Social Values at the Interface: Toward "Just" Human-Computer Ranking Designs at Scale

Leslie Wu
Stanford University
353 Serra Mall
Stanford, CA 94305, USA
lwu2@cs.stanford.edu

ABSTRACT

In the design of human-computer ranking systems for the adaptive display of information, designers often define a domain-specific scoring function which maps items such as people or information search results to numeric scores. Classic ranking systems typically display these items in a linear fashion, sorted by score. There are shortcomings to this approach: such ranking systems do not provide a diversity of results, and in aggregate the distribution of collective user attention is biased by the users' trust in the quality of these orderings. Furthermore, ranking systems based on sorted orders embody a property of chaotic systems, namely that small perturbations in the input—the underlying scoring functions—may have large effects in the output—the distribution of collective user attention. Thus, we propose an alternative human-computer ranking system called donkey sort, which strikes a balance between complete order and uniform randomness, performing probability sampling of the display permutations found in a Latin square design.

Keywords

sorting, ranking, values

1. INTRODUCTION

Human-computer ranking systems play an important role in directing the collective attention of a large number of Internet users through general Web portals such as MSN or Google, social networks such as Facebook and MySpace.com, or through domain-specific searches for news, commercial goods, and entertainment. In spite of this, there has been little work that explores the possible *social* impact of these data-driven user interfaces and their underlying algorithmic ranking designs. In contrast to design analyses that consider a single, perhaps normatively "average" user, interacting with a single user interface, we instead consider a large population of users, and suggest a theoretical model for how the choice of ranking designs may affect a society of such users.

Kranzberg, a professor of the History of Technology, has noted that "Technology is neither good nor bad; nor is it neutral." [4] We assert that the same is true with ranking designs, namely that human-computer ranking designs are not neutral and may implicitly espouse specific social values.

When we say *human-computer ranking design*, we refer to the way in which humans and computers co-adaptively modulate their attentions in certain contexts. For example, in a web directory such as the original Yahoo.com, or in some modern web portals, a small subset of human editors select a small subset of appropriate content–links, stories, and media–that are then consumed by a much larger audience. Another example includes the human-computer ranking design embodied in PageRank, the technology that served to spur a search engine named Google. In Brin's PageRank, researchers made use of the social ties and status markers implicit in the pre-Google web's node-link structure, constructing a human-computer ranking design where an automated algorithm defined by scientists is made to infer the relative social prestige of sites and pages on the web.

In the case of the original Yahoo, these rankings were categorical; Some pages were deemed impoverished, some others worthy of inclusions, and some other number, smaller still, deemed "cool" by an iconic representation of red-framed sunglasses. With Google's original PageRank, rankings were numeric, each page assigned a scalar valued score of its query-independent relevance.

A more recent example of this interplay between categorical and numeric, editor and algorithm, includes the human-computer ranking designs embedded within the display algorithms of social networks such as Facebook and MySpace.com. On the social networking site Facebook.com, engineers define algorithms (robots) that modulate who and what appears on a users' news feed, as the status updates of your Facebook friends are not thought to be all uniformly "newsworthy." On the networking site MySpace.com, every user's profile has a listing of that user's so-called "top friends." In MySpace's original "top 8" feature, users could only specify a fixed number of friends, as a way to, as danah boyd points out, "demarcate their identity and signal meaningful relationships with others."

What these examples share in common are processes for how human rankings—of social and political stories (newsworthiness), digital documents (social prestige or authority online), and even people (the strength of social ties)—and automated, computer-selected ranking algorithms (such as PageRank, collaborative filtering algorithms, or item-item clustering) are mixed, refined, and recombined into the rankings eventually perceived and acted upon by users on the web.

In this paper, we propose a theoretical and technical framework for considering the implications of such ranking designs. Specifically, we look at the social values which these designs may encode, and briefly consider three such values:

.

diversity, stability, and awareness.

Flanagan and Nissenbaum have explored activist game approaches, articulating a design methodology that incorporates activist social themes [1]. Inspired by this work, and with these three social values in mind, we apply such an approach towards an understanding of the implications of human-computer ranking design at scale, proposing a theoretical model of collective attention, a "justness" metric in this model, and a ranking design that optimizes for this metric.

1.1 Social Values at the Interface

For example, let us first consider a virtual society of bidders and merchants interacting with an auction web site such as eBay.com. These users may prefer a ranking design that encourages a diversity of merchants as well as awareness of differing items, rather than a ranking design that maximizes merchant profit on a few selected items. A ranking design that incorporates a static sorted order, for example, may direct a large subset of users to a small number of popular items, leading to bidding wars.

In a society that depends on data-driven ranking designs to direct its commercial attention, consumers may prefer a more stable marketplace, that is, one where "eyeballs" and clickstreams are directed to merchants in a manner that is proportional to their collective value to that society, rather than in a ranking design mechanism that encourages a winner-takes-all style of game theory. To be specific, significantly more attention is typically paid to items in the top slot of an ordering compared to items in the second or third slots [2]. Thus, if several items have similar scores, then small perturbations in the scoring functions may lead to large changes in the attention given to these nearly equivalent items.

Besides diversity and stability, consider the social value of awareness—the social consciousness of people in the news and events of contextual or global import. The use of strictly sorted orders in news sites or social networks may lead to a glut of socialite gossip and trivial news on the most popular social centers of attention, as care may not be taken to unbias users' attentional preference for highly ranked items. Such designs, then, may trade-off for social arousal at the cost of greater social awareness.

2. RANKING MODELS

A user's gaze is often used as a proxy measure for that user's attention. Recent eye tracking studies suggest that web searchers tend to trust the rankings provided by search engines, which suggests that the attention given to an item i may be largely a function of its ordinal rank k [2]. We now propose a simple model for the attentional bias of a population of users. In this model, we assume that we know the probability that a user is paying attention to slot k in a ranking order, and thus are given discrete probability distribution function $\sigma(i)$, which defines the amount of attention given to an item i.

We also assume that we are given, perhaps by an oracle, a scoring function f which assigns a strictly positive, numeric score to each item. The scoring function f then directly determines the discrete probability distribution function π, where $\pi(i)$ is the probability that a random user finds item i more relevant than all other items.

We now define a metric for "justness". We say that a ranking system is just if $\sigma \propto \pi$.

In other words, a system is just if it directs the attention of a population of viewers to ranked items in a way that is proportional to the numeric scores originally defined by the scoring function.

In contrast, let us consider two ranking systems, sorted ordering and uniform, random sampling. An adaptive display based on sorted ordering is "just" in the case that the distribution of scores matches the distribution of attention, but in general this is not the case. On the other hand, a system based on uniform, random sampling is "just" in the case when the scoring function evaluates to a constant, but again, this is generally not the case.

3. RANKING DESIGNS

To build some intuition, we first consider the design of a ranking system in the case of a user interface that has only a single "slot." In this case, an adaptive display based on sorted ordering simply displays the item with the highest score, and a system based on uniform, random sampling, simply samples from the set of items uniformly at random.

If we would like to construct a just ranking system in the case of a single slot, note that we can simply perform probability sampling, chosing an item with probability proportional to its score. By the "justness" metric introduced previously, this ranking system would be considered just.

Often, an adaptive display has multiple slots. In this case, we would like to sample from the set of $n!$ permutations such that the collective attention is justly distributed. Oommen et al. sampled permutations according to a distribution, allowing a designer to turn the entropy knob, so to speak, from complete order to uniform randomness [5]. However, Oommen's sampling approach is not easily modified to support our purposes.

3.1 donkey sort

We briefly sketch an algorithmic approach that addresses attentional bias.

We would like to define a ranking design for a list of n slots and a set of n items, each of which has an associated score. In a sorted order design, the items are sorted by score. In a random order design, items are ordered randomly, that is, they are sampled uniformly at random from the set of $n!$ possible permutations. "donkey sort" strikes a balance between these two extremes, attempting to maximize the "justness" metric previously defined.

Instead of sampling orders from all possible $n!$ re-orderings, a practical approach is to consider only n of these re-orderings, namely the n permutations found in a Latin square [6]. The sampling is done such that when you sum the "attentional bias" over all sampled orders, the bias can be cancelled out (or at least reduced). Done carefully, the collective attention given to items can be adjusted to be more in proportion with its previously defined score.

3.2 Related Work

Google has admitted that diversity is a factor involved in their ranking design [3]. However, Google and other search engines remain secretive about these designs. Guan and Cutrell have used eye trackers to study the effect of rank on users' web search behaviors, suggesting re-ordering approaches as further work [2]. donkey sort is the result of our attempt to address this previously anticipated need.

4. CONCLUSION AND FUTURE WORK

We looked at the fundamental technology of ranking design with a social justice lens, and found the choices of complete order and total chaos to be somewhat unappealing. In light of this, we defined a metric for "justness" in ranking and described a novel approach to provide a balance between sorted orders and uniform randomness. By sampling permutations from a Latin Square design, we aim to direct the collective attention of a society in a way that is consistent with the given scoring functions. Doing so, we argue that such a design may serve to explicitly incorporate social values such as awareness, diversity, and stability.

As ongoing work, we look to improve upon our ranking design, and providing a more formal analysis. Studies of the actual impact of ranking designs, with the help of eye tracking and web traffic data sets, could serve to clarify the importance of attentional bias and ranking design.

4.1 Acknowledgments

Thanks to Pat Hanrahan, Tim Roughgarden, and Scott Klemmer for their feedback and support.

5. REFERENCES

[1] M. Flanagan and H. Nissenbaum. A game design methodology to incorporate social activist themes. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 181–190, New York, NY, USA, 2007. ACM Press.

[2] Z. Guan and E. Cutrell. An eye tracking study of the effect of target rank on web search. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 417–420, New York, NY, USA, 2007. ACM Press.

[3] S. Hansell. Google keeps tweaking its search engine. *The New York Times*, June 2007.

[4] M. Kranzberg. Technology and history: "kranzberg's laws". *Technology and Culture*, 27(3):544–560, July 1986.

[5] B. J. Oommen and D. T. H. Ng. On generating random permutations with arbitrary distributions. In *CSC '89: Proceedings of the 17th conference on ACM Annual Computer Science Conference*, pages 27–32, New York, NY, USA, 1989. ACM Press.

[6] E. Williams. Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Chemistry*, 2(2):149–168, 1949.