Reviews, what are you hiding ?

Clément Lefebvre, Vincent Pollet, Nelson Antunes

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Introduction

Review systems are a key feature of most online shopping sites such as Amazon. They make available to the customers the experience of multiple other customers, emulating a word of mouth opinion circulation, but at a much larger scale. The scale of these systems is both their strength and their weakness: their strength because the massive number of reviews can be statistically averaged to a meaningful information, and their weakness because a customer typically reads only a small number of all the reviews written, and is therefore vulnerable to the eventual biases present in these reviews. Is it possible to compensate this drawback by providing the customer additional information about the reviews, to allow him to more objectively judge the content of a review?

Data Analysis

The analysis was conducted on the 5-core dataset of the 'Electronic' category.
Two common biases influencing a customer interpretation of a review are

- the reviewer expertise in the product category
- the reviewer grading exigency.

The reviewer expertise can be evaluated by the number of reviews written by the reviewer in the product category. The reviewer grading exigency can be assessed by two complementary means: the average rating that the reviewer gives and how this rating relates to the sentiment conveyed by the review.

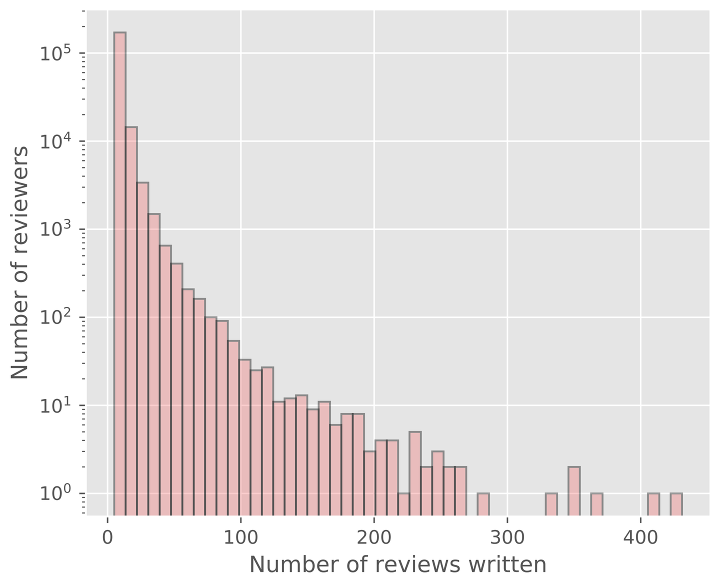

Figure 1: Distribution of the number of reviews per reviewer

The distribution of the number of reviews per reviewer (figure 1) follows a power law so the distinction between an expert reviewer and an average reviewer can be reasonably evaluated. The 5-core dataset is useful in this context, to ensure that the average statistics of a reviewer are computed on a minimal set of elements.

Both sentiment score and rating averages per reviewer are skewed toward high grades, however the ratings are extremely concentrated toward the maximal grade. Most of the reviewers give an excellent rating even though the sentiment conveyed by their review is more mitigated.

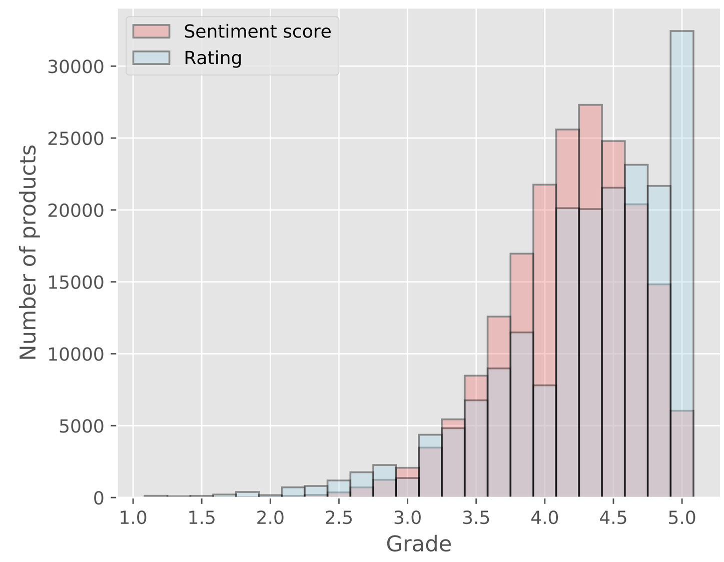

Figure 2: Sentiment scores and review ratings per product

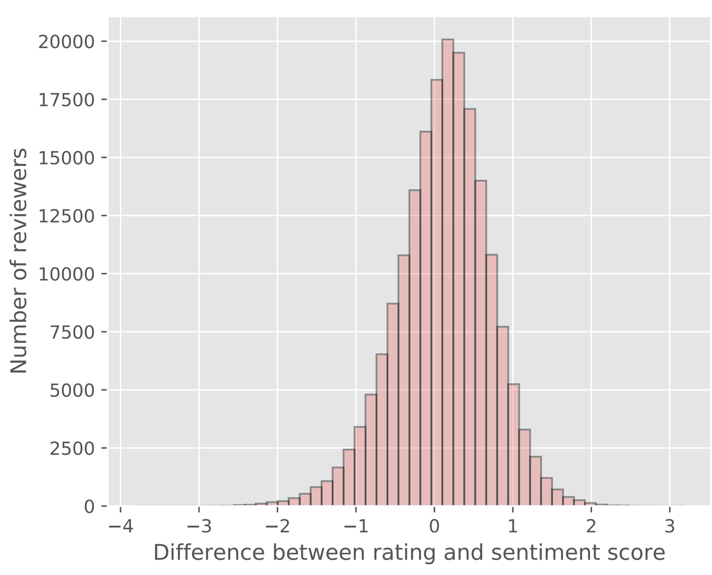

Figure 3: Average difference between reviewer's sentiment score and rating

For a majority of the reviewers, the average difference between their rating and their sentiment score (figure 3) is close to 0. However, for approximately 40% of the reviewers the difference is more than 0.5, and for 10% of them it is more than 1. These are significant difference on a rating between 1 and 5.

The reviewers tend to always give ratings in the same range of values; the sentiment scores follow the same trend (figure 4). This is a useful information to determine the exigency of a reviewer.

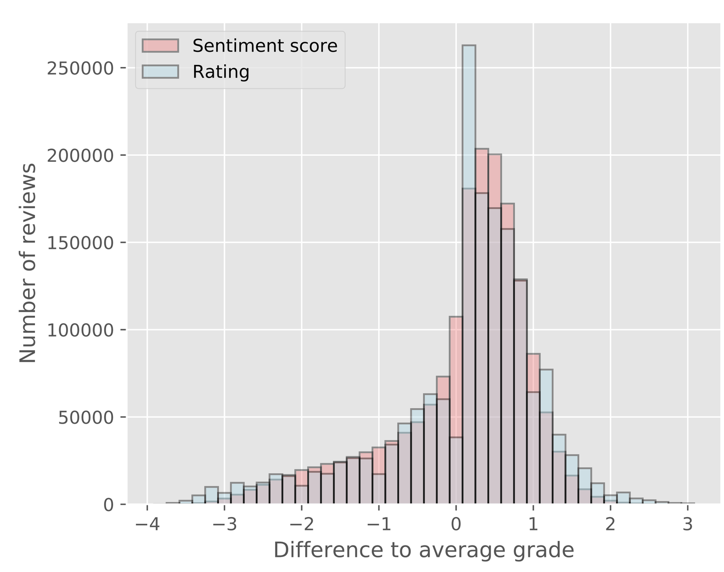

Figure 4: Difference to reviewer's Sentiment score and rating averages

The maths behind the new rating

- The new rating is divided in two parts: the sentiment of the reviewer, and the original rating of the review. Moreover, if the reviewer has given a higher rate than its average then we suppose that he particularly likes the product. With these assumptions, we derive the following formula to compute a new rating of a review:

$$N^R = w_1^s(w_1^{sa} S^{sc} + w_2^{sa}(S^{sc} - S_a^{sc})) + w_2^s(w_1^{ga} R^{sc} + w_2^{ga}(R^{sc} - R_a^{sc})) \quad (1)$$

  Where N^R is the new rating, $w_i^s, w_i^{sr}, w_i^{gr}$ are weights relative to the sentiments and original ratings. They are such that $\sum_{i=1}^{2} w_i = 1$ for all weights. S^{sc} is the sentiment score, R^{sc} is the rating score and S_a^{sc}, R_a^{sc} are the score compared to the average of the reviewers.

- To compute the average new rating of a product, we add another parameter: the reviewer expertise. If the reviewers has a lot of reviews in this category, then we make the assumptions that its reviews are more insightful. Let us compute the expertise score E^{sc} of a reviewer:

$$E^{sc} = 1 + 3\log(N^{rc}) \quad (2)$$

  N^{rc} is the number of reviews in the same product category for the reviewer. Finally a new product rating is computed by taking into account the new rating N^R and the expertise E^{sc} computed before for each review:

$$N^{RP} = \frac{1}{N_b}\sum_{k}^{N_b} w^e N_k^R + (1 - w^e) E_k^{sc} N_k^R \quad (3)$$

  Where N^{RP} is the average of the new rating of a product, N_b is the number of review for the product, w_e is the weight given to the expertise.

Example of an updated rating

- With the formulas presented above, we computed the new rating for one review (figure 5) and the average new rating for one product (figure 6).

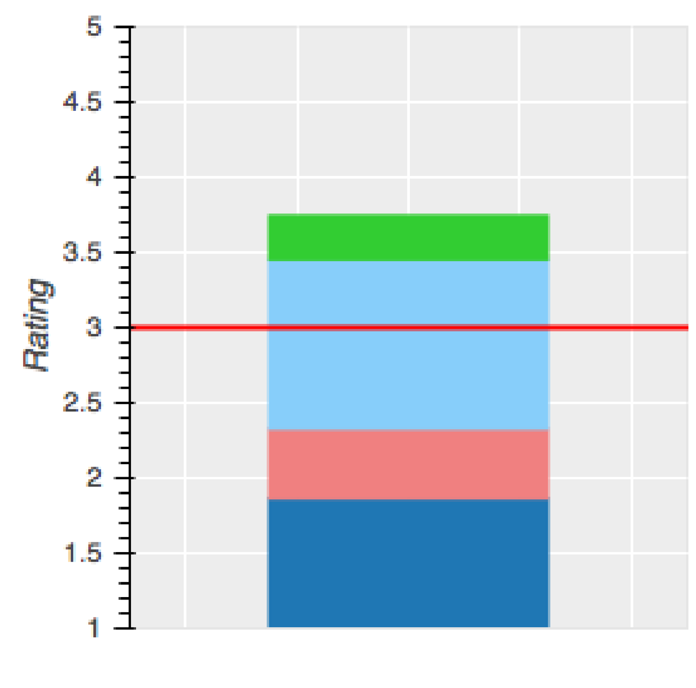

Figure 5: Computed new rating of a review with $w_1^s = 0.5, w_1^{sa} = 0.75, w_1^{ga} = 0.25$

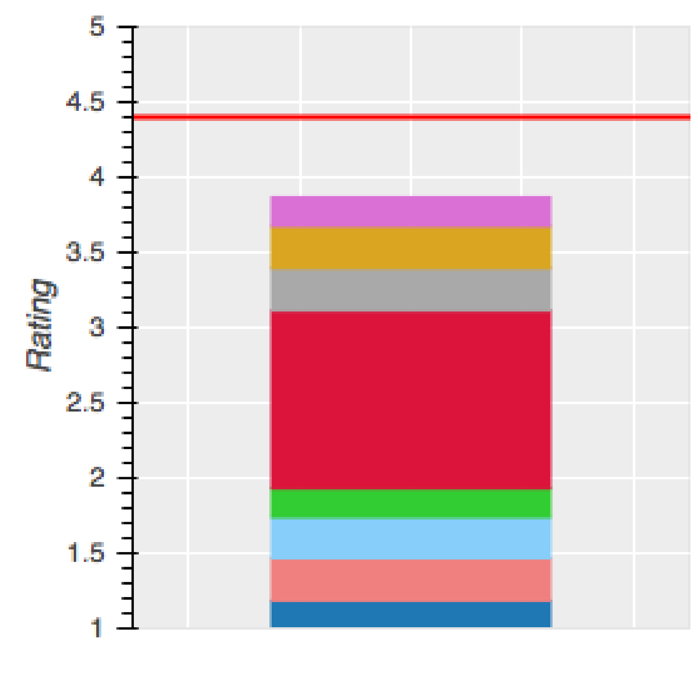

Figure 6: Computed new rating of a product with $w_1^e = 0.5, w_1^s = 0.75, w_1^{sa} = 0.75, w_1^{ga} = 0.5$

- On the two figures, the red line is the original rating of the review or product and the maximum of the bar chart is the new rating. The different colors of the bar charts each represents a different part of the rating as given in equations (1) and (3). For example the green is $w_1^s w_1^{sa} S^{sc}$, a more detailed description is available on [1].
- As a result our rating can be tuned following multiple weights according to the criteria of the customer. For example, the customer can give more or less weight to the expertise of the reviewers.

Conclusion

- With this project, we derived a way to get more insight on a given rating. The computation of the updated rating of a product requires to iterate over all the reviews of every reviewer of the product, which is computationally expensive.
- The model can be enriched with other features, for example we could take into account the length of the review and make the assumptions that a longer review means that the rating is potentially more insightful.

References

[1] Reviews, what are you hiding ?, https://c-lefebvre.github.io/ADA_Homeworks/
[2] Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering, R. He, J. McAuley, WWW, 2016
[3] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.