We Are Humor Beings: Understanding and Predicting Visual Humor

Shuai Wang

University of Toronto

March 29, 2016
An integral part but not understood in detail
Intro

- An integral part but not understood in detail
- An adult laughs 18 times a day
Intro

- An integral part but not understood in detail
- An adult laughs 18 times a day
- A good sense humor
  - is related to communication competence
  - helps raise an individual’s social status & popularity
  - even helps attract compatible mates
  - makes yourself happier :)

What makes an image funny?
Humor Techniques

- Animal doing something unusual
- Person doing something unusual
- Somebody getting hurt
- Somebody getting scared
Animal doing something unusual
Person doing something unusual
Somebody getting hurt
Somebody getting scared
Changing objects can alter the funniness of a scene
Removing Incongruities

An elderly person kicking a football while skateboarding is incongruous, but a young girl doing so is not.
Adding Incongruities

Add incongruities (and humor) by replacing the expected with the unexpected
Two Tasks to Understand Visual Humor

- Predicting how funny a given scene is (scene-level)
- Changing the funniness of a scene (object-level)
Object-level Features

- **Object embedding** (150-d): captures the context in which an object usually occurs
- **Local embedding** (150-d): weighted sum of object embeddings of all other instances
Scene-level Features

- **Cardinality** (150-d): bag-of-words representation of how many instances of each object are in the scene
- **Location** (300-d): horizontal and vertical coordinates of every object (closest to the center if multiple instance)
- **Scene embedding** (150-d): sum of object embeddings of all objects in the scene
Predicting Funniness Score

- Dataset: 6,400 scenes, with funny score from 1-5 labelled by workers from Amazon Mechanical Turk
Predicting Funniness Score

- Dataset: 6,400 scenes, with funny score from 1-5 labelled by workers from Amazon Mechanical Turk
- Support Vector Regressor (SVR) on scene-level features
Predicting Funniness Score

- Dataset: 6,400 scenes, with funny score from 1-5 labelled by workers from Amazon Mechanical Turk
- Support Vector Regressor (SVR) on scene-level features
- Metric: average relative error

$$\frac{1}{N} \sum_{i=1}^{N} \frac{|Predicted F_i - Ground Truth F_i|}{Ground Truth F_i}$$
Different feature subsets perform about the same: slightly better than baseline (average score of the training scenes)

<table>
<thead>
<tr>
<th>Features</th>
<th>Avg. Rel. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Prediction Baseline</td>
<td>0.3151</td>
</tr>
<tr>
<td>Embedding</td>
<td>0.2516</td>
</tr>
<tr>
<td>Cardinality</td>
<td>0.2450</td>
</tr>
<tr>
<td>Location</td>
<td>0.2400</td>
</tr>
<tr>
<td>Cardinality + Location</td>
<td>0.2435</td>
</tr>
<tr>
<td>Embedding + Location</td>
<td>0.2435</td>
</tr>
<tr>
<td>Cardinality + Embedding</td>
<td>0.2435</td>
</tr>
<tr>
<td>Embedding + Cardinality + Location</td>
<td>0.2400</td>
</tr>
</tbody>
</table>
Alter Funniness of a Scene

- Detect the objects that do (or do not) contribute to humor
- Identify which objects should replace the objects from step 1
Predicting Objects to be Replaced

- On average, the model replaces 3.67 objects (2.54 ground truth) → this bias towards replace ensures a large ‘margin’
Predicting Objects to be Replaced

- On average, the model replaces 3.67 objects (2.54 ground truth) → this bias towards replace ensures a large ‘margin’
- Animate objects like humans and animals are more likely sources of humor → tends to replace these objects
Funny → Unfunny

Old man dancing → young boy dancing
Hawk stealing meat → baseball
Funny → Unfunny

Cute puppy → Insect
Watermelon → Ax
Unfunny → Funny

Couple having dinner at the table → Puppies having dinner at the table
Unfunny → Funny

Cating playing around → Racoon driving motorcycle
Discussion

- Style/genre of an image or painting can make a difference
Discussion

- Style/genre of an image or painting can make a difference
- Dataset is small: 6,400 images
Discussion

- Style/genre of an image or painting can make a difference
- Dataset is small: 6,400 images
- Feature representation can be improved
THE END!