



SVM Based EVM for Open Space Problems

Yasuyuki Kaneko^(✉) and Mineichi Kudo

Hokkaido University, Sapporo 060-814, Japan
{yasuyuki,mine}@ist.hokudai.ac.jp

Abstract. In this paper, we consider a multi-class classifier for problems where unknown classes are included in testing phase. Previous classifiers consider the “closed-set” case where the classes used for training and the classes used for testing are the same. A more realistic case is the “open-set” recognition, in which only a limited number of classes appear at training time, and unknown classes appear during testing. To handle such problems, we need classifiers that accurately classify data belonging to not only known classes but also unknown classes. In this paper, We introduce a Support Vector Machine (SVM) based Extreme Value Machine (EVM) to determine a compact class region. Any data outside of such class regions is rejected as being in unknown classes. To construct a class region, we approach the class decision boundary found by SVM towards the samples, by removing some support vectors close to the boundary. This SVM based EVM resolves the three problems that EVM possesses: unfair size of class regions, excessive sensibility to certain points and fragmentation of a class region.

Keywords: Extreme value theory · Support vector machines · Open set recognition · Anomal detection

1 Introduction

A lot of high performance classifiers such as deep neural networks have been developed so far. These classifiers work most under a *closed condition* where classes to appear are known already in training time. A more realistic scenario is an *open condition* where unknown classes appear in testing time. Such a situation is also called Open Set Recognition (OSR) [10]. Examples are rare disease diagnosis and web application services where newly discovered diseases and newly released applications appear day by day.

A pioneer work in OSR was made by Scheirer *et al.* [10] who formalized OSR and proposed a 1-vs-set machine, using a linear SVM, by considering an open space risk in addition to an empirical risk. Although 1-vs-set machine reduces the region of a known class compared with that of the original SVM, the region is

still unbounded because of the linearity. The difficulty of OSR is that classifiers have to deal with samples belonging to unknown classes, that is, they have to be trained by no sample of the classes. There are some approaches to cope with this difficulty such as one-class SVM [11] and classifiers with a reject option [1, 3, 4, 13]. Among them, Extreme Value Machine (EVM) [8] is one of the most promising ones. EVM determines a class region as a set of hyper-balls centered at each of the samples. The radius of a ball is determined by Extreme Value Theorem (EVT) [6] applied to the nearest samples belong to the other classes. The region is now bounded but still suffers from several problems that will be introduced later.

2 Related Work

2.1 Algorithm for Open Space Recognition

Scheirer *et al.* proposed Weibull-calibrated SVM (WSVM) [9] for OSR. WSVM introduces a Compact Abating Probability (CAP) model that guarantees that the probability of samples becomes zero if they are away at a certain distance from any training sample of a class.

Junior *et al.* proposed Open Space Nearest Neighbor (OSNN) [7]. In OSNN, for an input sample s , we find the nearest two samples t, u belonging to different classes, if ratio $d(s, t)/d(s, u) > T$ holds for a threshold T , then the input sample s is assigned to an unknown class, otherwise recognized as the class of t or u .

2.2 EVT Based Algorithm

We explain the outline of EVM that we put our basis on. In EVM [8], we select one sample x_1 from the positive class, a *positive* class of interest (Fig. 1(a)). Then, as statistics, we consider the half distance m 's from x_1 to all the negative samples (Fig. 1(b)(c)). By multiplying -1 to these distances, we consider the max value. To estimate a distribution (EVD) of the extreme value of m 's, we collect τ maximum values (Fig. 1(d)). According to EVT [6], we use a Weibull distribution as the extreme value distribution because the values are upper-bounded by zero (Fig. 1(e)). With a parameter δ as a percentile (Fig. 1(e)), we determine the radius m_{ex} of the ball centered at x_1 (Fig. 1(f)). The ball centered at x_1 shows a local domain of x_1 (Fig. 1(e)). Collecting these local balls over all positive samples, we have a positive class region.

Applying this procedure to all classes in turn, we have their class regions. Classification is made by whether a test sample falls into one of the class regions (to be assigned to a known class) or not (to reject). EVM is a kernel-free nonlinear classification.

Unfortunately, EVM suffers from three problems (Fig. 2). First, a class can have a larger region than those of the other classes when it is apart from the other classes (*the problem of unfair size of class regions*). Second, a samples far

from the other samples of the same class can dominate the class region) (*the problem of excessive sensibility of individual samples*). Third, a class region can be divided into small connected regions (*the problem of fragmentation of a class region*). These problems are due to the independency of radii of balls and the isotropy by a ball.

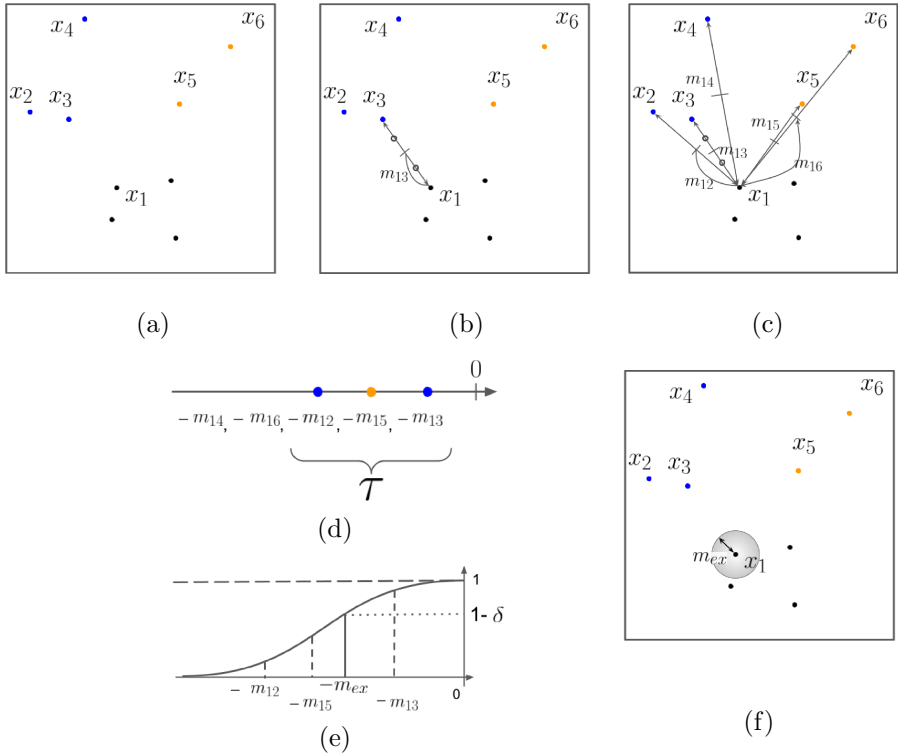


Fig. 1. Working flow of EVM. We consider to enclose a positive sample x_1 by a ball as a component of the class region. There are three classes (black, yellow and blue) in Fig. 1(a). We calculate distances between x_1 and all negative samples (x_2 to x_6) (samples of blue and orange), e.g., $m_{13} = \|x_1 - x_3\|/2$. To find the max value of m 's, we multiply -1 and sort these values (Fig. 1(d)). We estimate an extreme value distribution by the maximum τ values, and obtain m_{ex} with a percentile δ (Fig. 1(e)). Finally, we build a ball around x_1 with radius of m_{ex} (Fig. 1(f)). (Color figure online)

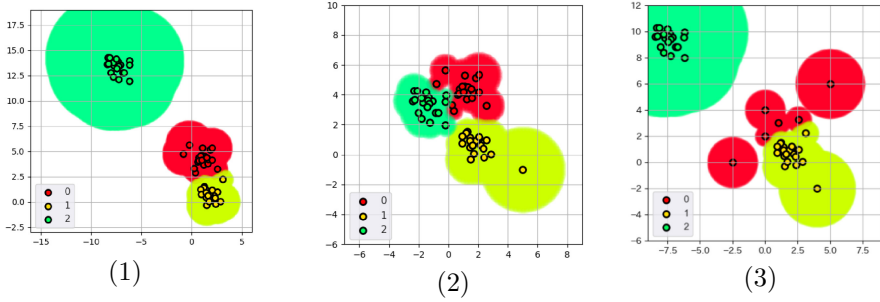


Fig. 2. The three problems in EVM [8]: (1) Some classes can have exceptionally larger regions than the other classes (the green region compared with red or yellow region) (**Unfair region size problem**), (2) An anomaly sample can affect much on the class region (a yellow sample at (5, -1)) (**Excessive sensibility problem**), and (3) A class region can be fragmented (red class) (**Fragmentation problem**) (Color figure online)

3 SVM Based EVM

To cope with these three problems possessed by EVM, we propose a way to determine a class region by non-linear SVM. With a kernel trick, we are able to have a more flexible region than those of hyper-planes (realized by 1-vs-set) or hyper-balls (realized by EVM). We call it *SVM based EVM* and denote by SVM-EVM. The class regions learned by SVM-EVM are in general narrower than those by EVM, although the size is controllable by a parameter δ .

3.1 Algorithm of SVM-EVM

First, taking a linear SVM as an example, we explain our idea of SVM-EVM. We first construct a decision boundary $\{\mathbf{x} \mid \omega^T \mathbf{x} + \omega_0 = 0\}$ by a linear SVM (Fig. 3(a)). Then, the margin m_0 is obtained as

$$m_0 = \frac{1}{\sqrt{\sum_{i \in \text{SV}} \lambda_i}}, \tag{1}$$

where λ_i is Lagrange multipliers determined by the formulation of SVM (for example, see [2]), and SV is the set of support vectors. As a next step, we estimate an Extreme Value Distribution (EVD) of $-m_0$ (the minimum distance of the positive samples to the decision). To collect another candidate extreme value, we remove one of the *positive* support vectors and reconstruct another SVM to have the second margin m_1 (Fig. 3(b)). We repeat this procedure according to a deletion ordering of samples (Fig. 4) to obtain a necessary number τ of margins. All these margins multiplied by -1 are dealt as extreme values for estimation of an EVD. We can think of Weibull distributions for because the values of $-m$'s is upper-bounded by zero. The parameter estimation is made

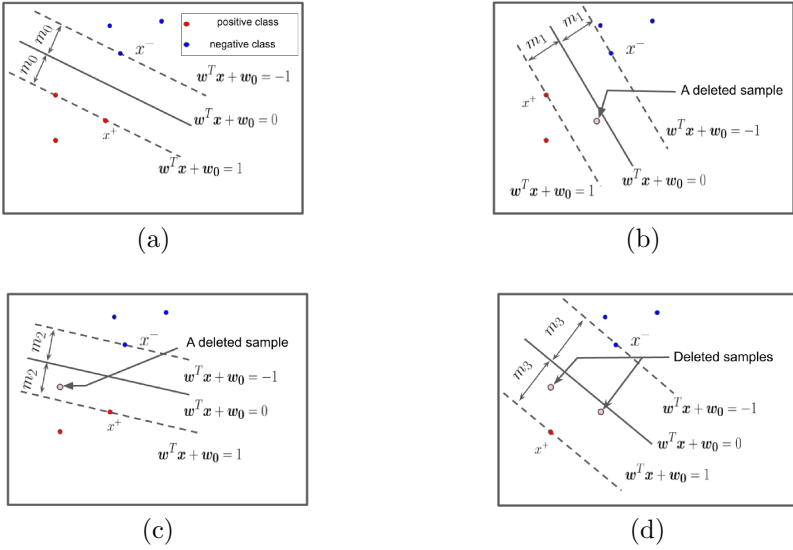


Fig. 3. Procedure for deleting support vectors (SVs) to obtain an extreme value and semi-extreme values. Positive class SVs are deleted one by one. First we find a decision boundary with margin m_0 by linear-SVM in (a). In (b), we delete a positive SV and construct another linear-SVM to find a new margin m_1 , then m_2, m_3, \dots . We continue this procedure until τ m 's are obtained. The removing order is shown in Fig. 4.

by Maximum Likelihood Estimate implemented in SciPy [14]. Last, with a user-specified percentile δ , we determine the values of m_{ex} . With this m_{ex} , we define the class region as the positive region:

$$R = \left\{ \mathbf{x} \mid \sum_{i \in SV} \lambda_i y_i K(\mathbf{x}_i, \mathbf{x}) + \mathbf{w}_0 - \frac{m_{ex}}{m_0} \geq 0 \right\}, \quad (2)$$

where $K(\mathbf{x}_i, \mathbf{x})$ is an RBF kernel. In our method, we set τ to 1% or less of total number of samples.

3.2 An Achievement of SVM-EVM

A simple experiment was conducted. In a two-dimension space, we considered three classes of 20 samples each. In Fig. 5, the class regions by SVM-EVM with $\delta \in \{0.0, 0.01, 0.5\}$ are shown.

3.3 Solving Three Problems

Here, we show some examples to demonstrate that the proposed SVM-EVM resolves three problems of EVM (Fig. 6). The proposed SVM-EVM obtains the class regions surrounded by nonlinear functions that are determined by a whole set of the training samples.

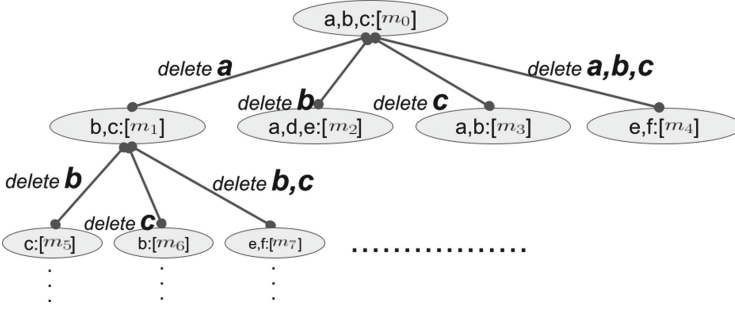


Fig. 4. Removal ordering of positive support vectors. By traversing this tree in width-first search, we determine the next sample to remove.

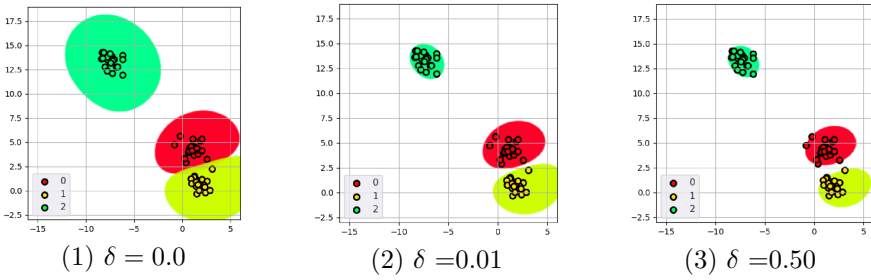


Fig. 5. Class regions obtained by SVM-EVM for $\delta = 0.0, 0.5$ and 0.01 . The case of $\delta = 0.0$ is equal to SVM.

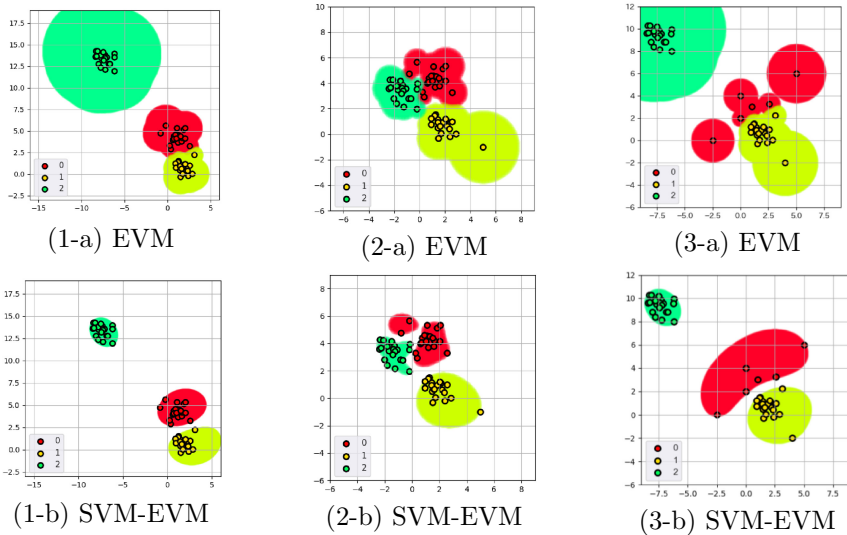


Fig. 6. Solutions by SVM-EVM to three problems in EVM (Fig. 2). From top to down, (1) **unfair region size problem**, (2) **Excessive sensibility problem** and (3) **Fragmentation problem**.

4 Experiments

An experiment was conducted to confirm the effectiveness of the proposed method. We used OLETTER dataset which is generated for an open space recognition problem and is made by modifying Letter dataset [5]. This data is of 20 000 black and white images of $N = 26$ capital letters in 16 different styles (Fig. 7).

4.1 Experimental Procedure

We carried out the following, referring to [7].

1. Choose $n \in \{3, 6, 9, 12\}$ known classes randomly from $N = 26$ classes and leave $N - n$ classes as unknown classes.
2. Choose randomly a half of all training samples from known classes to make a training set S_{tr_known} . By collecting the remaining samples of known classes we make a set S_{te_known} . With the set $S_{unknown}$ of samples of $N - n$ unknown classes, we make a test set S_{te_both} by $S_{te_both} = S_{te_known} \cup S_{unknown}$.
3. Apply EVM or SVM-EVM: training with S_{tr_known} and testing with S_{te_both} .
4. Repeat 10 times Steps 1 to 3.

We compared the proposed SVM-EVM with EVM. The RBF kernel $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ is used in SVM-EVM and the kernel parameter γ is chosen from $\{0.10, 0.11, \dots, 10.00\}$. We set the restriction parameter δ to 0.5 to determine m_{ex} as shown in Fig. 1(e). The larger δ is, the smaller the area is as shown in Fig. 5. The number τ of candidate extreme values was set to 2. This is less than 1% of the number of training samples in each class. In addition, SVM was forcibly set so as to have a hard margin. The parameters τ and δ of EVM, we used the author's setting [8].

As a metric for evaluation, we used “micro” f -measure denoted by f_μ [12] and defined by,

$$f_\mu = \frac{2 \times \text{precision}_\mu \times \text{recall}_\mu}{\text{precision}_\mu + \text{recall}_\mu}, \quad (3)$$



Fig. 7. Different font styles in OLETTER (10 styles in one letter)

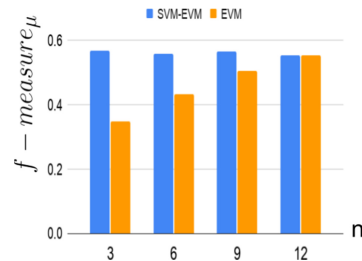


Fig. 8. Comparison between EVM and SVM-EVM in f -measure $_\mu$. The horizontal axis shows the number n of known classes. The vertical axis is the average f_μ value in 10 times.

where

$$precision_{\mu} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FP_i)}, \quad recall_{\mu} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FN_i)}, \quad (4)$$

where FP_i and FN_i are ‘‘False Positive’’ and ‘‘False Negative’’ when class i is regarded as the positive class and the other classes including unknown classes as the negative class as a whole. The micro f-measure is the accuracy on known classes taking into consideration unknown classes (note that FP_i or FN_i ($i=1, 2, \dots, n$) includes the samples from or to unknown classes ($n + 1$ th class)).

4.2 Result

The result is shown in Fig. 8. SVM-EVM is better than EVM regardless of the number n of known classes. It is more advantageous when n is small. We also show their confusion matrices in Table 1. These matrices are for when letters ‘L’, ‘M’ and ‘K’ are known classes ($n = 3$). The parameters were chosen in such a way that the same degree of correct prediction is made for samples of unknown classes: $\tau = 75$ (the recommended value in [8]), $\delta = 1.0 - 1.0^{-14}$ for EVM, and $\tau = 2, \delta = 0.5$ for SVM-EVM. The values of δ and γ are chosen so as to attain the best *f-measure* value. A large difference of δ between EVM and SVM-EVM comes from the difference of distances: the Euclidean distance in the former, while a distance in a reproducing kernel space in the latter. From this comparison, we see that the class regions found by SVM-EVM are more appropriate than those of EVM. We also examined the sensitivity of δ . As a result, it was revealed that δ is insensitive to the result, so that the authors recommend to use $\delta = 0.5$ in general.

Table 1. Confusion matrices of EVM and SVM-EVM for three known classes of ‘L’, ‘M’, ‘X’ ($n = 3$), and other 23 unknown classes. The parameters for EVM are chosen so as to show a comparable performance on unknown classes.

		EVM						SVM-EVM			
		Predicted						Predicted			
Actual		”L”	”M”	”X”	Unknown	Actual		”L”	”M”	”X”	Unknown
”L”		46	0	0	350	”L”		125	0	0	271
”M”		0	73	0	319	”M”		0	153	0	239
”X”		0	0	39	354	”X”		0	0	153	240
Unknown	0	0	0	15	17624	Unknown	0	2	8	8	17629

5 Discussion

SVM-EVM is superior to 1-vs-set machine in the shape of class regions, because a class region by the latter is a half-space, while that by the former has a non-linearly enclosed shape. SVM-EVM is superior to EVM in the treatment of data

because a class region by the latter is a collection of local regions associated to individual samples, while that by the former is a single region associated to all the samples. The reason why SVM-EVM does not depend on the value of the percentile δ so much is because the estimated EVD on the minus margins is so steep around a point, that the margin specified by a percentile δ does not change even if δ has changed. The steepness means that many samples are the support vectors due to the nonlinearity of SVM with RBF kernels.

6 Conclusion

We have presented a novel algorithm for open set classification. This algorithm, called SVM-EVM, determines a class region by a decision boundary generated by SVM but in such a way that the boundary is closer to the samples of the positive class than the original boundary. The extreme value theory is used to determine the degree to what the boundary is close to the samples. SVM-EVM has solved three problems that EVM held, and showed a better performance in an experiment. We will investigate more datasets to confirm the effectiveness of SVM-EVM.

Acknowledgment. This work was partially supported by JSPS KAKENHI Grant Number 19H04128.

References

1. Bartlett, P.L., Wegkamp, M.H.: Classification with a reject option using a hinge loss. *J. Mach. Learn. Res.* **9**(59), 1823–1840 (2008)
2. Bishop, C.M.: *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, Heidelberg (2006)
3. Chow, C.: On optimum recognition error and reject tradeoff. *IEEE Inf. Theory* **16**(1), 41–46 (1970)
4. Fischer, L., Hammer, B., Wersing, H.: Optimal local rejection for classifiers. *Neurocomputing* **214**, 445–457 (2016)
5. Frey, P.W.: Letter recognition using Holland-style adaptive classifiers. *Mach. Learn.* **6**, 161–182 (1991). <https://doi.org/10.1007/BF00114162>
6. Kotz, S., Nadarajah, S.: *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London (2000)
7. Mendes Júnior, P.R., et al.: Nearest neighbors distance ratio open-set classifier. *Mach. Learn.* **106**(3), 359–386 (2017)
8. Rudd, E.M., Jain, L.P., Scheirer, W.J., Boulton, T.E.: The extreme value machine. *IEEE Pattern Anal. Mach. Intell.* **40**(3), 762–768 (2018)
9. Scheirer, W.J., Jain, L.P., Boulton, T.E.: Probability models for open set recognition. *IEEE Pattern Anal. Mach. Intell.* **36**(11), 2317–2324 (2014)
10. Scheirer, W.J., de Rezende Rocha, A., Sapkota, A., Boulton, T.E.: Toward open set recognition. *IEEE Pattern Anal. Mach. Intell.* **35**(7), 1757–1772 (2012)
11. Schölkopf, B., Platt, J.C., Shawe-Taylor, J.C., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. *Neural Comput.* **13**(7), 1443–1471 (2001)

12. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **45**(4), 427–437 (2009)
13. Tax, D., Duin, R.: Growing a multi-class classifier with a reject option. *Pattern Recogn. Lett.* **29**(10), 1565–1570 (2008)
14. Virtanen, P., Gommers, R., et al.: SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020)