Data Mining and Knowledge Discovery
https://doi.org/10.1007/s10618-024-01026-x

Active learning with biased non-response to label requests

Thomas S. Robinson[1,2] · Niek Tax[2] · Richard Mudd[2] · Ido Guy[3]

Received: 5 December 2023 / Accepted: 16 April 2024
© The Author(s) 2024

Abstract
Active learning can improve the efficiency of training prediction models by identifying the most informative new labels to acquire. However, non-response to label requests can impact active learning's effectiveness in real-world contexts. We conceptualise this degradation by considering the type of non-response present in the data, demonstrating that biased non-response is particularly detrimental to model performance. We argue that biased non-response is likely in contexts where the labelling process, by nature, relies on user interactions. To mitigate the impact of biased non-response, we propose a cost-based correction to the sampling strategy–the *Upper Confidence Bound of the Expected Utility (UCB-EU)*–that can, plausibly, be applied to any active learning algorithm. Through experiments, we demonstrate that our method successfully reduces the harm from labelling non-response in many settings. However, we also characterise settings where the non-response bias in the annotations remains detrimental under UCB-EU for specific sampling methods and data generating processes. Finally, we evaluate our method on a real-world dataset from an e-commerce platform. We show that UCB-EU yields substantial performance improvements to conversion models that are trained on clicked impressions. Most generally, this research serves to both better conceptualise the interplay between types of non-response and model improvements via active learning, and to provide a practical, easy-to-implement correction that mitigates model degradation.

Keywords Active learning · Non-response · Missing data · E-commerce · CTR prediction

1 Introduction

Many real-life machine learning (ML) system deployments contain a human-in-the-loop component where the model is continuously or periodically updated based on new batches of labelled data. These new labels are typically obtained either by a pool of human annotators (e.g., through crowdsourcing) or originate from users'

Responsible editor: Rita P. Ribeiro.

Extended author information available on the last page of the article

Published online: 25 May 2024

interactions with an application. *Active Learning* (AL) (Settles 2009) is a research field focused on optimizing the efficiency of this labeling, by directing efforts towards instances whose labels are deemed most informative.

Most AL approaches make the critical assumption that every query (i.e., an attempt to acquire the label of an unlabelled example) receives a response. This assumption may not always hold in practice. For example, consider a fraud detection setting (Tax et al. 2021; Carcillo et al. 2018) where AL selects which items are to be sent for human review to be investigated for potential fraud or an integrity violation. The responses obtained from the human reviewer are then incorporated into the training data for an automated detection system. In practice, a reviewer may struggle to identify fraud in some instances more so than in others; for instance, emails soliciting payment details may be easier to review than emails containing phishing links. In the harder cases, where the reviewer is unsure, they may avoid returning a label (or deliberately mark it null to indicate that they were unable to reach a conclusion). Beyond highlighting the possibility of non-response, this example also underscores that non-response can be biased, i.e., non-response may be more likely for some unlabelled examples than for others.

Annotations may also arise as an implicit byproduct of users' interactions with an application, rather than from annotators whose main focus is to provide labels (and who are the predominant focus in the AL literature). In implicit labelling cases, AL methods can be used to provide an exploration component to ML applications (Elahi et al. 2016). As an example, consider a recommender system that uses AL to present an item to a user that may improve the recommender system's future ability to suggest relevant items *if* we obtain a label for that user-item pair. In such an application, a label may, for example, be obtained only if the user clicks on the recommendation. If, however, the user gets recommended the item but scrolls past it without interacting, this is plausibly an instance of non-response to an AL label request. In such contexts, non-response can be both substantial and also dependent on the user-item pair (i.e., biased). In Sect. 7 we present a real-life example of this scenario, where labels arise from the actions of users and non-response stems from failures to interact with advertised items.

Some works study *AL with abstention feedback* (Fang et al. 2012; Yan et al. 2015; Amin et al. 2021; Nguyen et al. 2022), which accounts for scenarios where the annotator provides no label. However, these studies often overlook the potential consequences of non-response *bias* in the abstention mechanism. While non-response bias is a well-studied problem in statistics, in Sect. 6 we show that its presence in the context of AL with abstention feedback yields new and unique challenges. To illustrate these challenges, consider again the recommender system example. An AL recommender system may query an item for which the model has high uncertainty about the item's relevance to the user. However, the uncertainty for this query may stem from the fact that users rarely interact with this item. Consequently, and in practice, the AL system's attempts to learn about the user's preferences may end up being wasteful as the item likely continues to receive little interaction, which contradicts the goal of AL to maximise model improvement efficiency.

In this paper, we introduce how biased non-response can affect the performance of AL, and demonstrate these mechanisms empirically. We then present an

algorithm to adjust AL to account for non-response, demonstrating both experimental and applied contexts in which it improves model performance. We also show specific contexts where model performance is still impacted negatively. In summary, we make three contributions to AL research:

- We conceptualise a mechanism for *how* biased non-response can undermine the supposed benefits of AL and consider important contexts where the non-response probabilities are very high.
- We propose a simple algorithmic correction for incorporating a model of non-response into AL.
- We demonstrate how the mechanics of AL can lead to reinforcing negative behaviour due to the unavailability of labels in specific regions of the possible feature space.

2 Background

This section introduces the theoretical framework on which our argument rests. We use lowercase letters to denote scalars (e.g., a), lowercase bold letters to denote vectors (e.g., $\mathbf{a}$), and uppercase bold letters to denote matrices (e.g., $\mathbf{A}$).

We consider the case where a researcher faces a common modelling problem: given a vector of features $\mathbf{x}$, what is the corresponding label y? To answer this question, typically the researcher fits a *target* model $\mathcal{M}$ on a training set $\mathbf{D}^{\text{Train}} = \{\mathbf{X}, \mathbf{y}\}$, and uses the resulting model to predict labels for new observations $\mathbf{X}'$:

$$\hat{\mathbf{y}}' = \mathcal{M}(\mathbf{X}' \mid \mathbf{D}^{\text{Train}}).$$

AL extends this logic by making iterative attempts to acquire new labelled examples to improve model performance efficiently. Let $\mathbf{X} \subseteq \mathbb{R}^d$ be a dataset of all instances that can potentially be labelled. At each point in time t, we have a training set $\mathbf{D}^{\text{Train}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{|T|}$, where each $x_i \in \mathbf{X}$ and $y_i \in \{0, 1\}$. The subset of $\mathbf{X}$ that is unlabelled at time t is the "pool" ($\mathbf{X}^{\text{Pool}}$). Given this context, AL first identifies which unlabelled example ($\mathbf{x}_t \in \mathbf{X}^{\text{Pool}}$) to query:

$$\mathbf{x}_t \sim \phi(\mathbf{X}^{\text{Pool}}, \mathcal{M}_{t-1}),$$

where ϕ is some AL querying strategy given the pool of unlabelled examples and the previous state of the model. Common criteria include *uncertainty sampling* (Lewis 1995) (i.e., identifying the next query include the entropy of the model output), or *Query-by-Committee* (QbC) (Seung et al. 1992; Freund et al. 1997) (i.e., maximising the disagreement between members of an ensemble).

Once the example has been chosen, an annotator (be it a human reviewer or other process) labels this datapoint. In conventional settings, we assume that we receive the observed value of every requested label, y_t. More formally, let $\Omega(\cdot)$ be some (unknown) function that determines the probability of receiving a response, such that $p_t = \Omega(\mathbf{x}_t)$ is the response probability for a specific datapoint. Consequently, let $r_t \sim \mathcal{B}(p_t)$, be a draw from the Bernoulli distribution, which indicates whether a

response was received for that example. In the conventional setting, therefore, AL assumes that $\Omega(\mathbf{x}_t) = 1,\ \forall \mathbf{x}_t$.

The results of this process are then combined with existing training examples from previous periods:

$$\mathbf{D}_t^{\text{Train}} = \begin{cases} \mathbf{D}_{t-1}^{\text{Train}} \cup \{\mathbf{x}_t, y_t\}, & \text{if } r_t = 1 \\ \mathbf{D}_{t-1}^{\text{Train}}, & \text{otherwise,} \end{cases}$$

and finally, using the updated training set, a new target model is trained.

3 Related literature

Problems associated with missing data have long been documented in the econometrics literature (Rubin 1976). In practice, data can be missing for a variety of reasons, including imperfect measurement, dataset corruption, and partial responses. In the survey sampling literature, non-response refers to the failure of individuals to reply to a questionnaire or survey (Hansen and Hurwitz 1946), and hence non-response is a particular form of missing data. This longstanding sampling problem has a clear analogue to AL in the form of failed attempts to query a label.

Theoretical work in this area has largely focused on how "missingness" in the data can *bias* parameter estimates (Rubin 1976; Mohan et al. 2013; Little and Rubin 2019). In particular, this work has yielded a typology of missingness: data *missing completely at random* (MCAR), where missing values are independent of observed and unobserved features of the data generating process (DGP) including the outcome; data *missing at random* (MAR), where missing values are independent of unobserved features but related to observed features of the DGP; and, data *missing not at random* (MNAR), where missing values are related to unobserved features of the DGP.

Our focus differs from the econometric treatment of missing data because AL is *inherently* biased by the deliberate acquisition of training data (Farquhar et al. 2021). By selecting the most informative labels, the data will not follow the underlying population distribution. Thus, our interest is not squarely on the inferential validity of the model, but rather on the relative performance of a model afflicted with missingness compared to the counterfactual context where there is no probability of non-response. Our work is, therefore, more aligned with work on missing data in machine learning contexts, where the goal is to account for missingness to ensure datasets are complete, rather than ascertain unbiased parameter estimates (Stekhoven and Buhlmann 2012).

A recent corpus of work has considered non-response in AL contexts (Fang et al. 2012; Yan et al. 2015; Amin et al. 2021; Nguyen et al. 2022). Of these contributions, two provide algorithmic improvements aimed at handling the sorts of non-response described above. Yan et al. (2015) propose repeatedly querying examples with non-response, which may be costly if a null label is highly likely for some informative

regions of the feature space. More recent work has sought, therefore, to incorporate the posterior predictive rate of abstention in the objective function (Nguyen et al. 2022). The Bayesian aspect of this work is, however, computationally taxing, and in practice requires approximate methods using *maximum a posteriori* estimation and regularised regression models. It is unclear, given these constraints, whether this method can be applied to ensemble-based AL strategies, like QbC, which can be more performant than single-hypothesis methods like uncertainty sampling. In contrast, the method proposed in this paper simply requires weighting the sampling "score" from any strategy.

Moreover, these works do not formalise directly how differences in the *types* of missingness that determine non-response impact the performance of AL. Work in this area has noted "knowledge blind-spots'" (Fang et al. 2012) and differences when non-response is close and far from the decision boundary (Nguyen et al. 2022), but not compared more general, theoretically derived mechanisms of missingness. Finally, we consider AL contexts where a non-response model can be trained separately, and often in advance, which may be especially beneficial in production systems and/or where the rate of non-response is particularly high.

One other proximal research area within AL research focuses on noisy or imperfect labels (Yan et al. 2016). Here, like in the case of non-response, model performance may be affected by differences between the returned and the true label. However, unlike noisy labels, we know when we get non-response (a null value is returned), whereas we often do not know which labels are noisy, leading to systematic differences in how these complications are handled. Work explicitly on noisy labels has focused on identifying instances where we are unsure about the label value, and re-labelling these points (Sheng et al. 2008; Zhao et al. 2011; Lin et al. 2016; Nguyen et al. 2020).

Finally, there are conceptual similarities between AL with non-response and research on multi-armed bandits (Lattimore and Szepesvári 2020), i.e., a class of algorithms that study the setting where an agent iteratively takes an action that results in an observed reward, where the goal is to maximise the total reward over a time window. Bandit algorithms face the choice of *exploring* actions where currently little is known, and *exploiting* existing knowledge. There is a strong connection between the AL problem and *best arm identification* (BAI) (Audibert et al. 2010), a bandit context where one is solely concerned with maximising the knowledge about an arm's (potentially context-dependent) reward distribution. One notable difference is that BAI is concerned with exploration for the sake of learning the reward distribution over actions depending on context while AL is concerned with exploration for the sake of minimising expected future prediction errors. Moreover, the *partial monitoring* literature (Bartók et al. 2014) considers bandit settings where for some actions the reward is never observed. This setting has similarities to an active learning context with zero response probability in certain regions of covariate space.

4 The impact of (biased) non-response

Labelling takes different forms, which can lead to different non-response mechanisms. First, consider the conventional context where an unlabelled example is sent to a human annotator who explicitly returns a value for that instance. Non-response in these contexts can arise for several reasons. Annotators may have capacity to review a fixed number of labels, smaller than the number requested, and thus some queries remain unreviewed. Or, the annotator may be unsure over the example's label and, rather than guess, abstain or return a null value. Similarly, if there are multiple annotators, a value may not be assigned if there is not majority agreement or consensus.

Not all labelling is explicit, however: annotations can be based on some implicit (user) behaviour. AL is a well-documented method for improving the performance of recommendation systems (Elahi et al. 2016). Real-life ranking systems often factorize the target prediction into several components (Ma et al. 2018). For example, video streaming services often train two models that separately estimate a *click-through rate* (CTR), $P(click = 1)$, and some quantity that depends on the user behavior after the click such as watch time ($\mathbb{E}[watch\ time \mid click = 1]$). This second model is referred to as the *post-click model*. A final ranking is created by sorting on the result of their multiplication: $\mathbb{E}[watch\ time] = \mathbb{E}[watch\ time \mid click = 1] \times P(click = 1)$ (Lin et al. 2023).

Similarly, advertisement ranking systems often estimate conversion probabilities of ads by factoring this quantity into a *CTR model* and a *post-click model*, i.e., $P(conversion) = P(conversion \mid click = 1)P(click = 1)$, where the post-click model estimates the conversion probabilities of clicked ads (Barbieri et al. 2016; Rosales et al. 2012; Ma et al. 2018). In the ranking systems of these examples we may seek to employ active learning to make the exploration of post-click models estimating $P(click = 1)$ and $\mathbb{E}[watch\ time \mid click = 1]$ more efficient. We discuss a practical example in more depth in Sect. 7.

Active learning systems that aim to improve post-click/interaction models are particularly affected by non-response since labels in these settings are contingent upon user interactions such as clicks. In such contexts, the probability of receiving a valid label may be very low as users only interact with a tiny share of the items. Consequently, the costs of AL may be considerably higher and outweigh the benefits of purposefully serving new advertisements or recommendations in order to improve the post-click model.

Beyond the labelling costs of non-response, the presence of missing or null values across both explicit and implicit AL contexts can affect model training, potentially introducing additional sources of bias into the model framework (Cortes et al. 2018). More formally, let $R_{\Omega(\mathbf{x})}$ be a random binary variable that denotes whether the labelling process returns a valid response (1) or non-response (0). Wherever the query is clear from the context, we will refer to this random variable as R_Ω, for simplicity. We assume that a query can be repeated in subsequent rounds if a label is not returned in the current round. As a result, the prediction model becomes conditional

not only on the AL-identified training set but also on the success of the labelling process itself, i.e.,

$$\hat{\mathbf{y}}' = \begin{cases} \mathcal{M}_t(\mathbf{X}' \mid \mathbf{D}_t^{\text{Train}}), & \text{if } R_\Omega = 1, \\ \mathcal{M}_t(\mathbf{X}' \mid \mathbf{D}_{t-1}^{\text{Train}}) & \text{otherwise.} \end{cases}$$

Hypothetically, we can quantify the impact that non-response has by considering the differential model performance (using some metric like the ROC AUC score) between (i) a model trained only on the data with responses and (ii) a full-response baseline where all queried examples are labelled. There are two pathways through which non-response may impact this performance:

1. Non-response leads to a reduction in the number of training samples for any round t. We call this the "*volume effect*". Since new examples are only added to the training data when there is a non-null label, the reduction in the volume of examples will reduce the ability of the model to improve, compared to the full-response AL model.
2. Non-response alters the distribution of training examples, relative to the full-response model, leading to an "*imbalance effect*". This imbalance affects the performance of the model at time t, though at the most general level it is unclear *how* it would do so: bias in the non-response could even (partially) cancel out the inherent selection bias of AL frameworks that we noted earlier.

Crucially, both the volume and imbalance mechanisms have knock-on effects on model performance in *subsequent* rounds of labelling, since the state of the model in period t determines the selection of new target labels in $t' > t$. Where non-response induces poor model performance on certain regions of the feature space, the querying function may seek to address this deficiency by oversampling this area in future rounds. For example, in a video recommender model, we only observe the watch time of a video when the user clicks it. The non-response rate, therefore, will be higher for items with low CTR. Oversampling these regions to try to get better data on watch time, by upranking these videos, simply results in low response rates in future queries.

Figure 1 summarises the hypothesised impact that non-response has on the AL framework. In short, a combination of volume and imbalance effects alters the selection of new unlabelled examples, which in turn leads to degradation in model performance. Given the sequential nature of AL training, this loss of performance has the potential to compound over multiple steps of the AL sequence.

4.1 Types of non-response

We contend that the mechanism(s) of non-response itself can impact *how* and the extent to which these two non-response dynamics – the training and selection effects – impact model performance. Here we focus on two types of non-response, applying intuitions from the wider inferential literature on missing data (Rubin 1976).

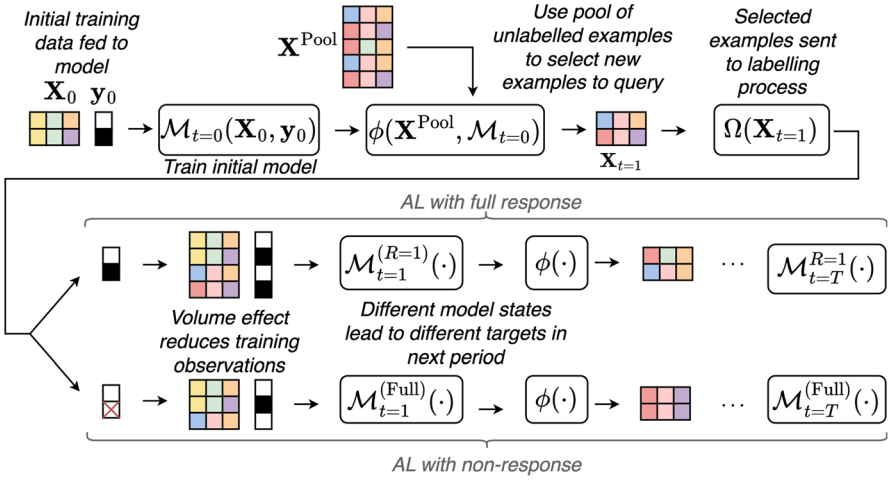

Fig. 1 The knock-on consequences of non-response on AL. From the same initial model, non-response leads to volume and imbalance effects in the AL sequence. Here, the result of these effects is the repeated querying of a non-responsive example. Colored blocks refer to data values and the red cross indicates a non-response label

Missing completely at random (MCAR) In the simplest case, the labelling process may be prone to corruption that is random in nature. For example, data loss incurred when streaming data due to unreliable networks or input errors made by human annotators may mean that labels are not always returned. Crucially, in these cases non-response is induced in ways unrelated to the subject of the measurement.

If non-response is MCAR, then:

$$P(R_\Omega = 1 \mid \mathbf{X}, \mathbf{y}, \phi) = P(R_\Omega = 1).$$

That is, the probability of missingness is distributed uniformly across the entire feature space and orthogonal to which unlabelled examples are queried. Therefore, while non-response will reduce the size of the training set, we would not expect any substantial imbalance effect as a result of non-response.

There are many reasons why AL may suffer from MCAR non-response. Consider the human annotator example discussed previously. If any requested labels are not logged by the end of the working day, suppose they are left unlabelled. If the order of labels given to the annotator is random, then this non-response is unrelated to feature values.

Missing at random (MAR) Alternatively, it may be that the non-response mechanism *is* related to $\mathbf{X}$ and/or $\mathbf{y}$. In these instances, the probability of non-response is not uniform across the feature space. Instead, and assuming non-response is explainable by observed features present in X, then:

$$P(R_\Omega = 1 \mid \mathbf{X}, \mathbf{y}, \phi) = P(R_\Omega = 1 \mid \mathbf{X}) \neq P(R_\Omega = 1).$$

Table 1 Summary of hypothesised effects of non-response in AL

		Effect	
Non-response	Cor(Informativeness, R_Ω)	*Volume*	*Imbalance*
MCAR	–	Yes	No
MAR	Negative	*Limited*	No
MAR	Positive	Yes	Yes

For example, consider the CTR and post-click setting described earlier. Our "query" might be an advertisement placed in a video carousel, where a label is assigned only if users click through from the advert to the product page. Importantly, features of the user may determine both whether the advert is shown (i.e., the user is queried) *and* whether the user clicks.

As Table 1 summarises, MAR non-response does not immediately introduce problematic bias into the model. Rather, the effect of missingness on model performance will depend on the interaction between the distribution of *informative* datapoints (at time t) and the (unknown) distribution of non-response.

On the one hand, the probability of non-response could be negatively related to the probability of being queried. In other words, some portion of the feature space may be more likely to return missing values, but also have low informativeness from an AL perspective. For example, younger users may be less likely to click through adverts, but given their age, the model is already confident about predicting these conversion outcomes. As a result, AL would favour requesting labels from other portions of the dataset, which are less likely to be affected by non-response. In which case, while there may be some limited volume effect due to a small number of non-responses, we would not expect a substantial biasing imbalance effect.

On the other hand, and perhaps more naturally, if some examples are both highly informative and have high rates of non-response, then *both* a volume and bias effect will impact model performance. This case can be acutely relevant to e-commerce and other content platforms. For example, if one set of items in the advertising model never gets clicked, then the model is likely to be uncertain about their conversion probabilities due to a lack of training examples, but to refine these estimates would exactly require recording clicks. More generally, higher non-response may be precisely why there is uncertainty in these regions, which over the course of sampling iterations leads to reduced informativeness in all parts of the feature space except here.

As in the MCAR case, non-response reduces the size of the training set, but now this effect is compounded by non-response occurring precisely for those unlabelled examples AL has identified as being most important for model improvement. In the case where the model is able to learn from other parts of the feature space, then over subsequent rounds the relative weight placed on querying this space will increase. Hence, the selection effect may force the model into repeatedly sampling from a region with high non-response rates, which may stall or degrade model improvement.

There is one form of missing data we do not consider in detail: data missing not-at-random (MNAR), where non-response is a result of *unobserved* features of the DGP. One particular manifestation of this phenomenon may be where the CTR model has features that the post-click model does not have. This case is particularly problematic as it would not be possible to model these relationships in the post-click model. We leave this case, as any correction applied to AL sequences will be dependent on the missing relationship being congenial with the data observed.

4.2 Performance difference effects under MAR and MCAR

In both the MAR *and* MCAR contexts, fewer responses from the querying function will impair the performance of the model (holding constant the number of training rounds). In the MAR case, however, the presence of local regions of non-response can further impact the model by unbalancing the training data (relative to the full-response model). The presence of this additional source of degradation, therefore, suggests that AL performance can be *worse* under MAR non-response compared to MCAR non-response.

The extent of this divergence, we hypothesise, depends on how skewed the non-response distribution is, and thus the extent to which the model is able to explore certain areas of the feature space. In the extreme case, suppose that there are inaccessible regions of the data that, if queried, never return a label – they act like "black holes" that absorb the entire exploration budget without returning any labels. In this case, the imbalance effect will be large, because despite the high informativeness of this region, the model is totally prohibited from improving in this area. As a result, holding constant the unconditional rate of non-response, we would expect a large differential in model performance between MAR and MCAR missingness.

This case can be contrasted against one where there is still imbalance in the non-response distribution, but where there is nevertheless *some* possibility of returning a label, and so some chance for model improvement. As the imbalance in the distribution of non-response lessens, and thus approaches the distribution under MCAR, then this differential should disappear. We demonstrate this expectation in the next section.

5 Adjusting for the probability of non-response

General approach In AL contexts without non-response, the "utility" of a label is its informativeness to subsequent model training, i.e., $U_{y,t+1} = \mathcal{I}(\mathbf{x} \mid \mathcal{M}_t)$. For example, in uncertainty sampling contexts, the most informative labels are defined as those from regions of the feature space which the model is most uncertain about. When there is the possibility of (random) non-response, however, the query utility should be conditional on both the probability of response and the resulting label's informativeness.

Hence, we propose a simple adjustment to how query targets are selected, by optimizing the *expected* utility (EU) of the informativeness score:

$$EU = \sum_{\mathbf{x} \in \mathbf{X}'} \mathcal{I}(\mathbf{x} \mid \mathcal{M}_t) \times \mathrm{P}(R_\Omega \mid \mathbf{x}), \; s.t. |\mathbf{X}'| = b, \tag{1}$$

where b is the total budgeted number of queries and $\mathbf{X}'$ are the query targets. Algorithm 1 details the implementation of this expected utility sampling strategy.[1] Note that non-response is realised "by nature".

Algorithm 1 Expected Utility Active Learning

Input $\mathbf{X}^{\text{pool}} \subset \mathbb{R}^{n \times k}$; $\mathbf{X}^{\text{train}} \subset \mathbb{R}^{m \times k}$; $\mathbf{y}^{\text{train}} \in \{0, 1\}^m$;

Target estimator $\mathcal{M}$ and non-response estimator $\mathcal{P}$; $n > 0$.

Output $\mathcal{M}_T$, a target model after T rounds

1: $\mathcal{M}_0 \leftarrow \mathcal{M}(\mathbf{X}^{\text{train}}, \mathbf{y}_{\text{train}})$
2: **for** t in 0:T **do**
3: $\quad \mathbf{u} \leftarrow \mathcal{I}(\mathbf{X}^{\text{pool}}, \mathcal{M}_t)$ // Calculate informativeness scores
4: $\quad \mathbf{p} = \mathcal{P}(\mathbf{X}^{\text{pool}})$ // Estimate response probabilities
5: $\quad \mathbf{i} \leftarrow \mathrm{TopM}\big(\log(\mathbf{u}) + \log(\mathbf{p}), n\big)$ // Choose indices with highest expected utility
6: $\quad \mathbf{X}^{\text{query}} \leftarrow \mathbf{X}^{\text{pool}}[\mathbf{i}]$
7: $\quad \mathbf{X}^{\text{train}} \leftarrow \mathbf{X}^{\text{train}} \cup \mathbf{X}^{\text{query}}[\mathbf{r}_t^{\text{query}} = 1]$ // Add successful queries to training data
8: $\quad \mathbf{y}^{\text{train}} \leftarrow \mathbf{y}^{\text{train}} \cup \mathbf{y}^{\text{query}}[\mathbf{r}_t^{\text{query}} = 1]$
9: $\quad \mathcal{M}_{t+1} = \mathcal{M}_t(\mathbf{X}^{\text{train}}, \mathbf{y}^{\text{train}})$ // Update model state
10: **end for**

By incorporating the cost of querying non-responsive regions of the feature space, this adaptation should prevent the model from "wasting" its budget on areas where labels are informative but the likelihood of observing one is very small.

The economic aspect of this correction has one further advantage: with sufficient training steps, the model will saturate in regions of the feature space where responses are likely. In these instances, the informativeness (however defined) will be so minimal that, despite high non-response rates, the model is able to switch to sampling solely from highly missing regions.

By design, therefore, this adjustment still grants the AL model leeway to explore regions of the feature space: as the model becomes more confident over regions with higher response rates, it becomes useful, and relatively less costly, to sample from low response regions. Model progress at this point will, clearly, be slow, but it may yield informative examples with a sufficient budget (and willingness to pay).

Non-response prediction error In practice, calculating the expected utility involves an *estimate* of the probability of response. In many settings, it may be possible to train or develop such a model before the active learning process. Note, for the purpose of training a CTR model, impressions without a click have a (negative) label. Therefore, the CTR model has more data available than the post-click model

[1] Following Lattimore (Lattimore and Szepesvári (2020) p.418), we use the TopM($\mathbf{a}, m$) operator to denote the largest m values in vector $\mathbf{a}$. This step has complexity $O(n \log m)$ for choosing m largest out of an array of length n, when implemented with a MinHeap. Note also line 5 sums the logged scores, which is equivalent to multiplying the estimates but more numerically stable.

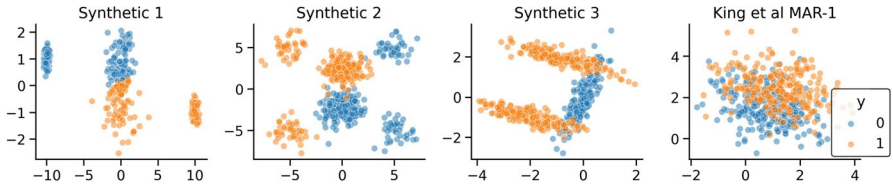

Fig. 2 Illustration of synthetic datasets used in AL experiments. Note: MAR-1 is a restricted view of the data, and only shows two of the five X dimensions

that we target in active learning. Hence, the CTR model may be accurate in regions where the post-click model is not.

That said, the non-response probabilities from such a model are nevertheless estimates. In some cases, these estimates may have (considerable) uncertainty. We may therefore want to explore these areas optimistically. This trade-off will be especially important, for example, where the non-response model is trained on a small number of observations, or where the non-response model is retrained iteratively during AL. We can address this issue by replacing the predicted probability of response $\mathbf{p}$ in Algorithm 1 with:

$$\mathbf{p} = \mathcal{Q}(\mathbf{X}^{pool}, 0.95),$$

where $\mathcal{Q}(\cdot, 0.95)$ returns the 95th quantile of the distribution of predictions for each $\mathbf{x} \in \mathbf{X}^{\text{pool}}$. This strategy resembles upper confidence bound (UCB) sampling, a common bandit algorithm. It requires an estimator that is capable of returning an (approximate) posterior distribution, like a Gaussian Process model or bootstrapped ensemble. For well-trained (i.e., precise) non-response models, note also that we would expect (and find) limited changes incorporating UCB, as the 95th quantile will be close to the predicted probability.

We refer to the final correction as the "*Upper Confidence Bound of the Expected Utility*" (UCB-EU). One major advantage of this strategy is that it is deliberately compatible with different base AL sampling strategies. Plausibly, any informativeness metric $\mathcal{I}$ and non-response estimator $\mathcal{P}$ can be plugged into Algorithm 1. As a result, the overall complexity class depends on the choices of $\mathcal{I}$ and $\mathcal{P}$. In our experiments, we demonstrate this approach using QbC, uncertainty, and random sampling strategies.

6 Empirical evidence of AL model degradation under non-response

6.1 Experimental setup

Synthetic data We focus on four synthetic data scenarios, as illustrated in Fig. 2. Synthetic 1 simulates a linear decision boundary with clusters offset in one dimension, such that $\mathbb{E}[Y] = 0.1$. Synthetic 2 and 3 are based on DGPs presented in (Huang et al. 2014), which have been demonstrated to prove challenging for AL strategies

like uncertainty sampling (Yang and Loog 2018). In Synthetic 2, the data contain six normally distributed clusters in two dimensions, with $\mathbb{E}[Y] = 0.5$. In Synthetic 3, the data are distributed in a rotated U-shape where the distribution of positive cases intersects the tails of the two other sides. Finally, King et al's MAR-1 dataset 2001 considers a multivariate normal distribution with moderate correlations between five dimensions, based on a simulation design used in inference-focused missing data studies (King et al. 2001; Lall and Robinson 2022). We convert this final scenario into a classification problem by implementing a non-linear decision boundary: $y_i \leftarrow 5X_0 - 4X_1 + 3X_2 - 2X_3 + X_4 + 0.5X_0^2 + 3X_1X_2 >= c$, setting c such that $\mathbb{E}[Y] = 0.1$.

Non-response mechanisms As noted in Sect. 4, we consider three non-response mechanisms. First, that there is no non-response and so we observe the full data (i.e., $R_\Omega = 1$ across the entire feature space). This mechanism is the benchmark, or ideal, data generating context and yields the full-response model. Second, that the data is MCAR. This mechanism helps understand the volume effect of training (relative to the full data mechanism), since there is no correlation between the probability of non-response and the feature space. To model MCAR missingness, we randomly induce non-response uniformly across the feature space. Third, that the data is MAR, where missingness is a function of the observed data. In our experiments, we partition each DGP's feature space into two regions at some threshold value along its first explanatory dimension. Therefore, the missingness is correlated with the value of this dimension. On one side of this threshold, the "low response region", we impose a high probability of *non*-response, $P(R_\Omega = 1) = 0.001$, and on the other side we impose a high probability of response. We hold constant the unconditional missingness probability (to match the MCAR mechanism), by adjusting the specific threshold for the two missing regions.[2]

Simulations For each scenario and non-response mechanism, we simulate 50 rounds of AL. Each model is seeded with two random examples. In each round, the model queries the 10 most informative unlabelled examples. We repeat each simulation 200 times to calculate the expected performance and 95% confidence intervals for each step of training, assessed using the ROC-AUC score on 1000 holdout examples.

We run identical versions of the experiment using QbC, uncertainty, and random sampling strategies. QbC is a commonly used query strategy, and it largely addresses documented deficiencies of simpler methods like uncertainty sampling (Settles 2009). Our QbC model uses a random forest classifier as the ensemble. We also benchmark uncertainty sampling as a simple form of AL strategy. The uncertainty sampler uses a linear support vector machine (SVM) as the learning algorithm. Finally, we include random sampling as a naive acquisition strategy to demonstrate the generality of our correction.[3]

[2] Given some defined probability of response in the first region, p_1, and targeted unconditional response probability $p*$, we assign a response probability $p_2 = 1 - p_1(1 - p*)$ to the other region.

[3] All simulations were run on a single server instance, using GPU processing equivalent to an NVIDIA Tesla P100 GPU with 16GB memory. The total time taken to run 200 simulations for every combination of missing mechanism, strategy, and DGP was 12 h. The non-response model for each DGP took approximately 5 min to train.

Non-response corrections We implement our proposed algorithm, for each AL strategy, as follows:

- For *QbC* sampling, we use McCallum et al. (McCallum et al. 1998)'s modified acquisition function, which is based on the log of the maximum Kullback–Leibler divergence for each label to the log of the UCB predicted probability of response
- For *uncertainty* sampling, we add the log entropy of the predictions over the pool examples to the log of the UCB predicted probability of response
- For *random* sampling, we generate the UCB predicted upper confidence bound and softmax these values so the scores sum to 1. n new observations are then randomly selected according to this vector of sampling probabilities

To model the probability of non-response, and to perform UCB sampling of a posterior, we pre-train Gaussian Process (GP) models for each simulation DGP.[4]

6.2 Results

We first consider the uncorrected impact of non-response on AL performance (Sect. 6.2.1). We then demonstrate how our solution improves on these results (Sect. 6.2.2), before exploring in more detail cases where non-response poses particularly acute problems for active learning (Sect. 6.3).

6.2.1 Uncorrected impacts of non-response

Figure 3 plots the simulation results without any correction. Across all four synthetic DGPs, we see that it is the MAR form of missingness that is more harmful to active learning than other non-response mechanisms. Over sequential training steps, the MAR-affected model either fails to improve or its performance even *deteriorates* as we collect more data. This effect is particularly clear in the Synthetic 3 scenario, where the inability to query parts of the feature space biases the model and leads to worse performance over sequential training, under both uncertainty and random sampling.

Across all scenarios, sequential steps yield model improvements under MCAR non-response, but the extent of the difference between this model and the full data model, on average, differs considerably. We do observe the effect of having a lower volume of training data under both the Synthetic 1 and MAR-1 DGPs, with MCAR performance significantly worse than the full data model. This difference is less pronounced under QbC, which is relatively less affected by MCAR non-response. In Synthetic 3, MCAR performance is substantially poorer, by 50 steps all three

[4] We implemented the GP models using `GPyTorch` (Gardner et al. 2018). We fit a constant prior mean function and scaled RBF kernel for the covariance matrix, using a Bernoulli Likelihood function. Since our MAR mechanism is relatively simple, we find these models prove to be very effective classifiers: all models have ROC-AUC values above 0.995 and mean absolute errors at or below 0.04.

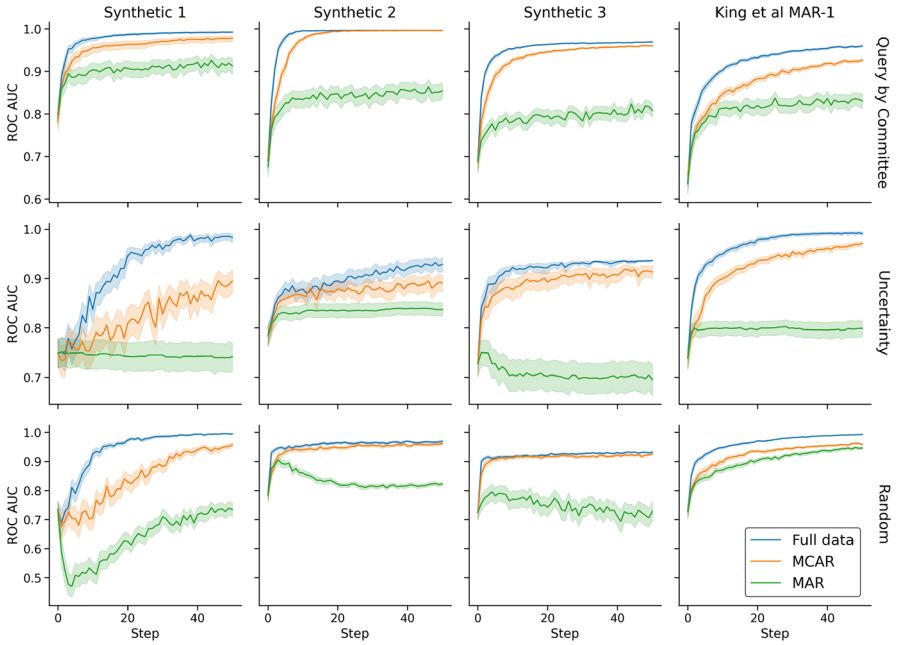

Fig. 3 AL model performance in the presence of non-response, using different sampling strategies. $\mathbb{E}[R] = 0.3$ across all simulations. Observations in the missing region had a 0.001 probability of response. Shaded areas show the 95% confidence interval over 200 separate simulations (per non-response mechanism)

sampling strategies see similar performance to the full data model, although there are distinguishable differences earlier in training.

We also hypothesised that the difference in model performance between MAR and MCAR non-response mechanisms results from an imbalance effect. In the most extreme case, when regions of the feature space preclude label acquisition entirely, AL strategies simply cannot learn about these regions and thus may perform poorly.

We test this hypothesis by adjusting the probabilities of non-response either side of the non-response threshold. To hold constant the volume effect inherent to increased missingness, we adjust the cut-off position to maintain the same unconditional probability of response ($\mathbb{E}[R] = 0.3$). We run 200 simulations per missingness mechanism for each of the five different non-response probability tuples. Where the severity of the non-response threshold is most acute (i.e., fully observed on one side, and never-observed on the other) we would expect the largest difference in performance. Conversely, as the severity of this threshold decreases, such that the non-response probabilities approach uniformity, we would expect MCAR and MAR models to converge in performance.

Figure 4 reports the average results over these simulations, using an uncertainty sampling strategy. The MCAR-afflicted model improves (slowly) because the missingness probability is constant across the *entire* feature space. By contrast, under MAR missingness it is *impossible* to learn about some portion of the feature space.

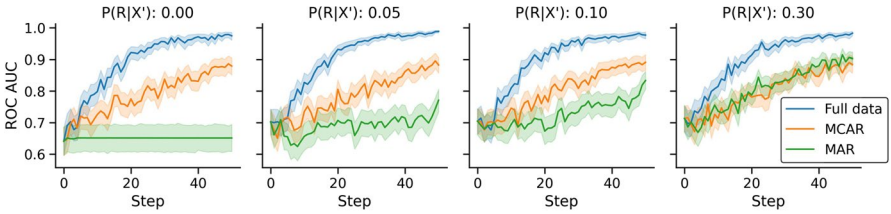

Fig. 4 The effect of imbalance on model performance between MAR and MCAR non-response mechanisms. The probabilities above each panel indicate the probability of response in the low response region of the feature space. Shaded areas show 95% confidence intervals over 200 simulations

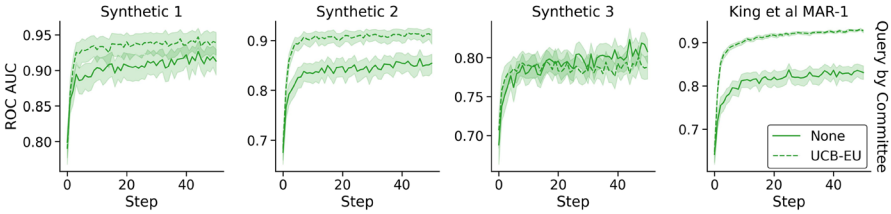

Fig. 5 Comparison of Query-by-Committee AL performance with and without UCB-EU correction, under MAR non-response. The DGP is identical to the results presented in Fig. 3

This becomes increasingly pronounced for low values of $P(R_\Omega = 1 \mid \mathbf{X})$ in the low-response region, in which case it results in flat learning curves and large gaps between MAR and MCAR. As we increase $P(R_\Omega = 1 \mid \mathbf{X})$ the MAR/MCAR gap decreases, such that in the final panel, where the probability of response in the low response region is the same as the marginal rate of response, MAR and MCAR performance are similar.

6.2.2 MAR impacts with UCB-EU correction

Compared to the uncorrected results, the modified algorithm shows better performance for many combinations of strategy and DGP. As shown in Fig. 5, implementing a cost to searching low response regions improves the ability of the QbC strategy to refine its selection of unlabelled examples to query. In all but the Synthetic 3 DGP, using the UCB-EU correction leads to substantially better model performance.

Figure 6 plots the results of using the UCB-EU correction for uncertainty and random sampling strategies, respectively. In the case of random sampling, the UCB-EU correction also yields considerably better model performance, although the naivety of the baseline strategy appears to add additional noise to the learning process early in training. By around 30 steps of AL training, and similar to the more performant QbC strategy, the UCB-EU corrected models outperform the baseline model under the Synthetic 1, 2, and MAR-1 DGPs.

In the case of uncertainty sampling, the performance improvements are less pronounced. However, we observe that uncertainty sampling consistently underperforms random sampling on these datasets. This result is not surprising given existing

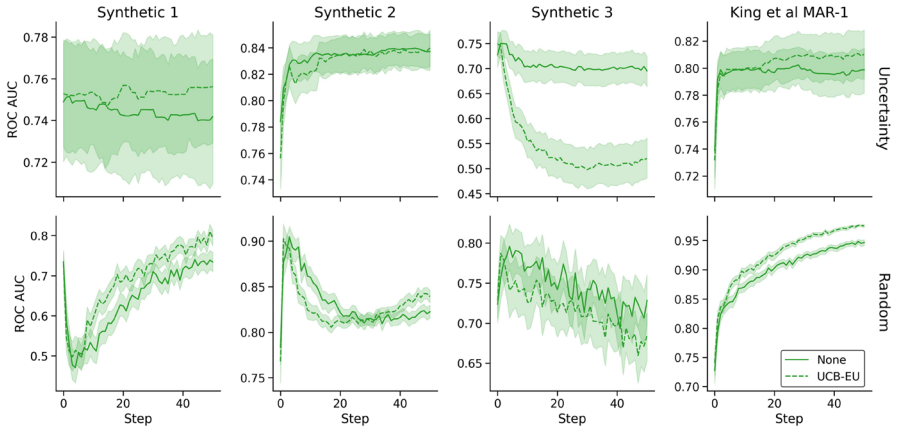

Fig. 6 Comparison of Uncertainty and Random AL performance with and without UCB-EU correction, under MAR non-response. The DGP is identical to the results presented in Fig. 3

literature: it is widely documented that uncertainty sampling does not always outperform random sampling (Attenberg and Provost 2011; Settles 2012; Yang and Loog 2018; Jin et al. 2022; Tifrea et al. 2023). We cannot expect UCB-EU non-response corrections to make a sampling strategy sample efficient under non-response settings, if that sampling strategy is not sample efficient without non-response to start with. Therefore, we believe this is failure of uncertainty sampling itself rather than a failure of the UCB-EU adjustment.

6.3 Where should the model query?

MAR non-response on Synthetic 3 leads to *worse* performance under our modified algorithm. This result is most pronounced in the case of uncertainty sampling. We conjecture that the degradation in performance observed under the Synthetic 3 DGP is a result of non-response forcing the model to fit on a low non-response part of feature subspace that is non-representative of the general population (and hence, of the test set distribution). In other words, modifying the AL sequence to discount the utility of low-response regions may narrow the model's focus too much, therefore learning a "local" decision boundary that is optimised only for the low non-response parts. This boundary is quite different from the global optimal boundary that would result in performance similar to the full data model.

We can illustrate this point in two ways. First, we abstract away from AL and simply assume areas of the feature space are fully observed ($R_{\Omega(\mathbf{x})} = 1$) or never observed ($R_{\Omega(\mathbf{x}')} = 0$). We take a large N sample and train a target model on this data, which should approximate the long-run performance of the AL model with many queries from the pool. We conduct this exercise four times, varying where the non-response threshold is applied, which has the effect of varying the proportion of the DGP that is observable.

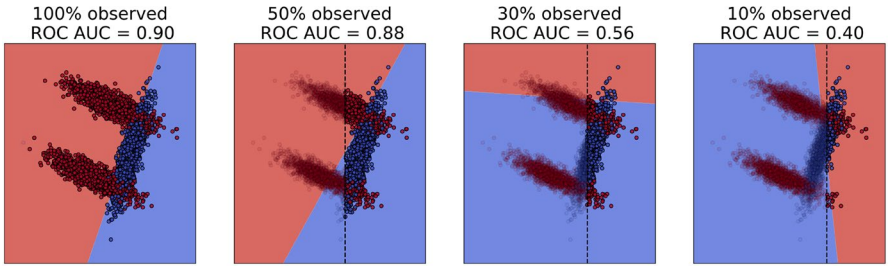

Fig. 7 The effect of non-response on estimated decision boundaries in Synthetic 3 DGP. Dashed vertical lines indicate the threshold between low-response (left) and full-response (right) regions of the covariate space. Blurred datapoints indicate specific examples of the underlying DGP that are not available to the model during fitting. All models are trained on 6000 observations, to minimise uncertainty over the decision boundary

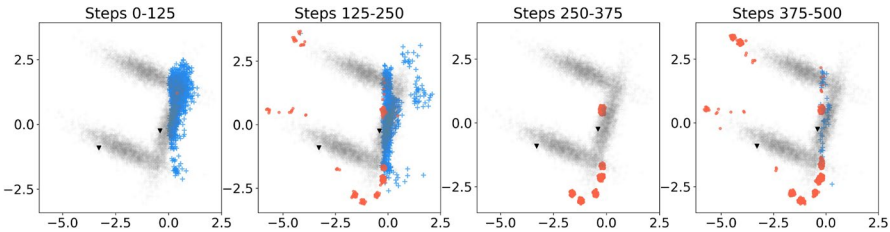

Fig. 8 Query history using the QbC strategy and UCB-EU correction with Synthetic 3 DGP. The model is trained for 500 steps with a batch size of 10 and two initial labelled examples (black daggers). The underlying DGP is shown as a faint distribution of grey dots in the background. Blue crosses indicate the queried label was observed, and red circles indicate non-response. Each facet shows the (binned) queries within the sequence of AL iterations

Figure 7 plots the DGP and the estimated decision boundaries in each case. The leftmost panel displays the optimal linear decision boundary with zero non-response. The dashed vertical line indicates the boundary between the low-response (left) and full-response (right) regions of covariate space. While a linear boundary cannot perfectly distinguish labels under this DGP's distribution, the ROC AUC score is nevertheless high with a 100% response rate. As the response rate decreases, the decision boundary rotates as it fits to the narrower set of points in the high-response region on the right side of the dashed line. The optimal decision boundary, conditional on the observed data, becomes increasingly different from the "true" decision boundary as the proportion of non-response increases. By the third panel, the decision boundary is mis-classifying the entire bottom wing of the DGP distribution. By the final panel, moreover, the classification boundary has inverted. Unsurprisingly, the ROC AUC scores decline substantially as non-response affects more of the area. These results confirm our conjecture that in the "ideal" world where an AL sequence focuses only on *observable* examples, this myopia can lead UCB-EU to exacerbate the selection bias effects that are always present when using active learning. When response probabilities

become infinitesimally small, the harm from this selection bias may outweigh the benefits of UCB-EU that are obtained from increased annotation volume relative to uncorrected AL.

Next, we verify this expectation by tracking the query targets in our AL experimentation setup, using the UCB-EU correction and QbC sampling strategy. Figure 8 plots each individual query attempt and whether the resulting label was observed. We bin the AL iterations into sequential facets, to show *when* in the sequence each query was made. In early iterations of training, the AL model pays almost exclusive attention to the high response region of the feature space. In particular, in steps 1 to 125, it focuses on the top-right section (right of where the dashed line was). This is the result of the UCB-EU correction: when the model has few observations, the most informative queries, and those with the highest response probabilities, are in this region. As the model reaches the point where it has exhausted the learning potential in this region, it switches its focus to attempt to learn in the low-response parts of covariate space where the model does not yet have much training data. Consistent with Fig. 7, we now observe selection bias induced by non-response. The UCB-EU correction prevents the AL sequence from “wasting" early queries on parts of the feature space where the probability of getting a valid label are very small, but which would ultimately yield a different (and better) decision boundary if observed.

7 Case study: taobao shopping behaviour modelling

Consider an e-commerce platform that ranks products that are presented as a list of items on its homepage. A click on an item in this ranking brings the user to a product details page with a checkout button. The aim is to rank products in a way that maximises the conversions on the platform. Imagine that the product ranking system is designed to factorize the estimation task of conversion probabilities into two separate machine learning models: $P(conversion) = P(click = 1)P(conversion \mid click = 1)$, i.e., a *CTR* model that estimates the probability that a user will click on a product in the ranking, and a so-called *post-click* model that estimates the probability that a user who clicked on the product and lands on the product details page will proceed to purchase the item. Factorization into separate CTR and post-click models is common in the e-commerce and online advertising industry (Ma et al. 2018; Lin et al. 2023; Barbieri et al. 2016; Rosales et al. 2012). The resulting quantity $P(conversion)$ yields the probability that a user will purchase an item in the ranking, and serves as our sorting criterion for ranking purposes. The CTR model can be trained on a dataset of all items that users viewed, while the post-click model can only be trained on a data set of page loads of the product details page (hence, on clicked items).

Imagine a setting where we aim to improve the *post-click* model component through AL exploration, and to achieve this exploration we can intervene in what products we display in the product ranking. In this setting, whether a user clicks the

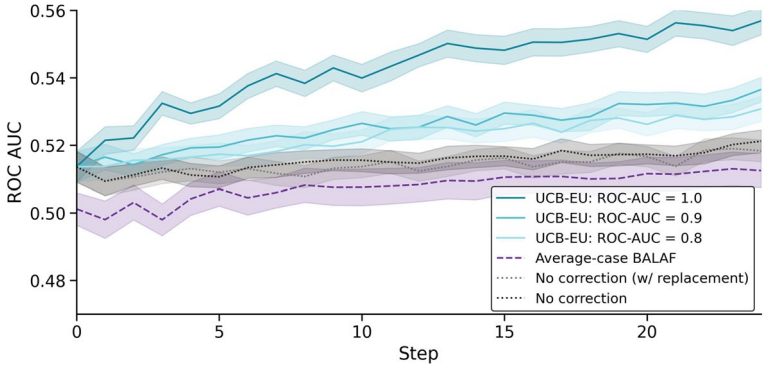

Fig. 9 Target model performance of UCB-EU AL-trained models compared to other algorithms. Lines plot the mean ROC-AUC at each step and the shaded areas plot the 95% confidence intervals

item or not constitutes the non-response mechanism, i.e., we do not obtain a label for the post-click model if the user does not click.[5]

We use the large-scale Taobao (Tianchi 2018) dataset that consists of product impressions, clicks, and conversions like in the setup that we described above. We set out to assess whether our proposed method helps correct for the potential non-response bias in this context. This data includes a random sample of 1.1 million users and their corresponding behaviours on the platform between 6-13 July 2017 (n=700 million behaviour logs). We use a combination of the user, product, and behaviour data so that we can record which users clicked which product (the labelling process) and which users purchased the product (the post-click target model).

To simulate the AL process, we initialise a random forest model with 50 random observations. We conduct 25 steps of AL, and at each step query 5000 new examples from the pool (n=12.3 million user-product-behaviour triples), to simulate impressions by users. We realise non-response using the *observed* CTR indicator from the data. At the end of each step, *all* queried observations are removed from the pool to mimic the temporary nature of the impression. We also test a version of this simulation, approximating the re-querying strategy of Yan et al. (2015), where queried observations with non-response are replaced in the pool rather than removed. We holdout a test set of 10,000 observations (*post-click* user-item pairs).

Finally, to assess how the performance of the non-response (i.e., CTR) model affects the improvement produced by UCB-EU, we run our simulation multiple times, implementing a series of "synthetic" CTR models where we deliberately vary the model's ROC-AUC score across simulations. These CTR models use, as their base prediction, the true click-through scores (from the pool). We then corrupt this vector, by inverting 1 − ROC-AUC labels randomly, to generate CTR models of

[5] The absence of a click should be interpreted as missing data for the $P(conversion \mid click = 1)$ estimator and not as a negative label. Including non-clicked impressions as negative labels would result in a direct estimation of the marginal $P(conversion)$. Our aim here, instead, is to estimate the conditional probability $P(conversion \mid click = 1)$ and then estimate $P(conversion)$ by multiplying the conditional quantity by the estimated CTR, as is common in industry.

varying performance. We compare our general method to the average case Bayesian Active Learning with Abstention Feedbacks (BALAF) proposed by Nguyen et al. (2022). This approach uses L_2-regularised logistic regression models to approximate learning the posterior of the target and non-response models, using *maximum a posteriori* estimation, and we pre-train the missing model using the entire pool.

Figure 9 displays the results of each scenario, averaged over 100 independent simulations of each AL process. Compared to the uncorrected QbC model, using UCB-EU for 25 steps of AL can yield up to a 10% improvement in the model's ROC-AUC relative to AL without the UCB-EU correction. However, the quality of the non-response model matters: less accurate non-response models result in smaller gains from UCB-EU as the adjustments become noisy.

Replacing queried datapoints has very little impact on model performance, and is still markedly worse than the UCB-EU correction. The average-case BALAF strategy results are also largely indistinguishable from the uncorrected AL model as the confidence intervals largely overlap. We believe that this result is unsurprising: as the authors themselves note, the performance of BALAF breaks down when the rate of non-response is high.

8 Discussion and conclusion

Non-response in the feature space limits model performance. In this paper, we extend our understanding of this process by demonstrating that the *form* of non-response, and its potentially *correlated* relation with the underlying DGP, leads to differential impacts on model performance, and can even degrade model performance over time. Biased non-response of this type may be particularly common where AL involves user interactions. We demonstrate that we can mitigate this loss in performance by adjusting the utility of querying labels by the estimated probability of non-response.

Importantly, we find, however, that non-response distributions can lead to local decision boundaries inconsistent with the theoretically global optima. This is a challenging problem for AL methods, which deserves attention in future research. It is also worth noting that, at any given step, our correction may yield labels that are less intrinsically informative but which avoid the model learning nothing by receiving a null label. Therefore, while our approach should have improved performance relative to a naive AL model, it faces the hard constraint that non-response may ultimately curtail the informativeness of the data available to the model.

Finally, our experiments use a high non-response probability. This severity creates a sharp threshold in the feature space, and likely contributes to the limited effect our correction has in some contexts (relative to random non-response). As we demonstrate, separately, where the probability of response is higher (even under MAR conditions), the imbalance effect on model performance becomes relatively less substantial. We do, however, think that contexts of high (or even

total) non-response are common in areas where, for example, labelling requires costly user behaviour.

Funding No funding was received to assist with the preparation of this manuscript.

Availability of data and materials All simulation data-generating processes are described in the paper. All data used in Sect. 7 is available at https://tianchi.aliyun.com/dataset/56.

Declarations

Competing interest The authors have no relevant financial/non-financial interests.

Ethics approval and consent to participate There are no ethical issues. This research was considered and approved by the London School of Economics' research ethics procedure (ref. 183654).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Amin K, DeSalvo G, Rostamizadeh A (2021) Learning with labeling induced abstentions. In: Advances in Neural Information Processing Systems, pp 12576–12586

Attenberg J, Provost F (2011) Inactive learning? difficulties employing active learning in practice. ACM SIGKDD Explorations Newsl 12(2):36–41

Audibert JY, Bubeck S, Munos R (2010) Best arm identification in multi-armed bandits. In: COLT, pp 41–53

Barbieri N, Silvestri F, Lalmas M (2016) Improving post-click user engagement on native ads via survival analysis. In: Proceedings of the 25th International Conference on World Wide Web, pp 761–770

Bartók G, Foster DP, Pál D et al (2014) Partial monitoring-classification, regret bounds, and algorithms. Math Oper Res 39(4):967–997

Carcillo F, Le Borgne YA, Caelen O et al (2018) Streaming active learning strategies for real-life credit card fraud detection: assessment and visualization. Int J Data Sci Anal 5:285–300

Cortes C, DeSalvo G, Gentile C, et al (2018) Online learning with abstention. In: International conference on machine learning, pp 1059–1067

Elahi M, Ricci F, Rubens N (2016) A survey of active learning in collaborative filtering recommender systems. Comput Sci Rev 20:29–50

Fang M, Zhu X, Zhang C (2012) Active learning from oracle with knowledge blind spot. In: Twenty-Sixth AAAI Conference on Artificial Intelligence

Farquhar S, Gal Y, Rainforth T (2021) On statistical bias in active learning: How and when to fix it. arXiv preprint arXiv:2101.11665

Freund Y, Seung HS, Shamir E et al (1997) Selective sampling using the query by committee algorithm. Mach Learn 28(2–3):133

Gardner J, Pleiss G, Weinberger KQ, et al (2018) GPyTorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration. In: Advances in neural information processing systems

Hansen MH, Hurwitz WN (1946) The problem of non-response in sample surveys. J Am Stat Assoc 41(236):517–529

Huang SJ, Jin R, Zhou ZH (2014) Active learning by querying informative and representative examples. IEEE Trans Pattern Anal Mach Intell 36(10):1936–1949

Jin Q, Yuan M, Li S et al (2022) Cold-start active learning for image classification. Inf Sci 616:16–36

King G, Honaker J, Joseph A et al (2001) Analyzing incomplete political science data: an alternative algorithm for multiple imputation. Am Polit Sci Rev 95(1):49–69

Lall R, Robinson T (2022) The midas touch: Accurate and scalable missing-data imputation with deep learning. Polit Anal 30(2):179–196

Lattimore T, Szepesvári C (2020) Bandit algorithms. Cambridge University Press

Lewis DD (1995) A sequential algorithm for training text classifiers: Corrigendum and additional data. In: ACM SIGIR Forum, pp 13–19

Lin C, Mausam M, Weld D (2016) Re-active learning: Active learning with relabeling. In: Proceedings of the AAAI Conference on Artificial Intelligence

Lin X, Chen X, Song L, et al (2023) Tree based progressive regression model for watch-time prediction in short-video recommendation. arXiv preprint arXiv:2306.03392

Little RJ, Rubin DB (2019) Statistical analysis with missing data, vol 793. John Wiley & Sons

Ma X, Zhao L, Huang G, et al (2018) Entire space multi-task model: An effective approach for estimating post-click conversion rate. In: Proceedings of the International ACM SIGIR Conference on Research & Development in Information Retrieval, pp 1137–1140

McCallum A, Nigam K, et al (1998) Employing EM and pool-based active learning for text classification. In: ICML, pp 350–358

Mohan K, Pearl J, Tian J (2013) Graphical models for inference with missing data

Nguyen CV, Ho LST, Xu H et al (2022) Bayesian active learning with abstention feedbacks. Neurocomputing 471:242–250

Nguyen VA, Shi P, Ramakrishnan J, et al (2020) CLARA: confidence of labels and raters. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp 2542–2552

Rosales R, Cheng H, Manavoglu E (2012) Post-click conversion modeling and analysis for non-guaranteed delivery display advertising. In: Proceedings of the fifth ACM international conference on Web search and data mining, pp 293–302

Rubin DB (1976) Inference and missing data. Biometrika 63(3):581–592

Settles B (2009) Active learning literature survey technical report. University of Wisconsin-Madison Department of Computer Sciences

Settles B (2012) Uncertainty sampling. In: Active Learning. Springer, p 11–20

Seung HS, Opper M, Sompolinsky H (1992) Query by committee. In: Proceedings of the fifth annual workshop on Computational learning theory, pp 287–294

Sheng VS, Provost F, Ipeirotis PG (2008) Get another label? improving data quality and data mining using multiple, noisy labelers. In: Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining, pp 614–622

Stekhoven DJ, Bühlmann P (2012) Missforest-non-parametric missing value imputation for mixed-type data. Bioinformatics 28(1):112–118

Tax N, de Vries KJ, de Jong M et al (2021) (2021) Machine learning for fraud detection in e-commerce: A research agenda. Deployable Machine Learning for Security Defense: Second International Workshop, MLHat 2021, Virtual Event, August 15. Springer, pp 30–54

Tianchi (2018) Ad display/click data on taobao.com. https://tianchi.aliyun.com/dataset/dataDetail?dataId=56

Tifrea A, Clarysse J, Yang F (2023) Margin-based sampling in high dimensions: When being active is less efficient than staying passive. In: International Conference on Machine Learning (ICML), vol 202. PMLR, pp 34222–34262

Yan S, Chaudhuri K, Javidi T (2015) Active learning from noisy and abstention feedback. In: 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp 1352–1357

Yan S, Chaudhuri K, Javidi T (2016) Active learning from imperfect labelers. In: Advances in Neural Information Processing Systems

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

Zhao L, Sukthankar G, Sukthankar R (2011) Incremental relabeling for active learning with noisy crowdsourced annotations. In: 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing, IEEE, pp 728–733

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Thomas S. Robinson[1,2] · Niek Tax[2] · Richard Mudd[2] · Ido Guy[3]

✉ Thomas S. Robinson
t.robinson7@lse.ac.uk

Niek Tax
niek@meta.com

Richard Mudd
rmudd@meta.com

Ido Guy
idoguy@meta.com

1 Department of Methodology, The London School of Economics and Political Science, London, UK

2 Central Applied Science, Meta, London, UK

3 Central Applied Science, Meta, Tel Aviv, Israel