Using Visibility to Control Collective Attention in Crowdsourcing

Kristina Lerman[1] and **Tad Hogg**[2]
1. USC Information Sciences Institute
2. Institute for Molecular Manufacturing

Abstract

Online crowdsourcing provides new opportunities for ordinary people to create original content. This has led to a rapidly growing volume of user-generated content, and consequently a challenge to readily identify high quality items. Due to people's limited attention, the presentation of content strongly affects how people allocate effort to the available content. We evaluate this effect experimentally using Amazon Mechanical Turk and show that it is possible to manipulate attention to accomplish desired goals.

Peer production systems allow ordinary people to contribute original content in the form of photos (e.g., Flickr, Instagram), videos (e.g., YouTube, Vimeo), news and text (e.g., Twitter, blogs), reviews (e.g., Amazon, Yelp), and much more. While the quantity of user-generated content has skyrocketed, its quality varies dramatically: on YouTube, for example, one can find home videos of a four year old's first violin recital, as well as virtuoso performances by accomplished violinists. Content providers have innovative methods to identify high quality content based on crowdsourcing or peer recommendation. Social news aggregator Digg, Flickr and Yelp for example, ask their users to recommend interesting news stories, photos and restaurants respectively, and prominently feature most recommended items.

Content providers face a dual goal: identify quality content while keeping users engaged by showing them quality content. These goals often conflict. To understand why, consider a simple content evaluation strategy in which the provider shows users a random selection of items and asks them to evaluate them, for example, by recommending interesting items. After enough people examine an item, its quality, or how interesting it is to people, should be reflected in the number of recommendations it receives (its popularity) (Salganik, Dodds, and Watts 2006). However, if there are only a few quality items, the selections seen by most people may not contain a single interesting item. When users continue to see low quality content, they may not return to the site. The provider could instead present the highest quality items that it knows about, but if users do not inspect all items, some high quality items may be missed.

Our ongoing work investigates these questions. We demonstrate that leveraging human cognitive biases can direct system behavior toward specific goals: in other words, to program human computation. Our approach exploits the fact that people pay more attention to items at the top of a list than those below them (Payne 1951; Buscher, Cutrell, and Morris 2009), what's known in psychology as position bias. As a result, the presentation order affects the choices people make about what document in a list of search results they click (Craswell et al. 2008), or the multiple choice answer they select in response to a question (Blunch 1984). Position bias affects how easily an item is found: what we refer to as its *visibility*. A high visibility item, for example, one at the top of a web page or a list, is easily seen, and therefore, receives most attention. The more effort it takes to find an item, for example, by scrolling down a list, the less attention it will receive. Therefore, one can change the visibility of items, or how much attention they will receive, simply by changing their presentation order.

By manipulating the visibility, we can direct users' attention as necessary: to high quality items, or items that have not yet been rated, or items whose quality estimates have low confidence values, or according to some other task requirement. We are conducting experiments on Amazon Mechanical Turk to evaluate various strategies for manipulating visibility with the goals of efficiently rating items and correctly ranking them according to their quality.

Experimental Design

We experimentally evaluated the effects of visibility on item recommendation using Amazon Mechanical Turk (AMT). Experiments provide a controlled setting to isolate the effects of visibility from the items' appeal to users.

In our experiments, we presented users with a list of one hundred science stories drawn from the Science section of the New York Times and science-related press releases from major universities (sciencenewsdaily.com). We asked each user to recommend articles they found interesting. Each article contains a title, short description, and a link to the full story. We recorded all user actions, including recommendations and URL clicks.

We tested four methods of ordering stories shown to users, which we refer to as "visibility interfaces". We assigned each user to one of these interfaces. The *random* interface

presented the stories in a random order, with a new ordering generated for each user. The *popularity* interface ordered stories by their popularity, i.e., in the decreasing order of the number of recommendations. The *activity* interface ordered stories in the chronological order of the latest recommendation they received, with the most recently recommended story at the top of the list. Finally, the *fixed* interface showed all stories in the same order to every user. These interfaces are common in online crowdsourcing and social media. For example, news stories on Digg's front page were by default ordered by the time of promotion. This ordering corresponds to the fixed layout, since every Digg visitor saw the stories in the same order. However, users could also sort stories by popularity, i.e., the number of votes they received. A Twitter stream, on the other hand, is similar to an activity interface, because each retweeted (i.e., recommended) item appears at the top of a follower's stream. The random interface serves as a control by averaging over the effect of ordering, so the number of recommendations an item receives is related to its quality (Salganik, Dodds, and Watts 2006).

Preliminary Results

(a) estimated visibility vs position in list

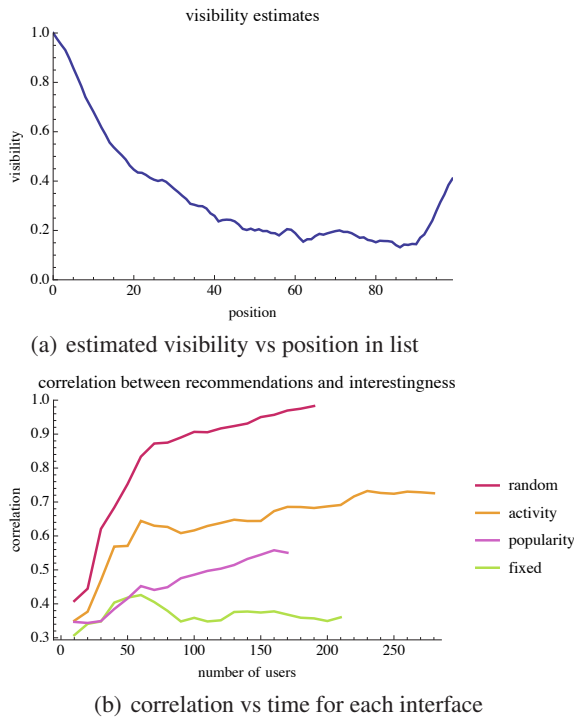(b) correlation vs time for each interface

Figure 1: Experimental results.

We assume that each story has an *interestingness* value, which characterizes its quality or appeal to users. Given that a user sees the story, interestingness gives the probability a user will recommend it. We apply a previously developed stochastic model of user behavior to jointly estimate, via maximum likelihood, story interestingness and position visibility based on user behavior in the random interface. In this model, the visibility estimates are a proxy for user attention,

which our experiments can not measure directly. Figure 1(a) shows that users pay five times more attention to stories in the top position than those in the middle of the list. Positions near the end of the list have greater visibility than those in the middle. This is likely due to "contrarian" users inspect the list from the end (Salganik and Watts 2009).

Controlling the visibility allows directing users' attention to stories of differing quality. This, in turn, has a dramatic effect on the evolution of their popularity. Figure 1(b) shows the correlation between the number of recommendations stories received and their interestingness, as a function of the number of recommending users. The Twitter-like activity interface is better at uncovering story quality (higher correlation with interestingness) than the Digg-like fixed interface. These results suggest that the common practice of displaying most popular items first is inferior to the activity interface that highlights recently recommended items.

Implications

Our experiments show that exploiting human cognitive biases can direct attention in a prescribed way, such as, to identify high quality items. This is achieved by manipulating the visibility of items merely by changing their position in a list shown to users. These results suggest new visibility interfaces could direct user attention according to other criteria, e.g., to items with uncertain quality estimates, thereby rapidly improving estimates for all items as a whole.

Our work suggests that visibility plays an important role in the popularity of items. Moreover, knowing visibility, it is possible to estimate the true quality of items by observing the evolution of their popularity. In addition, having estimates of visibility and quality allows predicting popularity growth of items, i.e., whether they will become blockbusters, or fail.

Our study has implications to the marketplace. Companies can use techniques describe in this paper to estimate quality of the products they offer, such as books or movies, in order to correctly price them or to more intelligently allocate marketing resources, including the order in which ads are shown to users.

References

Blunch, N. J. 1984. Position bias in multiple-choice questions. *Journal of Marketing Research* 21(2):pp. 216–220.

Buscher, G.; Cutrell, E.; and Morris, M. R. 2009. What do you see when you're surfing?: using eye tracking to predict salient regions of web pages. In *Proc. the 27th Int. Conf. on Human factors in computing systems*, 21–30.

Craswell, N.; Zoeter, O.; Taylor, M.; and Ramsey, B. 2008. An experimental comparison of click position-bias models. In *WSDM '08*, 87–94.

Payne, S. L. 1951. *The Art of Asking Questions*. Princeton University Press.

Salganik, M. J., and Watts, D. J. 2009. Web-Based experiments for the study of collective social dynamics in cultural markets. *Topics in Cognitive Science* 1(3):439–468.

Salganik, M. J.; Dodds, P. S.; and Watts, D. J. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856.