Collective Smile: Measuring Societal Happiness from Geolocated Images

ABSTRACT
The increasing adoption of social media has provided a unique opportunity to quantitatively characterize human behavior at a broader scale. Status updates from Twitter, in particular, have been aggregated for large scale sentiment analysis. While the methodologies are diverse in their applications and findings, they often focus on textual analysis and ignore significant media data, such as images.

In this work, we use geolocated images to determine patterns of happiness, and to sense underlying societal events and community characteristics. Using existing methods from computer vision, we detect smiling faces from nine million geolocated tweets posted between January 2012 and April 2013. As far as we know, this is the first approach using images for large scale happiness and sentiment analysis. We compare and correlate these patterns with various temporal events, stock prices and economic indices. We also use the findings to characterize different aspects of community — comparing the distribution of smiling images with crime rates across precincts in NYC, for example. Our results indicate that image content can be used as a reliable measure of happiness in a community. Given the consistency in facial expression to convey emotion, we believe that our system can be used for robust, cross-cultural public sentiment assessment which is often not possible in language dependent textual feature based systems.

1. RESULT
1.1 Smile and race distribution in NYC
The table 1 shows the correlation between racial distribution1 and the smile count per capita across zip areas in NYC. There is a strong2 positive correlation with fraction of white population and the smile count while the correlation with

<table>
<thead>
<tr>
<th>Race</th>
<th>Spearman Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>0.464114**</td>
</tr>
<tr>
<td>Black</td>
<td>−0.371355**</td>
</tr>
<tr>
<td>Asian</td>
<td>0.119849</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.054171</td>
</tr>
<tr>
<td>Black Majority (&gt; 50%)</td>
<td>−0.422906*</td>
</tr>
</tbody>
</table>

Table 1: Spearman Correlation between Race and number of images with smile per capita across zips in NYC.

black population is negative. The effect is more pronounced if we only consider areas with black majority population. There is no such relation with other races that is statistically significant.

1.2 Smile and median household income distribution in NYC
Using the American Community Survey data from U.S. Census Bureau, we also compare the distribution of median household income with smile distribution across zip areas in NYC. The median household income in NYC shows high degree of variation — the maximum is 232,031$ in Manhattan comparing to the minimum value 19,840$ from Bronx in our dataset.

<table>
<thead>
<tr>
<th>Overall</th>
<th>Spearman Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Majority Area (&gt; 50%)</td>
<td>−0.586325**</td>
</tr>
<tr>
<td>White Majority Area (&gt; 50%)</td>
<td>0.312572*</td>
</tr>
<tr>
<td>White Majority Area (&gt; 75%)</td>
<td>0.536020**</td>
</tr>
<tr>
<td>Black Majority Area (&gt; 75%)</td>
<td>−0.726471**</td>
</tr>
</tbody>
</table>

Table 2: Spearman Correlation between median house income and smile count per capita across Zip areas in NYC.

1.3 Crime

Table 2 shows the correlation between per capita smile count and median household income. While the overall relation does not show very strong correlation, when we take the racial distribution into consideration, a clear pattern emerges. The correlation goes from positive to negative in black majority areas where it is more positive in white majority areas.
Table 3: Spearman Correlation between major felony crime and smile count per capita across precincts in NYC.

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>White Majority (&gt; 50%)</th>
<th>Black Majority (&gt; 50%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crime</td>
<td>−0.109572**</td>
<td>0.515584*</td>
<td>−0.336264</td>
</tr>
</tbody>
</table>

Table 4: Spearman Correlation between number of coffee shop and smile count per capita across Zip areas in NYC.

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>White Majority Area (&gt; 50%)</th>
<th>White Majority Area (&gt; 75%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman Correlation</td>
<td>0.536572**</td>
<td>0.595560**</td>
<td>0.721890**</td>
</tr>
</tbody>
</table>

As neighborhood often gets characterized by safety and crime features, we use our dataset to have insight about such patterns as well. For this we use the total number of major felony crime reported in 2011 as a proxy for safety index for precincts.

As shown in table 3, the correlation between total number of crime is negatively correlated with smile count per capita. Interestingly, the correlation of crime and number of smile is positive for white majority population and negative (though statistically not significant) for black majority area.

Similar to this finding, in a recent work, Salesses et al. [6] concluded that after perception of safety, the reported violent crime like homicide and robbery occurs in relatively more upper class areas. We think this is an intriguing finding from socio-economic perspective that requires further investigation.

1.4 Gentrification and happiness index

The effect of gentrification for causing fundamental shift in neighborhood and communities has been an active topic in the theories of social change. In particular, the relationship between gentrification and change in crime rate has been a recurring theme in socio-political space.

To measure the effect of gentrification, Papachristos et al.[5] uses the number of coffee shop in the neighborhood arguing that the growth and spread of coffee shop provides on-the-ground measurement of economic development and consumption that follows gentrification. Similarly, to see if there is noticeable effect of gentrification on happiness index, we use the count of coffee shops as extracted from the public restaurant inspection data.

As table 4 shows there is a strong positive correlation with number of coffee shop and the happiness index. As number of coffee shop is an economic indicator of the society, the finding is consistent with the previous one about median household income. More interestingly, similar to [5], we find that the effect of gentrification differs according to racial composition of the communities. In the white majority population, the correlation is more pronounced, while there is no such statistically significant effect of gentrification and happiness for other races.

2. TEMPORAL SMILE DISTRIBUTION

To see if the public mood gets reflected in our dataset, we pick three major events across different places from 2012-13 timespan and compare the ratio of images containing smile during the events.

Figure 1 compares the smile ratio during the Boston marathon bombing and subsequent manhunt period — from April 15th, 2013 to April 19th, 2013. The mean of daily smile ratio from MA is 0.105($\sigma = 0.0368$) which indicates the average happiness index from our dataset falls during the bombing.

Similarly, Figure 2 shows that during Hurricane Sandy the smile ratio from NY is lower than the average value 0.107 ($\sigma = 0.0249$). Figure 3 compares the smile ratio during Presidential Inauguration in DC. For all of the events, the smile ratio is consistently higher during the New Year Eve.

How can I test for significance of the value change in this periods? Chow Test?
Figure 3: Smile ratio from DC.

2.1 Time Series Normalization
As Bollen et al. [1] suggests, to avoid systematic variance, for time series analysis we use Z Score instead of raw smile count. For the time period $[t_1, t_2]$, the normalized smile count for a day $d$ is calculated as:

$$\hat{S}_d = \frac{S_d - \mu_{[t_1, t_2]}}{\sigma_{[t_1, t_2]}}$$

where $S_d$ is the raw count of smile and $\mu_{[t_1, t_2]}$, $\sigma_{[t_1, t_2]}$ denotes the mean and the standard deviation during the $[t_1, t_2]$ time period.

Figure 5 and 4 shows the normalized count of smile in NYC and USA respectively.

Daily pattern of happiness
Figure 6 shows the daily distribution of smile count in NYC. Saturday and Sunday is the happiest days while there is a dramatic fall of happiness on Monday and it reaches minimum on Tuesday. The daily pattern is similar when the data

Figure 7: Distribution of smile (Z-normalized) over days in USA.
Hourly pattern

We also examine the hourly pattern of smile in our dataset. As shown in Figure 8, the smile count reaches a maximum after 10pm and gradually falls to minimum around 5AM. After that the happiness gradually increases through out the day. Similar pattern of hourly distribution in happiness can be observed when considering data across USA as shown in Figure 9.

The pattern of peak values around midnight and gradual decline to minimum around 5 AM is consistent with the pattern of positive affect found by Golder et al. [3] using word based indicator for enthusiasm, delight, activeness, and alertness from Twitter.

A weekday distribution of hourly pattern as Michael Macy did would be interesting.

is aggregated all over USA as shown in Figure 9 though the count reaches minimum on Wednesday. The daily pattern of happiness from smile is consistent with earlier works on Twitter [2] and blogs [4].

Table 5: Spearman correlation between smile count and DJIA index.

<table>
<thead>
<tr>
<th></th>
<th>DJIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>0.712264*</td>
</tr>
<tr>
<td>NYC</td>
<td>0.5546</td>
</tr>
</tbody>
</table>

3. ECONOMIC PHENOMENON

Given that public mood has often attributed to economic phenomenon, we explore the possibility of quantifiable relationship between happiness index from smile count and macro-economy.

3.1 Dow Jones Industrial Average

Figure 10 shows the DJIA values from corresponding to our time line. As shown in the table 5, both the smile count in NYC and USA is strongly correlated with the DJIA values during the time period.

Would it be interesting to do further period specific analysis of DJIA marked by rise and fall?

3.2 Index of Coincident Economic Indicators

We also use Index of Coincident Economic Indicators (ICEI) for measuring the relationship with regional happiness and economic activity. ICEI, published by New York State Department of Labor, is a monthly composite index consisting of key indicators of economic activity — private sector employment, the unemployment rate, average weekly work hours of manufacturing workers, and sales tax collections. Historically, ICEI has been a reliable indicator of economic condition of New York State. Figure 11 shows the monthly ICEI from January 2012 to April, 2013.

To compare with ICEI, we aggregate the smile count in NYC over month and perform a Z-Score normalization. The resulting graph is shown in Figure 12. As the table 6 shows the happiness index in NYC closely tracks ICEI.

A weekday distribution of hourly pattern as Michael Macy did would be interesting.

4. REFERENCES

Figure 11: Index of Coincident Economic Indicators (ICEI) for NYS.

Figure 12: Normalized monthly smile count (Z Score) from NYC.

Table 6: Spearman correlation between ICEI and smile index in NYC.

<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYC</td>
<td>0.955882</td>
</tr>
</tbody>
</table>


