

(3) 尝试通过学习隐式特征提升关键词提取效果。

近年来,表示学习技术得到了长足进步和广泛应用.特别是 Word2Vec^[71,76]的提出极大地促进了表示学习技术在自然语言处理领域的应用,尽管个别文献(如 WordAttractionRank 方法^[22])尝试在维基百科数据集上通过 SENNA 方法^[96]学习词向量特征,然后利用词向量特征提升关键词提取效果,然而在该文献中,词向量仅仅是用来计算词间的相似度特征的.该文献显然存在几方面的不足:词向量的学习应该优先尝试从目标文档中学习,而不是直接从维基百科中学习;在该文献发表时表示学习技术还没有得到长足的发展,现在在表示学习技术得到长足发展的情况下,可以尝试用不同的表示学习技术来学习词嵌入向量,进而提升提取效果;对词向量的使用过于简单,仅用来计算词间的相似度.由此可见,在这个方向上还有很大的研究空间,未来可能会成为一个重要的研究方向,且非常有可能出现突破性的研究成果。

(4) 针对主题动态演进的短文本提出有针对性的提取方法。

目前提出的关键词提取方法大都是针对主题静态的文本,即文档的主题是不变的.这些方法不能很好地适用于主题动态演进的短文本,为此,HG-RANK^[104]方法将超图用于推文的关键词中,将时间变化特征和社交网络中的社会属性添加到超图模型中.目前这方面的研究成果非常有限,随着对推文等短文本数据分析需求的不断增加以及关键词提取技术的不断发展,相信未来针对主题动态演进的短文本的关键词提取问题研究将成为一个重要的研究方向。

本文详细综述了面向文本的自动关键词提取技术,围绕各关键词提取方法中使用的特征,展开对现有研究成果进行系统的分析、对比和总结.这种以特征为驱动综述现有关键词提取方法,可使研究着力点由以往重点关注提取方法前移至重点关注特征的定义和对其的综合利用,这种视角的变化将有助于研究者在此基础上提出更好的解决方法.此外,还对一些可能的研究方向进行了简要的探讨,期望能借此推动国内对关键词提取技术的关注和深入研究。

致谢 作者感谢对本文提出宝贵修改意见的评审专家。

References:

- [1] Gutwin C, Paynter G, Witten I, Nevill-Manning C, Frank E. Improving browsing in digital libraries with keyphrase indexes. *Decision Support Systems*, 1999,27(1):81–104.
- [2] Kim SN, Medelyan O, Kan MY, Baldwin T. Automatic keyphrase extraction from scientific articles. *Language Resources and Evaluation*, 2013.47(3):723–742.
- [3] Hassaine A, Mecheter S, Jaoua A. Text categorization using hyper rectangular keyword extraction: Application to news articles classification. In: *Proc. of the ARAMiCS*. Cham: Springer-Verlag, 2015. 312–325. [doi: 10.1007/978-3-319-24704-5_19]
- [4] Zhao WX, Jiang J, He J, Song Y, Achananuparp P, Lim EP, Li X. Topical keyphrase extraction from twitter. In: *Proc. of the ACL*. Stroudsburg PA: ACL, 2011. 379–388.
- [5] He WM. Chinese social topic's keywords extraction algorithm [MS. Thesis]. Beijing: Beijing Jiaotong University, 2017 (in Chinese with English abstract).
- [6] Zhang WN, Ming ZY, Zhang Y, Liu TS, Chua TS. Exploring key concept paraphrasing based on pivot language translation for question retrieval. In: *Proc. of the AAAI*. Palo Alto: AAAI Press, 2015. 410–416.
- [7] Wu HC, Tian ZH, Wu W, Chen EH. An unsupervised approach for low-quality answer detection in community question-answering. In: *Proc. of the DASFAA*. Cham: Springer-Verlag, 2017. 85–101. [doi: 10.1007/978-3-319-55699-4_6]
- [8] Tang YX, Huang WL, Liu Q, Tung AKH, Wang XL, Yang JS, Zhang BB. QALink: Enriching text documents with relevant Q&A site contents. In: *Proc. of the CIKM*. New York: ACM, 2017. 1359–1368. [doi: 10.1145/3132847.3132934]
- [9] Zhang W, Feng W, Wang JY. Integrating semantic relatedness and words' intrinsic features for keyword extraction. In: *Proc. of the IJCAI*. San Francisco: Morgan Kaufmann Publishers Inc., 2013. 2225–2231.
- [10] Gollapalli SD, Caragea C. Extracting keyphrases from research papers using citation networks. In: *Proc. of the AAAI*. Palo Alto: AAAI Press, 2014. 1629–1635.

- [11] Hasan KS, Ng V. Automatic keyphrase extraction: A survey of the state of the art. In: Proc. of the ACL. Stroudsburg: ACL, 2014. 1262–1273.
- [12] Marujo L, Ling W, Trancoso I, Dyer C, Black AW, Gershman A, Matos DMD, Neto JP, Carbonell J. Automatic keyword extraction on twitter. In: Proc. of the ACL and IJCNLP. Stroudsburg: ACL, 2015. 637–643. [doi: 10.3115/v1/P15-2105]
- [13] Sterckx L, Demeester T, Deleu J, Devellder C. Topical word importance for fast keyphrase extraction. In: Proc. of the WWW. New York: ACM, 2015. 121–122. [doi: 10.1145/2740908.2742730]
- [14] Sterckx L, Caragea C, Demeester T, Devellder C. Supervised keyphrase extraction as positive unlabeled learning. In: Proc. of the EMNLP. Stroudsburg: ACL, 2016. 1924–1929.
- [15] Gollapalli SD, Li XL, Yang P. Incorporating expert knowledge into keyphrase extraction. In: Proc. of the AAAI. 2017. Palo Alto: AAAI Press, 3180–3187.
- [16] Florescu C, Caragea C. A position-biased pagerank algorithm for keyphrase extraction. In: Proc. of the AAAI. Palo Alto: AAAI Press, 2017. 4923–4924.
- [17] Meng R, Zhao SQ, Han SG, He DQ, Brusilovsky P, Chi Y. Deep keyphrase generation. In: Proc. of the ACL. Stroudsburg: ACL, 2017. 582–592.
- [18] Sparck-Jones K. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 1972, 28(1):11–21.
- [19] Salton G, Buckley C. Term-Weighting approaches in automatic text retrieval. *Information Processing & Management*, 1988,24(5): 513–523.
- [20] Turney PD. Learning algorithms for keyphrase extraction. *Information Retrieval*, 1999,2(4):303–336.
- [21] Frank E, Paynter GW, Witten IH, Gutwin C, Nevill-Manning CG. Domain-Specific keyphrase extraction. In: Proc. of the IJCAI. San Francisco: Morgan Kaufmann Publishers Inc., 1999. 668–673.
- [22] Wang R, Liu W, McDonald C. Corpus-Independent generic keyphrase extraction using word embedding vectors. In: Proc. of the Software Engineering Research Conf. 2014. 39.
- [23] Liu ZY, Research on keyword extraction using document topical structure [Ph.D. Thesis]. Beijing: Tsinghua University, 2011 (in Chinese with English abstract).
- [24] Ding ZY, Zhang Q, Huang XJ. Keyphrase extraction from online news using binary integer programming. In: Proc. of the IJCNLP. Stroudsburg: ACL, 2011. 165–173.
- [25] Figueroa G, Chen PC, Chen YS. RankUp: Enhancing graph-based keyphrase extraction methods with error-feedback propagation. *Computer Speech & Language*, 2018,47:112–131. [doi: 10.1016/j.csl.2017.07.004]
- [26] Rafiei-Asl J, Nickabadi A. TSAKE: A topical and structural automatic keyphrase extractor. *Applied Soft Computing*, 2017,58: 620–630. [doi: 10.1016/j.asoc.2017.05.014]
- [27] Zhao JS, Zhu QM, Zhou GD, Zhang L. Review of the research in automatic keyword extraction. *Ruan Jian Xue Bao/Journal of Software*, 2017,28(9):2431–2449 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5301.htm> [doi: 10.13328/j.cnki.jos.005301]
- [28] Boudin F, Mougard H, Cram D. How document pre-processing affects keyphrase extraction performance. In: Proc. of the COLING Workshop on Noisy User-Generated Text. Osaka: The COLING 2016 Organizing Committee, 2016. 121–128.
- [29] Toutanova K, Klein D, Manning CD, Singer Y. Feature-Rich part-of-speech tagging with a cyclic dependency network. In: Proc. of the ACL. Stroudsburg: ACL, 2003. 173–180.
- [30] Park Y, Byrd RJ, Boguraev BK. Automatic glossary extraction: Beyond terminology identification. In: Proc. of the ACL. Stroudsburg: ACL, 2002. 1–7.
- [31] Kumar N, Srinathan K. Automatic keyphrase extraction from scientific documents using n -gram filtration technique. In: Proc. of the 8th ACM Symp. on Document Engineering. New York: ACM, 2008. 199–208. [doi: 10.1145/1410140.1410180]
- [32] Hulth A. Improved automatic keyword extraction given more linguistic knowledge. In: Proc. of the ACL. Stroudsburg: ACL, 2003. 216–223.
- [33] Kim SN, Kan MY. Re-Examining automatic keyphrase extraction approaches in scientific articles. In: Proc. of the ACL Workshop on Multiword Expressions: Identification, Interpretation, Disambiguation and Applications. Stroudsburg: ACL, 2009. 9–16.

- [34] Wang LT, Li F. SJTULTLAB: Chunk based method for keyphrase extraction. In: Proc. of the ACL Workshop on Semantic Evaluation. Stroudsburg: ACL, 2010. 158–161.
- [35] Le TTN, Nguyen ML, Shimazu A. Unsupervised keyphrase extraction: introducing new kinds of words to keyphrases. In: Proc. of the AJCAI. Cham: Springer-Verlag, 2016. 665–671. [doi: 10.1007/978-3-319-50127-7_58]
- [36] Xie F, Wu XD, Zhu XQ. Efficient sequential pattern mining with wildcards for keyphrase extraction. Knowledge-Based Systems, 2017,115:27–39. [doi: 10.1016/j.knosys.2016.10.011]
- [37] Wang QR, Sheng VS, Wu XD. Keyphrase extraction with sequential pattern mining. In: Proc. of the AAI. Palo Alto: AAAI Press, 2017. 5003–5004.
- [38] Lovins JB. Development of a stemming algorithm. Mechanical Translation and Computational Linguistics, 1968,11:22–31.
- [39] Bird S. NLTK: The natural language toolkit. In: Proc. of the COLING/ACL on Interactive Presentation Sessions. Stroudsburg: ACL, 2006. 69–72.
- [40] Mihalec R, Tarau P. TextRank: Bringing order into texts. In: Proc. of the EMNLP. Stroudsburg: ACL, 2004. 404–411.
- [41] Zesch T, Gurevych I. Approximate matching for evaluating keyphrase extraction. In: Proc. of the RANLP. Stroudsburg: ACL, 2009. 484–489.
- [42] Wang R, Liu W, McDonald C. How preprocessing affects unsupervised keyphrase extraction. In: Proc. of the CILing. Berlin, Heidelberg: Springer-Verlag, 2014. 163–176.
- [43] Hofmann K, Tsagkias M, Meij E, De Rijke M. A comparative study of features for keyphrase extraction in scientific literature. 2009. <http://edgar.meij.pro/comparative-study-features-keyphrase-extraction>
- [44] Haddoud M, Mokhtari A, Lecroq T, Abdeddaïm S. Accurate keyphrase extraction from scientific papers by mining linguistic information. In: Proc. of the CLBib. 2015. 12–17.
- [45] Aquino GO, Lanzarini LC. Keyword identification in Spanish documents using neural networks. Journal of Computer Science & Technology, 2015,15.
- [46] Nguyen TD, Luong MT. WINGNUS: Keyphrase extraction utilizing document logical structure. In: Proc. of the ACL Workshop on Semantic Evaluation. Stroudsburg: ACL, 2010. 166–169.
- [47] Witten IH, Paynter GW, Frank E, Gutwin C, Nevill-Manning CG. KEA: Practical automatic keyphrase extraction. In: Proc. of the JCDL. New York: ACM, 1999. 254–255. [doi: 10.1145/313238.313437]
- [48] Li GY, Wang HF. Improved automatic keyword extraction based on textrank using domain knowledge. Communications in Computer & Information Science, 2014,(496):403–413.
- [49] Haddoud M, Abdeddaïm S. Accurate keyphrase extraction by discriminating overlapping phrases. Journal of Information Science, 2014,40(4):488–500. [doi: 10.1177/0165551514530210]
- [50] Caragea C, Bulgarov F, Godea A, Gollapalli SD. Citation-Enhanced keyphrase extraction from research papers: A supervised approach. In: Proc. of the EMNLP. Stroudsburg: ACL, 2014. 1435–1446. [doi: 10.3115/v1/D14-1150]
- [51] Zhang K, Xu H, Tang J, Li JZ. Keyword extraction using support vector machine. In: Proc. of the WAIM. Berlin, Heidelberg: Springer-Verlag, 2006. 85–96. [doi: 10.1007/11775300_8]
- [52] Tomokiyo T, Hurst M. A language model approach to keyphrase extraction. In: Proc. of the ACL Workshop on Multiword Expressions. Stroudsburg: ACL, 2003. 33–40. [doi: 10.3115/1119282.1119287]
- [53] Eichler K, Neumann G. DFKI KeyWE: Ranking keyphrases extracted from scientific articles. In: Proc. of the ACL Workshop on Semantic Evaluation. Stroudsburg: ACL, 2010. 150–153.
- [54] John AK, Di Caro L, Boella G. A supervised keyphrase extraction system. In: Proc. of the SEMANTICS. New York: ACM, 2016. 57–62. [doi: 10.1145/2993318.2993323]
- [55] Sarkar K. Automatic keyphrase extraction from medical documents. Pattern Recognition and Machine Intelligence, 2009, 273–278. [doi: 10.1007/978-3-642-11164-8_44]
- [56] Feng H, Chen K, Deng XT, Zheng WM. Accessor variety criteria for chinese word extraction. Computational Linguistics, 2004,30(1):75–93. [doi: 10.1162/089120104773633394]
- [57] Nguyen TD, Kan MY. Keyphrase extraction in scientific publications. In: Proc. of the ICADL. Berlin, Heidelberg: Springer-Verlag, 2007. 317–326. [doi: 10.1007/978-3-540-77094-7_41]

- [58] Medelyan O, Frank E, Witten IH. Human-Competitive tagging using automatic keyphrase extraction. In: Proc. of the EMNLP. Stroudsburg: ACL, 2009. 1318–1327.
- [59] Berend G. Exploiting extra-textual and linguistic information in keyphrase extraction. *Natural Language Engineering*, 2014,22(1): 73–95. [doi: 10.1017/S1351324914000126]
- [60] Joorabchi A, Mahdi AE. Automatic keyphrase annotation of scientific documents using Wikipedia and genetic algorithms. *Journal of Information Science*, 2013,39(3):410–426. [doi: 10.1177/0165551512472138]
- [61] Wang JB, Peng H. Keyphrases extraction from Web document by the least squares support vector machine. In: Proc. of the IEEE/WIC/ACM Int'l Conf. on Web Intelligence. Washington: IEEE, 2005. 293–296. [doi: 10.1109/WI.2005.87]
- [62] Zhang CZ, Wang HL, Liu Y, Wu D, Liao Y, Wang B. Automatic keyword extraction from documents using conditional random fields. *Journal of Computational Information Systems*, 2008,4(3):1169–1180.
- [63] Bhaskar P, Nongmeikapam K, Bandyopadhyay S. Keyphrase extraction in scientific articles: A supervised approach. In: Proc. of the COLING. Mumbai: The COLING 2012 Organizing Committee, 2012. 17–24.
- [64] Marujo L, Gershman A, Carbonell J, Frederking R, Neto JP. Supervised topical key phrase extraction of news stories using crowdsourcing, light filtering and co-reference normalization. In: Proc. of the LREC. European Language Resources Association, 2012. 1385–1389.
- [65] Fellbaum C. Wordnet: An electronic lexical database. *Computational Linguistics*, 1998,25(2):292–296.
- [66] Ercan G, Cicekli I. Using lexical chains for keyword extraction. *Information Processing & Management*, 2007,43(6):1705–1714. [doi: 10.1016/j.ipm.2007.01.015]
- [67] Suo HG, Liu YS, Cao SY. A keyword selection method based on lexical chains. *Zhong Wen Xin Xi Xue Bao/Journal of Chinese Information Processing*, 2006,20(6):25–30 (in Chinese with English abstract).
- [68] Turney PD. Coherent keyphrase extraction via Web mining. In: Proc. of the IJCAI. San Francisco: Morgan Kaufmann Publishers Inc., 2003. 434–439.
- [69] Liu W, Chung BC, Wang R, Ng J, Morlet N. A genetic algorithm enabled ensemble for unsupervised medical term extraction from clinical letters. *Health Information Science and Systems*, 2015,3(5):1–14. [doi: 10.1186/s13755-015-0013-y]
- [70] Zhou XH, Zhang XD, Hu XH. Maxmatcher: Biological concept extraction using approximate dictionary lookup. *PRICAI: Trends in Artificial Intelligence*, 2006, 1145–1149. [doi: 10.1007/978-3-540-36668-3_150]
- [71] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: Proc. of the Workshop at ICLR. 2013. 1–12.
- [72] Liu ZY, Li P, Zheng YB, Sun MS. Clustering to find exemplar terms for keyphrase extraction. In: Proc. of the EMNLP. Stroudsburg: ACL, 2009. 257–266.
- [73] Boudin F. A comparison of centrality measures for graph-based keyphrase extraction. In: Proc. of the IJCNLP. Nagoya: Asian Federation of Natural Language Processing, 2013. 834–838.
- [74] Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 2003,3:993–1022.
- [75] Liu ZY, Huang WY, Zheng YB, Sun MS. Automatic keyphrase extraction via topic decomposition. In: Proc. of the EMNLP. Stroudsburg: ACL, 2010. 366–376.
- [76] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: Proc. of the NIPS. New York: Curran Associates Inc., 2013. 3111–3119.
- [77] Lai SW. Word and document embeddings based on neural network approaches [Ph.D. Thesis]. Beijing: The University of Chinese Academy of Sciences, 2016 (in Chinese with English abstract).
- [78] Zhang Q, Wang Y, Gong YY, Huang XJ. Keyphrase extraction using deep recurrent neural networks on twitter. In: Proc. of the EMNLP. Stroudsburg: ACL, 2016. 836–845. [doi: 10.18653/v1/D16-1080]
- [79] Wang YL, Jin Y, Zhu XD, Goutte C. Extracting discriminative keyphrases with learned semantic hierarchies. In: Proc. of the COLING. Osaka: The COLING 2016 Organizing Committee, 2016. 932–942.
- [80] Papagiannopoulou E, Tsoumakas G. Local word vectors guide keyphrase extraction. arXiv Preprint arXiv:1710.07503, 2017.
- [81] Whitley D. The genitor algorithm and selection pressure: Why rank-based allocation of reproductive trials is best. In: Proc. of the 3rd Int'l Conf. on Genetic Algorithms. San Francisco: Morgan Kaufmann Publishers Inc., 1989. 116–121.

- [82] Hulth A. Reducing false positives by expert combination in automatic keyword indexing. In: Proc. of the RANLP. John Benjamins Publishing Company, 2003. 367–376. [doi: 10.1075/cilt.260.41hul]
- [83] Medelyan O, Witten IH. Thesaurus based automatic keyphrase indexing. In: Proc. of the JCDL. New York: ACM, 2006. 296–297. [doi: 10.1145/1141753.1141819]
- [84] Bulgarov F, Caragea C. A comparison of supervised keyphrase extraction models. In: Proc. of the WWW. New York: ACM, 2015. 13–14. [doi: 10.1145/2740908.2742776]
- [85] Krapivin M, Autayeu M, Marchese M, Blanzieri E, Segata N. Improving machine learning approaches for keyphrases extraction from scientific documents with natural language knowledge. In: Proc. of the JCDL. Berlin, Heidelberg: Springer-Verlag, 2010. 102–111. [doi: 10.1007/978-3-642-13654-2_12]
- [86] Jiang X, Hu YH, Li H. A ranking approach to keyphrase extraction. In: Proc. of the SIGIR. New York: ACM, 2009. 756–757. [doi: 10.1145/1571941.1572113]
- [87] Chen YQ, Zhou RQ, Zhu WH, Li MT, Yin J. Mining patent knowledge for automatic keyword extraction. Journal of Computer Research and Development, 2016,53(8):1740–1752 (in Chinese with English abstract).
- [88] Sarkar K, Nasipuri M, Ghose S. Machine learning based keyphrase extraction: Comparing decision trees, Naïve Bayes, and artificial neural networks. Journal of Information Processing Systems, 2012,8(4):693–712. [doi: 10.3745/JIPS.2012.8.4.693]
- [89] Boudin F. Reducing over-generation errors for automatic keyphrase extraction using integer linear programming. In: Proc. of the ACL Workshop on Novel Computational Approaches to Keyphrase Extraction. Stroudsburg: ACL, 2015. 19–24.
- [90] Pei J, Han JW, Mortazavi-Asl B, Wang JY, Pinto H, Chen QM, Dayal U, Hsu MC. Mining sequential patterns by pattern-growth: the prefixspan approach. IEEE Trans. on Knowledge and Data Engineering, 2004,16(11):1424–1440. [doi: 10.1109/TKDE.2004.77]
- [91] Frantzi K, Ananiadou S, Mima H. Automatic recognition of multi-word terms: The *c*-value/*nc*-value method. Int'l Journal on Digital Libraries, 2000,3(2):115–130. [doi: 10.1007/s007999900023]
- [92] Herbrich R, Graepel T, Obermayer K. Large margin rank boundaries for ordinal regression. In: Advances in Large Margin Classifiers. 2000. 115–132.
- [93] Shi W, Zheng WG, Yu JX, Cheng H, Zou L. Keyphrase extraction using knowledge graphs. In: Proc. of the APWeb and WAIM Joint Conf. on Web and Big Data. Cham: Springer-Verlag, 2017. 132–148. [doi: 10.1007/978-3-319-63579-8_11]
- [94] Page L, Brin S, Motwani R. The pagerank citation ranking: Bringing order to the Web. Technical Report, Stanford InfoLab, 1999.
- [95] Wan XJ, Xiao JG. Single document keyphrase extraction using neighborhood knowledge. In: Proc. of the AAAI. Palo Alto: AAAI Press, 2008. 855–860.
- [96] Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. Journal of Machine Learning Research, 2011,12:2493–2537.
- [97] Haveliwala TH. Topic-Sensitive PageRank: A context-sensitive ranking algorithm for Web search. IEEE Trans. on Knowledge and Data Engineering, 2003,15(4):784–796. [doi: 10.1109/TKDE.2003.1208999]
- [98] Teneva N, Cheng WW. Saliency rank: Efficient keyphrase extraction with topic modeling. In: Proc. of the ACL. Stroudsburg: ACL, 2017,2:530–535. [doi: 10.18653/v1/P17-2084]
- [99] Bougouin A, Boudin F, Daille B. TopicRank: Graph-Based topic ranking for keyphrase extraction. In: Proc. of the IJCNLP. Stroudsburg: ACL, 2013. 543–551.
- [100] Zhang YX, Chang YC, Liu XQ, Gollapalli SD, Li XL, Xiao CJ. MIKE: Keyphrase extraction by integrating multidimensional information. In: Proc. of the CIKM. New York: ACM, 2017. 1349–1358. [doi: 10.1145/3132847.3132956]
- [101] Wan XJ, Yang JW, Xiao JG. Towards an iterative reinforcement approach for simultaneous document summarization and keyword extraction. In: Proc. of the ACL. Stroudsburg: ACL, 2007. 552–559.
- [102] Ching WK, Fung ES, Ng MK. A multivariate markov chain model for categorical data sequences and its applications in demand predictions. IMA Journal of Management Mathematics, 2002,13(3):187–199. [doi: 10.1093/imaman/13.3.187]
- [103] Yan Y, Tan QP, Xie QZ, Zeng P, Li PP. A graph-based approach of automatic keyphrase extraction. Procedia Computer Science, 2017,107:248–255. [doi: 10.1016/j.procs.2017.03.087]
- [104] Bellaachia A, Al-Dhelaan M. HG-Rank: A hypergraph-based keyphrase extraction for short documents in dynamic genre. In: Proc. of the MSM. 2014. 42–49.

- [105] Bellaachia A, Al-Dhelaan M. Short text keyphrase extraction with hypergraphs. Progress in Artificial Intelligence, 2015,3(2):73–87. [doi: 10.1007/s13748-014-0058-1]
- [106] Grineva M, Grinev M, Lizorkin D. Extracting key terms from noisy and multitheme documents. In: Proc. of the WWW. New York: ACM, 2009. 661–670. [doi: 10.1145/1526709.1526798]
- [107] Medelyan O. Human-Competitive automatic topic indexing [Ph.D. Thesis]. The University of Waikato, 2009.
- [108] Krapivin M, Autaeu A, Marchese M. Large dataset for keyphrases extraction. Technical Report, University of Trento, 2008.
- [109] Augenstein I, Das M, Riedel S, Vikraman L, McCallum A. Semeval 2017 task 10: Scienceie-Extracting keyphrases and relations from scientific publications. arXiv Preprint arXiv:1704.02853, 2017.
- [110] Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval. Cambridge University Press, 2008.
- [111] Voorhees EM. The trec-8 question answering track report. In: Proc. of the TREC-8. 1999. 77–82.
- [112] Buckley C, Voorhees EM. Retrieval evaluation with incomplete information. In: Proc. of the SIGIR. 2004. 25–32. [doi: 10.1145/1008992.1009000]
- [113] Camacho JEP, Ledeneva Y, Hernández RAG. Comparison of automatic keyphrase extraction systems in scientific papers. Research in Computing Science, 2016,115:181–191.
- [114] Sterckx L, Demeester T, Deleu J, Develder C. Creation and evaluation of large keyphrase extraction collections with multiple opinions. Language Resources and Evaluation, 2017,52:503–532. [doi: 10.1007/s10579-017-9395-6]
- [115] Boudin F. PKE: An open source python-based keyphrase extraction toolkit. In: Proc. of the COLING. Osaka: The COLING 2016 Organizing Committee, 2016. 69–73.

附中文参考文献:

- [5] 何伟名. 中文社交媒体话题关键词抽取算法[硕士学位论文]. 北京: 北京交通大学, 2017.
- [23] 刘知远. 基于文档主题结构的关键词抽取方法研究[博士学位论文]. 北京: 清华大学, 2011.
- [27] 赵京胜, 朱巧明, 周国栋, 张丽. 自动关键词抽取研究综述. 软件学报, 2017, 28(9): 2431–2449. <http://www.jos.org.cn/1000-9825/5301.htm> [doi: 10.13328/j.cnki.jos.005301]
- [67] 索红光, 刘玉树, 曹淑英. 一种基于词汇链的关键词抽取方法. 中文信息学报, 2006, 30(6): 25–30.
- [77] 来斯惟. 基于神经网络的词和文档语义向量表示方法研究[博士学位论文]. 北京: 中国科学院大学, 2016.
- [87] 陈忆群, 周如旗, 朱蔚恒, 李梦婷, 印鉴. 挖掘专利知识实现关键词自动抽取. 计算机研究与发展, 2016, 53(8): 1740–1752.



常耀成(1992—), 男, 江苏淮安人, 硕士, 主要研究领域为自然语言处理.



万怀宇(1981—), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为社交网络挖掘, 用户画像.



张宇翔(1975—), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为自然语言处理, 网络数据分析.



肖春景(1978—), 女, 讲师, CCF 专业会员, 主要研究领域为推荐系统, 数据挖掘, 人工智能.



王红(1963—), 女, 教授, CCF 专业会员, 主要研究领域为智能信息处理, 大数据挖掘.