Learning Lexical Embeddings with Syntactic and Lexicographic Knowledge

Tong Wang[1], Abdel-rahman Mohamed[2], and Graeme Hirst[1]
1. University of Toronto 2. Microsoft Research

1. Lexical Embeddings

- Real-valued vector representations of words
- Vectors geometrically positioned per *The Distributional Hypothesis*[1]

Problem:

Window-based co-occurrence suffers from ...

The strongest rain ever recorded in India shut down the financial hub of Mumbai, snapped communication lines, ...

False positives
- *in India shut*
- *India shut down*
- *down the financial*

False negatives
- *rain ... shut*
- *rain ... snapped*

2. Proposed Solutions

Use syntactic association instead of window-based co-occurrence.

- *prep_in(recorded, India)*
- *nsubj(shut, rain)*
- *nsubj(snap, rain)*

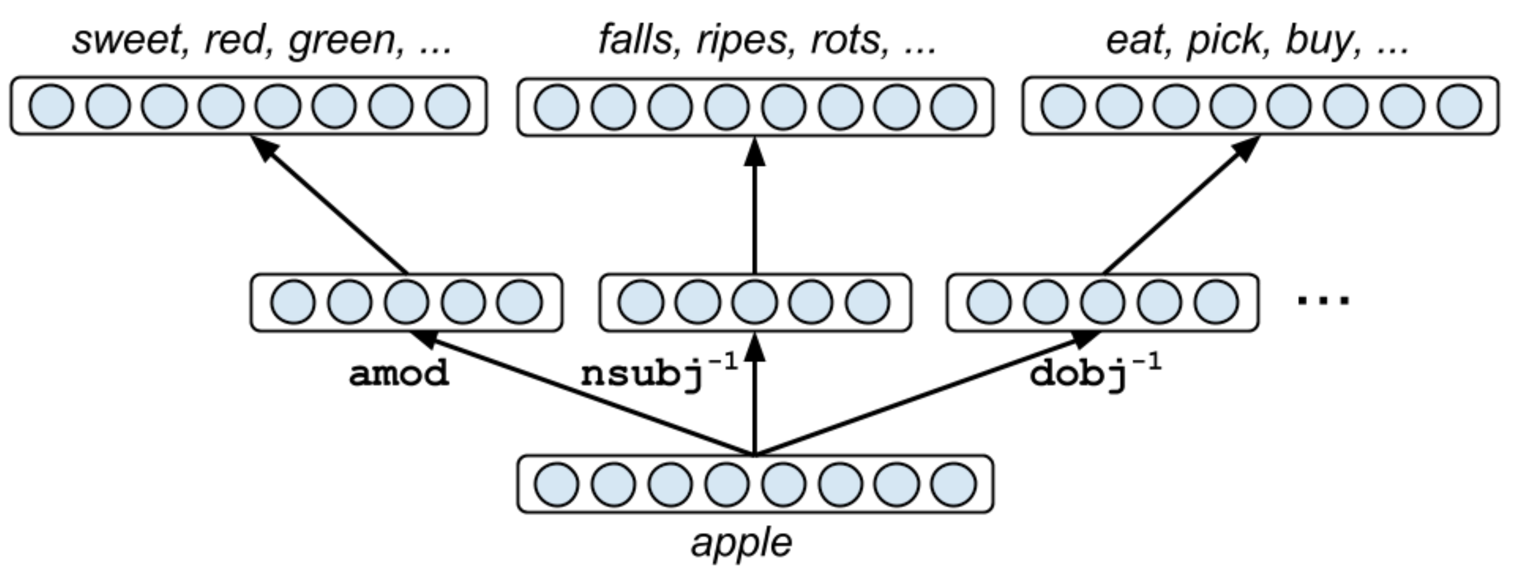
Use lexicographic resources (e.g., dictionary definitions) for lexical association.

The defining relation:
- *apple* ⇐ *fruit*
- *apple* ⇐ *rosaceous*

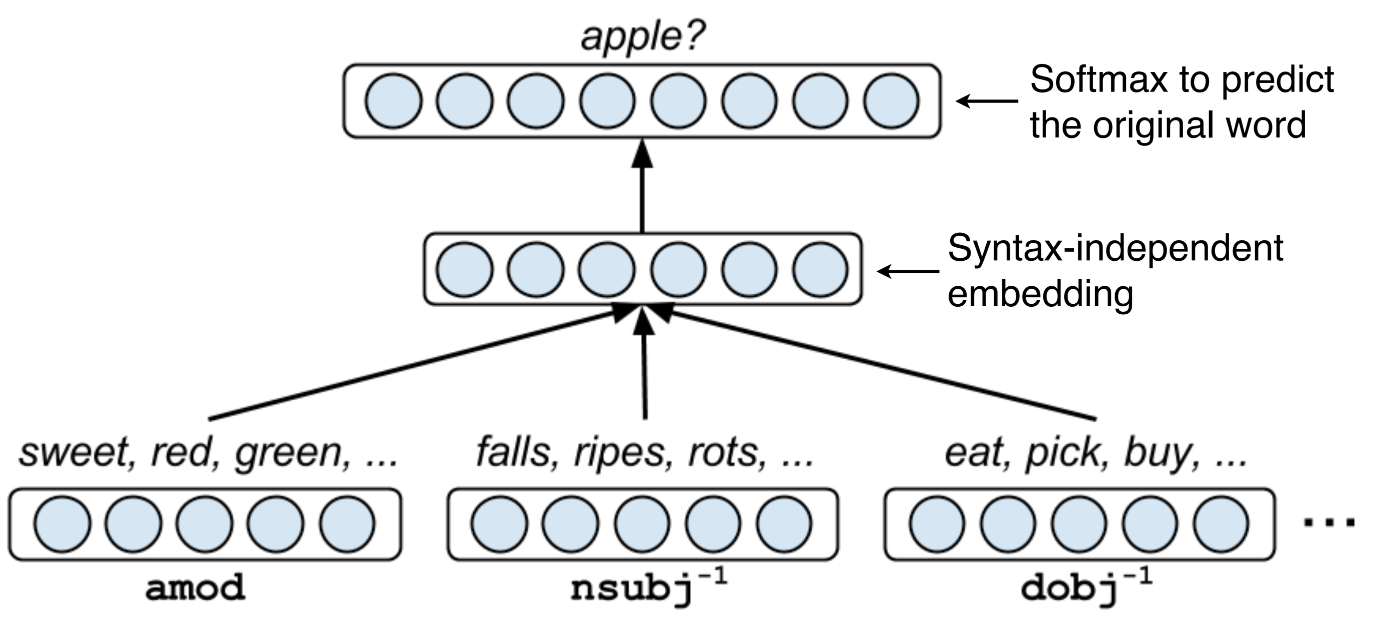
Or its inverse:
- *apple* ⇒ *cider*
- *apple* ⇒ *pippin*

3. Factorizing Syntactic Relations

- Syntax-dependent model



- Syntax-independent Model



4. Evaluation

Lexical similarity

- Performance measured by correlation between:
  (a) human judgment of similarity...

car – automobile: 3.92 *rooster – voyage:* 0.08
gem – jewel: 3.84 *noon – string:* 0.08

... and (b) system similarity score (i.e., *cosine* similarity between embeddings)

- Datasets:
  - *MC*[2], *RG*[3] – nouns; semantic
  - *FG*[4] (or *wordsim353*) – nouns; distributional
  - *SL*[5] (or *SimLex999*) – nouns (SL_n), adjectives (SL_a), and verbs (SL_v); strictly semantic

5. Results and Conclusions

1. Factorizing syntactic relations notably improves lexical embedding learning.

Model	Datasets					
	MC	*RG*	*FG*	SL_n	SL_v	SL_a
amod	**.766**	**.798**	.572	**.566**	.154	.466
amod^{-1}	.272	.296	.220	.218	.248	**.602**
nsubj	.442	.350	.376	.388	**.392**	.464
nn	.596	.620	.514	.486	.130	.068
Baselines						
DEP	.640	.670	.510	.400	.240	.350
w2v	.656	.618	**.600**	.382	.237	.560
GloVe	.609	.629	.546	.346	.142	.517

2. Combining pre-trained syntax-dependent embeddings alleviates sparsity issues on smaller dataset.

Rel. Dep. #1	.512	.486	.380	.354	.222	.394
Rel. Dep. #2	.390	.380	.360	.304	.206	.236
Rel. Indep.	**.570**	.550	.392	**.360**	**.238**	.338
Baselines						
DEP	.530	**.558**	.506	.346	.138	.412
w2v	.563	.491	**.562**	.287	.065	.379
GloVe	.306	.368	.308	.132	−.007	.254

3. Lexicographic knowledge from monolingual dictionaries helps produce high-quality lexical embeddings.

def	.640	.626	.378	.332	.320	.306
def^{-1}	.740	.626	.436	.366	.332	.376
Combined	**.754**	**.722**	.530	**.410**	**.356**	.412
w2v	.656	.618	**.600**	.382	.237	**.560**

References

[1] Zellig Harris. Distributional structure. *Word*, 10 (23):146–162, 1954.

[2] George Miller and Walter Charles. Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1):1–28, 1991.

[3] Herbert Rubenstein and John Goodenough. Contextual correlates of synonymy. *Communications of the ACM*, 8(10):627–633, 1965.

[4] Lev Finkelstein, Evgeniy Gabrilovich, Yossi Matias, Ehud Rivlin, Zach Solan, Gadi Wolfman, and Eytan Ruppin. Placing search in context: The concept revisited. In *Proceedings of the 10th International Conference on World Wide Web*, pages 406–414. ACM, 2001.

[5] Felix Hill, Roi Reichart, and Anna Korhonen. Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *arXiv preprint arXiv:1408.3456*, 2014.

Presented at *The 53rd Annual Meeting of the Association for Computational Linguistics*, Beijing, China, July 2015

UNIVERSITY OF TORONTO

Financially supported by: NSERC CRSNG