
Bibliography

- [1] C. Aggarwal. Data classification: Algorithms and applications, *CRC Press*, 2014.
- [2] C. Aggarwal. Data mining: The textbook. *Springer*, 2015.
- [3] C. Aggarwal. Recommender systems: The textbook. *Springer*, 2016.
- [4] C. Aggarwal. Outlier analysis. *Springer*, 2017.
- [5] C. Aggarwal. On the effects of dimensionality reduction on high dimensional similarity search. *ACM PODS Conference*, pp. 256–266, 2001.
- [6] C. Aggarwal, S. Gates, and P. Yu. On using partial supervision for text categorization. *IEEE Transactions on Knowledge and Data Engineering*, 16(2), 245–255, 2004. [Extended version of ACM KDD 1998 paper “On the merits of building categorization systems by supervised clustering.”]
- [7] C. Aggarwal and N. Li. On node classification in dynamic content-based networks. *SDM Conference*, pp. 355–366, 2011.
- [8] C. Aggarwal and C. Reddy. Data clustering: algorithms and applications, *CRC Press*, 2013.
- [9] C. Aggarwal and S. Sathe. Outlier ensembles: An introduction. *Springer*, 2017.
- [10] C. Aggarwal and K. Subbian. Event detection in social streams. *SDM Conference*, 2012.
- [11] C. Aggarwal, Y. Xie, and P. Yu. On Dynamic Link Inference in Heterogeneous Networks. *SDM Conference*, pp. 415–426, 2012.
- [12] C. Aggarwal and P. Yu. On effective conceptual indexing and similarity search in text data. *ICDM Conference*, pp. 3–10, 2001.
- [13] C. Aggarwal and P. Yu. On clustering massive text and categorical data streams. *Knowledge and Information Systems*, 24(2), pp. 171–196, 2010.
- [14] C. Aggarwal, and C. Zhai, Mining text data. *Springer*, 2012.

- [15] C. Aggarwal and P. Zhao. Towards graphical models for text processing. *Knowledge and Information Systems*, 36(1), pp. 1–21, 2013. [Preliminary version in *ACM SIGIR*, 2010]
- [16] C. Aggarwal, Y. Zhao, and P. Yu. On the use of side information for mining text data. *IEEE Transactions on Knowledge and Data Engineering*, 26(6), pp. 1415–1429, 2014.
- [17] E. Agichtein and L. Gravano. Snowball: Extracting relations from large plain-text collections. *ACM Conference on Digital Libraries*, pp. 85–94, 2000.
- [18] J. Allan, J. Carbonell, G. Doddington, J. Yamron, and Y. Yang. Topic detection and tracking pilot study final report. *CMU Technical Report*, Paper 341, 1998.
- [19] J. Allan, R. Papka, V. Lavrenko. Online new event detection and tracking. *ACM SIGIR Conference*, 1998.
- [20] A. Andreevskaia and S. Bergler. Mining WordNet for a Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses. *European Chapter of the Association for Computational Linguistics*, pp. 209–216, 2006.
- [21] R. Angelova and S. Siersdorfer. A neighborhood-based approach for clustering of linked document collections. *ACM CIKM Conference*, pp. 778–779, 2006.
- [22] V. Anh, O. de Kretser, and A. Moffat. Vector-space ranking with effective early termination. *ACM SIGIR Conference*, pp. 35–42, 2001.
- [23] V. Anh and A. Moffat. Inverted index compression using word-aligned binary codes. *Information Retrieval*, 8(1), pp. 151–166, 2005.
- [24] V. Anh and A. Moffat. Pruned query evaluation using pre-computed impacts. *ACM SIGIR Conference*, pp. 372–379, 2006.
- [25] V. Anh and A. Moffat. Improved word-aligned binary compression for text indexing. *IEEE Transactions on Knowledge and Data Engineering*, 18(6), pp. 857–861, 2006.
- [26] M. Antonie and O. Zaïane. Text document categorization by term association. *IEEE ICDM Conference*, pp. 19–26, 2002.
- [27] C. Apte, F. Damerau, and S. Weiss. Automated learning of decision rules for text categorization, *ACM Transactions on Information Systems*, 12(3), pp. 233–251, 1994.
- [28] C. Apte, F. Damerau, and S. Weiss. Text mining with decision rules and decision trees. *Conference on Automated Learning and Discovery*, Also appears as *IBM Research Report*, RC21219, 1998.
- [29] A. Asuncion, M. Welling, P. Smyth, and Y. Teh. On smoothing and inference for topic models. *Uncertainty in Artificial Intelligence*, pp. 27–34, 2009.
- [30] S. Baccianella, A. Esuli, and F. Sebastiani. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *LREC*, pp. 2200–2204, 2010.
- [31] R. Baeza-Yates, and B. Ribeiro-Neto. Modern information retrieval. *ACM press*, 2011.
- [32] R. Baeza-Yates, A. Gionis, F. Junqueira, V. Murdock, , V. Plachouras, and F. Silvestri. The impact of caching on search engines. *ACM SIGIR Conference*, pp. 183–190, 2007.

- [33] L. Baker and A. McCallum. Distributional clustering of words for text classification. *ACM SIGIR Conference*, pp. 96–103, 1998.
- [34] L. Ballesteros and W. B. Croft. Dictionary methods for cross-lingual information retrieval. *International Conference on Database and Expert Systems Applications*, pp. 791–801, 1996.
- [35] M. Banko and O. Etzioni. The tradeoffs between open and traditional relation extraction. *ACL Conference*, pp. 28–36, 2008.
- [36] R. Banchs. Text Mining with MATLAB. *Springer*, 2012.
- [37] M. Baroni, G. Dinu, and G. Kruszewski. Don’t count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *ACL*, pp. 238–247, 2014.
- [38] M. Baroni and A. Lenci. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4), pp. 673–721, 2010.
- [39] R. Barzilay and M. Elhadad. Using lexical chains for text summarization. *Advances in Automatic Text Summarization*, pp. 111–121, 1999.
- [40] R. Barzilay, N. Elhadad, and K. McKeown. Inferring strategies for sentence ordering in multidocument news summarization. *Journal of Artificial Intelligence Research*, 17, pp. 35–55, 2002.
- [41] R. Barzilay and K. R. McKeown. Sentence fusion for multidocument news summarization. *Computational Linguistics*, 31(3), pp. 397–328, 2005.
- [42] I. Bayer. Fastfm: a library for factorization machines. *arXiv preprint arXiv:1505.00641*, 2015. <https://arxiv.org/pdf/1505.00641v2.pdf>
- [43] H. Becker, M. Naaman, and L. Gravano. Beyond Trending Topics: Real-World Event Identification on Twitter. *ICWSM Conference*, pp. 438–441, 2011.
- [44] D. Beeferman, A. Berger, and J. Lafferty. Statistical models for text segmentation. *Machine Learning*, 34(1–3), pp. 177–210, 1999.
- [45] O. Bender, F. Och, and H. Ney. Maximum entropy models for named entity recognition. *Conference on Natural Language Learning at HLT-NAACL 2003*, pp. 148–51, 2003.
- [46] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida. Detecting spammers on twitter. *Collaboration, Electronic Messaging, Anti-abuse and Spam Conference*, 2010.
- [47] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin. A neural probabilistic language model. *Journal of Machine Learning Research*, 3, pp. 1137–1155, 2003.
- [48] D. Bertsekas. Nonlinear programming. *Athena Scientific*, 1999.
- [49] D. Bikel, S. Miller, R. Schwartz, and R. Weischedel. Nymble: a high-performance learning name-finder. *Applied Natural Language Processing Conference*, pp. 194–201, 1997.
- [50] C. M. Bishop. Pattern recognition and machine learning. *Springer*, 2007.

- [51] C. M. Bishop. Neural networks for pattern recognition. *Oxford University Press*, 1995.
- [52] D. Blei. Probabilistic topic models. *Communications of the ACM*, 55(4), pp. 77–84, 2012.
- [53] D. Blei and P. Moreno. Topic segmentation with an aspect hidden Markov model. *ACM SIGIR Conference*, pp. 343–348, 2001.
- [54] D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, pp. 993–1022, 2003.
- [55] D. Blei and J. Lafferty. Dynamic topic models. *ICML Conference*, pp. 113–120, 2006.
- [56] A. Blum, and T. Mitchell. Combining labeled and unlabeled data with co-training. *COLT*, 1998.
- [57] A. Blum and S. Chawla. Combining labeled and unlabeled data with graph mincuts. *ICML Conference*, 2001.
- [58] D. Boley, M. Gini, R. Gross, E.-H. Han, K. Hastings, G. Karypis, V. Kumar, B. Mobasher, and J. Moore. Partitioning-based clustering for Web document categorization. *Decision Support Systems*, Vol. 27, pp. 329–341, 1999.
- [59] E. Breck, Y. Choi, and C. Cardie. Identifying expressions of opinion in ontext. *IJCAI*, pp. 2683–2688, 2007.
- [60] L. Breiman. Random forests. *Journal Machine Learning archive*, 45(1), pp. 5–32, 2001.
- [61] L. Breiman. Bagging predictors. *Machine Learning*, 24(2), pp. 123–140, 1996.
- [62] L. Breiman and A. Cutler. Random Forests Manual v4.0, *Technical Report*, UC Berkeley, 2003. https://www.stat.berkeley.edu/~breiman/Using_random_forests_v4.0.pdf
- [63] S. Brin. Extracting patterns and relations from the World Wide Web. *International Workshop on the Web and Databases*, 1998. http://link.springer.com/chapter/10.1007/10704656_11#page-1
- [64] S. Brin, and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 30(1–7), pp. 107–117, 1998.
- [65] P. Bühlmann and B. Yu. Analyzing bagging. *Annals of Statistics*, pp. 927–961, 2002.
- [66] J. Bullinaria and J. Levy. Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, 39(3), pp. 510–526, 2007.
- [67] R. Bunescu and R. Mooney. A shortest path dependency kernel for relation extraction. *Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 724–731, 2005.
- [68] R. Bunescu and R. Mooney. Subsequence kernels for relation extraction. *NIPS Conference*, pp. 171–178, 2005.
- [69] C. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2), pp. 121–167, 1998.

- [70] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Huelender. Learning to rank using gradient descent. *ICML Conference*, pp. 86–96, 2005.
- [71] S. Butcher, C. Clarke, and G. V. Cormack. Information retrieval: Implementing and evaluating search engines. *The MIT Press*, 2010.
- [72] J. Callan. Distributed information retrieval. *Advances in Information Retrieval*, Springer, pp. 127–150, 2000.
- [73] M. Califf and R. Mooney. Bottom-up relational learning of pattern matching rules for information extraction. *Journal of Machine Learning Research*, 4, pp. 177–210, 2003.
- [74] Y. Cao, J. Xu, T. Liu, H. Li, Y. Huang, and H.-W. Hon. Adapting ranking SVM to document retrieval. *ACM SIGIR Conference*, pp. 186–193, 2006.
- [75] Z. Cao, T. Qin, T. Liu, M. Tsai, and H. Li. Learning to rank: from pairwise approach to listwise approach. *ICML Conference*, pp. 129–136, 2007.
- [76] J. Carbonell and J. Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. *ACM SIGIR Conference*, pp. 335–336, 1998.
- [77] D. Carmel, D. Cohen, R. Fagin, E. Farchi, M. Herscovici, Y. Maarek, and A. Soffer. Static index pruning for information retrieval systems. *ACM SIGIR Conference*, pp. 43–50, 2001.
- [78] D. Chakrabarti and K. Punera. Event Summarization Using Tweets. *ICWSM Conference*, 11, pp. 66–73, 2011.
- [79] S. Chakrabarti. Mining the Web: Discovering knowledge from hypertext data. *Morgan Kaufmann*, 2003.
- [80] S. Chakrabarti, B. Dom, R. Agrawal, and P. Raghavan. Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies. *The VLDB Journal*, 7(3), pp. 163–178, 1998.
- [81] S. Chakrabarti, B. Dom, and P. Indyk. Enhanced hypertext categorization using hyperlinks. *ACM SIGMOD Conference*, pp. 307–318, 1998.
- [82] S. Chakrabarti, S. Roy, and M. Soundalgekar. Fast and accurate text classification via multiple linear discriminant projections. *The VLDB Journal*, 12(2), pp. 170–185, 2003.
- [83] S. Chakrabarti, M. Van den Berg, and B. Dom. Focused crawling: a new approach to topic-specific Web resource discovery. *Computer Networks*, 31(11), pp. 1623–1640, 1999.
- [84] Y. Chali and S. Joty. Improving the performance of the random walk model for answering complex questions. *Annual Meeting of the Association for Computational Linguistics on Human Language Technologies*, pp. 9–12, 2008.
- [85] N. Chambers, S. Wang, and D. Jurafsky. Classifying temporal relations between events. *Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pp. 173–176, 2007.

- [86] Y. Chan and D. Roth. Exploiting syntactico-semantic structures for relation extraction. *ACL Conference: Human Language Technologies*, pp. 551–560, 2011.
- [87] C. Chang and C. Lin. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 27, 2011. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [88] Y. Chang, C. Hsieh, K. Chang, M. Ringgaard, and C. J. Lin. Training and testing low-degree polynomial data mappings via linear SVM. *Journal of Machine Learning Research*, 11, pp. 1471–1490, 2010.
- [89] O. Chapelle. Training a support vector machine in the primal. *Neural Computation*, 19(5), pp. 1155–1178, 2007.
- [90] O. Chapelle, B. Schölkopf, and A. Zien. Semi-supervised learning. *MIT Press*, 2010.
- [91] P. Cheeseman and J. Stutz. Bayesian classification (AutoClass): Theory and results. *Advances in Knowledge Discovery and Data Mining*, Eds. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthuruswamy. AAAI Press/MIT Press, 1996.
- [92] D. Chickering, D. Heckerman, and C. Meek. A Bayesian approach to learning Bayesian networks with local structure. *Uncertainty in Artificial Intelligence*, pp. 80–89, 1997.
- [93] J. Cho, H. Garcia-Molina, and L. Page. Efficient crawling through URL ordering. *Computer Networks*, 30(1–7), pp. 161–172, 1998.
- [94] K. Cho, B. Merriënboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *EMNLP*, 2014. <https://arxiv.org/pdf/1406.1078.pdf>
- [95] F. Choi. Advances in domain independent linear text segmentation. *North American Chapter of the Association for Computational Linguistics Conference*, pp. 26–33, 2000.
- [96] F. Choi, P. Wiemer-Hastings, and J. Moore. Latent semantic analysis for text segmentation. *EMNLP*, 2001.
- [97] Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. Identifying sources of opinions with conditional random fields and extraction patterns. *Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 355–362, 2005.
- [98] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv:1412.3555*, 2014. <https://arxiv.org/abs/1412.3555>
- [99] K. Church and P. Hanks. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), pp. 22–29, 1990.
- [100] F. Ciravegna. Adaptive information extraction from text by rule induction and generalisation. *International Joint Conference on Artificial Intelligence*, 17(1), pp. 1251–1256, 2001.
- [101] J. Clarke and M. Lapata. Models for sentence compression: A comparison across domains, training requirements and evaluation measures. *ACL Conference*, pp. 377–384, 2006.

- [102] W. Cohen. Fast effective rule induction. *ICML Conference*, pp. 115–123, 1995.
- [103] W. Cohen. Learning rules that classify e-mail. *AAAI Spring Symposium on Machine Learning in Information Access*, 1996.
- [104] W. Cohen. Learning with set-valued features. In *National Conference on Artificial Intelligence*, 1996.
- [105] W. Cohen, R. Schapire, and Y. Singer. Learning to Order Things. *Journal of Artificial Intelligence Research*, 10, pp. 243–270, 1999.
- [106] W. Cohen and Y. Singer. Context-sensitive learning methods for text categorization. *ACM Transactions on Information Systems*, 17(2), pp 141–173, 1999.
- [107] M. Collins and N. Duffy. Convolution kernels for natural language. *NIPS Conference*, pp. 625–632, 2001.
- [108] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12, pp. 2493–2537, 2011.
- [109] R. Collobert and J. Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. *ICML Conference*, pp. 160–167, 2008.
- [110] J. Conroy and D. O’Leary. Text summarization via hidden markov models. *ACM SIGIR Conference*, pp. 406–407, 2001.
- [111] J. Conroy, J. Schlessinger, D. O’Leary, and J. Goldstein. Back to basics: CLASSY 2006. *Document Understanding Conference*, 2006.
- [112] T. Cooke. Two variations on Fisher’s linear discriminant for pattern recognition *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2), pp. 268–273, 2002.
- [113] W. Cooper. Some inconsistencies and misnomers in probabilistic information retrieval. *ACM Transactions on Information Systems*, 13(1), pp. 100–111, 1995.
- [114] B. O’Connor, R. Balasubramanyan, B. Routledge, and N. Smith. From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, pp. 122–129, 2010.
- [115] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3), pp. 273–297, 1995.
- [116] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), pp. 1–27, 1967.
- [117] N. Cristianini, and J. Shawe-Taylor. An introduction to support vector machines and other kernel-based learning methods. *Cambridge University Press*, 2000.
- [118] W. B. Croft. Clustering large files of documents using the single-link method. *Journal of the American Society of Information Science*, 28, pp. 341–344, 1977.
- [119] W. B. Croft and D. Harper. Using probabilistic models of document retrieval without relevance information. *Journal of Documentation*, 35(4), pp. 285–295, 1979.
- [120] W. B. Croft, D. Metzler, and T. Strohman. Search engines: Information retrieval in practice, *Addison-Wesley Publishing Company*, 2009.

- [121] S. Cucerzan. Large-scale named entity disambiguation based on Wikipedia data. *EMNLP-CoNLL*, pp. 708–716, 2007.
- [122] A. Culotta and J. Sorensen. Dependency tree kernels for relation extraction. *ACL Conference*, 2004.
- [123] J. Curran and S. Clark. Language independent NER using a maximum entropy tagger. *Conference on Natural Language Learning at HLT-NAACL 2003*, pp. 164–167, 2003.
- [124] D. Cutting, D. Karger, J. Pedersen, and J. Tukey. Scatter/gather: A cluster-based approach to browsing large document collections. *ACM SIGIR Conference*, pp. 318–329, 1992.
- [125] W. Dai, Y. Chen, G. Xue, Q. Yang, and Y. Yu. Translated learning: Transfer learning across different feature spaces. *NIPS Conference*, pp. 353–360, 2008.
- [126] D. Das and A. Martins. A survey on automatic text summarization. *Literature Survey for the Language and Statistics II course at CMU*, 4, pp. 1–31, 2007.
- [127] S. Das and M. Chen. Yahoo! for Amazon: Extracting market sentiment from stock message boards. *Asia Pacific Finance Association Annual Conference (APFA)*, 2001.
- [128] J. Dean and S. Ghemawat. MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1), pp. 107–113, 2008.
- [129] G. DeJong. Prediction and substantiation: A new approach to natural language processing. *Cognitive Science*, 3(3), pp. 251–273, 1979.
- [130] H. Deng, B. Zhao, J. Han. Collective topic modeling for heterogeneous networks. *ACM SIGIR Conference*, pp. 1109–1110, 2011.
- [131] H. Deng, J. Han, B. Zhao, Y. Yu, and C. Lin. Probabilistic topic models with biased propagation on heterogeneous information networks. *ACM KDD Conference*, pp. 1271–1279, 2011.
- [132] I. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. *ACM KDD Conference*, pp. 269–274, 2001.
- [133] I. Dhillon and D. Modha. Concept decompositions for large sparse text data using clustering. *Machine Learning*, 42(1–2), pp. 143–175, 2001.
- [134] T. Dietterich. Machine learning for sequential data: A review. *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pp. 15–30, 2002.
- [135] C. Ding, X. He, and H. Simon. On the equivalence of nonnegative matrix factorization and spectral clustering. *SDM Conference*, pp. 606–610, 2005.
- [136] C. Ding, T. Li, and M. Jordan. Convex and semi-nonnegative matrix factorizations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1), pp. 45–55, 2010.
- [137] C. Ding, T. Li, and W. Peng. On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing. *Computational Statistics and Data Analysis*, 52(8), pp. 3913–3927, 2008.

- [138] C. Ding, T. Li, W. Peng, and H. Park. Orthogonal nonnegative matrix t-factorizations for clustering. *ACM KDD Conference*, pp. 126–135, 2006.
- [139] X. Ding, B. Liu, and P. S. Yu. A holistic lexicon-based approach to opinion mining. *WSDM Conference*, pp. 231–240, 2008.
- [140] P. Domingos and M. Pazzani. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29(2–3), pp. 103–130, 1997.
- [141] B. Dorr, D. Zajic, and R. Schwartz. Hedge Trimmer: A parse-and-trim approach to headline generation. *HLT-NAACL Workshop on Text Summarization*, pp. 1–8, 2003.
- [142] N. Draper and H. Smith. Applied regression analysis. *John Wiley & Sons*, 2014.
- [143] H. Drucker, C. Burges, L. Kaufman, A. Smola, and V. Vapnik. Support Vector Regression Machines. *NIPS Conference*, 1997.
- [144] R. Duda, P. Hart, W. Stork. *Pattern Classification*, Wiley Interscience, 2000.
- [145] S. Dumais. Latent semantic indexing (LSI) and TREC-2. *Text Retrieval Conference (TREC)*, pp. 105–115, 1993.
- [146] S. Dumais. Latent semantic indexing (LSI): TREC-3 Report. *Text Retrieval Conference (TREC)*, pp. 219–230, 1995.
- [147] S. Dumais, J. Platt, D. Heckerman, and M. Sahami. Inductive learning algorithms and representations for text categorization. *ACM CIKM Conference*, pp. 148–155, 1998.
- [148] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 41(6), pp. 391–407, 1990.
- [149] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3), pp. 211–218, 1936.
- [150] H. P. Edmundson. New methods in automatic extracting. *Journal of the ACM*, 16(2), pp. 264–286, 1969.
- [151] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2), pp. 407–499, 2004.
- [152] J. Eisenstein and R. Barzilay. Bayesian unsupervised topic segmentation. *Conference on Empirical Methods in Natural Language Processing*, pp. 334–343, 2008.
- [153] P. Elias. Universal codeword sets and representations of the integers. *IEEE Transactions on Information Theory*, 21(2), pp. 194–203, 1975.
- [154] G. Erkan and D. Radev. LexRank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22, pp. 457–479, 2004.
- [155] E. Erosheva, S. Fienberg, and J. Lafferty. Mixed-membership models of scientific publications. *Proceedings of the National Academy of Sciences*, 101, pp. 5220–5227, 2004.
- [156] A. Esuli, and F. Sebastiani. Determining the semantic orientation of terms through gloss classification. *ACM CIKM Conference*, pp. 617–624, 2005.

- [157] A. Esuli and F. Sebastiani. Determining term subjectivity and term orientation for opinion mining. *European Chapter of the Association of Computational Linguistics*, 2006.
- [158] O. Etzioni, M. Cafarella, D. Downey, A. Popescu, T. Shaked, S. Soderland, D. Weld, and A. Yates. Unsupervised named-entity extraction from the web: An experimental study. *Artificial Intelligence*, 165(1), pp. 91–134, 2005.
- [159] O. Etzioni, M. Banko, S. Soderland, and D. Weld. Open information extraction from the web. *Communications of the ACM*, 51(12), pp. 68–74, 2008.
- [160] A. Fader, S. Soderland, and O. Etzioni. Identifying relations for open information extraction. *Conference on Empirical Methods in Natural Language Processing*, pp. 1535–1545, 2011.
- [161] A. Fader, L. Zettlemoyer, and O. Etzioni. Paraphrase-Driven Learning for Open Question Answering. *ACL*, pp. 1608–1618, 2013.
- [162] A. Fader, L. Zettlemoyer, and O. Etzioni. Open question answering over curated and extracted knowledge bases. *ACM KDD Conference*, 2014.
- [163] C. Faloutsos and S. Christodoulakis. Signature files: An access method for documents and its analytical performance evaluation. *ACM Transactions on Information Systems*, 2(4), pp. 267–288, 1984.
- [164] R. Fan, K. Chang, C. Hsieh, X. Wang, and C. Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9, pp. 1871–1874, 2008. <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>
- [165] R. Fan, P. Chen, and C. Lin. Working set selection using second order information for training support vector machines. *Journal of Machine Learning Research*, 6, pp. 1889–1918, 2005.
- [166] T. Fawcett. ROC Graphs: Notes and Practical Considerations for Researchers. *Technical Report HPL-2003-4*, Palo Alto, CA, HP Laboratories, 2003.
- [167] R. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7: pp. 179–188, 1936.
- [168] R. Feldman and J. Sanger. The text mining handbook: advanced approaches in analyzing unstructured data. *Cambridge University Press*, 2007.
- [169] M. Fernandez-Delgado, E. Cernadas, S. Barro, and D. Amorim. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *The Journal of Machine Learning Research*, 15(1), pp. 3133–3181, 2014.
- [170] K. Filippova and M. Strube. Sentence fusion via dependency graph compression. *Conference on Empirical Methods in Natural Language Processing*, pp. 177–185, 2008.
- [171] D. Freitag and A. McCallum. Information extraction with HMMs and shrinkage. *AAAI-99 Workshop on Machine Learning for Information Extraction*, pp. 31–36, 1999.

- [172] C. Freudenthaler, L. Schmidt-Thieme, and S. Rendle. Factorization machines: Factorized polynomial regression models. *German-Polish Symposium on Data Analysis and Its Applications (GPSDAA)*, 2011. https://www.ismll.uni-hildesheim.de/pub/pdfs/FreudenthalerRendle_FactorizedPolynomialRegression.pdf
- [173] Y. Freund, and R. Schapire. A decision-theoretic generalization of online learning and application to boosting. *Computational Learning Theory*, pp. 23–37, 1995.
- [174] J. Friedman. Stochastic gradient boosting. *Computational Statistics and Data Analysis*, 38(4), pp. 367–378, 2002.
- [175] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics*, 28(2), pp. 337–407, 2000.
- [176] M. Fuentes, E. Alfonseca, and H. Rodriguez. Support Vector Machines for query-focused summarization trained and evaluated on Pyramid data. *ACL Conference*, pp. 57–60, 2007.
- [177] G. Fung and O. Mangasarian. Proximal support vector classifiers. *ACM KDD Conference*, pp. 77–86, 2001.
- [178] J. Fürnkranz and G. Widmer. Incremental reduced error pruning. *ICML Conference*, pp. 70–77, 1994.
- [179] M. Galley and K. McKeown. Lexicalized Markov grammars for sentence compression. *Human Language Technologies: The Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 180–187, 2007.
- [180] T. Gärtner. A survey of kernels for structured data. *ACM SIGKDD Explorations Newsletter*, 5(1), pp. 49–58, 2003.
- [181] Y. Goldberg. A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research (JAIR)*, 57, pp. 345–420, 2016.
- [182] Y. Goldberg and O. Levy. word2vec Explained: deriving Mikolov et al.’s negative-sampling word-embedding method. *arXiv:1402.3722*, 2014. <https://arxiv.org/abs/1402.3722>
- [183] I. Goodfellow, Y. Bengio, and A. Courville. Deep learning. *MIT Press*, 2016.
- [184] W. Greiff. A theory of term weighting based on exploratory data analysis. *ACM SIGIR Conference*, pp. 11–19, 1998.
- [185] E. Gaussier and C. Goutte. Relation between PLSA and NMF and implications. *ACM SIGIR Conference*, pp. 601–602, 2005.
- [186] L. Getoor, N. Friedman, D. Koller, and B. Taskar. Learning probabilistic models of link structure. *Journal of Machine Learning Research*, 3, pp. 679–707, 2002.
- [187] J. Ghosh and A. Acharya. Cluster ensembles: Theory and applications. *Data Clustering: Algorithms and Applications*, CRC Press, 2013.

- [188] D. Gillick, K. Riedhammer, B. Favre, and D. Hakkani-Tur. A global optimization framework for meeting summarization. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4769–4772, 2009.
- [189] S. Gilpin, T. Eliassi-Rad, and I. Davidson. Guided learning for role discovery (glrd): framework, algorithms, and applications. *ACM KDD Conference*, pp. 113–121, 2013.
- [190] M. Girolami and A. Kabán. On an equivalence between PLSI and LDA. *ACM SIGIR Conference*, pp. 433–434, 2003.
- [191] F. Girosi and T. Poggio. Networks and the best approximation property. *Biological Cybernetics*, 63(3), pp. 169–176, 1990.
- [192] Y. Gong and X. Liu. Generic text summarization using relevance measure and latent semantic analysis. *ACM SIGIR Conference*, pp. 19–25, 2001.
- [193] R. Grishman and B. Sundheim. Message Understanding Conference-6: A Brief History. *COLING*, pp. 466–471, 1996.
- [194] D. Grossman and O. Frieder. Information retrieval: Algorithms and heuristics, *Springer Science and Business Media*, 2012.
- [195] A. Graves. Supervised sequence labelling with recurrent neural networks *Springer*, 2012. <http://rd.springer.com/book/10.1007%2F978-3-642-24797-2>
- [196] A. Graves. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*, 2013. <https://arxiv.org/abs/1308.0850>
- [197] A. Graves, A. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. *Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6645–6649, 2013.
- [198] M. Gutmann and A. Hyvarinen. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. *AISTATS*, 1(2), pp. 6, 2010.
- [199] B. Hagedorn, M. Ciaramita, and J. Atserias. World knowledge in broad-coverage information filtering. *ACM SIGIR Conference*, 2007.
- [200] A. Haghighi and L. Vanderwende. Exploring content models for multi-document summarization. *Human Language Technologies*, pp. 362–370, 2009.
- [201] D. Hakkani-Tur and G. Tur. Statistical sentence extraction for information distillation. *Conference on Acoustics, Speech and Signal Processing*, 4, 2007.
- [202] E.-H. Han, G. Karypis, and V. Kumar. Text categorization using weighted-adjusted k -nearest neighbor classification, *PAKDD Conference*, 2001.
- [203] E.-H. Han and G. Karypis. Centroid-based document classification: Analysis and experimental results. *PKDD Conference*, 2000.
- [204] J. Han, M. Kamber, and J. Pei. Data mining: concepts and techniques. *Morgan Kaufmann*, 2011.
- [205] T. H. Haveliwala. Topic-sensitive pagerank. *World Wide Web Conference*, pp. 517–526, 2002.

- [206] T. Hastie, R. Tibshirani, and J. Friedman. The elements of statistical learning. *Springer*, 2009.
- [207] T. Hastie and R. Tibshirani. Discriminant adaptive nearest neighbor classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(6), pp. 607–616, 1996.
- [208] T. Hastie, R. Tibshirani, and M. Wainwright. Statistical learning with sparsity: the lasso and generalizations. *CRC Press*, 2015.
- [209] T. Hastie and R. Tibshirani. Generalized additive models. *CRC Press*, 1990.
- [210] V. Hatzivassiloglou, J. Klavans, M. Holcombe, R. Barzilay, M.-Y. Kan, and K. R. McKeown. SIMFINDER: A flexible clustering tool for summarization. *NAACL Workshop on Automatic Summarization*, pp. 41–49, 2001.
- [211] M. Hatzivassiloglou, and K. McKeown. Predicting the semantic orientation of adjectives. *European Chapter of the Association for Computational Linguistics*, pp. 174–181, 1997.
- [212] V. Hatzivassiloglou and J. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. *Conference on Computational Linguistics*, pp. 299–305, 2000.
- [213] M. Hearst. TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1), pp. 33–64, 1997.
- [214] D. Hiemstra. A linguistically motivated probabilistic model of information retrieval. *International Conference on Theory and Practice of Digital Libraries*, pp. 569–584, 1998.
- [215] M. Hearst. TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1), pp. 33–64, 1997.
- [216] S. Heinz and J. Zobel. Efficient single-pass index construction for text databases. *Journal of the American Society for Information Science and Technology*, 54(8), pp. 713–729, 2003.
- [217] G. Hinton. Connectionist learning procedures. *Artificial Intelligence*, 40(1–3), pp. 185–234, 1989.
- [218] G. Hinton and R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786), pp. 504–507, 2006.
- [219] G. Hirst and D. St-Onge. Lexical chains as representation of context for the detection and correction of malapropisms. In *WordNet: An Electronic Lexical Database and Some of its Applications*, MIT Press, 1998.
- [220] T. K. Ho. Random decision forests. *Third International Conference on Document Analysis and Recognition*, 1995. Extended version appears as “The random subspace method for constructing decision forests” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), pp. 832–844, 1998.
- [221] T. K. Ho. Nearest neighbors in random subspaces. *Lecture Notes in Computer Science*, Vol. 1451, pp. 640–648, *Proceedings of the Joint IAPR Workshops SSPR’98 and SPR’98*, 1998. <http://link.springer.com/chapter/10.1007/BFb0033288>

- [222] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8), pp. 1735–1785, 1997.
- [223] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, *A Field Guide to Dynamical Recurrent Neural Networks*, IEEE Press, 2001.
- [224] T. Hofmann. Probabilistic latent semantic indexing. *ACM SIGIR Conference*, pp. 50–57, 1999.
- [225] T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine learning*, 41(1–2), pp. 177–196, 2001.
- [226] K. Hornik and B. Grün. topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13), pp. 1–30, 2011.
- [227] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), pp. 359–366, 1989.
- [228] E. Hovy and C.-Y. Lin. Automated Text Summarization in SUMMARIST. in *Advances in Automatic Text Summarization*, pp. 82–94, 1999.
- [229] M. Hu and B. Liu. Mining opinion features in customer reviews. *AAAI*, pp. 755–760, 2004.
- [230] M. Hu and B. Liu. Mining and summarizing customer reviews. *ACM KDD Conference*, pp. 168–177, 2004.
- [231] A. Huang. Similarity measures for text document clustering. *Sixth New Zealand Computer Science Research Student Conference*, pp. 49–56, 2008.
- [232] M. Iyyer, J. Boyd-Graber, L. Claudino, R. Socher, and H. Daume III. A Neural Network for Factoid Question Answering over Paragraphs. *EMNLP*, 2014.
- [233] N. Jakob and I. Gurevych. Extracting opinion targets in a single-and cross-domain setting with conditional random fields. *Conference on Empirical Methods in Natural Language Processing*, pp. 1035–1045, 2010.
- [234] W. Jin, H. Ho, and R. Srihari. OpinionMiner: a novel machine learning system for Web opinion mining and extraction. *ACM KDD Conference*, pp. 1195–1204, 2009.
- [235] N. Jindal and B. Liu. Opinion spam and analysis. *WSDM Conference*, pp. 219–230, 2008.
- [236] J. Jiang. Information extraction from text. *Mining Text Data*, Springer, pp. 11–41, 2012.
- [237] J. Jiang and C. Zhai. A systematic exploration of the feature space for relation extraction. *HLT-NAACL*, pp. 113–120, 2007.
- [238] H. Jing. Sentence reduction for automatic text summarization. *Conference on Applied Natural Language Processing*, pp. 310–315, 2000.
- [239] H. Jing. Cut-and-paste text summarization. *PhD Thesis*, Columbia University, 2001. http://www1.cs.columbia.edu/nlp/theses/hongyan_jing.pdf

- [240] T. Joachims. Text categorization with support vector machines: learning with many relevant features. *ECML Conference*, 1998.
- [241] T. Joachims. Making Large scale SVMs practical. *Advances in Kernel Methods, Support Vector Learning*, pp. 169–184, MIT Press, Cambridge, 1998.
- [242] T. Joachims. Training Linear SVMs in Linear Time. *ACM KDD Conference*, pp. 217–226, 2006.
- [243] T. Joachims. A probabilistic analysis of the Rocchio algorithm with TFIDF for text categorization. *ICML Conference*, 1997.
- [244] T. Joachims. Optimizing search engines using clickthrough data. *ACM KDD Conference*, pp. 133–142, 2002.
- [245] C. Johnson. Logistic matrix factorization for implicit feedback data. *NIPS Conference*, 2014.
- [246] D. Johnson, F. Oles, T. Zhang, T. Goetz. A decision tree-based symbolic rule induction system for text categorization, *IBM Systems Journal*, 41(3), pp. 428–437, 2002.
- [247] I. T. Jolliffe. Principal component analysis. *John Wiley & Sons*, 2002.
- [248] I. T. Jolliffe. A note on the use of principal components in regression. *Applied Statistics*, 31(3), pp. 300–303, 1982. .
- [249] D. Jurafsky and J. Martin. Speech and language processing. *Prentice Hall*, 2008.
- [250] N. Kalchbrenner and P. Blunsom. Recurrent continuous translation models. *EMNLP*, 3, 39, pp. 413, 2013.
- [251] N. Kambhatla, Combining lexical, syntactic and semantic features with maximum entropy models for information extraction. *ACL Conference*, pp. 178–181, 2004.
- [252] J. Kamps, M. Marx, R. Mokken, and M. Rijke. Using wordnet to measure semantic orientations of adjectives. *LREC*, pp. 1115–1118, 2004.
- [253] R. Kannan, H. Woo, C. Aggarwal, and H. Park. Outlier detection for text data. *SDM Conference*, 2017.
- [254] H. Kanayama and T. Nasukawa. Fully automatic lexicon expansion for domain-oriented sentiment analysis. *Conference on Empirical Methods in Natural Language Processing*, pp. 355–363, 2006.
- [255] A. Karatzoglou, A. Smola, A. K. Hornik, and A. Zeileis. kernlab – An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, 11(9), 2004. <http://epub.wu.ac.at/1048/1/document.pdf> <http://CRAN.R-project.org/package=kernlab>
- [256] A. Karpathy, J. Johnson, and L. Fei-Fei. Visualizing and understanding recurrent networks. *arXiv preprint arXiv:1506.02078*, 2015. <https://arxiv.org/abs/1506.02078>
- [257] A. Karpathy. The unreasonable effectiveness of recurrent neural networks, *Blog post*, 2015. <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>
- [258] G. Karypis and E.-H. Han. Fast supervised dimensionality reduction with applications to document categorization and retrieval, *ACM CIKM Conference*, pp. 12–19, 2000.

- [259] H. Kim, K. Ganesan, P. Sondhi, and C. Zhai. Comprehensive Review of Opinion Summarization. *Technical Report*, University of Illinois at Urbana-Champaign, 2011. <https://www.ideals.illinois.edu/handle/2142/18702>
- [260] S. Kim and E. Hovy. Automatic identification of pro and con reasons in online reviews. *COLING/ACL Conference*, pp. 483–490, 2006.
- [261] Y. Kim. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*, 2014.
- [262] J. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5), pp. 604–632, 1999.
- [263] K. Knight and D. Marcu. Summarization beyond sentence extraction: A probabilistic approach to sentence compression. *Artificial Intelligence*, 139(1), pp. 91–107, 2002.
- [264] R. Kohavi and D. Wolpert. Bias plus variance decomposition for zero-one loss functions. *ICML Conference*, 1996.
- [265] E. Kong and T. Dietterich. Error-correcting output coding corrects bias and variance. *ICML Conference*, pp. 313–321, 1995.
- [266] A. Krogh, M. Brown, I. Mian, K. Sjolander, and D. Haussler. Hidden Markov models in computational biology: Applications to protein modeling. *Journal of Molecular Biology*, 235(5), pp. 1501–1531, 1994.
- [267] M. Kuhn. Building predictive models in R Using the caret Package. *Journal of Statistical Software*, 28(5), pp. 1–26, 2008. <https://cran.r-project.org/web/packages/caret/index.html>
- [268] J. Kupiec. Robust part-of-speech tagging using a hidden Markov model. *Computer Speech and Language*, 6(3), pp. 225–242, 1992.
- [269] J. Kupiec, J. Pedersen, and F. Chen. A trainable document summarizer. *ACM SIGIR Conference*, pp. 68–73, 1995.
- [270] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *ICML Conference*, pp. 282–289, 2001.
- [271] W. Lam and C. Y. Ho. Using a generalized instance set for automatic text categorization. *ACM SIGIR Conference*, 1998.
- [272] A. Langville, C. Meyer, R. Albright, J. Cox, and D. Duling. Initializations for the nonnegative matrix factorization. *ACM KDD Conference*, pp. 23–26, 2006.
- [273] J. Lau and T. Baldwin. An empirical evaluation of doc2vec with practical insights into document embedding generation. *arXiv:1607.05368*, 2016. <https://arxiv.org/abs/1607.05368>
- [274] Q. Le. Personal communication, 2017.
- [275] Q. Le and T. Mikolov. Distributed representations of sentences and documents. *ICML Conference*, pp. 1188–1196, 2014.

- [276] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, pp. 556–562, 2001.
- [277] D. Lee and H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755), pp. 788–791, 2001.
- [278] R. Lempel and S. Moran. Predictive caching and prefetching of query results in search engines. *World Wide Web Conference*, pp. 19–28, 2003.
- [279] J. Leskovec, N. Milic-Frayling, and M. Grobelnik. Impact of linguistic analysis on the semantic graph: coverage and learning of document extracts. *National Conference on Artificial Intelligence*, pp. 1069–1074, 2005.
- [280] J. Leskovec, A. Rajaraman, and J. Ullman. Mining of massive datasets. *Cambridge University Press*, 2012.
- [281] N. Lester, J. Zobel, and H. Williams. Efficient online index maintenance for contiguous inverted lists. *Information Processing and Management*, 42(4), pp. 916–933, 2006.
- [282] O. Levy and Y. Goldberg. Neural word embedding as implicit matrix factorization. *NIPS Conference*, pp. 2177–2185, 2014.
- [283] O. Levy, Y. Goldberg, and I. Dagan. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, 3, pp. 211–225, 2015.
- [284] O. Levy, Y. Goldberg, and I. Ramat-Gan. Linguistic regularities in sparse and explicit word representations. *CoNLL*, 2014.
- [285] D. Lewis. An evaluation of phrasal and clustered representations for the text categorization task. *ACM SIGIR Conference*, pp. 37–50, 1992.
- [286] D. Lewis. Naive (Bayes) at forty: The independence assumption in information retrieval. *ECML Conference*, pp. 4–15, 1998.
- [287] D. Lewis and M. Ringuette. A comparison of two learning algorithms for text categorization. *Third Annual Symposium on Document Analysis and Information Retrieval*, pp. 81–93, 1994.
- [288] F. Li, C. Han, M. Huang, X. Zhu, Y. Xia, S. Zhang, and H. Yu. Structure-aware review mining and summarization. *Conference on Computational Linguistics*, pp. 6563–661, 2010.
- [289] H. Li, and K. Yamanishi. Document classification using a finite mixture model. *ACL Conference*, pp. 39–47, 1997.
- [290] Y. Li and A. Jain. Classification of text documents. *The Computer Journal*, 41(8), pp. 537–546, 1998.
- [291] Y. Li, C. Luo, and S. Chung. Text clustering with feature selection by using statistical data. *IEEE Transactions on Knowledge and Data Engineering*, 20(5), pp. 641–652, 2008.

- [292] D. Liben-Nowell, and J. Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7), pp. 1019–1031, 2007.
- [293] E. Lim, V. Nguyen, N. Jindal, B. Liu, and H. Lauw. Detecting product review spammers using rating behaviors. *ACM CIKM Conference*, pp. 939–948, 2010.
- [294] C. Lin. Projected gradient methods for nonnegative matrix factorization. *Neural Computation*, 19(10), pp. 2756–2779, 2007.
- [295] C.-Y. Lin and E. Hovy. The automated acquisition of topic signatures for text summarization. *Conference on Computational linguistics*, pp. 495–501, 2000.
- [296] C.-Y. Lin and E. Hovy. From single to multi-document summarization: A prototype system and its evaluation. *ACL Conference*, pp. 457–464, 2002.
- [297] H. Lin and J. Bilmes. Multi-document summarization via budgeted maximization of submodular functions. *Human Language Technologies*, pp. 912–920, 2010.
- [298] X. Ling and D. Weld. Temporal information extraction. *AAAI*, pp. 1385–1390, 2010.
- [299] Z. Lipton, J. Berkowitz, and C. Elkan. A critical review of recurrent neural networks for sequence learning. *arXiv:1506.00019*, 2015. <https://arxiv.org/abs/1506.00019>
- [300] L. V. Lita, A. Ittycheriah, S. Roukos, and N. Kambhatla. Truecasing. *ACL Conference*, pp. 152–159, 2003.
- [301] D. Litman and R. Passonneau. Combining multiple knowledge sources for discourse segmentation. *Association for Computational Linguistics*, pp. 108–115, 1995.
- [302] C. Lin and Y. He. Joint sentiment/topic model for sentiment analysis. *ACM CIKM Conference*, pp. 375–384, 2009.
- [303] B. Liu. Web data mining: exploring hyperlinks, contents, and usage data. *Springer*, New York, 2007.
- [304] B. Liu. Sentiment Analysis and Subjectivity. *Handbook of Natural Language Processing*, 2, pp. 627–666, 2010.
- [305] B. Liu. Sentiment analysis: Mining opinions, sentiments, and emotions. *Cambridge University Press*, 2015.
- [306] B. Liu, W. Hsu, and Y. Ma. Integrating classification and association rule mining. *ACM KDD Conference*, pp. 80–86, 1998.
- [307] T.-Y. Liu. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, 3(3), pp. 225–231, 2009.
- [308] H. Lodhi, C. Saunders, J. Shawe-Taylor, N. Cristianini, and C. Watkins. Text classification using string kernels. *Journal of Machine Learning Research*, 2, pp. 419–444, 2002.
- [309] X. Long and T. Suel. Optimized query execution in large search engines with global page ordering. *VLDB Conference*, pp. 129–140, 2003.

- [310] Y. Lu, M. Castellanos, U. Dayal, and C. Zhai. Automatic construction of a context-aware sentiment lexicon: an optimization approach. *World Wide Web Conference*, pp. 347–356, 2011.
- [311] Y. Lu and C. Zhai. Opinion integration through semi-supervised topic modeling. *World Wide Web Conference*, pp. 121–130, 2008.
- [312] H. P. Luhn. The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2), pp. 159–165, 1958.
- [313] K. Lund and C. Burgess. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, 28(2). pp. 203–208, 1996.
- [314] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4), pp. 395–416, 2007.
- [315] A. Maas, R. Daly, P. Pham, D. Huang, A. Ng, and C. Potts. Learning word vectors for sentiment analysis. *Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pp. 142–150, 2011.
- [316] C. Mackenzie. Coded character sets: History and development. *Addison-Wesley Longman Publishing Co., Inc.*, 1980.
- [317] S. Madeira and A. Oliveira. Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 1(1), pp. 24–45, 2004.
- [318] R. Malouf. A comparison of algorithms for maximum entropy parameter estimation. *Conference on Natural Language Learning*, pp. 1–7, 2002.
- [319] O. Mangasarian and D. Musicant. Successive overrelaxation for support vector machines. *IEEE Transactions on Neural Networks*, 10(5), pp. 1032–1037, 1999.
- [320] I. Mani and G. Wilson. Robust temporal processing of news. *ACL Conference*, pp. 69–76, 2000.
- [321] C. Manning, P. Raghavan, and H. Schütze. Introduction to information retrieval. *Cambridge University Press*, Cambridge, 2008.
- [322] C. Manning and H. Schütze. Foundations of statistical natural language processing. *MIT Press*, 1999.
- [323] E. Marsi and E. Kraemer. Explorations in sentence fusion. *European Workshop on Natural Language Generation*, pp. 109–117, 2005.
- [324] J. Martens and I. Sutskever. Learning recurrent neural networks with hessian-free optimization. *ICML Conference*, pp. 1033–1040, 2011.
- [325] A. McCallum. Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering. <http://www.cs.cmu.edu/~mccallum/bow>, 1996.
- [326] A. McCallum, D. Freitag, and F. Pereira. Maximum entropy Markov models for information extraction and segmentation. *ICML Conference*, pp. 591–598, 2000.

- [327] A. McCallum and K. Nigam. A comparison of event models for naive Bayes text classification. *AAAI Workshop on Learning for Text Categorization*, 1998.
- [328] P. McCullagh and J. Nelder. Generalized linear models *CRC Press*, 1989.
- [329] R. McDonald, K. Hannan, T. Neylon, M. Wells, and J. Reynar. Structured models for fine-to-coarse sentiment analysis. *ACL Conference*, 2007.
- [330] G. McLachlan. Discriminant analysis and statistical pattern recognition *John Wiley & Sons*, 2004.
- [331] D. McClosky, M. Surdeanu, and C. Manning. Event extraction as dependency parsing. *Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pp. 1626–1635, 2011.
- [332] Q. Mei, D. Cai, D. Zhang, and C. Zhai. Topic modeling with network regularization. *World Wide Web Conference*, pp. 101–110, 2008.
- [333] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai. Topic sentiment mixture: modeling facets and opinions in weblogs. In *World Wide Web Conference*, pp. 171–180, 2007.
- [334] S. Melink, S. Raghavan, B. Yang, and H. Garcia-Molina. Building a distributed full-text index for the web. *ACM Transactions on Information Systems*, 19(3), pp. 217–241, 2001.
- [335] P. Melville, W. Gryc, and R. Lawrence. Sentiment analysis of blogs by combining lexical knowledge with text classification. *ACM KDD Conference*, pp. 1275–1284, 2009.
- [336] A. K. Menon, and C. Elkan. Link prediction via matrix factorization. *Machine Learning and Knowledge Discovery in Databases*, pp. 437–452, 2011.
- [337] D. Metzler, S. Dumais, and C. Meek. Similarity measures for short segments of text. *European Conference on Information Retrieval*, pp. 16–27, 2007.
- [338] L. Michelbacher, F. Laws, B. Dorow, U. Heid, and H. Schütze. Building a cross-lingual relatedness thesaurus using a graph similarity measure. *LREC*, 2010.
- [339] R. Mihalcea and P. Tarau. TextRank: Bringing order into texts. *Conference on Empirical Methods in Natural Language Processing*, pp. 404–411, 2004.
- [340] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K. Müller. Fisher discriminant analysis with kernels. *NIPS Conference*, 1999.
- [341] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv:1301.3781*, 2013. <https://arxiv.org/abs/1301.3781>
- [342] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. *NIPS Conference*, pp. 3111–3119, 2013.
- [343] T. Mikolov, M. Karafiat, L. Burget, J. Cernocky, and S. Khudanpur. Recurrent neural network based language model. *Interspeech*, Vol 2, 2010.

- [344] T. Mikolov, W. Yih, and G. Zweig. Linguistic Regularities in Continuous Space Word Representations. *HLT-NAACL*, pp. 746–751, 2013.
- [345] T. Mikolov, Q. Le, and I. Sutskever. Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*, 2013. <https://arxiv.org/abs/1309.4168>
- [346] D. Miller, T. Leek, and R. Schwartz. A Hidden Markov Model information retrieval system. *ACM SIGIR Conference*, pp. 214–221, 1999.
- [347] G. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography (special issue)*, 3(4), pp. 235–312, 1990. <https://wordnet.princeton.edu/>
- [348] M. Mintz, S. Bills, R. Snow, and D. Jurafsky. Distant supervision for relation extraction without labeled data. *Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*, pp. 1003–1011, 2009.
- [349] T. M. Mitchell. Machine learning. *McGraw Hill International Edition*, 1997.
- [350] T. M. Mitchell. The role of unlabeled data in supervised learning. *International Colloquium on Cognitive Science*, pp. 2–11, 1999.
- [351] A. Mnih and G. Hinton. Three new graphical models for statistical language modelling. *ICML Conference*, pp. 641–648, 2007.
- [352] A. Mnih and K. Kavukcuoglu. Learning word embeddings efficiently with noise-contrastive estimation. *NIPS Conference*, pp. 2265–2273, 2013.
- [353] A. Mnih and Y. Teh. A fast and simple algorithm for training neural probabilistic language models. *arXiv:1206.6426*, 2012. <https://arxiv.org/abs/1206.6426>
- [354] A. Moffat and J. Zobel. Self-indexing inverted files for fast text retrieval. *ACM Transactions on Information Systems*, 14(4), pp. 14(4), 1996.
- [355] F. Moosmann, B. Triggs, and F. Jurie. Fast Discriminative visual codebooks using randomized clustering forests. *NIPS Conference*, pp. 985–992, 2006.
- [356] T. Mullen and N. Collier. Sentiment analysis using support vector machines with diverse information sources. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 412–418, 2004.
- [357] J.-C. Na, H. Sui, C. Khoo, S. Chan, and Y. Zhou. Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews. *Conference of the International Society for Knowledge Organization (ISKO)*, pp. 49–54, 2004.
- [358] G. Nemhauser, L. Wolsey, and M. Fisher. An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14(1), pp. 265–294, 1978.
- [359] A. Nenkova and K. McKeown. Automatic Summarization *Foundations and Trends in Information Retrieval*, 5(2–3), pp. 103–233, 2011.

- [360] A. Nenkova and K. McKeown. A survey of text summarization techniques. *Mining Text Data*, Springer, pp. 43–76, 2012.
- [361] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. *NIPS Conference*, pp. 849–856, 2002.
- [362] T. Nguyen and A. Moschitti. End-to-end relation extraction using distant supervision from external semantic repositories. *ACL Conference*, pp. 277–282, 2011.
- [363] K. Nigam, J. Lafferty, and A. McCallum. Using maximum entropy for text classification. *IJCAI Workshop on Machine Learning for Information Filtering*, pp. 61–67, 1999.
- [364] K. Nigam, A. McCallum, S. Thrun, and T. Mitchell. Text classification with labeled and unlabeled data using EM. *Machine Learning*, 39(2), pp. 103–134, 2000.
- [365] H. Niitsuma and M. Lee. Word2Vec is a special case of kernel correspondence analysis and kernels for natural language processing, *arXiv preprint arXiv:1605.05087*, 2016. <https://arxiv.org/abs/1605.05087>
- [366] A. Ntoulas and J. Cho. Pruning policies for two-tiered inverted index with correctness guarantee. *ACM SIGIR Conference*, pp. 191–198, 2007.
- [367] M. Osborne. Using maximum entropy for sentence extraction. *ACL Workshop on Automatic Summarization*, pp. 1–8, 2002.
- [368] E. Osuna, R. Freund, and F. Girosi. Improved training algorithm for support vector machines, *IEEE Workshop on Neural Networks and Signal Processing*, 1997.
- [369] M. Ott, Y. Choi, C. Cardie, and J. Hancock. Finding deceptive opinion spam by any stretch of the imagination. *Association for Computational Linguistics: Human Language Technologies- Volume 1*, pp. 309–319, 2011.
- [370] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation engine: Bringing order to the web. *Technical Report*, 1999–0120, Computer Science Department, Stanford University, 1998.
- [371] C. D. Paice. Constructing literature abstracts by computer: techniques and prospects. *Information Processing and Management*, 26(1), pp. 171–186, 1990.
- [372] A. Pak and P. Paroubek. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *LREC*, pp. 1320–1326, 2010.
- [373] B. Pang and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *ACL Conference*, 2004.
- [374] B. Pang and L. Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *ACL Conference*, pp. 115–124, 2005.
- [375] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), pp. 1–135, 2008.
- [376] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 79–86, 2002.

- [377] R. Pascanu, T. Mikolov, and Y. Bengio. On the difficulty of training recurrent neural networks. *ICML*, (3), 28, pp. 1310–1318, 2013. <http://www.jmlr.org/proceedings/papers/v28/pascanu13.pdf>
- [378] M. Pazzani and D. Kibler. The utility of knowledge in inductive learning. *Machine Learning*, 9(1), pp. 57–94, 1992.
- [379] H. Paulheim and R. Meusel. A decomposition of the outlier detection problem into a set of supervised learning problems. *Machine Learning*, 100(2–3), pp. 509–531, 2015.
- [380] J. Pennington, R. Socher, and C. Manning. Glove: Global Vectors for Word Representation. *EMNLP*, pp. 1532–1543, 2014.
- [381] F. Pereira, N. Tishby, and L. Lee. Distributional clustering of English words. *ACL Conference*, pp. 183–190, 1993.
- [382] J. C. Platt. Sequential minimal optimization: A fast algorithm for training support vector machines. *Advances in Kernel Method: Support Vector Learning*, MIT Press, pp. 85–208, 1998.
- [383] J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, 10(3), pp. 61–74, 1999.
- [384] L. Polanyi and A. Zaenen. Contextual valence shifters. *Computing Attitude and Affect in Text: Theory and Applications*, pp. 1–10, Springer, 2006.
- [385] J. Ponte and W. Croft. A language modeling approach to information retrieval. *ACM SIGIR Conference*, pp. 275–281, 1998.
- [386] J. Ponte and W. Croft. Text segmentation by topic. *International Conference on Theory and Practice of Digital Libraries*, pp. 113–125, 1997.
- [387] A. Popescu and O. Etzioni. Extracting product features and opinions from reviews. *Natural Language Processing and Text Mining*, pp. 9–28, 2007.
- [388] J. Pritchard, M. Stephens, and P. Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2), pp. 945–959, 2000.
- [389] J. Pustejovsky *et al.* The timebank corpus. *Corpus Linguistics*, pp. 40, 2003.
- [390] J. Pustejovsky *et al.* TimeML: Robust specification of event and temporal expressions in text. *New Directions in Question Answering*, 3. pp. 28–34, 2003.
- [391] G. Qi, C. Aggarwal, and T. Huang. Towards semantic knowledge propagation from text corpus to web images. *WWW Conference*, pp. 297–306, 2011.
- [392] G. Qi, C. Aggarwal, and T. Huang. Community detection with edge content in social media networks. *ICDE Conference*, pp. 534–545, 2012.
- [393] L. Qian, G. Zhou, F. Kong, Q. Zhu, and P. Qian. Exploiting constituent dependencies for tree kernel-based semantic relation extraction. *International Conference on Computational Linguistics*, pp. 697–704, 2008.

- [394] G. Qiu, B. Liu, J. Bu, and C. Chen. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1), pp. 9–27, 2011.
- [395] J. Quinlan. C4.5: programs for machine learning. *Morgan-Kaufmann Publishers*, 1993.
- [396] J. Quinlan. Induction of decision trees. *Machine Learning*, 1, pp. 81–106, 1986.
- [397] L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), pp. 257–286, 1989.
- [398] D. Radev, H. Jing, and M. Budzikowska. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. *NAACL-ANLP Workshop on Automatic summarization*, pp. 21–30, 2000.
- [399] D. Radev, H. Jing, M. Stys, and D. Tam. Centroid-based summarization of multiple documents. *Information Processing and Management*, 40(6), pp. 919–938, 2004.
- [400] A. Ratnaparkhi. A maximum entropy model for part-of-speech tagging. *Conference on Empirical Methods in Natural Language Processing*, pp. 133–142, 1996.
- [401] R. Rehurek and P. Sojka. Software framework for topic modelling with large corpora. *LREC 2010 Workshop on New Challenges for NLP Frameworks*, pp. 45–50, 2010. <https://radimrehurek.com/gensim/index.html>
- [402] X. Ren, M. Jiang, J. Shang, and J. Han. Constructing Structured Information Networks from Massive Text Corpora (Tutorial), *WWW Conference*, 2017.
- [403] S. Rendle. Factorization machines. *IEEE ICDM Conference*, pp. 995–1000, 2010.
- [404] S. Rendle. Factorization machines with libfm. *ACM Transactions on Intelligent Systems and Technology*, 3(3), 57, 2012.
- [405] B. Ribeiro-Neto, E. Moura, M. Neubert, and N. Ziviani. Efficient distributed algorithms to build inverted files. *ACM SIGIR Conference*, pp. 105–112, 1999.
- [406] M. Richardson, A. Prakash, and E. Brill. Beyond PageRank: machine learning for static ranking. *World Wide Web Conference*, pp. 707–715, 2006.
- [407] R. Rifkin. Everything old is new again: a fresh look at historical approaches in machine learning. *Ph.D. Thesis*, Massachusetts Institute of Technology, 2002. <http://cbcl.mit.edu/projects/cbcl/publications/theses/thesis-rifkin.pdf>
- [408] A. Ritter, Mausam, O. Etzioni, and S. Clark. Open domain event extraction from twitter. *ACM KDD Conference*, pp. 1104–1102, 2012.
- [409] A. Ritter, S. Clark, Mausam, and O. Etzioni. Named entity recognition in tweets: an experimental study. *Conference on Empirical Methods in Natural Language Processing*, pp. 1524–1534, 2011.
- [410] S. Robertson. Understanding inverse document frequency: On theoretical arguments for IDF. *Journal of Documentation*, 60, pp. 503–520, 2004.
- [411] S. Robertson and K. Spärck Jones. Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27(3), pp. 129–146, 1976.

- [412] S. Robertson, H. Zaragoza, and M. Taylor. Simple BM25 extension to multiple weighted fields. *ACM CIKM Conference*, pp. 42–49, 2004.
- [413] J. Rodríguez, L. Kuncheva, and C. Alonso. Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10), pp. 1619–1630, 2006.
- [414] J. Rocchio. Relevance feedback information retrieval. *The Smart Retrieval System-Experiments in Automatic Document Processing*, G. Salton, Ed. Prentice Hall, Englewood Cliffs, NJ, pp. 313–323, 1971.
- [415] X. Rong. word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738*, 2014. <https://arxiv.org/abs/1411.2738>
- [416] B. Rosenfeld and R. Feldman. Clustering for unsupervised relation identification. *ACM CIKM Conference*, pp. 411–418, 2007.
- [417] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290, no. 5500, pp. 2323–2326, 2000.
- [418] M. Sahami and T. D. Heilman. A Web-based kernel function for measuring the similarity of short text snippets. *WWW Conference*, pp. 377–386, 2006.
- [419] T. Sakai and K. Spärck Jones. Generic summaries for indexing in information retrieval. *ACM SIGIR Conference*, pp. 190–198, 2001.
- [420] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes Twitter users: real-time event detection by social sensors. *World Wide Web Conference*, pp. 851–860, 2010.
- [421] G. Salton and J. Allan. Selective text utilization and text traversal. *Proceedings of ACM Hypertext*, 1993.
- [422] G. Salton, J. Allan, and C. Buckley. Approaches to passage retrieval in full text information systems. *ACM SIGIR Conference*, pp. 49–58, 1997.
- [423] G. Salton and C. Buckley. Term weighting approaches in automatic text retrieval, *Technical Report 87-881*, Cornell University, 1987. <https://ecommons.cornell.edu/bitstream/handle/1813/6721/87-881.pdf?sequence=1>
- [424] G. Salton and M. J. McGill. Introduction to modern information retrieval. *McGraw Hill*, 1986.
- [425] G. Salton, A. Singhal, M. Mitra, and C. Buckley. Automatic text structuring and summarization. *Information Processing and Management*, 33(2), pp. 193–207, 1997.
- [426] G. Salton, A. Wong, and C. Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11), pp. 613–620, 1975.
- [427] H. Samet. Foundations of multidimensional and metric data structures. *Morgan Kaufmann*, 2006.
- [428] R. Samworth. Optimal weighted nearest neighbour classifiers. *The Annals of Statistics*, 40(5), pp. 2733–2763, 2012.

- [429] P. Saraiva, E. Silva de Moura, N. Ziviani, W. Meira, R. Fonseca, and B. Ribério-Neto. Rank-preserving two-level caching for scalable search engines. *ACM SIGIR Conference*, pp. 51–58, 2001.
- [430] S. Sarawagi. Information extraction. *Foundations and Trends in Satabases*, 1(3), pp. 261–377, 2008.
- [431] S. Sarawagi and W. Cohen. Semi-markov conditional random fields for information extraction. *NIPS Conference*, pp. 1185–1192, 2004.
- [432] S. Sathe and C. Aggarwal. Similarity forests. *ACM KDD Conference*, 2017.
- [433] R. Sauri, R. Knippen, M. Verhagen, and J. Pustejovsky. Evita: a robust event recognizer for QA systems. *Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 700–707, 2005.
- [434] H. Sayyadi, M. Hurst, and A. Maykov. Event detection and tracking in social streams. *ICWSM Conference*, 2009.
- [435] F. Scholer, H. Williams, J. Yiannis, and J. Zobel. Compression of inverted indexes for fast query evaluation. *ACM SIGIR Conference*, pp. 222–229, 2002.
- [436] B. Schölkopf, A. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5), pp. 1299–1319, 1998.
- [437] M. Schuster and K. Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), pp. 2673–2681, 1997.
- [438] H. Schütze and C. Silverstein. Projections for Efficient Document Clustering. *ACM SIGIR Conference*, pp. 74–81, 1997.
- [439] F. Sebastiani. Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34(1), 2002.
- [440] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Galligher, and T. Eliassi-Rad. Collective classification in network data. *AI magazine*, 29(3), pp. 93, 2008.
- [441] G. Seni and J. Elder. Ensemble methods in data mining: Improving accuracy through combining predictions. *Synthesis Lectures in Data Mining and Knowledge Discovery*, Morgan and Claypool, 2010.
- [442] K. Seymore, A. McCallum, and R. Rosenfeld. Learning hidden Markov model structure for information extraction. *AAAI-99 Workshop on Machine Learning for Information Extraction*, pp. 37–42, 1999.
- [443] F. Shahnaz, M. Berry, V. Pauca, and R. Plemmons. Document clustering using nonnegative matrix factorization. *Information Processing and Management*, 42(2), pp. 378–386, 2006.
- [444] S. Shalev-Shwartz, Y. Singer, N. Srebro, and A. Cotter. Pegasos: Primal estimated sub-gradient solver for SVM. *Mathematical Programming*, 127(1), pp. 3–30, 2011.
- [445] A. Shashua. On the equivalence between the support vector machine for classification and sparsified Fisher’s linear discriminant. *Neural Processing Letters*, 9(2), pp. 129–139, 1999.

- [446] Y. Shinyama and S. Sekine. Preemptive information extraction using unrestricted relation discovery. *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pp. 304–311, 2006.
- [447] S. Siencnik. Adapting word2vec to named entity recognition. *Nordic Conference of Computational Linguistics, NODALIDA*, 2015.
- [448] A. Singh and G. Gordon. A unified view of matrix factorization models. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 358–373, 2008.
- [449] A. Siddharthan, A. Nenkova, and K. Mckeown. Syntactic simplification for improving content selection in multi-document summarization. *International Conference on Computational Linguistic*, pp. 896–902, 2004.
- [450] A. Singhal, C. Buckley, and M. Mitra. Pivoted document length normalization. *ACM SIGIR Conference*, pp. 21–29, 1996.
- [451] N. Slonim and N. Tishby. The power of word clusters for text classification. *European Colloquium on Information Retrieval Research (ECIR)*, 2001.
- [452] S. Soderland. Learning information extraction rules for semi-structured and free text. *Machine Learning*, 34(1–3), pp. 233–272, 1999.
- [453] K. Spärck Jones. A statistical interpretation of term specificity and its application in information retrieval. *Journal of Documentation*, 28(1), pp. 11–21, 1972.
- [454] K. Spärck Jones. Automatic summarizing: factors and directions. *Advances in Automatic Text Summarization*, pp. 1–12, 1998.
- [455] K. Spärck Jones. Automatic summarising: The state of the art. *Information Processing and Management*, 43(6), pp. 1449–1481, 2007.
- [456] K. Spärck Jones, S. Walker, and S. Robertson. A probabilistic model of information retrieval: development and comparative experiments: Part 2. *Information Processing and Management*, 36(6), pp. 809–840, 2000.
- [457] M. Stairmand. A computational analysis of lexical cohesion with applications in information retrieval. *Ph.D. Dissertation*, Center for Computational Linguistics UMIST, Manchester, 1996. <http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.503546>
- [458] J. Steinberger and K. Jezek. Using latent semantic analysis in text summarization and summary evaluation. *ISIM*, pp. 93–100, 2004.
- [459] J. Steinberger, M. Poesio, M. Kabadjov, and K. Jezek. Two uses of anaphora resolution in summarization. *Information Processing and Management*, 43(6), pp. 1663–1680, 2007.
- [460] G. Strang. An introduction to linear algebra. *Wellesley Cambridge Press*, 2009.
- [461] A. Strehl, J. Ghosh, and R. Mooney. Impact of similarity measures on web-page clustering. *Workshop on Artificial Intelligence for Web Search*, 2000. <http://www.aaai.org/Papers/Workshops/2000/WS-00-01/WS00-01-011.pdf>

- [462] Y. Sun, J. Han, J. Gao, and Y. Yu. itopicmodel: Information network-integrated topic modeling. *IEEE ICDM Conference*, pp. 493–502, 2011.
- [463] M. Sundermeyer, R. Schluter, and H. Ney. LSTM neural networks for language modeling. *Interspeech*, 2010.
- [464] I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. *NIPS Conference*, pp. 3104–3112, 2014.
- [465] C. Sutton and A. McCallum. An introduction to conditional random fields. *arXiv preprint*, arXiv:1011.4088, 2010. <https://arxiv.org/abs/1011.4088>
- [466] J. Suykens and J. Venderwalle. Least squares support vector machine classifiers. *Neural Processing Letters*, 1999.
- [467] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2), pp. 267–307, 2011.
- [468] K. Takeuchi and N. Collier. Use of support vector machines in extended named entity recognition. *Conference on Natural Language Learning*, pp. 1–7, 2002.
- [469] P.-N Tan, M. Steinbach, and V. Kumar. Introduction to data mining. *Addison-Wesley*, 2005.
- [470] J. Tang, Y. Chang, C. Aggarwal, and H. Liu. A survey of signed network mining in social media. *ACM Computing Surveys (CSUR)*, 49(3), 42, 2016.
- [471] J. Tang, S. Chang, C. Aggarwal, and H. Liu. (2015, February). Negative link prediction in social media. *WSDM Conference*, pp. 87–96, 2015.
- [472] M. Taylor, H. Zaragoza, N. Craswell, S. Robertson, and C. Burges. Optimisation methods for ranking functions with multiple parameters. *ACM CIKM Conference*, pp. 585–593, 2006.
- [473] J. Tenenbaum, V. De Silva, and J. Langford. A global geometric framework for non-linear dimensionality reduction. *Science*, 290 (5500), pp. 2319–2323, 2000.
- [474] A. Tikhonov and V. Arsenin. Solution of ill-posed problems. *Winston and Sons*, 1977.
- [475] M. Tsai, C. Aggarwal, and T. Huang. Ranking in heterogeneous social media. *WSDM Conference*, pp. 613–622, 2014.
- [476] J. Turner and E. Charniak. Supervised and unsupervised learning for sentence compression. *ACL Conference*, pp. 290–297, 2005.
- [477] P. Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. *ACL Conference*, pp. 417–424, 2002.
- [478] P. Turney and M. Littman. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21(4), pp. 314–346, 2003.
- [479] P. Turney and P. Pantel. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37(1), pp. 141–188, 2010.

- [480] C. J. van Rijsbergen. Information retrieval. *Butterworths*, London, 1979.
- [481] C.J. van Rijsbergen, S.E. Robertson, and M.F. Porter. New models in probabilistic information retrieval. *London: British Library. (British Library Research and Development Report, no. 5587)*, 1980. <https://tartarus.org/martin/PorterStemmer/>
- [482] V. Vapnik. The nature of statistical learning theory. *Springer*, 2000.
- [483] L. Vanderwende, H. Suzuki, C. Brockett, and A. Nenkova. Beyond SumBasic: Task-focused summarization with sentence simplification and lexical expansion. *Information Processing and Management*, 43, pp. 1606–1618, 2007
- [484] A. Vinokourov, N. Cristianini, and J. Shawe-Taylor. Inferring a semantic representation of text via cross-language correlation analysis. *NIPS Conference*, pp. 1473–1480, 2002.
- [485] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. Show and tell: A neural image caption generator. *CVPR Conference*, pp. 3156–3164, 2015.
- [486] E. Voorhees. Implementing agglomerative hierarchic clustering algorithms for use in document retrieval. *Information Processing and Management*, 22(6), pp. 465–476, 1986.
- [487] G. Wahba. Support vector machines, reproducing kernel Hilbert spaces and the randomized GACV. *Advances in Kernel Methods-Support Vector Learning*, 6, pp. 69–87, 1999.
- [488] H. Wallach, D. Mimno, and A. McCallum. Rethinking LDA: Why priors matter. *NIPS Conference*, pp. 1973–1981, 2009.
- [489] D. Wang, S. Zhu, T. Li, and Y. Gong. Multi-document summarization using sentence-based topic models. *ACL-IJCNLP Conference*, pp. 297–300, 2009.
- [490] H. Wang, H. Huang, F. Nie, and C. Ding. Cross-language Web page classification via dual knowledge transfer using nonnegative matrix tri-factorization. *ACM SIGIR Conference*, pp. 933–942, 2011.
- [491] S. Weiss, N. Indurkha, and T. Zhang. Fundamentals of predictive text mining. *Springer*, 2015.
- [492] S. Weiss, C. Apte, F. Damerau, D. Johnson, F. Oles, T. Goetz, and T. Hampp. Maximizing text-mining performance. *IEEE Intelligent Systems*, 14(4), pp. 63–69, 1999.
- [493] X. Wei and W. B. Croft. LDA-based document models for ad-hoc retrieval. *ACM SIGIR Conference*, pp. 178–185, 2006.
- [494] J. Weston, A. Bordes, S. Chopra, A. Rush, B. van Merriënboer, A. Joulin, and T. Mikolov. Towards ai-complete question answering: A set of pre-requisite toy tasks. *arXiv preprint arXiv:1502.05698*, 2015. <https://arxiv.org/abs/1502.05698>
- [495] J. Weston, S. Chopra, and A. Bordes. Memory networks. *ICLR*, 2015.

- [496] J. Weston and C. Watkins. Multi-class support vector machines. *Technical Report CSD-TR-98-04*, Department of Computer Science, Royal Holloway, University of London, May, 1998.
- [497] B. Widrow and M. Hoff. Adaptive switching circuits. *IRE WESCON Convention Record*, 4(1), pp. 96–104, 1960.
- [498] W. Wilbur and K. Sirotkin. The automatic identification of stop words. *Journal of Information Science*, 18(1), pp. 45–55, 1992.
- [499] J. Wiebe, R. Bruce, and T. O’Hara. Development and use of a gold-standard data set for subjectivity classifications. *Association for Computational Linguistics on Computational Linguistics*, pp. 246–253, 1999.
- [500] J. Wiebe and E. Riloff. Creating subjective and objective sentence classifiers from unannotated texts. *International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 486–497, 2005.
- [501] C. Williams and M. Seeger. Using the Nyström method to speed up kernel machines. *NIPS Conference*, 2000.
- [502] H. Williams, J. Zobel, and D. Bahle. Fast phrase querying with combined indexes. *ACM Transactions on Information Systems*, 22(4), pp. 573–594, 2004.
- [503] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. *Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 347–354, 2005.
- [504] T. Wilson, J. Wiebe, and R. Hwa. Just how mad are you? Finding strong and weak opinion clauses. *Computational Intelligence*, 22(2), pp. 73–99, 2006.
- [505] M. J. Witbrock and V. O. Mittal. Ultra-summarization: A statistical approach to generating highly condensed non-extractive summaries. *ACM SIGIR Conference*, pp. 315–316, 1999.
- [506] I. H. Witten, A. Moffat, and T. C. Bell. Managing Gigabytes: Compressing and indexing documents and images. *Morgan Kaufmann*, 1999.
- [507] K. Wong, M. Wu, and W. Li. Extractive summarization using supervised and semi-supervised learning. *International Conference on Computational Linguistics*, pp. 985–992, 2008.
- [508] W. Xu, X. Liu, and Y. Gong. Document clustering based on non-negative matrix factorization. *ACM SIGIR Conference*, pp. 267–273, 2003.
- [509] Z. Wu, C. Aggarwal, and J. Sun. The troll-trust model for ranking in signed networks. *WSDM Conference*, pp. 447–456, 2016.
- [510] J. Xu and H. Li. Adarank: a boosting algorithm for information retrieval. *ACM SIGIR Conference*, 2007.
- [511] J. Yamron, I. Carp, L. Gillick, S. Lowe, and P. van Mulbregt. A hidden Markov model approach to text segmentation and event tracking. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 333–336, 1998.

- [512] J. Yang, J. McAuley, and J. Leskovec. Community detection in networks with node attributes. *IEEE ICDM Conference*, pp. 1151–1156, 2013.
- [513] Q. Yang, Q., Y. Chen, G. Xue, W. Dai, and T. Yu. Heterogeneous transfer learning for image clustering via the social web. *Joint Conference of the ACL and Natural Language Processing of the AFNLP*, pp. 1–9, 2009.
- [514] T. Yang, R. Jin, Y. Chi, and S. Zhu. Combining link and content for community detection: a discriminative approach. *ACM KDD Conference*, pp. 927–936, 2009.
- [515] Y. Yang. Noise reduction in a statistical approach to text categorization, *ACM SIGIR Conference*, pp. 256–263, 1995.
- [516] Y. Yang. An evaluation of statistical approaches to text categorization. *Information Retrieval*, 1(1–2), pp. 69–90, 1999.
- [517] Y. Yang. A study on thresholding strategies for text categorization. *ACM SIGIR Conference*, pp. 137–145, 2001.
- [518] Y. Yang and C. Chute. An application of least squares fit mapping to text information retrieval. *ACM SIGIR Conference*, pp. 281–290, 1993.
- [519] Y. Yang and X. Liu. A re-examination of text categorization methods. *ACM SIGIR Conference*, pp. 42–49, 1999.
- [520] Y. Yang and J. O. Pederson. A comparative study on feature selection in text categorization, *ACM SIGIR Conference*, pp. 412–420, 1995.
- [521] Y. Yang, T. Pierce, and J. Carbonell. A study of retrospective and online event detection. *ACM SIGIR Conference*, pp. 28–36, 1998.
- [522] Y. Yue, T. Finley, F. Radlinski, and T. Joachims. A support vector method for optimizing average precision. *ACM SIGIR Conference*, pp. 271–278, 2007.
- [523] D. Zajic, B. Dorr, J. Lin, and R. Schwartz. Multi-candidate reduction: Sentence compression as a tool for document summarization tasks. *Information Processing and Management*, 43(6), pp. 1549–1570, 2007.
- [524] M. Zaki and W. Meira Jr. Data mining and analysis: Fundamental concepts and algorithms. *Cambridge University Press*, 2014.
- [525] O. Zamir and O. Etzioni. Web document clustering: A feasibility demonstration. *ACM SIGIR Conference*, pp. 46–54, 1998.
- [526] D. Zelenko, C. Aone, and A. Richardella. Kernel methods for relation extraction. *Journal of Machine Learning Research*, 3. pp. 1083–1106, 2003.
- [527] C. Zhai. Statistical language models for information retrieval. *Synthesis Lectures on Human Language Technologies*, 1(1), pp. 1–141, 2008.
- [528] C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to information retrieval. *ACM Transactions on Information Systems*, 22(2), pp. 179–214, 2004.

- [529] C. Zhai and S. Massung. Text data management and mining: A practical introduction to information retrieval and text mining. *Association of Computing Machinery/Morgan and Claypool Publishers*, 2016.
- [530] Y. Zhai and B. Liu. Web data extraction based on partial tree alignment. *World Wide Web Conference*, pp. 76–85, 2005.
- [531] J. Zhang, Z. Ghahramani, and Y. Yang. A probabilistic model for online document clustering with application to novelty detection. *NIPS Conference*, pp. 1617–1624, 2004.
- [532] J. Zhang, X. Long, and T. Suel. Performance of compressed inverted list caching in search engines. *World Wide Web Conference*, pp. 387–396, 2008.
- [533] M. Zhang, J. Zhang, and J. Su. Exploring syntactic features for relation extraction using a convolution tree kernel. *Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 288–295, 2006.
- [534] M. Zhang, J. Zhang, J. Su, and G. Zhou. A composite kernel to extract relations between entities with both flat and structured features. *International Conference on Computational Linguistics and the Annual Meeting of the Association for Computational Linguistics*, pp. 825–832, 2006.
- [535] S. Zhao and R. Grishman. Extracting relations with integrated information using kernel methods. *ACL Conference*, pp. 419–426, 2005.
- [536] Y. Zhao, G. Karypis. Empirical and theoretical comparisons of selected criterion functions for document clustering. *Machine Learning*, 55(3), pp. 311–331, 2004.
- [537] S. Zhong. Efficient streaming text clustering. *Neural Networks*, Vol. 18, 5–6, 2005.
- [538] Y. Zhou, H. Cheng, and J. X. Yu. Graph clustering based on structural/attribute similarities. *Proceedings of the VLDB Endowment*, 2(1), pp. 718–729, 2009.
- [539] Z.-H. Zhou. Ensemble methods: Foundations and algorithms. *CRC Press*, 2012.
- [540] Y. Zhu, Y. Chen, Z. Lu, S. J. Pan, G. Xue, Y. Yu, and Q. Yang. Heterogeneous transfer learning for image classification. *AAAI Conference*, 2011.
- [541] L. Zhuang, F. Jing, and X. Zhu. Movie review mining and summarization. *ACM CIKM Conference*, pp. 43–50, 2006.
- [542] J. Zobel and P. Dart. Finding approximate matches in large lexicons. *Software: Practice and Experience*, 25(3), pp. 331–345, 1995.
- [543] J. Zobel and P. Dart. Phonetic string matching: Lessons from information retrieval. *ACM SIGIR Conference*, pp. 166–172, 1996.
- [544] J. Zobel, A. Moffat, and K. Ramamohanarao. Inverted files versus signature files for text indexing. *ACM Transactions on Database Systems*, 23(4), pp. 453–490, 1998.
- [545] J. Zobel and A. Moffat. Inverted files for text search engines. *ACM Computing Surveys (CSUR)*, 38(2), 6, 2006.

- [546] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Stat. Methodology)*, 67(2), pp. 301–320, 2005.
- [547] <http://snowballstem.org/>
- [548] <http://opennlp.apache.org/index.html>
- [549] <https://archive.ics.uci.edu/ml/datasets.html>
- [550] http://scikit-learn.org/stable/tutorial/text_analytics/working_with_text_data.html
- [551] <https://cran.r-project.org/web/packages/tm/>
- [552] <https://www.ibm.com/developerworks/community/blogs/nlp/entry/tokenization?lang=en>
- [553] <http://www.cs.waikato.ac.nz/ml/weka/>
- [554] <http://nlp.stanford.edu/software/>
- [555] <http://nlp.stanford.edu/links/statnlp.html>
- [556] <http://www.nltk.org/>
- [557] <https://cran.r-project.org/web/packages/lsa/index.html>
- [558] <http://scikit-learn.org/stable/modules/generated/sklearn.decomposition.TruncatedSVD.html>
- [559] <http://weka.sourceforge.net/doc.stable/weka/attributeSelection/LatentSemanticAnalysis.html>
- [560] <http://scikit-learn.org/stable/modules/generated/sklearn.decomposition.NMF.html>
- [561] <http://scikit-learn.org/stable/modules/generated/sklearn.decomposition.LatentDirichletAllocation.html>
- [562] <https://cran.r-project.org/>
- [563] <http://www.cs.princeton.edu/~blei/lda-c/>
- [564] <http://scikit-learn.org/stable/modules/manifold.html>
- [565] <https://code.google.com/archive/p/word2vec/>
- [566] <https://www.tensorflow.org/tutorials/word2vec/>
- [567] <http://www.netlib.org/svdpack>
- [568] http://scikit-learn.org/stable/modules/kernel_approximation.html
- [569] http://scikit-learn.org/stable/auto_examples/text/document_clustering.html
- [570] <https://www.mathworks.com/help/stats/cluster-analysis.html>
- [571] <https://cran.r-project.org/web/packages/RTextTools/RTextTools.pdf>
- [572] <https://cran.r-project.org/web/packages/rotationForest/index.html>

- [573] <http://trec.nist.gov/data.html>
- [574] <http://research.nii.ac.jp/ntcir/data/data-en.html>
- [575] <http://www.clef-initiative.eu/home>
- [576] <https://archive.ics.uci.edu/ml/datasets/Twenty+Newsgroups>
- [577] <https://archive.ics.uci.edu/ml/datasets/Reuters-21578+Text+Categorization+Collection>
- [578] <http://www.daviddlewis.com/resources/testcollections/rcv1/>
- [579] <http://labs.europeana.eu/data>
- [580] <http://www.icwsm.org/2009/data/>
- [581] <https://www.csie.ntu.edu.tw/~cjlin/libmf/>
- [582] <http://www.lemurproject.org>
- [583] <https://nutch.apache.org/>
- [584] <https://scrapy.org/>
- [585] <https://webarchive.jira.com/wiki/display/Heritrix>
- [586] <http://www.dataparksearch.org/>
- [587] <http://lucene.apache.org/core/>
- [588] <http://lucene.apache.org/solr/>
- [589] <http://sphinxsearch.com/>
- [590] <https://snap.stanford.edu/snap/description.html>
- [591] <https://catalog.ldc.upenn.edu/LDC93T3A>
- [592] <http://www.berouge.com/Pages/default.aspx>
- [593] <https://code.google.com/archive/p/icsisumm/>
- [594] <http://finzi.psych.upenn.edu/library/LSAfun/html/genericSummary.html>
- [595] <https://github.com/tensorflow/models/tree/master/textsum>
- [596] <http://www.summarization.com/mead/>
- [597] <http://www.scs.leeds.ac.uk/amalgam/tagsets/brown.html>
- [598] https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html
- [599] <http://www.itl.nist.gov/iad/mig/tests/ace>
- [600] <http://www.biocreative.org>
- [601] <http://www.signll.org/conll>

- [602] <http://reverb.cs.washington.edu/>
- [603] <http://knowitall.github.io/ollie/>
- [604] <http://nlp.stanford.edu/software/openie.html>
- [605] <http://mallet.cs.umass.edu/>
- [606] <https://github.com/dbpedia-spotlight/dbpedia-spotlight/wiki>
- [607] <http://crf.sourceforge.net/>
- [608] <https://en.wikipedia.org/wiki/ClearForest>
- [609] <http://clic.cimec.unitn.it/composes/toolkit/>
- [610] <https://github.com/stanfordnlp/GloVe>
- [611] <https://deeplearning4j.org/>
- [612] <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>
- [613] <http://www.nltk.org/api/nltk.tokenize.html#nltk.tokenize.texttiling.TextTilingTokenizer>
- [614] <http://www.itl.nist.gov/iad/mig/tests/tdt/>
- [615] <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [616] <http://deeplearning.net/tutorial/lstm.html>
- [617] <http://machinelearningmastery.com/sequence-classification-lstm-recurrent-neural-networks-python-keras/>
- [618] <https://deeplearning4j.org/lstm>
- [619] <https://github.com/karpathy/char-rnn>
- [620] <https://arxiv.org/abs/1609.08144>

Index

Symbols

L_1 -Regularization, 169
 L_2 -Regularization, 165
 χ^2 -statistic, 120
 k -Means, 88
 n -Grams, 62, 308

A

Abstractive Summarization, 361, 362, 377
Accumulators, 268
AdaBoost, 220
Apache OpenNLP, 20
ASCII, 18
Associative Classifiers, 151
Authorities, 300

B

Backpropagation through Time, 344
Bag-of-Words Kernel, 62
Bag-of-Words Model, 24
Bagging, 138
Bayes optimal error rate, 134
Bernoulli Model (Classification), 123
Bernoulli Model (Clustering), 84
Bias Neuron, 322
Bidirectional Recurrent Networks, 354
Binary Independence Model, 281
BM25 Model, 283
Boolean Retrieval, 266
Boosting, 220

Bootstrapping (IE), 410
Bottom-Up Hierarchical Clustering, 92
BPTT, 344
Buckshot, 96

C

C99, 438
CBOW Model, 331
Champion Lists, 277
Character Encoding, 18
Classification, 10
Classifier Evaluation, 209
Cluster Digest, 90, 137
Cluster Purity, 106
Clustering, 8, 73
Clustering Evaluation, 105
Co-clustering, 82
Co-reference Resolution, 381
Coefficient of Determination, 225
Collective Topic Modeling, 252
Compression, 278
Conditional Entropy, 107, 119
Conditional Random Fields, 397
Conjugate Prior, 53
Constituency-Based Parse Tree, 385
Continuous Bag-of-Words Model, 331
Convolutional Neural Network, 347
Corpus, 3
Cosine Similarity, 7, 27
Cross-Lingual Text Mining, 248
Cross-Validation, 224

D

d-Gaps, 279
 Damping Functions, 25, 316
 DCG, 231
 Decision Boundary, 116
 Decision Trees, 11, 142
 Deductive Learner, 116
 Deep Learning, 321
 Delta Encoding, 279
 Dendrogram, 94
 Dependency Graphs, 403
 Dependent Variable, 11
 Determiner, 384
 Dimensionality Reduction, 7, 31
 Dirichlet distribution, 53
 Distance Graph, 319
 Distant Supervision, 410
 Distributional Semantic Models, 312
 Doc2vec, 70, 77, 79, 99, 341
 Document Parsing, 19
 Document-at-a-time Query Processing, 268
 Document-Term Matrix, 5
 DT-min10 Algorithm, 153

E

Elastic Net, 170
 Embedded Feature Selection, 118, 122, 170
 Embedding, 57
 Energy of a Matrix, 33
 Entropy, 107, 119
 Euclidean Distance, 26
 Event Detection, 15, 445
 External Validity Measures, 105
 Extractive Summarization, 361, 362

F

Feature Engineering, 32, 57, 77
 Feature Selection, 75
 Feature Weighting, 75
 Filter Models (Feature Selection), 76, 118
 First Story Detection, 444
 Fisher's Linear Discriminant, 123, 170
 Focussed Crawling, 290
 FOIL, 149
 FOIL's Information Gain, 149
 Fowlkes-Mallows Measure, 108
 Fractionation, 96
 Frobenius Norm, 33

G

Gaussian Radial Basis Kernel, 59, 194
 Gazetteer, 386
 Generalization Power, 115
 Generative Models, 8
 Gini Index, 107, 118
 GloVe, 316
 Graph-Based Summarization, 372

H

HAL, 314
 Headline Generation, 378
 Hidden Variable, 46
 Hierarchical Clustering, 92
 HITS, 300
 Hold-Out, 223
 HTTP, 287
 Hubs, 300
 Hypertext Transfer Protocol, 287

I

IDCG, 232
 Ideal Discounted Cumulative Gain, 232
 Image Captioning, 347
 Independent Variables, 11
 Index Compression, 278
 Inductive Learner, 116
 Information Extraction, 14, 381
 Information Fusion for Summarization, 378
 Information Gain, 108, 120
 Information Retrieval, 259
 Instance-Based Learners, 12, 133
 Internal Validity Measures, 105
 Inverse Document Frequency, 25
 Inverted Index, 133, 263
 ISOMAP, 70

J

Jaccard Similarity, 28
 Jelinek-Mercer Smoothing, 286

K

Kernel k -Means, 99
 Kernel Methods, 59, 313
 Kernel Regression, 168
 Kernel Trick, 100

L

Labels in Classification, 10
 Language Models in Information Retrieval, 285
 LASSO, 12, 169
 Latent Concepts, 32
 Latent Semantic Analysis, 35
 Latent Topics, 32
 Lazy Learners, 12, 133
 Learning Algorithm, 5
 Least Angle Regression, 170
 Least-Squares Regression, 165
 Leave-one-out, 133
 Leave-one-out Cross-validation, 224
 Left Eigenvector, 298
 Left Singular Vectors, 36
 Lemmatization, 24
 Lexical Chains for Summarization, 371
 Lexicon, 3
 LexRank, 373
 libFM, 252
 LIBLINEAR, 204
 LIBSVM, 204
 Linear Classifiers, 12, 159
 Linear Discriminant Analysis Metric, 141
 Linear Kernel, 59, 194
 Linear Least-Squares Fit, 175
 Linear Probing, 262
 Link Prediction, 243
 LLE, 70
 LLSF, 175
 Local Linear Embedding, 70
 Logarithmic Merging, 293
 Logistic Regression, 12, 187
 Loss Function, 12, 163
 Low-Rank Factorization, 31
 Luhn's Summarization Algorithm, 364

M

Machine Translation, 348
 MALLETT, 71, 154, 205, 410, 451
 Matrix Factorization, 7, 31
 Maximum Entropy Markov Models, 396
 Maximum Entropy Model, 192
 Memory Networks, 351
 Memory-Based Learners, 12, 133
 Mixed Membership Models, 74, 79
 Multiclass Learning with Linear Models, 163
 MultiGen, 378

Multilayer Neural Network, 326
 Multinomial Model (Classification), 126
 Multinomial Model (Clustering), 86
 Mutual Information, 108, 120

N

Naïve Bayes Classifier, 11, 123
 Named Entity Recognition, 386
 NDCG, 232
 Near Duplicate Detection, 291
 Nearest Centroid Classification, 136
 Nearest Neighbor Classifiers, 12, 133
 Neural Language Models, 320
 Neural Networks, 320
 Noise Contrastive Estimation, 337
 Nonlinear Dimensionality Reduction, 56
 Normalized Discounted Cumulative Gain, 232
 Normalized Mutual Information, 108
 Nymble, 392
 Nyström Technique, 66

O

Okapi Model, 283
 One-Against-All Multiclass Learning, 164
 One-Against-One Multiclass Learning, 164
 One-Against-Rest Multiclass Learning, 164
 Open Domain Event Extraction, 449
 Open Information Extraction, 410
 Opinion Lexicon, 415
 Opinion Lexicon Expansion, 415
 Opinion Mining, 413
 opinion Mining, 14
 Overlapping Clusters, 79

P

PageRank, 288, 295
 PageRank Algorithm, 13
 PageRank for Summarization, 373
 Parsing in Linguistics, 385
 Parts-of-Speech Tagging, 384
 Pegasos, 181
 Perceptron, 320, 321
 Pessimistic Error Rate, 150
 Plate Diagram, 47
 Pointwise Mutual Information, 119, 421
 Polynomial Kernel, 59, 194
 Porter's Stemming Algorithm, 24
 Positive Pointwise Mutual Information, 317

- Power-Iteration Method, 298
 PPMI, 317
 PPMI Matrix Factorization, 317
 Preferential Crawlers, 287
 Principal Component Analysis, 168
 Principal Components Regression, 167
- Q**
 Query Likelihood Models, 285
 Question Answering, 350
- R**
 Rand Index, 108
 Random Forests, 142, 146
 Random Walks, 295
 Ranking Algorithms, 295
 Ranking Outputs in Classification, 127
 Ranking Support Vector Machines, 274
 Recommender Systems, 246
 Recurrent Neural Networks, 342
 Regressand, 11
 Regression Modeling, 10, 115
 Regressor, 11
 Regularized Least-Squares Classification, 175
 Representer Theorem, 200
 Residual Matrix (Factorization), 35
 Retrieval Status Value, 282
 Right Eigenvector, 298
 Right Singular Vectors, 36
 RNN, 342
 Rocchio Classification, 136
 Rotation Forest, 123, 147
 Rule-Based Classifiers, 11, 147
 Rule-Based Named Entity Recognition, 387
- S**
 Search Engines, 259
 Segmentation of Text, 436
 Semi-supervised Learner, 116
 Sentence Compression, 378
 Sentiment Analysis, 14, 413
 Sequence-to-Sequence Learning, 348
 Sequential Minimal Optimization, 185
 SGNS, 338
 Shingling, 291
 Short Text Mining, 13
 Sigmoid Kernel, 59, 194
 Similarity Computation, 6, 26
 Similarity Forests, 145, 146
 SimRank, 299
 Singular Value Decomposition, 7, 35
 Skip Pointers, 276
 Skip-Grams, 62, 310
 SMO, 185
 Social Streams, 447
 Softmax, 192, 325
 Sparse Coding, 4
 Spectral Clustering, 102
 Spectral Decomposition of SVD, 37
 Spider Traps, 290
 Spiders, 287
 SPPMI Matrix Factorization, 318
 Stacking, 98
 Stemming, 23
 Stop Words, 6, 17, 22
 Streaming Clustering, 443
 String Subsequence Kernels, 62
 Subsampling, 66, 138
 Suffix Stripping, 24
 Summarization (Text), 361
 Supervised Learning, 10
 Supervised Segmentation, 439, 441
 Support Vector Machines, 12, 177
 Support Vectors, 180
 SVDPACK, 70
 SVM, 12, 177
 SVMPerf, 185
 Synsets, 370
- T**
 Tag Trees, 21
 Taxonomy, 73
 Tempex, 449
 Term Frequency, 25
 Term Strength, 75
 Term-at-a-time Query Processing, 268
 Test Data Set, 10
 Testing, 115
 Text Segmentation, 15
 TextRank for Summarization, 379
 TextTiling, 437
 tf-idf Model, 24
 Tiered Indexes, 277
 Tikhonov Regularization, 12, 165

Token, 19
Tokenization, 6, 19
Topic Detection and Tracking, 436
Topic Modeling, 32
Topic Signatures, 366
Topic-Sensitive PageRank, 298
Topical Crawling, 290
Training, 10, 115
Transductive Learner, 116
Tri-factorization, 81
Truecasing, 23

U
Unconstrained Matrix Factorization, 35
Unicode, 19
Unigram Language Model, 285
Universal Crawlers, 287
Unsupervised Information Extraction, 410
Unsupervised Learning, 8
UTF-8, 19

V
Valence Shifter, 419
Vapnik's Principle, 132
Variable Byte Codes, 279
Vector Space Representation, 18, 24
Visible Markov Models, 310
Viterbi Algorithm, 395

W
Web Crawling, 287
Web Resource Discovery, 287
Weston-Watkins Multi-Class SVM, 164, 192
WHISK, 389
Widrow-Hoff Learning, 176
Word2vec, 70, 77, 79, 99, 331
WordNet, 79, 370
Wrapper Models (Feature Selection), 76, 118

Z
Zoned Scoring, 272