Strategies for Truth Discovery under Resource Constraints

ABSTRACT
We present a decision-theory based approach for efficiently sampling information sources in resource-constrained environments, where there is uncertainty regarding source trustworthiness. We exploit diversity among sources to stratify the population into homogeneous subgroups to both minimise redundant sampling and mitigate the effect of certain biases (e.g., source collusion). After presenting our formal framework, we show empirically that our approach performs as well as existing truth discovery approaches with respect to accuracy, while significantly reducing sampling cost.

Categories and Subject Descriptors
H.4 [Information Systems Applications]: Miscellaneous

General Terms
Experimentation, Performance

Keywords
Trust, Truth Discovery, Diversity, Sampling

1. INTRODUCTION
The tremendous surge in the number and variety of information sources available to support decision-making calls for efficient methods of harnessing their potential. Information sources may be unreliable, and misleading reports can affect decisions. Truth discovery exists within this context, as an essential activity in multi-agent systems, whereby agents try to infer the true state of the world from potentially different or conflicting information [15].

Truth discovery mechanisms [16, 24, 25] rely on reports from as many sources as possible. In many real-world contexts, however, capturing and distributing evidence can be costly in terms of energy, bandwidth, and delay overheads. Relying on responses from all possible sources may adversely affect the utility of an agent such as when the cost of sampling some sources outweighs the value derived. In addition, there is often no guarantee that reports obtained from different sources are based on direct, independent observations. Social influence in online environments, for example, may bias collective opinions [17], introducing an extra challenge of distinguishing fact from rumour. The question then arises, and which we address in this paper, as to how to efficiently sample the source population to estimate ground truth.

The use of strategies such as stratified sampling [6] is widely employed in conducting surveys, and has been shown to perform well in applications such as social media analytics [18]. Stratified sampling involves partitioning a population into disjoint subgroups with members having similar values of one or more stratification variables (geography, culture, age, etc.). The values of these variables are usually known, and are assumed to be correlated with the target parameter. Samples are taken from each subgroup independently to estimate the target parameter. The advantage of this technique is that it usually provides a good representation of the population with fewer samples [7]. In addition, the diversity reflected in the stratification may be exploited to make better assessments, for example, by obtaining evidence from groups with potentially different perspectives on a problem. This has the potential to mitigate the risk of double-counting evidence due to correlated biases among group members.

In this research, we are interested in how to combine a form of stratification and trust assessment in order to optimise the selection of sources, an approach which allows us to effectively manage truth discovery under resource constraints. Taking stratified sampling as our point of departure, we exploit diversity among sources to form groups, made up of sources likely to provide similar reports. We then use reinforcement learning to identify effective sampling strategies across groups. In this way, we relax important limiting assumptions underlying truth discovery and trust approaches: the greater the number of reports acquired, the better the estimate; reports are independent; acquiring reports is cost-free. We show empirically that our model, hereafter referred to as DRIL (Diversity modelling and Reinforcement Learning), performs as well as classical trust approaches in estimating ground truth, but by sampling significantly fewer sources. If sampling has a concomitant cost, the combination of diversity modelling and learning sampling strategies in DRIL significantly increases the overall value of information [5] to an agent.

2. RELATED WORK
The majority of the approaches for addressing the truth discovery problem in multi-agent systems are variants of trust and reputation systems, the focus of which is to model sources’ behaviour and assess the quality of information they
provide. The Beta Reputation System (BRS) [12] uses Beta probability density functions to model the reliability (or trustworthiness) of a source. This probability can be used to normalise reports provided by the source. TRAVOS [24] uses personal observations about sources to estimate their trustworthiness. TRAVOS assumes that reports that do not align with the agent’s perspectives are biased. This assumption can be problematic, however, when considering the fact that an agent may not be in a position to observe the reported state of the world. Even if this were possible, an agent’s imperfect or subjective view cannot be regarded as a suitable metric for assessing the reliability of other sources with respect to reporting an objective fact. In addition, modelling the behaviour of every single source in large multi-agent systems may be impracticable in some settings.

Other approaches exploit the statistical properties of the reports themselves as a basis for truth discovery. These approaches tend to filter out, or discard reports that deviate from mainstream opinion. For example, [16] uses a clustering mechanism to filter outliers. [25] utilises an iterative filtering technique, which is based on the assumption that a source is reliable if it often provides true information, and a piece of information is likely to be reliable or true if it is provided by many “reliable” sources. A major limitation of these approaches is the reliance on large numbers of reports in order to obtain reasonable results in their estimates. Thus, they are usually ill-equipped to function in environments where costs are associated to queries. Also, by making a strong assumption about the proportion of reliable sources (i.e., the majority rule), these approaches may function poorly in some settings. For instance, outlying sources might not be liars, but rather individuals or groups with different (but not necessarily biased) viewpoints. In addition, the system may contain malicious sources, which collude to distort mainstream opinion [19].

Whereas trust and reputation mechanisms focus on maximising the quality of estimates of ground truth, cost minimisation or the need to balance the trade-off between information quality and cost has been the focus of other approaches. These tend to regard truth discovery as an optimisation problem. Notable examples in this area include [10, 14, 23], which formulate truth discovery as a Markov Decision Process (MDP). Although MDPs and related techniques (e.g., multi-armed bandits [2]) provide a sound mathematical framework for modelling the problem of choice under uncertainty, they suffer from complexity issues and tend to be overwhelmed by the explosion of the problem space. As a result, these approaches can only be applied to problems with limited scale. Still within this context, active learning [3] presents an interesting body of work, the goal of which is to sample a distribution (or group of sources) proportionate to the variance of members’ reports. Active learning has been applied to stratified sampling and multi-armed bandits for optimum allocation [2, 3, 9]. Given that these approaches do not take the reliability of groups into account, more sampling effort may be appropriated to unreliable groups, which may in turn impact truth discovery.

As mentioned, the focus of this research is to strategically sample information sources for opinions under resource constraints to acquire an accurate estimate of ground truth. In doing so, we overcome some of the challenges faced by current approaches briefly described above.

3. TRUTH DISCOVERY STRATEGIES

Figure 1 provides an overview of our model. Various information sources may be sampled, which we cluster into diverse groups (Layer 1). We then seek a strategy for sampling from these groups (Layer 2), which allows us to maximise the likelihood of determining an accurate estimate of ground truth. Finally, in (Layer 3) we combine the evidence obtained from sources sampled to estimate ground truth (truth discovery). In this section, we formalise our model, emphasising the core contribution of the paper: learning sampling strategies, which constitutes the function represented by Layer 2 in Figure 1.

We assume an agent that has the task of monitoring an environmental state, θ (e.g., the weather condition at a location, or the number of casualties following a disaster). The domain of the variable θ may be different for different query types, such as “is it snowing?” and “how many casualties?”. For a particular query, Θ represents the set of possible values of θ. The value of θ ∈ Θ can change over time, and the agent must therefore repeatedly update its estimate, ˆθ ∈ Θ, of the environmental state at time t, denoted ˆθt. Consequently, the estimate of θ at time t is denoted ˆθt.

To acquire an estimate of the environmental state, sources of varying trustworthiness may be queried, the result of which will be a set of reports from the selected sources. Let 𝑁 = {1, ..., n} denote the set of sources known to the agent. The report received from a single source, i ∈ 𝑁 at time t regarding θt is denoted oit. In our evaluation (see Section 4), we assume that reports are continuous (e.g., temperature readings), such that θt ∈ ℝ.

In querying the sources, 𝑁 ⊆ 𝑁 being a subset of the sources selected, the agent incurs cost. We define sampling cost as a function: cost : 2N → ℝ. In many settings, sampling costs are strictly additive: cost(N) = ∑i∈N cost({i}). Our goal is to optimise the selection of sources by learning subsets of 𝑁 (a diversity structure), and then deciding how to query those subsets.

As illustrated in Figure 1, our model attempts to cluster information sources into groups of similar sources (e.g., different individuals from the same organisation). Prior to presenting our method of learning sampling strategies, we discuss how such groups are identified.

3.1 Modelling Source Diversity

The diversity modelling we consider here is a technique through which a source population is clustered into homogeneous subgroups, or a diversity structure.
Definition 1. A diversity structure, \( DS \), is a stratification of the set of sources, \( \mathcal{N} \), into \( K \) exhaustive and disjoint groups. That is, \( DS = \{G_1, \ldots, G_K\} \) such that \( \bigcup_{k=1}^K G_k = \mathcal{N} \) and \( G_k \cap G_l = \emptyset \) for any \( k, l \in \{1, \ldots, K\} \) with \( k \neq l \).

Any grouping mechanism can be used to realise a diversity structure, the requirement being that sources in a group are similar with respect to the reports they are likely to provide each time they are queried. The approach we adopt in this paper is a refinement of the model proposed in [1].

In the grouping of the sources, we assume no prior knowledge of stratification variables. An agent exploits information from observable features of sources and their past reports to learn relevant metrics for stratification. We assume that each source, \( i \in \mathcal{N} \), can be described by a finite set of observable features, \( F_i \), and a history of reports that they have provided, \( H_i \). Thus, we model a source as a tuple \( \langle F_i, H_i \rangle \). Given this model of sources, we make use of machine learning techniques to learn associations between sets of features, and a target value (similarity value) defining the degree of similarity between sources. In particular, we employ the M5 model tree learning algorithm [20] to build a regression model to estimate the similarity between sources, and hierarchical clustering [8] to form a diversity structure. Input to the M5 algorithm is a collection of training instances. Each instance is specified by the value of a fixed set of attributes for all pairs \( i, j \in \mathcal{N} \). In our experiments, we assume numeric values for features. We compute the distance between each feature for a source pair as the absolute difference of the pair (e.g., \( |F_i^1 - F_j^1|, \ldots, |F_i^d - F_j^d| \), \( \forall i, j \in \mathcal{N} \)). Different types of feature values can be used in the model, but we use this absolute distance metric for the evaluation in Section 4. The class label for each training instance is the report similarity, \( \varphi_{ij} \), for the source pair over history \( H \), where \( \varphi_{ij} \rightarrow [0, 1] \). While computing the report similarity, we define a threshold \( \varphi \), and compute report similarity as a function of the number of similar reports, \( p_{ij} \), and conflicting reports, \( q_{ij} \), for the source pair for each time point \( t \in H \):

\[
(p_{ij}^t, q_{ij}^t) = \begin{cases} (1,0), & \text{if } |o_i^t - o_j^t| \leq \varphi \\ (0,1), & \text{if } |o_i^t - o_j^t| > \varphi \\ (0,0), & \text{otherwise} \end{cases}
\]

To compute the report similarity, \( \varphi_{ij} \), we use the Beta distribution [11]. The Beta distribution provides a means of forming opinions based on available evidence. For instance, opinions about the degree of report similarity of the source pair, \( i \) and \( j \) can be formed on the basis of similar, \( p_{ij} \), and conflicting, \( q_{ij} \), reports as obtained in Equation 1. The Beta distribution is denoted as \( \text{Beta}(p \mid \alpha, \beta) \), where \( \alpha \) and \( \beta \) are its two evidence parameters. The pair \( (p_{ij}, q_{ij}) \) provides a source of \( \alpha \) and \( \beta \) parameters of the Beta distribution, such that: \( \alpha_{ij} = p_{ij} + 1 \) and \( \beta_{ij} = q_{ij} + 1 \). The expected value of \( \text{Beta}(p \mid \alpha_{ij}, \beta_{ij}) \), hence the report similarity, \( \varphi_{ij} \), can be computed as \( \varphi_{ij} = \frac{(\alpha_{ij})}{(\alpha_{ij}) + (\beta_{ij})} \).

The learned similarity metric is used to cluster the sources to form a diversity structure. In particular, given the feature vector of any two arbitrary sources, a similarity score can be obtained relative to other sources (or groups), which is used to iteratively merge ‘close’ sources until a stopping criteria is met. This technique has been shown to be effective in modelling diversity [1], and allows us to generalise from similarity in sequences of reports from different sources to similarity of sources on the basis of their observable features.

3.2 Sampling Decision-Making

We formulate the sampling problem to exploit the similarity among sources as captured in a diversity structure, \( DS \). In exploiting the diversity structure, the agent is uncertain about the similarity (or variance) in the reports of sources, or how reflective of ground truth (trustworthy) they may turn out to be. If the agent knew these dynamics, it could sample from diverse groups in a more clever manner. For instance, the agent could sample less from groups that would provide very similar reports (low variance) to save cost, or it could sample more from groups that would guarantee a significant improvement in the estimate. We use reinforcement learning to model this problem of choice under uncertainty, and to guide the process of identifying effective strategies. Before presenting our sampling decision mechanism, we provide a précis of this technique.

3.2.1 Reinforcement Learning

Reinforcement learning (RL) [22] provides a framework by which an agent can learn optimal behaviour based on experience and rewards. The domain is modelled as a Markov decision process (MDP) \( \langle S, A, Pr, R \rangle \), where \( S \) is a set of states, \( A \) is a set of actions, \( Pr(s_{t+1} \mid s_t, a_t) \) is the probability of reaching state \( s_{t+1} \) when action \( a_t \) is taken in state \( s_t \), and \( R : S \rightarrow R \) is a reward function specifying the probability with which the agent obtains reward \( r_{t+1} \) when state \( s_{t+1} \) is reached. RL aims to maximise the expected discounted total reward, \( \mathcal{R} \), over an infinite-horizon:

\[
\mathcal{R} = E \left[ \sum_{t=0}^{\infty} \gamma^t r_{t+1} \right].
\]

where \( 0 \leq \gamma < 1 \) is a constant discount factor, and \( r_{t+1} \) denotes the reward received at time step \( t + 1 \). Pr and R are initially unknown, and the agent must learn an action selection strategy, or policy that maximises the expected discounted reward. We use SARSA [22], a simple and efficient model-free RL algorithm for learning a policy. SARSA maintains an estimate of how good it is to perform action \( a_t \) when in state \( s_t \), encoded by a Q-value function \( Q(s_t, a_t) \). It then selects actions based on their Q-values. These Q-values are then updated after taking an action:

\[
Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \eta(r_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)).
\]

where \( \eta \) and \( \gamma \) are the learning rate and the discount factor respectively. In learning a policy, the agent must find a balance between exploring current knowledge and exploiting alternative (and possibly unknown) action choices. Using Boltzmann exploration (Softmax) [22], the probability of selecting an action \( a_t \) in state \( s_t \), tempered by \( T_{mp} \), which decreases over time to favour exploration earlier in the learning process can be computed:

\[
P(a_t \mid s_t) = e^{Q(s_t, a_t)/T_{mp}} / \sum_{a_t \in A} e^{Q(s_t, a_t)/T_{mp}}.
\]

We present our decision-making mechanism, showing how the sampling problem can be mapped to RL.
3.2.2 States

A sampling state reflects knowledge available to the agent about the dynamics of a diversity structure, DS (i.e., trustworthiness and similarity of sources in diverse groups).

**Definition 2 (Sampling State).** Let $G_k \in DS$ denote a group in a diversity structure. Also, let $\tau_k$ and $\sigma_k$ denote the trust and similarity parameters of $G_k$ respectively. A sampling state, $s_t$ is a tuple $(T, \Sigma)$, where $T = \langle \tau_1, \ldots, \tau_K \rangle$ and $\Sigma = \langle \sigma_1, \ldots, \sigma_K \rangle$ are vectors corresponding to the trust and similarity levels of groups $G_k \in DS, \forall 1 \leq k \leq K, K = |DS|$.

We use evidence from past interactions involving sources in the different groups to constantly update these parameters as will be described in Section 3.3. We assume discrete RL algorithms. Therefore, $\tau_k$, the trust parameter, and $\sigma_k$, the similarity parameter, varies discretely. Similarly, a value of 0 for both parameters $	au_k$ and $\sigma_k$ denotes “untrustworthy” and “dissimilar” respectively. A sampling state $s_t$ is a tuple $\langle T, \Sigma \rangle$ where $T = \langle \tau_1, \ldots, \tau_K \rangle$, and $\Sigma = \langle \sigma_1, \ldots, \sigma_K \rangle$. A sampling state is denoted as $s_t$, $\forall 1 \leq k \leq K, K = |DS|$.

3.2.3 Actions

At each sampling time, $t$ the agent must decide how to sample from diverse groups. A sampling strategy or action is an allocation model, which specifies how to sample from diverse groups.

**Definition 3 (Sampling Action).** Let $G_k \in DS$ denote a group in a diversity structure, and $G = \{0, 1, \ldots, |G_k|\}$ be a finite set of possible allocations to $G_k$, $|G_k| > 0$. A sampling action, $a_t$ is a tuple $\langle g_1, \ldots, g_K \rangle$ s.t. $g_k \in G_k \forall 1 \leq k \leq K$, $K = |DS|$.

This definition implies that the agent may decide to sample zero up to all the sources in a group. The action space, $A$, dependent on $DS$, is given by the K-ary product over $K$ sets $G_1, \ldots, G_K$, which is the set of K-tuples:

$$A = \prod_{k=1}^{K} G_k$$

(5)

The action space represents all possible sampling combinations available to the agent. The expression: $A = \{(0, 0, 0), (0, 0, 1), \ldots, (3, 2, 5)\}$, is an example action space based on a DS with three groups, $G_1, G_2, G_3 \in DS$, having sizes $|G_1| = 3, |G_2| = 2, |G_3| = 5$, and $|A| = 72$. An action, $(2, 0, 1)$, where $|N| = 3$, implies that 2 sources are (randomly) sampled (without replacement) from the first group, $G_1$, 0 from the second group, $G_2$, and 1 from the third group, $G_3$. The tuple $(0, 0, 0)$ is the null action, where no sources are queried. In settings where there is a limitation, $\Phi$ in the number of sources to sample, we can prune the action space such that $\sum_k g_k \leq \Phi$.

3.2.4 Reward

When the agent takes a sampling action $a_t$ in sampling state $s_t$, it receives a reward $r_{t+1}$. We define this reward as a function of information quality or the deviation of the estimate, $\hat{\theta}$ from ground truth, $\theta$ (qual) and the cost of sampling the selected sources (cost). In computing the quality of information (Eq. 7), we assume that the agent can observe ground truth at some point down the line (too late for this to be useful to the agent). This multi-criteria optimisation problem (i.e., maximising both the quality of information and the negative cost of sampling) can be transformed into a single-criterion problem by employing scalarisation functions [4]. Thus, the reward function takes the form:

$$r_{t+1} = \lambda r_{t+1\text{qual}} + (1 - \lambda) r_{t+1\text{cost}},$$

(6)

where $r_{t+1\text{qual}}$ and $r_{t+1\text{cost}}$ represent rewards for the quality and the cost criteria respectively. The parameter, $\lambda \in [0, 1]$, is a coefficient that controls the trade-off between the different criteria, and intuitively captures the notion of a task constraint or preference in our decision model. Higher (or lower) weights may, for instance, be placed on a criteria to emphasise its relative importance to the agent’s task.

The reward signal is computed at the end of every time step using the outcome of the agent’s action. The quality component, $r_{t+1\text{qual}}$, is based on the system’s estimate of the environmental state:

$$r_{t+1\text{qual}} = 1 - \frac{\hat{\theta}_{t+1} - \theta^t}{\theta^t}, \theta^t \neq 0$$

(7)

$\hat{\theta}^t$ is computed as follows. The reports from sources within a group, $G_k$, are aggregated to form a group estimate, $\hat{\theta}_k^t$:

$$\hat{\theta}_k^t = \frac{\sum_{i \in G_k} \theta_i}{|G_k|}$$

(8)

where $g_k \subseteq G_k$ is regarded in this context as a subset of sources sampled from group $G_k \in DS$. The resulting estimates from each group sampled are then normalised by the parameter $w_k$ (Eq. 16). The normalised estimates are then combined to obtain the estimate, $\hat{\theta}$:

$$\hat{\theta} = \sum_{k \in G} \hat{\theta}_k \times w_k / \sum_{k \in G} w_k,$$

(9)

where $G$ in this context is regarded as a set that contains all the sampled groups.

The cost component, $r_{t+1\text{cost}}$ is computed as:

$$r_{t+1\text{cost}} = \frac{\text{cost}(N) - \text{cost}(N)}{\text{cost}(N)}$$

(10)

The reward $r_{t+1}$ is used to update the $Q$ value for the current state-action pair $(s_t, a_t)$ according to Equation 3 with the scalarisation function applied:

$$SQ(s_t, a_t) = \lambda Q(s_t, a_t, r_{t+1\text{qual}}) + (1 - \lambda) Q(s_t, a_t, r_{t+1\text{cost}}),$$

(11)

where $SQ$ denotes scalarised $Q$-values.

3.3 State Space Approximation

In this section, we describe how the agent can compute $T$ and $\Sigma$ in order to describe its state space. Since we assume discrete states, we define one possible function that maps continuous values of trust and similarity of groups to discrete state values. We describe how evidence about the trustworthiness and similarity of groups are obtained, and provide some background on the representative model which we adopt for similarity and trust.

3.3.1 Subjective Logic

Subjective logic (SL) [11] is a form of probabilistic logic that allows an agent to express opinions as degrees of belief, disbelief, and uncertainty about propositions. Let $\rho$ represent a proposition such as “source $i$ is trustworthy”. The binomial opinion about the truth of the proposition $\rho$, $\omega_\rho$, is represented as a tuple: $\langle b_\rho, d_\rho, u_\rho \rangle$, where $b_\rho$ is the belief...
that $\rho$ is true, $d_p$ is the belief that $\rho$ is false, and $u_\rho$ is the uncertainty, such that $b_\rho + d_p + u_\rho = 1.0$ and $b_\rho, d_p, u_\rho \in [0, 1]$. For convenience, we omit $\rho$ from the notation when it is implied by the context.

SL uses Beta probability density distributions to represent binomial opinions. Opinions are formed on the basis of positive and negative evidence observed in the past. Let $p$ and $q$ be the number of positive and negative experiences about the proposition $\rho$, respectively. Then, $b$, $d$, and $u$ are computed as:

\[
\begin{align*}
    b &= \frac{p}{p + q + 2}; \\
    d &= \frac{q}{p + q + 2}; \\
    u &= \frac{2}{p + q + 2}
\end{align*}
\]

(12)

Each opinion is associated with a base rate, $\alpha \in [0, 1]$, which is the a priori probability in the absence of evidence. SL often assumes a uniform prior ($\alpha = 0.5$). That is, before any positive or negative evidence has been received, both outcomes are considered equally likely. Using the base rate, an opinion's probability expectation value is computed as:

\[
E(\omega_p) = b + u \times \alpha.
\]

(13)

### 3.3.2 Group Trust

We assume that the agent can observe ground truth, $\theta^t$ at the end of each time step $t$ (i.e., after fusion and decision-making utilising $\hat{\theta}^t$). Let $p_{k,t}$ denote the number of times a group, $G_k$ is regarded as providing accurate information, which intuitively captures positive evidence about the group’s trustworthiness. $q_{k,t}$ denotes the corresponding amount of negative evidence. These parameters are updated at the end of each time step as follows:

\[
\begin{align*}
(\tau_k, \sigma_k^t) &= \begin{cases}
    (1, 0), & \text{if } |\hat{\theta}_k^t - \theta^t| \leq \nu \\
    (0, 1), & \text{if } |\hat{\theta}_k^t - \theta^t| > \nu \\
    (0, 0), & \text{otherwise},
\end{cases}
\end{align*}
\]

(14)

where $\nu$ is an acceptable error threshold for the agent, $\hat{\theta}_k^t$ is the estimate of group $k$ concerning $\theta^t$ (Eq. 8). If a group has not been sampled in the time step $t$ (i.e., if $|g_k| = 0$ in the chosen action), then its $p_{k,t}$ and $q_{k,t}$ at time $t + 1$ remain unchanged.

\[
\hat{\theta}_k^t = \sum_{i \in g_k} \Theta_i \left/ |g_k| \right.
\]

(15)

Finally, the weight $w_k$ for normalising estimates provided by a group, $k$ is computed using Equation 13:

\[
w_k = E(\omega_{k,t}) = b_{k,t} + u_{k,t} \times \alpha,
\]

(16)

where $b_{k,t}$ and $u_{k,t}$ are computed from the evidence parameters, $p_{k,t}$ and $q_{k,t}$, using Equations 12.

### 3.3.3 Group Similarity

Let $p_{k,\sigma}$ denote the amount of positive evidence for the similarity of two entities in a group $G_k$, and $q_{k,\sigma}$ be the corresponding negative evidence. The individual reports of sources in a group are used as evidence for the similarity of members of the group at the end of each time step.

We set $\omega_{k,\sigma} = 1$ for any group with $|G_k| = 1$, since there is absolute certainty about a source being similar to itself. For groups with $|G_k| > 1$ and $|g_k| = 1$ in the chosen action, $\alpha$,$,$ $p_{k,\sigma}, q_{k,\sigma} = (0, 0)$, since more than one source is required to effectively estimate similarity. In all other cases we update $(p_{k,\sigma}, q_{k,\sigma})$ using Equation 1.

### 3.3.4 Computing State Parameters

We can compute the parameters of the state space given the information about the trust and similarity of groups. In particular, associated with each group at time $t$ is information required to compute the group parameters $\tau_k$ and $\sigma_k$, given the sequence of observations until $t$; i.e. $p_{k,t}, q_{k,t}, p_{k,\sigma}$ and $q_{k,\sigma}$. We compute the expected trust, $\omega_{k,t}$ and similarity, $\omega_{k,\sigma}$ of a group $G_k \in DS$ using Equations 12 and 13:

\[
b_{k,t} = \frac{p_{k,t}}{p_{k,t} + q_{k,t} + 2}; \quad u_{k,t} = \frac{2}{p_{k,t} + q_{k,t} + 2}
\]

(17)

\[
b_{k,\sigma} = \frac{p_{k,\sigma}}{p_{k,\sigma} + q_{k,\sigma} + 2}; \quad u_{k,\sigma} = \frac{2}{p_{k,\sigma} + q_{k,\sigma} + 2}
\]

(18)

\[
\omega_{k,t} = b_{k,t} + u_{k,t} \times \alpha \
\omega_{k,\sigma} = b_{k,\sigma} + u_{k,\sigma} \times \alpha
\]

(19)

From these, we obtain the following values for $\tau_k$ and $\sigma_k$:

\[
\tau_k = \begin{cases}
    1 & \omega_{k,t} > 0.5 \\
    0 & \omega_{k,t} \leq 0.5
\end{cases} \\
\sigma_k = \begin{cases}
    1 & \omega_{k,\sigma} > 0.5 \\
    0 & \omega_{k,\sigma} \leq 0.5
\end{cases}
\]

(20)

![Figure 2: Different Truth Discovery Approaches](image)
Table 1: Experimental Parameters

<table>
<thead>
<tr>
<th>Parameter-value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N = 100 )</td>
<td>No. of sources in popl.</td>
</tr>
<tr>
<td>( L = 20 )</td>
<td>Learning interval</td>
</tr>
<tr>
<td>( \varpi = 0.1 )</td>
<td>Report similarity threshold</td>
</tr>
<tr>
<td>( \nu = 0.1 )</td>
<td>Report reliability threshold</td>
</tr>
</tbody>
</table>

Table 2: Source profiles

<table>
<thead>
<tr>
<th>ID</th>
<th>( f_1 )</th>
<th>( f_2 )</th>
<th>( f_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_1 )</td>
<td>( x )</td>
<td>( x )</td>
<td></td>
</tr>
<tr>
<td>( p_2 )</td>
<td>( x )</td>
<td>( x )</td>
<td></td>
</tr>
<tr>
<td>( p_3 )</td>
<td>( x )</td>
<td>( x )</td>
<td></td>
</tr>
</tbody>
</table>

multi-armed bandit technique [2], where each sampling action described in Section 3.2.3 is regarded as an arm using this algorithm. The aim is therefore to select the action or strategy that leads to the least variance in the estimate. Comparing our model with ADSS allows us to assess the effect of a diversity structure or source grouping only on outcomes.

CTRS, the strategy adopted by trust and reputation models [12, 24], focuses on accurately estimating the quality of information from sources rather than optimally selecting the sources to query. CTRS samples all sources in the population. To limit sampling, an arbitrary budget might need to be set, which may not necessarily be informed by trade-off. Thus, as shown in Figure 2, CTRS employs no explicit sampling decision model (captured as null in the figure). CTRS uses exactly the same trust model (Beta distribution) as DRIL to model trust, but for individual sources. In estimating ground truth, trust assessments are used to discount reports received during fusion. Comparing our model to CTRS allows us to assess the value of strategic sampling of sources to an agent’s utility, especially in environments where there is a concomitant cost to sampling.

4.1 Experimental Setup

Table 1 provides a summary of the experimental parameters. The simulated environment consists of 100 sources. Each source is randomly assigned to one of three profiles, which determines its reporting pattern in relation to other sources in the system. Each profile has three features, and for each feature, a distribution is defined from which feature values may be drawn for individual sources in the profile. Each feature value is drawn from a Gaussian distribution, with informative profile features having a small standard deviation \( N(\mu, 0.01) \), and uninformative feature values following a uniform distribution \( N(\mu, 1.0) \). In addition, each profile has a correlation parameter, \( P_c \), that specifies the degree to which reports of sources in a profile tend to be correlated. Therefore, with probability \( P_c \), a source will provide a similar report to other profile members, and with probability \( 1 - P_c \), it provides independent reports. In particular, a source that does not conform, deviates from mainstream opinion held by its profile members. A low \( P_c \) value means that more sources in a profile will report independently. A conforming source when reporting, first finds out about opinions maintained by its profile members. If any exists, it randomly selects one of these opinions to report, discarding its own private opinion or observation. In this way, we define the target relationships among groups of sources we wish to exploit. The \( P_c \) parameter adds an extra challenge to the learning algorithm, and allows us to evaluate the ability of our sampling decision model to cope with noise due to uncorrelated feature-behaviour similarity, hence variance in the reports of the sources. The \( P_c \) parameter is set at 0.8 for all profiles. A summary of the profiles is provided in Table 2.

In the figure, informative feature for defining similarity are marked with an “x”, while unmarked ones are noise features.

Each source has a reliability parameter, \( \rho \), that determines the type of reports it provides (i.e., honest, malicious). We define the following report types:

- **Reliable report:** This type of report is closer to ground truth, \( \theta' \), and is drawn from the distribution \( N(\theta' + 0, 0.01) \). Sources with high reliability ratio \( \rho \) are more likely to provide this type of report when queried.

- **Malicious report:** Reports of this kind are significantly deviated from ground truth, and follow the distribution \( N(\theta' + \varepsilon, 0.01) \), \( \varepsilon \in [0, 5] \). Sources with low \( \rho \) are more likely to provide this type of report, which, if left unmanaged, could potentially undermine an accurate estimate \( \theta' \) of the environmental state.

The report reliability threshold, \( \nu \), is set to 0.1, which reflects the intuition that information is still useful if it has a small amount of noise or is slightly discounted [21]. The system uses evidence obtained in the first \( L = 20 \) time steps to construct a diversity structure.

Finally, we instantiated the RL model with the following parameter values: the learning rate, \( \eta = 1.0 \); the discount factor, \( \gamma = 0.1 \); and the temperature, \( T_{mp} \) fixed at 0.1.

4.2 Results

Each instance of our simulation was repeated 20 times. Statistical significance of differences between strategies was computed using ANOVA at the 95% confidence interval. Analyses of significant differences between pairs of strategies was performed using Tukey’s HSD. We present and analyse the mean absolute error (accuracy) and total reward (utility) averaged over multiple runs for the different strategies considered, under different task constraint (\( \lambda \)) values.

In our first set of experiments, we assess the value of actively learning sampling strategies for an agent whose interest is to minimise costs. We ran our simulation with task constraint value \( \lambda = 0.2 \), which represents high preference for cost-minisation. Under this experimental condition, the agent is rewarded more for implementing strategies that lead to the use of fewer resources. In real-world contexts, higher preferences for cost-minimisation may reflect transactions or tasks associated with lower risk to the agent. The result of our simulation is shown in Figures 3a and 4a. Both DRIL and the adaptive sampling strategy (ADSS) employ learning mechanisms in order to decide how to sample the source population. These approaches significantly outperform the classical trust and reputation strategy (CTRS) in terms of the utility or reward (Figure 3a). This result was found to be statistically significant (\( p = 1.36 \times 10^{-32} \)). A post-hoc analyses suggest a significance difference both in the performance of DRIL over CTRS, and ADSS over CTRS. The adjusted \( p \ll 0.0001 \). Unlike DRIL and ADSS, CTRS does not actively learn sampling strategies. As the focus of CTRS is mainly on maximising quality, this approach samples all the sources in order to achieve this objective. Therefore, in settings where the emphasis is less on quality than cost-saving, CTRS-based approaches are bound to
perform poorly in terms of utility. It is however not immediately clear if DRIL has a superior sampling strategy to ADSS under this setting, as there is no significant difference in the rewards obtained by both approaches. As expected in Figure 4a, CTRS performs significantly better than ADSS in terms of accuracy ($p = 0.014$). Sampling all the sources gives CTRS an advantage in this context. DRIL on the other hand uses reinforcements from the system to adapt its sampling strategies towards actions that lead to better pay-offs. In particular, DRIL adopts sampling strategies that involve the selection of fewer sources from diverse groups. However, in doing so, the approach still maintains a relatively good balance between cost and quality. The post-hoc analysis (with an adjusted $p$-value of 0.14) suggests that the performance of CTRS in terms of quality is not significantly different from that of DRIL.

Figures 3b and 4b present the case when an equal preference is placed both on the quality and cost objectives ($\lambda = 0.5$). In this setting (Figure 3b), DRIL dominates CTRS in terms of reward. Furthermore, DRIL also shows a clear performance improvement over ADSS. This result was found to be statistically significant ($p = 1.33 \times 10^{-17}$). A post-hoc analysis suggests a significant difference both in the performance of DRIL over CTRS and DRIL over ADSS. In terms of accuracy, as measured by mean absolute error (Figure 4b), DRIL continues to perform robustly when compared to CTRS, and outperforms ADSS given that this strategy does not take trustworthiness of sources into consideration.

Finally, we consider the case when there is an extreme preference for information quality, as reflected by a high task constraint value, $\lambda = 0.8$. In this setting, it is assumed that the decision-maker is willing to incur a higher cost in order to acquire better quality information. While our DRIL model continues to perform better in terms of reward (Figure 3c), the performance gap between CTRS and ADSS is reduced. The result was found to be significant ($p = 7.70 \times 10^{-4}$). A post-hoc analysis suggests that the performance of DRIL is significantly better than CTRS and ADSS in terms of reward, but there is no significant difference in the performance of ADSS over CTRS. With a greater emphasis on quality, the savings made by ADSS are insufficient to yield higher pay-offs. Through a careful selection of sampling strategies, DRIL is able to outperform the other two approaches. In high $\lambda$ situations, DRIL tends to adjust its sampling strategy towards sampling more sources in order to improve its likelihood of acquiring more accurate estimates. This is reflected by the very similar result obtained by both DRIL and CTRS in Figure 4c. However, in implementing a more greedy strategy biased towards quality, the clear advantage previously experienced by DRIL in terms of reward is slightly diminished.

**Performance over time** To investigate the performance of the strategies over time, we kept the percentage of untrustworthy sources fixed (80), with a task constraint value, $\lambda = 0.9$, which reflects a high constraint on information quality. We report the net reward and mean absolute error at intervals through the experiments (Fig. 5). As expected, all the strategies have high estimation errors initially (error $\approx 0.64$) (Figure 5b). In the case of DRIL, the initial imperfect estimates of its state parameters impact its sampling deci-
The results are statistically significant (DRIL prove over time. In Fig. 5a, we see the effect of the superior source trustworthiness, and thus its performance fails to improve over time. In Fig. 5a, we see the effect of the superior sampling strategy employed by DRIL with respect to reward. The results are statistically significant (\( p = 2.85 \times 10^{-10} \) in Fig. 5a, and \( p = 5.19 \times 10^{-9} \) in Fig. 5b).

5. DISCUSSION

Reasoning about the trustworthiness of potential information sources is important for truth discovery in large and open systems. Reliance on trust alone is, however, insufficient for making effective sampling decisions. In resource-constrained environments, with a concomitant cost to sampling, an agent needs to devise means for optimally sampling the source population for evidence. Sampling all possible sources, an approach often adopted by trust and reputation mechanisms, may adversely affect the utility of the agent: the cost of sampling some sources may outweigh the value derived. The decision model we have proposed allows an agent to learn effective sampling strategies. Results of our evaluation show that by utilising a combination of diversity modelling and trust, an agent can better manage the trade-offs between quality and cost of information than existing approaches based only on trust or on some form of source diversification.

Our truth discovery framework, which is based on the idea of grouping sources, employs reinforcement learning to optimally allocate sampling resources to diverse groups. While reinforcement learning provides a principled means to learn optimal sampling strategies, there are known complexity issues [13]. In particular, the use of RL when no groups are formed (i.e., when sources are treated as individuals) is computationally expensive. For instance, in a similar environment to our evaluation, containing 100 sources, the learning algorithm would have \( 2^{100} \) different actions to select from in each decision state. Such a large action space would make it impracticable to identify a good strategy. Grouping of sources, for example, through the technique described in Section 3.1 enables this complexity to be managed. In fact, this method of using some form of clustering to structure some set of entities in an environment is a means to address scalability challenges for reinforcement learning and similar techniques. As highlighted in Section 3.2.3, the action space can be further pruned in settings where upper cost bounds are imposed. The approximation of the similarity and trust parameters that form the state space in our evaluation further allows us to manage the size and complexity of the reinforcement learning problem. As shown in our results, this approximation does not affect the model’s ability to make good estimates and source selection decisions, especially when compared to classical trust models without any behaviour approximations. In addition to the issues already discussed when considering the use of reinforcement learning settings without the formation of groups, we note that sources may collude or be biased in other ways. This increases the risk of sampling at greater cost, with possibly no improvement in estimates, thus making the formation of groups in an reinforcement learning context even the more pertinent.

6. CONCLUSIONS

In this research we have studied the problem of how to optimally sample a population of sources to determine ground truth. We proposed DRIL, a model that combines source diversification and reinforcement learning to identify effective sampling strategies. We have demonstrated that this approach performs as well as classical trust approaches in estimating ground truth. By sampling significantly fewer sources, DRIL significantly increases the utility to an agent. Unlike existing research, we relax the assumption that querying sources is cost-free. We believe that the problem we have addressed in this paper has significant impact in environments such as sensor networks where working within resource constraints is critical, and in social networks and many crowdsourcing problems where dependencies among information sources increase the risks of correlated biases.

7. REFERENCES


