

Proyecto 02 - Top2Vec: Distributed representations of topics
Giovanni Gamaliel López Padilla

Abstract

Índice

1. Introduction	1
2. Methods	1
3. Results	1
4. Conclusion	1
5. Referencias	1

1. Introduction

Natural language processing is a sub field of Artificial Intelligence and linguistics. It was formed in 1960 with the objective of studying problems en la automatización de la generación de texto y comprensión del lenguaje natural. Al inicio de la decada se obtuvo un desarrollo tecnológico el cual habilito el procesamiento de grandes cantidades de información con un nivel de eficiencia alto.

2. Methods

3. Results

4. Conclusion

5. Referencias

- [1] Angelov D. Top2Vec: Distributed Representations of Topics. CoRR. 2020;abs/2008.09470. Available from: <https://doi.org/10.48550/arXiv.2008.09470>.
- [2] Blei D, Ng A, Jordan M. Latent Dirichlet Allocation. In: Dietterich T, Becker S, Ghahramani Z, editors. Advances in Neural Information Processing Systems. vol. 14. MIT Press; 2001. p. 993–1022. Available from: <https://proceedings.neurips.cc/paper/2001/file/296472c9542ad4d4788d543508116cbc-Paper.pdf>.
- [3] Hofmann T, et al. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. ACM. 1999;p. 50–57.

- [4] Boyd-graber J, Blei D. Syntactic Topic Models. In: Koller D, Schuurmans D, Bengio Y, Bottou L, editors. *Advances in Neural Information Processing Systems*. vol. 21. Curran Associates, Inc.; 2008. p. 185–192. Available from: <https://proceedings.neurips.cc/paper/2008/file/8d3bba7425e7c98c50f52ca1b52d3735-Paper.pdf>.
- [5] Hinton GE, et al. Learning distributed representations of concepts. In: *Proceedings of the eighth annual conference of the cognitive science society*. vol. 1. Amherst, MA; 1986. p. 12.
- [6] Widdowson H. J.R. Firth, 1957, *Papers in Linguistics* 1934–51. *Int J Applied Linguistics*. 2007 nov;17(3):402–413. Available from: <https://doi.org/10.1111%2Fj.1473-4192.2007.00164.x>.
- [7] Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. *arXiv*; 2013. Available from: <https://doi.org/10.48550/arXiv.1301.3781>.
- [8] Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed Representations of Words and Phrases and their Compositionality. In: Burges CJ, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, editors. *Advances in Neural Information Processing Systems*. vol. 26. Curran Associates, Inc.; 2013. p. 3111–3119. Available from: <https://proceedings.neurips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf>.
- [9] Baroni M, Dinu G, Kruszewski G. Don’t count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*; 2014. p. 238–247. Available from: <http://aclweb.org/anthology/P/P14/P14-1023.pdf>.
- [10] Bairong Z, Wenbo W, Zhiyu L, Chonghui Z, Shinozaki T. Comparative analysis of word embedding methods for DSTC6 end-to-end conversation modeling track. In: *Proceedings of the 6th Dialog System Technology Challenges (DSTC6) Workshop*; 2017. Available from: http://workshop.colips.org/dstc6/papers/track2_paper18_zhuang.pdf.
- [11] Levy O, Goldberg Y. Neural Word Embedding as Implicit Matrix Factorization. In: Ghahramani Z, Welling M, Cortes C, Lawrence N, Weinberger KQ, editors. *Advances in Neural Information Processing Systems*. vol. 27. Curran Associates, Inc.; 2014. p. 2177–2185. Available from: <https://proceedings.neurips.cc/paper/2014/file/feab05aa91085b7a8012516bc3533958-Paper.pdf>.
- [12] Pennington J, Socher R, Manning CD. Glove: Global vectors for word representation. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*; 2014. p. 1532–1543. Available from: <https://aclanthology.org/D14-1162.pdf>.
- [13] Bengio Y, Ducharme R, Vincent P. A Neural Probabilistic Language Model. In: Leen T, Dietterich T, Tresp V, editors. *Advances in Neural Information Processing Systems*. vol. 13. MIT Press; 2000. Available from: <https://proceedings.neurips.cc/paper/2000/file/728f206c2a01bf572b5940d7d9a8fa4c-Paper.pdf>.
- [14] Mikolov T, Yih Wt, 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*; 2013. p. 746–751. Available from: <https://aclanthology.org/N13-1090.pdf>.
- [15] Levy O, Goldberg Y, Dagan I. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the association for computational linguistics*. 2015;3:211–225. Available from: https://direct.mit.edu/tacl/article-abstract/doi/10.1162/tacl_a_00134/43264.

- [16] Le Q, Mikolov T. Distributed representations of sentences and documents. In: International conference on machine learning. PMLR; 2014. p. 1188–1196. Available from: <http://proceedings.mlr.press/v32/le14.html?ref=https://githubhelp.com>.
- [17] Griffiths TL, Steyvers M, Tenenbaum JB. Topics in semantic representation. Psychological review. 2007;114(2):211. Available from: <https://psycnet.apa.org/record/2007-05396-001>.
- [18] Řehůřek R, Sojka P. Software Framework for Topic Modelling with Large Corpora. In: Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks. Valletta, Malta: University of Malta; 2010. p. 46–50. Available from: <http://www.fi.muni.cz/usr/sojka/presentations/lrec2010-poster-rehurek-sojka.pdf>.
- [19] Campello RJ, Moulavi D, Sander J. Density-based clustering based on hierarchical density estimates. In: Pacific-Asia conference on knowledge discovery and data mining. Springer; 2013. p. 160–172. Available from: https://link.springer.com/chapter/10.1007/978-3-642-37456-2_14.
- [20] McInnes L, Healy J. Accelerated hierarchical density based clustering. In: 2017 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE; 2017. p. 33–42. Available from: <https://ieeexplore.ieee.org/abstract/document/8215642/>.
- [21] McInnes L, Healy J, Astels S. hdbscan: Hierarchical density based clustering. J Open Source Softw. 2017;2(11):205. Available from: <https://doi.org/10.21105/joss.00205>.
- [22] McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:180203426. 2018; Available from: <https://arxiv.org/abs/1802.03426>.