ConceptNet at Twenty:
Reflections on structured common sense in an era of machine learning

Dr. Catherine Havasi, MIT Media Lab

Photo: Michael Vesia, CC-By
MEANWHILE, BACK IN MASSACHUSETTS

Photo: Eric Baetscher
COLLECTING COMMON-SENSE KNOWLEDGE

But then things moved on without us – users changed how they interacted with search engines.
How do people do it?
We have built models of how people think about the world in 73 languages – called ConceptNet.
“I don’t have to actually experience crashing my car into a wall a few hundred times before I slowly start avoiding to do so.”

- Andrej Karpathy, Open AI
Distributional semantics

• “You shall know a word by the company it keeps.” – J. R. Firth
• *Skipgrams* are one way to train a neural net to understand words by their contexts

Retrofitting

• A way of connecting knowledge graphs to traditionally created word embeddings

• Apply knowledge-based constraints after training distributional word vectors

• It works better than during training, for some reason

Retrofitting

• Terms that are connected in the knowledge graph should have vectors that are closer together

• Many extensions now, such as “antonyms should move farther apart” and morphology

Retrofitting just works*

• Until 11/2018, the top-performing systems almost always use retrofitting
• At Semeval 2018, about half of the top results used external information in order to get a top result.
• Other areas of ML were seeing similar results from adding structure information
• Word vectors with common sense built in
• Hybrid of ConceptNet and distributional semantics
• Multilingual by design
• Open source, open data

Building ConceptNet Numberbatch

Many data sources

Google News

Common Crawl

Open Subtitles

word2vec

GloVe

fastText

ConceptNet

Retrofit

Retrofit

Retrofit

Join

Reduce dimensionality

De-bias

ConceptNet Numberbatch
Word relatedness (SemEval 2017)

Source: Robyn Speer and Joanna Lowry-Duda: Extending Word Embeddings with Multilingual Relational Knowledge (Semeval 2017)
Distinguishing attributes using ConceptNet

- A task at SemEval 2018
- We got **74%** accuracy (2nd place) by directly querying ConceptNet Numberbatch
- Additional features trained on the data didn’t help on the test set
- All top systems used knowledge graphs

---

Source: *Robyn Speer and Joanna Lowry-Duda*: Distinguishing Attributes Using Text Corpora and Relational Knowledge (Semeval 2018)
What is domain Adaptation?

customer intents, product names, industry jargon, specific issues

common words, multiple languages, paraphrases, general sentiment

domain specific data

domain general data
Spin-out: Luminoso (2010-)

Graphs showing trends and themes in data over time.
What are other examples of transfer learning?

- Commercial transfer learning (Luminoso, etc)
- Pretraining
- Fine tuning and layer freezing for Elmo and Bert (and GPT-2)
- Fast.ai’s ULMFiT (http://nlp.fast.ai/)
Computational Creativity and Common Sense
Story understanding evaluations

- The Story Cloze Test evaluates common sense
- Five-sentence stories, two possible endings, only one makes sense
  - Previous state of the art (GPT-1): 86.5%
  - Jiaao Chen et al., adding ConceptNet as an input: 87.6%

Story understanding evaluations

- **SemEval-2018 task**: answer simple multiple-choice questions about a passage

**Text**: It was a long day at work and I decided to stop at the gym before going home. I ran on the treadmill and lifted some weights. I decided I would also swim a few laps in the pool. Once I was done working out, I went in the locker room and stripped down and wrapped myself in a towel. I went into the sauna and turned on the heat. I let it get nice and steamy. I sat down and relaxed. I let my mind think about nothing but peaceful, happy thoughts. I stayed in there for only about ten minutes because it was so hot and steamy. When I got out, I turned the sauna off to save energy and took a cool shower. I got out of the shower and dried off. After that, I put on my extra set of clean clothes I brought with me, and got in my car and drove home.

**Q1**: Where did they sit inside the sauna?
(a) on the floor  
(b) on a bench

**Q2**: How long did they stay in the sauna?
(a) about ten minutes  
(b) over thirty minutes
Story understanding using ConceptNet

- The winning system on the SemEval task (Yuanfudao Research) used ConceptNet to find unstated connections, on top of an LSTM + attention model

Narratarium

Intelligent Agents that focus on character-based interactions

Jason Alonso, Catherine Havasi

>>> Where have you been?
well, as far as i know, they've been having a good time in there.
>>> And what are you doing?
well, i'm trying to set up a three phase energy beam to neutralise the neutron
>>> Isn't it already neutralized?
no, it won't be neutralised.
>>> Why not?
well, er, we don't know. there's no known cure.
>>> For a neutron?
yes, for a neutron star.
>>> Oh.
and we are not ready for it.
>>>
Can we use this to bring experiences to life?

WITH: NINA LUTZ AND JASON ALONSO ~ 2020
Can you tell me how to get to Sesame Street?
BERT: predicting masked words and next sentences

Input = [CLS] the man went to [MASK] store [SEP]
he bought a gallon [MASK] milk [SEP]
Label = IsNext

Input = [CLS] the man [MASK] to the store [SEP]
penguin [MASK] are flight ##less birds [SEP]
Label = NotNext

And are more generalizable than previous models...

John Hewitt and Christopher D. Manning: A Structural Probe for Finding Syntax in Word Representations

Andy Coenen, Emily Reif, Ann Yuan, Been Kim, Adam Pearce, Fernanda Viégas, Martin Wattenberg. Visualizing and Measuring the Geometry of BERT
Was the knowledge inside all along?
Is BERT good at Common Sense?

Source: Jeff Da, Jungo Kasai: Cracking the Contextual Commonsense Code: Understanding Commonsense Reasoning Aptitude of Deep Contextual Representations
Common sense understanding is not the default

• Question-answering AIs learn to give typical answers like “Yes” and “2”
• Story-understanding AIs learn that the wrong answers are written in a different style
• Reading-comprehension AIs learn to copy the nearest answer-shaped phrase

Common sense understanding is not the default

Source: Janelle Shane, on Twitter, using Visual Chatbot
We understand each other perfectly
Training a transformer takes a lot of energy.
Often, there isn’t enough data for a domain or language to train a new model.
One Size (Pre-trained) Model Does **Not** Fit All:

- Language*
- Domain
- Cultural/Identity
The ensemble learning process

What is ConceptNet Numberbatch?

Example edges

What is ConceptNet?

uni
Align the results, and combine them into smaller, distributional source, keeping both vocabularies.

(pre-trained outputs of word2vec and GloVe).

Acquire some good distributional word embeddings

knowledge graph (ConceptNet 5.5).

Combine sources of relational knowledge into a embeddings that are signi-

ConceptNet Numberbatch

slept
fi
ed vectors using principal component analysis.

FormOf
ExternalURL
Synonym
IsA
Antonym
RelatedTo
Relation
motivated by goal

shows how to reason over ConceptNet using


LRA (2006) Searches the Web for word pairs

word2vec (2013) Distributional semantics .486


Method

System

Evaluation name As published

Performance of this system (ConceptNet 5.5 + Numberbatch) Performance on Turney (2003) dataset

Evaluation Score

ρ correlation)

Adding ConceptNet improves performance signi-

0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

Results: State-of-the-art semantic vectors

Rare Words

Languages in ConceptNet

English
6.5 million edges

French
4.9 million edges

German
1.6M

Italian
1.1M

Spanish
830k

Japanese
740k

Russian
620k

Portuguese
540k

Chinese
500k

Finnish
420k

Dutch
400k

Swedish
300k

bg
pl
cs
sh
eo
ms
sl
ar

Total: 24.6 million edges in 70+ languages
### Evaluations in Portuguese

#### Table 15. Accuracy for answering cloze questions.

<table>
<thead>
<tr>
<th>LKB</th>
<th>Relations</th>
<th>Algorithm</th>
<th>( \rho )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAPÉL</td>
<td>All</td>
<td>PR-Jac800</td>
<td>0.49</td>
</tr>
<tr>
<td>DA</td>
<td>All</td>
<td>PR-Jac800</td>
<td>0.38</td>
</tr>
<tr>
<td>Wikt. PT</td>
<td>All</td>
<td>PR-Jac800</td>
<td>0.42</td>
</tr>
<tr>
<td>OWN-PT</td>
<td>Syn + Hyp</td>
<td>Adj-Cos</td>
<td>0.44</td>
</tr>
<tr>
<td>PULO</td>
<td>Syn + Hyp</td>
<td>Adj-Cos</td>
<td>0.29</td>
</tr>
<tr>
<td>TeP</td>
<td>Syn + Hyp</td>
<td>Adj-Jac</td>
<td>0.36</td>
</tr>
<tr>
<td>OT-PT</td>
<td>Syn + Hyp</td>
<td>Adj-Cos</td>
<td>0.34</td>
</tr>
<tr>
<td>Port4Nooj</td>
<td>All</td>
<td>Adj-Jac</td>
<td>0.19</td>
</tr>
<tr>
<td>WN.Br</td>
<td>Syn + Hyper</td>
<td>Adj-Jac</td>
<td>0.04</td>
</tr>
<tr>
<td>ConceptNet</td>
<td>Syn + Hyper</td>
<td>Adj-Jac</td>
<td>0.43</td>
</tr>
<tr>
<td>CARTÃO</td>
<td>All</td>
<td>PR-CosV100</td>
<td>0.53</td>
</tr>
<tr>
<td>Redun3</td>
<td>Syn + Hyper</td>
<td>Adj-Jac</td>
<td>0.44</td>
</tr>
<tr>
<td>Redun2</td>
<td>Syn + Hyper</td>
<td>PR-Jac80</td>
<td>0.49</td>
</tr>
<tr>
<td>All</td>
<td>Syn + Hyper</td>
<td>PR-CosV80</td>
<td>0.57</td>
</tr>
<tr>
<td>All</td>
<td>Syn + Hyper</td>
<td>PR-CosV100</td>
<td>0.59</td>
</tr>
<tr>
<td>All</td>
<td>Syn + Hyper</td>
<td>PR-CosV200</td>
<td>0.61</td>
</tr>
<tr>
<td>All</td>
<td>Syn + Hyper</td>
<td>PR-CosV400</td>
<td>0.61</td>
</tr>
<tr>
<td>All</td>
<td>Syn + Hyper</td>
<td>PR-CosV100</td>
<td>0.61</td>
</tr>
<tr>
<td>All</td>
<td>Syn + Hyper</td>
<td>PR-CosV3200</td>
<td>0.60</td>
</tr>
<tr>
<td>All</td>
<td>Syn + Hyper</td>
<td>Adj-Cos</td>
<td>0.58</td>
</tr>
<tr>
<td>All</td>
<td>Syn + Hyper</td>
<td>Adj-Jac</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Could you make transformers more accurate using data like we did with retrofitting?
Transformers + Knowledge

Source: *Jeff Da*: BIG MOOD: Relating Transformers to Explicit Commonsense Knowledge (COIN 2019)
Transformers + Knowledge

Transformers + ConceptNet


Table 2: The detailed procedure of constructing a multi-choice question answering sample with the proposed AMS method. The * in the fourth step is a wildcard character. The correct answer for the question is underlined.
Transformers + Entity Graphs

Source: M. Peters, M. Neumann, R. Logan, R. Schwartz, V. Joshi, S. Singh, N. Smith: Knowledge Enhanced Contextual Word Representations
Can we teach a network to retrofit?

- Use a GAN to teach a network to add a knowledge graph to an embedding
- Essentially increasing the vocabulary of the kgraph
- Better downstream performance
- Potential for work in bi-lingual embedding space

Source: Edoardo M. Ponti, Ivan Vulic, Goran Glavaš, Nikola Mrkšić, Anna Korhonen
Can we use transformers to build knowledge graphs?
Can transformers give back to knowledge graphs?

- Big idea: Increase the density of a knowledge graph by adding additional links

Inferring common sense with CoMET

- Uses ConceptNet as a training set instead of a knowledge resource
- Fine-tune a GPT language model to generate ConceptNet statements

Go ahead,
I’m listening...
What about human-centric AI?
Knowledge graphs & explainability?
It’s happening again: today’s agents don’t communicate like us.
Let’s find a way to communicate with machines like we always dreamed of.
conceptnet.io – a browsable interface

**bicycle**

An English term in ConceptNet 5.5

**Sources:** Open Mind Common Sense contributors, DBPedia 2015, JMDict 1.07, OpenCyc 2012, German Wiktionary, English Wiktionary, French Wiktionary, and Open Multilingual WordNet

**Synonyms**

- tr: bisiklet
- en: wheel (n)
- ja: 銀輪 (n)
- it: bici
- ar: دراجة هوائية (n)
- fr: vélo
- en: cycle
- da: cykel (n)
- it: bicicletta

**Related terms**

- en: biker (n)
- fr: bécanne
- en: tricycle
- en: penny farthing (n)
- br: marc'h houarn (n)
- en: propel
- en: bicicleta (n)
- en: like riding bicycle
- en: gasp (n)

**bicycle is a type of...**

- en: a two wheel vehicle
- en: means of transportation
- en: a machine
- en: ride (v)
- en: an efficient form of human transportation
- en: toy
- en: transportation
- en: wheeled vehicle (n)

**bicycle is used for...**

- en: transportation
- en: riding
- en: Racing
- en: personal transport
- en: ride (v)
- en: travelling on
- en: rush (v)
- en: cause cultural change
- en: traveling
api.conceptnet.io – a Linked Data API

{
  "@context": [
    "http://api.conceptnet.io/ld/conceptnet5.5/context.ld.json",
    "http://api.conceptnet.io/ld/conceptnet5.5/pagination.ld.json"
  ],
  "@id": "/c/en/bicycle",
  "edges": [
    {
      "@id": "/a[/r/AtLocation/,/c/en/bicycle/,/c/en/garage/]",
      "dataset": "/d/conceptnet/4/en",
      "end": {
        "@id": "/c/en/garage",
        "label": "the garage",
        "language": "en",
        "term": "/c/en/garage"
      },
      "license": "cc:by/4.0",
      "rel": {
        "@id": "/r/AtLocation",
        "label": "AtLocation"
      }
    }
  ]
}
We’d like to work with you…

• Building and expanding ConceptNet
• Cross-resource alignment
• Bias in machine learning and structural resources
• Explainablity in kgs
• Other ideas
Thank you to all ConceptNet Collaborators

- Co-director Robyn Speer
- Jason Alonso, Charlotte Chen, Pedro Colon-Hernandez, Joanna Lowry-Duda, Nina Lutz, Katherine Xiao
- So many others over the years…

blog.conceptnet.io
Thank you to all of my crowds over the years.

Photo: Moses Namkung, CC-By
Thank you!

Dr. Catherine Havasi
havasi@media.mit.edu
@catherinehavasi