A preliminary study on viewer sentiment analysis of social media videos

Marc A. Kastner, Shin’ichi Satoh

mkastner@nii.ac.jp

Motivation

• Purpose: Find scenes which are funny, scary, sad ...
• Annotation expensive. No existing datasets!
• Can we use user comments to cluster sentiment of videos?

Approach

• Using videos from SNS (YouTube):
  • Crawl videos + their top-n comments
  • Analyze comments using NRC sentiment dictionaries
• Train model on:
  • X = [Visual features + Audio features]
  • Y = generated Emotion / VAD annotation

From comments to sentiment

• The comments are direct reactions to comments
  • Sentiment analysis of comments helps understanding videos
• Sentiment analysis to generate labels (majority decision)
  Emotion = {sad, happy, ...} VAD = {0.1, 0.5, 0.3}

Experiments

• Dataset: 17,112 videos with generated Emotion/VAD from their top-100 comments
  • Train separate models for each
• Results
  • Works, but not enough data for some emotions
  • Dataset imbalanced

Table 1: Results for VAD estimation.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>2.99</td>
<td>0.47</td>
<td>2.00</td>
<td>0.51</td>
<td>1.98</td>
<td>0.32</td>
</tr>
<tr>
<td>Audio</td>
<td>2.83</td>
<td>0.54</td>
<td>1.99</td>
<td>0.51</td>
<td>1.95</td>
<td>0.36</td>
</tr>
<tr>
<td>Combined</td>
<td>2.84</td>
<td>0.55</td>
<td>1.95</td>
<td>0.55</td>
<td>1.93</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Table 2: Results for emotion estimation.

<table>
<thead>
<tr>
<th>Features</th>
<th>Avg. Precision</th>
<th>Avg. Recall</th>
<th>Avg. F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>0.30</td>
<td>0.39</td>
<td>0.28</td>
</tr>
<tr>
<td>Audio</td>
<td>0.36</td>
<td>0.41</td>
<td>0.34</td>
</tr>
<tr>
<td>Combined</td>
<td>0.33</td>
<td>0.41</td>
<td>0.34</td>
</tr>
</tbody>
</table>
A preliminary study on viewer sentiment analysis of social media videos

Marc A. Kastner, Shin’ichi Satoh

Next steps

- Improve features
  - RGB / Audio currently simple average over all frames (Switch to RNN model)
  - Include audio sentiment, music mood, etc.
- Train separate models for different categories
  - Can we find per-community sentiment models?

Emotion

Relationship Emotion <> VAD

Our dataset

Dataset composition

- Categories of videos
- Generated emotion distribution

Used datasets

- Sentiment dictionaries
- YouTube video dataset: