



Outcome Prediction by Using Collaborative Approach for Crowdsourcing

Vinay V. Mandhare, Prof. Pragati Patil

Department of Mtech, Computer science and Engineering
R.T.M. Nagpur University, Maharashtra, India

Abstract— *Crowdsourcing has often been used in the past as a application in order to discover a solution. Crowdsourcing has been cited as the new way of research or a pattern in which online-crowd can be utilized to complete the specific task by using a new machine science technique. The crowd can be motivated further to change their behaviour by analyzing their own outcome. This paper Describe the new principal approach implementation to enhance the performance and time span for the prediction of the behavioural outcome.*

Keywords— *Crowdsourcing, Machine science, Outcome, Behavioural Outcome, Implementation.*

I. INTRODUCTION

Crowdsourcing is the process of obtaining needed services, ideas, or content by soliciting contributions from a large group of people, and especially from an online community, rather than from traditional employees or suppliers. [3]

Now days a Crowdsourcing has emerge as new pattern that completely based on the intelligence of “crowd” of people. Any problem can be solve by utilizing the specific crowd in the experimental form with a monetary reward [4][5][6].The crowdsourcing is the term invented by the Jeff Howe in 2006. Most recently it could be define as “obtain (information or input into a particular task or project) by enlisting the services of a number of people, either paid or unpaid, typically via the Internet.” By Howe “*Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call.*”

Crowdsourcing has prompted models and applications for problem solving and knowledge building in a wide range of domains, from science to enterprise to social networks. The development of crowdsourcing has been leveraged in a way that largely satisfies social demands for working with people in a larger geographic scope, as well as task demands for accomplishing “human intelligence tasks (HIT).” [7] Successful crowdsourcing applications, such as Amazon’s Mechanical Turk (AMT), Everything2, Wikipedia, etc., serve as innovative learning and network collaboration paradigms.

A. Machine science

Machine Science, or data driven research, is not only new but also interesting scientific methodology that use advance computational techniques to identify, retrieve, classify and analyse data in order to generate hypothesis and develop models. In machine science it is difficult to decide which variables is to be selected first nor easy to decide which variable to be automated. In recent robot scientist demonstrated that can carry out experiments as well as algorithms that cycle through hypothesis generation, experimental design, experiment execution and hypothesis refutation[8][9].

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, which becomes new independent variables when answer by other users in the modelling process. Thus to discover and populate independent variables will be done by user community.

B. Crowd sourcing

Most promising and interactive model that can be applied in different disciplines is the crowdsourcing. Now days in every field like academic, enterprise and social we found the crowdsourcing technique. There are few application based on the crowdsourcing

1. Design interfaces that can attract and sustain a large number of people to join in crowdsourcing.
2. Design mechanisms that support efficient crowdsourcing formation as well as avoid conflict among collaborators.
3. Design experience models that evaluate the “quality of experience in crowdsourcing process.” [10]
4. Design a mechanism for balancing communication and independency within crowdsourcing. [11]

5. Design in a way to increase the robustness of crowdsourcing in dealing with collaborations between unknown people within unknown situations and/or environments. [11]

Chung et al. [12] introduce the “Stress Outsourced (SOS)” system which is designed to enable people to relieve stress by sending each other “therapeutic massages” as an instance of the concept of crowdsourcing. They state: “The SOS system consists of members wearing customizable signal modules for sending and responding to an anonymous SOS call, and wearable message actuation modules to receive the message strokes sent by others.” Chung et al. discuss three innovative design concepts of the SOS system. Based on their prototype of the message module, they describe a design framework which can be useful to help better understand and develop scalable social media and haptic networks with crowdsourcing. The prototype of the SOS system is achieved by wearable garments (such as sweater, shirt etc.) that can send and receive signals through “an embedded or attachable signal module and a message module.”

Huberman et al. [13] describes social networking as a construction of the “interaction that people have with their friends, families and acquaintances.” From data gathered from Twitter, they found that the degree of people’s contributions to crowdsourcing depends on the attention people receive from others. Based on this notion, they describe two types of networks, “a very dense one made up of followers and followers” and “a sparser and simpler network of actual friends.” This distinction can be part of the design and research agenda for design online collaboration network based on social aspects.

Jeff Howe offers the following definition of crowdsourcing: “Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call” The designated agent is often called the requester, while people working on the Tasks are called crowd workers. The requester needs to assign the task to these workers with unknown capability and willingness to complete the task based on the goals to achieve.

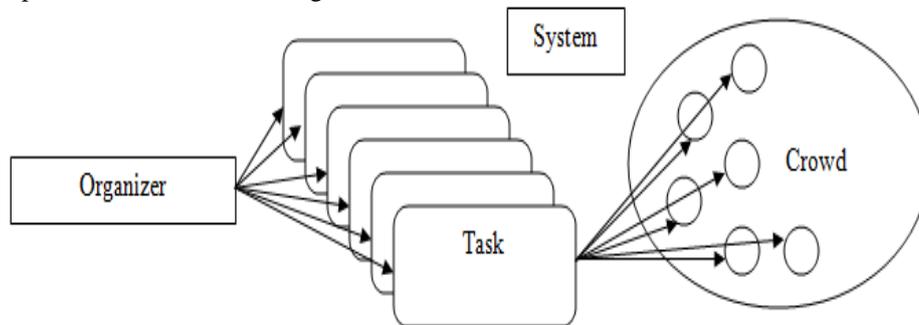


Fig. 1 General design of Crowdsourcing

The Crowdsourcing research is involved in variety of fields such as computer science, management, and many other domains which have discovered Crowdsourcing as a useful approach.

The effectiveness of crowdsourcing is proved by the best example, Amazon’s Mechanical Turk [3]. In this, it is explained that a “Human Intelligence Task” such as characterization of data, transcribing spoken language, or creating data visualizations can be accomplished with the help of group of people which is very difficult for a computer to accomplish alone.

II. PROPOSED WORKED

Our proposed method is consisting of a hierarchical regression approach instead of linear regression approach used in the existing system. In the proposed system principal approach for determining the outcome of user’s behaviour is utilised. As like the existing system instead of assuming a single model multiple models are generated from the same model to predict the behavioural outcome.

III. METHODOLOGY

One of the crowd sourcing based system is described below which wraps the human behaviour paradigm. Figure 2 explains the detail workflow of the system .Complete system is divided into three different classes.

1) Investigator model

The investigator model is nothing but the investigator who defines some web-based platform to predict the user’s behavioural outcome here we are considering the web based body mass index predictor.

2) User behavioural Model

The user behaviour model is based on the behaviour of the user and prediction of the result; generated from the model engine. Users are further motivated to answer the questions based on the inputs provided by the user. The result is compare to other user within the same group so that the behaviour of the user should be change by comparing with the other group.

3) Model engine

The backend working model is the main model which enables the system to avail the dynamic approach instead of linear approach in the existing system [base paper].The question pose by the user are first analysed with following different steps.

A) Similar question removal

If the question enters by the user found similar to the entered question in the database then the question is discarded by using sentence to sentence similar method by using the cosine similarity function each sentence S is compared with the other two different sentences. Considering the sentence Si and Sj

Having the terms t to n with weight Wi, Wj of the same. The Similarity formula is utilized as follows

$$sim(SiSj) = \frac{\sum_{t=1}^n Wit \times Wjt}{\sqrt{\sum_{t=1}^n Wit^2} \times \sqrt{\sum_{t=1}^n Wjt^2}} \quad (1)$$

B) Hateful word removal

If the entered question is having the word which is in the hateful word list then it is discarded instead storing to the database.

C) Users Identity removal

If any of the questions based on the identity of the user having noticed the name of the user the question is discarded.

D) Question type predictors

System is categories in two different groups to predict the outcome of interest.

1. Descriptive question

In this step the user's behaviour is based on the term weight of the answer stored in the database

2. Yes/no type questions or Numeric value based question

To perform the modelling first the relation between the scalar dependent and the explanatory variable should be expressed. And to achieve this system utilized the regression approach for prediction of the numerical data. The following formula explains the concept in details.

$$\hat{C} = Ax + B \quad (2)$$

Where

\hat{C} = is the predictive value

x = Predictor value

A & B are constants

The final result is the aggregation of the above results

$$\frac{\sum(C-\hat{C})^2}{N} \quad (3)$$

Where,

C= criterion variable.

N= number of active user.

IV. BODY MASS INDEX PREDICTOR

In order implement this approach for the outcome that is more readily available to the user are provided with the website in which model engine will predict the users body mass index of each participants. Generally Body mass index is calculated by the formula

$$\text{Body mass Index (BMI)} = \frac{\text{mass(kg)}}{(\text{Height(m)})^2} \text{ or } \frac{\text{mass(lb)}}{(\text{height(in)})^2} \times 703 \quad (4)$$

Though having a several limitations [14] is the most common measure to determine the level of patient's obesity. Every single users Body mass can be easily calculated by entering its height and mass. Users are provided with the option having the different options to enter data in both kilogram or pound also height in meter or foot. Since up to date the obesity is one of the major global public health challenges [15] it is investigated. And people with extreme BMI values are likely to have intuitions as to why they deviate so far from the norm.

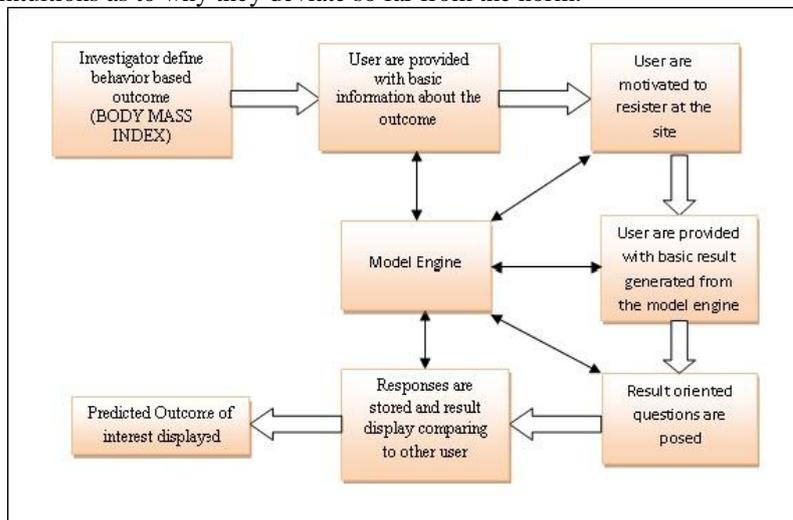


Fig.2. Workflow of the system

Participants arriving to the site are first motivated to enter the height and weight as per their convenience. Then participants are free to respond to the questioner and pose the question. Further for motivation users are displayed with the predicted outcome with the outcome compared to two different groups. Each group is dynamically generated from the resulted outcome obtained from the same user. First group is consists of a minimum number of users having the outcome Less than that of the user and next group of comparison is consists of a minimum number of users having the average outcome greater than that of the user. Also the responses from the both the compared group is displayed to the user. The compared output is displayed to user so that to motivate his behaviour comparing him with other user. Thus the behaviour of the user is change by simply applying the dynamic concept of the crowd sourcing. The peer group data were meant to help users compare how their lifestyle choices measured up to their most similar peers who were slightly healthier than themselves, and slightly less healthy than themselves. This approach in effect provides individualized suggestions to each user as to how slight changes in lifestyle choices may lead to improvements in the health indicator being measured.

V. CONCLUSION FUTURE WORK

In proposed scheme, hierarchical regression approach is utilized to increase the performance of the system instead of the linear regression algorithm. Instead of creating the number of models single model is updated to increase the processing speed. Compare to the existing system principal approach is utilized in investigation behavioural model. Our future work is to apply the same concept in analysing the home owner electricity consumption. In which users will motivate to compare the electricity utilization with the same residual surrounding.

REFERENCES

- [1] Josh C. Bongard, Member, IEEE, Paul D. Hines, Member, IEEE, Dylan Conger, Peter Hurd, and Zhenyu Lu, "Crowdsourcing Predictors of Behavioral Outcomes", IEEE Transactions on Systems, Man, and Cybernetics: systems, vol. 43, no. 1, January 2013.
- [2] Josh C. Bongard, "Crowdsourcing Predictors of Behavioral Outcomes," IEEE transactions on knowledge and data engineering, 2013.
- [3] "Crowdsourcing - Definition and More". Merriam-Webster.com. August 31, 2012. Retrieved 2014-02-03.
- [4] Maja Vukovic. 2009. "Crowdsourcing for Enterprises," in Proceedings of the 2009 Congress on Services - I (SERVICES '09). IEEE Computer Society, Washington, DC, USA, 686-692.
- [5] Osamuyimen Stewart, Juan M. Huerta, and Melissa Sader. 2009. "Designing crowdsourcing community for the enterprise," In Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '09), Paul Bennett, Raman Chandrasekar, Max Chickering, Panos Ipeirotis, Edith Law, Anton Mityagin, Foster Provost, and Luis
- [6] Thierry Bücheler, Rudolf M. Fuchslin, Rolf Pfeifer, Jan Henrik Sieg. (2010). In Proceedings of Artificial Life XII -- Proceedings of the Twelfth International Conference on the Synthesis and Simulation of Living Systems. Pages 679–686. MIT Press.
- [7] Nathan Eagle. 2009. "txteagle: Mobile Crowdsourcing," In Proceedings of the 3rd International Conference on Internationalization, Design and Global Development: Held as Part of HCI International 2009 (IDGD '09), Nuray Aykin (Ed.). Springer-Verlag, Berlin, Heidelberg, 447-456.
- [8] J. Bongard, V. Zykov, and H. Lipson, "Resilient machines through continuous self-modeling," Science, vol. 314, pp. 1118–1121, 2006
- [9] J. Evans and A. Rzhetsky, "Machine Science." Science (New York, N.Y.), vol. 329, no. 5990, pp. 399–400, Jul. 2010. [Online]
- [10] Maja Vukovic. 2009. "Crowdsourcing for Enterprises," in Proceedings of the 2009 Congress on Services - I (SERVICES '09). IEEE Computer Society, Washington, DC, USA, 686-692.
- [11] Thierry Bücheler, Rudolf M. Fuchslin, Rolf Pfeifer, Jan Henrik Sieg. (2010). In Proceedings of Artificial Life XII -- Proceedings of the Twelfth International Conference on the Synthesis and Simulation of Living Systems. Pages 679–686. MIT Press.
- [12] Keywon Chung, Carnaven Chiu, Xiao Xiao, and Pei-Yu (Peggy) Chi. 2009. "Stress outsourced: a haptic social network via crowdsourcing," In Proceedings of the 27th international conference extended abstracts on Human factors in computing systems (CHI '09). ACM, New York, NY, USA,.
- [13] Huberman, B. A., Romero, D. M., and Wu, F. "Social networks that matter: Twitter under the microscope," arXiv: 0812.1045v1, Dec 2008.
- [14] A. Romero-Corral, V. Somers, J. Sierra-Johnson, R. Thomas, M. Collazo-Clavell, J. Korinek, T. Allison, J. Batsis, F. Sert-Kuniyoshi, and F. Lopez-Jimenez, "Accuracy of body mass index in diagnosing obesity in the adult general population," *International Journal of Obesity*, vol. 32, no. 6, pp. 959–966, 2008.
- [15] L. Barness, J. Opitz, and E. Gilbert-Barness, "Obesity: genetic, molecular, and environmental aspects," *American Journal of Medical genetics Part A*, vol. 143, no. 24, pp. 3016–3034, 2007.