

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DAINF - DEPARTAMENTO ACADÊMICO DE INFORMÁTICA
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

GUILHERME PONTES PINTO
LEONARDO DE ASSIS DA SILVA
LUIZ FILIPE KLUPPEL CUNHA

**DETECÇÃO E ANÁLISE DE METÁFORAS EM *FAKE*
*NEWS***

TRABALHO DE CONCLUSÃO DE CURSO

CURITIBA
2019

GUILHERME PONTES PINTO
LEONARDO DE ASSIS DA SILVA
LUIZ FILIPE KLUPPEL CUNHA

**DETECÇÃO E ANÁLISE DE METÁFORAS EM *FAKE*
*NEWS***

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Sistemas de Informação da Universidade Tecnológica Federal do Paraná, como requisito parcial para a obtenção do título de Bacharel.

Orientador: Luiz Celso Gomes Jr.
DAINF - Departamento Acadêmico de Informática - UTFPR

Coorientadora: Josie Helen Siman
UNICAMP

CURITIBA
2019



Ministério da Educação
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
Câmpus Curitiba
Diretoria de Graduação e Educação Profissional
Departamento Acadêmico de Informática
Coordenação do Curso de Bacharelado em Sistemas de Informação



TERMO DE APROVAÇÃO

“DETECÇÃO E ANÁLISE DE METÁFORAS EM FAKE NEWS”

por

“GUILHERME PONTES PINTO, LEONARDO DE ASSIS DA SILVA E LUIZ FILIPE KLUPPEL CUNHA”

Este Trabalho de Conclusão de Curso foi apresentado no dia **02** de **JULHO** de **2019** como requisito parcial à obtenção do grau de Bacharel em Sistemas de Informação na Universidade Tecnológica Federal do Paraná - UTFPR - Câmpus Curitiba. O(a)s aluno(a)s foi(ram) arguido(a)s pelos membros da Banca de Avaliação abaixo assinados. Após deliberação a Banca de Avaliação considerou o trabalho

<hr/> Prof. Luiz Celso Gomes Jr. (Orientador - UTFPR/Curitiba)	<hr/> Prof. Thiago Henrique Silva (Avaliador 1 - UTFPR/Curitiba)
<hr/> Profa. Leyza Baldo Dorini (Avaliadora 2 - UTFPR/Curitiba)	<hr/> Profa. Leyza Baldo Dorini (Professora Responsável pelo TCC – UTFPR/Curitiba)
<hr/> Prof. Marcelo Mikosz Gonçalves (Coordenador do curso de Bacharelado em Sistemas de Informação – UTFPR/Curitiba)	

“A Folha de Aprovação assinada encontra-se na Coordenação do Curso.”

RESUMO

PINTO, Guilherme P.; DA SILVA, Leonardo de A.; CUNHA, Luiz F. K.; GOMES JR., Luiz C.; SIMAN, Josie H.. Detecção e análise de metáforas em *fake news*. 2019. 61 f. Trabalho de Conclusão de Curso – Curso de Bacharelado em Sistemas de Informação, Universidade Tecnológica Federal do Paraná. Curitiba, 2019.

A metáfora é uma figura de linguagem amplamente presente tanto nas conversas do dia a dia como nos textos formais. Além disso, acredita-se que as metáforas afetam a aquisição de novas informações ao formar uma nova estrutura conceitual para o alvo, semelhante às características da fonte com a qual ele está sendo comparado. Dado que as notícias falsas, um fenômeno que vem ganhando atenção acadêmica e social, são usadas para confundir e manipular leitores, procuramos investigar se as metáforas são um dos métodos utilizados para influenciar o conteúdo das reportagens. Para isso, realizamos a tarefa de detecção automática de metáforas através de um modelo de *Long Short-Term Memory* e analisamos as emoções associadas com as anotações resultantes para entidades específicas (Hillary Clinton e Donald Trump). Observamos que notícia falsa foi a classe com maior concentração de metáforas por palavras no corpo de seus artigos e que as emoções das metáforas usadas para descrever cada candidato diferem significativamente de acordo com o tipo de notícia. Acreditamos que o estudo de metáforas e das emoções que elas transmitem pode ser empregado em conjunto com os métodos atuais usados para detectar notícias falsas, a fim de melhorar o desempenho da classificação.

Palavras-chave: Detecção de Metáfora. Notícia Falsa. Análise de Emoção.

ABSTRACT

PINTO, Guilherme P.; DA SILVA, Leonardo de A.; CUNHA, Luiz F. K.; GOMES JR., Luiz C.; SIMAN, Josie H.. Metaphor Detection and Analysis on fake news. 2019. 61 f. Trabalho de Conclusão de Curso – Curso de Bacharelado em Sistemas de Informação, Universidade Tecnológica Federal do Paraná. Curitiba, 2019.

Metaphor is a figure of speech largely present both in day-to-day conversations and formal texts. Furthermore, it is believed that metaphors affect the acquisition of new information by forming a new conceptual structure for the target, similar to that of the source characteristics it is being compared to. Given that fake news, a phenomenon that has gained academic and social attention, is used to confuse and manipulate readers, we seek to investigate whether metaphors are one of the methods used to bias the content of news articles. For that purpose, we perform the task of automatically detecting metaphors through a Long Short-Term Memory model and analyze the emotions associated with the resulting annotations for specific entities (i.e., Hillary Clinton and Donald Trump). We have observed that fake news present the highest concentration of metaphors by words in the body of its articles and that the emotions of the metaphor used to describe each candidate differ according to the type of news. We believe that studying the metaphors and the emotions they convey could be used in addition to the current methods being employed to detect fake news in order to improve the classification performance.

Keywords: Metaphor Detection. Fake News. Emotion Analysis

LISTA DE FIGURAS

Figura 1 – Neurônio artificial.	22
Figura 2 – Rede Neural Feedforward completamente conectada de 1 camada escondida.	23
Figura 3 – Representação detalhada de uma RNN.	25
Figura 4 – Decomposição de Valor Único.	28
Figura 5 – Topologia botzen.	40
Figura 6 – Distribuição por tipo de notícia.	47
Figura 7 – Médias de metáforas por palavras.	49
Figura 8 – Média de metáforas por tipo de notícia para Trump e Hillary.	49
Figura 9 – Variação da medida F1 conforme o grau de certeza.	50
Figura 10 – Variação da medida F1 para as classes <i>fake</i> e <i>reliable</i> conforme o grau de certeza.	51
Figura 11 – Média de subjetividade para títulos de notícias, por tipo de notícias.	52
Figura 12 – Distribuição de emoções de notícias confiáveis e falsas.	53
Figura 13 – Distribuição de emoções para entidades Trump e Hillary.	53

LISTA DE TABELAS

Tabela 1 – Campos do <i>data set</i> Metaphor Corpus	36
Tabela 2 – Campos do <i>data set</i> Fake News	37
Tabela 3 – Subconjunto de campos do <i>data set</i> Fake News a serem explorados . .	37
Tabela 4 – Parâmetros botzen.	39
Tabela 5 – Resultados da Anotação de Entidades.	48

LISTA DE ACRÔNIMOS

ANN	Redes Neurais Artificiais
BPTT	<i>Back-propagation through time</i>
FFNN	Redes Neurais Feedforward
GRU	<i>Gated Recurrent Unit</i>
LSTM	<i>Long-Short Term Memory</i>
MIP	<i>Metaphor Identification Procedure</i>
MIPVU	<i>Metaphor Identification Procedure VU University Amsterdam</i>
NAACL	<i>North American Chapter of the Association for Computational Linguistics</i>
NER	Reconhecimento de Entidade Nomeada
NLP	Processamento de Linguagem Natural
RNN	Redes Neurais Recorrentes
TBPTT	<i>Truncated Back-propagation through time</i>
UNICAMP	Universidade Estadual de Campinas
UTFPR	Universidade Tecnológica Federal do Paraná
VUA	<i>Vrije Universiteit Amsterdam</i>

LISTA DE SÍMBOLOS

ϕ	Função de ativação
a_k	Somatório de entradas do neurônio k
b	<i>Bias</i> , viés de deslocamento do neurônio
D	Matriz de documentos
$h^{(t)}$	Estado da rede no passo de tempo t
M	Matriz tf-idf
o	Saída da rede no estado h antes da função de ativação
p_k	Vetor da palavra na posição k do Vocabulário
R	Matriz de espaço dimensional
T	Matriz de termos
U_k	Vetor de pesos das entradas do neurônio k
u_{ki}	Peso da aresta da entrada i do neurônio k
Voc	Vocabulário
V	Peso do estado da rede na saída
W	Peso da aresta recorrente
X	Vetor de entrada do neurônio
x_i	Entrada i de um neurônio
\hat{y}_k	Saída do neurônio k
y	Saída esperada

SUMÁRIO

1 – Introdução	12
1.1 Justificativa	14
1.2 Objetivo Geral	15
1.3 Objetivos Específicos	15
1.4 Estrutura do texto	15
2 – Referencial Teórico e Estado da Arte	17
2.1 Referencial Teórico	17
2.1.1 Linguagem Figurada	17
2.1.2 Metáfora	18
2.1.3 Redes Neurais Feedforward	21
2.1.4 Redes Neurais Recorrentes	23
2.1.5 Explosão e Desaparecimento de Gradiente	26
2.1.6 Redes Long Short-Term Memory	26
2.1.7 Word Embeddings	27
2.1.7.1 Representação One-hot	27
2.1.7.2 Representação com Análise de Semântica Latente	28
2.1.7.3 Representação com Redes Neurais	29
2.1.8 Reconhecimento de Entidades Nomeadas	30
2.1.9 Fake News	31
2.2 Estado da Arte	32
2.2.1 Análise Crítica	34
3 – Metodologia	35
3.1 Dados	35
3.1.1 Dados de Treinamento	35
3.1.2 Fake News Corpus	36
3.2 Workflow - Fluxo de trabalho	37
3.2.1 Reconhecimento de Entidades Nomeadas	38
3.2.2 Detector de Metáforas	39
3.2.3 Anotação de Emoções	41
3.3 Perguntas da Pesquisa	41
3.4 Casos de Análise	42
3.4.1 Cenário 1	42
3.4.2 Cenário 2	42
3.4.3 Cenário 3	42

3.5	Método de Avaliação	42
4	Recursos de Hardware e Software	44
4.1	Recursos de Hardware	44
4.2	Recursos de Software	44
4.2.1	Keras	44
4.2.2	Gensim - FastText	44
4.2.3	Sklearn - Metrics	45
4.2.4	NLTK e TextBlob	45
4.2.5	NumPy	45
4.2.6	DBpedia Spotlight	45
4.2.7	NRC Lexicon	46
5	Análise de Resultados	47
5.1	Análise de Entidades	47
5.2	Análise de Metáforas	48
5.2.1	Avaliação do Detector de Metáforas	50
5.3	Análise de Emoções	51
5.4	Análise de Metáforas Emotivas direcionadas à Entidades	53
6	Considerações Finais	55
6.1	Trabalhos Futuros	56
	Referências	57

1 Introdução

Processamento de Linguagem Natural (NLP) é uma área da Ciência da Computação que tem por objetivo estudar como computadores podem manipular ou processar linguagens naturais e dados não estruturados relacionados. Mesmo não sendo um campo de estudo criado recentemente, o processamento de linguagem natural tem ganhado mais espaço e atenção, devido à quantidade massiva de dados disponibilizados, sobretudo na internet. O aumento da disponibilidade de dados, também provocou um grande crescimento na quantidade de dados não estruturados disponíveis e, conseqüentemente, a lacuna existente entre os facilmente processáveis por computadores, e aqueles que não podem ser processados sem o conhecimento dessa área. Dessa forma, há um grande e crescente conjunto de problemas de importante aplicação que não seriam resolvidos sem os avanços dessa área de estudo (BELL; HEY; SZALAY, 2009). Dentre as principais tarefas atribuídas à *NLP*, pode-se citar aquelas ligadas mais diretamente ao conhecimento especializado linguístico que explora regras de sintaxe - buscando entender palavras e suas relações com as frases nas quais estão inseridas - e semântica - buscando entender o significado atrelado às entidades presentes nas frases, e aquelas ligadas ao uso de propriedades estatísticas referentes a corpus e ao documento (LIDDY, 2001).

Com o constante avanço das áreas ligadas à Inteligência Artificial, continuamente buscam-se modelos que permitam capturar efetivamente a linguagem falada por humanos, seja com intuito de gerá-la artificialmente, para melhor interpretar as necessidades de usuários que buscam informações, ou mesmo para permitir o completo entendimento e interpretação da linguagem falada (CHOWDHURY, 2003). As tarefas acima citadas são constantes alvos de pesquisa, visto que a resolução destas implicaria em avanços nos mais variados domínios. Embora grandes progressos tenham sido realizados nesta área, tais tarefas ainda não foram resolvidas. Ainda não há uma forma canônica de se interpretar ironia ou sarcasmo, por exemplo; o mesmo acontece para um amplo espectro expressões não literais - ou figuradas - visto que tais casos por vezes não são facilmente distinguíveis mesmo por seres humanos, e podem depender de contexto, ou de uma análise mais aprofundada do domínio em que estão inseridos (SCHULDER; HOVY, 2014).

O processamento de linguagem figurada, ou não literal, também é uma tarefa não resolvida no âmbito do processamento de linguagem natural. Preocupando-se com a distinção automática entre linguagem literal e figurada - ou que faz uso de figuras de linguagem - este procedimento é necessário para a resolução de vários problemas atualmente estudados neste domínio. Retomando os exemplos citados para a área de Inteligência Artificial, não se pode desenvolver um modelo efetivo de interpretação de necessidades de usuários através da linguagem casual se nele não for inserida uma forma de distinção entre declarações literais e não-literais; a falta dessa distinção certamente causaria interpretações

errôneas, levando a resultados indesejados, situações inesperadas ou mesmo cômicas, como é possível observar em alguns dos sistemas de assistência virtual no mercado considerados estado da arte (CARBONELL, 1982).

Assim, visto a relevância dos problemas acima discutidos, pesquisas desenvolvidas neste segmento, mesmo quando se limitam à uma classe específica de linguagem figurada, podem oferecer contribuições diretas em aplicações e ainda nortear subsequentes pesquisas em NLP e áreas relacionadas.

A linguagem figurada pode ser descrita como um recurso linguístico que enriquece a linguagem, permitindo associações ou comparações que intensificam ideias, e também podem tornar a comunicação mais clara no contexto em que são usadas. Um caso particular de linguagem figurada é a figura de linguagem denominada metáfora - objeto de estudo desta monografia. Metáfora pode ser definida 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 das ideias envolvidas na metáfora é explicado em termos de outra, que possui certa equivalência com a primeira. Metáforas são amplamente utilizadas nas diversas formas de literatura; não obstante, elas também aparecem constantemente no cotidiano, sendo parte integrante da comunicação oral e escrita, nos mais variados domínios. De forma semelhante, metáforas são amplamente utilizadas em conteúdos jornalísticos; tomando como exemplo sua aplicação em notícias, essa figura de linguagem pode ser responsável por expor ao leitor um conceito complexo ou desconhecido em termos de outro – metafórico – que compartilhe semelhanças com o primeiro, e desta forma poderá ser melhor entendido ou interpretado.

A área da linguística possui estudos buscando relações entre o uso de metáforas, a construção do pensamento e até mesmo a formação de opiniões (ver Seção 2.1.2). Nos estudos de Thibodeau e Boroditsky (2011), por exemplo, foram encontrados indícios de que a metáfora utilizada para introduzir um problema através de um conceito (*e.g* comparando o conceito de "crime" com uma besta selvagem; ou comparando-o com um vírus), poderia influenciar nas decisões tomadas pelo leitor quando este efetuava propostas para a resolução do problema. Uma ideia que pode ser obtida a partir dessa propriedade da linguagem metafórica, é que a sua utilização no contexto de notícias falsas poderia diferir em comparação ao uso em notícias verídicas, de forma a modificar a percepção do leitor em relação ao conteúdo exposto.

Computacionalmente, a manipulação de metáforas pode ser dividida em duas subcategorias: reconhecimento de metáfora e interpretação de metáfora. A tarefa de reconhecimento de metáfora busca explorar características semânticas e lexicais na descoberta de padrões para a identificação de tal fenômeno, seja em contexto formal como texto científicos, ou em conversas informais onde o texto muitas vezes não se conforma às normas da linguagem (GLUCKSBERG, 1989) e (SHUTOVA, 2010). Uma vez identificada a existência de uma metáfora, a tarefa de interpretação tenta encontrar o significado

concreto para a frase, isto é, qual o real significado pretendido através do uso da metáfora. Esta monografia concentra-se na tarefa de detecção – ou reconhecimento – de metáforas.

Técnicas computacionais de detecção de metáforas têm sido criadas desde a década de 80, sendo grande parte delas desenvolvidas focando-se em propriedades da língua inglesa. Enquanto diversas abordagens foram adotadas ao longo dos anos, a base teórica tomada pelos estudos parecem constantemente seguir os mapas conceituais metafóricos de (LAKOFF; JOHNSON, 2008) e/ou a violação de restrições seletivas de (WILKS, 1975). Entre as abordagens mais populares encontram-se a violação de preferência seletivas baseadas em estatísticas do domínio, similaridades semânticas baseadas em clusterização de coocorrência, identificação de nível de abstração e, mais recentemente, modelos conexionistas.

Ainda que tais abordagens possuam qualidades importantes, é necessário ressaltar que a tarefa de reconhecimento de metáfora é um problema em aberto, dado que as alternativas atuais ainda apresentam características indesejadas: necessidade de conhecimento especialista adicionado manualmente, dependência da aquisição de domínios anotados, lista de metáforas estática, eficácia apenas em tipos específicos de metáforas, dificuldade na aplicação em idiomas diferentes. Além disso, a tarefa exhibe desafios como a escassez de corpora anotado em idiomas diferentes do inglês, e pela própria característica de produção da linguagem e da dinamicidade das metáforas – isto é, a possibilidade de aparecimento de novas metáforas impede a elaboração de uma lista a priori que cubra todas as possíveis variações de metáforas.

Diante das abordagens atualmente existentes para a detecção de metáforas, optamos por utilizar redes neurais pela sua capacidade de adaptação para diferentes domínios e mudanças no uso da linguagem de acordo com os dados de treinamento. Com o intuito de testar a qualidade da abordagem escolhida, a detecção de metáfora é aplicada no domínio de textos jornalísticos. A preferência por notícias ocorre devido a abrangência de temas cobertos pela área, pelo formato relativamente mais estruturado quando comparado a conversas casuais e, principalmente, pela existência de diversos gêneros de notícias como notícias de ódio, políticas e *fake news*. Dada a possibilidade da existência de vies distintos nestes tipos de notícias, o domínio jornalístico apresenta bom potencial para investigarmos como metáforas podem ser usadas para influenciar a percepção do leitor. Para isto, exploramos ainda quais as emoções associadas as metáforas que se referem a sujeitos em particular.

1.1 Justificativa

O uso de metáforas permeia a linguagem humana em diversos níveis e sua aplicação pode variar do contexto de uma conversa casual ao raciocínio empregado na resolução de problemas complexos. A alta frequência de metáforas na linguagem parece estar presente em vários gêneros como o da literatura, jornalístico, história, entre outros analisados

por (SHUTOVA, 2010) em textos extraídos do British National Corpus (BNC), onde aproximadamente 31,66% das frases anotadas continham linguagem metafórica. No entanto, enquanto metáforas estão presentes em diferentes gêneros, o domínio do texto e o objetivo de sua mensagem parecem influenciar no tipo e frequência das metáforas utilizadas, conforme analisado em (SKORCZYNSKA; DEIGNAN, 2006). Portanto, o desenvolvimento de modelos computacionais capazes de identificar e interpretar texto metafórico de forma precisa e razoavelmente eficaz pode proporcionar avanços significativos para outras tarefas da *NLP*, como a compreensão de linguagem natural.

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 – (THIBODEAU; 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. Dessa forma, a melhoria de técnicas de processamento de metáforas em *NLP* em seu estado final podem levar a avanços em outros campos de estudo, como por exemplo em sistemas de raciocínio que consideram similaridades no conhecimento previamente obtido (e.g., *case based reasoning*) e no reaproveitamento de conhecimento obtido em um domínio em outro, discriminando as características não compartilhadas (i.e. *transfer learning*).

1.2 Objetivo Geral

Construir uma arquitetura de estado da arte de detecção de metáforas e verificar seu comportamento em notícias falsas.

1.3 Objetivos Específicos

- Identificar metáforas usando um modelo previamente treinado (STEMLE; ONYSKO, 2018).
- Analisar e compreender o uso de metáforas em um corpus jornalístico.
- Analisar como diferentes tipos de metáforas compõem os variados tipos de notícias falsas e como as mesmas se comportam quanto a seus usos em notícias verdadeiras.
- Investigar quais são as principais características relacionadas às metáforas detectadas, usando Reconhecimento de Entidade Nomeada (NER) e análise de sentimento.

1.4 Estrutura do texto

O restante do trabalho foi organizado da seguinte maneira: no Capítulo 2 são introduzidos os principais conceitos empregados no estudo, bem como as características dos principais trabalhos correlatos recentes. No Capítulo 3 é descrita a metodologia de

pesquisa. As configurações de hardware e de ambiente de desenvolvimento são discutidas no Capítulo 4. O capítulo 5, por sua vez, traz os resultados advindos das análises e as discussões e conjecturas por eles gerados e, por fim, apresentamos as considerações finais e perspectivas futuras no capítulo 6.

Ademais, mencionamos também que os resultados iniciais desta pesquisa foram compilados para um artigo, sendo este submetido e aprovado para a categoria Aplicações/Experiências no evento Escola Regional de Banco de Dados (ERBD 2019). O evento ocorre anualmente em cidades da Região Sul do Brasil, realizado pela Sociedade Brasileira de Computação (SBC) e organizado por Instituições de Ensino Superior da região; o tema abordado pelo evento no presente ano foi **Inteligência de Dados**.

2 Referencial Teórico e Estado da Arte

2.1 Referencial Teórico

Nesta seção são apresentados os principais conceitos envolvidos na abordagem analisada para a tarefa de detecção de metáforas.

2.1.1 Linguagem Figurada

Linguagem literal – também definida como denotativa – é aquela em que as palavras são usadas com seus significados convencionados, de forma objetiva. Já a linguagem figurada – ou conotativa – é aquela que em que seu significado é obtido fora daquele convencionado, frequentemente ligado ao contexto no qual está inserida.

Em uma tarefa de processamento de linguagem natural, linguagem literal pode ser manipulada mais facilmente do que linguagem figurada pois na primeira as palavras são usadas com seus sentidos convencionados (SHUTOVA, 2011). O contexto, portanto, faz-se importante para a interpretação da linguagem figurada; sendo essa uma informação que muitas vezes está além daquelas fornecidas à um sistema computacional, a interpretação automática desse tipo de linguagem é prejudicada. Por esse motivo, a interpretação deste tipo de linguagem requer maior análise e não é feita trivialmente. Embora a linguagem figurada seja empregada frequentemente na poesia, ressaltamos que não é exclusiva desse domínio; esta se faz presente na linguagem falada e escrita do cotidiano, como afirmam Roberts e Kreuz (1994).

Mais do que a frequente presença no cotidiano, estudos acerca de linguagens figuradas – datando da década de 1980 – apontam esse tipo de linguagem, e majoritariamente metáforas, como sendo parte integrante da formação do discurso, e talvez até mesmo do pensamento (GLUCKSBERG et al., 2001); questões sobre metáforas serão abordadas nas próximas seções. Embora as relações entre pensamento e linguagem estejam fora do escopo deste trabalho, é importante salientar que existem teorias e estudos que procuram relacioná-los.

Existem diversas subcategorias de linguagem figurada, denominadas figuras de linguagem ou figuras de estilo; algumas serão brevemente comentadas aqui.

- **Metáfora:** Comparação entre ideias, sendo em sua maioria, de maneira implícita (e.g. *Aquela pessoa é um anjo*).
- **Símile:** Um tipo de metáfora; dessa forma, também promove a comparação, através de um conectivo explícito (e.g. *Ele está parado como uma estátua*).
- **Metonímia:** Substituição de um termo por outro similar ou equivalente (e.g. *Tomamos três copos de água* – neste caso, o líquido foi substituído pelo seu recipiente).
- **Hipérbole:** Usada para expressar um exagero na ideia a ser transmitida (e.g. *Estava*

morrendo de fome).

- **Antítese:** Expressão de ideias contrastantes entre si (e.g. *O paciente estava entre a vida e a morte*).
- **Eufemismo:** Utilizada para suavizar expressões, tronando-as mais brandas e agradáveis (e.g. *O paciente, que estava muito doente, **partiu desta para uma melhor***).
- **Ironia:** Construção que expressa sentido oposto ao literal (e.g. *Que indivíduo educado. Entrou sem sequer pedir licença*).
- **Personificação:** Quando objetos, animais e outras entidades não humanas, recebem características presentes no ser humano (e.g. *Hoje o dia está triste*).
- **Sinestesia:** Remete à mistura de sentidos (e.g. *Seu perfume era doce...*).

Em (ROBERTS; KREUZ, 1994), os experimentos executados mostram que diferentes tipos de figuras de linguagem são usadas para atingir objetivos comunicativos distintos. Como exemplo, quando um discurso tem por objetivo o esclarecimento de uma ideia, a hipérbole, juntamente com metáforas e símiles são mais comumente utilizadas. Por outro lado, para enfatizar emoções negativas, o uso da ironia é mais frequente.

Assim, dada a grande importância e onipresença da linguagem figurada nos mais variados domínios, bem como sua aplicação em diferentes contextos, a tarefa de processamento desse tipo de linguagem torna-se cada vez mais relevante. No entanto, a tarefa de processamento de linguagem figurada não é trivial. Como descrito anteriormente, situações em que o contexto é necessário (e.g. na ironia) e, no entanto, não é fornecido, não poderão ser distinguidas entre literais ou não. Em função do frequente uso de metáforas, e de sua importância mostrada nos diversos estudos aqui citados, grande parte dos trabalhos publicados nesta área tentam resolver o problema de processamento de linguagem figurada para essa figura de linguagem. De forma semelhante, este trabalho concentra seus esforços no processamento de metáforas. Nas próximas seções, são descritas abordagens para tal tarefa, bem como a estrutura linguística geral de metáforas, as teorias que as cercam e seus usos diversos.

2.1.2 Metáfora

Depois de brevemente descrever as distinções entre linguagem figurada e literal, e assim expor diferentes figuras de linguagens, começamos a evidenciar aquela que é foco deste trabalho – a metáfora. Citamos aqui diferentes definições e tipos de metáforas, visto que esta área de estudo engloba teorias diversas sobre este assunto.

Para (MILLER, 1993), metáforas podem ser definidas como comparações implícitas entre declarações; ou ainda como símiles reduzidas. Como anteriormente comentado, símiles são figuras de linguagem que expressam comparações explícitas com a utilização conectivos. Miller ainda classifica metáforas em três categorias:

- **Nominais:** Construção mais simples; quando um conceito nominal é expressado por uma frase nominal usada metaforicamente; Em inglês, uma forma comum desta

categoria é, segundo o autor, "*x is the y of x'*" (e.g. *The lion is the king of beasts*); também citada em outros estudos como "*Noun is a noun*".

- **Predicativas:** Quando um conceito predicativo é expressado por uma frase predicativa (e.g. *The politician sailed into the crowd*).
- **Sentenciais:** Quando toda a sentença deve ter seu conceito inferido a partir do contexto (e.g. *The laborers returned to the fields* - ao descrever trabalhadores que voltam aos seus postos, depois de um período de descanso).

Segundo (KREUZ; ROBERTS, 1993), metáforas não somente constituem o tipo de figura de linguagem mais comum, mas estas também são as mais abordadas por estudos de campos como filosofia, teoria literária e outros. Além disso, afirma também que o estudo de metáforas é tão importante que os esforços de pesquisa são concentrados majoritariamente neste tipo de figura de linguagem.

Ainda em seu artigo, (KREUZ; ROBERTS, 1993) apresentam também a frequência relativa de figuras de linguagem distintas presentes dentre 38000 palavras, sendo estas extraídas de 32 textos (*contemporary American short stories*); dos oito tipos de figuras de linguagens consideradas, 29% era constituído de metáforas – sendo esta a figura de linguagem mais presente – seguida de hipérbole, com 27%.

Diversos estudos também apontam a divisão de frases metafóricas em duas partes; dentre as diversas descrições para as partes da divisão, Richards difunde os termos *tenor* – para tratar objeto principal a ser referenciado – e *vehicle* – correspondente ao domínio metafórico (BILSKY, 1952). Lakoff e Johnson definem a primeira delas como o alvo (target), e a outra, fonte (source). A fonte pode ser definida como o domínio do qual buscamos a metáfora utilizada para explicação de um conceito; enquanto o alvo representa o conceito que buscamos entender ou explicar. Por exemplo, na frase "*Este relacionamento é uma rua sem saída*", o relacionamento representa o alvo, enquanto a rua sem saída é a fonte (i.e. expressão metafórica utilizada para descrever o relacionamento).

Para (LAKOFF; JOHNSON, 2008), as metáforas constituem o objeto de estudo e o principal foco de suas teorias. Inicialmente abordando a presença regular de metáforas no cotidiano, os autores argumentam que elas não somente se fazem presentes frequentemente em nossa linguagem de forma geral, mas também em nossos pensamentos e ações. Ao relacionar metáforas à pensamentos e ações, os autores discorrem sobre o que denominam sistema conceitual – em tradução direta – referindo-se ao sistema humano, não físico, e que permite, por exemplo, a execução de tarefas cognitivas básicas, como memória, percepção e pensamento; afirmando que nosso sistema conceitual é de natureza metafórica, eles argumentam que, mesmo não sendo este um sistema evidente ou óbvio para nossa própria percepção, é o mesmo que permite a comunicação e essa pode ser utilizada para caracterizá-lo, como descrito em (LAKOFF; JOHNSON, 1980). Como exemplificado por Lakoff e Johnson, culturalmente tratamos o conceito de argumento – e o processo de argumentação – através da metáfora "*argumentação é uma guerra*". Ao argumentar, podemos ganhar ou

perder, planejamos as estratégias, contra-atacamos os argumentos de nossos "adversários" e defendemos nossa posição. Para contrastar com tal estruturação cultural do conceito, os autores colocam o seguinte cenário (citação mantida em inglês para preservar seu significado original):

Try to imagine a culture where arguments are not viewed in terms of war, where no one wins or loses, where there is no sense of attacking or defending, gaining or losing ground. Imagine a culture where an argument is viewed as a dance, the participants are seen as performers, and the goal is to perform in a balanced and aesthetically pleasing way. In such a culture, people would view arguments differently, experience them differently, carry them out differently, and talk about them differently. But we would probably not view them as arguing at all: they would simply be doing something different. It would seem strange even to call what they were doing "arguing". Perhaps the most neutral way of describing this difference between their culture and ours would be to say that we have a discourse form structured in terms of battle and they have one structured in terms of dance.

A situação descrita acima, ao especular sobre as implicações de uma estrutura metafórica alternativa do conceito de argumento, apresenta um cenário em que percebemos, tratamos e agimos de forma diferente a respeito da argumentação. É importante ressaltar que a argumentação não é um tipo de guerra, mas que seu conceito foi estruturado dessa forma, culturalmente, e isso nos permite pensar e interpretar esse conceito em função do outro.

O tipo de metáfora citada nos exemplos acima é denominada, por estes e outros autores, como *metáfora estrutural*, pois neste tipo, um conceito é estruturado metaforicamente a partir de outro; no entanto, esse não é o único presente na comunicação. Também são exemplificadas as metáforas intituladas *metáforas orientacionais – orientational metaphors*; Estas orientam conceitos de acordo com a percepção humana de orientação física. Dessa forma, esta categoria de metáforas é influenciada pelo ambiente, pelos mecanismos físicos que guiam o corpo humano, pelas suas interações com o próprio ambiente, sendo também culturalmente moldada. Nossos sentimentos, por exemplo, seguem tais tipos de orientação; quando estamos tristes, frequentemente dizemos que estamos nos sentindo "para baixo". Se temos uma boa auto-estima, ela está "para cima". Expressões em inglês mantêm as mesmas orientações: *I'm feeling up* ou *I'm feeling down*. Algumas orientações descritas são: Feliz é para cima, triste é para baixo; consciente é para cima, inconsciente para baixo; saúde e vida são para cima, enquanto doença e morte são para baixo. Além disso, algumas dessas expressões parecem possuir precedência sobre outras, quando o sentido entre elas é conflitante ou não coerente. Como no exemplo "A taxa de criminalidade está crescendo", e a partir da orientação "Bom" é "para cima", as ideias são contraditórias

(i.e. o crescimento da taxa de criminalidade não é considerado algo positivo); no entanto, a precedência é tomada pela orientação "Mais" é "para cima". Tais precedências, segundo os autores, dependem da subcultura na qual o indivíduo está inserido, como também de valores pessoais.

Ainda descrevendo os tipos de metáforas, embora ressaltem a importância e a riqueza que as metáforas orientacionais fornecem em termos de linguagem, os autores explicam que as demais percepções humanas, como sentimentos, experiências e outras não palpáveis, não podem ser capturadas por esses tipos de metáforas. Assim, tendemos a objetificar nossas experiências, tornando-as substanciais, para que elas possam ser capturadas metaforicamente – as metáforas criadas por esse fenômeno são chamadas de *metáforas ontológicas*.

Entender nossas experiências em termos de objetos e substâncias nos permite escolher partes de nossas experiências e tratá-las como entidades discretas ou substâncias de um tipo uniforme. Uma vez que podemos identificar nossas experiências como entidades ou substâncias, podemos nos referir a elas, categorizá-las, agrupá-las, e quantificá-las – e dessa forma, raciocinar sobre elas.

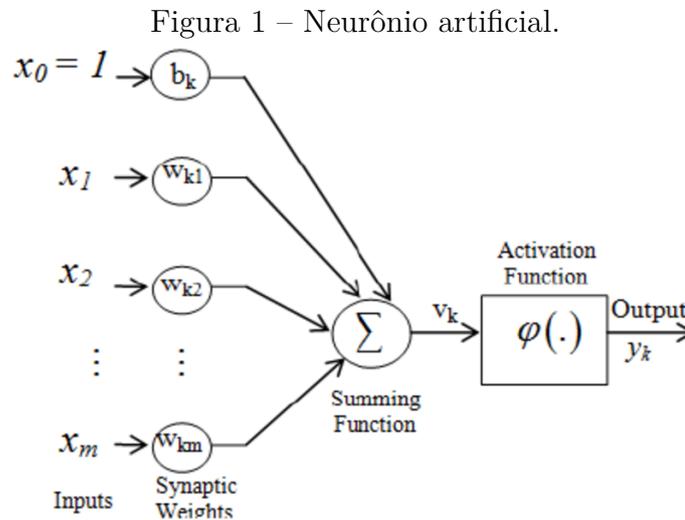
Dessa forma, ao discretizarmos fenômenos e experiências – que é uma prática comum, segundo os autores, pois também somos seres limitados por uma superfície – transformando-as em "objetos palpáveis", podemos efetuar diversas operações sobre elas, de forma a melhor entendê-las.

Dada a grande riqueza de linguagem trazida pelas metáforas, bem como sua onipresença na comunicação humana e estruturação do pensamento, diversas são as tentativas de desenvolvimento de métodos automáticos para sua detecção e interpretação. Dentre as já desenvolvidas – sobretudo para a língua inglesa – podemos citar desde abordagens estatísticas, que levam em conta o conhecimento do domínio sendo tratado, até aquelas que utilizam redes neurais complexas para atingir os mesmos objetivos. As principais ferramentas, modelos e ideias utilizadas para a realização dessa tarefa são descritas nas próximas seções.

2.1.3 Redes Neurais Feedforward

Redes Neurais Artificiais (ANN) são modelos conexionistas computacionais inspirados no cérebro humano. O processamento de novas informações ocorre quando um conjunto de entradas provoca estímulos pelos dendritos que conjuntamente são suficientemente potentes para ativar a sinapse de um neurônio, que por sua vez transfere por seu axônio uma porção desta energia adiante.

Computacionalmente, um neurônio pode ser visualizado como um vértice v de grafo direcionado que realiza uma soma ponderada de acordo com o peso das arestas dos vértices anteriores a si conectados. A Figura 1 exibe graficamente este neurônio.



Fonte: (HAYKIN, 1994)

Adaptando a notação exibida em (HAYKIN, 1994), podemos definir formalmente o estímulo a recebido por um neurônio k como o produto escalar entre o vetor de entrada X e o vetor de pesos U através da equação 1, onde os pesos determinam a inclinação do hiperplano da função aproximada, enquanto o *bias* b_k ajusta o deslocamento do mesmo.

$$a_k = \sum_{i=1}^m x_i * u_{ki} + b_k = X \cdot U_k + b_k \quad (1)$$

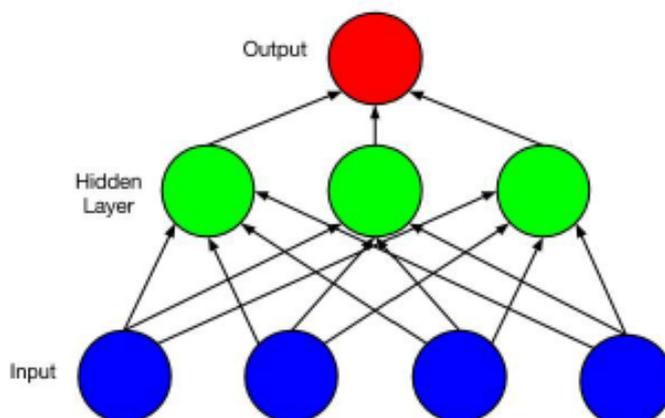
Enquanto a saída estimada \hat{y} é definida pela equação 2 através da função de ativação ϕ aplicada no somatório a_k . Funções de ativação servem para delimitar os valores de saída do neurônio em um intervalo desejado, exemplos populares de funções de ativação incluem a função Sigmoidal (0,1) e Tangente Hiperbólica (-1,1).

$$\hat{y}_k = \phi(a_k) \quad (2)$$

O caso especial no qual é possível representar uma rede de neurônios por um grafo acíclico organizado em fileiras/camadas de neurônios, de maneira que os nós apenas conectem-se com nós na camada a frente, é denominado *Feedforward*. A Figura 2 mostra um exemplo simples de Rede Neural Feedforward (FFNN) de duas camadas.

Topologias de redes neurais com uma ou mais camadas intermediárias entre a entrada e a saída são comumente denominadas *multilayer perceptron*. Uma característica peculiar deste tipo de rede é o fato de que o valor de saída é transparente apenas para os neurônios da última camada, por isso camadas intermediárias são chamadas também de *escondidas*.

Figura 2 – Rede Neural Feedforward completamente conectada de 1 camada escondida.



Fonte: (LIPTON, 2015)

Durante o treinamento, erros são corrigidos através da comparação da saída estimada com a esperada, tipicamente representada por $\hat{y} - y$. Como a rede dispõe apenas do resultado final para comparação, as camadas escondidas inserem um maior grau de dificuldade na correção dos pesos da rede. O algoritmo de *back-propagation* remedia esse problema retropropagando o erro proporcionalmente a quanto este neurônio influenciou na saída da rede. Ainda segundo (HAYKIN, 1994, p.191), *O algoritmo de retro-propagação fornece uma 'aproximação' da trajetória no espaço de peso computado pelo método de descida de gradiente*, ou seja, como o ajuste de pesos segue a direção oposta à do vetor gradiente, cada atualização tende a ocorrer de forma ótima em direção ao ponto mínimo, embora a descida de gradiente possa ficar presa em um mínimo local dependendo do ponto de inicialização.

A maior distinção das redes neurais quanto aos modelos simbólicos é a não existência de regras declaradas explicitamente, geralmente com o auxílio de especialistas no domínio estudado, e sim através da interação entre diferentes nós através dos pesos de suas arestas. Enquanto essa propriedade implica em diversos aspectos positivos como a diminuição da influência do viés do pesquisador nas regras do sistemas e a não necessidade de conhecimento do domínio e de todos os parâmetros de entrada para que partições satisfatórias no espaço de estados sejam alcançadas; fatores negativos como a dificuldade de busca no espaço de topologias e pela configuração final não ser amigável para a compreensão humana precisam ser considerados de acordo com a tarefa a ser executada.

2.1.4 Redes Neurais Recorrentes

Redes Neurais Recorrentes (RNN) podem ser definidas como uma extensão das FFNNs 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.

Tal característica torna as Redes Neurais Recorrentes particularmente interessantes na manipulação de dados sequenciais como palavras em trechos de texto e imagens em vídeos. Em tarefas de Processamento de Linguagem Natural, os benefícios desta abordagem podem ser notados, por exemplo, analisando o seguinte trecho: "(1) O cachorro do meu vizinho é um pastor-alemão ... , (2) enquanto o **meu** é um doberman". Respeitando a dependência dos elementos da sequência é possível identificar que a palavra **meu** na segunda passagem faz uma referência à uma raça de cachorro.

Além disso, as RNNs destacam-se por permitir que o número de entradas e de saídas sejam variáveis (LIPTON, 2015). Como textos frequentemente não apresentam padrão de comprimento, tratá-los através de redes neurais comuns pode não ser uma alternativa viável. De fato, tarefas como Tradução Automática onde línguas diferentes podem representar a mesma informação com diferentes número de caracteres e palavras, ou tarefas que exigem modelos generativos como a criação de uma letra de música a partir de um estilo musical podem se beneficiar de tais variações de arquitetura.

Adaptando a notação de (GOODFELLOW et al., 2016) para se conformar as definições apresentadas na seção 2.1.3, podemos calcular o estado h de uma RNN no passo t através das seguintes equações:

$$a^{(t)} = W_h \cdot h^{(t-1)} + U_x \cdot X^{(t)} \quad (3)$$

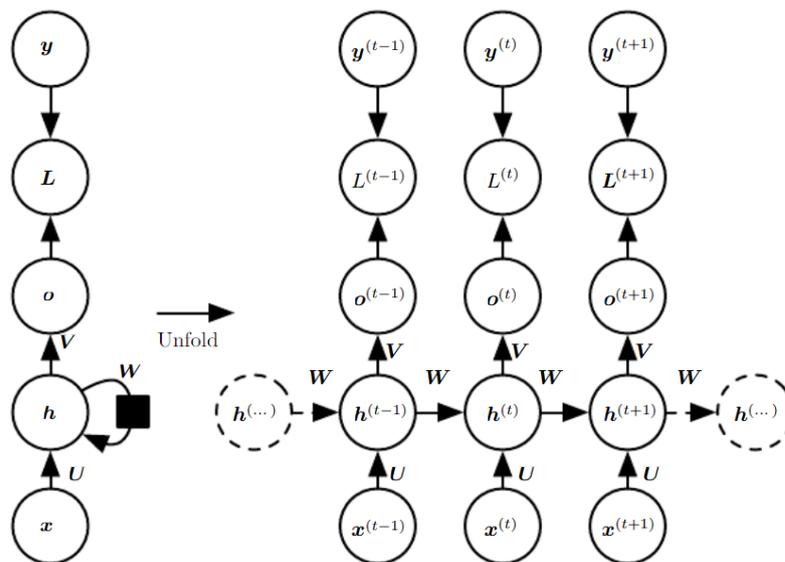
$$h^{(t)} = \phi(a^{(t)} + b_t) \quad (4)$$

Isto é, em vez de simplesmente calcularmos a saída a partir da entrada recebida diretamente no passo t , agora consideramos também uma fração do estímulo da entrada anterior proporcional ao peso da aresta recorrente W_h . A Figura 3 demonstra graficamente os elementos apresentados acima.

Conforme o exemplo de arquitetura, vemos que uma saída $o^{(t)}$ na rede é dada pelo estado h ponderado ao peso de saída V . Enquanto o erro L é calculado comparando-se a saída desejada y à saída estimada \hat{y} obtida aplicando-se uma função de ativação no valor de o .

A correção de erro em RNN ocorre através do ajuste de pesos de forma similar a uma FFNN, porém segundo (GOODFELLOW et al., 2016), o algoritmo de *back-propagation through time* (BPTT) possui duas desvantagens: impossibilidade de paralelização, já que a rede é inerentemente sequencial, e pela necessidade de armazenar o valor de cada estado durante a propagação para o cálculo da retro-propagação. Ambas apresentam custo $O(n)$, em tempo e memória, respectivamente, de acordo com o comprimento n de passos.

Figura 3 – Representação detalhada de uma RNN.



Fonte: (GOODFELLOW et al., 2016)

Além disso, de acordo com (LIPTON, 2015) o treinamento de RNNs é especialmente problemático devido à dificuldade de capturar longas dependências. Essa dificuldade ocorre por causa da chamada explosão e desaparecimento de gradiente, explicadas mais aprofundadamente na próxima subseção. Ainda segundo (LIPTON, 2015), uma alternativa para este problema é o chamado *truncated back-propagation through time* (TBPTT) que delimita uma janela para a retro-propagação do erro. No entanto, enquanto janelas menores amenizam o problema com maior eficácia, ao mesmo tempo estas também prejudicam a qualidade de captura de longas dependências.

Assim, arquiteturas onde a transferência de energia entre dois estados ocorre diretamente através de portões (*gates*) passaram a ser adotadas com o objetivo de diminuir o problema de explosão/desaparecimento de gradiente sem prejudicar a capacidade de capturar longas sequências, entre elas a *Gated Recurrent Units* (GRU) e *Long Short Term Memory* (LSTM).

Outro problema enfrentado por RNNs convencionais é a incapacidade de reconhecer dependências entre um elemento e seus posteriores em uma sequência, dado que o estado somente é propagado adiante. Logo, em frases como "(1) Ela buscou por um **banco** para realizar o pagamento" e "(2) Ela buscou por um **banco** em qual pudesse descansar" identificar se o termo **banco** refere-se a um objeto ou um local exige necessariamente a análise das palavras após o elemento de interesse. (SCHUSTER; PALIWAL, 1997) propuseram como alternativa uma Rede Neural Recorrente Bidirecional (BRNN) similar a arquitetura da Figura 3 porém com duas camadas escondidas h independentes, uma responsável por propagar os estados no sentido positivo (do passado ao futuro) e outra no

sentido negativo.

Portanto, Redes Neurais Recorrentes Bidirecionais do tipo LSTM se adequam bem à tarefa de detecção de metáforas por possibilitar tanto a captura de dependências mais distantes como a relação entre uma possível metáfora e os termos posteriores. Metáforas sentenciais onde o contexto é altamente relevante para sua identificação ou metáforas onde ambos *source* e *target* estão presentes no texto porém distantes podem ser reconhecidas através deste tipo de classificador.

2.1.5 Explosão e Desaparecimento de Gradiente

Como citado anteriormente, o modelo de Redes Neurais Recorrentes possui grande vantagem em problemas sequenciais (como é o caso de processamento de linguagem natural), devido à capacidade de utilizar conhecimento previamente adquirido, recebendo também como entrada a ativação de camadas anteriores.

Contudo, conforme o tamanho da rede aumenta, o treinamento fica mais difícil e pode até mesmo não convergir. Isso acontece porque em sequências maiores, como é o exemplo de uma frase ou parágrafo grande, o relacionamento entre elementos iniciais e finais pode ser relevante para o processamento, e a modelagem clássica de RNN considera muito mais a relevância local do que a distante. Ou seja, o modelo clássico de RNN pode acabar não relacionando corretamente palavras do início e do fim de uma sentença.

Portanto, perda de gradiente em Redes Neurais Recorrentes, levantado em (HOCHREITER, 1998), é um problema extremamente significativo para modelos sequenciais extensos (como sentenças mais compridas), resultando na perda da capacidade de relação entre elementos distantes na sequência, afetando os principais propósitos do modelo.

Alguns modelos propostos são capazes de solucionar este problema, como o Long-Short Term Memory (LSTM), (HOCHREITER; SCHMIDHUBER, 1997), e o Gated Recurrent Unit (GRU), (CHUNG et al., 2014).

2.1.6 Redes Long Short-Term Memory

O Long Short-Term Memory (LSTM) é um modelo motivado pelo problema apresentado anteriormente (desaparecimento de gradiente), que está relacionado com o distanciamento exponencial do erro retro-propagado a longo prazo nas RNNs, (GRAVES; SCHMIDHUBER, 2005).

O LSTM é um modelo com capacidade de aprendizagem mais limitada, porém poderoso, como citado em (SUNDERMEYER; SCHLÜTER; NEY, 2012). A arquitetura 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, empoderadas com *gating units* (*input*, *output* e *forget*, análogos a operações de leitura, escrita e *reset* de uma memória), que somam as ativações para cada *gate* das camadas anteriores com a atual, passando por uma função sigmoide.

Quando os *gates* estão fechados, (ativação próxima a 0), as entradas irrelevantes ou dados fora do padrão (*outliers*) são ignorados e a informação (conteúdo da célula de memória) é mantida, permitindo que dados relevantes sejam levados a longo prazo (GERS; SCHMIDHUBER; CUMMINS, 2000).

No contexto desse trabalho, o processamento de frases metafóricas pode demandar o relacionamento de palavras distantes em uma frase, e para tal, o LSTM é um modelo importante que permite a realização do processamento dessas relações.

2.1.7 Word Embeddings

A representação de palavras através de vetores contínuos, ou *word embeddings*, trata-se de uma forma de mapear as palavras de um vocabulário para um espaço de dimensões reduzido em comum. Esta representação permite que o conhecimento sobre as relações existentes entre as palavras seja reutilizado em outras tarefas de NLP, reduzindo assim suas complexidades. Entre as técnicas mais populares destacam-se as abordagens através de matrizes de coocorrência de termos devido a suposição de que termos similares aparecem em contextos mútuos e redes neurais que exploram as distribuições probabilísticas dos termos. No entanto, uma série de limitações ocorrem de acordo com a abordagem escolhida, como a incapacidade de reconhecer os diferentes significados apresentados por um mesmo termo e a incapacidade de aproveitar o conhecimento proveniente da ordem dos termos ou ainda sobre a semelhança entre palavras que compartilham mesma raiz. Algumas das principais técnicas para a descoberta de tais relações são apresentadas a seguir, introduzindo seu conceito com o auxílio de uma técnica mais simples de vetores discretos chamada *one-hot*.

2.1.7.1 Representação One-hot

Representar uma palavra através de um vetor numérico em vez de um conjunto de caracteres pode facilitar sua manipulação. Na codificação *one-hot*, ou um-quente em português, o vetor p_k de uma palavra $p \in Voc$ na posição k do vocabulário Voc é determinado por:

$$p_k = [0_1, 0_2, \dots, 0_{k-1}, 1_k, 0_{k+1}, \dots, 0_{n-1}, 0_n] \quad (5)$$

Onde o vetor possui apenas uma posição de valor 1, com o restante das $|Voc| - 1$ posições fixadas em 0. Por exemplo, em um vocabulário de $|Voc| = 4$ com as palavras $Voc = \{estudante, professor, prova, universidade\}$, o vetor que representaria a palavra **professor** é dado por $x_2 = [0, 1, 0, 0]$. Embora tal representação seja facilmente gerada a partir do vocabulário, o seu formato não permite a extrapolação de outras informações como relações de coocorrência e similaridade semântica de palavras no vocabulário.

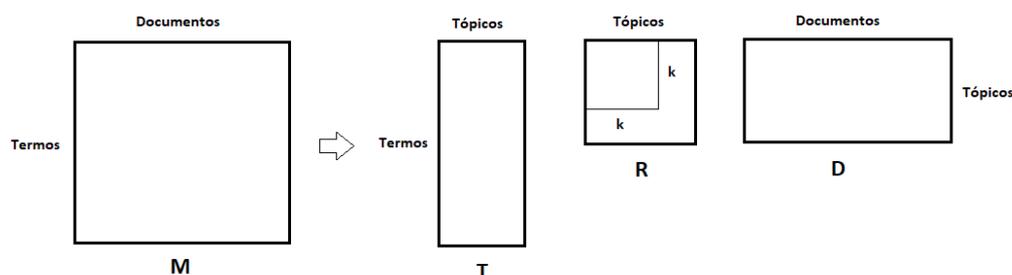
2.1.7.2 Representação com Análise de Semântica Latente

A Análise de Semântica Latente (LSA) é uma técnica estatística com maior poder de representação capaz de identificar as relações entre termos, documentos e temas através do uso da hipótese da semântica distribucional, isto é, de que palavras com sentidos semelhantes ocorrem em textos semelhantes. Para isso a técnica LSA utiliza uma matriz M onde cada linha corresponde a uma palavra, cada coluna a um documento ou passagem do texto e o valor da célula é dado tipicamente pelo peso *tf-idf* do termo, onde o peso cresce quanto mais frequente for este termo no documento e mais raro em relação ao demais documentos do corpus. Em seguida, essa matriz passa pela técnica de Decomposição de Valor Único (SVD), onde M é transformada no resultado do produto de 3 matrizes, conforme a fórmula 6:

$$M = TRD \quad (6)$$

Uma matriz T representando um termo a partir de sua relevância a cada um dos tópicos, outra matriz D representando cada documento de acordo com a sua importância em relação aos tópicos e uma terceira matriz R diagonal que mapeia a relevância de tópico um espaço dimensional reduzido k , obtido removendo os tópicos de menor relevância, como descrito em (LANDAUER; FOLTZ; LAHAM, 1998). Entre as vantagens desta técnica estão o potencial de aplicação para categorização de textos e por ser totalmente gerada a partir do próprio corpus analisado em vez de ser dependente de dicionários externos como WordNet, além de ser capaz de superar um dos principais problemas enfrentados pelo Modelo de Espaço Vetorial (VSM): a dificuldade em captar significados semelhantes expressos por termos diferentes (sinonímia), (MANNING; RAGHAVAN; SCHÜTZ, 2009). No entanto, tal abordagem ainda sofre algumas limitações como não conseguir distinguir diferentes significados de um mesmo termo (polissemia) e, por se tratar de uma técnica do tipo *bag of words*, não aproveitar a informação decorrente da ordem das palavras.

Figura 4 – Decomposição de Valor Único.



Como pode ser visto na Figura 4, diferente da representação *One-hot* onde cada palavra estava em uma dimensão diferente, quando o SVD é aplicado sobre a matriz original reduzindo o número de dimensões K , os termos e documentos são forçados a um espaço dimensional em comum e agrupados de acordo com uma similaridade em relação à um determinado tópico.

2.1.7.3 Representação com Redes Neurais

Abordagens que exploram a informação da ordem de palavras através da probabilidade conjunta de sequências de palavras foram então investigadas, entre elas o modelo n -gramas, no qual uma sequência longa é dada pela combinação de várias sequências curtas assumindo a propriedade de Markov de ordem $n - 1$, isto é, de que a probabilidade de uma palavra é afetada por seus n vizinhos anteriores. (BENGIO et al., 2003) propôs uma alternativa ao método tradicional de n -gramas que permitisse considerar: (1) contextos mais longos do que 1 ou 2 palavras e (2) similaridade entre as palavras. No entanto, segundos os autores a generalização de variáveis discretas (e.g., palavras em uma sentença) são tipicamente mais vulneráveis à maldição da dimensionalidade (VERLEYSSEN; FRANÇOIS, 2005), enquanto variáveis contínuas possuem a vantagem de que sua função de aprendizagem apresenta características locais suaves, de forma que mudanças nas variáveis observadas resultam em diferenças menores na função obtida.

A contribuição de (1) e (2) é destacada quando sequências não presentes no conjunto de treinamento são analisadas, demonstrada pelos autores na seguinte passagem:

Por exemplo, ter visto a frase "O gato está andando no quarto" no corpus de treinamento poderia nos ajudar a generalizar para que a frase "Um cachorro estava correndo em uma sala" seja tão provável quanto à primeira, simplesmente porque "cachorro" e "gato" (assim como "o" e "um", "sala" e "quarto", etc..) possuem papéis semânticos e gramaticais semelhantes.

De forma que a generalização para sequências similares de cada sentença observada no conjunto de treinamento corresponde a um número combinatorial de outras sentenças. Os autores resumem a técnica proposta em 3 tarefas: representar cada palavra do vocabulário por um vetor de propriedades de palavras distribuído, representar a probabilidade conjunta das sequências de palavras através de seus vetores individuais e a aprendizagem simultânea dos vetores e dos parâmetros da função de probabilidade.

Como o objetivo é determinar a palavra com maior probabilidade dada uma sequência de palavras como entrada, o desafio computacional é estimar uma função que represente de forma satisfatória a função de verossimilhança mantendo uma complexidade aceitável. Os autores implementaram essa função através de uma Rede Neural Feedforward onde a entrada foram as palavras da sequência, a primeira camada escondida possuía um neurônio para cada entrada – de forma a mapear o vetor de propriedades de palavra;

a segunda camada com um único neurônio para a combinação dos valores através de uma função tangente hiperbólica, e uma saída do tipo *softmax* que indicava a palavra de maior probabilidade. Resultados encontrados no artigo demonstraram que esta abordagem apresenta de 10% a 20% de melhoria de complexidade quando comparada ao modelo de *n*-gramas.

Mais recentemente, técnicas de *word embeddings* a partir de redes neurais inspiradas em (BENGIO et al., 2003) foram exploradas buscando maior eficiência computacional, dentre as quais três populares se destacam: Word2Vec por (MIKOLOV et al., 2013a), GloVe por (PENNINGTON; SOCHER; MANNING, 2014) e FastText por (BOJANOWSKI et al., 2016).

Como a abordagem analisada neste estudo é baseada em (STEMLE; ONYSKO, 2018), no qual a aplicação de *word embeddings* geradas a partir de FastText exibiram os melhores resultados, descrevemos aqui como o modelo proposto alcança melhor complexidade computacional.

Segundo (BOJANOWSKI et al., 2016), embora várias técnicas de *word embeddings* tenham sido estudadas, grande parte delas ainda não explora a informação a respeito da estrutura de uma palavra. Em idiomas morfolologicamente ricos, em especial, palavras podem apresentar um grande número de variações, de forma que a utilização de conhecimento sobre as sub-palavras pode auxiliar nas variações particularmente raras no conjunto de treinamento. Para isso os autores propõem uma alteração no modelo de *skip*-grama contínuo apresentado em (MIKOLOV et al., 2013b).

No modelo proposto, uma palavra é representada pela soma dos *n*-gramas de seus caracteres. Ademais, este modelo permite o uso de operações aritméticas de vetores para descoberta de analogias já que a relação da entrada e saída são lineares. Isto é, a diferença entre os vetores das palavras *rei* e *rainha* é aproximada da diferença entre os vetores de *homem* e *mulher*, sendo então possível encontrar o vetor de *rei* através da operação $HOMEM - MULHER + RAINHA$.

2.1.8 Reconhecimento de Entidades Nomeadas

Ao abordar problemas de processamento de linguagem natural e recuperação de informações, muito comumente há interesse em analisar as entidades referenciadas através de um conteúdo textual. A necessidade de reconhecimento e extração de informações como nomes de pessoas, de organizações e localidades geográficas de textos fez surgir um conjunto de técnicas que compõe a tarefa denominada Reconhecimento e Classificação de Entidades Nomeadas.

Segundo (NADEAU; SEKINE, 2007), os primeiros estudos acerca desta importante tarefa datam de 1991 – durante a sétima conferência IEEE de Aplicações da Inteligência Artificial – quando (RAU, 1991) descrevia um sistema baseado em regras e heurísticas para a identificação e extração de nomes de companhias. Já o termo *Named Entity* –

Entidade Nomeada – foi cunhado durante a *Message Understanding Conference - MUC-6*, na qual era pretendido estudar e aprimorar a extração de informações a partir de dados não estruturados. Após este evento, os esforços e estudos científicos acerca desta tarefa começaram a aumentar.

Visto os esforços concentrados nesta tarefa específica, é possível inferir que o Reconhecimento de Entidades Nomeadas não é trivial. Como há uma grande quantidade de termos que poderiam ser considerados entidades, não haveria capacidade de processamento computacional suficiente para que todas elas fossem suportadas em um sistema de tal propósito; dessa forma, e ainda segundo (NADEAU; SEKINE, 2007), para se tornar efetiva, a tarefa se concentra nos ditos *rigid designators* – designadores rígidos, em tradução livre. Estes, segundo a filosofia da linguagem, podem ser definidos como termos que se referem sempre à um mesmo objeto, em todos os *mundos possíveis* em que tal objeto existe. Os *rigid designators* incluem nomes próprios, espécies biológicas e substâncias; além dessas citadas, unidades monetárias, expressões temporais e algumas grandezas numéricas também são consideradas no processo de identificação de entidades.

Em uma *shared task* descrita em (SANG; MEULDER, 2003), foram estudados, desenvolvidos e testados sistemas de reconhecimento de entidades nomeadas para o inglês e o alemão. Os principais atributos utilizados para construção dos sistemas da competição foram atributos lexicais, *part-of-speech*, *n*-gramas, entidades nomeadas anteriormente preditas, informações ortográficas e *gazetteers* – dicionários geográficos. A respeito dos tipos de modelos usados, o Modelo de Máxima Entropia foi a escolha da maior parte dos grupos, se mostrando significativo neste tipo de tarefa, já que as equipes melhores colocadas utilizavam este tipo de modelo; o modelo oculto de Markov, e redes neurais também foram explorados por outras equipes.

2.1.9 Fake News

Atualmente, muito tem se discutido a respeito das mundialmente conhecidas *fake news* – ou notícias falsas em tradução direta. Sendo nomeado pelo *Collins English Dictionary* um dos principais termos do ano de 2017, é inegável que as *fake news* tenham ganhado visibilidade e causado grande impacto a respeito das informações por meio delas divulgadas. Embora o termo pareça claro em sua definição, devemos aqui comentar que estas podem ter múltiplos significados, além de serem organizadas em diversas subcategorias de acordo com sua origem e tipo de conteúdo.

O termo *fake news* pode ser definido como sendo usado para se referir a notícias que são intencionalmente e verificadamente falsas e que poderiam, dessa forma, enganar os leitores (ALLCOTT; GENTZKOW, 2017). Outra definição adotada para esses tipos de notícias as descreve como sendo informações fabricadas que imitam conteúdo midiático de notícias em sua forma, mas não em seus processos ou em suas intenções (LAZER et al., 2018). Através da revisão de estudos que utilizavam o termo *fake news* e analisando

suas definições em cada um deles, foi possível identificar pontos comuns à todas elas: notícias falsas buscam se apropriar da aparência de conteúdos reais, tanto em *websites*, como em artigos e até mesmo em imagens (JR; LIM; LING, 2018). Para o corpus utilizado, notícias falsas são reportagens divulgadas por *fontes que fabricam inteiramente informações, disseminam conteúdo enganoso ou que distorcem grosseiramente notícias reais*¹, isto é, o veículo jornalístico é o que determina a veracidade da informação, não o conteúdo em si.

Embora consigam atingir os mais diversos domínios, principalmente pelo seu poder de propagação, as notícias falsas são debatidas mais frequentemente nos âmbitos políticos e sociais. Um dos fatores que contribuem para a disseminação desse tipo de notícia é, segundo (LAZER et al., 2018), o fato de que indivíduos questionam a credibilidade e confiabilidade de informações somente quando estas vão de encontro ao seu ponto de vista e, dessa forma, tendem a aceitar aquelas que afirmam suas crenças e valores.

Em (VOSOUGHI; ROY; ARAL, 2018), é apresentado um estudo a respeito da disseminação de notícias verdadeiras e falsas na rede social Twitter, durante os anos de 2006-2017. Segundo os autores, notícias falsas eram compartilhadas mais rapidamente, e possuíam maior poder de difusão do que as verdadeiras. Segundo os autores, embora esforços para desenvolver estudos sobre o tema tenham ocorrido com frequência, poucos são aqueles que procuram analisar a dinâmica de propagação e aceitação de informações acerca de diferentes tópicos, ou que buscam entender as diferenças de disseminação entre notícias falsas e verdadeiras.

Ainda no contexto de redes sociais, um estudo conduzido em (GOTTFRIED; SHEARER, 2016) apontou que, na época, 44% dos americanos utilizavam o Facebook como principal fonte de notícias. O consumo de notícias em redes sociais torna-se problemático quando estas mostram-se como os principais vetores para a propagação de notícias falsas; como apontado em (GUESS; NYHAN; REIFLER, 2018), o Facebook era o maior disseminador de notícias falsas dentre várias mídias digitais consideradas, também em 2016.

2.2 Estado da Arte

A seguir são apresentados cinco trabalhos relacionados ao tema deste estudo. Primeiramente é descrito uma competição de detectores de metáforas sob uma mesma base de dados que buscou permitir a comparação direta de desempenho obtido entre diversos modelos conexionistas. Em seguida, dois modelos participantes da competição são discutidos, incluindo o modelo LSTM – base do presente trabalho. Por último, duas abordagens simbólicas que utilizam heurísticas linguísticas são descritas.

Em (LEONG; KLEBANOV; SHUTOVA, 2018), são descritos os resultados obtidos a partir de uma *shared task* sobre detecção de metáforas. A *shared task* tinha por objetivo

¹<https://github.com/several27/FakeNewsCorpus>

a comparação de modelos computacionais capazes de detectar uso metafórico da linguagem à nível de palavras. Para isso, os grupos participantes da competição dispunham de um corpus anotado contendo metáforas. Como comparativo, os organizadores forneceram aos participantes dois modelos *baseline* baseados em *WordNet*, níveis de concretude/abstração de palavras e outras *features*.

Durante o período da competição, diversos times submeteram seus modelos para avaliação, sendo que alguns deles obtiveram performance superior aos modelos *baseline* disponibilizados. A respeito dos tipos de modelos e arquiteturas utilizadas, apenas um dos sistemas não utilizou um modelo baseado em redes neurais; todos os modelos que superaram os *baselines* utilizaram *word embeddings* juntamente com outros atributos. Ademais, é importante ressaltar que os sistemas de detecção de metáforas obtiveram desempenho superior para algumas das subcategorias textuais contidas no corpus, o que pode indicar uma dinâmica diferente de utilização de linguagem figurada para cada uma delas, ou mesmo evidenciar limitações em relação aos métodos e tecnologias utilizadas para a identificação de diferentes tipos e usos de linguagem metafórica; também foi possível extrair algumas informações sobre a performance de detecção de metáforas em relação à classe gramatical das palavras. Foi verificado que todos os sistemas participantes obtiveram melhores resultados ao detectar metáforas em verbos; enquanto que os piores resultados concentraram-se nos advérbios; para adjetivos e substantivos, o desempenho foi equivalente.

Em (STEMLE; ONYSKO, 2018), um detector de metáforas foi implementado para a *shared task* sobre detecção de metáforas, descrita na seção anterior; este sistema de detecção foi utilizado neste trabalho para verificar a utilização de metáforas no domínio escolhido. Segundo os autores, embora a detecção manual de metáforas já tenha sido amplamente utilizada, a abordagem automatizada de detecção de metáforas surge com diversos desafios.

Os autores, ao utilizarem a ideia de que o uso de linguagem não literal, e especialmente a metafórica, muda à depender do grau de conhecimento – ou proficiência – na linguagem, efetuaram o treinamento das *word embeddings* em diferentes corpora, com parte delas contendo diferentes níveis de proficiência na língua inglesa. Assim, verificam que a arquitetura e as técnicas utilizadas são capazes de detectar metáforas com um desempenho consideravelmente superior à outras técnicas já utilizadas na abordagem do problema.

Em (SCHULDER; HOVY, 2014), é apresentada uma abordagem estatística para o problema de detecção de metáforas. A principal ideia seria utilizar raridade de metáforas como um indicativo para sua presença, ignorando o vocabulário comum dos candidatos à classificação. Palavras fora de contexto literal, ou sendo utilizadas de forma incomum, estarão presentes em menor frequência, e serão fortes candidatas à metáforas.

Para modelar o problema e capturar a raridade de palavras, uma ideia baseada na relevância de termos adaptada para o conceito de domínios foi utilizada. Como vantagem, tal ideia requer apenas dados do domínio literal do texto; assim, são exploradas as métricas

relevância de domínio – que mede a tipicidade de um termo em relação à um domínio – e *relevância comum* – que mede o uso comum ou frequente de um termo, apontando que ele não é um bom indicativo de uso metafórico. Como demonstrado, os modelos desenvolvidos baseados nas ideias descritas acima mostraram-se superiores aos métodos *baseline* da área; o modelo que utilizava apenas relevância de termos como *feature*, mostrou desempenho 58% melhor em relação ao *baseline*.

A abordagem apresentada por (NEUMAN et al., 2013) utiliza a definição de variedades de padrões linguísticos realizadas por (KRISHNAKUMARAN; ZHU, 2007), focando o trabalho em três tipos de frases metafóricas envolvendo substantivos. Entre eles o tipo 1, onde um sujeito é associado a um objeto, como "Deus é um rei", o tipo 2, onde o verbo é o foco do uso metafórico representando o ato de um sujeito em um objeto, como em "a guerra absorveu sua energia", e também o tipo 3, quando a metáfora envolve um adjetivo e um substantivo, como em "doce criança".

Para solucionar o problema de detecção de metáforas, os autores utilizam três algoritmos baseados em regras, ou seja, não precisam ser treinados em algum corpus anotado. Cada um dos algoritmos funciona para um dos tipos de metáfora. Os autores aplicaram o algoritmo em dois corpus distintos, sendo o *Reuters RCV1 data set* e também o *New York Times archive (1984)*, com foco em metáforas envolvendo os substantivos *God, Governance, Father, Mother e Government*.

2.2.1 Análise Crítica

Nesta subseção discutimos quais as possíveis deficiências apresentadas por modelos alternativos, bem como os diferenciais deste estudo em relação ao demais trabalhos do estado da arte.

Em relação aos modelos simbólicos que exploram conhecimentos especializados sobre propriedades linguísticas, tais abordagens são muitas vezes limitadas a casos específicos, alcançando bons resultados somente em determinados domínios, metáforas compostas por certas classes gramaticais, ou idiomas que compartilham determinadas características. 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 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 dar direcionamento à análise das metáforas, suas aplicações e impactos, em domínios específicos – neste caso, detecção de *fake news*.

3 Metodologia

Na Seção 3.1 são abordados cada um dos *data sets* utilizados, descrevendo suas origens, principais características, bem como uma breve descrição de seus campos. Em seguida, na Seção 3.2 são apresentados as técnicas aplicadas para gerar as anotações de entidades, metáforas e emoções. Na Seção 3.3 as questões a serem respondidas pelos experimentos são levantadas. Na Seção 3.4 são esboçados os cenários dos experimentos. Por último, na Seção 3.5 são apresentadas as métricas que permitem a quantização das respostas às questões anteriores.

3.1 Dados

A presente seção busca descrever os conjuntos de dados utilizados durante o desenvolvimento deste trabalho. Como o presente trabalho busca explorar um sistema já desenvolvido para a detecção de metáforas, descreveremos, também, os *data sets* por ele utilizados.

3.1.1 Dados de Treinamento

Como brevemente descrito na subseção 2.2, o modelo a ser aqui considerado para a detecção de metáforas, apresentado no artigo de (STEMLE; ONYSKO, 2018), foi desenvolvido para a *shared task* exposta em (LEONG; KLEBANOV; SHUTOVA, 2018). Buscando o *benchmarking* dos sistemas desenvolvidos para a *shared task*, o VU Amsterdam Metaphor Corpus foi utilizado.

O corpus acima citado foi proposto pela NAACL (*North American Chapter of the Association for Computational Linguistics*), durante o workshop de processamento de linguagem figurada, e serviu como base para o treinamento e avaliação dos modelos submetidos para a competição, como aquele a ser explorado nesta monografia; Sendo formado a partir do BNC – British National Corpus – que contém mais de 100 milhões de palavras, é constituído pelas subcategorias acadêmico, notícias, conversação e ficção; as metáforas presentes no corpus foram anotadas pelo chamado *Metaphor Identification Procedure* – um método para identificar palavras metafóricas utilizadas no discurso – explanado detalhadamente em (GROUP, 2007).

Os gêneros textuais contidos no corpus da competição, embora não possuam a mesma quantidade de textos, são formados por, aproximadamente, a mesma quantidade de *tokens*. Para fins de avaliação, 23% dos dados disponibilizados pela competição foram fornecidos aos participantes sem as anotações de metáforas, com o restante sendo utilizado no treinamento. Além disso, foram criadas duas categorias para os dados de classificação: *All Part-of-speech*, contendo metáforas de diferentes classes gramaticais; e *Verbs*, contendo

somente verbos empregados metaforicamente, sendo analisadas em nosso estudo todas as classes gramaticais. A tabela 1 apresenta os campos disponíveis no corpus.

Tabela 1 – Campos do *data set* Metaphor Corpus

Campo	Tipo	Descrição
txt_id	String	Identificador do texto
sentence_id	Inteiro	Identificador da sentença em relação ao texto
sentence_txt	String	Conteúdo textual

3.1.2 Fake News Corpus

Um dos objetivos específicos deste trabalho é a detecção e análise do uso de expressões e linguagem metafórica no âmbito das notícias falsas. Tornando-se um problema emergente e cada vez mais presente principalmente no meio digital, as notícias falsas mostram-se um terreno fértil para diversas análises e estudos. Buscamos aqui efetuar uma análise sobre a utilização de linguagem figurada – mais especificamente metafórica – na composição de diferentes tipos de *Fake News*. Para este fim, necessitamos de um corpus extenso e com uma grande variedade de notícias falsas.

Disponibilizado de forma aberta em (SZPAKOWSKI, 2018), o *data set* de *fake news* por nós utilizado – *Fake News Corpus* – é uma coleção de diferentes categorias de notícias extraídas de 745 domínios. Os dados foram adquiridos utilizando *data scraping*, isto é, coleta de dados brutos de forma automática, através da biblioteca *Scrapy* para a linguagem *python*; os domínios a partir dos quais as notícias são extraídas, foram fornecidos em *OpenSources* – contendo listas de fontes de informações, disponíveis para uso público. Adicionalmente, como o *OpenSources* não fornecia muitas fontes de notícias confiáveis, foram considerados também artigos do *NYTimes* e *WebHose English News Articles*. Depois do procedimento de coleta de dados, as notícias obtidas no formato *HTML* foram extraídas ao utilizar a biblioteca *newspaper*, também para a linguagem *python*.

Após ser formatado e estruturado no formato *CSV – comma separated values* – o corpus passa a contar com campos como domínio, url, conteúdo, título, autor, fonte e outros (conforme descritos na Tabela 2). Ademais, algumas das categorias de notícias falsas presentes no *data set* serão brevemente descritas abaixo.

- **Fake News** - Fontes que criam informações, ou efetuam distorções dos fatos reais.
- **Satire** - Fontes que utilizam humor, ironia, exagero e informações falsas para descrever eventos.
- **Extreme Bias** - Fontes que divulgam um ponto de vista específico.
- **Conspiracy Theory** - Fontes que disseminam teorias da conspiração.
- **Junk Science** - Fontes que promovem pseudociência, metafísica e outras notícias sobre falsa ciência.
- **Hate News** - Fontes que promovem diversos tipos de notícias de ódio.

- **Clickbait** - Fontes que fornecem informações confiáveis, mas com apelo em seus títulos, com exagero e conteúdo ambíguo.
- **Credible** - Fontes que fornecem notícias confiáveis.

Tabela 2 – Campos do *data set* Fake News

Campo	Tipo	Descrição
id	Inteiro	Identificador único para a notícia
domain	String	Domínio do qual a notícia foi extraída
type	String	Tipo da notícia - cada um dos descritos acima
url	String	URL da notícia
content	String	Conteúdo da notícia
scraped_at	Date Time	Data/Hora em que a notícia foi extraída
inserted_at	Date Time	Data/Hora em que a notícia foi inserida
updated_at	Date Time	Data/Hora em que a notícia foi atualizada
title	String	Título da notícia
authors	String	Autores da notícia
keywords	String	Palavras-chave da notícia
meta_keywords	String	Palavras-chave fornecidas pela página da notícia
meta_description	String	Descrição fornecida pela página da notícia
tags	String	Etiquetas da notícia
summary	String	Sumário da notícia
source	String	opensources, nytimes, or webhose

Embora o *data set* disponha de diversas informações obtidas durante a extração das notícias, neste trabalho nos concentraremos em um subconjunto de seus campos, dando enfoque principalmente às informações textuais nele contidas. A tabela 3 apresenta os campos que serão utilizados para as análises e tarefas de detecção de metáforas.

Tabela 3 – Subconjunto de campos do *data set* Fake News a serem explorados

Campo	Tipo	Descrição
id	Inteiro	Identificador único para a notícia
type	String	Tipo da notícia
content	String	Conteúdo da notícia
title	String	Título da notícia

3.2 Workflow - Fluxo de trabalho

O processamento de cada notícia envolve três etapas principais: o reconhecimento de entidades, a detecção de metáforas e a anotação das emoções correspondentes. Como cada etapa é composta pela aplicação de várias técnicas de NLP, estas são apresentadas separadamente nas subseções a seguir.

3.2.1 Reconhecimento de Entidades Nomeadas

Buscando um melhor entendimento da composição das notícias estudadas, a tarefa de reconhecimento de entidades nomeadas foi aplicada às sentenças de notícias do *data set* explorado. A aplicação deste procedimento também possibilitou o direcionamento das análises e permitiu uma nova perspectiva a respeito do uso de metáforas, com o intuito de se analisar o uso de metáforas a respeito de entidades diversas referenciadas pelas notícias.

A tarefa de reconhecimento e anotação de entidades nomeadas foi feita dividindo os corpos das notícias em sentenças e utilizando-as como entrada para a *API DBpedia Spotlight* – vide subseção 2.1.8. Nesta, a anotação foi feita através do *endpoint /annotate*, que efetua o procedimento em quatro passos (localização, mapeamento de candidatos, desambiguação e filtragem).

Abaixo, um exemplo de chamada do método *annotate* da *API DBpedia Spotlight* usando *cURL*. Para este trabalho utilizamos somente o argumento *text*, no entanto, existem outros argumentos que realizam filtrações nas entidades a serem detectadas.

```
1 curl -X GET
2     "https://api.dbpedia-spotlight.org/en/annotate
3         ?text=Cristiano Ronaldo plays a lot."
4     -H "accept: application/json"
```

O exemplo acima retorna:

```
1 {
2   "@text": "Cristiano Ronaldo plays a lot.",
3   "@confidence": "0.5",
4   "@support": "0",
5   "@types": "",
6   "@sparql": "",
7   "@policy": "whitelist",
8   "Resources": [
9     {
10      "@URI": "http://dbpedia.org/resource/Cristiano_Ronaldo"
11      ,
12      "@support": "1542",
13      "@types": "Http://xmlns.com/foaf/0.1/Person,Wikidata:
14      Q937857,Wikidata:Q5
15      ,Wikidata:Q24229398,Wikidata:Q215627,DUL:
16      NaturalPerson,DUL:Agent,
17      Schema:Person,DBpedia:SoccerPlayer,DBpedia:Person,
18      DBpedia:Athlete,
```

```

15         DBpedia:Agent",
16         "@surfaceForm": "Cristiano Ronaldo",
17         "@offset": "0",
18         "@similarityScore": "1.0",
19         "@percentageOfSecondRank": "0.0"
20     }
21 ]
22 }

```

Quanto ao retorno do método *annotate*, e a principal utilização neste trabalho, podemos citar a chave *Resources* e os objetos nela contidos, os quais representam entidades detectadas, seus tipos, a referência dentro do *DBPedia* para a entidade, a posição no texto e também as *scores* de similaridade da posição respectiva do texto (*offset*).

3.2.2 Detector de Metáforas

O modelo aplicado neste estudo trata-se de uma BRNN LSTM (consultar as Seções 2.1.4 e 2.1.6) desenvolvida por (STEMLE; ONYSKO, 2018), estudo abordado em detalhes na Subseção 2.2, doravante referido como **botzen**. Segundo o relato dos autores, o desempenho do modelo botzen foi otimizado através de um ajuste de parâmetros e de topologia de maneira *ad-hoc*, por meio de busca em *grid*, com validação cruzada. Seguimos aqui a configuração que forneceu os melhores resultados no trabalho original, conforme exibida na Tabela 4.

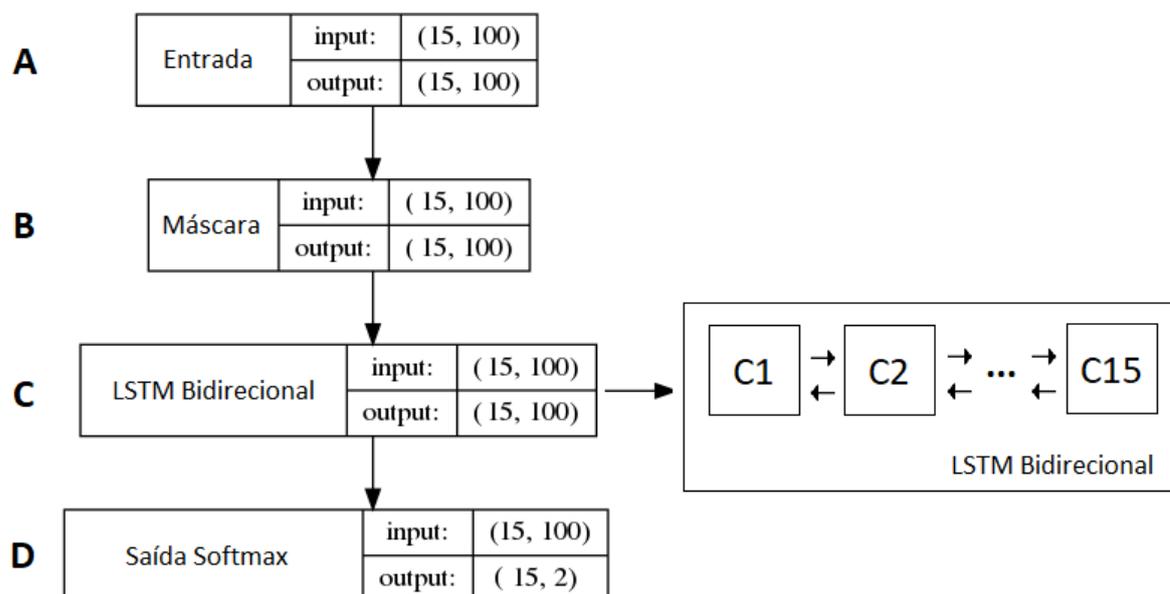
Tabela 4 – Parâmetros botzen.

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

O modelo é alimentado com sequências de palavras representadas pelos vetores de *features* (i.e., valores extraídos da lista de *word embeddings*) de cada palavra de uma sentença. Caso a sequência seja menor que o tamanho definido (15), a entrada é preenchida por *dummy information* (valores de preenchimento nulos). Se a sequência ultrapassar o comprimento delimitado, a mesma é separada em múltiplas sequências. As notícias são separadas em múltiplas sentenças através da ferramenta *sent tokenize* da biblioteca NLTK.

Em relação à topologia da rede, representada na Figura 5, o tamanho da camada de entrada **A** varia de acordo com as *features* utilizadas, dado que o modelo aceita múltiplas informações como classes gramaticais e diferentes representações de word embeddings (consultar Seção 2.1.7) e, portanto, a entrada precisa ser adaptada para tolerar várias quantidades de *features* para uma mesma palavra. Em seguida, esta camada é conectada a uma camada de máscara **B** que serve para ignorar os valores de preenchimentos nos passos seguintes. A informação é então transmitida para uma camada LSTM bidirecional que a transforma no estímulo da camada de saída **D**. Finalmente, a saída é gerada pela aplicação de uma função de ativação do tipo *softmax* que indica a probabilidade da palavra ter sido usada metaforicamente ou não.

Figura 5 – Topologia botzen.



Fonte: Autoria Própria

Para facilitar a compreensão é apresentado um trecho de notícia usado como entrada para o modelo botzen: "*The United States government and its agencies are in no position to be wasting hard-earned taxpayer dollars.*", ou em português, "O governo dos Estados Unidos e suas instituições não estão em posição de desperdiçar o dinheiro ganho com grande esforço pelos contribuintes". Cada palavra nesta sentença seria então representada por um vetor de valores reais com 100 dimensões.

A saída gerada para a sentença é uma lista contendo o identificador da notícia, o identificador da sentença na notícia, a posição da palavra na sentença, o grau de certeza de que a palavra se trata de uma metáfora e a palavra. Para o exemplo apresentado, o modelo detectou duas palavras com alto grau de serem metáforas: (2926943_1_11, 0.99, position) e (2926943_1_15, 0.76, hard-earned).

3.2.3 Anotação de Emoções

Com o intuito de investigar para qual a finalidade as metáforas foram empregadas, as palavras marcadas como metáfora na saída do botzen são buscadas no conjunto de dados *NRC Lexicon*. Este conjunto possui uma lista de oito emoções associadas manualmente a 10.170 palavras compostas por substantivos, verbos, adjetivos e advérbios comumente usados na língua inglesa (ver Seção 4.2.7).

Para garantir que palavras presentes no NRC Lexicon não sejam reconhecidas devido a variações morfológicas, a ferramenta *NLTK WordNetLemmatizer* é empregada para normalizar as diferentes aparições de um termo nas notícias e no conjunto de emoções para o mesmo formato. Cada palavra encontrada é representada por um vetor que indica se o termo invoca ou não as oito emoções: raiva, expectativa, aversão, medo, alegria, tristeza, surpresa e confiança. Palavras não listadas no NRC Lexicon são presumidas como nulas dado que a maioria das palavras anotadas também são nulas quanto à emoções de acordo com os voluntários responsáveis pela anotação do corpus, conforme descrito por (MOHAMMAD; TURNEY, 2010; MOHAMMAD; TURNEY, 2013).

Usando um segundo trecho da notícia apresentada na subseção anterior: "*Lets look at just a couple M(0.51)_glaring examples.*", onde o autor cita exemplos de ações tomadas pelo governo dos Estados Unidos vistas como desperdício de dinheiro público, a metáfora detectada foi razoavelmente associada à emoção de raiva.

3.3 Perguntas da Pesquisa

Com o intuito de melhor direcionar o delineamento dos cenários de análise e métodos de avaliação, nesta subseção são apontadas questões relacionadas ao problema central do trabalho.

1. Como podemos descrever cada classe de notícia segundo seu perfil de distribuição de metáforas?
2. De acordo com as métricas definidas, existem diferenças significativas no perfil de notícias autênticas e *fake news*?
3. Se foi possível representar o perfil de notícias através das metáforas nelas detectadas, como podemos descrever as emoções relacionadas à estas metáforas? Existe variação entre as classes?
4. Como a intensidade de sentimentos e a subjetividade expressas no texto estão relacionados aos tipos de notícias? Estas intensidades variam de acordo ao uso de metáforas?
5. Como se comportam diferentes tipos de notícias em relação às entidades específicas por elas referenciadas? (*e.g.*, quais são as entidades mais frequentemente referenciadas para cada uma das categorias de notícias).
6. O padrão encontrado para as respostas acima se mantêm quando analisadas somente

notícias referente à entidades específicas? Se não, existe variação entre entidades?

3.4 Casos de Análise

Abaixo são descrito brevemente os cenários para os quais são direcionados as diferentes análises das anotações resultantes do *workflow*.

3.4.1 Cenário 1

Gerar análise quantitativas e qualitativas sobre as anotações referentes as metáforas detectadas pelo botzen.

3.4.2 Cenário 2

Analisar quais as emoções que definem as metáforas que se referem à entidade Donald Trump durante o período pré-eleitoral de 2016.

3.4.3 Cenário 3

Analisar quais as emoções que definem as metáforas que se referem à entidade Hillary Clinton durante o período pré-eleitoral de 2016.

3.5 Método de Avaliação

A qualidade da classificação gerada pelo detector de metáforas é examinada através da comparação com a anotação manual de um conjunto de cem notícias aleatórias obtidas por amostragem estratificada. A anotação manual de metáforas foi realizada por pesquisadores da área de linguística da Universidade Estadual de Campinas (UNICAMP), que utilizaram três classes de marcação:

- Metáforas: palavras que claramente apresentam mapeamentos entre fonte e alvo, ou cujos critérios de classificação do método MIPVU (STEEN, 2010; KRENNMAYR, 2011) se aplicam com clareza.
- Figurativas: outros tipos de linguagem figurada onde não ocorrem transferência de significado entre domínios distintos.
- Duvidosas/Incertas: metáforas identificadas por seres humanos, mas que causaram dúvidas ao tentar classificá-las sob os critérios de padronização do método MIPVU. Os motivos das dúvidas foram diversas, entre eles, o uso de palavras cujo significado foram convencionados, especialmente palavras da área de computação como "mouse". No entanto, o método MIPVU aconselha manter a anotação inicial em casos de dúvidas, isto é, o uso desta classe é válido para a avaliação.

As métricas de precisão, *recall* e medida F1 são calculadas em dois cenários, o primeiro ignorando as palavras marcadas como duvidosas e o segundo incluindo-as. O objetivo é

descobrir se as metáforas que não são claramente cobertas pelo método de anotação padrão também resultam em piores resultados ao classificador. O desempenho da classificação é calculado para cada classe individualmente para identificar tipos de notícias possivelmente mais difíceis de serem anotadas.

4 Recursos de Hardware e Software

Neste capítulo são descritos os recursos físicos e virtuais que foram utilizados como meios para execução das tarefas desenvolvidas neste trabalho.

4.1 Recursos de Hardware

Inicialmente, os experimentos foram realizados em computadores pessoais comuns (com processadores quad/octa core, 8GB de memória RAM). Ao decorrer do desenvolvimento da pesquisa, observamos certos gargalos ao efetuarmos execuções diversas e análises em amostras maiores do *data set* estudado. Por esse motivo, foi necessário buscar por soluções alternativas que oferecessem uma infraestrutura compatível com as necessidades da escala do trabalho.

Para a realização dos processamentos, se fez necessário a contratação de uma máquina na nuvem (instância virtualizada da *Digital Ocean*), com 16GB de RAM e 6vCPUs, rodando o sistema operacional *Ubuntu Server 18.04*.

4.2 Recursos de Software

Para a exploração e análise dos *data sets*, bem como para a execução do modelo descrito detalhadamente nas seções anteriores, a linguagem *Python* foi utilizada. Sendo uma linguagem amplamente utilizada para tarefas de processamento de linguagem natural, esta possui uma grande e crescente comunidade ativa, além de uma ampla variedade de bibliotecas disponíveis para utilização em trabalhos relacionadas a esta e outras áreas correlatas.

4.2.1 Keras

Keras é uma biblioteca de alto nível que contém uma API para criação e utilização de redes neurais sobre outras bibliotecas, *frameworks* e *toolkits* de mesmo propósito. Ela foi utilizada para a criação do modelo apresentado em (STEMLE; ONYSKO, 2018), e descrito nas seções 2.2 e 3.2.2, aplicado neste trabalho para a detecção de metáforas nos *data sets* definidos.

4.2.2 Gensim - FastText

Um dos interesses ao modelar certos problemas de processamento de linguagem natural é a representação de palavras em um espaço multidimensional, utilizando vetores numéricos – as chamadas *Word Embeddings* – como descrito na seção 2.1.7. Para esse fim, Gensim – outra biblioteca para a linguagem Python, mostra-se importante. Desenvolvida

sobre outras bibliotecas que visam performance, ela implementa os principais algoritmos para a construção de modelos vetoriais, além de contar com outras funcionalidades, como *topic modeling*, por exemplo.

4.2.3 Sklearn - Metrics

Sklearn, ou *Scikit-learn*, trata-se de uma biblioteca python para aprendizado de máquina; conta com uma grande quantidade de algoritmos e demais ferramentas para as tarefas de classificação, regressão, agrupamento, redução de dimensionalidade, seleção de modelo, pré-processamento e métricas de avaliação. Para este trabalho, nos concentramos nas métricas para avaliação de modelos, a partir das quais podemos validar e verificar o desempenho dos modelos utilizados.

4.2.4 NLTK e TextBlob

NLTK – *Natural Language Toolkit* – e *TextBlob* são conjuntos de ferramentas e bibliotecas para o processamento de linguagem natural, desenvolvidos para a linguagem *Python*. Do NLTK serão utilizadas funcionalidades para análise lexical, geração de *n*-gramas, análise e visualização de estrutura sintática, reconhecimento de entidades nomeadas, *POS-tagging*, análise de sentimentos e outras; além disso, será utilizada para grande parte das atividades de pré-processamento e análise exploratória dos *data sets* deste trabalho. Já a utilização da biblioteca *TextBlob* será focada principalmente nas análises de polaridade e subjetividade que esta disponibiliza.

4.2.5 NumPy

NumPy é um conjunto de ferramentas e bibliotecas para criação e manipulação de *arrays* multidimensionais. Sendo amplamente utilizado para computação científica, também é utilizada como base para a criação de inúmeras bibliotecas para a mesma linguagem, incluindo algumas já citadas aqui. Desenvolvido para a linguagem *Python*, contém uma ampla quantidade de funções para cálculos numéricos e complexas operações sobre vetores de grandes dimensões.

4.2.6 DBpedia Spotlight

DBpedia é uma iniciativa que tem como o objetivo a extração de conteúdo da enciclopédia colaborativa online, a *Wikipedia*, e disponibilizar tal conteúdo de forma aberta na *web*; através da extração de conteúdo se faz possível a consulta por meio de uma linguagem de *queries*, de forma semelhante ao *SQL*. Além disso, existem times que buscam disponibilizar o conteúdo correspondente em outros idiomas. Seu funcionamento detalhado é descrito em (AUER et al., 2007). *DBpedia Spotlight*, por sua vez, assim como descrito em (MENDES et al., 2011), pode ser definido com um sistema criado com o

propósito de automaticamente anotar e efetuar o processo de desambiguação dos conteúdos estruturados advindos da *DBpedia*. Disponível abertamente por meio de um *Web Service*, um texto é enviado, juntamente com os tipos de entidades buscadas, e então o sistema efetua uma sequência de quatro passos para que a anotação aconteça. No primeiro deles, o sistema **localiza** palavras – ou conjuntos de palavras – definidas como *surface forms*, que possam ser referências a entidades; no segundo, é feita a seleção de um conjunto de *surface forms* do passo 1 e são buscados recursos do *DBpedia* que possam ser **candidatos** a referências feitas pelas *surface forms* selecionadas. Depois, ocorre a **desambiguação**, de forma a definir o melhor candidato para cada uma das *surface forms*. Finalmente, é feita a **filtragem** do conteúdo anotado para que o retorno siga as configurações determinadas pelo usuário.

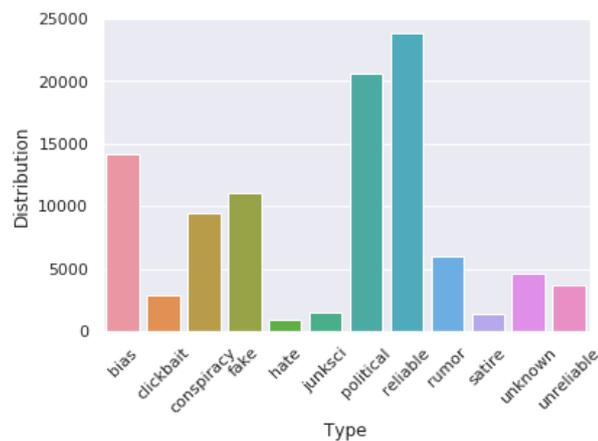
4.2.7 NRC Lexicon

O NRC Lexicon é um conjunto de 10.170 palavras anotadas através de colaboração coletiva pelo serviço *Mechanical Turk* da Amazon. Cada palavra é descrita por um vetor que representa a presença ou não de duas polaridades de sentimentos (positiva e negativa) e oito emoções (raiva, expectativa, aversão, medo, alegria, tristeza, surpresa e confiança). Essas emoções, definidas por (PLUTCHIK, 1980), são um superconjunto das seis frequentemente consideradas as emoções básicas capazes através de combinações mais complexas de cobrir toda a gama de emoções sentidas por seres humanos, propostas por (EKMAN, 1992). O superconjunto é utilizado por capturar ainda pares de emoções inversas: alegria-tristeza, raiva-medo, confiança-aversão e expectativa-surpresa, de maneira que é possível mapeá-las para um espaço dimensional em comum.

5 Análise de Resultados

Este capítulo apresenta a análise das anotações geradas para uma amostra de 100.000 notícias (em torno de 1% do corpus original). O uso de uma amostra foi necessário devido ao custo computacional da tarefa de detecção de metáforas dado que a taxa de anotação, na máquina remota locada, foi de 10.000 notícias por hora e, portanto, a anotação do corpus completo levaria em torno de 40 dias. Para a seleção de notícias foi utilizada amostragem estratificada para garantir a representabilidade original dos tipos de notícias conforme ilustrado na Figura 6.

Figura 6 – Distribuição por tipo de notícia.



Fonte: Autoria Própria

5.1 Análise de Entidades

As análises de entidades efetuadas seguiram os procedimentos descritos na subseção 3.2.1. A identificação de entidades tem o intuito de tornar mais concreto o possível viés no uso de metáforas, de forma a evidenciar diferenças quantitativas e qualitativas através de exemplos. As entidades mais frequentemente citadas nas notícias foram países, figuras ligadas à política e entidades relacionadas a redes de notícias.

Ao optar pela análise das entidades que, além de serem referenciadas com frequência nas notícias e conhecidas por serem figuras públicas de opiniões conflitantes, buscamos verificar os nomes Donald Trump e Hillary Clinton em relação a seus usos em termos gerais e também relacionando-os com metáforas. Além dos motivos acima citados, o fato de o período de eleições presidenciais americano ter sido marcado por grandes quantidades de notícias falsas a respeito desses candidatos também corroborou para a escolha destas personalidades.

A execução do procedimento de anotação de entidades resultou em 10.950.423 menções de entidades anotadas na amostra. Dentre estas, Trump foi citado 201.678 vezes, representando aproximadamente 1.8% do total de menções de entidades; já para Hillary, foram encontradas 33.248 menções, representado aproximadamente 0.3% – tais percentuais se mostram significativos, dada a quantidade de entidades detectáveis que poderiam ser citadas nas notícias. Os resultados da anotação são sumarizados na Tabela 5.

Tabela 5 – Resultados da Anotação de Entidades.

Entidade	Número de menções	Porcentagem
Hillary Clinton	33.248	0.3%
Donald Trump	201.678	1.8%
Todas entidades	10.950.423	100%

Sobre os tipos de notícias envolvidos nas menções dos candidatos, notícias confiáveis e políticas, respectivamente, foram as que mais citaram Hillary; já Trump foi citado majoritariamente em notícias políticas, seguido de notícias confiáveis e duvidosas – estas últimas em proporções semelhantes – mesmo as notícias confiáveis sendo a classe majoritária no corpus, isto é, Donald Trump é um exemplo de agente que apresenta proporcionalmente menos notícias confiáveis do que seria esperado. Para investigar esta desproporção de citação de Trump em notícias confiáveis, analisamos também as entidades mais frequentes em cada classe de notícia. Embora Trump esteja entre as entidades mais citadas no corpus em geral, em relação a notícias confiáveis ele apresentou relativamente um menor número de citações, abaixo das 5 entidades mais frequentes.

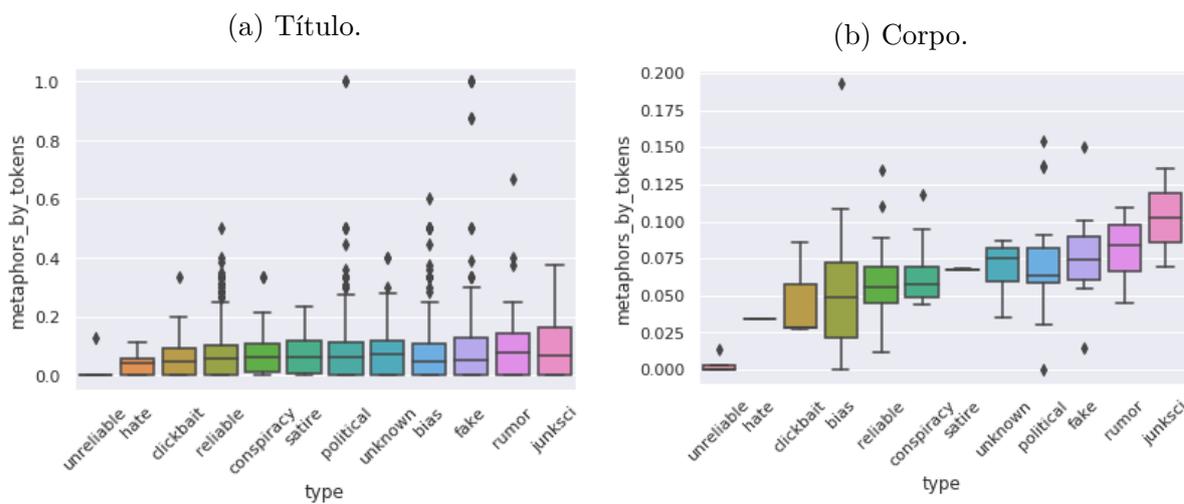
5.2 Análise de Metáforas

A média de metáforas empregada em cada classe de notícia é calculada para os títulos e para os corpos de texto separadamente. Como as notícias possuem tamanhos variados, as médias são normalizadas pelo número de palavras em cada sentença. Em termos gerais, a execução do detector de metáforas na amostra selecionada resultou na anotação de 4.816.838 palavras usadas metaforicamente.

Conforme apresentado na Figura 7, é possível verificar que títulos de todos os tipos de notícias possuem médias relativamente equilibradas de metáforas – ao contrário do que era esperado, já que títulos tendem a utilizar recursos que induzam o leitor a acessá-lo, e não necessariamente correspondem inteiramente ao conteúdo da notícia – bem como uma grande quantidade de *outliers*, dada a quantidade limitada de *tokens* neste tipo de conteúdo textual. Em relação aos corpos de notícias, é interessante notar que notícias falsas constituem o tipo de notícia que mais utilizam metáforas em relação à proporção de tokens no texto. As metáforas geralmente estão associadas as seguintes funções retóricas: (i) engajar emocionalmente o leitor (SAPOLSKY, 2010; CITRON; GOLDBERG, 2014); (ii) manipular a percepção do leitor sobre um determinado assunto

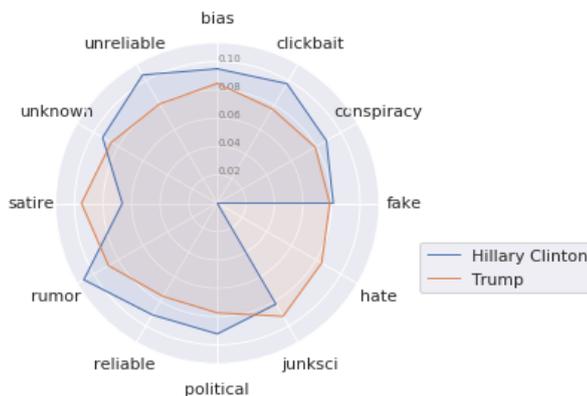
(THIBODEAU; BORODITSKY, 2011; THIBODEAU; HENDRICKS; BORODITSKY, 2017) ; (iii) exprimir uma ideia de forma mais viva; e/ou (iv) veicular uma maior quantidade de informação através de um termo (CHARTERIS-BLACK, 2006). Todas essas funções podem contribuir para aumentar o apelo do texto aos seus leitores. Além disso, ressaltamos que a hipótese de que a utilização de metáforas para compor um recurso de detecção de notícias falsas se faz plausível.

Figura 7 – Médias de metáforas por palavras.



Fonte: Autoria Própria

Figura 8 – Média de metáforas por tipo de notícia para Trump e Hillary.



Fonte: Autoria Própria

Em relação à média de metáforas para notícias que mencionavam as entidades políticas escolhidas, levando em conta os diferentes tipos de notícias, pudemos verificar que as referências à Donald Trump possuíam médias mais equilibradas entre todos os os tipos de notícias, com aproximadamente 0.8 metáforas por notícia. Já para Hillary, embora numericamente a diferença de médias não fosse altamente significativa, seu nome fora

mencionado com menos frequência em notícias do tipo sátira e ódio, e com sensivelmente mais frequência naquelas dos tipos rumor e não confiáveis.

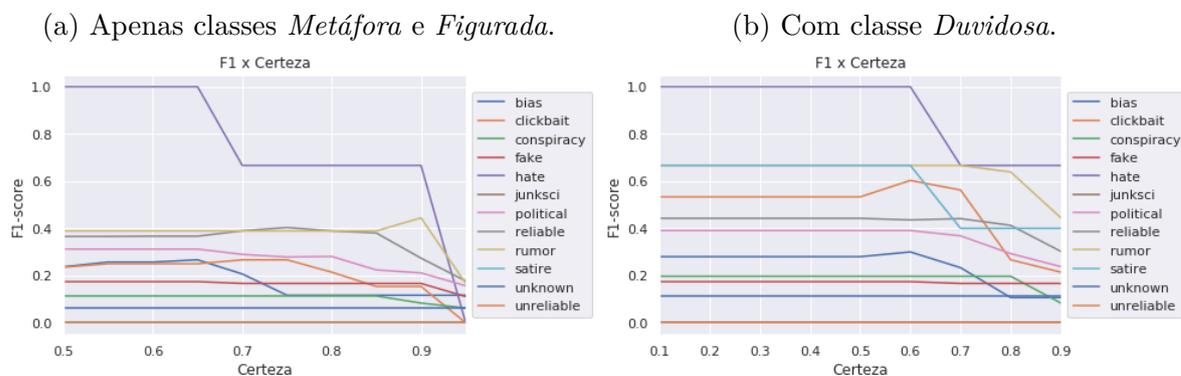
5.2.1 Avaliação do Detector de Metáforas

Como a saída do detector de metáforas fornece o grau de certeza da classificação, as métricas de avaliação podem ser calculadas considerando somente metáforas acima de determinados limiares, variando de 10% à 90%, com o objetivo de encontrar o parâmetro que maximiza o desempenho do modelo segundo a medida F1. Conforme a Figura 9 é possível notar que a medida F1 das diferentes classes se concentram na região entre 0.1-0.5, onde o grau de certeza entre 60% e 80% maximiza o desempenho geral das classes.

No entanto, ao analisarmos o gráfico para a variação de *recall* para este mesmo intervalo de 60%-80% notamos que existe uma queda conforme maior o grau de certeza utilizado. Assim, o grau de 65% parece ser o ponto ótimo para o classificador, o que é confirmado também no cálculo da medida F1.

Ao incluirmos na avaliação a classe de metáforas *Duvidosas*, isto é, metáforas difíceis de classificar usando as decisões arbitrárias do MIPVU, inesperadamente ocorre um aumento significativo de precisão. O intervalo ótimo de desempenho se mantém na faixa entre 60% e 70%, como ilustrado na Figura 10(b).

Figura 9 – Variação da medida F1 conforme o grau de certeza.



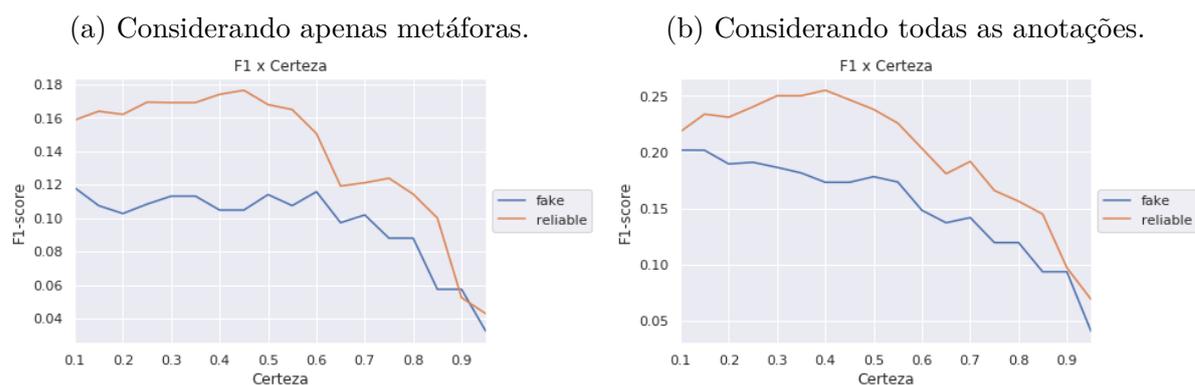
Fonte: Autoria Própria

Entretanto, podemos perceber uma queda drástica de *recall* em graus de certeza acima de 60%. Usamos a métrica F1 para confirmar o valor ótimo observado. Portanto, o limiar de certeza entre 60% e 65% apresentam melhores resultados tanto ao ignorar e ao incluir a classe de metáforas *Duvidosas*.

É importante ressaltar que como a amostra de cem notícias foi coletada de maneira estratificada, as classes *bias* e *unreliable* tiveram apenas uma notícia inclusa na amostra, o que as tornam suscetíveis a presença de anomalias dado o tamanho da amostra. Isso pode explicar o comportamento destoante da demais classes em todas as análises apresentadas.

Como as classes *fake* e *reliable* são o foco do trabalho e houve uma discrepância de resultados entre as duas classes que poderia tornar inviável a comparação das metáforas detectadas, uma nova anotação manual de 100 notícias com somente as duas classes foi realizada para minimizar a influências de anomalias no resultado. Nesta nova anotação houve um maior equilíbrio de desempenho entre as duas classes, com a classe de notícias confiáveis apresentando maior precisão ao considerarmos apenas a anotação de metáforas, mas a classe *fake* melhor considerando as anotações de metáforas, figurativas e duvidosas. Além disso, como a precisão diminui ao usarmos um grau de precisão muito alto, há indícios de que metáforas detectadas com grau de certeza entre 60% e 80% são mais precisas do que metáforas detectadas com 100% de certeza. Em relação a medida F1, a classe *fake* apresentou pior desempenho em ambos os casos, o que significa que existem metáforas nesta classe que não são reconhecidas pelo classificador.

Figura 10 – Variação da medida F1 para as classes *fake* e *reliable* conforme o grau de certeza.



Fonte: Autoria Própria

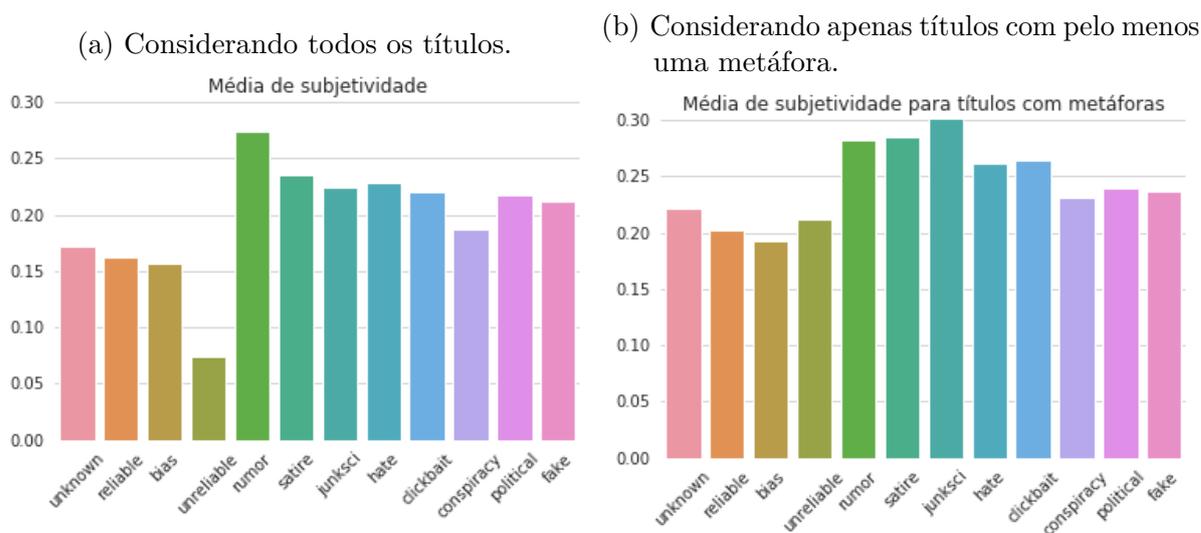
A análise das métricas de desempenho não somente ajuda a validar os resultados do experimentos, mas também a guiar qual o valor de certeza de metáfora deve ser considerado nas demais análises. Além disso, é possível perceber que notícias confiáveis utilizam mais metáforas "puras", isto é, metáforas onde seres humanos conseguem claramente encontrar os domínios de fonte e alvo da figura de linguagem. Uma possível explicação para este fato é que as notícias confiáveis podem apresentar mais metáforas convencionais do que as falsas. Ou seja, as notícias confiáveis podem apresentar um perfil mais tradicional estilisticamente.

5.3 Análise de Emoções

Além da análise quantitativa das metáforas empregadas em cada classe de notícias, verificamos aqui a natureza destas metáforas através da análise de emoções, sentimentos e subjetividade presentes nos textos.

Inicialmente, em uma pequena amostra de títulos de notícias do *dataset* original, efetuamos algumas análises de subjetividade, intensidade e polaridade de emoções utilizando os recursos próprios da biblioteca *python NLTK*. A subjetividade se define pela variação de uma ideia ou opinião a depender dos ideais de cada indivíduo. O índice de subjetividade-objetividade é dado através de um *lexicon* que possui adjetivos da língua inglesa, extraídos de avaliações de produtos e anotados – automática e manualmente – com um *score* que indica a subjetividade expressada pelo autor (SMEDT; DAELEMANS, 2012). Através da subjetividade, e relacionando-a às metáforas pudemos perceber que esta se mostrou mais significativa quando presente em títulos com metáforas. Como é possível verificar na figura 11, a média de subjetividade é relativamente alta para grande parte dos tipos de notícia; no entanto, ao limitar esta análise a apenas títulos com pelo menos uma metáfora, a média se eleva sensivelmente e a subjetividade tende a se equilibrar para todos os tipos de notícias. Isso indica que metáforas estão comumente relacionadas à subjetividade. Os demais aspectos das emoções não apresentaram resultados significativos.

Figura 11 – Média de subjetividade para títulos de notícias, por tipo de notícias.

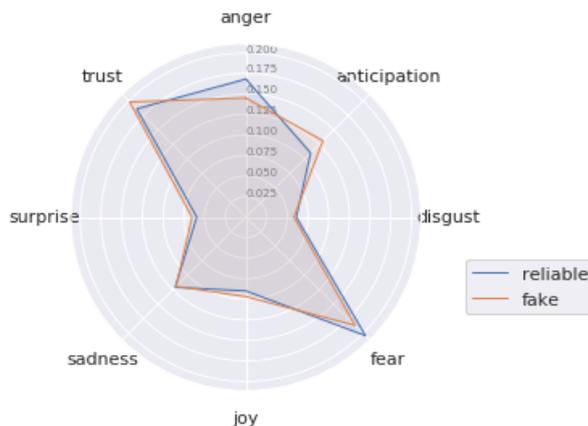


Fonte: Autoria Própria

A detecção de emoções nos corpos das notícias seguiu os procedimentos apresentados na subseção 4.2.7; seus resultados mostraram que 1.178.733 metáforas foram associadas a alguma das emoções – isso representa 24.47% do total de metáforas. A análise de emoções foi concentrada nas categorias *reliable* e *fake*, que recebem maior foco neste trabalho. A figura 12 mostra que esses tipos de notícias são compostos de forma semelhante pelos mesmos níveis de emoções entre eles; esse resultado é, de certa forma, inesperado, considerando que subjetividade se mostrou mais relacionado a metáforas, e estas aparecem com maior frequência em notícias falsas do que em aquelas confiáveis, sendo assim esperado um nível relativamente maior de emoções sendo aplicadas em *fake news*. Este fenômeno poderia ser explicado por uma possível tentativa de mascarar a notícia como confiável

através da cópia de seu estilo linguístico, assim como observado na média de metáforas em títulos.

Figura 12 – Distribuição de emoções de notícias confiáveis e falsas.

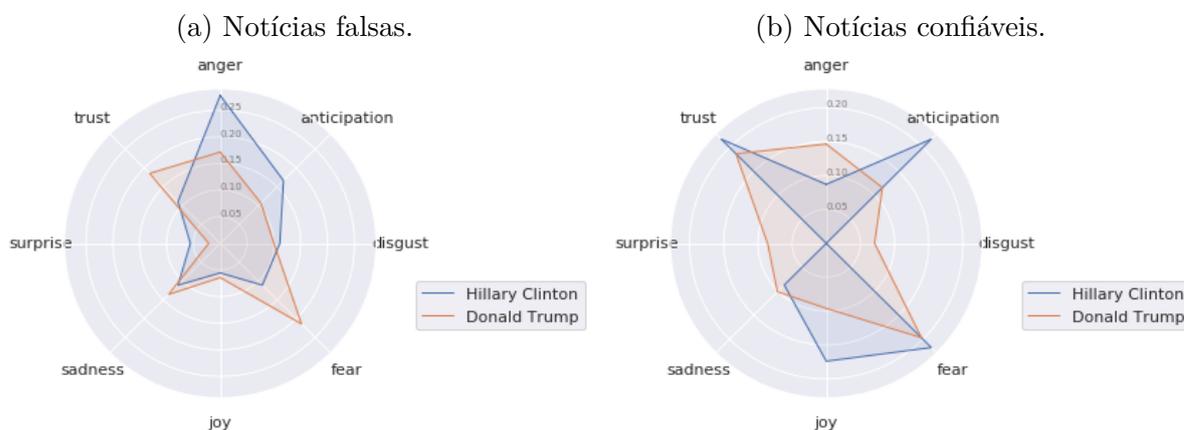


Fonte: Autoria Própria

5.4 Análise de Metáforas Emotivas direcionadas à Entidades

Dado que a classe de notícias *fake* possui uma maior média de metáforas, como indicado na Figura 8(b), a hipótese de que metáforas podem ser usadas para influenciar a interpretação do leitor parece ser verdadeira. Para investigar mais a fundo, analisamos as notícias onde foram detectadas emoções associadas às metáforas que co-ocorreram com as menções de duas entidades específicas: Donald Trump (5008 notícias) e Hillary Clinton (519 notícias).

Figura 13 – Distribuição de emoções para entidades Trump e Hillary.



Fonte: Autoria própria

Embora a distribuição de emoções referente às duas entidades para todos os tipos de notícias são similares, ao separarmos as classes, especialmente notícias falsas e confiáveis,

existe uma clara diferença de emoções usadas para se referir aos dois. Por exemplo, na Figura 14(a) notamos que a emoção mais frequentemente associada às metáforas para Hillary é a raiva, enquanto que para o Donald Trump é o medo. É difícil dizer se essas metáforas que co-ocorrem com o nome de Trump enfatizam o medo sobre situações no mundo que continuarão a ocorrer (ou venham a ocorrer) caso Trump não interfira, ou se refletem o temor por efeitos ruins possivelmente causados por eventuais políticas adotadas. Em todo caso, há um apelo ao medo associado ao nome de Trump. Por outro lado, o nome de Hillary Clinton aparece em contextos em que as metáforas refletem raiva. Isso sugere que as notícias falsas sobre Hillary Clinton expressam ataques abertos à candidata.

Em contraste, para as notícias confiáveis podemos conjecturar que a grande antecipação e alegria associadas a Hillary Clinton demonstram a expectativa de vitória na corrida eleitoral que existiu entre os maiores veículos jornalísticos que possuem maior credibilidade. Já para Donald Trump a emoção destoante foi a raiva, o que poderia ser explicado pelo Trump ter sido considerado um *outsider* sem apoio da mídia, com exceção de redes jornalísticas específicas. Essa análise é reforçada ao considerarmos que uma das principais fontes de notícias confiáveis do *FakeNewsCorpus* foi The New York Times, um jornal com maiores tendências democratas do que republicanas.

Uma avaliação provisória sobre a análise de emoções em metáforas presentes nos textos sobre Hillary Clinton e Trump nas notícias falsas e confiáveis nos permite inferir que o jornalismo confiável, de forma geral, apoiava Clinton, enquanto que no âmbito das Fake News, a candidata era mais frequentemente rejeitada.

6 Considerações Finais

Neste trabalho, procuramos estudar a utilização de metáforas na composição de conteúdos textuais do domínio jornalístico através de um modelo computacional estado da arte validado com o auxílio de uma especialista da área linguística. Além de apresentar os principais conceitos e definições relacionados ao domínio linguístico da metáfora, apontamos também os conhecimentos, técnicas e ferramentas necessárias para a abordagem computacional do problema.

Conforme as perguntas de pesquisa, buscamos criar experimentos que evidenciassem as características do corpus analisado: *(i) Como podemos descrever cada classe de notícia segundo seu perfil de distribuição de metáforas?* Tipos de notícias podem ser representadas de acordo com a média de metáforas por sentença de texto, considerando a extensão do notícia dado que notícias falsas tendem a ser mais extensas; *(ii) De acordo com as métricas definidas, existem diferenças significativas no perfil de notícias autênticas e fake news?* Ao considerarmos apenas os títulos encontramos uma pequena diferença, estatisticamente relevante, mas ao analisarmos os corpos de texto a diferença entre as duas classes se tornou mais evidente; *(iii) Se foi possível representar o perfil de notícias através das metáforas nelas detectadas, como podemos descrever as emoções relacionadas à estas metáforas? Existe variação entre as classes?* Usando as oito emoções primárias verificamos que existe um equilíbrio na distribuição de emoções entre todos os tipos de notícias; *(iv) Como a intensidade de sentimentos e a subjetividade expressas no texto estão relacionados aos tipos de notícias? Estas intensidades variam de acordo ao uso de metáforas?* Verificamos que para títulos, em relação à subjetividade e intensidade dos sentimentos, não foram encontradas diferenças significativas. Para os corpos de texto, a distribuição das medidas de sentimentos se mostraram semelhantes entre notícias verídicas e fake news. Entretanto, quando a análise é feita para entidades específicas, as mesmas medidas tornam-se significantes; *(v) Como se comportam diferentes tipos de notícias em relação às entidades específicas por elas referenciadas?* As entidades mais referenciadas em cada tipo de notícias tendem a ser iguais, com exceção da falta de Donald Trump entre as entidades mais citadas na classe confiável; *(vi) O padrão encontrado para as respostas acima se mantêm quando analisadas somente notícias referente à entidades específicas? Se não, existe variação entre entidades? Existe uma grande variação de quantidade de metáforas e emoções de acordo com a entidade analisada, além disso, embora as classes anteriormente apresentaram equilíbrio de emoções no geral, ao analisar os resultados para entidades específicas há uma disparidade entre as classes.*

A partir da observação de que metáforas não apenas estão frequentemente presentes nos processos de comunicação humana, como também são componentes estruturantes de nossos processos cognitivos básicos, inclusive enviesando a aquisição de novas informações,

levantamos a hipótese de que metáforas pode ser um dos métodos empregados para manipular a opinião de leitores de notícias. Para verificar nossa hipótese, analisamos as emoções associadas as metáforas encontradas em cada tipo de notícia com o intuito de identificar quais suas variações e como podemos relacioná-las ao objetivo da classe. Para tornar a investigação mais concreta, comparamos resultados quantitativos e qualitativos referentes ao Donald Trump e Hillary Clinton, para os quais diferenças significativas no padrão de metáforas e emoções ficaram evidentes, por exemplo a emoção mais frequentemente associada à metáforas presentes nas notícias falsas referentes a Hillary ter sido a raiva, enquanto para Trump tenha sido o medo. Isso pode ser explorado para obter um indicativo sobre a intencionalidade na divulgação de desinformação não somente de diferentes indivíduos, mas também em relação a temas específicos. Veículos jornalísticos poderiam ser avaliados quanto às metáforas emotivas usadas em notícias sobre, por exemplo, propostas políticas e eventos ocorridos dado que notícias falsas apresentaram uma maior concentração de metáforas por palavras. De maneira geral, os resultados apresentados sugerem a viabilidade de explorar características relacionadas as metáforas usadas em notícias para detectar não apenas viés na divulgação, mas também qual o direcionamento estimulado no leitor.

6.1 Trabalhos Futuros

Estudo subsequentes nesta mesma linha de pesquisa poderiam explorar arquiteturas de detecção de metáforas com redes profundas, além do uso de *word embeddings* diferentes para adaptar o modelo para outros domínios de texto e idiomas. Isso permitiria verificar se as características linguísticas que são exploradas em notícias falsas se repetem em diferentes contextos, ou se os resultados encontrados neste trabalho são limitados ao contexto cultural e linguístico abordado.

Propomos ainda investigar a intencionalidade presente nos textos, de maneira a estabelecer uma melhor compreensão sobre a motivação do autor no uso de metáforas. Caso os resultados positivos obtidos ocorram em diferentes cenários, explorar as características com variação significativa entre as classes de notícias em conjunto aos elementos usados pelos modelos de detecção de *fake news* do estado da arte poderia enriquecer o desempenho de classificação.

Referências

- ALLCOTT, H.; GENTZKOW, M. Social media and fake news in the 2016 election. **Journal of economic perspectives**, v. 31, n. 2, p. 211–36, 2017. Citado na página 31.
- AUER, S. et al. Dbpedia: A nucleus for a web of open data. In: **The semantic web**. [S.l.]: Springer, 2007. p. 722–735. Citado na página 45.
- BELL, G.; HEY, T.; SZALAY, A. Beyond the data deluge. **Science**, American Association for the Advancement of Science, v. 323, n. 5919, p. 1297–1298, 2009. Citado na página 12.
- BENGIO, Y. et al. A neural probabilistic language model. **Journal of machine learning research**, v. 3, n. Feb, p. 1137–1155, 2003. Citado 2 vezes nas páginas 29 e 30.
- BILSKY, M. Ia richards’ theory of metaphor. **Modern Philology**, University of Chicago Press, v. 50, n. 2, p. 130–137, 1952. Citado na página 19.
- BOJANOWSKI, P. et al. Enriching word vectors with subword information. **arXiv preprint arXiv:1607.04606**, 2016. Citado na página 30.
- CARBONELL, J. G. Metaphor: an inescapable phenomenon in natural-language comprehension. **Strategies for natural language processing**, v. 415, 1982. Citado na página 13.
- CHARTERIS-BLACK, J. Britain as a container: Immigration metaphors in the 2005 election campaign. **Discourse & Society**, Sage Publications Sage CA: Thousand Oaks, CA, v. 17, n. 5, p. 563–581, 2006. Citado na página 49.
- CHOWDHURY, G. G. Natural language processing. **Annual review of information science and technology**, Wiley Online Library, v. 37, n. 1, p. 51–89, 2003. Citado na página 12.
- CHUNG, J. et al. Empirical evaluation of gated recurrent neural networks on sequence modeling. **arXiv preprint arXiv:1412.3555**, 2014. Citado na página 26.
- CITRON, F. M.; GOLDBERG, A. E. Metaphorical sentences are more emotionally engaging than their literal counterparts. **Journal of cognitive neuroscience**, MIT Press, v. 26, n. 11, p. 2585–2595, 2014. Citado na página 48.
- EKMAN, P. An argument for basic emotions. **Cognition & emotion**, Taylor & Francis, v. 6, n. 3-4, p. 169–200, 1992. Citado na página 46.
- GERS, F. A.; SCHMIDHUBER, J.; CUMMINS, F. Learning to forget: Continual prediction with lstm. **Neural Computation**, MITP, v. 12, n. 10, p. 2451–2471, 2000. Citado na página 27.
- GLUCKSBERG, S. Metaphors in conversation: How are they understood? why are they used? **Metaphor and Symbol**, Taylor & Francis, v. 4, n. 3, p. 125–143, 1989. Citado na página 13.
- GLUCKSBERG, S. et al. **Understanding figurative language: From metaphor to idioms**. [S.l.]: Oxford University Press on Demand, 2001. Citado na página 17.

- GOODFELLOW, I. et al. **Deep learning**. [S.l.]: MIT press Cambridge, 2016. v. 1. Citado 2 vezes nas páginas 24 e 25.
- GOTTFRIED, J.; SHEARER, E. **News Use Across Social Media Platforms 2016**. [S.l.]: Pew Research Center, 2016. Citado na página 32.
- GRAVES, A.; SCHMIDHUBER, J. Framewise phoneme classification with bidirectional lstm and other neural network architectures. **Neural Networks**, Elsevier, v. 18, n. 5-6, p. 602–610, 2005. Citado na página 26.
- GROUP, P. Mip: A method for identifying metaphorically used words in discourse. **Metaphor and symbol**, Taylor & Francis, v. 22, n. 1, p. 1–39, 2007. Citado na página 35.
- GUESS, A.; NYHAN, B.; REIFLER, J. Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 us presidential campaign. **European Research Council**, 2018. Citado na página 32.
- HAYKIN, S. **Neural networks: a comprehensive foundation**. [S.l.]: Prentice Hall PTR, 1994. Citado 2 vezes nas páginas 22 e 23.
- HOCHREITER, S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. **International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems**, World Scientific, v. 6, n. 02, p. 107–116, 1998. Citado na página 26.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural computation**, MIT Press, v. 9, n. 8, p. 1735–1780, 1997. Citado na página 26.
- JR, E. C. T.; LIM, Z. W.; LING, R. Defining “fake news” a typology of scholarly definitions. **Digital Journalism**, Taylor & Francis, v. 6, n. 2, p. 137–153, 2018. Citado na página 32.
- KRENNMAYR, T. *Metaphor in newspapers*. Amsterdam: Vrije Universiteit, 2011. Citado na página 42.
- KREUZ, R. J.; ROBERTS, R. M. The empirical study of figurative language in literature. **Poetics**, Elsevier, v. 22, n. 1-2, p. 151–169, 1993. Citado na página 19.
- KRISHNAKUMARAN, S.; ZHU, X. Hunting elusive metaphors using lexical resources. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the Workshop on Computational approaches to Figurative Language**. [S.l.], 2007. p. 13–20. Citado na página 34.
- LAKOFF, G.; JOHNSON, M. The metaphorical structure of the human conceptual system. **Cognitive science**, Wiley Online Library, v. 4, n. 2, p. 195–208, 1980. Citado na página 19.
- LAKOFF, G.; JOHNSON, M. **Metaphors we live by**. [S.l.]: University of Chicago press, 2008. Citado 2 vezes nas páginas 14 e 19.
- LANDAUER, T. K.; FOLTZ, P. W.; LAHAM, D. An introduction to latent semantic analysis. **Discourse processes**, Taylor & Francis, v. 25, n. 2-3, p. 259–284, 1998. Citado na página 28.

- LAZER, D. M. et al. The science of fake news. **Science**, American Association for the Advancement of Science, v. 359, n. 6380, p. 1094–1096, 2018. Citado 2 vezes nas páginas 31 e 32.
- LEONG, C. W.; KLEBANOV, B. B.; SHUTOVA, E. **A Report on the 2018 VUA Metaphor Detection Shared Task**. [S.l.], 2018. 56–66 p. Disponível em: <<https://github.com/EducationalTestingService/metaphor>>. Citado 2 vezes nas páginas 32 e 35.
- LIDDY, E. D. Natural language processing. In **Encyclopedia of Library and Information Science, 2nd Ed.** NY. Marcel Decker, Inc, 2001. Citado na página 12.
- LIPTON, Z. C. A critical review of recurrent neural networks for sequence learning. **CoRR**, abs/1506.00019, 2015. Disponível em: <<http://arxiv.org/abs/1506.00019>>. Citado 3 vezes nas páginas 23, 24 e 25.
- MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. An introduction to information retrieval. In: . [S.l.]: Cambridge University Press, 2009. Citado na página 28.
- MENDES, P. N. et al. Dbpedia spotlight: shedding light on the web of documents. In: **ACM. Proceedings of the 7th international conference on semantic systems**. [S.l.], 2011. p. 1–8. Citado na página 45.
- MIKOLOV, T. et al. Efficient estimation of word representations in vector space. **CoRR**, abs/1301.3781, 2013. Disponível em: <<http://arxiv.org/abs/1301.3781>>. Citado na página 30.
- MIKOLOV, T. et al. Distributed representations of words and phrases and their compositionality. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2013. p. 3111–3119. Citado na página 30.
- MILLER, G. A. Images and models, similes and metaphors. In: _____. **Metaphor and Thought**. 2. ed. [S.l.]: Cambridge University Press, 1993. p. 357–400. Citado na página 18.
- MOHAMMAD, S. M.; TURNEY, P. D. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text**. [S.l.], 2010. p. 26–34. Citado na página 41.
- MOHAMMAD, S. M.; TURNEY, P. D. Crowdsourcing a word-emotion association lexicon. v. 29, n. 3, p. 436–465, 2013. Citado na página 41.
- NADEAU, D.; SEKINE, S. A survey of named entity recognition and classification. **Linguisticae Investigationes**, John Benjamins, v. 30, n. 1, p. 3–26, 2007. Citado 2 vezes nas páginas 30 e 31.
- NEUMAN, Y. et al. Metaphor identification in large texts corpora. **PloS one**, Public Library of Science, v. 8, n. 4, p. e62343, 2013. Citado na página 34.
- PENNINGTON, J.; SOCHER, R.; MANNING, C. Glove: Global vectors for word representation. In: **Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)**. [S.l.: s.n.], 2014. p. 1532–1543. Citado na página 30.

- PLUTCHIK, R. **A general psychoevolutionary theory of emotion**. [S.l.]: Emotion: Theory, research, and experience, 1980. Citado na página 46.
- RAU, L. F. Extracting company names from text. In: IEEE. [1991] **Proceedings. The Seventh IEEE Conference on Artificial Intelligence Application**. [S.l.], 1991. v. 1, p. 29–32. Citado na página 30.
- ROBERTS, R. M.; KREUZ, R. J. Why do people use figurative language? **Psychological science**, SAGE Publications Sage CA: Los Angeles, CA, v. 5, n. 3, p. 159–163, 1994. Citado na página 18.
- SANG, E. F.; MEULDER, F. D. Introduction to the conll-2003 shared task: Language-independent named entity recognition. **arXiv preprint cs/0306050**, 2003. Citado na página 31.
- SAPOLSKY, R. This is your brain on metaphors. **The New York Times**, v. 14, n. 11, p. 2010, 2010. Citado na página 48.
- SCHULDER, M.; HOVY, E. Metaphor detection through term relevance. **ACL 2014**, p. 18–26, 2014. Citado 2 vezes nas páginas 12 e 33.
- SCHUSTER, M.; PALIWAL, K. K. Bidirectional recurrent neural networks. **IEEE Transactions on Signal Processing**, IEEE, v. 45, n. 11, p. 2673–2681, 1997. Citado na página 25.
- SHUTOVA, E. Models of metaphor in nlp. In: **Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics**. [S.l.: s.n.], 2010. p. 688–697. Citado 2 vezes nas páginas 13 e 15.
- SHUTOVA, E. V. **Computational approaches to figurative language**. [S.l.], 2011. Citado na página 17.
- SKORCZYNSKA, H.; DEIGNAN, A. Readership and purpose in the choice of economics metaphors. **Metaphor and Symbol**, Routledge, v. 21, n. 2, p. 87–104, 2006. Citado na página 15.
- SMEDT, T. D.; DAELEMANS, W. Pattern for python. **Journal of Machine Learning Research**, v. 13, n. Jun, p. 2063–2067, 2012. Citado na página 52.
- STEEN, G. **A method for linguistic metaphor identification: From MIP to MIPVU**. [S.l.]: John Benjamins Publishing, 2010. v. 14. Citado na página 42.
- STEMLE, E.; ONYSKO, A. Using language learner data for metaphor detection. In: **Proceedings of the Workshop on Figurative Language Processing**. [S.l.: s.n.], 2018. p. 133–138. Citado 6 vezes nas páginas 15, 30, 33, 35, 39 e 44.
- SUNDERMEYER, M.; SCHLÜTER, R.; NEY, H. Lstm neural networks for language modeling. In: **Thirteenth annual conference of the international speech communication association**. [S.l.: s.n.], 2012. Citado na página 26.
- SZPAKOWSKI, M. **Fake News Corpus**. [S.l.]: GitHub, 2018. <<http://github.com/several27/FakeNewsCorpus/>>. [Online; accessed 21-October-2018]. Citado na página 36.

THIBODEAU, P. H.; BORODITSKY, L. Metaphors we think with: The role of metaphor in reasoning. **PLoS ONE**, v. 6, n. 2, p. 11, 2011. Citado 2 vezes nas páginas 15 e 49.

THIBODEAU, P. H.; HENDRICKS, R. K.; BORODITSKY, L. How linguistic metaphor scaffolds reasoning. **Trends in cognitive sciences**, Elsevier, v. 21, n. 11, p. 852–863, 2017. Citado na página 49.

VERLEYSSEN, M.; FRANÇOIS, D. The curse of dimensionality in data mining and time series prediction. In: SPRINGER. **International Work-Conference on Artificial Neural Networks**. [S.l.], 2005. p. 758–770. Citado na página 29.

VOSOUGHI, S.; ROY, D.; ARAL, S. The spread of true and false news online. **Science**, American Association for the Advancement of Science, v. 359, n. 6380, p. 1146–1151, 2018. Citado na página 32.

WILKS, Y. A preferential, pattern-seeking, semantics for natural language inference. **Artificial intelligence**, Elsevier, v. 6, n. 1, p. 53–74, 1975. Citado na página 14.