CS6742: NLP and Social Interaction, Spring 2011
1/25/2011: Assignment A1 (first reading and diagnostic assignment)

Introduction Learning to read technical papers effectively is a important research skill. One should be able to identify what is important or interesting about a paper, and use it as inspiration for future work.

We selected Eric Gilbert and Karrie Karahalios' "Understanding Deja Reviews", *Conference on Computer Supported Cooperative Work, CSCW 2010*, as your first reading because it (a) is short and does not require a great deal of technical background to get the gist of, (b) appeared in a good conference (acceptance rate 20%[1]), thus indicating that some portion of some community thought the work was worthwhile, (c) fits in well with the course theme and the first two lectures, (d) serves as a diagnostic (we can judge your current ability to understand technical writing, and if you feel that reading research papers is not at all to your liking, this may not be the right class for you), and (e) may serve as an inspiration for project ideas.

That being said, we do not necessarily endorse any or all of the views or findings found in the paper.

Instructions

Academic integrity policy for this assignment: you can discuss the paper with other students, but each of you must independently write up your answers to the questions. Group submissions are not allowed.

Remarks on general academic integrity: In no case, in this class or otherwise, should you copy phrases or larger chunks of text or paragraphs phrases or larger chunks of text from any source (online, from a text book, from another person, etc.) without citing the source from which you got the original. *To do otherwise is to commit fraud.*

By **3pm Wednesday Jan 26th**, make sure you can access the course CMS page, http://cms.csuglab.cornell.edu .[2]

Attached is the reading. You can also get an online version from the course homepage, http://www.cs.cornell.edu/courses/2011sp, which contains the text of the comments attached to the highlighted selections. For the purposes of this assignment, which is to be completed within a short time frame (**due date: 6am Thursday Jan 27th on the course CMS**), spend about an hour skimming the paper: read every word, but don't dwell on things you don't understand. Do expend some effort trying to make out what you can about the "Methods" section (you may not make much progress right now).

Then, answer the following questions on the CMS, *in your own words* (that is, do not just copy from or paraphrase from the paper (or any other source or person, for that matter); remember that one of the purposes of this assignment is diagnostic). You should type (or cut-and-paste) plain text into the answer fields in the CMS assignment.

- Question 1 (1-3 sentences suffices): What was the overall goal to which language-processing techniques were applied in this paper?
- Question 2: (1-3 sentences suffices): What improvements or alternate language-processing techniques would you propose to accomplish the same goal, assuming computational resources (and page allotments) were unlimited?

---

[1] Statistic from http://portal.acm.org/citation.cfm?id=1718918&CFID=6727150&CFTOKEN=40694591

[2] If you do not see the course "CS6742" by then when you login with your netid and password, please immediately email llee@cs.cornell.edu *and* cristian@cs.cornell.edu, subject line "CS6742 CMS registration request", giving your Cornell netID in the body (NOT your Cornell ID number, or your preferred email address, but your official Cornell NetID, which should be your initials followed by a number, e.g., LJL2).

Example response (which you cannot use in your answer, of course!): "Compensate for coverage effects. Reviews that will be closest to the cluster centroid would be ones that mention as many of the 'aspects' that are in the centroid as possible, which will thus probably be longer reviews. But, finding short reviews that are simply an echo of one of the (common) aspects of the centroid is important because [... something about the goal of the NLP techniques]."

- Question 3: What prior coursework or experience have you had in artificial intelligence, human-computer communication, machine learning, linguistics, natural language processing, psychology, or sociology?
- Question 4: Are you a PhD student, MEng, or undergraduate?
- Question 5: What are your main research interests?

Model questions for class discussions later in the course You do not need to answer these questions for the assignment. We present them here to exemplify the questions that arose in our minds as we read the paper.

1. (A question one should always start with.) What are the most important and/or interesting findings of the paper? Is there any interesting question left unanswered?
2. How accurate was the redundancy ranking? (How should one measure this? Does accuracy matter for the purposes of this paper?)
3. Is the (implicit) distinction between "saying the same thing" (holding the same opinion) vs. "saying the exact same thing" (copying) worth making in other contexts (e.g., Twitter, letters to congresspeople)?
4. Can "amateurs" and "pros" be distinguished by their language alone? Would the distinction be useful in other contexts?
5. To what extent does the "helpfulness vote" mechanism that Amazon provides ("was this review helpful to you?") already aid the goals of topic diversity and social navigation? (We recognize that helpfulness voting is already proposed as part of the suggested new social navigation in the paper.)
6. Is the lack of added information in the deja reviews reflected by their "helpfulness votes"? (i.e., are redundant reviews deemed less helpful than original reviews?)
7. What is an appropriate sample size of interviewees, keeping in mind the time it takes to subsequently "code" (analyze) the interview transcripts? Would it have been possible to structure the interviews so that pros would not have felt "maligned"?
8. One question brought up in the paper is if reviewers do in general read other reviews. We can further ask, are those reviewers that are aware of the other reviews (even if merely through the product average star rating) influenced by them? (see Wu and Huberman's "Opinion Formation Under Costly Expression" referenced below)

(OVER)

Companion readings

- David and Pinch (2006) also discuss in depth the phenomenon of Amazon reviews often being very similar to others.

  Shay David and Trevor J Pinch. ”Six Degrees of Reputation: The Use and Abuse of Online Review and Recommendation Systems.” *First Monday* 2006, Special Issue on Commercial Applications of the Internet. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1590/1505.

- Wu and Huberman (2010) explore the relation between early and late opinions expressed on Amazon and IMDB.

  Fang Wu and Bernardo A. Huberman. ”Opinion Formation Under Costly Expression.” ACM Transactions on Intelligent Systems and Technology, October 2010. http://www.hpl.hp.com/research/scl/papers/howopinions/wine.pdf

Understanding Deja Reviewers

Eric Gilbert and Karrie Karahalios
University of Illinois at Urbana-Champaign
[egilber2, kkarahal]@cs.uiuc.edu

ABSTRACT
People who review products on the web invest considerable time and energy in what they write. So why would someone write a review that restates earlier reviews? Our work looks to answer this question. In this paper, we present a mixed-method study of *deja reviewers*, latecomers who echo what other people said. We analyze nearly 100,000 Amazon.com reviews for signs of repetition and find that roughly 10–15% of reviews substantially resemble previous ones. Using these algorithmically-identified reviews as centerpieces for discussion, we interviewed reviewers to understand their motives. An overwhelming number of reviews partially explains deja reviews, but deeper factors revolving around an individual's status in the community are also at work. The paper concludes by introducing a new idea inspired by our findings: a self-aware community that nudges members toward community-wide goals.

Author Keywords
Online communities, ecommerce, product reviews, text

ACM Classification Keywords
H5.3. Group and Organization Interfaces; Asynchronous interaction; Web-based interaction.

General Terms
Human Factors

INTRODUCTION

At the time of this writing, 1,434 Amazon.com users have written a review of the Wii Fit. Earlier this year, the 1,000th review came in. The reviewer enthusiastically praised the balance games; so had 130 reviewers before him. Because you can really work up a sweat (19 earlier reviews), he prefers it to the drudgery of the gym (10 earlier reviews). Although he bought the Wii Fit as a Christmas gift for the whole family, his son enjoys it most. Seven reviewers before him had related essentially that same story.

This paper is about reviews and reviewers like this one. (Hereafter, we refer to them as *deja* reviews and reviewers.) What's in it for these reviewers? Why spend the time and energy to say what so many others have said? It's not outright copying [17]. Most deja reviewers invest considerable energy in their reviews, yet happen to converge almost entirely on what people have said before them. It is this very fact that enables recent systems [e.g., 10, 16] to mine and summarize discussions about products on the web. Without the sociotechnical phenomenon of convergence among reviewers, mining and summarization engines simply would not work. We looked at these systems and a question sprang to mind: Why do so many reviews look alike?

In this paper, we try to answer this question. We think it's a fascinating sociotechnical case study in its own right. Our results indicate that roughly 10–15% of all reviews substantially resemble earlier ones. More importantly, after a certain point, deja reviews are missed opportunities. Did we need a 130th review professing love for the Wii Fit's balance games? Or, could we have gained more from hearing about how the Wii Fit holds up over time? Of course, the reviewer gained something from the act of writing the review, but the community gained little it did not already know. As a potential buyer, it helps to read many people reporting on a single product feature, but when do we have enough? 130? 200? 500? More broadly, the question for successful communities is no longer "How do we get people to contribute?" but "How do we best use all these contributions?" We see deja reviews as an intriguing way to think about this question.

Relative to other highly successful online communities like Wikipedia [e.g., 4, 19], reviewing communities have received little attention. Machine learning and data mining researchers [9, 13] gravitated toward sites like Amazon for their linguistic data sets, but we are aware of no work focusing on reviewer motivations in these large, successful communities. In the universe of online communities, reviewing sites are relatively unique. They are not primarily social spaces. Unlike Wikipedia, reviewing communities do not pursue a grand agenda. And yet, Amazon and others have amassed tens of millions of reviews, reviews that significantly influence demand for many products [1].

This paper presents the results of a mixed-method study exploring the motivations of deja reviewers, latecomers who echo what earlier reviewers said. Amazon is our site for this study. As is perhaps to be expected, we find that the sheer number of reviews makes it hard for reviewers to fully grasp what everyone else has said. This contributes to the number of deja reviews. However, beyond this, we describe a remarkable trend in our data: *amateur* and *pro* reviewers think about deja reviews very differently, and we believe this bears on design. In the tradition of [4] and extending the work of [3, 8, 11], we conclude the paper by introducing an idea inspired by our findings: a self-aware community that knows what it wants.

CSCW 2010, February 6–10, 2010, Savannah, Georgia, USA.

METHODS

We employed a mixed-method design to study deja reviews. First, we applied a purely computational technique to identify deja reviews; next, we conducted interviews to understand the authors and the motivations behind them. In this regard, we like to think of our approach as a microscopic complement to the more macroscopic work of [21]. The list of every Amazon review ever written is quite long; to direct our search, we focused on bestselling products, which disproportionately attract reviews. We downloaded every review from 200 bestselling products belonging to 15 Amazon product categories. The corpus comprises 98,191 product reviews.

Text is complex, messy and highly multidimensional. Any computational technique to identify deja reviews necessarily results in an approximation. Our goal was not necessarily to develop the best algorithm for finding deja reviews, but to assemble a compelling list for subsequent interviews. Consequently, we applied the concept of centrality [15] to our corpus. It is a well-established method, and probably a lower bound for identifying deja reviews (i.e., it does not know about synonyms). The method looks for reviews closest to the core, or *centroid*, of what earlier reviews said. Closeness in this context is defined by cosine similarity [6], and the centroid is defined by statistically representative unigrams, bigrams and trigrams normalized to Google's 5-gram corpus [2].

As an example, the centroid for the iPod nano contains (among others) the unigrams *volume, Zune, shuffle*, the bigrams *my purse, first iPod, a pain,* and the trigrams *easy to use, a bigger screen* and *plenty of space.* Yet, this only measures the surface; certainly deeper aspects are at play. A reviewer often spends a good part of their review judging a product (e.g., "I *hate* the earbuds!"). To account for this, we relied on an ensemble of linguistic resources [5, 13, 14, 20] to track the positive and negative words appearing in the reviews. (We chose sentiment over star-ratings to capture nuances and to confine our analysis to review text.) In this work, a review is close to the core if its sentiment is close to the running average sentiment of earlier reviews.

Using these techniques, we were able to sort the review list by how much a review echoed earlier reviews. We scanned the highest 5% of this list to find reviewers that publicly disclosed an email address or website. (The top 5% represented reviews that nearly matched many earlier ones, but were not copies.) Of the 200 reviewers we contacted in this way, we conducted semi-structured interviews with 20. The interviews (14 email, 3 IM, 3 phone) took roughly 30 minutes. Interviews revolved around the particular deja review, participants' motivations for writing it and their relationship with other reviews. The email interviews consisted of questions similar to the IM and phone interviews, but we engaged in follow-up conversations to clarify certain points. To ground the conversations, we also provided a link to the deja review. Finally, the interviews were coded using a grounded-theory influenced approach where themes arise iteratively [7].

RESULTS

Across all participants, the sheer number of reviews made reading every single one impractical. The following was a common refrain:

> If there's 5, 10, I'll read those. If there's 50 or 60, I'll skim some titles. If a title looks like it has some wit or humor, I might read that one. That's about it. (P13)

Beyond the universal constraint of scale, our interview data clearly clustered into two groups: groups we will call *amateurs* and *pros*. We did not anticipate this; it emerged from our data. Amateurs (9 of 20) reviewed only occasionally, generally had reviewed fewer than 30 products and scarcely received helpful votes. (Amazon users can vote a review up or down by giving it "helpful votes.") Pros (11 of 20) reviewed many hundreds of products, had followings on Amazon and often received helpful votes. Motivations for reviewing, and reactions to seeing deja reviews, varied remarkably between these two groups. We structure our results in terms of this division.

Amateurs: what's the point of a unique review?

Amateur reviewers often write a review because of an almost visceral reaction to a product. Reading other reviews holds little appeal because it does not satisfy the need to share this powerful reaction. Routinely using words like "ecstatic," we might characterize their deja reviews as spontaneous, heartfelt appeals.

> I used this documentary in class when I taught the novel *Call of the Wild*. The review was written because I was so moved by this documentary, and as a wolf advocate, I wanted a few others to know about it … I don't care if I write the first review or the 5,000th. It's cathartic for me. (P3)

> I think people like to write the 500th review just because sometimes it's nice to get a chance to be heard or "read" and share … it helps people to vent in a way. (P16)

> It is sort of a soap box time for me … For me to write a review, I have to be either ecstatic about it, or really bummed after getting it. (P20)

Interestingly, other researchers have also reported what P20 describes in quantitative terms. Amazon's bimodal ratings distribution has two peaks: one for the people who love a product and one for those who hate it [18]. A main goal of this study is to better understand relationships between reviewers. Do reviewers read other reviews? How do they feel about them? When we asked our participants how they felt now knowing about their deja review, amateurs had little problem seeing themselves as part of a cluster of similar reviewers, part of the crowd.

> I felt very strongly that everyone should see/own this collection. Suppose felt compelled to add my two cents without any consideration of whether it would count for anything. (P4)

> I didn't know another review or other reviews were similar. Honestly, that doesn't bother me one bit. The more people who say they liked the product, the more I believe it. (P17)

> I liked the [gadget] because both my kids had them and they could play games together. Very innovative product. I'm sure that other reviewers noted the same features that appealed to me and my kids. A completely unique review wouldn't serve any real purpose, would it? (P5)

This is an important point. Being part of a crowd that converges on a few themes seemed to hold little stigma. We do not want to imply that being part of the crowd is inherently bad; we frame it this way to showcase how differently amateurs and pros think. To conclude, we propose a way to understand amateur reviewers: they pop into the community when a product grabs them (for better or worse) and do not mind identifying as part of a group.

Pros: building a brand above the din

Pro reviewers felt very differently. In fact, two pros thought we had maligned them: one accused us of trying to "stifle dissent" while another said that we implied she writes "just to see my own words on the internet." (We should note that we did not use the term *deja review* in the interviews; our language resembled "your review seems to look a lot like other reviews.") Many more pro reviewers, however, felt that we had only identified the worst review they had ever written.

> I do not think buyers would find that review [deja review] useful … I would not buy that camera based on my review … it was useless writing that would not influence a decision one way or the other. (P7)

> This one, [deja review] and some others I'd written, will surely get lost in the shuffle. I'd just seen the film before I wrote it, and it was more spontaneous than seriously thought-out. (P14)

P14 even followed up after the interview to tell us to go read his best reviews.

> [Deja review] was a review I wrote on a whim, without a lot of thought, I'd be happier if you read my reviews of [product], [product] or [product]. I thought those were among my best. (P14)

Where amateurs may avoid reading other reviews because their goal is self-expression, pro reviewers seem to avoid other reviews because they have elevated status in the community. We characterize this as "above the din."

> OK, so now I avoid other people's reviews. I'm really not interested in their reviews; I really don't want to be influenced by what they say. I sorta wanta put my opinion out there. The only time I might look at reviews, I tend to look at editorial reviews. (P12)

> [Do you read other reviews?] Absolutely not … I do not condone plagiarism and would never think to base my opinion on another reviewer's thoughts. I believe that this alone has led me to be rated highly in the review system. (P7)

> I actually seem to have almost like a fan base … So I figure there might be 500 or 600 reviews up on there already, but most of them look like they don't have anything under their name [a reputation] or I check on some of them and that might be their only review, or they're not rated at all. (P13)

This disregard for other reviews seems to stem ultimately from pro reviewers' goal to build a brand or identity within the reviewing community (also see P13 above).

> I think it is important to realize that there are two sides to this—the product being reviewed, and the person reviewing it. In my case, as in many others, I am trying to build a name for myself as an author & someone who gives quality reviews … My view is that if every other person on earth has said something similar that is fine, it in no way diminishes the value of my personal opinion. (P19)

> I am an author ([her book title]) and I can put a link back to my book with every review I write. I mostly write reviews for books that are either similar to mine in some way—and the similarity could be remote but explainable by me—or that I enjoyed reading. (P15)

> One of my goals has become to try and reach Amazon's Top 100 Reviewers. Right now, I'm close. You know, that's an ego thing. Something where I want that little badge that says "Top 100 Reviewers." And I think the only way I'm going to become a Top 100 reviewers is to get reviews on there. So, why not write one? (P13)

Reviewers might read a handful of other reviews, but every one of our reviewers ignored the rest. This is where the similarity between pros and amateurs ends. Where amateurs write to scratch an itch, pros write to advance a personal agenda. That personal agenda might mean satisfying a small fan base, climbing the reviewing ladder or promoting one's own product. This is the first striking difference between amateurs and pros. The second is the relationship to other reviewers: amateurs do not mind being a brick in the wall; pros want to stand out.

IMPLICATIONS

Consider the curse of success afflicting sites like Amazon. Whether by innovation or market forces, the site has a very large user base. The new question is "How do we best use all these contributions?" So far, the lone answer has come from the data mining community: take the reviews as a given and summarize them for new buyers [e.g., 10, 16]. This is a useful approach and we expect to see it in mainstream use soon. Yet, it ignores the opportunity that site designers have to shape a community as it develops.

We now introduce a new idea, one example of how to draw design inspiration from our findings: a self-aware community that knows what it wants. Imagine it is mid-to-late 2008. The Wii Fit is remarkably popular, attracting many hundreds of reviews. Roughly fifty people have already intensely praised its balance games, but few have commented on how it looks. The site knows [1] that new buyers generally care *a lot* about how a product looks. Because it knows what it wants, perhaps the site can ask for it: "If you can tell us anything about how the Wii Fit looks, that would help!" When is the right time to do this, and with whom? This might seem like a subtle point, but recall the intensely negative reaction from pros even at the suggestion of repetition. It would be unwise to rush in given the sensitivities around echoing and repetition.

First, we argue that the site should entirely avoid asking pros for particular information. This perhaps seems counter-intuitive, and we certainly would not have guessed it ahead of time. But pro reviewers in communities such as Amazon have personal agendas that transcend the particular product in question. We understand that this is a significant slash through the user base. The power-law curve that governs Amazon's reviewing community means that pros write a sizable proportion of the site's reviews.

Yet, in this case, we strongly advise a focus on amateur reviewers: they are primarily attached to *the product*, not their identity as a reviewer. So, when should a site ask for information about a certain product feature? Here we invoke our characterization of amateur reviews as spontaneous, heartfelt appeals. We suggest an approach that lets an amateur reviewer get some things off their chest first. Perhaps the site should even ask for particular pieces of information after the reviewer lets loose a few emotionally charged words. This allows both the site and the reviewer to achieve their goals.

This idea extends beyond topic diversity. Amazon and other sites realized that potential buyers had trouble navigating the huge number of reviews some products attracted. Amazon's solution is the question "Was this review helpful to you?" Other sites have implemented similar features. (The general technique is called social navigation.) One noted problem with this approach is that once it gets going, it reaches a steady state very fast: the reviews that attract a few votes get more and more, leaving everyone else in the dust. Recent work [12] has suggested introducing randomness into how customers see reviews, so that an unrated review might get out of the abyss. However, a site always wants to present its best content to a potential buyer. We think deja reviewers might be able to fill this gap. Again, we advise focusing on amateur reviewers. But now consider a site that automatically calculates a few highly similar deja reviews and asks the reviewer for a helpfulness vote in the sidebar. The reviewer is uniquely qualified to voice an opinion because they hold similar views. We are excited by this approach because it applies the resources of a motivated reviewer toward a classic social navigation problem: dampening the stampede toward only a few contributions among thousands of candidates.

The general theme behind these ideas is nudging members toward a community-wide goal without trampling on their reasons for contributing in the first place. We have presented two instantiations of it informed by our findings: topic diversification and improving social navigation. Certainly many others exist, and many open problems remain. We see this work as a first step toward the goal of making the most effective use of contributions in large online communities and look forward to deeper explorations along these lines.

REFERENCES

1. Archak, N., Ghose, A., et al. Show me the money!: deriving the pricing power of product features by mining consumer reviews. *Proc. KDD,* 2007. 56–65.
2. Brants, T. and Franz, A. Web 1T 5-gram Version 1. *Linguistic Data Consortium, Philadelphia*.
3. Cosley, D., Frankowski, D., et al. How oversight improves member-maintained communities. *Proc. CHI,* 2005. 11–20.
4. Cosley, D., Frankowski, D., et al. SuggestBot: using intelligent task routing to help people find work in wikipedia. Proc. IUI, 2007. 32–41.
5. Esuli, A. and Sebastiani, F. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. *Proc. LREC,* 2006.
6. Frakes, W. B. and Baeza-Yates, R. *Information Retrieval: Data Structures and Algorithms.* Prentice Hall, 1992.
7. Glaser, B. and Strauss, A. *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Aldine Transaction, 1967.
8. Harper, M. F., Frankowski, D., et al. Talk amongst yourselves: inviting users to participate in online conversations. *Proc. IUI,* 2007. 62–71.
9. Hu, M. and Liu, B. Mining and summarizing customer reviews. *Proc. KDD,* 2004. 168–177.
10. Liu, B., Hu, M., et al. Opinion observer: analyzing and comparing opinions on the Web. *Proc. WWW,* 2005. 342–351.
11. Ludford, P. J., Cosley, D., et al. Think different: increasing online community participation using uniqueness and group dissimilarity. *Proc. CHI,* 2004. 631–638.
12. Otterbacher, J. 'Helpfulness' in online communities: a measure of message quality. *Proc. CHI,* 2009. 955–964.
13. Pang, B., Lee, L., et al. Thumbs up?: sentiment classification using machine learning techniques. *Proc. EMNLP,* 2002. 79–86.
14. Pennebaker, J. W. and Francis, M. E. *Linguistic Inquiry and Word Count.* Lawrence Erlbaum, 1999.
15. Radev, D. R., Jing, H., et al. 2004. Centroid-based summarization of multiple documents. *Inf. Process. Manage., 40*(6), 919–938.
16. Scout Labs. http://scoutlabs.com. Accessed Jun 1, 2009.
17. Shay, D. and Pinch, T. 2006. Six degrees of reputation: The use and abuse of online review and recommendation systems. *First Monday Special Issue #6: Commercial applications of the Internet*.
18. Talwar, A., Jurca, R., et al. Understanding user behavior in online feedback reporting. *Proc. EC,* 2007. 134–142.
19. Thom-Santelli, J., Cosley, D., et al. What's mine is mine: territoriality in collaborative authoring. *Proc. CHI,* 2009. 1481–1484.
20. Thomas, M., Pang, B., et al. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. *Proc. EMNLP,* 2006. 327–335.
21. Wu, F. and Huberman, B. A. 2008. How Public Opinion Forms. *Proc. WINE,* 2008. C. Papadimitriou and S. Zhang, Eds. Lecture Notes In Computer Science, vol. 5385. Springer-Verlag, Berlin, 334–341.