University of Stuttgart
Germany

Second-order Co-occurrence Sensitivity of Skip-Gram with Negative Sampling

Dominik Schlechtweg Cennet Oguz Sabine Schulte im Walde

Second-order Co-occurrence

first order

(1a) As far as the Soviet Communist **Party** was concerned . . .

second order

(1b) It is well-known that Communist authorities hated rock culture . . .

(2a) . . . this is the approach taken by the British **Government**.

(2b) . . . rather than risking deportation to British authorities.

[Schütze 1998]

Research Question

- **Does SGNS capture second-order co-occurrence information?**

→ Yes – it is more sensitive to second- than first-order context overlap

- SGNS is similar to Truncated SVD in terms of capturing second-order co-occurrence structure [cf. Levy & Goldberg 2014]
- Capturing higher-order co-occurrence structure may explain superior performance of SGNS and SVD over PPMI

Exp. 1: Simulating Context Overlap

order	1ST	2ND	NONE
C1	a **c**	a c	a c
	a **d**	a d	a d
	b **c**	b e	b e
	b **d**	b f	b f
C2	c u	c **u**	c u
	c v	c **v**	c v
	d w	d **u**	d w
	d x	d **v**	d x

Table: Artificial co-occurrence pairs with context overlap in different orders of co-occurrence.

Hypothesis: SGNS makes vectors of words from the 2ND group more similar than vectors of words from the NONE group (although both groups have no first-order context overlap)

Side Hypothesis: SVD shows similar behavior as SGNS

Exp. 1: Results

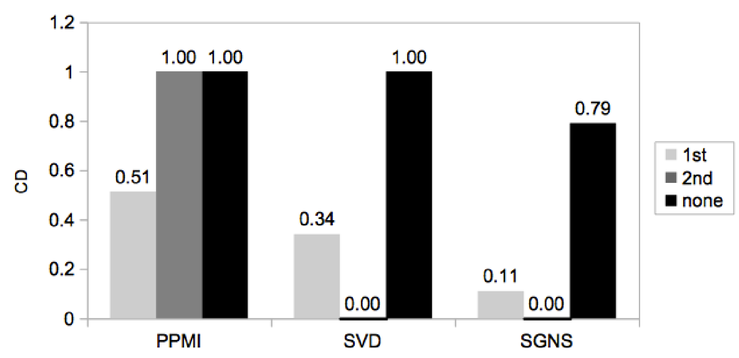


Figure: Results of simulation experiment. Values give average cosine distances across target words with different levels of context overlap.

Exp. 2: Propagating Second-order Context

- create very small corpus (10M tokens from ukWaC)

⇓

- extract first-and second-order word-context pairs

⇓

- add second- to first-order pairs for low-frequency words

⇓

- compare performance (WordSim353) on first-order vs. mixed training pairs

Hypothesis: Additional second-order information impacts PPMI representations positively and stronger than SVD and SGNS (because the latter already capture second-order information)

Exp. 2: Results

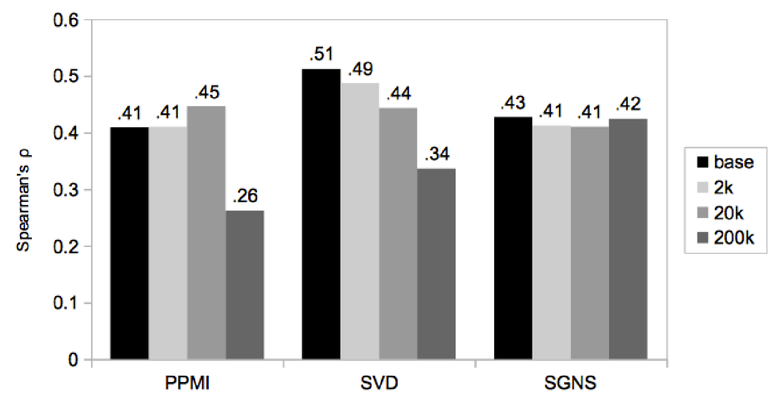
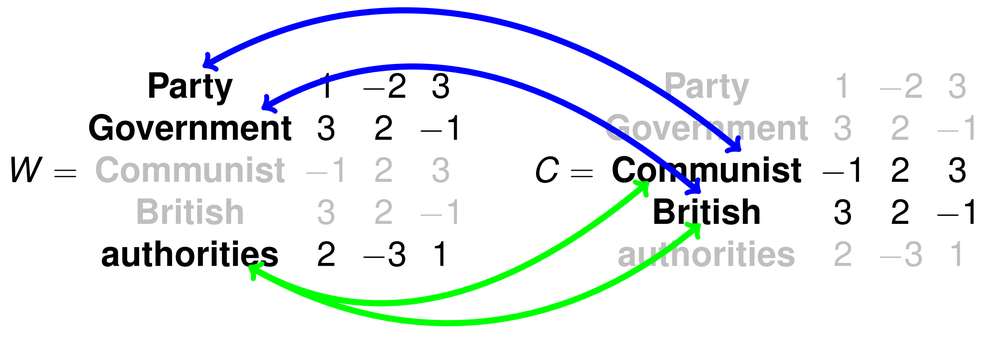


Figure: Results of experiment 2. Values give correlation (Spearman's ρ) of model predictions with human similarity judgments.

Explanation: Transitivity of SGNS



Transitivity of SGNS: The representation of context words as continuous objects (vectors) establishes transitivity of the similarity relation between word and context vectors. This enables SGNS to capture second-order co-occurrence information.

Open Questions

- Do other modern embeddings (GloVe, ELMo, BERT) capture second-order co-occurrence?

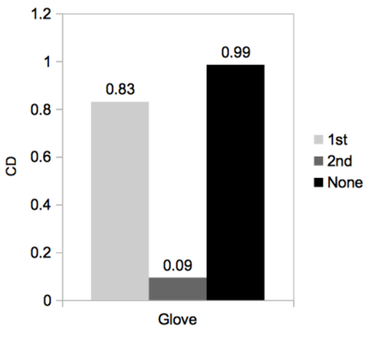


Figure: Results of simulation experiment with GloVe embeddings. [Pennington et al. 2014]

- How does second-order information relate to performance?

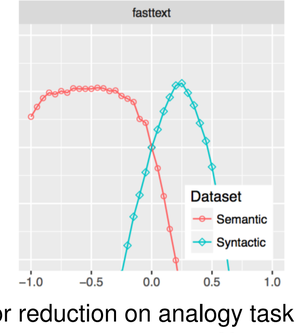


Figure: Error reduction on analogy task for levels of higher-order co-occurrence [Artetxe et al. 2018]

Acknowledgments

The first author was supported by the Konrad Adenauer Foundation and the CRETA center funded by the German Ministry for Education and Research (BMBF) during the conduct of this study. We thank Agnieszka Falenska, Sascha Schlechtweg and the anonymous reviewers for their valuable comments.

References

Artetxe, M., Labaka, G., Lopez-Gazpio, I., & Agirre, E. (2018). Uncovering divergent linguistic information in word embeddings with lessons for intrinsic and extrinsic evaluation. In *Proceedings of the 22nd conference on computational natural language learning* (pp. 282–291). Brussels, Belgium: Association for Computational Linguistics.

Levy, O., & Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. In *Proceedings of the 27th international conference on neural information processing systems - volume 2* (pp. 2177–2185). Cambridge, MA, USA: MIT Press.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing* (pp. 1532–1543). Doha, Qatar.

Schütze, H. (1998, March). Automatic word sense discrimination. *Comput. Linguist.*, *24*(1), 97–123.