object on a surface of fewer dimensions
dna: a nucleic acid that carries the genetic information in the cell
dna: the sequence of nucleotides determines individual hereditary characteristics
dna: defense nuclear agency
dna: did not answer
heathland: a tract of level wasteland uncultivated land with sandy soil and scrubby vegetation
matroid: a structure that captures the essence of a notion of independence that generalizes linear independence in vector spaces
polioptila: an isolated genus of oscine passerine birds typical of the subfamily polioptilinæ the american: gnatcatchers so called from the hoary edgings of the wings
polioptila: new world gnatcatchers
dangdut: a genre of indonesian music that combines elements of arab and malay folk music

Learning to Understand Phrases by Embedding the Dictionary (Hill et al. 2016)
abductor muscle that draw part
abductor someone who detain victim
abeam line which form angle
abecedarian germany who have disdain
abecedarian one who learn letter
abecedarian one who study alphabet
abecedarian one who teach alphabet
abecedarian one who teach letter
abecedarian work which use word
abelard héloise whom marry héloise
aberdeen knot which see canutus
abettor one who encourage another
abettor one who help another
abettor one who incite another
abhorrence that which excite repugnance
abhorrence that which excite servility
ability quality that facilitate accomplishment
ability quality that facilitate achievement
ability quality that permit accomplishment
## Development Results

<table>
<thead>
<tr>
<th>Model</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition</td>
<td>53.8</td>
</tr>
<tr>
<td>Neural Net (all holistic)</td>
<td>58.2</td>
</tr>
<tr>
<td>Neural Net (holistic + dict)</td>
<td>54.8</td>
</tr>
<tr>
<td>Ensemble</td>
<td>57.4</td>
</tr>
</tbody>
</table>
## Individual APs

<table>
<thead>
<tr>
<th>Term</th>
<th>ADD</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>garrison</td>
<td>43.45</td>
<td>49.55</td>
</tr>
<tr>
<td>mission</td>
<td>23.79</td>
<td>29.67</td>
</tr>
<tr>
<td>division</td>
<td>73.96</td>
<td>78.92</td>
</tr>
<tr>
<td>family</td>
<td>32.41</td>
<td>27.32</td>
</tr>
<tr>
<td>charity</td>
<td>93.87</td>
<td>95.58</td>
</tr>
<tr>
<td>army</td>
<td>52.41</td>
<td>48.19</td>
</tr>
<tr>
<td>navy</td>
<td>79.93</td>
<td>91.27</td>
</tr>
<tr>
<td>railway</td>
<td>47.42</td>
<td>48.10</td>
</tr>
<tr>
<td>religion</td>
<td>35.35</td>
<td>42.13</td>
</tr>
<tr>
<td>expert</td>
<td>31.17</td>
<td>28.96</td>
</tr>
<tr>
<td>follower</td>
<td>89.74</td>
<td>70.87</td>
</tr>
</tbody>
</table>
TERM: **navy 10**

- **navy**: organization that blockade port
- **navy**: organization that fleet destroy
- **navy**: organization that battleship fight
- **navy**: organization that defeat fleet
- **navy**: organization that establish blockade
- **navy**: organization that use submarine
- **navy**: organization that maintain blockade
- **navy**: organization that sailor join
- **garrison**: organization that army install
- **army**: organization that seize garrison
- **pub**: building that attract sailor
- **traveler**: person that take ship
- **navy**: organization that vessel serve
Definition Ranking Example

TERM: expert 10

specification: document that engineer develop
expert: person that author monograph
learner: person that develop skill
expert: person that have knowledge
intellectual: person that hold seminar
intellectual: person that publish study
expert: person that panel include
specification: document that set standard
expert: person that lawyer consult
golfer: player that bunker challenge
characteristic: quality that determine identity
charity: organization that celebrity represent
likelihood: quality that risk assess
learner: person that subject engage
expert: person that give tutorial
## Test Results

<table>
<thead>
<tr>
<th>Model</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADDITION</td>
<td>53.2</td>
</tr>
<tr>
<td>Neural Net (all holistic)</td>
<td>54.1</td>
</tr>
<tr>
<td>Neural Net (holistic + dict)</td>
<td>52.8</td>
</tr>
<tr>
<td>Ensemble</td>
<td>55.6</td>
</tr>
</tbody>
</table>
Conclusion

• Modelled function word separately from content words
• Difficult to beat the addition baseline (but we did it)
• “Is God trying to tell us something”?

Papers available at: http://www.cl.cam.ac.uk/~sc609/
The End

A man stands in a strange position on a rock high above some trees.