Journal of Engineering and Applied Sciences 13 (2): 353-360, 2018
ISSN: 1816-949X

Fuzzy C-Means with Improved Chebyshev Distance for Multi-Labelled Data

Aseel Mousa and Yuhanis Yusof
School of Computing, College of Arts and Sciences, Universiti Utara Malaysia (UUM),
06010 Sintok, Kedah, Malaysia

Abstract: Fuzzy C-Means (FCM) is one of the most well-known clustering algorithms, nevertheless its performance has been limited by the utilization of Euclidean as its distance metric. Even though there exist studies that applied FCM with other distance metrics such as Manhattan, Minkowski and Chebyshev, its performance can still be argued particularly on multi-label data. Various applications rely on data points that can be grouped into more than one class and this includes protein function classification and image annotation. This study proposes the employment of FCM that is implement using an improved Chebyshev distance metric. The proposed work eliminates correlation in data points and improve performance of clustering. The results show that the proposed FCM improves the performance of clustering as it produces minimum objective function value and with less iteration count. Such a result indicates that FCM with improved distance metric contributes in producing better clusters.

Key words: Chebyshev distance metric, fuzzy C-means clustering, multi label data, implement, clusters, FCM

INTRODUCTION

Recently, multi-label clustering arises in many applications. In contrast to traditional data classification, where each instance is assigned to only one label for multi-label classification, one instance may be simultaneously relevant to several labels (Nasierding and Kouzani, 2012). However, multi label data clustering has been an active research topic due to the urgent need to recognize datasets images, voice, etc. (Zhang *et al*., 2013). Several researches has been proposed in this field. In term of experimenting multi-label data using classification algorithm (Lou *et al*., 2012) propose a method for novel convex simi-supervised feature selection. The experimental results demonstrate that the proposed algorithm consistently achieve better result when compared with other semi-supervised feature selection algorithms as well as baseline using all features. It does not require graph construction and Eigen-decomposition, therefore, the computational cost is comparably low and the algorithm can be readily applied to large-scale multi label dataset. In Shi *et al*. (2011), the researcher study the problem of multi-label ensemble learning. They aim at improving the generalization ability of multi-label learning systems by constructing a group of multi-label base learners which are both accurate and diverse. They propose a novel solution, called EnML to effectively augment the accuracy as well as the diversity of multi-label base learners. In Fan and Lin (2007), the researcher propose adjusting thresholds in decision functions of the binary method of classification multi label data which significantly improves the performance. The results show that it marginally helps in multi label classification performance. Hence, optimizing each label once is enough in practice. Developed models and algorithms for multi-label classification (supervised learning) as well as for multi-assignment clustering (unsupervised learning). The proposed generative multi-label classifier yields lower classification errors than state-of-the-art algorithms. Algorithms derived from generative models were applied to two real world problems that includes multi label data and they outperformed state-of the-art methods.

In terms of fuzzy clustering property, the Fuzzy C-means (FCM) clustering algorithm which was suggested by Dunn (1973) and improved by Bezdek (1981) is the one most used and discussed. FCM has become the most powerful and well-known method for the analysis of clusters. However, the algorithm suffers from accuracy with multi-label data due to its employed distance metric (Lou *et al*., 2012). A good clustering algorithm should be both powerful and able to tolerate this situation which often occurs in real application systems (Zanaty, 2013). In order to improve the performance of standard FCM, a number of works have been reported in the form of multi-label data. By Lou *et al*. (2012), a distance regulatory factor was proposed to solve the problems of dependences of similarity measurement

Corresponding Author: Aseel Mousa, School of Computing, College of Arts and Sciences, Universiti Utara Malaysia (UUM), 06010 Sintok, Kedah, Malaysia

on data structure, he approved that this a lgorithm has a good tolerance to different densities and various cluster shapes (Zhang and Shen, 2014), improve FCM algorithm based on Particle Swarm Optimization (PSO) and shadowed sets to perform feature clustering to organize the wide variety of data sets automatically and acquire accurate classification. Experiments show that the proposed approach significantly improves the clustering effect (Zhu *et al*., 2009), introducing a novel membership constraint function, a new objective function. A generalized algorithm called GIFP-FCM for more effective clustering is proposed. Experimental results including a noisy image texture segmentation are presented to demonstrate its average advantage over FCM and IFP-FCM in both clustering and robustness capabilities. In this study, a new distance metric has been proposed to improve the performance of the standard FCM algorithm when applied on multi label dataset.

Literature review: Fuzzy C-Mens algorithm (FCM). FCM has the ability to determine and iteratively update the values of membership of a data point in clusters that are defined previously (Chattopadhyay *et al*., 2011). So, any data point can be related to all clusters with its membership value (Kanamori and Amano, 2007). The algorithm attempts to assign membership value to each data point contained in each cluster center (Cai *et al*., 2007). This is done by relying on the mean distance between each data point and group centroid point (Cai *et al*., 2007). Hence, the closer the data point is to the cluster center, the greater is its membership (Bezdek, 1981). The objective function of a clustering algorithm, which is formulated as a minimization problem is defined as in Eq. 1 (Cai *et al*., 2017):

$$J(u,v) = \sum_{i=1}^{n}\sum_{j=1}^{c} \mu_{ij}^{m} \left\| X_i - V_j \right\|^{q} \quad (1)$$

Where:

- m = Defined to any real number that is greater than 1
- μ_{ij} = The degree of membership of x_j in cluster
- j, x_i = The ith of d-dimensional measured data
- v_j = The dimension centroid of the cluster
- ||*|| = The norm expressing the similarity between any measured data and the centroid
- c = The number of cluster center
- n = The number of data points
- $\|x_i-v_j\|$ = The distance between ith data to jth cluster center

When a high membership values relates to a specific data point, the FCM objective function is minimized which is close to the centroid for their articular class while when they are far from the centroid, low membership values are reached. The FCM clustering algorithm gives a fuzzy membership value to each data point depending on its closeness to the cluster centroids in the feature space (Eq. 2) (Niu and Huang, 2011):

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(d_{ij}/d_{ik}\right)^{\left(\frac{2}{m-1}\right)}} \quad (2)$$

FCM algorithm also updates v_{ij} membership matrix and center of clustering as in Eq. 3 (Chattopadhyay *et al*., 2011). The standard FCM algorithm is formulated to minimize the objective function with respect to the membership values of the ith data jth cluster center:

$$V_j = \frac{\left(\sum_{i=1}^{n} (\mu_{ij})^m x_i\right)}{\left(\sum_{i=1}^{n} (\mu_{ij})^m\right)}, \forall_j = 1,2,3,....,c \quad (3)$$

where, D_{ij} represents the Euclidian distance between ith data to jth cluster center. The Euclidean Distance (ED) employed in FCM has the weakness of correlation is difficult to establish when there is a high noise-to-signal ratio (Cai *et al*., 2007). In addition, in FCM, random selection is performed for determining the cluster centers. This means a specific cluster may be out of the data set, as the cluster center value may be far from data points of datasets this will cause that the distance value is high leads to empty cluster (Chattopadhyay *et al*., 2011). Also, ED is sensitive to outliers, so high skew will throw off the mean and alter covariance (Jun and Tong, 2010). The nonlinear relationships of data will not be accurately measured, resulting in a low correlation. In a worst case scenario it could end up with correlation zero when it actually had a relationship of +1 and -1 in separate portions of the same dataset (Tsai and Lin, 2011). Upon the completion of clustering using FCM, each data point will be accompanied with a membership value for each class. Each data point is also supposed to be independent of the other data point and spatial interaction between data points is not considered. However, for medical data, there is strong correlation between neighboring pixels (Kannan *et al*., 2010).

The Multi Label Data (MLD) has more labels than features. Some MLDs have only a few labels per instance, while others have much of them. Most MLDs are imbalanced which means that some labels are very frequent while others are rarely represented (Zhang *et al*., 2013). The labels in an MLD can be correlated or not. Moreover, frequent labels and rare labels can appear together in the same instances. The most basic information that can be obtained from an MLD is the number of instances, attributes and labels (Song *et al*., 2013). For any MLD containing |D| instances, any instance D_i, i (1, ... , |D|) (Lou *et al*., 2012) will be the union of a set of attributes and a set of labels (X_i, Y_i), X_i∈X^1 x X^2 x, ..., xX^f, Y_i∈L, where f is the number of input features and X^j is the space of possible values for the jth attribute, j {1, f}. L being the full set of labels used in D, could be any subset of items in L (Nasierding and Kouzani, 2012). On the other hand, multi label classification assigns to each sample a set of target labels. This can be thought as predicting properties of a data-point that are not mutually exclusive, such as topics that are relevant for a document. A text might be about any of religion, politics, finance or education at the same time or none of these. Each instance x_i is associated with a subset of labels Y_i⊆L. Existing methods for multi-label classification can be grouped into two main categories) problem transformation methods and algorithm adaptation methods. Multi-label classification problems can be found in various domains (Cabral *et al*., 2011) including classifications of text document (Cabral *et al*., 2011; Duygulu *et al*., 2002), bioinformatics data Yeast (Cabral *et al*., 2011; Fan and Lin, 2007) and Genbase (Klimt and Yang, 2004), emotions related musical data (Han *et al*., 2011), scene images (Streich, 2010), textual email messages Enron (Klimt and Yang, 2004), image and video annotation (Das, 2013). The multi-label learning is a form of supervised learning where the classification algorithm is required to learn from a set of instances, each instance can belong to multiple classes. Multi Label Datasets (MLDs) are generated from text documents (Klimt and Yang, 2004), sets of images (Duygulu *et al*., 2002), music collections and protein attributes (Diplaris *et al*., 2005), among other sources. For each sample a set of features (input attributes) is collected and a set of labels (output label set) is assigned. In order to determine if a data set is of type multi label or vice versa, a classification process need to be applied on it to find the multi label data points. There are several classification methods, for example, Support Vector Machine (SVM), K-nearest neighbor, Decision trees, neural network, quadratic classifiers, bayesian classification, etc. In this study, the SVM classification method will be used where the aim is to determine the location of decision boundaries also known as hyperplane that produce the optimal separation of classes (Han *et al*., 2011). SVM generates high accuracy and work well even if data is not linearly separable (Cortes and Vapnik, 1995; Han *et al*., 2011). SVM is originally a binary classification method developed by Vapnik and colleagues at Bell laboratories (Das *et al*., 2014). For a binary problem, we have training data points:

$$\{x_i, y_i\}, i = 1, ..., 1, y_i, x_i \in R^d$$

Suppose, we have some hyperplane which separates the positive from the negative examples (a "separating hyperplane"). The points x which lie on the hyperplane satisfy w.x+b = 0 where w is normal to the hyperplane, b ||w|| is the perpendicular distance from the hyperplane to the origin and ||w|| is the Euclidean norm of w. Let, d+ (d-) be the shortest distance from the separating hyperplane to the closest positive (negative) example. Define the "margin" of a separating hyperplane to be d++d-. For the linearly separable case, the support vector algorithm simply looks for the separating hyperplane with largest margin (Kumar, 2011). This can be formulated as follows: suppose that all the training data satisfy the following constraints:

$$x_i, w+b \geq +1 \text{ for } y_1 = +1 \quad (4)$$

$$x_i, w+b \geq -1 \text{ for } y_1 = -1 \quad (5)$$

These can be combined into one set of inequalities:

$$Y(x_i.w+b) - 1 \geq 0v_1 \quad (6)$$

Chebyshev distance metric: The Chebyshev distance is also known as the maximum metric (Streich, 2010). Another term is the chessboard distance as it can be illustrated on the real number plane as the number of moves needed by a chess king to travel from one point to another (Das, 2013). The Chebyshev distance is the L8-norm of the difference, a special case of the Minkowski distance in which p goes to infinity (Tsai and Lin, 2011). The distance is measured by the following equation:

$$D_{xy} = MAX\left|X_{ik} - X_{jk}\right| \quad (7)$$

The infinity value of the distance causes that the data point goes far from the cluster center it belongs to, hence, data point will not relate to its real cluster which then affects the clustering result.

Many studies have employed Chebyshev distance metric with clustering algorithms (Cha, 2007) made a comprehensive survey on distance metrics between probability density functions, he use four distance metrics to solve some pattern recognition problems such as classification, clustering and retrieval problems. Comparison is made in both syntactic and semantic relationships in order to find which distance is suitable to specific application. Categorization conclude that Chebyshev distance is more suitable for medical application (Gomathi and Karthikeyan, 2014) on the other hand made a comparative evaluation among some distance metrics for clustering data points for organ segmentation. The comparison relays on the task, number of data and complexity of the task. Results concluded that no universal distance measure which can be best suited for all clustering applications (Kouser and Sunita, 2013) use K-means clustering algorithm to find the effect of distance functions on clustering. They apply various distance metrics such as Euclidean, Manhattan, Chebyshev. The functions used for analyzing the result is the number of iterations, overall accuracy, mean absolute error.

MATERIALS AND METHODS

Three datasets are used in this study including the STARKEY'93 that contains 88 trajectories (corresponding to 33 elk, 14 deer and 41 cattle). The whole dataset has 79,987 (x, y, t) observations, an average of 909 observations per trajectory. Figure 1 illustrates sample of the dataset. The second dataset is the Genbase dataset which is a biology dataset. It has 662 objects and 27 labels with density of 0.046. An example for the dataset is as in Fig. 1. Finally, the Yeast dataset, also a Biology dataset, has 2417 objects and 14 labels with 0.303 density and 4.237 cardinality. The dataset is presented in Fig. 1. All of the employed datasets are available at http://mulan.sourceforge.net/datasets.html. This study focuses on improving clustering performance on multi label data. Hence, the proposed study includes the following steps:

- Step 1: Apply SVM classifier on dataset
- Step 2: Apply improved FCM algorithm using output of Step 1
- Step 3: Initialize $\mu = (\mu_{ij})^0$ matrix of membership
- Step 4: Choose parameter >0 to stop the iteration. Set the iteration counting parameter equal to 0
- Step 5: Repeat k = 1, 2
- Step 6: At k-step calculate the centre vectors $v^k = [v_j]$ k is iteration step:

$$v_j\left(\sum_{i=1}^{n}\frac{(\mu_{ij})^m x_i}{\left(\sum_{i=1}^{n}(\mu_{ij})^m, \forall_j = 1, 2, ..., c\right]}\right)$$

- Step 7: Update the membership $(\mu_{ij})^k$, $(\mu_{ij})^{k+1}$ matrix by:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d}{d_{ik}}\right)^{\left(\frac{2}{m-1}\right)}}$$

where (d_{ij}) is the distance metric. The distance metrics applied are the Euclidean, Manhattan, Minkowski, original Chebyshev and improved Chebyshev distance, respectively.

(a) (b) (c)

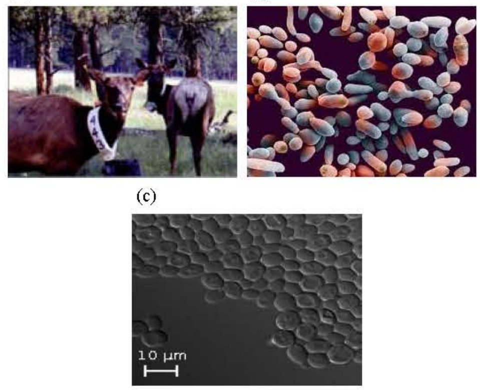

Fig. 1: Sample images of: a) STARKEY'93; b) Genbase and c) Yeast datasets

- Step 8: If $(\mu_{ij})^{k+1}-(\mu_{ij})^k>\varepsilon$ go to step 6 else
- Step 9: Stop
- Step 10: End if

The first step is to apply SVM classification in order to determine there exist multilabel points. The multi label is proved by the SVM concept of decision planes that separates between a set of objects having different class memberships (Zhang *et al.*, 2013). Figure 2 illustrates an example of applying SVM classifier on STARKEY'93 dataset. Figure 2a represents two dimension plot of the dataset, represents the decision boundary that separates data points into classes while indicates that there are misclassified data points (blue points that appeared in red point's area and vice versa). These misclassified data points are the multi label data which will be reprocessed in the clustering step.

The second step is performed by applying the improved FCM clustering algorithm. The improvement includes replacing FCM existing distance metric (i.e., Chebyshev) with an improved Chebyshev in order to produce a better clustering. The formulation of Chebyshev distance is as follows:

$$D_{xy} = MAX\left|x_{ik} - x_{ik}\right| \quad (8)$$

Nevertheless, in order to avoid the infinity that may occurred while using Chebyshev distance, this study proposes the following improvement:

$$D_{xy} = \lim_{p\to\infty}\left(max_k\left(\left|x_{ik} - x_{jk}\right)^{1/p}\right|\right)\infty<p>2 \quad (9)$$

The limit is used in two basic problems of calculus: the area problem and the tangent line problem (Horvath and Khoshnevisan, 2010). The expression

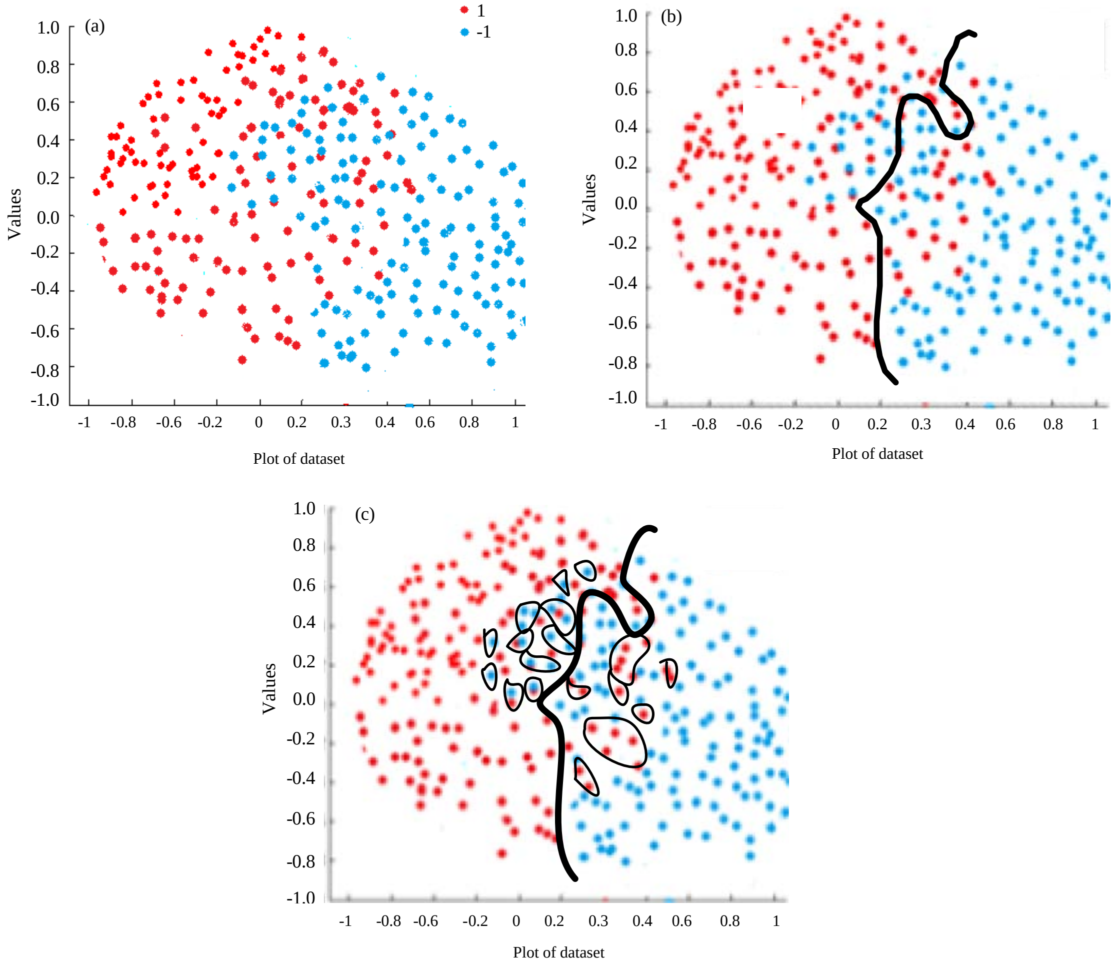

Fig. 2: SVM classification result on STARKEY'93: a) Plot of dataset; b) SVM decision boundary and c) Misclassification areas

Table 2: The value of Iteration Count (IC) that leads to the minimum Objective Function (OF)

	Euclidean		Manhattan		Minkowski		Chebyshev		Improved chebyshev	
Dataset	IC	OF	IC	OF	IC	OF	IC	OF	IC	OF
STARKEY 93	30	3.297862	40	1.882093	45	1.743560	30	5.8036730	10	0.453989
Genbase	15	0.467862	18	0.482093	20	1.853560	14	5.5036730	8	0.253989
Yeast	48	0.997862	43	0.382093	50	5.853560	40	1.1503673	20	0.018805

$\lim_{p\to\infty} f(p) = L$ indicates that if the value of p is close but not equal to 8, then f(p) will be close to the value L; moreover, f(p) gets closer and closer to L as p gets closer and c loser to 8 (Guichard, 2014). For means $\lim_{p\to\infty} f(p) = L^{1/p}$ that 1/p approach 0 as p approaches infinity (Heinbockel, 2007). The goal of improving the distance metric is to minimize the distance between any point and cluster center of the class. This requires that the value of the distance metric to be close to zero.

RESULTS AND DISCUSSION

In this study, the results of the experiments are presented. After applying the SVM classification on the three datasets, the improved FCM clustering algorithm is applied on the SVM output. Table 1 illustrates the iteration count and minimum objective function. From the table, it is noted that the FCM with improved Chebyshev distance requires about 50% less iteration than the other distances. Furthermore, it generates 30% less for objective

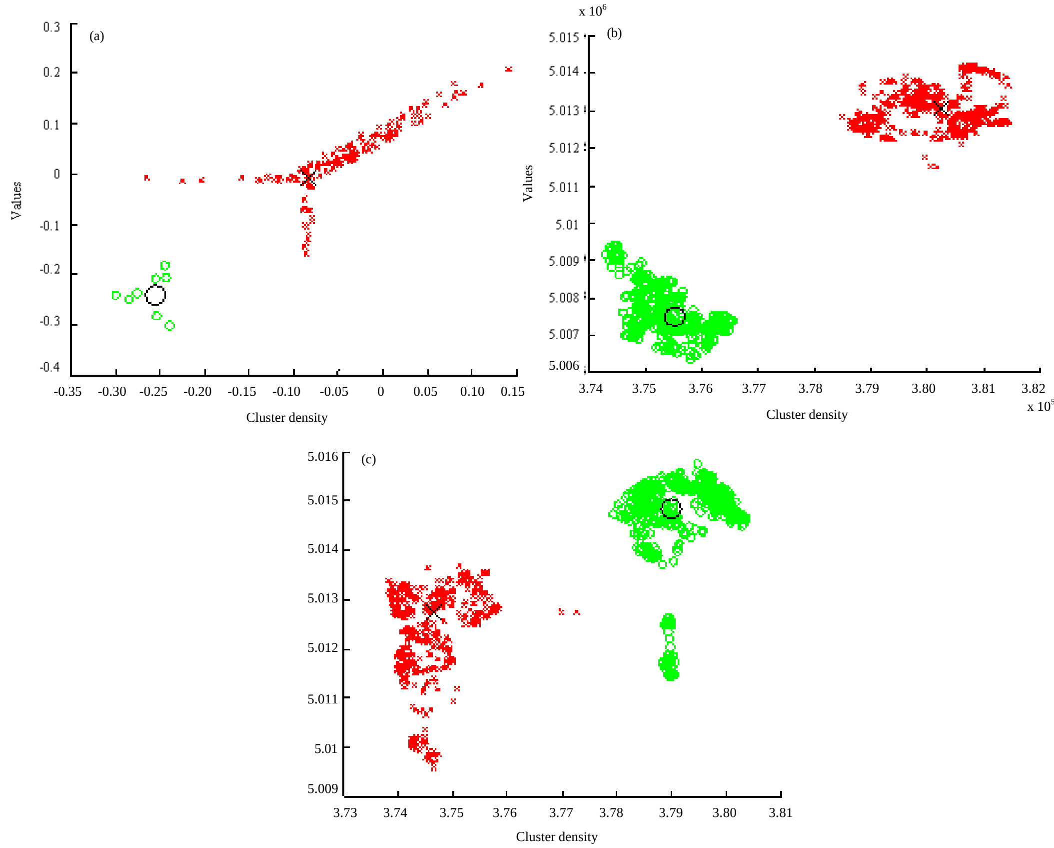

Fig. 3: Cluster density for: a) STARKEY'93; b) Genbase and c) Yeast datasets

function than achieved by applying FCM with the other four distances. On the other hand, Fig. 3 includes the graphical representation of cluster density for all the three datasets.

CONCLUSION

Multi-label classification approaches focus on exploiting the label correlations to improve the accuracy of the classifier by building an individual multi-label classifier or a combined based upon a group of single label classifier. In this study, a combination of classification and improved clustering approaches are used on multi label datasets. Although, FCM is one of the most well-known clustering algorithms, its performance has been limited by the utilization of Euclidean as its distance metric. This study proposes the replacement of Euclidean distance with an improved Chebyshev distance. Based on the results, the proposed method improves the performance of clustering by minimizing the objective function. The improved Chebyshev distance gave the minimum distance with fewer iterations while the other four distance metrics used in this research need higher number of iteration to reach the minimum objective function. Results also provide the insight that performance in clustering relies on the cluster density.

REFERENCES

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. 1st Edn., Kluwer Academic Publishers, Norwell, MA, USA.

Cabral, R.S., F. Torre, J.P. Costeira and A. Bernardino, 2011. Matrix completion for multi-label image classification. Proceedings of the 24th International Conference on Neural Information Processing Systems NIPS, December 12-14, 2011, Curran Associates, Inc, Granada, Spain, ISBN:978-1-61839-599-3, pp: 190-198.

Cai, W., S. Chen and D. Zhang, 2007. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. J. Pattern Recognit., 40: 825-838.

Cha, S.H., 2007. Comprehensive survey on distance/similarity measures between probability density functions. Int. J. Math. Models Meth. Applied Sci., 1: 300-307.

Chattopadhyay, S., D.K. Pratihar and D.S.C. Sarkar, 2011. A comparative study of fuzzy c-means algorithm and entropy-based fuzzy clustering algorithms. Computi. Inf., 30: 701-720.

Cortes, C. and V. Vapnik, 1995. Support-vector networks. Mach. Learn., 20: 273-297.

Das, A., U. Pal, M.A.F. Ballester and M. Blumenstein, 2014. A new efficient and adaptive sclera recognition system. Proceedings of the IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM), December 9-12, 2014, IEEE, Orlando, Florida, ISBN:978-1-4799-4532-0, pp: 1-8.

Das, S., 2013. Pattern recognition using the fuzzy C-means technique. Intl. J. Energy Inf. Commun., 4: 1-14.

Diplaris, S., G. Tsoumakas, P.A. Mitkas and I. Vlahavas, 2005. Protein classification with multiple algorithms. Proceedings of the Conference on Panhellenic Informatics, November 11-13, 2005, Springer, Volas, Greece, ISBN:978-3-540-29673-7, pp: 448-456.

Dunn, J.C., 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybernet., 3: 32-57.

Duygulu, P., K. Barnard, J.F.D. Freitas and D.A. Forsyth, 2002. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. Proceedings of the European Conference on Computer Vision, May 28-31, 2002, Springer, Copenhagen, Denmark, ISBN:978-3-540-43748-2, pp: 97-112.

Fan, R.E. and C.J. Lin, 2007. A study on threshold selection for multi-label classification. Master Thesis, National Taiwan University, Taipei, Taiwan.

Gomathi, V.V. and S. Karthikeyan, 2014. Performance analysis of distance measures for computer tomography image segmentation. Int. J. Comput. Technol. Applic., 5: 400-405.

Guichard., D., 2014. Calculus Early Transcendentals. Createspace Publishing, North Charleston, South Carolina, ISBN:9781501075582, Pages: 514.

Han, J., M. Kamber and J. Pei, 2011. Data Mining: Concepts and Techniques. 3rd Edn., Morgan Kaufmann Publishers, USA., ISBN-13: 9780123814791, Pages: 744.

Heinbockel, J.H., 2007. Introduction to the Variational Calculus. Vol. 1, Trafford Publishing, Bloomington, Indiana, ISBN:9781425103521, Pages: 350.

Horvath, L. and D. Khoshnevisan, 2010. Weight functions and path wise local central limit theorems. Stochastic Process. Appl., 59: 105-123.

Jun, W. and W.S. Tong, 2010. Double indices FCM algorithm based on hybrid distance metric learning. J. Software, 21: 1878-1888.

Kanamori, T. and H. Amano, 2007. Implementation and evaluation of a high speed license plate recognition system on an FPGA. Proceedings of the 7th IEEE International Conference on Computer and Information Technology CIT, October 16-19, 2007, IEEE, Aizu-Wakamatsu, Fukushima, Japan, ISBN:0-7695-2983-6, pp: 567-572.

Kannan, S.R., S. Ramathilagam and R. Pandiyarajan, 2010. Modified Bias Field Fuzzy C-Means for Effective Segmentation of Brain MRI. In: Transactions on Computational Science VIII, Gavrilova, M.L. and C.J.K. Tan (Eds.). Springer, Berlin, Germany, ISBN:978-3-642-16235-0, pp: 127-145.

Klimt, B. and Y. Yang, 2004. The enron corpus: A new dataset for email classification research. Proceedings of the European Conference on Machine Learning ECML, September 20-24, 2004, Springer, Pisa, Italy, pp: 217-226.

Kouser, K. and A. Sunita, 2013. A comparative study of K means algorithm by different distance measures. Intl. J. Innov. Res. Comput. Commun. Eng., 1: 2443-2447.

Kumar, C.N.R., 2011. Pattern classification. Master Thesis, JSS Science and Technology University, Mysore, India.

Lou, X., J. Li and H. Liu, 2012. Improved fuzzy C-means clustering algorithm based on cluster density. J. Comput. Inf. Syst., 8: 727-737.

Nasierding, G. and A.Z. Kouzani, 2012. Comparative evaluation of multi-label classification methods. Proceedings of the 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), May 29-31, 2012, IEEE, Sichuan, China, ISBN:978-1-4673-0025-4, pp: 679-683.

Niu, Q. and X. Huang, 2011. An improved fuzzy C-means clustering algorithm based on PSO. J. Software, 6: 873-879.

Rowland, M.M., P.K. Coe, R.J. Stussy, A.A. Ager and N.J. Cimon et al., 1998. The Starkey habitat database for ungulate research: Construction, documentation and use. Master Thesis, U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Oregon.

Shi, C., X. Kong, S.Y. Philip and B. Wang, 2011. Multi-label ensemble learning. Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, September 5-9, 2011, Springer, Athens, Greece, ISBN:978-3-642-23807-9, pp: 223-239.

Song, J., Z. Zhu, P. Scully and C. Price, 2013. Selecting Features for Anomaly Intrusion Detection: A Novel Method using Fuzzy C Means and Decision Tree Classification. In: Cyberspace Safety and Security, Guojun, W. I. Ray, D. Feng and M. Rajarajan (Eds.). Springer, Berlin, Germany, ISBN:978-3-319-03583-3, pp: 299-307.

Streich, A.P., 2010. Multi-label Classification and Clustering for Acoustics and Computer Security. ETH-Bibliothek Publisher, Zurich, Switzerland, Pages: 216.

Tsai, D.M. and C.C. Lin, 2011. Fuzzy C-means based clustering for linearly and nonlinearly separable data. Pattern Recogn., 44: 1750-1760.

Zanaty, E.A., 2013. An adaptive fuzzy C-means algorithm for improving MRI segmentation. Open J. Med. Imaging, 3: 125-135.

Zhang, J. and L. Shen, 2014. An improved fuzzy C-means clustering algorithm based on shadowed sets and PSO. Comput. Intell. Neurosci., 2014: 1-22.

Zhang, Z., M. Zhao and T.W. Chow, 2013. Binary-and multi-class group sparse canonical correlation analysis for feature extraction and classification. IEEE. Trans. Knowl. Data Eng., 25: 2192-2205.

Zhu, L., F.L. Chung and S. Wang, 2009. Generalized fuzzy C-means clustering algorithm with improved fuzzy partitions. IEEE. Trans. Syst. Man Cybern. Part B. Cybern., 39: 578-591.