MIS Quarterly

RESEARCH ARTICLE

DIFFERENTIAL INFLUENCE OF BLOGS ACROSS DIFFERENT STAGES OF DECISION MAKING: THE CASE OF VENTURE CAPITALISTS

Rohit Aggarwal
Operations and Information Systems, David Eccles School of Business, University of Utah,
Salt Lake City, UT 84112 U.S.A. {Rohit.Aggarwal@business.utah.edu}

Harpreet Singh
Information Systems and Operations Management, Naveen Jindal School of Business, University of Texas at Dallas,
Richardson, TX 75080 U.S.A. {Harpreet@utdallas.edu}

Appendix A

Detail of the Sentiment Analysis

In this work, we followed the sentiment analysis literature, and used supervised learning methods, which take manually classified data (corpus) as input and automatically extract features (combination of words and parts of speech of words) for sentiment analysis (Dave et al. 2003; Ghose and Ipeirotis 2011; Pang et al. 2002; Shanahan et al. 2006). These supervised methods do not rely on manually or semi-manually constructed discriminant-word lexicons. Prior research has shown that supervised methods perform better than lexicon-based approaches for sentiment analysis (Chaovalit and Zhou 2005; Pang et al. 2002).

We used two dimensions across which we developed in total four classifiers. One dimension used is the venture category provided by VCs, and the two categories provided are *software* and *IT services*. The second dimension used is the media type, and the two media types are *blog posts* and *media and press releases*. Next, taking an example of classifier for *blog posts* under venture category *software*, we provide details about how we constructed a classifier. We randomly picked 600 blog posts for ventures under the category *software*. Three raters (graduate students) then manually classified posts into one of the following three categories: positive, negative, and neutral. Each post was categorized by two randomly assigned raters out of three raters. In the case of disagreement, the third rater's judgment was considered as final. To check for inter-rater reliability, we calculated Cohen's kappa between each two pairs (Cohen 1960). Cohen's kappa is a statistical measure for examining the reliability of agreement between two raters when categorizing items. Cohen's kappa came out to be 0.807, 0.812, and 0.818, which connotes almost perfect agreement among raters (Landis and Koch 1977).

The manually classified corpus was then used to construct a classifier using a language-based model, namely, the dynamic language model (Carpenter 2005), using LingPipe. Typically past studies use n-gram (specifically, n = 8) based character language model with a generalized form of Witten-Bell smoothing to create a classifier from the manually classified corpus (Ghose and Ipeirotis 2011). Using this approach, the accuracy of our classifier was 70.12 percent, which is quite good. However, in this study, we needed to classify posts into three categories, and previous research on sentiment analysis has shown that in such situations, pair-wise coupling with stacking gives better results than a single multivariate classifier (Koppel and Schler 2006). The basic idea of pair-wise coupling is to convert *c*-class problem into series of binary classification problems by learning one classifier for each pair. This is accomplished by using training data of the pair alone and ignoring rest of the data. The predictions from these classifiers are then combined via a framework referred to as *stacking*—mapping of each of 2^c possible outcomes to a particular class *c* (Wolpert 1992). This approach requires us to built three binary classifiers— positive/negative, negative/neutral, and positive/neutral—and then combine possible outcomes into a single category (*class*) using optimal stack (Koppel and Schler 2006; Savicky and Fuernkranz 2003). The accuracy of this approach was 73.47 percent, which is better than the single multivariate classifier approach. We have reported model results using pair-wise coupling in the paper, and the results from the single multivariate classifier were also qualitatively similar.

Appendix B

Survey Conducted at Venture Capital Summit, New York on June 8–9, 2007

A total of 55 entrepreneurs participated in this survey. Some more information such as the industry-type of ventures, the development stage of ventures, and relevant industry experience of entrepreneurs, was also collected but is not reported here.

Questions	Yes / 1st option	No / 2nd option	Don't Know
(a) Other than blogs, is there any other communication medium such as discussion forums, and online posted reviews, that are important part of eWOM for new ventures?	4 (7.3%)	46 (83.6%)	5 (9.1%)
(b) Does volume of eWOM influence the amount financed to new ventures?	40 (72.7%)	8 (14.5%)	7 (12.7%)
(c) Does volume of eWOM influence the new ventures' valuations?	42 (76.4%)	6 (10.9%)	7 (12.7%)
(d) Does tone of eWOM influence the amount financed to new ventures?	39 (70.9%)	9 (16.4%)	7 (12.7%)
(e) Does tone of eWOM influence the new ventures' valuations?	41 (74.5%)	7 (12.7%)	7 (12.7%)

Appendix C

VCs Interviewed

VCs	Firm Name	Some of the Financed Ventures
Greg McAdoo	Sequoia Capital	Apple, Cisco Systems, Google, Oracle, Yahoo, YouTube, PayPal, Symantec, Palm Commerce, Paras Pharmaceuticals, SourceForge, Kayak, LinkedIn
Ajit Nazre	Kleiner Perkins Caufield Byers	Amazon, Sun, Genentech, Intuit, Verisign, Google
Andrew Braccia	Accel Partners	Facebook, Walmart.com, Macromedia, Real Networks, Kayak, Rhapsody Networks,
Keyur Patel	Velocity IG	NDTV, Next New Networks, Radar Networks, Broadband Enterprises
Sridar Iyengar	Bessemer Venture Partners	Skype, Staples, Dick' Clothing and Sporting Goods, Verisign, Sahara Networks, Gartner group, Hotjobs
Venkat A. Mohan	Norwest Venture Partners	Peoplesoft, SPSS, Kayak, SideStep, Nextel, Polycom, Lending Club, Cray Research
Joshua Goldman		
David Hornik	August Capital	Seagate Technology, Atheros Communications, Technorati, Six Apart, Shopping.com (formerly Epinions), Guru, Silicon Image
John W. Jarve	Menlo Ventures	Hotmail, MobiTV, 3Par, Acme Packet, Ascend, Catena
Ken Elefant	Opus Capital	DoubleClick, Genesys Labs, Phone.com
U. Haskell Crocker II	VIMAC Ventures	Half.com, Focus Enhancements, WebEd, Convergence Corporation

References

Carpenter, B. 2005. "Scaling High-Order Character Language Models to Gigabytes," in *Proceedings of the 2005 Association for Computational Linguistics Software Workshop*, Ann Arbor: Association for Computational Linguistics.

Chaovalit, P., and Zhou, L. 2005. "Movie Review Mining: A Comparison Between Supervised and Unsupervised Classification Approaches," in *Proceedings of the 38th Hawaii International Conference on System Sciences*, Los Alamitos, CA: IEEE Computer Society.

Cohen, J. 1960. "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement* (20:1), pp. 37-46.

Dave, K., Lawrence, S., and Pennock, D. M. 2003. "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews," in *Proceedings of the 12th International Conference on the World Wide Web*, May 20-24, Budapest, Hungary, pp. 519-528.

Ghose, A., and Ipeirotis, P. 2011. "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *IEEE Transactions On Knowledge and Data Engineering* (23:10), pp. 1498-1512.

Koppel, M., and Schler, J. 2006. "The Importance of Neutral Examples for Learning Sentiment," *Computational Intelligence* (22:2), pp. 100-109.

Landis, J. R., and Koch, G. G. 1977. "The Measurement of Observer Agreement for Categorical Data," *Biometrics* (33), pp. 159-174.

Pang, B., Lee, L., and Vaithyanathan, S. 2002. "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Philadelphia, July 7-12, pp. 79-86.

Saviky, P., and Fuernkranz, J. 2003. "Combining Pairwise Classifiers with Stacking," in *Advances in Intelligent Data Analysis*, V. M. R. Berthold, H. J. Lenz, E. Bradley, R. Kruse, and C. Borgelt (eds.), Berlin: Springer, pp. 219-229.

Shanahan, J. G., Qu, Y., and Wiebe, J. M. 2006. *Computing Attitude and Affect in Text: Theory and Applications*,. New York: Springer-Verlag.

Wolper, D. H. 1992. "Stacked Generalizatoin," *Neural Networks* (5:2), pp. 241-259.