Understanding Image Virality

Xuewen Yang

June 18 2018

Abstract

Virality of online content on social networking websites is an important but esoteric phenomenon often studied in fields like marketing, psychology and data mining. In this paper we study viral images from a computer vision perspective. We introduce three new image datasets from Reddit1 and define a virality score using Reddit metadata. We train classifiers with state-of-the-art image features to predict virality of individual images, relative virality in pairs of images, and the dominant topic of a viral image. We also compare machine performance to human performance on these tasks. We find that computers perform poorly with low level features, and high level information is critical for predicting virality. We encode semantic information through relative attributes. We identify the 5 key visual attributes that correlate with virality. We create an attribute-based characterization of images that can predict relative virality with 68:10% accuracy (SVM+Deep Relative Attributes) Cbetter than humans at 60:12%. Finally, we study how human prediction of image virality varies with different contexts in which the images are viewed, such as the influence of neighbouring images, images recently viewed, as well as the image title or caption. This work is a first step in understanding the complex but important phenomenon of image virality. Our datasets and annotations will be made publicly available.

1. Introduction

What graphic should I use to make a new startup more eye-catching than Instagram? Which image caption will help spread an under-represented shocking news? Should I put an image of a cat in my YouTube video if I want millions of views? These questions plague professionals and regular internet users on a daily basis. Impact of advertisements, marketing strategies, political campaigns, non-profit organizations, social causes, authors and photographers, to name a few, hinges on their ability to reach and be noticed by a large number of people. Understanding what makes content viral has thus been studied extensively by marketing researchers [4].

Many factors such as the time of day and day of week when the image was uploaded, the title used with the image, etc. affect whether an image goes viral or not. To what extent is virality dependent on these external factors, and how much of the virality depends on the image content itself? How well can state-of-the-art computer vision image features and humans predict virality? Which visual attributes correlate with image virality?

In this paper, we address these questions. We introduce three image databases collected from Reddit and a virality score. Our work identifies several interesting directions for deeper investigation where computer vision techniques can be brought to bear on this complex problem of understanding and predicting image virality.

2. Related Work

Most existing works [1] study how people share content on social networking sites after it has been posted. They use the network dynamics soon after the content has been posted to detect an oncoming snowballing effect and predict whether the content will go viral or not. We argue that predicting virality after the content has already been posted is too late in some applications. It is not feasible for graphics designers to try out various designs to see if they become viral or not. In this paper, we are interested in understanding the relations between the content itself (even before it is posted online) and its potential to be viral.

There exist several qualitative theories of the kinds of content that are likely to go viral [2]. Only a few works have quantitatively analyzed content, for instance Tweets and New York Times articles [3] to predict their virality. However, in spite of them being a large part of our online experience, the connections between content in visual media and their virality has not been analyzed. This forms the focus of our work.

Virality of text data such as Tweets has been studied in [5]. The diffusion properties were found to be dependent on their content and features like embedded URLs and hashtags. Generally, diffusion of content over networks has been studied more than the causes. The work of Leskovec *et*

(a) WTF (b) atheism

Figure 1. Examples of temporal contextual priming through blurring in viral images. Looking at the images on the left in both (a) and (b), what do you think the actual images depict? Did your expectations of the images turn out to be accurate?

al. models propagation of recommendations over a network of individuals through a stochastic model, while Beutel *et al.* approach viral diffusion as an epidemiological problem.

3. Understanding Image Virality

Consider the viral images of Figure 1, where face swapping, contextual priming, and scene gist make the images quite different from what we might expect at a first glance. An analogous scenario researched in NLP is understanding the semantics of Thats what she said! jokes. We hypothesize that perhaps images that do not present such a visual challenge or contradiction C where semantic perception of an image does not change significantly on closer examination of the image C are boring [3] and less likely to be viral. This contradiction need not stem from the objects or attributes within the image, but may also rise from the context of the image: be it the images surrounding an image, or the images viewed before the image, or the title of the image, and so on. Perhaps an interplay between these different contexts and resultant inconsistent interpretations of the image is necessary to simulate a visual double entendre leading to image virality. With this in mind, we define four forms of context that we will study to explore image virality.(1)Intrinsic context: This refers to visual content that is intrinsic to the pixels of the image.(2) Vicinity context: This refers to the visual content of images surrounding the image (spatial vicinity).(3) Temporal context: This refers to the visual content of images seen before the image (temporal vicinity).(4) Textual context: This non-visual context refers to the title or caption of the image. These titles can sometimes manifest themselves as visual content (e.g. if it is photoshopped). A word graffiti has both textual and intrinsic context, and will require NLP and Computer Vision for understanding.

References

[1] A. Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 2005. 1

[2] J. Berger. Arousal increases social transmission of information. *Psychological Science*, 2011. 1

[3] J. Berger and K. L. Milkman. What makes online content viral. *Journal of Marketing Research*, 2012. 1, 2

[4] J. Berger and E. M. Schwartz. What drives immediate and ongoing word of mouth? *Journal of Marketing Research*, 2011. 1

[5] M. Nagarajan, H. Purohit, and A. P. Sheth. A qualitative examination of topical tweet and retweet practices. In *ICWSM*, 2010. 1