Active Learning for kNN Using Instance Impact

Sayed Waleed Qayyumi(✉), Laurence A. F. Park, and Oliver Obst

Centre for Research in Mathematics and Data Science School of Computer, Data and Mathematical Sciences, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia
{s.qayyumi,l.park,o.obst}@westernsydney.edu.au
https://www.westernsydney.edu.au/crmds/

Abstract. Labelling unlabeled data is a time-consuming and expensive process. Labelling initiatives should select samples that are likely to enhance the classification accuracy of the classifier. Several methods can be employed to accomplish this goal. One of these techniques is to select samples with the highest level of uncertainty in their predicted labels. Experts then label these samples. Another option is to choose samples at random. This paper proposes three methods for identifying unlabeled samples to improve predictive accuracy when they are labelled. Our study explores how to select samples when we have very few labelled samples available from manifold distributed data sets. In order to assess performance, we have compared our approaches with uncertainty sampling and random sampling. We demonstrate that our methods outperform uncertainty sampling and random sampling by using public and real-world data sets.

Keywords: Active learning · Uncertainty sampling · Unlabelled sampling · Random sampling · Incremental learning · Few shot learning · Entropy · Uncertain labels

1 Introduction

To classify complex tasks, supervised machine learning models can be used to learn complex relationships between queries and responses. For instance, machine learning models can help detect tumours at an early stage. Upon finding a tumour by these models, a specialist can examine it further. Training data must contain queries and responses created by or evaluated by specialists to train models for specialized purposes. Therefore, such data can be challenging to obtain. Many queries are available (e.g. image scans, feature vectors, videos), yet, it is hard to receive accurate responses to each of these queries. In the case of specialized data, we must hire a specialist to examine each query and provide a response. A specialist must take time to do this, which is costly for both the data modeller and the specialist. If labelling is cost-prohibitive, a smaller sample of queries is forwarded to a specialist. The selection of the samples is either accomplished randomly or using uncertainty sampling.

H. Aziz et al. (Eds.): AI 2022, LNAI 13728, pp. 413–426, 2022.
https://doi.org/10.1007/978-3-031-22695-3_29

Table 1. Sampling/Labelling scenarios. With a large sample that is difficult to label, we resort to labelling a random sample, but is that best approach?

		Sampling	
		1: Simple	2: Difficult
Labelling	A: Simple	Label all	Label all
	B: Difficult	Label Random	Label all

This paper discusses how to sample queries for manual labelling to improve the accuracy of the machine learning models. In our research, we generally focus on small sample sizes (e.g., the few shot learning scenario and manifold distributed data). The article will proceed as follows: Sect. 2 discusses the current state of the art in the selection of the next best-unlabelled sample. Section 3 examines our approaches to the next best sample selection. Section 4 presents the results of experiments conducted on different public and real-life data sets in a few-shot learning and semi-supervised learning scenario. This section also compares our sampling techniques with active learning's uncertainty sampling and random sampling. Section 4.3 contains a list of our observations. Section 5 concludes this paper and discusses our future work.

2 Background and Related Work

In attempting to classify manifold distributed data with very few label samples, we investigated the problem of finding the best-unlabelled sample for labelling. In classifier training, there are four scenarios regarding data availability. Table 1 lists all these scenarios. This article discusses scenario 1B (easy to sample and difficult to label). In this scenario, we have to label more samples to achieve higher accuracy in classification. Labelling is a complex and costly endeavour, so choosing the right unlabelled sample is crucial. Imagine, for example, one million CT scans with only ten labels. To improve the accuracy of your classifier, you need to label another ten items. Choosing a sample that increases the accuracy of the classifier is crucial in such a scenario. This is the focus of our sample selection methods.

Active learning is the process of selecting an optimal unlabelled sample from a pool of unlabelled data. Unlabeled data is classified with a classifier, and then the observations with the most uncertain labels are identified. This process is known as uncertainty sampling. There are many methods that are available to estimate the uncertainty of a labelled sample. The active learning process consists of querying an information source, for example, an Oracle, to assign a new label to a data point. This algorithm attempts to choose the best possible sample to be labelled [16,18]. The term optimal experimental design can also refer to active learning in statistics. In situations, unlabeled data is readily available, but its labelling is costly. When such a scenario occurs, a learning algorithm can aid in identifying samples for labeling. This process is known as active learning. Choosing examples that the learner finds meaningful is generally more effective, which results in fewer examples needed than is necessary for supervised learning. Recent advances in active learning include multi-label active learning [24] and hybrid active learning [11]. These research areas combine machine learning concepts with incremental learning policies. There are three different scenarios or settings in which learners typically query instances' labels.

- The learner generates instances based on the underlying distribution in the membership query synthesis.
- In stream-based sampling, the assumption is that unlabeled samples are free to obtain. Thus, each unlabelled sample is selected one at a time. Upon reading an unlabelled instance, the learner can decide whether to query or reject. Acceptance or rejection of the instance is driven by its informativeness. A query strategy determines how informative the sample is.
- Pool-based sampling is based on the assumption that there is a large pool of unlabeled data. An informativeness measure can be applied to all samples in the pool to identify the best candidates for labeling. The proposed sampling methods described in this paper can also be referred to as pool-based sampling techniques.

The learner can utilize a variety of measures to identify the most appropriate sample. An example of one of these measures is uncertainty. The learner labels all unlabeled data using the available labelled data. Upon determining the uncertainty of each predicted label, the sample with the most uncertain label is selected and sent to Oracle for labeling. The following are three commonly used approaches to querying instances based on uncertainty sampling.

- Least Confidence: LC strategies let learners select the instance for which the learner is least confident in its most likely label.

$$U(x) = 1 - P(\hat{x}|x) \tag{1}$$

- Margin Sampling: A fundamental problem with the LC strategy is that it only considers the most probable label and disregards the other label probabilities. For this reason, the margin sampling strategy selects the instance with the minimal difference between the first and second most probable labels.

$$M(x) = P(\hat{x}_1|x) - P(\hat{x}_2|x) \tag{2}$$

- Highest Entropy: All the potential label probabilities can be computed using entropy. All instances are analyzed by calculating the entropy value of each instance and querying the instance with the highest value.

$$H(x) = -\sum_k p_k \log(p_k) \tag{3}$$

It is important to note that uncertainty sampling is dependent on predicted labels. In addition, calculating uncertainty is not straightforward for all classification methods. It is not easy, for example, to calculate the uncertainty in neural network setup [23].

Sampling plays a significant role in classifier training. In order to improve prediction accuracy, it is necessary to train a classifier with sufficient training data. Sample collection can help provide the necessary data. You can find a detailed description of most of the sampling techniques in Altmann et al. [1] and Etikan et al. [5]. In random sampling [15], each sample has an equal chance of being selected. A stratified random sampling method [14] involves dividing the population into subgroups called strata and selecting samples at random from each stratum. A systematic sample selection method, [12] is based on choosing a fixed interval and starting point. After establishing a starting point, subsequently, samples can be collected at regular intervals. Clustered sampling [6] allows drawing samples at random from some of the clusters. Clustered sampling draws samples from random groups, whereas stratified sampling selects samples from each stratum or group, allowing us to exclude entire groups from the study. The convenience sampling method [17] involves selecting a sample solely on the basis of its convenience for sampling purposes. Quota sampling [13] selects samples based on specific characteristics. There is also snowball sampling [8], which selects a sample based on the judgment of the experts who need it, and then uses it to select subsequent samples. Sampling methods are bound to be biased. A number of methods have been proposed to address bias in sampling [10,21]. All samples need not come from the same distribution. They may even come from a distribution similar to the one we study. If the main distribution is unavailable, importance sampling [22] is applied. In this scenario, we sample from another distribution by adjusting the weights of the distribution so that it represents the desired distribution. We can use information gain to select the samples. The information gain is the amount of entropy removed from the data set by splitting it. Therefore, a split with a higher information gain [2] is preferred.

Data samples are collected before classification models are built and trained. It is possible to construct classification models if enough data is available. When we do not have enough data, we can continually improve our classification models by retraining additional labelled data. During retraining, newly acquired labelled data is incorporated into the learning process. The method of learning is called incremental learning [7]. It is possible to apply several traditional classification methods to incremental learning [20]. In incremental learning, the goal is to acquire new knowledge based on new data without forgetting the existing knowledge derived from older data. The next best action recommendation is a popular marketing technique designed to retain customers. In order to determine what the best next step for a given customer is, it is necessary to compare their profile to a similar customer model [9]. Reinforcement learning determines the next best task based on this approach [4]. In a similar fashion to incremental learning, the next best task has been an active area of research [3].

Aside from uncertainty sampling, our work is also comparable to Transductive Semi-supervised Deep Learning (TSSDL) and Personalized next-best-action recommendation [3,19]. In particular, we discuss the topics of sampling, entropy, incremental learning, and recommendation of the next best task. We will discuss the relevance of these topics after providing a brief overview of these topics. We do not estimate labels for the unlabeled samples, but rather rank all unlabeled samples according to their potential influence on classification accuracy.

3 Measuring the Utility of an Instance for Training

As a rule of thumb, the performance of a classification model is contingent on how well the training data represent the population to be classified. Therefore, it is imperative to select a sufficient number of instances from the population of interest for manual labelling and inclusion in the training set. To accomplish this, we must be able to select the most appropriate sample of the population, and then manually label each instance within that sample. We should take as large a sample as possible if the labelling of each instance is a straightforward process. Furthermore, we should attempt to label all the observations in our training set. When both manual labelling and sampling are time-consuming and costly, fewer samples can be collected, and all instances will have to be labelled. There is a question regarding how to proceed when we have access to a large pool of unlabelled data but cannot label each instance. Therefore, we must determine which subset of that sample should be manually labelled and added to the training set. Our goal should be to select observations that will produce the highest increase in classification accuracy when used for training. However, the question remains as to how to choose the instances.

This section examines three candidate functions for evaluating the utility of including an unlabeled instance in a training set. Each instance is assigned a score based on its potential to influence classification accuracy. The article focuses on data with a relatively small training set (only a few cases were manually labelled), and which are manifold distributed. Thus, we use k Nearest Neighbors (kNN) as a classifier. Please also note that we have assumed that sampling from the population will be relatively straightforward, whereas labelling will be more complex. Thus, we can also safely presume that we have a large pool of candidate instances from which to choose and that it is possible to assess the utility of a selected instance; we refer to this pool of available data as the "test set". We have described our proposed methods in the following three subsections.

3.1 Neighbourhood Impact

In order for a new instance to have the potential to increase accuracy, it has to play a role in the classification of newly created instances. The training instance is only relevant if it is the nearest neighbour of the test instance in the kNN classification. Therefore, one measure of the utility of a candidate training instance is the number of data points it is closest to.

Given a set of manually labelled instances $\mathcal{X}$ and a set of unlabelled instances $\mathcal{U}$, let $N_k(u;\mathcal{X}) \subset \mathcal{X}$ be the set of k nearest neighbours of u chosen from the manually labelled set of instances, where $u \in \mathcal{U}$ and $|\mathcal{X}| > k$. We define the *Neighbourhood Impact* I of labelled instance x as

$$I(x) = \sum_{u \in \mathcal{U}} \mathbf{1}_{N_k(u;\mathcal{X})}(x). \tag{4}$$

for $x \in \mathcal{X}$, where $\mathbf{1}_A(x)$ is the indicator function ($\mathbf{1}_A(x) = 1$ if $x \in A$ or 0 otherwise).

To measure the neighbourhood impact of an unlabelled instance u, we must remove the instance from the set $\mathcal{U}$ to obtain $\mathcal{U} \setminus u$, and append it to the set of labelled instances $\{\mathcal{X}, u\}$. The neighbourhood impact for an unlabelled instance is

$$I(u) = \sum_{v \in \mathcal{U} \setminus u} \mathbf{1}_{N_k(v;\{\mathcal{X},u\})}(u). \tag{5}$$

Including an unlabelled instance in the training set will not cause the trained model's classification accuracy to improve if $I(u) = 0$. Including an unlabelled instance with a high $I(u)$ will affect the model's classification accuracy when included in the training set. The hypothesis is that if an unlabelled instance has a high $I(u)$ value, then manually labelling it and adding it to the training set will improve its accuracy.

3.2 Maximum Entropy

A neighbourhood impact refers to the potential of an instance based on its proximity to a neighbourhood. Furthermore, it is possible to examine whether the point may be able to alter the unlabelled class prediction. By calculating the entropy of the label distribution one can determine how robust the prediction is when there is a set of training labels. As a result, the notion of high entropy implies that one change in an instance label might alter a prediction, whereas the notion of low entropy requires many changes in order to change a prediction.

The class prediction for test instance u is the mode class of the set of k nearest neighbours from the labelled set $\mathcal{X}$. We define $L_k(u;\mathcal{X})$ as the set of class labels associated to the training instances $N_k(u;\mathcal{X})$. Using this, the predicted class label for instance u is $\text{mode}\,(L_k(u;\mathcal{X}))$ and the entropy of the neighbourhood distribution is $\text{Ent}\,(L_k(u;\mathcal{X}))$.

This potential for an unlabelled instance u to influence the class prediction is expressed by *Maximum Entropy.* Essentially, this can be defined as the maximum class distribution entropy if the example was included in the training set with a class label. The maximum entropy $H(u)$ of an unlabelled instance u can be defined as

$$H(u) = \max_{l_u \in \mathcal{L}} \sum_{v \in \mathcal{U} \setminus u} \text{Ent}\,(L_k(v;\{\mathcal{X},u\})). \tag{6}$$

where u is the candidate unlabelled instance, $\mathcal{U} \setminus u$ is the unlabelled set with the candidate instance removed, l_u is the label of the candidate instance, $\mathcal{L}$ is the set of all possible class labels and $\mathrm{Ent}(X)$ is the entropy of the categorical distribution X.

3.3 Delta in Prediction

It would be ideal if we could identify which of the unlabelled instances would be suitable for labelling and inclusion in the training set. The ideal training example is the one that offers the highest accuracy. Due to the lack of labelling, we cannot examine the increase in accuracy for each candidate instance.

Instead of measuring the increase in accuracy, we can instead measure the potential increase in accuracy. The classification accuracy for kNN using training set $\mathcal{X}$ and testing set $\mathcal{U}$ is

$$\frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} \mathbf{1}\left(\mathrm{mode}\left(L_k(u;\mathcal{X})\right) = l_u\right). \tag{7}$$

where $\mathbf{1}(A)$ is an indicator function (providing 1 if A is true and 0 is A is false) and l_u is the class label of instance u.

We define the *Delta in Prediction* of labelled instance x as

$$\Delta(u) = \max_{l_u\in\mathcal{L}} \frac{1}{|\mathcal{U}\setminus u|}\sum_{v\in\mathcal{U}\setminus u} \mathbf{1}\left(\mathrm{mode}\left(L_k(v;\{\mathcal{X},u\})\right) = l_v\right). \tag{8}$$

where u is the candidate unlabelled instance, $\mathcal{U} \setminus u$ is the unlabelled set with the candidate instance removed, l_u is the label of the candidate instance, and $\mathcal{L}$ is the set of all possible class labels. Thus, the $\Delta(u)$ represents the maximum classification accuracy that may be obtained by including u in the training set, concerning all class labels.

In this paper, we examine the relationship between each of these functions' scores and the accuracy of classification when choosing the associated instance.

4 Experimental Setup

We have only a small training set and wish to add to it. But manual labelling is challenging, so we should choose carefully when selecting which unlabeled instances are to be labelled. This study aims to answer the question: "Does the use of instance selection functions to refine sample selection result in better accuracy than random and uncertainty based selection?". We empirically investigate this question using the data from the UCI repository.

In each run of the experiment, we follow the steps below. A random sample of instances from a given data set is chosen as the training set containing manually assigned labels. The remainder of the instances are left unlabeled. Every unlabelled observation is assigned a selection score, and the sample with the highest

Table 2. Data used for evaluating instance selection functions.

Dataset	No of classes	Characteristics	Instances	Attributes	Features
Banknotes	2	Multivariate	13,72	Real	5
Satlog	6	Multivariate	6,435	Integer	36
Segmentation	7	Multivariate	2,310	Real	19
Heart disease	5	Multivariate	303	Real	14
Diabetes	2	Multivariate, Time-series	768	Real	9
Pendigits	10	Multivariate	10,992	Integer	16

Table 3. Comparison of average classification accuracy of random and uncertainty sampling with all our methods - average of 100 iterations

Dataset	Rand samp	Uncert samp	I	Δ	H	$I\Delta$	IH	ΔH	$I\Delta H$
Banknotes	0.59	0.60	$0.61^{\ddagger}$	0.56	0.58	0.56	$0.61^{\ddagger}$	$0.60^{\dagger}$	$0.61^{\ddagger}$
Satlog	0.45	0.45	$0.46^{\ddagger}$	$0.46^{\ddagger}$	$0.46^{\dagger}$	$0.46^{\dagger}$	$0.46^{\ddagger}$	$0.46^{\ddagger}$	$0.46^{\ddagger}$
Segmentation	0.35	0.36	$0.36^{\dagger}$	$0.35^{\star}$	$0.36^{\dagger}$	0.35	$0.36^{\ddagger}$	0.35^{*}	$0.36^{\ddagger}$
Heart disease	0.46	0.45	0.44	$0.48^{\ddagger}$	0.45	$0.48^{\ddagger}$	0.44	0.45	$0.48^{\ddagger}$
Diabetes	0.60	0.61	$0.63^{\ddagger}$	$0.63^{\ddagger}$	$0.64^{\ddagger}$	$0.63^{\ddagger}$	$0.63^{\ddagger}$	$0.63^{\ddagger}$	$0.63^{\ddagger}$
Pendigits	0.38	0.37	$0.39^{\dagger}$	0.38	0.38	0.37	$0.39^{\dagger}$	0.38	$0.39^{\dagger}$

Signif. codes: $\star$: $p < 0.05$, †: $p < 0.01$, ‡: $p < 0.001$.

score is added to the labelled training set. kNN accuracy is determined before and after the new point has been added to the training set.

The experiment variables are: the candidate instance selection functions {Random selection, Neighbourhood Impact, Maximum Entropy, Delta in Prediction}, the data (shown in Table 2), the initial training set size {4, 8, 16, 32, 64, 128}, and the number of instances chosen. Initial analysis shown in Fig. 1 showed that high accuracy instances are those that provide more central scores, so we selected the instance that provided the score closest to the mean score from all observations, to include in our training set. We also expanded the candidate instance selection function set to include the sum of each combination of the three candidate function scores. The selection methods are shown in the results as: Random sampling(adding a randomly chosen instance), uncertainty sampling, Δ (Delta in prediction), H (Maximum entropy), and I (Neighbourhood impact). Whenever two or more methods have been combined, the scores for the respective methods have been added.

4.1 Choosing One Instance

In the first experiment, we examine the results of selecting one instance from the unlabeled set to be manually labelled using a randomly chosen training set size of four. The experiment is paired, i.e. each method employs the same random

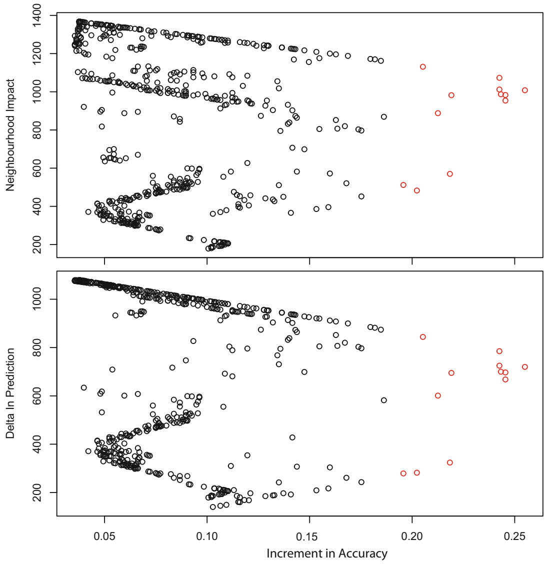

Fig. 1. Relationship of higher accuracy, neighbourhood impact and delta in prediction - Banknotes dataset. The figure shows that high accuracy is related to mean neighbourhood impact and delta in prediction

training sets. Figure 2 provides the accuracy of each method based on 100 runs, where the results are sorted by uncertainty sampling. Each point on the figure represents the average prediction accuracy after one hundred iterations. The graph has exactly 100 points, so each line represents ten thousand executions. The proposed selection technique performs much better than the benchmarks, namely uncertainty sampling and random selection. In light of our experimental findings, and the above demonstrations, we find that our proposed techniques perform very well in a few-shot learning environment. The p values of all techniques are compared in Table 3. This table also presents the average accuracy of each technique for different data sets in comparison with random sampling and uncertainty sampling.

4.2 Choosing n Instances

As the number of labelled samples increases, i.e. as we move from a few shot learning scenario to a semi-supervised learning scenario, Fig. 3 illustrates the average accuracy for Random, Oracle, uncertainty sampling and our methods. For all methods, accuracy is based on sample sizes of $4, 8, 16, 32, 64, 128$. The figure illustrates that our approaches are more accurate than random selection when the labelled sample size is small while at par with random selection as the labelled sample size increases. Our study found that as the number of labelled samples per class exceeds 32, the accuracy of selecting new unlabelled samples remains the same for all methods, including random sampling. Depending on the

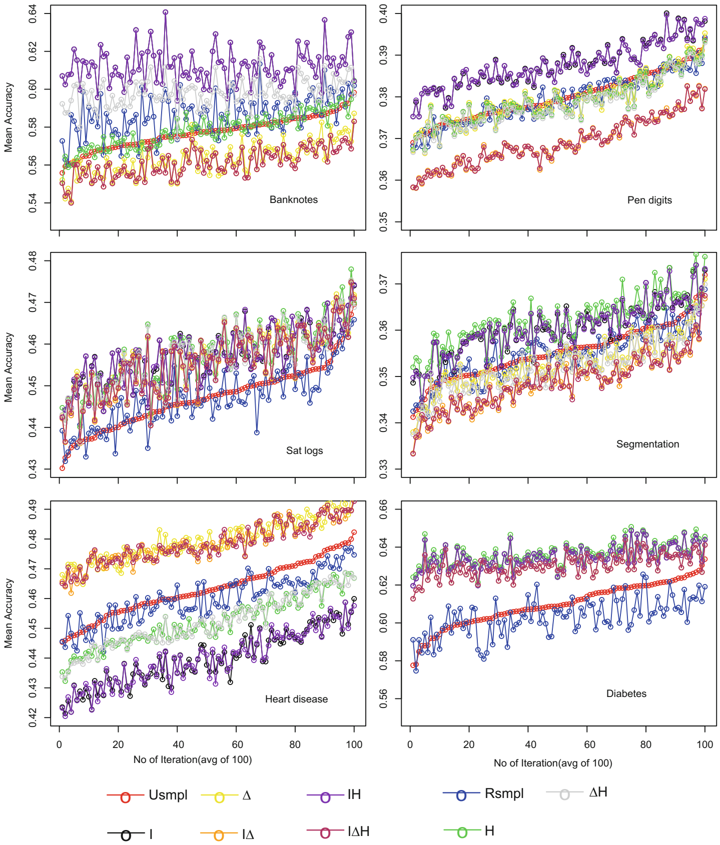

Fig. 2. Mean accuracy of classification – next best 1 sample selection – each point represent 100 executions

data set, this saturation point may vary, but it is typically close to 30 samples per class.

These techniques have been tested in a number of situations, including the next-best 1, n unlabelled samples and n labelled samples. Our results demonstrate that these techniques are on par with random selection in the next-best 1 and n unlabelled sample selection setting. Please refer to Fig. 4 that shows the results of a banknotes data set using the next-best $3, 5, 7$, and 9 unlabelled samples. Next-best 3 unlabelled sample setting is one where three 3 unlabelled samples are selected to compare their accuracy.

4.3 Semi-supervised Learning Scenario

In a setting with many labelled samples, we observed similar results. This test aimed to assess performance in a semi-supervised setting. In Fig. 3, we compare

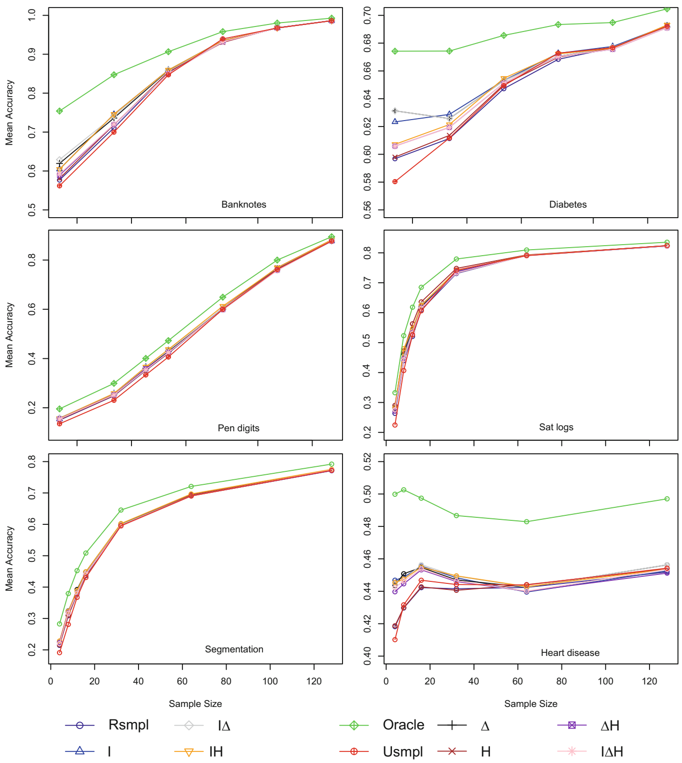

Fig. 3. Comparing the average accuracy (average of 100) for Oracle, Random, and our methods as the number of labelled samples increases

uncertainty sampling, random selection, and all of the techniques we propose using all of the data sets. A comparison is made between the average accuracy as we move from a small labelled data set (few-shot learning) to a bigger labelled data set (semi-supervised learning). The number of available samples doubles with each stage. This analysis shows that our approach performs better with 4,8,16, and 32 observations. Although our method outperforms when only a few labelled samples are available, it is still competitive when many labelled samples are available. Table 3, we present the results of ten thousand computations and compare uncertainty sampling and random with the proposed methods.

Based on our experiments conducted in few shot and semi-supervised settings, the following results were observed.

1. A systematic selection of an unlabeled sample is preferable when small labeled samples are available for the training of the classifier.
2. The methodologies we propose for systematic selection can also be applied in a semi-supervised setting where a large number of labels are available.

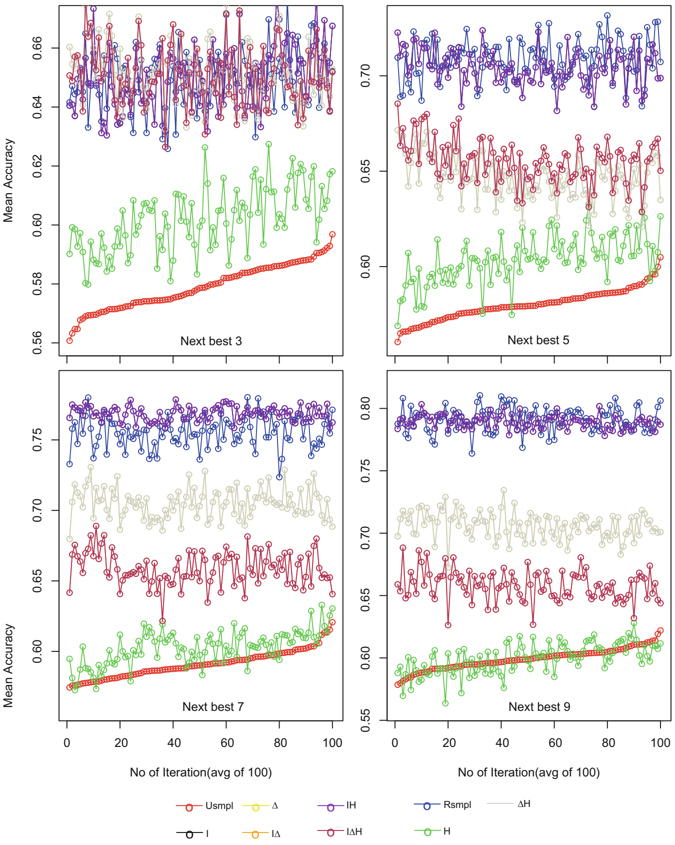

Fig. 4. Comparison of various methods for next best 3, 5, 7 and 9 sample selection – Banknotes data set

The cost of systematic selection in this case is higher than that of random selection.

3. There is a saturation point in terms of the number of samples that have been labeled. There are no differences between uncertainty sampling, random sampling and systematic selection beyond this point.

5 Conclusions

We present three novel approaches to selecting a good next sample in few-shot and semi-supervised learning situations. We evaluate our proposed methods using random sampling and uncertainty sampling as benchmarks. Performance is evaluated by comparing the accuracy of classification before and after including the selected samples in the training set. Our evaluation of real-life, publicly available data sets shows that our proposed sampling methods are preferable to uncertain sampling and random sampling when there are only a few labelled samples available. Furthermore, our method performs as well as the benchmarks when there are a lot of labelled samples.

References

1. Altmann, J.: Observational study of behavior: sampling methods. Behaviour **49**(3–4), 227–266 (1974)
2. Bestmann, S., et al.: Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. Curr. Biol. **18**(10), 775–780 (2008)
3. Cao, L., Zhu, C.: Personalized next-best action recommendation with multi-party interaction learning for automated decision-making. arXiv preprint arXiv:2108.08846 (2021)
4. Dunn, E., Frahm, J.M.: Next best view planning for active model improvement. In: BMVC, pp. 1–11 (2009)
5. Etikan, I., Bala, K.: Sampling and sampling methods. Biomet. Biostatist. Int. J. **5**(6), 00149 (2017)
6. Fraboni, Y., Vidal, R., Kameni, L., Lorenzi, M.: Clustered sampling: low-variance and improved representativity for clients selection in federated learning. arXiv preprint arXiv:2105.05883 (2021)
7. Giraud-Carrier, C.: A note on the utility of incremental learning. AI Commun. **13**(4), 215–223 (2000)
8. Goodman, L.A.: Snowball sampling. Ann. Math. Statist. **32**, 148–170 (1961)
9. Jenkinson, A.: What happened to strategic segmentation? J. Direct Data Digit. Mark. Pract. **11**(2), 124–139 (2009)
10. Kramer-Schadt, S., et al.: The importance of correcting for sampling bias in maxent species distribution models. Divers. Distrib. **19**(11), 1366–1379 (2013)
11. Lughofer, E.: Hybrid active learning for reducing the annotation effort of operators in classification systems. Pattern Recogn. **45**(2), 884–896 (2012)
12. Madow, W.G., Madow, L.H.: On the theory of systematic sampling, I. Ann. Math. Stat. **15**(1), 1–24 (1944)
13. Moser, C.A.: Quota sampling. J. R. Statist. Soc. Ser. A (General) **115**(3), 411–423 (1952)
14. Neyman, J.: On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. In: Kotz, S., Johnson, N.L. (eds.) Breakthroughs in Statistics, pp. 123–150. Springer Series in Statistics. Springer, New York, NY (1992). https://doi.org/10.1007/978-1-4612-4380-9_12
15. Olken, F.: Random sampling from databases. Ph.D. thesis, University of California, Berkeley (1993)
16. Rubens, N., Kaplan, D., Sugiyama, M.: Active learning in recommender systems. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 735–767. Springer, Boston (2011). https://doi.org/10.1007/978-0-387-85820-3_23
17. Sedgwick, P.: Convenience sampling. BMJ. **347**, 1–2 (2013)
18. Settles, B.: Active learning literature survey (2009)
19. Shi, W., Gong, Y., Ding, C., Ma, Z., Tao, X., Zheng, N.: Transductive semi-supervised deep learning using min-max features. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11209, pp. 311–327. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01228-1_19

20. Syed, N.A., Liu, H., Sung, K.K.: Incremental learning with support vector machines (1999)
21. Syfert, M.M., Smith, M.J., Coomes, D.A.: The effects of sampling bias and model complexity on the predictive performance of maxent species distribution models. PLoS ONE **8**(2), e55158 (2013)
22. Tokdar, S.T., Kass, R.E.: Importance sampling: a review. Wiley Interdiscipl. Rev. Comput. Statist. **2**(1), 54–60 (2010)
23. Van Amersfoort, J., Smith, L., Teh, Y.W., Gal, Y.: Uncertainty estimation using a single deep deterministic neural network. In: International Conference on Machine Learning, pp. 9690–9700. PMLR (2020)
24. Yang, B., Sun, J.T., Wang, T., Chen, Z.: Effective multi-label active learning for text classification. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 917–926 (2009)