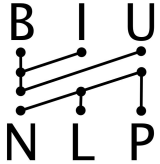


How Large are Lions? Inducing Distributions over Quantitative Attributes

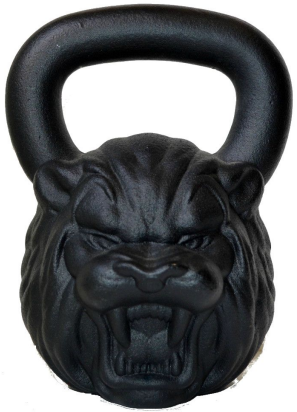
ISCOL, September 11, 2019

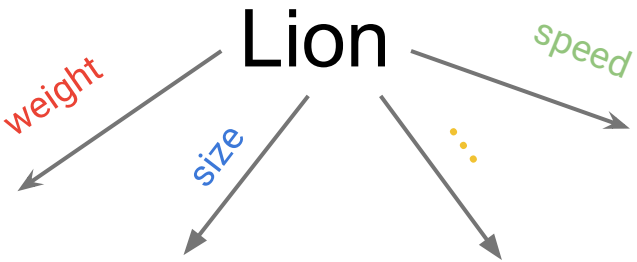
Yanai Elazar
Abhijit Mahabal
Deepak Ramachandran
Tania Bedrax-Weiss
Dan Roth

Quantitative Understanding

Quantitative Understanding

- Understanding numerical properties and the way they relate to words.

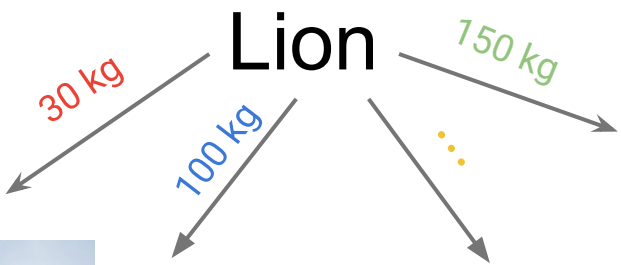




Physical attributes

Quantitative Understanding

- Understanding numerical properties and the way they relate to words.



Commonsense quantization attributes

Quantitative Understanding in Q&A

- “What is a fast but expensive way to send small cargo?”
  - Ship’s hold
  - Boat
  - Airplane

Talmor et. al 2019

Quantitative Understanding in Q&A

- “What is a **fast** but **expensive** way to send small cargo?”
  - Ship’s hold → Slow
  - Boat → Slow
  - Airplane → **Fast** and **Expensive**

Talmor et. al 2019

Other Quantitative Work

Shameless plug

Since the summer of '99 I haven't eaten apples like this one

Other Quantitative Work

Shameless plug

I haven't eaten apples like this **one** since the summer of '**99**

YEAR

TACL 19'

Other Quantitative Work

Shameless plug

spaCy plugin

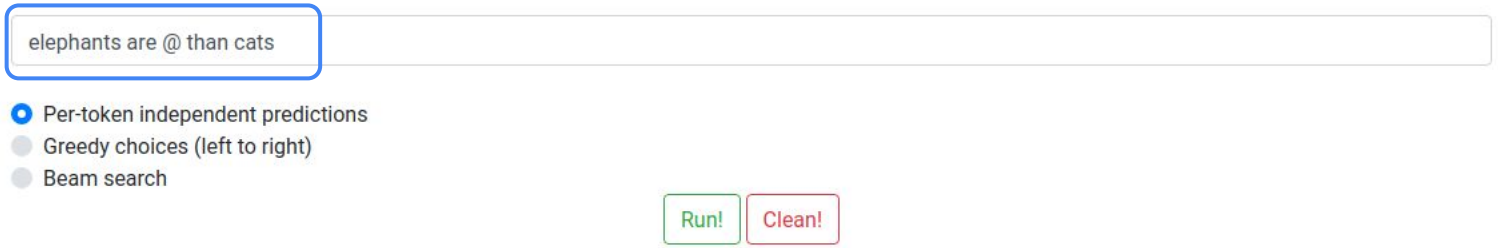
TACL 19'

```python
import spacy
from num_fh import NFH
nlp = spacy.load('en_core_web_sm')
nfh = NFH(nlp)
nlp.add_pipe(nfh, last=True)

doc = nlp("Since the summer of '99 I haven't eaten apples like this one")

assert doc[5]._.is_nfh == True
assert doc[5]._.nfh_head == 'YEAR'
assert doc[-1]._.is_nfh == True
assert doc[-1]._.nfh_head == doc[10]  # doc[10].text == 'apple
```

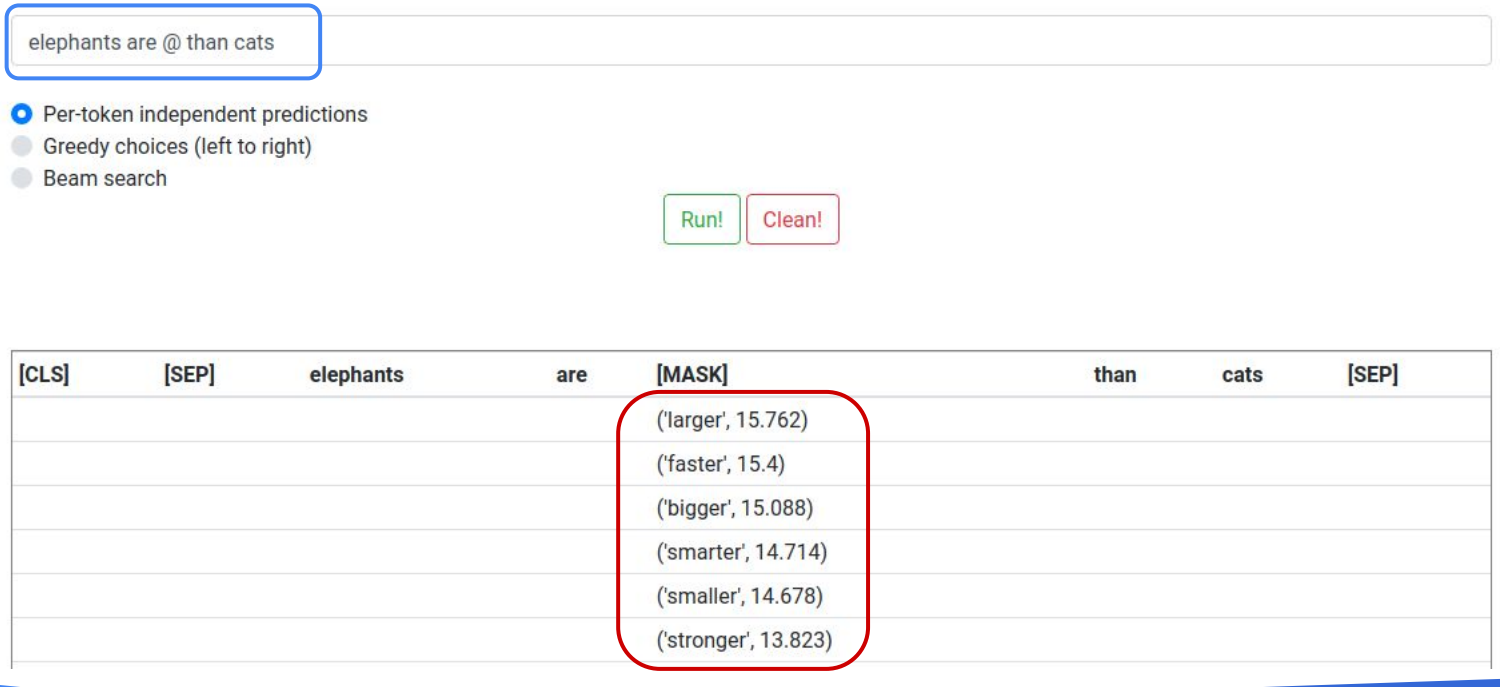
Quantitative Understanding

- It is hard to generalize numerical quantization and common sense from datasets alone.
- Running End-to-End distributional solutions on these tasks is not enough to solve them.

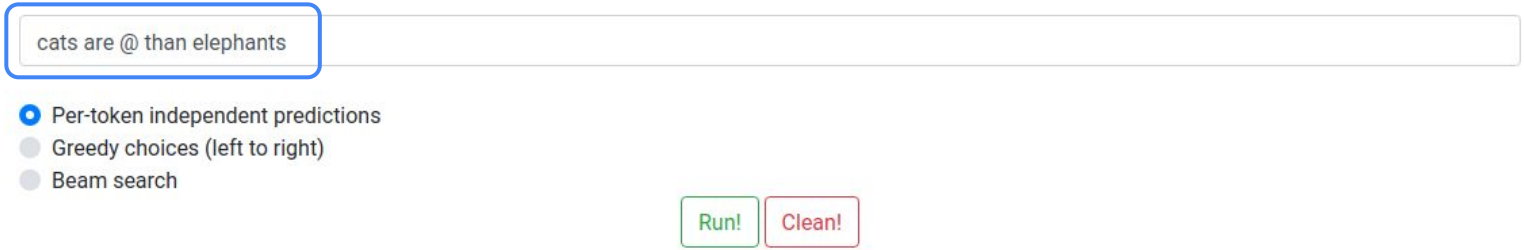
Quantitative Understanding



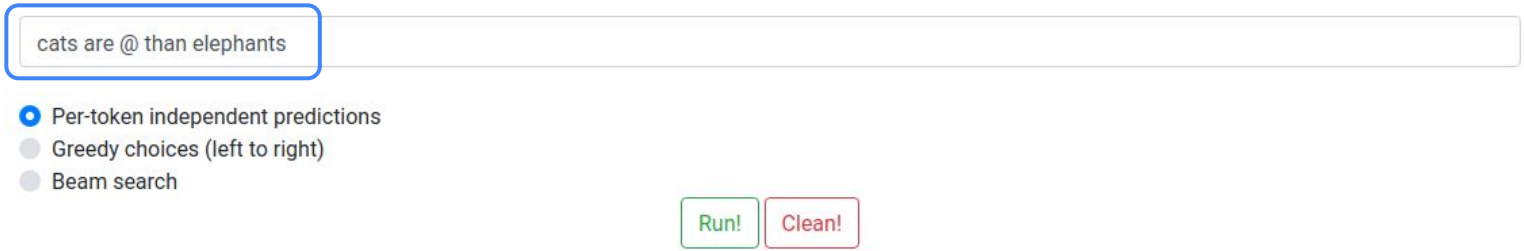
Quantitative Understanding



Quantitative Understanding



Quantitative Understanding



[CLS]	[SEP]	cats	are	[MASK]	than	elephants	[SEP]
				('larger', 17.199)			
				('faster', 15.985)			
				('smaller', 15.976)			
				('bigger', 15.79)			
				('smarter', 14.794)			
				('stronger', 14.579)			

Scalable Attributes of Objects

Google

BIU NLP

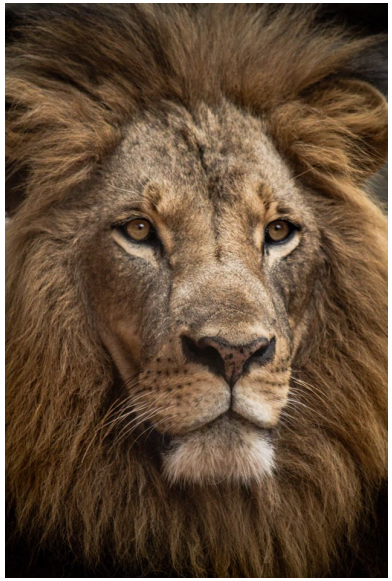
Let’s ground our “Measurable World”

We focus on…

- Items which can be measured objectively

How Big is a...

https://en.wikipedia.org/wiki/Mouse



https://unsplash.com/photos/IPRFX7CVVoU

https://www.thisisinsider.com/homes-popular-style-us-2017-10

How big is Big?

Let’s ground our “Measurable World”

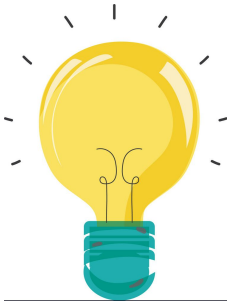
- These can be object’s attributes, but also other things, like adjective, verbs, etc...

Solution - Counting!

Google

BIU NLP

The Idea



- Count co-occurrences of measurements with the words that appear in their context
- By using a large text corpora

Counting can be useful!

- Google Books NGram

Counting can be useful!

- Google Books NGram
- Google Syntactic NGram

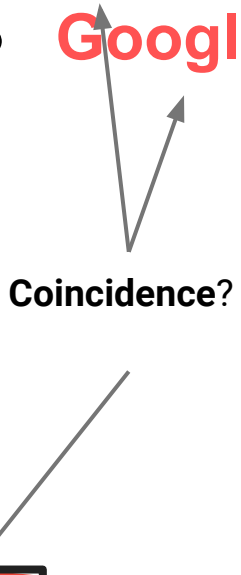
A Dataset of Syntactic-Ngrams over Time from a Very Large Corpus of English Books

Yoav Goldberg
Bar Ilan University*
yoav.goldberg@gmail.com

Jon Orwant
Google Inc.
orwant@google.com

Counting can be useful!

- **Google** Books NGram
- **Google** Syntactic NGram



A Dataset of Syntactic-Ngrams over Time from a Very Large Corpus of English Books

Yoav Goldberg
Bar Ilan University*
yoav.goldberg@gmail.com

Jon Orwant
Google Inc.
orwant@google.com

Google

B I U
N L P

Example - Walk Through The Process

Example - Input Sentence

“These breeds can vary in size and weight from a 0.46 kg teacup poodle ...”

Source: Wikipedia

Example - Measurement Detection

“These breeds can vary in size and weight from a

***0.46 kg** teacup poodle ...”*

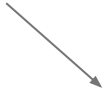
Source: Wikipedia

We detect numerical measurements using a set of rules:
kg/kgs/kilogram -> MASS

Example - Measure Normalization

“These breeds can vary in size and weight from a

***0.46 kg** teacup poodle ...”*



Source: Wikipedia

460 gram

Using the units and the measurement type to normalize the number

Example - Co-Occurring objects

Noun *Verb* *Noun* *Noun*

“These breeds can vary in size and weight from a

***0.46 kg** teacup poodle …”*

NP

Source: Wikipedia

460 gram

We detect objects of interest (Nouns, Adjectives and Verbs) using a POS tagger.

Example - Aggregating Measurements

Noun *Verb* *Noun* *Noun*

"These breeds can vary in size and weight from a

***0.46 kg** teacup poodle ..."*

NP

460 gram

Source: Wikipedia

```
objects_distribution['poodle']['mass'] += [460]
objects_distribution['breeds']['mass'] += [460]
...
```

Example - Aggregating Measurements

```
In  [2]: objects_distribution
Out [2]: {'poodle': {
                'mass' : [460, 400, 350, 800, 16000],
                'speed': [5, 8, 1.5],
                ...
              },
           'car': {
                 'speed': [100, 80, 50, 50, 120, 40],
                 ...
           },
           ...
         }
```

Graham Neubig
@gneubig

Following

One commonly cited argument about the difficulty of learning common-sense reasoning is that "no-one writes down common sense". A counter-argument is "well, the web is big": instructables.com/id/How-To-Open…

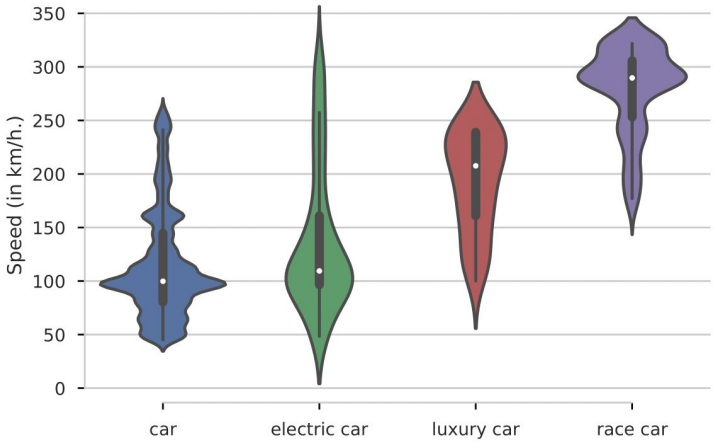
How to Open a Door

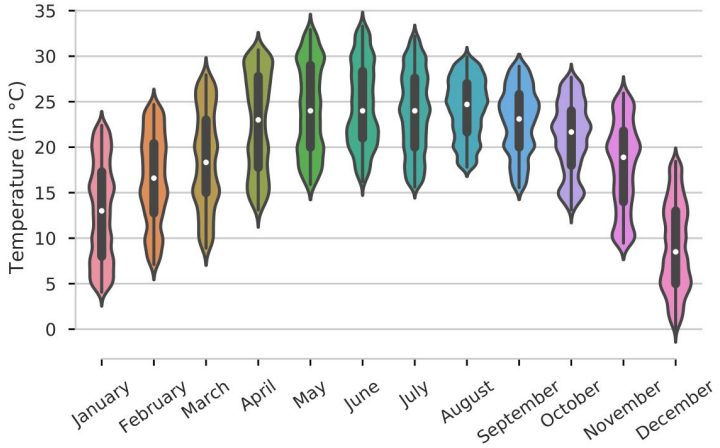
Step 1: Locate Desired Door

Step 2: Locate Door Handle or Knob

Step 3: Turn Knob or Handle and Pull or Push

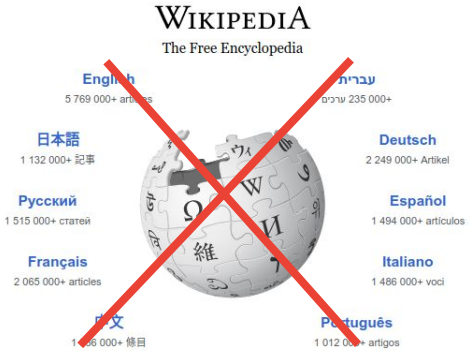
Example - Aggregating Measurements





Underlying Resource

Resource Statistics - Origin



Not enough data

Billion of web pages! (in English)

Resource Statistics - DoQ

- We present: **D**istributions **o**ver **Q**uantities (DoQ)
- A very large and diverse resource

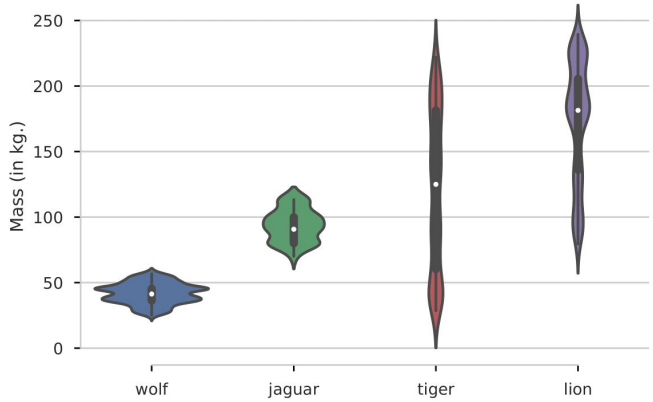
- ~120M Unique tuples (object, measurement)
  - ~350K with >= 1000 occurrences

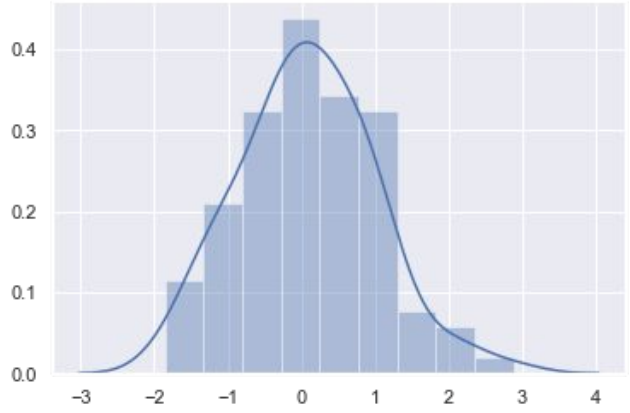
Resource Statistics - DoQ

- Measurement types:
  - Length
  - Mass
  - Currency
  - Temperature
  - …
- 27 In total (But not all are useful)

Using DoQ

- We collected a bunch of numbers for each key
- Which in turn creates: Distributions!





Using DoQ

- Given two objects and a scale, we can compare them using their corresponding distributions
- By:
  - Comparing the Mean - Noisy
  - Comparing the Median - Better
  - Comparing a Statist - Doesn’t make much difference, but returns a probability

Quantitative Evaluation

Comparable Objects

- Comparing 2 objects on a given dimension
- Nouns
  - 3 different datasets (including a new one we created)
- Adjectives
  - 2 different datasets

Comparable Objects

VERB PHYSICS: Relative Physical Knowledge of Actions and Objects

Maxwell Forbes **Yejin Choi**
Paul G. Allen School of Computer Science & Engineering
University of Washington
{mbforbes,yejin}@cs.washington.edu

- A dataset of ~3.6K object pairs, compared on 5 dimension (e.g. speed, weight, size)

Comparable Objects

Extracting Commonsense Properties from Embeddings with Limited Human Guidance

Yiben Yang [1], Larry Birnbaum[2], Ji-Ping Wang[1], Doug Downey[2]

[1]Department of Statistics, Northwestern University, Evanston, IL, 60208, USA
[2]Department of Electrical Engineering & Computer Science, Northwestern University, Evanston, IL, 60208, USA
[1]{yiben.yang,jzwang}@northwestern.edu
[2]{l-birnbaum,d-downey}@northwestern.edu

- Learning a transformation over pre-trained word embedding to infer relations

Comparable Objects

Are Elephants Bigger than Butterflies?
Reasoning about Sizes of Objects

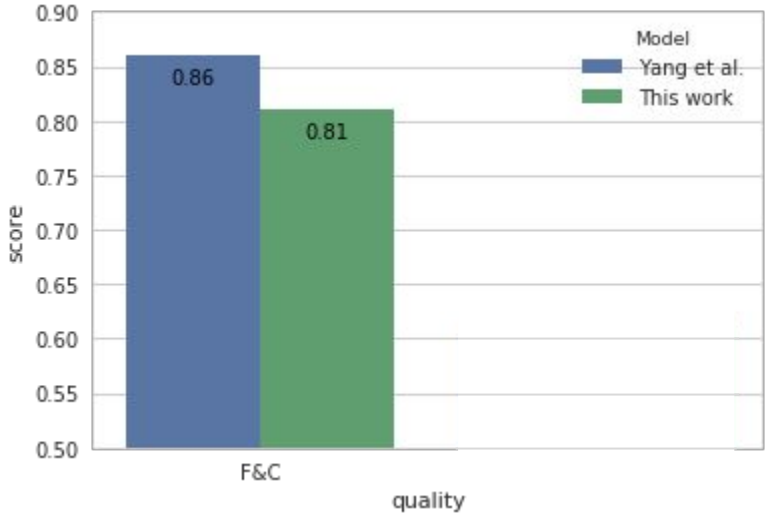
Hessam Bagherinezhad[†] and **Hannaneh Hajishirzi**[†] and **Yejin Choi**[†] and **Ali Farhadi**[†‡]
[†]University of Washington, [‡]Allen Institute for AI
{hessam, hannaneh, yejin, ali}@washington.edu

- Dataset for size comparison
- A combination of Images and texts to infer sizes

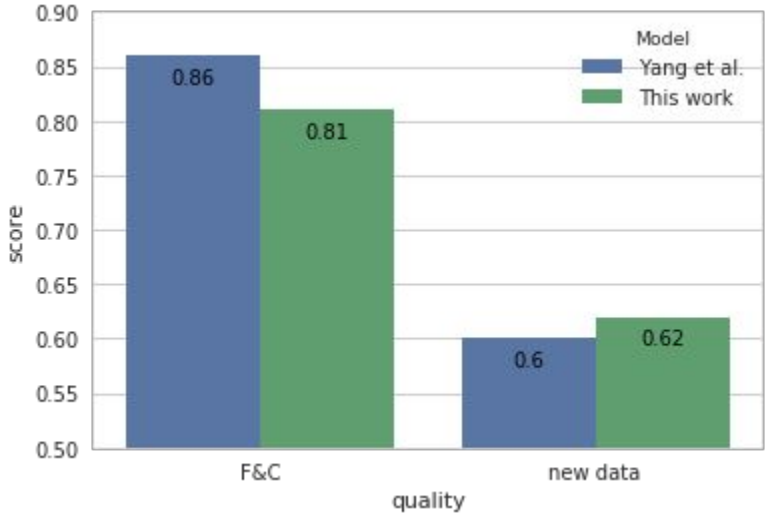
Comparable Objects

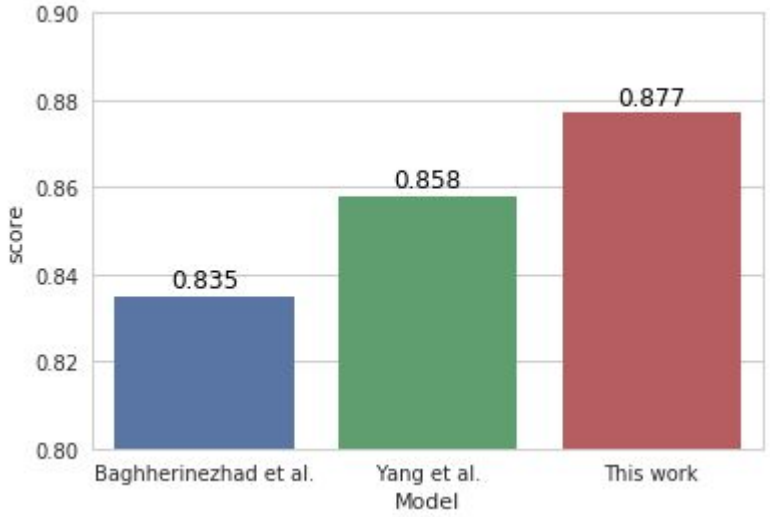
- In this work, we introduce a new dataset for object comparison
- 4 dimension (including Currency, which wasn’t evaluated on before)
- High agreement score (77.1 Kappa)

Comparable Objects - Results



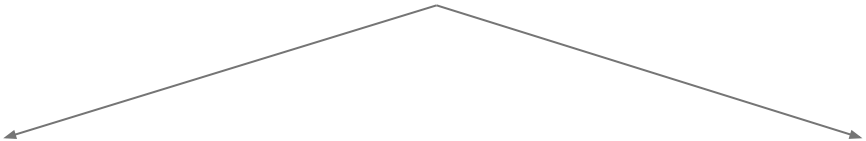
Comparable Objects - Results





Comparable Adjectives - Intensifiers

Freezing < Cold < Warm < Hot



Freezing < Cold

Warm < Hot

Comparable Adjectives - Intensifiers

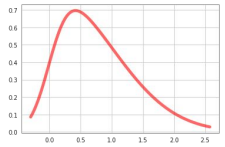
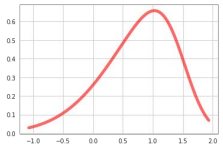
- Previous work used Open-IE style methods to infer relations between two objects
  - E.g *“hot and almost scorching”*

Comparable Adjectives - Intensifiers

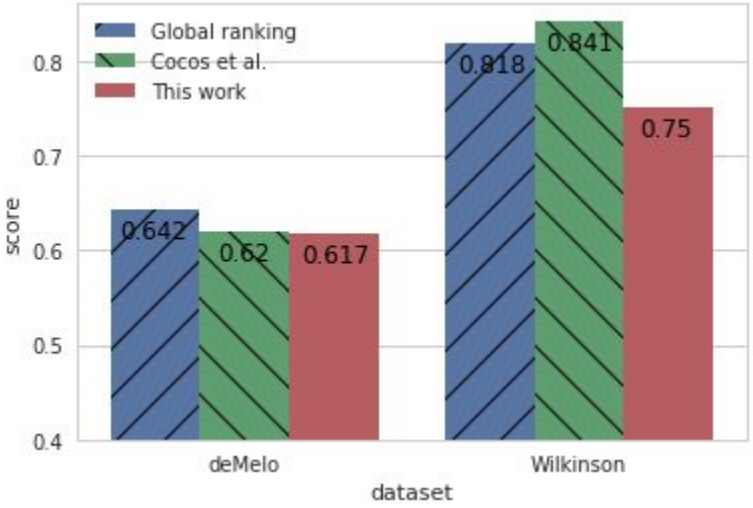
- Previous work used Open-IE style methods to infer relations between two objects
  - E.g *“hot and almost scorching”*

X < Y

- We have concrete individual distributions for each term, so we don’t rely on specific comparisons

X  <  Y

Comparable Adjectives Inference

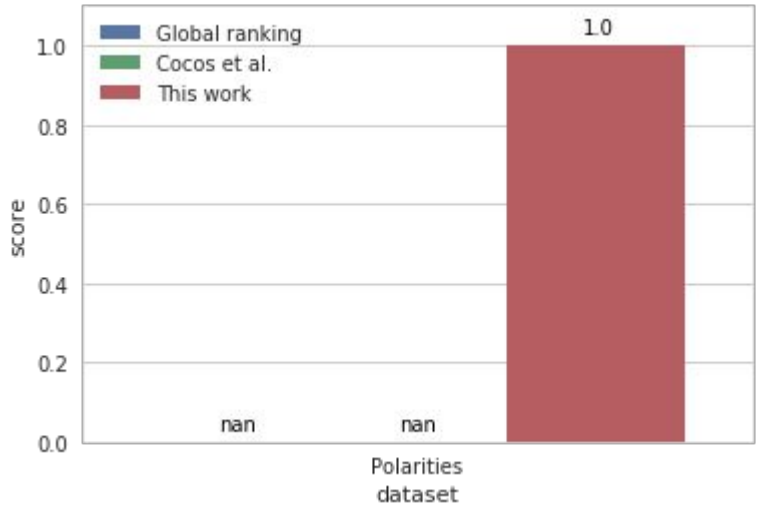


Comparable Adjectives - Polarities

“hot and almost scorching”

“hot and almost freezing”

Freezing, Cold < Warm, Hot



Intrinsic Evaluation

Intrinsic Evaluation

- Extract the median of “popular” noun distributions
- Expand to a range
  - 20 mm → 10-100 mm

Intrinsic Evaluation

- Extract the median of “popular” noun distributions
- Expand to a range
  - 20 mm → 10-100 mm
- Ask annotators if the item fits the range

Intrinsic Evaluation

- Extract the median of “popular” noun distributions
- Expand to a range
  - 20 mm → 10-100 mm
- Ask annotators if the item fits the range
  - “Is the usual length of a screw between 10-100mm?”

Intrinsic Evaluation

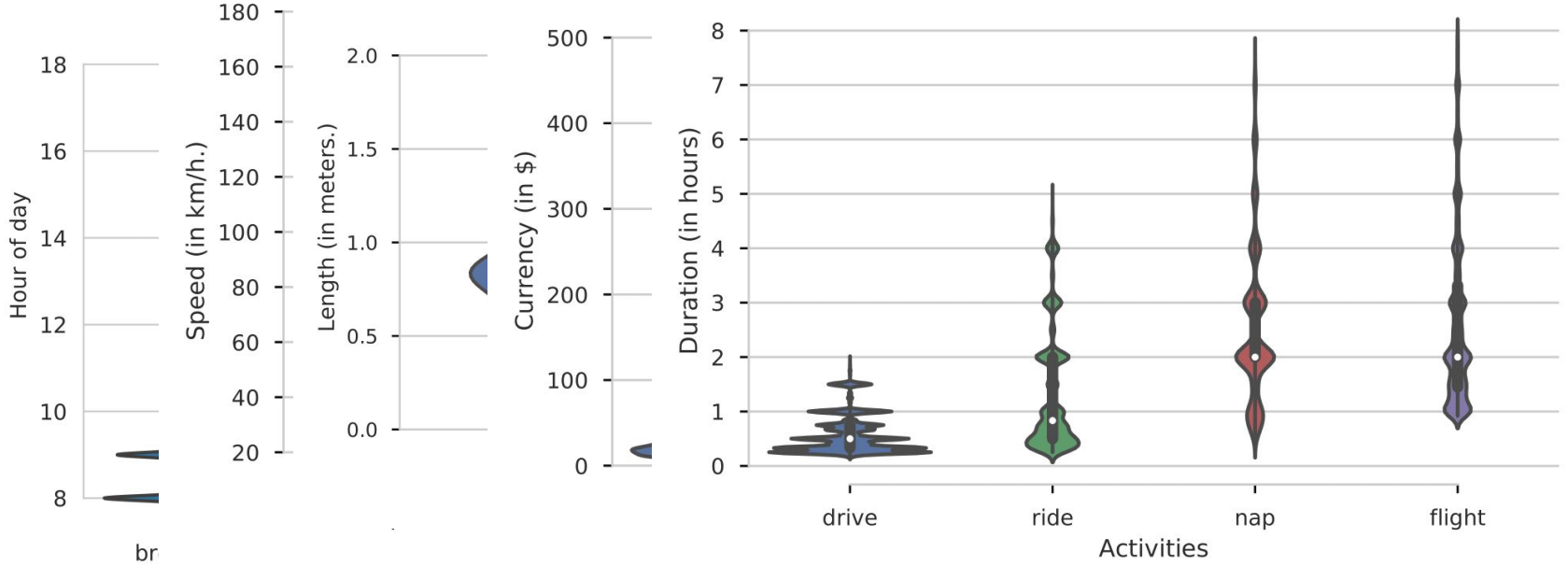
- 69% agreement with our predictions
- Not perfect, but a good start for acquiring such knowledge

Qualitative Analysis

Comparable Objects - Cool Results

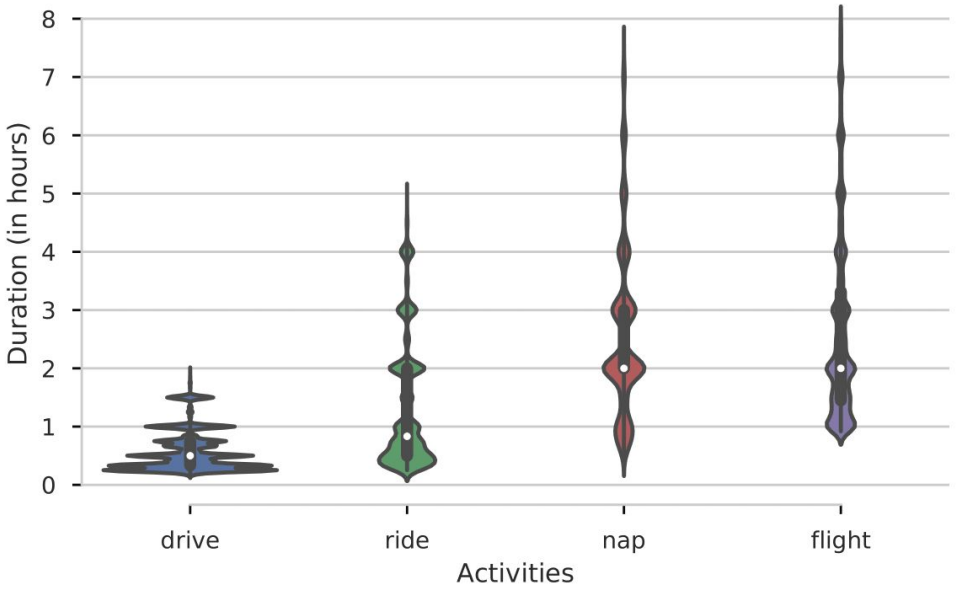
- Many (many) cool and accurate examples

Comparable Objects - Cool Results

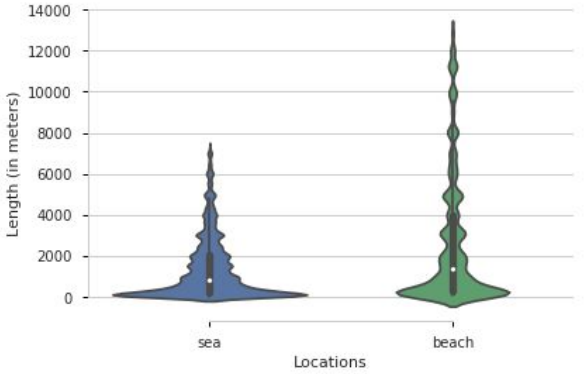


Comparable Objects - Cool Results

We will focus on the errors



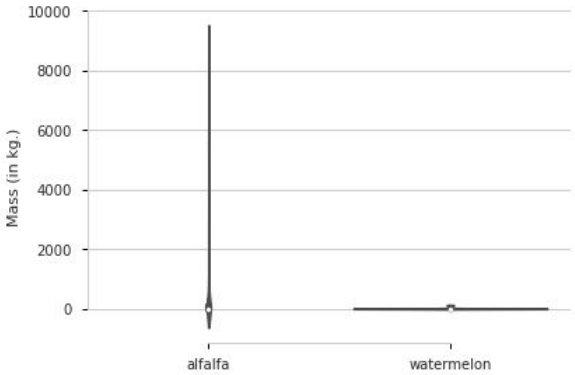
Comparable Objects - Some Issues



“Elevation ranges from **3,000 feet** ... above ***sea*** level.”

That’s a small sea!

Comparable Objects - Some Issues



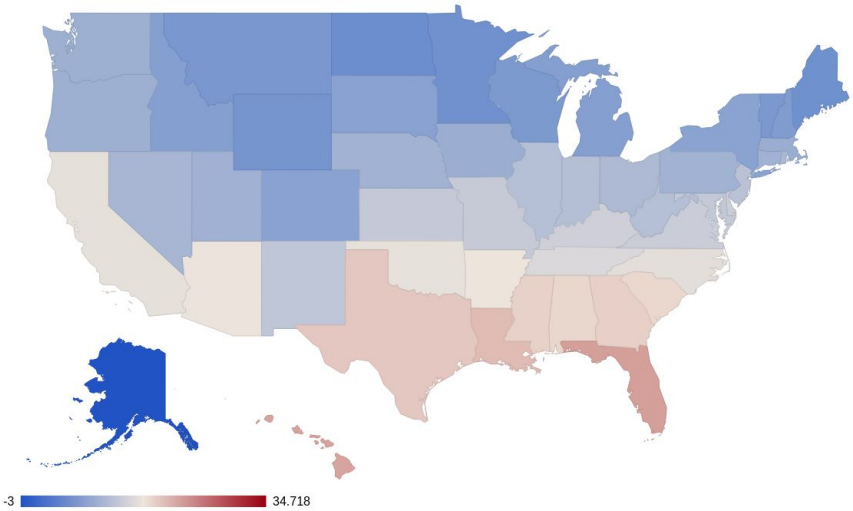
That’s a heavy alfalfa

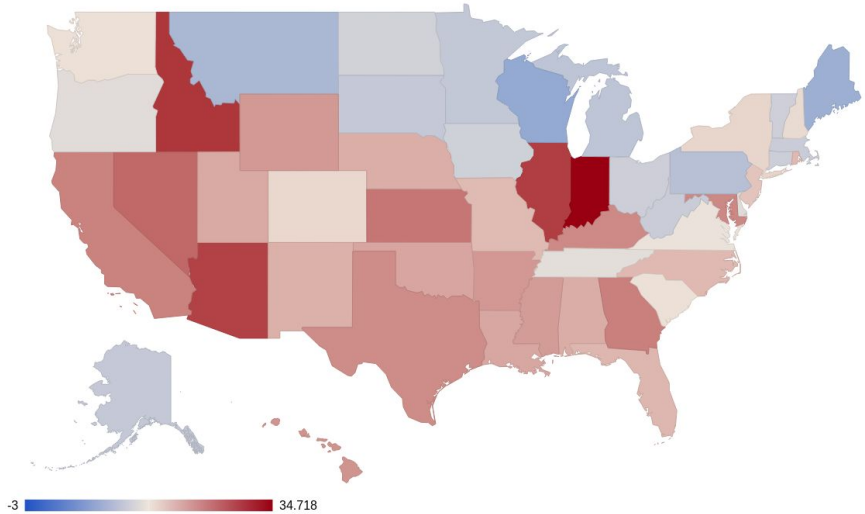
“***Alfalfa*** is the most cultivated legume ... reaching around **454 million tons** ...”

https://alivebynature.com/the-right-way-to-eat-alfalfa-sprouts/

Comparable Objects - Case Study

Collected temperatures of US States





“Real” average

Predicted median

Summary

- A simple method for collecting measure attribution
- Obtaining distribution for a various of objects
- Releasing a big, new and unique resource
- Releasing a refined annotation for an existing dataset and a new one.

Thanks