P.O.Box 8718, Beijing 100080, China	Journal of Software, June 2004,15(6):858-869
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn	Copyright © 2004 by The Editorial Department of Journal of Software

基于划分的模糊聚类算法

张 敏, 于 剑

张 敏, 于 剑, (北京交通大学 计算机与信息技术学院,北京 100044)
作者简介: 张敏(1979－),女,江西南昌人,硕士生,主要研究领域为数据挖掘,聚类分析;于剑(1969－),博士,副教授,主要研究领域为计算智能,模式识别,数据挖掘.
联系人: 于 剑 Phn: +86-10-51688055, E-mail: jianyu@center.njtu.edu.cn, http://cit.njtu.edu.cn
Received 2003-07-16; Accepted 2003-11-11

Abstract
Fuzzy partitional clustering algorithms are widely used in pattern recognition field. Until now, more and more research results on them have been developed in the literature. In order to study these algorithms systematically and deeply, they are reviewed in this paper based on c-means algorithm, from metrics, entropy, and constraints on membership function or cluster centers. Moreover, the advantages and disadvantages of the typical fuzzy partitional algorithms are discussed. It is pointed out that the standard FCM algorithm is robust to the scaling transformation of dataset, while others are sensitive to such transformation. Such conclusion is experimentally verified when implementing the standard FCM and the maximum entropy clustering algorithm. Finally, the problems existing in these algorithms and the prospects of the fuzzy partitional algorithms are discussed.

Zhang M, Yu J. Fuzzy partitional clustering algorithms. *Journal of Software*, 2004,15(6):858~869.
http://www.jos.org.cn/1000-9825/15/858.htm

摘要
在众多聚类算法中,基于划分的模糊聚类算法是模式识别中最常用的算法类型之一.至今,文献中仍不断有关于基于划分的模糊聚类算法的研究成果出现.为了能更为系统和深入地了解这些聚类算法及其性质,本文从改变度量方式、改变约束条件、在目标函数中引入熵以及考虑对聚类中心进行约束等几个方面,对在C-均值算法的基础上得到的基于划分的模糊聚类算法作了综述和评价,对各典型算法的优缺点进行了实验比较分析.指出标准FCM算法被广泛应用的原因之一是它对数据的比例变化具有鲁棒性,而其他类似的算法对这种比例变化却很敏感,并以极大熵方法为例进行了比较实验.最后总结了基于划分的模糊聚类算法普遍存在的问题及其发展前景.

基金项目: Supported by the National Natural Science Foundation of China under Grant No.60303014 (国家自然科学基金); the Scientific Key Project of Ministry of Education of China under Grant No.02031 (教育部科学技术研究重点项目)

References:

[1] Han I, Kamber M. Data Mining: Concepts and Techniques. Berlin: Morgan Kaufmann Publishers, 2000. 335~389.

[2] Yang MS. A survey of fuzzy clustering. Mathl. Comput. Modelling, 1993,18(11):1~16.

[3] Gustafson DE, Kessel WC. Fuzzy clustering with a fuzzy covariance matrix. In: Proc. of the IEEE CDC. San Diego: IEEE, 1979. 761~766.

[4] Wu KL, Yang MS. Alternative c-means clustering algorithms. Pattern Recognition, 2002,35(10):2267~2278.

[5] Yang MS. On a class of fuzzy classification maximum likelihood procedures. Fuzzy Sets and Systems, 1993,57(3):365~375.

[6] Lin JS. Fuzzy clustering using a compensated fuzzy hopfield network. Neural Processing Letters, 1999,10(1):35~48.

[7] Ozdemir D, Akarun L. A fuzzy algorithm for color quantization of images. Pattern Recognition, 2002,35(8):1785~1791.

[8] Menard M, Courboulay V, Dardignac P. Possibilistic and probabilistic fuzzy clustering: Unification within the framework of the non-extensive thermostatistics. Pattern Recognition, 2003,36(6):1325~1342.

[9] Hathaway RJ, Bezdek JC, Hu YK. Generalized fuzzy c-means clustering strategies using Lp norm distances. IEEE Trans. on Fuzzy Systems, 2000,8(5):576~582.

[10] Bobrowski L, Bezedek JC. C-means clustering with the l1 and l(norms. IEEE Trans. on SMC, 1991,21(3):545~554.

[11] Dave RN, Bhaswan K. Adaptive fuzzy c-shells clustering and detection of ellipses. IEEE Trans. on Neural Networks, 1992,3(5): 643~662.

[12] Krishnapuram R, Nasraoui O, Frigui H. The fuzzy c spherical shells algorithm: A new approach. IEEE Trans. on Neuaral Network, 1992,3(5):663~671.

[13] Kaymak U, Setne M. Fuzzy clustering with volume prototypes and adaptive cluster merging. IEEE Trans. on Fuzzy Systems, 2002, 10 (6):706~712.

[14] Yang MS, Wu KL, Yu J. A novel fuzzy clustering algorithm. In: Proc. of the 2003 IEEE Int'l Symp. on Computational Intelligence in Robotics and Automation. Kobe: IEEE, 2003. 647~652.

[15] Dave RN. Characterization and detection of noise in clustering. Pattern Recognition Letters, 1991,12(11):657~664.

[16] Frigui H, Krishnapu R. A robust algorithm for automatic extraction of an unknown number of clusters from noisy data. Pattern Recognition Letters, 1996,17(12):1223~1232.

[17] Cherkassky V, Mulier F. Learning From Data Concepts, Theory, and Methods. John Wiley & Sons, Inc., 1998. 413~417.

[18] Pedrycz W. Conditional fuzzy c-means. Pattern Recognition Letters, 1996,17(6):625~632.

[19] Barni M, Cappellini V, Mecocci A. Comments on 'A possibilistic approach to clustering'. IEEE Trans. on Fuzzy Systems, 1996,4 (3):393~396.

[20] Krishnapuram R, Keller JM. A possibilistic approach to clustering. IEEE Trans. on Fuzzy Systems, 1993,1(2):98~110.

[21] Schneider A. Weighted possibilistic clustering algorithms. In: Proc. of the 9th IEEE Int'l Conf. on Fuzzy Systems. Texas: IEEE, 2000,1:176 ~180.

[22] Krishnapuram R, Keller JM. The possiblistic means algorithms: Insights and recommendation. IEEE Trans. on Fuzzy Systems, 1996,4 (3):98~110.

[23] Karayiannis NB. Generalized fuzzy c-means algorithms. In: Proc. of the 5th IEEE Int'l Conf. on Fuzzy Systems, Vol 2. New Orleans: IEEE, 1996. 1036~1042.

[24] Pal RN, Pal K, Bezedek JC. A mixed c-means clustering model. In: Proc. of the 6th IEEE Int'l Conf. on Fuzzy Systems, Vol 1. Barcelona：IEEE, 1997. 11~21.

[25] Liu SH, Lin JS. Vector quantization in DCT domain using fuzzy possibilistic c-means based on penalized and compensated constraints. Pattern Recognition, 2002,35(10):2201~2211.

[26] Li RP, Mukaidon M. A maximum entropy approach to fuzzy clustering. In: Proc. of the 4th IEEE Int'l Conf. on Fuzzy System. Yokohama: IEEE, 1995. 2227~2232.

[27] Karayiannis NB. MECA: Maximum entropy clustering algorithm. IEEE World Congress on Computational Intelligence, Vol 1. In: Proc. of the 3rd IEEE Conf. on Fuzzy Systems, Vol 2. Orlando: IEEE, 1994. 630~635.

[28] Ichihashi H, Miyagishi K, Honda K. Fuzzy C-means clustering with regularization by K-L information. In: Proc. of the 10th IEEE Int'l Conf. on Fuzzy Systems, Vol 2. Melbourne: IEEE, 2001. 924~927.

[29] Yasuda M, Furuhashi T, Matsuzaki M, Okuma S. A study on statistical mechanical characeteristics of fuzzy clustering. In: Proc. of IEEE Int'l Conf. on Systems, Man and Cybemetics, Vol 4. Tucson: IEEE, 2001. 2415~2420.

[30] Wei C, Fahn C. The multisynapse neural network and its application to fuzzy clustering. IEEE Trans. on Neural Networks, 2002, 13 (3):600~618.

[31] Rose K. Deterministic annealing for clustering, compression, classification, regression and Related optimization problems. Proc. of the IEEE, 1998,86(11):2210~2239.

[32] Rose K, Gurewitz E, Fox GC. Constrained clustering as optimization method. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1993,15(8):785~794.

[33] Yasuda M, Furnuhashi T, Okuma S. Fuzzy clustering using deterministic annealing method and its statistical mechanical characteristics. In: Proc. of the 10th IEEE Int'l Conf. on Fuzzy Systems, Vol 2. Melbourne: IEEE. 2001. 797~800.

[34] Ozdemir D, Akarun L. Fuzzy algorithms for combined quantization and dithering. IEEE Trans. on Image Processing, 2001,10(6): 923~931.

[35] Timm H, Kruse R. A modification to improve possibilistic fuzzy cluster analysis. In: Proc. of the 2002 IEEE Int'l Conf. on Fuzzy Systems, Vol 2. Honululu: IEEE, 2002. 1460~1465.

[36] Timm H, Borgelt C, Dorring C, Kruse R. Fuzzy cluster analysis with cluster repulsion. In: Proc. of the Eruopean Symp. on Intelligent Technologies, Tenerife. 2001. CD-ROM

[37] Yu J, Yang MS. A study on generalized FCM. RSFDGrC Lecture Notes in Computer Science, 2003,2639:390~393.

[38] Yu J, Huang HK. A new weighting fuzzy c-means algorithms. In: Proc. of the 12th IEEE Int'l Conf. on Fuzzy Systems, Vol 2. St. Louis: IEEE, 2003. 896~901.

[39] Yu J. On the Fuzziness index of the FCM algorithm. Chinese Journal of Computers, 2003,26(8):974~981 (in Chinese with English abstract).

[40] Yu J, Cheng QS. A note on the weighting exponent m of FCM algorithm. Acta Electronica Sinica, 2003,31(3):478~480 (in Chinese with English abstract).

[41] Gao XB, Xie WX. Advances in theory and application of fuzzy clustering. Chinese Science Bulletin, 1999,44(22):961~969 (in Chinese with English abstract).

[42] Jain AK, Mutty MN, Flynn PJ. Data clustering: A review. ACM Computing Surveys, 1999,31(3):264~323.

[43] Yu J. Cluster validity and its application [Ph.D. Thesis]. Beijing: School of Mathematic Science, Peking University, 2000 (in Chinese with English abstract).

[44] Selim SZ, Ismail MA. K-means-type algorithms: A generalized convergence theorem and characterization of local optimality. IEEE Trans. on PAMI, 1984,6(1):81~86.

[45] Bezdek JC, Hathaway RJ, Sabin MJ, Tucker W. Convergence theory for fuzzy c-means: Counter-examples and repairs. IEEE Trans. on Systems, Man, and Cybernetics-part B: Cybernetics, 1987,17(5):873~877.

[46] Bezedek JC. A convergence theorem for the fuzzy ISODATA clustering algorithms. IEEE Trans. on PAMI, 1980,2(1):1~7.

[47] Yu J, Shi HB, Huang HK. Counterexamples to convergence theorem of maximum entropy clustering algorithm. Science in China, Series F, 2003,46(5):321~326.

[48] Miyamoto S, Agusta Y. Algorithms for L1 and Lp fuzzy c-means and their convergence. In: Proc. of the IFCS'96. 1996. 295~302.

[49] Yu J, Huang HK, Tian SF. An efficient optimality test for the fuzzy c-means algorithm. In: Proc. of the 2002 IEEE Int'l Conf. on Fuzzy Systems. Honululu: IEEE, 2002,1(2):98~103.

[50] Wei W, Mendel JM. Optimality tests for the fuzzy c-means algorithm. Pattern Recognition, 1994,27(11):1567~1573.

[51] Selim SZ, Ismail MA. On the local optimality of the fuzzy ISODATA clustering algorithm. IEEE Trans. on PAMI, 1996,8(2): 284~288.

[52] Jajuga K. L1 norm based fuzzy clustering. Fuzzy Sets and Systems, 1991,39(1):43~50.

[53] Dave RN, Krishnapuram R. Robust clustering methods: A unified view. IEEE Trans. on Fuzzy Systems, 1997,5(2):270~293.

[54] Yu J, Cheng QS, Huang HK. Analysis of the weighting exponent in the FCM. IEEE Transactions on Systems, Man and Cybernetics-part B: Cybernetics, 2004, 34(1):634~639.

[55] Yu J. General c-means clustering model and its applications, In: Proc. of the 2003 IEEE Computer Society Conference on CVPR, Vol 2. Madison: IEEE, 2003. 122~127.

[56] Zhang JS, Leung YW. Improved possibilistic c-means clustering algorithms. IEEE Trans. on Fuzzy Systems, 2004,12(2):209~217.

附中文参考文献:
[39] 于剑,论模糊C均值算法的模糊指标.计算机学报,2003,26(8):974~981.

[40] 于剑,程乾生.关于FCM算法中的权重指数m的一点注记.电子学报,2003,31(3):478~480.

[41] 高新波,谢维新.模糊聚类理论发展及应用的研究发展.科学通报,1999,44(22):961~969.

[43] 于剑.聚类有效性及其应用[博士学位论文].北京:北京大学数学科学学院,2000.