

Li, Dandan; Summers-Stay, Douglas

Dual embeddings and metrics for word and relational similarity. (English) Zbl 1444.68268
Ann. Math. Artif. Intell. 88, No. 5-6, 533-547 (2020).

Summary: Word embedding models excel in measuring word similarity and completing analogies. Word embeddings based on different notions of context trade off strengths in one area for weaknesses in another. Linear bag-of-words contexts, such as in word2vec, can capture topical similarity better, while dependency-based word embeddings better encode functional similarity. By combining these two word embeddings using different metrics, we show how the best aspects of both approaches can be captured. We show state-of-the-art performance on standard word and relational similarity benchmarks.

MSC:

68T50 Natural language processing
68T30 Knowledge representation

Keywords:

word embeddings; dual embeddings; word similarity; relational similarity

Software:

BERT; Simlex-999; StanfordCoreNLP; word2vec

Full Text: [DOI](#)

References:

- [1] Agirre, E., Alfonseca, E., Hall, K., Kravalova, J., Pasca, M., Soroa, A.: A study on similarity and relatedness using distributional and wordnet-based approaches. In: NAACL '09 Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 19-27. Boulder (2009)
- [2] Alvarez, M.A., Lim, S.J.: A graph modeling of semantic similarity between words. In: Proceedings of the Conference on Semantic Computing, pp. 355-362 (2007)
- [3] Bengio, Y.; Ducharme, R.; Vincent, P.; Jauvin, C., A neural probabilistic language model, *J. Mach. Learn. Res.*, 3, 1137-13155 (2003) · [Zbl 1061.68157](#)
- [4] Bici, E., Yuret, D.: Clustering word pairs to answer analogy questions. In: Proceedings of the Fifteenth Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN 2006). Akyaka (2006)
- [5] Boteanu, A., Chernova, S.: Solving and explaining analogy questions using semantic networks. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, pp. 1460-1466 (2015)
- [6] Bruni, E.; Tran, N.K.; Baroni, M., Multimodal distributional semantics, *J. Artif. Intell. Res.*, 49, 1, 1-47 (2014) · [Zbl 1361.68287](#)
- [7] Collobert, R., Weston, J.: A unified architecture for natural language processing: Deep neural networks with multitask learning. In: Proceedings of the 25 th International Conference on Machine Learning, pp. 160-167 (2008)
- [8] Deerwester, S.; Dumais, S.T.; Furnas, G.W.; Landauer, T.K.; Harshman, R., Indexing by latent semantics analysis, *J. Assoc. Inf. Sci. Technol.*, 41, 6, 391-407 (1990)
- [9] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: arXiv:1810.04805 (2018)
- [10] Faruqi, M., Dodge, J., Jauhar, S.K., Dyer, C., Hovy, E., Smith, N.A.: Retrofitting word vectors to semantic lexicons. In: The 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HL 2015), Denver (2015)
- [11] Faruqi, M., Tsvetkov, Y., Rastogi, P., Dyer, C.: Problems with evaluation of word embeddings using word similarity tasks. arXiv:160502276 (2016)
- [12] Finkelstein, L.; Evgeniy, G.; Yossi, M.; Ehud, R.; Zach, S.; Gadi, W.; Eytan, R., Placing search in context: The concept revisited, *ACM Trans. Inf. Syst.*, 20, 1, 116-131 (2002)
- [13] Firth, J.R.: A synopsis of linguistic theory 1930-1955. In: Studies in Linguistic Analysis, pp 1-32. Blackwell, Oxford (1957)
- [14] Gatti, L., Özbal, G., Stock, O., Strapparava, C.: To sing like a mockingbird. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, pp. 298-304 (2017)

- [15] Halawi, G., Dror, G., Gabrilovich, E., Koren, Y.: Large-scale learning of word relatedness with constraints. In: Proceedings of The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1406-1414 (2012)
- [16] Han, L., Kashyap, A.L., Finin, T., Mayfield, J., Weese, J.: Umbc ebiquty-core: Semantic textual similarity systems. In: Proceedings of the Second Joint Conference on Lexical and Computational Semantics. Association for Computational Linguistics (2013)
- [17] Harris, ZS, Distributional structure, *Word*, 10, 23, 146-162 (1954)
- [18] Herdagdelen, A., Baroni, M.: Bagpack: A general framework to represent semantic relations. In: Proceedings of the EACL 2009 Geometrical Models for Natural Language Semantics (GEMS) Workshop, pp. 33-40 (2009)
- [19] Hill, F., Reichart, R., Korhonen, A.: Simlex-999: Evaluating semantic models with (genuine) similarity estimation. In: arXiv:1408.3456, pp. 1-23 (2014)
- [20] Hughes, T., Ramage, D.: Lexical semantic relatedness with random graph walks. In: Proceedings of EMNLP-CoNLL-2007, pp. 581-589 (2007)
- [21] Iacobacci, I., Pilehvar, M.T., Navigli, R.: Sensembed: Learning sense embeddings for word and relational similarity. In: ACL-IJCNLP 2015: The 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, pp 95-105, Beijing (2015)
- [22] Jurgens, D.A., Mohammad, S.M., Turney, P.D.: Semeval-2012 task 2: Measuring degrees of relational similarity. In: *SEM 2012: The First Joint Conference on Lexical and Computational Semantics, pp. 356-364. Montreal (2012)
- [23] Levy, O., Goldberg, Y.: Dependency-based word embeddings. In: Proceedings of the 52nd Annual Meeting of the Associations for Computational Linguistics (short papers), pp. 302-308 (2014)
- [24] Levy, O., Goldberg, Y.: Linguistic regularities in sparse and explicit word representations. In: Proceedings of the 18th Conference on Computational Natural Language Learning, pp. 171-180 (2014)
- [25] Li, D., Summers-Stay, D.: Dual embeddings and metrics for relational similarity. In: Proceedings of the 12th International Conference on Computational Semantics — Short papers, pp. 1-7 (2017)
- [26] Luong, M.T., Socher, R., Manning, C.D.: Better word representations with recursive neural networks for morphology. In: Proceedings of the 17th Conference on Computational Natural Language Learning, pp. 1-7 (2013)
- [27] Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D.: The Stanford corenlp natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60 (2014)
- [28] Melamud, O., McClosky, D., Patwardhan, S., Bansal, M.: The role of context types and dimensionality in learning word embeddings. In: Proceedings of NAACL-HLT, pp. 1030-1040 (2016)
- [29] Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv:1301.3781 (2013)
- [30] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Proceedings of the 26th International Conference on Neural Information Processing Systems, pp. 3111-3119. Nevada (2013)
- [31] Mikolov, T., Yih, W.T., Zweig, G.: Linguistic regularities in continuous space word representations. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 746-751. Atlanta (2013)
- [32] Miller, GA; Charles, WG, Contextual correlates of semantic similarity, *Ling. Cogn. Process.*, 6, 1, 1-28 (1991)
- [33] Parker, R., Graff, D., Kong, J., Chen, K., Maeda, K.: English gigaword, 5th edn. In: Linguistic Data Consortium, LDC2011T07 (2011)
- [34] Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations. In: Proceedings of NAACL-HLT 2018, pp. 2227-2237 (2018)
- [35] Quesada, J., Kintsch, W., Mangalath, P.: Analogy-making as prediction using relational information and lsa vectors. In: Proceedings of the 26th Annual Meeting of the Cognitive Science Society, p. 1623. Austin (2004)
- [36] Radinsky, K., Agichtein, E., Gabrilovich, E., Markovitch, S.: A word at a time: Computing word relatedness using temporal semantic analysis. In: Proceedings of the 20th International Conference on World Wide Web, pp. 337-346 (2011)
- [37] Rubenstein, H.; Goodenough, JB, Contextual correlates of synonymy, *Commun. ACM*, 8, 10, 627-633 (1965)
- [38] Santus, E., Chersoni, E., Lenci, A., Huang, C.R., Blache, P.: Testing apsyn against vector cosine on similarity estimation. In: Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation, pp. 229-238 (2016)
- [39] Santus, E., Chiu, T.S., Lu, Q., Lenci, A., Huang, C.R.: What a nerd! Beating students and vector cosine in the esl and toefl datasets. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (2016)
- [40] Santus, E., Wang, H., Chersoni, E., Zhang, Y.: A rank-based similarity metric for word embeddings. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, pp. 552-557 (2018)
- [41] Strapparava, C., Valitutti, A., Stock, O.: Dances with words. In: Proceedings of the Twentieth International Joint Conference on Artificial Intelligence, pp. 1719-1724 (2007)
- [42] Turney, P.D.: Expressing implicit semantic relations without supervision. In: Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (Coling/ACL-06), pp 313-320. Sydney (2006)
- [43] Turney, PD, Similarity of semantic relations, *Comput. Ling.*, 32, 3, 379-416 (2006) · [Zbl 1234.68434](#)

- [44] Turney, PD, The latent relation mapping engine: Algorithm and experiments, *J. Artif. Intell. Res.*, 33, 615-655 (2008) · [Zbl 1182.68324](#)
- [45] Turney, P.D.: A uniform approach to analogies, synonyms, antonyms, and associations. In: *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pp 905-912, Manchester (2008)
- [46] Turney, PD, Domain and function: A dual-space model of semantic relations and compositions, *J. Artif. Intell. Res. (JAIR)*, 44, 533-585 (2012) · [Zbl 1280.68273](#)
- [47] Turney, PD, Distributional semantics beyond words: Supervised learning of analogy and paraphrase, *Trans. Assoc. Comput. Ling. (TACL)*, 1, 353-366 (2013)
- [48] Turney, PD; Littman, ML, Corpus-based learning of analogies and semantic relations, *Mach. Learn.*, 60, 1-3, 251-278 (2005)
- [49] Veale, T.: Wordnet sits the sat: A knowledge-based approach to lexical analogy. In: *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI 2004)*, pp 606-612, Valencia (2004)
- [50] Yang, D., Powers, D.M.: Measuring semantic similarity in the taxonomy of wordnet. In: *Proceedings of the Twenty-eighth Australasian Conference on Computer Science*, pp. 315-322 (2005)
- [51] Yang, D., Powers, D.M.W.: Verb similarity on the taxonomy of wordnet. In: *Proceedings of the 3rd International WordNet Conference* (2006)
- [52] Zesch, T., Muller, C., Gurevych, I.: Using wiktionary for computing semantic relatedness. In: *Proceedings of the 23rd National Conference on Artificial Intelligence*, pp. 861-866 (2008)
- [53] Zhila, A., Yih, W.T., Meek, C.: Combining heterogeneous models for measuring relational similarity. In: *Proceedings of NAACL-HLT*, pp. 1000-1009 (2013)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.