Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017)

From Camera to Deathbed: Understanding Dangerous Selfies on Social Media

Hemank Lamba,[1] **Varun Bharadhwaj,**[3] **Mayank Vachher,**[2] **Divyansh Agarwal,**[2]
Megha Arora,[1] **Niharika Sachdeva,**[2] **Ponnurangam Kumaraguru**[2]
[1]Carnegie Mellon University, USA, [2]Indraprastha Institute of Information Technology, Delhi, India
{hlamba@cs,marora@andrew}.cmu.edu, {mayank13059,divyansha,niharikas,pk}@iiitd.ac.in
[3]National Institute of Technology, Tiruchirappalli
var6595@gmail.com

Abstract

Selfie culture has emerged as a ubiquitous instrument for self-portrayal in recent years. To portray themselves differently and attractive to others, individuals may risk their life by clicking selfies in dangerous situations. Consequently, selfies have claimed 137 lives around the world since March 2014 until December 2016. In this work, we perform a comprehensive analysis of the reported selfie-casualties and note various reasons behind these deaths. We perform an in-depth analysis of such selfies posted on social media to identify dangerous selfies and explore a series of statistical models to predict dangerous posts. We find that our multimodal classifier using combination of text-based, image-based and location-based features performs the best in spotting dangerous selfies. Our classifier is trained on 6K annotated selfies collected on Twitter and gives 82% accuracy for identifying whether a selfie posted on Twitter is dangerous or not.

Introduction

A selfie is defined as *a photograph that one has taken of oneself, typically taken with a smartphone or webcam and shared via social media* (Taslim and Rezwan 2013). In 2015, Google estimated that 24 billion selfies were uploaded to Google Photos [1] and the number of selfies posted on Instagram increased by 900 times between 2012 and 2014 (Souza et al. 2015). Selfie, nowadays has become a ubiquitous tool for self-representation on social media. However, in some cases, selfie culture may promote dangerous behavior posing significant moral, mental and physical health implications on the individuals clicking selfies (hereafter referred as "selfie-er")(Adamkolo and Elmi-Nur 2015). Users click multiple selfies and post on social media aesthetically altered versions that make them look more attractive (Marwick 2015). In extreme cases, users engage in behaviors that portray them to be adventurous or enhance their appearance to others while risking their own physical well-being (Leary 1994). As many as 137 individuals have been reported to be killed since 2014 till December 2016 while attempting to take selfies. Considering the hazardous implications of taking selfies, Russian authorities published public posters, indicating the dangers of taking selfies,[2]and Indian authorities including Mumbai police and Indian Railways issued warning for taking selfies at dangerous locations.[3]

Despite the increase in incidents where selfies were the reason behind physical harm caused to individuals, few research works explore factors that may result into dangerous selfies. Studies have indicated clicking selfies at dangerous locations as one of the reasons for selfie-related casualties (Bhogesha, John, and Tripathy 2016). Social media has emerged as a powerful medium to share and gain attention through such dangerous selfies (Souza et al. 2015). Given the popularity of the selfie culture, and increasing number of selfie deaths, it is crucial to characterize and predict the behavior of taking/posting dangerous selfies on social media. However, this remains largely unexplored. In this work, we try to identify features that can be derived from selfies posted on social media to predict dangerous selfies.

We formulate our specific research goals as:

1. Analyze incidents associated with reported fatal selfie casualties to understand the reasons behind the deaths and characterize such selfie-ers.
2. Investigate the content measures derived from social media that are predictive of dangerous selfies.

In this paper, we first comprehensively analyze the deaths that have occurred due to the victims trying to take selfies. We dig up all the news articles related to selfie deaths from credible news sources and manually annotate them to understand the causes of such deaths. We further propose multimodal features including text, images and other meta-data available on social media, based on the attribute analysis of the reported fatal selfie casualties. These features are leveraged to build a model, allowing us to identify dangerous selfies posted on social media.

Selfie Deaths Characterization

In this section, we discuss the robust method adopted to identify and characterize reported selfie-casualties.

Identifying selfie-casualties: We collected every news article that reported any selfie casualty. We used a keyword

[1]https://googleblog.blogspot.in/2016/05/google-photos-one-year-200-million.html

[2]https://www.theguardian.com/world/2015/jul/07/a-selfie-with-a-weapon-kills-russia-launches-safe-selfie-campaign

[3]http://metro.co.uk/2016/02/25/mumbai-orders-selfie-ban-after-19-people-die-5716731/

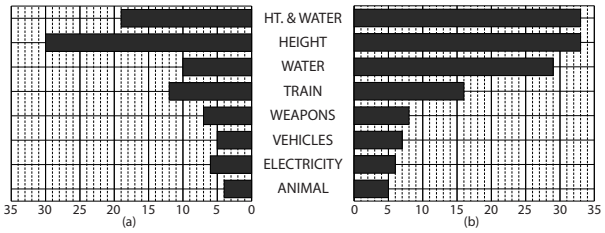


Figure 1: (a) Number of incidents and (b) Number of deaths due to various reasons. "HT" is height.

Reason	Number of Dangerous Selfies
Water Related	297
Vehicle Related	149
Height Related	135
Height & Water Related	105
Road Related	30
Animal Related	28
Train Related	11
Weapons Related	4
Electricity Related	0

Table 1: Reasons marked by annotators for a selfie being dangerous.

based extensive web searching mechanism to identify these articles. Further, we considered only those articles as credible which were hosted on websites having Global Alexa ranking less than 5,000, or a country specific Alexa rank less than 1,500. The earliest article reporting a selfie death that we were able to collect was published in March 2014. Using this method, we found 137 selfie-related casualties from 93 incidents reported since March 2014 until December 2016.

Characterizing selfie-casualties: Two annotators manually annotated the articles to identify the reason for death, demographics of the victims, and the location where the selfie casualty occurred. Overall, we were able to find 8 unique reasons behind the deaths. Figure 1 shows the number of casualties for various reasons of selfie deaths. We observed that the most common reason for selfie casualties was clicking selfie at an elevated region. These involved, people falling off buildings or mountains while trying to take a selfie. The least number of casualties was caused while clicking selfies with animals.[4]

Based on our understanding from the reported selfie-casualties, in our work, we define a selfie-related casualty vis a vis dangerous selfie as *a casualty to an individual or a group of people that may occur while the individual(s) attempts to take a selfie*.

Selfie Dataset Curation

For data collection process, we chose Twitter, as it is a popular social media platform observing the selfie culture. We carefully collected tweets from Twitter related to selfies by searching words like *selfie* or its immediate variants (#myselfie, #selfie). The data was collected between August 1, 2016 and September 27th, 2016. Through this method, we obtained 138K unique tweets posted by 78K individual users. After filtering out the tweets that did not contain any images, we obtained 91,059 tweets containing images. We further filtered geo-located tweets from 91,059 tweets that contained images to obtain 9,444 geocoded tweets.

Preprocessing and Manual Annotation: We want to distinguish selfie images from non-selfie images in our dataset. We used CNN model Inception-v3 (Szegedy 2015) for this. For training the model, we used 2.1K randomly chosen images from our dataset and manually annotated them into 1.3K selfies and 800 non-selfies. Using this classifier, we obtained labels for all the 9,444 geocoded images. This process yielded a candidate set of 6,842 tweets which were potential selfie tweets.

The final step for identifying dangerous selfies involved human annotations on the selfie candidate set of 6,842 tweets. For the purpose of annotation, we developed a web interface and provided each annotator with an authentication login. We recruited annotators via posting a request for participation on the mailing list of different universities.

Each selfie was annotated by 3 distinct annotators. The inter-annotator agreement rate, using the Fleiss Kappa metricwas 0.58, thus indicating moderate agreement between the annotators. We used majority voting to decide the final label for a given selfie. We found that from the selfie candidate set of 6,842 tweets, our annotators agreed that 6,460 tweets contained selfies. Among these 623 were dangerous selfie tweets, corresponding to 579 users with 547 users posting one dangerous selfie and one user posting 5 dangerous selfies. We conduct all our future analysis on this set of 6,460 tweets.

On analyzing annotators' possible reason to mark a selfie as dangerous, we found water-related reason to be the most common, followed by vehicle-related (see Table 1).

Identifying Dangerous Selfies

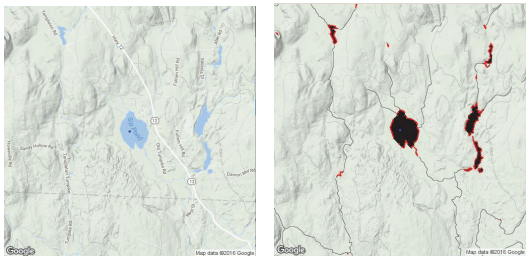
Based on the analysis of selfie casualties we did in the previous section, we operationalize different features (explained below) which can potentially help us to design models for identifying dangerous selfies.

Location based Features: From our dataset of reported selfie casualties from news sources, we observed 33 selfie casualties occurred because of selfie-ers falling from an elevated location. Thus, we believe that the elevation of the terrain around the location might be indicative of whether the given selfie is dangerous or not. We used Google elevation API to estimate the elevation of the location.[5] We define the neighborhood as K sampled points within a radius of r metres around the location of the selfie. We generated the following features: (a) Elevation of the exact location of the selfie, (b) maximum elevation in the neighborhood, (c) maximum difference in elevation of the selfie-location and sampled points in the neighborhood and (d) the maximum ele-

[4]More statistics are available at http://labs.precog.iiitd.edu.in/killfie/analysis

[5]https://developers.google.com/maps/documentation/elevation/

(a)



(b) (c)

Figure 2: Segmentation Example: Different Stages of processing to get the final segmented image distinguishing between the water and land.

vation difference between any 2 points in the neighborhood. For sampling locations, the choice of radius r and number of locations K was made on the basis of the lowest p-value of 2-sampled Kolmogorov-Smirnov (KS) test on the specified feature between dangerous and non-dangerous samples. To evaluate the efficiency (or the discriminative power) of these features, we computed the KS test for all the above features between dangerous and non-dangerous points. We obtained p-value < 0.01 for all, except for the elevation feature.

The second highest number of casualties were related to drowning in water. Therefore, we used the location of the selfie-er to determine how far he/she is from a water body when clicking a selfie. Consider the selfie in Figure 2(a) which was taken in the middle of a water body. We mapped the exact location of the selfie as obtained from the geo-tagged tweet to Google Maps and considered 500×500 pixel image pertaining to level 13 zoom factor on Google Maps. The image after this step looked like Figure 2(b). We applied image segmentation to identify the contour of all the water bodies as shown in Figure 2(c). We used two water related features - (a) minimum distance to a water body from the location of the image and (b) fraction of the pixels in the segmented image. We observed that for both the water features, the distribution of water-related dangerous and non-dangerous selfies was considerably different (p-value < 0.01). This indicates that the features can potentially help distinguish between dangerous and non-dangerous selfies. Besides the features mentioned above, we also took into account other location-based features such as distance from train/railway tracks, and distance from a major highway.

Image-based Features: Classifying an image as dangerous or not requires extensive understanding of the context and the elements in the image. Therefore, we first extracted the salient regions in images and then generated captions for each of those regions. Based on these captions, an understanding of the context and the elements in the image can be formed which can then be used to identify dangerous selfies. To extract informative regions in images and for the caption-generating process, we used DenseCap (Johnson, Karpathy, and Fei-Fei 2016). DenseCap is start-of-the-art deep learning based captioning technique for regions in an image. It outperforms other models such as Full Image RNN (Recurrent Neural Network) and Region RNN on both the tasks: dense captioning and as well as image retrieval. An example of the output of the DenseCap on a selfie in our dataset is shown in Figure 3. We treated the generated captions as text describing the image in natural language. From the text, we computed natural language features such as unigrams and bigrams to determine if the content of the image was dangerous or not.

Text-based Features: The content of the tweet can be a useful source for indicating if the image accompanying it is a dangerous selfie. Users tend to provide context to the image either directly in the tweet text or through hashtags. We used both (tweet text and hashtags) to generate our text-based features. After pre-processing the text, we used TF-IDF over the set of unigrams and bigrams as features.

Classifier

Considering the annotations performed in the section above as ground truth, we evaluated the performance of our classifier on the task of classifying a selfie as dangerous. The problem of classifying dangerous selfies is a highly unbalanced problem. We had only 623 (roughly 9%) dangerous selfies in comparison to the remaining 5,837 non-dangerous selfies. Imbalance in annotated data is a common problem in many machine learning applications. In these cases, applying a classifier on the data as is, leads to a classification algorithm to simply predict the majority class label for all the samples. To avoid this, many methods have been proposed in the literature for balancing such data sets (He and Garcia 2009). For our task, we experimented with random down-sampling (randomly removing samples from majority class).

As mentioned earlier, our feature space can be easily divided into 3 categories - text, image, and location-based. To compare all feature types, we built and tested the classifiers for every possible combination of the features. For all our experiments, we performed 10-fold cross validation. Furthermore, we used grid search to find the ideal set of hyperparameters for each classifier by doing 3-fold cross-validation on the training set. We tested the performance of our method using 4 different classification algorithms - Random Forests, Nearest Neighbors, SVM and Decision Trees. Each of the classifiers was trained and tested on the similar dataset using the same feature configuration. Figure 4 shows the ROC curves by using various classification techniques over different combinations of our feature space.

Multimodal features are important: Based on the re-

Figure 3: An example of the DenseCap on one of the images (Left) from our dataset. We use the captions produced by DenseCap (Right) to come up with text based features.

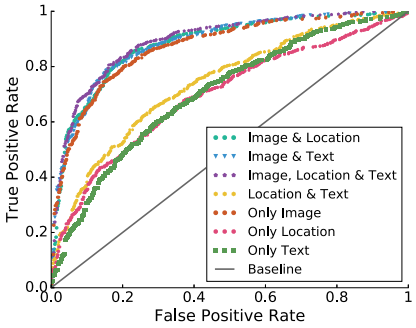


Figure 4: Receiver Operating Characteristic (ROC) curves corresponding to the statistical models for identifying dangerous selfies. "Dangerous" selfie is the positive class.

sults shown in Figure 4, we can observe that when all the classes of features are used, the accuracy is the highest. This validates our approach of using multimodal features. It can also be seen that the combination of image and text features perform better than the image and location features. This might indicate that the context and content of the selfies are far better predictors than the location of the selfie.

Image features perform well: Further analyzing the results, we can clearly see that image-based features performed the best out of all the classes of features. Therefore, even in the absence of location of a selfie, a model based only on the image based features can perform relatively well in finding dangerous selfies. This can be helpful in cases, where the user's post is not geocoded, or in an application case when location information is not available due to GPS being turned off or unavailable.

Discussion

In this paper, we create a novel dataset of reported selfie casualties to describe the subtleties of the situations where such accidents may occur. Our work demonstrates the viability of using selfies and content posted on Twitter as an instrument to quantify and characterize *dangerous selfies* that may cause casualty to selfie-ers. Further, we present a multimodal classifier that uses various features such as - text-, image-, and location-based features to identify dangerous selfies. In this work, we demonstrate that measuring the multimodal subtleties (image, text, and location) of selfie tweets available on social media can help to identify physical harm possibilities to selfie-ers. We show that location-based features can be customized to detect the common reasons such as water-related, height-related factors pertaining to selfie casualties. We adopt state of the art deep learning techniques such as DenseCap to determine the content of the selfie. The approach demonstrated in our work, suggests that even in absence of one or more of the above mentioned features, technologies can be developed to identify dangerous selfies. We believe that there is an opportunity to extend our approach for identifying selfie-ers who are at high risk of selfie-related casualties.

Limitations: Our work explores a set of Twitter users, who are explicit about sharing selfies and mention hashtags such as #selfies and #myselfie in their posts. However, we acknowledge that these users may not be representative of the entire Twitter or general social media population. There could be a section of users who may not be explicit about sharing selfies using hashtags or keywords. We also acknowledge, that there may be a section of selfie-ers who may not be sharing their selfies on social media. There might be an inherent selection bias towards selfie-ers who prefer to use Twitter as a platform to share selfies.

References

Adamkolo, M., and Elmi-Nur, H. 2015. Communicating 'the self' through digital images: Gender bias and mental health risks associated with selfie use on social network sites. *Global Media Journal*.

Bhogesha, S.; John, J. R.; and Tripathy, S. 2016. Death in a flash: selfie and the lack of self-awareness. *Journal of Travel Medicine*.

He, H., and Garcia, A., E. 2009. Learning from imbalacned data. *IEEE TKDE*.

Johnson, J.; Karpathy, A.; and Fei-Fei, L. 2016. Densecap: Fully convolutional localization networks for dense captioning. In *CVPR*.

Leary, M. R. e. a. 1994. Self-presentation can be hazardous to your health: impression management and health risk. *Health Psychology*.

Marwick, A. E. 2015. Instafame: Luxury selfies in the attention economy. *Public Culture*.

Souza, F.; de Las Casas, D.; Flores, V.; Youn, S.; Cha, M.; Quercia, D.; and Almeida, V. 2015. Dawn of the selfie era: The whos, wheres, and hows of selfies on instagram. In *COSN*. ACM.

Szegedy, C. e. a. 2015. Rethinking the inception architecture for computer vision. *CoRR* abs/1512.00567.

Taslim, I., and Rezwan, M. Z. 2013. Selfie re-de-fined: Self-(more/less). *Wizcraft Journal of Language and Literature*.