

# Active Learning for One-Class Classification Using Two One-Class Classifiers

Patrick Schlachter and Bin Yang

Institute of Signal Processing and System Theory

University of Stuttgart, Germany

{patrick.schlachter, bin.yang}@iss.uni-stuttgart.de

**Abstract**—This paper introduces a novel, generic active learning method for one-class classification. Active learning methods play an important role to reduce the efforts of manual labeling in the field of machine learning. Although many active learning approaches have been proposed during the last years, most of them are restricted on binary or multi-class problems. One-class classifiers use samples from only one class, the so-called target class, during training and hence require special active learning strategies. The few strategies proposed for one-class classification either suffer from their limitation on specific one-class classifiers or their performance depends on particular assumptions about datasets like imbalance. Our proposed method bases on using two one-class classifiers, one for the desired target class and one for the so-called outlier class. It allows to invent new query strategies, to use binary query strategies and to define simple stopping criteria. Based on the new method, two query strategies are proposed. The provided experiments compare the proposed approach with known strategies on various datasets and show improved results in almost all situations.

## I. INTRODUCTION

A general challenge of supervised machine learning is the acquisition of hand-labeled samples. Since labeling often requires human experts, it is a time-consuming and costly task. Therefore, it is highly desirable to reduce the manual labeling effort. This is also the case for one-class classification.

One-class classification tries to recognize samples of one given class, the target class, against samples belonging to the other class, the outlier class. Its specialty lies in the fact that only samples from the target class are used to train a one-class classifier (OCC). Samples of the outlier class are rarely available or, if they are available, not representative to describe different types of outliers and thus cannot be used. In contrast to binary classification, an OCC does not try to separate samples from two classes but encompasses the training samples of the target class. During the evaluation of an OCC, samples from both classes are necessary of course. An overview of one-class classification and various OCCs was shown in [1], [2].

Although one-class classification was first motivated by the absence of outlier samples during training, it is also deployed in other scenarios. For example, OCCs are used in some binary classification applications whose training data is either imbalanced or only representative of one of both classes. An example of the latter case is fault detection where possible types of faults are unknown in advance.

Despite requiring only samples from one class during training, the problem of a high labeling effort is still present.

An expert, a so-called *oracle*, has to manually select target samples from a pool of unlabeled data which are subsequently used to train an OCC. Regarding one-class classification, the problem is aggravated by the fact that many one-class applications have to strictly avoid false negative (FN) errors, i.e. outliers classified as targets. Consequently, decision boundaries of OCCs are normally chosen tight to training samples of the target class. This can lead to many false positive (FP) errors, i.e. targets classified as outliers, at which human experts typically have to take a closer look. If a target is found, it can be used to iteratively retrain the OCC.

Active learning is a technique to tackle the problem of high labeling efforts by querying only the samples that are most valuable for the training of a classifier. There exist various query strategies [3] which differ in the interpretation of the term "valuable". However, most of the proposed query strategies are limited to binary or multi-class classifiers. Active learning for one-class classification is more challenging because the decision boundary of an OCC is only supported by samples from one class. Hence, labeled outliers, i.e. samples that the oracle marked as outliers, cannot be used for iterative training.

Some former works proposed active learning strategies specially tailored to one-class classification. In general, they are a trade-off between *exploration* and *exploitation* of the feature space. Querying a sample in a non-sampled area of the feature space corresponds to exploration and helps to extend the decision boundary to the non-sampled area. In contrast, selecting a sample in a sampled area of the feature space contributes to exploitation and helps to refine the decision boundary. The query strategy uncertainty sampling presents those samples to an oracle about which a classifier is the most uncertain [4]. Juszczak et al. adopted uncertainty sampling to one-class classification by introducing confidence values [5]. They proposed four query strategies from which the "lh" strategy works best for almost all experiments. It presents estimated targets with low confidence and estimated outliers with high confidence to the oracle. In contrast to uncertainty sampling, Barnabé-Lortie et al. proposed to query samples about which an OCC is the most certain that they are outliers [6]. Ghasemi et al. presented a different approach which makes use of the distribution of target and unlabeled samples and does not consider the classification results of an OCC for active learning. Based on kernel density estimation, they proposed the two query strategies expected margin sampling

and entropy sampling [7]. There exist further query strategies which are not considered in this paper due to their limitation to special OCCs or high computational cost. Some of them are based on kernel density estimation [8], [9], support vector data description [10], [11] or the variation in label assignments [9]. Finally, it was shown in [12] that active learning for one-class classification can be improved by applying dimensionality reduction.

Most of the presented active learning strategies for OCCs have one major drawback: If a sample is labeled as an outlier by the oracle, both the OCC and the active learning strategy receive no information gain. During training, the reason for having no information gain obviously lies in the nature of one-class classification. However, labeled outliers can improve active learning.

In this paper, we propose a new active learning method for one-class classification which is able to take labeled outliers into account. Our method is generic in the sense that it enables to define various query strategies. Furthermore, it is universally applicable, i.e. it is not limited to specific OCCs, and it is adjustable in multiple manners like the trade-off between exploration and exploitation. Finally, the proposed method does not rely on specific assumptions on a dataset or on the distribution of the outlier class. Based on the generic method, two query strategies are proposed.

During our experiments, we made use of a popular one-class classifier: the one-class support vector machine (OCSVM) proposed by Schölkopf et al. [13]. It is based on the maximization of the margin between a hyperplane and the origin of the feature space. The solution of an estimated class label  $\hat{y}$  for a given feature vector  $\mathbf{x}$  and a training set  $\{\mathbf{x}_i\}_{i=1}^n$  is given by

$$\hat{y}(\mathbf{x}) = \text{sgn}\left(\underbrace{\sum_{i=1}^n \alpha_i K(\mathbf{x}, \mathbf{x}_i)}_{DV} - \rho\right). \quad (1)$$

Alg. 1: Proposed active learning method for one-class classification based on two OCCs

```

1: given: target set  $\mathcal{T}$ , outlier set  $\mathcal{O}$ , unlabeled set  $\mathcal{U}$ 
2:  $\mathbf{x}_i = \emptyset$ ,  $y_i = \text{target}$ ,  $DV_u^t = DV_u^o = 1 \forall u = 1, \dots, |\mathcal{U}|$ 
3: loop
4:   if  $y_i = \text{target}$  then
5:     add sample  $\mathbf{x}_i$  to  $\mathcal{T}$ 
6:     retrain target OCC with  $\mathcal{T}$ 
7:     test target OCC on  $\mathcal{U} \Rightarrow \{DV_u^t, u = 1, \dots, |\mathcal{U}|\}$ 
8:   else
9:     add sample  $\mathbf{x}_i$  to  $\mathcal{O}$ 
10:    retrain outlier OCC with  $\mathcal{O}$ 
11:    test outlier OCC on  $\mathcal{U} \Rightarrow \{DV_u^o, u = 1, \dots, |\mathcal{U}|\}$ 
12:   end if
13:    $\mathbf{x}_i = \arg \max_u \{\text{agg}(DV_u^t, DV_u^o), u = 1, \dots, |\mathcal{U}|\}$ 
14:   query label  $y_i$  of sample  $\mathbf{x}_i$  from oracle
15:   remove sample  $\mathbf{x}_i$  from  $\mathcal{U}$ 
16: end loop

```

The value of the discriminant function whose sign defines the class affiliation is called *decision value*  $DV$ . It is based on Lagrange multipliers  $\alpha_i$ , an offset  $\rho$  and a kernel  $K$ . As our method is not limited to OCSVMs, decision values may emerge from other OCCs. In general, decision values are negative for samples that belong to the training class and positive for the other class. The higher the absolute value of a decision value is, the more certain an OCC is about the estimated class affiliation.

The remaining parts of this paper are structured as follows: In the next section, we present our main contribution, a novel active learning method for one-class classification. Section 3 evaluates the method on multiple datasets using an OCSVM as OCC and compares it to other query strategies. Finally, we will conclude our work in Section 4.

## II. PROPOSED METHOD

The proposed generic active learning method for one-class classification is based on two OCCs: one for the target class and one for the outlier class. The target-class classifier is used for classification and both OCCs are used for active learning.

### A. Basic idea

The basic idea behind this method is to make use of labeled outliers during active learning. As mentioned in the introduction, OCCs are also deployed in applications where outliers occur frequently but are not necessarily representative of the whole distribution of the outlier class. Therefore, switching to a binary classifier after obtaining some labeled outliers is not always reasonable. Indeed, binary classification carries the risk of producing many FN errors, whereas an OCC leads to mostly FP errors which is more desirable in many one-class applications. Although outlier samples are insufficient for classification, an OCC trained on them can support active learning.

We notice that combining OCCs is well known for both multi-class classification [14], [15] and to improve one-class classification in terms of performance and robustness [16]. However, for the purpose of active learning, joining two one-class classifiers is novel.

### B. Algorithm

Alg. 1 outlines the proposed active learning process for an OCC.  $y_i$  denotes the class label of the previously added sample  $\mathbf{x}_i$  and is initialized to the target label. If  $y_i$  is a target, then the target OCC is retrained on the increased target set  $\mathcal{T}$  and is used to test the samples in the unlabeled set  $\mathcal{U}$ . If  $y_i$  is an outlier, then the outlier OCC is retrained on the increased outlier set  $\mathcal{O}$  and is used again to test all unlabeled samples. Together with the respective second stored decision boundary, this results in two decision values  $DV_u^t$  and  $DV_u^o$  for each unlabeled sample which are subsequently aggregated using an *aggregation function*  $\text{agg}$ . Finally, the sample with maximal aggregate is presented to an oracle. Corresponding to other active learning approaches, this querying sequence is repeated until a defined stopping criterion is fulfilled.

The method has several advantages. First, it is able to consider labeled outliers during active learning. In addition, only

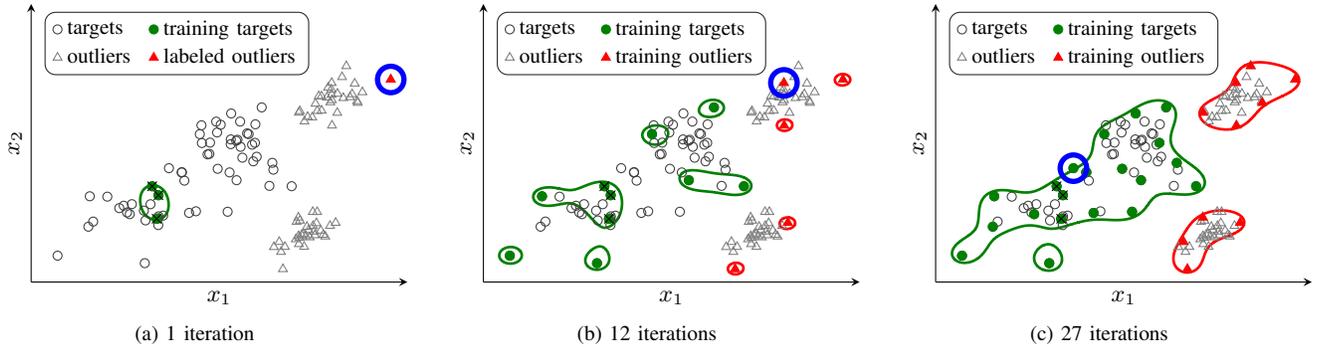


Fig. 1: Exploration sampling on artificial dataset after 1, 12 and 27 iterations. The three initial training targets are marked by crosses. Decision boundaries of target OCC and outlier OCC are depicted as green and red lines, respectively. The currently queried sample is marked by a thick blue circle.

one of both OCCs has to be retrained per query step which makes it computationally efficient. Moreover, the generic method can be adapted by choosing different aggregation functions  $\text{agg}$ . Further adjustments are conceivable by using different hyperparameters for both OCCs.

The following two subsections propose two meaningful aggregation functions which result in the query strategies *exploration sampling* and *similarity sampling*.

### C. Exploration sampling

The query strategy *exploration sampling* is based on the generic method from above and uses the following aggregation function:

$$\text{agg}_{\text{exp}} = DV^t \cdot DV^o. \quad (2)$$

In contrast to outlier sampling [6], exploration sampling presents samples to an oracle that lie not only furthest from the decision boundary of the target OCC but also furthest from the decision boundary of the outlier OCC. This leads to a high exploration which is crucial in many one-class applications that either have to be operational based on only a few training samples or in which asking an oracle is expensive. The risk of presenting many true outliers to an oracle is reduced by the fact that real-world outliers are clustered and not uniformly distributed in the feature space. For example, in fault detection, outliers are typically clustered due to multiple kinds of errors. Unlike outlier sampling [6], exploration sampling quickly presents outliers from different clusters to an oracle by producing smaller aggregates  $\text{agg}$  for samples that lie near already sampled regions in the feature space. Consequently, an expert can faster detect different outlier causes.

Fig. 1 illustrates the incremental development of the optimal decision boundary by using exploration sampling on an artificial, two-dimensional dataset consisting of 60 targets and 60 outliers. Two OCSVMs [13] were used as OCCs. Fig. 1b demonstrates how exploration sampling leads to uniformly exploring the feature space after 12 iterations. Finally, Fig. 1c shows an optimal classification result after only 27 iterations and no more than 10 presented outliers.

### D. Similarity sampling

The usage of a second OCC allows to design the query strategy *similarity sampling* which is similar to classical uncertainty sampling [4] and differs only in the kind of uncertainty measure. Instead of using decision values from a binary classifier as uncertainty, similarity sampling uses the absolute difference between the decision values of both OCCs as uncertainty measure. The following aggregation function  $\text{agg}_{\text{sim}}$  leads to querying samples which lie in the middle between the decision regions of the target and outlier class, i.e. samples with similar values of  $DV^t$  and  $DV^o$ :

$$\text{agg}_{\text{sim}} = -|DV^t - DV^o|. \quad (3)$$

It focuses more on exploitation than on exploration. Thereby we assume decision values that are normalized to a maximum of 1 in order to be consistent with the initial decision values in Alg. 1.

### E. Stopping criterion

In contrast to other active learning approaches, the proposed method easily allows to define the following stopping criterion: The active learning stops if, for all samples from  $\mathcal{U}$ , at least one of both decision values is below its respective user-predefined threshold  $T^t$  or  $T^o$ :

$$\forall x_u \in \mathcal{U} : DV_u^t < T^t \vee DV_u^o < T^o. \quad (4)$$

In other words, active learning is finished if for each sample at least one of both OCCs is sure enough that the sample belongs to the respective class. Regarding OCSVMs, a reasonable choice is  $T^t = T^o = 0$ .

## III. EXPERIMENTS

### A. Setup

We evaluated our proposed active learning strategies on 15 datasets which are listed in Table I. All datasets originate from the Outlier Detection DataSets (ODDS) Library [17]. Some of them contain many possibly representative outliers to deploy binary classification. Nevertheless, we used all datasets for one-class classification. The selected datasets differ in its

TABLE I: Datasets used in experiments

Dataset	Features	Targets	Outliers	Batch size	Iterations
BreastW	9	444	239	1	40
Cardio	21	1655	176	1	40
Glass	9	205	9	1	40
Letter	32	1500	100	5	100
Lympho	18	142	6	1	10
Mammography	6	10923	260	30	50
Mnist	100	6903	700	10	40
Musk	166	2965	97	1	60
Optdigits	64	5066	150	5	50
Pendigits	16	6714	156	1	50
Satellite	36	4399	2036	10	50
Thyroid	6	3679	93	1	50
Vowels	12	1406	50	3	50
WBC	30	357	21	1	20
Wine	13	119	10	1	20

absolute number of samples, number of features and relative amount of outliers.

We deployed an OCSVM [13] with radial basis function kernel as OCC. Its optimal hyperparameters were determined by a grid search based on splitting a dataset into training and validation set. Both OCCs in the proposed active learning method were trained using the same hyperparameters.

Our evaluation procedure was similar to [6]. We performed a 10-fold cross-validation. The first fold was used as initial training set, the folds 2-6 were considered as pool of unlabeled samples and the remaining four folds were used as test set. The split into 10 folds was only done once. The mapping between fold number and real fold was circularly shifted 10 times. In order to start all active learning strategies with the same amount of information, the outliers in the initial training set were ignored.

Besides the three proposed query strategies, our experiments considered the active learning approaches "lh" uncertainty sampling [5], expected margin sampling [7], entropy sampling [7], outlier sampling [6] and random sampling. The two kernel density estimation strategies of [7] were implemented using the publicly available code by Ghasemi [18]. In the case of large datasets, batches consisting of multiple samples were queried after each iterative training step in order to reduce the computational effort. The batch size for each dataset is listed in Table I.

The metric balanced accuracy (BACC)

$$\text{BACC} = \frac{1}{2} \left( \frac{TN}{N} + \frac{TP}{P} \right) \quad (5)$$

was calculated to evaluate the results.  $N$  denotes the number of targets and  $P$  the number of outliers.  $TN$  represents the number of true negatives, i.e. the number of correctly classified targets.  $TP$  denotes the number of true positives, i.e. the number of correctly classified outliers.

Attention should be paid to the evaluation of active learning strategies because their performance depends on the number of queries. Due to the practical relevance, we compared query strategies after a few iterations. Thus, the number of presented samples was manually chosen in such a way that

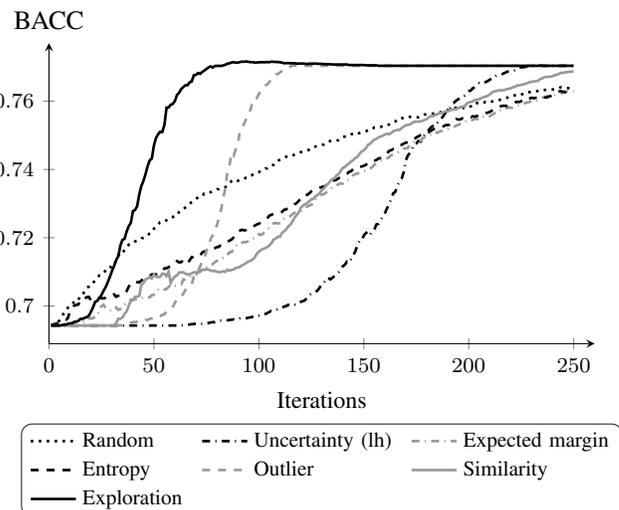


Fig. 2: Balanced accuracy over number of iterations on Satellite dataset using various active learning strategies

the performance improved noticeably, but did not yet saturate. The total number of presented samples equals the number of iterations times the batch size. These values are listed in Table I.

### B. Results

At first, we compared all query strategies depending on the number of iterations by taking the exemplary dataset Satellite. Fig. 2 shows the resulting graph. The plot reveals that exploration sampling performs best after a few iterations until the performance saturates. Thus, exploration plays an important role at the beginning, when only few labeled samples are available. Expected margin and entropy sampling lead both to an approximately linear increase of the BACC on the exemplary dataset. Random sampling performed best for very few iterations and was always better than expected margin and entropy sampling. Similarity sampling performed worse than random sampling for almost all considered iterations. A possible reason is that it focuses on exploitation which seems to be not as important as exploration in one-class classification. Outlier sampling performed poorly for few iterations. This is because it found mostly true outliers which have no impact on the target OCC used for classification. After enough iterations, its performance improved distinctly, because informative targets were found which contribute to a high exploration. Finally, "lh" uncertainty sampling lead to a bad performance indicating a low exploration and exploitation. One reason for this is the fact that estimated targets with low confidence contribute little to iterative learning. Furthermore, estimated outliers with high confidences are often true outliers which do not affect the classification result.

Table II shows the results of the experiments on all datasets by listing the mean and standard deviation of the BACC calculated over all folds. The table evaluates the query strategies after a practically relevant number of iterations during which exploration is most important. The results confirm that

TABLE II: Experimental results (mean and standard deviation of BACC calculated over 10-fold cross-validation)

Dataset	Initial	Random	Uncertainty (lh) [5]	Expected margin [7]	Entropy [7]	Outlier [6]	Similarity	Exploration
BreastW	0.925 ( $\pm$ 0.023)	0.938 ( $\pm$ 0.014)	0.925 ( $\pm$ 0.024)	0.931 ( $\pm$ 0.013)	0.936 ( $\pm$ 0.013)	0.925 ( $\pm$ 0.024)	0.944 ( $\pm$ 0.016)	<b>0.953</b> ( $\pm$ <b>0.010</b> )
Cardio	0.842 ( $\pm$ 0.014)	0.860 ( $\pm$ 0.012)	0.846 ( $\pm$ 0.016)	0.863 ( $\pm$ 0.014)	0.862 ( $\pm$ 0.015)	0.856 ( $\pm$ 0.015)	0.844 ( $\pm$ 0.015)	<b>0.878</b> ( $\pm$ <b>0.016</b> )
Glass	0.600 ( $\pm$ 0.031)	0.762 ( $\pm$ 0.027)	0.681 ( $\pm$ 0.036)	0.758 ( $\pm$ 0.023)	0.755 ( $\pm$ 0.027)	0.817 ( $\pm$ 0.018)	0.792 ( $\pm$ 0.031)	<b>0.823</b> ( $\pm$ <b>0.014</b> )
Letter	0.603 ( $\pm$ 0.016)	0.765 ( $\pm$ 0.026)	0.784 ( $\pm$ 0.026)	0.747 ( $\pm$ 0.030)	0.748 ( $\pm$ 0.031)	<b>0.799</b> ( $\pm$ <b>0.014</b> )	0.736 ( $\pm$ 0.017)	<b>0.799</b> ( $\pm$ <b>0.014</b> )
Lympho	0.906 ( $\pm$ 0.054)	0.938 ( $\pm$ 0.023)	0.972 ( $\pm$ 0.015)	0.941 ( $\pm$ 0.041)	0.941 ( $\pm$ 0.041)	<b>0.985</b> ( $\pm$ <b>0.009</b> )	0.973 ( $\pm$ 0.029)	0.984 ( $\pm$ 0.010)
Mammography	0.729 ( $\pm$ 0.007)	0.767 ( $\pm$ 0.011)	0.748 ( $\pm$ 0.010)	0.763 ( $\pm$ 0.013)	0.768 ( $\pm$ 0.018)	0.783 ( $\pm$ 0.015)	0.773 ( $\pm$ 0.019)	<b>0.786</b> ( $\pm$ <b>0.017</b> )
Mnist	0.817 ( $\pm$ 0.013)	0.830 ( $\pm$ 0.014)	0.822 ( $\pm$ 0.014)	0.831 ( $\pm$ 0.016)	0.831 ( $\pm$ 0.016)	0.837 ( $\pm$ 0.014)	0.794 ( $\pm$ 0.016)	<b>0.851</b> ( $\pm$ <b>0.012</b> )
Musk	0.779 ( $\pm$ 0.012)	0.821 ( $\pm$ 0.017)	0.779 ( $\pm$ 0.013)	0.818 ( $\pm$ 0.017)	0.817 ( $\pm$ 0.012)	0.797 ( $\pm$ 0.015)	0.797 ( $\pm$ 0.015)	<b>0.856</b> ( $\pm$ <b>0.016</b> )
Optdigits	0.629 ( $\pm$ 0.007)	0.711 ( $\pm$ 0.007)	0.686 ( $\pm$ 0.009)	0.702 ( $\pm$ 0.007)	0.702 ( $\pm$ 0.007)	0.734 ( $\pm$ 0.010)	0.700 ( $\pm$ 0.006)	<b>0.755</b> ( $\pm$ <b>0.011</b> )
Pendigits	0.937 ( $\pm$ 0.006)	0.941 ( $\pm$ 0.006)	0.937 ( $\pm$ 0.006)	0.943 ( $\pm$ 0.005)	0.943 ( $\pm$ 0.005)	0.940 ( $\pm$ 0.006)	0.937 ( $\pm$ 0.006)	<b>0.967</b> ( $\pm$ <b>0.003</b> )
Satellite	0.694 ( $\pm$ 0.015)	0.723 ( $\pm$ 0.009)	0.694 ( $\pm$ 0.016)	0.703 ( $\pm$ 0.011)	0.709 ( $\pm$ 0.013)	0.697 ( $\pm$ 0.016)	0.709 ( $\pm$ 0.017)	<b>0.747</b> ( $\pm$ <b>0.010</b> )
Thyroid	0.898 ( $\pm$ 0.009)	0.900 ( $\pm$ 0.011)	0.899 ( $\pm$ 0.008)	0.903 ( $\pm$ 0.013)	0.905 ( $\pm$ 0.012)	0.900 ( $\pm$ 0.009)	0.871 ( $\pm$ 0.025)	<b>0.909</b> ( $\pm$ <b>0.011</b> )
Vowels	0.735 ( $\pm$ 0.018)	0.839 ( $\pm$ 0.021)	0.880 ( $\pm$ 0.015)	0.842 ( $\pm$ 0.022)	0.837 ( $\pm$ 0.023)	0.938 ( $\pm$ 0.019)	0.835 ( $\pm$ 0.020)	<b>0.940</b> ( $\pm$ <b>0.015</b> )
WBC	0.810 ( $\pm$ 0.026)	0.839 ( $\pm$ 0.034)	0.822 ( $\pm$ 0.033)	0.852 ( $\pm$ 0.049)	0.844 ( $\pm$ 0.043)	0.874 ( $\pm$ 0.023)	0.828 ( $\pm$ 0.060)	<b>0.882</b> ( $\pm$ <b>0.023</b> )
Wine	0.581 ( $\pm$ 0.035)	0.737 ( $\pm$ 0.034)	0.690 ( $\pm$ 0.049)	0.728 ( $\pm$ 0.059)	0.743 ( $\pm$ 0.042)	0.798 ( $\pm$ 0.037)	0.713 ( $\pm$ 0.043)	<b>0.815</b> ( $\pm$ <b>0.043</b> )

exploration sampling performed best for most of the datasets. Furthermore, it appears that outlier sampling achieves a slow exploration for datasets with relatively many outliers like BreastW or Satellite due to our previous considerations. If only few outliers are present, its performance can be comparable to or even slightly better than exploration sampling. As expected, similarity sampling generally performed bad, because it leads to a low exploration. The remaining query strategies could not compete with exploration and outlier sampling either.

#### IV. CONCLUSION

We presented a new active learning method for one-class classification. It is able to take labeled outliers into account by employing a second OCC. Hence, our method can develop its full potential in applications in which some outliers are available and one-class classification is preferred over binary classification. The decision values of both OCCs are combined through an aggregation function which is chosen based on a desired trade-off between exploration and exploitation. We proposed two example aggregation functions which focus more on exploration and exploitation, respectively. The experiments showed that exploration sampling outperforms the other active learning strategies for almost all datasets after a practically relevant number of queries. Thus, exploration seems to be more important than exploitation in active learning for one-class classification. Future implications of this paper may include research on combining or applying other aggregation functions and using different hyperparameters for both OCCs. Finally, it is conceivable to apply the method to binary or multi-class classification.

#### REFERENCES

- [1] Shehroz S. Khan and Michael G. Madden, "One-class classification: Taxonomy of study and review of techniques," *CoRR*, vol. abs/1312.0049, 2013.
- [2] David M.J. Tax, *One-class classification*, PhD thesis, Delft University of Technology, Delft, 2001.
- [3] Burr Settles, "Active learning literature survey," Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.
- [4] David D. Lewis and William A. Gale, "A sequential algorithm for training text classifiers," in *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, 1994, SIGIR '94, pp. 3–12, Springer-Verlag New York, Inc.
- [5] Piotr Juszczak and Robert P. W. Duin, "Uncertainty sampling methods for one-class classifiers," in *In Proceedings of the ICML03 Workshop on Learning from Imbalanced Data Sets, 2003. Sigkdd Explorations. Volume 6, Issue 1 - Page 5*, 2003.
- [6] Vincent Barnabé-Lortie, Colin Bellinger, and Nathalie Japkowicz, "Active learning for one-class classification," in *14th IEEE International Conference on Machine Learning and Applications, ICMLA 2015, Miami, FL, USA, December 9-11, 2015*, 2015, pp. 390–395.
- [7] Alireza Ghasemi, Hamid R. Rabiee, Mohsen Fadaee, Mohammad T. Manzuri, and Mohammad H. Rohban, "Active learning from positive and unlabeled data," in *Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops*, Washington, DC, USA, 2011, ICDMW '11, pp. 244–250, IEEE Computer Society.
- [8] A. Ghasemi, M. T. Manzuri, H. R. Rabiee, M. H. Rohban, and S. Haghiri, "Active one-class learning by kernel density estimation," in *2011 IEEE International Workshop on Machine Learning for Signal Processing*, Sept 2011, pp. 1–6.
- [9] Piotr Juszczak, Robert P. W. Duin, and Marcel J.T. Reinders, *Learning to recognise: A study on one-class classification and active learning*, PhD thesis, Delft University of Technology, 2006.
- [10] Nico Görnitz, Marius Kloft, and Ulf Brefeld, *Active and Semi-supervised Data Domain Description*, pp. 407–422, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [11] Zhigang Liu and Zhichao Sun, "Active one-class classification of remote sensing image," 2008.
- [12] Mohsen Ghazel and Nathalie Japkowicz, *Improving Active Learning for One-Class Classification Using Dimensionality Reduction*, pp. 39–44, Springer International Publishing, Cham, 2017.
- [13] Bernhard Schölkopf, John C. Platt, John C. Shawe-Taylor, Alex J. Smola, and Robert C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, 2001.
- [14] B. Hadjadji, Y. Chibani, and Y. Guerbai, "Multiple one-class classifier combination for multi-class classification," in *2014 22nd International Conference on Pattern Recognition*, Aug 2014, pp. 2832–2837.
- [15] Bartosz Krawczyk, Michał Woźniak, and Francisco Herrera, "On the usefulness of one-class classifier ensembles for decomposition of multi-class problems," *Pattern Recogn.*, vol. 48, no. 12, pp. 3969–3982, Dec. 2015.
- [16] David M. J. Tax and Robert P. W. Duin, "Combining One-Class Classifiers," *Lecture Notes in Computer Science*, vol. 2096, pp. 299–308, 2001.
- [17] Shebuti Rayana, "ODDS library," 2016.
- [18] Alireza Ghasemi, "Alpud: Active learning from positive and unlabeled data," <https://github.com/aghasemi/alpud>, 2016.