Human heuristic based Drop-out Mechanism for Active Learning

Author information scrubbed for double-blind reviewing

Abstract. Active Learning (AL) operates on the concept that machine learning algorithms can achieve greater accuracy with fewer training labels if they can choose the data points from which they learn. Typically, human annotators provide the labels, but their decisions are often influenced by heuristics and biases. Research in social psychology indicates that humans tend to make decisions based on a limited set of attributes, sometimes relying on a single attribute (referred to as Take the Best). This paper presents a closed-form expression that provides the probability of a data point being mislabeled by the human heuristic. Expanding on this, we also introduce a new drop-out mechanism that impacts the human labeller's attribute selection, thereby nearly doubling the effectiveness of Active Learning.

Keywords: Active Learning · Human-in-the-loop · Human behaviour · Biases · Take the best heuristic

1 Introduction

There are various scenarios where it is significantly expensive to obtain labels of instances as opposed to their input attribute values. Take the following scenarios: [Scenario 1]A Hospital wants to decide on the patients to whom Intensive care must be provided, [Scenario 2] A Bank intends to decide on the customers to whom the loan must be sanctioned, or [Scenario 3]An IT Firm needs to shortlist applicants for a particular role. In all these scenarios, it is clear that for a prediction model to be created, obtaining the attribute information would be an undemanding task compared to their labels. Active Learning(AL) has the leverage of choosing the data points to be queried at each instance, reaching the benchmark accuracy with fewer queries(labelled instances). Starting with a few labelled examples, a typical Active Learner queries the oracle to obtain labels for one or more unlabeled samples and chooses further points to query based on labels obtained on previous queries [1]. A substantial set of AL-based querying[9], including the three standard use case scenarios(Scenarios 1,2 & 3), involves Human-in-the-loop learning (HIL/HITL) systems where the annotator is a human.

Many of the previously implemented studies in Active Learning assume the oracle to be a bias and error-free annotator[2–4]. However, published psychological works have established that humans tend to be biased, resulting from a heuristic when making decisions[5]. Hence, the labels provided to the Active

Learner could also result from a particular bias possessed by the annotator, thereby reducing the performance of the model trained[6].

Our work tries to decrease the impact of such biases by following a query approach mechanism where specific data would be hidden while querying the annotator for certain instances prone to be mislabelled. This forces the human oracle to provide labels without information on the values of a few sets of attributes for each instance. In addition to the decrease in the possibility of the human oracle to base all its decisions on a particular set of attributes alone, it also increases the quality of labels obtained.

The rest of the paper is structured as follows. Section 2 describes a few published works related to our study. Section 3 details the methodology followed. Sections 4 and 5 include the human heuristic(Take the best) and the drop-out mechanism proposed in our study. The Theoretical Validation and Experimental Results encompass the subsequent sections, followed by the Conclusion.

2 Related Works

A few published works in active learning consider human errors in providing labels and have found AL performs poorly when an oracle happens to be human[12]. This error has also been assumed to result from an unbiased label noise[10, 11]. Earlier works tried to tackle this by obtaining feedback on the features in addition to the labels from the human annotator [13, 14].

However, literature from behavioural economics and psychology informs us that humans follow various heuristic strategies when making decisions in both simulated and real settings[7, 8]; providing labels for AL should be no exception. Thus, the assumption made by traditional active learning literature of unbiased label noise might not hold.

Our work addresses the impact of human heuristics on an active learning setup by proposing a novel drop-out mechanism. Dropping specific nodes randomly while performing predictions is a technique used in deep learning to avoid over-fitting[15]. More specifically, a drop-out-based active learning for regression was proposed involving a tailored neural network employed with a drop-out mechanism that works by dropping out certain neurons, thus disabling certain parts while training the model[16]. Even though this methodology helps in regularizing overfitting in the model, it doesn't directly impact the bias included in the dataset by the oracle that our model achieves.

3 Methodology

As shown in figure 3, the methodological framework includes a standard active learning algorithm(entropy sampling[17]) that chooses the data point(x_n) from a pool of unlabeled data points to query(X_{pool}). From the data point chosen, the proposed drop-out mechanism drops the attribute values that must not be presented to the oracle. This results in x_n^{oracle}, which contains a subset of attributes present in x_n that are to be sent to the oracle for labelling. In our study,

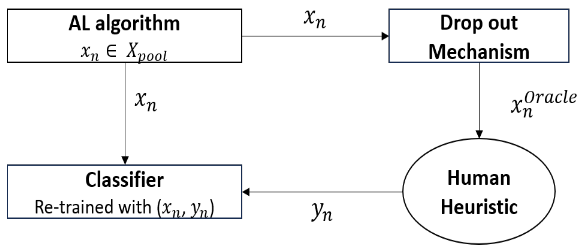

Fig. 1. Methodological framework

we mimic the functionality of an oracle by using Take the best heuristic as the decision strategy used by the oracle. The heuristic finally provides the label(y_n), which is then used to train the underlying Logistic Regression classifier. Take the best heuristic and the proposed drop-out mechanism, which has been detailed in subsequent sections.

4 Take the best heuristic

Well-established behavioural science literature's[7, 8] have formulated the existence of various heuristics humans use while making a decision. We consider one such heuristic, Take the Best(TTB), for our study due to its frugal nature, high usage, and good performance.

This heuristic is based on the scenario that a human would make a decision based on the best attribute, i.e., the attribute that the human believes to influence the most. Thus, when a data point is queried, the heuristic looks at a particular feature value alone and assigns a label if the value is greater than the population median. In our study, the accuracy of the decision made by each attribute was evaluated to find the best-performing attribute.

5 Drop-out Mechanism

The quality of a label provided for a query depends on the attribute chosen by the heuristic for labelling. This selection can be influenced by concealing attribute information during querying. To determine which attributes to drop during each query, we propose the following hypothesis:

Hypothesis: The probability a data-point queried would provide an incorrect label can be computed using their attribute values(a,b) as follows:

$$P(\frac{error}{a,b}) = 0.5 - \frac{1}{90}\tan^{-1}\frac{min(a,b)}{max(a,b)} \tag{1}$$

Theoretical proof of this hypothesis is provided in the subsequent section. While querying, the designed mechanism uses the probability of error computed using eqn(1) for every possible attribute combination to drop attributes before querying the heuristic. This prevents the heuristic from relying on those in its decision-making process.

The mechanism thus involves the following steps:

Input: Active Learner provides the data point to be queried.
1: The below are computed during each query.

$$(a^*, b^*) = \arg\max_{(a,b)} P\left(\frac{\text{error}}{a, b}\right)$$

$$P_{\max} = P\left(\frac{\text{error}}{a^*, b^*}\right)$$

2: if $P_{max} \geq 0.8$:
The attribute whose value min(a*,b*) is dropped
3: The oracle is then queried by providing information on the remaining attributes
Output: The Label thus obtained is trained with all attribute values of the instance

The above steps are repeated after the selection of subsequent instances by the Active Learner.

6 Theoretical Validation

Take the best heuristic involves decision-making by the human oracle based on a single attribute. Let us consider a prediction task involving attributes A and B, as shown in Figure 2. The origin here represents the median values. Hence, the Heuristic decision boundary(HDB) will be along the A axis if B is used by the heuristic and vice versa. This study assumes the True decision boundary(TDB) is linear and could make any angle with A.

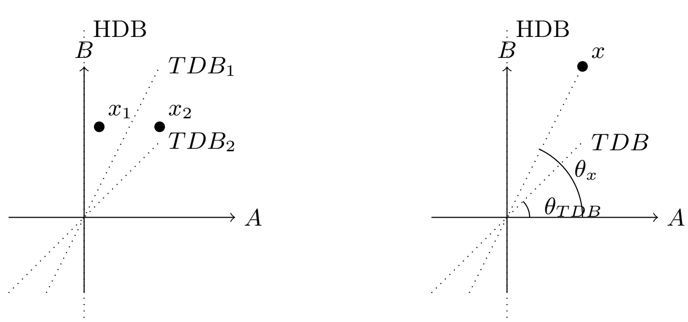

Fig. 2. L-Figure showing the intuitive reason behind x_1 having more chances of being mislabeled by heuristic. R-Figure generalizing the scenario with suitable variables

The probability of a data point being misclassified depends on the number of True decision boundary scenarios where they would be misclassified i.e. where they would lie between the TDB and HDB. In Figure 2, x_1 would be misclassified if the True decision boundary happens to be any one among the TDB_1, TDB_2. However, x_2 would be misclassified only if TDB_2 happens to be the decision boundary, making x_1 more susceptible to obtaining incorrect labels from the annotator in comparison to B. This scenario would be mirrored for the data points having a key-attribute value less than the splitting value. In Figure 2(R), let x be a datapoint, making an intercept of a and b with Axes A and B respectively.

Hence,

$$P(\frac{error}{a,b}) = 1 \text{ if } \theta_{TDB} < \theta_x \text{ and 0 otherwise} \tag{2}$$

Here, θ_x is the angle made by datapoint x with axis A. θ_{TDB} is the angle made by True decision boundary with axis A that ranges between $\frac{\pi}{4}$ and $\frac{\pi}{2}$. Hence probability of error can be shown as:

$$P(\frac{error}{a,b}) = \frac{\int_{\theta_x}^{\frac{\pi}{4}} d\theta}{\frac{\pi}{2}} \tag{3}$$

We must note that $\theta_x = \tan^{-1}\frac{b}{a}$.

The above can thus be shown as:

$$P(\frac{error}{a,b}) = 0.5 - \frac{1}{90}\tan^{-1}\frac{b}{a} \tag{4}$$

To bound the equation within 0 and 1:

$$P\left(\frac{\text{error}}{a,b}\right) = \max\left(0, \min\left(1, 0.5 - \frac{1}{90}\tan^{-1}\frac{b}{a}\right)\right) \tag{5}$$

It must be noted that equation 4 is valid only when the HDB is along A (Attribute B is used by the heuristic). When HDB is along B (Attribute A is used by the heuristic), the $\frac{b}{a}$ will be replaced by $\frac{a}{b}$. This enunciates that when the heuristic picks attribute A in decision-making, the error probability is 0 for data points with b>a and vice versa.

Since the attribute picked by the heuristic is not known to the active learner, we take the worst-case scenarios to compute the probabilities:

$$P(\frac{error}{a,b}) = 0.5 - \frac{1}{90}\tan^{-1}\frac{min(a,b)}{max(a,b)} \tag{6}$$

As discussed in the previous sections, the proposed drop-out mechanism drops attributes when the probability of error is above 80%, i.e. datapoints inclined within 27^o with A or B axis. For datapoint x_1, Attribute A would be dropped while querying, forcing the heuristic to use B in decision-making. The error probability, when computed using eqn.5, would be null since b>a for x_1. This theoretically validates the increased quality of labels obtained due to the proposed drop-out mechanism.

7 Experimental Results

To test the effectiveness of the proposed drop-out mechanism, we follow the methodology stated in previous sections to perform experiments on four datasets[18] with a budget of 350 data points to query. The learning curves(No. of data points queried vs. prediction accuracy) averaged over ten different iterations are shown in Fig.3.

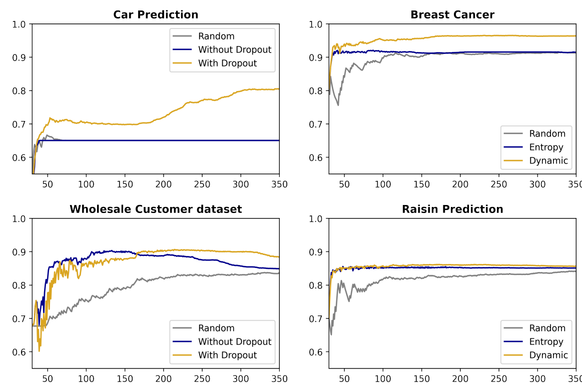

Fig. 3. Learning curves

The area under the learning curves for all the prediction tasks is shown in Table 1 of the Appendix. It can be seen that the proposed drop-out mechanism shows an improvement in performance at all the prediction tasks considered. The effectiveness of the AL algorithm, i.e., the difference in the performance of the AL algorithm and random sampling, has been nearly doubled due to the usage of the drop-out mechanism, thereby advocating the usage of the same.

8 Conclusion

Starting from the common-sense observation that sometimes the labels needed for active learning must be provided by a human, who might be biased, we model the oracle as a fast-and-frugal heuristic. This paper provided a theoretical formulation that provides the probability of data points being mislabeled by the heuristic. Based on this inference, we proposed a novel drop-out mechanism. This mechanism prevents the heuristic from picking certain attributes while querying every instance, which could lead to incorrect labels when used in heuristic decision-making. Experimental results showed that coupling the drop-out mechanism with active learning nearly doubled their effectiveness. This motivates the development of various human-heuristic-aware mechanisms that could enhance the performance of prediction models, as demonstrated in this study.

Acknowledgments. A project "Improving active learning performance in the context of human heuristics and biases" - SB20210345CPAMEXAMEHOC was funded internally by the Amex DART Lab (Data Analytics, Risk and Technology lab). A laboratory within IIT Madras

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cohn, D., Ghahramani, Z. & Jordan, M. Active learning with statistical models. *Journal Of Artificial Intelligence Research.* **4** pp. 129-145 (1996)
2. Hoi, S., Jin, R. & Lyu, M. Large-Scale Text Categorization by Batch Mode Active Learning. *Proceedings Of The 15th International Conference On World Wide Web.* pp. 633-642 (2006), https://doi.org/10.1145/1135777.1135870
3. Arora, S., Nyberg, E. & Rosé, C. Estimating Annotation Cost for Active Learning in a Multi-Annotator Environment. *Proceedings Of The NAACL HLT 2009 Workshop On Active Learning For Natural Language Processing.* pp. 18-26 (2009,6), https://aclanthology.org/W09-1903
4. Guo, Y. & Schuurmans, D. Discriminative Batch Mode Active Learning. *Advances In Neural Information Processing Systems.* **20** (2007), https://proceedings.neurips.cc/paper/2007/file/ccc0aa1b81bf81e16c676ddb977c5881-Paper.pdf
5. Ariely, D. Predictably irrational : the hidden forces that shape our decisions. (First Harper Perennial edition. New York : Harper Perennial, 2010.,2010)
6. Agarwal, D., Covarrubias-Zambrano, O., Bossmann, S. & Natarajan, B. Impacts of Behavioral Biases on Active Learning Strategies. *2022 International Conference On Artificial Intelligence In Information And Communication (ICAIIC).* pp. 256-261 (2022)
7. Katsikopoulos, K., Şimşek, Buckmann, M. & Gigerenzer, G. Classification in the Wild: The Science and Art of Transparent Decision Making. (MIT Press,2021,2)
8. Czerlinski, J., Gigerenzer, G. & Goldstein, D. How good are simple heuristics?. *Simple Heuristics That Make Us Smart.* pp. 97-118 (1999,1)
9. Robinson, T., Tax, N., Mudd, R. & Guy, I. Active learning with biased nonresponse to label requests. *Data Mining And Knowledge Discovery.* (2024,5,25), https://doi.org/10.1007/s10618-024-01026-x
10. Settles, B. Active Learning Literature Survey. *University Of Wisconsin, Madison.* **52** (2010,7)
11. Twomey, N., Diethe, T. & Flach, P. Bayesian Active Learning with Evidence-Based Instance Selection. (2015), European Conference on Machine Learning and Knowledge Discovery (ECML PKDD)
12. Baum, E. & Lang, K. Query learning can work poorly when a human oracle is used. *International Joint Conference On Neural Networks.* **8** pp. 8 (1992)
13. Raghavan, H., Madani, O. & Jones, R. Active Learning with Feedback on Features and Instances. *Journal Of Machine Learning Research.* **7**, 1655-1686 (2006), http://jmlr.org/papers/v7/raghavan06a.html
14. Du, J. & Ling, C. Active Learning with Generalized Queries. *2009 Ninth IEEE International Conference On Data Mining.* pp. 120-128 (2009)

15. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Drop-out: A Simple Way to Prevent Neural Networks from Overfitting. *Journal Of Machine Learning Research*. **15**, 1929-1958 (2014), http://jmlr.org/papers/v15/srivastava14a.html
16. Tsymbalov, E., Panov, M. & Shapeev, A. Dropout-Based Active Learning for Regression. *Analysis Of Images, Social Networks And Texts*. pp. 247-258 (2018)
17. Shannon, C. Mathematical theory of communication. *The Bell System Technical Journal*. **27** pp. 379-423 (1948)
18. Blake, C. & Merz, C. UCI repository of machine learning databases. (University of California, Irvine, School of Information,1998),

9 Appendix

Table 1 shows the area under the Accuracy vs. data points queried for all the prediction tasks considered.

Dataset	**Random**	**Without Drop-out**	**With Drop-out**
Car Condition	228.16	227.85	**260.34**
Raisin Dataset	286.26	297.01	**299.08**
Wholesale Customer	279.64	304.29	**306.21**
Breast Cancer	313.86	319.17	**333.88**
Average	276.98	287.08	**299.66**

Table 1. Area under the learning curves

Increase in effectiveness$=\frac{Avg.AUC_{withdropout}-Avg.AUC_{withoutdropout}}{Avg.AUC_{withoutdropout}-Avg.AUC_{Random}}$
Increase in effectiveness $= 1.2(120\%)$