Using Coreference Resolution to Measure Sentiment in Media Coverage

Farhan Iqbal
farhan.iqbal@uga.edu

July 30, 2021 – Academy of Management Annual Conference
The role of the media

The media are key stakeholders for organizations as:

Information Intermediaries

Social Arbiters
If variance in media coverage is meaningful, we should capture it.
How should we measure sentiment?

The best-in-class way to measure sentiment is by using human raters.

...but that’s not always feasible.
So we use computer-aided text analysis (CATA) to help us
1) Identify articles that include the organization’s name.
2) Read the first paragraph to determine if the article is relevant.*
   1) Discard articles including > X organizations, where X ≈ 1-4
3) Run the entire article through LIWC (or another “bag of words” program that counts the occurrence of words within a corpus); the program classifies text as positive and/or negative.

*if sample size permits

Bundy et al., 2021; Gamache et al., 2015; Pfarrer et al., 2010; Pollock et al., 2008; Zavyalova et al., 2012
Assessing these models

**Pros**

1) Provides reliable, consistent measurement.
2) Requires no technical knowledge or ability.
3) Is accessible and user-friendly.
4) Has become a standard in applied fields such as management.

**Cons**

1) Misattributes words to the chosen focal org that refer to other objects (i.e., “type I” errors).
2) Does not attribute words to other orgs in your sample that refer to them (i.e., “type II” errors).
3) Relatedly, we are not capturing within-article variance between coverage across multiple orgs.

*Bundy et al., 2021; Gamache et al., 2015; Pfarrer et al., 2010; Pollock et al., 2008; Zavyalova et al., 2012*
Some have created custom sentiment classifiers by:

1) Identifying articles that include the organization’s name.
2) Manually classifying a subset of sentences that include the names/keywords.
3) Training a natural language processing algorithm against the manually labeled data.
4) Applying the trained algorithm to the remaining text.

While this process appears to mitigate “type I” errors, it does little to address “type II” errors—and, as we’ll see, “type I” errors persist.

Choudhury et al., 2019; Suarez et al., 2021
An example helps illustrate the measurement error

**Headline: Following Walmart’s lead, Kroger asks customers not to openly carry firearms in stores (Washington Post; Sept 4, 2019)**

“Kroger is respectfully asking that customers no longer openly carry firearms into our stores, other than authorized law enforcement officers,” Jessica Adelman, group vice president of corporate affairs, said in a statement to CNBC on Tuesday. “We are also joining those encouraging our elected leaders to pass laws that will strengthen background checks and remove weapons from those who have been found to pose a risk for violence.”

[...]

The announcement came hours after Walmart’s. But the Bentonville, Ark.-based retailer went further, saying it would stop selling ammunition for military-style weapons and complete its exit from the handgun business. The company had been under pressure from gun-control advocacy groups, politicians and its own employees since the two store shootings. Roughly 40 white-collar workers in California walked off the job to protest Walmart’s gun policies last month, and e-commerce workers in Portland, Ore., and Brooklyn urged the company to stop selling firearms and organized a Change.org petition, which has since garnered more than 140,000 signatures.
An example helps illustrate the measurement error

**Headline:** Following Walmart’s lead, Kroger asks customers not to openly carry firearms in stores (Washington Post; Sept 4, 2019)

“Kroger is respectfully asking that customers no longer openly carry firearms into our stores, other than authorized law enforcement officers,” Jessica Adelman, group vice president of corporate affairs, said in a statement to CNBC on Tuesday. “We are also joining those encouraging our elected leaders to pass laws that will strengthen background checks and remove weapons from those who have been found to pose a risk for violence.”

**The problem(s):** While the first paragraph would suggest this article is about Kroger, it includes text that refers to another company. Additionally, the companies are referred to by words beyond their names.

The announcement came hours after Walmart’s. But the Bentonville, Ark.-based retailer went further, saying it would stop selling ammunition for military-style weapons and complete its exit from the handgun business. The company had been under pressure from gun-control advocacy groups, politicians and its own employees since the two store shootings. Roughly 40 white-collar workers in California walked off the job to protest Walmart’s gun policies last month, and e-commerce workers in Portland, Ore., and Brooklyn urged the company to stop selling firearms and organized a Change.org petition, which has since garnered more than 140,000 signatures.
In linguistics, coreference occurs when two or more expressions in a text refer to the same entity.
Coreference resolution is actually a series of independent, entity-specific neural networks.

Lee et al. (2013)
A great example from the authors:

Lee et al. (2013)
Coreference Resolution: A Practical Example

**Headline: Following Walmart’s lead, Kroger asks customers not to openly carry firearms in stores (Washington Post; Sept 4, 2019)**

“Kroger is respectfully asking that customers no longer openly carry firearms into our stores, other than authorized law enforcement officers,” **Jessica Adelman, group vice president of corporate affairs**, said in a statement to **CNBC** on Tuesday. “We are also joining those encouraging our elected leaders to pass laws that will strengthen background checks and remove weapons from those who have been found to pose a risk for violence.”

[...] “A year ago, Kroger made the conscious decision to completely exit the firearm and ammunition business when we stopped selling them in our Fred Meyer stores in the Pacific Northwest,” **Adelman** said in the statement to **CNBC**. “Kroger has demonstrated with our actions that we recognize the growing chorus of Americans who are no longer comfortable with the status quo and who are advocating for concrete and common sense gun reforms.”

Cincinnati-based **Kroger** is the second-largest grocer in the United States and counts Harris Teeter and Mariano’s among its many brands. **It** operates nearly 2,800 grocery stores in 35 states, many of which also include on-site pharmacy and jewelry stores.
Coreference Resolution: A Practical Example (continued)

*Headline: Following Walmart’s lead, Kroger asks customers not to openly carry firearms in stores (Washington Post; Sept 4, 2019)*

[...] The announcement came hours after Walmart’s. But the Bentonville, Ark.-based retailer went further, saying it would stop selling ammunition for military-style weapons and complete its exit from the handgun business. The company had been under pressure from gun-control advocacy groups, politicians and its own employees since the two store shootings. Roughly 40 white-collar workers in California walked off the job to protest Walmart’s gun policies last month, and e-commerce workers in Portland, Ore., and Brooklyn urged the company to stop selling firearms and organized a Change.org petition, which has since garnered more than 140,000 signatures.

“In a complex situation lacking a simple solution, we are trying to take constructive steps to reduce the risk that events like these will happen again,” Walmart chief executive Doug McMillon said in a memo to employees on Tuesday. “The status quo is unacceptable.”

The decision was a blow to gun rights advocates, some of whom had been showing up at Walmart locations carrying guns on their hips in the hope that the retailer would not shift its policies.
Why use coreference resolution?

Once properly trained, we can classify sentiment for multiple organizations within one article – mitigating both types of errors.

Beefing it up: By pairing coreference resolution with part-of-speech tagging, we can also identify and remove potentially irrelevant sentences, e.g.:

The decision was a blow to gun rights advocates, some of whom had been showing up at Walmart locations carrying guns on their hips in the hope that the retailer would not shift its policies.

This decision, of course, is dependent on your research question.
Let’s implement CR on our running example

**Headline:** Following Walmart’s lead, Kroger asks customers not to openly carry firearms in stores (Washington Post; Sept 4, 2019)

- 22 sentences total* (7 Kroger-specific; 7 Walmart-specific; 569 total words).

- LIWC calculates the percentage of words falling into “positive emotion” and “negative emotion”.

- Devil’s advocate: Most NLP classifiers are trained on higher levels of data than is LIWC (e.g., sentence, paragraph, corpus). So, differences in sentiment scores are expected—not because of CR, but because of the classifier.

<table>
<thead>
<tr>
<th></th>
<th>Score**</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIWC positive</td>
<td>1.41</td>
</tr>
<tr>
<td>LIWC negative</td>
<td>1.93</td>
</tr>
<tr>
<td>Walmart – positive</td>
<td>57.14</td>
</tr>
<tr>
<td>Walmart – negative</td>
<td>28.57</td>
</tr>
<tr>
<td>Kroger – positive</td>
<td>71.43</td>
</tr>
<tr>
<td>Kroger - negative</td>
<td>14.28</td>
</tr>
</tbody>
</table>

*Sentence classifiers can sometimes consider multiple sentences as just one if there is continuity between them (e.g., a quotation by a person that extends to two sentences).

**LIWC scores represent the percentage of words in the article reflecting the corresponding variable; the scores reported for this analysis have thus been comparably manipulated.
Let’s implement CR on our running example

<table>
<thead>
<tr>
<th></th>
<th>Score</th>
<th>Z-score*</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIWC positive</td>
<td>1.41</td>
<td>-0.814</td>
</tr>
<tr>
<td>LIWC negative</td>
<td>1.93</td>
<td>0.216</td>
</tr>
<tr>
<td>LIWC (W) positive</td>
<td>0.84</td>
<td>-1.270</td>
</tr>
<tr>
<td>LIWC (W) negative</td>
<td>1.26</td>
<td>-0.374</td>
</tr>
<tr>
<td>LIWC (K) positive</td>
<td>1.98</td>
<td>-0.360</td>
</tr>
<tr>
<td>LIWC (K) negative</td>
<td>1.58</td>
<td>-0.092</td>
</tr>
</tbody>
</table>

The negative affect appears to stem from parts of the article not specifically related to Kroger or Walmart.

Even when we run the parsed data through LIWC, we see variance in its scores, suggesting that coreference resolution leads to differential evaluations of media coverage within an article.

*Based on a subsample of 226,126 articles published in the third quarter of 2019 across ~30 media outlets
When does this variance matter?

- When you’re working with text sources that can be attributed to multiple focal entities.
  - In other words, this is not a problem with CEO/firm communications such as letters to shareholders, quarterly earnings calls, tweets, etc.

- When you’re studying behaviors that are more likely to have varied coverage.
  - In my case, sociopolitical acts by firms are more likely to be covered differentially by media outlets (even within one article).

- When you’re studying behaviors that incorporate a breadth of organizations.
  - E.g.: If you’re studying product launches, it is less likely that a given article discussing the launch will feature multiple companies (and, if multiple companies are mentioned, it is unlikely that their mentions are substantive).
Closing thoughts

Garbage in, garbage out: The importance of having quality data, and preprocessing it properly, can never be overstated.

Know what you’re doing: Make sure you know what’s going on “under the hood,” and be able to explain it (especially to reviewers).
Thank you!
References


