Who Judges Books by Their Covers

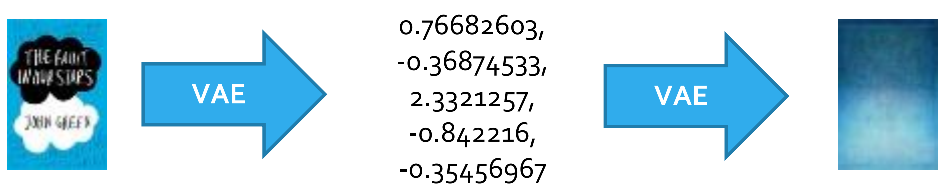
A detailed look at using cover images to predict an individual's enjoyment of a book.

Do you judge books by their covers? I do not mean this metaphorically, (of course you'd never judge a person on how they look), but rather quite literally: do the covers of books dictate your enjoyment of them? While you may be saying "no, of course not" the data suggest otherwise! Using nothing more than an individual's ratings of 10 books that they have previously read, and the book's cover images it is quite likely that a simple model can predict how they will feel about books into the future.

Book Covers as Data

The obvious hurdle in conducting an analysis with images as data is finding a way to represent them. Models (typically) require numerical data, and work best if this numerical data is relatively small and informative. Pictures seem to be the complete opposite. Even when expressed numerically a relatively small image (say 75-by-50) will be comprised of many (11,250) numbers! Moreover, if a pixel is (for example) blue, it is likely that the neighboring pixels also are blue, and so the information is repetitive.

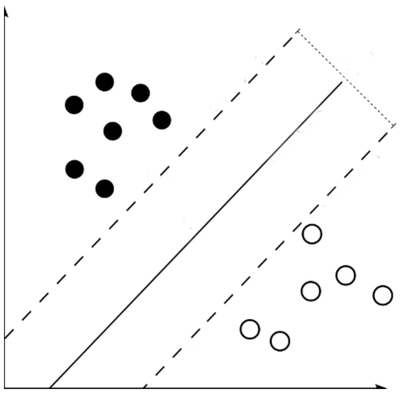
Luckily, we can make use of techniques for "Dimensionality Reduction." Dimensionality reduction is a class of broadly applicable methods in data science which try to find condensed representations of our data, without losing much information. Imagine the process of studying for a test. You have an entire textbook's worth of material that you are required to know, which is obviously a lot to deal with. Instead of trying to memorize the entire contents, you likely write highly condensed notes or flashcards containing what you think are the most important bits of information, summarized in your own words. Your goal with these flashcards is to maintain as much of the original information that you deem important, while greatly reducing the size of the material that you must memorize. Then, when you go into the test, you do your best to recreate the original textbook contents, from your memory of the flashcards. That is the goal of dimensionality reduction. In this analysis, our "textbook" is the original images, and our flash cards will be a small set of numbers that capture this information. While there are very many techniques available, for this analysis I use a deep learning structure known as a "Variational Autoencoder" (VAE). As a concrete example, consider this image which is a 75-by-50 book cover (and is thus made-up of 11,250 numbers). Here, I tell the VAE to try to find a set of 5 numbers that can most accurately re-create the original image (as though you limited yourself to five flash cards to study).



11,250 Numbers to represent **Projected down to 5 numbers** **Projected back up to 11,250**

While it is obvious that a much of the information is lost in this process (note that the re-created image has not text, and less well-defined shapes), the reduced representation is 0.04% the size of the initial image – an astonishing reduction.

Determining Whether A Cover Matters



After embedding the relevant book covers into the lower-dimensional space, we need to decide how to go about determining whether this information is predictive. This can be framed as a straightforward classification problem. I decided to use Support Vector Machines (SVMs) and created a multiclass SVM for each user, attempting to predict the rating each user would give based on the cover image.

For the unfamiliar, an SVM works by trying to find the "mathematically optimal" line to separate two groups of data. Consider the image to the left. If the black and white circles represent two categories of objects, then we can see that both dashed lines, as well as the solid line, accurately divide the two groups. However, the solid line is "mathematically optimal" as it maximizes the space between the groups. If we used either dashed line, then it is more likely that an additional point we have not yet observed will fall on the wrong side of the line, compared to the using the solid line. A multiclass SVM extends this idea to more than two groups.

Do People Judge Books by their Covers?

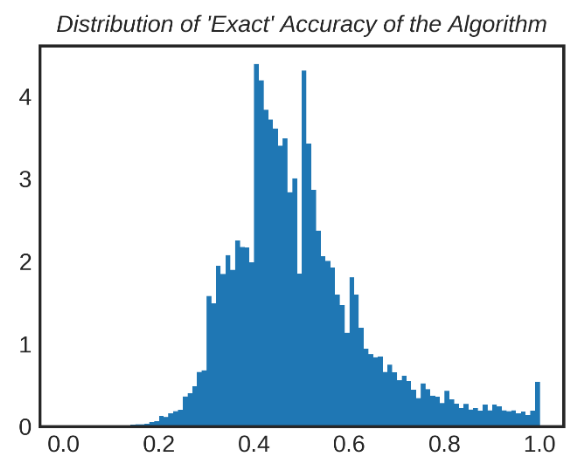
The results of the relatively simple analysis are quite surprising. This is a five-class prediction problem, and so in the event of completely random guessing we would expect that approximately 20% of our guesses are correct and approximately 52% of our guesses are ±1 of the truth. Over the entire dataset, the algorithm correctly predicted the rating that the user would give in **50%** of cases and was within one in **88%** of the cases.

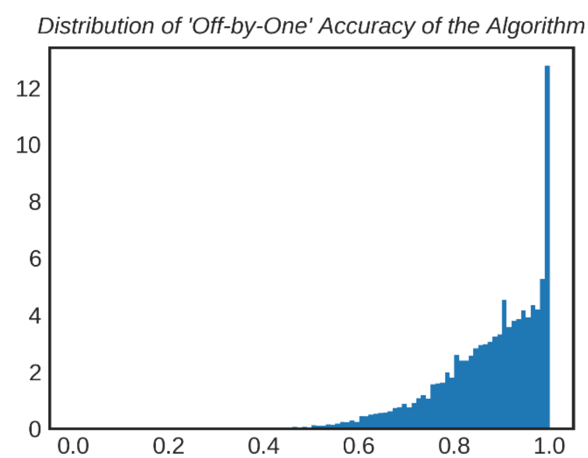
Looking at the complete distributions, for most individuals the algorithm gets somewhere between 40% and 60% of the recommendations 'exactly' correct. Interestingly, we see a positive skew, meaning that there are more individuals who have extremely high rates of accuracy than extremely low.

The distribution of the 'off-by-one' accuracy favours the algorithm even more strongly. There are a tremendous number of individuals for whom all predictions are within one of the true rating, and there appears to be effectively no individuals with less than 60% of the predictions within one unit.

Clearly the use of book covers generates predictions that work better than random chance, but the aggregated data are far less interesting than looking at individuals.

Algorithm's Results





Individual Case Studies

One interesting component of this analysis is the ability to personalize the results. We need not say that people generally judge books by their covers, as we are able to say whether certain individuals do or do not. This personalization is one of the great promises of data science, and so I consider two examples here.

Users #21500: Frequently Judges by the Covers

User #21500 is someone who seems to frequently judge a book by its cover. In **56.3%** of the 80 predictions made, the algorithm got the exact result and in **85%** it was within one. What exactly is the algorithm seeing with this user?

The classifier was trained on the following samples of book covers – there was one book which received a rating of 2, 4 books which received ratings of 3, 4, and 5 each. The following figure shows both the original covers for these books, and the reconstructed covers that the algorithm is working with.

When run through the classification stage, the algorithm never predicted a rating of 2 for the user. In the following figure you can see three covers which had a correctly predicted rating of 3, 4, and 5, and one incorrect prediction each.

Included with the data of current ratings were "read later" lists, containing books that each user had indicated that they were interested in. If user #21500 happens to be reading this, I would recommend that you start with one of ***Every Day*** by David Levithan, ***The Time Keeper*** by Mitch Album, or ***Madame Tussaud: A Novel of the French Revolution*** by Michelle Moran – the algorithm figures you will rate each of these a 5. Moreover, you should probably avoid ***Goliath, The Secret Hour,*** and ***Peeps*** all by Scott Westerfeld, as you seem likely to only rate these a 3.

User #10897: Never Judges by the Covers

In complete contrast user #10897 appears to almost never judge a book by its cover. In fact, the algorithm scores are worse than random. While the algorithm exactly guessed the score in 20% of cases (in line with random chance), the algorithm was within one in only 35% of the cases, well below random guessing. For this user, 13 books were selected (the data shown below) to train the classifier. Predictions were then made for an additional 74 books.

Unfortunately, (or perhaps fortunately if you are an author), the algorithm could pick-up no concrete patterns in this data. Thus, if you are reading this, I can make no firm recommendations as to whether you should pick up ***Restless*** by William Boyd, ***To the Lighthouse*** by Virginia Wolf, or ***Treasure Island*** by Robert Louis Stevenson next. The algorithm seems to think that you will rate each of these as a 5, so perhaps it is wise to stay away from all of them!

Conclusions

While it is not clear that we can say all users judge books by their covers, it is certainly true that some do. More accurately we can say that there is valuable information contained in the covers of books, and this may be a useful signal in assessing the perceived quality. It seems likely that a recommendation system, for instance, ought to be based on more than just the covers, though they should not be ruled out entirely (perhaps that is an interesting problem for you to tackle!).

For a complete look at the code that was used, a more in-depth technical report, and some specific caveats to the results obtained check-out https://github.com/DylanSpicker/judging-covers.