Dont judge a book by its cover!

Mike Bardet
mike.bardet@epfl.ch

Diana Petrescu
diana.petrescu@epfl.ch

Abstract

In the consumer society in which we live today, how do appearances influence our critical thinking and opinions? Does something as simple as a color could influence our perception of something ? To answer these questions we have chosen to analyze a concrete case, namely the impact of book covers on people's opinions. More precisely, we analyze the impact of colors and brightness of book covers on the reviews and ratings of the book.

1 Introduction

Our ancestors often relied on appearances to judge situations and react quickly in case of danger. Nowadays, we have kept this instinct and tend to be easily influenced by the visual aspects. However, in this mass production and consumption society, this way of reacting can become a weakness and make us susceptible to manipulation. Do we still manage in general to keep our critical mind and not always trust the appearances ? Even in an intellectual context such as reading, are we still influenced by the visual? This is the question we will answer in this report. In particular, we will analyze the impact of the colors and brightness of book covers on the score and reviews given by the consumer. We will also address the question of whether there is an effect of the category of the book on the ratings and colors.

2 Dataset

2.1 Data Collection

In order to have relevant and enough data to answer our questions, we used a dataset based on the Amazon.com, Inc. marketplace. For all the basic information on books, we used the data sets provided in the ADA cluster. More precisely, we used the three amazon_reviews_us_Books_* tsv. We also had to to get more information about book covers and book categories. To do this, we used another data set also (BookCoverDataset, 2018).

2.2 Dataset Description

Given the data collected and described above, we decided to keep only the books that had at least 5 reviews in order for the ratings to be relevant and based on several opinions. After doing this, we had a total of 1786163 reviews for 47976 different books and 47963 distinct book covers given by their urls (given the fact that some of the books had several volumes or edition and the same cover for example).

There were 32 book categories. However, for our analysis 32 categories were too much so we grouped some of them together. Moreover, in order to estimate the opinion of the readers, we wanted to base our thinking on the star-rating of the product. However, when analyzing the distribution of the star-ratings, we noticed that 71% of the reviews had the maximal score (5 stars) so it couldn't help us discriminate between the books. We discuss in more detail how we addressed these two problems in Subsection 3.1 and Subsection 3.5.

When analyzing the reviews over time, we notice an exponential increase in the number of reviews. One the cause may be the increasing publication of books over time, which is not a problem for us. What could be a problem is that this review distribution may also be due on one hand by the increasing importance of Amazon, and on the second hand by the democratization of internet, which allows more people to give reviews.

For more detailed information about each column of our data set, see the ReadMe on Github.

3 Methodology

3.1 Reviews Scores

As already explained above, a lot of reviews have 5 stars (maximal grade). We would prefer something more uniformly distributed, because it is likely that people tend to give very high ratings, even though they do not consider the book being perfect. So, it’s harder for us to distinguish between good and bad books. Therefore, we need another way to discriminate a bit more between the books scores.

We could think about training a model based on the reviews, but the problem is we do not have a training set with ”real” ratings, in the sense ratings that represents perfectly what people feel about the books. As we couldn’t train a model based on the reviews, we use the already trained sentiment intensity analyser of VADER (Valence Aware Dictionary and sEntiment Reasoner) Python package. It is a great tool to analyze social media text as it allows to take into account how people usually write on Internet and takes into account the punctuation or abbreviation words like ”wtf”, ”:)”, ”!!!” which are useful for judging the sentiment of a text (Vader, 2017).

Using VADER, we found a new score based on the intensity of sentiment (negative, neutral, positive) of the reviews. This score has a correlation of 0.3, which is not that high, but is statically significative because of the high number of reviews. So we weighted the old score of the review by this score with a weight of 0.5 to each score.

Even though the total amount of 5 stars is a little bit reduced, the adjusted score has a similar distribution as the star rating. We conclude that the star rating may be not so bad and could reflect pretty well the real feeling of people about the rated books. Of course, the vader analyser is not perfect and we cannot make strong conclusion about that.

3.2 Covers’ dominant colors extraction

We first had to find the dominant colors of every book cover. We first naively thought of averaging the colors of the pixels to find the dominant color. However, this technique didn’t work at all and the color obtained was not the color a human perceived in the cover image. We also thought of taking the most frequent colors. However as this worked well on some of the covers on other, it was a complete failure. Finally, we decided to do use k-Means clustering. K-means build k clusters of pixel that have similar color. We used sklearn.cluster.Kmeans with $k = 10$ as it provides the best results on the cover images and was the most stable (when launching several times, we obtained the same results independently of the initial random points) (Kmeans, 2017).

At first, we just wanted to take a dominant color. However, when taking a closer look at the covers we saw that the best way to describe every cover was by taking the two most common dominant colors.

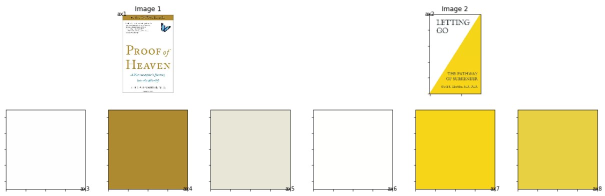

Figure 1: 3 dominant colors - Images 1 and 2

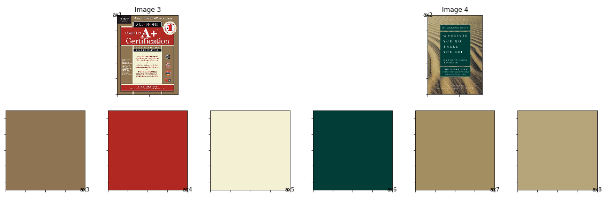

Figure 2: 3 dominant colors - Images 3 and 4

Indeed, we can see in the image above that if we only took one dominant colors, covers like the second image or the third image wouldn’t be well represented. It would be good for covers like the third one. However, some of the covers that were really ”simple” wouldn’t be well described by the third or more dominant color and often the third most dominant color was similar to the the second one.

However, finding the dominants colors take between 4 and 15 seconds depending on the cover image (due to the stabilization part of the k-mean algorithm). Therefore, we thought of reducing the size of the image and creating thumbnails to work on them. The results with thumbnail were a good approximation and we could calculate all the dominants colors in just one hour. However, as they were less precise, we did the calculation with the whole images so we had to distribute the work and launch several program in parallel on different parts of the data set.

3.3 Covers' dominant colors classification

As there are a lot of different shade of colors, we had to classify the dominant colors obtained into categories of colors. We decided to classify them into 16 different distinct colors which could represent the spectrum of colors namely: red, orange, yellow, lime, green, cyan, blue, purple, magenta, grey, black, white, maroon, brown, teal, navy (DistinctColors, 2017). To find the category of the dominant colors found, we first thought of using the euclidian distance between the RGB of the dominant color and the RGB of the categories and to take the minimum. However, as we were interested in how people perceived the colors of the cover, doing so wouldn't be relevant because some colors are really similar in terms of RGB but as human beings perceive them as quite different. Therefore we decided to represent the colors using the CIELAB color space which is relevant for this kind of analysis (CIELABColorSpace,).

3.4 Covers' brightness

From our results concerning the color distribution, we decided to also explore the relation between the brightness of images and the ratings. Again, there are several ways to compute the brightness. Acccording to (Finley, 2006), we decided to use the perceived brightness. The perceived brightness is a weighted mean of the RGB values, considering that green is perceived as bright and blue as dark, so they high a high and low weight correspondingly. The following image shows the difference between the original image on the left, the brightness as computed by HSL, and the perceived brightness on the right.

Figure 3: Perceived brightness versus simple brightness

3.5 Book categories

In the original data set there were 32 books categories but we decided to group the categories that were similar together to obtain 8 categories. Here are the final categories obtained from the initial ones. "Health lifestyle" = ["Cookbooks, Food Wine", "Health, Fitness Dieting", "Sports Outdoors", "Self-Help", "Travel"]. "History Religion" = ["History", "Biographies Memoirs", "Religion Spirituality", "Christian Books Bibles"]. "Arts" = ["Arts Photography", "Crafts, Hobbies Home", "Calendars"]. "Engineering Science" = ["Engineering Transportation", "Business Money", "Science Math", "Law", "Test Preparation", "Computers Technology", "Medical Books"]. "Literature" = ["Literature Fiction", "Mystery, Thriller Suspense", "Science Fiction Fantasy", "Romance"]. "Entertainement" = ["Children's Books", "Humor Entertainment", "Comics Graphic Novels", "Teen Young Adult"]. "Social sciences" = ["Politics Social Sciences", "Parenting Relationships", "Education Teaching", "Gay Lesbian"]. "Other" = ["Reference"] ("Reference" could not be classified in other categories as it contained a lot of different books).

4 Results

4.1 Influence of book covers

The first thing we noticed when analyzing the covers colors is that black and white appear very often as dominant color. This is mainly because a very dark or very bright background is often used for book covers. Also, sometimes the title font is big and white, over a dark background, which lead to black and white as dominant colors.

So, we are interested in the color distribution concerning the very low (1 star) reviews and the very high (5 stars) reviews. We plot the difference between the color distribution of 5 stars and 1 stars.

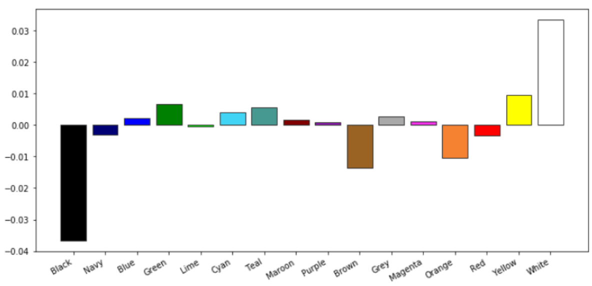

Figure 4: Color distribution difference (%) between 5-star and 1-star rating

It appears that black, maroon, orange and red color are more present when the rating is low. In the psychology of colors, red and black are related to darkness, aggressiveness, while white, blue and green are related to something more peaceful. We can indeed see that green and blue are more present in the 5-star ratings.

In the same line, it appears that bright colors seem more related to high ratings and dark colors to low ratings. We explore this idea by computing the perceived brightness for each class (low and high ratings). We found 138.3 for 1-star and 143.7 for 5-star. This does not seem very high, but we need to do a test. A Levene test tells us the variances are heterogeneous, so we perform a Welch t-test. Note that the distribution of brightness does not follow a normal distribution, but since the observation count if pretty high, this is not a problem and we can go on with t-test which gives us a significant result. We can conclude the brightness is not the same for low rating and high rating.

4.2 Influence of book categories

First we explore how the average rating change between categories. The average does not change much between categories, we oscillate between 4.3 and 4.6. Still we have to notice that Entertainment and Arts have a significantly better average rating than Engineering and Science. We can suppose that since the first ones are more for fun and the last ones for work, it's more likely that people will appreciate something that is suppose to entertain them.

Concerning the color distribution between categories, there are no huge differences, but according to figure 5, we have to notice that bright color are more present for the categories "Health Lifestyle", "Social sciences", and "Other". The category "Literature" has very dark color on average, comparing to the others. Why these differences ? We argue that the color choice is important when someone design a book cover. The literature represents mostly fictional stories. We want to trigger curiosity, so the cover has to convey mystery and deepness. On the opposite side, "Health Lifestyle" is about us, about how to get better, it has to convey trust and happiness, which are better represented by bright colors.

Finally, it is important to notice that in almost all categories, the covers corresponding to 5-star rating have a higher brightness on average, comparing to the covers corresponding to 1-star rating. This suggests that categories do not influence the

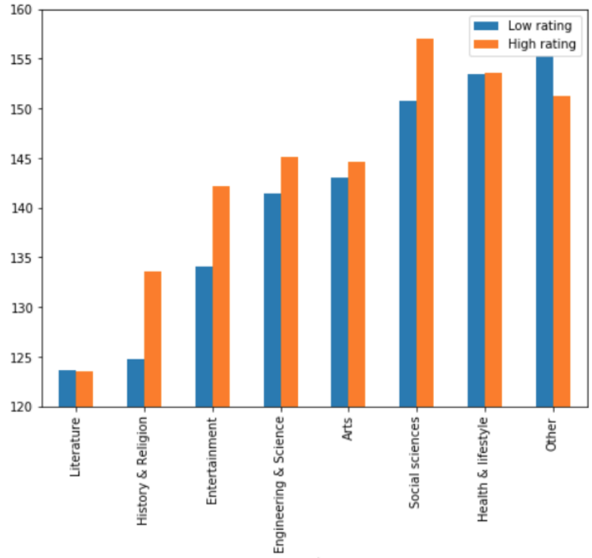

Figure 5: Perceived brightness depending on the category and rating

relation between rating and brightness.

4.3 Influence of time

5 Conclusion

Our main result concerns the relation between perceived brightness and ratings. It appears that high ratings are more likely given to bright books rather than dark books. Unfortunately, it is hard to talk about causality. If someone rates a book, it means he has read it, so he should not be influenced that much by the cover. The cover color influence could be explored with a question as "Is a book more often bought if its cover is bright", but in our case, we do not expect this kind of strong influence. However, the results are there. It would also be weird to say that a bright color is more likely given to good books. So we can conclude that the brightness of a book appears to have an influence on the rating. Confirming this theory would be difficult, because we would need something like two different color for the same book. In this case we could directly compare the average rating between people who have the dark book and people who have the bright book, but on practice this is not feasible on large scale.

Do not be fooled by appearances and keep your critical mind !

References

Adam Spannbauer. 2018. *Finding and Using Images' Dominant Colors using Python & OpenCV*. https://adamspannbauer.github.io/2018/03/02/app-icon-dominant-colors/.

Human Interface Laboratory, Kyushu University. 2018. *Book Cover Dataset*. Kyushu University. https://github.com/uchidalab/book-dataset.

Jodie Burchell. 2017. *Using VADER to handle sentiment analysis with social media text*. http://t-redactyl.io/blog/2017/04/using-vader-to-handle-sentiment-analysis-with-social-media-text.html.

Sasha Trubetskoy. 2017. *List of 20 Simple, Distinct Colors*. https://sashat.me/2017/01/11/list-of-20-simple-distinct-colors/.

Wikipedia. *CIELAB color space*. https://en.wikipedia.org/wiki/CIELAB_color_space.

Darel Rex Finley. 2006. *HSP Color Model Alternative to HSV (HSB) and HSL*. http://alienryderflex.com/hsp.html.