Combate automático às Fake News nas mídias sociais virtuais: uma revisão do estado da arte

Paulo M. S. Freire, Ronaldo R. Goldschmidt*
Instituto Militar de Engenharia (IME)
Praça General Tibúrcio, 80, 22290-270, Praia Vermelha, Rio de Janeiro, RJ, Brasil.
***ronaldo.rgold@ime.eb.br**

RESUMO: Esta pesquisa mostrará o atual estado da arte dos estudos relacionados ao uso de ferramentas computacionais para o combate às Fake News em mídias sociais virtuais. O problema de combater as Fake News não é recente, mas sua complexidade aumentou devido ao uso de mídias sociais digitais. Com base na suposição de que o combate às Fake News é uma tarefa dividida em detecção e intervenção, este artigo fornecerá uma coleção sobre as áreas relacionadas ao combate às Fake News em mídias sociais virtuais, bem como ferramentas, conjuntos de dados e questões em aberto.

PALAVRAS-CHAVE: Fake News. Mídias sociais virtuais. Ferramentas computacionais.

ABSTRACT: This survey will show the current state-of-the-art of studies related to the usage of computational tools for Fake News combat on virtual social media. The problem of combating Fake News is not recent, but its complexity has increased due to the use of digital social media. Based on the assumption that Fake News combat is a task divided into detection and intervention, this paper will provide a collection about the areas pertaining to Fake News combat on virtual social media as well as tools, datasets and open issues.

KEYWORDS: Fake News. Virtual social media. Computacional tools.

1. Introdução

Historicamente, a divulgação de notícias era restrita às mídias tradicionais, tais como: rádio, tv e veículos de comunicação impressos. Atualmente, há uma tendência das pessoas divulgarem e consumirem mais notícias online do que aquelas disponibilizadas por mídias tradicionais [1]. Essa migração para as mídias sociais virtuais tem como uma das principais causas, a crescente facilidade de acesso a baixo custo. Por outro lado, essa maior acessibilidade possibilita a qualquer um, independentemente da credibilidade da fonte, divulgar notícias falsas com um intenso poder de espalhamento [2].

Dessa forma, a divulgação de notícias falsas, apesar de ser um problema antigo, teve a sua complexidade aumentada, de forma significativa, com o uso das mídias sociais virtuais [3]. Um segmento ainda mais preocupante abrange as *Fake News*, que são as notícias falsas publicadas de forma intencional [2], pois uma inverdade premeditada tende a ser mais bem elaborada e, consequentemente, mais eficaz no seu principal objetivo que é a mudança de opinião. Essa ploliferação de notícias intencionalmente falsas, normalmente, não atinge somente a integridade jornalística, mas também causam perturbações em áreas sociais, políticas, econômicas, culturais, da saúde e da segurança [4,5, 41].

Com base no exposto, surge um importante desafio que é combater a propagação de *Fake News* em mídias sociais virtuais [6,7,8,9,10], pois não basta verificar se a notícia é falsa, mas também é preciso determinar a intenção do divulgador em ludibriar os receptores [2,11]. Esse combate apresenta-se como não trivial, tanto pelo volume e velocidade das publicações, quanto pela notória dificuldade do homem em avaliar a veracidade das notícias [12,13,14]. Assim, o emprego de ferramentas computacionais, devido a sua maior velocidade de atuação, vem se destacando no combate às *Fake News* nas mídias sociais virtuais [13].

Apesar de ser um assunto relevante, até onde foi

possível observar, apenas duas pesquisas [2,3] fizeram um estudo sobre o estado da arte no combate às *Fake News* em mídias sociais virtuais. Como essa área é relativamente recente, essa pesquisa busca considerar alguns aspectos importantes não abordados pelos referidos trabalhos anteriores.

2. Caracterização de *Fake News*

A utilização do termo *Fake News* é relativamente recente, por isso surgem algumas divergências relativas ao seu significado. As *Fake News* são publicações em que a falsidade intensional é verificada [2,3,12,15]. Assim, o aspecto proposital é fundamental, haja vista que para uma determinada notícia *n* ser rotulada como *Fake News* há a necessidade de que *n* seja intencionalmente falsa.

Para enfatizar a diferença entre uma notícia falsa e uma intencionalmente falsa, pode-se utilizar dois termos denominados *misinformation* e *disinformation* [16,10]. A *misinformation* corresponde às notícias falsas publicadas pela falta da informação verdadeira, enquanto que a disinformation diz respeito às notícias propositalmente falsas, denominadas de *Fake News* [6].

Apesar da originalidade da expressão, as *Fake News* não surgiram com o uso das mídias sociais virtuais, pois, mesmo com as mídias tradicionais, já existiam pessoas ou Instituições procurando, de forma proposital, divulgar notícias falsas. Basicamente, essa divulgação pode ter o objetivo de ridicularizar e/ou enganar os receptores [16].

Independente do objetivo, a recente proliferação de notícias falsas e mal-intencionadas tem sido uma fonte de preocupação generalizada. Essa apreensão se deve pela constatação do poder de influência das *Fake News* na sociedade [8].

Somente nos Estados Unidos da América (EUA), mais de sessenta e dois porcento dos adultos recorrem às mídias sociais virtuais para receberem notícias. Como consequência desse elevado percentual, pode-se destacar o caso ocorrido nos três meses finais das eleições presidenciais americanas de 2016. Nessa ocasião as notícias falsas publicadas no Facebook, que favoreceram qualquer um dos dois candidatos, foram compartilhadas 37 milhões de vezes [15].

Inclusive casos relacionados às *Fake News* não se limitam aos EUA, pois, em 2018 na Índia, o *WhatsApp*, após notícias falsas terem, supostamente, levado a linchamentos, anunciou um limitador para a quantidade de encaminhamentos de mensagem[1].

Como um último e atual exemplo do poder de influência das *Fake News*, pode-se destacar o caso da pandemia de COVID-19 (*Coronavirus Disease*-19), causada pelo patógeno SARS-Cov-2 e que já matou milhares de pessoas no ano de 2020 ao redor do mundo[2], inúmeras *Fake News* têm sido divulgadas em mídias sociais virtuais [41]. Essas divulgações têm dificultado, de forma significativa, o esclarecimento da população sobre a disseminação da pandemia e sobre as devidas medidas de enfrentamento da doença a serem adotadas a cada instante.

O poder de influência das *Fake News* na sociedade se potencializa por fatores inerentes ao ser humano [2], dentre eles podemos destacar que as pessoas:

- preferem receber informações que confirmem as suas opiniões sem, necessariamente, verificarem a veracidade da notícia;
- tendem a aceitar as informações não pela análise da verdade, mas pela relação de ganhos e perdas que a notícia vai trazer para elas;
- tendem a avaliar as informações não pela busca da veracidade, pois acabam acompanhando a aceitação dos outros.

Além da influência das *Fake News* ser potencializada pelas características humanas, outra razão que as fortalece nas mídias sociais virtuais é a facilidade dos usuários, também denominados de agentes, publicarem e/ou propagarem as notícias [3]. Um aspecto importante inerente a essa facilidade é a criação de contas digitais maliciosas por meio de

[1] BBC News - https://www.bbc.com
[2] www.who.int/emergencies/diseases/novel-coronavirus-2019

agentes mal-intencionados de natureza humana e/ou computacional [2]. Esses agentes subdividem-se em:

- Bot que são robôs responsáveis por publicar e/ou propagar informações intencionalmente falsas;
- Trolls que são humanos com os mesmos objetivos dos Bots;
- Cyborgs já são mecanismos híbridos (humano/bot) que buscam disparar *Fake News*.

Existem ainda dois fatores importantes para incentivar essa disseminação das *Fake News*. O primeiro está relacionado à falta de legislação punitiva, devido à alegação de que as referidas leis poderiam cercear a liberdade de expressão. O segundo fator é o potencial ganho financeiro com a propagação da notícia [6].

Com base na facilidade de se publicar e propagar as notícias intencionalmente falsas nas mídias sociais virtuais, uma das principais formas de criação e disseminação de *Fake News* é se infiltrar em uma comunidade de pessoas engajadas em discutir um determinado assunto. Segundo MUSTAFARA [5], devem ser realizados os seguintes passos: Criar um domínio falso (*website*), criar contas anônimas, identificar comunidades e agentes interessados em um determinado assunto, contaminar esses agentes com a notícia falsa e incentivar a discussão para que as *Fake News* sejam espalhadas.

A partir da premissa de que a notícia falsa seja publicada e, em seguida, propagada, é importante caracterizar que uma notícia falsa pode ser criada no momento da publicação e, consequentemente, ser potencializada pela sua disseminação. Entretanto, uma notícia não *fake* pode ser publicada e se tornar *fake*, a partir do seu espalhamento, de acordo com as contribuições intencionalmente falsas feitas durante a sua propagação.

Com base na caracterização explicitada, o combate às *Fake News* pode se subdividir em detectar a notícia intencionalmente falsa e, posteriormente, intervir sobre a mesma. Essa intervenção objetiva mitigar os efeitos nocivos causados pela notícia na mídia social virtual.

3. Áreas relacionadas com *Fake News*

Como visto anteriormente, a utilização das mídias sociais virtuais não tem somente vantagens. Assim sendo, existem segmentos que atuam no combate aos problemas provenientes do uso dessas mídias. Portanto, esses serviços se tornam campos importantes de estudo devido ao seu relacionamento com o combate às *Fake News*. Algumas dessas áreas se encontram listadas a seguir, onde se procurou ordenar, de forma decrescente, essa lista de acordo com a similaridade das áreas com a detecção de *Fake News*, embora não exista consenso a respeito:

- *Fact Checking* (Checagem de fatos) - são *websites* ou *frameworks* responsáveis pela verificação, normalmente realizada com a ajuda de especialistas, da veracidade de fatos divulgados em mídias sociais virtuais [23,13,24]. A verificação da verdade dos fatos pode ser utilizada na tarefa de detecção de *Fake News*, assim como na criação de datasets;
- *Reputation and Trust System* (Sistemas de Reputação e Confiança) - são sistemas que buscam determinar o nível de confiança em mídias sociais virtuais baseados na obtenção de graus de reputação [25,26]. Com base na confiança, podem ser determinados riscos e recomendações, conforme ilustra a **figura 1**. A determinação de níveis de reputação e confiança, consequentemente riscos e recomendações, pode ser utilizada na tarefa de identificação de *Fake News*;
- *Rumor Classification* (Classificação de Rumores) - Rumor é uma informação em circulação cuja veracidade não foi verificada no momento da publicação. Um rumor pode ser classificado como verdadeiro, falso ou ainda não verificado [2,17,1,18]. Um rumor identificado como falso, após a sua publicação, é caracterizado como *Fake News*, caso haja intenção, conforme ilustra a **figura 2**. A tarefa mais relacionada com o combate às *Fake News* é a classificação de veracidade dos rumores;
- *Bot Detection* (Detecção de Bots) – procura identificar o envio automático de informações na mídia social por meio de robôs [20]. Esses envios podem acabar causando a propagação das *Fake News* [2,21,22,9];

- *Truth Discovery* (Descoberta da Verdade) - é a descoberta da verdade de fatos conflitantes entre diferentes fontes [2,19]. O combate às *Fake News* pode se beneficiar da Descoberta da Verdade para determinar a veracidade das afirmações, sendo, posteriormente, caracterizadas como *Fake News* as afirmações propositalmente falsas, conforme ilustra a **figura 2**;
- Clickbait Detection (Detecção de Iscas de Cliques) – procura identificar, nas páginas *WEB*, as chamadas iscas de cliques que, praticamente, forçam o agente a selecionar a opção apresentada. Nesse caso, o corpo do texto (bodytext) do artigo é, frequentemente, pobre e essa discrepância tem sido usada para identificar a inconsistência entre as linhas do cabeçalho (headlines) e o conteúdo da notícia, em uma tentativa de detectar *Fake News*. Sendo assim, o Clickbait pode ser usado como um indicador de *Fake News* [2].

Fig. 1 – Relação entre reputação e confiança.

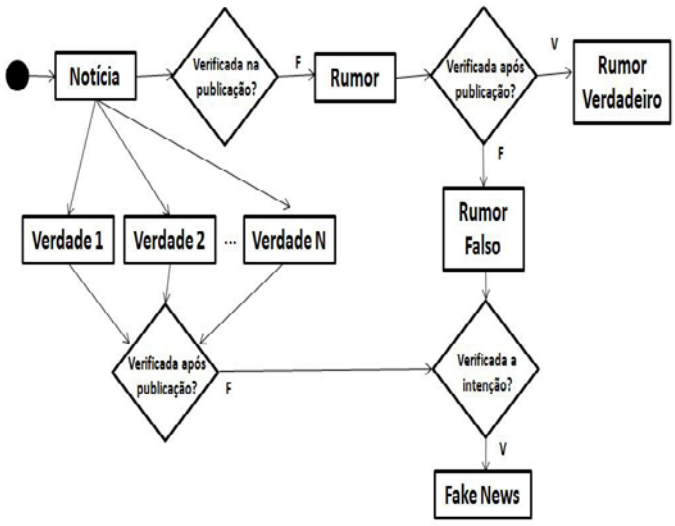

Fig. 2 – Relação entre descoberta da verdade, rumor, e *Fake News*.

4. Trabalhos relacionados ao combate automático às *Fake News*

Nesta seção são apresentados alguns trabalhos que desenvolveram ferramentas de combate automático às *Fake News*. Cabe destacar que a maioria desses trabalhos se enquadram na área de aprendizado de máquinas (AM) ao utilizarem os métodos tradicionais da AM [39,40]. Os referidos trabalhos são brevemente descritos, podendo seus detalhes serem consultados nas respectivas referências:

- *Automatic Detection of Fake News* [27]: Esse trabalho cria uma ferramenta de detecção de *Fake News* por classificação com *Support Vector Machines* (SVM), combinando informações léxicas, sintáticas, semânticas e de legibilidade. Também compara os resultados com a detecção humana;
- *Automatic Detection of Fake News on Social Media Platforms* [12]: Esse artigo implementa a detecção com os classificadores Binários *Logistic Regression*, *Support Vector Machines* (SVM), *Decision Tree*, *Random Forest* e *Extreme Gradient Boosting*. O referido trabalho compara os resultados entre os classificadores;
- *Automatically Identifying Fake News in Popular Twitter Threads* [28]: O trabalho apresenta um método para detecção de *Fake News* no *Twitter* que aculuma, ao longo do tempo, as características de rede, agente e conteúdo da notícia com análise léxica e sintática para gerar uma regressão linear. Assim, a abordagem realiza a sua análise, levando em consideração os aspectos temporais relacionados à notícia;
- *Combining Neural, Statistical and External Features for Fake News Stance Identification* [29]: Nesse estudo a ferramenta, desenvolvida para o primeiro desafio (FNC-1) [3], não tem o objetivo de detectar se a notícia é *Fake News*. Nessa abordagem, as notícias são classificadas de acordo com a relação existente entre a manchete e o corpo do texto. A ferramenta combina as abordagens neural e estatística com recursos externos. Para isto, a solução implementa um modelo profundo

[3] http://www.fakenewschallenge.org/

recorrente *Neural Embedding*, um modelo ponderado de características estatísticas *n-gram bag-of-words* e recursos externos criados à mão com a ajuda de uma heurística de engenharia de atributos. Por fim, usando uma rede neural profunda, todas as referidas abordagens são combinadas. Os resultados foram comparados com as demais ferramentas participantes do referido desafio;

- CSI: A *Hybrid Deep Model for Fake News Detection* [13]: O trabalho procura melhorar a acurácia na detecção de *Fake News* por meio de um modelo híbrido de rede neural profunda chamado CSI que utiliza três características: o texto da notícia, a resposta do agente que recebeu a notícia e o agente fonte da notícia. Esse modelo trabalha com o comportamento temporal de agentes e artigo, como também o comportamento em grupo dos agentes propagadores. Esse modelo se divide em três partes: *Capture, Score e Integrate*. O primeiro módulo é baseado na resposta e no texto, por meio de uma rede neural recorrente (LSTM) para capturar um padrão temporal de atividades do agente sobre o artigo e a representação Doc2Vec do texto gerado nessa atividade. O segundo usa uma rede neural para aprender as características da fonte, baseado no comportamento dos agentes de acordo com suas interações, gerando um score por meio de um grafo. Os dois módulos são integrados com o terceiro para caracterizar ou não o artigo como *Fake News*. O trabalho pode ser usado em diferentes domínios, inclusive em bancos de dados. Os resultados foram comparados com técnicas criadas para detecção de rumores;
- Data *mining applied in fake news classification through textual patterns* [37]: Esse artigo propõe uma detecção de *Fake News* a partir da análise do texto da notícia. Portanto, essa pesquisa classifica uma notícia através de uma análise gramatical, léxica e de polaridade do texto. Para tanto, foram utilizados o léxico Sentilex (polaridade), a biblioteca *Spacy* (gramatical), assim como, os classificadores *Naive Bayes*, SVM e *AdaBoost*;
- *DistrustRank: Spotting False News Domains* [30]: Essa solução propõe uma estratégia de aprendizagem semi-supervisionada para separar automaticamente notícias falsas a partir de fontes não confiáveis de notícias. O trabalho utiliza como fonte experts de portais de checagem de fatos para classificar manualmente as notícias. A partir disso é criado um grafo de pesos com os ranks de confiança sobre os sites e as arestas representam sua similaridade. A pesquisa computa a centralidade, utilizando o *PageRank*, em busca de uma similaridade entre os sites não confiáveis. O resultado da análise é a classificação em *Trust* ou *Distrust* para a fonte da notícia. Essa abordagem comparou os seus resultados, executando outras técnicas sobre o mesmo dataset;
- *Evaluating Machine Learning Algorithms for Fake News Detection* [31]: Esse artigo explora técnicas de linguagem natural para a detecção de *Fake News*. O trabalho aplicou term *frequency-inverse document frequency* (TF-IDF) de bi-grams e probabilistic context free grammar (PCFG) para um conjunto de 11.000 artigos em um dataset obtido pela *Signal Media* e uma lista de fontes da *OpenSources.com*. Esse dataset foi testado com os algoritmos de classificação *Support Vector Machines, Stochastic Gradient Descent, Gradient Boosting, Bounded Decision Trees* e *Random Forests*. Os modelos com melhor desempenho foram os *Stochastic Gradient Descent*, treinados apenas no conjunto de recursos do TF-IDF;
- *Exploiting Tri-Relationship for Fake News Detection* (TriFN) [32]: Esse artigo explora, simultaneamente, as correlações da postura da notícia, o bias e engajamento do agente. Assim, é apresentado um Tri-Relacionamento para detecção de *Fake News* (TriFN). O trabalho propõe que, tanto informações partidárias, quanto níveis de confiança do agente nas mídias sociais virtuais podem trazer benefícios adicionais para prever notícias falsas. Além disso, os agentes tendem a formar relacionamentos com pessoas afins que podem aumentar o espalhamento das *Fake News*. Essa abordagem cita e compara os seus resultados com os trabalhos RST-SVM, LIWC e *Information Credibility*;
- *Fake News Detection in Social Networks via Crowd Signals* [33]: A ferramenta desenvolvida trabalha na detecção e, consequente, intervenção da *Fake News*. Essa solução possui um algoritmo, chamado de *Detective* que usa inferência Bayesiana para detectar *Fake News* a partir da opinião do usuário e da sua respectiva reputação em opinar. O objetivo é, com base nas opiniões dos usuários, detectar, de forma antecipada, uma notícia falsa e bloqueá-la. Os resultados foram comparados

com as soluções denominadas pelo artigo como Opt, Oracle, Fixed-CM e No-Learn;

- *Fake News Mitigation Via Point Process Based Intervention* [15]: Nesse artigo, o enfoque está na intervenção da Fake News. A proposta é intervir, mitigando a notícia falsa, fornecendo recompensas na forma de notícias verdadeiras para quem recebeu a *Fake News*. O modelo utilizado foi baseado em l*east-squares temporal difference learning* (LSTD). Um dos experimentos foi real, com a criação de cinco contas no Twitter;
- *Liar, Liar Pants on Fire*: A *New Benchmark Dataset for Fake News Detection* [4]: Essa abordagem cria uma técnica de detecção de *Fake News* híbrida, usando redes neurais convolucionais (CNNs) para analisar, não somente textos mas também os dados do agente. O artigo obteve os melhores resultados, comparados com os de outros três detectores implementados com *Logistic Regression Classifier* (LR), *Support Vector Machine Classifier* (SVM) e *bi-directional long short-term memory networks model* (Bi-LSTMs);
- *Ranking-based Method for News Stance Detection* [7]: Mais uma pesquisa relacionada ao primeiro desafio (FNC-1). A solução do artigo é criada a partir de uma rede neural *Multi-Layer Perceptron*. Os resultados foram comparados com as demais ferramentas participantes do referido desafio;
- *Real-time Detection of Content Polluters in Partially Observable Twitter Networks* [22]: Essa pesquisa procura encontrar um tipo específico de bots, chamados de poluidores de conteúdo (*content polluters*), para poder distinguir notícias verdadeiras de *Fake News*. Segundo o artigo, o estado da arte de detecção de bots é analisar padrões de comportamento, sentimento, difusão, textual e temporal da rede social. Dessa forma, os dados são clusterizados para que os agentes possam ser classificados como bots pela análise dos respectivos perfis e a frequência dos tweets. Os resultados do trabalho foram comparados com os obtidos por uma ferramenta citada pelo artigo, denominada *Truthy*;
- *Towards News Verification: Deception Detection* Methods *for News Discourse* [14]: O trabalho propõe a ferramenta RST-SVM que analisa a notícia para extrair o estilo por meio da combinação do *Rhetorical Structure Theory* (RST) e *Vector Space Modeling* (VSM) para Clusterização. A classificação da notícia em enganosa ou real é feita por meio do SVM. Os resultados obtidos não foram significativamente melhores do que a detecção humana;
- *Tracing Fake-News Footprints*: *Characterizing Social Media Messages by How They Propagate* [34]: Esse trabalho foca a detecção de *Fake News* modelando a propagação da notícia na rede social por meio de mineração de grafos em Florestas (difusão da informação). Segundo o artigo, classificar *Fake News* pelo conteúdo da notícia é muito difícil, pois os respectivos criadores já estão se preocupando em divulgar as *Fake News* com os mesmos padrões das notícias verdadeiras. Em contra partida, as notícias falsas tendem a ter as mesmas fontes, pessoas e sequências. O trabalho propõe a ferramenta paralelizável chamada *TraceMiner* que utiliza *Recurrent Neural Networks* (LSTM-RNNs), para classificar o caminho de propagação das mensagens do Twitter. O artigo comparou os seus resultados com técnicas de análise de conteúdo criadas, usando SVM e *XGBoost*.

5. *Datasets*

Para treinar e/ou avaliar a acurácia das técnicas para combate às *Fake News* em mídias sociais virtuais são utilizadas métricas obtidas, normalmente, a partir da aplicação de modelos em datasets.

Até onde foi possível observar, não existe um dataset considerado como *benchmark*, pois os poucos disponíveis não contemplam as diferentes características que podem ser utilizadas na tarefa de combate às *Fake News*. Alguns datasets públicos se encontram descritos abaixo:

- *BS Detector* [2]: Esse dataset é coletado de uma extensão de *browser* chamado BS Detector que foi desenvolvido para checagem da veracidade de notícias. Os rótulos existentes são "*Fake news*", "*Satire*", "*Extreme bias*", "*Conspiracy theory*", "*Rumor mill*", "*State news*", "*Junk science*", "*Hate group*" e "*Clickbait*";
- *BuzzFeedNews* (2016-10-*facebookfact-check*) [2]: Esse *dataset* compreende as notícias no *Facebook* de nove agências para a eleição presidencial americana de 2016. Os eventos e artigos ligados foram checados por jornalistas do *BuzzFeed*. Ele contêm 1.627 artigos rotulados como "*Mostly true*", "*Mixture of true and false*", "*Mostly false*" e

"*No factual content*";

- *BuzzFeedNews*(2016-10-*facebookfact-check* modificado) [12]: Conjunto de dados criado a partir do *BuzzFeedNews* (2016-10-*facebookfact-check*), contudo os artigos são rotulados com "*Fake*" e "*Non-Fake*";
- *Celebrity* [27]: Esse *dataset* fornece os dados da notícia para análise léxica, sintática e semântica. As notícias verdadeiras e falsas foram retiradas da *Web*, sendo relacionadas com assuntos de celebridades;
- *CREDBANK* [2]: Conjunto de dados criado a partir do cruzamento de várias fontes, com aproximadamente 60 milhões de tweets, que cobrem 96 dias, iniciados em outubro de 2015. Todos os tweets são relacionados com mais de 1.000 eventos de notícias. Cada evento foi avaliado, pela credibilidade, por 30 anotadores da *Amazon Mechanical Turk*. Os rótulos existentes são "[-2]*Certainly inaccurate*", "[-1]*Probably inaccurate*", "[0] *Uncertain* (*doubtful*)", "[+1] *Probably accurate*", "[+2] *Certainly accurate*)";
- *DataSet Emergent* [7,29]: Nesse repositório, as notícias são rotuladas como "*Agree*" (o texto do corpo concorda com a manchete), "*Disagree*" (o texto do corpo discorda da manchete), "*Discuss*" (o texto do corpo discute a mesma afirmação que o título, mas não toma uma posição) e "*Unrelated*" (o texto do corpo discute uma alegação diferente do título). Essa base faz parte do primeiro desafio (FNC-1) e foi criado a partir do dataset para detecção de rumor chamado *Emergent*;
- *DistrustRank Datasets* [30]: Foram desenvolvidos dois datasets. O primeiro, gerado com sites confiáveis, por meio do Similar*Web*[4], tem 502 domínios e 396.422 URLs de notícias. O segundo, obtido com sites não confiáveis, através do *Wikipedia's list of prominent Fake News*) [5], possui 47 dominios e 37.320 URLs de notícias. As URLs das notícias foram obtidas no *Internet Archive*;
- *Facebook* para Detective [33]: Fonte de dados que considera os círculos sociais do *Facebook*, consistindo de 4.039 agentes (nós) e 88.234 arestas. Essas informações foram geradas, usando um aplicativo do Facebook, para identificar círculos sociais;
- *Fake*.Br [38]: *Corpus* com notícias em português que foram rotuladas manualmente. Esse dataset contém 7.200 notícias, sendo 3.600 rotuladas como "true" e 3.600 rotuladas como "*fake*". Cada uma dessas notícias é composta pelo seu respectivo texto e os metadados relacionados aos seus dados textuais;
- *Fake News vs Satire* [16]: *DataSet* para diferenciar *Fake News* e Sátiras onde as notícias são codificadas manualmente. A base, oriunda de diversas fontes, é composta por 283 relatos rotulados como *Fake* news e 203 como Satirical. Esses relatos são compostos pelo título, texto e um link para cada artigo;
- *FakeNewsAMT* [27]: Esse repositório de dados fornece os dados da notícia para análise léxica, sintática e semântica. As notícias falsas foram criadas a partir de notícias reais coletadas de assuntos diversos;
- *FakeNewsNet* [2,32]: Essa base de dados fornece notícias rotuladas contendo características linguísticas, visuais e dos agentes da publicação /propagação, incluindo dados relacionados à rede;
- *Kaggle*: O conjunto de dados contém texto e metadados de 244 sites e representa 12.999 postagens no total. Os dados foram extraídos usando a API *webhose.io*. Cada site foi rotulado de acordo com o *BS Detector*, sendo que as fontes de dados sem rótulo foram categorizadas como "Bs";
- KV [35]: Nessa base as notícias têm sujeito, predicado e objeto. Cada notícia tem um rótulo que indica a probabilidade de ser verdadeira. A ferramenta, por meio de uma fusão de conhecimentos, cria um grafo relacionando sujeito com objeto para medir a quantidade de interações e gerar automaticamente o *dataset*;
- LIAR [4]: Essa base de dados é coletada de um *website* de checagem de fatos chamado *PolitiFact*. Ele inclui 12.836 notícias de vários contextos como entrevistas de rádio, televisão e discursos de campanha. Os dados foram rotulados manualmente como "*Pants-fire*", "*False*", "*Barely-true*", "*Half-true*", "*Mostly true*" e "*True*". Cabe salientar que os dados referentes ao agente se resumem ao nome do autor da postagem;
- RST-SVM *Dataset* [14]: Essa base de dados foi criada a partir de codificadores, usando notícias do *Bluff the Listener*. Esse repositório consiste de 144 notícias

4 https://www.similarweb.com/top-websites/category/News-and-media
5 https://en.wikipedia.org/wiki/List-of-fake-News-websites

selecionadas, aleatoriamente, de 2010 até 2014;
- *Signal Media para Evaluating Machine Learning Algorithms for Fake News Detection* [31]: *Dataset* rotulado com "*Fake*" ou "Não *fake*" criado a partir de uma base de notícias da *Signal Media* e uma lista do repositório de confiança de fontes *OpenSources.co*. O citado dataset contém 11.051 artigos, sendo 3.217 categorizados com falsos;
- Twitter e Sina *Weibo* para CSI [13]: *Dataset* criado com 2.811 artigos rotulados como "*Fake*" e 2.845 como "*True*". A citada base de dados foi obtida a partir do repositório, para detecção de rumores, gerado no artigo [36];
- Twitter para *Automatically Identifying Fake News* [28]: Base de dados que utilizou os datasets PHEME (rumor no Twitter), *CredBank* (credibilidade no Twitter) e *BuzzFeed News Fact-Checking Dataset* (Checagem de fatos no Facebook). Os três datasets precisaram ser alinhados com as mesmas características e rótulos;
- Twitter para *Content Polluters* [22]: Repositório de dados criado para detecção de bots. Esse dataset, obtido a partir do Twitter, foi rotulado manualmente como "Bot" ou "Não Bot";
- Twitter para *TraceMiner* [34]: Conjunto de dados gerado pela coleta de informações do Twitter com rotulação a partir do site de checagem de fatos Snopes. Nessa base, os rótulos atribuídos às notícias são "*Real news*" ou "*Fake news*".

6. Problemas em aberto

O combate automático às *Fake News* em mídias sociais virtuais é uma nova e emergente área de pesquisa onde se podem destacar alguns problemas em aberto:
- Carência de datasets que forneçam, de forma suficiente, os diferentes dados necessários para detectar e/ou intervir nas *Fake News* em mídias sociais virtuais;
- Trabalhos que levem em consideração aspectos temporais do ciclo de vida da *Fake News* e que, consequentemente, possam intervir mais rapidamente na sua propagação;
- Os trabalhos de detecção de *Fake News* normalmente se limitam a verificar a veracidade das notícias, ignorando o aspecto intencional;
- Extração de características a partir de imagem e/ou áudio, limitando-se assim as análises de mídia somente em texto;
- Métodos que abordem características baseadas na rede que representa a propagação da notícia na mídia social. Nesse caso, podem ser aplicadas técnicas baseadas em grafos;
- Abordagens que procurem agregar diferentes dados para a geração de pesos que podem ser usados, por exemplo, para identificação de reputação;
- Pesquisas que, em vez de realizarem uma classificação binária (*true* ou *false*) de *Fake News*, utilizem probabilidades e/ou pertinências na detecção. Essa linha de trabalho se baseia no fato de que, normalmente, uma notícia intencionalmente falsa é uma mistura de afirmações falsas e verdadeiras;
- Utilização de um comitê de classificadores para determinar se uma notícia é uma *Fake News*. Dessa forma, pode-se agregar diferentes técnicas de classificação durante a detecção;
- Utilização de modelos não supervisionados ou semi-supervisionados devido à carência de datasets rotulados que possuam variedade de dados;
- Estudo sobre o diferente comportamento da *Fake News* em diferentes comunidades (escolar, trabalho e etc) e/ou mídias sociais virtuais (*Weibo*, *WhatsApp* e etc). Isto se deve pela possível mudança de forma de atuação das notícias intencionalmente falsas de acordo com a mídia social utilizada;
- Classificar os agentes de *Fake News* com o objetivo de identificar o seu tipo (*trolls, bots e cyborgs*). Isto se deve pela possível alteração de comportamento das notícias propositalmente falsas de acordo com o tipo de agente;
- Trabalhos relacionados à intervenção de *Fake News*, tanto para bloqueio, quanto para mitigação. Haja vista que o combate às *Fake News* não se limita à detecção, sendo necessária, também, a intervenção sobre a mesma;
- Ferramentas de combate às *Fake News* que atuem em tempo real e/ou descentralizada na rede. Essa atuação se destaca, pois, quanto mais rápido e extensivo for o combate, menor serão os efeitos nocivos das referidas notícias;
- Abordagens que utilizem o assunto para a análise da notícia, pois assuntos relevantes, normalmente, motivam a criação de notícias intencionalmente falsas;
- Pesquisas que utilizem a reputação dos agentes, tanto

para detecção, quanto para intervenção da notícia propositalmente falsa. Nesse tipo de abordagem é considerado que agentes com baixa reputação sejam potenciais divulgadores de *Fake News*.

7. Considerações Finais

Com a crescente popularidade das mídias sociais virtuais, cada vez mais pessoas consomem notícias online, em vez dos tradicionais meios de comunicação. No entanto, as mídias sociais virtuais também são usadas para divulgar notícias intencionalmente falsas, as chamadas *Fake News*, que podem causar fortes impactos negativos. Nesse artigo é explorado o combate automático às *Fake News* em mídias sociais virtuais. Para tal, a literatura existente foi revisada objetivando, por meio de um levantamento do estado da arte, fornecer subsídios para pesquisas que busquem desenvolver ferramentas para o combate automático às *Fake News* em mídias sociais virtuais. Tendo como base essa revisão da literatura, dois aspectos significativos podem ser destacados. O primeiro é a carência de datasets, rotulados com *fake* e não *fake*, que disponibilizem não somente os dados da publicação, mas, também, as informações relacionadas à propagação das notícias na mídia social virtual. O segundo aspecto é que as ferramentas computacionais, voltadas para a detecção, vêm se adaptando de acordo com as mudanças nas características das *Fake News*. Uma dessas mudanças é a maior similaridade nas características de escrita, presentes no texto, entre as notícias *fake* e não *fake*. Portanto, as ferramentas que não utilizam somente o conteúdo da notícia na tarefa de detecção de *Fake News* têm se sobressaído. Nesse grupo particular de ferramentas, aquelas baseadas na reputação dos usuários das mídias sociais virtuais se apresentam como uma alternativa para detecção de notícias intencionalmente falsas.

Referências Bibliográficas

[1] VOSOUGHI, *S. et al*. (2017). *Rumor gauge: Prediction the veracity of rumors on twitter.* ACM *Transactions on Knowledge Discovery from Data*, 11(4):50:1–50:35.

[2] SHU, *K. et al*. (2017). *Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter*, 19(1):22–36.

[3] CONROY, N. J. et al. (2015). *Automatic deception detection: Methods for finding fake news. Association for Information Science and Technology*, 52:1–4.

[4] Wang, *W. Y*. (2017). "*liar, liar pants on fire*": *A* n*ew benchmark dataset for fake news detection*. *Association for Computational Linguistics, page*s 422–426.

[5] MUSTAFARAJ, *E. and Metaxas*, P. T. (2017). *The fake news spreading plague:was it preventable? In Web Science Conference, pages* 236–239.

[6] KSHETRI, *N. and Voas*, J. (2017). *The economics of "fake news*". IT *Professional*, 19.

[7] ZHANG, Q. et al. (2018). *Ranking-based method for news stance detection. In Companion Proceedings of the The Web Conference 2018*, WWW '18, pages 41–42, *Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee. Achtert*, *E. et al. Global Correlation Clustering Based on the Hough Transform. Statistical Analysis and Data Mining*. vol 1(3), pp. 111-127. 2008.

[8] FLINTHAM, M. *et al*. (2018). *Falling for fake news: Investigating the consumption of news via social media. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, *pages* 376:1–376:10, *New York*, NY, USA. ACM.

[9] WANG, P. *et al*. (2018). *Is this the era of misinformation yet: Combining social bots and fake news to deceive the masses. In Companion Proceedings of the The Web Conference* 2018, WWW '18, *pages* 1557–1561, *Republic and Canton of Geneva, Switzerland. InternationalWorldWideWeb Conferences Steering Committee*.

[10] CAMPAN, *A. and Cuzzocrea*, A. (2017). *Fighting fake news spread in online social networks: Actual trends and future research directions. 2017 IEEE International Conference on Big Data, Boston, MA, pages* 4453-4457.

[11] RUBIN, V. L. *et al*. (2015). *Deception detection for news: three types of fakes. Association for Information Science and Technology*, 52:83:1–83:4.

[12] JANZE, *C. and Risius*, M. (2017). *Automatic detection of fake news on social media platforms. In Pacific Asia Conference on Information Systems, pages* 2–16.

[13] RUCHANSKY, *N. et al*. (2017). *Csi: A hybrid deep model for fake news detection. In CIKM - ACM International Conference on Information and Knowledge Management, pages* 797–806.

[14] V. *L. Rubin et al. (2015). Towards news verification: Deception detection methods for news discourse*. HICSS2015.

[15] FARAJTABAR, M. *et al*. (2017). *Fake news mitigation via point process based intervention*. In ICML - *International Conference*

on Machine Learning, pages 1–18
[16]GOLBECK, J. *et al*. (2018). *Fake news vs satire*: A *dataset and analysis. In Proceedings of the 10th ACM Conference on Web Science, WebSci '18, pages* 17–21, *New Yo*rk, NY, USA. ACM.
[17]LIU, *Y. and Xu*, S. (2016). D*etecting rumors through modeling information propagation networks in a social media environment. IEEE Transactions on Computational Social Systems*, 3:46–62.
[18]MA, *J. et al*. (2015). *Detect rumors using time series of social context information on microblogging websites. In ACM International on Conference on Information and Knowledge Management, pages* 1751–1754.
[19]LI, *Y. et al*. (2015). *A survey on truth discovery. ACM SIGKDD Explorations Newsletter*, 17:1–16.
[20]BRAZ, *P. and Goldschmidt*, R. (2017). Um método para detecção de bots sociais baseado em redes neurais convolucionais aplicadas em mensagens textuais. In SBSeg 2017, pages 501–508. 10/11/2017. *Buntain*, C. *and Golbeck*, J. (2017).
[21]FERRARA, E. *et al*. (2016). *The rise of social bots. Commun*. ACM, 59(7):96–104.
[22]NASIM, M. et al. (2018). *Real-time detection of content polluters in partially observable twitter networks. In Companion Proceedings of the The Web Conference* 2018, WWW '18, *pages* 1331–1339, *Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee*.
[23]CIAMPAGLIA, G. L. *et al*. (2015). *Computational fact checking from knowledge networks. PLOS ON*E, 1:1–13.
[24]SETHI, R. J. (2017). *Crowdsourcing the verification of fake news and alternative facts. In ACM Conference on Hypertext and Social Media, pages* 315–316.
[25]VAVILIS, *S. et al*. (2014). *A reference model for reputation systems. Decision Support Systems*, 1:147–154.
[26]F.*Hendrikx*, K. B. and *Chard*, R. (2015). *Reputation systems: A survey and taxonomy. Journal of Parallel and Distributed Computing, pages* 184–197.
[27]V. *P´erez-Rosas et al*. (2018). *Automatic detection of fake news. In Proceedings of the 27th International Conference on Computational Linguistics*, pages 3391-3401.
[28]BUNTAIN, C. and *Golbeck*, J. (2017). *Automatically identifying fake news in popular twitter threads. In 2017 IEEE International Conference on Smart Cloud* (SmartCloud) pages 208–215.
[29]BHATT, *G. et al*. (2018). *Combining neural, statistical and external features for fake news stance identification. In Companion Proceedings of the The Web Conference* 2018, WWW '18, *pages* 1353–1357, *Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee*.
[30]WOLOSZYN, *V. and Nejdl, W*. (2018). *Distrustrank*: *Spotting false news domains. In Proceedings of the 10th* ACM *Conference on Web Science, WebSc*i '18, pages 221–228, New York, NY, USA. ACM.
[31]GILDA, S. (2017). *Evaluating machine learning algorithms for fake news detection*. I*n* 2017 IEEE 15th Student *Conference on Research and Development* (SCOReD), *page*s 110–115.
[32]K. Shu, *et al.* (2017). *Exploiting tri-relationship for fake news detection. ArXiv abs/1712.07709 (2017): n. pag*.
[33]TSCHIATSCHEK, *S. et al*. (2018). *Fake news detection in social networks via crowd signals. In Companion Proceedings of the The Web Conference* 2018, WWW '18, *pages* 517–524, *Republic and Canton of Geneva, Switzerland. InternationalWorldWideWeb Conferences Steering Committee*.
[34]WU, *L. and Liu*, H. (2018). *Tracing fake-news footprints: Characterizing social media messages by how they propagate. In Proceedings of the Eleventh ACM Internationa*l *Conference on Web Search and Data Mining*, WSDM '18, pages 637–645, New York, NY, USA. ACM.
[35]DONG, *X. et al*. (2014). *Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 601–610.
[36]MA, J. *et al*. (2016). *Detecting rumors from microblogs with recurrent neural networks. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16). AAAI Press, pages 3818–3824*
[37] MORAES, M. C. *et al*. (2019). *Data mining applied in fake news classification through textual patterns. In Proceedings of the 25th Brazillian Symposium on Multimedia and the Web (WebMedia '19). Association for Computing Machinery, New York*, NY, USA, 321–324.
[38]MONTEIRO, R. A. *et al*. (2018). *Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results: 13th International Conference*, PROPOR 2018, Canela, *Brazil*, *September* 24–26, 2018.
[39]FACELI, K. *et al* (2011). Inteligência Artificial: uma Abordagem de Aprendizado de Máquina. LTC. isbn = {9788521618805}.
[40]GOLDSCHMIDT, R. R. *et al* (2015). *Data Mining*. Conceitos, técnicas, algoritmos, orientações e aplicações. *Elsevier.* isbn = {9788535278224}.
[41]MEJOVA, Y. a*nd Kalimeri*, K. (2020). *Advertisers Jump on Coronavirus Bandwagon: Politics, News, and Business*. ArXiv. Vol 2003.00923.