Using Deep Learning to Classify Images of Retail Goods

Amrit Saxena, Neal Khosla, Vignesh Venkatraman

Stanford University | CS 229: Machine Learning | Autumn 2013

Project Milestone

**Introduction**

In response to the relatively tepid rate of technological innovation in the retail industry, we are leveraging machine learning to build an algorithm that can classify images of retail goods. The goal of this project is to be able to eventually recommend retail goods to consumers by analyzing images of goods that they like. We believe that successfully doing so can significantly improve the way in which people shop both physically and online as it will allow for more customized and accurate recommendations. This is a problem that many people are trying to tackle in the real world right now with varying levels of success. A clothing recommendation engine, Kaleidoscope, initially tried to recommend similar clothing items to its users by utilizing machine learning-based classification methods but ended up settling for manual classifications due to an inability to properly classify the stylistic elements of retail goods that they were assessing.

The problem that we are seeking to solve is admittedly a difficult one, and we want to ensure that we start with a retail product that is not overly variable or complex, in order to maximize our chances of substantive initial results. As such, for our CS 229 final project, we are starting by analyzing images of shoes to generate these recommendations. We feel that shoes are an appropriate good to begin with as there is not an overwhelming amount of variability in terms of the styles of shoes as most shoes have fairly similar elements (e.g., shoelaces, logos, relatively flat soles, relatively consistent coloring patterns, etc.).

**Dataset**

We have created a dataset by scraping the images of all **[insert number here]** shoes that are listed on Zappos.com. We chose to start with the images on Zappos as it contains a fairly exhaustive set of shoes among its product listings and has images with fairly consistent lighting and positioning for all of its products. Along with each of these images, we have scraped information about each of the associated products to use to verify that our eventual classifications are correct. The information that we have stored includes color, brand, style, and price.

**Analysis**

We have stated our analysis by trying to classify two features of all of the images in our dataset: color and height. We believe that these are critical features to assess since color is an important aspect to consider when recommending similar items and height can be used to help classify the style of a shoe in an image. We describe our processes for both of these classifications below:

Color

We classify the colors of the shoes by decomposing each image into a matrix of pixels mapped to their rgb values. We then utilize a modified form of k-means clustering to identify the most prevalent color clusters in a given image. Subsequently, we ascertain the overall color of the shoe by weighting each cluster center by the total number of points that it includes, and we then map the most prevalent color cluster(s) in each image to the color range(s) that they fall into to compute the color of the shoe. Thus far, our clustering process has successfully generated the colors of all of the items that we have analyzed.

Height

We ascertain the heights of shoes by computing the distance between the top of the shoe and the bottom of the shoe at the rear end of the shoe. The way that we do this is we calculate the highest and lowest non-white pixels at the rear end of the shoe in the image, and we set the height of the shoe equal to the Euclidean distance between these two pixels’ coordinates. Since all of our images are standardized and are scaled and positioned in the same manner on the same white backdrop, we are able to use the resultant height in pixels as a feature to compute the style of the shoe without any further modification.

Now that we are able to classify the color and height of shoes, we are starting to determine other features that can be extracted from these images, such as logo, width, presence of shoelaces, shoelace orientation, texture of the sides of shoes, and texture of the soles of shoes to complete our feature collection process. Once we have finished aggregating features, we will utilize Naïve Bayes, Support Vector Machines, and Neural Networks to calculate the recommendations on the basis of these features.

**Next Steps**

1. Ascertain features that we wish to assess for the purposes of classifying the images (e.g., color(s), style(s), brand, price, etc.)
2. Create a classification algorithm that outputs said features – our initial plan is to train a layered model of neural nets (i.e. Deep Learning) to recognize features in the images and then classify them as described above
3. Train the algorithm on our training set of images
4. Iterate upon and optimize the algorithm
5. Leverage the classification algorithm to calculate similarities between goods
6. Output similarity recommendations
7. Try to apply this methodology to other retail goods