International Journal of Science Engineering and Advance Technology

Crowd sourced modeling approach of human behavior modeling

1Thumu Nagaraju, 2JagadeeswaraRao P, 3SayeedYasin
Dept. of CSE, Nimra Institute of Science & Technology, Vijayawada, AP, India

Abstract:

Generating models from large data setsand determining which subsets of data to mine is becoming increasingly automated. However choosing what data to collect in the first place requires human intuition or experience, usually supplied by a domain expert. This paper describes a new approach to machine science which demonstrates for the first time that non-domain experts can collectively formulate features, and provide values for those features such that they are predictive of some behavioral outcome of interest. This was accomplished by building a web platform in which human groups interact to both respond to questions likely to help predict a behavioral outcome and pose new questions to their peers. This results in a dynamically-growing online survey, but the result of this cooperative behavior also leads to models that can predict user's outcomes based on their responses to the user-generated survey questions.

1. Introduction:

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.

This paper introduces, for the first time, a method by which non domain experts can be motivated to formulate independent variables as well as populate enough of these variables for successful modeling. In short, this is accomplished as follows. Users arrive at a website in which a behavioral is to be modeles. Users provide their own outcome and then answer questions that may be predictive ofthat outcome.

2. Existing System:

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 . 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 need for the involvement of domain experts can become a bottleneck to new insights. However, if the wisdom of crowds could be harnessed to produce insight into difficult problems, one might see exponential rises in the discovery of the causal factors of behavioral outcomes, mirroring the exponential growth on other online collaborative communities. Thus, the goal of this research was to test an alternative approach to modeling in which the wisdom of crowds is harnessed to both propose potentially predictive variables to study by asking questions, and respond to those questions, in order to develop a predictive model.

3. Proposed System:

This paper introduces, for the first time, a method by which non domain experts can be motivated to formulate independent variables as well as populate enough of these variables for successful modeling. In short, this is accomplished as follows.

Users arrive at a website in which a behavioral outcome is to be modeled. Users provide their own outcome and then answer questions that may be predictive of that outcome. Periodically, models are constructed against the growing data sets that predict each user's behavioral outcome. Users may also pose their own questions that, when answered by other users, become new independent variables in the modeling process. In essence, the task of discovering and populating predictive independent variables is outsourced to the user community.

The rapid growth in user-generated content on the Internet is an example of how bottom-up interactions can, under some circumstances, effectively solve problems that previously required explicit management by teams of experts. Harnessing the experience and effort of large numbers of individuals is frequently known as "crowdsourcing" and has been used effectively in a number of research and commercial applications.

The most important datum shown to each user after respondingto each question was the value of their actual outcome asthey entered it (bi) as well as their outcome as predicted bythe current model ($\hat{}bi$). Fig. 2 illustrates that visitors to theBMI site were shown their actual BMIand their predicted BMI. The models were able to predict eachuser's outcome before they had responded to every question bysubstituting in missing values.

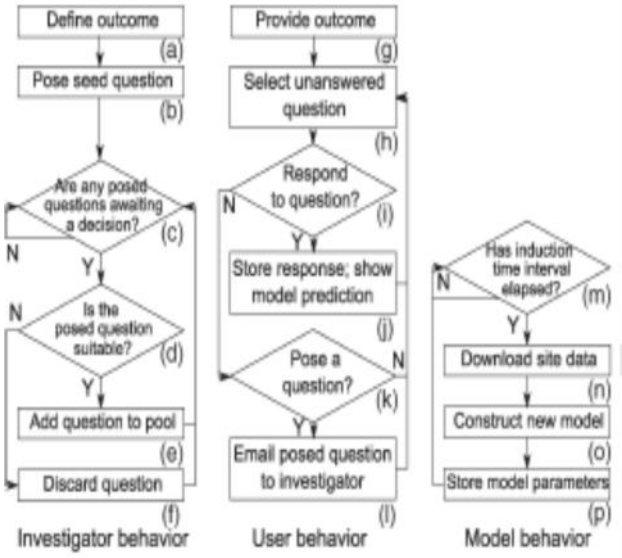

Fig. 2.Overview of the system. (a)–(f) The investigator is responsible forinitially creating the Web platform and seeding it with a starting question.

3.1 ENERGY EFFICIENCY INSTANTIATION:

We develop aWeb-based social network to model residential electric energyconsumption. Because of policy efforts to increase energyefficiency many are working to provide consumers withbetter information about their energy consumption. Researchon consumer perception of energy efficiency indicates thatelectricity customers often misjudge the relative importance ofvarious activities and devices to reducing energy consumption.

The *Answer Page*asked participants to respond to questions and provided themwith information about each answered question including thedistribution of answers within the social network. Finally a*Ranking Page* showed users their energy consumption, relativeto that of others in the group.

4. Performance evolution:

Fig. 2 shows the relative predictive power of the ten questions.The results show that the most highly correlated factors($Q3$,$Q5$, and $Q6$) were posed after the initial two seed questions[see Fig. 3(a)] and a weak correlation between the response rateand the $r2$ values, indicating that more answers to questionswould have likely produced improved results. Panels (c) and (d)show the distributions of $r2$ values and the number of responses,to facilitate comparison with the BMI results.

While the small sample size in this study limits the generalityof these results, this initial trial provided useful informationabout the crowdsourced modeling approach. First, we foundthat participants were reluctant or unable to provide accurateoutcome data due to the challenge of finding one's electricbills.

In order to further motivate the participants, in addition todisplaying their predicted outcome, users were also shown howtheir responses compared to two peer groups. For each user, thepeer groups were constructed as follows. The first peer groupwas composed of ten other users who had BMI values as closeto but below that of the user; the second group was composes of ten other users who had BMI values as close to but abovethat of the user.

The peer group data were meant to help users compare howtheir lifestyle choices measured up to their most similar peerswho were slightly more healthy than themselves, and slightlyless healthy than themselves.

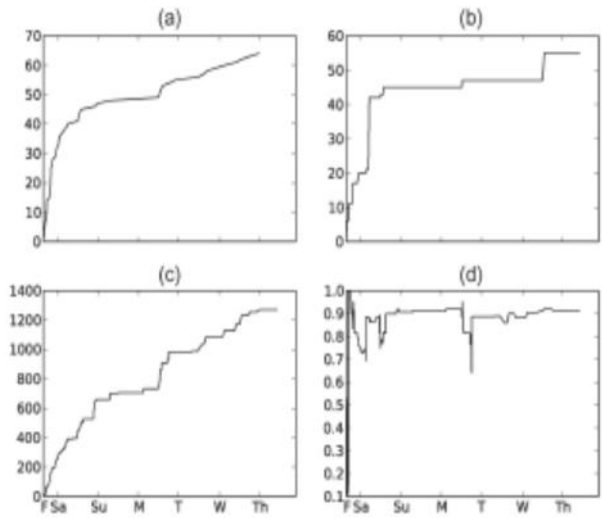

Fig. 2.User behavior on the BMI site. The BMI site was maintained forslightly less than seven days.

5.1 Conclusion:

This paper introduced a new approach to social science modeling in which the participants themselves are motivated to uncover the correlates of some human behavior outcome, such as homeowner electricity usage or body mass index. In both cases participants successfully uncovered at least one statistically significant predictor of the outcome variable. For the body mass index outcome, the participants successfully formulated many of the correlates known to predict BMI, and provided sufficiently honest values for those correlates to become predictive during the experiment. While, our instantiations focus on energy and BMI, the proposed method is general, and might, as the method improves, be useful to answer many difficult questions regarding why some outcomes are different than others.

Despite this initial success, much work remains to be done to improve the functioning of the system, and to validate its performance. The first major challenge is that the number of questions approached the number of participants on the BMI website.

6. REFERENCES

1. J. Bongard and H. Lipson, "Automated reverse engineering of nonlineardynamical systems," *Proc. Nat. Acad. Sci.*, vol. 104, no. 24, pp. 9943–9948, Jun. 2007.
2. J. Evans and A. Rzhetsky, "Machine science," *Science*, vol. 329, no. 5990,pp. 399–400, Jul. 2010.
3. R. D. King, K. E. Whelan, F. M. Jones, P. G. K. Reiser, C. H. Bryant,S. H. Muggleton, D. B. Kell, and S. G. Oliver, "Functional genomichypothesis generation and experimentation by a robot scientist," *Nature*,vol. 427, no. 6971, pp. 247–252, Jan. 2004.
4. R. King, J. Rowland, S. Oliver, M. Young, W. Aubrey, E. Byrne,M. Liakata, M. Markham, P. Pir, L. Soldatova, A. Sparkes, K. E. Whelan,and A. Clare, "The automation of science," *Science*, vol. 324, no. 5923,pp. 85–89, Apr. 2009.
5. J.Bongard,V. Zykov, andH. Lipson, "Resilientmachines through continuousself-modeling," *Science*, vol. 314, no. 5802, pp. 1118–1121,Nov. 2006.
6. J. Giles, "Internet encyclopedias go head to head," *Nature*, vol. 438,no. 15, pp. 900–901, Dec. 2005.
7. D. C. Brabham, "Crowdsourcing as a model for problem solving,"*Convergence*, vol. 14, no. 1, pp. 75–90, Feb. 2008.
8. A. Sorokin and D. Forsyth, "Utility data annotation with Amazon MechanicalTurk," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. PatternRecognit. Workshops*, Anchorage, AK, 2008, pp. 1–8.
9. M.Marge, S. Banerjee, and A.Rudnicky, "Using the AmazonMechanicalTurk for transcription of spoken language," in *Proc. IEEE Int. Conf.Acoust., Speech, Signal Process.*, Dallas, TX, 2010, pp. 5270–5273.\
10. N. Kong, J. Heer, and M. Agrawala, "Perceptual guidelines for creatingrectangular treemaps," *IEEE Trans. Vis. Comput. Graphics*, vol. 16, no. 6,pp. 990–998, Nov. 2010.
11. A. Kittur, E. Chi, and B. Suh, "Crowdsourcing user studies with MechanicalTurk," in *Proc. 26th Annu. SIGCHI Conf. Human Factors Comput.Syst.*, Florence, Italy, 2008.
12. D. Wightman, "Crowdsourcing human-based computation," in *Proc.6th Nordic Conf. Human-Comput. InteractExtending Boundaries*, Reykjavik, Iceland, 2010.

THUMU NAGARAJU is a M.tech(CSE) student of NIMRA Institute of science & Technology an affiliated college ofJNTUK, Kankinada,INDIA. He has done her B.Tech(cse) in SSNENGINEERINGcollege AT ongoleof JNTU Kakinada,INDIA.

JAGADEESWARA RAO P received his M.Tech from CSE-SRMUNIVERSITY,KATTANKULTHR, CHENNAI. He is currently working as an Assistant Professor in NIMRA Institute of Science & Technology in the Department of Computer Science & Engineering, Ibrahimpatnam, Vijayawada. He has two years of Experience in teaching. E-Mail :pjagadeeswararao@gmail.com

SAYEEdYASIN received hisMTECH in Computer Science &Enggfrom JNTU Hyderabad. He is pursuingPh.D in RayalaseemaUniversity,Kurnool. He is currently working asAssoc. Professor & HOD in Nimra Institute of Science & Technology the Department of Computers Science and Engineering, Vijayawada.He has more than Eight years of experience in teaching.His area of interests are wireless