IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 9, Issue 2, Ver. III (Mar - Apr. 2014), PP 111-114
www.iosrjournals.org

A Framework for Prediction of Crowdsoursing Behavioural Outcomes

Jayshree D. Abhonkar [*1], Prof. P.R.Barapatre[*2]
Department of Computer Engineering, University of Pune
SKN Sinhgad Institute of Technology & Sciences, Lonavala, Pune, Maharashtra, India

Abstract: *To retrieve the information from huge amount of data and finding what kind of data to be mined is becoming progressively computerized. On the other hand selecting what data to gather requires human association or practice, generally delivered by field expert. This system describe that for prediction of some behavioral outcomes, non-field experts can jointly formulate structures and then provide values. Existing system not give the accurate behavioral model, this achieved in system by constructing a web site in which peoples respond to questions and pose new questions to others. This results in a dynamically-growing online survey. There are two different approaches for the system in which first can predict users monthly electricity consumption; the other predict users body mass index.*
Index terms: *Crowdsourcing, Human Behavior modeling, Survey, Prediction.*

I. Introduction

To develop predictive models between predictor variables and an outcome, there are many problems. When the set of predictive covariates and the model structure are pre-specified, tool like neural networks provide advanced methods for calculating model parameters. Now, current research is providing new tools for gathering the essential form of non-linear predictive models of given good input -output data. Though, the task of selecting which potentially predictive variables to study is largely a qualitative task that requires essential domain expertise. For example, to select questions that will find predictive covariates a survey designer must have domain expertise. In order to find which variables can be logically adjusted and to optimize performance an engineer must improve significant familiarity with a design.

If the wisdom of crowds is coupled to produce difficult problems then exponential growth occurs in the causal factors of behavioral outcomes. To overcome this problem system predicts human behavioral outcomes. Thus, the goal of this study was to check an another approach to modelling in which the crowds predict variables by asking questions and respond to those questions, in order to improve a predictive model.

The proposed system is framework for predicting human behavior outcomes. This paper introduces a method by which non field experts can be inspired to frame independent variables. This is accomplished as follows; Users open a website, provide their own outcome and then answer questions that could be predictive of that outcome. Sometimes when models are created against the data sets that predict each user's behavioral outcome, User fake their own questions to other users, then it becomes new independent variables in the modelling process.

Section 1 introduces the paper, Section 2 describes Literature Survey, Section 3 explains proposed system, Section 4 shows results and Section 5 and 6 are conclusion and acknowledgement respectively.

II. Literature Survey

The proposed methodology based on Natural Language Processing, Questions Classification & Model Identification. There are many problems in which one seeks to develop predictive models to map between a set of predictor variables and an outcome. Statistical tools such as multiple regression or neural networks provide mature methods for computing model parameters when the set of predictive covariates and the model structure are pre-specified. Furthermore, recent research is providing new tools for inferring the structural form of non-linear predictive models, given good input and output data [1]. However, the task of choosing which potentially predictive variables to study is largely a qualitative task that requires substantial domain expertise. For example, a survey designer must have domain expertise to choose questions that will identify predictive covariates. An engineer must develop substantial familiarity with a design in order to determine which variables can be systematically adjusted in order to optimize performance.

The system described here wraps a human behavior modeling paradigm in cyber infrastructure such that:
(1) The investigator defines some human behavior-based outcome that is to be modeled
(2) Data is collected from human volunteers
(3) Models are continually generated automatically

(4) The volunteers are motivated to propose new independent variables.

How the investigator, participant group and modeling engine work together to produce predictive models of the outcome of interest. The investigator begins by constructing a web site and defining the human behavior outcome to be modeled. In this paper a financial and health outcome were investigated: the monthly electric energy consumption of an individual homeowner (Sect. III), and their body mass index (Sect. IV). The investigator then initializes the site by seeding it with a small set (one or two) of questions known to correlate with the outcome of interest. For example, based on the suspected link between fast food consumption and obesity [2], [3], we seeded the BMI website with the question "*How many times a week do you eat fast food?*" Users who visit the site first provide their individual value for the outcome of interest, such as their own BMI. Users may then respond to questions found on the site their answers are stored in a common data set and made available to the modeling engine. Periodically the modeling engine wakes up (Fig. 1m) and constructs a matrix A n×k and outcome vector b of length n from the collective responses of n users to k questions. Each element aij in A indicates the response of user i to question j, and each element bi in b indicates the outcome of interest as entered by user i. In the work reported here linear regression was used to construct models of the outcome but any model form could be employed. The modeling process outputs a vector c of length k+1 that contains the model parameters. It also outputs a vector d of length k that stores the predictive power of each questioned stores the r2 value obtained by regressing only on column j of A against the response vector b. These two outputs are then placed in the data store [1].

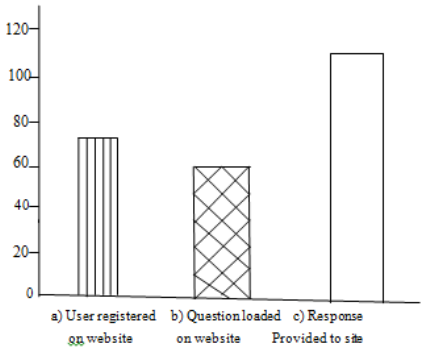

Fig 1: User behavior on body mass index site

Fig. shows that body mass index site is attracted near about 70 users register on site (a) they collectively loaded a 60 questions (b) and got 110 responses to those question.

Crowdsourcing is an online, distributed problem-solving and production model that has emerged in recent years. Notable examples of the model include Threadless, iStockphoto, Inno-Centive, the Goldcorp Challenge, and user-generated advertising contests. This article provides an introduction to crowdsourcing, both its theoretical grounding and exemplar cases, taking care to distinguish crowdsourcing from open source production. This article also explores the possibilities for the model, its potential to exploit a crowd of innovators, and its potential for use beyond forprofit sectors. Finally, this article proposes an agenda for research into crowdsourcing. In this article I have provided an introduction to crowdsourcing through definitions established by its pioneers and illustrated through a collection of case examples. Crowdsourcing can be explained through a theory of crowd wisdom, an exercise of collective intelligence, but we should remain critical of the model for what it might do to people and how it may reinstitute long-standing mechanisms of oppression through new discourses.[4]

Crowdsourcing is a newly developed term which refers to the process of outsourcing of activities by a firm to an online community or crowd in the form of an 'open call'. Any member of the crowd can then complete an assigned task and be paid for their efforts. Although this form of labor organization was pioneered in the computing sector, businesses have started to use 'crowdsourcing' for a diverse range of tasks that they find can be better completed by members of a crowd rather than by their own employees. This paper examines how firms are utilizing crowdsourcing for the completion of marketing-related tasks, concentrating on the three broad areas of product development, advertising and promotion, and marketing research. It is found that some firms are using crowdsourcing to locate large numbers of individuals willing to complete largely menial repetitive tasks for limited financial compensation. Other firms utilize crowdsourcing to solicit solutions to particular tasks from a crowd of diverse and/or expert opinions. Conclusions are drawn regarding the advantages and the limitations of crowdsourcing and the potential for the future use of crowdsourcing in additional marketing-related applications. The advantages of crowdsourcing are that it gives firms access to a potentially huge amount of labour outside of the firm which can complete necessary tasks often in a fraction of the time and at a fraction of the cost than if the same activities were conducted in-house. Some of the available 'crowd' may have limited skills but they will be willing to take on repetitive, menial tasks which cannot easily be performed by computers. On the other hand selected crowds may have a degree of expertise not available within the firm

which can work to solve more complex issues or tasks. With particular applicability to the marketing field, crowdsourcing allows firms to harvest ideas from a wide and diverse collection of individuals with experiences and outlooks different from those that exist within the firm. [5]

III. Implementation Details

The proposed system shown in Fig. 2 in which the system is divided in to four parts that are - Pre-processing, Question Pre-processing, Question vector and Model Behavior.

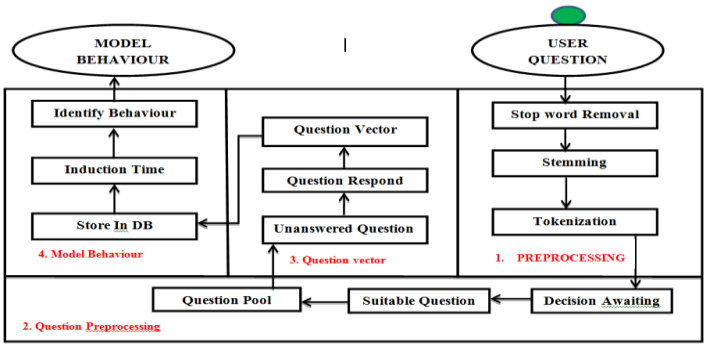

Figure 2: Architecture of Proposed System

User question:

User is most important part of the system, the system begins with user question, in which user can easily respond the question to predict a behavioral outcomes and get the appropriate result. Non-domain experts can also responds to questions.

So that first input or data set for our system is that user natural thinking in the form of natural/general questions. These questions are further handle by the other parts of system which described in detail below.

1) Pre-processing:

Pre-processing is the first step in a system which takes the user input in the form of natural language and evaluate that input in to an appropriate format give to the next step of system which is question Pre-processing. Pre-processing perform the basic operation on users input question.

Following steps involved in the Pre-processing

a) Stop word removal: in this stop word processing, useful and meaningful words will be sorted out and questions will be filtered out. Typically stop word lists contain words that don't carry as much meaning, such as determiners and prepositions Words like *the*, *is*, *at*, *which*, and *on*. By this way we get only main required worlds which are really useful for answering to the user's question.

b) Stemming: Stemming is a process of removing prefixes and suffixes from words. A stemming is a process of linguistic normalization, where variant forms of a word are reduced to a common normal form. We can use a stemming for increasing a performance of the system to provide an exact result from a system. Variables having a ending part or any suffix taking of it is known as a Stemming. For example –ion, –ions, –ive, –ed, –ing.

c) Tokenization: Firstly we are having a raw state which is our systems inputs comes from user's minds in the form of natural language. By using a tokenization process without changing its meaning whole text, it is segmented into sequential manner of words and sentences which represent a token.

2) Question Pre-processing:

In this question is either selected or reject. The question is suitable or not is find in decision awaiting process then it is transfer to question pool and then given to the next step that is question vector. The investigator is responsible for initially creating the web platform, and seeding it with a starting question. Then, as the experiment runs they filter new survey questions generated by the users. However, once posed, the question was filtered by the investigator as to its suitability. A question was deemed unsuitable if any of the following conditions were met:

(1) The question revealed the identity of its author (e.g. "*Hi, I am John Doe. 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 (e.g. "*What is your BMI?*"). If the question was deemed suitable it was added to the pool of questions available on the site; otherwise the question was discarded.[1]

3) Question vector:
Users respond the question that are already loaded in site and also pose their new question to others that unanswered questions is transferred to question vector for further processing.

4) Model Behavior:
Users who visit the site first provide their individual value for the outcome of interest. Users may then respond to questions found on the site. Their answers are stored in a common data set and made available to the modeling engine.
At any time a user may elect to pose a question of their own devising. Users could pose questions that required a yes/no response, a five-level Likert rating, or a number. Users were not constrained in what kinds of questions to pose.

IV. Results

Data Set:

The system deals with the prediction of human behavior model. User can directly pose new questions to other users. Question should be in appropriate format so that machine can understand and which gives an exact results according to the uses need.
So we are having data set with the natural human minds questions and the appropriate result will be created in a result set.
For example - the types of question are:
1) For the body mass index questions like - how many times you eat spicy food in a week, do you exercise per day, how many times do you work per week?
2) For the electricity usage questions like – how many members in your house? Do you have heater, dryer and refrigerator?

Result set:

The systems produce a result according to the users need and where user gets exact result with its natural behavior by natural thinking. The result will satisfies a usability functionality of system, final result is prediction of human behavior.

V. Conclusion

In this paper participants are encouraged to relate some human behavior outcome, such as homeowner electricity usage or body mass index. In which participants effectively revealed at least one statistically substantial predictor of the outcome variable.

The system is useful to answer many difficult questions regarding why some outcomes are different than others. By considering example of fast food and behavior of children we can predict "why better quality and tasty food is available only in certain hotels, but not others, "why do hobbies and behavior differ in children's even if they have same edge.

Acknowledgement

I am very thankful to my guide Prof. P. R. Barapatre, who has been very concerned and has aided for all the material essential for the preparation of this paper, he helped me to explore this vast topic in an organized manner and provided me with all the ideas on how to work towards a research oriented venture.
I am also thankful to Prof. S. B. Sarkar, P.G Coordinator and Prof. Preeti Sharma for the motivation and Inspiration that triggered me for the thesis work.

References

[1] Josh C. Bongard, "Crowdsourcing Predictors of Behavioral Outcomes," IEEE transactions on knowledge and data engineering, 2013.
[2] S. Bowman, S. Gortmaker, C. Ebbeling, M. Pereira, and D. Ludwig, "Effects of fast-food consumption on energy intake and diet quality among children in a national household survey," *Pediatrics*, vol. 113, no. 1, p. 112, 2004.
[3] J. Currie, S. Della Vigna, E. Moretti, and V. Pathania, "The effect of fast food restaurants on obesity and weight gain," *American Economic Journal: Economic Policy*, vol. 2, no. 3, pp. 32–63, 2010.
[4] D. C. Brabham, "Crowdsourcing as a model for problem solving," *Convergence*, vol. 14, pp. 75–90, 2008.
[5] Paul Whitla, "Crowdsourcing and Its Application in Marketing Activities", Contemporary Management Research Pages 15-28, Vol. 5, No. 1, March 2009