Table 4.4: Similarity Threshold Value For Separate The Relevant Features

Threshold for selection of useful features	Similarity threshold value	Cluster quality	
		relief+knn	relief+parzen
0.1	0.90	0.5558	0.6066
0.12	0.80	0.6032	0.5131
0.203	0.70	0.6319	0.4881
0.180	0.65	0.7237	0.6579
0.254	0.543	0.7040	0
0.30	0.44	0.8371	0.10
0.497	0.10	1.2744	0.09

5. Discussion

In this paper we are applying algorithms on discontinuous raw data. This paper mainly uses the combinations of Relief and kNN algorithms. The output of this paper is clustered data. There is a scope for performing same operations on continuous data, that is another scenario.

6. Conclusion

In this work, comparing two feature weighting algorithms. So the selected relevant features are showing in clusters by using some clustering algorithms for better validation, limitations of the well known clustering techniques for large datasets and details of the proposed clustering method, Leaders-Subleaders have been presented. Our experimental results on numerical datasets show that Leaders-Subleaders algorithm performs well. The representatives of the Subleaders help in improving the classification accuracy. Devies-Bouldin index showed a good performance to the results were equivalent, even with different radius.

References

[1] Guangtao Wang, Qinbao Song, Heli Sun, Xueying Zhang, "A Feature Subset Selection Algorithm Automatic Recommendation Method." Journal of Artificial Intelligence Research 47 (2013) 1-34 Submitted 10/2012; published 05/2013
[2] Marjan Kuchaki Rafsanjani, Zahra Asghari Varzaneh, Nasibeh Emani Chukanlo, "A Survey of Hierarchical Clustering Algorithms." The Journal of Mathematics and Computer Science Vol .5 No.3 (2012) 229-240
[3] Yogita Rani, Dr.Harish Rohil, "A Study of Hierarchical Clustering Algorithm." International Journal of Information and Computation Technology. ISSN 0974- 2239 Volume 3, Number 11 (2013), pp. 1225-1232 © International Research Publications House
[4] N.Sudhakar Reddy, KVN Sunitha, "Efficient Clustering Algorithm for Large Dataset." Volume 2, Issue 1, January 2012 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering.
[5] Ankita Choubey, Dr.Sadhna K.Mishra, "Enhanced Clustering Algorithm for Processing Online Data" IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278- 8727Volume 12, Issue 5 (Jul. - Aug. 2013), PP 24-29
[6] Srinivasulu M, Kotilinswara Rao, Baji Mohammed, RVSP Kumar, Data Clustering With Leaders and Subleaders Algorithm." IOSR Journal of Engineering (IOSRJEN) e- ISSN: 2250-3021, p-ISSN: 2278-8719, Volume 2, Issue 11 (November2012), PP 01-07
[7] Kagan Tumer, Joydeep Ghosh, "Bayes Error Rate Estimation Using Classifier Ensembles"
[8] Ali Mustafa Qamar, Eric Gaussier, "RELIEF Algorithm and Similarity for k-NN". International Journal of Computer Information Systems and Industrial Management Applications. ISSN 2150-7988 Volume 4 (2012) pp. 445-458 MIR Labs.
[9] Yijun Sun, Jian Li, "Iterative RELIEF for Feature Weighting." Interdisciplinary Center for Biotechnology Research,Department of Electrical & Computer Engineering, University of Florida, Gainesville, FL 32611, USA
[10] Marko Robnik-Sikonja, Igar Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF." University of Ljubljana, Faculty of Computer and Information Science, Tr˘za˘ska 25, 1001 Ljubljana, Slovenia
[11] Shuang-Hong Yang and Bao-Gang Hu, Senior Member, IEEE, "Discriminative Feature Selection by Nonparametric Bayes Error Minimization."IEEE Transactions on knowledge and data engineering, vol.24, no.8, August 2012
[12] Kenji Kira, Larry A.Rendell, "The Feature Selection Problem: Traditional Methods and a New Algorithm." From: AAAI-92 Proceedings. Copyright ©1992, AAAI. All rights reserved.
[13] Yuxuan SUN, Xiaojun LOU, Bisai BAO, "A Noval Relief Feature Selection Algorithm Based on Mean-Variance Model." Journal of Information & Computational Science 8: 16 (2011) 3921–3929.
[14] K.Saranya, T.Deepa, "Discriminative Clustering Based Feature Selection and Nonparametric Bayes Error Minimization and Support Vector Machine(SVM)." International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064.

Author Profile

Mr. Choudari Anantha is currently pursuing his M.Tech(SE) in Department of Computer Science and Engineering at Vignan Institute of Information Technology, Visakhapatnam. He is Received B.Tech (IT) From Swarnandhra College of Engineering and Technology, Narsapuram, West Godavari; A.P. His area of interests includes Data Mining and Information Security.

Dr. Devarapalli Dharmaiah is currently working as an Associate Professor in Computer Science and Engineering department, Vignan Institute of Information Technology, Visakhapatnam and the teaching and research experience of about 11 years. He also guided various dissertation works for both UG and PG students of VIGNAN'S IIT and other Colleges. He has taught various subjects of Computer Science and Engineering for both of UG & PG students such as C programming, Data Structures, Java, Operating Systems, Compler Design, Linux, UNIX, Data Structures and Bioinformatics. He has a very good reputation among the students and faculty community for his proficiency in subjects. He is a life member of CSI. He has published many papers in National, International conferences and leading International Journals. His area of interest is Bioinformatics, Neural Networks, Data mining and computer networks. He received Best Engineering Teachers award 2013, PRATIBHA AWARD, Sasatra Award, Best Project Award in Computer Science and Engineering Department.