Fake News Detection Based on Explicit and Implicit Signals of a Hybrid Crowd: Proposal, Impacts and Perspectives

Paulo Márcio Souza Freire
Ronaldo Ribeiro Goldschmidt
paulomsfreire@ime.eb.br
ronaldo.rgold@ime.eb.br
Military Institute of Engineering (IME)
Rio de Janeiro, RJ

Abstract

The problem of automatic Fake News detection in digital media of news distribution (DMND - e.g., social networks, online newspaper) has become even more relevant. Among the main detection approaches, the one based on crowd signals from DMND users has stood out by obtaining promising results. Although promising, the Crowd Signals approach has a significant limitation: it depends on the explicit user opinion (which is not always available) about the classification of news. Facing this limitation, the present work raises the hypothesis that it is possible to build models of Fake News detection with a performance comparable to the Crowd Signals based approach, avoiding the dependence on the explicit opinion of DMND users. To validate this hypothesis, the present work proposes *HCS*, an approach based on crowd signals that considers implicit user opinions instead of the explicit ones. The implicit opinions are inferred from the behavior of users concerning the dissemination of the news. Inspired in Meta-Learning, the *HCS* can also use the explicit opinions from machines (news classification models) to complement the implicit user opinions by means of hybrid Crowds. Experiments presented significant evidence that confirms the raised hypothesis.

Keywords: artificial intelligence, crowdsourcing, disinformation, social networks and social media

1 Introduction

Despite their advantages, some digital media of news distribution (DMND), such as social networks, allow any person, regardless of her credibility, to disseminate the news with an intense power of propagation [6]. Such permissiveness has amplified the dissemination of Fake News, a particular type of false news whose dissemination happens intentionally (i.e., with intention to cause disinformation) [4].

Given this scenario, machine learning approaches based on reputation for detecting intentionally false news in DMND have been proposed. Among them, approaches based on Crowd Signals of DMND users has been highlighted for obtaining promising results [7]. In essence, in order to classify a recent piece of news n as fake or not fake, such an approach explores the collective sense by combining opinions (signals, i.e., votes about the classification of n). Opinions about n are explicitly provided by the users through a DMND functionality and pondered by the reputation of these users. The reputations are inferred from right and wrong opinions given by the users to news previously evaluated [7].

Although promising, the Crowd Signals based approach has a significant limitation that impairs its application to most DMND: it depends on the user's explicit opinion (which is not always available) about the news. This unavailability may be caused by two reasons. The first is that generally DMND do not provide a functionality to collect user opinion about the news. The second, and most important, is that, even when a collection functionality is available, the approach depends on the goodwill of users to give their opinion about every news accessed in digital media [3].

Given the above, the following research issue is posed: *Is it possible to detect Fake News in DMND through signals of the members (users) of a crowd and from their reputations, without depending on the explicit opinion of these users?*

To answer the above question, the present work raises the hypothesis that it is possible to build Fake News detection models, with a performance comparable to the Crowd Signals based approach, without depending on the explicit opinions of DMND users. In order to validate such hypothesis, this work proposes *HCS (Hybrid Crowd Signals*). *HCS* is an approach based on crowd signals that considers *implicit* user opinions instead of the explicit ones. According to *HCS*, implicit opinions are inferred from the behavior of users concerning the dissemination of the analyzed news. This inference is based on the principle that when a user

In: IV Concurso de Teses e Dissertações (CTD 2022), Curitiba, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2022.

ISSN 2596-1683

disseminates news in digital media, he wants, maliciously or not, to demonstrate that he considers the news to be true.

Inspired by meta-learning, *HCS* can also use the explicit opinions of machines (i.e., news classification models that already exist in the literature) to complement the implicit opinions of users. To this end, when a machine classifies a piece of news as fake or not fake, this classification represents the machine's explicit opinion about a given piece of news. It is also important to highlight that machines with different information demands can detect Fake News from different types of media (e.g., text, image, and audio). In this way, *HCS* uses the formation of a hybrid Crowd, since Crowd members can be both disseminating users and machines.

Experiments carried out in five datasets presented significant evidence that confirms the raised hypothesis. Besides that, the results also revealed a performance improvement of *HCS* when the implicit opinions of the users were combined with the explicit opinions of the machines.

In line with the abovementioned results, this work contributed to the state-of-the-art in the areas of Multimedia Systems and the Web, more specifically to the following interest topics of WebMedia 2022 Main Track: Artificial Intelligence, Crowdsourcing, Disinformation, Social Networks and Social Media.

This article is organized as follows. The proposed approach is detailed in Section 2. Following that, Section 3 describes the experiments and the results obtained. Finally, Section 4 depicts final considerations, as well as the possibilities regarding future researches.

2 Proposed Approach

The proposed approach *HCS (Hybrid Crowd Signals)* is based on crowd signals that considers *implicit* user opinions instead of the explicit ones. According to *HCS*, implicit opinions are inferred from the behavior of users concerning the dissemination of the analyzed news. This inference is based on the principle that when a user disseminates news in digital media, he wants to, maliciously or not, demonstrate that he considers the news to be true. The fact that user u posts the news n is an implicit signal that, in the opinion of u, n is not fake. As such, when a user u decides to disseminate n in digital media, maliciously or not, u wants to show to other users in digital media that they consider n as true. Such a principle is inspired by the quote from philosopher *Habermas*, according to which every communicative action carries with it an, inevitable, claim to truth [5]. Based on the hits or misses of the disseminating users in giving implicit opinions about the news already evaluated and whose labels are known, the reputation of these users is obtained and used to weight their respective opinions about the news to be evaluated. In this way, unlike the Crowd Signals based approach to detect Fake News, the *HCS* does not need that the digital media provides a functionality for the user to express his or her opinion about the news. Just as, it is not necessary to rely on the users' goodwill in giving their opinion.

Inspired in meta-learning, *HCS* yet allows the formation of a hybrid Crowd, once Crowd members may be, in a digital media, disseminating users as well as machines (models of news classification) made available for *HCS* usage. These machines are implementations of classification methods to detect Fake News already accessible in literature. Therefore, when a machine m classifies a piece of news n as fake (resp. not fake), the explicit opinion of this member m in a Crowd is that n is fake (resp. not fake). In this sense, the reputation of every machine, used to weigh their respective opinions, is obtained by means of their hits or misses when giving explicit opinions about already classified news. When using opinions from trained machines, and with the reputation already measured, *HCS* seeks to complement the users' implicit opinions. This complementation aims at avoiding the *cold-start* problem. This problem may be basically materialized when these users do not have their reputation sufficiently measured, as they historically disseminated few news.

Given the above, to detect Fake News in DMND based on explicit and implicit signals presented by members of a hybrid Crowd, *HCS* has three macro-functional stages (*Crowd composition*, *Member Reputation Admeasurement* and *News Classification*) that are executed for each news n^D belonging to the set of news to be detected N^D.

Hence, *HCS* initiates by the *Crowd composition* stage has the function of composing, dynamically, a hybrid Crowd C, formed by disseminating users and/or by machines M. Next, *Member Reputation Admeasurement* stage is responsible for measuring the reputation of each member $c_i \in C$. Thus, the reputation of each c_i is expressed by the probability of hitting and, consequently, missing the opinions already provided by c_i about fake or not fake news examined before. The admeasurement of C members reputation depends on the availability of a set N^L of news disseminate in DMND and already labeled (i.e., checked by a reliable source) as fake (f) or not fake ($\overline{f}$). *HCS* having calculated these probabilities, it is capable of storing in C^R each $c_i \in C$ with their respective reputation. Finally, in *News Classification* stage, for each $c_i^R \in C^R$, *HCS* utilizes c_i and its respective reputation (probabilities of hitting or missing) to conclude whether n^D is fake or not, in which n^D becomes n^C and is stored in the set of news already classified N^C. A detailed view of approach can be obtained in Freire et al. [2].

3 Experiments and Results

3.1 Experimental Methodology

The choice of the datasets (i.e., *Gossip, PolitiFact, Gossip2. FakeNewsSet and FakeBr*) was guided by one main reason: the five datasets were used and made available by recent and relevant publications [1].

We chose the *Detective* method as the baseline of our experiments because it was proposed by Tschiatschek et al. [7], which, so far as it was possible to observe, is the major study that has followed the Crowd Signals based approach for Fake News detection.

For the *Detective*, the experimental methodology used in our experiments was similar to the optimal methodology (*Optimal Detective*) followed by Tschiatschek et al. [7]. Although less realistic, this methodology leads to the highest precision results produced by the *Detective*.

Concerning *HCS*'s configuration for the experiments, the probabilities were calculated automatically from the implicit opinions of disseminating users and/or explicit opinions of machines about news stored in N^L. Besides, the set of available machines was $M = \{SL\_RF, SL\_XGBOOST, SL\_SVM, FNE, DMText\}$. Each machine is represented by an existing traditional or specific Fake News detection method. A detailed view of each machine can be obtained in Freire et al. [2].

In order to train the machines of the hybrid crowd to be used in the experiments and evaluate *HCS* and *Optimal Detective*, it was necessary to divide each dataset into two disjoint subsets with a fifty-fifty data proportion. While the first subset (N^{Tr}) was used to train the machines in M, the second one (N^{Ex}) was used to evaluate and compare *HCS* and *Optimal Detective*.

To assess *HCS* and *Optimal Detective* in each dataset, a *10-fold cross-validation* was applied to the corresponding N^{Ex} set. Thus, for each of the 10 rounds, two temporary subsets were created: N^L and N^D. N^L was formed by 90% of N^{Ex} (i.e., news to measure Crowd members reputation) and N^D was composed by the remaining 10% (i.e., news to be classified). Accuracy, precision, recall and F1 were the adopted metrics for performance evaluation.

Besides the baseline method (i.e., Optimal Detective) and the machines in M, the experiments encompassed two methods that implement *HCS* (i.e., *HCS-I* and *HCS-F*). While *HCS-F* combines the implicit opinions from users with the explicit opinions from machines in M, *HCS-I* only takes into account the implicit opinions from users.

3.2 Results and Discussion

This subsection discusses the results of the experiments under two perspectives. The first analyzes and compares the performances of the *HCS-I* and *HCS-F* methods with the ones achieved by the machines in M. The idea behind this analysis is to evaluate whether the collective sense implemented by crowds to detect Fake News can overcome the individual machines. The second perspective tries to evaluate whether the hypothesis raised by this work is valid, i.e., *it is possible to build Fake News detection models, with a performance comparable to the Crowd Signals based approach, without depending on the explicit opinions.* To this end, it compares the results produced by the *HCS-I* and *HCS-F* methods with the results obtained by the *Optimal Detective*.

Following the first perspective, Table 1 presents the values obtained individually by each machine with the achieved by the *HCS-I* and *HCS-F* methods. Note that, for *HCS-F*, the machines used in crowd are informed between parentheses.

Table 1. Results of Machines x HCS methods in each dataset

Dataset	Method	Accuracy $[\mu \pm \sigma]$	Precision $[\mu \pm \sigma]$	Recall $[\mu \pm \sigma]$	F1 $[\mu \pm \sigma]$
Gossip	SL_RF	0.9304 ± 0.0083	0.9556 ± 0.0100	0.9009 ± 0.0134	0.9274 ± 0.0095
	SL_XGBOOST	0.9308 ± 0.0094	0.9653 ± 0.0112	0.8920 ± 0.0151	0.9272 ± 0.0109
	SL_SVM	0.9083 ± 0.0094	0.9651 ± 0.0120	0.8446 ± 0.0173	0.9008 ± 0.0131
	HCS-I	0.9389 ± 0.0094	0.9937 ± 0.0031	0.8822 ± 0.0205	0.9345 ± 0.0106
	HCS-F (SL_RF,SL_XGBOOST, SL_SVM)	**0.9671 ± 0.0094**	**0.9940 ± 0.0029**	**0.9391 ± 0.0207**	**0.9657 ± 0.0100**
PolitiFact	SL_RF	0.7134 ± 0.0802	0.7041 ± 0.0859	0.7485 ± 0.0943	0.7220 ± 0.0723
	SL_XGBOOST	0.6884 ± 0.0637	0.6675 ± 0.1062	0.7456 ± 0.1047	0.7000 ± 0.0908
	SL_SVM	0.6075 ± 0.0881	0.5694 ± 0.0838	**0.9344 ± 0.0890**	0.7027 ± 0.0713
	HCS-I	0.9013 ± 0.0484	0.9761 ± 0.0391	0.8283 ± 0.0897	0.8927 ± 0.0506
	HCS-F (SL_RF,SL_XGBOOST, SL_SVM)	**0.9109 ± 0.0506**	**0.9796 ± 0.0336**	0.8437 ± 0.0916	**0.9034 ± 0.0537**
Gossip2	SL_RF	0.8079 ± 0.0228	0.8311 ± 0.0383	0.7814 ± 0.0346	0.8047 ± 0.0236
	SL_XGBOOST	0.8140 ± 0.0344	0.8501 ± 0.0342	0.7730 ± 0.0572	0.8082 ± 0.0319
	SL_SVM	0.6279 ± 0.0324	0.7910 ± 0.0707	0.3654 ± 0.0398	0.4983 ± 0.0402
	HCS-I	0.9078 ± 0.0238	0.9708 ± 0.0218	0.8444 ± 0.0395	0.9027 ± 0.0255
	HCS-F (SL_RF,SL_XGBOOST, SL_SVM)	**0,9104 ± 0.0249**	**0.9731 ± 0.0236**	**0.8473 ± 0.0436**	**0.9052 ± 0.0288**
FakeNewsSet	SL_RF	0.8389 ± 0.0631	0.8379 ± 0.0909	0.8449 ± 0.0887	0.8377 ± 0.0860
	SL_XGBOOST	0.8117 ± 0.0864	0.8375 ± 0.0884	0.7699 ± 0.1287	0.7983 ± 0.0953
	SL_SVM	0.6295 ± 0.0818	0.9264 ± 0.1202	0.0288 ± 0.1006	0.4303 ± 0.1215
	DMText	0.8868 ± 0.0432	0.9300 ± 0.8541	0.9405 ± 0.0588	0.8930 ± 0.0356
	FNE	0.9098 ± 0.0423	0.9521 ± 0.8709	**0.9629 ± 0.0551**	0.9123 ± 0.0440
	HCS-I	0.9179 ± 0.0397	**0.9832 ± 0.0404**	0.8542 ± 0.0852	0.9109 ± 0.0424
	HCS-F (SL_RF,SL_XGBOOST, SL_SVM, DMText, FNE)	**0.9639 ± 0.0281**	0.9831 ± 0.0274	0.9459 ± 0.0511	**0.9631 ± 0.0250**
FakeBr	DMText	0.9330 ± 0.0109	0.9354 ± 0.0139	0.9296 ± 0.0101	0.9325 ± 0.0108
	FNE	0.9339 ± 0.0105	0.9315 ± 0.0126	0.9364 ± 0.0113	0.9339 ± 0.0093
	HCS-I	0.9987 ± 0.0028	**0.9986 ± 0.0045**	0.9990 ± 0.0032	0.9988 ± 0.0026
	HCS-F (DMText, FNE)	**0.9989 ± 0.0026**	0.9980 ± 0.0046	**1.0000 ± 0.0000**	**0.9990 ± 0.0023**

Overall, *HCS*'s methods overcame the machines used individually. In the case of *HCS-I*, the results provide experimental evidences that reinforce the current literature's belief that consider the implicit *Crowd* approach as a promising solution [2]. Another aspect to be mentioned is related to the results that indicate *HCS-F* as an interesting way to combine the outcomes (i.e., opinions) of different existing Fake News detection methods.

In line with the second perspective, Table 2 summarizes the results produced by *HCS*'s methods and by the *Optimal Detective*. These results are analyzed in detail bellow.

The first aspect to be highlighted is the close performance reached by both *HCS-I* and *Optimal Detective* in all datasets. The intersections between the corresponding intervals $[\mu \pm \sigma]$ of each evaluation metric are significative. These values indicate that, although being submitted to a more realistic experimental methodology and inferring user opinions, *HCS-I* reached results comparable to the ones produced by *Optimal Detective*, a method that, different from the proposed approach, demands explicit opinions of those users. In other words, the results obtained provide experimental evidence that *implicit Crowd Signals* may be used to detect Fake News in DMND, dismissing the demand for explicit opinions.

Another relevant aspect worthy to note is that *HCS-F* and *HCS-I* outperformed *Optimal Detective* in datasets *FakeBr* and *Gossip*. One possible reason for such results is that, different from the others, these datasets contain a high number

Table 2. Results of Optimal Detective x HCS in each dataset

Dataset	Method	Accuracy [$\mu \pm \sigma$]	Precision [$\mu \pm \sigma$]	Recall [$\mu \pm \sigma$]	F1 [$\mu \pm \sigma$]
Gossip	Optimal Detective	0.9554 ± 0.0080	**1.0000** ± 0.0000	0.9099 ± 0.0153	0.9528 ± 0.0085
	HCS-I	0.9389 ± 0.0094	0.9937 ± 0.0031	0.8822 ± 0.0205	0.9345 ± 0.0106
	HCS-F (SL_RF,SL_XGBOOST, SL_SVM)	**0.9671** ± 0.0094	0.9940 ± 0.0029	**0.9391** ± 0.0207	**0.9657** ± 0.0100
PolitiFact	Optimal Detective	**0.9896** ± 0.0169	**1.0000** ± 0.0000	**0.9814** ± 0.0306	**0.9904** ± 0.0158
	HCS-I	0.9013 ± 0.0484	0.9761 ± 0.0391	0.8283 ± 0.0897	0.8927 ± 0.0506
	HCS-F (SL_RF,SL_XGBOOST, SL_SVM)	0.9109 ± 0.0506	0.9796 ± 0.0336	0.8437 ± 0.0916	0.9034 ± 0.0537
Gossip2	Optimal Detective	**0.9967** ± 0.0043	**1.0000** ± 0.0000	**0.9940** ± 0.0078	**0.9970** ± 0.0039
	HCS-I	0.9078 ± 0.0238	0.9708 ± 0.0218	0.8444 ± 0.0395	0.9027 ± 0.0255
	HCS-F (SL_RF,SL_XGBOOST, SL_SVM)	0,9104 ± 0.0249	0.9731 ± 0.0236	0.8473 ± 0.0436	0.9052 ± 0.0288
FakeNewsSet	Optimal Detective	**0,9808** ± 0.0264	**0.9955** ± 0.0144	**0.9679** ± 0.0451	**0.9810** ± 0.0254
	HCS-I	0.9179 ± 0.0397	0.9832 ± 0.0404	0.8542 ± 0.0852	0.9109 ± 0.0424
	HCS-F (SL_RF,SL_XGBOOST, SL_SVM, DMText, FNE)	0.9639 ± 0.0281	0.9831 ± 0.0274	0.9459 ± 0.0511	0.9631 ± 0.0250
FakeBr	Optimal Detective	0,9048 ± 0.0151	0.9014 ± 0.0209	0.9079 ± 0.0218	0.9045 ± 0.0167
	HCS-I	0.9987 ± 0.0028	**0.9986** ± 0.0045	0.9990 ± 0.0032	0.9988 ± 0.0026
	HCS-F (DMText, FNE)	**0.9989** ± 0.0026	0.9980 ± 0.0046	**1.0000** ± 0.0000	**0.9990** ± 0.0023

of users with measured reputation. It can be explained by the fact that once both datasets have a higher amount of news per user, there is a lower probability that a user, when partitioning the datasets for cross-validation, will have a small number of news published in the subset used to calculate the reputations (i.e., in N^L). Hence, in these datasets, fewer users would not have their reputations assessed. It is important to note that, similarly to what occurred with the other three datasets, in real situations (non-experimental), this lack of reputation for users can be caused by the *cold-start* problem, as described in Section 2.

Another aspect probably related to the *cold-start* problem is reflected by the better results of *HCS-F* when compared to *HCS-I* in most of the evaluation metrics. These values point to an improvement in the performance of *HCS* when implicit opinions of the disseminating users are complemented with machines' explicit opinions. We believe that, once *HCS-F* can count on the opinions and reputations of the machines, complementation provided by the machine's opinions may have contributed to reduce the impact of the opinion of users whose reputation is poorly or even not assessed.

Finally, in order to verify the existence of a statistically significant difference between the results of the *Optimal Detective*, *HCS-I* and *HCS-F* methods, we applied the *Wilcoxon Signed Ranks* test with significance level $\alpha = 0.05$ and null hypothesis H_0 stating that the results are statistically identical in the five datasets considered. The test between *HCS-I* and *Optimal Detective* (*W-value* = 3.5), as well as between *HCS-F* and *Optimal Detective* (*W-value* = 6) did not reject H_0, indicating no significant difference between the performance of the *HCS* and *Optimal Detective* methods. Such results confirm this work's hypothesis that the user's implicit opinion (inferred by his/her behavior facing the news) may enable the construction of *Fake News* detection models, with performance similar to that obtained by the state of the art of methods based on *Crowd Signals*, without depending on the explicit opinion of these users. In addition, the test between *HCS-I* and *HCS-F* produced *W-value* = 0 and, hence, rejected H_0, indicating the existence of a significant difference between the performances of the two *HCS* methods. This result is another indication of the improvement in performance of *HCS* by complementing the users' implicit opinions with the machines' explicit opinions.

4 Conclusion

One of the main approaches to detect Fake News automatically is the Crowd Signals based one. This approach combines opinions (signals) manifested by a high number of DMND users (crowd) in order to indicate whether a piece of news is fake or not. Although promising, this approach has an important limitation: it depends on the (not always available) user's explicit opinion about the news.

To overcome this difficulty, this work presented *HCS*, an approach based on crowd signals that considers implicit user opinions instead of the explicit ones to detect Fake News. *HCS* can also use the explicit opinions from machines (news classification models) to complement the implicit user opinions by means of hybrid Crowds.

A set of statistically supported experimental evidences that even without considering DMND users' explicit opinions, *HCS* may lead to results comparable to the ones produced by the state-of-the-art method of the Crowd Signals based approach. Besides that, the experiments also revealed a performance improvement of HCS when users' implicit opinions were combined with machines' explicit opinions.

Our initiatives for future works include experiments with other datasets, the investigation of different forms of inferring users' implicit opinions, and the application of a meta-classifier to combine machines' explicit opinions.

References

[1] Flávio R. M. da Silva, Paulo M. S. Freire, Marcelo P. de Souza, Gustavo de A. B. Plenamente, and Ronaldo R. Goldschmidt. 2020. FakeNewsSetGen: A Process to Build Datasets That Support Comparison Among Fake News Detection Methods *(WebMedia '20)*. 241–248.

[2] Paulo M. S. Freire, Flávio R. M. da Silva, and Ronaldo R. Goldschmidt. 2021. Fake news detection based on explicit and implicit signals of a hybrid crowd: An approach inspired in meta-learning. *Expert Systems with Applications* 183 (2021).

[3] Paulo M. S. Freire and Ronaldo R. Goldschmidt. 2019. Fake News Detection on Social Media via Implicit Crowd Signals *(WebMedia '19)*. 521–524.

[4] Paulo M. S. Freire and Ronaldo R. Goldschmidt. 2019. Uma Introdução ao Combate Automático às Fake News em Redes Sociais Virtuais *(34th SBBD)*. SBC, 38–67.

[5] Jürgen Habermas. 1982. *Teoria de la acción comunicativa: complementos y estudios prévios.* Madrid: Cátedra.

[6] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *SIGKDD Explor. Newsl.* 19, 1 (Sept. 2017), 22–36.

[7] Sebastian Tschiatschek, Adish Singla, Manuel G. Rodriguez, Arpit Merchant, and Andreas Krause. 2018. Fake News Detection in Social Networks via Crowd Signals *(WWW '18)*. 517–524.