

Extinguindo o 171: classificação e detecção de fake news através de redes neurais artificiais

Extinguishing the bullshit: classification and detection of fake news through artificial neural networks

RESUMO

Felipe Augusto Stark
stark@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Toledo, Paraná, Brasil

Daniel Cavalcanti Jeronymo
danielc@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Toledo, Paraná, Brasil

O avanço da tecnologia possibilitou que as notícias sejam transmitidas de forma extremamente rápida e eficiente, não somente em portais de jornalismo mas também nas redes sociais. O uso dessas redes sociais traz consigo um problema, a dificuldade em diferenciar o que de fato é verdade do que pode ser uma mentira, as fake news, que nada mais são do que notícias falsas a respeito de alguém ou algo. A fim de contribuir na filtragem destas notícias, o presente trabalho explora como podemos usar a computação e a inteligência artificial em prol da veracidade. Implementou-se um algoritmo de rede neural de múltiplas camadas com procedimentos de processamento de linguagem natural resultando em um modelo de predição com o intuito de analisar um conjunto de dados de notícias falsas e reais para que se estabeleça um padrão textual reconhecido nos dois tipos de notícias. Com este modelo de predição estabelecido temos uma aplicação de inteligência artificial capaz de realizar uma previsão a sobre a confiabilidade de textos diversos. Aplicando a implementação da rede neural nos textos de teste e comparando o resultado com a classificação verdadeira do dataset, obtemos uma precisão de 76,5% na detecção de produções textuais com informações falsas.

PALAVRAS-CHAVE: Fake News. Rede neural artificial. Aprendizagem profunda.

Recebido: 19 ago. 2019.

Aprovado: 01 out. 2019.

Direito autorial: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



ABSTRACT

The advancement of technology has made it possible for news to be transmitted quickly and efficiently, not only on journalism portals but also on social networks. Social networks brings with it a problem, a difficulty in differentiating what is true and what may be a lie, the fake news, which is more important than fake news about someone or some topic. In order to suppress the news, this paper explores how computation and artificial intelligence for the sake of truth. Through the Python programming language and the Tensorflow framework, a multilayer neural network algorithm with the ability to process natural data with a prediction model was implemented in order to study a false and real data set. a standard requirement written on both types of news. With this prediction model, you can have an artificial intelligence feature capable of predicting various information. Applying the implementation of the neural neural in the texts of the testing of comparison and result obtained with the real class dataset, we obtained a precision of 76.5% in the detection of textual productions with false and doubtful information.

KEYWORDS: Fake news. Artificial neural network. Deep learning.

INTRODUÇÃO

O advento da internet possibilitou um aceleramento na velocidade e um aumento no volume das informações transmitidas pelo mundo, com cada vez mais pessoas tendo acesso a esses dados. Esse fácil acesso aliado a este grande volume, ao lucro de anúncios, mídias sociais e a polarização política implicam na disseminação de informações falsas e de origens questionáveis, as ditas *fake news*.

As *fake news* consistem em uma forma de imprensa marrom que é a publicação de boatos em manchetes ou outros veículos de comunicação com o objetivo de enganar os leitores. Na maioria dos casos essas notícias possuem cunho sensacionalista e podem ser exageradas, para que chamem a atenção dos leitores aumentando o número de acessos. (HUNT, 2016).

Pensando nas dificuldades que esta prática proporciona a imprensa como um todo e a comunicação nas redes sociais, aliado a visão computacional de resolução de problemas e ao estado da arte do ferramental de inteligência artificial, este trabalho tem como objetivo propor uma abordagem computacional de identificação e classificação de *fake news* através redes neurais artificiais.

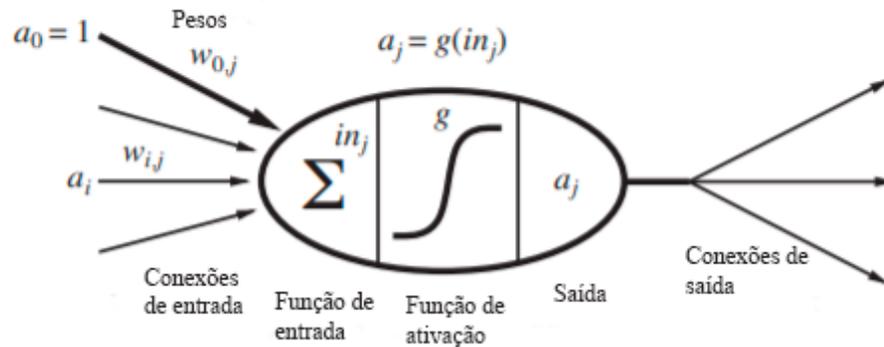
O restante deste trabalho está organizado da seguinte maneira. Na seção de material e método é descrita a metodologia de criação do sistema inteligente utilizando redes neurais artificiais em configuração profunda. Na seção de resultados são apresentados os resultados obtidos de classificação e identificação de *fake news*. Na seção de conclusão são apresentadas considerações finais e possibilidades de trabalhos futuros.

MATERIAL E MÉTODOS

Para atingir o objetivo de classificar os textos que possuímos precisamos entender como funciona uma rede neural construída com este intuito, além de entender alguns métodos de processamento de linguagem natural.

Uma rede neural artificial é um modelo computacional inspirado no sistema nervoso central. As redes neurais são apresentadas como conjuntos de neurônios, a estrutura básica de uma rede neural é um nó, chamado de *perceptron*, conectado diretamente a outros nós, onde cada ligação tem peso numérico associado que determina a força da conexão. Este *perceptron* recebe um dado de entrada, que passa pela função de ativação juntamente com o peso associado ao nó vai definir a saída de cada nó, dando continuidade à rede. (RUSSEL, NORVIG, 2010).

Figura 1 – Modelo matemático de um neurônio



Fonte: Russel, Norvig (2010, adaptado).

Uma rede neural antes de ser posta a teste passa por um processo de treinamento, onde ela recebe como entrada vários exemplos com saída conhecida do tipo de dado que deseja que a rede neural manipule, a medida que os dados são processados a rede atualiza os valores dos pesos de cada nó. Para determinar a taxa de aprendizado desta rede comparamos o valor de saída com o valor esperado dos dados, utiliza-se então uma métrica para avaliar este desempenho do algoritmo.

Para que a rede neural consiga processar o texto precisamos adequar este objeto para que a entrada de dados seja entendível pelo algoritmo, neste ponto aplicamos métodos de processamento de linguagem natural, para converter uma linguagem compreendida pelo ser humano em uma representação mais formal, facilmente manipulada pelo computador.

Terminado o treinamento da rede neural, aliado aos métodos de processamento das entradas de dados na rede, obtemos um modelo de predição, este modelo é o que efetivamente utiliza-se para tratar os casos desejados.

No caso deste trabalho a entrada da rede neural serão textos diversos, o tratamento desses dados para que sejam passados a rede consiste em converter cada palavra do texto para um número que se chama *index*, em seguida criar uma matriz para cada texto, onde a célula será preenchida com 1 caso a palavra esteja presente no texto e 0 caso não esteja.

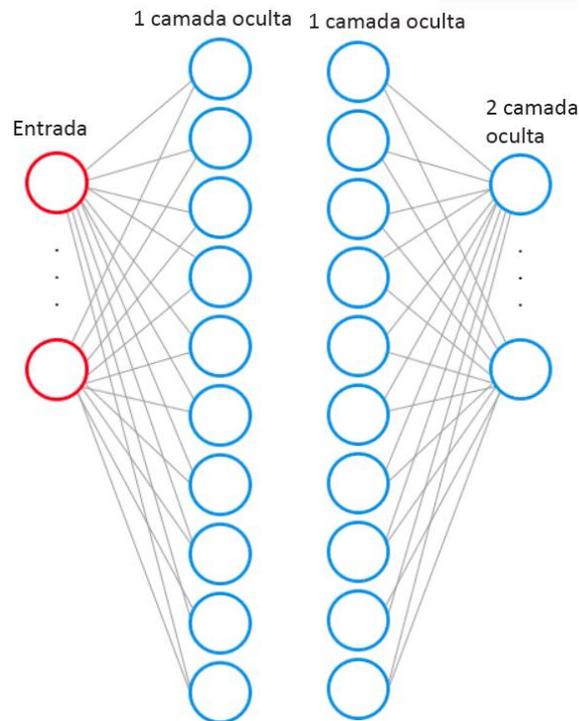
Figura 2 – Exemplo de texto processado

entrada: artigo do sicite
saída: [1. 1. 1.]

Fonte: Acervo do autor (2019).

A rede construída para este caso consiste em uma arquitetura de camadas ocultas, a função destas camadas é processar o dado de entrada de forma que a saída sirva como entrada para a próxima camada.

Figura 3 – Exemplificação das camadas ocultas



Fonte: Mesquita (2017, adaptado).

Esta rede foi implementada e testada com 2 camadas ocultas por 3 vezes, com 70, 90 e 100 neurônios na primeira camada e 110, 120, 150 na segunda camada, em cada teste respectivamente, utilizou-se um total de 15 ciclos de treinamento, com 4 mil textos para esta etapa.

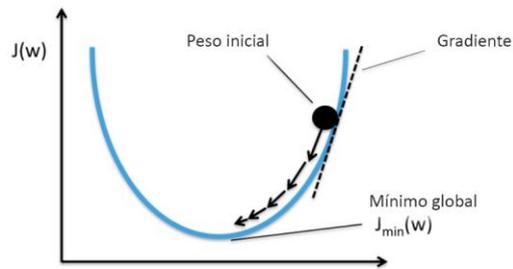
Ao executar o algoritmo pela primeira vez, os pesos foram definidos utilizando a distribuição normal, um método de distribuição probabilística utilizado para modelar fenômenos naturais matematicamente, conforme equação (1).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

O método de definição da taxa de aprendizado da rede neural foi o *cross entropy error*, que mede a performance de classificação com valores entre 0 e 1, este número decresce conforme o valor encontrado diverge do valor esperado. A taxa total de perda do desempenho do algoritmo foi calculada como a média de erros de cada unidade.

Como o objetivo após cada iteração é minimizar a perda de desempenho na classificação, utilizamos o método de gradiente descendente estocástico para calcular os novos pesos dos nós com base nas iterações anteriores. Este método consiste em obter os valores dos pesos por função da perda de desempenho através da minimização destes parâmetros, obtendo um valor ótimo para os pesos de modo que minimizem as perdas.

Figura 4 – Método do Gradiente Descendente Estocástico



Fonte: Raschka (Adaptado).

Após finalizar o cálculo dos novos pesos, o algoritmo atualiza os mesmos para dar início a outra etapa de treinamento. Ao fim de 15 ciclos deste treino, finalizamos o modelo de predição.

RESULTADOS E DISCUSSÃO

Com o modelo de predição em mãos é possível testar o algoritmo e avaliar sua precisão, comparando o resultado obtido pela rede com o resultado esperado para cada texto ao avaliar o *dataset* de teste com aproximadamente 2 mil textos.

A tabela a seguir mostra a precisão dos 3 modelos de predição com a quantidade de 70 neurônios na primeira camada oculta e 110 na segunda no primeiro teste, 90 neurônios na primeira camada e 120 na segunda para o segundo teste e 100 neurônios na primeira camada e 150 na segunda para o terceiro teste.

Tabela 1 – Resultados

Teste	Precisão (%)
Teste 1	76.5
Teste 2	71.4
Teste 3	68.7

Fonte: Acervo do autor (2019).

Observa-se que os valores do primeiro teste resultaram no modelo de predição mais preciso possível para este algoritmo, resultando em 76.5% de acurácia na classificação de textos de acordo com sua provável veracidade.

CONCLUSÃO

Apesar de a evolução da comunicação e da tecnologia trazerem consigo dificuldades e problemas, essa mesma evolução é a que possibilita buscar soluções para estes e outros obstáculos. Observamos ao exemplo deste trabalho que os métodos computacionais podem trabalhar a nosso favor e ao lado da imprensa e do jornalismo, buscando e analisando com rapidez o material que circula nos meios de comunicação para que retire de atividade parte das informações falsas que alguns veículos divulgam, principalmente no caso das redes sociais devido a seu fácil acesso.

REFERÊNCIAS

- Bottou, Léon. "SGD tricks." *Neural Networks: Tricks of the Trade*. Springer Berlin Heidelberg, 2012. 421-436
- HUNT, ELLE. What is fake news? How to spot it and what you can do to stop it. **The Guardian**. 2016. Disponível em : <<https://www.theguardian.com/media/2016/dec/18/what-is-fake-news-pizzagate>>. Acesso em: 15 ago. 2019.
- JARMUL, KATHARINE. Detecting Fake News with Scikit-Learn. **Datacamp**. 2017. Disponível em : < <https://www.datacamp.com/community/tutorials/scikit-learn-fake-news>>. Acesso em: 20 mai. 2019.
- MCCAFFREY, JAMES. Why You Should Use Cross-Entropy Error Instead Of Classification Error Or Mean Squared Error For Neural Network Classifier Training. 2013. Disponível em : < <https://jamesmccaffrey.wordpress.com/2013/11/05/why-you-should-use-cross-entropy-error-instead-of-classification-error-or-mean-squared-error-for-neural-network-classifier-training/>>. Acesso em: 20 jul. 2019.
- MESQUITA, DÉBORAH. Classificando textos com Redes Neurais e TensorFlow. **Medium**. 2017. Disponível em : <<https://medium.com/@dehmesquita/classificando-textos-com-redes-neurais-e-tensorflow-5063784a1b31>>. Acesso em: 20 mai. 2019.
- RASCHKA, SEBASTIAN. Fitting a model via closed-form equations vs. Gradient Descent vs Stochastic Gradient Descent vs Mini-Batch Learning. What is the difference?. Disponível em : < <https://sebastianraschka.com/faq/docs/closed-form-vs-gd.html>>. Acesso em: 15 ago. 2019.
- RUSSEL, S.; NORVIG, P.; **Artificial Intelligence: A Modern Approach**. Prentice-Hall, 3. ed, 2010. p. 727-730.