#DemocratsAreDestroyingAmerica: Rumour Analysis on Twitter During COVID-19

Lin Tian[a], Xiuzhen Zhang*[a] and Jey Han Lau[b]

[a]RMIT University, Melbourne, Australia
[b]The University of Melbourne, Melbourne, Australia

Abstract

COVID-19 has brought about significant economic and social disruption, and misinformation thrives during this uncertain period. In this paper, we apply state-of-the-art rumour detection systems that leverage both text content and user metadata to classify COVID-19 related rumours, and analyse how users, topics and emotions of rumours differ from non-rumours. We found that a number of interesting insights, e.g. rumour-spreading users have a disproportionately smaller number of followers compared to their followees, rumour topics largely involve politics (with an abundance of party blaming), and rumours tend to be emotionally charged (anger) but *reactions* towards rumours exhibit disapproving sentiments.

Keywords

Rumour Detection, Rumour Analysis, COVID-19, Twitter

1. Introduction

COVID-19, a novel disease that was first identified in China, is an ongoing pandemic that has brought about significant impact to global economy and created hitherto unseen social disruption. Since late Feburary 2020, the pandemic has come to dominate both traditional news and social media platforms,[1] and misinformation such as fake news, conspiracy theories and rumours thrive during these uncertain times [1].

For example, in Italy we saw rumours being spread to blame the outbreak on migrants and refuges by making the implicit connection between migration/movement with the spread of the virus.[2] Hydroxychloroquine, a drug that was rumoured to be a COVID-19 treatment despite lacking robust scientific evidence about its effectiveness [2, 3], is another popular topic on social media.[3] These rumours can have serious consequences, e.g. misinformation about hydroxychloroquine has lead to the death of a man in Arizona.[4]

Social media provides a perfect platform for misinformation propagation as they are largely unregulated. To identify misinformation or fake news, we may rely on general fact-checking websites,[5] or COVID-19 specific ones.[6] However, due to the evolving circumstances of a pandemic it is unlikely fact-checking or debunking websites will have the capacity to keep themselves up-to-date.

As such, early detection of potentially malicious rumours and understanding what or how rumours are being spread during a crisis is an important task [4]But what is a "rumour"? We adopt a widely used definition which defines it as *a story or a statement with unverified truthful value* [5].

In this paper, we seek to understand what sorts of COVID-19 rumours are being spread on Twitter. To this end, we train state-of-the-art rumour detection systems on out-of-domain labelled rumour data and apply them to COVID-19 related tweets to detect rumours. We analyse several characteristics that differentiate rumours from non-rumours in this COVID-19 data, such as their propagation patterns, users, topics, and emotions. Our rumour detection systems leverage both message content and user characteristics, and our analyses reveal a number of interesting insights. For example, rumour-speaders

The 5th International Workshop on Mining Actionable Insights from Social Networks (MAISoN 2020) - Special edition on Dis/Misinformation Mining from Social Media
EMAIL: s3795533@student.rmit.edu.au (L. Tian);
Corresponding author: xiuzhen.zhang@rmit.edu.au (X. Zhang);
jeyhan.lau@gmail.com (J.H. Lau)
ORCID: 0000-0001-5558-3790 (X. Zhang*)

[1]https://www.vox.com/recode/2020/3/12/21175570/coronavirus-covid-19-social-media-twitter-facebook-google.

[2]https://time.com/5789666/italy-coronavirus-far-right-salvini/.

[3]https://abcnews.go.com/Health/tracking-hydroxychloroquine-misinformation-unproven-covid-19-treatment-ended/story?id=70074235.

[4]https://edition.cnn.com/2020/03/23/health/arizona-coronavirus-chloroquine-death/index.html.

[5]E.g. https://www.snopes.com/ and https://www.factcheck.org/.

[6]E.g. https://www.fema.gov/coronavirus/rumor-control and https://www.defense.gov/Explore/Spotlight/Coronavirus/Rumor-Control/

tend to have low follower but high followee count, rumours tend to talk about politics (mostly party blaming) and are more emotionally charged (e.g. anger), but reactions towards them are also disproportionately more disapproving. We also provide a website[7] to share our latest findings and up-to-date rumour tracking data analysis.

2. Related Work

Rumour detection approaches can generally be categorised into text-based or non-text-based methods. Text-based methods focus on rumour detection using the textual content, which may include the original source document/message and user comments/replies. Shu et al. [6] introduce linguistic features to represent writing styles and other features based on sensational headlines from Twitter to detect misinformation. To detect rumours as early as possible, Zhou et al. [7] incorporate reinforcement learning to dynamically decide how many responses are needed to classify a rumour.

Non-text-based methods utilise features such as user profiles or propagation patterns for rumour detection. For example, Gupta et al. [8] propose a semi-supervised approach to evaluate the credibility of tweets using hand-crafted features based on tweet and user metadata. Castillo et al. [9] leverage user registration age and number of followers to assess credibility. Following studies explore more complex features such as belief/intention for rumour prediction [10], where users are categorised based on their "support" or "deny" attitudes toward a piece of news.

In terms of emotion analysis on social media, Larsen et al. [11] propose using principle component analysis to predict emotions of tweets, and introduces a real-time system that analyses global and regional emotional signals on Twitter. More recently, Farruque et al. [12] formulate the emotion detection task as a multi-label classification problem and use an LSTM model with attention for emotion prediction.

For analysis of COVID-19 on Twitter, Li et al. [13] explore using multi-lingual BERT [14] to analyse public mental health using tweets. Sharma et al. [15] present analysis of COVID-19 misinformation based on news sources from fact-checking sites rather than automatic classification and contrast analysis of rumours versus non-rumours.

[7] https://xiuzhenzhang.github.io/rmit-covid19/

Table 1
Rumour classification training data.

	Twitter15	Twitter16	PHEME	SemEval
#source tweets	1,490	818	6,425	446
#all tweets	624,458	363,535	105,354	42,195
#users	426,501	251,799	50,593	5,666
#rumours	1,118	613	2,402	446
#non-rumours	372	205	4,022	0

3. Methodology and Data

3.1. Rumour Classification

We focus on the detection of rumours vs. non-rumours, rather than the *veracity* (truthfulness) of rumours. In other words, truthful, untruthful and unverified rumours are all rumours in our definition — they exhibit novelty/surprise in terms of content and tend to be spread by users — while non-rumours are traditional news stories and non-news related conversations. The task of rumour detection can therefore be formulated as a binary classification problem, and we explore both textual information and user metadata as input features.

Consider a set of n source tweets $S = \{s_1, s_2, ..., s_n\}$. Each source tweet is associated with a label l indicating the tweet is rumour ($l = 1$) or non-rumour ($l = 0$). Each source tweet s_i also has a set of m reactions: $R_i = \{r_{i1}, r_{i2}, ..., r_{im}\}$. Reactions are retweets, replies and quotes. Each reaction r_{ij} is represented with a tuple $r_{ij} = (t_{ij}, u_{ij})$, which includes the following information: t_{ij} is the textual content of the reaction, and u_{ij} the metadata features of the user who creates the reaction tweet.

In terms of rumour classification models, we explore two methods based on: (1) text [16]; and (2) user metadata [17]. The text-based model is implemented with BERT [14] and uses a pre-trained user stance prediction model to classify the veracity of a rumour. We adapt the model to our task which treats rumour classification as a binary classification task. For the user-based model, it uses a convolutional network to process user metadata features extracted from their Twitter profile and a recurrent network to combine a set of user features in the propagation path. We extend the original eight features to sixteen features.[8] We limit the processing of user features in the propagation path to the first 50 users.

[8] The extended integer user features are: length of user screenname, count of posts and favourite posts; and the binary features are: whether the profile is protected, has URL, profile image, uses default profile and default profile image.

Table 2
Filtered data statistics.

#tweets	30,077,742
#source tweets	60,550
#users	8,692,422
mean #reactions	497
max #reactions	165,592
mean #replies	28
max #replies	2,177

To combine both text and user models for rumour detection, we create an ensemble model that takes the output of both models to make the final prediction. As both models produce a probability value for the rumour class in each source tweet, we compute the mean probability and tune a threshold p to separate rumours from non-rumours.[9]

3.2. Labelled Rumour Data

We use Twitter15, Twitter16 [18], PHEME [19], and SemEval2019 [20] as training data to train our binary rumour classification models. For Twitter15, Twitter16 and PHEME, there are originally 4 classes: truthful rumours, untruthful rumours, unverified rumours and non-rumours; we collapse the truthful, untruthful and unverified rumours into the rumour class. SemEval2019 focuses on veracity classification and as such has only 3 classes (truthful, untruthful and unverified); they are all treated as the rumour class. Statistics of the datasets is presented in Table 1.

3.3. COVID-19 Twitter Data

We use a public COVID-19 Twitter dataset [21] for our analyses.[10] We use version 4 of the dataset, which contains tweets from 1st January 2020 to 5th April 2020. The dataset is regularly updated, and collects tweets for several languages (English, French, Spanish and German) based on COVID-19 keywords.

As we are interested in rumour analyses in English, we filter the data to keep only source tweets that are in English (based on Twitter metadata) and also have at least 10 replies (since those with few reactions are of little significance for rumour analysis). Table 2 presents some statistics of our filtered dataset. We have approximately 30M tweets post-filtering, and 60K of them are source tweets (the remaining tweets are "reaction tweets": retweets, replies or quotes).[11]

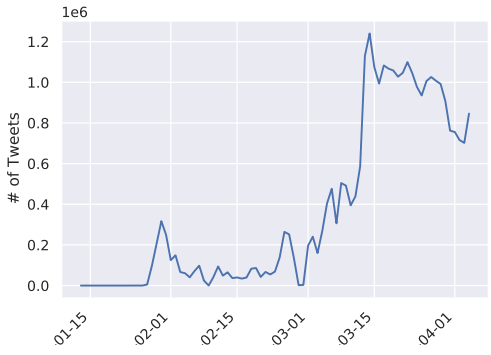


Figure 1: Filtered English Tweets Volume

Figure 1 shows the volume of filtered English tweets over time. We can see there is some traffic of COVID-19 related tweets from late January 2020, although it doesn't really pick up until mid-March. We suspect the spike of activity may be due the World Health Organisation declaring it as a pandemic on 12th March.[12].

In terms of pre-processing, we tokenise the tweets with the `TweetTokenizer` [22] package of NLTK, and lowercase and lemmatise all words with the `WordNetLemmatizer` package, as well as remove digits, non-Latin characters and @usernames. We also filter stopwords based on an extended NLTK stopword list, which includes COVID-19 specific stopwords, such as *covid19* or *coronavirus*. Hyperlinks are encoded with a special token for rumour classification (Section 4.1) or removed for topic analysis (Section 4.3).

4. Results and Analysis

4.1. Rumour Classification

To assess the quality of the rumour classification models, we first evaluate the in-domain performance of Twitter15, Twitter16 and PHEME. For each dataset, we randomly split the full data in 60%/20%/20% to create the training, validation and test partitions. In-domain classification performance is presented in Table 3 (in-domain performances are those where "Train" and "Test" are from the same domain).[13]

[9] That is, the ensemble model labels a source tweet as rumour if the mean probability $\geq p$.

[10] https://github.com/thepanacealab/covid19_twitter.

[11] Quote is similar to retweet, except that it contains some response to the original tweet. Both retweets and quotes are displayed on the user's home page, while replies are not.

[12] https://twitter.com/WHO/status/1237777021742338049

[13] For the ensemble model, we tune the threshold p based on the validation set, and p ranges from 0.7 to 0.8.

Table 3
In-domain and cross-domain classification results. "P", "T15" and "T16" denote the PHEME, Twitter15, and Twitter 16 datasets respectively.

Test	Train	Model	Accuracy
T15	T15	user	0.85
		text	0.88
		user+text	0.88
	P+T16	user	0.73
		text	0.71
		user+text	0.80
T16	T16	user	0.82
		text	0.86
		user+text	0.92
	P+T15	user	0.75
		text	0.78
		user+text	0.82
P	P	user	0.63
		text	0.92
		user+text	0.81
	T15+T16	user	0.65
		text	0.70
		user+text	0.78

Table 4
User statistics. Top half of the table is median statistics, bottom half mean.

	Rumour	Non-rumour
#Follower	151,521	223,651
#Following	1,486	976
# $\frac{\text{Follower}}{\text{Following}}$ Ratio	63	121
#Post	31,433	28,644
Account Age	2,992	3,119
Geo Enabled	51%	57%

Overall, we can see the text model does better than the user model, but the ensemble model ("user+text") performs best.

We next evaluate cross-domain performance. Given a test domain (e.g. Twitter15), we train the rumour classification models using a combination of all out-of-domain data (e.g. Twitter16 and PHEME), and assess their accuracy on the test domain. This is an arguably more difficult setting, as there is little or no topic overlap between the different domains.

Unsurprisingly, we see a dip in accuracy compared to the in-domain performance. Encouragingly, however, with the ensemble model we are still getting at least 78% accuracy over all domains, suggesting that the model is robust for cross-domain rumour detection.

Given these results, we next train an ensemble model on all datasets (Twitter15+Twitter16+PHEME), and use it to classify tweets on our filtered English COVID-19 data (Section 3.3).[14]

In total, out of the 60K source tweets (Table 2) 15K are classified as rumours. These rumours (and non-rumours) will serve as the basis for user, topic and emotion analyses in subsequent experiments.

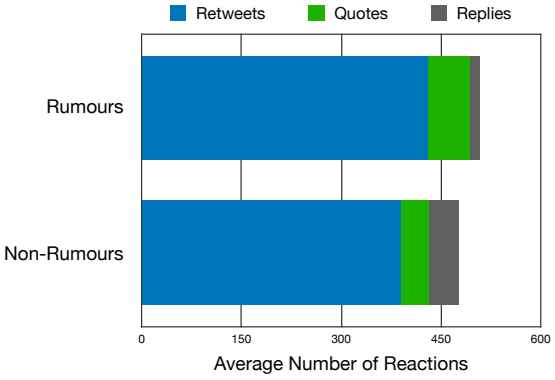


Figure 2: Reaction types.

Figure 3: Reaction speed.

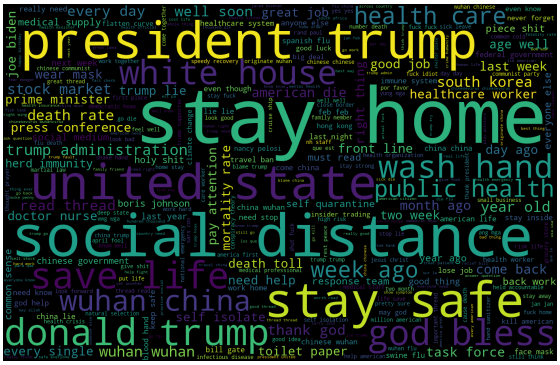
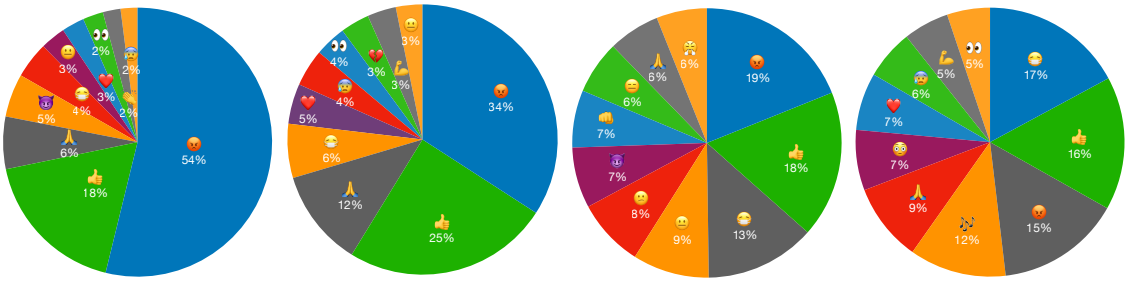


Figure 4: Bigram word cloud.

[14] We set the threshold p to 0.85, which is marginally higher than the thresholds we used in the cross-domain experiments to improve precision. Note that the COVID-19 data does not include user metadata, so we crawl them using the official Twitter API.

Table 5
Salient hashtags, unigrams and bigrams in rumour and non-rumour tweets.

Rumour	Hashtag	#WuhanVirus, #MOG, #OneVoice1, **#FoxNews**, **#DemocratsAreDestroyingAmerica**, #KAG2020, **#ChinaVirus**, **#Hydroxychloroquine**, #IWillStayAtHome, #ChinaLiedPeopleDied, #MasksNow, #TheMoreYouKnow, #TheResistance, #StopAiringTrump, #VoteRedToSaveAmerica, **#WuhanHealthOrganisation**, **#CCP_is_terrorist**, #DemCast, **#BillGates**, **#TrumpIsTheWORSTPresidentEVER**, #TrumpOwnsEveryDeath, **#5G**
	Unigram	**trump**, pelosi, bill, democrat, **fox**, gop, american, blame, president, briefing, joe, lie, **hoax**, medium, fail, governor, response, **china**, vote, drug, **hydroxychloroquine**
	Bigram	nancy pelosi, chinese chinese, jared kushner, **chinese communist**, trump response, held accountable, **trump supporter**, trish regan, speaker pelosi, joe biden, **bill gate**, **china lie**, task gown, deep state, **blame trump**, fox business
Non-Rumour	Hashtag	#BREAKING, **#StaySafe**, #CoronaUpdate, **#CoronavirusLockdown**, #IndiaFightsCorona, #CoronaOutbreak, #DonaldTrump, #COVID19PH, #COVID19Pandemic, #covid19australia, #TakeResponsibility, #21daylockdown, #CoronavirusPandemic, #Covid19usa, **#StayHomeStaySafe**, #StayAtHome, **#coronapocalypse**, #flu, #Italia, #COVID19OhioReady, #COVID_19uk, #masks, #china, **#StrongerTogether**
	Unigram	positive, confirm, total, india, march, **symptom**, health, minister, due, nigeria, **lockdown**, update, death, infect, old, donate, day, negative, cancel, **wash**, **hand**, social, hour, announce, today, data, stay, worker, isolation, **quarantine**
	Bigram	bring total, march march, year old, **total number**, relief fund, **number confirm**, patient positive, **prime minister**, travel history, premier league, **wash hand**, hubei province, first death, **cruise ship**, health condition, **social care**



(a) Rumour source tweets (b) Non-rumour source tweets (c) Rumour reply tweets (d) Non-rumour reply tweets

Figure 5: Emoji Distribution for rumour vs. non-rumour tweets.

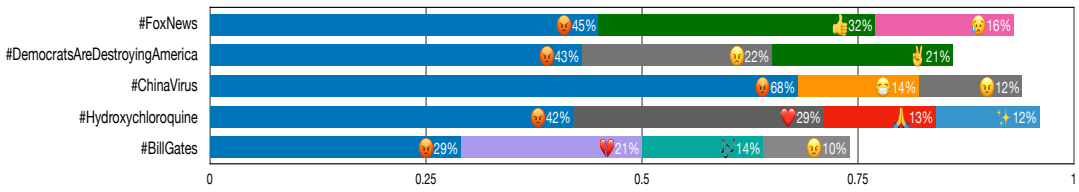
4.2. User Analysis

More than 8M users are involved in the conversations around COVID-19 in our filtered English dataset (Table 2). We focus *only* on users who published the source tweets in this analysis. Table 4 presents some statistics of these users for rumours and non-rumours.

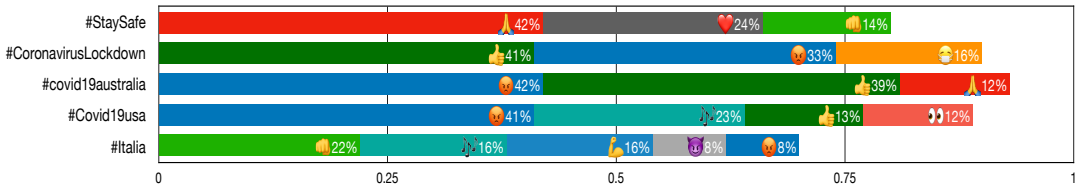
Interestingly, users who are involved in rumour creation tend to tweet more (higher post counts) and follow more users but have less followers, resulting in a substantially lower $\#\frac{\text{Follower}}{\text{Following}}$ ratio. Their account is also generally younger (7.7% rumour accounts are created during January to May 2020, as opposed to 6.1% for non-rumour accounts).

Figure 2 presents the average volume of different reactions toward rumours and non-rumours. While the majority of the reactions for both are retweets, we can see that retweets and quotes are much more popular as a response to rumours. This suggests that non-rumours tend to attract more discussion/replies than rumours.

Rumours tend to have high novelty in their content so as to attract propagation [23], and we can see this in Figure 3, which shows the average volume of

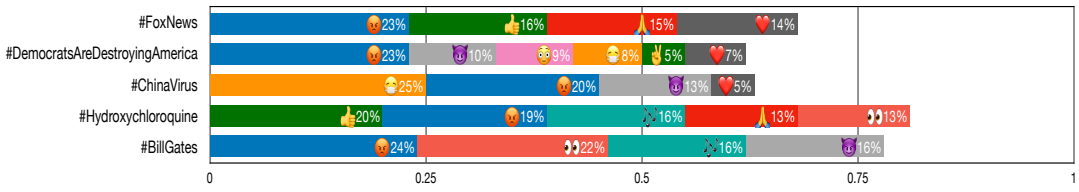


(a) Rumour

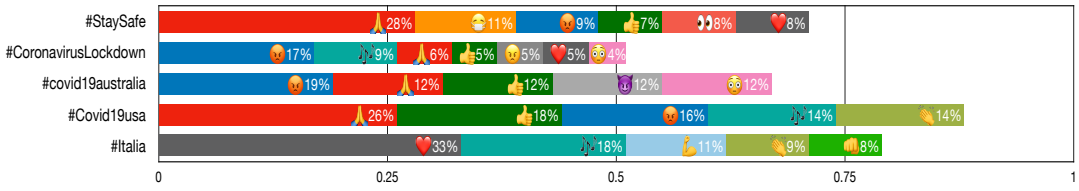


(b) Non-Rumour

Figure 6: Emoji Distribution for salient hashtags in source tweets



(a) Rumour



(b) Non-Rumour

Figure 7: Emoji Distribution for salient hashtags in responses

reactions over time for rumours and non-rumours. Although rumours tend to attract more reactions in the first 24 hours, we see a convergence after 48 hours.

4.3. Topic Analysis

To understand the popular topics discussed in Twitter, we first present a bigram wordcloud in Figure 4. We see several broad topics: (1) health advice (*social distance*, *stay home*, *wash hand*, and *wear mask*); (2) US politics (*president trump* and *joe biden*); (3) UK politics (*prime minister*, *boris johnson*, and *herd immunity*[15]); (4) blame on China (*wuhan china* and *china lie*), (5) status reports (*death toll* and *death rate*), (6) healthcare (*doctor nurse* and *health worker*); (7) panic buying (*toilet paper*[16]); and others.

To better understand the topical difference between rumours and non-rumours, we compute log-likelihood ratio [24] of unigrams, bigrams, hashtags and display the most salient words in Table 5.[17]

To ease readability, we highlight some of the salient words in the table. For rumours, US politics is one of the major topics, with both parties putting blame on each other (*#DemocratsAreDestroyingAmerica* and

[15]https://www.theatlantic.com/health/archive/2020/03/coronavirus-pandemic-herd-immunity-uk-boris-johnson/608065/.

[16]https://www.bbc.com/news/world-australia-53196525.

[17]We include both source tweets and reactions to construct the rumour and non-rumour "corpora", and use NLTK's `BigramAssocMeasures` to compute the loglikelihood ratio. To decide whether a word is salient for rumour or non-rumour, we look at its normalised frequency.

#TrumpIsTheWORSTPresidentEVER). Unsurprisingly, Fox News (*#FoxNews* and *fox*) are associated with rumours.[18] China is another topic, and the hashtags/bigrams suggest blaming (*#ChinaVirus*, *#CCP_is_terrorist*, *#WuhanHealthOrganisation* and *china lie*). We also see also some of the well-known COVID-19 rumours/hoaxes: *#Hydroxychloroquine*, *#BillGates*,[19], and *#5G*.[20]

Looking at non-rumours, the topics are very different: they are mostly related to health advice (*#CoronavisuLockdown*, *#StayHomeStaySafe* and *wash hand*) and status updates (*total number*, *number confirm*), and more neutral/positive in tone (*#StrongerTogether* and *#coronapocalypse*). Politics is rare, although we see *prime minister*, which may be related to UK politics. Another interesting non-rumour topic observed here is the cruise ship outbreaks (*cruise ship*).

4.4. Emotion Analysis

To understand the public sentiment during the COVID-19 crisis, we explore using an emotion prediction system to classify the emotion of tweets in our data. We experiment with DeepMoji [25], a Bi-LSTM with attention model trained on a large number of emoji occurrences in tweets. We use their pre-trained model to label our data with 63 predefined emojis.

Figure 5 illustrates the distribution of emojis for source and reply tweets in rumours and non-rumours. Looking at the emotions of source tweets (Figure 5(a) and (b)), "anger" dominates both rumours and non-rumours, but substantially more in rumours than non-rumours (54% vs. 34%). Non-rumours also see more "thumbs up" (encouragement), although the difference is less severe (25% vs. 18%).

For reply tweets (Figure 5(c) and (d)), we see a similar distribution for the top-3 emotions ("anger", "thumbs up" and "mask face"), but the interesting observation here is the emojis for the rest (left half of the pie chart): the reply tweets for rumours display disapproving sentiments (e.g. "punch" and "frown"), while that of non-rumours are generally positive and encouragement in tone ("pray", "love" and "biceps").

We next present the emoji distribution for some of the salient hashtags for the source and reaction tweets in Figure 6 and 7 respectively, to see how public attitude towards different topics vary across rumours and non-rumours. For rumour source tweets (Figure 6(a)), anger dominates all hashtags, although #ChinaVirus source tweets are substantially "angrier" (68%!). Anger in non-rumour source tweets (Figure 6(b)) is a little more toned down; interestingly the dominant emotion for the global lockdown (#CoronavirusLockdown) is more positive than negative (41% "thumbs up" vs. 33% "angry").

Moving over to the emoji distribution for reactions towards rumour tweets (Figure 7(a)), we see anger in all hashtags, but some of the other emotions are rather curious, e.g. "thumbs up" (approval) for #Hydroxychloroquine, and "googly eyes" (attention drawing) for #BillGates. Unsurprisingly though, reactions for all non-rumour hashtags (Figure 7(b)) are dominated by "prayers" and approval emojis ("thumbs up" and "biceps"), suggesting that despite the general doom and gloom atmosphere of COVID-19, there is still a sense of positivity.

5. Conclusion

We explored an ensemble model combining text-based and user-based rumour detection models to classify COVID-19 related rumours on Twitter. We presented quantitative evaluation to demonstrate its robustness in cross-domain rumour detection, analyse the users, topics and emotions of rumours vs. non-rumours, and found a number of insights.

Acknowledgements

This work is partially supported by the Australian Research Council Discovery Project DP200101441.

References

[1] S. Vieweg, A. L. Hughes, K. Starbird, L. Palen, Microblogging during two natural hazards events: what Twitter may contribute to situational awareness, in: Proceedings of the SIGCHI conference on human factors in computing systems, 2010, pp. 1079–1088.

[2] E. A. Meyerowitz, A. G. Vannier, M. G. Friesen, S. Schoenfeld, J. A. Gelfand, M. V. Callahan, A. Y. Kim, P. M. Reeves, M. C. Poznansky, Rethinking the role of hydroxychloroquine in the treatment of COVID-19, The FASEB Journal 34 (2020) 6027–6037.

[3] D. N. Juurlink, Safety considerations with chloroquine, hydroxychloroquine and azithromycin in the management of SARS-CoV-2 infection, CMAJ 192 (2020) E450–E453.

[18] https://www.nytimes.com/2020/03/31/opinion/coronavirus-fox-news.html.

[19] https://www.bbc.com/news/52847648.

[20] https://www.reuters.com/article/uk-factcheck-coronavirus-5g/false-claim-coronavirus-is-a-hoax-and-part-of-a-wider-5g-and-human-microchipping-conspiracy-idUSKBN22P22I.

[4] B. Wang, J. Zhuang, Crisis information distribution on Twitter: a content analysis of tweets during Hurricane Sandy, Natural hazards 89 (2017) 161–181.

[5] G. W. Allport, L. Postman, The psychology of rumor. (1947).

[6] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media: A data mining perspective, ACM SIGKDD Explorations Newsletter 19 (2017) 22–36.

[7] K. Zhou, C. Shu, B. Li, J. H. Lau, Early rumour detection, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pp. 1614–1623.

[8] A. Gupta, P. Kumaraguru, C. Castillo, P. Meier, Tweetcred: Real-time credibility assessment of content on Twitter, in: International Conference on Social Informatics, Springer, 2014, pp. 228–243.

[9] C. Castillo, M. Mendoza, B. Poblete, Information credibility on Twitter, in: Proceedings of the 20th international conference on World wide web, ACM, 2011, pp. 675–684.

[10] X. Liu, A. Nourbakhsh, Q. Li, R. Fang, S. Shah, Real-time rumor debunking on Twitter, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, ACM, 2015, pp. 1867–1870.

[11] M. E. Larsen, T. W. Boonstra, P. J. Batterham, B. O'Dea, C. Paris, H. Christensen, We feel: mapping emotion on Twitter, IEEE journal of biomedical and health informatics 19 (2015) 1246–1252.

[12] N. Farruque, C. Huang, O. Zaiane, R. Goebel, Basic and depression specific emotion identification in Tweets: multi-label classification experiments, in: The 20th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing), 2019.

[13] I. Li, Y. Li, T. Li, S. Alvarez-Napagao, D. Garcia, What are we depressed about when we talk about COVID19: Mental health analysis on tweets using natural language processing, arXiv preprint arXiv:2004.10899 (2020).

[14] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota, 2019, pp. 4171–4186.

[15] K. Sharma, S. Seo, C. Meng, S. Rambhatla, A. Dua, Y. Liu, Coronavirus on social media: Analyzing misinformation in Twitter conversations, arXiv preprint arXiv:2003.12309 (2020).

[16] L. Tian, X. Zhang, Y. Wang, H. Liu, Early detection of rumours on Twitter via stance transfer learning, in: European Conference on Information Retrieval, Springer, 2020, pp. 575–588.

[17] Y. Liu, Y.-F. B. Wu, Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[18] J. Ma, W. Gao, K.-F. Wong, Detect rumors in microblog posts using propagation structure via kernel learning, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2017, pp. 708–717.

[19] E. Kochkina, M. Liakata, A. Zubiaga, All-in-one: Multi-task learning for rumour verification, arXiv preprint arXiv:1806.03713 (2018).

[20] G. Gorrell, E. Kochkina, M. Liakata, A. Aker, A. Zubiaga, K. Bontcheva, L. Derczynski, Semeval-2019 task 7: Rumoureval, determining rumour veracity and support for rumours, in: Proceedings of the 13th International Workshop on Semantic Evaluation, 2019, pp. 845–854.

[21] J. M. Banda, R. Tekumalla, G. Wang, J. Yu, T. Liu, Y. Ding, G. Chowell, A large-scale COVID-19 Twitter chatter dataset for open scientific research–an international collaboration, arXiv preprint arXiv:2004.03688 (2020).

[22] S. Bird, E. Klein, E. Loper, Natural Language Processing with Python, 1st ed., O'Reilly Media, Inc., 2009.

[23] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, Science 359 (2018) 1146–1151.

[24] T. E. Dunning, Accurate methods for the statistics of surprise and coincidence, Computational linguistics 19 (1993) 61–74.

[25] B. Felbo, A. Mislove, A. Søgaard, I. Rahwan, S. Lehmann, Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 2017, pp. 1615–1625.