

# Incremental Maximum Gaussian Mixture Partition For Classification

*Xianbin Hong, Jiehao Zhang, Sheng-Uei Guan, Di Yao, Nian Xue, Xuan Zhao, Xin Huang\**

*\*Department of Computer Science and Software Engineering  
Xi'an Jiaotong-Liverpool University  
Suzhou, China  
Xianbin.Hong12@student.xjtlu.edu.cn*

**Keywords:** *Classification, Gaussian Function, K-means*

## Abstract

In the field of classification, the main task of most algorithms is to find a perfect decision boundary. However, most decision boundaries are too complex to be discovered directly. Therefore, in this paper, we proposed an Incremental Maximum Gaussian Mixture Partition (IMGMP) algorithm for classification, aiming to solve those problems with complex decision boundaries. As a self-adaptive algorithm, it uses a divide and conquer strategy to calculate out a reasonable decision boundary by step. An Improved K-means clustering and a Maximum Gaussian Mixture model are used in the classifier. This algorithm also has been tested on artificial and real-life datasets in order to evaluate its remarkable flexibility and robustness.

## 1. Introduction

**I**N the field of machine learning, Artificial Neuron Network, Support Vector Machine [1] (SVM) and other plenty of algorithms are used for solving classification problem [2]. A common idea of these algorithms is trying to find a decision boundary among two classes [3]. However, in practice, discovering of decision boundary is always a tricky question [4, 5]. First of all, most of datasets are non-linear separated and the decision boundaries are complex [6, 7]. If a dataset's dimension is more than two, its decision boundary is in fact a decision surface and becomes more complex [6, 8]. In addition, the boundaries of different classes could intersect in some cases. Though many algorithms can discover complex decision boundaries, they still cannot satisfy all requirements. For instant, SVM [1, 9] can improve kernel function to approximately close to the real decision boundary, but few kernel functions cannot satisfy various tasks[10]. In order to flexibly represent the decision boundary, researchers try to combine a set of boundaries to simulate the actual boundary. For instance, Jinghao and Binge use a set of hyper-planes[11] or hyper-spheres [12, 13] (HSP [13] use PSO algorithm, and ILEGA [12] use GA algorithm ) to form a complex decision hyper-surface. In generally, the patterns of a class are scatted over the space and form a group of dispersed clusters [14].

Thus, if we find the clusters and their corresponding boundaries, the decision hyper-surface can be defined. IHPP (Incremental Hyper-Plane Partitioning)[11] employs hyper-planes to incrementally slice the space until the area of target class can be distinguished from other classes. IHSP (Incremental Hyper-Spheres Partitioning)[12, 13] uses GA or PSO to dynamically search the whole data space to find the clusters of each class and then a group of hyper spheres will be employed to represent the clusters as classifiers. Research also describes above method as an expert system. Each cluster will be assigned to an expert, which knows the boundary of the corresponding cluster and the whole experts imply the intact decision surfaces. An expert indicates a range of a class's patterns, and the pattern in this range will be regarded as the same class.

Above methods make sense in the practice, but they still have some limitations. First of all, the shape of clusters is complex, so it is not enough if we only use hyper-plane and hyper-sphere to represent it. Secondly, the boundaries of experts are absolute. Once the boundaries of the two classes are mixed, there will be some problems [15]. Finally, there are gap areas outside the experts and it is difficult to judge whether the patterns fall in the gap.

To solve the problems above, IMGMP of this paper uses the ideal of fuzzy learning [16, 17] in the expert system. IMGMP will employ Gaussian model [18] as classifiers in the expert system instead of a specific shape. With this approach, we do not need to consider the shape of cluster [17]. Also, it is not necessary to consider whether a pattern is inside or outside the range of expert, since we only calculate the probability that the pattern is inside [2, 16]. In following parts, we will firstly introduce the algorithm design and secondly present the experimental results. Finally, we summarize the algorithm in the conclusion section, and propose the future work.

## 2. Algorithm Design

The basic procedure of IMGMP is incrementally training each class. For each class, we use clustering method to find the clusters and use Gaussian models to represent them. All the Gaussian models form an expert system, and for each test pattern, this system will find which cluster has the highest

probability for it. Finally, we assign the pattern to the corresponding class of the cluster.

### A. Overview

#### 1) Training phase

IMGMP has two basic training phases, the first stage aims to find the clusters of each class by improved K-means. The second stage will use multivariate Gaussian functions to create Gaussian models for all clusters.

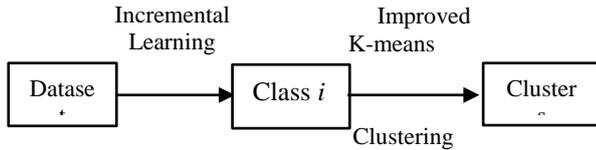


Figure 1. Flowchart of First Stage of Training Phase

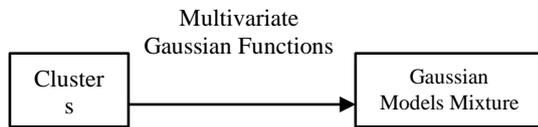


Figure 2. Flowchart of First Stage of Training Phase

#### 2) Test phase

For a test pattern, IMGMP will calculate the probability for each cluster that contains the pattern. Then, it will find the cluster with highest probability and assign the class label to the test pattern.

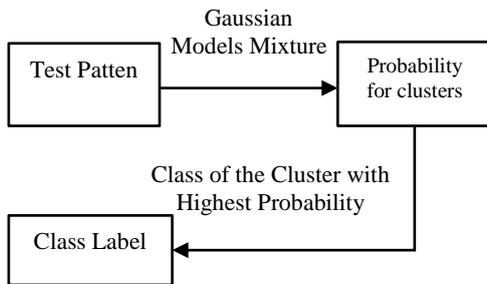


Figure 3. Flowchart of Test Phase

### B. Clustering, Improved K-means

Clustering is generally used in the unsupervised learning field to group the similar patterns together [2, 5]. In classification field, the label is given so that we does not need clustering. However, by using the character of clustering, we can effectively find the clusters in the space. Because clustering methods are innumerable, we choose the most widely applied one—K-means [19, 20].

### C. Canopy K-means

The canopy clustering algorithm is an unsupervised pre-clustering algorithm. It is often used as a preprocessing step for the K-means algorithm or the Hierarchical clustering algorithm. It intends to speed up clustering operations on large data sets while using another algorithm may be impractical due to the size of the data set.

In Canopy K-means, the initial centers will not be random, it will use the result of Canopy clustering algorithm [21, 22]. This will make the initial center more reasonable.

The algorithm proceeds as follows, using two thresholds  $T_1$  (the loose distance) and  $T_2$  (the tight distance), where  $T_1 > T_2$ .

Algorithm: Canopy Clustering Algorithm
<b>Input:</b> A set of data points $X_i$ , thresholds $T_1$ and $T_2$ $C \leftarrow \phi$ ( $C$ is a set of canopies) $\Sigma \leftarrow S$ ( $\Sigma$ is a set of center candidates) <b>while</b> $\Sigma \neq \phi$ <b>do</b> $C \leftarrow$ get a point from $\Sigma$ at random ( $c$ is a center) $C \leftarrow \phi$ <b>for</b> $x \in S$ <b>do</b> <b>if</b> $d(x, c) \leq T_1$ <b>then</b> $C \leftarrow C \cup \{x\}$ (a canopy $C$ includes a point $x$ ) <b>end</b> <b>if</b> $d(x, c) \leq T_2$ <b>then</b> $\Sigma \leftarrow \Sigma - \{x\}$ remove $x$ from the candidates <b>end</b> <b>end</b> $C \leftarrow C \cup \{C\}$ <b>end</b> <b>return</b> $C$

### D. Maximum Gaussian mixture model

As mentioned above, IMGMP will not use specific shapes as partition tools. Instead, it will use statistic and probability model. This way could help algorithm to represent any clusters flexibly with lower cost. Thus, once the clusters gained, we could use Gaussian model to represent each cluster. As Gaussian distribution is the most widespread used distribution model, this paper adopts it.

### E. Maximum Gaussian mixture model

In order to use Gaussian model in the expert systems mentioned above, this paper modifies the GDA, which uses Gaussian model to every cluster rather than a class [2, 23]. We define the Maximum Gaussian mixture model, that is:

$$f(x|\theta) = \max\{N(X|\mu_k, \Sigma_k)\}, k = 1, 2, \dots, K_c$$

Where  $K_c$  is the number of Gaussians function used to describe the class  $c$ . We can use the improved K-means to solve this model for each class. In order to make two Gaussians in different class comparable, we need to set each Gaussians have similar training patterns. For instance,

assuming class 1 and class 2 has  $N_1$  and  $N_2$  points, then  $\frac{N_1}{N_2} = \frac{K_1}{K_2}$ .

For a test pattern  $x$ , we can calculate the probability for each class, that is:

$$p(x|\theta_c) = \frac{\max\{N(x|\mu_{kc}, \Sigma_{kc})\}}{\sum_{j=1}^C \max\{N(x|\mu_{kj}, \Sigma_{kj})\}}$$

The class  $c$  that acquires the maximum  $p$  is what we wanted.

### 3. Experiments

In the experimental section, several datasets including UCI benchmark datasets and artificial datasets were used to evaluate IMGMP. For each dataset, a 10-fold cross validation approach was used to ensure the accuracy of the results.

#### A. Artificial datasets

In the validation period, we created two artificial datasets: CL2 dataset and CLS datasets to evaluate the performance of IMGMP.

CL2 is an artificial two-dimensional dataset with 3401 instances. Two circles partition it into two classes. One locates in a wide ring and another one in the center and corners. Although CL2 includes 6 clusters, the clusters still have clear boundaries and easy to classify.

CLS is a two-dimensional dataset with 3344 instances. Five circles partition it into two classes with five rings and four triangles. CLS has 9 clusters and is more complex than CL2. The ring of CLS is also narrower than CL2 and some parts of boundaries are not clear.

The experiment result is shown in Table 1. ILEGA & HSP are two algorithms mentioned above, which use hyper-spheres to create classifiers.

TABLE I. CLASSIFICATION ACCURACY and UPGRADE RATE COMPARISON upon ARTIFICIAL DATASETS

	ILEGA	HSP	IMGMP	IMGMP vs ILEGA	IMGMP Vs HSP
<b>CL2</b>	0.6678	0.9382	0.9724	31.26%↑	3.51%↑
<b>CLS</b>	0.4587	0.8594	0.9243	50.37%↑	7.02%↑

Table I shows that our approach, IMGMP performs best upon both CL2 and CLS. Especially, the vantage of IMGMP is more obvious in CLS than CL2, which means IMGMP is more suitable for complex situation than hyper-sphere approaches.

For CL2, hyper-spheres approaches can perfectly classify the center area, but it will meet problems in the ring and corners. For CLS, the difficulty increases for hyper-sphere approaches because there are more rings. For a non-sphere area, hyper-sphere approaches must use spheres to fill the area as much as possible. On the one hand, if the spheres are too large, classifiers will ignore many blank areas among spheres. In the ILEGA, the algorithm will discard all small clusters of patterns and cause serious problems. On another hand, if the spheres are too small, it can easily lead to overfitting. Although HSP use more and smaller spheres than ILEGA and significantly increase the precision, the precision of test period also observably decrease. The experiments results prove that IMGMP has more flexibility and robustness than hyper-sphere approaches.

#### B. Real life datasets

Four UCI real life benchmark datasets: Cancer, Wine, Sonar and Glass were used to evaluate the performance of IMGMP.

TABLE II. CLASSIFICATION ACCURACY and UPGRADE RATE COMPARISON upon REAL-LIFE DATASETS

	ILEGA	HSP	IMGMP	IMGMP vs ILEGA	IMGMP vs HSP
<b>Cancer</b>	0.9059	0.9177	0.9633	6.34%↑	4.98%↑
<b>Wine</b>	0.8806	0.9287	0.9722	10.41%↑	4.69%↑
<b>Sonar</b>	0.4638	0.5333	0.8279	10.41%↑	55.22%↑
<b>Glass</b>	0.4766	0.4455	0.6720	40.98%↑	50.84%↑

Cancer (Breast Cancer Wisconsin (Original)) dataset records 4-year-old clinical cases, and has 683 instances (16 instances with missing attributes were discard). It is a binary dataset and has 9 attributes. Its decision boundary is quite simple people can even use a hyper-plane to represent it [24].

Wine dataset includes three classes of 178 instances with 13 attributes. Each attribute represents a chemical attribution of wine, using for determining the origin of wine.

Sonar dataset (Connectionist Bench (Sonar, Mines vs. Rocks)) contains 208 instances to classify mines and rocks by sonar signals. It has 60 attributes and is a high dimension dataset.

Glass dataset has 214 instances and 10 dimensions. This dataset is difficult to classify because it has 6 classes and some of them have high similarity. For instances, two classes of it are the same type glasses with different manufacturing technique. In the experiments, we still use ILEGA and HSP to compare with the performance of IMGMP.

From Table II, it is obvious that IMGMP performs much better than hyper-sphere approaches upon Sonar and Glass datasets. Sonar is a high-dimensional dataset and Glass is a multi-class dataset, which means IMGMP has advantages upon high-dimensional or multi-class datasets. In high-dimensional cases, the patterns are sparse in the data space and far from others. In order to involve all patterns of a cluster, the radius of hyper-sphere should be very large. If the radius is too small, the hyper-sphere will only contain a few patterns and some of them may be discarded. This tricky situation means that hyper-sphere approaches may face problems in high-dimensional environment. In multi-class case, the patterns of Glass distribute as bands and are always mixed together. It has the most complex distribution among the four real life dataset. In addition, since the dataset only has 214 instances, most of classes only have 20~40 patterns and one class even has less than 10 patterns. In this case, it is hard for hyper-sphere approaches to create suitable hyper-sphere to classify patterns.

#### 4. Conclusion And Further Work

In this paper, we proposed an Incremental Maximum Gaussian Mixture Partition (IMGMP) approaches for classification. IMGMP extends the previous research of IHPP [5] and IHSP [12, 13]. It breaks through the restriction of classifiers' shape and raises a self-adaptive method to create classifiers. Comparing with hyper-plane or hyper-sphere approaches, IMGMP is able to apply for datasets with any pattern distribution. From low dimension to high dimension, from two classes to multi-classes, IMGMP performs better than previous approaches and shows a preferable adjustability.

However, we still discontent with the clustering methods in the first stage of training phase. Although the Improved K-means has been adopted, the clustering results are still imperfect. The cluster size will significantly influence the performance of IMGMP, but there is still not a standard to determine a suitable size. Currently, the researchers just set a default possible cluster size for IMGMP. Obviously, a default size is not suitable for all datasets and IMGMP still could be improved. In further research, we will progress the clustering method and implement a more robust approach to automatic determine the size of cluster.

#### ACKNOWLEDGEMENT

This research is supported by Jiangsu Provincial Science and Technology under Grant No. BK20131182, China.

#### References.

[1] H. Zhang, A. C. Berg, M. Maire, and J. Malik, "SVM-KNN: Discriminative nearest neighbor classification for visual category recognition," in *Computer Vision and Pattern Recognition*, 2006 IEEE Computer Society Conference on, 2006, pp. 2126-2136.

[2] C. M. Bishop, "Pattern Recognition," *Machine Learning*, 2006.

[3] T. G. Dietterich, "Machine learning for sequential data: A review," in *Structural, syntactic, and statistical pattern recognition*, ed: Springer, 2002, pp. 15-30.

[4] D. Michie, D. J. Spiegelhalter, and C. C. Taylor, "Machine learning, neural and statistical classification," 1994.

[5] P. Arabie, L. J. Hubert, and G. De Soete, *Clustering and classification*: World Scientific, 1996.

[6] J. Quionero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, *Dataset shift in machine learning*: The MIT Press, 2009.

[7] C. Lee and D. A. Landgrebe, "Decision boundary feature extraction for neural networks," *Neural Networks, IEEE Transactions on*, vol. 8, pp. 75-83, 1997.

[8] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144-152.

[9] K.-B. Duan and S. S. Keerthi, "Which is the best multiclass SVM method? An empirical study," in *Multiple classifier systems*, ed: Springer, 2005, pp. 278-285.

[10] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik, "Feature selection for SVMs," in *NIPS*, 2000, pp. 668-674.

[11] T. Yang, S.-U. Guan, J. Song, B. Zheng, M. Cao, and T. Yu, "Incremental Hyperplane Partitioning for Classification," *International Journal of Applied Evolutionary Computation (IJAEC)*, vol. 4, pp. 67-79, 2013.

[12] J. Song, S.-U. Guan, and B. Zheng, "Incremental Hyper-Sphere Partitioning for Classification," *International Journal of Applied Evolutionary Computation (IJAEC)*, vol. 5, pp. 72-88, 2014.

[13] B. Zheng, S.-U. Guan, and J. Song, "A PSO Based Incremental Hyper-Sphere Partitioning Approach to Classification," *International Journal of Applied Evolutionary Computation*, vol. 5, pp. 58-71, 2014.

[14] C. Lee and D. A. Landgrebe, "Decision boundary feature extraction for nonparametric classification," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 23, pp. 433-444, 1993.

[15] S. I. Reynolds, "Adaptive resolution model-free reinforcement learning: Decision boundary partitioning," in *Proceedings of the Seventeenth International Conference on Machine Learning*, 2000, pp. 783-790.

[16] R. R. Yager and L. A. Zadeh, *An introduction to fuzzy logic applications in intelligent systems* vol. 165: Springer Science & Business Media, 2012.

[17] N. B. Karayiannis and J. C. Bezdek, "An integrated approach to fuzzy learning vector quantization and fuzzy c-means clustering," *Fuzzy Systems, IEEE Transactions on*, vol. 5, pp. 622-628, 1997.

[18] C. E. Rasmussen, "Gaussian processes for machine learning," 2006.

[19] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl, "Constrained k-means clustering with background knowledge," in *ICML*, 2001, pp. 577-584.

[20] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern recognition letters*, vol. 31, pp. 651-666, 2010.

[21] S. Owen, R. Anil, T. Dunning, and E. Friedman, *Mahout in action*: Manning Shelter Island, 2011.

[22] J. Wang, J. Wang, Q. Ke, G. Zeng, and S. Li, "Fast approximate k-means via cluster closures," in *Multimedia Data Mining and Analytics*, ed: Springer, 2015, pp. 373-395.

[23] J.-L. Gauvain and C.-H. Lee, "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," *Speech and audio processing, IEEE transactions on*, vol. 2, pp. 291-298, 1994.

[24] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert systems with applications*, vol. 36, pp. 3240-3247, 2009.