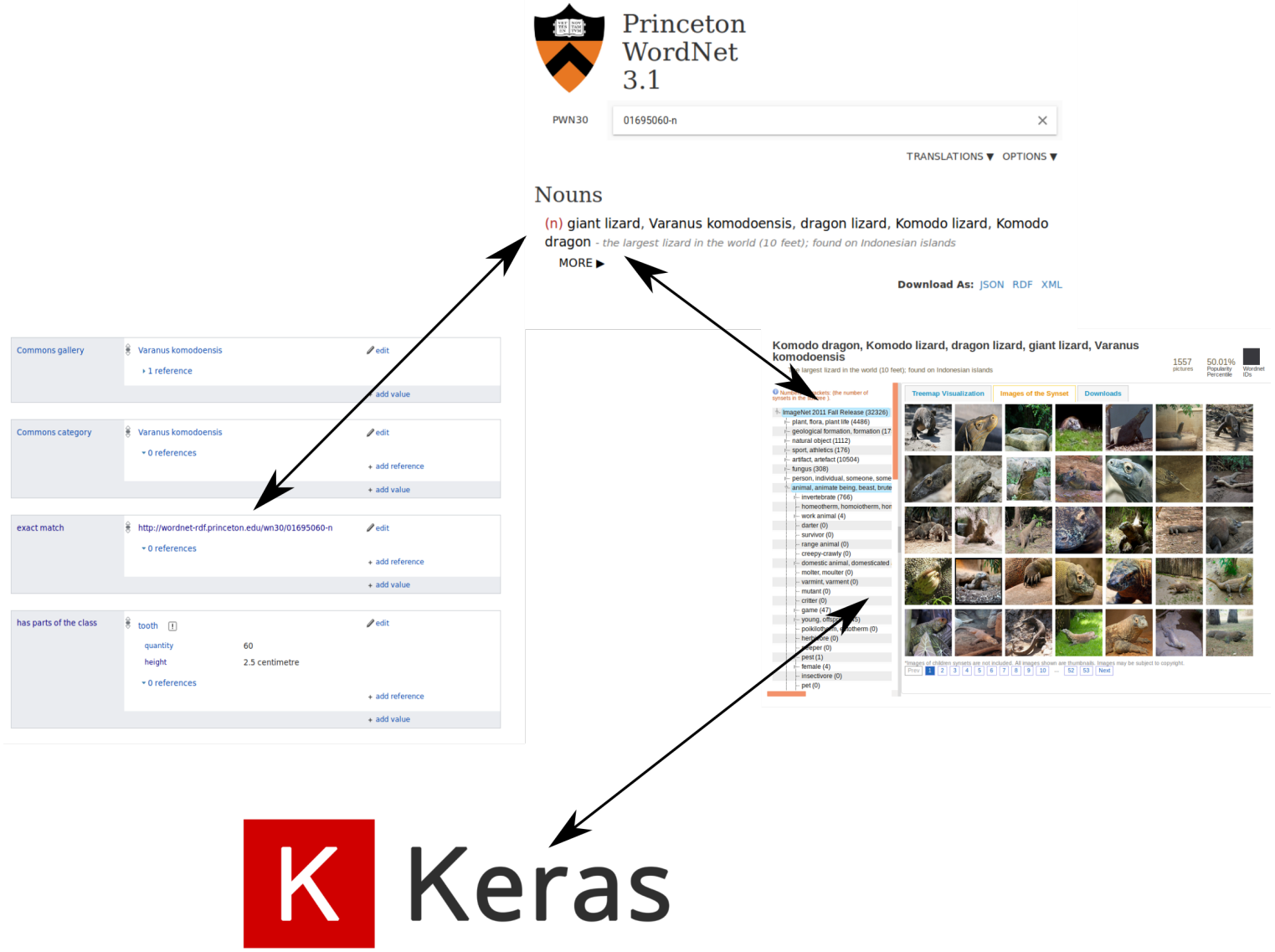
Inferring visual semantic similarity with deep learning and Wikidata: Introducing imagesim-353

Finn Årup Nielsen, Lars Kai Hansen

Cognitive Systems, DTU Compute, Technical University of Denmark

4 June 2018

Wikidata–WordNet–ImageNet–Keras



Wikidata–WordNet–ImageNet–Keras

We got a linkage from the Wikidata knowledge graph through the WordNet semantic network and ImageNet to pre-trained deep learning models (Nielsen, 2018).

Wikidata–WordNet–ImageNet–Keras

We got a linkage from the Wikidata knowledge graph through the WordNet semantic network and ImageNet to pre-trained deep learning models (Nielsen, 2018).

Can we use that for something interesting?

Wikidata–WordNet–ImageNet–Keras

We got a linkage from the Wikidata knowledge graph through the WordNet semantic network and ImageNet to pre-trained deep learning models (Nielsen, 2018).

Can we use that for something interesting?

Beyond using Wikidata as a multilingual labeler, as in (Nielsen, 2018).

Wikidata–WordNet–ImageNet–Keras

We got a linkage from the Wikidata knowledge graph through the WordNet semantic network and ImageNet to pre-trained deep learning models (Nielsen, 2018).

Can we use that for something interesting?

Beyond using Wikidata as a multilingual labeler, as in (Nielsen, 2018).

Visual semantic similarity?

Wikidata–WordNet–ImageNet–Keras

We got a linkage from the Wikidata knowledge graph through the WordNet semantic network and ImageNet to pre-trained deep learning models (Nielsen, 2018).

Can we use that for something interesting?

Beyond using Wikidata as a multilingual labeler, as in (Nielsen, 2018).

Visual semantic similarity? Utilizing the links in the Wikidata knowledge graph?

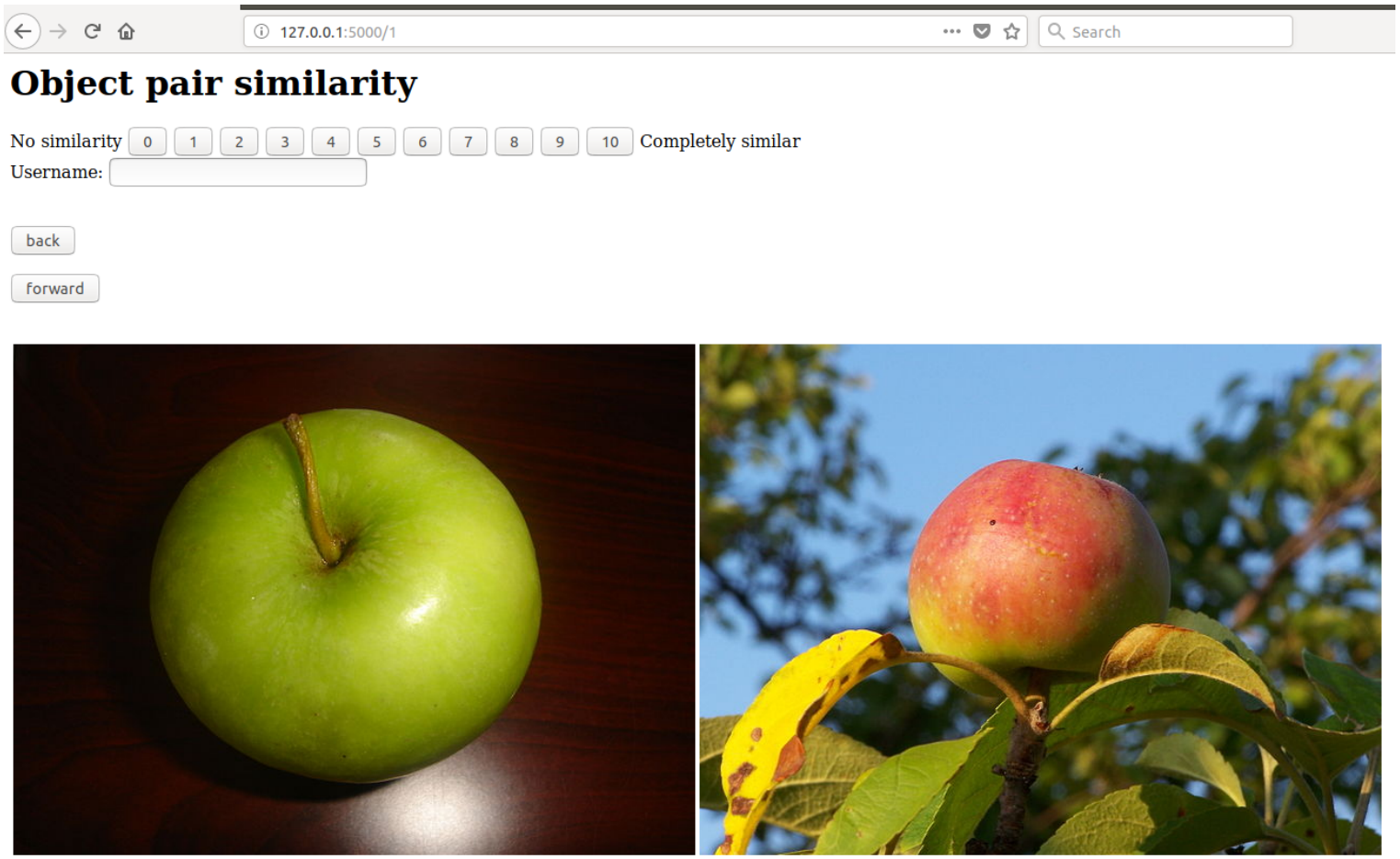
Textual semantic similarity

WordSim-353 (Finkelstein et al., 2002): 353 word pairs with score.

Word 1	Word 2	Human (mean)
love	sex	6.77
tiger	cat	7.35
tiger	tiger	10.00
book	paper	7.46
⋮		
smart	student	4.62
stock	phone	1.62
stock	CD	1.31
stock	jaguar	0.92
⋮		

Several other word lists exist, e.g., one that stresses the difference between relatedness and similarity (Hill et al., 2014). Widely used to evaluate textual semantic models .

Visual semantic similarity labeler



Rudimentary Python-Flask webapp built for labeling pairs of images for semantic similarity with a score from 0 to 10.

imagesim-353 Dataset

We collected 353 free images from Wikimedia Commons.

Scored 353 image pairs for similarity.

Deep learning for image classification

Use a pre-trained model: The ResNet-50 model supplied from Keras.

Scale the image to the input of the target size, i.e., the input of the deep neural network. Preprocess.

Predict with `y = model.predict(x)` where the output is probabilities over class labels.

Building a bag-of-properties matrix

474 results in 4427 ms </> Code Download Link

Search

item	synset	properties
wd:Q4504	n01695060	P10 P18 P31 P105 P138 P141 P171 P181 P183 P225 P227 P244 P373 P462 P462 P462 P508 P627 P646 P685 P815 P830 P838 P846 P935 P1034 P1417 P1843 P1843 P1843 P1843 P1843 P1843 P1843 P1843 P2040 P2043 P2052 P2067 P2250 P2670 P2833 P2888 P2924 P2974 P2974 P2974 P2974 P3031 P3151 P3387 P3417 P3465 P3827 P4024 P4728
wd:Q4439	n03492542	P18 P156 P227 P279 P279 P279 P373 P487 P508 P575 P646 P910 P935 P1330 P1482 P2167 P2581 P2581 P2888 P3365 P3381 P3417 P4746 P5008
wd:Q653	n02085620	P18 P31 P279 P373 P495 P528 P646 P935 P1417 P1417 P2581 P2888 P2888
wd:Q144	n02084071	P3417 P3465 P3479 P3485 P3827 P3984 P4733 P5008 P5063 P18 P18 P31 P227 P279 P279 P349 P366 P366 P366 P366 P373 P417 P417 P417 P487 P487 P508 P646 P902 P910 P935 P1036 P1036 P1051 P1245 P1319 P1343 P1417 P1843 P1843 P1843 P1843 P1843 P1843 P1843 P1843

Building a bag-of-properties SPARQL

```
SELECT ?item (SAMPLE(?synset_) AS ?synset)
       (GROUP_CONCAT(?pid) AS ?properties)
WITH {
  SELECT ?item ?uri WHERE {
    ?item wdt:P2888 ?uri .    # P2888 is "exact␣match"
    FILTER STRSTARTS(STR(?uri),
      "http://wordnet-rdf.princeton.edu/wn30/")
  }
} AS %items
WHERE {
  INCLUDE %items
  ?item ?p [] .
  ?property wikibase:directClaim ?p .
  BIND(SUBSTR(STR(?p), 37) AS ?pid)
  BIND(CONCAT('n', SUBSTR(STR(?uri), 39, 8)) AS ?synset_)
}
GROUP BY ?item
```

Bag-of-properties matrix

	P10 video	P18 image	P31 instance	P105 taxon rank	...
Varanus komodoensis	1	1	1	1	
backpack	0	1	0	0	
automobile	0	1	0	0	
Cricetinae	0	1	0	1	
⋮					

This bag-of-properties matrix, $\mathbf{W}$, is of size 1000 times 192.

This matrix can be scaled in various ways. We have used *tfidf* from scikit-learn: $\tilde{\mathbf{W}}$.

Match between similarity score results

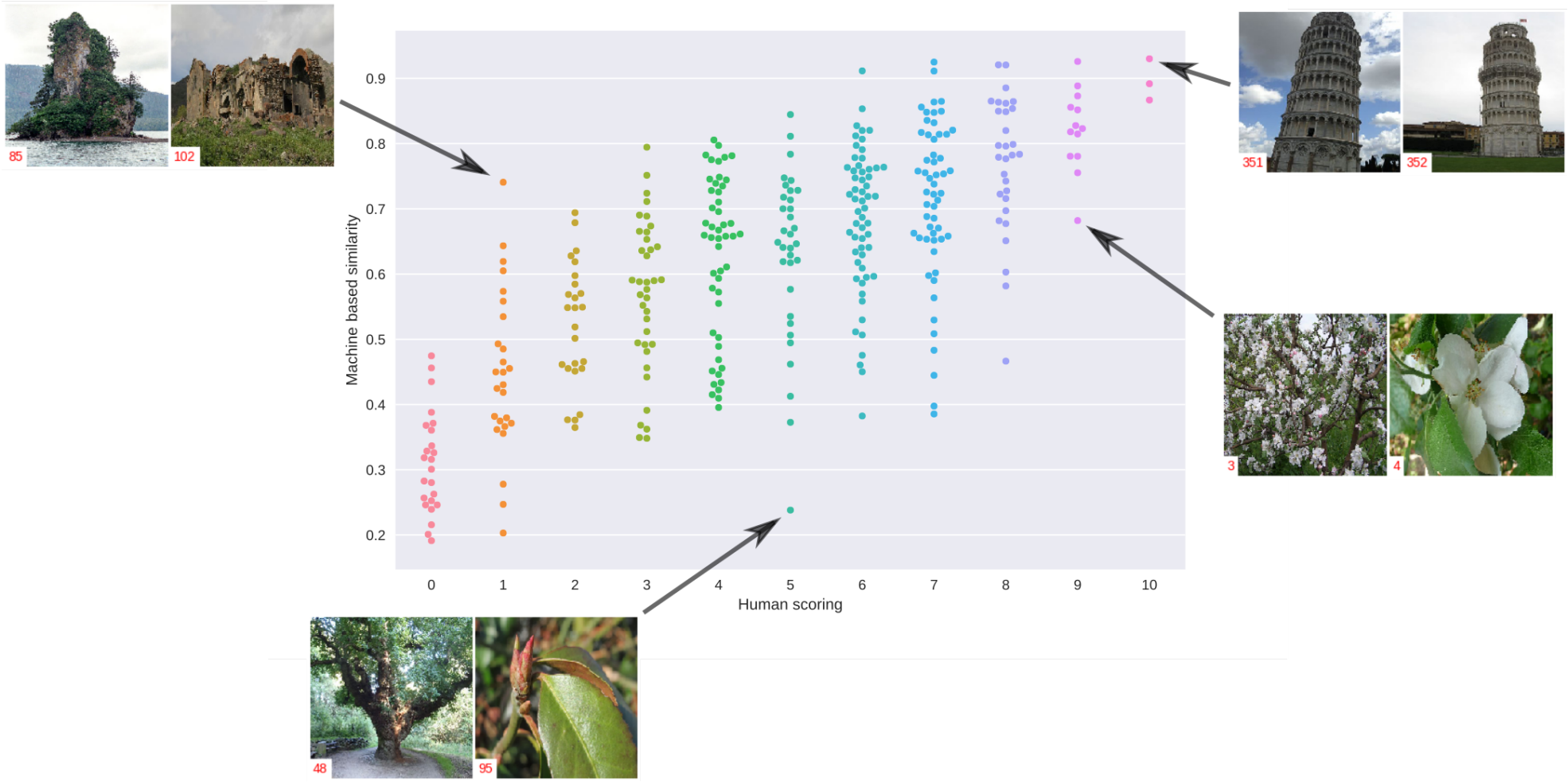
Description	Equation	Correlation
Pixel values (baseline)	$\mathbf{X}$	0.12
Untransformed NN output	$\mathbf{Y}$	0.62
Log-transformed NN output	$\ln(\mathbf{Y}) = \tilde{\mathbf{Y}}$	0.67
Log-transformed NN output and BoP	$\tilde{\mathbf{Y}}\tilde{\mathbf{W}} = \mathbf{Z}$	0.60
Concatenated model	$[\tilde{\mathbf{Y}}, \ln(\mathbf{Y}\tilde{\mathbf{W}})]$	0.70

The machine computes the similarity between two images as the correlation coefficient between the two representations.

Spearman correlation coefficient between human and machine-estimated visual semantic similarity.

Restricting the computation of similarity to the top class (the class with the highest probability) performed poorly.

Scatterplot of concatenated model



Summary

Small dataset established for evaluation of methods for visual semantic similarity.

We have used no learning as we have not much data.

Possibly a slight improvement in performance of visual semantic similarity estimation by using the Wikidata knowledge graph.

References

Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., and Ruppin, E. (2002). Placing search in context: the concept revisited. *ACM Transactions on Information Systems*, 20:116–131. DOI: 10.1145/503104.503110.

Hill, F., Reichart, R., and Korhonen, A. (2014). SimLex-999: Evaluating Semantic Models with (Genuine) Similarity Estimation. *Computational Linguistics*, 41:665–695. DOI: 10.1162/COLI_A_00237.

Nielsen, F. Å. (2018). Linking ImageNet WordNet Synsets with Wikidata. *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France*, pages 1809–1814. DOI: 10.1145/3184558.3191645.

Copyright and license