6th International Workshop on Climate Informatics
September 22-23, 2016
CI 2016
Hosted by the National Center for Atmospheric Research in Boulder, CO

DETECTING AND PREDICTING BEAUTIFUL SUNSETS USING SOCIAL MEDIA DATA

Emma Pierson[1]

***Abstract*—Beautiful sunsets are one of the few pleasures freely available to everyone, but due to their subjective nature are difficult to quantitatively study on a large scale. Here, we use 1.2 million sunset posts on Instagram, a picture-sharing platform, to detect beautiful sunsets in 10 American cities over 7 months. We show that our metric of sunset quality correlates with human assessments, make sunset quality scores publicly available to allow more systematic study of sunsets, and use this dataset to answer a number of basic questions. Do some locations have more beautiful sunsets than others? Are there meteorological features which predict beautiful sunsets? Does a beautiful sunset today predict a beautiful sunset tomorrow? Is it possible to detect beautiful sunsets early enough to notify people so they can go outside to enjoy them? What visual features are people responding to when they call a sunset beautiful? We validate a widely used sunset prediction model developed by meteorologists, produce an algorithm which can visually discriminate between beautiful and mediocre sunsets, and provide a messaging service and web interface to notify users of beautiful sunsets.**

I. MOTIVATION

If beautiful sunsets could be reliably predicted, or detected in real time, millions of people could enjoy a free daily light show. In spite of this, beautiful sunsets have long been more the realm of artists and poets than of scientists; large-scale, quantitative data is not readily available, and previous research is sparse and based on small datasets or general application of physical laws [1], [2], [3] rather than analysis of large-scale datasets. Our contribution in this paper is two-fold: we use the image-sharing website Instagram to collect and make publicly available what is to our knowledge the first large-scale dataset on beautiful sunsets, comprising sunset quality scores in 10 large American cities over 7 months; second, we use this dataset to derive principles of beautiful sunset prediction and detection.

Corresponding author: Emma Pierson, epierson@cs.stanford.edu
[1]Department of Computer Science, Stanford University

II. METHOD

A. Dataset

Instagram is a social media platform with more than 300 million daily users on which people can post images with tags to describe image content (eg, "beautiful #sunset"). From October 2015 - May 2016 we used the Instagram Search API [4] to collect public Instagram posts tagged with latitude and longitude with tags relating to sunset: eg "#sunset" and "#instasunset". To mitigate cultural differences in Instagram usage, we confined our analysis to posts within the United States. In total we collected 1.2 million posts, each with a location and sunset picture. Because reliably detecting spikes in Instagram activity requires a large number of posts, we focused our analysis on 10 large American cities: Los Angeles, New York, Boston, Chicago, Washington DC, Miami, San Diego, Seattle, Philadelphia, and San Francisco. For each city c, we computed the total number of Instagram posts n_{cd} within 0.5 degrees of the city on each day d.

Although it is intuitive that more people will take sunset photographs on days with beautiful sunsets, we performed three additional tests to validate our metric. First, we confirmed by hand inspection that images collected under these tags were sunset-related. Second, we confirmed that sunset posting activity does spike dramatically at sunset (Fig. 1A), implying that people are in fact reacting to local conditions as opposed to posting previously taken pictures. Third, we had three research assistants hand-code sunset quality in five cities in our dataset (Los Angeles, Miami, New York, Seattle, and Boston) as follows. For each city, we randomly selected 10 sunset pictures from the five days with the highest n_{cd}, and 10 pictures from days near the median n_{cd}; we refer to these below as "beautiful" and "average" sunsets. We presented the hand coders with one pair of sunset pictures at a time – one beautiful sunset, and one average sunset – in random order, and had them choose the one they felt had the more beautiful sunset. The three hand-coders agreed 86% of the time on average, indicating consensus about what constituted a beautiful sunset. Hand-coders overwhelmingly preferred sunsets from days with high n_{cd} to days with median n_{cd}: in cases where all three hand-coders agreed, they preferred the high n_{cd} sunset 80% of the time, and on average individual hand-coders preferred the high n_{cd} sunset 75% of the time. (We suspect that, in fact, the gap between the high n_{cd} and median n_{cd} sunsets is even more dramatic; low-quality pictures on high n_{cd} days appeared to be due to lack of photographer skill, not sunset quality. We provide the pictures used in the test in the Supplementary Information[1] (SI).) While assessing sunset quality is a gloriously subjective enterprise which we invite our readers to partake in, these quantitative checks confirm that n_{cd} do offer the first large-scale scores of sunset quality which correlate with human assessments.

We further increased the utility of our sunset scores by controlling time-of-year effects and differences in city size using the following procedure:

1) To account for differences in city size, we divided each city's daily count of sunset posts by the mean number of sunset posts in that city, so each city had mean $n_{cd} = 1$.
2) Social media usage varies by weekday and time of day: a sunset that occurs at 8 PM on Sunday may get more posts than one at 4 PM on Tuesday even if both are of the same quality. To account for this, we ran a linear regression where the dependent variable was n_{cd} and the covariates were a categorical indicator variable for each weekday and a second-order polynomial in sunset time. (We used a second-order polynomial because it captured the fact that the number of posts may increase or decrease non-linearly in sunset time, but does not provide too many degrees of freedom, which could produce overfitting). We set s_{cd}, the normalized sunset score, equal to the residual: ie, the variation in sunset quality which time effects did not account for. (We accounted for city size effects separately, rather than in our regression, because they are likely multiplicative rather than additive.)
3) Finally, we computed a binary "beautiful sunset" indicator variable b_{cd} which was 1 if a sunset had a score s_{cd} in the top 15% for city c and 0 otherwise. We focus our analysis on b_{cd} rather than s_{cd} because it is not that important to discriminate between sunsets at the 20th and 40th percentiles (neither is worth going outside for); the goal of interest is to find beautiful sunsets.

Features that correlate with b_{cd} predict that a city is unusually likely to have a sunset in the top 15% when controlling for time of sunset. While this metric is interpretable and relevant, future work should also investigate whether some cities have better sunsets than others. To facilitate future analysis, we make the dataset of our sunset quality scores publicly available (SI).

[1] All supplementary information is available at http://cs.stanford.edu/~emmap1/sunsets/supplementary_information.zip

B. Beautiful sunset prediction

As a first illustration of the utility of our dataset, we evaluated the accuracy of a widely used sunset prediction algorithm. SunsetWx [5], which predicts sunset quality using the 4km NAM, is used by more than 20 TV stations across the country to forecast beautiful sunsets, but its developers have had to rely on anecdotal reports of sunset quality as validation [6]. They provided early access to their API, and we compared their predicted sunset quality to our sunset quality scores b_{cd} on 58 days across 10 cities (a total of 580 datapoints).

SunsetWx computes real-valued sunset quality predictions which it stratifies into "Poor", "Fair", "Good", and "Great"; very few sunsets are predicted to be "Great", so we exclude them from our analysis. We found that SunsetWx's stratified scores have predictive value. $b_{cd} = 1$ for 6% of sunsets predicted to be "Poor"; 15% of sunsets predicted to be "Fair", and 31% of sunsets predicted to be "Good".

We suspect, however, that predictions from SunsetWx could be improved if a systematic machine learning approach was used to develop a predictive algorithm, since even very simple predictors exhibit comparable accuracy. For example, using humidity as a univariate predictor of b_{cd} and running a logistic regression yields an out-of-sample AUC of 0.72, which is better than the SunsetWx AUC of 0.67. (AUC [7], or area under the curve, is a standard measure for assessing performance on a binary classification task – for example, discriminating between beautiful and average sunsets – and is defined as the probability that the model assigns a higher score to a positive example than to negative example. Higher AUCs denote better performance.) We thus believe there is room for improvement, and it is worth noting that SunsetWx's prediction algorithm has evolved since we collected the data for this paper. We believe that the most successful prediction algorithms will use a large number of sunset quality scores, such as those we have collected here, to train and assess a machine learning algorithm. Prediction algorithms which are tuned and developed using anecdotal reports of good sunsets are unlikely to yield optimal performance.

While developing a prediction algorithm that fully uses meterological data is a topic for future work, we evaluated the predictive utility of 20 weather features using meteorological data from ForecastIO [8] and pollution data from the EPA [9]. Days with temperatures above 60 degrees are statistically significantly more likely to have $b_{cd} = 1$; so are days with lower cloud cover (with zero cloud-cover days having the highest probability), lower humidity, and calm winds (for detailed graphs, see SI). Interestingly, we also found that higher levels of pollution correlate positively with b_{cd}, with higher levels of carbon monoxide and nitrogen dioxide showing statistically significant correlations ($p < .01$); (we also assessed ozone and sulfur dioxide, which showed positive but not statistically significant correlations). This provides some evidence in the debate between those who claim that clear skies produce more brilliant sunsets and those who disagree [3], [10] although it is worth noting that our analysis is confined to cities and rural, unpolluted areas may have better sunsets in general, even if cities have better sunsets on days with more pollution.

We also find that a good predictor of today's sunset quality is *yesterday's* sunset quality. If yesterday's sunset was beautiful, the probability that today's sunset will be beautiful is 34%, as compared to 12% if yesterday's was not beautiful. We confirm that this is correlation is primarily due to short-term effects, not to longer-term time of year effects; the predictive value is much reduced if a week's separation rather than a day's separation is used.

In Table 1, we compare the predictive power of the features discussed above. (To reduce noise, we use our full dataset, rather than just the subset for which we have SunsetWx data, so the AUCs for these predictions are not directly comparable to the AUCs for SunsetWx predictions). For each feature set, we fit a logistic regression model on a train set comprising a random half of the dataset and assess model performance (measured by AUC) on a test set comprising the remaining half. (Using a test set avoids overfitting.) The strongest weather predictors are humidity and cloud cover, and the strongest pollution predictors are nitrogen dioxide and carbon monoxide. We find that weather features are more predictive than pollution features, and that a combined model using both weather and pollution features (and the previous day's sunset quality) slightly outperforms weather features alone. It is worth noting that we assess predictive power using features measured *at the time of sunset*, and that a true prediction algorithm would have to use the values of those features predicted ahead of time, rather than their measured values.

C. Locations with beautiful sunsets

Merely plotting the density of sunset posts will not identify locations with beautiful sunsets because Instagram usage is heavily correlated with population density.

Features used in model	AUC
Humidity	0.64 ± 0.03
Cloud cover	0.64 ± 0.03
Temperature	0.57 ± 0.02
Visibility	0.57 ± 0.01
Wind speed	0.57 ± 0.03
Pressure	0.48 ± 0.02
Nitrogen dioxide	0.58 ± 0.02
Carbon monoxide	0.57 ± 0.05
Ozone	0.54 ± 0.03
Sulfur dioxide	0.53 ± 0.03
Previous day	0.58 ± 0.02
All features combined	0.71 ± 0.02
Weather features only	0.69 ± 0.02
Pollution features only	0.64 ± 0.02

Table I

How well does each model predict b_{cd}? AUCs reported are computed using a held-out test set, with the model fit on a separate train set. Errors are the standard deviation in AUC over multiple bootstrapped train sets.

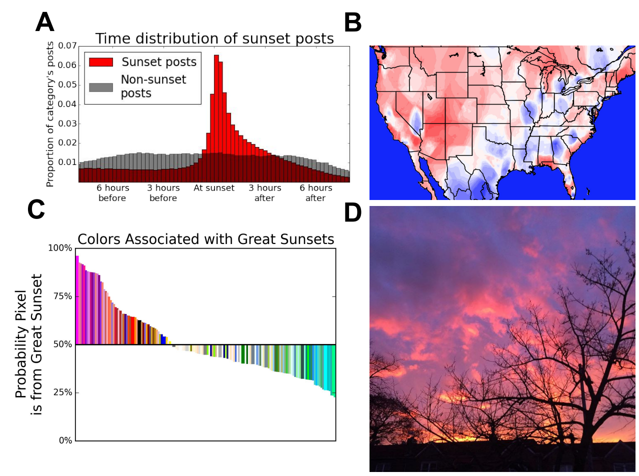

Fig. 1. A: Instagram sunset posts spike at sunset, while control posts do not. B: Density of sunset posts (red denotes higher densities) relative to control posts. C) colors associated with excellent sunsets. The vertical axis is the fraction of pixels of a particular color which are sampled from beautiful sunsets (as opposed to mediocre ones); D, the sunset algorithmically classified as most beautiful in a test set of 227 images.

We mitigate this problem by collecting a second dataset of Instagram posts with common tags (eg, "friends" and "smile") which we refer to as "control" posts; in Figure 1B, we plot the density of sunset posts relative to control posts across the United States. Sunset posts are particularly dense (red areas) near the coasts, in the mountains, and across much of the western United States. Results controlling for population density were qualitatively similar. (Because both methods control only imperfectly for location-based variation in social media usage, we used them only for this exploratory analysis rather than for the analysis in section B above.)

D. Visual features of beautiful sunsets

We next investigated whether we could quantitatively identify the visual features people associate with beautiful sunsets. We trained a classifier to classify sunsets as beautiful or average using a balanced dataset of beautiful and average sunsets selected as described in section A. We followed the following procedure: we randomly sampled 1,000 pixels from each image, computed the RGB representation of each pixel, and trained a K-nearest-neighbors classifier [11] to predict whether each pixel came from a beautiful or average sunset. (We used k-nearest-neighbors, which classifies a color depending on what fraction of the pixels nearest to it come from a beautiful sunset, because it allowed for flexible partitions of the color cube). In Figure 1C, we plot colors ranked by how strongly they are associated with beautiful sunsets (left, upward facing bars) or average sunsets (right, downward facing bars). The ranking matches intuition: pinks, purples, oranges, and reds are associated with beautiful sunsets, while greens and blues are associated with average sunsets. Strikingly, even this simple algorithm can classify a sunset as beautiful or average with accuracy comparable to our research assistants: on a test set of sunsets not used to train the algorithm, its AUC – ie, the probability that it assigns a higher score to a beautiful sunset than to an average sunset – is 0.70. (As discussed previously, the algorithm is unable to perfectly classify sunsets in part because photographer quality varies: some people take bad pictures on beautiful sunset days, or good pictures on mediocre sunset days.) In Figure 1D we show the picture in the test set to which the algorithm assigns the highest probability of being beautiful. It is easy to think of potential extensions and applications of this algorithm: it could be used to rapidly find particularly beautiful sunsets and improved by using a more sophisticated image-processing algorithm, like a convolutional neural network [12].

E. Real-time sunset notification

Beautiful sunset prediction is difficult because beautiful sunsets are rare: even if an algorithm can identify days when beautiful sunsets are twice as likely, they are still quite unlikely. An easier problem may be beautiful sunset *detection*: is it possible to detect beautiful sunsets as they are happening, using real-time social media data? We present a pilot of such a system, Sunset Nerd: we have created a website, http://sunsetfinder.herokuapp.com/, which monitors the real-time number of social media posts about the sunset in each city and assigns the sunset a quality score in real time by comparing to the historical number of sunset posts. We have built interfaces on Twitter and Facebook Messenger to allow users to receive beautiful sunset notifications so they can go outside. We will present results as they become available.

III. Discussion and Future Work

Though the sun sets on our present analysis of beautiful sunsets, many questions await a new dawn – and future work. Is it possible to use more sophisticated metereological features, like data on types of clouds, to improve sunset prediction? Will the real-time sunset detection system offer useful notifications? These open questions, and the analyses already performed, illustrate the utility of large-scale quantitative data in beautiful sunset prediction and detection.

Supplementary Information

All supplementary information is available at http://cs.stanford.edu/~emmap1/sunsets/supplementary_information.zip.

Acknowledgments

The author thanks the Hertz Foundation and NDSEGF for financial support; Justin Lowery and the other founders of SunsetWX for providing access to their API; Nick Mangus for EPA data; and Shengwu Li, Katherine Maslyn, and Andrew Suciu for research assistance and helpful conversations.

References

[1] C. Zerefos, P. Tetsis, A. Kazantzidis, V. Amiridis, S. Zerefos, J. Luterbacher, K. Eleftheratos, E. Gerasopoulos, S. Kazadzis, and A. Papayannis, "Further evidence of important environmental information content in red-to-green ratios as depicted in paintings by great masters," *Atmospheric Chemistry and Physics*, vol. 14, no. 6, pp. 2987–3015, 2014.

[2] R. E. Peterson, "Striking southern sunsets," *Weatherwise*, 1983.

[3] S. F. Corfidi, "The colors of sunset and twilight," 2014.

[4] "Instagram api." https://www.instagram.com/developer/, 2016.

[5] B. Reppert, J. DeFlitch, S. Hallett, and J. Lowery, "Sunsetwx: Innovation beyond the horizon." https://sunsetwx.com/, 2016.

[6] B. Reppert, J. DeFlitch, and S. Hallett, "Case study: Sunset model." https://www.sunsetwx.com/casestudy.pdf, 2016.

[7] J. Huang and C. X. Ling, "Using auc and accuracy in evaluating learning algorithms," *IEEE Transactions on knowledge and Data Engineering*, vol. 17, no. 3, pp. 299–310, 2005.

[8] "Forecastio." https://developer.forecast.io/, 2016.

[9] "Epa air quality data." https://www3.epa.gov/airdata/ad_data.html, 2016.

[10] C. Ballantyne, "Fact or fiction: Smog creates beautiful sunsets," *Scientific American*, 2007.

[11] K. Fukunaga and P. M. Narendra, "A branch and bound algorithm for computing k-nearest neighbors," *IEEE transactions on computers*, vol. 100, no. 7, pp. 750–753, 1975.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.