Eric Shen
Sharona Badloe

Language as Data
Assignment 2

Sentiment annotation guidelines

1. Introduction
For this project, we annotate sentences from two tech-news websites: The News Lens and Techradar. We have compiled 100 articles per website using the keywords "artificial intelligence", and from those articles we pulled 60 sentences that contain any of our three target terms "Privacy," "Security" and "Data". The sentences will be annotated on sentiment according to the following guidelines.

2. Labels and explanation
Each sentence will be annotated on three possible sentiments: positive, negative and neutral. For this project, only the language that the speaker uses will be considered. We will not make any inferences about the speaker's actual emotional state in our annotations.

The following sentiment questionnaire found in A practical Guide to Sentiment Annotation (saif 2016), will be used as a base approach:

What kind of language is the speaker using?

1. The speaker is using positive language, for example, expressions of support, admiration, positive attitude, forgiveness, fostering, success, positive emotional state.

2. The speaker is using negative language, for example, expressions of criticism, judgment, negative attitude, questioning validity/competence, failure, negative emotion.

3. The speaker is using expressions of sarcasm, ridicule, or mockery.

4. The speaker is using positive language in part and negative language in part.

5. The speaker is neither using positive language nor using negative language.

Option 1 will be labelled as positive. Option 2 will be labelled as negative. Option 5 will be labelled as neutral. Options 3 and 4 will be judged on a case by case basis, discussed further in section 4: problematic cases.

3. Examples
First we will provide some straightforward examples, and then we will discuss some of the more problematic ones.

Negative sentences:
Negative sentiment is fairly easy to spot in our dataset. The following sentence is labelled as negative because it expresses judgement towards organizations and their data policies:

"Mounting privacy concerns and customer skepticism could be a big hurdle and seriously slow down the progress of AEI application in marketing."

Positive sentences:
Positive sentences in our dataset either express positive feelings. The following sentence is unambiguously positive because it praises AI:

"This is where artificial intelligence (AI) comes into play â€" it can benefit enterprises with their data in three fundamental ways."

Neutral sentences:
We consider neutral sentences, those that use neither positive nor negative language. These sentences are, in our dataset, often explanations such as the following:

"This is important because incomplete data will result in incomplete insights."

4. Problematic cases
These are the approaches we agreed upon when dealing with some of the more ambiguous cases.

Sentences that use both positive and negative language:
In sentences that express two sentiments, the two parts of the sentence are often linked with a transition word. We then assume that the part of the sentence that the transition word refers to is the more important one linguistically. We therefore annotate the sentiment expressed in that part of the sentence. For instance, in the following sentence the transition word 'while' refers to 'it also creates risks.' Therefore, the sentence will be annotated as negative.

"While extracting value from data is undoubtedly beneficial for organizations, it also creates risks."

Marketing/selling phrases:
Our dataset has a lot of sentences geared towards selling a product. We noticed that some annotators are inclined to annotate these kinds of sentences as neutral, because the speaker might not truly experience positive emotions when repeating a slogan or selling a product. However, since we are not taking the speaker's emotional state into account, we look at only the language that is being used. Since a commercial sentence like the following expresses praise towards a product, it will be annotated as positive.

"Tapping into a wealth of new and archived footage, player data and historical records, fans can now benefit from personalized recommendations and highlights reels."

Sarcasm:
Sentences that contain sarcasm will be either positive or negative depending on the content of the sentence. For these sentences, we will trust in the annotaters speaker's intuition to make the right choice on a case by case basis.

Citations

Mohammad, Saif. (2016). A Practical Guide to Sentiment Annotation: Challenges and Solutions. 174-179. 10.18653/v1/W16-0429