A Survey on Crowdsourcing and Behavioral Outcome

Vinay V. Mandhare[*]   **Vinod Nayyar**
Department of M.Tech CSE   *Department of M.Tech CSE*
R.T.M. Nagpur University, India   *R.T.M. Nagpur University, India*

Abstract— large datasets tends to developed models and determine which subset of data to mine is now becoming automated. However what kind of data to be select and place in first place requires human experience which is supplied by domain expert. This paper gives new approach to machine science demonstrating that non domain experts can collectively formulate features and provide values for those features so that they are predictive of some behavioural outcome interest. This was accomplishing by web platform where group of people get interact with each other by responding to question which help to predict behavioural outcome. Which result in dynamically growing online survey.

Keywords—machine Science, surveys, outcome, Behavioral outcome, human modeling

I. INTRODUCTION

To develop predictive model which map between set of outcomes and predictor variables create many problems. While the model structure are pre-specified with set of predictive covariates statistical tools provides mature methods to compute model parameter. The goal of the research is to test an alternative way to modelling in which online crowd can be used to define potentially predictive variables to study by asking and getting response to question, so that an predictive model is developed .

A. Machine Science

Machine science is a new scientific technology which very interesting to find, analyse, classify the data to generate hypothesis and develop models [1] Machine Science is a hot topic in the theory and philosophy of modern science, with recent claims that "within a decade, even more powerful tools will enable automated, high-volume hypothesis generation to guide high-throughput experiments in biomedicine, chemistry, physics, and even the social sciences" [1].

This paper introduces a method in which non domain experts can be motivated to formulate independent variables but also populate enough of variables to form successful modelling. This can be well explained as follows. Users or user visit a site based on behavioural outcomes (The behavioural outcomes could be a body mass index or daily electricity consumption) is to be modelled. The user will provide their own outcomes (like their own consumption of electricity) and answer the questions that may be predictive of that outcome (Such as how much electricity they use daily).By ordinarily, different models are constructed in oppose to growing data sets predicting user's behavioural outcome. User can also post their own questions that, which becomes new independent variables when answer by other users in the modeling process. Thus to discover and populate independent variables will be done by user community.

B. Crowdsourcing

Now days there is a rapid growth in user generated contend on the internet is one of the best example that how the user interactions can effectively solve the problem under the explicit management by team of experts [2].Crowdsourcing can be defined as "Harnessing the experience and efforts of large number of individuals is frequently known as the Crowdsourcing". The Crowdsourcing has been used in number of research and commercial applications. For an example the Wikipedia in which set of information is being published and later on number of people can add and enhanced the information finally the best of the best enhanced data is available for the user which cannot be done by using a single computer alone and could be expensive to achieve through expert domain process.

Wikipedia is the best example that explains how the online collaborative approach can solve the difficult problem very easily without any expenses. This survey paper reports on task with direct motivation: one could be a house hold energy usage task in which user are able to understand their own energy consumptions which helps to improve their energy saving; and another could be the body mass indexing task which helps user to understand their lifestyle choice so that they can live a healthy life .These both instantiations include comparative approaches in which participants compare with each other and also allow to predict quality of question that participant provide.

In comparison with the top-down system the collaborative systems are generally more scalable [4]. At last the Crowdsourcing can tend to develop a creative solution that is substantially different from the experts. The crowdsourced poem translation task is surprising and preferable than the expert translation [5]

II. METHODOLOGY

This system is wrapping a human behavioral paradigm in cyberinfrastructure Such that: (1) the human behavioral based outcome is defined by the investigator which is to be modeled; (2) The data collection is done with the help of

human volunteers; (3)the models are then continually generated with automatic approach. (4) The volunteers are also motivated to propose new independent variables. Basically the system is divided into group of task like first group of task is that which include the working of the investigator, participant group and modeling engine this group work together and produces the predictive model of the outcome of interest which is the second group of task. First of all the investigator creates a web site which define the human behavioral outcome which is to be modeled. The paper deals with the financial and health outcome which were investigated. First outcome is the monthly electric energy conservation and second one is the body mass index. The investigator then initializes the site with set of some questions which are related to the outcome of interest. For example in case of fast food consumption and obesity, we seeded the BMI website with the question "How many time in a week do you eat fast food?"

The users who visit the site individually provide the values for the outcome of interest such as their own BMI(Body mass Index).The user then respond to the questions display on the site. A datasets are used to store the information provided by the user forward to the modeling engine. After Receiving the data modeling engines start working and construct a single matrix $A \in R^{n \times k}$ and outcome vector **b** having length n from the collective responses of n users to k questions. **A** is combination of a_{ij} elements indicating the i user's response to question j, and each element b_i in **b** is the i users outcome of interest. Here the linear regression model was used to achieve the model of outcome. The output of the modeling process is the vector c of length k+1 having models parameters. It also gives the output vector d having k length which stores the predictive power of each question: d_j stores the r^2 values obtained by regression only on column j of **A** against the response vector b. These two outputs are stored in the data store

At any time user can elect to pose a question of their own design. User can also pose a yes/No response questions, five level likart rating, or a number. Users were not impelling in what kind of questions to pose. Once the user poses a question the suitability is check by the investigator. A questions was suitable or unsuitable if it holds the following conditions.(1) the question revealed the identity of its author (e.g. *"Hi, I am Sandy Ray. I would like to know if..."*) thereby contravening the Institutional Review Board approval for these experiments ;(2) the question contained profanity or hateful text; (3) the question was inappropriately correlated with the outcome. If the question was specifically suitable it was added to the pool of questions available on the site; otherwise the question was discarded. Each time a user visit to the site, they were shown a new unanswered question as well as additional data so that interest in the site, their participation in the experiment should be maintain. After answering all the available questions, they were shown a listing of the questions, their responses, and contextual information to indicate how their responses compared to those of their peers.

The most important information shown to each user after responding to each question was the value of their actual outcome as they entered it (bi) as well as their outcome as predicted by the current model (ˆbi). After each response from a user

$$\hat{b}_i = c_0 + c_1 a_{i1} + c_2 a_{i2} + ... + c_k a_{ik} + \varrho_i \quad (1)$$

if user i has not yet responded to question j then where $a_{ij} = 0$ otherwise aij is set to the user's response.

III. PROPOSED APPROACH
- In our proposed method, we are applying the hierarchical regression approach instead of using the linear regression approach.
- In the existing system the main model used is the single model m, Instead of using the single model by applying n number of models we can increased the processing system and result analysis of the system

IV. CONCLUSIONS
This paper has introduced a new approach to social science modeling in which human behavioral outcome is generated by motivating the participants. In this paper participate is visiting the web site and answering to the questions which he wants to and not wants to be leads to hectic for the participant. So by applying principle approach using the different regression model question ordering could be made principally so that the user could not face the questions that he don't want to answer. Also instead of using the single model the working of the system could be enhance by applying the n number of models. This is the potentially a new way to do science. This new approach to science could result the exponential growth which is found tin another online collaborative communities.

REFERENCES
[1] J. Evans and A. Rzhetsky, "Machine science",*Science,* vol.329,no.5990,p.399,2010
[2] J. Giles, "Internet encyclopedias go head to head," *Nature*, vol. 438,no. 15, pp. 900–901, 2005.
[3] D. C. Brabham, "Crowdsourcing as a model for problem solving,"*Convergence*, vol. 14, pp. 75–90, 2008.
[4] D. Wightman, "Crowdsourcing human-based computation," in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction:*
[5] A. Kittur, "Crowdsourcing, collaboration and creativity," *XRDS*, vol. 17, no. 2, pp. 22–26, 2010.