Preprints of the 18th IFAC World Congress
Milano (Italy) August 28 - September 2, 2011

Gaussian Based Classification with Application to the Iris Data Set [⋆]

H. Chang [*] **A. Astolfi** [*,**]

[*] *Department of Electrical and Electronic Engineering, Imperial College London, UK*
[**] *Dipartimento di Informatica, Sistemi e Produzione, Università di Roma Tor Vergata, Via del Politecnico 1, 00133 Roma, Italy*
(e-mail: hj.chang05@ic.ac.uk, a.astolfi@ic.ac.uk)

Abstract: The paper describes a novel method to group data points into clusters. Our method is derived from the assumption of Gaussian distribution of the points. We apply the clustering method to the well-known Iris flower data set and we suggest a nonlinear discriminant analysis of the data for a more sophisticated classification. To the best knowledge of the authors this paper shows the best classification result for the Iris data so far.

Keywords: Clustering, Classification, Gaussian Distribution, Discriminant Analysis.

1. INTRODUCTION

Clustering is the partition of a set into subsets so that the elements in each subset share some common treat. For some entities such as convoys of vehicles, crowds of people, and dust clouds, data clustering is an important procedure and it is at the core of pattern recognition and classification. Measurements of such entities are usually represented by points in a well-defined space. The points can represent observations of position, activity or other features, related to the data. In most cases the initial task in understanding such data is to study relations among the points to classify into groups with similar attributes.

A clustering of data is required in a number of disciplines such as marketing research (Punj and Stewart (1983)), gene expression study (D'haeseleer (2005)), and image processing (Shi and Malik (2000)). The literature for the problems of clustering and classification is huge. To solve classification/clustering problems there are a number of algorithms based on various approaches such as deterministic approach (Huang (1998)), probabilistic approach (Lauritzen (1995)), simulated annealing based approach (Selim and Alsultan (1991)), evolutionary process based approach (Krishna and Narasimha Murty (1999)), tabu-search based approach (Pan and Cheng (2007)), hierarchical approach (Karypis et al. (1999)), density based approach (Ester et al. (1996)). In particular one of the authors has recently reported Hamiltonian based clustering algorithm and its applications in Casagrande and Astolfi (2008); Casagrande et al. (2009a); Casagrande and Astolfi (2009); Casagrande et al. (2009b). The core idea of the Hamiltonian approach is to regard the clustering function as a Hamiltonian function and to determine the level lines as the trajectories of the corresponding Hamiltonian system.

In this paper we propose a novel clustering/classification algorithm. Within our framework we firstly use the definition of cluster with notions from graph theory (see Augustson and Minker (1970); Zahn (1971); Wu and Leahy (1993); Aksoy and Haralick (1999); Shi and Malik (2000); Shi et al. (2005) for examples of graph theoretic approaches to clustering problem). Then we develop a classification method with the assumption of Gaussian distribution of the data. The proposed method yields two advantages: we do not need to tune any parameter and the method is applicable to measured data in high dimensional space.

Throughout the paper we discuss the Iris data as an application example of our classification method. Anderson (1935) has proposed this data set which presents the geographic variation of Iris flowers and these data have been used as a benchmark for the linear discriminant analysis in Fisher (1936). The data set has been considered many times in the literature of clustering research such as Cannon et al. (1998); Domany (1999); Dubnov et al. (2002); Paivinen (2005); Vathy-Fogarassy et al. (2006). To the best knowledge of the authors the classification result of the paper is the best for the Iris data so far.

The paper is organised as follows. In Section 2 we give an introduction to the Iris flower data set. In addition we give a definition of cluster based on graph theory and describe the problem of so-called 'touching clusters' studied in Zahn (1971). Section 3 introduces the clustering idea for the Iris data with the assumption of Gaussian distribution and illustrates the classification result. Particularly, in Subsection 3.1, we make use of nonlinear discriminant analysis to deal with the limitation of the previous approach. Finally Section 4 provides a summary and some conclusions.

[⋆] This work has been supported by the Engineering and Physical Sciences Research Council (EPSRC) and the MOD University Defence Research Centre on Signal Processing (UDRC) project "Hamiltonian-based cluster-tracking and dynamical classification".

Table 1. Anderson's Iris flower data set.

Iris Setosa	Iris Versicolor	Iris Virginica
(5.1, 3.5, 1.4, 0.2)	(7.0, 3.2, 4.7, 1.4)	(6.3, 3.3, 6.0, 2.5)
(4.9, 3.0, 1.4, 0.2)	(6.4, 3.2, 4.5, 1.5)	(5.8, 2.7, 5.1, 1.9)
(4.7, 3.2, 1.3, 0.2)	(6.9, 3.1, 4.9, 1.5)	(7.1, 3.0, 5.9, 2.1)
$\vdots$	$\vdots$	$\vdots$

2. THE IRIS DATA SET AND A GRAPH THEORETIC APPROACH

We consider the Anderson's Iris flower data set of Anderson (1935) to show an application of the clustering/classification method of the paper. This data set has points in $\mathbb{R}^4$ describing sepal length, sepal width, petal length, and petal width. In the set there are three species of Iris, i.e., Iris Setosa, Iris Versicolor, and Iris Virginica (see Fig. 1, Fig. 2, and Fig. 3). Each species is represented by 50 points, so the Iris data set has 150 points. Some samples of the data set are shown in Table 1. As a method of the separation of Iris Setosa we suggest a graph theoretic approach (e.g. Zahn (1971)) since this approach is applicable to clustering problems in high dimensional spaces.

Fig. 1. Example: Iris Setosa (Radomil (2005))

Fig. 2. Example: Iris Versicolor (Langlois (2005))

Fig. 3. Example: Iris Virginica (Mayfield (2007))

A graph G is defined as a pair (X, E), where X and E are called vertices and edges, respectively. In this paper X is considered as a set of the elements to be clustered and let $X = \{x_1, x_2, \cdots, x_N\}$ be the set of the data which corresponds to N objects. Thus x_i represents the i-th objects which has n variables, i.e. $x_i \in \mathbb{R}^n$. E is a collection of pairs of X hence an element of E (i.e. an edge) is $e_{i,j} = (x_i, x_j)$ where $x_i, x_j \in X$. We then define a ball $B_i(r_T) = \{x \in \mathbb{R}^n : \|x - x_i\| \leq r_T\}$ for a positive parameter r_T and we assume the metric is the Euclidean distance.

An edge $e_{i,j}$ exists in E if and only if $B_i(r_T)$ and $B_j(r_T)$ overlap. (Note that we can equivalently define an edge $e_{i,j} \in E$ if and only if $x_j \in B_i(2r_T)$ or if and only if $d_{i,j} < 2r_T$ where $d_{i,j}$ $(= d(v_i, v_j))$ is the distance between x_i and x_j.) In this section a cluster is defined as a *maximal connected subgraph* of the graph G.

Proposition 1. Consider a set of points $X = \{x_1, \cdots, x_N\}$ and a positive r_T. We define upper triangular adjacency matrix $A = [A_{c1}, A_{c2}, \cdots, A_{cN}]$ where $A(i,j) = 1$ if $d_{i,j} \leq 2r_T$ and $j > i$, and otherwise $A(i,j) = 0$. Note that A_{cj} and $A(i,j)$ correspond to the j^{th} column and $(i,j)^{th}$ entry of A, respectively. All the points in X describe one cluster if there are one or more "1" elements in each of the columns of $A_{c2}, A_{c3}, \cdots, A_{cN}$.

Proof. For the case $N = 2$, the two points in X are in one cluster if $A(1,2) = 1$ which implies that there is 1 element in the column of A_{c2}. Assume that the claim holds with $N = k$, where $k \geq 2$. This implies that the points $x_1, x_2, \cdots, x_k$ are in one cluster. Now consider $N = k+1$ with a new data point x_{k+1} and its corresponding column vector $A_{c(k+1)}$. The sufficient condition that this point x_{k+1} is included in the cluster of $x_1, x_2, \cdots, x_k$ is that the column vector $A_{c(k+1)}$ has one or more 1 elements, which means that this new data x_{k+1} has at least one edge to the cluster of $x_1, x_2, \cdots, x_k$. Thus the proof is completed by mathematical induction.

Corollary 2. Consider a set of points X, a positive r_T, and an upper triangular adjacency matrix $A = [A_{1r}, A_{2r}, \cdots, A_{Nr}]^T$ where $A(i,j) = 1$ if $d_{i,j} \leq 2r_T$ and $j > i$, and otherwise $A(i,j) = 0$. Note that A_{ir} corresponds to the i^{th} row of A. All the points in X describe one cluster if there are one or more 1 elements in each of the rows of $A_{1r}, A_{2r}, \cdots, A_{(N-1)r}$.

Versions of Proposition 1 and Corollary 2 can be given using lower triangular adjacency matrices instead of upper triangular adjacency matrices. Proposition 1 and Corollary 2 can be used to identify clusters in block diagonal upper triangular adjacency matrix.

Compared to Zahn (1971) we do not use minimal spanning trees (MST) to identify clusters. Instead we define a cluster as maximal connected subgraph. To show this graph theoretic approach intuitively, we consider the simple case of Fig. 4, a flock of geese flying in a 'V' shape formation. For an appropriate value of r_T the geese are clustered into 3 groups in Fig. 5. By inspection the clusters are recognised as overlapping circles.

Remark This graph-theoretic method is applicable to dynamic classification when the time evolution is considered as an axis in a multi-dimensional space. For example we are currently investigating dynamic clustering with the placement data of moving pedestrians measured by two Light Detection and Ranging (LIDAR) sensors.

Fig. 4. A flock of geese flying with a formation of ‘V’.

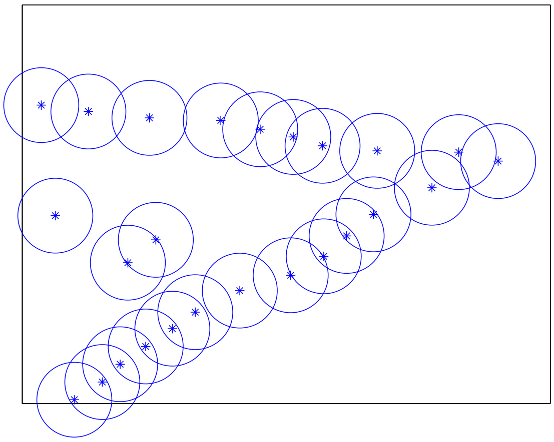

Fig. 5. A clustering identifying the formation of a flock of geese. The geese are clustered into 3 groups for an appropriate size of the circles.

For the case of the Iris data set, we see that $n = 4$ and $N = 150$ for the graph vertices X. Although the graph theoretic approach can be applied to the cases in multi-dimensional space we show a projection of the data onto the 2 dimensional space for simplicity of presentation. The projection of the Iris data point into the plane of petal length and petal width is shown in Fig. 6. In the figure Iris Setosa, Iris Versicolor, and Iris Virginica are plotted in dark grey, light grey, and black, respectively.

This projection provides enough information to cluster Iris Setosa out of the Iris data set using graph theoretic approach (Fig. 7). Setosa is clearly a distinguished cluster while Versicolor and Virginica are overlapping (Zahn (1971)). A certain diagnosis of the two species, Iris Versicolor and Iris Virginica cannot be accomplished using only 4 measurements (Fisher (1936)).

3. GAUSSIAN BASED CLASSIFICATION

The Anderson’s Iris flower data set cannot be classified completely by the approach of the previous section. To address this classification problem we assume that we have the information of the data points of a cluster *a priori* except the data point which we want to classify. Thus we

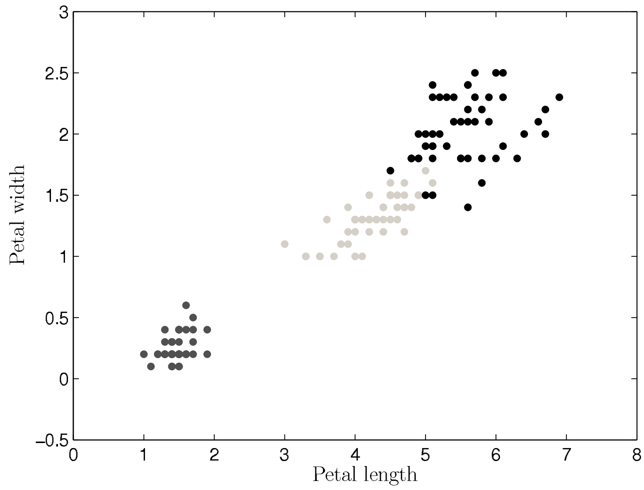

Fig. 6. Projection of the Iris data point into the plane of petal length and petal width. Iris Setosa, Iris Versicolor, and Iris Virginica are plotting with the points of dark grey, light grey, and black, respectively.

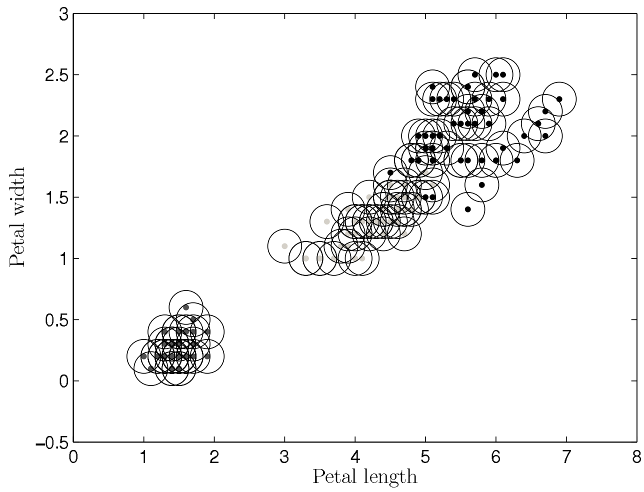

Fig. 7. Clustering with a graph theoretic approach to the Iris data set. It is clearly observed that the cluster of Iris Setosa is separated from the clusters of other Iris species with an appropriate size of circle r_T in the projection of petal length and petal width plane (Fig. 6).

can obtain the centre point of this known cluster and then the standard deviation of the distances between the data points and the centre point of the cluster. Now we assume that the distribution of the distances between each data point and the centre point is normal in the cluster (this assumption will be further discussed in Subsection 4.1).

Consider a new data point. Assume that this point has to be classified to one of the already known clusters. Then we know each centre point of the clusters and standard deviation from its centre point to the data points of each cluster, henceforth we can use these informations. For the new data point its probability to be in a known cluster is statistically estimated with its distance to the centre point of the cluster by integrating the probability density function (see Fig. 8). With the comparison of these

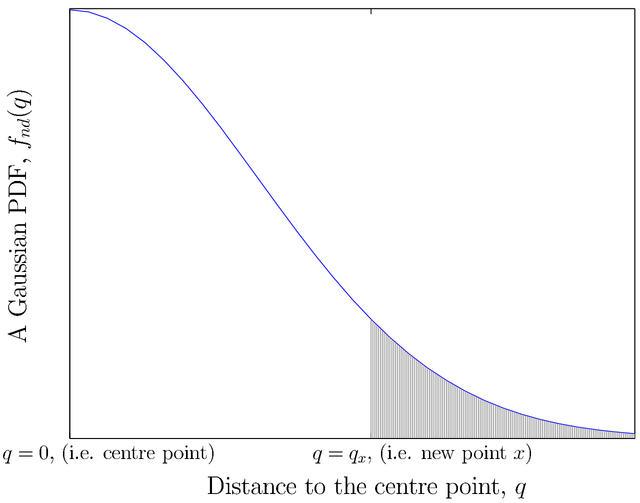

Fig. 8. Example of a one-sided Gaussian probability density function, $f_{nd}(\cdot)$ where $\int_0^\infty f_{nd}(q)\,\mathrm{d}q = 1$. If a new data point is found q_x away from the centre point of a cluster, we suggest an estimation for the probability to be in the cluster as $\int_{q_x}^\infty f_{nd}(q)\,\mathrm{d}q$ $(= 1 - \int_0^{q_x} f_{nd}(q)\,\mathrm{d}q)$.

statistic estimations to the candidate clusters, we can try to classify the new data point into one of the clusters.

Note that the statistic estimation does not correspond to a *probability* even if we obtain this estimation by means of the integration of a probability density function. The estimation does not have the meaning of probability and sum of these estimations of a new data point for several different clusters can be more than 1. However this estimation value can provide a comparison of possibilities for a new data point to be classified into considered clusters, and it is evaluated on the basis of the data set of known clusters.

Consider a Gaussian distribution function $f_{nd}(q)$ for a cluster where q is the distance to the centre point of the cluster. Thus, to define the function $f_{nd}(q)$, we should have the centre point and the standard deviation of the distances from the centre point to the data points of the cluster. Note that, due to the positivity of the distance to centre point, the Gaussian distribution function should be one-sided, hence

$$\int_0^\infty f_{nd}(q)\,\mathrm{d}q = 1.$$

The suggested probability-like estimation can be obtained as

$$\int_{q_x}^\infty f_{nd}(q)\,\mathrm{d}q = 1 - \int_0^{q_x} f_{nd}(q)\,\mathrm{d}q,$$

where a new data point to be classified is located q_x away from the centre point of the considered cluster (see Fig. 8). This estimation is proportional to the closeness of the new data point to the centre point. By the suggested estimation, the closer a new point is to the centre point of a cluster, the more likely the new point is included in the cluster.

Now we consider only the 100 Iris Versicolor and Iris Virginica data points in the Anderson's Iris flower data set (the light grey and the black points plotted in Fig. 6). The 50 Iris Setosa data points are not considered in this section since the Iris Setosa can be completely classified by means of several methods such as Casagrande and Astolfi (2008) or the graph-theoretic method in Section 2. Note that the perfect classification of Iris Setosa is also achieved by the method of this section while this is not further discussed in the paper due to page limitation.

We study the problem of estimating the probability that a data point is classified in one of the two categories, Iris Versicolor or Iris Virginica, based on the remaining 99 data points, which we assume known.

For a cluster, the light grey or the black points in Fig. 6, we obtain the centre points and the standard deviations of distances of all available 99 points in each cluster to its centre point. For the data point which we are interested in, we obtain the distances to the centre points of the two clusters. Then the likeliness of the data point to be included in the clusters can be estimated on the basis of each probability density function.

The classification result is shown in Fig. 9. The light grey circles and the black circles correspond to the Anderson's data points of Iris Versicolor and Iris Virginica, respectively, as in Fig. 6. Note that Fig. 9 represents a part of Fig. 6 or Fig. 7. The centre points of the light grey and the black clusters are $(4.260, 1.326)$ and $(5.552, 2.026)$, respectively. The light grey and black clusters have standard deviations of the distances of the data to each centre point, 0.2786 and 0.2878, respectively.

We apply the suggested classification method to the 100 points one by one and the decision result is indicated with light grey and the black 'star' mark. The decision is made based on the comparison of two estimations to each cluster. Light grey star and black star imply that the data point is regarded as Iris Versicolor and Iris Virginica, respectively. Thus a light grey star mark in a light grey circle and a black star mark in a black circle show that the decision is consistent with the Anderson's Iris data set.

In Fig. 9 three inconsistent star marks are presented: two black stars in two light grey circles and one light grey star in one black circle. However some of the data points in the figure are overlapping. With this Gaussian based method 5 data samples are incorrectly classified, which implies that 145 samples are correctly classified.

This 96.67% classification success rate is a very good result for the Anderson's Iris data set, as stated in Vathy-Fogarassy et al. (2006). Note that we do not need to set up any parameters in this methodology. To achieve the performance of our Gaussian based classification some parameters have to be fine-tuned in Vathy-Fogarassy et al. (2006).

3.1 Nonlinear Discriminant Analysis

In the previous approach we use only the information of petal length and petal width. Particularly the Iris Versicolor group and the Iris Virginica group have a $(4.8, 1.8)$ point and two $(4.8, 1.8)$ points in the plane of petal length versus petal width, respectively. Thus we have to use other information on the Iris data set, sepal length or sepal width, to identify cluster of such points projected to $(4.8, 1.8)$ in the plane.

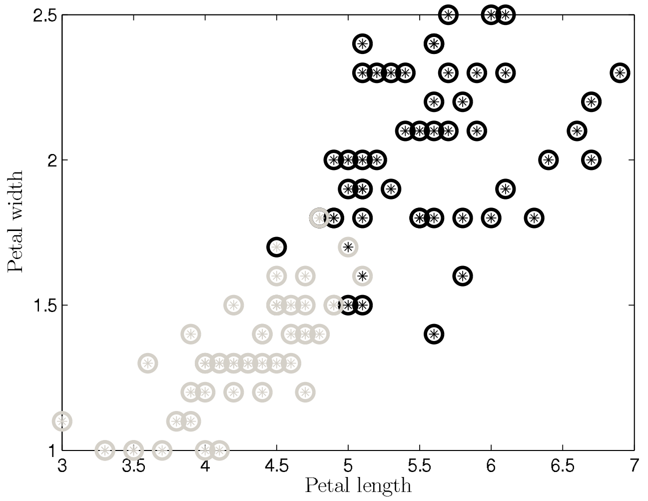

Fig. 9. Classification result of the 100 Iris data points of the two species, Iris Versicolor and Iris Virginica in the Anderson's Iris flower data set. Note that this figure is part of Fig. 6 or Fig. 7. The centre points of the light grey and the black clusters are $(4.260, 1.326)$ and $(5.552, 2.026)$, respectively.

In addition, based on the inspection of the pictures of the two species (Fig. 2 and Fig. 3) it is clear that the sepal length or sepal width could be a key discriminant factor to distinguish these two species. To overcome the limitation of the Gaussian based approach, we now suggest a plane with sepal width and 'petal area' (e.g. the product of petal length and petal width) for a nonlinear discriminant analysis. This idea is initiated by inspection of Fig. 2 and Fig. 3. See Fisher (1936) for an example of the use of linear discriminant analysis in the solution of this problem and see Baudat and Anouar (2000); Mika et al. (1999) for a general introduction on nonlinear discriminant analysis.

Fig. 10 shows an implementation of the approach. The Iris data points are plotted in the plane of sepal width versus the product of petal width and petal length. The light grey points and the black points correspond to the groups of Iris Versicolor and Iris Virginica, respectively. The dotted line shows that this nonlinear discriminant approach can be used to discriminate between two clusters of Iris Versicolor and Iris Virginica. Although this dotted line is linear in the plane, we can call it a nonlinear discriminant analysis since an axis of the plane corresponds to 'petal length $\times$ petal width'.

Only 3 misclassifications are found in Fig. 10, the 3 light grey points above the dotted line. Thus, with this nonlinear discriminant analysis, classification correctness is improved up to 98% (3 misclassifications of the 150 point of the Iris data set). Particularly only one of the 3 points projected to the $(4.8, 1.8)$ points in the plane of petal length versus petal width is misclassified whereas two of the points are in the Gaussian approach.

4. CONCLUSION

We employ the idea of using the centre data point and the standard deviation of a known cluster with assumption of Gaussian distribution of the data point. In our Gaussian

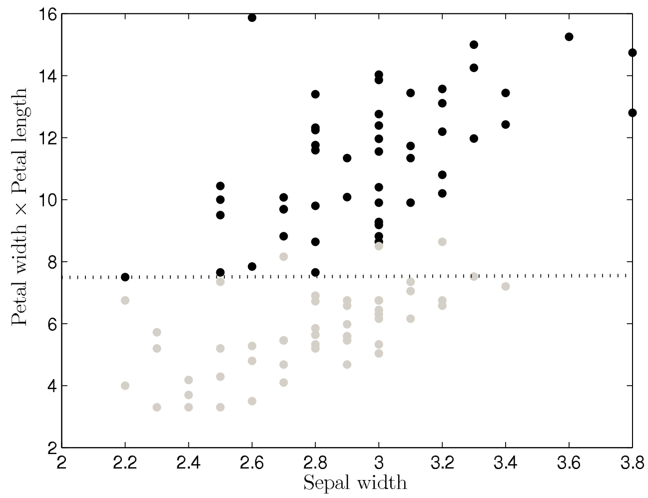

Fig. 10. Plot of the Iris data point in the plane of sepal width versus the product of petal width and petal length. The light grey points and the black points correspond to the groups of Iris Versicolor and Iris Virginica, respectively. The dotted line discriminates between the two groups only with 3 misclassifications.

based classification we do not require to tune any parameter.

The algorithm of Dubnov et al. (2002) has classified correctly 124 samples for the Iris data set. They claim that the result is comparable with results obtained by other algorithms. Also it is claimed that, in Dubnov et al. (2002), the Super-Paramagnetic pairwise clustering algorithm of Blatt et al. (1996) has classified correctly 125 samples in the data, which is the second best result next to the minimum spanning tree algorithm, and the best performing algorithm on this Iris data set example is the minimum spanning tree. More recently an improved minimum spanning tree algorithm of Vathy-Fogarassy et al. (2006) has shown only 5 misclassifications in the application to the Iris data example with a fine-tuning clustering step of their procedure.

Note that we achieve the same classification correctness 96.67% of Vathy-Fogarassy et al. (2006) without any tuning in the classification procedure in Section 3 and moreover we improve this classification correctness up to 98% by introduction of nonlinear discriminant analysis in Subsection 3.1. Thus, to the best knowledge of the authors, this paper gives the best result for the clustering problem of the Iris flower data set.

4.1 Future Work

In Section 2 we introduce a graph-theoretic approach based on an appropriate r_T and adjacency matrix. For the next step of this approach we need to study algorithms to decide the value of r_T and to manipulate the matrix to obtain the corresponding block diagonal upper triangular adjacency matrix. Then we will research cluster identification method in the block of the matrix by Proposition 1 or Corollary 2 and we can study dynamic clustering of the LIDAR data set which describes moving pedestrians.

In Section 3 we assume that the data point distribution to its centre point is Gaussian and this implies that the data points approximately consist of a sphere formation in the metric space. Although we can show a good classification result for the Iris data set with this assumption, it would be unrealistic for some geometric distribution, such as ring (doughnut) formation, 'V' formation, and so forth. Further study on this assumption is needed to apply the classification method to various kinds of data examples.

REFERENCES

Aksoy, S. and Haralick, R. (1999). Graph-theoretic clustering for image grouping and retrieval. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, 68 Vol. 1.

Anderson, E. (1935). The irises of the gaspe peninsula. *Bulletin of the American Iris Society*, 59, 2–5.

Augustson, J. and Minker, J. (1970). An analysis of some graph theoretical cluster techniques. *Journal of the Association for Computing Machinery*, 17(4), 571–588.

Baudat, G. and Anouar, F. (2000). Generalized discriminant analysis using a kernel approach. *Neural Computation*, 12(10), 2385–2404.

Blatt, M., Wiseman, S., and Domany, E. (1996). Clustering data through an analogy to the Potts model. In *Advances in Neural Information Processing Systems 8*, 416–422. MIT Press.

Cannon, A., Cowen, L., and Priebe, C. (1998). Approximate distance classification. In *Proceedings of the 30th Symposium on the Interface: Computer Science and Statistics*.

Casagrande, D. and Astolfi, A. (2008). A hamiltonian-based algorithm for measurements clustering. In *Proc. of 47th IEEE Conference on Decision and Control*, 3169 –3174.

Casagrande, D. and Astolfi, A. (2009). Time-dependent hamiltonian functions and the representation of dynamic measurements sets. In *Proc. of the IFAC European Control Conference*.

Casagrande, D., Sassano, M., and Astolfi, A. (2009a). Estimation of the dynamics of two-dimensional clusters. In *Proc. of the 2009 IEEE American Control Conference*, 5428 –5433.

Casagrande, D., Sassano, M., and Astolfi, A. (2009b). A hamiltonian approach to moments-based font recognition. In *Proc. of 48th IEEE Conference on Decision and Control*, 5906 –5911.

D'haeseleer, P. (2005). How does gene expression clustering work? *Nature Biotechnology*, 23(12), 1499–1501.

Domany, E. (1999). Superparamagnetic clustering of data - the definitive solution of an ill-posed problem. *Physics A*, 263, 158–169.

Dubnov, S., El-Yaniv, R., Gdalyahu, Y., Schneidman, E., Tishby, N., and Yona, G. (2002). A new nonparametric pairwise clustering algorithm based on iterative estimation of distance profiles. *Machine Learning*, 47, 35–61.

Ester, M., Kriegel, H., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. of 2nd International Conference on Knowledge Discovery and Data Mining*, 226–231. AAAI Press.

Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(7), 179–188.

Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, 2, 284–304.

Karypis, G., Han, E., and Kumar, V. (1999). Chameleon: hierarchical clustering using dynamic modeling. *Computer*, 32(8), 68 –75.

Krishna, K. and Narasimha Murty, M. (1999). Genetic k-means algorithm. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 29(3), 433 –439.

Langlois, D. (2005). Iris versicolor, photo by D. Langlois in july 2005 at the forillon national park of canada, GFDL. [Online], 22 Sept 2010. Available at `http://en.wikipedia.org/wiki/File:Iris_versicolor_3.jpg`.

Lauritzen, S. (1995). The EM algorithm for graphical association models with missing data. *Computational Statistics & Data Analysis*, 19(2), 191–201.

Mayfield, F. (2007). Iris virginica, uploaded from Flickr upload bot. [Online], 22 Sept 2010. Available at `http://en.wikipedia.org/wiki/File:Iris_virginica.jpg`.

Mika, S., Ratsch, G., Weston, J., Scholkopf, B., and Mullers, K. (1999). Fisher discriminant analysis with kernels. In *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*, 41 –48.

Paivinen, N. (2005). Clustering with a minimum spanning tree of scale-free-like structure. *Pattern Recognition Letters*, 26(7), 921 – 930.

Pan, S. and Cheng, K. (2007). Evolution-based tabu search approach to automatic clustering. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(5), 827 –838.

Punj, G. and Stewart, D. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, 20(2), 134–148.

Radomil (2005). Iris setosa, photo by Radomil, GFDL. [Online], 22 Sept 2010. Available at `http://en.wikipedia.org/wiki/File:Kosaciec_szczecinkowaty_Iris_setosa.jpg`.

Selim, S. and Alsultan, K. (1991). A simulated annealing algorithm for the clustering problem. *Pattern Recognition*, 24(10), 1003 – 1008.

Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8), 888 –905.

Shi, R., Shen, I.F., and Yang, S. (2005). Supervised graph-theoretic clustering. In *Neural Networks and Brain, 2005. ICNN B '05. International Conference on*, 683 –688.

Vathy-Fogarassy, A., Kiss, A., and Abonyi, J. (2006). Hybrid minimal spanning tree and mixture of gaussians based clustering algorithm. In *Foundations of Information and Knowledge Systems*, volume 3861 of *Lecture Notes in Computer Science*, 313–330. Springer Berlin / Heidelberg.

Wu, Z. and Leahy, R. (1993). An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(11), 1101 –1113.

Zahn, C. (1971). Graph-theoretical methods for detecting and describing gestalt clusters. *Computers, IEEE Transactions on*, C-20(1), 68 – 86.