Anchoring bias in online voting

ZIMO YANG¹, ZI-KÉ ZHANG² and TAO ZHOU¹(a)

¹ Web Sciences Center, University of Electronic Science and Technology of China - Chengdu 610054, PRC
² Institute of Information Economy, Alibaba Business College, Hangzhou Normal University
Hangzhou 310036, PRC

received 8 September 2012; accepted in final form 26 November 2012
published online 4 January 2013

PACS 89.20.Hh – World Wide Web, Internet
PACS 89.20.Ff – Computer science and technology
PACS 89.65.-s – Social and economic systems

Abstract – Voting online with explicit ratings could largely reflect people’s preferences and objects’ qualities, but ratings are always irrational, because they may be affected by many unpredictable factors like mood, weather and other people’s votes. By analyzing two real systems, this paper reveals a systematic bias embedding in the individual decision-making processes, namely people tend to give a low rating after a low rating, as well as a high rating following a high rating. This so-called anchoring bias is validated via extensive comparisons with null models, and numerically speaking, the extent of bias decays with voting interval in a logarithmic form. Our findings could be applied in the design of recommender systems and considered as important complementary materials to previous knowledge about anchoring effects on financial trades, performance judgments, auctions, and so on.

Uncovering human behavioral patterns, such as bursty nature of temporal activity [1–3], scaling laws of human travel [4–6], different selecting patterns of different kinds of users [7,8] on different kinds of objects [9,10], is significant to understand many socioeconomic phenomena and to provide high-quality services. Here we investigate online voting, which contains huge business value in e-commerce. Take recommender systems as an example. Via analyzing online votes, recommender systems can automatically find out suitable products for every customer [11]. So the performance of recommender systems could be largely improved if we make clear the way people vote [12,13].

In some systems, votes are confined to only two extremes —like or dislike, while in some other systems, people can vote with explicit ratings —usually from one star to five stars. Explicit ratings, however, do NOT signify rational judgments. Indeed, people’s votes may be largely affected by prior votes [14] and social pressure (like suggestions from friends) [15]. However, we do not consider the aforementioned biases in this paper. Indeed, in the systems considered in this paper, neither other people’s votes nor social network services are provided for every user. In spite of this, by comparing with null models, considerable voting bias are still observed, which originate from internal decision-making processes of individuals, that is to say, people strongly tend to give a low rating again after voting on a low-quality object, as well as to give a high rating again after voting on a high-quality object. We name such bias as anchoring bias, since it is similar to the well-known anchoring effects in purchases [16,17], auctions [18,19], judgments [20,21] and estimations [22,23] (see also the review article [24] and the references therein).

Previous experiments [24] showed that even a randomly assigned initial value of an object could remarkably affect our estimation on its real value. This paper indicates that a prior vote on another object could affect our current vote because we may take that prior vote as an anchor.

In this paper, we consider two real data sets, MovieLens and WikiLens. MovieLens is a movie rating system with five stars (i.e., ratings can be 1, 2, 3, 4 and 5). The WikiLens is a generalized collaborative recommender system that allows its community to define object types (e.g., beer) and categories (e.g., microbrews), and vote on objects. Ratings in WikiLens can be 1, 1.5, . . . , 4.5, 5. Both the two data sets can be found in GroupLens research web (http://grouplens.org/), and their basic statistics are summarized in table 1.

A recommender system with explicit ratings can be described by a weighted bipartite network where each vote is represented by an edge connecting the corresponding

(a)E-mail: zhutou@ustc.edu
Fig. 1: Distributions of user degrees, which obey the stretched exponential form. We therefore plot the cumulative distribution \( P(k_u) \) instead of \( p(k_u) \) and show the linear fittings of \( \ln(-\ln P(k_u)) \) vs. \( \ln k_u \) in the insets.

Fig. 2: Distributions of object degrees which also obey the stretched exponential form.

| Data Set       | \( N \) | \( M \)   | \( V \)   | \( \rho \) | \langle r \rangle |
|----------------|---------|-----------|-----------|------------|------------------|
| MovieLens      | 6040    | 3952      | 1000292   | 0.042      | 3.58             |
| WikiLens       | 289     | 4951      | 26937     | 0.019      | 3.71             |

Table 1: Basic statistics of MovieLens and WikiLens. \( N, M \) and \( V \) denote the number of users, objects and ratings, respectively, \( \rho = \frac{V}{NM} \) denotes the sparsity of the data, and \( \langle r \rangle \) is the average rating over all votes.
Table 2: Basic statistics of outliers. The six columns from the second to the last one are the number of low-quality outliers (#LQO), the number of votes right after votes on low-quality outliers (#A\(^{-}\)), the number of high-quality outliers (#HQO), the number of votes right after votes on high-quality outliers (#A\(^{+}\)), the number of votes after votes on outliers (#A\(^{-}\)&#A\(^{+}\)), and the percentage of these after-outlier votes in all votes.

| Data Set   | #LQO | #A\(^{-}\) | #HQO | #A\(^{+}\) | #A\(^{-}\)&#A\(^{+}\) | Percentage |
|------------|------|-------------|------|-------------|----------------|------------|
| MovieLens  | 97   | 8526        | 14   | 10713       | 19239          | 1.92%      |
| WikiLens   | 12   | 268         | 13   | 370         | 638            | 2.37%      |

Table 3: Statistics of votes after outliers for MovieLens. \(\langle r \rangle\), \(\langle d_o \rangle\) and \(\langle d_u \rangle\) respectively denote the average rating, the average difference to object average and the average difference to user average.

| MovieLens | \(\langle r \rangle\) | \(\langle d_o \rangle\) | \(\langle d_u \rangle\) |
|-----------|----------------------|------------------------|------------------------|
| A\(^{-}\)  | 2.72                 | -0.054                 | -0.635                 |
| A\(^{+}\)  | 4.16                 | 0.033                  | 0.449                  |

the anchoring bias exists, a vote on an outlier will become the anchor of the next vote, and thus in average we will give high rating after voting on a high-quality object and low rating after a low-quality object.

We use the average rating to estimate an object’s quality, and to reduce the possible errors caused by personalized tastes and unreasonable votes, we only consider the objects getting more than ten votes. Although ratings cannot perfectly reflect qualities, they are correlated with qualities and can be naturally treated as anchors by users. For both MovieLens and WikiLens, an object (with more than ten votes) is distinguished as low-quality or high-quality outlier if its average rating is lower than 2.0 or higher than 4.5.

Denote by \(r_{i\alpha}\) the rating from user \(i\) onto object \(\alpha\), and for an arbitrary user \(i\), all her \(k_i\) ratings are ordered by time as \(r_{iO_1}, r_{iO_2}, r_{iO_3}, \ldots, r_{iO_{k_i}}\), where \(O_1, O_2, O_3, \ldots, O_{k_i}\) are the objects having been voted by \(i\). \(r_{iO_i}\) is the oldest rating, and \(r_{iO_{k_i}}\) is the most recent rating. If \(O_i(l < k_i)\) is a low-quality outlier, \(r_{iO_{l+1}}\) is an after-low-quality-outlier rating (\(A^-\) rating for short), while if \(O_i\) is a high-quality outlier, \(r_{iO_{l+1}}\) is an \(A^+\) rating. According to the above criterion and definition, as shown in table 2, there are in total 19239 (1.92% of all ratings) after-outlier ratings for MovieLens and 638 (2.37% of all ratings) after-outlier ratings for WikiLens. It can be observed that the high-quality outliers get more votes in average, which is in accordance with our common sense that better objects are more popular.

We next compare the votes after low-quality outliers and those after high-quality outliers, namely \(A^-\) and \(A^+\) votes. The average rating among all \(A^-\) votes is defined as

\[
\langle r^- \rangle = \frac{1}{|A^-|} \sum_{r_{i\alpha} \in A^-} r_{i\alpha}.
\]  

In addition, we would like to know the difference between a rating \(r_{i\alpha} \in A^-\) and the average rating on the object \(\alpha\), as well as the difference between \(r_{i\alpha}\) and the average rating by the user \(i\). Accordingly, we define the average difference to object average as

\[
\langle d_o \rangle = \frac{1}{|A^-|} \sum_{r_{i\alpha} \in A^-} (r_{i\alpha} - \langle r_{i\alpha} \rangle),
\]  

where \(\langle r_{i\alpha} \rangle\) denotes the average rating on \(\alpha\), and the average difference to user average as

\[
\langle d_u \rangle = \frac{1}{|A^-|} \sum_{r_{i\alpha} \in A^-} (r_{i\alpha} - \langle r_i \rangle),
\]  

where \(\langle r_i \rangle\) denotes the average rating by \(i\). Analogously, we can define \(\langle r^+ \rangle\), \(\langle d_o^+ \rangle\), and \(\langle d_u^+ \rangle\) for \(A^+\) votes.

Table 3 and table 4 show the remarkable difference between people’s votes after low-quality and high-quality outliers. The results indicate the possible presence of anchoring bias, that is, people tend to give a low rating if the prior-visited object is not good, and vice versa. However, the above evidence is not solid enough since it covers only a tiny fraction of votes, and thus we will further analyze the full rating series of every user.

To get rid of the effects of different voting standards of users (e.g., some users are good-tempered and tend to give high ratings than others) and different deserving ratings of objects (e.g., some objects are of better qualities and should be voted with high ratings), we regulate each rating \(r_{i\alpha}\) in the following four ways: i) to eliminate the average rating over all votes as \(r'_{i\alpha} = r_{i\alpha} - \langle r \rangle\); ii) to eliminate the average rating over all votes on the corresponding object as \(r''_{i\alpha} = r_{i\alpha} - \langle r_{i\alpha} \rangle\); iii) to eliminate the average rating over all votes of the corresponding user as \(r'''_{i\alpha} = r_{i\alpha} - \langle r_{i} \rangle\); iv) to eliminate both average ratings as \(r''''_{i\alpha} = (r_{i\alpha} - \langle r_{i\alpha} \rangle) + (r_{i\alpha} - \langle r_{i} \rangle)\). Readers are easy to reproduce all the following experiments and will find that the four cases lead to qualitatively the same results, and
is defined as user to quantify the aggregation phenomenon for an arbitrary user to measure the memory effect of a time series \[32\], lower rating after a prior low rating. Similar to the method after a prior high rating while they are likely to give a behavior, namele people are likely to give a high rating kind of aggregation reveals the anchoring bias involving voting

We simply divide ratings into two classes—positive ratings and negative ratings, and display them without explicit values in fig. 3(b), where one could—positive ratings and negative ratings, and display them in MovieLens. We simply divide ratings into two classes—positive ratings and negative ratings, and display them in MovieLens who has voted 106 movies in total. All these 106 ratings are displayed according to the voting time in panel (a), and the positive and negative ratings are, respectively, represented by light-green and dark-green lines in panel (b). Notice that the regulated ratings are not necessarily to be integer.

thus we only present the results of case iv) hereinafter and without specific statement, the term rating(s) stands for regulated rating(s) of case iv).

Figure 3(a) presents the rating series of a typical user in MovieLens. We simply divide ratings into two classes—positive ratings and negative ratings, and display them without explicit values in fig. 3(b), where one could observe that ratings in the same class are aggregated. This kind of aggregation reveals the anchoring bias in voting behavior, namely people are likely to give a high rating after a prior high rating while they are likely to give a low rating after a prior low rating. Similar to the method used to measure the memory effect of a time series \[32\], to quantify the aggregation phenomenon for an arbitrary user i, we calculate the Pearson correlation coefficient \(R_i\) of two series \(r_{iO1}, r_{iO2}, \ldots, r_{iOn}\) and \(r_{iO2}, r_{iO3}, \ldots, r_{iOn+1}\), where the Pearson correlation coefficient for two finite series \(x_1, x_2, \ldots, x_n\) and \(y_1, y_2, \ldots, y_n\) is defined as

\[
R(x, y) = \frac{\sum_{i=1}^{n}(x_i - \langle x \rangle)(y_i - \langle y \rangle)}{\sqrt{\sum_{i=1}^{n}(x_i - \langle x \rangle)^2} \sqrt{\sum_{i=1}^{n}(y_i - \langle y \rangle)^2}}
\]  

(7)

According to the definition, a positive \(R_i\) indicates that the user \(i\) may have the anchoring bias.

We compare the empirical results with those of a null model, in which each user’s rating series is randomly reshuffled. That is to say, for an arbitrary user i, the rating series \(\{r_{iO1}, r_{iO2}, \ldots, r_{iOn}\}\), is reordered. Then, for each user i, we calculate both Pearson coefficients for the original series and reshuffled series. Figure 4 compares the distributions of Pearson correlation coefficients \(R\) of the original data and the reshuffled data (each distribution consists of N Pearson coefficients from all users). Clearly, the distributions of the null model (i.e., reshuffled data) peak at about zero, while the empirical distributions peak at a positive value. In addition, the empirical distributions, as a whole, lie in the right of the distributions of the null model. From the cumulative distributions, one could see that for the null model, less than 50% of users are of positive \(R\), while for the empirical data, more than 70% of users are of positive \(R\). To further validate the statistical significance, we apply the bootstrapping approach and denote by \(q_i\) the probability that the Pearson coefficient of the original series of user \(i\) is larger than the Pearson coefficient of a reshuffled series of user \(i\). For each user \(i\), \(q_i\) is numerically estimated by shuffling the original series 1000 times in an independent manner. Accordingly, if \(q_i\) equals 0.211, it means that in total 211 among the 1000 reshuffled series have less Pearson coefficients than the one of the original series of \(i\).

As shown in fig. 5(a) (b), for the majority of users, the \(q\)-values are larger than 0.5. Precisely, the average \(q\)-values over all users for MovieLens and WikiLens are 0.733 and 0.696, respectively. Since for small-degree users the fluctuations are generally large and the statistical results are less reliable, we further show how the average \(q\)-value, \(q(k_u)\), changes with the user-degree \(k_u\) in fig. 5(c) and (d), where \(q(k_u)\) is defined as the average \(q\)-value over all users of degree \(k_u\). One can observe a strongly positive correlation between \(q(k_u)\) and \(k_u\), and for large-degree users, their \(q\)-values is close to 1. Statistically speaking, data from the large-degree users are more reliable, and thus the increasing tendency of \(q(k_u)\) further demonstrates our hypothesis on the existence of an anchoring bias.

In a word, the aforementioned comparisons verify the significance of the anchoring bias in empirical data.

Fig. 3: (Color online) Rating series of a typical user in MovieLens who has voted 106 movies in total. All these 106 ratings are displayed according to the voting time in panel (a), and the positive and negative ratings are, respectively, represented by light-green and dark-green lines in panel (b).

Fig. 4: (Color online) Distributions of users’ Pearson correlation coefficients \(R\). (a) and (c) are for MovieLens, while (b) and (d) are for WikiLens. (a) and (b) show the histograms where \(p(R)\) is the probability density of \(R\), while (c) and (d) show the cumulative distributions where \(P(R)\) denotes the fraction of users whose Pearson correlation coefficients are less than \(R\). In each panel, results from the real data and the null model are, respectively, colored in red and blue.
Anchoring bias in online voting

Fig. 5: Cumulative distributions of users’ $q$-values for MovieLens (a) and WikiLens (b), where $P(q)$ denotes the fraction of users whose $q$-values are less than $q$. Average $q$-values over users with degree $k_u$ vs. $k_u$ for MovieLens (c) and WikiLens (d).

If the number of votes of a user $i$ is larger than $L$, we could extend the Pearson correlation coefficient $R_i$ to an $L$-dependent coefficient $R_i(L)$ as the Pearson correlation coefficient of two series $r_{i01}$, $r_{i02}$, $r_{i03}$, ..., $r_{i0L-1}$, $r_{i0L+1}$, $r_{iO2}$, $r_{iO3}$, ..., $r_{iOm}$. As shown in fig. 6, the average value of the $L$-dependent Pearson correlation coefficient, $\langle R(L) \rangle$, of the empirical data over all users is remarkably larger than that of the null model for small $L$, and the difference between $\langle R(L) \rangle$ of the empirical data and the null model decays in a logarithmic way as

$$\Delta\langle R(L) \rangle \approx A - B \log L,$$

where $A \approx 0.08$ and $B \approx 0.04$ for both MovieLens and WikiLens. This result again indicates the existence of the anchoring bias. Moreover, it suggests that this bias will last a considerable time period, which is in accordance with previous experimental results on the duration of anchoring effects [33].

Combining those aforementioned experiments, the existence and significance of the anchoring bias in online voting is obviously validated, whose pattern, as shown in fig. 3, is very similar to the memory-embedded time series [32]. The extent of the anchoring bias, quantified by the difference of the average regulated rating from the null model, decays in a logarithmic form and will last a considerable duration. The most known literature on anchoring effects considered people’s judgments, evaluations, estimations and predictions in offline world, meanwhile the quick development of Internet and the data processing technologies allow us to study the rich social psychological phenomena in the online world. Quantitative analysis and statistical description based on BIG DATA may build up a new paradigm for social psychology and facilitate the birth of a new branch of psychology, probably called Internet psychology. This work is an elementary attempt that tries to uncover underlying decision-making processes based on extensive statistical analysis. Our findings are helpful in understanding the online-voting pattern and improving the performance of recommender systems.

***

This work is partially supported by the Fundamental Research Funds for the Central Universities of China and the National Natural Science Foundation of China under Grant Nos. 11222543 and 11105024.

REFERENCES

[1] Barabási A.-L., Nature (London), 435 (2005) 207.
[2] Vázquez A., Oliveira J. G., Dezső Z, Goh K.-I., Kondor I. and Barabási A.-L., Phys. Rev. E, 73 (2006) 036127.
[3] Zhou T., Kiet H. A. T., Kim B. J., Wang B.-H. and Holme P., EPL, 82 (2008) 28002.
[4] Brockman D., Hufnagel L. and Geisel T., Nature (London), 439 (2006) 462.
[5] González M. C., Hidalgo C. A. and Barabási A.-L., Nature (London), 453 (2008) 779.
[6] Yan X.-Y., Han X.-P., Wang B.-H. and Zhou T., arXiv:1211.2874.
[7] Shang M.-S., Lü L., Zhang Y.-C. and Zhou T., EPL, 90 (2010) 48006.

68002-p5
Zhang C.-J. and Zeng A., *Physica A*, **391** (2012) 1822.

Vig J., Sen S. and Riedl J., *Proceedings of the 16th International Conference on Intelligent User Interfaces*, (ACM Press, New York) 2011, pp. 93–102.

Chen L. and Pu P., *User Model. User-Adapt. Interact.*, **22** (2012) 125.

Lü L., Medo M., Yeung C.-H., Zhang Y.-C., Zhang Z.-K. and Zhou T., *Phys. Rep.*, **519** (2012) 1.

Chen L. and Pu P., *User Model. User-Adapt. Interact.*, **22** (2012) 125.

Lü L., Medo M., Yeung C.-H., Zhang Y.-C., Zhang Z.-K. and Zhou T., *Phys. Rep.*, **519** (2012) 1.

Koren Y., Commun. ACM, **53**, issue No. 4 (2010) 89.

Koren Y. and Sill J., *Proceedings of the 5th ACM International Conference on Recommender Systems* (ACM Press, New York) 2011, pp. 117–124.

Alon N., Babaioff M., Karidi R., Lavi R. and Tennenholtz M., *Proceedings of the 13th ACM International Conference on Electronic Commerce* (ACM Press, New York) 2012.

Huang J., Cheng X.-Q., Shen H.-W., Zhou T. and Jin X., *Proceedings of the 5th ACM International Conference on Web Search and Data Mining* (ACM Press, New York) 2012, pp. 573–582.

Simonson I. and Drolet A., *J. Consum. Res.*, **31** (2004) 681.

Wu C.-S., Cheng F.-F. and Yen D. C., *Behav. Inf. Technol.*, **31** (2012) 829.

Ku G., Galinsky A. D. and Mumighan J. K., *J. Pers. Soc. Psychol.*, **90** (2006) 975.

Beggs A. and Graddy K., *Am. Econ. Rev.*, **99** (2009) 1027.

Hunt W. A. and College W., *Am. J. Psychol.*, **54** (1941) 395.

Thorsteinson T. J., Breier J., Atwell A., Hamilton C. and Privette M., *Organ. Behav. Hum. Decis. Process.*, **107** (2008) 29.

Kaustia M., Alho E. and Puttonen V., *Financ. Manag.*, **37** (2008) 391.

Cen L., Hilary G., Wei K. C. J. and Zhang J., *Soc. Sci. Res. Netw.*, no. 1572855 (2010).

Furnham A. and Boo H. C., *J. Socio-Econ.*, **40** (2011) 35.

Laherrere J. and Sornette D., *Eur. Phys. J. B*, **2** (1998) 525.

Ivanov P. Ch., Yuen A., Podobnik B. and Lee Y., *Phys. Rev. E*, **69** (2004) 056107.

Zhou T., Wang B.-H., Jin Y.-D., He D.-R., Zhang P.-P., He Y., Su B.-B., Chen K., Zhang Z.-Z. and Liu J.-G., *Int. J. Mod. Phys. C*, **18** (2007) 297.

Clauset A., Shalizi C. and Newman M. E. J., *SIAM Rev.*, **51** (2009) 661.

Newman M. E. J., *Proc. Natl. Acad. Sci. U.S.A.*, **98** (2001) 404.

Ren X.-Z., Yang Z., Wang B.-H. and Zhou T., *Chin. Phys. Lett.*, **29** (2012) 038904.

Lambiotte R. and Aeusloos M., *Phys. Rev. E*, **72** (2005) 066107.

Goh K.-I. and Barabási A.-L., *EPL*, **81** (2008) 48002.

Mussweiler T., *Eur. J. Soc. Psychol.*, **31** (2001) 431.