

Binary Classification using Pairs of Minimum Spanning Trees or N-ary Trees

Riccardo La Grassa¹[0000-0002-4355-0366], Ignazio Gallo¹[0000-0002-7076-8328],
Alessandro Calefati¹[0000-0003-3860-4785], and Dimitri
Ognibene²[0000-0002-9454-680X]

¹ University of Insubria, Varese, Italy

² University of Essex, Colchester, Essex

Abstract. One-class classifiers are trained only with target class samples. Intuitively, their conservative modeling of the class description may benefit classical classification tasks where classes are difficult to separate due to overlapping and data imbalance. In this work, three methods leveraging on the combination of one-class classifiers based on non-parametric models, N-ary Trees and Minimum Spanning Trees class descriptors (MST-CD) are proposed. These methods deal with inconsistencies arising from combining multiple classifiers and with spurious connections that MST-CD creates in multi-modal class distributions. Experiments on several datasets show that the proposed approach obtains comparable and, in some cases, state-of-the-art results.

Keywords: One-class Classifiers · Minimum Spanning Tree · Instance-based approaches · Non-Parametric models

1 Introduction

With the rise of social platforms and internet, data produced by users has grown exponentially enabling the use of data greedy machine learning algorithms in several applications, still in many domains of practical interests, data are still scarce and require more data efficient methods especially for non-trivial classification tasks where classes are difficult to separate due to overlapping and data imbalance. One-class classifiers are trained with target class only samples under the strong assumption that data from the other classes are not available or have low quality. Intuitively, their conservative modeling of the data distribution may benefit classical classification tasks if they were combined in a effective manner.

In this work, three methods are proposed which leverage on the combination of one-class classifiers based on non-parametric models, K-Nearest Neighbour, Trees and Minimum Spanning Trees class descriptors (MST-CD), to tackle binary classification problems.

In the first model, we train classifiers using Minimum Spanning Tree Class Descriptor (MST-CD) in the training step and then apply a new technique to provide a more reliable prediction. The second model creates a more powerful classifier based on MST-CD combining results according to an ensemble method.

The third model is very similar to the previous one but uses a tree starting to the closest neighbour to the target pattern for each classifier and finally it leverages on the ensemble technique.

In the next Section related works are shown, the proposed approach is described in Section 4, Experiments are in Section 6 and Conclusions in 7.

2 Related work

In the feature selection field, many approaches have been proposed and used with classifiers to obtain better accuracy [2, 7, 1, 12]. Krawczyk *et al.* [8] proposes a generic model that improves the performance of many common classifiers showing des standard and des-threshold methodologies. They are based on a k-Nearest Neighbour (k-NN) technique that assigns a pattern to a class on the basis of the class of its nearest k neighbours. The approach shows good results and authors compare it with classical models. However, they do not consider the case in which no classifier is activated and, thus, when it is not possible to obtain a prediction for a new instance x . Duin *et al.* [14] proposes a simple approach to assign these refused objects to the class with largest prior probability, but they do not describe a method in the scenario where the two decision boundaries overlap. Our approach combines part of the approach described in [8][14] using MSTs with other methodologies to improve the accuracy and considering also the overlapping. Abpeykar *et al.* [1] proposes a survey that sums up the performance of many classifiers on well-known datasets from UCI repository. A milestone on one-class classifier comes from Pekalska *et al.* [6] with their MST descriptor. The original idea was to try to search a pattern from all training sets in order to create a MST that represents the model on which will be done some geometrical operations with the goal to generate a border for a specific class. Segui *et al.* [11] focuses on the research of noise within a target class and removes it in order to have better accuracy at testing time. They confirm that a graph-based one-class classifier, like MST_CD obtains good results than other approaches, especially dealing with small samples cardinalities and high data dimensionalities. Quinlan [9] proposes a general method that allows predictions using both mixed approach of instance-based and model-based learning. He proves that these composite methods often produce better results in term of predictions than using only a single methodology.

3 Minimum Spanning Tree class descriptor

As widely described in [6] a MST_CD is a non-parametric classifier able to create a structure, seeing only data of the single class of interest. This structure is based on Minimum Spanning Tree, basic elements of this classifier are not only vertices but also edges of the graph, giving a richer representation of the data. Considering edges of the graph as target class objects, additional virtual target objects are generated. The classification of a new object x is based on the distance to the nearest vertex or edge. The key of this classifier is to track a shape around

the training set not considering only all instances of the training but also edges of the graph, in order to have more structured information. Therefore, in prediction phase, the classifier considers two important elements:

- Projection of point x on a line defined by vertices x_i, x_j
- Minimum Euclidean distance between (x, x_i) and (x, x_j)

The Projection of x is defined as follow:

$$p_{e_{i,j}}(x) = x_i + \frac{(x_j - x_i)^T(x - x_i)}{\|x_j - x_i\|^2}(x_j - x_i)$$

if $p_{e_{i,j}}(x)$ lies on the edge $e_{i,j} = (x_i, x_j)$, we compute $p_{e_{i,j}}(x)$ and the Euclidean distance between x and $p_{e_{i,j}}(x)$, more formally:

$$0 \leq \frac{(x_j - x_i)^T(x - x_i)}{\|x_j - x_i\|^2} \leq 1$$

then

$$d(x|e_{i,j}) = \|x - p_{e_{i,j}}(x)\|$$

Otherwise we compute the Euclidean distance of x and pairs (x_i, x_j) , precisely:

$$d(x|e_{i,j}) = \min(\|x - x_j\|, \|x - x_i\|)$$

Therefore, a new object x is recognized by MST_CD if it lies in proximity of the shape built in training phase, otherwise the object is considered as outlier. The decision whether an object is recognized by classifier or not is based on threshold of the shape created during the training phase, more formally:

$$d_{MST_CD}(x|X) \leq \theta$$

Authors set the threshold θ as the median of the distribution of the edge weights $w_{ij} = \|e_{ij}\|$ in the given MST. Given $\hat{e} = (\|e_1\|, \|e_2\|, \dots, \|e_n\|)$ as an ordered edge weights values, they define θ as $\theta = \|e_{[\alpha n]}\|$, where $\alpha \in [0, 1]$. For instance, with $\alpha = 0.5$, we assign the median value of all edge weights into the MST.

4 Proposed Approach

The main objective of a one-class classifier is to recognize instances of a selected class from a set of samples. All instances that are not classified by this model will be considered as outliers (or alien class), while others will be recognized as belonging to the same class of the training set. In this context, we cannot say anything about the refused objects (outlier), but if we have a one-class classifier for each label of the dataset we can not have outliers because if a classifier refuses an object, it should be accepted by the others classifiers. In this work, we use two one-class classifiers trained on two different classes. We assume a discriminant function f_{ab} on a binary classification problem to classify a new

object considering acceptance and rejection from both our classifiers. Given X a MST and x a new object, we define a function that assigns a label such that:

$$f_{ab}(x, X) = \begin{cases} 1 & \text{if } d_{MST.CD_0}(x|X) \leq \theta \text{ and } d_{MST.CD_1}(x|X) > \theta_1 \\ 0 & \text{if } d_{MST.CD_0}(x|X) > \theta \text{ and } d_{MST.CD_1}(x|X) \leq \theta_1 \end{cases} \quad (1)$$

However, we may have two anomaly cases:

1. both classifiers refuse the pattern to be predicted
2. both classifiers accept the pattern to be predicted

In this specific case we can apply a simple technique to assign a pattern to two possible classes depending on its distance to clusters. More formally, given x a target vector, i and j are all elements of the dataset belonging to both classes, we define:

$$\begin{aligned} d(x|i) &= (|x - i_0|, |x - i_1|, \dots, |x - i_{n_0}|) \\ d'(x|j) &= (|x - j_0|, |x - j_1|, \dots, |x - j_{n_1}|) \end{aligned}$$

where n_0 and n_1 are cardinalities of first and second class respectively. Then we take the k nearest elements for each class, such that:

$$k \leq \min(n_0, n_1)$$

We compute the vector difference as:

$$Vector_{diff} = [(d_1 - d'_1), (d_2 - d'_2), \dots, (d_k - d'_k)]$$

Finally, We assume a new function f'_{ab} to classify the object as:

$$f'_{ab}(x, Vector_{diff}) = \begin{cases} 1 & \text{if } |Vector_{diff}(k \leq 0)| > |Vector_{diff}(k > 0)| \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

4.1 First approach

Our first approach combines two one-class classifiers based on Minimum Spanning Tree class descriptor and solve both above mentioned issues. When one of the two issues appear, e.g. both classifiers accept (classifiers overlap) or reject (uncovered) the input sample, an approach similar to K-NN majority vote is applied. Using the already computed euclidean distances between the sample and the elements of the two MSTs, the K elements of each MST (MST_1, MST_2) nearer to the sample are selected to check which of the two MSTs is consistently closer to the sample. This is done by internally sorting the two sets of K elements in increasing distance order from the sample then subtracting the 2 corresponding K -ary vectors (D_1, D_2) of distances and finally counting how many positive elements are in the resulting vector $R = D_1 - D_2$. If R contains more positive than negative elements the sample is associated to MST_2 as its elements are closer to the sample. This method allows to integrate the generalization capabilities of the MST with the robustness offered by the K-NN vote strategy to deal with binary classification strategies. See the pseudo-code in Algorithm 1.

Algorithm 1 MST CD OVA

```

1: for  $v \in \mathcal{G}_0$  do
2:   all euclidean distances  $\leftarrow \|x - v\|$ 
3: end for
4: NodeX = Take min(all euclidean distances)
5: EdgeNodeX  $\leftarrow$  Search inc/out edge nodeX and return  $(u, v)$ 
6: for  $u, v \in \mathcal{E}(\text{NodeX})$  do
7:   if  $0 \leq \frac{(x_j - x_i)^T * (x - x_i)}{\|x_j - x_i\|^2} \leq 1$  then
8:      $P_{e_{ij}}(x) = x_i + \frac{(x_j - x_i)^T * (x - x_i)}{\|x_j - x_i\|^2} * (x_j - x_i)$ 
9:      $d(x|e_{ij}) \leftarrow \|x - P_{e_{ij}}(x)\|$ 
10:  else
11:     $d(x|e_{ij}) \leftarrow \min\{\|x - x_i\|, \|x - x_j\|\}$ 
12:  end if
13: end for
14: Repeat line 1-13 for graph  $G_1$ 
15: min dist0 =  $\min(d(x|e_{ij}))$ 
16: min dist1 =  $\min(d_1(x|e_{ij}))$ 
17:  $1 \leftarrow$  if  $d_{MST-CD_0}(x|X) \leq \theta$  and  $d_{MST-CD_1}(x|X) > \theta_1$ 
18:  $0 \leftarrow$  if  $d_{MST-CD_0}(x|X) > \theta$  and  $d_{MST-CD_1}(x|X) \leq \theta_1$ 
19: if min dist0  $\leq \theta$  and min dist1  $\leq \theta_1$  then
20:   knn weight1=order( $d_1(x|u)$ ) and take  $k_1$ -elements
21:   knn weight0=order( $d_0(x|v)$ ) and take  $k_1$ -elements
22:   euclidean distance vectors = (knn weight1 - knn weight0)
23:   positive = Count  $n_i > 0$  in euclidean distance vectors
24:   negative = Count  $n_i < 0$  in euclidean distance vectors
25:   if negative  $\geq$  positive then
26:     prediction  $\leftarrow 1$ 
27:   else
28:     prediction  $\leftarrow 0$ 
29:   end if
30: end if
31: if min dist0  $> \theta$  and min dist1  $> \theta_1$  then
32:   The approach is equal to lines [20-29]
33: end if

```

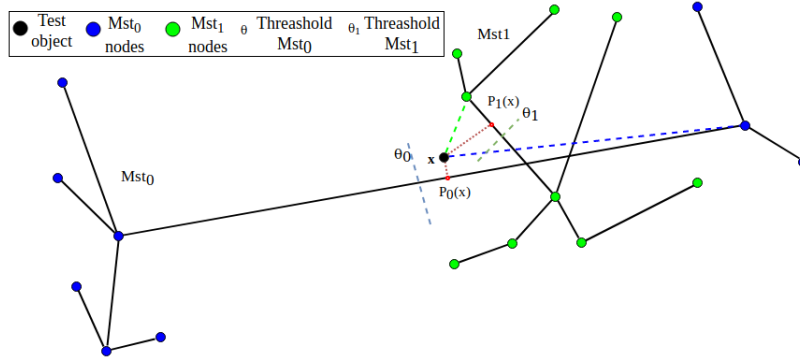


Fig. 1: MST_CD model: In this figure, we show the problem about the competence area into two different classifiers in a toy scenario. In this case the object x will be predicted as blue class instead of green class. This happens because the orthogonal projection of x is nearer to an edge that contains an outlier and the object is located into a non-competence area of the blue-classifier making a wrong prediction.

4.2 Second approach

In a second approach, we mix K-NN and MST in the reverse order. Using a K-NN like approach, we initially select the K training samples from each class closest to the X sample to classify. This results in two sets S_1 and S_2 , we then create two K -elements MST for each of this sets ($kMST_1, kMST_1$). After this we classify X as we did for approach 1. This second approach has two advantages over the previous. One, is that it does not need to perform the expensive creation of the two large MST covering the whole dataset (quadratic in the number of samples) but only deals with a K -elements MSTs. Second, it avoids MST spurious edges between elements of distant and unrelated nodes of a class distribution. See the pseudo-code in Algorithm 2 and Fig.2.

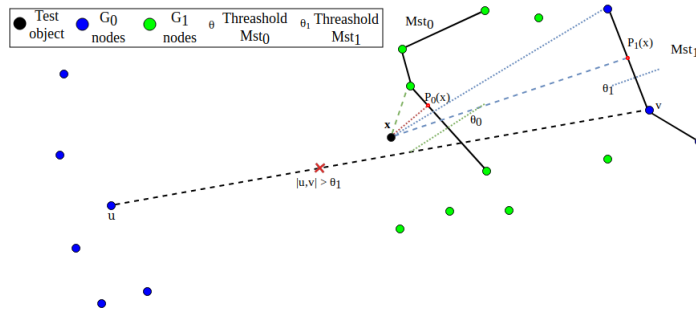


Fig. 2: MST_CD_GP model: A different approach on the same toy example considering a combination built on two small MST starting from nodes with a minimum distance from target pattern.

Algorithm 2 MST CD with Gamma parameters

```

1: function CREATE SMALL MST(g0 weight sorted, g1 weight sorted)
2:   smallG0 = first gamma index sorted values in g0 weight sorted
3:   edges couple ← all combinations nodes small g0
4:   for u, v ∈  $\mathcal{I}dgescouple$  do
5:     small G0 ← (node u, node v, weight = (u, v))
6:   end for
7:   small MST0 = ComputeMST(small G0)
8:   Same approach for MST1
9:   e(small MST0) = (||e0||, ||e1||, ..||en||)
10:  e(small MST1) = (||e0||, ||e1||, ..||en||)
11:   $\theta_0 = ||e_{(\alpha_n)}||$ 
12:   $\theta_1 = ||e_{(\alpha_n)}||$ 
13: return small MST0, smallMST1
14: end function

```

4.3 Third approach

The third is similar to the second one but instead of selecting the K nearest elements to the sample X to classify, it selects only the nearest element E to the sample and then selects the $K-1$ elements of the training set nearest to E . It then creates a tree having E as root and it uses it as a classifier like in previous cases. This further extends the robustness to outliers and spurious connections between far nodes. See the pseudo-code in Algorithm 3 and Fig.3.

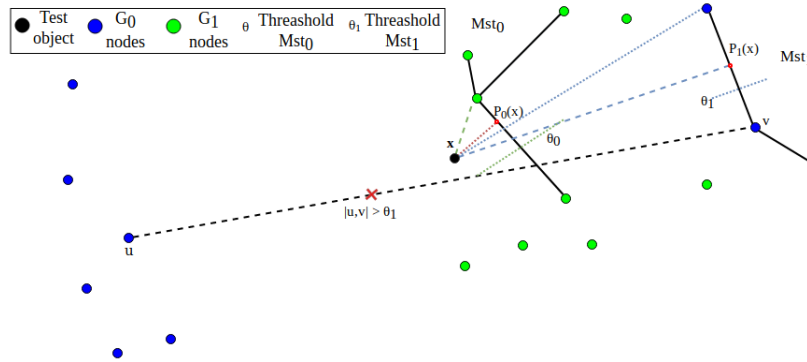


Fig. 3: N-ary model: A different approach on the same toy example. Using a N-ary model makes less complex the prediction phase ignoring outliers from the structure.

5 Datasets

We use five low-dimensional datasets (*Hill, Sonar, Australian, Mofn, Pima*) and three high-dimensional datasets (*Arcene, Gisette, Madelon*), all taken from the

Algorithm 3 N-ary model

```

1: function CREATE_N_ARY(subG0 weight sorted, subG1 weight sorted, gamma index)
2:   n_ary0 = first gamma index sorted values in g0 weight sorted
3:   n_ary1 = first gamma index sorted values in g1 weight sorted
4:   for u, v ∈ subG0weightedsorted do
5:     n_ary0 ← (nodeu, nodev, weight = (u, v))
6:   end for
7:   for u, v ∈ subG1weightedsorted do
8:     n_ary1 ← (nodeu, nodev, weight = (u, v))
9:   end for
10:  return n_ary0, n_ary1
11: end function

```

UCI repository [3]. All of them have two classes. Table 1 shows the details of all datasets. To evaluate our approach in a more robust way, we selected datasets with a huge variability in term of number of features and number of instances. For all the datasets we use 5-fold Cross Validation. We replace the missing values in the datasets with the average value for the missing features.

Table 1: Number of features, classes, instances and positive-negative samples for all the datasets used in our experiment.

Datasets	Features	Classes	Instances	pos-neg
Arcene	10000	2	100	44-56
Gisette	5000	2	6000	3000-3000
Madelon	500	2	2000	1000-1000
Hill	101	2	606	305-301
Sonar	60	2	208	97-111
Australian	14	2	690	307-383
Mofn	10	2	1324	292-1032
Pima	8	2	688	305-301

6 Experiments

In this section we report two groups of experiments,

- to study the effect on the parameters in our models,
- to compare the proposed models with the results available in the literature on some well known benchmarks.

6.1 Parameters evaluation

For each classifier, we extract some common metrics to evaluate the performance, such as Sensitivity, Precision and F1 Score. Starting from these values we create confusion matrix and obtain the final accuracy considering True/False positive and True/False negative samples.

Table 2: These are the results of both classifiers using our MST_CD model about Sensitivity, Precision, F1 Score and Total accuracy on Sonar dataset. Last line contains the average for each column.

MST_CD							
<i>Fold</i>	<i>Sensitivity</i> ⁰	<i>Sensitivity</i> ¹	<i>Precision</i> ⁰	<i>Precision</i> ¹	<i>F1</i> ⁰	<i>F1</i> ¹	<i>Accuracy</i>
1	0,867	0,900	0,867	0,900	0,867	0,900	0,886
2	0,833	0,944	0,938	0,850	0,882	0,895	0,889
3	0,800	1,000	1,000	0,833	0,889	0,909	0,900
4	0,688	0,880	0,786	0,815	0,733	0,846	0,805
5	0,714	0,850	0,769	0,810	0,741	0,829	0,794
<i>Average</i>	0,780	0,915	0,872	0,842	0,822	0,876	0,855

Table 3: Details of our third model N-ary considering the same measures than above table. We show an accuracy of 87.3% with gamma=20 and gamma=10 in N-ary model against 85.5% in our first model. Results of each lines are extracted using an average on k=5 fold-validation.

N-ary							
<i>Gamma</i>	<i>Sensitivity</i> ⁰	<i>Sensitivity</i> ¹	<i>Precision</i> ⁰	<i>Precision</i> ¹	<i>F1</i> ⁰	<i>F1</i> ¹	<i>Accuracy</i>
30	0,791	0,915	0,873	0,850	0,829	0,881	0,860
20	0,793	0,915	0,873	0,850	0,830	0,881	0,861
10	0,820	0,915	0,879	0,869	0,847	0,890	0,873
8	0,830	0,906	0,869	0,876	0,847	0,890	0,873
6	0,816	0,880	0,849	0,854	0,830	0,865	0,851
4	0,808	0,882	0,852	0,848	0,826	0,863	0,848
3	0,787	0,861	0,830	0,832	0,804	0,844	0,829

6.2 Comparison

We evaluated the proposed models on the well known benchmark datasets described in Section 5. In particular, we compared our methods using the results published in [7, 1] and shown the comparison in Tables 5, 6 and 7. In our work, we have two one-class classifiers and the rate of True/Negative samples has been computed considering two different models. For instance if an object has been predicted correctly True positive by a classifier the others must be refuse and label it as True negative, otherwise, we consider all the possible combinations (2^2) in our case (two classifiers). The final accuracy is obtained from average of Cross-Validation. In our experiments, we set γ parameter into the second and third model within the range [2,30] nodes and take the maximum accuracy ob-

Table 4: Parameters set

Parameters	Arcene	Madelon	Gisette	Mofn	Australian	Pima	Sonar	Hills
Threshold	0.25	0.25	0.2	0.05	0.05	0.05	0.1	0.9
Neighbours	4	12	4	2	6	6	2	2

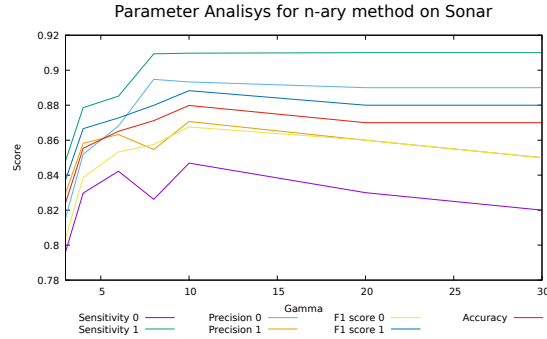


Fig. 4: Best Gamma parameter search for method using the Sonar dataset. We use the average accuracy to select the best parameter.

tained from Cross-Validation methods considering always the same dimension training/test (80%-20%) see Table 3. Threshold measures and k-neighbours are set as we show into Table 4 on each specific dataset. In Table 5 we compare results of our models with many references used to examine the performance of classifiers (Moradi & Rostami, 2015). Our models do not use feature selection techniques, therefore for each instance we exploit all available features. Comparing our models with the ensemble methods known as AdaboostM1 [4], Bagging [15], Dagging [15], LogitBoost [5], MODLEM [13], Decorate (Melville & Mooney, 2003), Grading [10], MultiBoostAB [16] and StackingC (Seewald, 2002), our models overcome the accuracy of other ensembles classifier (except for Arcene dataset where Bagging, LogitBoost and Modlem have higher accuracy than our models).

Table 5: 5-fold Cross-Validation accuracy results on 6 datasets. Results obtained with our three models (MST_CD, MST_CD.GP, N-ary) are compared with results published in [7] on same datasets.

Dataset	Krakovna <i>et al.</i> [7]										Our models		
	Bart	c5.0	Cart	Lasso	LR	NB	RF	SBFC	SVM	TAN	CD	CD_GP	N-ary
Arcene	71.6	66	63	65.6	52	69	71.8	72.2	72	-	79.6	77.7	80.3
Madelon	76	75.8	78.2	60.7	60	59.8	67.1	63.4	62	54.2	75	75.2	75.3
Gisette	97.7	94.8	90.8	97.2	88.1	90.3	97	95.2	96.9	-	96.8	-	-
Mofn	100	84.8	83.9	100	100	86.4	92.4	86.2	94.6	92.1	100	100	100
Australian	86.9	86.7	84.2	85.6	86.8	85.7	87.8	86.9	86	86.8	69.0	70.6	70.3
Pima	78.2	76.8	75.7	78	78.3	78	77.8	78.9	78.1	78.9	70.1	71.7	74.1

Table 6: 5-fold Cross-Validation accuracy results on five different datasets. Results obtained with our three models (MST_CD, MST_CD_GP, N-ary) are compared with the results of different ensemble methods published in [1].

Dataset	Abpeykar <i>et al.</i> [1]									Our models		
	AdaB.	Bagg.	Dagg.	LogitB.	Mod.	Decor.	Grad.	Mt.B	Stack.C	CD	CD_GP	N-ary
Arcene	79.5	82.5	74.5	85.5	86.0	-	56.0	80.0	56.0	79.6	77.7	80.3
Madelon	63.4	75.0	57.2	63.0	52.5	73.3	50.1	61.7	50.1	75	75.2	75.3
Sonar	71.6	76.9	69.7	79.3	70.6	84.1	53.3	74.5	53.3	85.4	85.2	88
Hills	50.4	50.2	50.4	50.4	-	-	50.4	50.4	50.4	58.1	57.9	61.1
Gisette	88.9	75.0	82.2	89.4	-	82.2	48.1	82.7	48.1	96.8	-	-

Table 7: 5-fold Cross-Validation accuracy results on five datasets. The results obtained with our three models are compared with the results of different machine learning methods published in [1].

Datasets	Abpeykar <i>et al.</i> [1]					Our models		
	Dt	k-NN	NB	RF	SVM	MST_CD	MST_CD_GP	N-ary
Arcene	67.2	66.7	69.7	71.4	71.4	79.6	77.7	80.3
Madelon	82.7	74.8	79.3	77.9	78.4	75	75.2	75.3
Sonar	79.5	80.4	77.6	81.4	82.3	85.4	85.2	88.0

7 Conclusion

The presented results show that the solutions proposed are competitive both with ensemble and classical classifiers. The second and third method that aimed at increasing the robustness to outliers and spurious MST connections were proved to consistently ameliorate accuracy. Also the latter methods avoid the initial computation of expensive large MST while requiring the computation of small MST at runtime and provide higher accuracy. This makes the latter methods more convenient with large datasets for which the complete MST computation may be too expensive. We are also confident that these methods can be highly optimized using caching methods to further improve online computation performance. Future work will tackle the issue of feature scaling and selection as well as the possibility to combine the approach with ensemble methods.

References

1. Abpeykar S., G.M., H., Z.: Ensemble decision forest of rbf networks via hybrid feature clustering approach for high-dimensional data classification. *Computational Statistics and Data Analysis* (2019)
2. Budak, H., Taabat, S.: A modified t-score for feature selection. *Anadolu universitesi Bilim Ve Teknoloji Dergisi A-Uygulamal Bilimler ve Mhendislik* (2016)
3. Dua, D., Graff, C.: UCI machine learning repository (2019), <http://archive.ics.uci.edu/ml>
4. Eibl, G., Pfeiffer, K.P.: How to make adaboost. m1 work for weak base classifiers by changing only one line of the code. In: *European Conference on Machine Learning*, pp. 72–83. Springer (2002)

5. Frank, E., Holmes, G., Kirkby, R., Hall, M.: Racing committees for large datasets. In: International Conference on Discovery Science. pp. 153–164. Springer (2002)
6. Juszczak, P., Tax, D.M., Pe, E., Duin, R.P., et al.: Minimum spanning tree based one-class classifier. *Neurocomputing* **72**(7-9), 1859–1869 (2009)
7. Krakovna, V., Du, J., Liu, J.S.: Interpretable selection and visualization of features and interactions using bayesian forests. *Statistics and Its Interface* **11**, 503–513 (2018)
8. Krawczyk, B., Galar, M., Woźniak, M., Bustince, H., Herrera, F.: Dynamic ensemble selection for multi-class classification with one-class classifiers. *Pattern Recognition* **83**, 34–51 (2018)
9. Quinlan, J.R.: Combining instance-based and model-based learning. In: Proceedings of the Tenth International Conference on Machine Learning. pp. 236–243 (1993)
10. Seewald, A.K., Fürnkranz, J.: An evaluation of grading classifiers. In: International symposium on intelligent data analysis. pp. 115–124. Springer (2001)
11. Segui, S., Igual, L., Vitrià, J.: Weighted bagging for graph based one-class classifiers pp. 1–10 (2010)
12. Singh, B., Vyas, O.: Maximum spanning tree based redundancy elimination for feature selection of high dimensional data. *International Arab Journal of Information Technology* **15**, 831–841 (09 2018)
13. Stefanowski, J.: The rough set based rule induction technique for classification problems. In: In Proceedings of 6th European Conference on Intelligent Techniques and Soft Computing EUFIT. vol. 98 (1998)
14. Tax, D.M., Duin, R.P.: Using two-class classifiers for multiclass classification. In: Object recognition supported by user interaction for service robots. vol. 2, pp. 124–127. IEEE (2002)
15. Ting, K.M., Witten, I.H.: Stacking bagged and dagged models (1997)
16. Webb, G.I.: Multiboosting: A technique for combining boosting and wagging. *Machine learning* **40**(2), 159–196 (2000)