Ranking and Classifying Attractiveness of Photos in Folksonomies

Jose San Pedro Wandelmer (jsanpedro@mac.com)  
The University of Sheffield, UK

Stefan Siersdorfer (siersdorfer@l3s.de)  
L3S Research Center, Hannover, Germany
Motivation

Web 2.0 is hot!

- Easy
- Collaborative
- Dynamic

Massive amount of shared resources

- pose problems for effective search & retrieval
- especially critical for multimedia information
  - ‘Semantic gap’ limits the effectiveness of content-based techniques
- Finding relevant content becomes a difficult task
Motivation

Vast amounts of user generated data available

- Metadata
  - tags, descriptions, comments, etc.
- User feedback
  - implicit: n. views, n. comments, etc.
  - explicit: ratings, favorite assignments, etc.
- Not always reliable: irregularities, sparsity, etc

Can we combine user generated data and content features to enhance retrieval?

- ‘Community knowledge’ can help to learn about the content
- Content can help overcome irregularities of user data
Problem Setting

Scenario
- Photo sharing - Flickr

Objective
- Determine \textit{attractiveness} of shared photos

Reasons
- Direct application for photo retrieval enhancement
- Subjective concept
  - but there is a whole community providing judgements
- Image semantics are not critical
  - enables efficient use of content-based visual features
Problem Setting

We propose

- A methodology for classification and ranking of images based on their visual appeal

Inputs

- Flickr Photo Stream
- #views
- #comments
- #favorites
- ...

Community Feedback (photo’s interestingness)

Content (visual features)

Metadata (textual features)

Classification & Regression Attractiveness Models Generator
Attractiveness of Images

Factors that influence human perception of attractiveness?

Landscape  Portrait  Flower
Attractiveness Visual Features

Human visual perception mainly influenced by

- Color distribution
- Coarseness

These are complex concepts

- Convey multiple orthogonal aspects
- Necessity to consider different low level features
Attractiveness Visual Features

Color Features

- Brightness
- Contrast
  - Luminance, RGB
- Colorfulness
- Naturalness

- Saturation
  \[ S = \max(R, G, B) - \min(R, G, B) \]
  - Mean, Variance
  - Intensity of the colors
  - Saturation is 0 for grey scale images
Attractiveness Visual Features

Color Features

- Brightness
- Contrast
- Luminance, RGB

\[ \text{Contrast} = \max(R, G, B) - \min(R, G, B) \]
Attractiveness Visual Features

Color Features

- Brightness
- Contrast
  - Luminance, RGB
- Colorfulness
- Naturalness
- Saturation
  - Mean, Variance
  - Intensity of the colors
  - Saturation is 0 for grey scale images

\[ \text{min}(R, G, B) \]
Attractiveness Visual Features

Color Features
- Brightness
- Contrast
  - Luminance, RGB
- Colorfulness
- Naturalness
- Saturation
  - Mean, Variance
  - Intensity of the colors

Saturation is 0 for grey scale images
Visual Features

Coarseness

- Resolution + Acutance
- Sharpness
  - Critical importance for final appearance of photos [Savakis 2000]

\[ Sh = \sum_{x,y} \frac{L(x,y)}{\mu_{xy}}, \quad \text{with} \quad L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \]
Visual Features

Coarseness

Resolution + Acutance

Sharpness

Critical importance for final appearance of photos [Savakis 2000]

\[ L(x, y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \]
Visual Features

Coarseness
Resolution + Acutance
Sharpness

Critical importance for final appearance [Savakis 2000]

\[ L(x, y) \]
Textual Features

We consider user generated meta data

- Correlation of topics with image appealing
- Tags seem appropriate to capture this information

<table>
<thead>
<tr>
<th>Technical</th>
<th>Judgements</th>
<th>Awards</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>macro</td>
<td>coolest</td>
<td>diamondclassphotographer</td>
<td>nature</td>
</tr>
<tr>
<td>portrait</td>
<td>bravo</td>
<td>flickrdiamond</td>
<td>sky</td>
</tr>
<tr>
<td>canon</td>
<td>art</td>
<td>blueribbonwinner</td>
<td>sunset</td>
</tr>
<tr>
<td>landscape</td>
<td>supershot</td>
<td>colorphotoaward</td>
<td>woman</td>
</tr>
<tr>
<td>hdr</td>
<td>superbmasterpiece</td>
<td></td>
<td>sea</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dates</th>
<th>Events</th>
<th>Places</th>
<th>People</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>wedding</td>
<td>madagascar</td>
<td>matt</td>
</tr>
<tr>
<td>2006</td>
<td>graduation</td>
<td>prague</td>
<td>ian</td>
</tr>
<tr>
<td>may</td>
<td>honeymoon</td>
<td>china</td>
<td>dad</td>
</tr>
<tr>
<td></td>
<td>birthday</td>
<td>vegas</td>
<td></td>
</tr>
</tbody>
</table>
Classification & Regression Models

Classification

- Automatic categorization of photos
  - attractive vs unattractive
- Supervised learning paradigm
  - Training set of photos represented as feature vectors
    - numeric values of visual features
    - tfidf weights of tags
- We consider different vector representations
  - Text features only - Visual features only - Text+Visual features
- Necessity of a sufficiently large training set of labeled photos
  - Flickr provides large photo collections with social feedback
  - Number of favorites (NumFav) as interestingness indicator
    - Distinct thresholds for minimum NumFav
Classification & Regression Models

Classification

- **Formally:** \( \{(\vec{p}_1, l_1), \ldots, (\vec{p}_n, l_n)\} \)  
  \[ l_i = \begin{cases} 
  1, & \text{NumFav} > \text{thr} \\
  -1, & \text{Otherwise} 
\end{cases} \]

- **Linear SVMs**
  - **Training:**  
    \[ w \cdot x + b = 0 \]
  - **Classification:**  
    \[ \vec{w} \cdot \vec{y} + b > 0? \]

Regression

- **Formally:** \( \{(\vec{p}_1, r_1), \ldots, (\vec{p}_n, r_n)\} \)

- **SVM Regression**
- **Find a function to assign continuous relevance values**
Experiments

Data

- Sample of Flickr photos
- Uploaded between June 1st and 7th 2007
- Flickr API for new photos - 20 minutes time interval
- Collection size: 2.2M - N. Users: 185k

Test sets

- Positive examples
  - N. Fav >= 2
  - Size: 35,000
- Negative examples
  - N. Fav = 0
  - Size: 40,000
Experiments

Classification

○ Setup
  ○ Attractive defined at various restrictiveness levels:
    ○ NumFav >= 2, 5, 10, and 20
  ○ Random selection of training photos:
    ○ Set size = 500, 2000, 8000, and 20000

○ Results
  ○ T=8000, NumFav >= 5
  ○ BEP for different configurations
    ○ Text Features : 0.7843
    ○ Visual Features: 0.6664
    ○ Text + Visual : 0.8363
Experiments

Classification Setup

Attractive defined at various restrictiveness levels:
- NumFav >= 2, 5, 10, and 20
- Random selection of training photos:
  - Set size = 500, 2000, 8000, and 20000

Results

T=8000, NumFav >= 5

BEP for different configurations:
- Text Features: 0.7843
- Visual Features: 0.6664
- Text + Visual: 0.8363
Experiments

Classification

Setup

Attractive defined at various restrictiveness levels:

- NumFav >= 2, 5, 10, and 20
- Random selection of training photos:
  - Set size = 500, 2000, 8000, and 20000

Results

T=8000, NumFav >= 5

BEP for different configurations

Text Features: 0.7843

Visual Features: 0.6664

Text + Visual: 0.8363
Experiments

Classification Setup
Attractive defined at various restrictiveness levels: NumFav >= 2, 5, 10, and 20
Random selection of training photos: Set size = 500, 2000, 8000, and 20000

Results
T=8000, NumFav >= 5
BEP for different configurations
Text Features : 0.7843
Visual Features: 0.6664
Text + Visual : 0.8363

Precision Recall

visual

(P=0.84, R=0.1)
(P=0.79, R=0.3)
Experiments

Classification

Precision

Recall

- text
- visual
- visual+text

BEP
Experiments

Ranking

- SVM Regression
- Training set:
  - 20,000 random photos with NumFav $\geq 2$
  - 20,000 random photos with NumFav $= 0$
- Test set:
  - Remaining (disjoint) set
- Test set ranked and compared to decreasing NumFav sorted list
  - Kendall’s Tau-b is used to compare ranked lists
Experiments

### Ranking

<table>
<thead>
<tr>
<th>Method</th>
<th>Kendall’s Tau-b</th>
</tr>
</thead>
<tbody>
<tr>
<td>brightness</td>
<td>0.0006</td>
</tr>
<tr>
<td>contrast</td>
<td>-0.0172</td>
</tr>
<tr>
<td>RGB contrast</td>
<td>0.0288</td>
</tr>
<tr>
<td>saturation</td>
<td>0.1064</td>
</tr>
<tr>
<td>saturation variation</td>
<td>0.0472</td>
</tr>
<tr>
<td>colorfulness</td>
<td>-0.0497</td>
</tr>
<tr>
<td>sharpness</td>
<td>0.0007</td>
</tr>
<tr>
<td>sharpness variation</td>
<td>-0.0914</td>
</tr>
<tr>
<td>naturalness</td>
<td>0.0143</td>
</tr>
<tr>
<td>text</td>
<td>0.3629</td>
</tr>
<tr>
<td>visual</td>
<td>0.2523</td>
</tr>
<tr>
<td>text + visual</td>
<td>0.4841</td>
</tr>
</tbody>
</table>

\[ \tau_b \in [-1, 1] \]
Conclusions

Successful strategy to mine

- Community Feedback
- Metadata
- Visual Content features

Specifically

- Classification & Ranking models of image attractiveness using:
  - Textual features from meta data: Tags
  - Visual features from image content
- Ground-truth (Class labels and relevance values)
  - Community feedback: Number of favorite assignments
Conclusions

Results show

- Hybrid approach (text + visual) offers the best performance
  - High precision-recall - BEP = 0.8363, T=8000, N.Fav>=5
- Visual models provide applicable results
  - Lower BEP : 0.6664
  - Higher flexibility
    - Local domain
    - Recently updated pictures (no feedback still available)

Future Directions

- Development & Evaluation of enhanced photo search tools
- Extension to other kinds of media
  - Videos (YouTube)
Thank You