Footwear Image Prediction

BY: QIN CHE, PEIYU LI, JUN SONG, ZIMEI XU

| Advanced Business Analytics | February 28, 2017

Setting

Image recognition has been a popular topic these years, and its potential in the business area has been yet to exploit.

Image recognition has a more formal term which is computer vision. Computer vision is an interdisciplinary field that deals with how computers can be made to gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do.[1] (definition from wiki)

Image recognition technique can be very powerful in the e-commerce business. In our project, the primary objective is to develop an Image Predictive Model for identifying categories of goods. In this case, we are using Women's Shoes on 6pm to fit the model. This model can be applied to the following two use cases:

- To reduce the workload of categorizing new shoes by human eyes and labeling them manually.
- To help customers browsing the similar category by uploading the image of their interesting shoes.

Data

We scraped all our data from www.6pm.com. Our algorithm collected 3000 images of women's shoes from categories that contain at least 200 products to be our full dataset. For images in each category, we label them with the corresponding category name (e.g. Heels, Boots, etc.).

The challenge we encountered during scraping data was regarding the wrong labels. By browsing part of the 3,000 images, we found lots of images were labeled incorrectly from what we saw on the website. However, there was no problem in our code for this process after diagnosis. At last, we narrowed down the issue to the URLs of browsing pages, which are varying among different categories. To fix the problem, we expanded our code to also get all the category information and the corresponding URLs of their browsing pages automatically, instead of just downloading the images on specific pages.

Results

We applied two approaches, Classification Tree (with a maximum depth of 10) and Support Vector Machine, to fit and predict. (We also tried K-means Clustering, but it didn't work well because it is an unlabeled predictive method.) Under each predictive approach, we validated the model by calculating cross validation scores and overall accuracy with both test and full dataset.

Cross validation score: using cross_val_score function from sklearn package to calculate 5-folds cross validation accuracy.

Overall accuracy: using the following equation to calculate the overall accuracy.

Accuracy = the number of correct predictions / the number of total observations

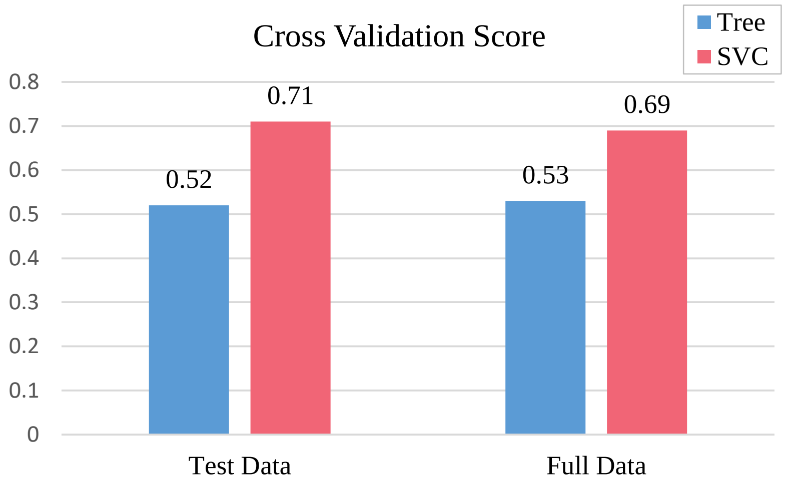
[1] https://en.wikipedia.org/wiki/Computer_vision

To get the number of correct predictions, we can construct a new data frame with two columns, the true category in the dataset and the predicted category of the dataset. Then we check whether the values in these two columns are equal to each other row by row and insert the results in a third column. We assign 1 as matched and 0 as unmatched. The number of correct predictions is simply the sum of the values in the third column.

Following are the comparison results of accuracies for different models and datasets:

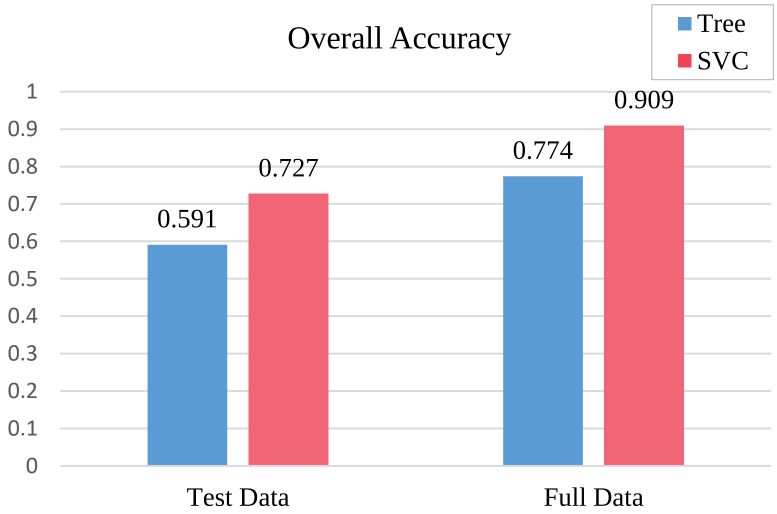
1. Cross Validation Score

   Classification Tree has an average score around 50%, whereas SVM has one around 70%. By comparing cross validation scores, we can see that the accuracy of Classification Tree is nearly 20% lower than that of SVM, with both test data and full data.



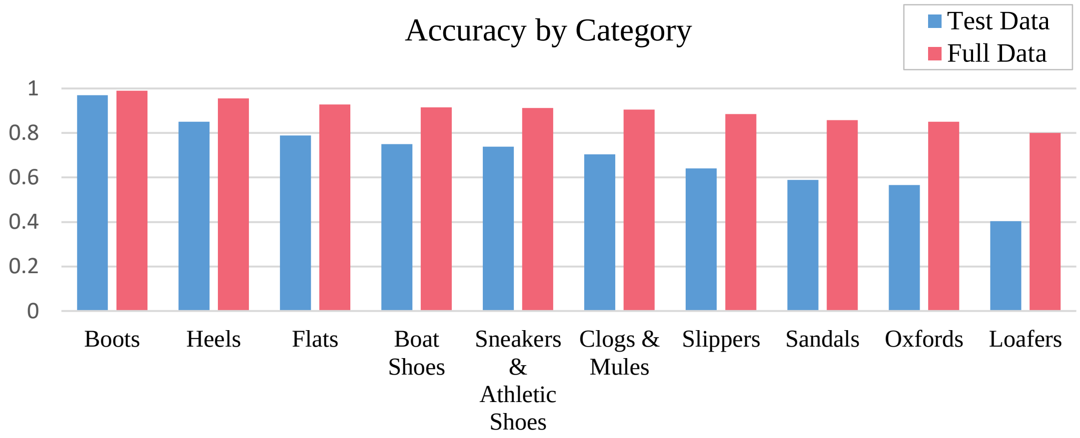
2. Overall Accuracy

   Following the computation of overall accuracy, we can also find the fact that the accuracy of Classification Tree is lower than that of SVM, with both test data and full data. Although the gap gets smaller than that from the cross validation approach, because the overall accuracies tend to be higher. In particular, the overall accuracy of SVM can be as high as 91% with the full dataset.



3. Overall Accuracy by Category

   By drilling down the overall accuracy to each category, we can see the ranking of the accuracies of the ten categories are in the same order for both test and full data. The difference is only between the values for the sake of the impact from training data in the full dataset. The order of the ten categories are: Boots, Heels, Flats, Sneakers & Athletic Shoes, Boat Shoes, Clogs & Mules, Slippers, Oxfords, Sandals, and Loafers.



Analysis and Discussion

Based on the cross validation score and the overall accuracy we calculated above, the SVM model predicts more accurately than classification tree with a maximum depth of 10. The result might change as we try different parameters to fit a better tree. Nonetheless, it will become less efficient for us to do this. In contrast, the SVM approach can fit a highly accurate model without lots of trials. Thereby, we choose this approach for further analysis.

Moreover, among all the categories, prediction accuracies for Boots and Heels are apparently higher than those for other categories. One intuitive explanation for this finding is that the shapes of shoes in these two categories are easier to distinguish than those in other categories. For example, Slippers, Sandals, and Boat shoes are so similar with each other in both shape and color that we even cannot perfectly categorize them manually. No wonder the accuracies of these categories are lower.

Our predictive model has been effective in predicting the label of shoes. For future improvement, we can try our model on a larger dataset and in more categories. To implement it further in e-commerce application, we can expand our algorithm from predict similar category to predict similar single image when a consumer search for one image. Furthermore, this algorithm can be incorporated to other recommendation systems such as collaborative filtering and item-based filtering to help consumers exploring the warehouse, and in turns enlarge the ARPU (Average Revenue Per User) for the company.

References:

1. https://en.wikipedia.org/wiki/Computer_vision (see footnote 1)