Title: Analyse text by computational methods in social research

Trends of computational text analysis

Cambria, Erik, and Bebo White. 2014. "Jumping NLP Curves: A Review of Natural Language Processing Research." *IEEE Computational Intelligence Magazine* 9(2):48–57. doi: 10.1109/mci.2014.2307227.

Khurana, Diksha, Aditya Koli, Kiran Khatter, and Sukhdev Singh. 2023. "Natural Language Processing: State of the Art, Current Trends and Challenges." *Multimedia Tools and Applications* 82(3):3713–44. doi: 10.1007/s11042-022-13428-4.

General descriptions

Church, Kenneth Ward, and Patrick Hanks. 1989. "Word Association Norms, Mutual Information, and Lexicography." in *Proceedings of the 27th annual meeting on Association for Computational Linguistics*. Morristown, NJ, USA: Association for Computational Linguistics.

Egami, Naoki, Christian J. Fong, Justin Grimmer, Margaret E. Roberts, and Brandon M. Stewart. 2022. "How to Make Causal Inferences Using Texts." *Science Advances* 8(42):eabg2652. doi: 10.1126/sciadv.abg2652.

Fairclough, Norman. 1992. "Discourse and Text: Linguistic and Intertextual Analysis within Discourse Analysis." *Discourse & Society* 3(2):193–217. doi: 10.1177/0957926592003002004.

Feder, Amir, Katherine A. Keith, Emaad Manzoor, Reid Pryzant, Dhanya Sridhar, Zach Wood-Doughty, Jacob Eisenstein, Justin Grimmer, Roi Reichart, Margaret E. Roberts, Brandon M. Stewart, Victor Veitch, and Diyi Yang. 2022. "Causal Inference in Natural Language Processing: Estimation, Prediction, Interpretation and Beyond." *Transactions of the Association for Computational Linguistics* 10:1138–58. doi: 10.1162/tacl_a_00511.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis: An Annual Publication of the Methodology Section of the American Political Science Association* 21(3):267–97. doi: 10.1093/pan/mps028.

Hirschberg, Julia, and Christopher Manning. 2015. "Advances in Natural Language Processing." *Science* 349(6245):261–66.

Keith, Katherine, David Jensen, and Brendan O'Connor. 2020. "Text and Causal Inference: A Review of Using Text to Remove Confounding from Causal Estimates." in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics.

Otter, Daniel W., Julian R. Medina, and Jugal K. Kalita. 2021. "A Survey of the Usages of Deep Learning for Natural Language Processing." *IEEE Transactions on Neural Networks and Learning Systems* 32(2):604–24. doi: 10.1109/TNNLS.2020.2979670.

Text analysis methods

Classification of word and word count

Applications

Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *The American Political Science Review* 97(02). doi: 10.1017/s0003055403000698.

De Choudhury, Munmun, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. "Predicting Depression via Social Media." Pp. 128–37 in *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*.

Data mining

Techniques

Frey, Brendan J., and Delbert Dueck. 2007. "Clustering by Passing Messages between Data Points." *Science* 315(5814):972–76. doi: 10.1126/science.1136800.

Hearst, Marti A. 1999. "Untangling Text Data Mining." in *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics -*. Morristown, NJ, USA: Association for Computational Linguistics.

Applications

Ghazouani, Dhouha, Luigi Lancieri, Habib Ounelli, Chaker Jebari. 2019. "Assessing Socioeconomic Status of Twitter Users: A Survey." in *Proceedings - Natural Language Processing in a Deep Learning World*. Incoma Ltd., Shoumen, Bulgaria.

Ikeda, Kazushi, Gen Hattori, Chihiro Ono, Hideki Asoh, and Teruo Higashino. 2013. "Twitter User Profiling Based on Text and Community Mining for Market Analysis." *Knowledge-Based Systems* 51:35–47. doi: 10.1016/j.knosys.2013.06.020.

Li, Jiwei, Alan Ritter, and Eduard Hovy. 2014. "Weakly Supervised User Profile Extraction from Twitter." in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg, PA, USA: Association for Computational Linguistics.

Topic modelling

Techniques

Blei, David M., Andrew Y. Ng, Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.

Chang, Jonathan, Jordan Boyd-Graber, Chong Wang, Sean Gerrish, and David M. Blei. 2009. "Reading Tea Leaves: How Humans Interpret Topic Models." Pp. 288–96 in *Proceedings of the 22nd International Conference on Neural Information Processing Systems*.

Hofmann, Thomas. 1999. "Probabilistic Latent Semantic Indexing." in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-ended Survey Responses: Structural Topic Models for Survey Responses." *American Journal of Political Science* 58(4):1064–82. doi: 10.1111/ajps.12103.

Applications

Guo, Yue, Stuart J. Barnes, and Qiong Jia. 2017. "Mining Meaning from Online Ratings and Reviews: Tourist Satisfaction Analysis Using Latent Dirichlet Allocation." *Tourism Management* 59:467–83. doi: 10.1016/j.tourman.2016.09.009.

Hall, David, Daniel Jurafsky, and Christopher D. Manning. 2008. "Studying the History of Ideas Using Topic Models." in *Proceedings of the Conference on Empirical Methods in Natural Language Processing - EMNLP '08*. Morristown, NJ, USA: Association for Computational Linguistics.

Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin, and Dragomir R. Radev. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54(1):209–28. doi: 10.1111/j.1540-5907.2009.00427.x.

Word embedding

Techniques

Bengio, Yoshua, Holger Schwenk, Jean-Sébastien Senécal, Fréderic Morin, and Jean-Luc Gauvain. 2006. "Neural Probabilistic Language Models." Pp. 137–86 in *Innovations in Machine Learning*. Berlin/Heidelberg: Springer-Verlag.

Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. "Indexing by Latent Semantic Analysis." *Journal of the American Society for Information Science* 41(6):391–407.

Gong, Yihong, and Xin Liu. 2001. "Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis." in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Distributed Representations of Words and Phrases and Their Compositionality." in *Proceedings of the 26th International Conference on Neural Information Processing Systems*. Lake Tahoe, Nevada, Curran Associates Inc.

Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. "Glove: Global Vectors for Word Representation." in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Stroudsburg, PA, USA: Association for Computational Linguistics.

Peters, Matthew, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. "Deep Contextualized Word Representations." in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Stroudsburg, PA, USA: Association for Computational Linguistics.

Applications

Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings." *American Sociological Review* 84(5):905–49. doi: 10.1177/0003122419877135.

Preoţiuc-Pietro, Daniel, Vasileios Lampos, and Nikolaos Aletras. 2015. "An Analysis of the User Occupational Class through Twitter Content." in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Stroudsburg, PA, USA: Association for Computational Linguistics.

Preoţiuc-Pietro, Daniel, Svitlana Volkova, Vasileios Lampos, Yoram Bachrach, and Nikolaos Aletras. 2015. "Studying User Income through Language, Behaviour and Affect in Social Media." *PloS One* 10(9):e0138717. doi: 10.1371/journal.pone.0138717.

Sentiment analysis

Techniques

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. "Thumbs up?: Sentiment Classification Using Machine Learning Techniques." in *Proceedings of the ACL-02*

conference on Empirical methods in natural language processing - EMNLP '02.
Morristown, NJ, USA: Association for Computational Linguistics.

Applications

Go, Alec, Richa Bhayani, and Lei Huang. 2009. *Twitter Sentiment Classification Using Distant Supervision*. CS224N project report, Stanford, 1(12):2009

Melville, Prem, Wojciech Gryc, and Richard D. Lawrence. 2009. "Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification." in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM.

Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. 2005. "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis." in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05*. Morristown, NJ, USA: Association for Computational Linguistics.

Unclassified

Techniques

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." in *Proceedings of NAACL-HLT*. Minneapolis, Minnesota: Association for Computational Linguistics.

Hopkins, Daniel J., and Gary King. 2010. "A Method of Automated Nonparametric Content Analysis for Social Science." *American Journal of Political Science* 54(1):229–47. doi: 10.1111/j.1540-5907.2009.00428.x.

Applications

Monroe, Burt L., Michael P. Colaresi, and Kevin M. Quinn. 2008. "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict." *Political Analysis: An Annual Publication of the Methodology Section of the American Political Science Association* 16(4):372–403. doi: 10.1093/pan/mpn018.