

# Classification for Overlapping Classes Using Optimized Overlapping Region Detection and Soft Decision

Wenyin Tang, K. Z. Mao

School of Electrical & Electronic Engineering  
Nanyang Technological University  
Singapore

[wenyintang@ntu.edu.sg](mailto:wenyintang@ntu.edu.sg), [ekzmao@ntu.edu.sg](mailto:ekzmao@ntu.edu.sg)

Lee Onn Mak, Gee Wah Ng

DSO National Laboratories  
Singapore

[mleeonn@dso.org.sg](mailto:mleeonn@dso.org.sg), [ngeewah@dso.org.sg](mailto:ngeewah@dso.org.sg)

**Abstract** – In many real applications such as target detection and classification, there exist severe overlaps between different classes due to various reasons. Traditional classifiers with crisp decision often produce high rates of misclassifications for patterns in overlapping regions. In this paper, we propose to use soft decision strategy with an optimized overlapping region detection to address the overlapping class problem. In contrast to crisp decision that assigns a single label to a pattern, the soft decision strategy provides multiple decision options to system operators for further analysis, which is believed to be better than producing a wrong classification. The effectiveness of the proposed method has been tested on both artificial and real-world problems.

**Keywords:** pattern recognition, overlapping classes

## 1 Introduction

As a well developed technology, pattern classification has been applied in a variety of applications such as target detection and classification. Figure 1 shows a typical target classification system, where the raw data is collected from multiple sensors and is aligned to form a track, which is then classified by the classifier built upon the training data generated from the known target library. The classification results are finally output for further analysis.

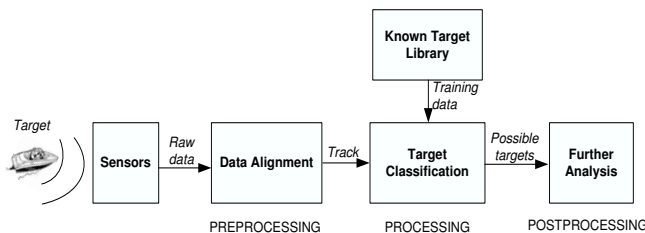


Figure 1: Target Classification System.

The known target library in Figure 1 contains the updated targets' characteristics either from long period of collection or provided by manufacturers. In the feature space, the char-

acteristics define a specific region for each of the known targets. Ideally, the regions occupied by different targets are well separated. In real scenarios, however, there might exist overlaps between different regions due to various reasons. For example, a sensor (same brand from a manufacturer) can be deployed in different target platforms with modified sensor parameter values in a range given by the manufacturer, which may generate either partial or full overlaps.

When classifying a track (i.e. a pattern) that falls into the overlapping regions of the target classes, the traditional classifiers usually go into dilemma of deciding which target class is the probable outcome. The crisp decision usually outputs a single class under the “winner-take-all” rule, and this will inevitably cause high rates of misclassifications (i.e. false alarm and false negative) around decision boundary.

In the literature, just a few work addressed the overlapping class problem. In [1], complementary features were introduced to separate overlapping classes; in [2], multistage classifier or multiple classifiers were employed; while in [3], overlapping classes were simply merged to form meta-classes. All the above work is under the framework of crisp decision, where a single class label is assigned to a pattern, regardless of the location of the pattern (i.e. in overlapping regions or non-overlapping regions), and the classification error directly correlates with the overlapping degree between classes.

A more effective solution to the overlapping class problem is to use soft decision strategy. In contrast to crisp decision that assigns a single label to a pattern, the soft decision assigns multiple labels to patterns falling into the overlapping regions. The rationale of soft decision is that providing multiple decision options to system operators for further analysis is better than providing a wrong classification. The idea of soft decision was applied to fuzzy ARTMAP in our previous study [7], which demonstrated strong capacity of soft decision to manage category proliferation caused by overlapping classes. But the predefined vigilance parameter therein often results in over-sized overlapping regions.

In this paper, we propose to deal with overlapping class problem using soft decision combined with an optimized

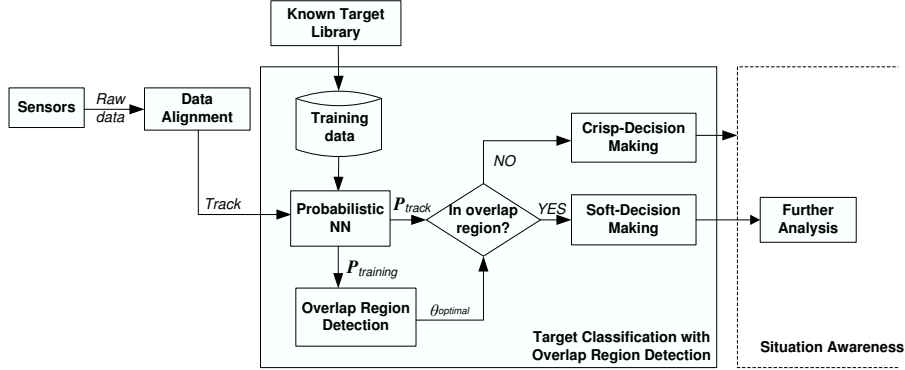


Figure 2: Diagram of classification system with optimized overlapping region detection and soft decision

overlapping region detection algorithm. The overlapping region detected optimizes a performance index that balances the classification accuracy of crisp decisions and the cost of soft decisions (i.e. the proportion of patterns with soft decision. Since these patterns need further analysis of system operators, they are considered as cost). The optimized overlapping regions divide the whole feature space into two parts with “low” and “high” confidence of correct classification respectively. For test data falling into overlapping regions, multiple decision options as well as measures of confidence (e.g. posterior or error probabilities) are produced for further analysis, while for test data falling into non-overlapping regions, crisp decisions are made.

The rest of the paper is organized as follows. The classification system with optimized overlapping region detection will be introduced in Section 2. Section 3 will discuss the evaluation issue for pattern classification under soft decision framework. Experiment results will be presented in Section 4. Section 5 will conclude the paper.

## 2 Pattern Classification with Optimized Overlapping Region Detection and Soft Decision

Figure 2 shows the pattern classification system with optimized overlapping region detection and soft decision. The training procedure first builds a probabilistic model using the training data generated from the known target library. The probabilistic information of training data ( $P_{training}$ ) is then used to search an optimal threshold ( $\theta^*$ ) that defines the overlapping region. In the testing stage, the incoming track (test data) is first processed to find its location (i.e. in overlapping regions or non-overlapping regions), and crisp or soft decision is then made accordingly.

A probabilistic model is well-suited for soft decision making because of its probabilistic classification output. This paper adopts probabilistic neural network (PNN) [4, 5] as the probabilistic model, since PNN is able to handle multi-class problems.

In the following two subsections, the optimized overlapping region detection and the soft decision strategy will be

introduced respectively.

### 2.1 Optimized Overlapping Region Detection

In feature space  $X$ , the overlapping region  $R(\theta)$  is centered at the decision boundary with a margin at each side of the boundary. The width of the margin is determined by the threshold  $\theta$  as follows:

$$R(\theta) = \{X : |P(\omega_1|X) - (P(\omega_2|X))| < \theta\} \quad (1)$$

where  $P(\omega_i|X)$  is the posterior probability of class  $i$ , given  $X$ . The location of decision boundary, i.e.  $P(\omega_1|X) = P(\omega_2|X)$ , is determined by the class distribution of training data and the probabilistic model employed. In overlapping region detection, the key problem is how to find a suitable value for threshold  $\theta$ .

In overlapping region detection, two considerations should be taken into account. On one hand, the overlapping region should be large enough to cover most potentially misclassified data so that the classification in non-overlapping regions is highly accurate. On the other hand, the overlapping region should be compact so as to avoid making soft decision to too many patterns, because patterns with soft decision will finally be verified by system operators and hence increase cost. To implement the above considerations in overlapping region detection, two criteria, i.e. the classification accuracy in non-overlapping regions  $acc_{\bar{o}}$  and the cost (i.e. the proportion of patterns in non-overlapping regions)  $r_{\bar{o}}$ , are used and defined as follows:

$$acc(\theta) = \frac{\text{No. correctly classified patterns}}{\text{No. patterns in non-overlapping region}} \quad (2)$$

$$r_{\bar{o}}(\theta) = \frac{\text{No. patterns in non-overlapping region}}{\text{No. training patterns}} \quad (3)$$

To find a good tradeoff between  $acc$  and  $r_{\bar{o}}$ , an aggregate performance evaluation criterion of  $acc$  and  $r_{\bar{o}}$  is needed. Here we use the weighted harmonic mean, as Eq. (4):

$$J_\beta(\theta) = \frac{1}{\frac{\beta}{acc(\theta)} + \frac{1-\beta}{r_{\bar{o}}(\theta)}} \quad (4)$$

where  $\beta$  is the weight ranged from 0 to 1. This weight parameter can be predefined by a user who wants to attend to the accuracy in the non-overlapping region  $\frac{\beta}{1-\beta}$  times as important as the volume of non-overlapping region. The default value of  $\beta$  is set to 0.85, since the decreasing rate of  $r_{\bar{o}}$  is always faster than the increasing rate of  $acc_{\bar{o}}$ . The optimal threshold  $\theta^*$  is the one that maximizes criterion  $J_\beta(\theta)$ :

$$\theta^* = \arg \max_{\theta} J_\beta(\theta) \quad (5)$$

By using this optimization criterion, the optimal volume of the overlapping region is able to adapt to various data distributions and overlapping degrees as shown in experimental study in Section 4.

The above criterion can be easily extended to multiple overlapping classes, by simply replacing the overlapping region defined in Eq. (1) with a generalized form as follows:

$$R(\theta) = \{X : P(\omega_{rank1}|X) - P(\omega_{rank2}|X) < \theta\} \quad (6)$$

where  $\omega_{rank1}$  and  $\omega_{rank2}$  are the classes obtaining the maximum and  $2^{nd}$  maximum posterior probabilities given  $X$ . Note that, using Eq. (6), the overlapping region is defined as the region occupied by at least, but not limited to, two classes. Next, we will further discuss the determination of the number of overlapping classes for any unknown data falling in the overlapping region.

## 2.2 Soft Decision Strategy

In two-class problems, soft decision for the test data falling in the overlapping region is just class 1 and 2, ranked by the value of posterior probabilities. In multi-class problems, however, the soft decision, which decides the possible classes of a pattern in overlapping regions, is not so straightforward.

In this work, we search the most suitable number of top classes for the input testing data using a novel method called ideal center matching (ICM) in the transformed  $L$ -dimensional prediction space, where the  $l^{th}$  dimension corresponds to the  $l^{th}$  largest posterior probability in prediction ranking list. For example, in a 3-class problem, the ideal centers for 2-class and 3-class overlapping regions are  $[1/2, 1/2, 0]$  and  $[1/3, 1/3, 1/3]$ , respectively, as shown in Figure 3. The testing data will be assigned to the  $M$ -class overlapping region if  $\mathbf{m}_M$  is the nearest ideal center of the data in the prediction space.

Generally, the ideal centers of 2-class to  $L$ -class overlapping regions are defined as follows.

$$\begin{bmatrix} \mathbf{m}_2 \\ \mathbf{m}_3 \\ \vdots \\ \mathbf{m}_L \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & \cdots & 0 \\ 1/3 & 1/3 & 1/3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1/L & 1/L & 1/L & 1/L & \cdots & 1/L \end{bmatrix} \quad (7)$$

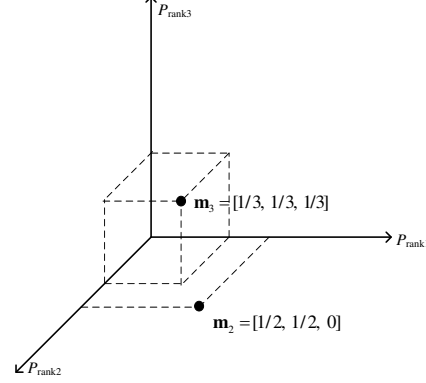


Figure 3: Ideal centers for 2-class and 3-class overlapping regions in prediction space spanned by the ranked posterior probabilities

For an input data  $\mathbf{x}$ , the original PNN will output the set of a posteriori probabilities, i.e.  $\mathbf{p}_\mathbf{x} = [P(\omega_1|\mathbf{x}), P(\omega_2|\mathbf{x}), \dots, P(\omega_L|\mathbf{x})]$ . Based on the optimal overlapping region  $\theta^*$ ,  $\mathbf{x}$  will fall in either the overlapping region or non-overlapping region. As  $\mathbf{x}$  falling in the non-overlapping region, the topmost class label with the highest posterior probability will be assigned to the testing data, based on MAP principle. As  $\mathbf{x}$  falling in the overlapping region, ICM will search the most possible number of overlapping classes in the following three steps:

**Step 1:** Sort the  $L$  classes in a descending order of a posteriori probabilities,  $\mathbf{p}_\mathbf{x} \rightarrow \mathbf{p}_\mathbf{x}^{sort}$ , where the topmost labels are more likely to be the multiple class labels of the input data  $\mathbf{x}$ .

**Step 2:** Calculate the Euclidean distance between  $\mathbf{p}_\mathbf{x}^{sort}$  and  $\mathbf{m}_l$  ( $l = 2, \dots, L$ ), and the most possible number of classes  $l^*$  is given by:

$$l^* = \arg \min_l \|\mathbf{p}_\mathbf{x}^{sort} - \mathbf{m}_l\|; \quad (8)$$

where,  $\mathbf{m}_l$  is the cluster center of the overlapping region occupied by the topmost  $l$  classes, as shown in Eq. (7).

**Step 3:** Output the top- $l^*$  classes as the multiple class label of the input pattern  $\mathbf{x}$  associated with the corresponding a posteriori probabilities.

The computational complexity of ICM is low because, for each incoming testing data, only  $(L-1)$  Euclidean distances are calculated.

## 3 Evaluation of Soft Decision

In traditional classification with crisp decision, both ground truth and prediction are single-labeled, therefore, the classification accuracy/recognition rate is simply the ratio of correctly classified patterns to the total patterns. Evaluating the performance of classification with soft decision is

not so straightforward since soft decision produces multi-labeled predictions, while the ground truth is single-labeled. As shown in Table 1, for example, when the ground truth of a test data is class  $\omega_1$ , there are three possible categories of predictions, including:

- hit - have only one prediction and hit the target.
- recognize - in the soft decision, the true target is recognized. Note that different ranks of the true class in the prediction ranking list correspond to different costs.
- miss - the target is missed. The number of output will also affect the corresponding cost.

Table 1: Three categories of predictions in target classification under soft decision framework.

Ground truth class: $\omega_1$				
Prediction	Category	Rec	No. output	Cost
$\omega_1$	hit	1	1	0
$\omega_1, \omega_2$	recognize	1	2	0
$\omega_2, \omega_3, \omega_1$	recognize	1	3	2
$\omega_3$	miss	0	1	1
$\omega_3, \omega_4$	miss	0	2	2

Obviously, these three categories of prediction are different in either classification accuracy or cost, or both of them. By using soft decision making strategy, the previously missed targets in the overlapping region may be recognized. Therefore, the recognition-rate (*Rec*) of test data is increased. On the other hand, the cost in this case will also increase by giving extra labels.

Assuming a test data set  $D = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ , where the ground truth of  $\mathbf{x}_i$  is  $y_i$  and the prediction ranking list is  $\mathbf{r}_{\mathbf{x}_i}$ , the recognition-rate and cost are defined as follows.

$$Rec_D = \frac{1}{m} \sum_{i=1}^m h_{\mathbf{x}_i} \quad (9)$$

$$h_{\mathbf{x}_i} = \begin{cases} 1 & \text{if } y_i \in \mathbf{r}_{\mathbf{x}_i} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

$$Cost_D = \frac{1}{m} \sum_{i=1}^m c_{\mathbf{x}_i} \quad (11)$$

$$c_{\mathbf{x}_i} = \begin{cases} rank(\mathbf{x}_i, y_i) - 1 & \text{if } y_i \in \mathbf{r}_{\mathbf{x}_i} \\ |\mathbf{r}_{\mathbf{x}_i}| & \text{otherwise} \end{cases} \quad (12)$$

where  $rank(\mathbf{x}_i, y_i)$  is the rank of  $y_i$  on the prediction ranking list  $\mathbf{r}_{\mathbf{x}_i}$ . The definition of cost for any testing data  $\mathbf{x}_i$  counts the number of misclassifications the system operator need to go through before finding the true class, which accounts for the cost of the system operator who does the further analysis based on the classification output. When these two criteria are used for the traditional pattern classification with crisp

decision, the definition of  $Cost_D$  is the same with the traditional definition of 0-1 loss function, which satisfies the  $Cost_D = 1 - Rec_D$ .

In our classification system using soft decision, when the volume of the overlapping region is increased, the recognition-rate will increase while the cost will increase too. A system with good performance of classification is expected to have a high recognition-rate under a reasonable cost. Therefore, we propose an decision-quality (DQ) criterion as defined in Eq. (13) which combines both recognition-rate and cost, to evaluate the overall performance of the classification system.

$$DQ_\alpha(D) = \frac{1}{\frac{\alpha}{Rec_D} + \frac{1-\alpha}{1-Cost_D}} \quad (13)$$

where  $\alpha$  ranged from 0 to 1 represents the user's *emphasis* on recognition-rate and can be predefined by the user.

## 4 Experimental Study

### 4.1 Data

In this experimental study, the proposed method is tested on both artificial data and real-world problems from UCI machine learning data repository [6].

#### 4.1.1 Artificial Data

Artificial overlapping classes with various data distributions and degrees of overlapping are generated, as shown in Table 2. 200 training samples are generated for each class. The goal of experimenting on artificial data is to test the effectiveness of the proposed overlapping region detection (ORD) algorithm (presented in Section 2) under various situations.

Table 2: Different settings in generating artificial data

Class distributions	Overlapping degree
normal	no overlap
XOR	slight overlap
spherical	moderate overlap
	severe overlap

#### 4.1.2 Real-World Data

Eight real-world multi-class problems from UCI machine learning repository [6] are used to test classification with soft decision. The eight data are summarized in Table 3. These datasets show diversity in the number of classes, samples and features.

## 4.2 Result

### 4.2.1 Overlapping Region Detection for Artificial Data

In this experiment, overlapping region detection (ORD) algorithm is tested on the artificial data with different data distributions and different overlapping degrees, as shown in Table 2. Figures 4-6 show the experimental results for normal, uniform and XOR distributions. In each figure, the

Table 3: Statistics of real-world data

data	no. classes	no. samples	no. features
iris	3	150	4
wine	3	178	13
zoo	7	101	16
ecoli	8	336	7
segment	7	2310	19
vehicle	4	846	18
vowel	11	990	10
yeast	10	1848	8

first row shows the original training data and the second row shows the optimized overlapping region detected from the training data. To explain the result of ORD, we take Figure 4 (d) as an example. The white and grey regions represent the overlapping and non-overlapping regions identified by ORD. The thick line is the classification boundary, while the two fine lines are the boundary of overlapping and non-overlapping regions detected. In each figure, From left to right, the overlapping degree increases gradually from (a) no overlap, (b) slight overlap, (c) moderate overlap, to (d) severe overlap.

Figures 4-6 show that the ORD developed in this work is adaptable to various data distributions and overlapping degrees. For each data distribution, the overlapping region is enlarged with the increasing overlapping degree of data. When there is no overlap between different classes (Figure 4.a, 5.a, 6.a), the volume of the overlapping region is automatically narrowed down to zero. In other words, classification with soft decision will automatically change to the traditional classification with crisp decision for non-overlapping classes.

#### 4.2.2 Comparative Study of Soft Decision with Crisp Decision

The goal of the second experiment is to test soft decision and compare it with crisp decision. Eight real-world multi-class problems used for testing are as shown in Table 3. We evaluate the classification performance of crisp and soft decisions by the recognition-rate, cost and the DQ criterion (Eq.13). For DQ criterion, which is the weighted harmonic average of recognition-rate and cost, the weight parameter  $\alpha$  is varied from 0.1 (emphasizing on cost) to 0.8 (emphasizing on recognition-rate). Note that for the traditional classifiers that make crisp decisions, the value of cost is always equal to one minus recognition-rate, and as a result, the DQ criterion is invariant over the weight  $\alpha$ . Since the traditional classifiers are unable to adjust the tradeoff between recognition-rate and cost, the value of DQ criterion is invariant.

Table 4 shows the experimental results of soft and crisp decisions on eight real-world datasets. Here we use the DQ criterion as the primary criterion and for each data, the higher one is written in bold. From this table, we can observe that when  $\alpha = 0.1$ , i.e. strongly inclining to low cost, the crisp decision performs slightly better than soft decision,

by winning in 4 out of 8 data. When  $\alpha \geq 0.3$ , soft decision performs much better than crisp decision, by winning 7 out of 8 data. When  $\alpha = 0.8$  which shows a greater emphasis on recognition-rate, soft decision outperforms crisp decision in all 8 datasets.

The good performance of soft decision relies on the optimization of the overlapping region, where the tradeoff between recognition-rate and cost has been taken into account when searching the best threshold to detect the overlapping region during the training phase.

#### 4.2.3 Comparative Study with Soft Decision-Based Modified Fuzzy ARTMAP

In the third experiment, we compare the new work with our previous work on modified Fuzzy ARTMAP (m-FAM), which adopts soft decision strategy but does not optimize overlapping regions. [7].

FAM is a supervised clustering algorithm that can be used for classification tasks [8], where each cluster learnt from FAM is mapped to one and only one of  $L$  classes. When it is applied to overlapping classes, the FAM suffers from category (cluster) proliferation. To solve this problem, [7] proposed a modified Fuzzy ARTMAP (m-FAM) by adopting soft decision strategy.

Table 5 shows the experimental results of the new work and the m-FAM. We can observe that the new work outperforms m-FAM on all datasets, and the improvement is more significant for data with severe class overlaps, such as *segment*, *vehicle* and *yeast*. The improved performance of the new work is because of its optimized overlapping region detection module, which leads to an adaptable volume of the overlapping region in different situations. From this table, it is also observed that the volume of overlapping regions detected by the new work is more stable than those found by m-FAM. For data with slight overlaps, such as *wine* and *zoo*, the  $P_{overlap}$  (the percentage of the testing samples classified into overlapping regions) from m-FAM is 0.1% and 0, which is too small. For data with severe overlaps, such as *vehicle* and *yeast*,  $P_{overlap}$  is 62.74% and 92.48%, which is too big and much higher than 19.53% and 45.47% of the new work.

## 5 Conclusion

In real-world problems, the overlapping problem is ubiquitous, due to the imperfectness of features. Traditional classifiers which use the crisp decision strategy may suffer from a significant number of misclassifications in the overlapping region. To solve this problem, a novel framework of pattern classification, including a probabilistic neural network-based probability estimation, a search procedure for an optimized overlapping region detection and a soft decision strategy, has been proposed. A series of evaluation criteria have been defined for soft decision. In experimental study, the proposed method has been tested extensively and three experimental results can be concluded. Firstly, the proposed overlapping region detection algorithm

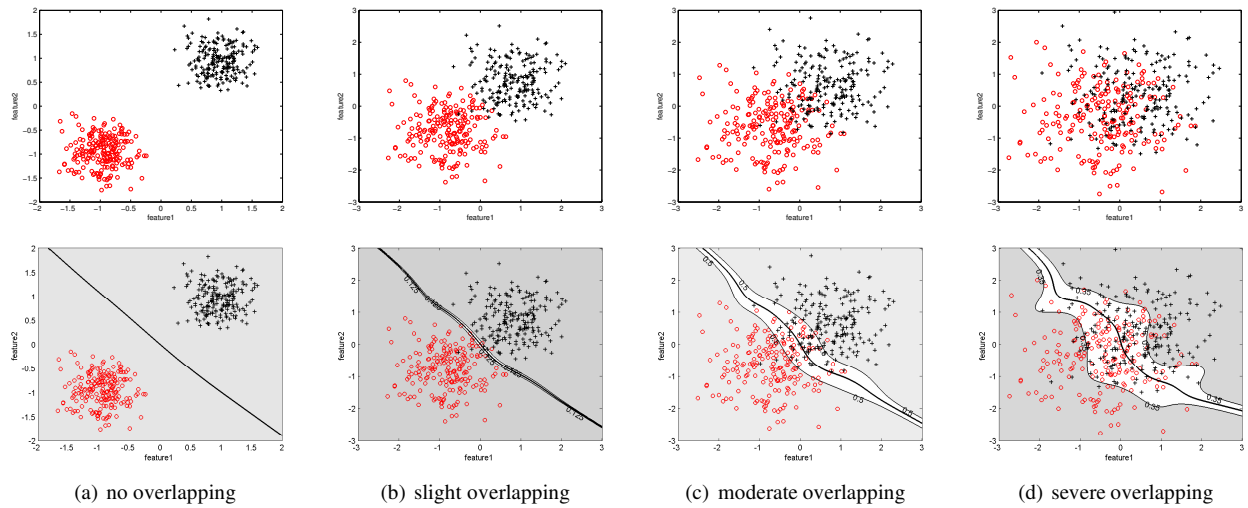


Figure 4: The optimized overlapping regions learnt from the artificial data with normal distributions.

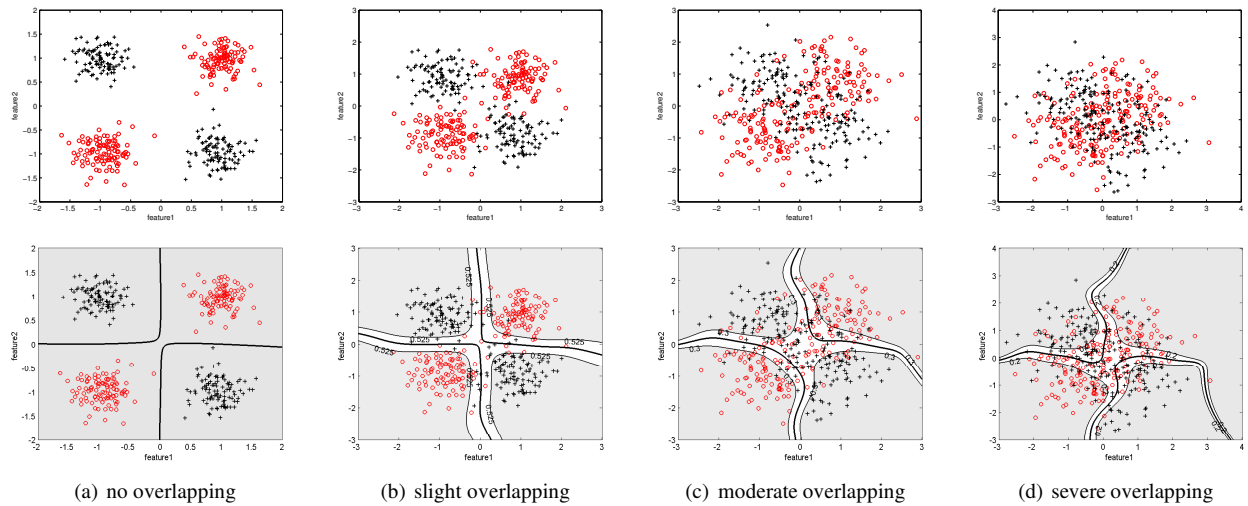


Figure 5: The optimized overlapping regions learnt from the artificial data with XOR distributions.

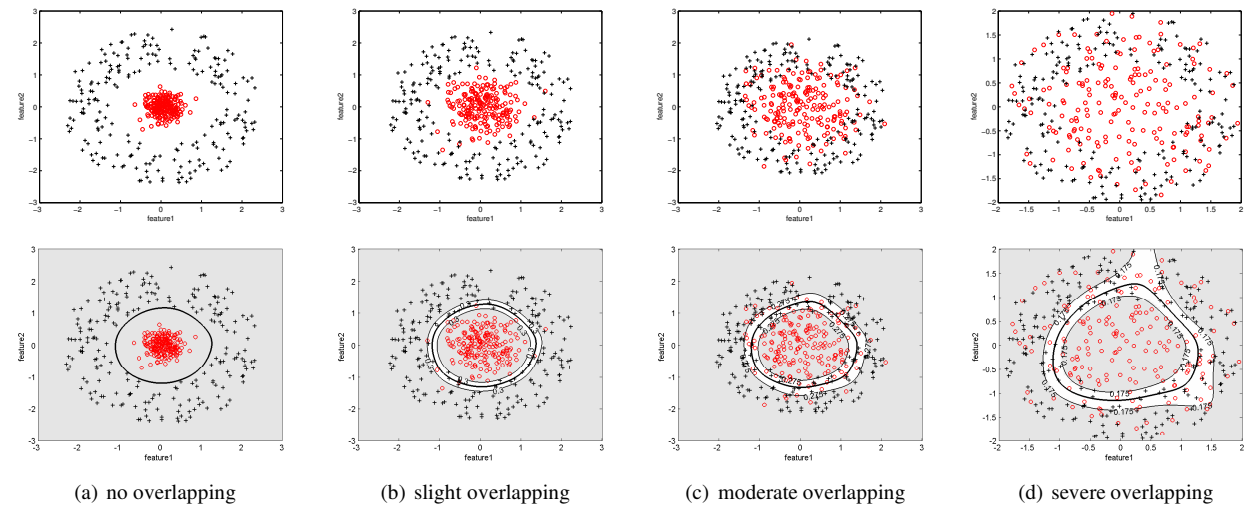


Figure 6: The optimized overlapping regions learnt from the artificial data with spherical distributions.

Table 4: Experimental results of classification with soft decision and the traditional classification with crisp decision on real-world multi-class problems.

Data	Methods	Rec	Cost	$DQ_\alpha$			
				$\alpha=0.1$	$\alpha=0.3$	$\alpha=0.5$	$\alpha=0.8$
iris	crisp decision	0.9497	0.0503	0.9497	0.9497	0.9497	0.9497
	soft decision	0.9812	0.0503	<b>0.9528</b>	<b>0.9589</b>	<b>0.9652</b>	<b>0.9747</b>
wine	crisp decision	0.9560	0.044	0.9560	0.9560	0.9560	0.9560
	soft decision	0.9611	0.044	<b>0.9565</b>	<b>0.9575</b>	<b>0.9585</b>	<b>0.9601</b>
zoo	crisp decision	0.9475	0.0525	0.9499	0.9499	0.9499	0.9499
	soft decision	0.9550	0.0507	0.9499	<b>0.9510</b>	<b>0.9521</b>	<b>0.9539</b>
ecoli	crisp decision	0.8696	0.1304	<b>0.8696</b>	0.8696	0.8696	0.8696
	soft decision	0.9186	0.1372	0.8681	<b>0.8788</b>	<b>0.8898</b>	<b>0.9069</b>
segment	crisp decision	0.9155	0.0845	<b>0.9155</b>	0.9155	0.9155	0.9155
	soft decision	0.9685	0.1006	0.9059	<b>0.9191</b>	<b>0.9327</b>	<b>0.9538</b>
vehicle	crisp decision	0.7099	0.2901	<b>0.7099</b>	0.7099	0.7099	0.7099
	soft decision	0.8050	0.3209	0.6899	<b>0.7125</b>	<b>0.7367</b>	<b>0.7762</b>
vowel	crisp decision	0.9761	0.0239	0.9761	0.9761	0.9761	0.9761
	soft decision	0.9882	0.0243	<b>0.9769</b>	<b>0.9794</b>	<b>0.9819</b>	<b>0.9857</b>
yeast	crisp decision	0.5971	0.4029	<b>0.5971</b>	<b>0.5971</b>	<b>0.5971</b>	0.5971
	soft decision	0.7883	0.5340	0.4859	0.5311	0.5857	<b>0.6925</b>
win/tie/loss:				3/1/4	7/0/1	7/0/1	8/0/0

Table 5: Experimental results of our method, TC-soft, and m-FAM [7] on real-world multi-class problems

Data	Methods	Rec	Cost	$DQ_\alpha$				$P_{overlap}$	No. Predict
				$\alpha=0.1$	$\alpha=0.3$	$\alpha=0.5$	$\alpha=0.8$		
iris	m-FAM	0.9607	0.0757	0.9278	0.9349	0.9421	0.9532	7.24%	1.0724
	soft decision	0.9812	0.0503	<b>0.9528</b>	<b>0.9589</b>	<b>0.9652</b>	<b>0.9747</b>	8.49%	1.0849
wine	m-FAM	0.9044	0.0961	0.9039	0.9040	0.9041	0.9043	0.10%	1.0010
	soft decision	0.9611	0.0440	<b>0.9565</b>	<b>0.9575</b>	<b>0.9585</b>	<b>0.9601</b>	1.07%	1.0107
zoo	m-FAM	0.9269	0.0731	0.9269	0.9269	0.9269	0.9269	0	1
	soft decision	0.9550	0.0507	<b>0.9499</b>	<b>0.9510</b>	<b>0.9521</b>	<b>0.9539</b>	1.4%	1.0144
ecoli	m-FAM	0.8813	0.4410	0.5802	0.6279	0.6841	0.7902	40.28%	1.5214
	soft decision	0.9186	0.1372	<b>0.8681</b>	<b>0.8788</b>	<b>0.8898</b>	<b>0.9069</b>	1.69%	1.1857
segment	m-FAM	0.8871	0.1377	0.8647	0.8696	0.8745	0.8820	19.82%	1.2070
	soft decision	0.9685	0.1006	<b>0.9059</b>	<b>0.9191</b>	<b>0.9327</b>	<b>0.9538</b>	13.23%	1.2019
vehicle	m-FAM	0.8833	0.4597	0.5622	0.6116	0.6705	<b>0.7838</b>	62.74%	1.9456
	soft decision	0.8050	0.3209	<b>0.6899</b>	<b>0.7125</b>	<b>0.7367</b>	0.7762	19.53%	1.2742
vowel	m-FAM	0.8221	0.3004	0.7102	0.7324	0.7560	0.7943	24.28%	1.3364
	soft decision	0.9882	0.0243	<b>0.9769</b>	<b>0.9794</b>	<b>0.9819</b>	<b>0.9857</b>	03.63%	1.0166
yeast	m-FAM	0.9151	1.7831	-	-	-	-	92.48%	3.7467
	soft decision	0.7883	0.5340	<b>0.4859</b>	<b>0.5311</b>	<b>0.5857</b>	<b>0.6925</b>	45.47%	1.7323
win/tie/loss:				8/0/0	8/0/0	8/0/0	7/0/1		

adapts well to various data distributions and various overlapping degrees. Secondly, pattern classification using optimized overlapping region detection and soft decision outperforms the traditional classification with crisp decision on eight real-world multi-class problems, by balancing the tradeoff between recognition-rate and cost. Thirdly, comparing to another soft decision based classifier, i.e. the modified Fuzzy ARTMAP, the new work performs better in all the eight problems, and the improvement is more significant for problems with severe overlapping between classes. This might be because the modified Fuzzy ARTMAP does not optimize the overlapping region and the overlapping region detected by the modified Fuzzy ARTMAP is not as compact as ours.

## References

- [1] B. L. Lu and M. Ito, "Task decomposition and module combination based on class relations: a modular neural network for pattern classification." *IEEE Transactions on Neural Networks*, vol. 10, no. 5, 1999.
- [2] I. T. Podolak, "Hierarchical classifier with overlapping class groups," *Expert Systems with Applications*, vol. 34, no. 1, pp. 673–682, 2008.
- [3] C.-L. Liu, "Partial discriminative training for classification of overlapping classes in document analysis," *International Journal on Document Analysis and Recognition*, vol. 11, 2008.
- [4] D. F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, pp. 109–118, 1990.
- [5] D. F. Specht, "Probabilistic neural networks and the polynomial adaline as complementary techniques for classification," *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 111–121, 1990.
- [6] P. M. Murphy and D. W. Aha, "UCI repository of machine learning databases," *Irvine, CA: University of California*, 1995.
- [7] W. Y. Sit, L. O. Mak, and G. W. Ng, "Managing category proliferation in fuzzy artmap caused by overlapping classes," *IEEE Transactions on Neural Networks*, vol. 20, no. 8, pp. 1244–1253, 2009.
- [8] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy artmap: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 698–713, 2004.