

# Detecção e Análise de Metáforas usadas em *Fake News* – resultados iniciais

Guilherme Pontes Pinto<sup>1</sup>, Leonardo de Assis da Silva<sup>1</sup>,  
Luiz Filipe Kluppel Cunha<sup>1</sup>, Luiz Gomes-Jr<sup>1</sup>

<sup>1</sup>DAINF – UTFPR – Curitiba – PR – Brasil

{guilherme.2015,leosil,luizcunha}@alunos.utfpr.edu.br, lcjunior@utfpr.edu.br

**Resumo.** Este artigo apresenta resultados preliminares em detecção automática de metáforas e análise do seu uso em textos do domínio jornalístico. Para a anotação das metáforas nas notícias, empregamos um modelo de rede neural recorrente bidirecional LSTM. O objetivo principal é analisar a existência de diferenças nas características de notícias falsas em relação à notícias confiáveis no uso de linguagem metafórica. Neste artigo apresentamos resultados iniciais da anotação e análise com foco nos títulos das notícias de um corpus de artigos em inglês.

## 1. Introdução

Notícias falsas, ou *fake news*, ganharam notoriedade nos últimos anos devido à influência em processos eleitorais de diversos países. Além da área política, notícias falsas são ainda frequentemente associadas a entendimentos equivocados em relação a doenças, vacinas, etc. Dada a gravidade dos problemas gerados a partir da disseminação de desinformação, amplia-se o interesse em técnicas para a detecção automática de notícias falsas de forma a analisar o enorme volume de notícias disponíveis online [Lazer et al. 2018].

Uma alternativa ainda pouca explorada diz respeito a análise das figuras de linguagem empregadas nas notícias, particularmente metáforas. Metáforas podem ser definidas como uma figura de linguagem que busca associar ideias através da comparação das mesmas, de forma a criar, na frase, um sentido não literal. De forma geral, uma metáfora pode ser definida como uma ideia explicada em termos de outro conceito, que possui certas características equivalentes com a primeira.

O uso de metáforas permeia a linguagem humana; sua aplicação pode variar do contexto de uma conversa casual ao raciocínio empregado na resolução de problemas complexos. Além disso, conforme estudos nos campos de psicologia e ciências cognitivas, o uso de metáforas pode influenciar a maneira como pessoas criam estruturas conceituais para a resolução de problemas abstratos e concretos – [H. Thibodeau and Boroditsky 2011] – isto é, características de experiências anteriores não apenas podem ser reaproveitadas em novas situações, como podem inserir um viés na maneira como novas informações são interpretadas.

Seguindo tais resultados, supomos que o padrão de uso de metáforas em notícias poderia fornecer um indicativo quanto a intencionalidade do autor em influenciar a maneira como os leitores recebem a informação. Conforme as hipóteses estabelecidas em [Horne and Adali 2017], autores tendem a utilizar dois tipos de persuasão de acordo com o tipo da notícia. Enquanto notícias confiáveis tendem convencer os leitores através de

argumentos lógicos mais extensos usando linguagem técnica, notícias falsas costumam recorrer a associações em textos curtos e repetitivos. A constatação de que notícias falsas empregam um maior número de metáforas, por exemplo, poderia indicar uma tentativa de associar conceitos negativos ao tópico alvo. Desta forma, notícias poderiam ser classificadas baseadas em seu padrão de uso de metáforas. Este artigo apresenta resultados iniciais da detecção de metáforas em notícias, focando inicialmente na análise dos títulos.

## 2. Trabalhos Correlatos

Embora metáforas, ao melhor de nosso conhecimento, ainda não tenham sido exploradas como aplicação no domínio de notícias falsas, diversos estudos têm pesquisado técnicas computacionais para a detecção automática de metáforas, seja através de conhecimento linguístico especializado ou modelos de redes neurais.

Particularmente, na competição descrita em [Wee Leong et al. 2018] foi comparado o desempenho de vários modelos computacionais capazes de detectar uso metafórico de linguagem à nível de palavra em textos extraídos do British National Corpus (BNC). Como *baseline* foram usados modelos baseados em *WordNet*, níveis de concretude/abstração de palavras e outras *features* adicionais. Todos os modelos que superaram os *baselines* exploraram *word embeddings*. Os sistemas de detecção de metáforas apresentaram diferença de desempenho dependendo da subcategoria textual analisada, o que indica a existência de uma disparidade no tipo e uso de linguagem metafórica de acordo com a categoria do texto, por exemplo, a frequência e tipos de metáforas exibidas em textos acadêmicos não necessariamente correspondem a padrões encontrados em notícias.

Um dos modelos que obtiveram melhor resultado, [Stemle and Onysko 2018], aplicou uma rede neural recorrente bidirecional, tipo *Long Short-Term Memory* (LSTM), seguindo a suposição de que a proficiência no uso de linguagem metafórica varia dependendo do grau de conhecimento do idioma. Dessa forma, o treinamento da rede neural responsável por aprender as *word embeddings* foi realizado através de corpus de diferentes níveis de proficiência na língua inglesa.

Em relação a modelos de regras que exploram conhecimentos especializados sobre propriedades linguísticas, como a violação de restrições seletivas, tais abordagens são limitadas a casos específicos, alcançando bons resultados somente em determinados domínios, ou metáforas compostas apenas por um pequeno subconjunto de classes gramaticais, ou idiomas que possuem características gramaticais similares. Como nosso trabalho explora uma arquitetura conexionista, isto é, seu desempenho está diretamente ligado à variáveis como topologia, parâmetros e dados de treinamento em vez de regras específicas, este poderia ser adaptado, por exemplo, para outros idiomas e contextos conforme os textos utilizados no treinamento. Ao melhor de nosso conhecimento, o presente estudo trata-se do primeiro a buscar aplicar um detector de metáforas para análise de uso e impactos em um domínio específico, no caso detecção de *fake news*.

## 3. Modelo

Redes Neurais Recorrentes (RNN) podem ser definidas como uma extensão das redes neurais convencionais *feedforward* tornadas em grafos cíclicos ao introduzir aresta recorrentes entre cada passo de tempo. Segundo [Lipton 2015], diferente das arquiteturas

*feedforward* onde o estado da rede é perdido a cada iteração, nas RNNs ocorre a transferência direta de influência de uma entrada para a(s) seguinte(s) de acordo com a força de sua ativação. A arquitetura LSTM trata-se de uma adaptação de RNN convencional na qual a camada escondida não representa mais um nó simples, mas sim uma ou mais células de memória, que mantêm um valor a longo prazo. O modelo LSTM bidirecional e parâmetros utilizado nesta pesquisa, retirados do trabalho [Stemle and Onysko 2018], foram implementados usando a biblioteca Keras<sup>1</sup> em Python 3.

O detector de metáforas LSTM é alimentado na camada de entrada A por uma sentença pré processada pela remoção das *stop words*. Cada palavra na sequência de texto é representada por um vetor contínuo de 100 *features*. Vetores contínuos, ou *word embeddings*, trata-se de uma forma de mapear palavras de um vocabulário para um espaço de dimensões reduzido em comum. Esta representação permite que o conhecimento sobre relações existentes entre as palavras, como similaridades semânticas e gramaticais, seja reutilizado em diferentes tarefas de NLP. No experimento foram utilizados vetores treinados através de rede neural fastText<sup>2</sup> no *dataset* BNC, contendo 100 milhões de tokens.

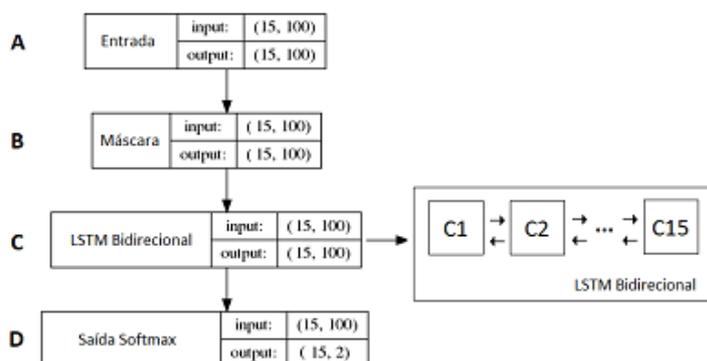


Figura 1. Topologia detector de metáforas.

Parâmetro	Valor
Tamanho da sequência	15
Features por palavra	100
Partições	3
Tamanho por batelada	32
Épocas de treinamento	20
Taxa de dropout	0.25
Word Embedding	BNC
Arquitetura	LSTM
Saída	Softmax

Figura 2. Parâmetros.

Na camada B sentenças menores que 15 palavras são completas por vetores nulos, enquanto sequências que ultrapassam o comprimento delimitado são separadas e analisadas separadamente. A camada seguinte é composta pelas células C1 à C15 do tipo LSTM, cada uma conectada ao vizinho anterior e sucessor, sendo responsável por processar a probabilidade da palavra ser uma metáfora dependendo de suas características capturadas pelo *word embeddings* e pelo estímulo recebido das palavras vizinhas.

#### 4. Experimento

Disponibilizado de forma aberta<sup>3</sup>, o *dataset* de notícias falsas é uma coleção de diferentes categorias de notícias extraídas de 745 sites de notícias (30GB de conteúdo). Além das classes confiável e falsa, o corpus inclui ainda categorias como sátira, *clickbait* e notícia de ódio. Para o treinamento foi utilizado o *dataset* padrão de metáforas da competição citada na Seção 2.

O experimento inicial realizado considerou somente os títulos das notícias, com o objetivo de verificar se estes seriam descritores suficientemente bons para a separação

<sup>1</sup><https://keras.io/>

<sup>2</sup><https://embeddings.sketchengine.co.uk/static/index.html>

<sup>3</sup><https://github.com/several27/FakeNewsCorpus/>

das classe de notícias analisadas. Os resultados foram obtidos ao analisar uma amostra aleatória de 40000 títulos (aproximadamente 4MB). A execução do modelo sobre a amostra demandou 45 minutos de processamento em um processador Intel Core i3 (2Ghz) e 8GB de memória RAM. Os títulos possuem em média 8.2 palavras e 0.7 palavras marcadas como de uso metafórico.

Mesmo ao usar uma amostra pequena do *dataset*, foram observadas certas tendências como a da classe rumores apresentar maior uso de metáforas (média 1.53, p-value < 0.01) quando comparada à classe de notícias confiáveis (média 0.65). Na comparação das notícias falsas (média 0.73) com as notícias confiáveis, a diferença foi menor mas ainda estatisticamente relevante (p-value < 0.01).

Esse resultado pode ser um indicativo de que títulos de notícias falsas tendem a simular características como estilo linguístico empregado em títulos de notícias confiáveis, com o objetivo de enganar leitores ao esconder o viés presente no corpo da notícia.

## 5. Considerações Finais

Neste trabalho, procuramos estudar a utilização de metáforas na composição de textos do domínio jornalístico através de um modelo computacional conexionista. Em relação às classes gramaticais de metáforas, decidimos restringir a análise preliminar para verbos e substantivos, já que estas possuem maior frequência nas metáforas do corpus de treinamento. No entanto, o uso de metáforas de classes gramaticais menos frequentes poderia melhorar a classificação de notícias específicas.

Nos próximos passos do trabalho esperamos que a execução das análises nos corpos das notícias possa trazer resultados mais significativos, como padrões de usos de metáforas e de suas classes gramaticais característicos para os diferentes tipos de notícias ou para diferentes fontes. Tais distinções abririam espaço para análises mais aprofundadas, como por exemplo a verificação da existência de tipos de notícias similares em relação ao uso de metáforas, detecção da intenção do autor de acordo com o tipo da metáfora e correlações entre uso de metáforas e análise de sentimento no texto.

## Referências

- H. Thibodeau, P. and Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. *PLoS ONE*, 6(2):11.
- Horne, B. D. and Adali, S. (2017). 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 *Eleventh International AAAI Conference on Web and Social Media*.
- Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., et al. (2018). The science of fake news. *Science*, 359(6380):1094–1096.
- Lipton, Z. C. (2015). A critical review of recurrent neural networks for sequence learning. *CoRR*, abs/1506.00019.
- Stemle, E. and Onysko, A. (2018). Using language learner data for metaphor detection. In *Proceedings of the Workshop on Figurative Language Processing*, pages 133–138.
- Wee Leong, C., Beigman Klebanov, B., and Shutova, E. (2018). A report on the 2018 via metaphor detection shared task. Technical report.