Sentribute: Image Sentiment Analysis from a Mid-level Perspective

Jianbo Yuan, Quanzeng You, Sean McDonough, and Jiebo Luo
Department of Computer Science
University of Rochester

@WISDOMKDD'13
August 11, 2013, Chicago

Why Images?
• Photography is the only language understood in all parts of the world
• "A picture is worth 1000 words"

Sentiment: Why from a Mid-level Perspective

Image Sentiment Analysis
• Textual contents contain sentiment information
• Images also contain:
  • Aesthetics [Joshi et al.]
  • Emotion

Why is Image Sentiment Important?
• Social multimedia platforms present great opportunities, for example, images constitute about 36 percent of all the shared links on Twitter
• Sentiments expressed in image tweets are correlated with regular textual tweets [You et al.]

Image Sentiment Analysis in Social Media
• Sentiment is arguably the most important signal from social media
• Most existing sentiment analysis are based on textual information only
  – Comments, reviews, textual tweets, and status updates
• Twitter
  – Easy? 36% tweets contain images
  – Difficult? sentiment labels are provided through AMT mechanism
• Questions
  – Do users express themselves only using text?


Image Sentiment Analysis

Why Images?

Figure 1: Examples of Image Sentiments

**Sentribute: The Framework**

- Scene descriptors are used to generate mid-level attributes [Farhadi et al.]
- Mid-level attributes are then used to predict sentiment labels
- Face detection and eigenface-based facial sentiment detection are used to enhance the sentiment prediction results for images containing salient people

**Low-level Features and Mid-level Attributes**

- Low-level visual features
  - HOG: object and human recognition
  - GIST: scene recognition
  - SSIM: invariant scene layout
  - GEO-COLOR-HIST: robust histogram features
- Mid-level attributes (102):
  - Materials: such as metal, vegetation;
  - Functions: such as playing, cooking;
  - Surface properties: such as rusty, glossy;
  - Spatial Envelope [Oliva et al.]: such as natural, man-made, enclosed


**Image Sentiment Dataset**

- Collected from Twitter and labeled by AMT [Columbia University]
- An unbalanced dataset with 660 positive images and 150 negative images
- Asymmetric bagging is employed

**Facial Sentiment Detection using Eigenfaces**

- Images containing faces constitute an important part in all images and are effective in evoking sentiments
- Training with frontal faces of emotions under consistent lighting conditions
  - Dataset: Karolinska Directed Emotional Faces (KDEF)

**Mid-level Attribute Classification**

- Dataset: SUN dataset (MIT)
  - 102 mid-level attributes [Patterson & Hays]
  - Each is labeled by 3 individuals, making votes ranking for 0 to 3
  - Consider images with votes of more than 1 as positive samples as Soft Decision (SD), and votes of more than 2 as positive as Hard Decision (HD)

Genevieve Patterson, James Hays. SUN Attribute Database: Discovering, Annotating, and Recognizing Scene Attributes. CVPR 2012.
Linking Attributes to Sentiments

- Mutual Information (MI) analysis
- Attributes with the 10 highest MI values for SD and HD:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Soft Decision</th>
<th>Hard Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>congregating</td>
<td>railing</td>
</tr>
<tr>
<td>2</td>
<td>flowers</td>
<td>hiking</td>
</tr>
<tr>
<td>3</td>
<td>aged/worn</td>
<td>gaming</td>
</tr>
<tr>
<td>4</td>
<td>vinyl/linoleum</td>
<td>competing</td>
</tr>
<tr>
<td>5</td>
<td>still water</td>
<td>trees</td>
</tr>
<tr>
<td>6</td>
<td>natural light</td>
<td>metal</td>
</tr>
<tr>
<td>7</td>
<td>glossy</td>
<td>Tiles</td>
</tr>
<tr>
<td>8</td>
<td>open area</td>
<td>direct sun/sunny</td>
</tr>
<tr>
<td>9</td>
<td>glass</td>
<td>aged/worn</td>
</tr>
<tr>
<td>10</td>
<td>ice</td>
<td>Constructing</td>
</tr>
</tbody>
</table>

Sentiment Classification

- State-of-the-art approaches
  - Textual information based sentiment analysis, as well as online sentiment dictionary
  - Sentiment analysis based on low-level visual features

- Baselines
  - Low-level visual feature based approach
  - Textual content based approach
  - Online sentiment dictionary SentiStrength

Results: Sentiment Classification

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Precision</th>
<th>Recall</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>82.6</td>
<td>56.8</td>
<td>55.2</td>
</tr>
<tr>
<td>HD</td>
<td>85.7</td>
<td>59.1</td>
<td>61.4</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>84.3</td>
<td>54.7</td>
<td>54.8</td>
</tr>
<tr>
<td>HD</td>
<td>88.1</td>
<td>58.8</td>
<td>61.2</td>
</tr>
</tbody>
</table>

Results: Advantage over Low-level Methods

- Low-level visual feature based method [S. Siersdorfer et al.]
  - SIFT, Global Color Histogram, Local Color Histogram as features
  - Linear SVM as Classifier

- Mid-level attribute based approach and easy to interpret, and amenable to modular learning

Results: Facial Sentiment Detection

- About half of the images in our dataset contain faces
- Testing on 153 images and 113 are correctly classified
- Examples of our results:

Correctly Classified

Falsey Classified

Unclassified

Examples of our results:

- Aged/worn
- Constructing
- Stressful
Results: Decision Fusion
• Combine the results based on mid-level attribute classification and Eigenface based facial sentiment detection

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid-level Based Prediction</td>
<td>64.71%</td>
</tr>
<tr>
<td>Facial Sentiment Detection</td>
<td>73.86%</td>
</tr>
<tr>
<td>(After Decision Fusion)</td>
<td>82.35%</td>
</tr>
</tbody>
</table>

Results: Examples of Classification Results
• Examples of
  - True Positive
  - True Negative
  - False Positive
  - False Negative

Conclusions
• Social network platforms become increasingly popular and thus present great opportunities for sentiment analysis
• It is not effective to interpret the sentiment of social images based on low-level visual features. Mid-level attribute based analysis provides a new perspective to bridge semantic gap and intent gap, and is also amenable to modular learning
• Facial emotion detection based on Eigenfaces is a simple but useful method for classifying images with salient frontal faces and contributes to a 18% gain in accuracy for images containing faces

Future Work
• Better ways to combine textual information and visual information
  – Learning-based approaches
  – Topic and user specific approaches
• Introduce mid-level attributes to other areas (aesthetics analysis)
• Potential applications of mid-level based image sentiment analysis
  – Psychology, public opinion analysis and online activity analysis

Thanks