Possibilistic rough fuzzy c-means algorithm in data clustering and inlage segmentation

Article · March 2014

3 authors, including:

B.K. Tripathy
Vellore Institute of Technology University

Anurag Tripathy

Possibilistic Rough Fuzzy C-Means Algorithm in Data Clustering and Image Segmentation

B.K.Tripathy
SCSE, VIT University,Vellore-632014,
TN, India
tripathybk@vit.ac.in

Anurag Tripathy
SCSE, VIT University,Vellore-632014,
TN, India
anurag7642@gmail.com

Kosireddy Govinda Rajulu
VITAM,Berhampur
Odisha, India
govinda_rajulu@rediffmail.com

***Abstract*—Several data clustering techniques have been developed in literature. It has been observed that the algorithms developed by using imprecise models like rough sets, fuzzy sets and intuitionistic fuzzy sets have been better than the crisp algorithms. Also, the hybrid models provide far better clustering algorithms than the individual models. Several such models have been developed by using a combination of fuzzy set introduced by Zadeh, the rough set introduced by Pawlak and the intuitionistic fuzzy introduced by Atanassov. Notable among them being the Rough Fuzzy C-Means (RFCM) introduced by Mitra et al and the rough intuitionistic fuzzy c-means algorithm (RIFCM) introduced and studied by Tripathy et al Krishnapuram and Keller observed that the basic clustering algorithms have the probabilistic flavour; for example due to the presence of the constraint on the memberships used in the fuzzy C-Means (FCM) algorithm. So, they introduced the concept of possibilistic approach and developed a possibilistic fuzzy C-means (PFCM) algorithm. Another approach to PFCM is due to Pal et al. In this paper, we improve a possibilistic rough C-Means (PRCM) algorithm introduced by Anuradha et,al. and introduce a new algorithm, which we call as possibilistic rough fuzzy C-Means (PRFCM) and compare its efficiency with the improved PRCM and the basic PRFCM algorithm to establish experimentally that this algorithm is comparatively better than PRCM and the corresponding RCM algorithm. We perform the experimental analysis by taking different types of numerical datasets and images as inputs and using standard accuracy measures like the DB and the D-index.**

***Keywords*—clustering; fuzzy sets; rough sets; rough fuzzy sets;possibilistic approach;DB index; D index**

I. INTRODUCTION (*Heading 1*)

A cluster consists of a set of objects which are similar to each other with respect to certain criteria. Clustering is the process of dividing a given dataset into several clusters. Putting similar objects into groups called clusters and putting dissimilar objects into different clusters. In the other parallel concept of classification a large set of training tuples patterns is necessary to model the groups and label them. In contrast the labeling is done after the clustering process is over, which leads to partitioning the objects into groups. The necessity of clustering arises from the fact that it is extremely difficult to analyze the large volumes of data generated from several sources every day before performing any specific task with them. This necessitates the use of some approach to classify the data so that only the classes of interest can be used for analysis instead of considering the entire dataset which may be huge. Clustering is used in several application areas, like statistical data analysis, machine learning, pattern recognition, image analysis, information retrieval and bioinformatics. It is widely used in exploratory data mining. Over the years several clustering algorithms have been developed, studied and applied in various fields of application.

Algorithms using crisp techniques cannot be applied to datasets having imprecision inherent in them. As a result imprecision based clustering techniques have been developed over the years [1]. The imprecise models used for this purpose are fuzzy set introduced by Zadeh [2], rough set by Pawlak [3], intuitionistic fuzzy set by Atanassov [4]. The corresponding algorithms are the fuzzy c-means (FCM) [5, 6], rough c-means (RCM) [7] and the intuitionistic fuzzy c-means (IFCM) [8]. It is observed that the hybrids of individual models are more efficient than the individual models [9, 10]. Hybrid clustering algorithms by combining fuzzy sets, rough sets or intuitionistic fuzzy sets have been proposed and studied. Some of these algorithms are the rough fuzzy c-means (RFCM) [11, 12] and the rough fuzzy intuitionistic fuzzy c-means (RIFCM) [13]. As expected the RIFCM performs better than all the other algorithms in this family like the FCM, RCM and RFCM as established in [13] through several experiments. They have used efficiency measuring indices like the Davies-Bouldin (DB) [14] and Dunn (D) indexes [15] to compare the accuracy of performance of these algorithms by using different types of images and datasets for the purpose. Many applications of hybrid algorithms can be found in [16, 17, 18, 19].

In the FCM algorithm it is assumed that the membership of data points across the classes must sum to 1. This constraint is meant to avoid trivial solution of all memberships being equal to 0 and it gives meaningful results in applications where it is appropriate to interpret memberships as probabilities or degrees of sharing. Since the memberships generated by this constraint are relative numbers, they are not suitable for applications in which the memberships are supposed to

978-1-4799-3975-6/14/$31.00

Table 2: D values for the two images

	Football Player			Metal coin image		
	3	4	5	3	4	5
PRCM	0.1368	0.14715	0.14619	0.13481	0.13486	0.17454
PRFCM	0.14045	0.15025	0.19438	0.17153	0.1532	0.17312
RCM	0.0862	0.0914	0.1219	0.12481	0.099137	0.0845
RFCM	0.09467	0.0928	0.1421	0.13832	0.1291	0.12775

B. Numeric Dataset

In table 3 and table 4 below, we present the values of the DB and D indices obtained by taking the soybean and the zoo dataset for the four algorithms; RCM, RFCM, PRCM and PRFCM. It has been observed that the DB value for PRFCM is the lowest as well as the D value for PRFCM is the highest. Also, figure 5 and 6 provide the graphical representation of their performances.

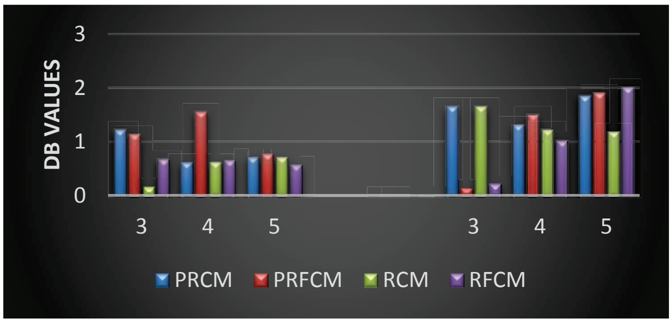

Figure 3: Bar diagram for Table 1

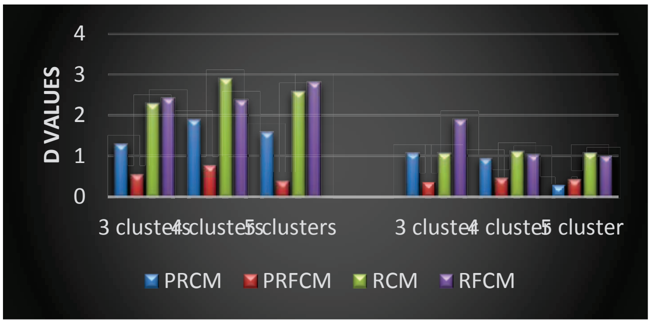

Figure 4: Bar diagram for Table 2

Table 3: DB values for the two numeric datasets

	Soybean			Zoo		
	3	4	5	3	4	5
PRCM	1.2916396	1.90339	1.58309	1.06987	0.9441	0.2824208
PRFCM	0.5514072	0.767325	0.3852521	0.359305	0.450571	0.41906
RCM	2.291639	2.903385	2.583098	1.0698706	1.111279	1.068061
RFCM	2.4163548	2.3751442	2.8128	1.89695	1.03732	1.00235

Table 4: D values for the two numeric datasets

	Soybean			Zoo		
	3	4	5	3	4	5
PRCM	1.227	0.62119	0.71839	1.66123	1.31483	1.85842
PRFCM	1.136	1.5623	0.7736	0.14025	1.50571	1.906
RCM	0.166	0.62119	0.71839	1.66123	1.2269	1.18773
RFCM	0.682	0.66084	0.57043	0.21735	1.0232	2.012

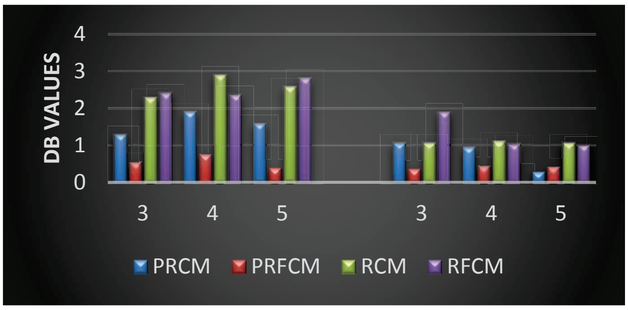

Figure 5: Bar diagram for Table 3

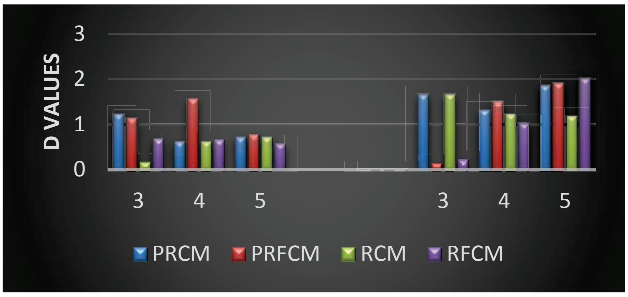

Figure 6: Bar diagram for Table 4

V. CONCLUSION

In this paper we concentrated on the possibilistic algorithms for data clustering. For this purpose, we improved an earlier algorithm developed by Anuradha et al on PRCM and put forth a new algorithm called the PRFCM. Through experimental analysis we established that in addition to the possibilistic algorithms performing better than the basic algorithms, the PRFCM provides the best performance. We have used two datasets (Zoo and Soybean) from the UCI repository and two images (A football player and a metal coin) as the inputs and use two standard accuracy measures (DB and D index) in support of our observations.

REFERENCES

[1] Swarnalatha P, Tripathy B.K., Nithin, P. L. and D. Ghosh, "Cluster Analysis Using Hybrid Soft Computing Techniques", Proceedings of Fifth CNC-2014, (2014), pp. 516-524.

[2] Zadeh, L. A. ," Fuzzy Sets, Information and Control", vol.8, no.11, (1965), pp. 338 – 353.

[3] Pawlak, Z. ," "Rough sets", Int. jour. of Computer and Information Sciences, 11, (1982), pp.341-356.

[4] Atanassov, K.T. ," Intuitionistic Fuzzy Sets", Fuzzy sets and Systems 20, 1, (1986), pp. 87-96. 1986.

[5] Bezdek, J.C. ," Pattern Recognition with Fuzzy Objective Function Algorithms", Kluwer Academic Publishers, 1981.

[6] Ruspini, Enrique H. ," A new approach to clustering", Information and control, 15, 1, (1969), pp. 22-32.

[7] Lingras, P. and West, C. ," Interval set clustering of web users with rough k-mean", Journal of Intelligent Information Systems, 23(1), (2004), pp. 5-16.

[8] Chaira, T. and Anand, S. ," A Novel Intuitionistic Fuzzy Approach for Tumor/Hemorrhage Detection in Medical Images", Journal of Scientific and Industrial Research, 70(6), (2011).

[9] Dubois, D. and Prade, H. ," Rough fuzzy sets model", International journal of General Systems, vol. 46, no.1, (1990), pp. 191 – 208.

[10] Tripathy, B.K., Ghosh, S.K. and Jena, S.P, "Intuitionistic Fuzzy Rough Sets", Notes on Intuitionistic Fuzzy Sets (Bulgaria), 8, (2002), no.1, pp.1-18.

[11] Maji, P. and Pal, S.K. ," RFCM: A Hybrid Clustering Algorithm using rough and fuzzy set", Fundamenta Informaticae 80.4 (2007), pp.475-496.

[12] Mitra, S., Banka, H. and Pedrycz, W. ," Rough-Fuzzy Collaborative Clustering, System, Man, and Cybernetics", Part B: Cybernetics, IEEE Transactions on 36.4 (2006), pp.795-805

[13] Bhargav, R., Tripathy, B.K., Tripathy, A., Dhull, R, Verma, E. and Swarnalatha, P. ," Rough Intuitionistic Fuzzy C-Means Algorithm and a Comparative Analysis", ACM conference, Compute 2013, (2013), ACM 978-1-4503-2545-5/13/08.

[14] Davis, D. L. and Bouldin, D.W. ," A cluster separation measure, IEEE Transactions on Pattern Analysis and Machine Intelligence", vol.PAMI-1, no.2, (1979), pp.224 – 227.

[15] Dunn, J. C. ," A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters", (1973), pp. 32-57.

[16] Swarnalatha, P. and Tripathy, B.K.," A Comparative Study of RIFCM with Other Related Algorithms from Their Suitability in Analysis of Satellite Images using Other Supporting Techniques", Kybernetes, vol.43, no.1,(2014), pp. 53-81.

[17] Tripathy, B.K and Ghosh, A., "Data Clustering Algorithms using Rough sets", Handbook of research on Computational Intelligence for engineering, Science and Business, IGI Global Publications, (2013), pp.297-327

[18] B.K.Tripathy and R. Bhargav,"Kernel Based Rough-Fuzzy C-Means, PReMI, ISI Calcutta, December, LNCS 8251, (2013), pp.148-157

[19] Tripathy, B.K, Tripathy, A., Govindarajulu, K. and Bhargav, R., "On kernel Based rough Intuitionistic Fuzzy C-means algorithm and a comparative analysis, ICACNI 2014, Advanced computing networking and informatics, vol.1, Smart innovation systems and technologies, vol.27, (2014), pp. 349-359.

[20] Krishnapuram, R. and Keller, J. M.," A Possibilistic Approach to Clustering", IEEE Transaction on Fuzzy Systems, vol,1, no.2, 1993, pp.98 – 110.

[21] Krishnapuram, R. and Keller, J. M.," The Possibilistic C-Means Algorithm: Insights and recommendations", IEEE Transactions on Fuzzy Systems, vol.4, no.3, 1996, pp.385- 393.

[22] Pal, N.R., Pal, K., Keller, J. M. and Bezdek, J. C.," A Possibilistic Fuzzy C-Means Clustering Algorithm", IEEE Transactions on Fuzzy Systems, vol.13, no.4, 2005, pp.517-530.