



Word embedding

In natural language processing, a **word embedding** is a representation of a word. The embedding is used in text analysis. Typically, the representation is a real-valued vector that encodes the meaning of the word in such a way that the words that are closer in the vector space are expected to be similar in meaning.^[1] Word embeddings can be obtained using language modeling and feature learning techniques, where words or phrases from the vocabulary are mapped to vectors of real numbers.

Methods to generate this mapping include neural networks,^[2] dimensionality reduction on the word co-occurrence matrix,^{[3][4][5]} probabilistic models,^[6] explainable knowledge base method,^[7] and explicit representation in terms of the context in which words appear.^[8]

Word and phrase embeddings, when used as the underlying input representation, have been shown to boost the performance in NLP tasks such as syntactic parsing^[9] and sentiment analysis.^[10]

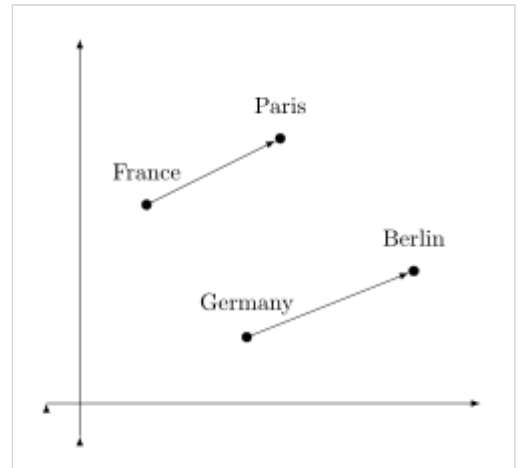


Illustration of word embedding. Each word is a point in some space. The word embedding enables to perform semantic operator like obtaining the capital of a given country.

Development and history of the approach

In distributional semantics, a quantitative methodological approach for understanding meaning in observed language, word embeddings or semantic feature space models have been used as a knowledge representation for some time.^[11] Such models aim to quantify and categorize semantic similarities between linguistic items based on their distributional properties in large samples of language data. The underlying idea that "a word is characterized by the company it keeps" was proposed in a 1957 article by John Rupert Firth,^[12] but also has roots in the contemporaneous work on search systems^[13] and in cognitive psychology.^[14]

The notion of a semantic space with lexical items (words or multi-word terms) represented as vectors or embeddings is based on the computational challenges of capturing distributional characteristics and using them for practical application to measure similarity between words, phrases, or entire documents. The first generation of semantic space models is the vector space model for information retrieval.^{[15][16][17]} Such vector space models for words and their distributional data implemented in their simplest form results in a very sparse vector space of high dimensionality (cf. curse of dimensionality). Reducing the number of dimensions using linear algebraic methods such as singular value decomposition then led to the introduction of latent semantic analysis in the late 1980s and the random indexing approach for collecting word co-occurrence contexts.^{[18][19][20][21]} In 2000, Bengio et al. provided in a series of papers titled "Neural probabilistic language models" to reduce the high dimensionality of word representations in contexts by "learning a distributed representation for words".^{[22][23][24]}

A study published in NeurIPS (NIPS) 2002 introduced the use of both word and document embeddings applying the method of kernel CCA to bilingual (and multi-lingual) corpora, also providing an early example of self-supervised learning of word embeddings.^[25]

Word embeddings come in two different styles, one in which words are expressed as vectors of co-occurring words, and another in which words are expressed as vectors of linguistic contexts in which the words occur; these different styles are studied in Lavelli et al., 2004.^[26] Roweis and Saul published in Science how to use "locally linear embedding" (LLE) to discover representations of high dimensional data structures.^[27] Most new word embedding techniques after about 2005 rely on a neural network architecture instead of more probabilistic and algebraic models, after foundational work done by Yoshua Bengio^[28] and colleagues.^{[29][30]}

The approach has been adopted by many research groups after theoretical advances in 2010 had been made on the quality of vectors and the training speed of the model, as well as after hardware advances allowed for a broader parameter space to be explored profitably. In 2013, a team at Google led by Tomas Mikolov created word2vec, a word embedding toolkit that can train vector space models faster than previous approaches. The word2vec approach has been widely used in experimentation and was instrumental in raising interest for word embeddings as a technology, moving the research strand out of specialised research into broader experimentation and eventually paving the way for practical application.^[31]

Polysemy and homonymy

Historically, one of the main limitations of static word embeddings or word vector space models is that words with multiple meanings are conflated into a single representation (a single vector in the semantic space). In other words, polysemy and homonymy are not handled properly. For example, in the sentence "The club I tried yesterday was great!", it is not clear if the term *club* is related to the word sense of a club sandwich, clubhouse, golf club, or any other sense that *club* might have. The necessity to accommodate multiple meanings per word in different vectors (multi-sense embeddings) is the motivation for several contributions in NLP to split single-sense embeddings into multi-sense ones.^{[32][33]}

Most approaches that produce multi-sense embeddings can be divided into two main categories for their word sense representation, i.e., unsupervised and knowledge-based.^[34] Based on word2vec skip-gram, Multi-Sense Skip-Gram (MSSG)^[35] performs word-sense discrimination and embedding simultaneously, improving its training time, while assuming a specific number of senses for each word. In the Non-Parametric Multi-Sense Skip-Gram (NP-MSSG) this number can vary depending on each word. Combining the prior knowledge of lexical databases (e.g., WordNet, ConceptNet, BabelNet), word embeddings and word sense disambiguation, Most Suitable Sense Annotation (MSSA)^[36] labels word-senses through an unsupervised and knowledge-based approach, considering a word's context in a pre-defined sliding window. Once the words are disambiguated, they can be used in a standard word embeddings technique, so multi-sense embeddings are produced. MSSA architecture allows the disambiguation and annotation process to be performed recurrently in a self-improving manner.^[37]

The use of multi-sense embeddings is known to improve performance in several NLP tasks, such as part-of-speech tagging, semantic relation identification, semantic relatedness, named entity recognition and sentiment analysis.^{[38][39]}

As of the late 2010s, contextually-meaningful embeddings such as ELMo and BERT have been developed.^[40] Unlike static word embeddings, these embeddings are at the token-level, in that each occurrence of a word has its own embedding. These embeddings better reflect the multi-sense nature of words, because occurrences of a word in similar contexts are situated in similar regions of BERT's embedding space.^{[41][42]}

For biological sequences: BioVectors

Word embeddings for n -grams in biological sequences (e.g. DNA, RNA, and Proteins) for bioinformatics applications have been proposed by Asgari and Mofrad.^[43] Named bio-vectors (BioVec) to refer to biological sequences in general with protein-vectors (ProtVec) for proteins (amino-acid sequences) and gene-vectors (GeneVec) for gene sequences, this representation can be widely used in applications of deep learning in proteomics and genomics. The results presented by Asgari and Mofrad^[43] suggest that BioVectors can characterize biological sequences in terms of biochemical and biophysical interpretations of the underlying patterns.

Game design

Word embeddings with applications in game design have been proposed by Rabii and Cook^[44] as a way to discover emergent gameplay using logs of gameplay data. The process requires transcribing actions that occur during a game within a formal language and then using the resulting text to create word embeddings. The results presented by Rabii and Cook^[44] suggest that the resulting vectors can capture expert knowledge about games like chess that are not explicitly stated in the game's rules.

Sentence embeddings

The idea has been extended to embeddings of entire sentences or even documents, e.g. in the form of the thought vectors concept. In 2015, some researchers suggested "skip-thought vectors" as a means to improve the quality of machine translation.^[45] A more recent and popular approach for representing sentences is Sentence-BERT, or SentenceTransformers, which modifies pre-trained BERT with the use of siamese and triplet network structures.^[46]

Software

Software for training and using word embeddings includes Tomáš Mikolov's Word2vec, Stanford University's GloVe,^[47] GN-GloVe,^[48] Flair embeddings,^[38] AllenNLP's ELMo,^[49] BERT,^[50] fastText, Gensim,^[51] Indra,^[52] and Deeplearning4j. Principal Component Analysis (PCA) and T-Distributed Stochastic Neighbour Embedding (t-SNE) are both used to reduce the dimensionality of word vector spaces and visualize word embeddings and clusters.^[53]

Examples of application

For instance, the fastText is also used to calculate word embeddings for text corpora in Sketch Engine that are available online.^[54]

Ethical implications

Word embeddings may contain the biases and stereotypes contained in the trained dataset, as Bolukbasi et al. points out in the 2016 paper “Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings” that a publicly available (and popular) word2vec embedding trained on Google News texts (a commonly used data corpus), which consists of text written by professional journalists, still shows disproportionate word associations reflecting gender and racial biases when extracting word analogies.^[55] For example, one of the analogies generated using the aforementioned word embedding is “man is to computer programmer as woman is to homemaker”.^{[56][57]}

Research done by Jieyu Zhou et al. shows that the applications of these trained word embeddings without careful oversight likely perpetuates existing bias in society, which is introduced through unaltered training data. Furthermore, word embeddings can even amplify these biases .^{[58][59]}

See also

- Embedding (machine learning)
- Brown clustering
- Distributional–relational database

References

1. Jurafsky, Daniel; H. James, Martin (2000). *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition* (<https://web.stanford.edu/~jurafsky/slp3/>). Upper Saddle River, N.J.: Prentice Hall. ISBN 978-0-13-095069-7.
2. Mikolov, Tomas; Sutskever, Ilya; Chen, Kai; Corrado, Greg; Dean, Jeffrey (2013). "Distributed Representations of Words and Phrases and their Compositionality". arXiv:1310.4546 (<https://arxiv.org/abs/1310.4546>) [cs.CL (<https://arxiv.org/archive/cs.CL>)].
3. Lebre, Rémi; Collobert, Ronan (2013). "Word Emdeddings through Hellinger PCA". *Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. Vol. 2014. arXiv:1312.5542 (<https://arxiv.org/abs/1312.5542>).
4. Levy, Omer; Goldberg, Yoav (2014). *Neural Word Embedding as Implicit Matrix Factorization* (<http://papers.nips.cc/paper/5477-neural-word-embedding-as-implicit-matrix-factorization.pdf>) (PDF). NIPS.
5. Li, Yitan; Xu, Linli (2015). *Word Embedding Revisited: A New Representation Learning and Explicit Matrix Factorization Perspective* (<http://ijcai.org/papers15/Papers/IJCAI15-513.pdf>) (PDF). Int'l J. Conf. on Artificial Intelligence (IJCAI).
6. Globerson, Amir (2007). "Euclidean Embedding of Co-occurrence Data" (<http://static.googleusercontent.com/media/research.google.com/en//pubs/archive/34951.pdf>) (PDF). *Journal of Machine Learning Research*.

7. Qureshi, M. Atif; Greene, Derek (2018-06-04). "EVE: explainable vector based embedding technique using Wikipedia". *Journal of Intelligent Information Systems*. **53**: 137–165. arXiv:1702.06891 (<https://arxiv.org/abs/1702.06891>). doi:10.1007/s10844-018-0511-x (<https://doi.org/10.1007/s10844-018-0511-x>). ISSN 0925-9902 (<https://search.worldcat.org/isbn/0925-9902>). S2CID 10656055 (<https://api.semanticscholar.org/CorpusID:10656055>).
8. Levy, Omer; Goldberg, Yoav (2014). *Linguistic Regularities in Sparse and Explicit Word Representations* (<https://levyomer.files.wordpress.com/2014/04/linguistic-regularities-in-sparse-and-explicit-word-representations-conll-2014.pdf>) (PDF). CoNLL. pp. 171–180.
9. Socher, Richard; Bauer, John; Manning, Christopher; Ng, Andrew (2013). *Parsing with compositional vector grammars* (https://web.archive.org/web/20160811041232/http://www.socher.org/uploads/Main/SocherBauerManningNg_ACL2013.pdf) (PDF). Proc. ACL Conf. Archived from the original (http://www.socher.org/uploads/Main/SocherBauerManningNg_ACL2013.pdf) (PDF) on 2016-08-11. Retrieved 2014-08-14.
10. Socher, Richard; Perelygin, Alex; Wu, Jean; Chuang, Jason; Manning, Chris; Ng, Andrew; Potts, Chris (2013). *Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank* (http://nlp.stanford.edu/~socherr/EMNLP2013_RNTN.pdf) (PDF). EMNLP.
11. Sahlgren, Magnus. "A brief history of word embeddings" (<https://www.linkedin.com/pulse/brief-history-word-embeddings-some-clarifications-magnus-sahlgren/>).
12. Firth, J.R. (1957). "A synopsis of linguistic theory 1930–1955". *Studies in Linguistic Analysis*: 1–32. Reprinted in F.R. Palmer, ed. (1968). *Selected Papers of J.R. Firth 1952–1959*. London: Longman.
13. Luhn, H.P. (1953). "A New Method of Recording and Searching Information". *American Documentation*. **4**: 14–16. doi:10.1002/asi.5090040104 (<https://doi.org/10.1002/asi.5090040104>).
14. Osgood, C.E.; Suci, G.J.; Tannenbaum, P.H. (1957). *The Measurement of Meaning*. University of Illinois Press.
15. Salton, Gerard (1962). "Some experiments in the generation of word and document associations". *Proceedings of the December 4-6, 1962, fall joint computer conference on - AFIPS '62 (Fall)*. pp. 234–250. doi:10.1145/1461518.1461544 (<https://doi.org/10.1145/1461518.1461544>). ISBN 9781450378796. S2CID 9937095 (<https://api.semanticscholar.org/CorpusID:9937095>).
16. Salton, Gerard; Wong, A; Yang, C S (1975). "A Vector Space Model for Automatic Indexing". *Communications of the ACM*. **18** (11): 613–620. doi:10.1145/361219.361220 (<https://doi.org/10.1145/361219.361220>). hdl:1813/6057 (<https://hdl.handle.net/1813/6057>). S2CID 6473756 (<https://api.semanticscholar.org/CorpusID:6473756>).
17. Dubin, David (2004). "The most influential paper Gerard Salton never wrote" (<https://web.archive.org/web/20201018193507/https://www.thefreelibrary.com/The+most+influential+paper+Gerard+Salton+never+wrote.-a0125151308>). Archived from the original (<https://www.thefreelibrary.com/The+most+influential+paper+Gerard+Salton+never+wrote.-a0125151308>) on 18 October 2020. Retrieved 18 October 2020.
18. Kanerva, Pentti, Kristoferson, Jan and Holst, Anders (2000): *Random Indexing of Text Samples for Latent Semantic Analysis* (<https://cloudfront.escholarship.org/dist/prd/content/qt5644k0w6/qt5644k0w6.pdf>), Proceedings of the 22nd Annual Conference of the Cognitive Science Society, p. 1036. Mahwah, New Jersey: Erlbaum, 2000.
19. Karlgren, Jussi; Sahlgren, Magnus (2001). Uesaka, Yoshinori; Kanerva, Pentti; Asoh, Hideki (eds.). "From words to understanding". *Foundations of Real-World Intelligence*. CSLI Publications: 294–308.
20. Sahlgren, Magnus (2005) *An Introduction to Random Indexing* (http://eprints.sics.se/221/1/R_I_intro.pdf), Proceedings of the Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering, TKE 2005, August 16, Copenhagen, Denmark

21. Sahlgren, Magnus, Holst, Anders and Pentti Kanerva (2008) Permutations as a Means to Encode Order in Word Space (<http://eprints.sics.se/3436/01/permutationsCogSci08.pdf>), In Proceedings of the 30th Annual Conference of the Cognitive Science Society: 1300–1305.
22. Bengio, Yoshua; Réjean, Ducharme; Pascal, Vincent (2000). "A Neural Probabilistic Language Model" (https://proceedings.neurips.cc/paper_files/paper/2000/file/728f206c2a01bf572b5940d7d9a8fa4c-Paper.pdf) (PDF). *NeurIPS*.
23. Bengio, Yoshua; Ducharme, Réjean; Vincent, Pascal; Jauvin, Christian (2003). "A Neural Probabilistic Language Model" (<https://www.jmlr.org/papers/volume3/bengio03a/bengio03a.pdf>) (PDF). *Journal of Machine Learning Research*. **3**: 1137–1155.
24. Bengio, Yoshua; Schwenk, Holger; Senécal, Jean-Sébastien; Morin, Frédéric; Gauvain, Jean-Luc (2006). "A Neural Probabilistic Language Model". *Studies in Fuzziness and Soft Computing*. Vol. 194. Springer. pp. 137–186. doi:10.1007/3-540-33486-6_6 (https://doi.org/10.1007/3-540-33486-6_6). ISBN 978-3-540-30609-2.
25. Vinkourov, Alexei; Cristianini, Nello; Shawe-Taylor, John (2002). *Inferring a semantic representation of text via cross-language correlation analysis* (<https://proceedings.neurips.cc/paper/2002/file/d5e2fbef30a4eb668a203060ec8e5eef-Paper.pdf>) (PDF). Advances in Neural Information Processing Systems. Vol. 15.
26. Lavelli, Alberto; Sebastiani, Fabrizio; Zanolli, Roberto (2004). *Distributional term representations: an experimental comparison*. 13th ACM International Conference on Information and Knowledge Management. pp. 615–624. doi:10.1145/1031171.1031284 (<http://doi.org/10.1145/1031171.1031284>).
27. Roweis, Sam T.; Saul, Lawrence K. (2000). "Nonlinear Dimensionality Reduction by Locally Linear Embedding". *Science*. **290** (5500): 2323–6. Bibcode:2000Sci...290.2323R (<https://ui.adsabs.harvard.edu/abs/2000Sci...290.2323R>). CiteSeerX 10.1.1.111.3313 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.111.3313>). doi:10.1126/science.290.5500.2323 (<https://doi.org/10.1126/science.290.5500.2323>). PMID 11125150 (<https://pubmed.ncbi.nlm.nih.gov/11125150>). S2CID 5987139 (<https://api.semanticscholar.org/CorpusID:5987139>).
28. יהושע בנג'יו: [he](#)
29. Morin, Fredric; Bengio, Yoshua (2005). "Hierarchical probabilistic neural network language model" (<http://proceedings.mlr.press/r5/morin05a/morin05a.pdf>) (PDF). In Cowell, Robert G.; Ghahramani, Zoubin (eds.). *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*. Proceedings of Machine Learning Research. Vol. R5. pp. 246–252.
30. Mnih, Andriy; Hinton, Geoffrey (2009). "A Scalable Hierarchical Distributed Language Model" (<http://papers.nips.cc/paper/3583-a-scalable-hierarchical-distributed-language-model>). *Advances in Neural Information Processing Systems*. 21 (NIPS 2008). Curran Associates, Inc.: 1081–1088.
31. "word2vec" (<https://code.google.com/archive/p/word2vec/>). *Google Code Archive*. Retrieved 23 July 2021.
32. Reisinger, Joseph; Mooney, Raymond J. (2010). *Multi-Prototype Vector-Space Models of Word Meaning* (<https://www.aclweb.org/anthology/N10-1013/>). Vol. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Los Angeles, California: Association for Computational Linguistics. pp. 109–117. ISBN 978-1-932432-65-7. Retrieved October 25, 2019.
33. Huang, Eric. (2012). *Improving word representations via global context and multiple word prototypes*. OCLC 857900050 (<https://search.worldcat.org/oclc/857900050>).
34. Camacho-Collados, Jose; Pilehvar, Mohammad Taher (2018). "From Word to Sense Embeddings: A Survey on Vector Representations of Meaning". arXiv:1805.04032 (<https://arxiv.org/abs/1805.04032>) [cs.CL (<https://arxiv.org/archive/cs.CL>)].

35. Neelakantan, Arvind; Shankar, Jeevan; Passos, Alexandre; McCallum, Andrew (2014). "Efficient Non-parametric Estimation of Multiple Embeddings per Word in Vector Space". *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Stroudsburg, PA, USA: Association for Computational Linguistics. pp. 1059–1069. arXiv:1504.06654 (<https://arxiv.org/abs/1504.06654>). doi:10.3115/v1/d14-1113 (<https://doi.org/10.3115%2Fv1%2Fd14-1113>). S2CID 15251438 (<https://api.semanticscholar.org/CorpusID:15251438>).
36. Ruas, Terry; Grosky, William; Aizawa, Akiko (2019-12-01). "Multi-sense embeddings through a word sense disambiguation process". *Expert Systems with Applications*. **136**: 288–303. arXiv:2101.08700 (<https://arxiv.org/abs/2101.08700>). doi:10.1016/j.eswa.2019.06.026 (<https://doi.org/10.1016%2Fj.eswa.2019.06.026>). hdl:2027.42/145475 (<https://hdl.handle.net/2027.42%2F145475>). ISSN 0957-4174 (<https://search.worldcat.org/issn/0957-4174>). S2CID 52225306 (<https://api.semanticscholar.org/CorpusID:52225306>).
37. Agre, Gennady; Petrov, Daniel; Keskinova, Simona (2019-03-01). "Word Sense Disambiguation Studio: A Flexible System for WSD Feature Extraction" (<https://doi.org/10.3390%2Finfo10030097>). *Information*. **10** (3): 97. doi:10.3390/info10030097 (<https://doi.org/10.3390%2Finfo10030097>). ISSN 2078-2489 (<https://search.worldcat.org/issn/2078-2489>).
38. Akbik, Alan; Blythe, Duncan; Vollgraf, Roland (2018). "Contextual String Embeddings for Sequence Labeling" (<https://www.aclweb.org/anthology/C18-1139>). *Proceedings of the 27th International Conference on Computational Linguistics*. Santa Fe, New Mexico, USA: Association for Computational Linguistics: 1638–1649.
39. Li, Jiwei; Jurafsky, Dan (2015). "Do Multi-Sense Embeddings Improve Natural Language Understanding?". *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA: Association for Computational Linguistics. pp. 1722–1732. arXiv:1506.01070 (<https://arxiv.org/abs/1506.01070>). doi:10.18653/v1/d15-1200 (<https://doi.org/10.18653%2Fv1%2Fd15-1200>). S2CID 6222768 (<https://api.semanticscholar.org/CorpusID:6222768>).
40. Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina (June 2019). "Proceedings of the 2019 Conference of the North" (<https://aclanthology.org/N19-1423/>). *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics: 4171–4186. doi:10.18653/v1/N19-1423 (<https://doi.org/10.18653%2Fv1%2FN19-1423>). S2CID 52967399 (<https://api.semanticscholar.org/CorpusID:52967399>).
41. Lucy, Li, and David Bamman. "Characterizing English variation across social media communities with BERT." *Transactions of the Association for Computational Linguistics* 9 (2021): 538-556.
42. Reif, Emily, Ann Yuan, Martin Wattenberg, Fernanda B. Viegas, Andy Coenen, Adam Pearce, and Been Kim. "Visualizing and measuring the geometry of BERT." *Advances in Neural Information Processing Systems* 32 (2019).
43. Asgari, Ehsaneddin; Mofrad, Mohammad R.K. (2015). "Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4640716>). *PLOS ONE*. **10** (11): e0141287. arXiv:1503.05140 (<https://arxiv.org/abs/1503.05140>). Bibcode:2015PLoSO..1041287A (<https://ui.adsabs.harvard.edu/abs/2015PLoSO..1041287A>). doi:10.1371/journal.pone.0141287 (<https://doi.org/10.1371%2Fjournal.pone.0141287>). PMC 4640716 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4640716>). PMID 26555596 (<https://pubmed.ncbi.nlm.nih.gov/26555596>).

44. Rabii, Younès; Cook, Michael (2021-10-04). "Revealing Game Dynamics via Word Embeddings of Gameplay Data" (<https://ojs.aaai.org/index.php/AIIDE/article/view/18907>). *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*. **17** (1): 187–194. doi:10.1609/aiide.v17i1.18907 (<https://doi.org/10.1609%2Faiide.v17i1.18907>). ISSN 2334-0924 (<https://search.worldcat.org/issn/2334-0924>). S2CID 248175634 (<https://api.semanticscholar.org/CorpusID:248175634>).
45. Kiros, Ryan; Zhu, Yukun; Salakhutdinov, Ruslan; Zemel, Richard S.; Torralba, Antonio; Urtasun, Raquel; Fidler, Sanja (2015). "skip-thought vectors". arXiv:1506.06726 (<https://arxiv.org/abs/1506.06726>) [cs.CL (<https://arxiv.org/archive/cs.CL>)].
46. Reimers, Nils, and Iryna Gurevych. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 3982-3992. 2019.
47. "GloVe" (<http://nlp.stanford.edu/projects/glove/>).
48. Zhao, Jieyu; et al. (2018) (2018). "Learning Gender-Neutral Word Embeddings". arXiv:1809.01496 (<https://arxiv.org/abs/1809.01496>) [cs.CL (<https://arxiv.org/archive/cs.CL>)].
49. "Elmo" (<https://allennlp.org/elmo>). 16 October 2024.
50. Pires, Telmo; Schlinger, Eva; Garrette, Dan (2019-06-04). "How multilingual is Multilingual BERT?". arXiv:1906.01502 (<https://arxiv.org/abs/1906.01502>) [cs.CL (<https://arxiv.org/archive/cs.CL>)].
51. "Gensim" (<http://radimrehurek.com/gensim/>).
52. "Indra" (<https://github.com/Lambda-3/Indra>). *GitHub*. 2018-10-25.
53. Ghassemi, Mohammad; Mark, Roger; Nemati, Shamim (2015). "A visualization of evolving clinical sentiment using vector representations of clinical notes" (<http://www.cinc.org/archive/s/2015/pdf/0629.pdf>) (PDF). *2015 Computing in Cardiology Conference (CinC)*. Vol. 2015. pp. 629–632. doi:10.1109/CIC.2015.7410989 (<https://doi.org/10.1109%2FCIC.2015.7410989>). ISBN 978-1-5090-0685-4. PMC 5070922 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5070922>). PMID 27774487 (<https://pubmed.ncbi.nlm.nih.gov/27774487>).
54. "Embedding Viewer" (<https://web.archive.org/web/20180208004241/https://embeddings.sketchengine.co.uk/>). *Embedding Viewer*. Lexical Computing. Archived from the original (<https://embeddings.sketchengine.co.uk/>) on 8 February 2018. Retrieved 7 Feb 2018.
55. Bolukbasi, Tolga; Chang, Kai-Wei; Zou, James; Saligrama, Venkatesh; Kalai, Adam (2016). "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings". arXiv:1607.06520 (<https://arxiv.org/abs/1607.06520>) [cs.CL (<https://arxiv.org/archive/cs.CL>)].
56. Bolukbasi, Tolga; Chang, Kai-Wei; Zou, James; Saligrama, Venkatesh; Kalai, Adam (2016-07-21). "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings". arXiv:1607.06520 (<https://arxiv.org/abs/1607.06520>) [cs.CL (<https://arxiv.org/archive/cs.CL>)].
57. Dieng, Adji B.; Ruiz, Francisco J. R.; Blei, David M. (2020). "Topic Modeling in Embedding Spaces" (https://direct.mit.edu/tac/article/doi/10.1162/tac_l_a_00325/96463/Topic-Modeling-in-Embedding-Spaces). *Transactions of the Association for Computational Linguistics*. **8**: 439–453. arXiv:1907.04907 (<https://arxiv.org/abs/1907.04907>). doi:10.1162/tac_l_a_00325 (https://doi.org/10.1162%2Ftac_l_a_00325).
58. Zhao, Jieyu; Wang, Tianlu; Yatskar, Mark; Ordonez, Vicente; Chang, Kai-Wei (2017). "Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints" (<https://aclanthology.org/D17-1323/>). *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. pp. 2979–2989. doi:10.18653/v1/D17-1323 (<https://doi.org/10.18653%2Fv1%2FD17-1323>).

59. Petreski, Davor; Hashim, Ibrahim C. (2022-05-26). "Word embeddings are biased. But whose bias are they reflecting?" (<https://doi.org/10.1007%2Fs00146-022-01443-w>). *AI & Society*. **38** (2): 975–982. doi:10.1007/s00146-022-01443-w (<https://doi.org/10.1007%2Fs00146-022-01443-w>). ISSN 1435-5655 (<https://search.worldcat.org/issn/1435-5655>). S2CID 249112516 (<https://api.semanticscholar.org/CorpusID:249112516>).
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Word_embedding&oldid=1300900105"