

Homework 10

Name: Hrishikesh Telang

SUID: 889489533

Questions:

1. Extraction – (Please review: Entities, Keywords, Concepts, Relations)

a. How accurate and effective was the extraction process?

The extraction process looks accurate except in the Entities section wherein it misclassifies 'Indian' as location and not a person. In terms of keywords, it also doesn't consider 'Black people' as one of the Concepts whereas it assumes 'Stanford University' to be one.

b. Did it miss anything important?

In the keyword section, it misses 'bias' as an important keyword. Also AI is a concept and not a keyword, thus, the concept misses 'AI' as the part of the list.

c. Were you surprised by anything that it picked up on?

I am surprised by how the extraction tool perfectly captures all the relationships between the entities. The tool does it perfectly.

2. Classification – (Please review: Sentiment and Emotion)

a. Sentiment

i. Which entities were positive and which were negative?

Positive Entities: East Asian, Barcelona Spain, Timnit Gebru, Stanford University, A.I.

Negative Entities: AI Community, Margaret Mitchell, Bloomberg News, Microsoft

ii. Why do you think the classifier scored each the way it did?

The classifier checks the score what is given to be termed as the 'Sentiment Analysis' of the classifier. It generally involves analyzing the general semantic processing of the text towards certain target phrases. If those phrases have a positive tone, the tool categorizes them as positive entities. On the other hand, any negative connotations of the sentences that have a critical

or derisive remark about a particular object are classified as negative entities.

b. Emotion

i. For the full document,

1. Which emotion were most predominant?

Joy was the most predominant emotion as it contributed to 44.74% of the emotion in the document.

2. Why do you think they were scored as they were?

This is because the classifier majorly decides the mood of the overall sentences based on the kind of keywords it uses. The classifier compares these keywords with the existing target parameter. This extraction tool analyzes sadness, joy, fear, disgust, and anger of the text, and since the overall document does not have words that have a critical emphasis and has a more informatory tone, the tool classifies the emotion as joy, however it is less than 50%.

ii. For individual entities

1. Which entities tended to have the highest Sadness and why?

Bloomberg News tended to have the highest sadness with its percentage as 64.53% because the keyword was more closely associated with 'suffering' and 'problem' which contributed to sadness scale much more than usual.

2. Which entities tended to have the highest Joy and why?

The word 'artificial intelligence' has the highest joy with its percentage as 89.08%. In the text, is more closely associated with 'improved' and 'important' which contributes to the high Joy percentage denoted to it by the extraction tool.

3. Which entities tended to have the highest Fear and why?

The word 'AI Community' has the highest fear with its percentage as 39.92% because the keyword is more

closely associated with 'insularity and 'arrogance' which contributes to the high fear percentage denoted to it by the extraction tool.