Why Watching Movie Tweets Won’t Tell the Whole Story?

Felix Ming-Fai Wong, Soumya Sen, Mung Chiang

EE, Princeton University

Twitter: A Gold Mine?

Could Twitter Data Replace Opinion Polls?

Twitter Predicts Box-Office Sales Better Than a Prediction Market [Updated]

A Prediction: Twitter to Predict the Future

Apps: Mombo, TwitCritics, Fflick
Or Too Good to be True?

Tweets as Poll Data? Be Careful

You Can't Use Twitter to Predict Election Results

Representativeness???
Why representativeness?

- Selection bias
  - Gender
  - Age
  - Education
  - Computer skills
  - Ethnicity
  - Interests

- The *silent* majority
Why Movies?

- Large online interest
- Relatively unambiguous
- Right in Timing: Oscars
Key Questions

- Is there a bias in Twitter movie ratings?
- How do Twitters compare to other online site?
- Is there a quantifiable difference across types of movies?
  - Oscar nominated vs. non-nominated commercial movies
- Can hype ensure Box-office gains?
Data Stats

- 12 Million Tweets
  - 1.77M valid movie tweets

- February 2 – March 12, 2012

- 34 movie titles
  - New releases (20)
  - Oscar-nominees (14)
### Example Movies

<table>
<thead>
<tr>
<th>Commercial Movies (non-nominees)</th>
<th>Oscar-nominees</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Grey</td>
<td>The Descendants</td>
</tr>
<tr>
<td>Underworld: Awakening</td>
<td>The Artist</td>
</tr>
<tr>
<td>Contraband</td>
<td>The Help</td>
</tr>
<tr>
<td>Haywire</td>
<td>Hugo</td>
</tr>
<tr>
<td><strong>The Woman in Black</strong></td>
<td>Midnight in Paris</td>
</tr>
<tr>
<td>Chronicle</td>
<td>Moneyball</td>
</tr>
<tr>
<td>The Vow</td>
<td>The Tree of Life</td>
</tr>
<tr>
<td>Journey 2: The Mysterious Island</td>
<td>War Horse</td>
</tr>
</tbody>
</table>

Oscar-nominees for Best Picture or Best Animated Film
Rating Comparison

- Twitter movie ratings
  - Binary Classification: Positivity/Negativity

- IMDb, Rotten Tomatoes ratings

- Rating scale: 0 – 10, 0 – 5

- Benchmark definition
  - Rotten Tomatoes: Positive - 3.5/5
  - IMDb: Positive: 7/10
  - Movie score: weighted average
  - High mutual information
Challenges

- Common Words in titles
  - e.g., “Thanks for the help”, “the grey sky”

- No API support for exact matching
  - e.g., “a grey cat in the box”

- Misrepresentations
  - e.g., “underworld awakening”

- Non-standard vocabulary
  - e.g., “sick movie”

- Non-English tweets

- Retweets: approves original opinion
# Tweet Classification

<table>
<thead>
<tr>
<th>Class</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Relevance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irrelevant (I)</td>
<td>Non-English (possibly relevant), or irrelevant from the context</td>
<td>“thanks for the help”</td>
</tr>
<tr>
<td>Relevant (R)</td>
<td>Otherwise</td>
<td>“watched The Help”</td>
</tr>
<tr>
<td><strong>Sentiment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative (N)</td>
<td>Contains <em>any</em> negative comment</td>
<td>“liked the movie, but don’t like how it ended”</td>
</tr>
<tr>
<td>Positive (P)</td>
<td><em>Unanimously</em> and <em>unambiguously</em> positive</td>
<td>“the movie was awesome!”</td>
</tr>
<tr>
<td>Mention (M)</td>
<td>Otherwise</td>
<td>“the movie was about wolves”</td>
</tr>
<tr>
<td><strong>Temporal Context</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After (A)</td>
<td>After watching as inferred from context</td>
<td>“had a good time watching the movie”</td>
</tr>
<tr>
<td>Before (B)</td>
<td>Before watching movie</td>
<td>“can’t wait to see the movie!”</td>
</tr>
<tr>
<td>Current (C)</td>
<td>Tweeted when person was already inside the cinema</td>
<td>“at cinema about to watch the movie”</td>
</tr>
<tr>
<td>Don’t know (D)</td>
<td>Otherwise</td>
<td>“have you seen the movie?”</td>
</tr>
</tbody>
</table>
Steps

- Preprocessing
  - Remove usernames
  - Conversion:
    - Punctuation marks (e.g., “!” to a meta-word “exclmark”)
    - Emoticons (e.g., 😊 or :, 😞 etc.), URLs, “@”

- Feature Vector Conversion
  - Preprocessed tweet to binary feature vector (using MALLET)

- Training & Classification
  - 11K non-repeated, randomly sampled tweets manually labeled
  - Trained three classifiers
Classifier Comparison

- SVM based classifier performance is significantly better
- Numbers in brackets are balanced accuracy rates
**Temporal Context**

- Most "current" tweets are "mentions".
- LBS
- "Before" or "After" tweets are much more positive.

---

**Fraction of tweets in joint-classes**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>P</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.045</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>B</td>
<td>0.011</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>C</td>
<td>0.0019</td>
<td>0.019</td>
<td>0.090</td>
</tr>
<tr>
<td>D</td>
<td>0.0097</td>
<td>0.034</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Twitters are overwhelmingly more positive for almost all movies.
Rotten Tomatoes vs. IMDb?

(a) Rotten Tomatoes vs IMDb ratings (new)
(b) Rotten Tomatoes vs IMDb ratings (Oscars)

-good match between RT and IMDb (Benchmark)
Twitters vs. IMDb vs. RT ratings (1)

- New (non-nominated) movies score more positively from Twitter users
Oscar-nominees rate more positively on IMDb and RT
Quantification Metrics (1)

Positeness ($P$): \[ P = \frac{x^* + y^*}{2} \]

- $(x^*, y^*)$ are the medians of proportion of positive ratings in Twitter and IMDb/RT

Bias ($B$): \[ B = 1 - \tan^{-1}\left(\frac{y^*}{x^*}\right) / \frac{\pi}{4} \]
Inferrability ($I$):

$$I = \sum_{i=1}^{m} \sum_{j=1}^{m} p_{XY}(i,j) \log_2 \frac{p_{XY}(i,j)}{p_X(i)p_Y(j)},$$

where

$$p_X(i) = \#\{(x_k, y_k) : x_k \in b_i\}/n$$

$$p_Y(j) = \#\{(x_k, y_k) : y_k \in b_j\}/n$$

$$p_{XY}(i,j) = \#\{(x_k, y_k) : x_k \in b_i, y_k \in b_j\}/n.$$

- If one knows the average rating on Twitter, how well can he/she predict the ratings on IMDb/RT?
Oscar-nominees have higher positiveness

RT-IMDb have high mutual inferrability and low bias

Twitter users are more positively biased for non-nominated new releases
Hype-approval vs. Ratings

\[ \mathcal{H} = \frac{BP}{AP} = \frac{\# \text{ Positive tweets before watching}}{\# \text{ Positive tweets after watching}} \]

(a) Online Ratings vs. \( \mathcal{H} \)  
(b) Online Ratings vs. BP

Higher (lower) Hype-approval \( \neq \) Higher (lower) IMDb/RT ratings
Hype-approval vs. Box-office

- $\frac{BP}{AP} > 1$?
- IMDb Rating $\geq 0.7$?
- Box Office $\geq$ $50M$?

- High Hype + High IMDb rating: Box-office success
- Other combination: Unpredictable

Journey 2
Summary

- Tweeters aren’t representative enough
- Need to be careful about tweets as online polls
- Tweeters’ tend to appear more positive, have specific interests and taste
- Need for quantitative metrics

Thanks!