Supplementary File for Authorship Identification of Microtext Using Capsule Network

1 Analysis

We have performed a rigorous analysis of heat-maps created by our developed systems, in order to study the important words for the identification. Pre-processing is hurting the performance of our model, thus we have not removed the URLs given in the tweet. Presence of URLs can also be helpful in learning the writing style of the author. Thus, we have not trimmed out the URLs present in the tweet. For both of the categories of authors, i.e., bot-like, and human-like, we have discussed the heat-maps below.

1.1 Bot-like Authors

We analysed the heat map for two bot-authors which are mentioned below:

1.1.1 For Author-1

The tweets for this author are similar to great extent, all the words are almost repeated with just change in the last number which is also common for all the tweets. For this author, we have analysed heat maps for our M-I-B, M-I-C, M-II-C, M-II-D, M-II-B and M-II-A models which are shown in Figure 1. In this way, it can be seen that, our developed models learn the writing style of the user based on the words present in it.

1.1.2 For Author-II

The tweets for this author are similar but there are some variations too. In all the tweets, author repeats some of the words and a URL at the end but author also has a unique style of writing a name in the middle of his tweet. This name can be different for different tweets (we came to know about this unique style of author by going through his tweets manually). For this author also we have analysed heat maps for our M-I-B , M-I-C , M-II-C , M-II-D , M-II-B and M-II-A models which are shown in Figure 2.
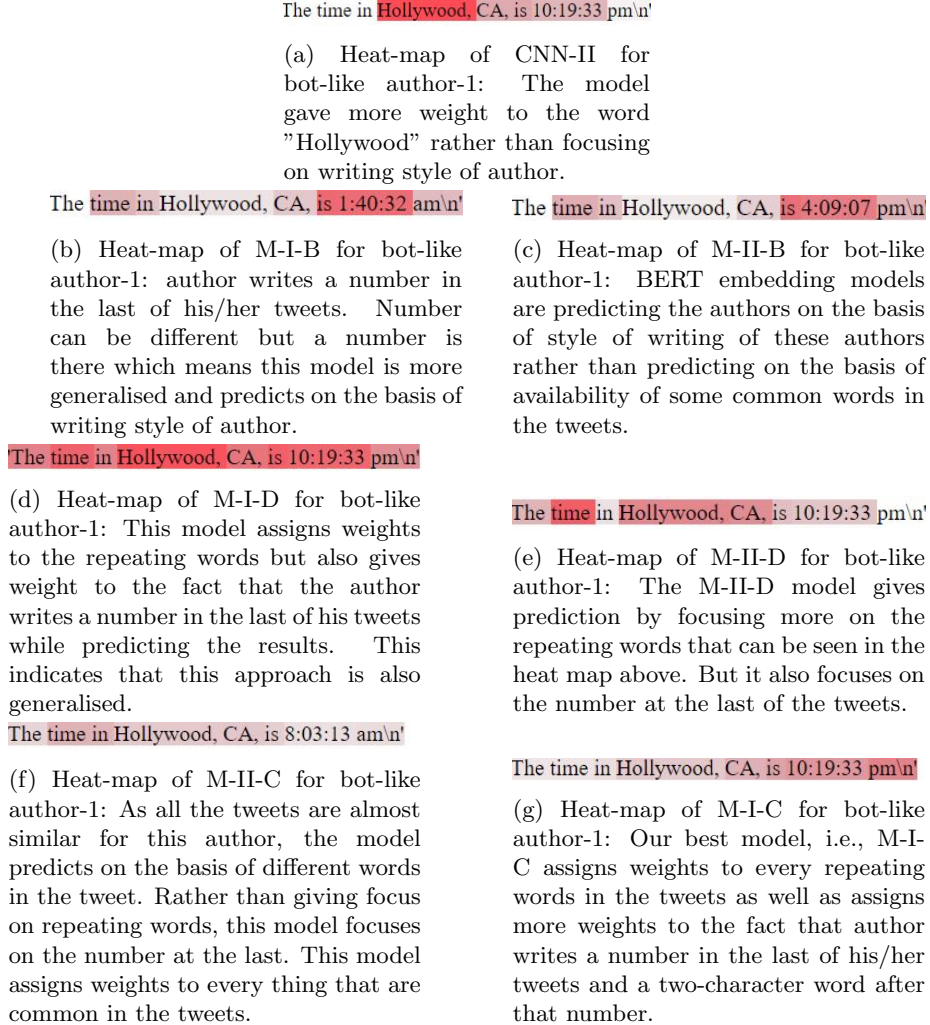
The time in Hollywood, CA, is 10:19:33 pm\n'

(a) Heat-map of CNN-II for bot-like author-1: The model gave more weight to the word "Hollywood" rather than focusing on writing style of author.

The time in Hollywood, CA, is 1:40:32 am\n'

(b) Heat-map of M-I-B for bot-like author-1: author writes a number in the last of his/her tweets. Number can be different but a number is there which means this model is more generalised and predicts on the basis of writing style of author.

The time in Hollywood, CA, is 10:19:33 pm\n'

(d) Heat-map of M-I-D for bot-like author-1: This model assigns weights to the repeating words but also gives weight to the fact that the author writes a number in the last of his tweets while predicting the results. This indicates that this approach is also generalised.

The time in Hollywood, CA, is 8:03:13 am\n'

(f) Heat-map of M-II-C for bot-like author-1: As all the tweets are almost similar for this author, the model predicts on the basis of different words in the tweet. Rather than giving focus on repeating words, this model focuses on the number at the last. This model assigns weights to every thing that are common in the tweets.

The time in Hollywood, CA, is 4:09:07 pm\n'

(c) Heat-map of M-II-B for bot-like author-1: BERT embedding models are predicting the authors on the basis of style of writing of these authors rather than predicting on the basis of availability of some common words in the tweets.

The time in Hollywood, CA, is 10:19:33 pm\n'

(e) Heat-map of M-II-D for bot-like author-1: The M-II-D model gives prediction by focusing more on the repeating words that can be seen in the heat map above. But it also focuses on the number at the last of the tweets.

The time in Hollywood, CA, is 10:19:33 pm\n'

(g) Heat-map of M-I-C for bot-like author-1: Our best model, i.e., M-I-C assigns weights to every repeating words in the tweets as well as assigns more weights to the fact that author writes a number in the last of his/her tweets and a two-character word after that number.

Figure 1: Heat-maps generated from different models for bot author:1

1.2 Human-like authors

Analysing heat maps for real users is a tough task, as to find correlation for the tweets of a real author is difficult. We have gone through different tweets of an author (let say author-1) for whom we are analysing the results. Some of the points that were common in tweets are i) the words like 'girl', 'toy' were frequently used in the tweets; ii) the author also has a unique style that is writing same words more number of times in his tweets, i.e., the same word is repeated in the author's tweets. We have analysed heat maps for our M-I-B, M-I-C, M-II-C, M-II-D, M-II-B and M-II-A models which are discussed in Figure

Now playing: Phantasia - Hold Me Now. Tune in: http://stream.laut.fm/eurodance.m3u'

(a) Heat-map of CNN-II for bot-like author-2: For CNN-II framework the model assigns more weight to the word "Now" which is also common in tweets of this author and some weight is also assigned to the style of the author in writing a URL in the last of his/her tweets.

Now playing: Axel Coon - Close to You. Tune in: http://stream.laut.fm/eurodance.m3u\n'

(b) Heat-map of M-I-B for bot-like author-2: The M-I-B model also assigns weight to the name "Axel" which is unique for that author; along with this model also assigns weights to the repeating words and the style of author mentioning a URL in the last. This model is generalized for the tweets of that author and predicts that author after considering the overall style of writing tweets.

Now playing: Axel Coon - Close to You. Tune in: http://stream.laut.fm/eurodance.m3u\n'

(c) Heat-map of M-II-B for bot-like author-2: The results in this case, are also similar to the results for the model M-I-B; the model has also learned that the author writes names in his tweet and the model predicts on the basis of writing a URL in the last by the author.

Now playing: Sash! - I Believe. Tune in: http://stream.laut.fm/eurodance.m3u'

(d) Heat-map of M-I-D for bot-like author-2: The M-I-D the model learns the fact that authors mention names in the tweet but during prediction it does not assign weight to the fact that author mentions a URL in his/her tweets at the last, but assigns weights to the repeating words by the author.

Now playing: Mario Lopez - Alone. Tune in: http://stream.laut.fm/eurodance.m3u'

(e) Heat-map of M-II-D for bot-like author-2: The M-II-D model attains results similar to M-I-D model. While predicting the author, it assigns more weight to the repeating word "playing" but fails to recognise the habit of the author in writing URL in the last.

Now playing: Sash! - I Believe. Tune in: http://stream.laut.fm/eurodance.m3u\n'

(f) Heat-map of M-II-C for bot-like author-2: The M-II-C model assigns more weight to the URL appearing at the end as well as assigns weights to the repeating words. This model was also able to recognise the unique characteristic of authors in mentioning names in the tweets.

Now playing: Mario Lopez - Alone. Tune in: http://stream.laut.fm/eurodance.m3u\n'

(g) Heat-map of M-I-C for bot-like author-2: The M-I-C model assigns weight to the name and the URL mentioned at the last which is common for all the tweets of the author; this model is more generic and does not predict on the basis of repeating words in the tweets.

Figure 2: Heat-maps for bot-like-author-2

3.

'FREE Girls Bicycle: Free Girl's purple bicycle. Roadmaster MT CLimber with 21 inch rims. For girls 6-14 years old...

(a) Heat-map of CNN-II for Human-like author-1: The CNN-II model predicts the author on the basis of the words he/she uses frequently; here the word "Girl" is assigned more weight as it is used by the author frequently but as mentioned above the author repeated the word "girl" thrice number of times in his tweet but all the words are not given equal weight-age by the model.

FREE Girls Bicycle: Free Girl's purple bicycle. Roadmaster MT CLimber with 21 inch rims. For girls 6-14 years old...

(b) Heat-map of M-I-B for Human-like author-1: The M-I-B correctly finds the unique characteristic of the author in repeating similar words in the tweet; in this tweet the "Girl" word is used thrice and the model has assigned equal weight-age to all the occurrences of the word.

'FREE Girls Bicycle: Free Girl's purple bicycle. Roadmaster MT CLimber with 21 inch rims. For girls 6-14 years old...

(c) Heat-map of M-II-B for Human-like author-1:Similar to the M-II-A, this model also assigns equal weight-ages to all the occurrences of the same word while predicting the author and this also means the model learns the style of writing of the author.

free day bed: free day bed come pick up 29107 loddington st spring tx fox run first come gets it ill erase when .. http://bit.ly/7VFd8

(d) Heat-map of M-I-D for Human-like author-1: The M-I-D model also assigns weight to the repeating word "free" which is repeated twice in the tweet; along with this, weight is also assigned to the URL written in the last which is common in some of the tweets of the author.

free items (copperfield/northwest): chair, old laptop, dvd,vcr,printer,toys, and various items...see the picture.

(e) Heat-map of M-II-D for Human-like author-1: The M-II-D model assigns weight to the word "toy" which was very frequent in the tweets for this author. So this model predicts on the basis of occurrences of words in the tweet which are frequent considering all other tweets.

old toys (Seabrook): Girl toys. Old. Unsure of condition. Bagged and ready. I can drop off if you live close enough... http://bit.ly/6UjiVu

(f) Heat-map of M-II-C for Human-like author-1: This model also assigns more weight to the frequent words such as "Girl" and "Old". Similar to M-II-D model, current one also predicts the author on the basis of occurrences of such frequent words.

free day bed: free day bed come pick up 29107 loddington st spring tx fox run first come gets it ill erase when .. http://bit.ly/7VFd8

(g) Heat-map of M-I-C for Human-like author-1: The model assigns weight to the repeating word that is "day"; it also assigns weight to the URL mentioned in the last which is common in other tweets of the author.

Figure 3: Heat-maps for Human-like author-1