Parametric Prediction from Parametric Agents

Yuan Luo
The Chinese University of Hong Kong
ly011@ie.cuhk.edu.hk

Nihar B. Shah
UC Berkeley
nihar@eecs.berkeley.edu

Jianwei Huang
The Chinese University of Hong Kong
jwhuang@ie.cuhk.edu.hk

Jean Walrand
UC Berkeley
walrand@berkeley.edu

Prediction algorithms are often designed under the assumption that the training data is provided to the algorithm, and that the algorithm has no control over the quality of the training data. In many situations, however, the training data is collected by surveying people, for instance, in the prediction of the future demand for a product by surveying a number of potential customers, or the prediction of the winner of an election by surveying potential voters. Collecting data from people is much cheaper, easier and faster today due to the emergence of several commercial crowdsourcing platforms such as Amazon Mechanical Turk and others. In such situations, it is possible to monetarily incentivize the respondents to provide higher quality inputs.

In any realistic setup, the responses obtained from people ("the agents") are noisy: one cannot expect a naive customer to gauge the sales of a product accurately. Moreover, every individual has a different expertise and ability, and will likely react differently to the amount of money paid per task. For example, some people may be active users of the surveyed product, therefore have a better understanding of its anticipated usage. We assume that the surveyor ("the principal") has no knowledge of the behavior of individual agents. It is therefore important to design an appropriate incentive mechanism for the prediction procedure that exploits the heterogeneity of the agents, motivating them to participate and exert suitable levels of effort. An appropriate incentive will provide higher quality data and as result, a superior prediction performance. This requirement motivates the problem at the interface between statistical estimation and mechanism design considered in this paper.

As compared to problems that tackle only one of the prediction and the mechanism design problems, the problem of joint design poses a significantly greater challenge. From the statistical prediction point of view, the challenge is that every sample is drawn from a different distribution, whose properties are unknown apriori to the principal. From the mechanism design perspective, the challenge is that the incentivization procedure not only needs to ensure that agents report truthfully, but also needs to ensure that each agent exerts an effort that minimizes the overall prediction error. In this paper, we formulate and optimally solve a "parametric" form of this joint design problem. More specifically, the principal desires to predict a parameter of a known distribution. Each agent is modeled in a parametric fashion, with her work quality (or expertise) governed by a single parameter that is the agent's private information. While each agent aims to maximize her own expected payoff (i.e., the revenue minus the cost of effort), the principal must optimize a joint utility that trades off the prediction error and the monetary costs

In this paper, we design a mechanism (called "COPE", for COst and Prediction Elicitation) that jointly optimizes the principal's payoff in terms of the payments made to the agents and the prediction error incurred. COPE provides a systematic way for the principal to incentivize all participating agents to report their estimations truthfully and exert appropriate amounts of effort based on their respective capabilities. The mechanism incorporates and exploits the heterogenity of the agents in terms of their capabilities and costs, in order to minimize the prediction error.

Our COPE mechanism operates under a fairly general framework, encompassing a wide range of distributions of the parameters for prediction as well as the cost functions of the agents. In this paper, however, we focus on two special settings to gain engineering insights towards the design of such mechanisms. We investigate the special scenario where the noise follows a Gaussian distribution, and study the impact of two specific cost functions on the principal's decision.

Our results show that when the costs incurred by the agents are linear in the amount of exerted effort, the principal should conduct a *crowd-tender*, soliciting service of only the agent with the lowest reported cost. On the other hand, when the costs are quadratic in the exerted effort, the optimal mechanism is that of *crowd-sourcing*, where the principal recruits multiple agents to complete the task.

The intuition is as follows. With a linear cost function, the agent's marginal cost is positive even when agent exerts no effort. As the principal needs to compensate agents' costs in order to incentivize them to put effort, the principal is willing to choose the most capable one to finish the task. With a quadratic cost function, however, an agent's marginal cost depends on the agent's effort level. In such a case, the principal is willing to explore the heterogeneity of agents and incentivize appropriate effort levels from them.

We refer the reader to [1] for more details.

1. REFERENCES

[1] [Online] http://eecs.berkeley.edu/~nihar/publications/netecon2015.html