“You are no Jack Kennedy”: On Media Selection of Highlights from Presidential Debates

Chenhao Tan
University of Colorado Boulder
chenhao@chenhaot.com
@ChenhaoTan

Joint work with Hao Peng and Noah Smith at University of Washington
Presidential debates have become a focal point of election campaigns

The first 2016 Presidential debates set the record as the most-watched debate in television history, with 84 million viewers.
Presidential debates have become a focal point of election campaigns

“Senator, you are no Jack Kennedy”
News media select highlights from presidential debates

"Yes, I did. And you know what else I prepared for? I prepared to be president." —Clinton

"you’ve been doing this for 30 years. Why haven’t you been able to implement any of these ideas?" —Trump

"He tried to change from looks to stamina, but this is a man who has called women pigs, slobs and dogs,” —Clinton
News media select highlights from presidential debates

These highlights can play an important role in shaping the public’s opinion of presidential debates and candidates

(Fridkin et al, 2008; Hillygus and Jackman, 2003; Hwang et al., 2007)
Media selection of highlights

- The effect of wording on media choices
- How well can **humans/machines** predict highlights based on texts?
- Media preferences over time
Dataset

- Debate transcripts from 1980-2016 (the American Presidency Project)
  - both primary and general debates
- Quotes in newspapers from 1980-2016 (LexisNexis)

Debate quotes are increasingly used in news articles
A binary classification framework

- Control for the speaker
- Control for the situation in the debate
- Control for the length

SANDERS: Do I consider myself part of the casino capitalist process by which so few have so much and so many have so little by which Wall Street's greed and recklessness wrecked this economy? [...] CLINTON: [...] I think what Senator Sanders is saying certainly makes sense in the terms of the inequality that we have. [...] SANDERS: [...] So what we need to do is support small and medium-sized businesses, the backbone of our economy, but we have to make sure that every family in this country gets a fair shake [...]

Human performance

- Human accuracy is 60%
- Top factors do not directly align with existing theories (narrative relevance, conspicuousness, extractability) (Clayman, 1995)
- Only 3% of the participants mentioned that context matters

<table>
<thead>
<tr>
<th>Rationale for choices</th>
<th>%subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>circular (sound bite, newsworthy)</td>
<td>30.0%</td>
</tr>
<tr>
<td>provocative, sensational</td>
<td>25.5%</td>
</tr>
<tr>
<td>surprising, funny</td>
<td>17.0%</td>
</tr>
<tr>
<td>issues, informative</td>
<td>16.0%</td>
</tr>
</tbody>
</table>
Quantitative Representation

• Modeling conversations
• Sentence-alone features
Modeling conversations

- Similarity with neighboring turns from the speaker
- Similarity with neighboring turns from other participants

previous turns (minus)  |  future turns (plus)
Highlights are “locally” distinct from one’s own turns and get echoed more...
Sentence-alone features

- Informativeness
- Emotions
- Contrast
- Personal pronouns
- Uncertainty
- Strong emphasis
- Generality
- Language model
- Parallelism
Sentence-alone features

- Informativeness
- Emotions
- Contrast
- Personal pronouns
- Uncertainty
- Strong emphasis
- Generality
- Language model
- Parallelism
Sentence-alone features

- Informativeness
- Emotions
- Contrast
- Personal pronouns
- Uncertainty
- Strong emphasis
- Generality
- Language model
- Parallelism

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Related theories/intuitions and brief description</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informativeness</td>
<td>We use length as a proxy of informativeness. Longer sentences are more likely to be highlighted despite our control on length as discussed in [3.1]. This echoes findings in Tan et al. [56, 58].</td>
<td>length: $\uparrow \uparrow \uparrow \uparrow$</td>
</tr>
<tr>
<td>Emotions</td>
<td>We consider positive and negative words in Pennebaker et al. [40]. Highlighted sentences use significantly more negative words, while there is no difference in positive words. This is consistent with negativity bias [64] and the negativity found in the news media [14].</td>
<td>pos: $\uparrow \uparrow \uparrow \uparrow$, neg: $\uparrow \uparrow \uparrow \uparrow$</td>
</tr>
<tr>
<td>Contrast</td>
<td>In general, highlighted sentences use more personal pronouns except first person plural and third person plural. Our explanation: the contrast between I and we is that we choose to focus on statements about ourselves whereas they focus on statements using we.</td>
<td>$i$, you, she, he, they: $\uparrow \uparrow \uparrow \uparrow$, use: $\uparrow \uparrow \uparrow \uparrow$</td>
</tr>
<tr>
<td>Personal pronouns</td>
<td>In general, highlighted sentences use more personal pronouns except first person plural and third person plural. Our explanation: the contrast between I and we is that we choose to focus on statements about ourselves whereas they focus on statements using we.</td>
<td>$i$, you, she, he, they: $\uparrow \uparrow \uparrow \uparrow$, use: $\uparrow \uparrow \uparrow \uparrow$</td>
</tr>
<tr>
<td>Uncertainty/subjectivity</td>
<td>In the debate context, hedges may also represent subjectivity.</td>
<td>$\uparrow \uparrow \uparrow \uparrow$, hedges: $\uparrow \uparrow \uparrow \uparrow$</td>
</tr>
<tr>
<td>Strong emphasis</td>
<td>Superlatives represent the extreme form of an adjective or an adverb and can be used to put emphasis on a statement. Surprisingly, highlighted sentences do not use more superlatives.</td>
<td>superlatives: $\uparrow \uparrow \uparrow \uparrow$</td>
</tr>
<tr>
<td>Generality</td>
<td>We count indefinite articles to measure generality. Our findings are consistent with Danescu-Niculescu-Mizil et al. [13], Shahaf et al. [45], and Tan et al. [54].</td>
<td>indef articles: $\uparrow \uparrow \uparrow \uparrow$</td>
</tr>
<tr>
<td>Language model</td>
<td>To capture surprise or completeness, we compute language model scores based on NYT texts and unigram part-of-speech (POS) tags in the WE$^2$ portion of Penn Treebank. However, the only significant feature is $\text{unigram}$ is that highlighted sentences are more similar to NYT texts in unigram usage. This finding is consistent with message sharing [54] but is different from memorable movie quotes [13].</td>
<td>POS: $\uparrow$, 2, 3-gram: $\uparrow \uparrow \uparrow \uparrow$</td>
</tr>
<tr>
<td>Parallelism</td>
<td>Using parallel sentence structure is a rhetorical technique, e.g., the first sentence in Table 1 and I've never willed in my life, and I've never worried in my life. We use average longest common subsequences between sub-sentences to measure it [64].</td>
<td>parallelism: $\uparrow \uparrow \uparrow \uparrow$</td>
</tr>
</tbody>
</table>

Table 3: Testing results of sentence-alone features. Upward arrows indicate that highlighted sentences have larger scores in that feature, while downward arrows suggest the other way around ($\uparrow \uparrow \uparrow \uparrow$, $p < 0.0001$, $\uparrow \uparrow$, $0.001 < p < 0.01$, $\uparrow$, $p < 0.05$, the same for downward arrows; $p$ refers to the $p$-value after the Bonferroni correction).
Highlights uses more uncertainty, but not more strong emphasis

Uncertainty (hedges, e.g., I am not sure)

Strong emphasis (superplatives, e.g., greatest)
Machines consistently outperform humans

The graph shows the accuracy (%) of random, human, and all top percentage predictions. The accuracy decreases as the top percentage increases. The random model consistently has the highest error rate, followed by human models, and then all models.
Media preferences over time

Niculae et al., 2015
Media preferences over time

- Min-cut: bipartisan coverage
- Clustering quality: media fragmentation
Media preferences over time

- **Min-cut: bipartisan coverage**
- Clustering quality: media fragmentation
Decreasing bipartisan coverage

![Graph showing decreasing bipartisan coverage from 1992 to 2016. The graph plots cut fraction (%) against years (1992, 2000, 2008, 2016). The graph includes lines for 'count', 'positive (*)', and 'negative'.]
Take away

• It is challenging for the public (at least the turkers) to distinguish highlights from non-highlights.

• Machines can better predict highlights.

• Highlights are locally distinct and get echoed during the debate.

• Bipartisan coverage is decreasing over time.

• Data & more at https://chenhaot.com/papers/debate-quotes.html

Thank you!