Journal of Theoretical and Applied Information Technology
10th October 2015. Vol.80. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FUZZY KERNEL K-MEDOIDS ALGORITHM FOR MULTICLASS MULTIDIMENSIONAL DATA CLASSIFICATION

[1]ZUHERMAN RUSTAM, [2]AINI SURI TALITA

[1] Senior Lecturer, Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Indonesia, INDONESIA

[2] Associate Lecturer, Faculty of Computer Science and Information Technology, Gunadarma University, INDONESIA

E-mail: [1]rustam@ui.ac.id, [2]ainisuri@staff.gunadarma.ac.id

ABSTRACT

The success of the classification method is highly dependent on how to specify initial data as the initial prototype, dissimilarity functions that we used and the presence of outliers among the data. To overcome these obstacles, in this paper we present Fuzzy Kernel k-Medoids (FKkM) algorithm that we claim to be robust against outliers, invariant under translation and data transformation, as the combined development of Fuzzy LVQ, Fuzzy k-Medoids and Kernel Function. Based on the experiments, it provides a better accuracy than Support Vector Machines, Kernel Fisher Discriminant and RBF Neural Network for multiclass multidimensional data classification.

Keywords: *Classification*, *Fuzzy LVQ*, *Fuzzy K-Medoids*, *Kernel Function*, *Multiclass Multidimensional Data*

1. INTRODUCTION

Classification methods are used to classify data set $X = \{x_1, x_2, \ldots, x_n\}$, $x_i \in \mathbb{R}^d, i = 1, 2, \ldots, n$ into several classes (clusters). Data will be placed at the same class C_j if they have the same characteristics. Each class is represented by a vector $v_j \in \mathbb{R}^d$. One of the classification technique is Vector Quantization (VQ). It maps the data $X = \{x_1, x_2, \ldots, x_n\}$ into $V = \{v_1, v_2, \ldots, v_c\} \subset \mathbb{R}^d$. V is called a prototype set, medoid, signature, or codebook. After the prototype was set, we build optimization model involving two sets of unknown, i.e. Membership Functions set and Prototype set. Alternating Minimization Algorithm (AMA) [1] is used to solve the model by iteratively updating membership function and prototype sets until the optimum solution is obtained.

Several clustering techniques that based on VQ and AMA are Fuzzy C-Mean (FCM) [1], Possibilistic C-Mean (PCM) [2], and Generalized Fuzzy C-Mean [3]. Fuzzy Clustering with Proportional Membership (FCPM) is a clustering technique that based on VQ [4]. On FCPM, membership function is updated based on the structure of cluster that was generated at every iteration. It is different with the prior three clustering techniques that ignore the structure of cluster that was generated at every iteration. Fuzzy k-Medoids (FkM) was generated as another development of VQ-based clustering method [5]. While Chu et al. [6] combine FkM with Genetic algorithm.

Researchers keep finding new clustering technique to increase its accuracy at solving clustering problems. The accuracy is dependent on the data that were used. VQ-based clustering technique is a sequential clustering technique where its accuracy depends on the initial data and is also susceptible to outliers.

K-Medoids - based classification method is an outlier (noise) robust clustering technique [6]. Moreover, it is also invariant to the translation and transformation of data. K-Medoids classification optimization model use dissimilarity function to compare similarity (closeness) between prototype and input data. In general dissimilarity function that was used at VQ-based classification technique is Euclidean norm. However, Hathaway et al. [7] use

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

a non-Euclidean norm, Minkowski norm, on Fuzzy C-Means with higher classification accuracy than using Euclidean norm.

The utilization of kernel methods by Vapnik [8] on multidimensional classification problems and its development by Scholkopf et al. [9] on Principal Component Analysis provide a very significant contribution to the classification technique. By applying kernel technique on Support Vector Machine method (SVM) [10, 11] results a higher accuracy compare to non-kernel classification methods.

With robustness of Fuzzy k-Medoids on outliers existence, its invariant property under translation and data transformation and also kernel function capability while dealing with multidimensional data, in this paper we develop a combination of Fuzzy k-Medoids method and kernel method, which is called Fuzzy Kernel k-Medoids (FKkM).

At the second part of this paper, we will discuss about Fuzzy k-Medoids algorithm, while Kernel function discussed at the third part. Fuzzy Kernel k-Medoids which is our main result will be discussed at Section 4, and the last section contains conclusion of this paper.

2. FUZZY K-MEDOIDS

Some of classification methods minimize objective function by using least square criteria that is susceptible to outliers. Existence of outliers will affect its results. Fuzzy k-Method would overcome that problem.

For a data set $X = \{x_1, x_2, \dots, x_n\}$, $x_i \in \mathbb{R}^d, i = 1, 2, \dots, n$, set $U = [u_{ij}]$, $n \times c$ matrix, $1 \le i \le n$, $1 \le j \le c$ and $V = \{v_1, v_2, \dots, v_c\} \subset \mathbb{R}^d$ a medoid set. To overcome outlier problem, Krishnapuram [5] use objective function:

$$J(U,V) = \sum_i^n \sum_j^c u_{ij} u_{ij}{}^m r(x_i, v_j) \quad (1)$$

To update membership, FkM use:

$$u_{ij} = \frac{r^b(x_i, v_j)}{\sum_{k=1}^c r^b(x_i, v_k)}, 1 \le i \le n, 1 \le j \le c \quad (2)$$

where $r(x_i, v_j)$ is a dissimilarity function between x_i and all prototype $v_j \in V$, m is the fuzziness degree, and $b = -\frac{1}{m-1}$.

After the membership is generated, medoid $v_j \in V$ is updated by using formula:

$$v_j = x_p \quad (3)$$

where $1 \le j \le c$, $p = \arg\min_{1 \le q \le n} u_{ij}{}^m r(x_q, x_i)$.

Fuzzy k-Medoids Algorithm can be expressed as:

1) Initialize $V = [v_1, v_2, \dots, v_c]$, where every $v_j, 1 \le j \le c$, comes from different cluster.
2) Update membership u_{ij} by using Eq. (2)
3) Update medoid $v_j \in V$ by using Eq. (3)
4) The updating process is finished when medoid remain unchanged or the number iteration already reached stopping criteria.

3. KERNEL FUNCTION

As it mentioned before, Kernel technique that was found by Vapnik [8] and later expanded by Scholkopf [9] and Christianini [10] can be used to improve performances of classification methods. This section is dedicated to explain about kernel function.

Set a non-linear mapping ϕ from input space $\mathbb{R}^d$ to feature space F where dimension of F is far greater then dimension of input space, $\phi: \mathbb{R}^d \to F$ or $x \mapsto \phi(x)$. After sample data are mapped into feature space, classification process take place at feature space F other than input space $\mathbb{R}^d$. One of the problem with performing classification at feature space is how to measure distance between $\phi(x)$ and $\phi(y)$, $x, y \in \mathbb{R}^d$. This problem arise because it is not known how to define the appropriate function ϕ. Other problem arise when dimension of $\phi(x)$ is greater than x, which means that the computation cost such as memory usage and time computation will increase. To overcome these problems, define a function K, where

Journal of Theoretical and Applied Information Technology
10th October 2015. Vol.80. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$K(x_i, x_j) = \phi(x_i)\phi^t(x_i)$. This function is called Kernel Function.

Distance function can be defined as:

$$d^2(\phi(x), \phi(y)) = \|\phi(x) - \phi(y)\|^2$$
$$= \phi(x)\phi^t(x) - 2\phi^t(x)\phi(y) + \phi^t(y)\phi(y)$$
$$= K(x,x) - 2K(x,y) + K(y,y) \qquad (4)$$

For simplicity, notation $K_{ij} = K(x_i, v_j) = d(x_i, v_j)$ is used to express distance between x_i and v_j. This distance function will be used to replace dissimilarity function on Fuzzy k-Medoids algorithm.

Kernel functions that will be used for experiments at the next section are:

1) RBF kernel function:
$$K(x,y) = \exp\left(-\frac{\|x-y\|^2}{\sigma^2}\right)$$
2) dth-degree polynomial kernel function:
$$K(x,y) = (x^t y + 1)^d$$
3) Linear kernel function:
$$K(x,y) = x^t y$$

4. RESULTS AND DISCUSSION

4.1. Fuzzy Kernel K-Medoids Algorithm

Karayiannis [12] use different fuzziness degree m for each iterations at Fuzzy Learning VQ algorithm. Fuzziness degree at each iterations is $m = m_i + \frac{t}{T}(m_f - m_i)$, where m_f and m_i are respectively initial value and end value of m. T is maximum number of iterations ant t is the counter of iteration.

If we set value of m_i quite large and m_i quite small (or vice versa), it is expected that the value of m will decrease (increase) on every iterations. Based on experiments, these different values of m are necessary due to various input data.

By combining Fuzzy k-Medoids [5], Kernel Function [8], and fuzziness degree [12], we build Fuzzy Kernel k-Medoids (FKkM) Algorithm as follows:

Input: $X, c, m_i, m_f, \varepsilon, T$
Output: U and V
1. Initialization $V^0 = [v_1, v_2, \ldots, v_c], v_j \in C_j$
2. For $t = 1$ to T
   3. $m = m_i + \frac{t(m_f - m_i)}{T}$
   4. Calculate membership:
   $$U^1 = [u_{ij}], 1 \le i \le n, 1 \le j \le c$$
   $$u_{ij} = \frac{K^b(x_i, v_j)}{\sum_{k=1}^{c} K^b(x_i, v_k)}$$
   5. Calculate medoid:
   $$v_j = x_p, i \le j \le c$$
   Where
   $$p = \arg\min_{1 \le q \le n} {u_{ij}}^m K(x_q, x_i).$$
   6. If $E = \sum_{j=1}^{c} K^2(v_{jt}, v_{jt-1}) \le \varepsilon$ then stop.
   Else
   $t = t + 1$

Algorithm 1: Fuzzy Kernel k-Medoids Algorithm

Two sequences $\{U^t\}$ and $\{V^t\}$ are convergent to minimum value of $J(U, V)$ [1].

4.2. Experiment Result

For this experiment we use sample data from [13], which are:

1. Breast Cancer Wisconsin Data

   It contains of 699 data, in which each data consist of 10 attributes (10 dimensional data). 65.5% of these data are cancer Benign case while the rest are cancer Malign. At this experiment, we use 30% data from each class as training data and the rest as testing data.

2. Diabetes Data

   It consists of 768 of 8 dimensional data that were classified into two classes.

Journal of Theoretical and Applied Information Technology
10th October 2015. Vol.80. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3. German Data

It consists of 1000 of 24 dimensional data that were classified into two classes.

4. Image Segmentation Data

It consists of 2100 data that were classified into 7 classes. The dimension of each data is 19.

5. Sonar Data

Sonar Data consists of 208 of 60 dimensional data that were classified into two classes.

6. Thyroid Data

Based on 5 laboratory experiments on blood sample of 215 patients, we will determine if a patient has one type of thyroid case: *euthyroidism*, *hypothyroidism*, or *hyperthyroidism*.

7. Iris Data

It consists of 150 data from three classes: *Iris Setosa*, *Iris Versicolour*, and *Iris Virginica*. Each data is a 4 dimensional data with attributes: *sepal length*, *sepal width*, and *petal width*.

8. Wine Data

It consists of 178 data that were classified into three classes. Each data is a 13 dimensional data which represents chemical analysis on every types of wine. The data is classified based on location of the wineries.

9. Soybean Data

It consists of 307 of 35 dimensional data that were classified into 19 classes.

On each experiment, we use p% of data as training data and the rest as testing data to validate the algorithm, where p= 10, 20, … , 90. The experiment was repeated 10 times and at every experiment the data were chosen randomly. Classification accuracy was measured by:

$$CR = \frac{TP+TN}{TP+FN+FP+TN} \quad (5)$$

Where TP = true positive, TN = true negative, FN = false negative, and FP = false positive.

We use personal computer with i-7 processor, 8 GB RAM, 1 TB hard disk, and software Matlab 2012a.

Table 1 contains experiment results by using FKkM and experiment comparison by Muller et al. [11] that using Support Vector Machine (SVM), Kernel Fisher Discriminant (KFD), and Neural Network RBF. The results by using FKkM algorithm use Gaussian kernel with $\sigma = 0.05$.

Table 1: Error percentage for Classification using SVM, KFD, RBF, and FKkM

Data	SVM	KFD	RBF	FKkM
Breast Cancer	26	25.8	27.6	20.2
Diabetes	23.5	23.2	24.3	18.3
German	23.6	23.7	24.7	20.6
Image	3	3.3	3.3	3
Sonar	32.4	33.2	34.4	28.8
Thyroid	4.8	4.2	4.5	2.3
Iris	-	-	-	2.1
Wine	-	-	-	3.4
Soybean	-	-	-	1.2

As we can see at Table 1, error percentages for classification by using FKkM for 6 types of data are smaller than other methods. While Muller et al. [11] does not include Iris, Wine, and Soybean data, we are succeeded to use FKkM to classify these data with error percentages are under 4%.

5. CONCLUSION AND FUTURE WORKS

On this paper we develop a new classification method by combining basic ideas from Fuzzy k-Medoids, Kernel Functions, and Fuzziness degree from Fuzzy LVQ which aims to preserve its advantages and overcome its deficiency. Based on experiments, Fuzzy Kernel k-Medoids gives a better accuracy than SVM, KFD, and RBF methods for sample data.

For FKkM algorithm, we use Gaussian kernel with $\sigma = 0.05$ for the best result. For future work, it is possible to find a non-parametric kernel that gives a better result.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REFERENCES

[1] J. Bezdek, Pattern Recognition With Fuzzy Objective Function Algorithms, New York, Plenum Press, 1981.

[2] R. Krishnapuram and J. M. Keller, "The Possibilistic C-Means Algorithm: Insights and Recommendations", *IEEE Trans. Fuzzy Syst*., Vol. 4, No 3, 1996, pp. 385-393.

[3] N. B. Karayiannis, "Soft Learning Vector Quantization and Clustering Algorithms Based on Ordered Weighted Aggregation Operators", *IEEE Trans. Neural Networks*, Vol. 11, No. 5, 2000, pp. 1093-1105.

[4] S. Nascimento, B. Mirkin, and F. Moura-Pires, "Modeling Proportional Membership in Fuzzy Clustering", *IEEE Trans. Fuzzy Systems*, Vol. 11, No. 2, 2003, pp. 173-186.

[5] R. Krishnapuram, A. Joshi, and L. Yi, "A Fuzzy Relative of the k-Medoids Algorithm with Application to Web Document and Snippet Clustering", *Proceedings of IEEE International Fuzzy Systems Conference*, Seoul (South Korea), Augustus 22-25, 1999, Vol. 3, pp. 1281-1286.

[6] S. C Chu, J. F. Roddick, and J. S. Pan, "An Incremental Multi - Centroid, Multi - Run Sampling Scheme for k – Medoids - based Algorithms – Extended Report", *Technical Report KDM-02-003*, Flinders University of South Australia, 2002.

[7] R. J. Hathaway, J. C. Bezdek, and J. Hu, "Generalized Fuzzy C-Means Clustering Strategies using L_p Norm Distances", *IEEE Transactions on Fuzzy Systems*, Vol. 8, No. 5, 2000, pp. 576-582.

[8] V. N. Vapnik, Statistical Learning Theory, New York, Wiley, 1998.

[9] B. Scholkopf, A. Smola, and K. R. Muller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem", *Neural Computation*, Vol. 10, No. 5, 1998, pp. 1299-1319.

[10] N. Cristianini, J. S. Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, 2000.

[11] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An Introduction to Kernel-Based Learning Algorithms", *IEEE Trans. On Neural Networks*, Vol. 12, No. 2, 2001, pp. 181-201.

[12] N. B. Karayiannis and J. C. Bezdek, "An Integrated Approach to Fuzzy Learning Vector Quantization and Fuzzy C-Means Clustering", *IEEE Trans. Fuzzy Systems*, Vol. 5, No. 4, 1997, pp. 622-628.

[13] UCI Benchmark Repository, http://www.ics.uci.edu/~mlearn/MLSummary.html