NAME: AVISHEK MUKHERJEE

EMAIL: amukherjee3@albany.edu

DATE: 3/7/2019

PAPER TITLE: Semi-Supervised recognition of Sarcastic sentences in Twitter and Amazon

1. What is the main problem or issue that the authors are addressing?

Ans: The authors are trying to make a semi-supervised algorithm for the identification of Sarcasm on Twitter and Amazon. Sarcasm is a complex form of speech in which the statements are made in an implicit way, with the intent to annoy someone. State of the art NLP algorithms are unable to detect sarcasm because of the positive connotations of the words used to make an implicit meaning.

1. Provide a short summary of the authors’ approach/argument.

Ans: The author uses 5.9 million tweets collected from Twitter and 66,000 product reviews from Amazon, using the Amazon Mechanical Turk. They created what they call as gold-standard sample by getting the data annotated by 3 annotator obtaining a F-SCORE of 0.78 on the product reviews dataset and 0.83 on the Twitter Dataset. They also used the Hashtag of Sarcasm to determine the sarcasm in the data.

They used a semi-supervised algorithm; the input is a relatively small seed of labelled sentences from 1 to 5, where 5 is explicitly sarcasm. They also used the assumption on the Amazon data, that sarcastic sentences usually appear in groups with other sarcastic sentences. To assign score to each statement they also used the KNN (K-Nearest Neighbor Strategy to label the data). They also used ‘Inter Annotator Agreement’ and reduced the 1 to 5 classification into binary classification, where 1 to 2 is non-sarcastic, 3 as hint of sarcasm, 4-5 as clearly sarcastic.

Star-Sentiment Baseline: Many studies on Sarcasm suggest that it arises from the gap between expected and actual utterances. They implemented a baseline designed to capture the notion of sarcasm as reflected by the models. There was a strong correlation between words with strong positive connotation and sarcasm. For example, ‘top’, ‘excellent’, ‘exciting’, ‘scintillating’ etc.

The model performs on the Amazon dataset with an accuracy of 0.947 on the ROC AUC curve. For the Twitter dataset, which is less structured the precision is high as 0.727, although the recall drops to 0.436 and the F-Score is 0.545.

1. What are the main strengths and/or weaknesses of the approach?

Ans: The weakness of the approach lies in the fact that 100% Inter-Annotator agreement couldn’t be obtained and given the subjective nature of sarcasm there is no way to clearly identify if a sentence is Sarcasm or not. Moreover, the Twitter dataset is clearly tagged with hashtag and the Amazon could be compared to the ratings.

In inter-personal communication, sarcasm is accompanied by behavior like shrugging, eye-rolling, smiling, laughing etc. Humans write as they speak, and as a result these qualities get muted. In this scenario, how effective would it be to identify sarcasm in real-world, as we can see that the model performs poorly on the Twitter dataset, and attributes it to the unstructured sentences etc.

1. Provide at least 1 question regarding the paper that you’d like to address during class discussion.

Ans: I would like to discuss how sarcasm recognition for review ranking and brand monitoring can be applied to inter-personal sarcasm identification. Because in conversation we neither use rating nor do we use a hashtag that made it easier to identify sarcasm.