Predicting News Source Credibility

Article · August 2019

CITATIONS
0

READS
854

3 authors:

Ahmet Aker
University of Duisburg-Essen
100 PUBLICATIONS **1,183** CITATIONS
SEE PROFILE

Vincentius Kevin
University of Duisburg-Essen
5 PUBLICATIONS **9** CITATIONS
SEE PROFILE

Kalina Bontcheva
The University of Sheffield
274 PUBLICATIONS **8,387** CITATIONS
SEE PROFILE

Predicting News Source Credibility

Ahmet Aker[1,2], **Vincentius Kevin**[1], **Kalina Bontcheva**[2]
University of Duisburg-Essen, Duisburg, Germany[1]
University of Sheffield, Sheffield, England[2]
a.aker@is.inf.uni-due.de, vincentius.kevin@stud.uni-due.de, K.Bontcheva@sheffield.ac.uk

Abstract

Assessing the credibility of a source of information is important in combating with misinformation. In this work we tackle the source credibility assessment as regression task. For this purpose we release a dataset containing around 700 news sources along with detailed credibility and transparency scores. These scores are manually assigned to every news source. We merge these scores to have final credibility score for every news source. The merged scores are then used to train prediction models. Our results show highly satisfactory performances in predicting the merged credibility scores. Along with the dataset we also plan to release our models to allow the use for a wider community.

1 Introduction

The WEB has been never that big than it is now. It contains tremendous amount of information such as standard web documents, videos, images, blog and social media posts and many other entries. One of the reason of this massive growth is that it is not anymore shaped by only few experts or dedicated people or institutions but by everyone who has access to it. Although this has led to immense information richness, alternative views and diversity however, it has also brought new challenges. It has stripped out the traditional information providers from their gate-keeping role (Baly et al., 2018) and has left the public in a jungle of web content with varying quality from reliable and true information to misinformation i.e., facts that are not true. A web user walking through in this jungle is likely to be mis-led, manipulated in her belief towards a specific group's interest, political party, a theory, etc. and psychologically attacked to capture her attention and lead her towards actions harmful to herself but also for the society. A famous example involving harmful action is the "Pizzagate" incident, which was provoked by misinformation shared on social media about 2016 presidential candidate Hillary Clinton's connection to a child pornography ring acting in a pizzeria that ended up with gun shootings (Allcott and Gentzkow, 2017; Berghel, 2017).

Misinformation can be interchangeably used with the term fake news. Douglas et al. refer to fake news as a "deliberate publication of fictitious information, hoaxes and propaganda" (Douglas et al., 2017), and is similarly defined by others (Klein and Wueller, 2017). Misinformation can be combat either on item level, i.e. determining that e.g. a social media post or news article is fake (Markowitz and Hancock, 2014; Hardalov et al., 2016; Rashkin et al., 2017; Zubiaga et al., 2018) or try to rank its origin/source in terms of credibility (Abbasi and Liu, 2013; Weng et al., 2010; Cha et al., 2010; Yamaguchi et al., 2010; Mukherjee and Weikum, 2015; Mukherjee et al., 2014; Abbasi and Liu, 2013) and with that make readers aware that what ever comes from that source might require a special treatment such as fact checking and should only be digested with care.

In this work we focus on the second option and rank sources according to their credibility. Our focus is on news sources. In our approach we use supervised learning with gold standard consisting of news sources and their credibility scores. We represent each news source with a rich set of features commonly used by e.g. search engines to rank web-sites reputations. Our results show our model can indeed act as surrogate of intensive manual efforts to rank news sources for credibility. In summary the contributions of the paper are:

- Provide a new gold standard dataset containing sources along with detailed credibility and transparency scores scores[1]. (Baly

[1]https://github.com/ahmetaker/sourceCredibility

et al., 2018) also reports a dataset containing sources with factual reporting scores. Note that this dataset focus only on how factual the contents are. Unlike this, our dataset contains credibility (includes factual reporting) aspect but also transparency scores computed by journalists. The transparency gives details about how open the source is towards its readers. Furthermore, each of our news source comes along with detailed credibility and transparency scores instead of just one score as it is the case with the dataset reported by (Baly et al., 2018).

- Investigate regression models along with rich feature sets to act as surrogate for human judges.

- Our gold standard data contains sources where we know the political orientation as well as whether those sources are known for e.g. producing fake news. We look at the link between the source credibility and political orientation as well as content quality production.

- We also plan to release our models to allow their use by the wider research community.

In the following we first describe our data collection process. In 3 we describe the features we use in our prediction models. The models and the different analysis we perform are described in Section 4. We also investigate whether e.g. low scoring sources are those that usually produce fake news. Furthermore, we compute credibility scores of sources from different locations in the world. Our findings from these are presented in Section 5. We conclude the paper with Section 6.

2 Data Collection

In our work the data is composed of sources reporting news. For these sources we collect their credibility scores. To collect the news sources we first used the Media Bias Fact Check (MBFC) [2] and recorded all the sources mentioned by MBFC. We also used a pre-collected list of sources from Poynter.org[3]. Next for each of these news sources, we manually recorded the scores reported by NewsGuard.[4] This section will describe the process in further detail.

2.1 NewsGuard

NewsGuard has manually and methodically reviewed thousands of news sources which are mostly based in the US. NewsGuard is available as a Chrome extension which can show these information when such news sources are opened in the browser or appear in some web searches.

NewsGuard provides nine labels for each news source, and counts credibility or transparency scores for each criteria it fulfills. The criteria is listed below.

Credibility criteria:

- Does not repeatedly publish false content (22 points)

- Gathers and presents information responsibly (18 points)

- Regularly corrects or clarifies errors (12.5 points)

- Handles the difference between news and opinion responsibly (12.5 points)

- Avoids deceptive headlines (10 points)

Transparency criteria:

- Website discloses ownership and financing (7.5 points)

- Clearly labels advertising (7.5 points)

- Reveals who's in charge (5 points)

- The site provides the names of content creators, along with either contact information or biographical information (10 points)

The total of credibility and transparency scores is 100 at maximum, and a news website is considered "safe" if it has at least 60 points.

Using a Chrome browser with the NewsGuard extension installed, we visited each news source in our dataset (see next sections) and recorded the labels shown by NewsGuard.

[2] https://mediabiasfactcheck.com/

[3] https://www.poynter.org

[4] https://www.newsguardtech.com/

2.2 Media Bias Fact Check

Media Bias Fact Check (MBFC) provides lists of news sources under ten categories. MBFC states that news sources under the "Least Biased" and "Pro-Science" categories are reliable/legitimate, while the "Left/Right Bias", "Conspiracy" and "Questionable Sources" categories may contain unreliable sources.

Furthermore, for each news source, MBFC provides its name, URL and a 'Factual Reporting' score (Very Low, Low, Mixed, High, Very High), as well as some further human-readable information. For news under the category "Questionable Sources", instead of a 'Factual Reporting' score, MBFC provides a list of reasons for the categorization, such as the news being "Propaganda", "Extreme Left/Right" and/or "Fake News".

All data from MBFC are publicly available in their website. At the time of writing, the lists contain in total 3007 news sources. NewsGuard had scores for only 673 of those. The score distribution is shown in Figure 1.

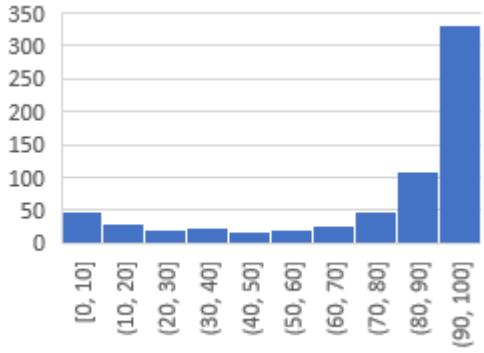

Figure 1: Histogram of NewsGuard Scores (x-axis) on the MBFC Dataset. Y-axis shows the sample count.

2.3 Unreliable News Sources DataSet

The Unreliable News Sources (UNS) dataset contains 513 news sources, 282 of which overlaps with the MBFC dataset. From the remaining, only 28 had scores on NewsGuard. The UNS dataset was obtained from a pre-collected list of sources provided to use by Poynter.org [5].

All sources in this dataset are categorized as unreliable. In contrast to the MBFC dataset, NewsGuard gives as expected low scores for most of the UNS sources, although a few of them got scores in the green range too (60 and above), as shown in Figure 2.

[5] https://www.poynter.org

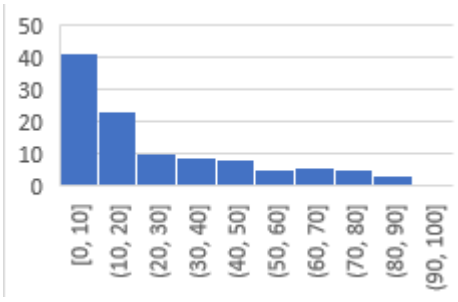

Figure 2: Histogram of NewsGuard Scores (x-axis) on the UNS Dataset. Y-axis shows the sampel count.

dataset	no. samples	NG
MBFC	3007	673
UNS	513	111
overlap	282	83
combined	3238	701

Table 1: Number of samples in the datasets and the overlap size. The "NG" column shows the number of samples for which NewsGuard scores are available.

3 Automatic Features

We gathered features which can be automatically obtained. A trained regression model would be able to use these features to automatically predict the NewsGuard score of a news source, even if NewsGuard has not analysed that particular news source.

CheckPageRank[6] (cPR) provides a free online tool which can report page rank (Page et al., 1999) score, alexa rank, and a few other domain analysis results for any given website. We used this tool to gather the features. Note, the tool does not provide any exact definition or information on how the scores are calculated. However, it provides scores which seem to be taken from non-free services such as Moz SEO and Majestic SEO tools. While these tools highly limits usage for free users to ten queries per month and a few queries per day respectively (as of year 2019), cPR allows one query every thirty seconds, although it does not provide the full information available in the other tools.

Below are very likely explanations we found for the features provided by cPR, either because the feature name is self-explanatory or the supposed underlying services give exact or very close scores compared to what is displayed by cPR.

- Google Page Rank: A score from 0 to 10

[6] checkpagerank.net

which estimates the importance of the website based on the quantity and quality of links to it from other websites.

- cPR Score: This is shown visually as one of the most important scores in checkpagerank.net, albeit without any given definition. We presume that 'cPR' simply stands for 'checkPageRank' and cPR score is calculated with a proprietary formula or algorithm.

- Citation Flow and Trust Flow: These two scores are most probably from Majestic[7], an SEO (Search Engine Optimization) tool. Like Page Rank, these metrics lets pages influence the scores of other pages it links to, recursively and with decaying effect. According to Majestic's glossary[8], citation flow focuses on the quantity of links to the website, taking into account the count and influential power of the links, while in contrast, trust flow focuses on links from manually reviewed trusted sites.

- Topic Value: this score also most likely come from Majestic. Majestic provides a "Topical Trust Flow" score, which, according to their glossary "shows the relative influence [...] in any given topic or category." It is a likely explanation that cPR show only the topic for which the website has the best Topical Trust Flow, since the topic names and value range are exactly the same in cPR and Majestic.

- Backlinks: External backlinks mean links from other websites to the subject website. This excludes internal links, which usually exist to let users navigate within the same website.

- Referring domains: this is the number of domains which contains backlink(s) to the subject website.

- EDU and GOV backlinks and domains: Majestic also provides the counts of educational and governmental backlinks and domains.

- Domain Authority and Page Authority: the Moz[9] SEO tool describe these scores as "the ranking potential in search engines based on an algorithmic combination of all link metrics". While MozRank is not used directly by search engines, it is similar and highly correlated to Google PageRank. We tested a few websites and confirmed that cPR shows exactly the same scores as Moz.

- Spam Score: This most likely represents the Moz SEO spam flags explained in their website[10]. The flags represent internal and external features of websites that are indicative of 'spam websites' and have been found to be penalized or banned by Google.

- Alexa Rank: Alexa Rank is described as a popularity measure which "is calculated using a proprietary methodology that combines a site's estimated traffic and visitor engagement over the past three months."[11]

- Alexa Reach Rank: this score is based specifically on the estimated number of people each website is able to reach.

4 Regression Analysis

We investigate linear and random forest regression to predict the source credibility scores. We evaluate the performance of models using leave-one-out as well as cross dataset validations by measuring the root mean squared error (RMSE) in predicting the NewsGuard scores of news sources. The leave-one-out scenario is a more in-domain evaluation where the model sees during training samples similar to the testing instances. Unlike this is the cross dataset evaluation where the trainer sees only instances which have very little in common with the testing instances. With this experiment the robustness of the model is challenged. We also use the best performing model to turn it to a source credibility label, similar to NewsGuard.

Note, in both linear and random forest regression we use a merged scored obtained through the credibility and transparency scores: $3 \times credibility + transparency$[12]. In the following sections we use credibility to refer to this merged score.

[7] majestic.com
[8] https://majestic.com/help/glossary
[9] moz.com
[10] https://moz.com/blog/spam-score-mozs- new-metric-to-measure-penalization-risk
[11] blog.alexa.com
[12] https://www.newsguardtech.com/ratings/rating-process-criteria/

4.1 Feature Scaling, Missing Values and Score Imbalances

4.1.1 Logarithmic Scaling

Assuming that website links follow the pattern of a scale-free network (preferential attachment) (Barabási and Pósfai, 2016), features such as backlink and domain counts are expected to follow the power law distribution, instead of being equally distributed within the value range. Therefore, we apply a logarithm scaling before feeding these features to our models (important for the linear model). We also did the same for Alexa Rank and Alexa Reach Rank, because ranking ratio would be a better measure than ranking difference (e.g. the difference between rank 10 and 20 is worth the same as the difference between ranks 1,000 and 2,000).

4.1.2 Handling Missing Values

Often times, not all of the features are available. For example, the trust flow score is not available for cosmopolitan.com. The ability to use features which are not always available is important in order to learn from the dataset as well as to predict the scores of news sources for which some features are missing.

We tested three different ways to handle missing feature values (Gelman and Hill, 2006, p. 531). Firstly, we simply removed data samples which contain missing values (complete-case analysis). This results in only 416 samples out of 673. This results in smaller RMSE since it is only tested on complete samples, but the model cannot handle missing values and is only applicable when all features of a news source are available.

Secondly, we replace missing values with the average value for that column from all other samples (imputing mean). The imputed values are treated as if they were original data by the models, thus also has drawbacks such as pulling the correlation with the output label towards zero (Gelman and Hill, 2006, p. 533).

Thirdly, for each test sample (since we are using leave-one-out), we retrain the model only on the other samples which have all the features available on the test sample. The disadvantages are smaller effective training set and having to retrain for every combination of available features on the test sample. In our experiment, the retrain approach's performance seems slightly worse than imputing means.

After experimenting with linear and random forest models with the three approaches above on the MBFC dataset (without sample weights), we found imputing mean to perform better than retraining on available features only. Results of this experiment are shown in Table 2. In the subsequent experiments we use the imputing mean strategy to handle missing feature values.

missing	model	data	rmse
complete	linear regr	416	15.572
complete	rnd. forest 10	416	16.505
complete	rnd. forest 100	416	15.997
impute	linear regr	673	18.655
impute	rnd. forest 10	673	18.006
impute	**rnd. forest 100**	**673**	**17.773**
impute	rnd. forest 1000	673	17.775
retrain	linear regr	673	18.417
retrain	rnd. forest 10	673	19.335
retrain	rnd. forest 100	673	18.395

Table 2: RMSE of regression models on MBFC dataset using different ways of handling missing values.

4.1.3 Score Imbalances

The MBFC dataset is not balanced in terms of NewsGuard scores. This is likely because MBFC and NewsGuard focus more on popular news sources, and popular sources tend to be more credible. Therefore it can be beneficial to weigh samples differently based on their score to prevent over- and under representation of certain score ranges.

In classification problems, class weights are commonly used to balance the classes. We applied a similar strategy by first grouping the samples into n bins based on their actual NewsGuard score. Each bin is then assigned a 'class weight' reciprocally proportional to the number of samples in that bin. We tried with 5, 10 and 20 bins. Too many bins will result in some bins being empty and we suspect that it may amplify noise in the dataset by applying rather unpredictable sample weights.

We also looked into the possibility of using the features to determine sample weights, as is done for instance in rim weighting technique (Sharot, 1986), but decided otherwise since it is difficult to determine which features to balance and it is unclear if we have enough data for such weighting schemes to be worthwhile.

4.2 Experiments

4.2.1 Leave-one-out Experiment

For this experiment, the MBFC and UNS datasets are combined with duplicates removed to form a bigger dataset. The mean imputation is done after combining the datasets. We experimented with both linear and random forest models. For random forest, we used the default parameters in sklearn v0.20.0, except for the number of trees, for which we tried 10, 100, and 1000. We run each model twice, once with and once without using the sample weights and measure the RMSE on leave-one-out predictions on the full combined dataset. Since linear regression model can overshoot, its outputs are clamped between 0 and 100 before the RMSE is calculated.

When sample weights are used, the RMSE of both models increase (see Table 3, column loo). This is probably due to the imbalance of samples in different score ranges as shown in Figure 3. Since most sources have a high score, the models without sample weighting would be biased to giving higher scores to minimize the RMSE. Overall both linear and random forest perform almost equally in terms of RMSE. However, the difference in the models are more apparent in the cross dataset evaluation which will be explained in the following section.

Random forest model performs asymptotically better with increasing number of trees and will not overfit due to too many trees (Breiman, 2001). In this leave-one-out experiment, we saw no improvement with 1000 trees compared to 100, suggesting that 100 is probably more than enough, but since the only drawback of using more trees is computation time, which is not part of our evaluation, we stayed with 100 trees for all subsequent experiments.

4.2.2 Cross Dataset Experiment

In the cross dataset evaluation we train the models on the MBFC dataset and test on the UNS dataset. We do this experiment to see if the model can generalize to other datasets, since MBFC and UNS have very different score distribution. Since these datasets have a big overlap, we test two approaches to make this truly a cross dataset evaluation. Firstly, we try removing the overlapping sources from MBFC (train set) and thus testing on the full UNS dataset. This results in a worse performance compared to the cross validation results

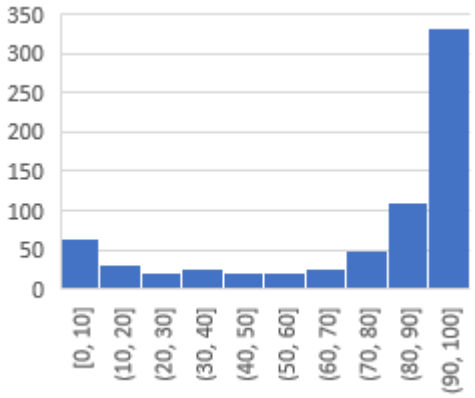

Figure 3: Histogram of NewsGuard Scores on the Combined Dataset (MBFC + UNS)

model	bins	loo	cs1	cs2
linear	-	20.26	39.64	33.34
	5	22.67	25.55	23.88
	10	22.17	25.93	24.40
	20	22.67	26.64	26.19
rnd. forest	-	18.60	39.28	28.72
	5	22.50	29.76	23.37
	10	23.26	30.28	26.88
	20	23.86	33.56	29.19
(train size)		701	590	673
(test size)		-	111	28

Table 3: Root mean squared error (RMSE) of regression models on predicting NewsGuard scores.
bins: number of bins the sample data is split into to determine the sample weights (see section 4.1.3), a minus sign means no sample weights were used.
loo: leave-one-out on MBFC+UNS.
cs1: model was trained on MBFC with overlaps removed, and tested on the whole UNS dataset.
cs2: model was trained on the whole MBFC dataset, and tested on UNS with overlaps removed.

as shown in Table 3 (column cs1). We suspect that this approach suffers because the training set now lacks low-scoring news sources. This is confirmed by the fact that using sample weights boosted the performance considerably.

Due to this, we also try removing the overlapping sources from the UNS (test set) and train on the full MBFC dataset. This results in a small test set (28 samples), but both models score better as expected (see Table 3, column cs2). As opposed to leave-one-out performance, weight sampling significantly boosts the performance in the cross dataset scenario, even more so with the linear model that it performs better than the random forest model.

4.2.3 Color Label Prediction

After finding the best regression model, we also try to automatically predict the icon (red or green) NewsGuard would give to a news source. NewsGuard marks a news source with a red icon if the score is below 60 and green otherwise. The red icon serves as a warning for the user to read with caution.

In the real application, NewsGuard only shows these icons at first, while further details are shown when the user hover or click on the icons. This may be more convenient for users such as during web searches. Therefore, even though the regression model's predicted score can be used as it is, we still explored the possibility of predicting the color.

The regression problem can be simply transformed into classification using a binary threshold, i.e. the predicted icon is red if the predicted score is below 60, and green otherwise. We evaluate the precision and recall of the model with this approach, and the results are shown in column '60' in Table 4.

Threshold	60	30-85
RED Precision	0.819	0.984
RED Recall	0.809	0.389
GREEN Precision	0.942	0.985
GREEN Recall	0.946	0.750

Table 4: Precision and Recall on classifying news sources as green and red.

This simple approach ignores the fact that the model is more confident about its prediction when its predicted score is far from the threshold (60). The precision and recall measures also do not handle the fact that samples in the dataset are not equally distributed in all ranges. For instance, the 85-100 score range covers more than half of the news sources (whether actual label range or predicted score range), so the precision and recall of green news sources are boosted by identifying those 'easier' cases, while the actual accuracy for news sources closer to the 60-points threshold is much lower.

Since global precision/recall may be misleading, in Table 5 we separate the samples into relatively small bins based on the model's prediction value to show the distribution across prediction ranges and the proportion of actual green vs red news sources in each bin, which also represents the model's confidence level we used as binary threshold.

predicted score	green	red	# samples
0-5	0%	100%	0
5-10	0%	100%	2
10-15	0%	100%	6
15-20	0%	100%	17
20-25	5%	95%	22
25-30	0%	100%	15
30-35	20%	80%	15
35-40	38%	63%	16
40-45	19%	81%	21
45-50	22%	78%	18
50-55	29%	71%	14
55-60	67%	33%	9
60-65	69%	31%	13
65-70	81%	19%	16
70-75	70%	30%	20
75-80	87%	13%	31
80-85	84%	16%	45
85-90	96%	4%	84
90-95	98%	2%	130
95-100	99%	1%	179

Table 5: Percentage of news sources labeled as green vs red for each range of predicted score.

Table 5 shows that the model cannot predict the label with a high accuracy when the predicted score is around the range 30-85. Therefore, it may be more useful in practice to show the red/green icons only when the model's prediction is outside this middle range, i.e. only provide high precision results, while a third icon can be used if the prediction is within the range. When 30-85 boundaries are used the precision scores are boosted close to 99% (Table 4, column 30-85).

5 Other Results

In this section we first present our analysis about the correlation between credibility scores and MBFC news source categorization. Next we look from what locations in the world the most credible sources come from.

5.1 Correlation Credibility Scores and Source Category

Table 6 shows the average scores of news sources in each MBFC category. NewsGuard scores which are available for 673 of the sources highly agree with MBFC's description which explains that cen-

ter and pro-science categories are the most credible categories while biased sources are less credible, and that sources in the conspiracy and fake-news categories are often unreliable. We however also see a slight tendency that left-biased sources are scoring higher than the right-biased ones.

We run our model on those 673 sources (left "pred" column in the table), as well as on all 3007 sources from MBFC, which means including samples for which NewsGuard scores are unavailable. The model shows a similar but with lower score trend as NewsGuard on each category. The prediction results on 3007 sources for which NewsGuard does not have the scores have again similar pattern. The model assigns higher prediction values to categories like pro-science, left-center, right-center and low values to fake-news, conspiracy and satire sources.

category	labeled only			all samples	
	#	NG	pred.	#	pred.
left	85	77	65	262	53
left-center	185	94	79	454	67
center	122	94	75	398	65
right-center	76	92	74	219	64
right	60	61	53	138	45
pro-science	27	94	88	310	76
conspiracy	39	30	44	284	39
fake-news	76	24	38	124	34
satire	3	5	46	463	39

Table 6: Average scores for sources in each category in MBFC dataset.
labeled only: only news sources which are available on NewsGuard.
all samples: all news sources including ones for which NewsGuard scores are unavailable.
#: number of samples
NG: average NewsGuard score (ground truth).
pred.: average predicted scores.

5.2 Credibility Scores by Region

Wikipedia lists in total 732 news sources as of writing[13]. Those news sources are divided into 9 categories based on region (except the "collection" category which contains sources such as Google News and Yahoo News). While each region contains both credible and questionable sources, on average America and Oceania scored the highest while Africa and India scored the lowest.

[13] https://en.wikipedia.org/wiki/Wikipedia:News_sources

region	count	score
America	67	67
Oceania	69	64
collection	54	61
else	17	58
Asia	153	57
China	13	57
Europe	241	56
India	43	54
Africa	75	53

Table 7: Number of news sources listed in Wikipedia and average predicted score per region category.

6 Conclusion

In this paper we tackled the problem of predicting the credibility scores of news sources. We provide a dataset containing news sources along with their credibility scores. This dataset is manually recorded using the NewsGuard plugin on MBFC's lists of news sources and UNS dataset. The MBFC dataset contains also news source categories. We analysed random forest and linear regression and along with a rich set of features. We performed leave-one-out as well as a cross data evaluation. Our results show that linear regression is more robust for the cross data evaluation. We used this model to turn it to a source credibility labeler. While considerable errors exist, the model seems reliable when its output is outside the middle range. With the grey color to represent this middle range, the new coloring system (green/grey/red) can be used to label all news websites around the world with high confidence. We also showed that our regression model is able to foresee the MBFC categories. Finally we showed credibility scores of sources coming from different locations in the world. The results of this small study show that American and Oceania sources tend to have the highest credibility scores and Africa and India the lowest.

In our future work we will investigate further futures to minimize the RMSE errors. We will turn the linear regression model to a service so it can be accessed by online readers and help them to have a critical thinking about the articles they are reading by simple assessing the credibility of their sources. We will also perform a deeper feature analysis (using ablation experiments) to understand the importance of individual features. Finally we also plan to investigate features which do not come from

commercial web sites but rather implemented in house such as those commonly used in research works.

Acknowledgements

This work was partially supported by the European Union under grant agreements No. 654024 SoBigData, No. 825297 WeVerify and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - GRK 2167, Research Training Group "User-Centred Social Media".

References

Mohammad-Ali Abbasi and Huan Liu. 2013. Measuring user credibility in social media. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, pages 441–448. Springer.

Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36.

Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and Preslav Nakov. 2018. Predicting factuality of reporting and bias of news media sources. *arXiv preprint arXiv:1810.01765*.

Albert-László Barabási and Márton Pósfai. 2016. *Network science*. Cambridge University Press, Cambridge.

Hal Berghel. 2017. Lies, damn lies, and fake news. *Computer*, (2):80–85.

Leo Breiman. 2001. Random forests. *Machine learning*, 45(1):5–32.

Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, P Krishna Gummadi, et al. 2010. Measuring user influence in twitter: The million follower fallacy. *Icwsm*, 10(10-17):30.

K. Douglas, C. S. Ang, and F. Deravi. 2017. Farewell to truth? conspiracy theories and fake news on social media. *The Psychologist*.

Andrew Gelman and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Momchil Hardalov, Ivan Koychev, and Preslav Nakov. 2016. In search of credible news. In *International Conference on Artificial Intelligence: Methodology, Systems, and Applications*, pages 172–180. Springer.

David O Klein and Joshua R Wueller. 2017. Fake news: A legal perspective. *Journal of Internet Law*, 20(10):6–13.

David M Markowitz and Jeffrey T Hancock. 2014. Linguistic traces of a scientific fraud: The case of diederik stapel. *PloS one*, 9(8):e105937.

Subhabrata Mukherjee and Gerhard Weikum. 2015. Leveraging joint interactions for credibility analysis in news communities. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 353–362. ACM.

Subhabrata Mukherjee, Gerhard Weikum, and Cristian Danescu-Niculescu-Mizil. 2014. People on drugs: credibility of user statements in health communities. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 65–74. ACM.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2931–2937.

Trevor Sharot. 1986. Weighting survey results. *Journal of the Market Research Society*, 28(3):269–84.

Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. 2010. Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 261–270. ACM.

Yuto Yamaguchi, Tsubasa Takahashi, Toshiyuki Amagasa, and Hiroyuki Kitagawa. 2010. Turank: Twitter user ranking based on user-tweet graph analysis. In *International Conference on Web Information Systems Engineering*, pages 240–253. Springer.

Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. 2018. Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR)*, 51(2):32.