A More Robust Baseline for Active Learning by Injecting Randomness to Uncertainty Sampling

Po-Yi Lu [1] **Chun-Liang Li** [2] **Hsuan-Tien Lin** [1]

Abstract

Active learning is important for human-computer interaction in the domain of machine learning. It strategically selects important unlabeled examples that need human annotation, reducing the labeling workload. One strong baseline strategy for active learning is uncertainty sampling, which determines importance by model uncertainty. Nevertheless, uncertainty sampling sometimes fails to outperform random sampling, thus not achieving the fundamental goal of active learning. To address this, the work investigates a simple yet overlooked remedy: injecting some randomness into uncertainty sampling. The remedy rescues uncertainty sampling from failure cases while maintaining its effectiveness in success cases. Our analysis reveals how the remedy balances the bias in the original uncertainty sampling with a small variance. Furthermore, we empirically demonstrate that injecting a mere 10% of randomness achieves competitive performance across many benchmark datasets. The findings suggest randomness-injected uncertainty sampling can serve as a more robust baseline and a preferred choice for active learning practitioners.

1. Introduction

In the intersection between Artificial Intelligence (AI) and Human-Computer Interaction (HCI), active learning continues to be a powerful tool. Active learning aims to reduce the labeling burden on humans by having computers strategically select the most crucial unlabeled examples that require human annotation. Consequently, the reduced burden can enhance the efficiency of deploying AI models in real-world applications where collecting human annotations can be expensive.

[1]National Taiwan University [2]Google Cloud Research. Correspondence to: Hsuan-Tien Lin <htlin@csie.ntu.edu.tw>.

AI & HCI Workshop at the 40^{th} International Conference on Machine Learning (ICML), Honolulu, Hawaii, USA. 2023.

Active learning contrasts with passive learning, or random sampling, which randomly selects examples for annotation. The efficiency of active learning is thus often evaluated with its performance gain over random sampling under the same amount of annotation budget. A strong active learning strategy is expected to achieve a positive gain often. One such strategy, uncertainty sampling, attempts to annotate the most uncertain example to the current model. It is a simple and effective strategy as shown by several benchmark studies (Yang & Loog, 2018; Karamcheti et al., 2021; Ji et al., 2023), with a solid physical intuition of reducing the model's uncertainty through annotation. The simplicity and effectiveness make uncertainty sampling an essential baseline in the active learning literature (Roy & McCallum, 2001; Houlsby et al., 2011; Sener & Savarese, 2018; Wang et al., 2018).

Uncertainty sampling not only proves to be effective on benchmark datasets, but also is consistently observed to be competitive. In fact, despite numerous research efforts to devise more sophisticated active learning strategies, uncertainty sampling outperforms many of those on benchmark datasets (Cawley, 2011; Yang & Loog, 2018; Karamcheti et al., 2021). The competitiveness has made uncertainty sampling the preferred choice for practitioners, including drug design (Reker & Schneider, 2015; Ding et al., 2021), medical image analysis (Liebgott et al., 2016; Smailagic et al., 2018), and fraud prediction (Leite, 2020).

Despite being generally competitive, it has been observed that uncertainty sampling *sometimes* fails to outperform random sampling (Yang & Loog, 2018; Karamcheti et al., 2021; Munjal et al., 2022). This issue introduces a non-trivial risk for active learning practioners, as they may not effectively reduce the labeling burden when choosing uncertainty sampling over random sampling. Figure 1 illustrates such a case, where an initial partitioning of the *PHONEME* dataset caused uncertainty sampling to perform worse than random sampling, despite its overall competiveness (as seen in Table 1). Several active learning studies have developed more sophisticated strategies to enhance uncertainty sampling (Donmez et al., 2007; Dasgupta, 2009; Huang et al., 2014; Li et al., 2015). Somehow those strategies do not always remain competitive across benchmark datasets (Zhan

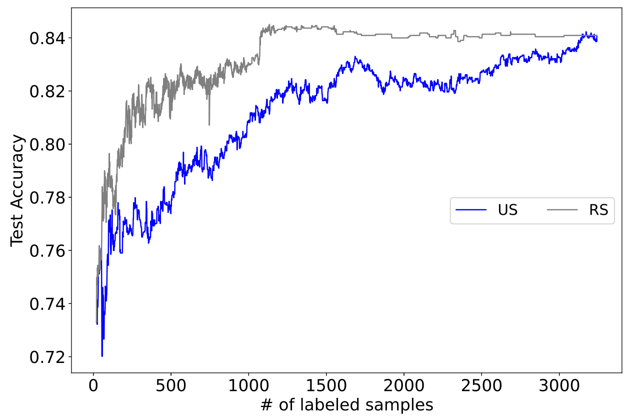

Figure 1. Learning curves on a specific initial labeled pool (seed 11) of the *PHONEME* dataset where uncertainty sampling is inferior to random sampling

et al., 2021). Moreover, the sophisticated nature makes it hard to tune those strategies as a stable baseline. Some other studies have investigated the reasons for the failure cases (Mussmann & Liang, 2018; Karamcheti et al., 2021; Tifrea et al., 2022), but have not proposed any alternative baselines. As a result, a more robust active learning baseline to replace uncertainty sampling remains unknown.

This work explores an alternative baseline to address the gap. Compared to sophisticated query strategies (e.g., QUIRE is time-consuming in Table 1), we investigate a straightforward yet often overlooked strategy known as ϵ-uncertainty sampling. This strategy involves conducting random sampling ϵ percent times while employing uncertainty sampling for the remaining $(1-\epsilon)$ percent in a stochastic manner. In other words, it injects a fraction of random sampling into uncertainty sampling. The strategy offers a lightweight modification of uncertainty sampling, which is easier to tune than more sophisticated strategies (Nguyen & Smeulders, 2004; Huang et al., 2014; Li et al., 2015).

By carefully analyzing the core principles and sensitivity of ϵ-uncertainty sampling, our empirical results demonstrate that injecting a mere 10% of randomness into uncertainty sampling results in a superior baseline. In particular, ϵ-uncertainty sampling with $\epsilon = 10\%$ generally performs no worse than uncertainty sampling while occasionally even outperforming it and rescuing it from inferior performance over random sampling.

Our contributions can be summarized as follows:

- We verify ϵ-uncertainty sampling could make uncertainty sampling more robust with extensive experiments on benchmark datasets.
- We analyze the strength of ϵ-uncertainty sampling with bias-varianc analysis and discover that the injected randomness can balance the bias in uncertainty sampling with a small variance for better performance.
- Our investigation suggests that ϵ-uncertainty sampling with $\epsilon = 10\%$ is competitive to both random sampling and uncertainty sampling and is ready to serve as a more robust baseline for active learning.

The rest of this paper is organized as follows. Section 2 defines pool-based active learning and introduces uncertainty sampling as well as other active learning strategies. Section 3 compares ϵ-uncertainty sampling with the original uncertainty sampling, random sampling, and other active learning strategies, and presents our analysis of the results. Finally, Section 4 concludes the study.

2. Query Strategies for Pool-Based Active Learning

The standard pool-based active learning (Settles, 2012) consists of a labeled pool D_ℓ, a large unlabeled pool D_u, and a model (hypothesis set) $\mathcal{H}$ at the beginning. A labeled pool contains examples $D_\ell = \{(x_1, y_1), \ldots, (x_n, y_n), \ldots, (x_N, y_N)\}$, where x_n are d-dimension features $x_n \in \mathbb{R}^d$ and y_n are corresponding binary labels $y_n \in \{-1, +1\}$; on the other hand, an unlabeled pool contains examples without labels $D_u = \{x_{n+1}, x_{n+2}, \ldots, x_m, \ldots, x_M\}$. In the machine (supervised) learning paradigm, we train a model based on a labeled pool, i.e., selecting the best hypothesis h from model $\mathcal{H}$ that performs well on the labeled pool.

The active learning process is to iteratively select examples from the unlabeled pool until running out of budget. At each round t of the query process, a query strategy $\mathcal{Q}$ is employed to acquire a new example x_m from the unlabeled pool D_u. This newly acquired example is then labeled as (x_m, y_m) by a human oracle and added to the labeled pool D_ℓ for subsequent use. Finally, a new hypothesis h is selected from the model $\mathcal{H}$ based on the updated labeled pool D_ℓ.

We aim to identify an effective query strategy that enhances model performance on the testing set D_{tst} with fewer query times. Random sampling (RS) and uncertainty sampling (US) are two commonly used baselines. Random sampling uniformly selects the example from the unlabeled pool without replacement, assuming equal importance for each example. In contrast, uncertainty sampling adopts a straightforward approach by selecting the most confusing example to the current model. Specifically, focusing on binary classification tasks, the 'margin score' is defined as the difference between the probabilities of the top predicted class and the second predicted class:

$$U(x) = -\left[P\left(h(x) = y^{(1)}\right) - P\left(h(x) = y^{(2)}\right)\right],$$

where $y^{(1)}$ and $y^{(2)}$ are the top and the second class of the model outputs, and $U(x)$ means the score function. The selection of the most uncertain example is determined by maximizing the uncertainty score $\mathcal{Q} = \arg\max_x U(x)$, which prefers to query the example close to the current model's decision boundary.

Uncertainty sampling exhibits instability in certain datasets (Table 1) and initial labeled pools (Figure 1) resulting in poorer performance compared to random sampling. The well-known problem for uncertainty sampling is 'sampling bias' that selects a non-representative example during the query process (Dasgupta, 2009; Huang et al., 2014; Yang et al., 2015; Shui et al., 2019). Several works suggested considering the representative information to deal with the sampling bias, such as Core-Set (Sener & Savarese, 2018), DWUS (Nguyen & Smeulders, 2004), QUIRE (Huang et al., 2014) and HintSVM (Li et al., 2015). Although various query strategies were designed to overcome the insufficiency of uncertainty sampling, the benchmark results in Table 1[1] demonstrates that none of these strategies consistently outperform uncertainty sampling. Furthermore, these strategies do not tackle the challenges posed by *BANANA* dataset.

To address the insufficiency of uncertainty sampling and achieve robust results on benchmarking datasets, we study ϵ-uncertainty sampling, which is a straightforward and efficient alternative approach, but less attention is given to the community. ϵ-uncertainty sampling queries the most uncertain example with probability $1-\epsilon$ and uniformly queries the example with probability ϵ at each round. The core idea behind ϵ-uncertainty sampling is treating random sampling as representative sampling, which encourages pure uncertainty sampling to explore the diverse (not close to the decision boundary) regions in an unlabeled pool.

Tifrea et al. (2022) also compared the ϵ-uncertainty sampling through adjusting ϵ within the $\epsilon \in \{0, 0.25, 0.5, 0.75, 1\}$ creates a spectrum between pure uncertainty sampling and random sampling. However, ϵ-uncertainty sampling yielded inferior results to random sampling under their scenario. In contrast with Tifrea et al. (2022), we carefully compare ϵ on the comprehensive scenarios. Our results show that with $\epsilon = 0.1$, ϵ-uncertainty sampling performs similarly to random sampling when uncertainty sampling fails but also performs well on most benchmarks. Therefore we provide the aspect that ϵ-uncertainty sampling could become the robust baseline for active learning with proper ϵ.

[1]We update the random sampling and uncertainty sampling because Zhan et al. (2021) misused the codebase of Google active learning playground. Please see `https://github.com/ariapoy/active-learning-benchmark`.

3. Experiments

This section verifies that ϵ-greedy could achieve a competitive and robust standard pool-based active learning on the existing benchmark. Then, we investigate the hyperparameter ϵ to realize how randomness affects the active learning process.

3.1. Experimental Setup

This work follows the recent comprehensive benchmark for the standard pool-based active learning (Zhan et al., 2021). We consider a random split of all dataset into 60% training set ($D_\ell \cup D_u$) and 40% testing set D_{tst} and randomly sample 20 as the initial labeled pool D_ℓ from the training set for each dataset. We repeated the experiments with 150 times for the $n < 1000$ datasets and 15 times for the remaining datasets with fixed seeds.

We follow Zhan et al. (2021) to use Radial Basis Function kernel Support Vector Machine (SVM(RBF)) as the model $\mathcal{H}$ and implement all query strategies $\mathcal{Q}$ based on the libact (Yang et al., 2017). Our evaluation metric is the area under the learning curve (AULC) where the learning curve represents the progression of testing accuracy (y-axis) for the model trained on the labeled pool across each round (x-axis) of evaluation. Following the standard pool-based active learning (Settles, 2012) and previous benchmark (Zhan et al., 2021), we focus on the binary classification problems, set the query batch size as 1 and maximum query times to be the size of the unlabeled pool.

3.2. Benchmarking Results

We verify that $\epsilon = 0.1$ stably achieves good performance on all binary classification datasets in existing benchmark (Zhan et al., 2021) with the following steps: First, we calculate the differences in average AULC between a query strategy from random sampling to show the improvement in Table 1. Second, we <u>underline</u> the dataset when uncertainty sampling performs less effectively than random sampling. Third, we **highlight** the improvement of a query strategy is better or equal to the improvement of uncertainty sampling over random sampling.

Table 1 shows $\epsilon = 0.1$ (10%-US), ϵ-uncertainty sampling performs similarly on 11 datasets. Moreover, we improve uncertainty sampling on the *CLEAN1*, *AUSTRALIAN CHECKERBOARD*, and *BANANA* datasets.

Besides revealing the differences in average AULC between a query strategy and random sampling, we plot the learning curves for *BANANA* and *SPLICE*. Figure 2 and Figure 3, which achieve the maximum improvement and de-

Table 1. Benchmark: Mean AULC for RS and differences of other query strategies to RS over 150 or 15 trials

(%)	RS (AVG)	US	10%-US	20%-US	30%-US	DWUS	QUIRE	HINTSVM	CORE-SET
APPENDICITIS	83.95	+0.59	**+0.60**	+0.54	+0.42	+0.26	+0.04	-0.05	+0.03
SONAR	74.63	+2.99	+2.85	+2.81	+2.34	-0.58	+0.12	-1.06	-0.43
PARKINSONS	83.05	+2.26	+2.23	+2.17	+2.03	-0.31	+0.00	-1.27	+0.51
EX8B	88.53	+1.28	+1.27	+1.21	+1.13	-0.13	-0.68	-1.54	+0.63
HEART	80.51	+1.06	+1.05	+1.02	+0.81	+0.06	+0.52	-0.12	+0.54
<u>HABERMAN</u>	73.08	-0.13	**-0.13**	-0.16	-0.15	**+0.04**	-0.64	-0.49	-0.41
IONOSPHERE	91.80	+1.20	+1.18	+1.17	+1.01	-3.87	-1.65	-2.16	-0.46
CLEAN1	81.83	+2.42	**+2.58**	**+2.45**	+2.37	+0.00	-0.03	-4.88	-2.86
BREAST	96.16	+0.16	+0.15	+0.14	+0.14	-0.12	+0.07	+0.07	+0.10
WDBC	95.39	+1.13	+1.11	+1.10	+1.05	-0.35	+0.44	+0.19	+0.47
AUSTRALIAN	84.83	+0.21	**+0.24**	+0.18	+0.14	-0.10	-0.07	-0.39	-0.05
DIABETES	74.24	+0.55	+0.49	+0.43	+0.46	-1.97	+0.46	+0.32	**+0.67**
MAMMOGRAPHIC	81.30	+0.47	**+0.47**	+0.43	**+0.47**	-1.31	+0.28	-0.25	+0.32
EX8A	85.39	+2.49	+2.47	**+2.49**	+2.45	-6.28	-4.67	-4.03	+0.01
TIC	87.18	+0.02	+0.01	+0.01	+0.01	-0.08	-0.19	+0.01	-0.02
GERMAN	73.40	+0.77	+0.70	+0.62	+0.65	-0.72	+0.17	-0.35	+0.25
SPLICE	80.75	+1.54	+1.41	+1.38	+1.20	-4.97	-0.31	-2.92	-5.57
GCLOUDB	89.52	+0.29	**+0.29**	+0.28	+0.28	-0.96	-1.76	-2.04	-0.32
GCLOUDUB	94.37	+1.23	**+1.23**	+1.21	+1.18	-0.73	-1.08	-4.82	-5.08
CHECKERBOARD	97.81	+0.66	**+0.73**	**+0.74**	**+0.79**	-7.36	-3.44	-5.39	**+0.93**
SPAMBASE	91.03	+1.02	+1.01	+1.01	+1.00	+0.00	(> 3 DAYS)	-1.18	-0.51
<u>BANANA</u>	89.26	-1.39	**+0.30**	**+0.40**	**+0.45**	-7.62	-6.27	-4.16	**+0.04**
PHONEME	82.54	+1.01	**+1.02**	**+1.06**	+0.99	-1.17	-0.71	-1.71	-0.14
RINGNORM	97.76	+0.10	**+0.10**	+0.09	+0.09	-4.30	(> 3 DAYS)	-0.61	-2.99
TWONORM	97.53	+0.11	**+0.11**	+0.10	+0.10	-0.22	(> 3 DAYS)	-0.17	+0.02
PHISHING	93.82	+0.78	+0.76	+0.76	+0.76	-4.59	(> 3 DAYS)	-0.86	+0.24

cline, show that injecting small randomness is a good choice for two scenarios.

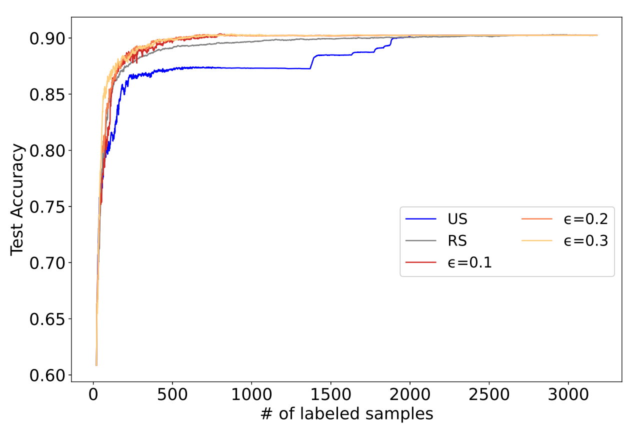

Figure 2. Learning curves with average accuracy over 15 trials on *BANANA*

3.3. Sensitivity Analysis of ϵ-Uncertainty Sampling

To realize that injecting small ϵ could robustly achieve competitive performance, we compare different ϵ on two scenarios to see how randomness affects the query process. The first scenario is uncertainty sampling performs worse than random sampling on average such as *BANANA* dataset. The second scenario is uncertainty sampling outperforms random sampling such as *SPLICE* dataset.

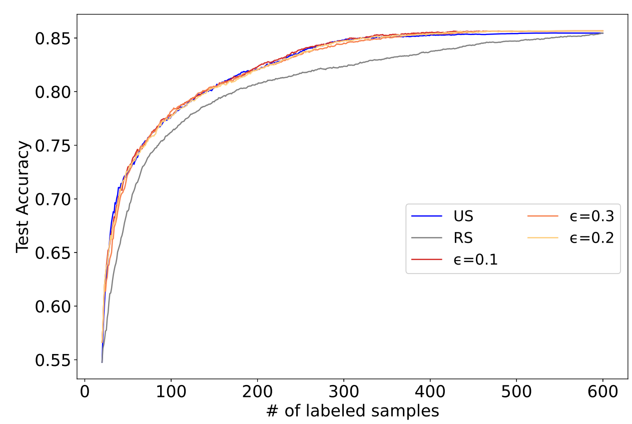

Figure 3. Learning curves with average accuracy over 150 trials on *SPLICE*

The *BANANA* scenario (Figure 4) demonstrates that the performance of uncertainty sampling (left) is unstable with low AULC and has a lower average AULC than random sampling (right). Figure 4 shows the improvement of the average AULC after injecting the randomness from small to large (ϵ at middle) and the improvement from $\epsilon = 0$ (pure uncertainty sampling) to $\epsilon = 0.1$ is significant.

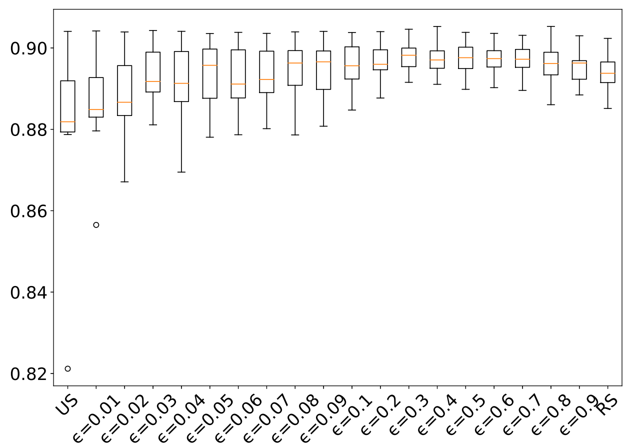

Figure 4. AULC (acc) with mean and standard deviation over 15 trials on *BANANA* for different ϵ

The *SPLICE* scenario (Figure 5) shows injecting randomness declines slowly in small ϵ. For example, even if we enlarge ϵ to 0.2, the hurt for uncertainty sampling is negligible. This evidence shows that injecting small ϵ could benefit when uncertainty sampling performs worse than random sampling and bring minor negative effects when uncertainty sampling performs well.

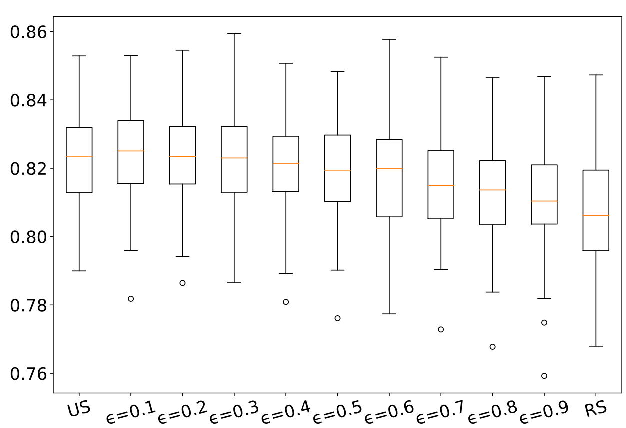

Figure 5. AULC (acc) with mean and standard deviation over 150 trials on *SPLICE* for different ϵ

Besides comparing fixed ϵ during the query process, we also design the dynamic ϵ to verify that the order of exploration and exploitation brings a minor effect to small randomness. According to previous results in Table 1, we fix the expected value of overall $\epsilon = 0.1$ with different starting and ending ϵ. For example, we set $\epsilon = 0.2$ at the beginning and $\epsilon = 0$ at the last round, which means that about 10% randomness is assigned more at the early round (explore first and exploit later). Both scenarios, *BANANA* (Figure 6) and *SPLICE* (Figure 7), demonstrate the difference between different settings is not obvious with a fixed expected value of ϵ. The results show that the order of exploration and exploitation has limited impact under $\epsilon = 0.1$.

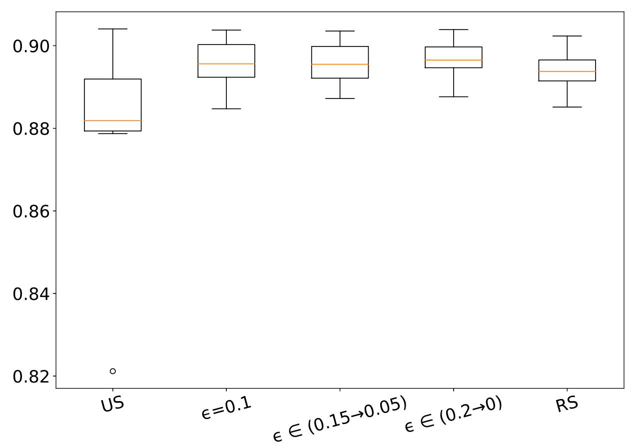

Figure 6. AULC (acc) with mean and standard deviation over 15 trials on *BANANA* for different settings of linear schedulers

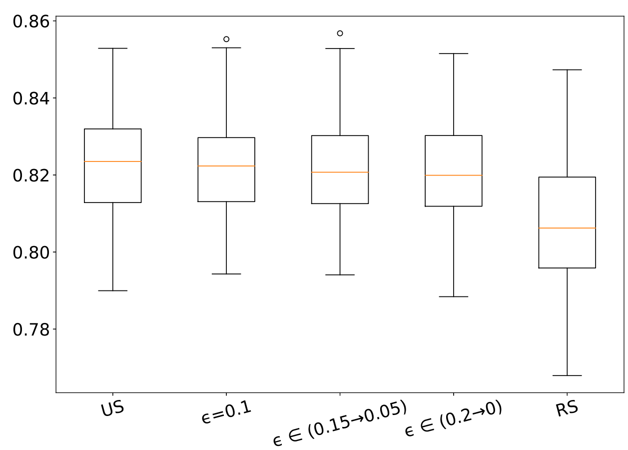

Figure 7. AULC (acc) with mean and standard deviation over 150 trials on *SPLICE* for different settings of linear schedulers

3.4. Bias-Variance Analysis of Randomness

Now, we look more carefully at the benefits of small randomness. Following previous scenarios, we study how ϵ affects the training error through the bias-variance analysis on *BANANA* and *SPLICE* datasets. We fix the seed for the training, testing, and initial labeled pool to reduce confounder. We decompose the training error (error on $D_\ell \cup D_u$) for the hypothesis h trained on the labeled pool at round t with the squared error to bias and variance terms. In the experiments, we check the round at $\{10\%, 33.3\%, 50\%\}$ size of unlabeled pool, which are critical rounds observed in Figure 2.

Figure 8 demonstrates the bias (blue bar) and variance (orange bar) of training error for each query strategy at $t = 10\%|D_u|$ round on *BANANA* dataset. The result shows pure uncertainty sampling contributes significantly high bias to the training error on *BANANA*. Randomness could reduce the bias and enlarge the variance.

Figure 9 and Figure 10 shows the bias and variance at $t \in \{16.7\%, 33.3\%, 50\%\}$ rounds for *BANANA* dataset. We observe that uncertainty sampling could overcome the bottleneck of high bias after collecting more than half of the unlabeled pool, which is less effective than introducing randomness at early rounds.

In *SPLICE* (Figure 11, Figure 12 and Figure 13), the reduction in bias was smaller than the increase in variance at initial rounds, it even led to higher bias in later. Therefore, we prefer small randomness ($\epsilon = 0.1$) to balance the tradeoff when uncertainty sampling brings bias at early rounds.

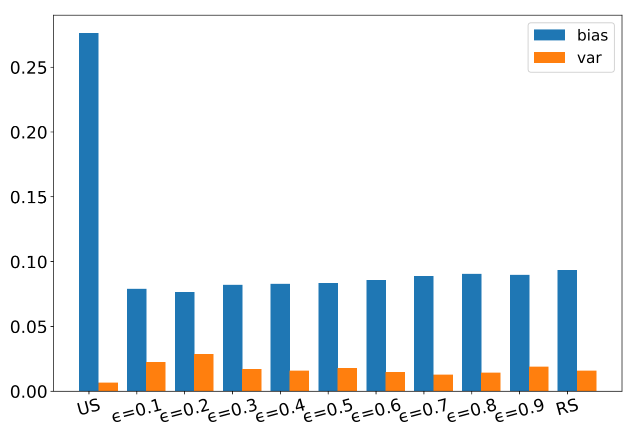

Figure 8. Bias and variance of training error on *BANANA* when using 10% $|D_u|$

3.5. Claim about ϵ-Uncertainty Sampling

We conclude our findings as follows:

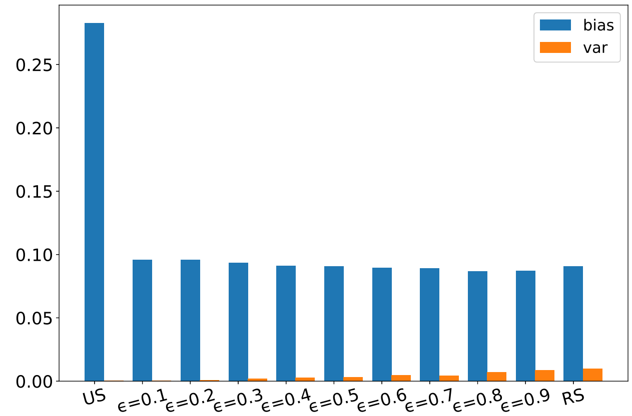

Figure 9. Bias and variance of training error on *BANANA* when using 33.3% $|D_u|$

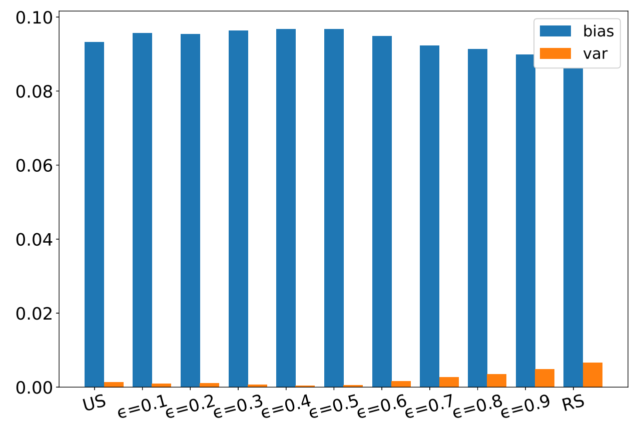

Figure 10. Bias and variance of training error on *BANANA* when using 50% $|D_u|$

- The uncertainty sampling might fail on some datasets, even on some seeds for a dataset, i.e., training, testing and initial labeled pool splits (Figure 1), although it outperforms other complicated query strategies on average on the benchmark.

- Injecting the randomness can reduce the bias in the early round for pure uncertainty sampling. For practical usage, set $\epsilon = 0.1$ can perform similarly to the pure uncertainty sampling in most scenarios.

After our careful investigation, our suggestion for future studies of pool-based active learning is that besides using uncertainty sampling as the baseline, using $\epsilon = 0.1$ is sufficient to become the stable baseline in realistic.

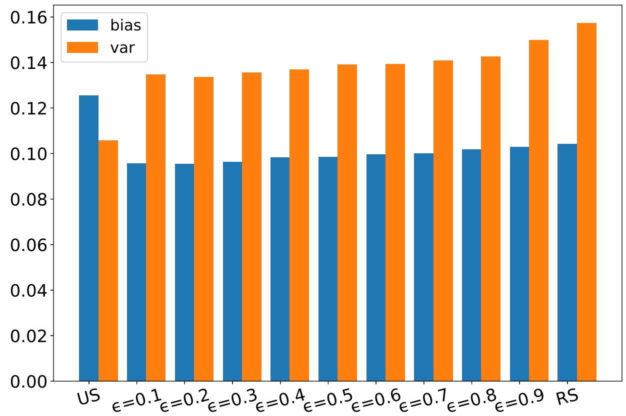

Figure 11. Bias and variance of training error on *SPLICE* when using 10% $|D_u|$

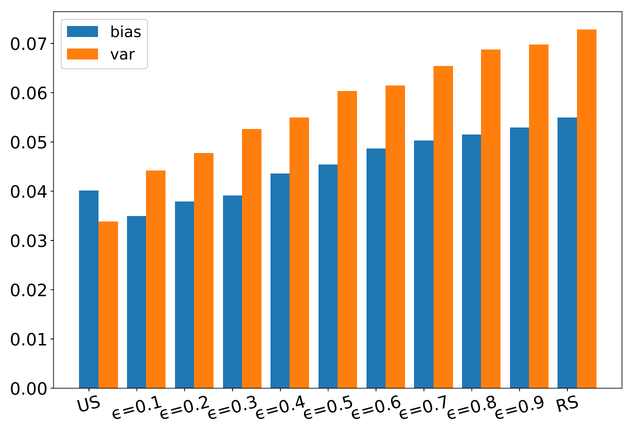

Figure 12. Bias and variance of training error on *SPLICE* when using 33.3% $|D_u|$

4. Conclusion

This work aims to robustify uncertainty sampling towards a stronger baseline and a preferred choice for practitioners across different kinds of applications. We put our attention on ϵ-uncertainty sampling, an overlooked variant of uncertainty sampling. We discover that the injected randomness in ϵ-uncertainty sampling can decrease the bias in the original uncertainty sampling, therefore improving its performance. Our careful study on the sensitivity of ϵ concludes that injecting 10% of randomness is sufficient to warrant competitive performance across existing pool-based active learning benchmarks for binary classification. Future work could consider active learning for multi-class classificati5on and other learning tasks. We hope that our modification and remedy to uncertainty sampling inspire the community to fundamentally study simpler strategies like ϵ-uncertainty sampling for practitioners.

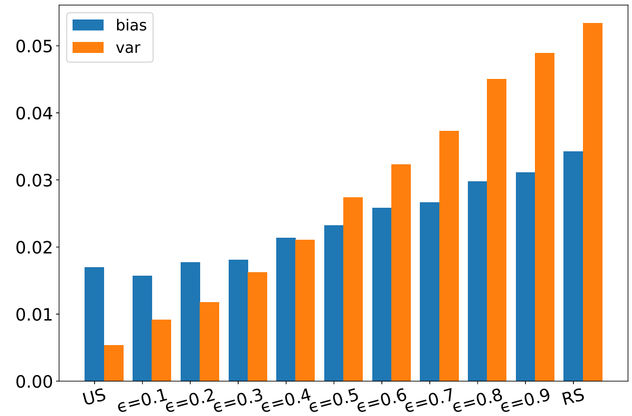

Figure 13. Bias and variance of training error on *SPLICE* when using 50% $|D_u|$

Acknowledgements

We thank the anonymous reviewers for their valuable comments. This work is supported by National Taiwan University Center for Data Intelligence via NTU-112L900901 as well as the Ministry of Science and Technology in Taiwan via MOST 111-2628-E-002-018 and 112-2628-E-002-030. We thank National Center for High-performance Computing (NCHC) in Taiwan for providing computational and storage resources.

References

Cawley, G. C. Baseline methods for active learning. In *Active Learning and Experimental Design workshop In conjunction with AISTATS 2010*, pp. 47–57. JMLR Workshop and Conference Proceedings, 2011.

Dasgupta, S. The two faces of active learning. In *Discovery Science*, pp. 35–35. Springer Berlin Heidelberg, 2009.

Ding, X., Cui, R., Yu, J., Liu, T., Zhu, T., Wang, D., Chang, J., Fan, Z., Liu, X., Chen, K., et al. Active learning for drug design: a case study on the plasma exposure of orally administered drugs. *Journal of Medicinal Chemistry*, 64 (22):16838–16853, 2021.

Donmez, P., Carbonell, J. G., and Bennett, P. N. Dual strategy active learning. In *Machine Learning: ECML 2007: 18th European Conference on Machine Learning, Warsaw, Poland, September 17-21, 2007. Proceedings 18*, pp. 116–127. Springer Berlin Heidelberg, 2007.

Houlsby, N., Huszár, F., Ghahramani, Z., and Lengyel, M.

Bayesian active learning for classification and preference learning. *ArXiv*, abs/1112.5745, 2011.

Huang, S.-J., Jin, R., and Zhou, Z.-H. Active learning by querying informative and representative examples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(10):1936–1949, 2014.

Ji, Y., Kaestner, D., Wirth, O., and Wressnegger, C. Randomness is the root of all evil: More reliable evaluation of deep active learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3943–3952, 2023.

Karamcheti, S., Krishna, R., Fei-Fei, L., and Manning, C. Mind your outliers! investigating the negative impact of outliers on active learning for visual question answering. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 7265–7281. Association for Computational Linguistics, 2021.

Leite, M. L. P. *Active learning for fraud detection*. PhD thesis, 2020.

Li, C.-L., Ferng, C.-S., and Lin, H.-T. Active learning using hint information. *Neural Computation*, 27(8):1738–1765, 2015.

Liebgott, A., Küstner, T., Gatidis, S., Schick, F., and Yang, B. Active learning for magnetic resonance image quality assessment. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 922–926, 2016.

Munjal, P., Hayat, N., Hayat, M., Sourati, J., and Khan, S. Towards robust and reproducible active learning using neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 223–232, 2022.

Mussmann, S. and Liang, P. On the relationship between data efficiency and error for uncertainty sampling. In *International Conference on Machine Learning*, pp. 3674–3682. PMLR, 2018.

Nguyen, H. T. and Smeulders, A. Active learning using preclustering. In *Proceedings of the twenty-first international conference on Machine learning*, pp. 79. Association for Computing Machinery, 2004.

Reker, D. and Schneider, G. Active-learning strategies in computer-assisted drug discovery. *Drug discovery today*, 20(4):458–465, 2015.

Roy, N. and McCallum, A. Toward optimal active learning through monte carlo estimation of error reduction. In *International Conference on Machine Learning*, 2001.

Sener, O. and Savarese, S. Active learning for convolutional neural networks: A core-set approach. In *International Conference on Learning Representations*, 2018.

Settles, B. *Active Learning*, volume 6. Springer, 2012.

Shui, C., Zhou, F., Gagn'e, C., and Wang, B. Deep active learning: Unified and principled method for query and training. In *International Conference on Artificial Intelligence and Statistics*, 2019.

Smailagic, A., Costa, P., Noh, H. Y., Walawalkar, D., Khandelwal, K., Galdran, A., Mirshekari, M., Fagert, J., Xu, S., Zhang, P., et al. Medal: Accurate and robust deep active learning for medical image analysis. In *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, pp. 481–488. IEEE, 2018.

Tifrea, A., Clarysse, J., and Yang, F. Uniform versus uncertainty sampling: When being active is less efficient than staying passive. *arXiv preprint arXiv:2212.00772*, 2022.

Wang, H., Chang, X., Shi, L., Yang, Y., and Shen, Y.-D. Uncertainty sampling for action recognition via maximizing expected average precision. In *IJCAI International Joint Conference on Artificial Intelligence*, pp. 964–970. International Joint Conferences on Artificial Intelligence Organization, 2018.

Yang, Y. and Loog, M. A benchmark and comparison of active learning for logistic regression. *Pattern Recognition*, 83:401–415, 2018.

Yang, Y., Ma, Z., Nie, F., Chang, X., and Hauptmann, A. G. Multi-class active learning by uncertainty sampling with diversity maximization. *International Journal of Computer Vision*, 113:113–127, 2015.

Yang, Y.-Y., Lee, S.-C., Chung, Y.-A., Wu, T.-E., Chen, S.-A., and Lin, H.-T. libact: Pool-based active learning in python. *arXiv preprint arXiv:1710.00379*, 2017. URL https://github.com/ntucllab/libact.

Zhan, X., Liu, H., Li, Q., and Chan, A. B. A comparative survey: Benchmarking for pool-based active learning. In *IJCAI*, pp. 4679–4686. International Joint Conferences on Artificial Intelligence Organization, 2021.