Toward a Content-Driven Approach for Improving the Product Review Pipeline

Mark Cusick
North Carolina State University
mbcusick@ncsu.edu

Benjamin Watson
North Carolina State University
bwatson@ncsu.edu

Oleg Veryovka
Amazon, Social Shopping
olegv@amazon.com

Michael Rappa
North Carolina State University
mrappa@ncsu.edu

ABSTRACT

Product reviews are a vital part of the online retail shopping experience, and retailers are constantly trying to improve them. Unfortunately, online retailers do not fully understand what makes reviews helpful, and therefore still seek to improve their methods for soliciting and displaying helpful reviews. Earlier work often relied on customer-provided helpfulness ratings, but they have proven to be at best a partial solution. Text mining is the obvious complement to ratings, but although the field has researched various aspects of review helpfulness, previous work lacks the scalability, product category coverage, and textual feature coverage (e.g., lexical, subjectivity, sentiment, topic) required by retailers. Previous visualization research has provided a variety of unique visuals, but they are often meant for consumption by experts, and none of them are based on features that customers find most helpful. We suggest an approach driven by customer-centric helpful content for gathering and presenting product evaluation data in order to improve the product review pipeline.

Author Keywords

Text mining; semantic analysis; text visualization; product reviews

ACM Classification Keywords

H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing - Linguistic Processing; H.4.m [Information Systems Applications]: Miscellaneous; H.5.2 [Information Interfaces and Presentation]: User Interfaces - Graphical User Interfaces.

INTRODUCTION

Product reviews are a key element of the online retail purchase experience for consumers, and for retailers they have been shown to increase sales (Chatterjee 2001; Chen, Dhanasobhon, and Smith 2008, Li and Hitt 2008). But for new or less popular products, reviews are often absent. Conversely, the most popular or desired products often have hundreds or thousands of reviews, which create an unhelpful overload of content for consumers. Online retailers want to increase the number of helpful reviews for less popular products and provide helpful summarization or guidance to consumers when reviews are too numerous.

Solving these problems will require addressing three important research challenges. *First, we must understand what makes a review helpful*. This is not a simple problem, given that what customers find helpful differs by product category (Mudambi and Schuff 2010). Given such an understanding, we can tackle *the second challenge: helping customers to contribute helpful product reviews*. This will involve a detailed analysis of the helpfulness of existing reviews, real time analytics to understand how the topics customers are currently addressing might improve helpfulness, as well as informational and persuasive appeals to guide them toward topics that would be most helpful. *The third challenge, is clearly displaying the meaning of product reviews to shoppers*. Since meaning in this case is essentially what is helpful, understanding helpfulness is again a crucial component of this goal. This might involve the techniques of both visualization and graphic design. Below, we describe each of these challenges in more detail.

CHALLENGE 1: UNDERSTANDING HELPFULNESS

At many retailers, customers can vote for a review as helpful or unhelpful. With enough votes, we can create a helpfulness ratio. Unfortunately this metric is not reliable. Accumulating a sufficient number of votes to stabilize the helpfulness ratio takes several months of displaying a review online. Even when the ratio stabilizes, it can be quite noisy, with customers voting for different and conflicting reasons. In addition, the majority of reviews will never receive any votes. Due to these issues, recent research no longer uses helpfulness voting as ground truth (Liu et al. 2013). There is still value in the helpfulness voting, but it should be paired with manual classification and automated approaches.

Given the limited utility of voting, text analysis is a natural alternative for measuring helpfulness. Unfortunately, text analysis of reviews is more challenging than analysis of other text in many ways. While product reviews have the advantage of context, they are completely "free text" with almost no structure imposed by online stores. As long as a customer’s review adheres to a few simple guidelines (e.g. family friendly language), most retailers will post it online. Thus reviews differ greatly in length and style. Length ranges from a few words to several pages; while style can vary from twitter abbreviation and brevity, to colloquial speech, to well written essays.

Not surprisingly then, previous research in text analysis offers only partial solutions. Ghose and Iperiotis (2011) used subjectivity and readability analysis as well as the reviewer's history to predict review helpfulness and expected sales effect. Their techniques predicted helpfulness with 77% to 89% accuracy for a small set of product categories, with a few thousand reviews. Zhang (2008) used different combinations of lexical and semantic analysis on datasets and product categories of similar scale. Cao, Duan, and Gan (2011) investigated the predictive power of basic review features (e.g., author and star rating), word and character counts, and Latent Semantic Analysis (LSA). Their work focused on one particular product category, and their work is further limited by the use of LSA, which does not scale well to large data and often produces topic clusters that are difficult to understand. There has been work using semantic analysis to determine the positivity or negativity in a review, but this work does not attempt to predict helpfulness (Turney 2002; Pang, Lee, and Vaithyananthan 2002). Another aspect of this research is using semantic analysis to determine a five star rating for a review, which again lacks association to helpfulness (Pang and Lee 2005; Goldberg and Zhu 2006; Zhang and Varadarajan 2006). Research has been done to summarize reviews, but focuses on sentiment analysis rather than predicting helpfulness (Dave, Lawrence, and Pennock 2003; Hu and Liu 2004; Chaovalit and Zhou 2005).

While useful, the existing research does not provide a complete solution for online retailers. We see three sub-challenges that require further research:

- *Handling review scale and product breadth:* Methods available today cannot process the breadth of product categories addressed by retailer reviews, nor handle the rate at which those reviews are submitted. Major online retailers receive millions of reviews per month. Maintaining an accurate model of helpfulness requires a much greater focus on scalability and speed, especially in modern distributed service-oriented architectures.
- *Accurately modeling helpfulness experience:* Which reviews do customers believe are helpful, and why? If our algorithms do not accurately model this experience, they will be be of little use. Only recently has the field of text analysis begun validating its algorithms with comparisons to human judgments; in the retail context, such validation may prove crucial. Traditional psychological methods such as card sorting are one way to generate a psychological testbed for comparison; when combined with modern online tools such as Mechanical Turk and live online surveys, they may prove particularly powerful.
- *Enabling reviewer guidance:* to help customers produce helpful reviews more quickly, they will need automated guidance based on a comparison of the text they've written so far compared to existing reviews. Obviously such guidance cannot be based on indirect measures of task helpfulness such as stars and reviewer history. Instead, they should be offered a concrete direction along lines of "describe your impression of the product's interface."

CHALLENGE 2: IMPROVING PRODUCT REVIEWS

Online retailers want to assist their customers in writing the most helpful reviews possible, so that a product's reviews become useful as quickly as possible. Most customers struggle writing reviews longer than 20 words. Online retailers believe this is because customers are unsure what to write or only write briefly about one topic or experience. We see two sub-challenges that require further research:

- *Real time review analysis:* Retailers have found that if customers are given simple guidance, then reviews will increase in length and include content reflective of the guidance suggestions. We envision a system that provides such guidance in real time: it monitors the state of a customer's partial review, and suggests improvements to the partial review so that it better complements existing reviews. We are not aware of any text or helpfulness analyses methods currently capable of such speed and functionality.
- *Applying the techniques of behavioral persuasion:* Research has consistently shown that informational solicitations alone are weak motivators of behavior change, such as writing longer or more carefully composed reviews. Successful solicitations are well timed, easily performed and made to motivated individuals. In this context, real time review analysis should allow well timed solicitations, while accurate helpfulness modeling should permit the submission of short but still useful reviews. Persuasive visualizations of topic coverage will identify topic gaps and help motivate customers to contribute specific topics. Online retailers are in a particularly good position to examine the effectiveness of these techniques, since they both send the solicitations and accept any resulting reviews.

CHALLENGE 3: COMMUNICATING PRODUCT REVIEW MEANING

Since the most popular products on retail websites have hundreds or thousands of reviews, online retailers require an effective way to visualize helpful content. While companies like Amazon have added visualizations such as text summarization quotes and aggregate star rating histograms, these visualizations are not directly associated with helpful content. Some research has created visuals based on key terms (Alper et al. 2011; Hu et al. 2013). Other research uses semantic analysis to create visuals, sometimes including key terms or feature selection (Gregory et al. 2006; Liu, Hu, and Chang 2005; Chen et al. 2006; Gamon et al. 2005; Oelke et al. 2009; Wu et al. 2010; Yatani et al. 2011). We see four sub-challenges that require further research:

- *Summarization of large text collections:* Reviews for a single product can comprise a fairly large collection of text. While existing text summarization and visualization techniques can accurately convey the gist of such collections to those familiar with the topics being visualized, we are not aware of any techniques that successfully communicate in depth meaning, or even surface meaning to viewers unfamiliar with the visualized topics.
- *Summarization for mobile:* Traffic to e-commerce sites from tablets and smartphones has nearly doubled in the past year reaching an average share of 25%. In Q1 2013, tablets exceeded traditional desktop devices for conversion rates for the first time suggesting that people are increasingly comfortable with the experience of buying via tablets[1]. Presenting product reviews on mobile devices in a compelling manner has many unique challenges including: limited display surface, tap based interaction, and varied connectivity bandwidth. Previous research has not focused on mobile visualizations, but as the mobile share continues to increase, online retailers require elegant solutions.
- *Visualization for consumers and retailers:* There is a divide between visualizations useful to consumers and those useful to retailers. Many visualization techniques are designed for use by subject matter experts; however, some recent work has focused on consumers (Yatani et al. 2011; Hu et al. 2013). Retail customers are not generally experts and certainly not experts across all product categories. One challenge then is to make visualizations of reviews understandable by most customers, while also conveying helpful amounts of information. Graphic designers have much more experience with this sort of visual communication and may be excellent collaborators in this effort. As many designers would say, the particular constraints of this problem may offer opportunity: we might specialize the visualization to those aspects of the reviews that are most helpful.
- *Persuasive visualization:* Retailers are visualizing reviews not only to encourage buying, but also to inspire the submission of additional reviews. There is a growing interest in the field of visualization in visualization rhetoric, the use of data visuals to influence opinion and change behavior. As we noted above, online reviews are a particularly good venue for studying the effectiveness of such appeals.

CONCLUSION

Improving online product reviews is a surprisingly complex problem, involving text analysis and summarization, marketing, as well as persuasion and visualization. We suggest an approach that involves understanding helpful review content, guiding users to write more of this content, and providing visuals built on this content. If this can be completed at scale and across a complete range of product categories, user experience and online retail sales both stand to benefit. But to engineer this approach, research needs to continue in text analytics, text visualization and their relationship to marketing, persuasion and user experience. This research needs to consider the overall customer experience from understanding what is important to shoppers, to helping product owners contribute helpful product information, to displaying this information on the website.

[1] http://www.smartinsights.com/mobile-marketing/mobile-marketing-analytics/mobile-marketing-statistics/

ACKNOWLEDGMENTS

We would like to thank the Amazon Community teams for their help and support while working on this document.

REFERENCES

1. Alper, B., Yang, H., Haber, E., & Kandogan, E. 2011. OpinionBlocks: Visualizing Consumer Reviews. In *Proceedings of the IEEE VisWeek Workshop on Interactive Text Analytics for Decision Making*.
2. Cao, Q., Duan, W., and Gan, Q. 2011. Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decis. Support Syst.* 50, 2 (January 2011), 511-521.
3. Chaovalit, P., and Zhou, L. 2005. Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches. In *Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 4 - Volume 04* (HICSS '05), Vol. 4. IEEE Computer Society, Washington, DC, USA, 112.3-.
4. Chatterjee, P. 2001. Online reviews: do consumers use them?. *Advances in consumer research 28* 129-133.
5. Chen, C., Ibekwe-SanJuan, F., SanJuan, E. and Weaver, C. 2006. Visual analysis of conflicting opinions. In *IEEE Symposium On Visual Analytics Science And Technology* 35 – 42.
6. Chen, P., Dhanasobhon, S., and Smith, M. 2006. All reviews are not created equal: The disaggregate impact of reviews and reviewers at amazon.com *Com* (May 2008).
7. Dave, K., Lawrence, S., and Pennock, D. 2003. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web* (WWW '03). ACM, New York, NY, USA, 519-528.
8. Gamon, M., Aue, A., Corston-Oliver, S. and Ringger, E. 2005. Pulse: Mining customer opinions from free text. In *International Symposium on Intelligent Data Analysis* 121–132.

9. Ghose, A., and Iperiotis, P. 2011. Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics. *IEEE Trans. on Knowl. and Data Eng*. 23, 10 (October 2011), 1498-1512.
10. Goldberg, A., Zhu, X. 2006. Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization. In *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing* (TextGraphs-1). Association for Computational Linguistics, Stroudsburg, PA, USA, 45-52.
11. Gregory, M. L., Chinchor, N., Whitney, P., Carter, R., Hetzler, E., and Turner, A. 2006. User-directed sentiment analysis: Visualizing the affective content of documents. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text* 23-30. Association for Computational Linguistics.
12. Hu, M., and Liu, B. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (KDD '04). ACM, New York, NY, USA, 168-177.
13. Hu, M., Yang, H., Zhou, M. X., Gou, L., Li, Y., and Haber, E. 2013. OpinionBlocks: A Crowd-Powered, Self-improving Interactive Visual Analytic System for Understanding Opinion Text. In *Human-Computer Interaction–INTERACT* 116-134.
14. Li, X., and Hitt, L. 2008. Self-selection and information role of online product reviews. *Information Systems Research* 19.4: 456-474.
15. Liu, B., Hu, M. and Cheng, J. 2005. Opinion observer: analyzing and comparing opinions on the web. In *International Conference on World Wide Web* 342–351.
16. Liu, Y., Jin, J., Ji, P., Harding, J.A., Fung, R.Y.K. 2013. Identifying helpful online reviews: A product designer's perspective. *Comput. Aided Des.* 45, 2 (February 2013), 180-194.
17. Mudambi, S. and Schuff, D. 2010. What makes a helpful online review? a study of customer reviews on amazon.com. *MIS Q.* 34, 1 (March 2010), 185-200.
18. Oelke, D., Hao, M., Rohrdantz, C., Keim, D. A., Dayal, U., Haug, L. E., and Janetzko, H. 2009. Visual opinion analysis of customer feedback data. In *IEEE Symposium On Visual Analytics Science And Technology* 187–194.
19. Pang, B., Lee, L., Vaithyananthan, S. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10* (EMNLP '02), Vol. 10. Association for Computational Linguistics, Stroudsburg, PA, USA, 79-86.
20. Pang, B., Lee, L. 2005. Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* (ACL '05). Association for Computational Linguistics, Stroudsburg, PA, USA, 115-124.
21. Turney, P. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (ACL '02). Association for Computational Linguistics, Stroudsburg, PA, USA, 417-424.
22. Wu, Y., Wei, F., Liu, S., Au, N., Cui, W., Zhou, H. and Qu, H. 2010. OpinionSeer: Interactive Visualization of Hotel Customer Feedback. In *IEEE Transactions on Visualization and Computer Graphics* 16,6, 1109–1118.
23. Yatani, K., Novati, M., Trusty, A., and Truong, K. N. 2011. Review spotlight: a user interface for summarizing user-generated reviews using adjective-noun word pairs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* ACM, 1541-1550.
24. Zhang, Z., and Varadarajan, B. 2006. Utility scoring of product reviews. In *Proceedings of the 15th ACM international conference on Information and knowledge management* (CIKM '06). ACM, New York, NY, USA, 51-57.
25. Zhang, Z. 2008. Weighing Stars: Aggregating Online Product Reviews for Intelligent E-commerce Applications. *IEEE Intelligent Systems* 23, 5 (September 2008), 42-49.