ISSN 2321-6905
January-2015

International Journal of Science Engineering and Advance Technology

Crowdsourcing: A New Approach Of Non-Domain Experts To Provide Predictive Behavioral Results

Yalamanchili Pavan

Abstract:
Bringing about models from great data sets and influential which subsets of data to mine is becoming more and more automated. Though preferring what data to gather in the first place necessitate human intuition or experience typically complete by a domain expert. In paper we explain a new approach to machine science which makes obvious for the first time those non-domain experts can communally put together features and present values for those features such that they are predictive of some behavioral outcome of interest. This was completed by building a web platform in which human groups interrelate to both respond to questions likely to help predict a behavioral outcome and pose new questions to their peers. These consequences are in a vibrant growing online survey the result of this supportive behavior also leads to replica that can forecast user's outcomes based on their answer to the user-generated survey questions.

Keywords: Crowd sourcing, machine science, surveys, social media, human behaviour modelling.

Introduction:

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 models. Users provide their own outcome and then answer questions that may be predictive of that outcome. Statistical tools 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-specific. 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. A survey designer must have domain expertise to choose questions that will identify predictive covariates.

Related Work:

The rapid growth in user-generated content on the Internet is an example of how bottom-up interactions 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 "crowd sourcing" and has been used effectively in a number of research and commercial applications. For an instance of how crowd sourcing can be useful, consider Amazon's Mechanical Turk. In this crowd sourcing tool a human describes a "Human Intelligence Task" such as characterizing data transcribing spoken language or creating data visualizations. By involving large groups of humans in many locations it is possible to complete tasks that are difficult to accomplish with computers alone and would be prohibitively expensive to accomplish through traditional expert-driven processes.

Existing Method:

The need for the involvement of domain experts can become a bottleneck to new insights. 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.

Disadvantages:

Although arguably not strictly a crowd sourced system, the rapid rise of Wikipedia illustrates

ISSN 2321-6905
January-2015

how online collaboration can be used to solve difficult problems without financial incentives.

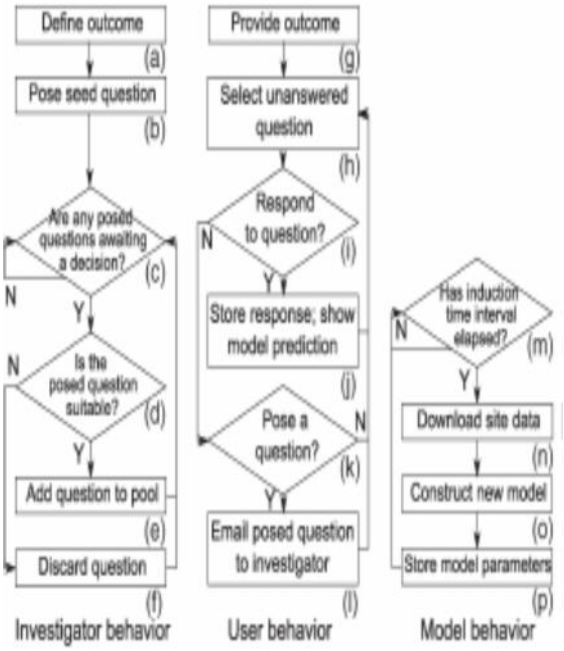
Proposed Method:

The rapid growth in user-generated content on the Internet is how bottom-up interactions can under some circumstances effectively solve problems that previously required explicit management by teams of experts. 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.

Advantages:

Visitors to the BMI site were shown their actual BMI and their predicted BMI. The models were able to predict each user's outcome before they had responded to every question by substituting in missing values.

System Architecture:



It shows the listing that was shown to those users who participated in the BMI site; illustrates that visitors to the BMI site were shown their actual BMI (as entered by them) and their predicted BMI. The models were able to predict each user's outcome before they had responded to every question by substituting in missing values. The most important datum 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).

Body Mass Index Instantiation:

Participants arriving for the first time at the BMI site were asked to enter their height and weight in feet, inches and pounds respectively In order to further motivate the participants, in addition to displaying their predicted outcome, users were also shown how their responses compared to two peer groups. For each user the peer groups were constructed as follows. The first peer group was composed of 10 other users who had BMI values as close to but below that of the user; the second group was composed of 10 other users who had BMI values as close to but above that of the user The peer group data were meant to help users compare how their lifestyle choices measured up to their most similar peers who were slightly more healthy 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.

Energy Efficiency Instantiation:

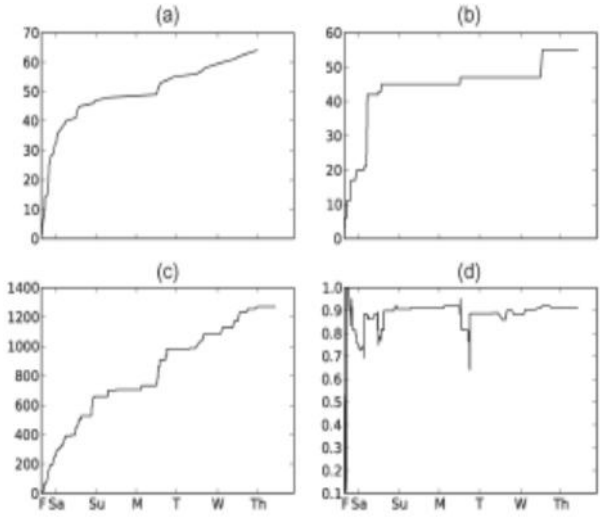
The *Answer Page* asked participants to respond to questions and provided them with information about each answered question including the distribution of answers within the social network. Finally a *Ranking Page* had shown users their energy consumption relative to that of others in the group. We develop a Web-based social network to model residential electric energy consumption. Because of policy efforts to increase energy efficiency many are working to provide consumers with better information about their energy consumption. Research on consumer perception of energy efficiency indicates that electricity customers often misjudge the relative importance of various activities and devices to reducing energy consumption.

Experimental Results:

The results show that the most highly correlated factors (*Q*3,*Q*5, and *Q*6) were posed after the

ISSN 2321-6905
January-2015

initial two seed questions and a weak correlation between the response rate and the *r*2 values, indicating that more answers to questions would have likely produced improved results. Panels (c) and (d) show the distributions of *r*2 values and the number of responses, to facilitate comparison with the BMI results. The study limits the generality of these results. This initial trial provided useful information about the crowd sourced modeling approach. For each user, the peer groups were constructed as follows. The first peer group was composed of ten other users who had BMI values as close to but below that of the user the second group was composes of ten other users who had BMI values as close to but above that of the user.



Conclusion:

It is 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 methods participants successfully uncover 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 provide 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.

References:

[1] J. Bongard and H. Lipson, "Automated reverse engineering of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 104, no. 24, pp. 9943–9948, 2007.

[2] J. Evans and A. Rzhetsky, "Machine science," *Science*, vol. 329, no. 5990, p. 399, 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 genomic hypothesis generation and experimentation by a robot scientist," *Nature*, vol. 427, pp. 247–252, 2004.

[4] R. King, J. Rowland, S. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L. Soldatova *et al.*, "The automation of science," *Science*, vol. 324, no. 5923, p. 85, 2009.

[5] J. Bongard, V. Zykov, and H. Lipson, "Resilient machines through continuous self-modeling," *Science*, vol. 314, pp. 1118–1121, 2006.

[6] J. Giles, "Internet encyclopedias go head to head," *Nature*, vol. 438, no. 15, pp. 900–901, 2005.

[7] D. C. Brabham, "Crowdsourcing as a model for problem solving," *Convergence*, vol. 14, pp. 75–90, 2008.

[8] A. Sorokin and D. Forsyth, "Utility data annotation with amazon mechanical turk," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008.

[9] M. Marge, S. Banerjee, and A. Rudnicky, "Using the amazon mechanical turk for transcription of spoken language," in *Proc. IEEE International Conference on Acoustics Speech and Signal Processing*, 2010.

[10] N. Kong, J. Heer, and M. Agrawala, "Perceptual guidelines for creating rectangular treemaps," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, 2010.

[11] A. Kittur, E. Chi, and B. Suh, "Crowdsourcing user studies with mechanical turk," in *Proc.Twenty-sixth annual SIGCHI*

ISSN 2321-6905
January-2015

conference on human factors in computing systems, 2008.
[12] D. Wightman, "Crowdsourcing human-based computation," in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, 2010.
[13] B. Fitzgerald, "The transformation of open source software," *Management Information Systems Quarterly*, vol. 30, no. 3, pp. 587–598, 2006.
[14] J. Howe, *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business*. Crown Business, 2009.
[15] N. Thurman, "Forums for citizen journalists? adoption of user generated content initiatives by online news media," *New Media and Society*, vol. 10, no. 1, 2008.