Personalized Peer Truth Serum for Eliciting Multi-Attribute Personal Data

Naman Goel
Artificial Intelligence Laboratory
École Polytechnique Fédérale de Lausanne
Lausanne, CH 1015

Boi Faltings
Artificial Intelligence Laboratory
École Polytechnique Fédérale de Lausanne
Lausanne, CH 1015

Abstract

Several peer consistency mechanisms have been proposed to incentivize agents for honestly solving crowdsourcing tasks. These game-theoretic mechanisms evaluate the answers provided by an agent based on the correlation with answers provided by other agents ("peers") *who solve the same tasks*. In this paper, we consider the problem of eliciting personal attributes (for e.g. body measurements) of the agents. Since attributes are personal in nature, the tasks can not be shared between two agents. We show for the first time how to extend a peer consistency incentive mechanism, the Logarithmic Peer Truth Serum, to this setting for collecting personal attributes. When individuals report combinations of multiple personal data attributes, the correlation between them can be exploited to find peers. This new mechanism applies, for example, to collecting personal health records and other multi-attribute measurements at private properties such as smart homes. We provide a theoretical analysis of the incentive properties of the new mechanism and show the performance of the mechanism on several public datasets, which confirm the theoretical analysis.

1 INTRODUCTION

Crowdsourcing is a promising method to collect data in an inexpensive way. The data can be, for example, subjective opinions such as restaurant reviews or objective measurements such as pollution levels in a city and image labels. Measurements tasks are particularly important in collecting features which are useful in supervised and unsupervised machine learning. However, there is always a concern about the reliability of the data thus obtained. While some crowdworkers (henceforth called agents) will do their best to provide accurate data, many are not motivated to make the effort to obtain and report the data properly. This degrades the quality of the data. For example, if the data is to be observed with a sensor device, many may not be willing buy and maintain the device to obtain correct measurements. One approach to address this problem is providing the agents with incentives that cover the cost of their effort and encourage them to provide high quality data. The incentives have to be contingent on the quality of the data, for example, based on spot-checking the data for agreement with a trusted ground truth. In many of the most interesting applications, however, the ground truth is not accessible. Peer consistency is an elegant idea for designing incentive mechanisms in this situation. The output agreement (Waggoner and Chen, 2014), the Bayesian truth serum (Prelec, 2004; Witkowski and Parkes, 2012a), the peer prediction (Miller et al., 2005), the peer truth serum (Radanovic et al., 2016) and the correlated agreement (Shnayder et al., 2016; Dasgupta and Ghosh, 2013) are all examples of the peer consistency mechanisms.

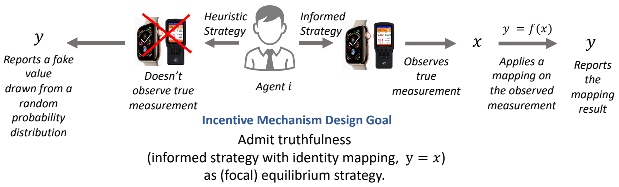

y
Reports a fake value drawn from a random probability distribution
Doesn't observe true measurement
Heuristic Strategy
Agent i
Informed Strategy
Observes true measurement
x
$y = f(x)$
Applies a mapping on the observed measurement
y
Reports the mapping result
Incentive Mechanism Design Goal
Admit truthfulness
(informed strategy with identity mapping, $y = x$)
as (focal) equilibrium strategy.

Figure 1: *Strategic agents in crowdsourcing.*

There is a lot of interest in extending this approach to collecting information such as records of personal sports activity, physiological measurements or diet. Other examples of personal information include sensor measurements observed at private properties such as smart homes or hotels. However, a fundamental limitation of the existing peer consistency mechanisms is that they require that a group of agents, called peers, observes the same data or a noisy version of it. For example, a group of agents should label the same image, measure pollution at the same location or give opinion about the same service. This does not work with personal data, since every agent is reporting data about a different object. However, we can extend the idea of peer consistency to a setting where agents report a *combination* of attributes that are known to be correlated with one another, even if the correlation structure is not known. When rewarding the report about one of the attributes, we can identify peers based on similarity in the *other* attributes reported at the same time.

In this paper, we show how one such mechanism, the *Logarithmic Peer Truth Serum (LPTS)* (Radanovic and Faltings, 2015), can be extended to elicit multi-attribute personal data from a crowd. We call this novel mechanism the *Personalized Peer Truth Serum (PPTS)*. We introduce task settings for eliciting continuous valued personal features from agents, exploit them to develop the *PPTS* and discuss the theoretical properties and practical applicability of the new mechanism. Our mechanism works in large scale settings when there are multiple agents reporting data and there exist (unknown) groups of agents sharing some personal characteristics. We show the validity of our assumptions on real datasets. *We also show that even when these groups are estimated from the data reported by agents, the incentive compatibility of the mechanism is not affected.*

1.1 Related Work

Different information elicitation mechanisms exist in the literature for two major settings. Techniques such as proper scoring rules (Gneiting and Raftery, 2007) and prediction markets (Wolfers and Zitzewitz, 2004) can be used to elicit truthful beliefs about events that are to be realized in future, if the realized outcomes of the events are observable by the mechanism. When such verification is not possible, peer-prediction mechanisms are a well-known solution for truthful information elicitation. In this paper, we are interested in the truthful mechanisms for information elicitation without verification.

The original peer-prediction method (Miller et al., 2005) is a mechanism for information elicitation without verification. The mechanism uses proper scoring rules to reward agents for reports that are predictive of other agents' reports and admits truth-telling as a Nash equilibrium. Several other methods (Jurca and Faltings, 2009) don't use proper scoring rules and instead use an "automated mechanism design" approach to determine adaptive payment rules that are incentive compatible. However, these mechanisms are not detail-free in the sense that they require agents' beliefs to be known.

The Bayesian Truth Serum (BTS) (Prelec, 2004) is another classic mechanism for information elicitation without verification. BTS doesn't use the knowledge of common beliefs to compute rewards, but collects two reports from each agent - an 'information' report (agent's own observation) and a 'prediction' report (agent's prediction about the distribution of information reports from other agents). The reward mechanism of the BTS ensures that truthful reporting is the highest-reward Nash equilibrium as number of agents solving a task tend to infinity. The Robust BTS of (Witkowski and Parkes, 2012b; Radanovic and Faltings, 2013) generalize the BTS to small populations in binary and non-binary answer settings respectively. These mechanisms are not minimal in the sense that they ask the agents to submit additional information than desired.

Several minimal and detail-free game theoretic incentive mechanisms have been developed recently for crowdsourcing. The seminal work in this category is due to (Dasgupta and Ghosh, 2013). The main idea in this work is to exploit multi-task settings, in which every agent solves multiple tasks. The mechanism rewards the agents for agreeing on a shared task and penalizes them for agreeing on a non-shared task. This mechanism ensures that truth-telling is a focal equilibrium in binary answer settings. The Correlated Agreement mechanism (Shnayder et al., 2016) generalizes the mechanism of (Dasgupta and Ghosh, 2013) to non-binary answer spaces with additional assumptions on the correlation structure of workers' observations. Both these mechanisms require that workers solve multiple tasks. The Logarithmic Peer Truth Serum (Radanovic and Faltings, 2015), which is based on an information theoretic principle, requires no such assumptions and ensures strong-truthfulness in non-binary answer spaces. (Kong and Schoenebeck, 2019) provide further complementary analysis for this information-theoretic framework. The guarantees of the mechanism are ensured in the limit (when every task is solved by an infinite number of workers). The Peer Truth Serum (PTSC) of (Radanovic et al., 2016) doesn't require even this assumption for the theoretical guarantees and works with a bounded number of tasks overall. The Deep Bayesian Trust mechanism (Goel and Faltings, 2019) ensures dominant strategy incentive compatibility and also computes fair rewards in large scale crowdsourcing by using both peer answers and some gold standard answers. The fundamental assumption in all of the above mechanisms is that the task solved by an agent can be shared with another agent, who submits independent noisy observation. (Agarwal et al., 2017) extend the Correlated Agreement mechanism to the settings where agents belong to one of the k possible categories of rating behaviors (for e.g. strict and lenient rating behavior). They cluster the agents with similar rating behavior to apply the CA mechanism. However, in this mechanism too, the assumption of shared tasks remains.

All these mechanisms are inherently inapplicable to elicit personal data. This is because when workers are asked to report measurements about the personal objects she owns (for example, her body or house), no other worker can share that task (because no worker can access the personal object owned by another worker). We extend the Logarithmic Peer Truth Serum to this setting while using a concept similar to that of "peers". In such a setting, these peers can not be distinguished using the 'shared task' definition. Our mechanism approximates them from the data reported by the workers while guaranteeing truthful equilibrium.

1.2 Our Contributions

From a technical standpoint, we address three main challenges in this work:

1. Define which agents can act as peers for one another in settings when agents can't share tasks.
2. Show that even if such peers are estimated from the reports submitted by the agents, the incentive compatibility is not affected.
3. Extend the mechanism to handle continuous data values instead of only discrete answers.

The summary of our contributions in the paper is as follows:

- We propose a novel incentive mechanism to elicit **continuous valued, multi-attribute and personal** data from crowd.
- We analyze and present several interesting theoretical properties of the mechanism. Our mechanism ensures that truthful reporting is an equilibrium and other undesired equilibria are less attractive. We also provide a practically useful and theoretically sound test to judge the applicability of our mechanism on a new type of data to be elicited.
- We show the performance of the mechanism on three real datasets, which are publicly available and are relevant to the settings of the paper.

2 SETTINGS

We consider the settings in which a requester (center) is interested in collecting data from a large number of agents W ($|W| = n \to \infty$) with some personal characteristics. The data being elicited consists of a set of attributes A ($|A| = d \geq 2$). The attributes A are personal characteristics such as body measurements of the agents. Agents independently obtain measurements for their attributes and report them to the center. The center in turn rewards them based on the quality of their reports. We assume the agents to be rational, seeking to maximize their expected rewards. The agents choose a reporting strategy to maximize their expected rewards. In a heuristic reporting strategy, they save the effort of even measuring the attribute and just report a random measurement drawn from an arbitrary probability distribution. In an informed reporting strategy, they obtain the measurement but report a mapping of the obtained measurement. Our aim is to formulate our incentive mechanism as a Bayesian game between the agents (who have probabilistic beliefs about the measurements of one another) and make truthful reporting (i.e. informed reporting with identity mapping) a profitable equilibrium strategy of the game for all agents. The strategic setting is described in Figure 1.

2.1 Belief Model

We model the beliefs of an agent i using thee continuous random variables for each attribute j. The first random variable X_{ij} is the attribute measurement itself. $P(X_{ij})$[1] is agent i's prior belief about measurements for the attribute j. The second random variable G_j models the global factors that affect the value of the j^{th} attribute of any random agent. $P(G_j)$ is the agent's prior belief about the global factors before obtaining her measurement for attribute j and $P(G_j|X_{ij})$ is her posterior belief after obtaining the measurement. The third random variable models the local factors that are personal to the agent and affect her attribute value. For every agent i, we model a set of other agents $N_i \subset W (1 << |N_i| << |W|)$, called cluster of agent i which share only these personal factors. Note that this is a much weaker modeling condition as compared to that of sharing personal measurements. Further, the clusters are unknown to the mechanism. The random variable for personal factors is denoted by L_{kj}, k being the cluster to which agent i belongs. In the rest of the paper, we will simply use notation L_{ij} for L_{kj} such that L_{ij} are equal for all i in the same cluster k. The $P(L_{ij})$ is the agent's prior belief about the personal factors before taking measurement for attribute j and $P(L_{ij}|X_{ij})$ is the posterior belief after taking measurement. L_{ij} and G_j are related through the conditional distribution $P(L_{ij}|G_j)$. It is easy to show that, in this model, the global distribution $P(X_{ij}|G_j)$ can be modeled by a mixture distribution as follows:

[1] In the paper, we use $P(\cdot)$ for density functions to keep notations simple.

$$P(X_{ij}|G_j) = \sum_{k=1}^{K} \alpha_k \cdot P(X_{ij}|L_{kj})$$

where K ($<< N$) is the number of distinct clusters in the population and α_k is the mixing probability of k^{th} cluster. The model is summarized in Figure 2.

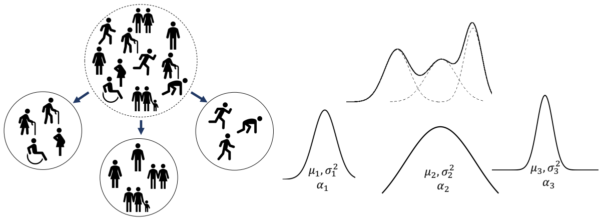


Figure 2: **(Belief Model)** *Left* - Agent population contains clusters of agents with similar characteristics. *Right* - Agents' measurements can be modeled using Gaussian mixture.

In this paper, we will use (Lyon, 2014) normal distribution[2] to model X_{ij}'s dependence on L_{ij}, i.e.,

$$P(X_{ij}|L_{ij}) = \mathcal{N}(\mu_{L_{ij}}, \sigma^2_{L_{ij}})$$

3 THE PPTS MECHANISM

The center collects reports from all agents for all their attributes. It then assigns each agent to its corresponding cluster described in agent i's belief. The cluster assignment step is discussed in Section 5. For now, let's assume this as an oracle that provides the mechanism with every agent's **true** cluster label. We define the j^{th} attribute score of agent i for reporting $X_{ij} = y$ as :

$$r_{ij} = \log \frac{f(y|\hat{\mu}_{L_{ij}}, \hat{\sigma}^2_{L_{ij}})}{\sum_{k=1}^{K} \hat{\alpha}_k \cdot f(y|\hat{\mu}_{L_{kj}}, \hat{\sigma}^2_{L_{kj}})} \tag{1}$$

where f is the Gaussian function given by

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\hat{\mu}_{L_{ij}}$ and $\hat{\sigma}^2_{L_{ij}}$ are the mean and variance of values reported for attribute j by agents in the cluster N_i. $\hat{\alpha}_k$ is the empirical relative mixing frequency of cluster k.

Agent i finally gets a cumulative reward (CR) equal to the average of attribute scores r_{ij} for all attributes $j \in \{1, 2...d\}$. More formally,

$$CR(i) = \frac{\sum_{j=1}^{d} r_{ij}}{d}$$

Example: As an example for calculation of attribute scores, consider agent i who reports its wrist measurement as 4.5 units. The reported wrist measurements of agents in the cluster of agent i have mean 4 and s.d. 3. If there are 2 distinct clusters in the population, another with mean 5 and s.d. 5 (with equal mixing frequencies), then the wrist attribute score of agent i is given by:

$$\begin{aligned} r_{ij} &= \log \frac{f(4.5|4, 3^2)}{0.5 \cdot (f(4.5|4, 3^2) + f(4.5|5, 5^2))} \\ &\approx \log \frac{0.1311}{0.5 \cdot (0.1311 + 0.0793)} \approx 0.22 \end{aligned}$$

On the other hand, if the means and s.d. of reports in the cluster of i are 0 and 1 respectively, then wrist attribute score of i is

$$\begin{aligned} r_{ij} &= \log \frac{f(4.5|0, 1^2)}{0.5 \cdot (f(4.5|0, 1^2) + f(4.5|5, 5^2))} \\ &\approx \log \frac{0.00002}{0.5 \cdot (0.00002 + 0.0793)} \approx -8.2 \end{aligned}$$

4 ANALYSIS

Intuitively, the numerator of the fraction inside the logarithm in Equation 1 measures how common (likely) a report is in its cluster while the denominator measures how likely a report is globally. Thus, similar to the Bayesian Truth Serum, PPTS rewards 'surprisingly common' reports. In the following theorems, we formally discuss the incentive compatibility and other properties of the mechanism. For better understanding, we first discuss the theoretical properties treating the cluster assignment step as a black box oracle and show in Section 5 how the mechanism obtains the clusters while preserving incentive compatibility. Because of space constraints, proofs are provided in the supplementary material, available on the first author's website[3]. The proof of theorem 1 is provided at the end of this paper to give an idea of the techniques used in the proofs.

We call a mechanism Bayes-Nash incentive compatible if truthful reporting is an equilibrium of the mechanism i.e. if other agents report their observations truthfully, no agent has an incentive to deviate from the truthful strategy for any observation of the agent. This is sometimes also called as the ex-post subjective equilibrium (Witkowski and Parkes, 2012a) since the beliefs of the agents are different (subjective).

Theorem 1. *The PPTS mechanism is Bayes-Nash incentive compatible, with strictly positive expected payoffs in the truthful reporting equilibrium.*

[2] This assumption simplifies the analysis of the mechanism and is not crucial for the main results presented in the paper.

[3] http://lia.epfl.ch/~goel/

The theorem states that given other agents are truthful, it is the best strategy for any agent to be truthful. The sketch of our information-theoretic proof is that from many independent and identically distributed truthful observations of other agents, the mechanism obtains maximum likelihood estimates of the true global and personal factors. A simple application of Bayes rule then shows that the mechanism rewards a report for its informativeness in predicting the personal factors, and the reward is maximized and is strictly positive for a truthful report.

While a truthful equilibrium is a desired outcome, there are other (non-truthful) equilibria that the mechanism admits - which is a common feature in the peer-consistency methods. It is important to ensure that such equilibria are not more profitable than the truthful equilibrium. They include heuristic reporting strategy equilibria. As discussed in Section 2, in heuristic reporting strategy, agents save the effort of even making an observation and report a random sample drawn from a probability distribution.

Theorem 2. *Heuristic reporting equilibria result in zero expected payoff in the mechanism.*

This is because when agents draw independently from a random distribution, both local and global MLEs converge to common values and it results in a reward of $\log 1 = 0$.

There are also informed non truthful equilibria, where agents do take the measurements but use a mapping to transform their actual measurements x into their reports y. Consider linear transformation mappings, where agents use a function $y = g(x) = ax + b$ to get their reports from their measurements x. In the real world, this strategy corresponds to agents systematically over reporting or under reporting their measurements.

Theorem 3. *In the PPTS mechanism, an equilibrium strategy profile defined by a function $g(x) = ax+b$ is not in expectation more profitable than the truthful strategy.*

The proof uses the observation that if agents use linear transformation to report, the MLE estimates also change accordingly and reward remains unchanged. Such equilibria don't give higher expected reward but choosing same g requires a lot of coordination among the agents and hence are unlikely to be played. Agents unilaterally choosing a different linear g' get lower scores than if they stay with g as well and thus such profile is not in equilibrium.

Next, we look at the ex-ante expected score of a truthful agent i.e. expected score before taking the measurement.

Theorem 4. *The ex-ante expected score of a truthful agent is equal to the conditional mutual information (CMI) of the attribute measurements and the personal factors given the global factors.*

The CMI (Cover and Thomas, 1991) is the expected value of the mutual information of two random variables given the value of a third, where the mutual information of two random variables measures the mutual dependence between two random variables. Since, CMI is always non-negative, the ex-ante expected score of a truthful agent is always non-negative. When the CMI is 0 i.e. when the attribute is independent of the personal factors, the mechanism can't be used to elicit truthful information because the expected payment is 0 regardless of the report. We discuss an interesting use of this theorem in further sections.

5 CLUSTERS APPROXIMATION

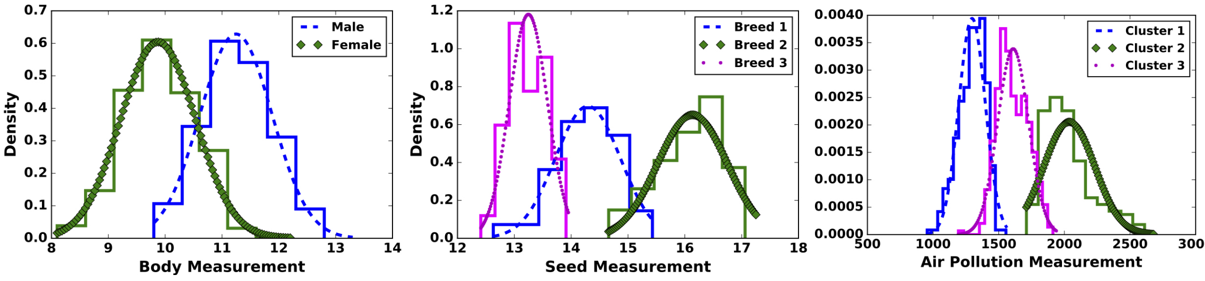
A crucial step in the mechanism described in Section 3 was to assign every agent to its correct cluster. We now describe how the mechanism achieves this without affecting the incentive compatibility. In the absence of the oracle, naturally the only option available to the center is to use the reports of the agents themselves to approximate the clusters. However, the question is whether doing this is game theoretically sound and preserves incentive compatibility?

Definition 1. ***(ε-Correct Clustering Algorithm)*** *A clustering algorithm is called ε-correct, if given true reports, it assigns a true report to a wrong cluster with probability at most ε and ε is such that as $|N_k| \to \infty$, the MLE estimates $\{\hat{\mu}_{kj}, \hat{\sigma}^2_{kj}\}$ converge to $\{\mu_{kj}, \sigma^2_{kj}\}$ and $\hat{\alpha}_k$ converge to α_k, $\forall k$.*

Note that the definition doesn't require every point to be assigned to correct clusters but only the approximated cluster parameters to converge to correct parameters. The conditions required for correct estimation of Gaussian mixture parameters from a finite sample are discussed in (Kalai et al., 2010),(Moitra and Valiant, 2010). The conditions include a lower bound on the mixing probabilities and the statistical distance between the cluster distributions. This implies that the more separated the clusters are, the better are the approximations of cluster parameters with fewer samples.

Theorem 5. *Given an ε- correct clustering algorithm, the PPTS is Bayes-Nash incentive compatible even if the clusters are approximated from the reports.*

The main insight of this theorem is the following : the fact, that the mechanism doesn't know the cluster labels but instead uses an ε-correct clustering algorithm to cluster the reports of the agents, doesn't provide any agent with a more profitable non-truthful strategy to deviate from the truthful equilibrium. This result addresses the concern that agents may strategically manipulate their report to get assigned to a different cluster and get a better



(a) Body Data

(b) Seed Data

(c) Air Data

Figure 3: Cluster Distribution in Datasets

reward. Hence, an ε-correct clustering algorithm can be applied to assign the clusters while preserving incentive compatibility.

Implementation and Practical Considerations

In this paper, we implement the *PPTS* mechanism by using the following technique to approximate the clusters. Consider approximating the cluster for calculation of the j^{th} attribute score of the agents. Let A_{-j} be the set of all attributes excluding attribute j i.e. $A_{-j} = A \setminus \{j\}$. We then apply k-means clustering algorithm on attribute sets A_{-j} to obtain the clusters used in calculating of the j^{th} attribute score.

It remains to discuss how one can judge if the clusters found using the above technique are indeed fit for being used with the *PPTS* mechanism in practice. For this, we make use of Theorem 4. The theorem says that if the conditional mutual information $I(X_{ij}; L_{ij}|G_j)$ is close to 0, then the mechanism can't be used for truthful elicitation. If some trusted prior data (i.e. some true observations X_{ij}) is available to the center for analysis, CMI estimators (Vejmelka and Paluš, 2008),(Ver Steeg, 2000) can be used to estimate $I(X_{ij}; L_{ij}|G_j)$ by using $\hat{\mu}_{L_{ij}}$ from the approximated clusters in place of L_{ij}. A low value of this CMI estimate suggests the unsuitability of the clusters for the mechanism. In the next section, we demonstrate this method on real datasets. To understand this in a more intuitive manner, recall that we use attribute set A_{-j} for approximating the clusters. If all attribute pairs are independent, observations for attribute j will be independent of the cluster approximated using A_{-j}, which means that the estimated clusters can't be used with the mechanism. Therefore, to find suitable clusters, we need to elicit interdependent attributes.

6 EXPERIMENTAL EVALUATION

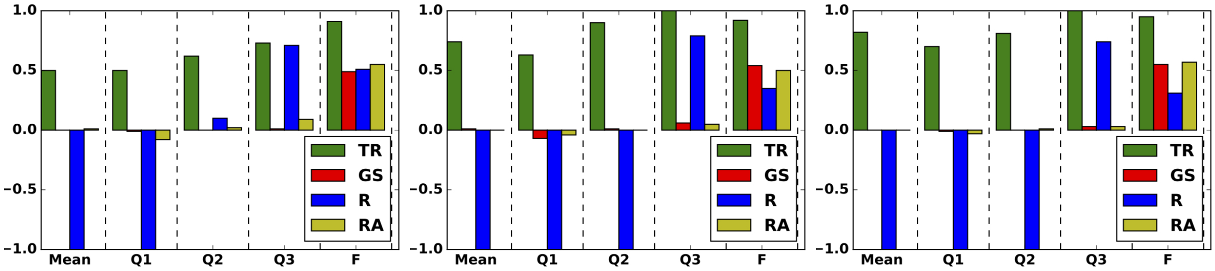
A real world validation of our mechanism by using it to collect new personal data is perhaps not feasible in the absence of ground truth for performance evaluation. However, the manipulation resistant properties of the mechanism can be best verified through simulations on real datasets. We simulate, on three real datasets, the strategies that agents may adopt and discuss the rewards that our mechanism decides for them.

Dataset	CMI Estimate
Body Measurements	0.41559387
Air Quality	0.98769209
Seed	0.98322659
Census Income	0.0194241

Table 1: Average CMI estimates for different datasets

6.1 Datasets

We selected three datasets from different domains for evaluating the mechanism through simulations. The *Body Measurements* (Heinz et al., 2003) dataset contains 21 body dimension measurements as well as age, weight, height, and gender of 507 individuals. The 247 men and 260 women were mainly young adults, with a few older men and women. The *Seed* (Charytanowicz et al., 2010) dataset consists of 7 measurements of 210 seeds of wheat. It has 70 samples each of three varieties of seeds (with labels). The *Air Quality* (De Vito et al., 2008) dataset consists of 9358 instances (852 complete instances) of hourly averaged responses from an array of 5 metal oxide chemical sensors embedded in an air quality multi-sensor device. The *Air Quality* dataset was not collected at different places but at a single place at different times. Another dataset that we considered for evaluation was extracted from U.S. 1994 census data. This *Census Income* (Kohavi, 1996) has 15 personal information attributes (continuous and categorical) about the

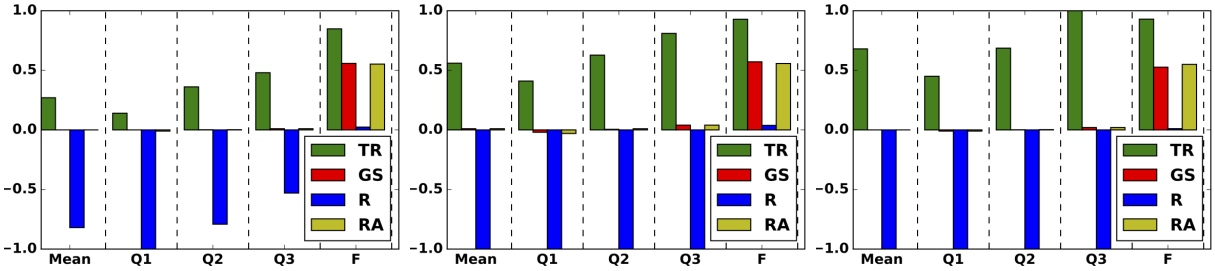


(a) Body Data

(b) Seed Data

(c) Air Data

Figure 4: Statistics of Attribute Scores

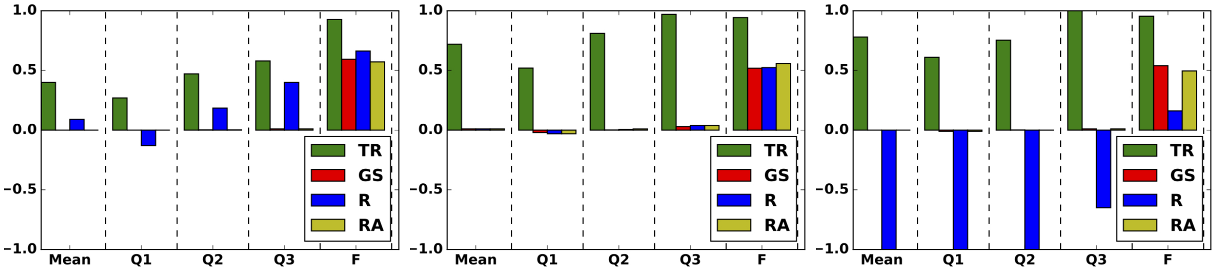


(a) Body Data

(b) Seed Data

(c) Air Data

Figure 5: Statistics of Cumulative Rewards (Average of attribute scores of all attributes)



(a) Body Data

(b) Seed Data

(c) Air Data

Figure 6: Statistics of Cumulative Rewards (Median of attribute scores of all attributes)

population such as salary class, education level, working hours, native country, age, sex, race, occupation etc. For simulations, we can assume each instance (row) in a given dataset to be reported by a different agent and each instance having multiple attributes. For example, in the *Air Quality* dataset, an instance has 5 attributes corresponding to the 5 metal oxide sensors. The datasets act as true private observations of agents. In *Seed* and *Body* datasets, clusters capture similarity between different individuals and seeds. In the *Air Quality* dataset, clusters capture the temporal similarity between pollution measurements. As datasets with more personal attributes are hardly available publicly, these public datasets do a good job at simulating the task settings we target i.e. elicitation of continuous valued unique personal attributes with normal like distribution. Figure 3 shows one attribute each in the *Body Measurements* dataset, the *Seed* dataset and the *Air* dataset along with their normal approximations in the clusters. The *body* and *seed* datasets are labeled but labels are used only for visualization and not for other experiments reported in the paper. *Air* dataset is unlabeled, hence we used our approximated clusters for visualization also.

6.2 Cluster Fitness Evaluation

To evaluate the fitness of the clusters approximated by the k-means algorithm on these datasets, we make use of Theorem 4. The average CMI estimates (of all attributes) from the four datasets are shown in Table 1. Note that for *Census Income*, the average CMI estimate is very small (close to 0). Hence, we can not use the clusters for eliciting the attributes of this dataset for reasons explained in Section 5.

Results

For better understanding, we present the results in two parts - attribute scores in Section 6.3 and cumulative rewards in Section 6.4. We will be discussing the following statistics of scores/rewards - mean (average of scores/rewards), $Q1$ (1^{st} quartile of scores/rewards), $Q2$ (2^{nd} quartile), $Q3$ (3^{rd} quartile) and F (fraction of agents receiving strictly positive score/reward), under different simulated strategies.

6.3 Attribute Score

We simulate the following reporting strategies that can be used by agents :

1. TR - All agents report all attributes truthfully.
2. RA - All agents report j^{th} attribute randomly within its true range and all other attributes truthfully.
3. R - All agents report all attributes truthfully except agent i, who reports j^{th} attribute randomly within its true range but other attributes truthfully.
4. GS - All agents collude to report j^{th} attribute using a Gaussian distribution with true mean and variance of the attribute and report all other attributes truthfully.

Figures 4a, 4b and 4c show statistics of attribute scores for j^{th} attribute in each dataset under different reporting strategies of agents. This $j^{th} attribute$ is 'height' in the *Body Measurements* data, 'kernel length' in the *Seed* data and 'PT08.S2' in the *Air Quality* data . Figure 4b shows results for the *Seed Measurements* data. The first important point to note is that the fraction of agents getting strictly positive score is more than 0.92 when agents report truthfully but hardly goes above 0.5 in other non-truthful strategies, which means that non-truthful strategic agents do no better in expectation than a random guesser. The other thing to note is that the mean score when agents are non-truthful is not positive, whereas for truthful agents, it is strictly positive with sufficient value to distinguish it from a 0 score. A similar trend can be observed for other statistics such as $Q1, Q2$ and $Q3$, where the score for truthful reporting is always greater than that for non-truthful strategies. In particular, we can observe that $Q2$ (i.e. the median) is also strictly positive for truthful agents and not more than 0 for non-truthful agents. Similar results can be seen in Figures 4a and 4c for *Body* and *Air* datasets respectively. It is worth mentioning here that the scores can be appropriately scaled to cover the cost of participation and satisfying budget constraints without affecting the incentive-compatibility of the mechanism.

Also to confirm our earlier conclusion of the clusters not being useful for the *Census Income* dataset, we computed the rewards of agents for reporting this data truthfully and found that only about 32% of the agents get positive score with mean score approaching 0.

6.4 Cumulative Reward

Here, we report simulation results for the following reporting strategies :

1. TR - All agents report all attributes truthfully.
2. RA - All agents report all attributes randomly within true ranges of respective attributes.
3. R - All agents report all attributes truthfully except agent i, who reports all attributes randomly within true ranges of respective attributes.
4. GS - All agents collude to report all attributes using Gaussian distributions with true means and variances of respective attributes.

In section 3, we defined the cumulative reward of a agent as the average of all attribute scores of this agent. Figures 5a, 5b and 5c show statistics of final or cumulative rewards. Figure 5b shows the results for *Seed* data. Similar to attribute scores discussed in Section 6.3, the fraction of agents with strictly positive cumulative reward is 0.93 when they report truthfully and is hardly more that 0.5 when they report non-truthfully. The mean cumulative reward for truthful reporting strategy is strictly positive and is not more than 0 for non-truthful strategies, attesting Theorem 1 and Theorem 2. In Figures 6a, 6b and 6c, we show statistics of cumulative rewards calculated as the *median* of the attribute scores instead of average of attribute scores, i.e.,

$$CR(i) = \underset{j \in \{1 \dots d\}}{\text{median}} \{r_{ij}\}$$

The median is another way to calculate CR from attribute scores and makes it robust to outliers in attribute scores. We also find the median to perform better in simulations as it makes the minimum reward of truthful agents non-negative.

7 CONCLUSIONS

In this paper, we investigated the problem of incentivizing agents to honestly report their personal attributes such as physiological measurements. We distinguish this problem from the problem of incentivizing agents where multiple agents can solve a common task such as labeling a common image. We thus extend the applicability of the peer based incentive mechanisms from discrete labels for shared objects to real valued multi-dimensional personal features. We propose the *Personalized Peer Truth Serum (PPTS)* to address the problem. The PPTS shows desired properties by making the honest reporting equilibrium more profitable than heuristic reporting equilibria. We further investigate the problem of finding peer agents against whom the report of an agent is to be evaluated and propose to exploit other reports of the agent to estimate its peers. We guarantee that the incentive compatibility of the mechanism continues to hold while doing so. We provide a theoretically sound practical test to determine the applicability of PPTS for a given set of attributes by estimating the ex-ante expected payment. We empirically analyze the performance of PPTS using estimated peers on real datasets. The PPTS is able to incentivize/penalize simulated honest and heuristic reporting strategies with a good accuracy.

A PROOF OF THEOREM 1

Proof. The attribute score of agent i is given by :

$$\log \frac{f(y|\hat{\mu}_{L_{ij}}, \hat{\sigma}^2_{L_{ij}})}{\sum_{k=1}^{K} \alpha_k \cdot f(y|\hat{\mu}_{L_{kj}}, \hat{\sigma}^2_{L_{kj}})}$$

Given that all other agents report truthfully, the attribute score becomes :

$$\log \frac{f(y|\mu_{L_{ij}}, \sigma^2_{L_{ij}})}{\sum_{k=1}^{K} \alpha_k \cdot f(y|\mu_{L_{kj}}, \sigma^2_{L_{kj}})}$$

This is because the maximum likelihood estimates $\{\hat{\mu}_{L_{ij}}, \hat{\sigma}^2_{L_{ij}}\}$ converge to $\{\mu_{L_{ij}}, \sigma^2_{L_{ij}}\}$ as $n, |N_i| \to \infty$ under the assumptions of conditional independence and statistical similarity. We can write it as:

$$r_{ij} = \log \frac{P(X_{ij} = y|L_{ij})}{P(X_{ij} = y|G_j)}$$

The expected attribute score R of agent i, who observed $X_{ij} = x$ and reported $X_{ij} = y$ is then given by :

$$R = \int\limits_{L_{ij},G_j} P(L_{ij}, G_j|X_{ij} = x) \log \frac{P(X_{ij} = y|L_{ij})}{P(X_{ij} = y|G_j)} dL_{ij} dG_j$$

where $P(L_{ij}, G_j|X_{ij} = x)$ is agent's posterior belief about L_{ij} and G_j conditional on observing $X_{ij} = x$.

Under the assumption that attribute value X_{ij} is conditionally independent of global factors G_j given the personal factors L_{ij}, i.e., $P(X_{ij} = y|L_{ij}) = P(X_{ij} = y|G_j, L_{ij})$, we get

$$R = \int\limits_{L_{ij},G_j} P(L_{ij}, G_j|x) \cdot \log \frac{P(y|G_j, L_{ij})}{P(y|G_j)} dL_{ij} dG_j \quad (2)$$

However, we know (using Bayes' rule) that,

$$\frac{P(y|G_j, L_{ij})}{P(y|G_j)} = \frac{P(L_{ij}|y, G_j)}{P(L_{ij}|G_j)} \quad (3)$$

Using Equations 2 and 3,

$$\begin{aligned} R &= \int\limits_{L_{ij},G_j} P(L_{ij}, G_j|x) \cdot \log \frac{P(L_{ij}|y, G_j)}{P(L_{ij}|G_j)} dL_{ij} dG_j \\ &= \int\limits_{L_{ij},G_j} P(L_{ij}|x, G_j) \cdot P(G_j|x) \log \frac{P(L_{ij}|y, G_j)}{P(L_{ij}|G_j)} dL_{ij} dG_j \\ &= \int\limits_{L_{ij},G_j} P(L_{ij}|x, G_j) \cdot P(G_j|x) \cdot \\ &\qquad \log \frac{P(L_{ij}|y, G_j) \cdot P(L_{ij}|x, G_j)}{P(L_{ij}|G_j) \cdot P(L_{ij}|x, G_j)} dL_{ij} dG_j \end{aligned}$$

which can be rearranged as,

$$\begin{aligned} R = \int\limits_{G_j} P(G_j|x) \Big[&\int\limits_{L_{ij}} -P(L_{ij}|x, G_j) \log \frac{P(L_{ij}|x, G_j)}{P(L_{ij}|y, G_j)} dL_{ij} \\ &+ \int\limits_{L_{ij}} P(L_{ij}|x, G_j) \log \frac{P(L_{ij}|x, G_j)}{P(L_{ij}|G_j)} dL_{ij} \Big] dG_j \quad (4) \end{aligned}$$

for brevity,

$$R = \int\limits_{G_j} P(G_j|x)[-KL_1 + KL_2] dG_j$$

where, KL_1 and KL_2 are KL-divergences and hence non-negative. It is easy to see that R is uniquely maximized when $KL_1 = 0$, which happens only when $y = x$. The expected attribute score at $y = x$ is

$$R_{Truth} = \int\limits_{G_j} P(G_j|x) KL_2 dG_j \quad (5)$$

which is strictly positive. □

References

Agarwal, A., Mandal, D., Parkes, D. C., and Shah, N. (2017). Peer prediction with heterogeneous users. In *Proceedings of the 18th ACM Conference on Economics and Computation (EC-2017)*.

Charytanowicz, M., Niewczas, J., Kulczycki, P., Kowalski, P. A., Łukasik, S., and Żak, S. (2010). Complete gradient clustering algorithm for features analysis of x-ray images. In *Information technologies in biomedicine*, pages 15–24. Springer.

Cover, T. M. and Thomas, J. A. (1991). Elements of information theory.

Dasgupta, A. and Ghosh, A. (2013). Crowdsourced judgement elicitation with endogenous proficiency. In *Proceedings of the 22nd international conference on World Wide Web*, pages 319–330. ACM.

De Vito, S., Massera, E., Piga, M., Martinotto, L., and Di Francia, G. (2008). On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical*, 129(2):750–757.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Goel, N. and Faltings, B. (2019). Deep Bayesian Trust: A dominant and fair incentive mechanism for crowd. In *AAAI Conference on Artificial Intelligence*.

Heinz, G., Peterson, L. J., Johnson, R. W., and Kerk, C. J. (2003). Exploring relationships in body dimensions. *Journal of Statistics Education*, 11(2).

Jurca, R. and Faltings, B. (2009). Mechanisms for making crowds truthful. *Journal of Artificial Intelligence Research*, 34:209–253.

Kalai, A. T., Moitra, A., and Valiant, G. (2010). Efficiently learning mixtures of two gaussians. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 553–562. ACM.

Kohavi, R. (1996). Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid. In *KDD*, volume 96, pages 202–207.

Kong, Y. and Schoenebeck, G. (2019). An information theoretic framework for designing information elicitation mechanisms that reward truth-telling. *ACM Transactions on Economics and Computation*, 7(1):2:1–2:33.

Lyon, A. (2014). Why are normal distributions normal? *The British Journal for the Philosophy of Science*, 65(3):621–649.

Miller, N., Resnick, P., and Zeckhauser, R. (2005). Eliciting informative feedback: The peer-prediction method. *Management Science*, 51(9):1359–1373.

Moitra, A. and Valiant, G. (2010). Settling the polynomial learnability of mixtures of gaussians. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 93–102. IEEE.

Prelec, D. (2004). A bayesian truth serum for subjective data. *Science*, 306(5695):462–466.

Radanovic, G. and Faltings, B. (2013). A robust bayesian truth serum for non-binary signals. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI" 13)*.

Radanovic, G. and Faltings, B. (2015). Incentive schemes for participatory sensing. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*.

Radanovic, G., Faltings, B., and Jurca, R. (2016). Incentives for effort in crowdsourcing using the peer truth serum. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(4):48.

Shnayder, V., Agarwal, A., Frongillo, R., and Parkes, D. C. (2016). Informed truthfulness in multi-task peer prediction. EC '16, pages 179–196. ACM.

Vejmelka, M. and Paluš, M. (2008). Inferring the directionality of coupling with conditional mutual information. *Physical Review E*, 77(2):026214.

Ver Steeg, G. (2000). Non-parametric entropy estimation toolbox (npeet).

Waggoner, B. and Chen, Y. (2014). Output agreement mechanisms and common knowledge. In *Second AAAI Conference on Human Computation and Crowdsourcing*.

Witkowski, J. and Parkes, D. C. (2012a). Peer prediction without a common prior. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 964–981. ACM.

Witkowski, J. and Parkes, D. C. (2012b). A robust bayesian truth serum for small populations. In *AAAI Conference on Artificial Intelligence*.

Wolfers, J. and Zitzewitz, E. (2004). Prediction markets. *Journal of economic perspectives*, 18(2):107–126.