Data Driven Ad Insights Using Computer Vision and Natural Language Processing Techniques

Madmen AI Takehome Assignment

Gautham Gorti
gauthamgorti@gmail.com

1 Introduction

We wish to design an automated system that can identify trends over large swaths of creatives. We are provided a dataset consisting of advertisements from a collection of grooming companies – each ad consists of an image and various accompanying texts. Using ideas from Computer Vision, Natural Language Processing, Statistics, and Machine Learning, we present a data driven approach to identify both widespread and novel marketing strategies for grooming products.

2 Featurization

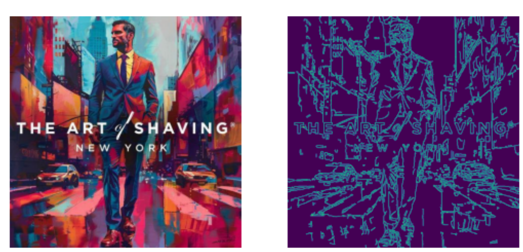
Before we can apply machine learning or statistical analysis, we must first extract quantitative features from each ad – a process known as featurization.

2.1 Bag of Words

Each ad has several important text fields and may even contain text in the accompanying image. To featurize this data, we collect all text from the *body, cta_text, title* and *link_description* fields. We also collect the text from the image using an off-the-shelf Optical Character Recognition model (Pytesseract). Then, we encode the text using a Bag of Words (BOW) scheme – the number of appearances of each word is counted and then stored as a new feature column.

BOW is simple but very effective when analyzing short text data. Other more powerful encoding schemes exist involving the use of Deep Neural Networks, such as the kind used by LLMs [1]. However, these methods are not appropriate for our project for several reasons. First, they require vast amounts of compute resources outside of the scope of this assignment. Second, they are black-box methods that yield features that are generally uninterpretable by the human data scientist, and therefore cannot provide meaningful insights. Thus, BOW is the most reasonable approach for featurizing the text data.

(a) Simple visual design with mean edge detection of 11.3



(b) Complex visual design with mean edge detection of 28.0.

Figure 1

2.2 Edge Detection

We wish to quantify some of the design aspects of the provided images. To do so, we use Canny Edge Detection [2]. The Canny Edge Detector applies a sequence of human-designed convolutional filters to detect sharp lines and shapes in images. Once the edge detector is applied to the whole image, the mean detection of the image should indicate how busy or complicated the design is – higher detection scores indicate ads with many shapes and objects that draw attention to the human eye, while low detection scores indicate simpler ad designs. We provide an illustrative example in figure 1.

2.3 Color Clustering Variance

We also wish to quantify how the ads are colored. We use pixel clustering to measure the overall color diversity of each ad image. To featurize the ad image, we apply k-means clustering (k = 4) to the RBG pixel color values. We sum the within-cluster variances to generate a single score that indicates color diversity of the image. The idea is that ads with less diverse coloring can be accurately summarized by the clustering and have low cluster variances, but very colorful ads cannot be and thus have higher cluster variances. We provide some examples in figure 2.

(a) Simple coloring can be reconstructed with four colors Color clusters variance of 5159.9

(b) The image has too many colors to be accurately reconstructed. Color clusters variance of 11712.8

Figure 2

2.4 Face Detection

Some companies use human models or faces in their advertisements. To detect this, we feed each ad image into an off-the-shelf face detection network [cite] and then store a boolean flag to indicate whether or not a human face is present in a given ad image.

3 Analysis

Once the features are generated, we analyze them using techniques from statistics and machine learning to yield insights.

3.1 Company Posterior Probabilities

Let C and W denote the set of all companies in the dataset and words in the vocabulary, respectively. Given $c \in C$ and $w \in W$, we can use the BOW data to estimate $p(c \mid w)$, the posterior probability of a word w having been sampled from company c, given that word. The calculation:

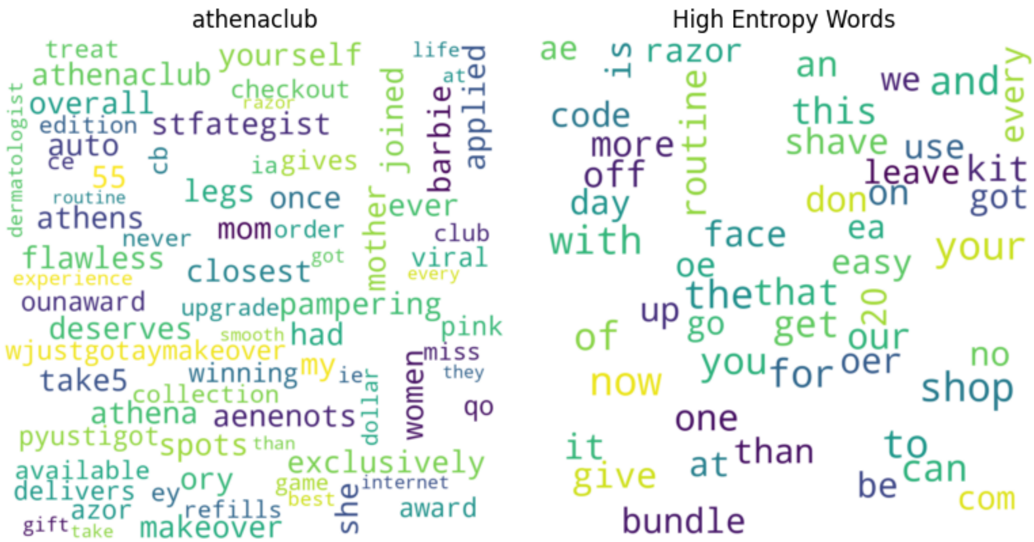
$$\begin{aligned} p(c \mid w) &= \frac{p(c)p(w \mid c)}{\sum_{c' \in C} p(c')p(w \mid c')} && \text{using bayes rule} \\ &= \frac{p(w \mid c)}{\sum_{c' \in C} p(w \mid c')} && \text{assuming uniform prior,} \end{aligned}$$

where

$$p(w \mid c) = \frac{\text{word count of } w \text{ in all ads from company } c}{\text{total number of words in all ads from company } c}.$$

High values of $p(c \mid w)$ imply that word w strongly indicates that it was sampled from an ad from company c. Thus, we can reveal novel strategies by observing words that yield the highest posterior probabilities for a specific company. A word cloud visualization of high-posterior-probability words for athenaclub is presented in figure 3a.

We can also use the posterior probabilities to study widespread trends. For each word, we can calculate the entropy of its posterior distribution over the companies – high entropy indicates that the word's usage is widespread throughout several companies. In figure 3b, we provide a word cloud consisting of high entropy words to reveal widespread trends.



(a) High-posterior-probability words for athenaclub (b) High entropy words over all companies

Figure 3: Word clouds using calculated posterior probabilities.

3.2 Image Feature Visualization

We have two continuous features that describe characteristics of the ad image: edge detection score, and color cluster variance. We plot each ad by these two features on a 2d plane and color-label by company. We also plot a fitted 2d gaussian random variable for each company. Simply examining the plots yields widespread insight.

We can also select only the ads that use human faces and generate the same plots. This reveals novel insights about the few companies that prominently use human models in their ads, and how they use these human models in separate ways to implement their unique marketing strategies.

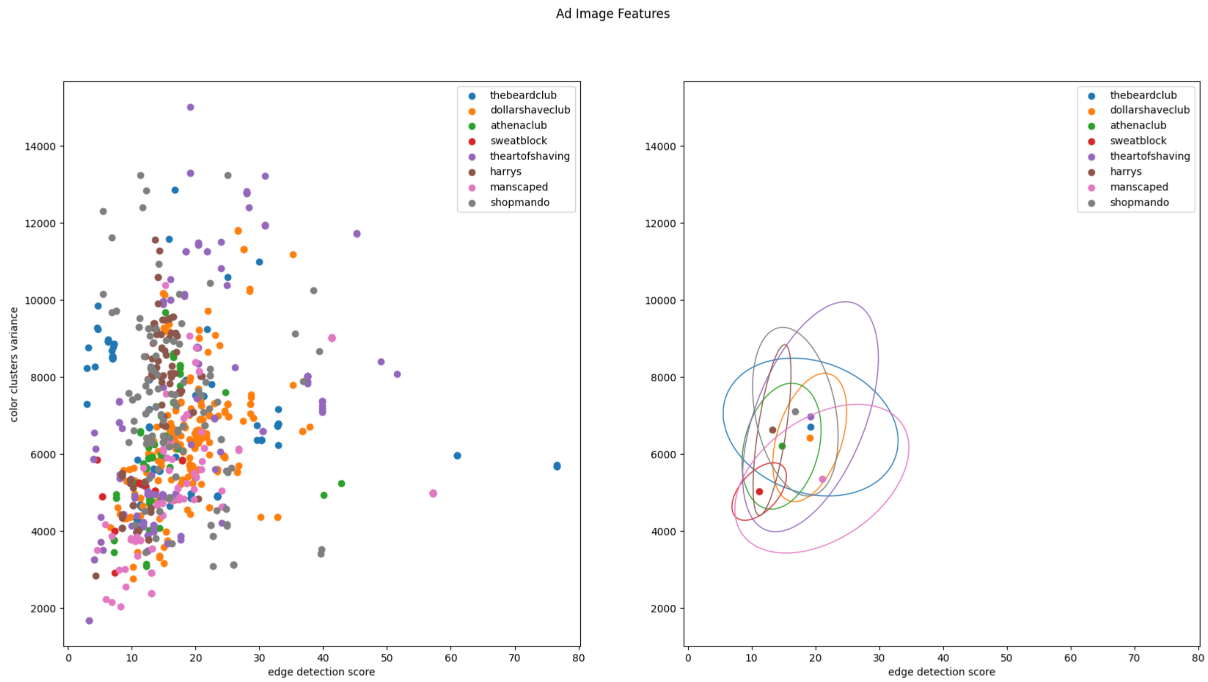


Figure 4: Example visualization of ad image features.

References

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.

[2] John Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986.