

The "sticky" review project

As a frequent online shopper, I rely heavily on online reviews. I have sometimes noticed similarities between consecutive reviews. This is surprising, as a review is supposed to reflect a person's private experience, and it is considered independent of other reviews.

There are 3 possible reasons for similarities between consecutive reviews:

1. The reviews reflect the genuine condition of the product. In the case of bad reviews of otherwise highly-scoring product, a repetition of low score or bad reports reflect a repeating problem with the product.
2. Competition attack: in the case of bad reviews, the reviews are written by a competitor to increase his/her sales.
3. The review expresses an opinion in a way that influences others opinion and encourages imitation.

I suggest testing for this phenomenon and identifying its causes.

A temporal decrease in a product score is naturally very important to the seller, whether it is caused by a problem in the product, a competition attack or a bad review that influences others to imitate it. Identifying a repeating problem with the product is interesting to the buyer, especially if bad reviews are buried in a volume of otherwise highly scoring reviews.

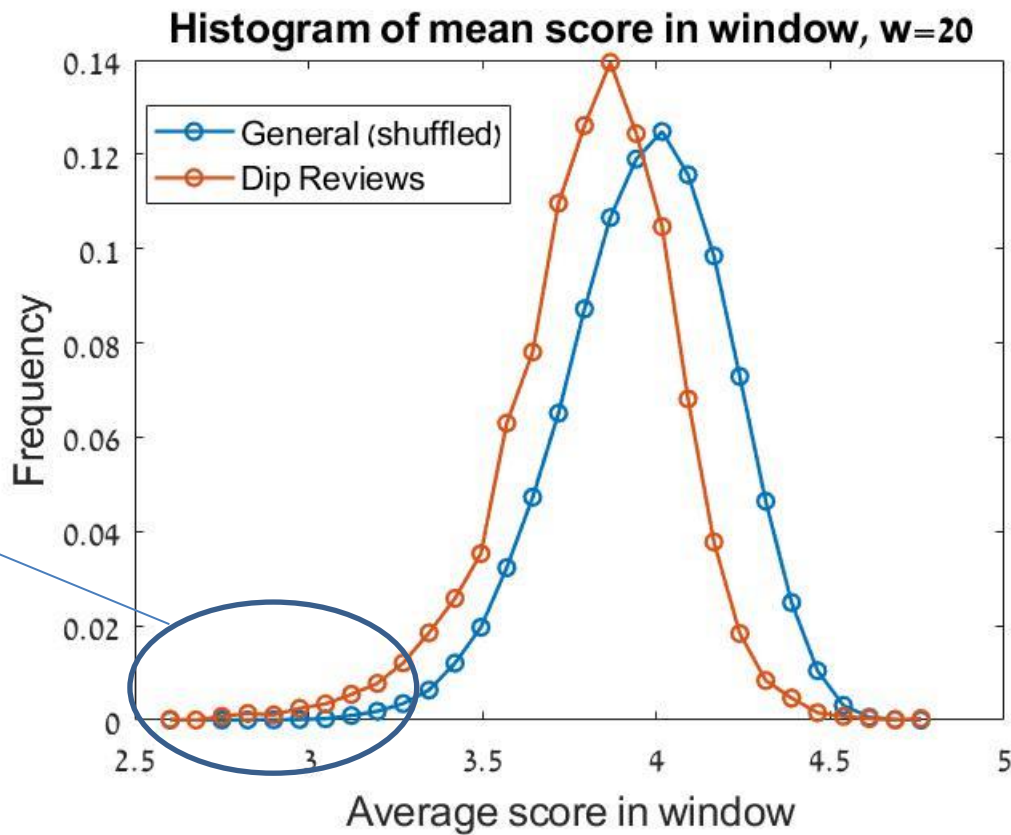
Part 1: identifying "sticky" reviews

I started by testing option 3: a review that influences others to imitate it.

I used Amazon reviews data of electronic products. I chose to concentrate on electronics because they tend to have a higher prices and presumably people invest some time and effort to research products before they buy when prices are high.

I focused on highly scoring products (average score 3.5 and up), and identified two types of reviews:

- Dip reviews: reviews with an exceptional low score (1.5 points lower than that product's average score).
- Sticky reviews: dip reviews that have an exceptionally low score in a review-window after it.



In Graph 1 I show a histogram of the average score in a window of 20 reviews after the dip review (red), and a histogram of the average score in a window of 20 reviews for the shuffled (bootstrapped) data (blue).

I defined as "sticky" all reviews falling at the left area of the histogram (bins 1 to 10, 223 reviews total). A review picked at random from this group has a much higher probability of being genuinely sticky (probability of 80%) than simply a statistical fluke (probability 20%).

Technical details:

I restricted my analysis to reviews with 100 reviews or more, as preliminary research showed that the review-window of interest is of size ~20.

I tested other window sizes: w=10, 30, 40 with similar results.

The probability to see this low window-score at random is the area under the blue graph.

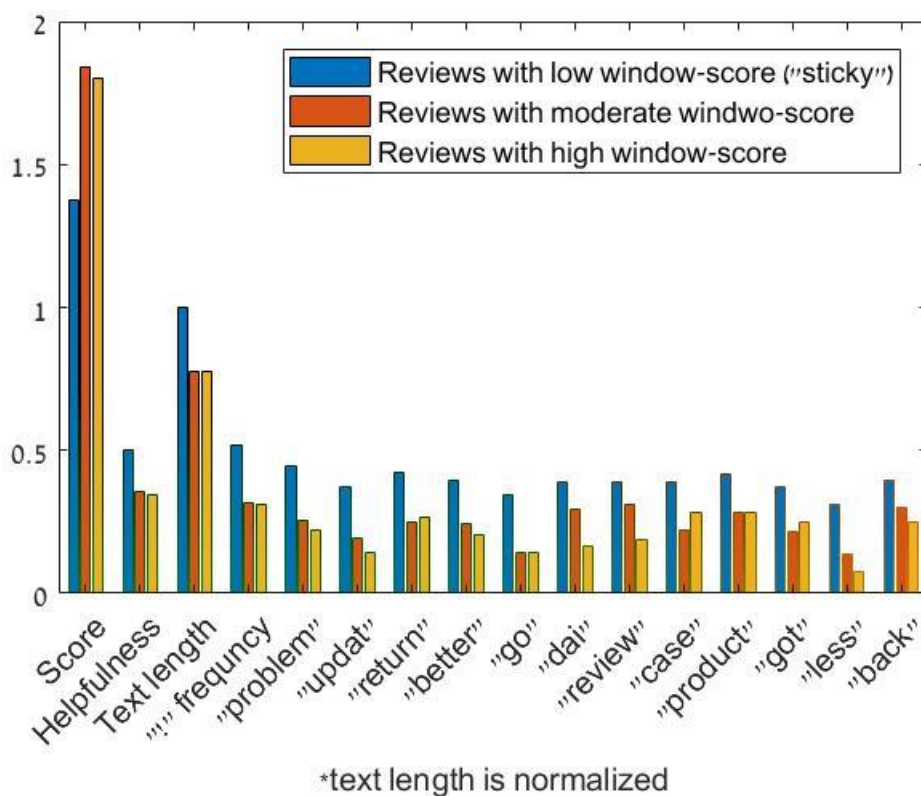
For the group of sticky reviews I defined, this probability is 20%.

Part 2: Content of sticky reviews and the reviews that follow them

Text analysis of the sticky reviews identified some tendencies, but no decisive features.

Comparing "Sticky" reviews with dip reviews falling on the middle and right part of the histogram, I have found that for sticky reviews (Graph2):

- The average score is lower. Note that this is a different measure from the window average score with which I defined the sticky reviews.
- The helpfulness measure is higher. Note that the sticky reviews are identified independently of the helpfulness score.
- The average text length is higher. This is a rough measure of the information communicated in the review.
- The frequency of exclamation mark ("!") is higher. This is a rough measure of the level of emotion expressed in the review.



A further analysis of word frequencies, even after the over-all frequencies of words is taken into account (tfidf calculation) renders no obvious conclusions. Also, a basic lda model could not identify unique topics for the sticky reviews.

Summary:

1. It seems that the reviews reflect genuine problems with the product. Texts tend to concentrate on specific problems and not general dissatisfaction.
2. Competition attack: the average time between reviews in general is 95 hours, while the average time between reviews after sticky reviews is 250 hours. This suggests that the reviews are genuine, but I have not yet analyzed the text in the review-windows.
3. No evidence to suggest that the reviews text influenced the following reviews. No "sticky" reviews.

I conclude that these temporal decreases in average score are not caused by a sticky reviews but by either a genuine problem with the product or a competitor attack.

