

References:

- [1] Zhao W, Ma H, He Q. Parallel K -means clustering based on MapReduce. In: Proc. of the Int'l Conf. on Cloud Computing. Springer-Verlag, 2009. 674–679. [doi: 10.1007/978-3-642-10665-1_71]
- [2] He J, Pan W. A DENCLUE based approach to neuro-fuzzy system modeling. In: Proc. of the Int'l Conf. on Advanced Computer Control. IEEE, 2010. 42–46. [doi: 10.1109/ICACC.2010.5487269]
- [3] Chen N, Chen A, Zhou LX. An incremental grid density-based clustering algorithm. Journal of Software, 2002,13(1):1–7 (in Chinese with English abstract). http://www.jos.org.cn/ch/reader/create_pdf.aspx?file_no=20020101&journal_id=jos
- [4] Chowdhury GG. Natural language processing. Annual Review of Information Science and Technology, 2003,37(1):51–89.
- [5] Cao Z, Wei F, Dong L, Li SJ, Zhou M. Ranking with recursive neural networks and its application to multi-document summarization. In: Proc. of the AAAI. 2015. 2153–2159.
- [6] Silverstein C, Marais H, Henzinger M, Moricz M. Analysis of a very large Web search engine query log. In: Proc. of the Sigir Forum. ACM Press, 1999. 6–12. [doi: 10.1145/331403.331405]
- [7] Rodriguez A, Laio A. Clustering by fast search and find of density peaks. Science, 2014,344(6191):1492–1496. [doi: 10.1126/science.1242072]
- [8] Owoputi O, O'Connor B, Dyer C, Gimpel K, Schneider N, Smith NA. Improved part-of-speech tagging for online conversational text with word clusters. In: Proc. of the Association for Computational Linguistics. 2013.
- [9] Amini A, Wah TY, Saboohi H. On density-based data streams clustering algorithms: A survey. Journal of Computer Science and Technology, 2014,29(1):116–141. [doi: 10.1007/s11390-014-1416-y]
- [10] Sander J, Ester M, Kriegel HP, Xu XW. Density-Based clustering in spatial databases: The algorithm gbscan and its applications. Data Mining and Knowledge Discovery, 1998,2(2):169–194. [doi: 10.1023/A:1009745219419]
- [11] Ankerst M, Breunig MM, Kriegel HP, Sander J. OPTICS: Ordering points to identify the clustering structure. Proc. of the ACM SIGMOD Record, 1999,28(2):49–60. [doi: 10.1145/304182.304187]
- [12] Hinneburg A, Keim DA. An efficient approach to clustering in large multimedia databases with noise. Proc. of the KDD, 1998,98: 58–65.
- [13] Lee CH, Chien TF. Leveraging microblogging big data with a modified density-based clustering approach for event awareness and topic ranking. Journal of Information Science, 2013,39(4):523–543. [doi: 10.1177/0165551513478738]
- [14] Lee CH. Mining spatio-temporal information on microblogging streams using a density-based online clustering method. Expert Systems with Applications, 2012,39(10):9623–9641. [doi: 10.1016/j.eswa.2012.02.136]
- [15] Yang CC, Ng TD. Analyzing and visualizing Web opinion development and social interactions with density-based clustering. IEEE Trans. on Systems, Man and Cybernetics, Part A: Systems and Humans, 2011,41(6):1144–1155. [doi: 10.1109/TSMCA.2011.2113334]
- [16] Yu Z, Han G, Li L, Liu J, Zhang J. Adaptive noise immune cluster ensemble using affinity propagation. In: Proc. of the IEEE Int'l Conf. on Data Engineering. IEEE Computer Society, 2016. 1454–1455. [doi: 10.1109/TKDE.2015.2453162]
- [17] Yu Z, Li L, Gao Y, You J, Liu J, Wong HS, Han GQ. Hybrid clustering solution selection strategy. Pattern Recognition, 2014,47(10):3362–3375. [doi: 10.1016/j.patcog.2014.04.005]
- [18] Yu Z, Luo P, You J, Wong HS, Leung H, Wu S, Zhang J, Han FQ. Incremental semi-supervised clustering ensemble for high dimensional data clustering. IEEE Trans. on Knowledge and Data Engineering, 2016,28(3):701–714. [doi: 10.1109/TKDE.2015.2499200]
- [19] Wang Z, Yu Z, Chen C, You J, Gu T, Wong HS, Zhang J. Clustering by local gravitation. IEEE Trans. on Cybernetics, 2017,99: 1–14. [doi: 10.1109/TCYB.2017.2695218]
- [20] Wang G, Song Q. Automatic clustering via outward statistical testing on density metrics. IEEE Trans. on Knowledge and Data Engineering, 2016,28(8):1971–1985. [doi: 10.1109/TKDE.2016.2535209]
- [21] Dean J, Ghemawat S. MapReduce: Simplified data processing on large clusters. In: Proc. of the Conf. on Symp. on Operating Systems Design and Implementation. Berkeley: USENIX Association, 2004. 137–150.

- [22] Lü Z, Hu Y, Zhong H, Wu JP, Li B, Zhou H. Parallel K -means clustering of remote sensing images based on MapReduce. In: Proc. of the Web Information Systems and Mining. Berlin, Heidelberg: Springer-Verlag, 2010. 162–170. [doi: 10.1007/978-3-642-16515-3_21]
- [23] Jin C, Liu R, Chen Z, Hendrix W, Agrawal A, Choudhary A. A scalable hierarchical clustering algorithm using spark. In: Proc. of the IEEE 1st Int'l Conf. on Big Data Computing Service and Applications. IEEE Computer Society, 2015. 418–426. [doi: 10.1109/BigDataService.2015.67]
- [24] Sinha A, Jana PK. A novel K -means based clustering algorithm for big data. In: Proc. of the Int'l Conf. on Advances in Computing, Communications and Informatics. IEEE, 2016. 1875–1879. [doi: 10.1109/ICACCI.2016.7732323]
- [25] Lu WM, Du CY, Wei BG, Shen CH, Ye ZC. Distributed near neighbor propagation clustering algorithm based on MapReduce. Computer Research and Development, 2012,49(8):1762–1772 (in Chinese with English abstract).
- [26] Fahad A, Alshatri N, Tari Z, Alamri A, Khilil I, Zomaya AY, Fofou S, Bouras A. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. IEEE Trans. on Emerging Topics in Computing, 2014,2(3):267–279. [doi: 10.1109/TETC.2014.2330519]

附中文参考文献:

- [3] 陈宁,陈安,周龙骧.基于密度的增量式网格聚类算法.软件学报,2002,13(1):1–7. http://www.jos.org.cn/ch/reader/create_pdf.aspx?file_no=20020101&journal_id=jos
- [25] 鲁伟明,杜晨阳,魏宝刚,沈春辉,叶振超.基于 MapReduce 的分布式近邻传播聚类算法.计算机研究与发展,2012,49(8):1762–1772.



王岩(1992 -),男,吉林白城人,硕士生,主要研究领域为数据挖掘,Web 挖掘,机器学习.



韩佳育(1989 -),女,博士生,CCF 学生会会员,主要研究领域为数据挖掘,社交网络,推荐系统.



彭涛(1977 -),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为数据挖掘,Web 挖掘,机器学习.



刘露(1989 -),女,博士,主要研究领域为数据挖掘,Web 挖掘,机器学习.