Textual Characteristics of News Title and Body to Detect Fake News: A Reproducibility Study

Anu Shrestha(✉) and Francesca Spezzano

Computer Science Department, Boise State University, Boise, ID, USA
anushrestha@u.boisestate.edu, francescaspezzano@boisestate.edu

Abstract. Fake news, a deliberately designed news to mislead others, is becoming a big societal threat with its fast dissemination over the Web and social media and its power to shape public opinion. Many researchers have been working to understand the underlying features that help identify these fake news on the Web. Recently, Horne and Adali found, on a small amount of data, that news title stylistic and linguistic features are better than the same type of features extracted from the news body in predicting fake news. In this paper, we present our attempt to reproduce the same results to validate their findings. We show which of their findings can be generalized to larger political and gossip news datasets.

Keywords: Misinformation detection on the web · Fake news · Linguistic analysis

1 Introduction

Social media and online news sources have become the major source of news diet for the increasingly large population instead of traditional media. In 2019, the Pew Research Center reported that more than half (55%) of American adults consume news from online platforms often or sometimes, which is 8% increase since 2018 [13]. With its increase in popularity, social media have also been proven to be an effective platform for fake news proliferation due to its lower cost and convenience of further sharing [16], which has attracted the attention of researchers, making it a global topic of interest. Several studies have been carried out to determine the validity of news relying on linguistic cues derived from the readability and lexical information of the news content [7,11,12].

Horne and Adali [7] conducted a study to understand and analyze the associated language patterns of the title and content of fake news. This paper has gained a lot of attention by the research community, with over 200 citations according to Google Scholar, and became the reference reading to understanding textual content differences between real and fake news. Horne and Adali witnessed that the general assumption about fake news that it is written to camouflage with real news and deceive the reader who does not care about the

D. Hiemstra et al. (Eds.): ECIR 2021, LNCS 12657, pp. 120–133, 2021.
https://doi.org/10.1007/978-3-030-72240-1_9

news sources' veracity is actually not true. In fact, they found the fake news is more similar to satire than to real news, and the focus of fake news is on users who are unlikely to read beyond the title. This sheds light on the necessity of research to understand the significant difference between the title of fake and real news separately from the news body content to mitigate the possible diffusion of the fake news. However, these claims were established based on a small data used in which labels were assigned according to the credibility of the news source, instead of fact-checking, which does not consider the fact that a news source can have mixed credibility and publish both real and fake information.

Thus, we decided to reproduce the paper by Horne and Adali [7] to validate their findings on larger state-of-the-art datasets with labels provided by professional journalists who have fact-checked the news, namely PolitiFact and GossipCop [15] and BuzzFeedNews [12]. Because the news trends continuously evolve, we analyze, similarly to Horne and Adali, news text (from body and title) by focusing on linguistic style, text complexity, and psychological aspects of the text, rather than topic-dependent representations of documents (e.g., [3]). In addition, we expanded the set of emotion features considered in the original paper to explore this aspect of the text further, given that Ghanem et al. [4] recently showed emotions play a key role in detecting false information. We also compare the classification performance of different classifiers beyond linear SVM (the only model used in [7]), and we discuss textual differences between two news domains, namely political and gossip news.

Our experiments confirm most of the original paper's findings regarding title and body feature differences between fake and real news, e.g., fake political news packs a lot in the title. However, differently from Horne and Adali, we found that fake titles contain more stop words than real titles. When using linear SVM to classify fake vs. real news, we confirm that title features outperform body features, but we observe the opposite results if we consider a non-linear and more expressive classifier such as Random Forest.

Furthermore, we show new patterns that were not present in the paper by Horne and Adali, namely fake news title and body express more negative emotions and sentiment than real news, and real news articles are more descriptive than fake news ones. Also, we highlight some differences between two different news domains: political and gossip. For instance, among stylistic, psychology, and complexity features in the news title, psychology features are the most important group of features for gossip news, while the most important group for political news is the one containing stylistic features. This shows how gossip news titles tend to be more persuasive than other news domains.

2 Overview of the Paper by Horne and Adali

In this section, we provide an overview of the approach, features, and findings by Horne and Adali [7].

2.1 Approach

Horne and Adali conducted a content analysis to study fake news by analyzing three small datasets: (i) a dataset (DS1) created by Buzzfeed leading to the 2016 U.S. elections which contains 36 real news stories and 35 fake news stories; (ii) a dataset (DS2) created by using Zimdars' list of fake and misleading news websites [18] and fact-checking website like snopes.com [7], containing 75 stories for each category: real, fake and satire sources; (iii) a dataset (DS3) containing 4000 real and 233 satire articles from a previous study [2]. During the experiments, they considered features from both news body and title for determining the veracity of news and comparing real news vs. fake news vs. satire.

2.2 Features

This research focused on three groups of features, including stylistic features (syntax, text style, and grammatical elements measured by 2015 Linguistic Inquiry and Word Count (LIWC) [10] and the Python Natural Language Toolkit Part of Speech tagger [1]), complexity features to capture details about how complex the article or title is (e.g., words per sentence, syntax tree depth determined by the Stanford Parser and readability level of text), and psychological features to capture emotional (positive/negative), social, and cognitive processes incorporated in news body or title computed by using the LIWC tool. Sentiment analysis was done through SentiStrength [17].

Feature Selection and Anaysis. The goal of feature selection is to avoid overfitting and increase generalizability. Because the datasets were small and the features generated were large, Horne and Adali performed feature selection by leveraging the one-way ANOVA test for those normally distributed features and the Wilcoxon rank-sum test for those that did not pass the normality test. This feature selection concluded with the selection of top 4 features for news body (number of nouns, lexical diversity (TTR), word count, and number of quotes) and news title (percentage of stop words, number of nouns, average word length, and Flesh-Kincaid Grade Readability Index).

Besides, they also used the above mentioned statistical tests to uncover statistically significant feature value differences among news with different labels (fake, satire, and real). If the value of a feature was higher (on average) for real news articles as compared to fake news articles, they denoted this by $R > F$ (and $F > R$ vice versa). We used the same notation while reproducing this experiments in Tables 2 and 3.

2.3 Observation and Evaluation

Horne and Adali's findings show how real news is different from fake and satire news and that fake news and satire have a lot in common across several dimensions. Regarding real vs. fake news (which is the scope of our reproducibility paper), they found that:

(f1) fake news articles tend to be shorter in terms of content, but use repetitive language,[1] smaller words, less punctuation, and fewer quotes (these results is consistent between datasets DS1 and DS2);
(f2) fake news articles require a lower educational level to read, use fewer analytic words, use more personal pronouns and adverbs, but fewer nouns (this result is not consistent between datasets DS1 and DS2 and it is less significative);
(f3) fake titles are longer, contain shorter words, use more all capitalized words, fewer stop words, and fewer nouns overall but more proper nouns (these results is consistent between datasets DS1 and DS2);
(f4) titles are a strong differentiating factor between fake and real news. They performed a binary classification of real vs. fake news separately on news body content and title on dataset DS2. They used the top 4 features from the feature selection process to run a linear SVM model with 5-fold cross-validation. The classification results show 71% accuracy for news body content and 78% accuracy for the title. Thus, they argued that the title is more important in predicting fake vs. real news, and the title and the body of the news should be analyzed separately.

3 Reproducibility

In this section, we describe in detail our attempt to reproduce and generalize findings (f1)–(f4) shown by Horne and Adali in their paper [7].

3.1 Datasets

There is generally limited availability of large scale benchmarks for fake news detection, especially where the ground truth labels are assigned via fact-checking, which is a time-consuming activity. FakeNewsNet [15] and BuzzFeedNews [12] are the only publicly available datasets having fact-checked labels. Thus, in this paper, we use these datasets to conduct our study (Table 1).

FakeNewsNet: PolitiFact and GossipCop. FakeNewsNet consists of two datasets, PolitiFact and GossipCop, from two different domains, i.e., politics and entertainment gossip, respectively. Thus, we used these two datasets separately in our study. Each of these datasets contains details about news content, publisher information, and social engagement information. We only used news content information in this paper.

[1] Repetitive language is measured by using the Type-Token Ratio (TTR) which is the number of unique words in the document by the total number of words in the document. A low TTR means more repetitive language, while a high TTR means more lexical diversity. Horne and Adali claim fake news has more repetitive language but show the opposite result in their paper, i.e., TTR is on average higher for fake than real news (cf. Table 4 in [7]), indicating more lexical diversity for fake than real news. Our results confirms more lexical diversity for fake news as shown in Table 2.

Table 1. Size of datasets used in our study.

Dataset	# Total News	# Fake News	# Real News
PolitiFact	838	378	460
BuzzFeedNews	1,561	299	1,262
GossipCop	19,759	4,734	15,025

PolitiFact contains news with known ground truth labels collected from the fact-checking website PolitiFact.[2] After cleaning the dataset from missing news bodies or titles, we obtained a total of 838 news articles, 378 fake and 460 real.

The GossipCop dataset contains fake news collected from GossipCop[3], which is a fact-checking website for entertainment stories and real news collected from E!Online,[4] a trusted media website for entertainment stories. After cleaning the dataset from missing news bodies or title, we obtained a total of 19,759 news articles, 4,734 fake and 15,025 real.

BuzzFeedNews Dataset. The BuzzFeedNews dataset contains news regarding the 2016 U.S. election published on Facebook by nine news agencies. This dataset[5] contains 1,262 articles that are mostly true, 212 that are a mixture of true and false, and 87 that are false, after cleaning the dataset from missing news bodies or titles. Ground truth is derived from professional journalists at BuzzFeed who have fact-checked the news in the dataset. As also done in the other datasets, we considered false news and news with a mixture of true and false as fake news and mostly true news as real news.

3.2 Features

This section describes the set of features we used in the paper to analyze real vs. fake news. In our implementation, we consider features similar to Horne and Adali [7], namely stylistic features, text complexity features, and psychology features. These features are computed for both the title and body text of the news.

Stylistic Features. We used the subset of LIWC features that represent the functionality of text, including word count (WC), words per sentence (WPS), time orientation (e.g., focus on past (focuspast) and focus on future (focusfuture)), number of personal (I, we, you, she/he – one feature each) and impersonal pronouns, number of quantifying words (quant), number of comparison words

[2] https://www.politifact.com/.
[3] https://www.gossipcop.com/.
[4] https://www.eonline.com/ap.
[5] The BuzzFeedNews dataset is available at https://zenodo.org/record/1239675#.X5riw0JKgXA.

(compare), number of exclamation marks (exlam), number of negations (negate), e.g., no, never, not, number of swear words (swear), number of online slang terms (netspeak), e.g., lol, brb, number of interrogatives, e.g., how, what, why (interrog), number of punctuation symbols (allPunc), number of quotes (quote).

Regarding the part of speech features, we used the Python Natural Language Toolkit part of speech (POS) tagger to compute the number of nouns (NN), proper nouns (NNP), personal pronouns (PRP), possessive pronouns (PRP$), Wh-pronoun (WP), determinants (DT), Wh-determinants (WDT), cardinal numbers (CD), adverbs (RB), interjections (UH), verbs (VB), Adjective (JJ), past tense verbs (VBD), gerund or present participle verbs (VBG), past participle verbs (VBN), non-3rd person singular present verbs (VBP), and third-person singular present verbs (VBZ).

This stylistic group of features also includes the upper case word count (all caps) and percent of stop words (per_stop).

Psychology Features. We computed these features by using the LIWC tool and include the number of analytic words (analytic), insightful words (insight), causal words (cause), discrepancy words (discrep), tentative words (tentat), certainty words (certain), differentiation words (differ), affiliation words (affil), power words, reward words, risk words, personal concern words (work, leisure, religion, money, home, death – one each), anxiety-related words (anx), emotional tone words (tone), and negative (negemo) and positive (posemo) emotional words. This group of features also includes positive (pos) and negative (neg) sentiment metrics as computed by the VADER sentiment analysis tool [5]. We also investigated the importance of features describing emotions expressed through the text, as Ghanem et al. [4] recently showed emotions play a key role in deceiving the reader and can successfully be used to detect false information. Thus, in addition to some emotion features provided by the LIWC tool (as described above), we computed additional emotion features such as anger, joy, sadness, fear, disgust, anticipation, surprise, and trust by using the Emotion Intensity Lexicon (NRC-EIL) [9] and the approach proposed in [8].

Complexity Features. The complexity of text in natural language processing depends on how easily the reader can read and understand a text. We used popular readability measures as complexity features in our analysis: Flesh Kincaid Grade Level (FK), Gunning Fog Index (GI), Simple Measure of Gobbledygook Index (SMOG). Higher scores of these readability measures indicate that the text is easier to read. This group of features also includes lexical diversity or Type-Token Ratio (TTR) and the average length of each word (avg wlen).

3.3 Analysis

Considering all the features from each group, we have a total of 68 features, which can still be too many for the size of the considered datasets (PolitiFact, BuzzFeedNews, and GossipCop) to perform a real vs. fake news articles classification. Therefore, we used the same statistical tests (ANOVA and Wilcoxon rank-sum) used by Horne and Adali to perform feature selection and analysis. For each dataset, features are sorted by F-value in descending order to determine the importance, and only features where the two averages (real vs. fake) were significantly different according to the statistical test (p-value < 0.05) were considered. Among these features, we selected a number of features up to the square root of the training set size (rule of thumb) for both news body and title to feed the classification algorithm.

Instead of just using the linear SVM classifier as done by Horne and Adali, we compared the performances of different classification algorithms, namely Logistic Regression (LR) classifier with L2 regularization, linear Support Vector Machine (SVM), and Random Forest (RF), with default parameters. As the datasets we considered are not balanced, we used class weighting to deal with class imbalance, stratified 5-fold cross-validation, and results are reported by using AUROC and average precision (AvgP).

3.4 Results

Feature Statistical Analysis. We start our analysis by checking whether Horne and Adali's findings (f1), (f2), and (f3) reported in Sect. 2.3 are confirmed in the three larger datasets we considered, namely PolitiFact and BuzzFeed (political news datasets), and GossipCop (gossip news dataset). To analyze these findings we refer to the results reported in Table 2 for news body text and Table 3 for news title.

Regarding finding (f1) (cf. Table 2), we confirm that fake news articles have a shorter content (WC) and use less punctuation (allPunc) than real news articles in all the three datasets we considered, and fake political articles have more lexical diversity (TTR) than real political articles. Our analysis does not allow us to generalize the finding that fake news articles use smaller words (avg wlen) and fewer quotes (true in BuzzFeedNews, but not in Politifact and GossipCop).

Regarding finding (f2) (cf. Table 2), we can generalize the finding that fake news articles use fewer analytic words (true in BuzzFeedNews and GossipCop). We found that fake news articles require a lower educational level to read (as measured by FK, GI, and SMOG readability indexes) only in one dataset (BuzzFeedNews) while the opposite trend holds for GossipCop dataset; the use of more personal pronouns (PRP), adverbs (RB), and proper nouns (NNP) in fake news articles is not confirmed in our analysis. We observe fake titles containing more proper nouns (NNP) in all the three datasets considered.

Regarding finding (f3) (cf. Table 3), we confirm that fake titles have more proper nouns (NNP) than real titles in all the three datasets we considered and have fewer nouns (NN) in BuzzFeedNews and GossipCop. Also, we confirm that fake political titles are longer (WC and WPS), use more capitalized words (all caps) (they also use more possessive pronouns – PRP$), and contain shorter words (avg wlen). Our analysis does not confirm the fact that fake titles contain

Table 2. Features that differ in body of news content. All differences are statistically significant ($p < 0.05$).

Features	PolitiFact	BuzzFeed	GossipCop	Features	PolitiFact	BuzzFeed	GossipCop
allPunc	$R > F$	$R > F$	$R > F$	analytic	$F > R$	$R > F$	$R > F$
exclam	$F > R$	$F > R$	$F > R$	quote	$F > R$	$R > F$	$F > R$
tone	$R > F$	$R > F$	$R > F$	WC	$R > F$	$R > F$	$R > F$
WPS		$R > F$	$R > F$	affect		$F > R$	$R > F$
affil	$R > F$		$F > R$	cause		$F > R$	$F > R$
certain		$F > R$	$F > R$	all caps	$R > F$	$R > F$	$R > F$
differ	$R > F$	$F > R$	$F > R$	discrep	$R > F$	$F > R$	$F > R$
FK		$R > F$		focusfuture			$F > R$
GI		$R > F$	$F > R$	i			$R > F$
insight		$F > R$		interrog			$R > F$
leisure	$F > R$		$R > F$	TTR	$F > R$	$F > R$	
money	$R > F$			negate		$F > R$	$F > R$
netspeak			$R > F$	JJ	$R > F$	$R > F$	$R > F$
RB	$R > F$	$R > F$		CD	$R > F$	$R > F$	$R > F$
DT	$R > F$	$R > F$	$R > F$	UH	$R > F$		
NN	$R > F$	$R > F$	$R > F$	NNP	$R > F$	$R > F$	$R > F$
PRP	$R > F$	$R > F$	$R > F$	$\mathrm{PRP\$}$	$R > F$	$R > F$	
VBD	$R > F$	$R > F$	$R > F$	VBG	$R > F$	$R > F$	
VBN	$R > F$	$R > F$		VBP	$R > F$	$R > F$	$R > F$
VBZ	$R > F$	$R > F$		VB	$R > F$	$R > F$	$R > F$
WP	$R > F$	$R > F$	$R > F$	WDT	$R > F$	$R > F$	$R > F$
per_stop	$F > R$	$F > R$	$F > R$	power		$R > F$	$R > F$
quant	$R > F$			relig	$F > R$	$F > R$	$R > F$
reward			$R > F$	risk			$F > R$
sheshe	$F > R$		$F > R$	SMOG		$R > F$	$F > R$
swear	$F > R$	$F > R$		tentat		$F > R$	$F > R$
we	$R > F$		$R > F$	avg wlen		$R > F$	
work	$R > F$	$R > F$		you	$R > F$	$F > R$	$R > F$
compare		$R > F$		focuspast	$F > R$		$F > R$
neg	$F > R$	$F > R$	$F > R$	surprise	$F > R$		
disgust	$F > R$	$F > R$	$F > R$	negemo	$F > R$	$F > R$	$F > R$
pos	$R > F$		$R > F$	fear	$F > R$	$F > R$	
posemo	$R > F$		$R > F$	anx	$F > R$	$F > R$	$F > R$
sadness	$F > R$	$F > R$	$F > R$	anger		$F > R$	$F > R$
trust			$F > R$	joy			$F > R$

fewer stop words (per_stop). Similarly, we observe that fake news articles contain more stop words.

Furthermore, our results in Tables 2 and 3 highlight new patterns that were not present in the analysis performed by Horne and Adali. Specifically, we found that real news articles use a more positive tone and more nouns (NN), determinants (DT), wh-determinants (WDT), verbs (VB), past tense verbs (VBD), Wh-pronouns (WP), and adjectives (JJ) in all the three datasets considered.

Table 3. Features that differ in the title of news content. All differences are statistically significant ($p < 0.05$).

Features	PolitiFact	BuzzFeed	GossipCop	Features	PolitiFact	BuzzFeed	GossipCop
WC	$F > R$	$F > R$		avg wlen	$R > F$	$R > F$	$F > R$
quote	$F > R$	$F > R$	$F > R$	allPunc	$R > F$		$F > R$
exclam	$F > R$	$F > R$	$F > R$	tone	$R > F$	$R > F$	$R > F$
WPS	$F > R$	$F > R$	$R > F$	affect	$F > R$		$R > F$
affil			$F > R$	compare	$F > R$		$R > F$
differ			$F > R$	discrep	$F > R$		$F > R$
focusfuture	$F > R$		$F > R$	focuspast	$F > R$	$F > R$	
insight		$F > R$		interrog			$R > F$
leisure			$R > F$	TTR	$F > R$		$F > R$
money		$R > F$		negate			$F > R$
netspeak	$R > F$		$R > F$	JJ		$R > F$	$R > F$
UH			$F > R$	GI	$F > R$		$F > R$
FK	$F > R$		$F > R$	SMOG	$F > R$		$F > R$
analytic		$R > F$	$R > F$	all caps	$F > R$	$F > R$	
NN		$R > F$	$R > F$	NNP	$F > R$	$F > R$	$F > R$
PRP	$F > R$	$F > R$		PRP$	$F > R$	$F > R$	$R > F$
DT			$R > F$	RB	$F > R$		$F > R$
VBD	$F > R$			VBG	$F > R$		$F > R$
VBN	$F > R$			VBP	$F > R$	$F > R$	
VBZ	$F > R$	$R > F$		VB	$F > R$		$F > R$
WP		$F > R$		per_stop	$F > R$	$F > R$	
quant			$R > F$	relig	$F > R$	$F > R$	
reward			$R > F$	risk			$F > R$
work	$R > F$	$R > F$		i	$F > R$		$R > F$
you			$R > F$	shehe	$F > R$	$F > R$	
CD			$R > F$	fear	$F > R$	$F > R$	$F > R$
neg	$F > R$	$F > R$	$F > R$	sadness	$F > R$	$F > R$	$F > R$
surprise	$F > R$	$R > F$		anger	$F > R$	$F > R$	$F > R$
negemo	$F > R$		$F > R$	trust	$R > F$		$R > F$
disgust	$F > R$	$F > R$	$F > R$	pos			$R > F$
posemo			$R > F$	anx			$F > R$
joy			$R > F$				

This indicates that real news articles are more descriptive than fake news articles. Also, fake news titles and bodies use more exclamation marks (exclam) than real news titles (true in all the three datasets considered).

In addition, we observe that fake titles express more negative emotions (anger, sadness, fear, and disgust) and negative sentiment (neg) than real titles consistently across all the three considered datasets. This pattern is also true for fake news body. In contrast, real titles tend to express more positive emotions (trust, posemo, joy) and positive sentiment (pos), but this is less consistent across datasets. When selecting information, people have a sensitivity to negative information [6]. This negativity bias induces people to pay more attention

Table 4. News title vs. news body features for detecting fake news on the PolitiFact, BuzzFeedNews, and GossipCop datasets: stylistic, psychology, and complexity features. Best results for both news title and body are in bold. Best overall results between news title and body are shaded.

	PolitiFact		BuzzFeedNews		GossipCop	
Features	AUROC	AvgP	AUROC	AvgP	AUROC	AvgP
News body (SVM)	0.583	0.466	0.614	0.257	0.623	0.327
News body (LR)	0.855	0.809	0.728	0.351	0.703	0.437
News body (RF)	**0.911**	**0.878**	**0.785**	**0.417**	**0.782**	**0.630**
News Title (SVM)	0.833	0.804	0.669	0.317	0.588	0.309
News Title (LR)	0.849	0.813	0.787	0.423	0.663	0.380
News Title (RF)	**0.867**	**0.823**	**0.812**	**0.424**	**0.715**	**0.490**

Table 5. News title vs. news body features for detecting fake news on the PolitiFact, BuzzFeedNews, and GossipCop datasets: same four features as in Horne and Adali [7] – NN, TTR, WC, and Quote for news body and FK, NN, per_stop, and avg wlen for title. Best results for both news title and body are in bold. Best overall results between news title and body are shaded.

	PolitiFact		BuzzFeedNews		GossipCop	
Features	AUROC	AvgP	AUROC	AvgP	AUROC	AvgP
News Body (SVM)	0.544	0.445	0.678	0.292	0.500	0.232
News Body (LR)	0.754	0.663	0.691	0.297	0.534	0.251
News Body (RF)	**0.861**	**0.803**	**0.708**	**0.342**	**0.631**	**0.42**
News Title (SVM)	0.649	0.531	0.713	**0.342**	0.528	0.250
News Title (LR)	0.643	0.530	**0.716**	**0.342**	0.530	0.251
News Title (RF)	**0.735**	**0.612**	0.706	0.330	**0.582**	**0.332**

to negative news, hence fake news tiles, bodies, and even associated images [14] express negative emotions to be catchier and circulate more among people.

Furthermore, there are some differences between political and gossip news. We found that fake political news articles have more religion-related words (relig) than real political news articles, while fake gossip news articles have fewer religion-related words; fake political news titles contain shorter words (avg wlen), and more words per sentence (WPS) and possessive pronouns (PRP$) than real political news titles, while this is the opposite for gossip news titles.

Real vs. Fake News Classification. Finding (f4) by Horne and Adali claims that title features are more informative (i.e., achieve higher accuracy) than news body features in classifying fake vs. real news with a linear SVM. Table 4 shows our classification results by comparing three classifiers, and when we used a number of features up to the square root of the training set size. We observe

Table 6. Feature group ablation for news title and body when the best classifier (Random Forest) is used on the PolitiFact, BuzzFeedNews, and GossipCop datasets. Best results for both news title and body are in bold.

Features	PolitiFact		BuzzFeedNews		GossipCop	
	AUROC	AvgP	AUROC	AvgP	AUROC	AvgP
News body						
Stylistic (RF)	**0.882**	**0.838**	**0.753**	**0.382**	**0.752**	**0.590**
Psychology (RF)	0.723	0.662	0.681	0.319	0.713	0.509
Complexity (RF)	0.804	0.708	0.630	0.285	0.000	0.000
News title						
Stylistic (RF)	**0.819**	**0.729**	**0.805**	**0.433**	0.634	0.365
Psychology (RF)	0.791	0.691	0.645	0.320	**0.651**	**0.407**
Complexity (RF)	0.583	0.486	0.555	0.257	0.553	0.287

that when we consider the linear SVM classifier, finding (f4) is confirmed, i.e., AUROC and average precision scores are higher for the title than the news body. However, Random Forest is the best classifier for both news body and title and outperforms linear SVM. When we consider Random Forest as the classifier, finding (f4) is reversed, i.e., AUROC and average precision scores are higher for news body than news title (this is true for two out of three of the datasets considered). We observe a similar trend also when we consider only the four features chosen by Horne and Adali to perform the classification (see results reported in Table 5). Of course, considering more than four features as we did in Table 4 results in better AUROC and average precision in all the three datasets.

Thus, our experiments reveal that whether or not the title is more informative than the news body depends on the chosen classifier. A non-linear classifier such as Random Forest has higher expressive power and outperforms linear SVM. Thus, if we choose the best classifier, namely Random Forest, finding (f4) does not hold in the larger datasets we considered. Having more information helps the Random Forest classifier to increase classification performances.

In addition, we performed feature ablation by feature group (style, psychology, and complexity) when the best classifier (Random Forest) is used. Results are reported in Table 6. We observe that stylistic features are the most important features in both title and news body for political news. For gossip news, stylistic features are the most important news body features, while psychology features are the most important features in title. Interestingly, this validates the definition of gossip as “small talk” that is originated from evolutionary psychology and has the basic intent to share information about third persons to indulge people in some discussion. Also, the reason people like gossip is because it is tempting and fun. Thus, the news title of gossip stories are written with more psychological words like tone and affect, e.g., “Angelina Jolie Can’t Get Over

Heartbreak Of Losing Brad Pitt—Real Reason For Fury, Says Source" to catch readers attention even though the body text is not that engaging.

4 How to Reproduce Our Experiments

For reproducibility propose, we made our code available in a GitHub repository.[6] Because we did not directly collect the datasets, we are not uploading them in our repository, but we provide instructions on finding and downloading them. In our repository, we make our code available for extracting the features that are considered in this paper, including complexity, stylistic and psychology features extracted using NLTK part-of-speech, VADER Sentiment Analyser and the Emotion Intensity Lexicon (NRC-EIL),[7] except LIWC features as the LIWC tool has proprietary dictionaries whose licence should be purchased. LIWC features can be computed in two ways: (1) by using the software tool to compute the features, or (2) by downloading the dictionary provided by the tool for which we have provided code to extract features using the dictionary. In addition, we also provide code for the statistical test performed in this paper to reproduce Tables 2 and 3. Likewise we also provide code for the classification to reproduce Tables 4, 5 and 6.

5 Conclusions

In this paper, we reproduced the study by Horne and Adali [7] of the relative importance of news body and title in detecting fake news. We extended their experimental setting by using larger real and fake news datasets with ground truth at the news level, considering additional features describing emotions expressed through the text, comparing different classification algorithms, and highlighting differences between political and gossip news domains. Our experiments have shown that some of the original paper's observations are not the same as the trend of news writing is continuously evolving. For instance, the finding that the news title is more informative and plays an important role in discerning the news's veracity is confirmed if we use the same classifier, linear SVM, as in [7], but using a non-linear classifier such as Random Forest reverses the finding. Finally, we provide evidence that fake news title and body attract readers' attention with more negative emotions and sentiment, while real news articles are more descriptive.

Acknowledgements. This work has been supported by the National Science Foundation under Award no. 1943370. We thank Ashlee Milton and Maria Soledad Pera for providing us the code used in their paper [8] to compute emotional features.

[6] https://github.com/shresthaanu/ECIR21TextualCharacteristicsOfFakeNews.

[7] The NRC-EIL lexicon should be downloaded at https://www.saifmohammad.com/WebPages/AffectIntensity.htm.

References

1. Bird, S.: NLTK: the natural language toolkit. In: Proceedings of the COLING/ACL on Interactive Presentation Sessions, pp. 69–72 (2006)
2. Burfoot, C., Baldwin, T.: Automatic satire detection: are you having a laugh? In: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, pp. 161–164 (2009)
3. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
4. Ghanem, B., Rosso, P., Rangel, F.: An emotional analysis of false information in social media and news articles. ACM Trans. Internet Technol. (TOIT) **20**(2), 1–18 (2020)
5. Gilbert, C., Hutto, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Eighth International Conference on Weblogs and Social Media (ICWSM 2014), vol. 81, p. 82 (2014)
6. Hills, T.T.: The dark side of information proliferation. Perspect. Psychol. Sci. **14**(3), 323–330 (2019)
7. Horne, B.D., Adali, S.: This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In: The 2nd International Workshop on News and Public Opinion at ICWSM (2017)
8. Milton, A., Batista, L., Allen, G., Gao, S., Ng, Y., Pera, M.S.: "Don't judge a book by its cover": exploring book traits children favor. In: RecSys 2020: Fourteenth ACM Conference on Recommender Systems, Virtual Event, Brazil, 22–26 September 2020, pp. 669–674. ACM (2020)
9. Mohammad, S.: Word affect intensities. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 7–12 May 2018 (2018)
10. Pennebaker, J.W., Boyd, R.L., Jordan, K., Blackburn, K.: The development and psychometric properties of LIWC2015. Technical report (2015)
11. Pérez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R.: Automatic detection of fake news. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 3391–3401 (2018)
12. Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., Stein, B.: A stylometric inquiry into hyperpartisan and fake news. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, 15–20 July 2018, Volume 1: Long Papers, pp. 231–240 (2018)
13. Shearer, E., Grieco, E.: Americans are wary of the role social media sites play in delivering the news (2019)
14. Shrestha, A., Spezzano, F., Gurunathan, I.: Multi-modal analysis of misleading political news. In: van Duijn, M., Preuss, M., Spaiser, V., Takes, F., Verberne, S. (eds.) MISDOOM 2020. LNCS, vol. 12259, pp. 261–276. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-61841-4_18
15. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H.: Fakenewsnet: a data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data **8**(3), 171–188 (2020)
16. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: a data mining perspective. ACM SIGKDD Explor. Newsl. **19**(1), 22–36 (2017)

17. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A.: Sentiment strength detection in short informal text. J. Am. Soc. Inform. Sci. Technol. **61**(12), 2544–2558 (2010)
18. Zimdars: False, misleading, clickbait-y, and satirical news sources (2016). https://docs.google.com/document/d/10eA5-mCZLSS4MQY5QGb5ewC3VAL-6pLkT53V_81ZyitM/preview