What Can We Expect from Active Class Selection?[*]

Mirko Bunse[1] and Katharina Morik[1]

TU Dortmund, AI Group, 44221 Dortmund, Germany
{firstname.lastname}@tu-dortmund.de

Abstract. The promise of active class selection is that the proportions of classes can be optimized in newly acquired data. In this short paper, we take a step towards the identification of properties that data sets must meet in order to make active class selection (potentially) successful. Also, we compare the conceivable benefit of active class selection to that of active learning and we identify open research issues. It becomes apparent that active class selection is a tough task, in which informed strategies often exhibit only minor improvements over random sampling.

Keywords: Active class selection · Active learning · Classification.

1 Introduction

Active class selection (ACS) [8] seeks to optimize the proportions of classes in newly acquired data. This process is taken out sequentially: In each iteration, the most promising proportion of classes is selected and instances are generated according to these proportions. Due to this iterative collection of training data, there is a certain similarity between ACS and active learning (AL) [10]. However, the data acquisition is fundamentally different between these paradigms: Where AL selects un-labeled instances to be labeled, ACS selects classes for which new instances are to be generated.

This distinction reveals the contrasting assumptions which underlie AL and ACS with regard to the data generating process: AL assumes an external oracle which is able to assign labels to observations, e.g. a human annotator. ACS assumes a data generator which produces observations from label queries. One prominent example of such generator is the artificial nose experiment, where a vapor (the label) must be selected before a sensor array can record data [8]. In both cases, it is assumed that each query is costly. Therefore, ACS and AL try

---

to minimize the amount of training data by selecting only the most promising examples. Let us thus narrow the question raised above: Given that new training data can be generated from label queries, can we expect ACS to make optimal use of a limited data generation budget? Which preconditions must hold to make ACS a success? Our contribution with respect to these questions is three-fold:

- We identify common properties of the data used in ACS publications.
- We compare the potential benefit of ACS to that of AL.
- We recognize open issues in ACS research.

The first one of these contributions is detailed in Sec. 2. The second and third ones are presented in Sec. 3 and in Sec. 4. Finally, Sec. 5 concludes our findings.

2   Data Used in ACS

Despite the potential relevance of ACS, we could make out only two papers which suggest algorithms for this task. Lomasky et al. [8], who also introduced ACS, present five approaches, the most successful of which seek to stabilize the empirical error of each class. Kottke et al. [7] compare these approaches to a framework with which AL methods can be adapted to ACS. Namely, they use AL to score pseudo instances and aggregate the scores for each class. Both papers use random sampling (proportional and uniform) as a baseline.

Table 1. The winning ACS strategies for each evaluated data set.

data set	no. classes	no. examples	PAL-ACS [7]	Redistricting [8]	Improvement [8]	Inverse [8]	Proportional [8]	Uniform [8]
3clusters [7]	3	60 / ∞	✓	✗		✗	*	✗
spirals [7]	3	120 / ∞	✓	✗		✗	*	✗
bars [7]	3	120 / ∞	✓	✓		✓	*	✓
vehicle [4]	4	80 / 946	✓	✗		✗	*	✓
vertebral [4]	3	60 / 310	✓	✗		✗	*	✓
yeast [4]	5 / 8	60 / 1150	✓	✗		✗	*	✓
land cover [2]	11	≈ 28000		✓	✗	✗	✗	✗
artificial nose [8]	8	≈ 1250		✗	✗	✗	✓	✗

*Proportional $\equiv$ Uniform, due to a uniform class distribution in the test set

Tab. 1 summarizes the results that have been reported for these methods. The columns with upright names indicate whether a method clearly outperforms its competitors (✓) or not (✗). Missing values denote that a method has not been evaluated. Please consider that the qualification of a "winner" must remain somewhat subjective. We therefore declare multiple methods as winners wherever a single winner cannot be made out from the published plots and tables.

One observation to make is that the random strategies "proportional" and "uniform" perform highly competitive. In this overview, they win on five out of eight data sets. Moreover, they come for free, whereas the informed (i.e. non-random) strategies imply a certain computational overhead which needs to be justified with the data acquisition cost. Also, one may be concerned about the applicability of (informed) active sampling in general [1]. Note that proportional sampling assumes that the correct label proportions of the test set are known at training time, which may not hold in some use cases.

All of the data sets used so far distinguish between at least three classes. Moreover, we see that the predictability differs among their classes. The synthetic data sets for instance are modeled so that one class can easily be distinguished from the other two classes, which in turn are hard to distinguish from each other. For the UCI data sets [4], we provide the confusion matrices in Tab. 2. Displayed are the mean values over 50 trials, using proportional sampling and the classifier from the ACS experiments. Each row is scaled to unit sum to account for class imbalance. We see that the yeast data exhibits large differences among class difficulties (78.7% vs 41.4% class-wise accuracy). The differences on the vertebral data are smaller, yet considerable (74.0% vs 56.1%).

Table 2. Confusion matrices of the Parzen window classifier [3].

	predicted class				
	0.526	0.348	0.007	0.052	0.066
	0.431	0.414	0.006	0.051	0.098
truth	0.002	0.002	0.787	0.2	0.008
	0.054	0.051	0.306	0.463	0.126
	0.077	0.13	0.028	0.105	0.66

yeast data set [4]

	predicted class		
	0.561	0.308	0.131
truth	0.268	0.691	0.041
	0.161	0.098	0.74

vertebral data set [4]

3 The Potential Benefits of ACS and AL

Given a data set consisting of at least three classes of varying difficulty, what is the improvement that we can expect from ACS? How does it relate to the improvement AL methods achieve? To answer these questions, we reproduce some of the experiments described in [7]. We add one strategy to these experiments that is optimal for the spirals data—it uses only a single example from the easy class and randomly samples from the difficult classes. It is "optimal" with regard to the overall accuracy because a single example is already enough to achieve 100% accuracy on the easy class. Even though this strategy does not adapt to any other data set, it shows how well ACS could potentially perform. Moreover, we extend the experiments by evaluating an AL strategy, namely the probabilistic active learning (PAL) [6] which is also used inside of PAL-ACS. Fig. 1 presents the results of these extensions, specifically the mean error over 500 trials.
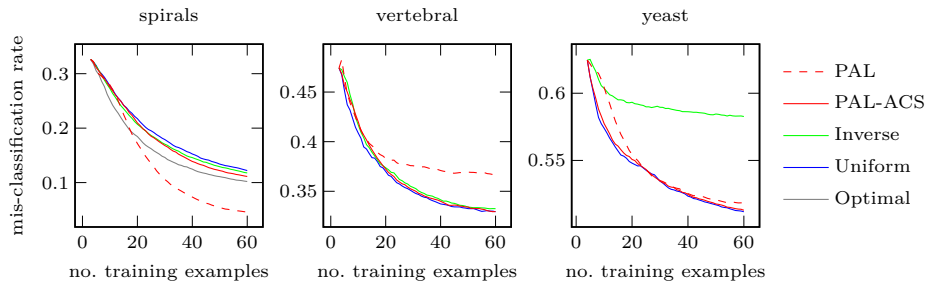
Fig. 1. The learning curves of ACS strategies and the AL strategy PAL in comparison.

The optimal strategy indicates that there is still room for improving ACS methods. In particular, knowing the difficulty of the classes in advance allows us to outperform the other strategies on the spirals data set. However, PAL is even better than that. Knowing which examples are available thus allows us to improve even further. These observations can not be made on the two UCI data sets. On both of them, neither uniform sampling nor PAL-ACS are clear winners—a finding we deem consistent with the original experiments. What is probably surprising is that the AL strategy performs worse than ACS. We conjecture that the identification of relevant examples is not necessarily easier than, but considerably different from the identification of relevant classes.

4   Open Issues in ACS

It remains open whether the current limits of (informed) ACS stem from the problem itself—i.e. from sequentially optimizing only the class proportions— or from the methods proposed to date. We suggest to approach this question by studying relaxations of "pure" ACS. Indeed, example generators are often not only controlled by class proportions but also from auxiliary parameters. In the artificial nose experiment, for instance, not only a vapor (the label) must be selected before data can be recorded, but also the vapor's concentration [9]. Optimizing the data generation only with respect to the class proportions means to limit the actual task artificially—and maybe even detrimentally.

An issue that has been neglected in ACS so far is the problem of imbalanced data [5]. This problem refers to situations in which one class is abound and another one is scarce, typically leading to the degradation of classifiers and evaluation metrics. It has also been argued that within-class imbalances, i.e. abound and scarce sub-groups of single classes, can hinder learning [11]. In ACS, we are free to choose how balanced the data is, but only with respect to the label. Methods for imbalanced learning could therefore guide ACS by constraining the class proportions for between-class balance and they may also correct the effects of within-class imbalances.

5 Conclusion

ACS addresses use cases which distinguish between at least three classes of varying predictability. However, this precondition does not necessarily lead to a successful application of ACS. Experiments suggest that a random sampling of classes is hard to beat with informed strategies. We expect future advances to be made by (i) queries which combine the label with auxiliary parameters that control the data generator and by (ii) accounting for data imbalances.

References

1. Attenberg, J., Provost, F.J.: Inactive learning? Difficulties employing active learning in practice. SIGKDD Explorations **12**(2), 36–41 (2010)
2. Brodley, C.E., Friedl, M.A.: Improving automated land cover mapping by identifying and eliminating mislabeled observations from training data. In: Int. Geoscience and Remote Sensing Symp. vol. 2, pp. 1382–1384. Citeseer (1996)
3. Chapelle, O.: Active learning for Parzen window classifier. In: Proc. of the AISTATS 2005. Society for Artificial Intelligence and Statistics (2005)
4. Dua, D., Graff, C.: UCI machine learning repository (2017), `http://archive.ics.uci.edu/ml`
5. Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B., Herrera, F.: Learning from Imbalanced Data Sets. Springer (2018)
6. Kottke, D., Krempl, G., Lang, D., Teschner, J., Spiliopoulou, M.: Multi-class probabilistic active learning. In: Proc. of the ECAI 2016. Frontiers in Artificial Intelligence and Applications, vol. 285, pp. 586–594. IOS Press (2016)
7. Kottke, D., Krempl, G., Stecklina, M., von Rekowski, C.S., Sabsch, T., Minh, T.P., Deliano, M., Spiliopoulou, M., Sick, B.: Probabilistic active learning for active class selection. In: Proc. of the NIPS Workshop on the Future of Interactive Learning Machines (2016)
8. Lomasky, R., Brodley, C.E., Aernecke, M., Walt, D., Friedl, M.A.: Active class selection. In: Proc. of the ECML 2007. LNCS, vol. 4701, pp. 640–647. Springer (2007)
9. Rodriguez-Lujan, I., Fonollosa, J., Vergara, A., Homer, M., Huerta, R.: On the calibration of sensor arrays for pattern recognition using the minimal number of experiments. Chemometrics and Intelligent Laboratory Systems **130**, 123–134 (2014)
10. Settles, B.: Active Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool Publishers (2012)
11. Weiss, G.M.: Mining with rarity: A unifying framework. SIGKDD Explorations **6**(1), 7–19 (2004)