

فصل ششم:

منابع

1. Manyika, J., et al., *Big data: The next frontier for innovation, competition, and productivity*. 2011.
2. Hathaway, R.J. and J.C. Bezdek, *Extending fuzzy and probabilistic clustering to very large data sets*. *Computational Statistics & Data Analysis*, 2006. **51**(1): p. 215-234.
3. Madden, S., *From databases to big data*. *IEEE Internet Computing*, 2012. **16**(3): p. 4-6.
4. Beyer, M.A. and D. Laney, *The importance of 'big data': a definition*. Stamford, CT: Gartner, 2012: p. 2014-2018.
5. Tole, A.A., *Big data challenges*. *Database Systems Journal*, 2013. **4**(3): p. 31-40.
6. Brachman, R.J. and T. Anand. *The process of knowledge discovery in databases*. in *Advances in knowledge discovery and data mining*. 1996. American Association for Artificial Intelligence.
7. Orgad, S., *How can researchers make sense of the issues involved in collecting and interpreting online and offline data?* 2009.
8. Hallikainen, H., E. Savimäki, and T. Laukkanen, *Fostering B2B sales with customer big data analytics*. *Industrial Marketing Management*, 2020. **86**: p. 90-98.
9. García, S., et al., *Big data preprocessing: methods and prospects*. *Big Data Analytics*, 2016. **1**(1): p. 9.
10. Dasarathy, B.V., *Information fusion—what, where, why, when, and how?* 2001, Elsevier.
11. Bossé, É., J. Roy, and S. Wark, *Concepts, models, and tools for information fusion*. 2007: Artech House.
12. Foo, P.H. and G.W. Ng, *High-level Information Fusion: An Overview*. *J. Adv. Inf. Fusion*, 2013. **8**(1): p. 33-72.
13. Nilsson, M., *Human decision making and information fusion: Extending the concept of decision support*. 2007, Institutionen för kommunikation och information.
14. Hall, D.L. and J. Llinas, *An introduction to multisensor data fusion*. *Proceedings of the IEEE*, 1997. **85**(1): p. 6-23.
15. Esteban, J., et al., *A review of data fusion models and architectures: towards engineering guidelines*. *Neural Computing & Applications*, 2005. **14**(4): p. 273-281.
16. Dasarathy, B.V., *Decision fusion*. Vol. 1994. 1994: IEEE Computer Society Press Los Alamitos, CA.
17. Bedworth, M. and J. O'Brien, *The Omnibus model: a new model of data fusion?* *IEEE Aerospace and Electronic Systems Magazine*, 2000. **15**(4): p. 30-36.
18. Endsley, M.R., *Toward a theory of situation awareness in dynamic systems*. *Human factors*, 1995. **37**(1): p. 32-64.
19. Endsley, M.R. and D. Garland, *Theoretical underpinnings of situation awareness: A critical review*. *Situation awareness analysis and measurement*, 2000: p. 3-32.
20. Al-Badrashiny, M., et al. *SPLIT: Smart Preprocessing (Quasi) Language Independent Tool*. in *LREC*. 2016.
21. Nurmi, P., et al. *An Architecture for Distributed Agent-Based Data Preprocessing*. in *AIS-ADM*. 2005. Springer.
22. Othman, Z.A., et al. *Agent based preprocessing*. in *Intelligent and Advanced Systems, 2007. ICIAS 2007. International Conference on*. 2007. IEEE.
23. Jiang, J., et al. *A Positive Preprocessing Framework for Mapping Traditional Replica Selection Algorithms*. in *Computing, Communication, Control, and Management, 2008. CCCM'08. ISECS International Colloquium on*. 2008. IEEE.

24. Maedche, A., A. Hotho, and M. Wiese, *Enhancing preprocessing in data-intensive domains using online-analytical processing*. Data warehousing and knowledge discovery, 2000: p. 258-264.
25. Ramentol, E., et al., *SMOTE-RSB*: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory*. Knowledge and information systems, 2012. **33**(2): p. 245-265.
26. Hu, F., et al., *A parallel oversampling algorithm based on NRSBoundary-SMOTE*. JOURNAL OF INFORMATION & COMPUTATIONAL SCIENCE, 2014. **11**(13): p. 4655-4665.
27. Yan, J., et al., *Effective and efficient dimensionality reduction for large-scale and streaming data preprocessing*. IEEE transactions on Knowledge and Data Engineering, 2006. **18**(3): p. 320-333.
28. Qu, Z., et al., *An attribute reducing method for electric power big data preprocessing based on cloud computing technology*. Automation of electric power systems, 2014. **8**(8): p. 67-71.
29. Kumar, P., V. Chandna, and M.S. Thomas, *Fuzzy-genetic algorithm for preprocessing data at the RTU*. IEEE transactions on power systems, 2004. **19**(2): p. 718-723.
30. Momm, H., G. Easson, and D. Wilkins. *Genetic programming as a preprocessing tool to aid multi-temporal imagery classification*. in *Proceedings of the ASPRS 2006 Annual Conference*. 2006.
31. Zhao, F. and T.-Y. Leong. *A data preprocessing framework for supporting probability-learning in dynamic decision modeling in medicine*. in *Proceedings of the AMIA Symposium*. 2000. American Medical Informatics Association.
32. Fritscher, K.D. and R. Schubert. *A software framework for preprocessing and level set segmentation of medical image data*. in *Proc. of SPIE Vol.* 2005.
33. Arnold-Bos, A., J.-P. Malkasse, and G. Kervern. *A preprocessing framework for automatic underwater images denoising*. in *European Conference on Propagation and Systems*. 2005.
34. Bogorny, V., P.M. Engel, and L.O. Alvares. *GEOARM: an Interoperable Framework to Improve Geographic Data Preprocessing and Spatial Association Rule Mining*. in *SEKE*. 2006.
35. Jeffery, S.R., et al. *A pipelined framework for online cleaning of sensor data streams*. in *Data Engineering, 2006. ICDE'06. Proceedings of the 22nd International Conference on*. 2006. IEEE.
36. Ingvaldsen, J.E. and J.A. Gulla. *Preprocessing support for large scale process mining of SAP transactions*. in *International Conference on Business process management*. 2007. Springer.
37. Ahmad, N.F., D.B. Hoang, and M.H. Phung. *Robust preprocessing for health care monitoring framework*. in *e-Health Networking, Applications and Services, 2009. Healthcom 2009. 11th International Conference on*. 2009. IEEE.
38. Figo, D., et al., *Preprocessing techniques for context recognition from accelerometer data*. Personal and Ubiquitous Computing, 2010. **14**(7): p. 645-662.
39. Carvalho, B.S. and R.A. Irizarry, *A framework for oligonucleotide microarray preprocessing*. Bioinformatics, 2010. **26**(19): p. 2363-2367.
40. Messaoud, I.B., et al. *A design of a preprocessing framework for large database of historical documents*. in *Proceedings of the 2011 Workshop on Historical Document Imaging and Processing*. 2011. ACM.

41. Liu, K., et al. *Clinical data preprocessing and case studies of POMDP for TCM treatment knowledge discovery*. in *e-Health Networking, Applications and Services (Healthcom), 2012 IEEE 14th International Conference on*. 2012. IEEE.
42. Zliobaite, I. and B. Gabrys, *Adaptive preprocessing for streaming data*. IEEE transactions on knowledge and data Engineering, 2014. **26**(2): p. 309-321.
43. Massouh, N., et al. *Experimental study on luminance preprocessing for energy-aware HTTP-based mobile video streaming*. in *Visual Information Processing (EUVIP), 2014 5th European Workshop on*. 2014. IEEE.
44. Danubianu, M. *A data preprocessing framework for students' outcome prediction by data mining techniques*. in *System Theory, Control and Computing (ICSTCC), 2015 19th International Conference on*. 2015. IEEE.
45. Ghayekhloo, M., M. Menhaj, and M. Ghofrani, *A hybrid short-term load forecasting with a new data preprocessing framework*. Electric Power Systems Research, 2015. **119**: p. 138-148.
46. Shi, W., et al., *An integrated data preprocessing framework based on Apache spark for fault diagnosis of power grid equipment*. Journal of Signal Processing Systems, 2017. **86**(2-3): p. 221-236.
47. Sreenivas, P. and C. Srikrishna. *An analytical approach for data preprocessing*. in *Emerging Trends in Communication, Control, Signal Processing & Computing Applications (C2SPCA), 2013 International Conference on*. 2013. IEEE.
48. Wang, M. and J. Wang. *A data preprocessing framework of geoscience data sharing portal for user behavior mining*. in *Geoinformatics, 2015 23rd International Conference on*. 2015. IEEE.
49. Frantz, D., et al., *An operational radiometric landsat preprocessing framework for large-area time series applications*. IEEE Transactions on Geoscience and Remote Sensing, 2016. **54**(7): p. 3928-3943.
50. Dubey, A.K., A. Kumar, and R. Agrawal, *An efficient ACO-PSO-based framework for data classification and preprocessing in big data*. Evolutionary Intelligence, 2021. **14**(2): p. 909-922.
51. Triguero, I., et al., *Transforming big data into smart data: An insight on the use of the k-nearest neighbors algorithm to obtain quality data*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2019. **9**(2): p. e1289.
52. Zasoba, E., et al., *Architecture of the Platform for Big Data Preprocessing and Processing in Medical Sector*. 2020.
53. Alcalde-Barros, A., et al., *DPASF: a flink library for streaming data preprocessing*. Big Data Analytics, 2019. **4**(1): p. 1-17.
54. Andreoni Lopez, M., et al., *Toward a monitoring and threat detection system based on stream processing as a virtual network function for big data*. Concurrency and Computation: Practice and Experience, 2019. **31**(20): p. e5344.
55. Ma, B.B., S. Fong, and R. Millham. *Data stream mining in fog computing environment with feature selection using ensemble of swarm search algorithms*. in *2018 Conference on Information Communications Technology and Society (ICTAS)*. 2018. IEEE.
56. García-Gil, D., et al., *Enabling smart data: noise filtering in big data classification*. Information Sciences, 2019. **479**: p. 135-152.
57. Vamsi, B., et al., *Comparative Analysis of Prevalent Disease by Preprocessing Techniques Using Big Data and Machine Learning: An Extensive Review*. Machine Intelligence and Soft Computing, 2021: p. 27-38.

58. Chakrabarty, S., et al. *Preprocessing of clinical neuro-oncology MRI studies for big data applications*. in *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*. 2020. International Society for Optics and Photonics.
59. Peralta, D., et al., *Evolutionary feature selection for big data classification: A mapreduce approach*. *Mathematical Problems in Engineering*, 2015. **2015**.
60. Triguero, I., et al., *ROSEFW-RF: the winner algorithm for the ECBDL'14 big data competition: an extremely imbalanced big data bioinformatics problem*. *Knowledge-Based Systems*, 2015. **87**: p. 69-79.
61. Zhao, Z., et al., *Massively parallel feature selection: an approach based on variance preservation*. *Machine learning*, 2013. **92**(1): p. 195-220.
62. Tan, M., I.W. Tsang, and L. Wang, *Towards ultrahigh dimensional feature selection for big data*. 2014.
63. Singh, S., et al. *Parallel large scale feature selection for logistic regression*. in *Proceedings of the 2009 SIAM International Conference on Data Mining*. 2009. SIAM.
64. Ordozgoiti, B., S.G. Canaval, and A. Mozo. *Massively parallel unsupervised feature selection on spark*. in *East European Conference on Advances in Databases and Information Systems*. 2015. Springer.
65. Wang, J., et al., *Online feature selection and its applications*. *IEEE Transactions on Knowledge and Data Engineering*, 2014. **26**(3): p. 698-710.
66. Kumar, M. and S.K. Rath, *Classification of microarray using MapReduce based proximal support vector machine classifier*. *Knowledge-Based Systems*, 2015. **89**: p. 584-602.
67. Sun, Z. and Z. Li. *Data intensive parallel feature selection method study*. in *Neural Networks (IJCNN), 2014 International Joint Conference on*. 2014. IEEE.
68. Kai, C., W.-q. Wan, and L. Yun, *Differentially private feature selection under mapreduce framework*. *The Journal of China Universities of Posts and Telecommunications*, 2013. **20**(5): p. 85103-90.
69. Chao, P., W. Bin, and D. Chao. *Design and implementation of parallel term contribution algorithm based on mapreduce model*. in *Open Cirrus Summit (OCS), 2012 Seventh*. 2012. IEEE.
70. He, Q., et al. *Parallel feature selection using positive approximation based on mapreduce*. in *Fuzzy Systems and Knowledge Discovery (FSKD), 2014 11th International Conference on*. 2014. IEEE.
71. Tanupabrungsun, S. and T. Achalakul. *Feature reduction for anomaly detection in manufacturing with mapreduce GA/kNN*. in *Parallel and Distributed Systems (ICPADS), 2013 International Conference on*. 2013. IEEE.
72. Dalavi, M. and S. Cheke. *Hadoop MapReduce implementation of a novel scheme for term weighting in text categorization*. in *Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014 International Conference on*. 2014. IEEE.
73. Meena, M.J., et al., *An enhanced ACO algorithm to select features for text categorization and its parallelization*. *Expert Systems with Applications*, 2012. **39**(5): p. 5861-5871.
74. Hodge, V.J., S. O'Keefe, and J. Austin, *Hadoop neural network for parallel and distributed feature selection*. *Neural Networks*, 2016. **78**: p. 24-35.
75. Del Río, S., et al., *On the use of MapReduce for imbalanced big data using Random Forest*. *Information Sciences*, 2014. **285**: p. 112-137.

76. del Río, S., J.M. Benítez, and F. Herrera. *Analysis of data preprocessing increasing the oversampling ratio for extremely imbalanced big data classification*. in *Trustcom/BigDataSE/ISPA, 2015 IEEE*. 2015. IEEE.
77. Triguero, I., et al. *Evolutionary undersampling for extremely imbalanced big data classification under apache spark*. in *Evolutionary Computation (CEC), 2016 IEEE Congress on*. 2016. IEEE.
78. Triguero, I., et al. *Evolutionary undersampling for imbalanced big data classification*. in *Evolutionary Computation (CEC), 2015 IEEE Congress on*. 2015. IEEE.
79. Park, S.H. and Y.G. Ha. *Large imbalance data classification based on mapreduce for traffic accident prediction*. in *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2014 Eighth International Conference on*. 2014. IEEE.
80. Wang, X., X. Liu, and S. Matwin. *A distributed instance-weighted SVM algorithm on large-scale imbalanced datasets*. in *Big Data (Big Data), 2014 IEEE International Conference on*. 2014. IEEE.
81. Park, S.-h., S.-m. Kim, and Y.-g. Ha, *Highway traffic accident prediction using VDS big data analysis*. *The Journal of Supercomputing*, 2016. **72**(7): p. 2815-2831.
82. Bhagat, R.C. and S.S. Patil. *Enhanced SMOTE algorithm for classification of imbalanced big-data using Random Forest*. in *Advance Computing Conference (IACC), 2015 IEEE International*. 2015. IEEE.
83. Zhai, J., S. Zhang, and C. Wang, *The classification of imbalanced large data sets based on mapreduce and ensemble of elm classifiers*. *International Journal of Machine Learning and Cybernetics*, 2017. **8**(3): p. 1009-1017.
84. Zhang, J., et al., *A parallel matrix-based method for computing approximations in incomplete information systems*. *IEEE Transactions on Knowledge and Data Engineering*, 2015. **27**(2): p. 326-339.
85. Chen, F. and L. Jiang. *A parallel algorithm for datacleansing in incomplete information systems using mapreduce*. in *Computational Intelligence and Security (CIS), 2014 Tenth International Conference on*. 2014. IEEE.
86. Ramírez-Gallego, S., et al. *Distributed entropy minimization discretizer for big data analysis under apache spark*. in *Trustcom/BigDataSE/ISPA, 2015 IEEE*. 2015. IEEE.
87. Ramírez-Gallego, S., et al., *Data discretization: taxonomy and big data challenge*. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2016. **6**(1): p. 5-21.
88. Zhang, Y., J. Yu, and J. Wang. *Parallel implementation of chi2 algorithm in mapreduce framework*. in *International Conference on Human Centered Computing*. 2014. Springer.
89. Triguero, I., et al., *MRPR: a MapReduce solution for prototype reduction in big data classification*. *neurocomputing*, 2015. **150**: p. 331-345.
90. Krempl, G., et al., *Open challenges for data stream mining research*. *ACM SIGKDD explorations newsletter*, 2014. **16**(1): p. 1-10.
91. Goldstein, J., et al., *Null stream analysis of Pulsar Timing Array data: localisation of resolvable gravitational wave sources*. arXiv preprint arXiv:1712.03975, 2017.
92. Lohmar, T., D. Catrein, and T. Einarsson, *Streaming with optional broadcast delivery of data segments*. 2017, Google Patents.
93. Gholmieh, R.A., et al., *Further Device Timing Adjustments and Methods for Supporting DASH Over Broadcast*. 2016, Google Patents.

94. Mao, Y., F. Ulupinar, and D.W. Gillies, *Determining available media data for network streaming*. 2016, Google Patents.
95. Spark, A., *Apache spark*. Retrieved January, 2018. **17**: p. 2018.
96. Meng, X., et al., *Mllib: Machine learning in apache spark*. arXiv preprint arXiv:1505.06807, 2015.
97. Garg, N., *Apache kafka*. 2013: Packt Publishing Birmingham.
98. George, L., *HBase: the definitive guide*. 2011: " O'Reilly Media, Inc."
99. Lam, C., *Hadoop in action*. 2010: Manning Publications Co.
100. Han, J., et al. *Survey on NoSQL database*. in *Pervasive computing and applications (ICPCA), 2011 6th international conference on*. 2011. IEEE.
101. Hive, A., *Apache Hive*. 2013.
102. Schelter, S., *Large scale graph processing with apache giraph*. Invited talk at GameDuell Berlin 29th May, 2012.
103. Huai, Y., et al. *Major technical advancements in apache hive*. in *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. 2014. ACM.
104. Miller, J.J. *Graph database applications and concepts with Neo4j*. in *Proceedings of the southern association for information systems conference, Atlanta, GA, USA*. 2013.
105. Cassandra, A., *Apache cassandra*. Website. Available online at <http://planetcassandra.org/what-is-apache-cassandra>, 2014. **13**.
106. Anil, R., et al., *Apache Mahout: Machine Learning on Distributed Dataflow Systems*. *J. Mach. Learn. Res.*, 2020. **21**: p. 127:1-127:6.
107. Hoffman, S., *Apache Flume: distributed log collection for Hadoop*. 2013: Packt Publishing Ltd.
108. Correia, J., C. Costa, and M.Y. Santos. *Challenging SQL-on-Hadoop Performance with Apache Druid*. in *International Conference on Business Information Systems*. 2019. Springer.
109. Vohra, D., *Apache Sqoop*, in *Practical Hadoop Ecosystem*. 2016, Springer. p. 261-286.
110. Shahi, D., *Apache Solr*. 2016: Springer.
111. CouchDB, A., *Apache CouchDB*. URL <https://couchdb.apache.org>.
112. Iqbal, M.H. and T.R. Soomro, *Big data analysis: Apache storm perspective*. *International journal of computer trends and technology*, 2015. **19**(1): p. 9-14.
113. Shin, J.E., B.H. Jung, and D.H. Lim, *Big data distributed processing system using RHadoop*. *Journal of the Korean Data and Information Science Society*, 2015. **26**(5): p. 1155-1166.
114. Felleisen, M. *On the expressive power of programming languages*. in *European Symposium on Programming*. 1990. Springer.
115. Sozeau, M. and N. Oury. *First-class type classes*. in *International Conference on Theorem Proving in Higher Order Logics*. 2008. Springer.
116. Grgic, H., B. Mihaljević, and A. Radovan. *Comparison of garbage collectors in Java programming language*. in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. 2018. IEEE.
117. Ogori, A. and P. Buneman. *Type inference in a database programming language*. in *Proceedings of the 1988 ACM conference on LISP and functional programming*. 1988.
118. Garey, M.R., D.S. Johnson, and L. Stockmeyer, *Some simplified NP-complete graph problems*. *Theoretical computer science*, 1976. **1**(3): p. 237-267.

119. Karger, D.R., P.N. Klein, and R.E. Tarjan, *A randomized linear-time algorithm to find minimum spanning trees*. Journal of the ACM (JACM), 1995. **42**(2): p. 321-328.
120. Esposito, R. and L. Saitta. *Boosting as a Monte Carlo algorithm*. in *Congress of the Italian Association for Artificial Intelligence*. 2001. Springer.
121. Karaboga, D. and B. Basturk. *Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems*. in *International fuzzy systems association world congress*. 2007. Springer.
122. Bezdek, J.C., R. Ehrlich, and W. Full, *FCM: The fuzzy c-means clustering algorithm*. Computers & geosciences, 1984. **10**(2-3): p. 191-203.
123. Chang, C.-C. and C.-J. Lin, *LIBSVM: a library for support vector machines*. ACM transactions on intelligent systems and technology (TIST), 2011. **2**(3): p. 1-27.
124. Griffin, G., A. Holub, and P. Perona, *Caltech-256 object category dataset*. 2007.
125. Winn, J. and N. Jojic. *Locus: Learning object classes with unsupervised segmentation*. in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*. 2005. IEEE.
126. Gosain, A. and S. Sardana. *Handling class imbalance problem using oversampling techniques: A review*. in *2017 international conference on advances in computing, communications and informatics (ICACCI)*. 2017. IEEE.
127. Zhang, Q., et al., *PPHOPCM: Privacy-preserving high-order possibilistic c-means algorithm for big data clustering with cloud computing*. IEEE Transactions on Big Data, 2017.
128. *CityPulse Smart City Datasets*. 2017; Available from: <http://iot.ee.surrey.ac.uk:8080/index.html>.
129. Kalman, R.E., *A new approach to linear filtering and prediction problems*. 1960.