One Representation per Word

Does it make Sense for Composition?

Thomas Kober, Julie Weeds, John Wilkie, Jeremy Reffin & David Weir

TAG Laboratory
University of Sussex

Outline

- Distributional composition as contextualisation
- Phrase similarity
- Word sense discrimination
- Analysis
- Summary

Outline

- **Distributional composition as contextualisation**
- Phrase similarity
- Word sense discrimination
- Analysis
- Summary

Distributional composition as contextualisation

- Refers to the specific meaning of a lexeme in context
- A polysemous lexeme can be sense discriminated based on the context in which it occurs
    - e.g. **river** bank vs. bank **account**
- Distributional composition can act as a sense discriminator
- No need for *a priori* disambiguation

Experimental Setup

- Evaluation on a standard phrase similarity task (Mitchell & Lapata, 2010), and a novel word sense discrimination task

- Comparing off-the-shelf word embeddings

  - word2vec (Mikolov et al., 2013)

  - dep2vec (Levy & Goldberg, 2014)

  - SensEmbed (Iacobacci et al., 2015)

Outline

- Distributional composition as contextualisation
- Phrase similarity
- Word sense discrimination
- Analysis
- Summary

Outline

- Distributional composition as contextualisation
- **Phrase similarity**
- Word sense discrimination
- Analysis
- Summary

Phrase Similarity (1)

- Compare AN, NN, VO phrase pairs to human similarity judgements (Mitchell & Lapata, 2010)
- High-frequency lexemes with some degree of polysemy (e.g. “company” or “state”)
- Composition by pointwise addition for all models
- Closest sense strategy for SensEmbed

Phrase Similarity (2)

Model	AN	NN	VO	Average
word2vec	0.47	<u>0.46</u>	<u>0.45</u>	<u>0.46</u>
dep2vec	<u>0.48</u>	<u>0.46</u>	<u>0.45</u>	<u>0.46</u>
SenseEmbed:max	0.39	0.39	0.32	0.37

Phrase Similarity (3)

- Pointwise addition with word2vec and dep2vec works very well

- Closest sense strategy (max) tends to overestimate similarities, problematic for dissimilar phrase pairs

- Performance of SensEmbed can be improved by using the mean instead of max, but it is still worse than word2vec and dep2vec

- Distributional composition as contextualisation

Outline

- Distributional composition as contextualisation
- Phrase similarity
- Word sense discrimination
- Analysis
- Summary

Outline

- Distributional composition as contextualisation
- Phrase similarity
- **Word sense discrimination**
- Analysis
- Summary

Word Sense Discrimination (1)

- Example:
    - *Target:* The **head** of state was under pressure.
    - *Option 1:* He got hit right in the **head**.
    - *Option 2:* I was pulled in by the **head** of HR.
- Goal is to identify the Option 2 as expressing the same sense of **head** as the given target sentence
- Different task setups of varying difficulty (2-5 senses)
- Data for 3 different parts of speech (Adjectives, Nouns, Verbs)
- Accuracy as evaluation metric

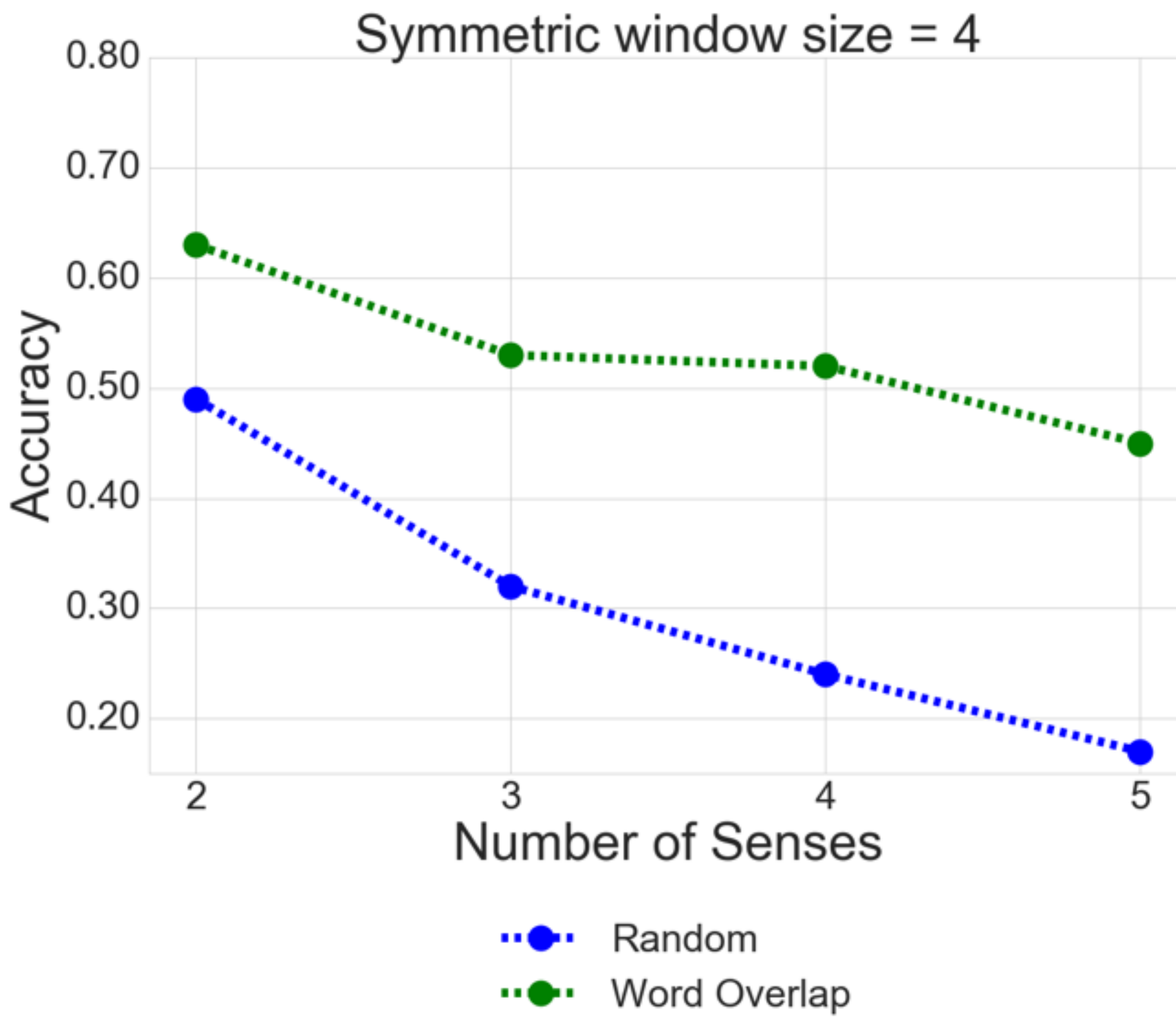
Word Sense Discrimination (2)

	2 senses (#dev/#test)	3 senses	4 senses	5 senses
Adjectives	66/209	47/170	37/137	28/115
Nouns	170/618	125/499	100/412	74/345
Verbs	127/438	71/354	72/295	56/256
Total	363/1265	263/1023	209/844	164/716

Word Sense Discrimination (3)

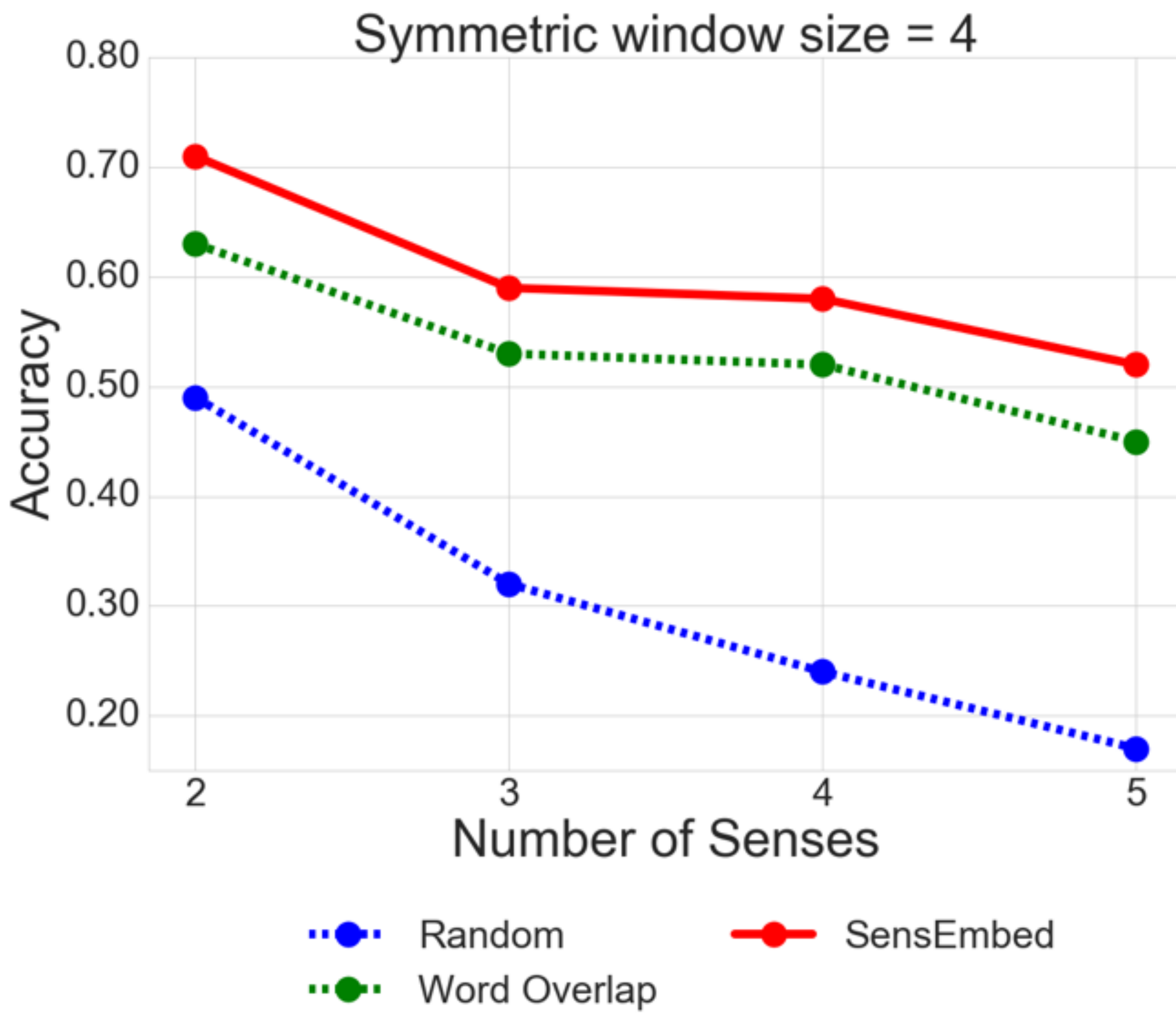
- Identify that two polysemous lexemes express the same sense
- No sense induction step
- No sense labelling step
- Example sentences taken from english dictionaries
- Sentential context provides enough information to discriminate the expressed sense of a polysemous lexeme
- Good testbed for evaluating contextualisation

Word Sense Discrimination (4)



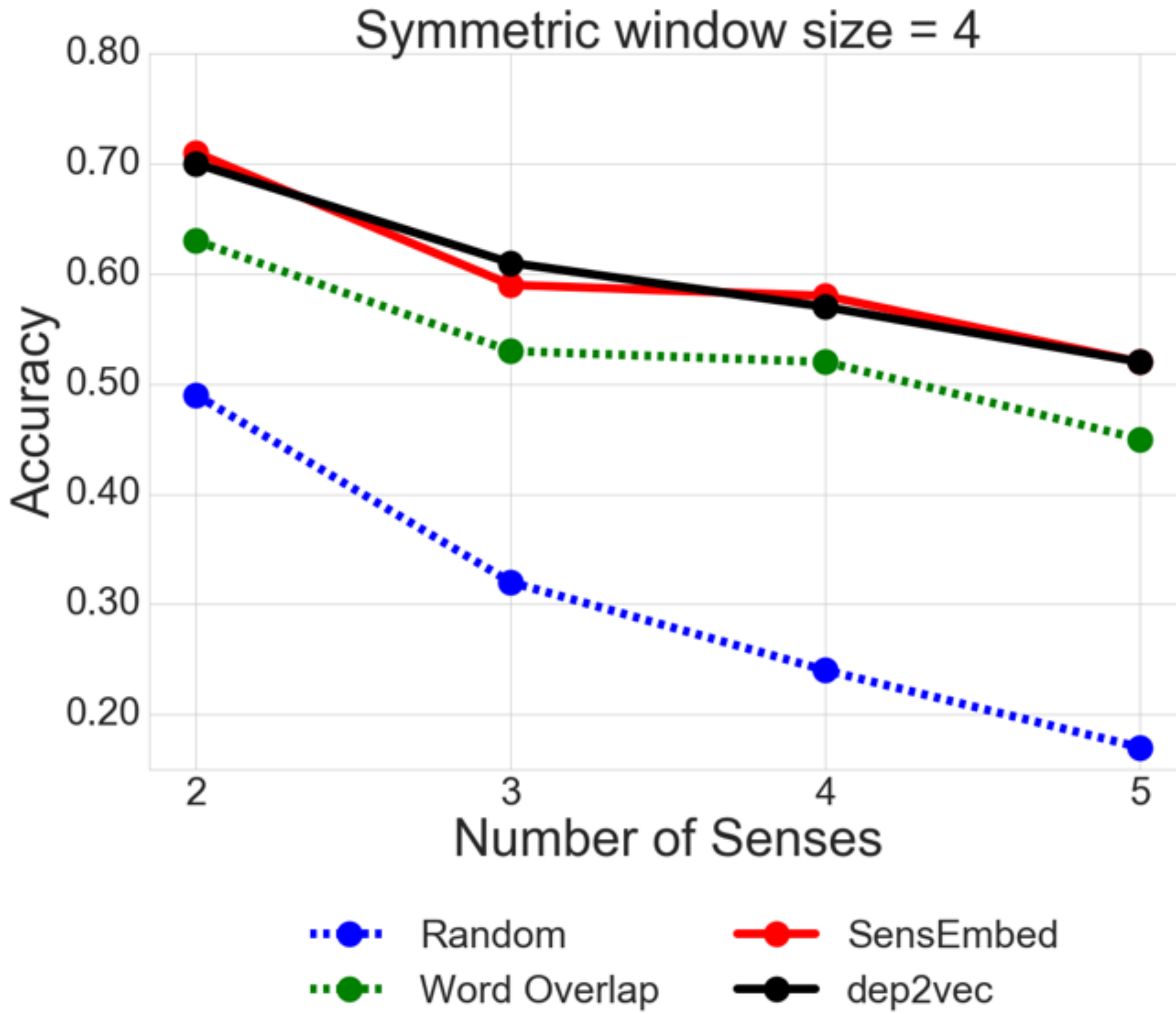
All differences significant at the $p < 0.01$ level

Word Sense Discrimination (4)



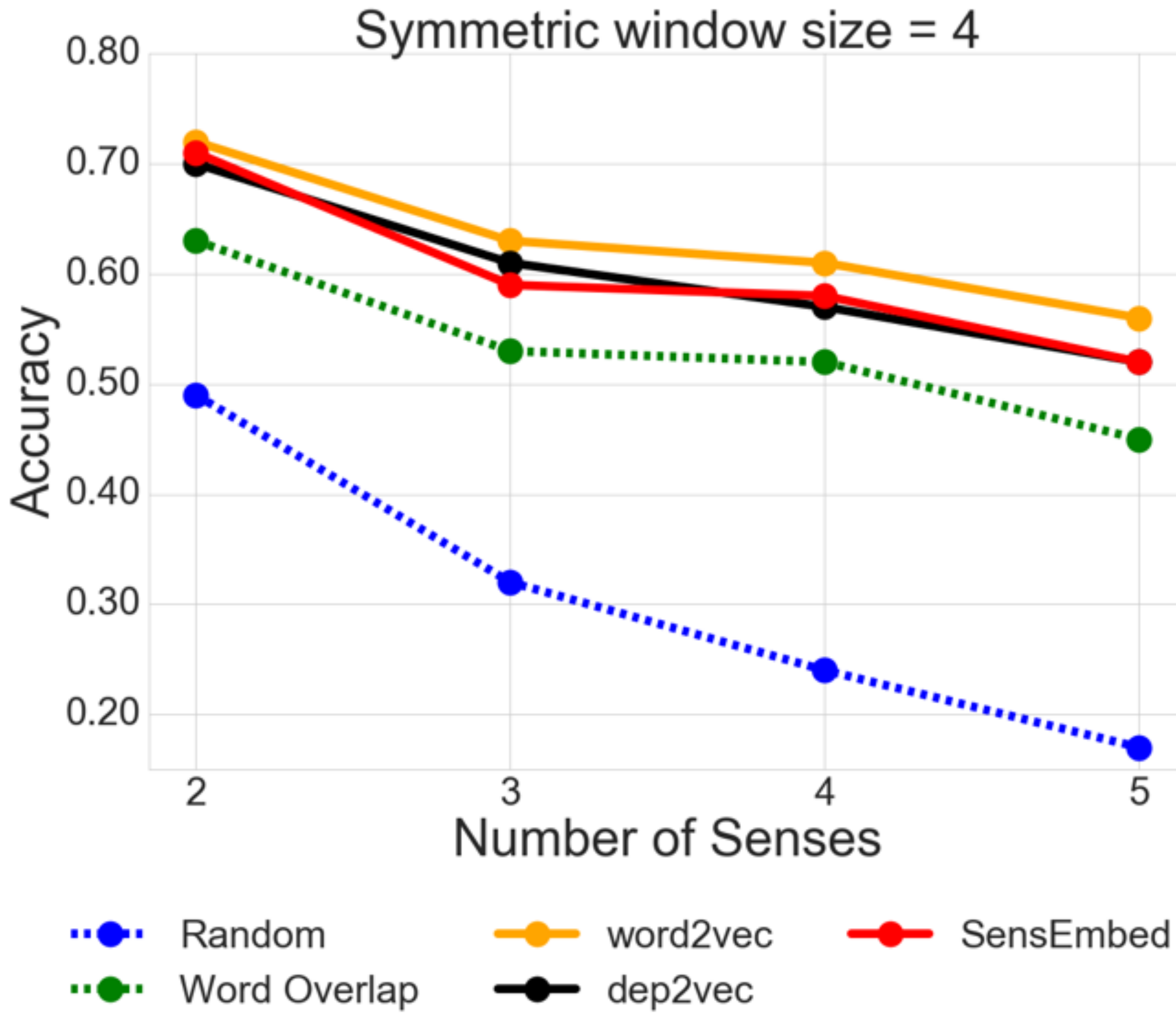
All differences significant at the p < 0.01 level

Word Sense Discrimination (4)



Differences significant w.r.t. to Word Overlap at the $p < 0.01$ level for 2,3 & 5 senses; significant at $p < 0.05$ for 4 senses

Word Sense Discrimination (4)



Differences significant w.r.t. to Word Overlap at the p < 0.01 level

Word Sense Discrimination (5)

- Pointwise addition with word2vec and dep2vec works very well
- No benefit from sense-level representations
- Distributional composition as contextualisation

Outline

- Distributional composition as contextualisation
- Phrase similarity
- Word sense discrimination
- Analysis
- Summary

Outline

- Distributional composition as contextualisation
- Phrase similarity
- Word sense discrimination
- **Analysis**
- Summary

Distributional composition as contextualisation (1)

- Pointwise addition in neural word embeddings approximates a feature intersection (Tian et al., 2015)
- Different senses of a polysemous lexeme reside in a linear substructure of the embedding (Arora et al., 2016)
- Sense specific meaning recoverable by composition

Distributional composition as contextualisation (2)

Phrase	word2vec	SensEmbed
desert <u>rock</u>	rock, desert, rocks, desolate expanse, arid desert	desert, rock, the desert, deserts, badlands
<u>rock</u> band	rock, band, rockers, bands, indie rock	band, rock, group, the band, rock group
river <u>bank</u>	bank, river, creek, lake, rivers	bank, river, stream, creek, river basin
<u>bank</u> account	account, bank, accounts, banks, citibank	bank, banks, the bank, pko bank polski, handlowy

Outline

- Distributional composition as contextualisation
- Phrase similarity
- Word sense discrimination
- Analysis
- Summary

Outline

- Distributional composition as contextualisation
- Phrase similarity
- Word sense discrimination
- Analysis
- **Summary**

Summary

- Distributional composition is contextualisation
- Composition acts as a sense discriminator and is able to recover sense specific information remarkably well
- Open question how to best leverage sense embeddings for distributional composition

Q & A

- Contact: t.kober@sussex.ac.uk
- Task: https://github.com/tttthomasssss/sense2017

References

- Sanjeev Arora, Yuanzhi Li, Yingyu Liang, Tengyu Ma, and Andrej Risteski. 2016. *Linear algebraic structure of word senses, with applications to polysemy*. arXiv: https://arxiv.org/abs/1601.03764.

- Ignacio Iacobacci, Mohammad Taher Pilehvar, and Roberto Navigli. 2015. *Sensembed: Learning sense embeddings for word and relational similarity*. In Proceedings of ACL, pages 95–105, Beijing, China, July. Association for Computational Linguistics.

- Omer Levy and Yoav Goldberg. 2014. *Dependency-based word embeddings*. In Proceedings of ACL, pages 302–308, Baltimore, Maryland, June. Association for Computational Linguistics.

- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. *Distributed representations of words and phrases and their compositionality*. In Proceedings of NIPS, pages 3111–3119. Curran Associates, Inc.

- Jeff Mitchell and Mirella Lapata. 2010. *Composition in distributional models of semantics*. Cognitive Science, 34(8):1388–1429.

- Ran Tian, Naoaki Okazaki, and Kentaro Inui. 2015. *The mechanism of additive composition*. arXiv: https://arxiv.org/abs/1511.08407.