Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 6166–6173
Marseille, 11–16 May 2020
© European Language Resources Association (ELRA), licensed under CC-BY-NC

A Real-Time System for Credibility on Twitter

Adrian Iftene[1], Daniela Gîfu[1,2], Andrei-Remus Miron[1], Mihai-Ștefan Dudu[1]
Alexandru Ioan Cuza University of Iași, Faculty of Computer Science[1], Romanian Academy – Iași branch, Institute of Computer Science[2]
General Berthelot, 16, Iaşi 700483, Romania[1], Codrescu 2, Iași 700481, Romania[2]
{adiftene, daniela.gifu}@info.uaic.ro

Abstract

Nowadays, social media credibility is a pressing issue of each of us who are living in an altered online landscape. The speed of news diffusion is striking. Given the popularity of social networks, more and more users began posting pictures, information, and news about personal life. At the same time, they started to use all this information to get informed about what their friends do or what is happening in the world, many of them arousing much suspicion. The problem we are currently experiencing is that we do not currently have an automatic method of figuring out in real-time which news or which users are credible and which are not, what is false or what is true on the Internet. The goal of this is to analyze Twitter in real-time using neural networks in order to provide us key elements about both the credibility of tweets and users who posted them. Thus, we make a real-time heatmap using information gathered from users to create overall images of the areas from which this fake news comes.

Keywords: neural networks, credibility, real-time, Twitter

1. Introduction

In present, the risk of running into misinformation is not negligible, especially on Social Media. For this reason, mining the credibility of both user and message itself constitute nowadays a major problem (Gînsca et al., 2015). The speed of news diffusion is striking worldwide. Given the increasing popularity of social networks, sharing your life on the Internet has become a natural activity for most of the users. The news is read quickly, marked with opinions (see Facebook), retransmitted (retweet on Twitter, share on Facebook) without having to check many times whether they are true or false news (Atodiresei et al., 2018). Researchers have begun in recent years to address the issue of identifying fake news and their credibility on television and YouTube (Ciampaglia et al., 2015; Clark, 2009), Twitter ((El Ballouli et al., 2017; Cusmuliuc et al., 2018; Iftene et al., 2017; Castillo et al., 2011), (Chu et al., 2010), (Cook et al., 2013), (Iozzio, 2018), Facebook (Allcott, and Gentzkow, 2017), (Chen et al., 2015) even showing sites that train and help users who want to identify false news (TenQuestionsForFakeNews). There is still a lot of talk about the influence of Twitter on the US elections in 2016 (Bovet and Makse, 2019) and the fact that there is a lot of fake news on Twitter (Brummette et al., 2018).

Below we will see the most used technologies in identifying false news, part of them presented in the paper (Conroy et al., 2015). In this paper, the notion of detecting false news is defined as the task of classifying news across a continuum of veracity with an associated measure of certainty. Also, the paper provides a typology of several authenticity assessment methods that come from two major categories - linguistic approaches (with machine learning) and social networking approaches. The hybrid approach, combining machine learning in computational linguistics with social networking approaches, seems very promising and it is very used in the last years. A design system for detecting false news is not a simple matter. The most promising directions for conceiving an efficient system were almost the same after 2015: 1. *linguistic approaches*, based on involuntary “leaks” of speakers, and existing methods are trying to catch such anomalies (Mihalcea and Strapparava, 2009) focused on (a) representation of data – typically uses statistics on n-grams, which are analyzed to identify fake information (Hadeer et al., 2017; Hadeer, 2017); (b) advanced linguistic structures – sentences are transformed into more advanced forms of information representation (such as parsing trees), which then analyze probabilities attached to identify anomalies (Conroy et al., 2015), (Perez-Rosas et al., 2017); (c) semantic analysis – it analyzes semantically the contents of a user’s statements, constructs pairs of the attribute form: descriptor and calculates compatibility scores (Shu et al., 2017); (d) rhetorical structures and utterance analysis – relations between the linguistic elements are built, which help determine the proximity to the centers of truth or deception (Popoola, 2017; Rubin and Lukoianova, 2014; Rubin et al., 2016); (e) classifiers – SVM classifiers or Naive Bayesian-type classifiers are used to predict future clutter-based fraud and distances (Rubin et al., 2016), (Singh et al., 2017); (f) deep learning - use neural networks that identify fast the fake news (Bajaj, 2017; Sneha et al., 2017); 2. *social networking approaches* using (a) linked data – knowledge networks are exploited to identify the lie (Conroy et al. 2015; Idehen, 2017); (b) the behavior of users on social networks – the fact that users are forced to authenticate when using the social network, provides increased confidence in the data that appears here (Shu et al., 2017; Wu and Liu, 2018).

In this context, our application implements a neural network to identify both fake users and fake news, and it provides real-time results and it offers statistics for the evolution of fake news over countries and continents. The application also creates a heatmap to display and filter between credible and not credible tweets.

The paper is structured as follows: chapter 2 describes a learning method used to develop the system, including details about training data. Chapter 3 presents the experiments and use cases performed on users and on tweets and Chapter 4 presents relevant statistics. Chapter 5 analyses the problems occurred and suggests some possible solutions, before drawing some conclusions in the last section.

2. Data set and Method

In order to present the specificity of this system, this section addresses the following question: how difficult is to find an optimal solution in analyzing credibility on Social Media? We propose two methods to automatically assess tweet credibility by using sentiment analysis and neural network models.

2.1 Training Data

In order to compute credibility scores for tweets, we needed training data[1]. For that, firstly, we collected more than 2,500 tweets (from 50 users with at least a few thousands of followers). Then, with the help of five human annotators, we manually assigned them a number - 0 for tweets that we did not consider credible and 1 for tweets that we did consider credible. From the initial collection of tweets we eliminate ambiguous tweets (tweets on which annotators have held controversial discussions whether it is credible or not) and we remain with 2,270 (of which 1,248 are not credible and 1,022 are credible). Retweets were considered not credible because we were not able to efficiently retrieve additional information used when computing the credibility score (Twitter API impose some limits that were too low to be usable within the terms of usage at that time).

For each individual tweet, we collected the following types of information: (1) *retweetsNo* – the number of times the current tweet was retweeted; (2) *favoritesNo* – the number of times the current tweet was marked as favorite; (3) *creationDate* – the date this tweet was posted; (4) *wordsNo* – the number of words in the current tweet (excluding stopwords); (5) *relevantWordsRatio* – ratio between the number of words within text that are not stopwords nor punctuations and the total number of words; (6) *charactersNo* – the number of characters in the current tweet.

For example, the tweet with ID 1111965027483951105: "*In honor of his past service to our Country, Navy Seal #EddieGallagher will soon be moved to less restrictive confinement while he awaits his day in court. Process should move quickly! @foxandfriends @RepRalphNorman*", tweeted by Donald Trump, it had at the moment it was collected: (1) *retweetsNo*: 19,919, (2) *favoritesNo*: 61,840, (3) *creationDate*: 2019-03-30 05:14 AM, (4) *wordsNo*: 33, (5) *relevantWordsRatio*: 22, (6) *charactersNo*: 183.

Additionally, for each user we collected the following information: (1) The most recent 40 tweets posted by that user; (2) *hasLocation* – *true*, if user filled the location field, *false* otherwise; (3) *hasDescription* – *true*, if user filled the description field, *false* otherwise; (4) *hasGeo* – *true*, if user turned on geolocation, *false* otherwise; (5) *isVerified* – *true*, if user was verified by Twitter, *false* otherwise; (6) *creationDate* – the date when account was created; (7) *followersNo* – the number of followers.

For example, for the user @realDonaldTrump we collected: (1) The most recent 40 tweets: 1111965027483951105, etc. (2) *hasLocation*: true; (3) *hasDescription*: true; (4) *hasGeo*: true; (5) *isVerified*: true; (6) *creationDate*: 18.03.2009; (7) *followersNo*: 59,600,617.

The users for our experiments were selected from different fields, such as politics (*Donald Trump, Barack Obama, Hillary Clinton*, etc.), business (*Tim Cook, Bill Gates*, etc.), companies (*Google, Microsoft,* etc.), organizations (*Discovery, NASA*, etc.), television (*CNN, NatGeo*, etc.), music (*Eminem, Justin Timberlake, Miley Cyrus*, etc.), sport (*Maria Sharapova, Simona Halep*), other (*Android, Kim Kardashian, Dalai Lama,* etc.).

2.2 Sentiment Analysis. Measure User's Credibility and Credibility for New Tweets

First, we tried to manually compose a formula that we thought would compute a relevant credibility score both for the user and for the tweet[2] (as we understood credibility – how likely is the fanbase of a user to trust a tweet, either by liking it or retweeting it). For clarity, the metrics presented in this paper are more useful for the measuring user "engagement". Social Media became an extremely favorable environment to spread information credible or not.

To compute the credibility of tweet, we consider the formula (1):

$$\text{TweetCredibilityscore} = w_R \times T_R + w_F \times T_F + w_W \times T_W + w_S \times T_S \qquad (1)$$

where T_R represents the retweets score (the number of retweets divided by the number of reachable followers of the author, we considered that a tweet reaches 3% of the followers base just by posting it), T_F represents the favorites score (number of this the tweet was marked as favorite divided by number of reachable followers), T_W represents the ratio of relevant words contained by tweet's text, T_S represents the sentiment score (cumulative sentiment score for tweet's text computed using the Stanford Sentiment Analysis component; very negative words weighted 0.75, negative and very positive words weighted 0.50, positive words weighted 0.25, neutral words weighted 0.00).

After a lot of experiments performed with the scope to identify the best distribution for the weights from tweet' credibility formula (2), we came up with the following weights:

$$w_R = 0.1, w_F = 0.3, w_W = 0.5, w_S = 0.1 \qquad (2)$$

To compute the user's credibility, we considered the following parameters for it (1) *the location*, (2) *the URL*, (3) *the description*, (4) *if he is verified or not*, (5) *the geolocation*, (6) *the creation date*, and (7) *the most recent 20 tweets tweeted by this user*. We applied the formula (3) over the 50 considered users from our database and we saved the result for later comparison in order to find the best values for weights that respected the associated credibility to these users and to their tweets.

$$\text{UserCredibilityscore} = w_L \times U_L + w_U \times U_U + w_D \times U_D + w_V \times U_V + w_G \times U_G + w_C \times U_C + w_{A20} \times U_{A20} \qquad (3)$$

where U_L is 1 if the user has location set, else value is 0, U_U is 1 if the user has URL set, else value is 0, U_D is 1 if the user has description set, else value is 0, U_V is 1 if the user's account is verified, else value is 0, U_G is 1 if the user has geolocation enabled, else value is 0, U_C is the division between the number of months from when the account was

[1] Human-annotated training data, after cleaning phase - https://bitbucket.org/Sakresys/credibility-on-twitter/src/master/annotated%20data/

[2] See the source code to compute both credibility types - https://bitbucket.org/Sakresys/credibility-on-twitter/src/master/

created until the date the user's credibility is computed and the number of months from 15 July 2006 (the day Twitter went public) until the date the user's credibility is computed, U_{A20} is the average credibility of the last 20 tweets of the current user.

After a lot of experiments performed with the scope to identify the best distribution for the weights from users' credibility formula (4), we came up with the following weights:

$$\mathrm{w_L} = 0.01, \mathrm{w_U} = 0.01, \mathrm{w_D} = 0.03, \mathrm{w_V} = 0.01, \mathrm{w_G} = 0.08, \mathrm{w_C} = 0.07, \mathrm{w_{A20}} = 0.7 \qquad (4)$$

As we can see the most significant value for the users' credibility is the U_{A20} (the average credibility of the last 20 tweets of this user). We conclude that these formulas can judge satisfactorily the perceived credibility of a tweet/user.

2.3 Learning Method

For an advanced approach, we decided to use a neural network and compare the results to both manually annotated scores and to the ones obtained by a manually created formula. Thus we trained a neural network model.

A machine learning model represents the experience obtained after training an important amount of training data, data that should be annotated as good as possible, ideally without any noise. In fact, this definition is not well suited for any type of machine learning algorithm, but for those types of algorithms that use labeled data as in our case. Specifically, a supervised neural network model is a mathematical function whose weights are refined iteration after iteration. After training, any further inputs are processed using the formula with the adjusted weights resulting in correct or incorrect decisions depending on how well trained in the model.

Our neural network model uses *five input neurons* for: (1) the *retweet score*, (2) the *favorite score*, (3) the *relevant words ratio*, (4) the *number of hashtags* and (5) the *number of hashtags characters*; and *one output neuron* that produces a value between 0 and 1. A value above 0.6 means the tweet is *credible*, else it is *not credible*. We decided to use a single hidden layer with thirteen neurons because our model does one thing - computes the credibility of a tweet. We decided to use thirteen neurons after doing some testing and not go higher to avoid overfitting. The number of neurons in the hidden layer is picked after the formula (5) that helps determine the upper bound of hidden neurons such that the training won't result in overfitting:

$$N_h = \frac{N_s}{\alpha \times (N_i + N_o)} \qquad (5)$$

where N_S is the *number of samples in training dataset*, N_i is the *number of input neurons*, N_o is the *number of output neurons*, *α* is an *arbitrary scaling factor* usually between 2 and 10.

In an attempt to get the best results possible out of the available data we considered to train several versions of the model with different inputs and various combinations. In the following section, we discuss each different configuration. In the end, we benchmarked all configurations and chose the one that provided the closest scores to the ones manually annotated.

Configurations

We created the following variations of the neural network model:

1. *Basic* (C1) - represents the *basic configuration* for which we consider the *retweets score*, the *favorites score*, and the *ratio of the relevant words*.
2. *BasicWithNoRetweets* (C2) - at basic configuration we completely *exclude retweets* from the training dataset.
3. *BasicWithSentiment* (C3) - additionally to the basic configuration, we add the *sentiment score* of the tweet.
4. *BasicWithSentimentHashTagLength* (C4) - at the basic configuration, we add the *sentiment score* and the *hashtag length*.
5. *BasicWithSentimentHashTagCount* (C5) - at the basic configuration, we add the *sentiment score* and the *hashtag count*.
6. *BasicWithHashtagsCountAndHashtagsLength* (C6) - a combination of the previous two models.

Basically, we constructed a base configuration over which we added or removed elements in order to find relevant connections between the content of the tweets and the credibility score. After training the presented models on 500, 1,000 and 1,500 tweets, we calculated the accuracy using one-third of the training data from the rest of the tweets from the dataset. The results for the above six configurations are presented in Figure 1, where we add also the results for the system based on the formula.

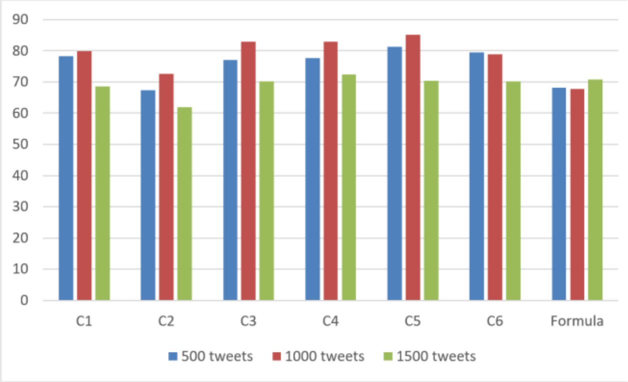

Figure 1: The results for considered configurations.

We observed that by removing the retweets from the basic configuration, we get noticeable lower results on accuracy (with around 10%). Information about sentiments and hashtags (length and count) helps us get better results. The highest rate of success that we obtained is 85.2% for configuration 5 with 1,000 tweets considered for training. An interesting aspect is that the increase in the number of tweets for the training data of more than 1,000 does not improve our results anymore; in fact, it leads to a decrease in quality.

We also tweaked the number of training iterations and we settled to about 100,000 training iterations for an optimal balance between training time and success rate. After 100,000 iterations the success rate increase is insignificant, moreover, it starts decreasing after a certain point.

3. Experiments

3.1 Tweets Monitoring

One of the components of our system allows us to specify a tweet and for it, we can monitor the evolution of credibility in time. Next, we will see different types of behavior for credible tweets caught by this component (with scores close to 1), but we have similar behavior for not credible tweets (with scores close to 0).

3.1.1 Use Case 1 - Constant Behavior

We monitored one of the tweets of user @realdDonaldTrump with id: 946731576687235072: “*The Democrats have been told, and fully understand, that there can be no DACA without the desperately needed WALL at the Southern Border and an END to the horrible Chain Migration & ridiculous Lottery System of Immigration etc. We must protect our Country at all cost!*”

Being posted by Donald Trump, we obtained linear credibility of 0.85, meaning that this tweet is a highly credible one (Figure 2a). The reason why this tweet has constant credibility over the hours it was monitored by us, is because that moment in time was far away from the creation date of the tweet. The conclusion would be that the interest in it has diminished drastically in the last period.

3.1.2 Use Case 2 - Variable Decreasing Behavior

We take a tweet from the user @NASA, with id 1112414759419371527: “*It's gettin' hot in here! “A Engineers recently conducted a static hot-fire test of our @NASA_Orion spacecraft to ensure it’s ready for missions to explore the Moon. Watch us turn up the heat: https://go.nasa.gov/2FEXwlx*”.

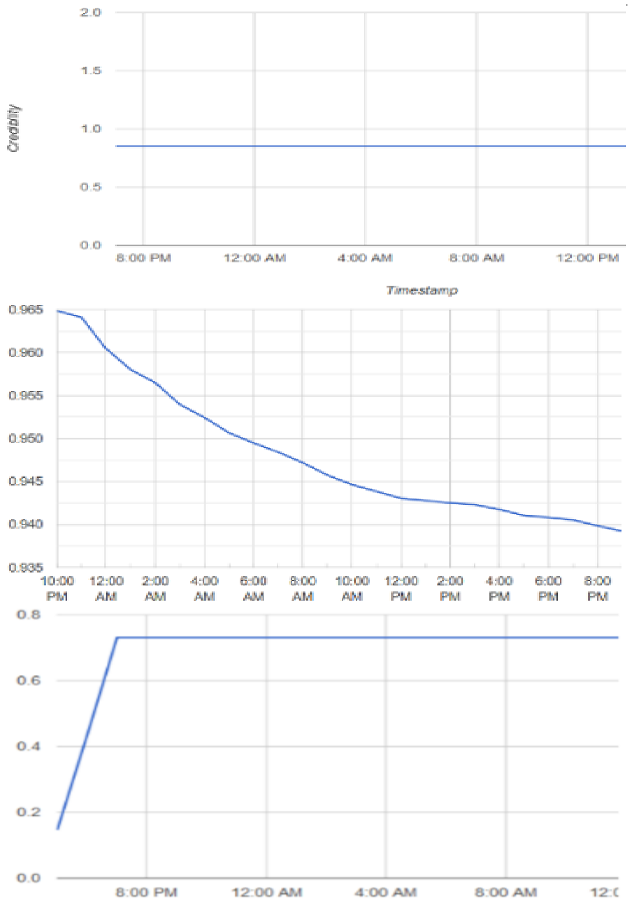

Figure 2: (a) Evolution of tweet credibility in time: constant (top), (b) variable decreasing (middle) and (c) constant growing (down).

In this case, our system captures the decreasing of the credibility of the tweet, which fluctuates from almost 0.965 to 0.935, mainly because it was monitored after the interest in this news has diminished (Figure 2b).

3.1.3 Use Case 3 - Constant Growing Behavior

We choose another tweet from the user @realdDonaldTrump with id 947461470924820480: “*Why would smart voters want to put Democrats in Congress in 2018 Election when their policies will totally kill the great wealth created during the months since the Election. People are much better off now not to mention ISIS, VA, Judges, Strong Border, 2nd A, Tax Cuts & more?*”

In this case, the increase of the credibility is more abrupt at the beginning (Figure 2c). The reason for this to happen would be that the system monitored the tweet’s last hours before becoming irrelevant or followers focusing on news tweets.

Our system classifies as not credible tweets, short posts without meaning, with many punctuation signs, posted by users without many followers without retweets and without interest from other Twitter users.

Tweets	Score
Just a moment prior to being told to “piss off” in classic style. #makingfriendsinmanchester @… https://t.co/AV9QvT26v8	0.000046
IT HURTS NOW???? BUT WHAT DOESN'T KILL YOU MAKES YOU STRONGER??IN THE END??? #brokenfamily… https://t.co/avuxg2L2zS	0.000125
The patient voice in cancer research @sysbioire #patientsinvolved https://t.co/qb7OP6nB0A	0.00034
Everyone knows what I look like, not even one of them knows me?? #everybodyhatesme… https://t.co/Ad37jfCIjA	0.001889

Table 1: Credible tweets

3.2 Users Monitoring

Another component of our system allows us to monitor a user and to see the evolution of his credibility in time. Next, we will illustrate different types of behavior for different users, according to their activity on Twitter.

3.2.1 Use Case 1 – Donald Trump

We monitored the credibility of Donald Trump over 10 hours that varied systematically according to the opinions of his supporters. (Figure 3).

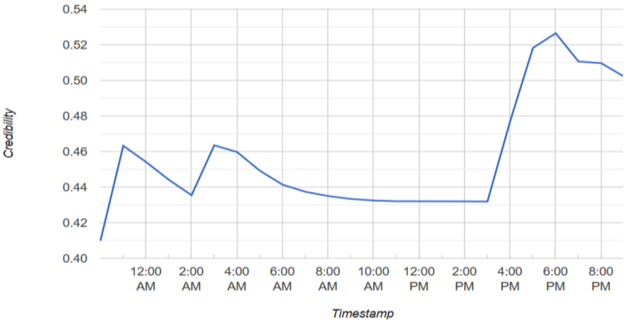

Figure 3: Evolution of **user credibility** in time for Donald Trump.

An important drop of credibility happened around 2:00 AM, but the credibility of the president of the US returned close to its previous scores. This was caused by one of his tweets that causes divided opinions through his supporters. After a while, we can observe linear credibility, meaning that Donald Trump had no activity between those hours and/or his followers' base was probably inactive on that period of time. Around 3:00 PM, there is a remarkable increase in credibility due to another tweet posted by the president, which led to a sudden increase in credibility, followed by a constant drop.

3.2.2 Use Case 1 - Justin Bieber

Monitoring Justin Bieber (Figure 4) we can see a continuous decrease of credibility over 10 hours.

Even so, the difference of credibility from the beginning of the monitoring until the end of it illustrates an insignificant decline (from 0.46674 to 0.46666).

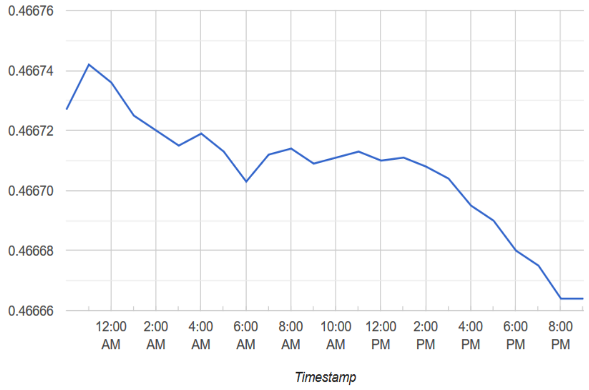

Figure 4: Evolution of **user credibility** in time for Justin Bieber.

The reason could be that he posted new tweets, but it took some time until these tweets got relevant up to the point the credibility of the user stabilized. On the other hand, for Donald Trump, the scores decreased and then increased, but the difference of credibility from the beginning of the monitoring until the end of it is higher (from 0.41 to 0.53).

3.3 Use Cases Conclusions

When monitoring tweets we observed that their credibility fluctuates after being posted. The credibility stabilizes after a few hours - when a tweet becomes outdated. A tweet that does not become viral or is not posted by a user with many followers may have a less volatile credibility score or does not fluctuate at all.

In the case of monitoring users, their credibility fluctuations are more noticeable than in the case of tweets, because the most recent 20 tweets are accounted for when calculating the credibility score every 60 minutes from the moment of monitoring. However, a user may also have low or no credibility fluctuations, if his activity is low or the number of followers is not high enough so the user's probability of watching the tweets is high.

In conclusion, the biggest changes in the credibility of a tweet/user occur in the early hours of posting. Another observation would be that if we increase the number of training tweets, we will not get significantly better results, because messages often have similar features.

4. Statistics

In this section, we will present and we will analyze the results obtained by our system on 50 selected users and on 3,004 tweets (collected in period March-April 2019) using the best neural network presented in the previous chapter. The credible/not credible label was assigned for the tweets using the NN described earlier in this article. 1,344 tweets were labeled as not credible and 1,660 tweets were labeled as credible.

4.1 Users Credibility

In Figure 5, we can see the overall credibility of all users, and we can remark: (1) in politics *Barack Obama* enjoys greater credibility compared to *Hillary Clinton* and *Donald Trump*; (2) in business, *Bill Gates* has a highest value of credibility and he is more credible than *Tim Cook*, which has also a good value of credibility; (3) for companies, Google is more credible than Microsoft, but both have medium values of credibility; (4) for organizations, *Discovery* and *NASA* have the lowest values of credibility; (5) for televisions, *NatGeo* and *Foxnews* have close and very good values for credibility; (6) in music, *Justin Timberlake* has the highest value in comparison with *Snoop Dogg, Rihanna, Lady Gaga, Eminem* and we can deduce that he is the most in vogue artist out of those who have been monitored; (7) in sport, *Sharapova* has a better credibility in comparison with *Simona Halep*. This is due to the fact she returned after a pause in which she was suspended, and Simona lost 1st place in the WTA rankings and her activity and followers are more active on Twitter. It is interesting how *Android, Tim Berners-Lee*, and *Dalai Lama* have the lowest value of credibility, due to the low activity on Twitter in the last period.

There are many ambiguous situations with the credibility of around 0.6, which makes the decision as a user is credible or not difficult to take.

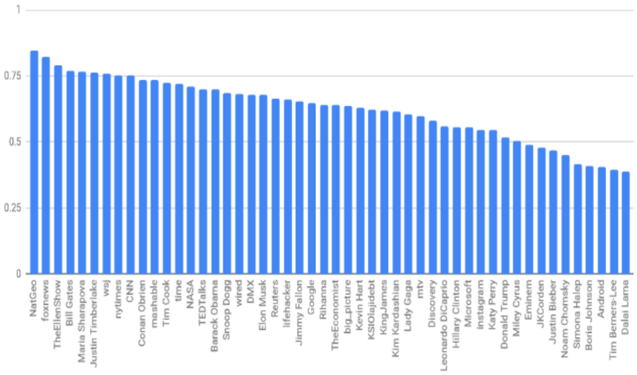

Figure 5: A statistic of the monitored users.

But we note that the information provided by our system is very useful when we want to compare two Twitter users to figure out which one is more credible.

What is interesting is the evolution of these users' credibility over time: when we started in 2018 to collect information about some of these users, (1) *Donald Trump* had a much greater credibility compared to other politicians, which means that lately his credibility has been affected by his posts on Twitter as well as by his political activity, (2) *Simona Halep* had more credibility than *Maria Sharapova*, who was suspended at that time, but the loss of her first position and Maria's return to the circuit made the hierarchy change.

Our plans for the future aim to make clearer when the credibility of a user increases or decreases and try to provide justification for these changes. The reasons we have identified so far for decreased credibility are poor network activity, or posting a controversial tweet, or moving attention to someone else who may have better results in the same domain of activity, etc. Reasons for increasing credibility are continued work within the network, notable results achieved in the domain of activity, posts on Twitter supported by network followers, etc.

4.2 Credibility by Continents and Countries

Figures 6 and 7 show the number of tweets (both credible and not credible) by continents and by countries.

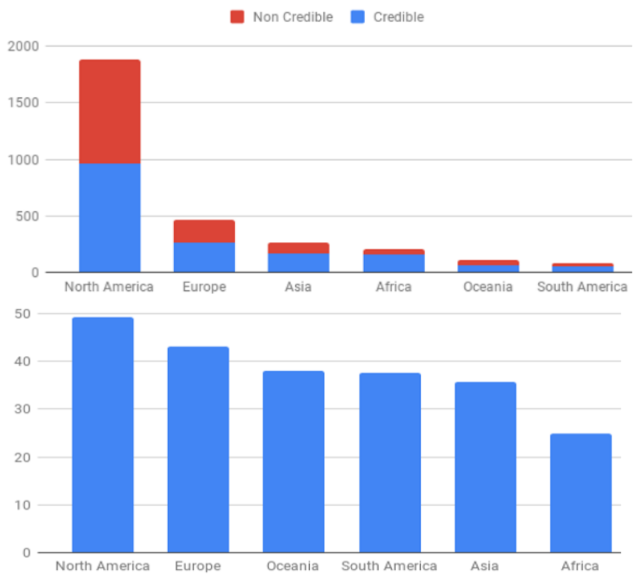

Figure 6: Statistics by continents, by the number of credible/not credible tweets (up) and by the percentage of credible/not credible tweets (down).

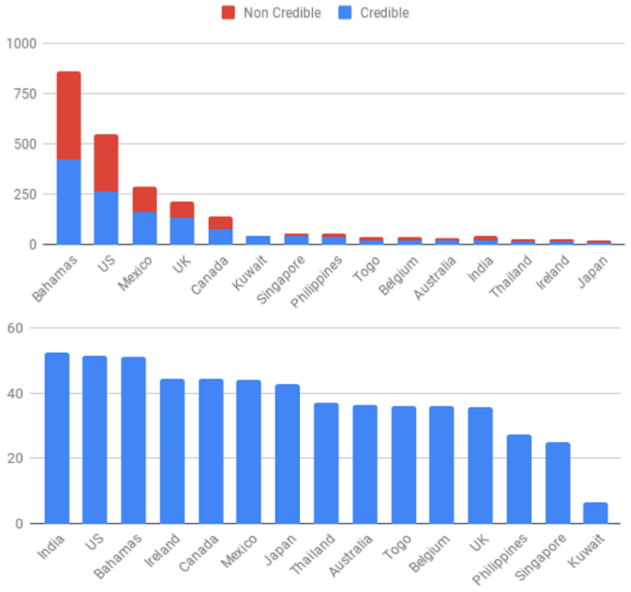

Figure 7: Statistics by countries, by the number of credible/not credible tweets (up) and by the percentage of credible/not credible tweets (down).

As can be deduced from Figure 6, the zones with most tweets are North America and Europe and from these continents come the most number of not credible tweets.

In terms of percentage, these continents have also the highest number of not credible tweets, and their number is around 5% greater than in Oceania, South America, and Asia. Our application, allow the user to display these tweets over a heatmap in real-time in order to see the distribution of credible and not credible tweets. Figure 8 contains the visual representation of the 1,344 not credible tweets (the red areas).

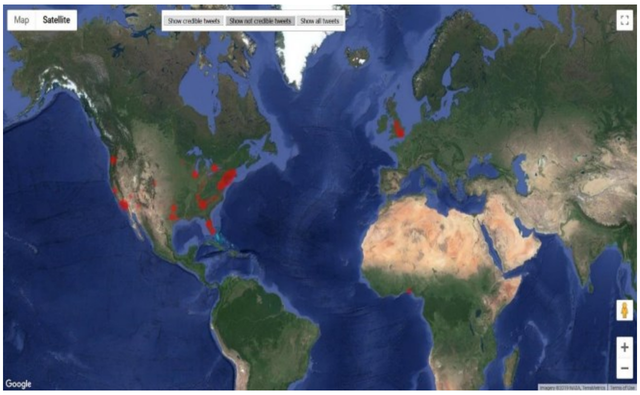

Figure 8: Heatmap of not credible tweets.

5. Error Analysis

The small number of relevant tweets compared to the total number of tweets that can be collected from Twitter is due to the fact that many tweets don't have geolocation information. This information is very important for our system because the statistics for a user based on continents or on countries or the heatmap need mandatory this information.

The biggest number of irrelevant tweets comes from the Bahamas, a small country, where most messages are by type advertising or job announcements, automatically added to twitter by bots, containing no information besides the user's location or advertisement. These tweets are so frequent that they make up almost 30% of tweets collected by us. For the future, we need to pay more attention to how we collect information for our system to avoid such tweets coming into our database.

Regarding the quality of the neural network approach, we still investigate ways to improve it. The main problem here is related to the fact that we want to keep a balance between the speed of the system, which works now in real-time, and the quality of the results. When we want to see in real-time the evolution of the credibility for users and we need that to collect more tweets from Twitter and then for everyone we need crawling, processing, classifying, updating statistics and maps, every delay may affect the quality of the user experience. Also, we want to come with more relevant data in our training dataset and we investigate what kind of data can be useful for our system.

Another problem come from the component that assigns a tweet to countries or to continents when the tweet is posted on Twitter in a region which is very close to a border between two countries. Our intention is to assign a country parameter to every Twitter user, which represents the most common country from which he posts the last 10 tweets. This value will be used instead of the geolocation of the tweet in limit situations when the user is near a border. Our experiments show until now that this value can be used with success.

6. Conclusions

This data set enriches the Corpus of contemporary Romanian Language (CoRoLa), one of the activities of the Resources and Technologies for developing human-machine interfaces in Romanian (ReTeRom) project. Due to the texts naturalness (e.g. tweets) and to the annotation it contains, this corpus is useful to developers of applications based on natural language.

Big companies like Google and Facebook attach great importance to the phenomenon of the appearance and spreading of false news by real or false users. Linguistic and social networking approaches allow us to build systems that assess the credibility of users and the information they post.

What we have presented in this study provides a way of classifying users and the information they post in credibility classes. To classify tweets and users, we trained a neural network model using a collection of tweets that were manually annotated by human users. Also, in our experiments, we consider users from different fields, such as politics, business, companies, organizations, television, music, sport, and others.

The current work comes with a new proposal to display in real-time statistics and information related to credible and not credible tweets on Google Maps using heatmap. Compared to previous approaches, which displays all the tweets collected over a certain period, we can manage to display credible or not credible tweets, statistics on countries and on continents in real-time. In the future, we have to find a more efficient way to classify tweets by credible and not, while having real-time processing of data. These experiments showed that the information credibility on Twitter will become a key component for solving social problems, for instance. That is why in future any application that will use data from social networks, no matter how small it is, it should make a difference between credible and not credible data.

Acknowledgements

This work was supported by a grant of the Ministry of Research and Innovation, Program PN-III-P1-1.2.- CCDI, nr. 73/2018, as part of the ReTeRom project. It was also partially supported by POC-A1-A1.2.3-G-2015 program, as part of the PrivateSky project (P_40_371/13/ 01.09.2016)

Bibliographical References

Allcott, H. and Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. In *Journal of Economic Perspectives*, 31(2): 211-236.

Atodiresei, C. S., Tănăselea, A. and Iftene, A. (2018). Identifying Fake News and Fake Users on Twitter. In *Proceedings of International Conference on Knowledge Based and Intelligent Information and Engineering Systems*, KES2018, 3-5 September 2018, Belgrade, Serbia. Procedia Computer Science, Vol. 126, 451-461.

Bajaj, S. (2017). The Pope Has a New Baby! Fake News Detection Using Deep Learning. *Stanford Report* https://web.stanford.edu/class/cs224n/reports/2710385.

Bovet, A. and Makse, H. (2019). Influence of fake news in Twitter during the 2016 US presidential election. In *Nature Communications* 10(1): 1-14.

Brummette, J., DiStaso, M., Vafeiadis, M. and Messner, M. (2018). Read All About It: The Politicization of "Fake News" on Twitter. In *Journalism & Mass Communication Quarterly*, 95(2): 497-517.

Castillo, C., Mendoza, M. and Poblete, B. (2011). Information Credibility on Twitter. In *Proceedings of the WWW 2011 – Session: Information Credibility*. 675-684, Hyderabad, India.

Chen, Y., Conroy, N. J. and Rubin, V. L. (2015). News in an Online World: The Need for an Automatic Crap Detector. In the *Proceedings of the Association for Information Science and Technology Annual Meeting* (ASIST2015), St. Louis.

Chu, Z., Gianvecchio, S., Wang, H. and Jajodia, S. (2010). Who is tweeting on Twitter: Human, Bot, or Cyborg? In the *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10, 21-30.

Ciampaglia, G., Shiralkar, P., Rocha, L., Bollen, J. Menczer, F. and Flammini, A. (2015). Computational fact checking from knowledge networks. *Journal Plos One*, https://doi.org/10.1371/journal.pone.0128193

Clark, C. S. (2009). Fake news? A survey on video news releases and their implications on journalistic ethics, integrity, independence, professionalism, credibility, and commercialization of broadcast news. *The University of Alabama*, Tuscaloosa, Alabama.

Conroy, N. J., Rubin, V. L. and Chen, Y. (2015). Automatic Deception Detection: Methods for Finding Fake News. In *Proceedings of the Association for Information Science and Technology Annual Meeting* (ASIST 2015), St. Louis, MO, USA.

Cook, D., Waugh, B., Abdipanab, M, Hashemi, O. and Rahman, S. (2013). Twitter Deception and Influence: Issues of Identity, Slacktivism and Puppetry. *Journal of Information Warfare*, 13(1): 58-71.

Cusmuliuc, C. G., Coca, L. G. and Iftene, A. (2018). Identifying Fake News on Twitter using Naive Bayes, SVM and Random Forest Distributed Algorithms. In *Proceedings of The 13th Edition of the International Conference on Linguistic Resources and Tools for Processing Romanian Language*, 177-188.

El Ballouli, R. El-Haki, W., Ghandour, A., Elbassuoni, S., Hajj, H. M. and Shaban, K. (2017). CAT: Credibility Analysis of Arabic Content on Twitter. In *Proceedings of the Third Arabic Natural Language Processing Workshop*, Valencia, Spain, 62-71.

Gînscă, A. L., Popescu, A., Lupu, M., Iftene, A. and Kanellos, I. (2015). Evaluating User Image Tagging Credibility. Experimental IR meets Multilinguality, Multimodality, and Interaction. LNCS, Vol. 9283, 41-52. In *Proceedings of 6th International Conference of the CLEF Association*, CLEF'15 Toulouse, France

Hadeer, A. (2017). Detecting opinion spam and fake news using n-gram analysis and semantic similarity. *Library of University of Victoria* http://dspace.library.uvic.ca:8080/handle/1828/8796.

Hadeer, A., Issa, T. and Sherif, S. (2017). Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. In *International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments* ISDDC 2017, 127-138.

Idehen, K. U. (2017). Exploitation of a Semantic Web of Linked Data, for Publishers. *Open Link Virtuoso Universal Server*.

Iftene, A., Dudu, M. S. and Miron, A. R. (2017). Scalable system for opinion mining on Twitter data. Dynamic visualization for data related to refugees' crisis and to terrorist attacks. At the *26th International Conference on Information Systems Development*, Larnaca, Cyprus.

Iozzio, C. (2012). Reuters built a bot that can identify real news on Twitter. *Popular Science*. December 2 http://www.popsci.com/artificialintelligenceidentifyrealnewsontwitterfacebook.

Mihalcea, R. and Strapparava, C. (2009). The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language. In the *Proceedings of the ACL-IJCNLP Conference Short Papers*, 309-312.

Perez-Rosas, V., Kleinberg, B., Lefevre, A. and Mihalcea, R. (2017). Automatic Deception Detection: Methods for Finding Fake News. *Cornell University Library* https://arxiv.org/abs/1708.07104.

Popoola, O. (2017). Using Rhetorical Structure Theory for Detection of Fake Online Reviews. In *Proceedings of the 6th Workshop Recent Advances in RST and Related Formalisms*, 58-63, Santiago de Compostela, Spain, ACL.

Rubin, V. and Lukoianova, T. (2014). Truth and Deception at the Rhetorical Structure Level. In J*ournal of the American Society for Information Science and Technology*, 66 (5).

Rubin, V.R., Conroy, N. J., Chen, Y. and Cornwell, S. (2016). Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News. In *Proceedings of the Workshop on Computational Approaches to Deception Detection at the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (NAACL-CADD2016), San Diego, California.

Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. In *SIGKDD Explor. Newsl.* 19 (1): 22-36.

Singh, V., Dasgupta, R., Sonagra, D., Raman, K., and Ghosh, I. (2018). Automated Fake News Detection Using Linguistic Analysis and Machine Learning. *In Proceedings of SBP-BRiMS*.

Sneha, S., Nigel, F. and Shrisha, R. (2017). 3HAN: A Deep Neural Network for Fake News Detection. In *24th International Conference on Neural Information Processing* (ICONIP 2017), Springer International Publishing AG 2017, Part II, LNCS 10635, 1-10.

Șusnea, E. and Iftene, A. (2018). The Significance of Online Monitoring Activities for the Social Media Intelligence (SOCMINT). In *Proceedings of the Conference on Mathematical Foundations of Informatics MFOI'2018*, Chisinau, Republic of Moldova, 230-240.

Wu, L. and Liu, H. (2018). Tracing Fake-News Footprints: Characterizing Social Media Messages by How They Propagate. In *Proceedings of the WSDM 2018*, Marina Del Rey, CA, USA, ACM.