Machine Learning Algorithms for Multiple Autonomous Unmanned Vehicle Operations

Using Support Vector Machine with Adaptively Asymmetric Misclassification Costs for Mine-Like Objects Detection

Xiaoguang Wang, Hang Shao, Xuan Liu and Nathalie Japkowicz
Nathalie Japkowicz Consulting Services

Prepared By:
Nathalie Japkowicz Consulting Services
5714 Queen Mary Rd
Hampstead, Quebec
H3X 1X6
Contract Project Manager: Nathalie Japkowicz
PWGSC Contract Number: W7714-115078/001/SV
CSA: F.-A. Bourque, Defence Scientist, 613-992-3206

DRDC CORA CR 2013–121
August 2013

Defence R&D Canada
Centre for Operational Research and Analysis

Maritime Operational Research Team

National Defence Défense nationale

Canada

Machine Learning Algorithms for Multiple Autonomous Unmanned Vehicle Operations

Using Support Vector Machine with Adaptively Asymmetric Misclassification Costs for Mine-Like Objects Detection

Xiaoguang Wang, Hang Shao, Xuan Liu and Nathalie Japkowicz
Nathalie Japkowicz Consulting Services

Prepared By:
Nathalie Japkowicz Consulting Services
5714 Queen Mary Rd
Hampstead, Quebec
H3X 1X6
Contractor's Document Number: [if used]
Contract Project Manager: Nathalie Japkowicz
PWGSC Contract Number: W7714-115078/001/SV
CSA: F.-A. Bourque, Defence Scientist, 613-992-3206

Defence R&D Canada – CORA

Contract Report
DRDC CORA CR 2013-121
August 2013

Principal Author

Original signed by Xiaoguang Wang

Xiaoguang Wang

Research Scientist

Approved by

Original signed by Dr. R. E. Mitchell

Dr. R. E. Mitchell

Head, Maritime and Air Systems Operational Research

Approved for release by

Original signed by Paul Comeau

Paul Comeau

Chief Scientist

Defence R&D Canada – Centre for Operational Research and Analysis (CORA)

Abstract

Real world data mining applications such as Mine Countermeasure Missions (MCM) involve learning from imbalanced data sets, which contain very few instances of the minority classes and many instances of the majority class. For instance, the number of naturally occurring clutter objects (such as rocks) that are detected typically far outweighs the relatively rare event of detecting a mine. In this paper we propose support vector machine with adaptive asymmetric misclassification costs (instances weighted) to solve the skewed vector spaces problem in mine countermeasure missions. Experimental results show that the given algorithm could be used for imbalanced sonar image data sets and makes an improvement in prediction performance.

Résumé

Les applications liées aux données réelles de mines, comme les missions de lutte contre les mines, impliquent d'apprendre à partir d'ensembles de données déséquilibrés qui ne contiennent qu'une poignée de détections des classes mineures et un grand nombre de la classe majeure. Par exemple, la grande majorité du fouillis détecté est généralement d'origine naturelle (comme des roches), les mines étant plutôt rares. Dans le présent rapport, nous proposons une machine de vecteurs de support (*Support Vector Machine — SVM*) qui repose sur les coûts adaptatifs liés aux erreurs de classification asymétrique (occurrences calculées) pour résoudre le problème d'espaces vectoriels asymétriques touchant les missions de lutte contre les mines. Les résultats expérimentaux montrent que l'algorithme pourrait servir aux ensembles de données d'images sonar déséquilibrés et améliorer le rendement des prévisions.

This page intentionally left blank.

Executive summary

Machine Learning Algorithms for Multiple Autonomous Unmanned Vehicle Operations: Using Support Vector Machine with Adaptively Asymmetric Misclassification Costs for Mine-Like Objects Detection

X. Wang; H. Shao; X. Liu; N. Japkowicz; DRDC CORA CR 2013-121; Defence R&D Canada – CORA; August 2013.

Introduction: Autonomous Underwater Vehicles (AUVs) are powerful tools that perform undersea tasks such as underwater-mine countermeasure missions. In these missions, the number of naturally occurring clutter objects (such as rocks, shipwrecks or fishes) that are detected typically far outweighs the relatively rare event of detecting a mine. As a result, the traditional classification approaches that do not account for severe class imbalances often lead to poor classification performance.

The machine learning community has addressed the issue of class imbalance in two different ways in order to solve the skewed vector spaces problem. The first method, which is classifier-independent, is to balance the original dataset. The second way involves modifying the classifiers in order to adapt them to the data sets.

Results: In this paper, we have proposed a Support Vector Machine (SVM) with adaptively asymmetric misclassification costs method for imbalanced problems of classification on MCM datasets. Experimental results show that this method can improve the prediction accuracy of the classification. We have extended Morik et al. [7] and Shawe-Taylor & Cristianini [8]'s algorithm by choosing a more general setting which is to assign each instance of a Mine-Like Object (MLO) a regularization parameter C_i. This method gives us more flexibility than adjusting the cost of false positives vs. false negatives. Our future work will include investigating the relationship of the process of hyper parameter optimization and the performance of classifiers, and will also include hyper parameter optimization on all parameters including the original cost factor c and the kernel parameters.

Sommaire

Machine Learning Algorithms for Multiple Autonomous Unmanned Vehicle Operations: Using Support Vector Machine with Adaptively Asymmetric Misclassification Costs for Mine-Like Objects Detection

X. Wang; H. Shao; X. Liu; N. Japkowicz ; DRDC CORA CR 2013-121 ; R & D pour la défense Canada – CARO; août 2013.

Introduction : Les véhicules sous-marins autonomes (AUV) constituent de puissants outils destinés aux tâches sous-marines, comme des missions de lutte contre les mines. Durant ces missions, la grande majorité du fouillis détecté est généralement d'origine naturelle (roches, épaves, poissons, etc.), les mines étant plutôt rares. C'est pourquoi la méthode de classification classique qui ne tient pas compte des déséquilibres considérables touchant les classes produit souvent de piètres résultats.

Le milieu de l'apprentissage machine a abordé ces déséquilibres de deux façons, afin de résoudre le problème d'espaces vectoriels asymétriques. La première méthode, indépendante du classificateur, consiste à équilibrer l'ensemble de données original et la seconde, à modifier des classificateurs en vue de les adapter aux ensembles de données.

Résultats : Dans le présent rapport, nous avons proposé une machine à vecteurs de support (*Support Vector Machine — SVM*) qui repose sur une méthode adaptative liée aux coûts des erreurs de classification et vise les problèmes déséquilibrés de la classification d'ensembles de données des missions de lutte contre les mines. Les résultats expérimentaux montrent que cette méthode peut améliorer l'exactitude des prévisions de classification. Nous avons étendu l'algorithme de Morik et coll. [7] et Shawe-Taylor et Cristianini [8] en choisissant un réglage plus général qui consiste à attribuer un paramètre de régularisation C_i à chaque détection d'un objet similaire à une mine (*Mine-Line Object — MLO*). Nous disposons ainsi d'une plus grande adaptativité que celle obtenue en réglant le coût de faux positifs par rapport aux faux négatifs. Les recherches futures comprendront l'étude de la relation entre le traitement de l'optimisation hyperparamètre et le rendement de classificateurs, de même que l'optimisation hyperparamètre de tous les paramètres, dont ceux du noyau et du facteur c du coût original.

Table of contents

List of figures

List of tables

This page intentionally left blank.

1 Introduction

A dataset is imbalanced if the classes are not approximately equally represented. In imbalanced data sets, the number of negative examples is often much larger than the number of positive examples. In this situation, a default classifier always predicts “negative”. In practice, one would like to penalize errors on positive examples more strongly than errors on negative examples. There have been attempts to deal with imbalanced data sets in domains such as fraudulent telephone calls, telecommunications management, text classification and images classification, disease detection and other areas. In this paper, we are focusing on Mine Countermeasure Missions (MCM).

Autonomous Underwater Vehicles (AUVs) are powerful tools that perform undersea tasks such as underwater-mine countermeasure missions. In these missions, the number of naturally occurring clutter objects (such as rocks, shipwrecks or fishes) that are detected typically far outweighs the relatively rare event of detecting a mine. As a result, the traditional classification approaches that do not account for severe class imbalances often lead to poor classification performance.

The machine learning community has addressed the issue of class imbalance in two different ways in order to solve the skewed vector spaces problem. The first method, which is classifier-independent, is to balance the original dataset. The second way involves modifying the classifiers in order to adapt them to the data sets.

The simplest way to balance a data set is by under-sampling or over-sampling (randomly or selectively) the majority class while keeping the original population of the minority class [1], [2]. As one of the most popular pre-processing methods to balance a data set, Synthetic Minority Over-sampling Technique (SMOTE) [1] is a technique to over-sample the minority class by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbors. Depending upon the amount of over-sampling required, neighbors from the k nearest neighbors are randomly chosen.

Working with classifiers to adapt data sets could be another way to deal with the imbalanced data problem. Assigning distinct costs to the training examples seems to be the best approach of this kind and various experimental studies of this type have been performed using different kinds of classifiers ([4], [5]). In terms of Support Vector Machine (SVM), several attempts have been made to improve their class prediction accuracy ([6], [7], [8], [9] and [10]). For underwater-mine classification, Williams et al. [3] used infinitely imbalanced logistic regression to solve the imbalanced data problem.

In this paper we extended the work by Morik et al. [7] and Shawe-Taylor & Cristianini [8]. Section II reviews the algorithms of [7] and [8]. In Section III, we introduce and discuss our new algorithm in detail. In Section IV, we provide the details for image processing in our particular domain of application. After these developments, we present an experimental section (Section V) that illustrates the efficiency of our algorithm and we make some observations about its performance. Section VI is the conclusion followed by an acknowledgment and references.

2 Preliminaries

Shawe-Taylor & Cristianini [8] show that the distance of a test point from the boundary is related to its probability of misclassification. This observation motivated a related technique which is used in their paper. The technique is to provide a more severe penalty if an error is made on a positive example than if it is made on a negative example. By using the cost factors and adjusting the cost of false positives and false negatives, such penalties can be directly incorporated into the SVM algorithm.

Morik et al. [7] and Shawe-Taylor & Cristianini [8] proposed an algorithm to use the L1 norm (k=1). Two cost-factors are chosen so that the potential total cost of the false positives equals the potential total cost of the false negatives. This means that the parameters of the SVM are selected such that they obey the ratio: $\frac{C^+}{C^-} = \frac{N^- C^+}{N^+ C^-} = \frac{N^-}{N^+}$ where C^+ and C^- are cost factors to adjust the cost of false positive vs. false negatives and N^+ and N^- are the number of positive instances and negative instances respectively. By increasing the margin on the side of the smaller class, this method provides a way to induce a decision boundary which is much more distant from the "critical" class than it is from the other. But in this model, the balance between sensitivity and specificity cannot be controlled adaptively resulting in over-fitting.

This can be achieved using cost factors C^+ and C^- to adjust the cost of false positives vs. false negatives. Such cost factors can be directly incorporated into the SVM. Finding this hyper-plane can be translated into the following optimization problem:

$$\text{minimise}_{\xi,\omega,b} \|\omega\|^2 + C^+ \sum_{i:y_i=1}^{l} \xi_i + C^- \sum_{i:y_i=-1}^{l} \xi_i \qquad (1)$$

$$\text{Subject to} \quad y_i(\omega^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \ldots, l$$

where x_i is the feature vector of example i. y_i equals +1(-1), if example I is in class +(-).

The corresponding Lagrangian for the 1-norm soft margin optimization problem is:

$$
\begin{aligned}
L(\omega, b, \xi, \alpha, \mu) = \frac{1}{2}\|\omega\|^2 + C^+ \sum_{i:y_i=1}^{l} \xi_i + C^- \sum_{i:y_i=-1}^{l} \xi_i \\
- \sum_{i=1}^{l} \alpha_i [y_i(\omega^T \phi(x_i) + b) + \xi_i] - \sum_{i=1}^{l} \mu_i \xi_i
\end{aligned} \tag{2}
$$

where $\alpha_i \geq 0$ and $\mu_i \geq 0$.

The dual problem of problem (1) is

$$
\begin{aligned}
\max_{\alpha} L(\omega, b, \xi, \alpha, \mu) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j k(x^i, x^j) \\
\text{Subject to} \quad 0 \leq \alpha_i \leq C_+, \text{ if } y_i = 1, \\
0 \leq \alpha_i \leq C_-, \text{ if } y_i = -1, \\
\textstyle\sum_{i=1}^{l} y_i \alpha_i = 0.
\end{aligned} \tag{3}
$$

Instead of using the L1 norm for the loss measure, Veropoulos et al. [9] use the square of the L2 norm (k=2). This method enables the algorithm to control the balance between sensitivity and specificity, not adding any information. Experimental results (Veropoulos et al. [9]) show that this method has the power to effectively control the sensitivity and not the specificity of the learning machine.

3 Proposed Approach

The motivation of our approach comes from Morik et al. [7] and Shawe-Taylor & Cristianini [8]'s algorithm. Instead of using cost factors $C^{+}$and C^{-} to adjust the cost of false positive vs. false negatives, we choose a more general setting which is to assign each instance a regularization parameter C_i. These parameters give us more flexibility for adjusting the cost of false positives vs. false negatives. The function of finding the hyper-plane (1) can be translated into the following optimization problem:

$$\text{minimise}_{\xi,\omega,b}\|\omega\|^2 + \sum_{i:y_i=1}^{l} C_i\xi_i + \sum_{j:y_j=-1}^{l} C_j\xi_j \tag{4}$$
$$\text{Subject to} \quad y_i(\omega^T\phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \ldots, l$$

The corresponding Lagrangian for the 1-norm soft margin optimization problem is:

$$\begin{aligned} L(\omega, b, \xi, \alpha, \mu) = \frac{1}{2}\|\omega\|^2 + \sum_{i:y_i=1}^{l} C_i\xi_i + \sum_{j:y_j=-1}^{l} C_j\xi_j - \sum_{i=1}^{l} \alpha_i[y_i(\omega^T\phi(x_i) + b) + \xi_i] \\ - \sum_{i=1}^{l} \mu_i\xi_i \end{aligned} \tag{5}$$

where $\alpha_i \geq 0$ and $\mu_i \geq 0$.

The dual problem of problem (4) is

$$\begin{aligned} &\max_{\alpha} L(\omega, b, \xi, \alpha, \mu) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j k(x^i, x^j) \\ &\text{Subject to} \quad 0 \leq \alpha_i \leq C_i, \text{ if } y_i = 1, \\ &\qquad\qquad\quad 0 \leq \alpha_j \leq C_j, \text{ if } y_j = -1, \\ &\qquad\qquad\quad \textstyle\sum_{i=1}^{l} y_i\alpha_i = 0. \end{aligned} \tag{6}$$

In this function, each instance is assigned a regularization parameter C_i. Our target is to choose the best set of parameters C_i so that the classifier can accurately predict unknown data (i.e. testing data). It is not known beforehand what kinds of C_i's are best for a given problem. In our approach, we use a cross-validation strategy to identify the best parameter combination. In k-fold cross-validation, we first divide the training set into k subsets of equal size. One subset is tested sequentially using the classifier trained on the remaining subsets. Therefore each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified.

Since in learning extremely imbalanced data, a trivial classifier that predicts every case as the majority class can still achieve very high accuracy, the overall classification accuracy is often not an appropriate measure of performance. We choose G-mean [2] as the measure for our algorithm and experiment. The definition of G-mean is listed below.

Table 1: Confusion matrix

	Predicted Positive Class	Predicted Negative Class
Actual Positive class	TP (True Positive)	FN (False Negative)
Actual Negative class	FP (False Positive)	TN (True Negative)

$$\text{Specificity: true Negative Rate } Acc^- = \frac{TN}{TN+FP}$$
$$\text{Sensitivity: true Positive Rate } Acc^+ = \frac{TP}{TP+FN}$$
$$\text{G-mean}=(Acc^- \times Acc^+)^{1/2}$$

Figure 1 shows the separating hyper plane and the margin of this classification problem where each instance is assigned a regularization parameter C_i. Each parameter of the point labeled ξ_i can affect the hyper plan and margin. The points labeled ξ_i are said to be on the wrong side of their margin with the cost factor C_i.

In [12], Hsu et al. provide grid search through a subset of the hyper parameter space of a learning algorithm to solve the problem of hyper parameter optimization. In [13], Bergstra et al. compare different strategies and prove that random search is generally better than grid search for hyper-parameter optimization. In our algorithm, we choose a mixture of grid-search and random search as our strategy for hyper-parameter (C_i) optimization. Here we define the biased cost factor C as $C = c \times C'$ (c is the original cost and C' is the cost weight). Experientially the original cost c is set to the imbalance ratio of the minority instance number to the majority instance number [9] [11]. Therefore we just need to optimize the cost weight C' to get the optimal cost factor C.

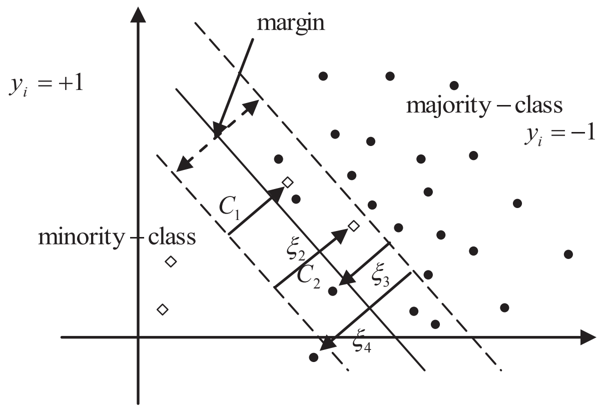

Figure 1: The non-separable classification problem together with the separating hyper plane and the margin.

If we optimize all cost weight C'_i, the cost will be too high and time wasting. Since from (4) we can see that adjusting the cost factor can affect both the positive margin and the negative margin, we just choose to optimize the minority class cost factors (C'^{+}_i). This strategy can highly reduce the cost of hyper-parameter (C'_i) optimization.

- Firstly, we use grid-search with G-mean as the measurement to optimize the cost factors weight of false positive and false negatives $C'^{+}$and C'^{-}. The purpose of this step is to reduce the optimization range of cost factors C'_i.

- Secondly, in each optimization round, we randomly choose n instances from the minority class ($0 \leq n \leq l^{+}$) where l^{+} is the number of instances in the minority class and keep the parameters of all the other instances (including both the minority class instances left and all the majority class instances) without any change. We choose a range of C'^{+}_i and value at each optimization step. Using the random search method to get the best parameters of the chosen instances within the threshold of optimization steps.

- Finally we use the classifier with optimized parameters on the testing data and get the prediction accuracy on all instances.

4 Data Processing

In MCMs, sonar images collected by AUVs will convey important information about the underwater conditions. How to properly process the sonar images will have a huge impact on the following MLOs detection and classification stage.

In the MCMs, a large part of sonar images collected by AUVs represent the background—seabed. In MLOs detection and classification, we are more interested in the object that lies on the seabed rather than the background. The areas from the images with only background information can be simply discarded. Image segmentation is a widely used image processing technique to detect target objects and segment the original images into small pieces that contain the target objects. The foreground objects are assumed to have a more complex texture than the seabed. Thus, the foreground object areas are obtained by using local range and standard deviation filters.

Instead of dealing with the whole sonar image, image segmentation allows us to only process the smaller pieces, reducing the future computational load. In this step, our goal is to delete image data that contain only background information and reduce the amount of data to be processed. Therefore whether the size, shape and location of the target object are accurately found is not a main concern in this step.

The objective of the image processing procedures at this point is data reduction rather than MLOs detection. Thus, a relatively high false alarm rate is acceptable.

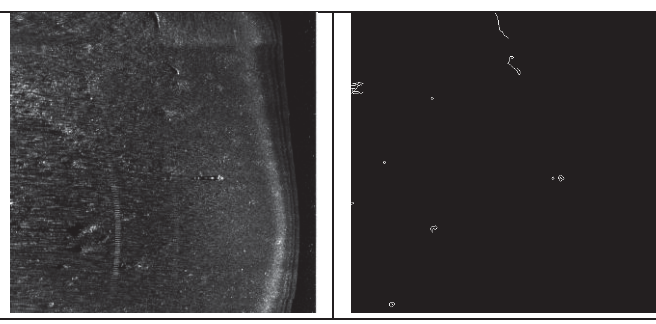

Figure 2: Example of an image processing result on an image provided by the Ocean System Lab, Heriot-Watt University

Figure 2 illustrates the extraction of foreground objects from a sonar image which was provided by the Ocean Systems Lab, Heriot-Watt University. Areas that do not have a reasonable size will be ignored.

For object detection tasks, an object should be detected through a single view, no matter where and how it lies on the seabed. Therefore, the feature used should be robust to the location and orientation of the object. The grayscale histogram, a simple but informative statistical feature, is considered. In many image recognition systems, many complex features are used, but such features will inevitably increase the computational complexity, impeding the real time detection. The histogram is easy to calculate and robust to rotation. The distribution of the grayscale value can be well described by this feature.

In our experiment, the grayscale value (0-255) is divided into 16 bins with width 16. The grayscale histograms are normalized to the frequency that a pixel value falls into each bin. The MLOs are labeled as the positive examples.

5 Experimental Results

The data sets used in our experiment were collected by an AUV fitted with side scan sonar from a trail on Loch Earn (Scotland) on November 10th and November 11th, 2010. The images were provided by the Ocean Systems Lab, Heriot-Watt University.

Table 2: Dataset details

Name	Data0	Data1	Data2
Number of features	16	16	16
Number of total instances	3367	1205	2192
Number of positive instances	35	105	142
Number of negative instances	3332	1100	2050
Imbalance ratio	1:95.2	1:10.5	1:14.4

Table 2 gives the details of the target data sets. The imbalance ratios of these three data sets are 95.2, 10.5 and 14.4 respectively. In our experiment, we choose a 5 fold validation strategy. Each data set is divided equally into five subsets. Each sub data set is a testing dataset while the left four sub data sets build the training data sets. For each training dataset, we choose 10 fold validation strategy as the method for hyper-parameter ($C_t'^+$, $t \subset l$) optimization. This means that each training data set is divided into 10 equal sub data sets, each data set is a validation dataset and the left 9 sub data sets build the primary training dataset. Here the choice of a fold validation number is not optimized but choosing large fold validation numbers may cause the minority instance numbers of the validation dataset to be too small to get enough information for optimization, so 5-10 fold validation strategies are reasonable. Using this strategy, we employ the proposed algorithm provided in Section III and get the optimized hyper-parameters ($C_t'^+$) for each training dataset. Then we use these optimized hyper-parameters ($C_t'^+$) on the training data sets and get the optimized classifiers, in the last step we apply these classifiers on the testing data sets and get the prediction accuracy (G-mean).

We choose the LibSVM [11] [12] as the software for support vector machine classification. Modifications were made to the original source code based on the LibSVM license [11]. In our experiment we chose a Gaussian kernel with parameters c=1 and γ =1/num_features since the Gaussian kernel is the most popular kernel function. To compare our proposed algorithm with other benchmark algorithms, the parameter c and γ will be set to the same values in every experiment.

Table 3: Experiment result on Data0 (1, γ = 1/num_features)

	TP	FN	FP	TN	Acc+	Acc-	Gmean
SMOTE	3135	197	4	31	0.941	0.886	0.9129
SMOTE+US	2839	493	3	32	0.852	0.914	0.8826
BP_SVM	3136	196	5	30	0.941	0.857	0.8982
AAC_SVM	3185	147	1	34	0.956	0.971	0.9636

Table 4: Experiment result on Data1 (kernel: Gaussian, c=1, γ = 1/num_features)

	TP	FN	FP	TN	Acc+	Acc-	Gmean
SMOTE	830	270	32	73	0.754	0.695	0.7243
SMOTE+US	772	328	31	74	0.702	0.705	0.7033
BP_SVM	830	270	32	73	0.754	0.695	0.7243
AAC_SVM	926	174	28	77	0.842	0.733	0.7857

Table 5: Experiment result on data2 (kernel: Gaussian, c=1, γ = 1/num_features)

	TP	FN	FP	TN	Acc+	Acc-	Gmean
SMOTE	1688	362	29	113	0.823	0.796	0.8095
SMOTE+US	1627	423	33	109	0.794	0.768	0.7805
BP_SVM	1695	355	30	112	0.827	0.789	0.8076
AAC_SVM	1930	120	14	128	0.942	0.901	0.9212

We compare our method with other popular methods: SMOTE, SMOTE with under-sampling (denotes as SMOTE+US), Morik et al. [7] and Shawe-Taylor & Cristianini [8]'s algorithm (biased penalty on SVM which is denotes as BP_SVM). Our adaptively asymmetric misclassification costs method is denoted as AAC_SVM.

Table 3 to Table 5 give the experimental results and from these results, we can find that our method can improve prediction accuracy (G-mean) on both of specificity and sensitivity on these given data sets. These experimental results show that our method gets the benefits of adjusting parameters on each minority instance. SMOTE can introduce some potential information to the classifier but it also creates noise while our method does not introduce any noise. BP_SVM is a special case of our method that makes all costs of instances in the same class the same value. Our method extends the ability of BP_SVM by adjusting much more parameters.

Obviously, one shortcoming of our method is that the internal cross-validation and optimization leads to higher computational complexity compared with BP_SVM. Since introducing weights on instances adaptively may cause over-fitting when the number of minority instances is very low, combining our method with ensemble methods may be a good way to go, although the computational complexity could be even higher.

6 Conclusion and Future Work

In this paper, we have proposed a SVM with adaptively asymmetric misclassification costs method for imbalanced problems of classification on MCM datasets. Experimental results show that this method can improve the prediction accuracy of the classification. We have extended Morik et al. [7] and Shawe-Taylor & Cristianini [8]'s algorithm by choosing a more general setting which is to assign each instance a regularization parameter C_i. This method gives us more flexibility than adjusting the cost of false positives vs. false negatives. Our future work will include investigating the relationship of the process of hyper parameter optimization and the performance of classifiers, and will also include hyper parameter optimization on all parameters including the original cost factor c and the kernel parameters.

This page intentionally left blank.

References

[1] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16: 321-357, 2002.

[2] Kubat, M., & Matwin, S. (1997). Addressing the curse of imbalanced training sets: One-sided selection. Proceddings of the Fourteenth International Conference on Machine Learning, 179-186.

[3] D. Williams, V. Myers, and M. Silvious, "Mine Classification with Imbalanced Data," IEEE Geoscience and Remote Sensing Letters, Vol. 6, No. 3, pp. 528-532, July 2009.

[4] Chen C., Liaw, A., and Breiman, L. (2004). Using random forest to learn unbalanced data. Technical Report 666, Statistics Department, University of California at Berkeley.

[5] H. Guo and HL Viktor (2004). Learning from Imbalanced Data Sets with Boosting and Data Generation: The DataBoost-IM Approach, ACM SIGKDD Explorations, 6(1), 30-39.

[6] Akbani, R., Kwek, S., and Japkowicz, N. (2004), Applying Support Vector Machines to Imbalanced Datasets,in the Proceedings of the 2004 European Conference on Machine Learning (ECML'2004).

[7] Morik, K., Brockhausen, P., Joachims, T. (1999). Combining Statistical Learning with a Knowledge-Based Approach - A Case Study in Intensive Care Monitoring. ICML: 268-277

[8] Shawe-Taylor, J. and Cristianini, N. (1999) Further results on the margin distribution. In Proceedings of the 12th Conference on Computational Learning Theory, 278-285

[9] Veropoulos, K., Campbell, C., & Cristianini, N. (1999). Controlling the sensitivity of support vector machines. Proceedings of the International Joint Conference on Artificial Intelligence, 55–60.

[10] Wang, B. X. and Japkowicz, N., Boosting Support Vector Machines for Imbalanced Data Sets, in the Proceedings of the 17th International Symposium on Methodologies for Intelligent Systems (ISMIS'2008) .

[11] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1-27:27, 2011.

[12] Chin-Wei Hsu, Chih-Chung Chang and Chih-Jen Lin (2010). A practical guide to support vector classification. Technical Report, National Taiwan University.

[13] Bergstra, James; Bengio, Yoshua (2012). "Random Search for Hyper-Parameter Optimization". J. Machine Learning Research 13: 281-305.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.)

Nathalie Japkowicz Consulting Services
5714 Queen Mary Rd
Hampstead, Quebec
H3X 1X6

2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.)

UNCLASSIFIED
(Non-Controlled Goods)
DMC: A
Review: GCEC June 2010

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Machine Learning Algorithms for Multiple Autonomous Unmanned Vehicle Operations: Using Support Vector Machine with Adaptively Asymmetric Misclassification Costs for Mine-Like Objects Detection

4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)

Wang, X.; Shao, H.; Liu, X.; Japkowicz, N.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.)	6b. NO. OF REFS (Total cited in document.)
August 2013	28	13

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Contract Report

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence R&D Canada – CORA
101 Colonel By Drive
Ottawa, Ontario K1A 0K2

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
10bz04	W7714-115078

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
[if used]	DRDC CORA CR 2013-121

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

Autonomous Underwater Vehicles (AUVs) are planned to conduct Mine Countermeasure mission in the future. With the help of high resolution imagery produced by the sonar systems mounted on AUVs, the mines and other objects of interest can be detected. In this work, the existing approaches for Mine-Like Objects (MLOs) detection are first reviewed, then considering the limitation of the exiting works, a novel machine learning method is designed for MLOs detection. The experimental result on real side scan sonar images shows that the new learning method is able to result in reliable and fast MLOs detection.

L’utilisation de véhicules sous-marins autonomes (VSA) est prévue pour de futures missions de lutte contre les mines. En effet, des mines et d’autres objets présentant un intérêt peuvent être détectés à l’aide des images à haute résolution produites par les systèmes de sonar installés sur des VSA. Dans le cadre des présents travaux, les approches existantes en matière de détection d’objets ressemblant à une mine sont d’abord examinées, puis, en tenant compte des limites des travaux existants, une nouvelle méthode d’apprentissage automatique est conçue pour la détection de tels objets. Le résultat expérimental sur de vraies images de sonar latéral démontre que la nouvelle méthode d’apprentissage peut permettre une détection fiable et rapide des objets ressemblant à une mine.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

AUV; Unmanned Vehicles; Autonomous Vehicles; Machine Learning; Imbalanced data sets; Support vector machines; Adaptive Asymmetric Misclassification cost; G-mean

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

DEFENCE DÉFENSE

www.drdc-rddc.gc.ca