Aspect of Blame in Tweets: A Deep Recurrent Neural Network Approach

Herman Wandabwa*, M. Asif Naeem, Farhaan Mirza
School of Engineering, Computer & Mathematical Sciences
Auckland University of Technology
Auckland, New Zealand
herman.wandabwa@aut.ac.nz, mnaeem@aut.ac.nz, farhaan.mirza@aut.ac.nz

ABSTRACT

Twitter as an information dissemination tool has proved to be instrumental in generating user curated content in short spans of time. Tweeting usually occurs when reacting to events, speeches, about a service or product. This in some cases comes with its fair share of blame on varied aspects in reference to say an event. Our work in progress details how we plan to collect the informal texts, clean them and extract features for *blame detection*. We are interested in augmenting Recurrent Neural Networks (RNN) with self-developed association rules in getting the most out of the data for training and evaluation. We aim to test the performance of our approach using human-induced terror-related tweets corpus. It is possible tailoring the model to fit natural disaster scenarios.

Keywords

Aspect extraction; Recurrent Neural Networks, Deep Learning, NLP

1. INTRODUCTION

Twitter[1] has risen to position itself as a de facto information dissemination and citizen journalism tool. This presents a perfect chance for information flow from certain entities (users) to other users in their networks. Criticism and blame exist on the platform especially for those with contrary opinions. Merriam-Webster dictionary [2] defines blame as an expression of disapproval or reproach or simply the thought of someone taking responsibility for something bad.

Our research aims at leveraging on aspect extraction techniques especially in tweets by precisely identifying what the blame aspects are in the contents. We aim to evaluate the approach in a terror event related tweets corpus that we developed. It's worth mentioning that with the availability of more data, other natural events e.g. earthquakes can also be learnt. Information relayed at such times is normally laden with sentimental emotions more so anger and despair. There will always be a tendency to pass the blame to entities involved in causing it or displeasure in measures being undertaken to deal with the situation.

Our contribution in this progressive research can be summarized as follows. First we'll present our own Blame Aspect Related Twitter(BART) dataset. This was developed preceding terror attacks in Kenya (Mpeketoni), Pakistan (Peshawar) and Australia (Sydney). Secondly, we also propose a deep RNN based model to extract and classify explicit blame related aspects. The rest of the paper is organized as follows. Section 2 reviews related work while section 3 presents our approach as well as the algorithm. We evaluate an example as proof of concept in section 4 while the conclusion and future work is elicited in the last section.

[1] `www.twitter.com`
[2] `https://www.merriam-webster.com/dictionary/blame`

WWW 2017 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3051157

2. LITERATURE REVIEW

Our work closely relates to the use of Twitter in crises as well as in detection of reference aspects in tweets. To the best of our knowledge there is no specific work tackling blame aspects extraction and classification in such a scenario via deep RNNs. Works by [1] closely relate to what we envision less for the methodology. Blame classification is subjectively related to *sentiment analysis*, an area that has been researched on for quite a while more so for a Twitter related perspective. Deep learning for crisis response using social media information is one area of application [4]. Aspects and sentiments from tweets can be extracted via a hybrid approach, leveraging *hashtags* and sentiment lexicon as well as by deep learning [6],[5],[2].

3. PROPOSED APPROACH WITH FRAMEWORK AND ALGORITHM

Our aspect model comprises of a two layered neural network. The first layer is a fully-connected layer and the second outputs a softmax distribution. Our training set is not very extensive as it just has about 5000 sentences thus difficult to train word vectors on. The Google News 300 dimensional word vectors [3] as well as vectors of our labeled data will be the choice.

3.1 Aspect Extraction and Classification

Each tweet is represented as an average of the word vector i.e. as is the case in the bag of words model. Output of the model is a probability distribution over the blame aspects. We identified 5 aspects in the context of the dataset. Key

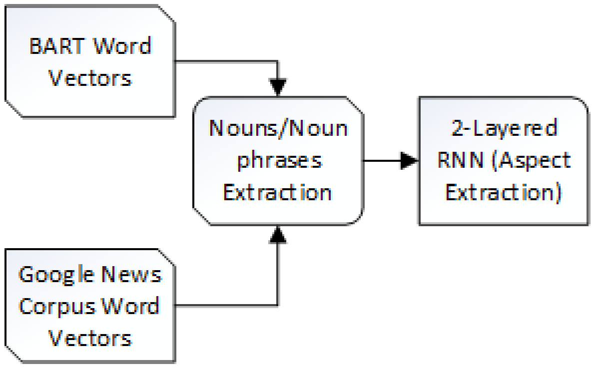

Figure 1: Aspect Extraction Model. Inputs are embeddings trained on Google News and BART Word Vectors. POS Tagger extracts nouns and noun phrase to be learnt by a 2-Layered RNN

ones are "GOVERNMENT","TERRORISTS","AID AGENCIES","OTHERS". "NONE" aspect is introduced as prediction in sentences where the aspect output is below the threshold. The output say w is defined as $w_i = \frac{1}{x}$ where x is the total number of aspects per sentence and i as the specific aspect in the sentence. Otherwise $w_i = \theta$. A threshold has to be exceeded in output w_i for a prediction to be made that the sentence has aspect i. Our Blame Aspect Related Twitter dataset is already labeled in this context. Classification of any kind is not in our scope in at the moment. We'll use the Stanford POS Tagger to identify nouns as well as noun phrases which are our aspect targets.

3.2 Aspect Model Summary

The model in Figure 1 summarizes our approach. Inputs are Google News word vectors and the Blame Aspect Related Twitter(BART) dataset. The are fed in the Stanford POS Tagger for noun and noun-phrase extraction. The 2-layered Recurrent Neural Network is where prediction output is computed over the defined aspects as in Algorithm 1.

Algorithm 1 Aspect Extraction

Input: Google and BART word vectors.
Output: Explicit aspects (w)
1: **for** $i = 1; i \leq I$ **do**
2: (i) Input features to the RNN for learning.
3: (ii)Predicted output w given $w_i = \frac{1}{x}$, x = sum of
4: all explicit aspects in a sentence.
5: Repeat step 3 until the model converges or a fixed
6: number of iterations reaches.

4. EVALUATION EXAMPLE

An sample evaluation of the below tweets is as follows.

1. `Just like westgate terrorist attack the government is not telling us the truth about MpeketoniAttack`
2. `Yet I acknowledge there is politics here alShabaab tapping into strong underlying narratives of exclusion`

Each tweet features and feature weights are learnt in vector form via RNN layers. Association rules then match the most relevant aspects in the tweets to the defined aspects through probability distribution of the tagged aspects. Our assumption is that all the tagged aspects are explicit. Expected output should be as below:-

```
1: westgate"NONE" , terrorist
"TERRORISTS" governments["GOVERNMENT"],
mpeketoniattack["NONE"]
2: politics["GOVERNMENT"],alshabaab
["TERRORIST"],narratives"NONE"
```

5. CONCLUSIONS AND FUTURE DIRECTIONS

Deep Recurrent Neural Networks have shown good performance when it comes to NLP especially in sentiment analysis. Tweets are sparse in nature i.e. their word dictionaries are not equivalent to conventional grammatically correct wordings thus quite a challenge when subjected to classifiers. We are confident that our approach when fully tested will perform well compared to other baseline approaches. We are in the process of finishing up the implementation and testing phases where full results will be published. This will be important in disaster management by identifying pertinent aspects in tweets in both blame and non-blame scenarios. Our scope is narrow at the moment i.e. human induced terror related events. We aim to broaden this to related aspects i.e. natural disasters after we gather enough datasets to train an all inclusive model.

6. REFERENCES

[1] N. D. Gitari, Z. Zuping, and W. Herman. Detecting polarizing language in twitter using topic models and ml algorithms. *International Journal of Hybrid Information Technology*, 9(9):211–222, 2016.

[2] K. W. Lim and W. L. Buntine. Twitter opinion topic model: Extracting product opinions from tweets by leveraging hashtags and sentiment lexicon. *CoRR*, abs/1609.06578, 2016.

[3] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[4] D. T. Nguyen, S. Joty, M. Imran, H. Sajjad, and P. Mitra. Applications of online deep learning for crisis response using social media information. *arXiv preprint arXiv:1610.01030*, 2016.

[5] B. Wang and M. Liu. Deep learning for aspect-based sentiment analysis, 2015.

[6] N. Zainuddin, A. Selamat, and R. Ibrahim. Improving twitter aspect-based sentiment analysis using hybrid approach. In *Asian Conference on Intelligent Information and Database Systems*, pages 151–160. Springer, 2016.