MIS Quarterly

RESEARCH ARTICLE

LEVERAGING PHILANTHROPIC BEHAVIOR FOR CUSTOMER SUPPORT: THE CASE OF USER SUPPORT FORUMS

Wael Jabr
Haskayne School of Business, University of Calgary,
Calgary, AB, T2N 1N4 CANADA {wjabr@ucalgary.ca}

Radha Mookerjee
Naveen Jindal School of Management, University of Texas at Dallas,
Richardson, TX 75080 U.S.A. {radham@utdallas.edu}

Yong Tan
Michael G. Foster School of Business, University of Washington,
Seattle, WA 98195 U.S.A. {ytan@uw.edu}

Vijay S. Mookerjee
Naveen Jindal School of Management, University of Texas at Dallas,
Richardson, TX 75080 U.S.A. {vijaym@utdallas.edu}

Appendix A

Summary Statistics Per Forum

Table A1. SAP

Variable	Obs	Mean	Std. Dev.	Min	Max
Contribution	279078	6.0977	27.5936	0	1720
Helpful Posts	279078	8.4931	21.6105	0	100
Correct Posts	279078	3.0169	10.9153	0	100
Thanks Posts	279078	3.9769	12.8371	0	100
Solution Score	279078	44.7060	168.0105	0	9420
Log Post Score	279078	4.7075	1.2368	0	9.7927
Others' Solution Posts	279078	6.5051	21.9278	0	95.2949
Distinct Solvers	279078	2584.1360	1631.6450	356	8891
Institutional Recognition	279078	8.0728	30.9391	0	100
Solver with Solver Exposure	279078	1.1413	7.2797	0.4377	31.87
Solver with Questioner Exposure	279078	2.8432	9.1447	0	429
Role	279078	0.3661	0.4817	0	1
Subforums	279078	0.7203	1.3048	1	37
Non Question Posts	279078	6.3095	28.2920	0	1715
Age	279078	32.4515	12.5480	0	84.63

Table A2. Apple

Variable	Obs	Mean	Std. Dev.	Min	Max
Contribution	236997	14.8876	86.1974	0	1787
Helpful Posts	236997	1.9917	10.6855	0	100
Correct Posts	236997	1.7151	8.6407	0	100
Thanks Posts	236997	4.7812	29.4485	0	100
Solution Score	236997	126.5910	793.5888	0	26464
Log Post Score	236997	1.9219	1.8024	0	10.1828
Others' Solution Posts	236997	76.7652	14.4829	29.7653	86.8556
Distinct Solvers	236997	3226.6040	741.5436	1930	4665
Institutional Recognition	236997	16.6845	30.6425	0	100
Solver with Solver Exposure	236997	1.1173	5.9185	0	27.73149
Solver with Questioner Exposure	236997	6.9270	32.6762	1	956
Role	236997	0.1996	0.3997	0	1
Subforums	236997	1.0735	0.6262	1	41
Non Question Posts	236997	11.4884	90.5642	0	8873
Age	236997	23.9562	23.2531	0	94.23

Table A3. SUN

Variable	Obs	Mean	Std. Dev.	Min	Max
Contribution	241972	24.3724	85.5228	0	846
Helpful Posts	241972	2.0423	10.5092	0	100
Correct Posts	241972	1.7785	8.3608	0	100
Thanks Posts	241972	3.8986	16.7202	0	100
Solution Score	241972	62.1308	292.8857	0	8752
Log Post Score	241972	4.0251	1.0944	0	10.0498
Others' Solution Posts	241972	74.8666	14.2593	27.7653	85.1969
Distinct Solvers	241972	3144.3810	897.8769	1930	4665
Institutional Recognition	241972	17.1113	30.8622	0	100
Solver with Solver Exposure	241972	2.0493	16.7808	0	21.16773
Solver with Questioner Exposure	241972	1.9073	5.5835	1	956
Role	241972	0.2015	0.4011	0	1
Subforums	241972	1.6906	2.4955	1	41
Non Question Posts	241972	29.1247	58.0988	0	4222
Age	241972	26.6174	29.3123	0	94.23

Table A4. ORACLE

Variable	Obs	Mean	Std. Dev.	Min	Max
Contribution	221919	13.5794	54.2582	0	899
Helpful Posts	221919	2.5986	14.0817	0	100
Correct Posts	221919	3.7383	15.1527	0	100
Thanks Posts	221919	2.5208	12.2326	0	100
Solution Score	221919	161.2927	627.0715	0	11747
Log Post Score	221919	4.3518	1.7329	0	9.6745
Others' Solution Posts	221919	56.9703	15.0856	13.9236	72.4778
Distinct Solvers	221919	490.4739	304.8221	37	1094
Institutional Recognition	221919	7.0236	24.5911	0	100
Solver with Solver Exposure	221919	1.2363	13.2786	0	17.644
Solver with Questioner Exposure	221919	6.3469	27.7884	0	824
Role	221919	0.1989	0.3991	0	1
Subforums	221919	1.2099	0.8781	1	27
Non Question Posts	221919	10.5474	44.3864	0	1395
Age	221919	33.6359	37.4173	0	144.83

Appendix B

Text Mining Performance in Each of the Four Forums for Each of the Three Classifiers

The text mining approach we adopt is a classification one that assigns a post to one of two categories. The classifier takes posts as input and returns a classified output based on the first-best guess for the category of each input item.

Specifically, and for each of our three variables, the text mining approach is implemented as follows. First, we construct a training data set of labeled posts by manually labeling each post in the data set with its appropriate class. In our setup, this is a binary classification (e.g., a post is a solution or not). To derive the *Contribution Level* variable, which is a count of the number of solutions, we compile a set of 6,000 posts (equally drawn from each of the four forums) that have been manually labeled as solutions. We then manually label another 6,000 posts as non-solution (equally drawn from each of the four forums as well). We can now use all 12,000 posts to classify posts as a solution or not. We follow the same approach for the other two variables by manually labeling posts as *Thanks Posts* or not (12,000 training set in total), and *Non Question* posts or not (12,000 training set in total). Second, we use the LingPipe toolkit (Alias-i 2009) to create a Dynamic Language Model classifier. LingPipe provides a classification facility that takes training sets and learns how to classify further documents (posts) using what it learned with such language models. We use 10-fold cross-validation to train and validate the classifier. Third, after adjusting relevant parameters to ensure that each of the three classifiers (for the three variables) have reasonably high accuracy (ranging between 88.9% and 91.3%), we evaluate the classifiers by running them on a completely different validation set of data. Our classification of the full data set is then done in a sequential manner. We begin with a data set of 12,000 posts, each of which has been manually pre-labeled as *Question* or *Non-Question*, *Solution* or *Non-Solution*, and *Thanks* or *Non-Thanks*. We then run our first classifier on the entire data set and classify the data set as one of two types—one containing *Non-Question* posts and the other *Question* posts. We then run the second classifier on the data set with *Non-Question* posts and classify the set into *Solution* posts and *Non-Solution* posts. We finally run the last classifier to classify the *Non-Solution* posts into *Thanks* and *Non-Thanks* posts. Table B1 details the performance of the text mining approach along with additional robustness checks.

Further, we ensure that each of the classifiers is *unbiased*. An unbiased classifier is one that has the same error rate in both classes. Since we use the text miner to derive the dependent variable as well as some of the independent variables used in the regression analysis, ensuring that the classifier is unbiased avoids the problem of measurement error. This also enables us to satisfy regression assumptions and avoid inflation in the overall regression error as well as the variance of estimators.

Table B1. Performance of Text Mining with Additional Robustness Checks

Classifier	Accuracy	Sensitivity (Recall)	Specificity	Precision
Solution	90.7%	90.1%	91.1%	91.2%
Question	89.4%	89.9%	89.0%	88.9%
Thanks	90.8%	91.3%	90.2%	90.0%

Note that we use well-established metrics to evaluate the text mining performance. For a binary classification into two types, positive (P) and negative (N), with the classifier making correct classification (True; T) or incorrect one (False; F), *accuracy* is the number of correct identifications divided by the total number of instances, *precision* is the ratio TP/(TP+FP), *recall* is the ratio of TP/(TP+FN), and specificity is the ratio of TN/(TN+FP).

Appendix C

Negative Binomial Model Selection

The dependent variable is a count variable. The Poisson regression model implements the canonical regression specification for a variable Y that is a count of events:

$$Prob[Y = y_{it} | x_{it}] = \frac{exp(-\lambda_i)\lambda_i^{y_{it}}}{\Gamma(1+y_{it})}$$

where $\lambda_i = exp(\alpha + \boldsymbol{x}_{it}'\beta)$, y_i - 0,1, …, $i = 1, \ldots N$, $t = 1, \ldots T$,
$\boldsymbol{x}_{it}$ is a vector of covariates,
$i = 1, \ldots N$ indexes the N observations and $t = 1, \ldots T$ indexes the T time periods.

Here the conditional mean is $E[y_{it}|\boldsymbol{x}_{it}, \boldsymbol{\varepsilon}_{it}] = \lambda_i$ and its equidispersion is $Var[y_{it}|\boldsymbol{x}_{it}] = \lambda_i$.

Since observed data will almost always display pronounced overdispersion, allowing for overdispersion dictates the use of the negative binomial model (Cameron et al. 1986; Greene 2008). Thus, the negative binomial model is employed as a functional form that relaxes the equidispersion restriction of the Poisson model. A useful way to motivate the model is through the introduction of latent heterogeneity in the conditional mean of the Poisson model. Thus, we write

$$E[y_{it}|\boldsymbol{x}_{it}, \boldsymbol{\varepsilon}_{it}] = exp(\alpha + \boldsymbol{x}_{it}'\beta) = h_i\lambda_i$$

where $h_i = exp(\boldsymbol{\varepsilon}_{it})$ is assumed to have a one parameter gamma distribution, G(θ,θ) with mean 1 and variance $1/\theta = \kappa$.

After integrating h_i out of the joint distribution, we obtain the marginal negative binomial (NB) distribution,

$$Prob[Y = y_{it} | x_{it}] = \frac{\Gamma(\theta + y_{it})r_i^{\theta}(1-r_i)^{y_{it}}}{\Gamma(1+y_{it})\Gamma(\theta)}$$

with $y_{it} = 0, 1, \ldots, \theta > 0$, $r_i = \theta/(\theta + \lambda_i)$.

The latent heterogeneity induces overdispersion while preserving the conditional mean:

$$E[y_{it}|\boldsymbol{x}_{it}, \boldsymbol{\varepsilon}_{it}] = \lambda_i$$
$$Var[y_{it}|\boldsymbol{x}_{it}] = \lambda_i[1 + (1/\theta)\lambda_i] = \lambda_i[1 + \kappa\lambda_i]$$

where $\kappa = Var[h_i]$.

Maximum likelihood estimation of the parameters of the NB model (α, β, θ) is straightforward.

We use fixed effect estimation for the model specified. Fixed effects regression is the appropriate model to use when we aim to control for omitted variables that differ between cases but are constant over time (Baltagi 2008).

Appendix D

Correlation Matrix Per Forum

Table D1. SAP

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. HelpfulPosts	1												
2. CorrectPosts	0.0422	1											
3. ThanksPosts	0.1615	0.1372	1										
4. SolScore	0.0672	0.026	0.0057	1									
5. LogPostScore	0.0345	0.0066	0.0305	0.3686	1								
6. OtherSol	-0.0105	-0.0004	-0.0067	0.0152	-0.084	1							
7. DistSol	-0.0616	-0.0692	-0.0376	0.0656	0.1705	0.0282	1						
8. Institutional Recognition	0.0562	0.0179	-0.0005	0.3724	0.3492	0.0177	0.0792	1					
9. Solver_Solver Exposure	0.0339	0.0344	0.0547	-0.0441	-0.0036	0.0308	0.103	-0.0428	1				
10. Solver_Questioner Exposure	0.067	0.0221	-0.0018	0.3657	0.363	0.0197	0.0777	0.3843	-0.0485	1			
11. Role	-0.0745	-0.0798	-0.0237	-0.0157	0.0362	0.0275	0.21	0.0134	-0.0027	-0.0067	1		
12. Subforums	0.1504	0.0832	0.043	0.4033	0.2864	0.0122	0.1379	0.4818	-0.0353	0.4244	-0.0106	1	
13. NonQPosts	0.065	0.0262	0.0031	0.3722	0.3372	0.012	0.0605	0.3781	-0.0477	0.3578	-0.0113	0.4626	1
14. Age	0.0235	0.0728	0.0731	-0.0396	0.2878	-0.2202	-0.0427	-0.0532	0.0525	-0.0462	-0.1853	-0.0024	-0.0425

Table D2. Apple

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. HelpfulPosts	1												
2. CorrectPosts	0.0831	1											
3. ThanksPosts	0.11	0.4019	1										
4. SolScore	0.0592	0.0714	0.3462	1									
5. LogPostScore	0.1695	0.171	0.2723	0.3702	1								
6. OtherSol	0.0239	0.018	0.0094	-0.0092	-0.0094	1							
7. DistSol	0.012	0.0141	0.002	-0.0186	-0.0403	0.4082	1						
8. Institutional Recognition	0.269	0.3162	0.1881	0.0729	0.2446	0.0232	0.03	1					
15. Solver_Solver Exposure	0.1162	0.1304	0.3764	0.3333	0.3657	0.0096	0.0087	0.1587	1				
16. Solver_Questioner Pres	0.1169	0.1314	0.3876	0.2371	0.3707	0.0095	0.0091	0.1548	0.3764	1			
9. Role	0.0791	0.0591	0.0449	0.0213	0.1273	-0.0013	0.0217	0.4045	0.103	0.0962	1		
10. Subforums	0.0864	0.0973	0.3047	0.3963	0.2401	0.0027	-0.0013	0.1017	0.2387	0.2599	0.0398	1	
11. NonQPosts	0.0797	0.0937	0.4618	0.2497	0.2991	0.0073	0.0038	0.1054	0.4429	0.3138	0.0573	0.4104	1
12. Age	-0.0286	-0.0222	0.0344	0.1132	0.2953	-0.2627	-0.3675	-0.0589	0.0307	0.0345	-0.0449	0.0474	0.0278

Table D3. SUN

	1	2	3	5	6	7	8	9	10	11	12	13	14
1. HelpfulPosts	1												
2. CorrectPosts	0.081	1											
3. ThanksPosts	0.1394	0.3648	1										
4. SolScore	-0.0125	-0.0064	-0.0056	1									
5. LogPostScore	0.1726	0.1612	0.2624	0.3766	1								
6. OtherSol	0.0309	0.028	0.0314	-0.0068	-0.0067	1							
7. DistSol	0.0145	0.0173	0.0206	-0.0177	-0.0291	0.4247	1						
8. Institutional Recognition	0.2797	0.3361	0.3092	-0.0335	0.2243	0.0335	0.0405	1					
17. Solver_Solver Exposure	0.0419	0.0254	0.0451	-0.0029	0.0571	0.0159	-0.0164	0.0793	1				
18. Solver_Questioner Pres	0.1869	0.1059	0.1965	0.0306	0.4539	0.0221	0.0145	0.1364	0.0313	1			
9. Role	0.0806	0.0573	0.0895	-0.0381	0.0769	0.0015	0.0298	0.3783	0.0707	0.0924	1		
10. Subforums	0.1874	0.1491	0.1913	-0.0176	0.3717	0.0268	0.0089	0.2211	0.0293	0.4589	0.133	1	
11. NonQPosts	0.2075	0.1676	0.2849	0.023	0.3412	0.0314	0.026	0.2461	0.05	0.3974	0.164	0.2946	1
12. Age	-0.0503	-0.0428	-0.0329	0.1098	0.2416	-0.2924	-0.3638	-0.074	0.0116	-0.0026	-0.0564	0.0133	-0.0185

Table D4. ORACLE

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. HelpfulPosts	1												
2. CorrectPosts	0.0717	1											
3. ThanksPosts	0.311	0.0688	1										
4. SolScore	0.0087	0.0012	0.0136	1									
5. LogPostScore	0.0293	-0.0407	0.0419	0.4786	1								
6. OtherSol	0.055	0.0601	0.0391	-0.1193	-0.1631	1							
7. DistSol	0.192	0.2682	0.1726	0.0128	-0.074	0.4012	1						
8. Institutional Recognition	0.0089	0.0083	0.0162	0.4848	0.3425	-0.1607	0.1201	1					
19. SolverWithSolver Exposure	0.0389	0.0413	0.0304	0.0017	0.0234	0.0017	0.0803	0.0153	1				
20. Solverwith QuestionerPres	0.0075	0	0.0149	0.4979	0.3279	-0.1024	0.0775	0.1083	0.0055	1			
9. Role	0.1067	0.1574	0.1081	0.0185	0.0359	-0.2565	0.212	0.1578	0.065	0.0827	1		
10. Subforums	-0.0026	0.0093	0.0144	0.2941	0.2456	-0.1236	0.1037	0.403	0.0233	0.3744	0.1484	1	
11. NonQPosts	0.0043	-0.0014	0.0105	0.4493	0.3172	-0.1374	0.1035	0.3474	0.0125	0.3218	0.1294	0.4715	1
12. Age	-0.0204	-0.0322	-0.0146	0.1851	0.2525	-0.0522	-0.0893	0.0672	-0.0162	0.072	-0.1214	0.0551	0.0646

Appendix E

Propensity Score Analysis

Table E1. Propensity Score (Probit) Estimation with the Dependent Variable Being the Participation in the Forum

Independent Variable	Probit Estimation		
	Coefficient	Standard Error	Significance
Correct Posts	0.0743	0.0434	***
Helpful Posts	0.0354	0.0437	**
Thanks Posts	0.0312	0.0247	**
Solver with Solver Exposure	0.5105	0.2528	***
Solver with Questioner Exposure	0.0020	0.0354	**
Membership	-0.0003	0.0005	NS
Question Posts	-0.0457	0.0494	**
Role	0.0151	0.0195	*
Number of Subforums	-0.0006	0.0012	NS
Total Posts	0.0003	0.0041	NS

The dependent variable is the participation in the forum with feedback-based recognition mechanism (DV = 1) or quantity-based recognition mechanism (DV = 0). As displayed in this table, the sets of variables related to responsiveness to the community and exposure within the community are sensible (with respect to significance and sign) for the propensity estimation.

Table E2. Regression Model Estimation for Difference in Contribution Across Treatments

		Difference in Contribution Across Treatment Clones		
		Coefficient	**Standard Error**	**Significance**
Peer Recognition	**Helpful Posts**			
	Lag 1	0.0009	0.0006	**
	Lag 2	0.0001	0.0006	*
	Lag 3	5.04E-05	0.0008	NS
	Correct Posts			
	Lag 1	0.0039	0.0017	**
	Lag 2	0.0015	0.0013	NS
	Lag 3	0.0006	0.0015	NS
	Thanks Posts			
	Lag 1	0.0010	0.0007	***
	Lag 2	0.0013	0.0010	*
	Lag 3	0.0009	0.0008	NS
Image Motivation	**Solution Score**			
	Lag 1	0.0006	0.0001	**
	Log Post Score			
	Lag 1	-0.0400	0.0000	***
	Institutional Recognition			
	Lag 1	-0.0058	0.0014	NS
Social Comparison	**Others' Solution Posts**			
	Lag 1	-6.19E-05	0.0004	NS
	Distinct Solvers			
	Lag 1	-5.46E-06	7.39E-06	NS
Social Exposure	**Solver with Solver Presence**			
	Lag 1	0.0003	0.0004	***
	Solver with Questioner Presence			
	Lag 1	0.0048	0.0017	***

References

Alias-i. 2009. LingPipe 3.8.1 (http://alias-i.com/lingpipe; accessed June 2009).

Baltagi, B. 2008. *Econometric Analysis of Panel Data* (4th ed.), Chichester, UK: John Wiley and Sons.

Cameron, A. C., and Trivedi, P. K. 1986. "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests," *Journal of Applied Econometrics* (1:1), pp. 29-53.

Greene, W. 2008. *Econometric Analysis* (7th ed.), Upper Saddle River, NJ: Prentice Hall.