Fake News - Detecção de Postura

Andressa Gabrielly Macedo Marçal
262878
a262878@dac.unicamp.br

Cicera Vanessa Marques Sampaio Sidrim
230304
c230304@dac.unicamp.br

Renato Marinho Alves
262890
r262890@dac.unicamp.br

***Abstract*—Vivemos a era das redes sociais onde as notícias são disseminadas de modo rápido, possibilitando a democratização do acesso a informação. Por outro lado essa forma de consumo de noticias permite a ampla divulgação das chamadas *fake news*, ou seja noticias de baixa qualidade contendo intencionalmente informações falsas. A alta propagação das *fake news* tem potencial extremamente negativos sobre o individuo e a sociedade, tornando assim pesquisas relacionadas a identificação de notícias falsas emergentes. Com o objetivo de explorar o uso da inteligencia artificial para detecção automatica de *fake news* foi criado o *Fake News Challenge* (FNC-1), um desafio que promove o desenvolvimento de ferramentas para tarefa de classificação de postura, um dos primeiros passos crucial no processo de identificação de notícias falsas. Neste relatório iremos explorar sobre a competição, o conjunto de dados utilizados no desafio e o funcionamento de uma das soluções propostas pelos participantes.**

I. Introdução

As tecnologias sociais facilitam o compartilhamento rápido e em larga escala da informação, permitindo a disseminação e democratização do acesso a informação, porém este ecossistema tem se tornado um local propícios à relativização do conceito de verdade [7], tornando a identificação do que é verdadeiro e falso um dos maiores desafios enfrentados na atualidade por jornalistas e setores de notícias [2].

As *fake news*, assim como são chamada não se trata apenas de uma informação pela metade ou mal apurada, mas de uma informação falsa intencionalmente divulgada, para atingir interesses de indivíduos ou grupos [10] gerando desinformação e causando impactos negativos na sociedade.

A tarefa de avaliar a veracidade de reportagens ainda é muito exigente e complexa, até mesmo para especialistas treinados. O processo de automatizar significativamente partes dos procedimentos utilizados por verificadores humanos para determinar se uma história é real ou uma farsa, chamou atenção considerável em várias comunidades de pesquisa tornando-se um assunto recorrente e bastante pesquisado.

A classificação de uma reportagem como verdadeira ou falsa pode ser dividido em etapas. O primeiro passo útil neste processo envolve estimar a perspectiva (ou posição) relativa de duas partes do texto em relação a um determinado tópico, reivindicação ou problema. Esse estagio é conhecido como detecção de postura.

Em 2017 organizado por um grupo de acadêmicos e colaboradores da indústria do jornalismo, surgiu o *Fake News Challenge* (FNC-1) ação desenvolvida para explorar como as tecnologias de inteligência artificial, particularmente o aprendizado de máquina e o processamento de linguagem natural, podem ser aproveitadas para combater o problema das notícias falsas. O estágio 1 do desafio se concentra na detecção de postura.

Um total de 50 equipes participaram ativamente da competição gerando várias soluções. Posteriormente foram divulgados os conjunto de dados de treinamento e teste utilizados no desafio, para incentivar novos desenvolvimentos.

Continuaremos este relatório da seguinte forma: Na seção 2 discutiremos os conceitos fundamentais e relevante da literatura para entendimento da solução proposta como também alguns trabalhos relacionados. Na seção 3 será abordada a arquitetura utilizada e sua implementação, em seguida na seção 4 discutiremos os experimentos executados, apresentando a base de dados, o sistema de pontuação, a execução e os resultados obtidos, finalizando com a seção 5 com as conclusões e os trabalhos futuros.

II. Fundamentos

A. Redes neurais recorrentes

Ao se trabalhar com textos que nada mais são do que dados sequencias, a ordem como as informações estão dispostas é um fator importante, ou seja, ao se processar um dado textual devemos realizar-lo em sequencia onde para cada palavra nova a ser aprendida devemos analisar os dados processados anteriormente para um entendimento adequado. Para isto utilizamos de redes neurais recorrentes (RNN) (Fig. 1) um tipo de rede neural especializada para processar uma sequência de valores x (1). . . , x (t). [4]

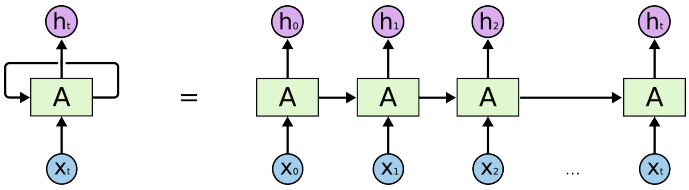


Figure 1. Processamento sequencial na RNN. Fonte: Imagem retirada do blog de Christopher Olah

Estes tipos de redes foram introduzidas na década de 80 e sua capacidade de manter a memória de entradas passadas abriu novos domínios de problemas para redes neurais.

B. RNNs bidirecionais

RNNs Bidirecionais (BiRNNs) são redes recorrentes baseadas na concepção de que o estado y_t (Fig. 2) não depende apenas das informações presentes nos elementos anteriores da

sequência de entrada [1], levando também em consideração o contexto das informações posteriores.

Portanto em uma RNN bidirecional, consideramos 2 sequências separadas, vindas de ordens inversa, possuindo como saídas os resultados gerados do processo de concatenação das sequencias de palavras.

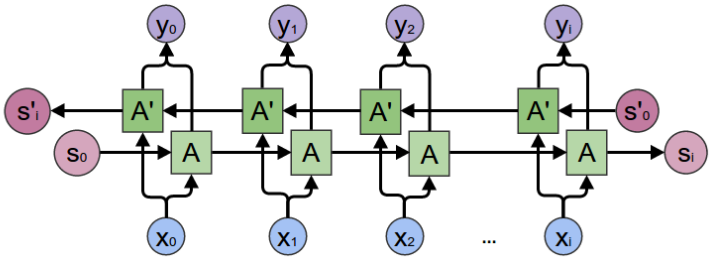


Figure 2. Estrutura geral das redes neurais recorrentes bidirecionais. Fonte: Imagem retirada do blog de Christopher Olah

C. LSTM

Desenvolvidas como umas das soluções para o problema de memória de curto prazo das RNNs anteriores. As redes de memória de longo prazo (LSTM) são um tipo especial de RNN, capaz de aprender dependências de longo prazo. Ela possui mecanismos internos chamados portões que podem regular o fluxo de informações (Fig. 3)

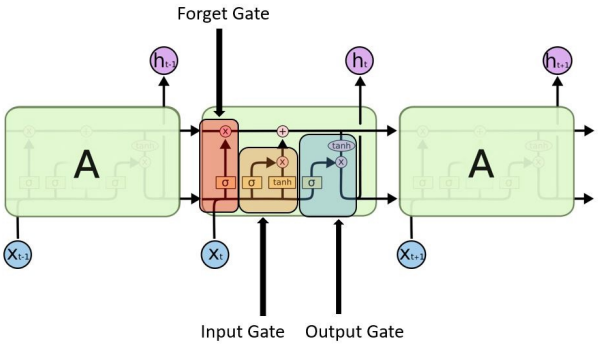
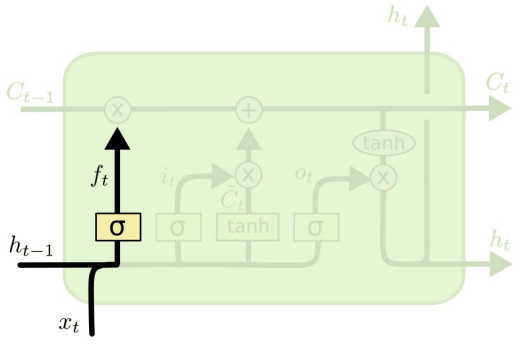


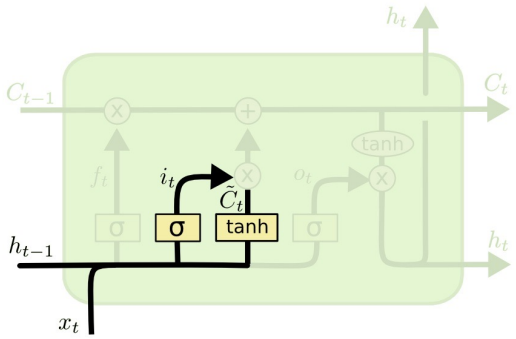
Figure 3. Processamento sequencial no LSTM. Fonte: Imagem retirada do blog de Christopher Olah

Estas portões nada mas são do que módulos de repetição onde cada um destes seguem um processo baseado em 3 etapas (Fig. 4), sendo estas: .

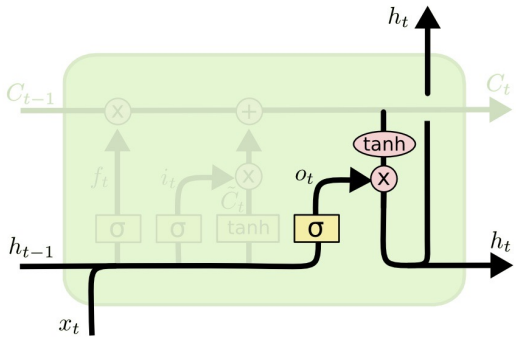
- *Forget Gate:* Decide quanto do passado você deve se lembrar, ou seja quais informações serão omitidas da célula.
- *Input Gate:* Decide quanto desta unidade é adicionada ao estado atual.
- *Output Gate:* Decide qual parte da célula atual chega à saída.



(a) Forget Gate



(b) [Input Gate



(c) [Output Gate

Figure 4. Etapas LSTML. Fonte: Imagem retirada do blog de Christopher Olah

D. Inferência de linguagem natural

A inferência de linguagem natural (NLI) consiste em uma tarefa de processamento de linguagem natural amplamente estudada, que determina se uma afirmação (premissa) semanticamente envolve outra afirmação (hipótese).

Existem dois conjuntos de dados em larga escala, contendo inferências anotadas por humanos bastante conhecidos para execução desta tarefa, são eles o SNLI[1] e o MultiNLI[2].

Um exemplo de dados contido no conjunto SNLI é:

- **Premissa:** Uma mulher que vende varas de bambu conversando com dois homens em uma doca de carregamento.

[1]https://nlp.stanford.edu/projects/snli/
[2]https://www.nyu.edu/projects/bowman/multinli/

- **Hipótese de implicação:** Há pelo menos três pessoas em uma doca de carregamento.
- **Hipótese neutra:** Uma mulher está vendendo varas de bambu para ajudar a sustentar sua família.
- **Hipótese de contradição:** Uma mulher não está recebendo dinheiro por nenhum de seus gravetos.

O corpus do SNLI é uma coleção de 570k pares de frases em inglês [12] e o MultiNLI é uma coleção de 433k pares de frases anotadas com informações de vinculação textual. O corpus MultiNLI é modelado no corpus do SNLI, mas difere no que diz respeito a uma variedade de gêneros de texto falado e escrito e suporta uma avaliação distinta de generalização de gênero [18].

E. Detecção de postura

A detecção de postura pode ser aplicada a diversos contexto textuais, como por exemplo, analise de *twitters* [13], debates *online* [3] e artigos de noticias [5].

No contexto do FNC, a detecção de postura pode ser definida como a tarefa de rotular o relacionamento entre o corpo de texto de uma reportagem e uma manchete, reivindicando especificamente, se o corpo concorda, discorda ou discute da manchete ou se não existe nenhuma relação entre as partes (Fig. 5).

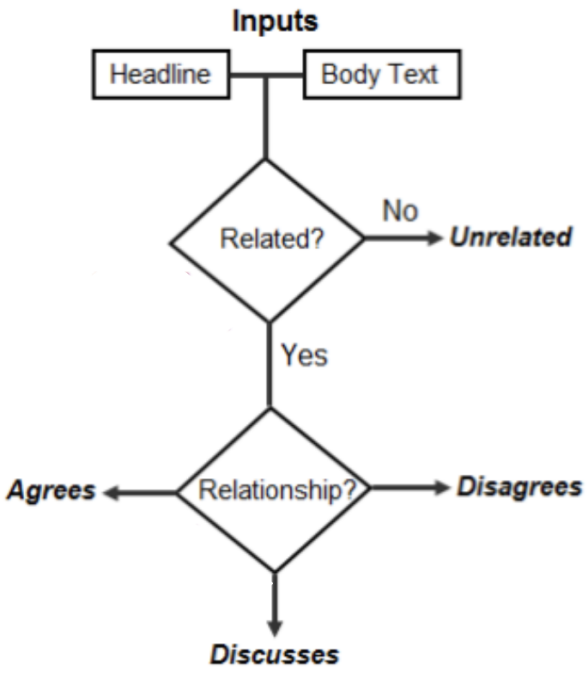


Figure 5. Classificações de Postura do FNC. Fonte: Imagem retirada do site do *Fake News Challenge*

F. Trabalhos Relacionados

O desafio *Fake News Challenge* teve um total de 50 equipes participantes. A primeira posição na competição foi ocupada pela equipe *SOLAT in the SWEN*[3] que utiliza de um conjunto de submodelos baseado em uma média ponderada de 50/50 entre árvores de decisão impulsionadas por gradiente e uma rede neural convolucional (CNN) unidimensional (Fig. 6). O modelo atingiu uma taxa de precisão de 82.02%.

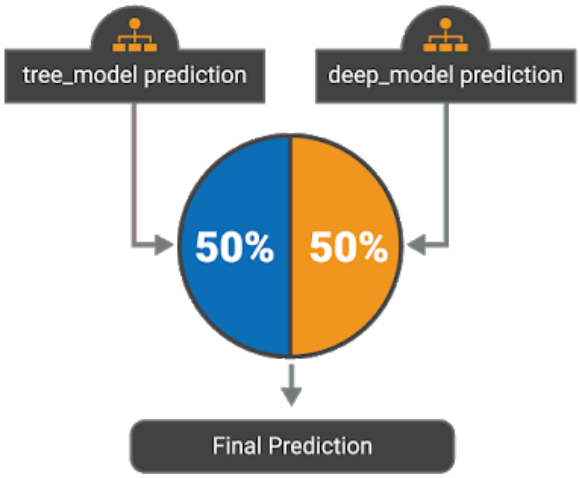


Figure 6. Perceptron multicamada usado no FNC-1 pela equipe Athene. Fonte: Imagem retirada do site da equipe *SOLAT in the SWEN*

O segundo lugar foi concedido a equipe *Athenas*[4] após atingir uma taxa de 81.97% de precisão utilizando um modelo baseado no uso de um perceptron multicamada juntamente com recursos de *Bag of Words* (BoW). (Fig. 7).

Ocupando a terceira posição o time *UCL Machine Reading* [11] atingiu a precisão de 81.72% com a construção de um modelo simples *percepton* multicamada com uma camada oculta e uma camada final de ativação *softmax*. (Fig. 8)

III. Solução Proposta

A. Arquitetura

Utilizamos como base para solução proposta a arquitetura elaborada por pesquisadores da Universidade de Lisboa [2] que por conseguinte se baseou nos modelos de rede sugeridos por Yang et al. [19] e por Nie e Bansal [17]. (Fig. 9).

Nesta arquitetura a manchete é processada através de um codificador de frases (Fig.10 que usa como entrada uma sequência de palavras-chave w_l com $l \in [0, L]$ onde L corresponde ao comprimento da sequência, substituindo cada palavra por uma incorporação *GloVe* [5] pré-treinada. O corpo do texto é codificado utilizando o mesmo procedimento usado para codificação do titulo e a sequencia resultante de vetores de sentença é processada através de uma RNN bidirecional, seguida de um *max-pooling*.

Um terceiro ramo compara o título com as duas primeiras frases do corpo, alavancando o codificador de frases para

[3]https://blog.talosintelligence.com/2017/06/talos-fake-news-challenge.html

[4]https://medium.com/@andre134679/team-athene-on-the-fake-news-challenge-28a5cf5e017b

[5]Algoritmo de aprendizado não supervisionado desenvolvido pela universidade de *Stanford* para gerar incorporações de palavras, agregando uma matriz global de co-ocorrência de palavras e palavras de um corpus

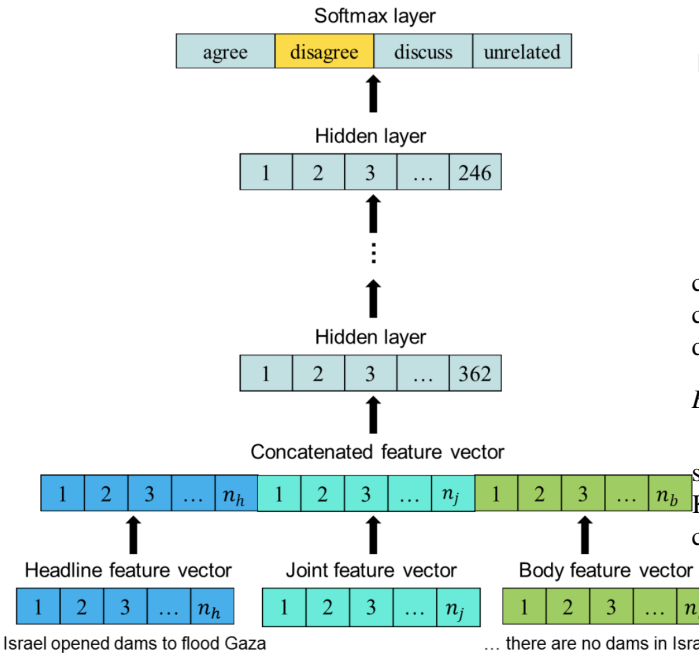


Figure 7. Perceptron multicamada usado no FNC-1 pela equipe Athene. Fonte: Imagem retirada do blog da equipe *Athena*

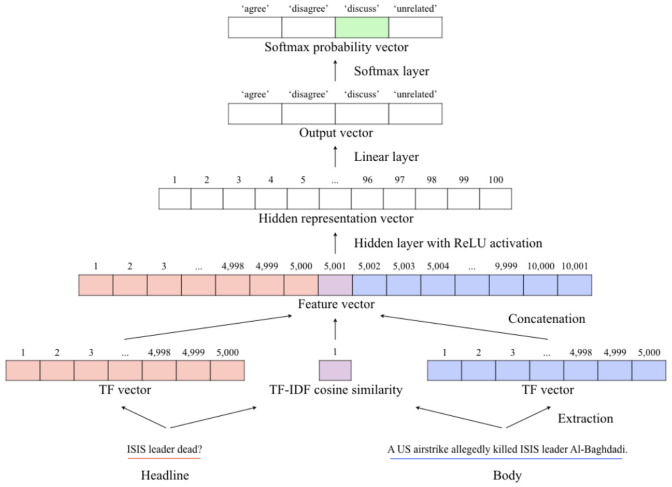


Figure 8. Diagrama da solução proposta pela equipe UCL Machine Reading. Fonte: [11]

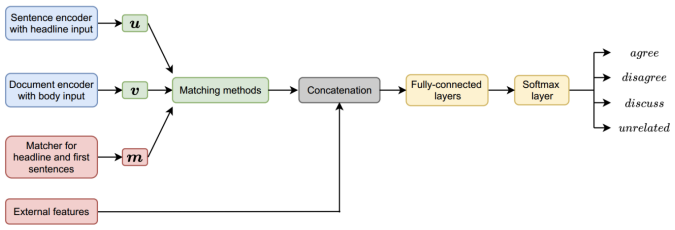


Figure 9. Proposta de solução. Fonte: [5]

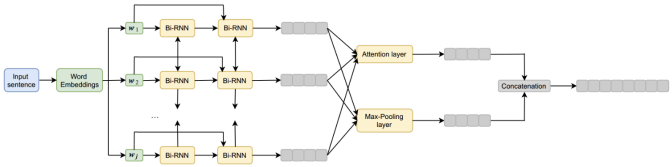


Figure 10. Codificador de frase. Fonte: [5]

construir as representações. Esse combinador, assim como o codificador de frases, foram pré-treinado com os conjuntos de dados SNLI e MultiNLI [5].

B. Implementação

A rede foi implementada utilizando a linguagem Python na sua versão 3 e a biblioteca de rede neural de código aberto Keras 2.2.3 [6]. Algumas bibliotecas utilizadas no processo de codificação foram:

- **NLTK[7]:** Utilizada para processamento simbólico e estatístico de linguagem natural.
- **ROUGE[8]:** Biblioteca para geração de métrica baseadas na analise de textos.
- **Jellyfish[9]:** Biblioteca de aproximação e correspondência fonética de strings.

A rede utilizada é a LSTM bidirecional composta por duas camadas com estados ocultos de 300 dimensões (como as representações são bi-direcionais teremos 600 dimensões ao total) juntamente com duas camadas *feed-forward* com 300 e 600 neurônios respectivamente na sua dimensão temporal, posicionadas antes da camada *softmax*.

Dadas as implementações para RNNs disponíveis no *Keras*, cada sentença (ou seja, manchetes, premissa, conjuntos de dados da NLI e corpo de textos das notícias) é preenchida com zero ou truncada para ter 50 tokens. Todo corpo de artigo de notícias também era preenchido com zero ou truncado para ter até 30 frases.

A rede neural profunda conta com representações para os tokens de palavras com base em encaixes GloVe pré-treinados com 300 dimensões. Palavras fora do vocabulário no conjunto de dados de teste foram representadas pelo GloVe incorporando sua palavra mais semelhante.

A rede ainda possui 3 camadas densas e 5 *dropouts* e uma camada *softmax* na sua saída. Foi definido o *batch*. O processo de otimização é realizado utilizado o otimizador Adam. O código do projeto está disponibilizado no Github.

IV. Experimentos

A. Base de dados

A base de dados fornecida pelo FNC-1 contém um conjunto de 2.587 corpos de texto (documento) e 2.587 títulos (manchetes) de noticias, ou seja, cada artigo de noticia foi

[6]https://keras.io/
[7]https://www.nltk.org/
[8]https://www.aclweb.org/anthology/W04-1013.pdf?forcedefault=true
[9]https://jellyfish.readthedocs.io/en/latest/

resumido em uma manchete que reflete sua posição de todo o corpo de texto, tratando de 300 tópicos diferentes cada um composto de 5 a 20 reportagens.

A tarefa principal de desenvolvimento deste desafio consiste no aprendizado de um classificador $f : (c, m) \mapsto r$ que prevê um dos quatro rótulos $r \in R = Agree, Disagree, Discuss, Unrelated$ para um corpo de texto c e uma manchete m. Os rótulos podem ser definidos da seguinte forma:

- **Agree:** O texto do corpo discute o mesmo tópico do título e se posiciona a favor.
- **Disagree:** O texto do corpo discute o mesmo tópico do título e se posiciona contra.
- **Discuss:** O texto do corpo discute o mesmo tópico do título, mas não se posiciona.
- **Unrelated:** O texto do corpo discute um tópico diferente do título.

Para gerar a classe *Unrelated*, títulos e documentos pertencentes a diferentes tópicos são correspondidos aleatoriamente. Um exemplo destas classificações pode ser visualizado na Fig. 11.

Headline: Hundreds of Palestinians flee floods in Gaza as Israel opens dams	
Agree (AGR)	GAZA CITY (Ma'an) – Hundreds of Palestinians were evacuated from their homes Sunday morning after Israeli authorities opened a number of dams near the border, flooding the Gaza Valley in the wake of a recent severe winter storm. The Gaza Ministry of Interior said in a statement that civil defense services and teams from the Ministry of Public Works had evacuated more than 80 families from both sides of the Gaza Valley (Wadi Gaza) after their homes flooded as water levels reached more than three meters [..]
Discuss (DSC)	Palestinian officials say hundreds of Gazans were forced to evacuate after Israel opened the gates of several dams on the border with the Gaza Strip, and flooded at least 80 households. Israel has denied the claim as "entirely false". [..]
Disagree (DSG)	Israel has rejected allegations by government officials in the Gaza strip that authorities were responsible for released storm waters flooding parts of the besieged area. "The claim is entirely false, and southern Israel does not have any dams," said a statement from the Coordinator of Government Activities in the Territories (COGAT). "Due to the recent rain, streams were flooded throughout the region with no connection to actions taken by the State of Israel." At least 80 Palestinian families have been evacuated after water levels in the Gaza Valley (Wadi Gaza) rose to almost three meters. [..]
Unrelated (UNR)	Apple is continuing to experience 'Hairgate' problems but they may just be a publicity stunt [..]

Figure 11. Trechos de manchete e texto dos corpos dos documentos com as suas respectivas classificações. Fonte: [5]

Os dados estão separados em conjunto de treinamento, contendo 200 tópicos diferentes de pares documente-manchete e em em dados de testes com 100 tópicos, portanto, os dados não são compartilhados entre as duas divisões dos dados. O conjunto de dados de treinamento contém 49.972 instâncias (ou seja, pares de títulos e corpos de texto) classificados com uma determinada postura. Já o conjunto de dados de teste rotulado contém 25.419 instâncias. O tamanho dos dados e a distribuição dos rótulos estão distribuídos conforme mostrado na Fig. 12.

Property	Training Split	Testing Split
Number of instances	49,972	25,413
Number of different news headlines	1,648	893
Number of different news article bodies	1,683	899
Headline average length (tokens)	13	12
Body average length (tokens)	428	396
Percentage of *unrelated* pairs	73.131%	72.203%
Percentage of *discuss* pairs	17.828%	17.566%
Percentage of *agree* pairs	7.360%	7.488%
Percentage of *disagree* pairs	1.681%	2.742%

Figure 12. Distribuição dos dados da base de dados do FNC. Fonte: [2]

B. Sistema de pontuação

Devido a relevância da distinção entre concordância, discordância e discussão ser muito mais expressiva para a detecção de notícias falsas, os organizadores do FNC-1 sugeriram um sistema de pontuação ponderada. Se uma instância de teste não estiver relacionada e o modelo a rotular corretamente, a pontuação será incrementada em 0,25. Caso o modelo escolha o rótulo correto de uma instância de teste relacionada, a pontuação será incrementada em 0,75 adicionais [2]. Resumindo, a equação da métrica de precisão ponderada proposta é a seguinte:

$$Acc_{FNC} = 0.25 \times Acc_{Related,Unrelated} + 0.75 \times Acc_{Agree,Disagree,Discuss}$$

C. Execução

O processo de execução durou em torno de 1 dia para geração das *features* e execução do treino e teste, utilizando 5 épocas e um *batch* de tamanho 64.

Ao total a rede possui 63.094.203 parâmetros, sendo destes 3.094.203 treináveis e 60.000.000 não treináveis.

D. Resultados

Ao final obtivemos uma acurácia de 80%(0.8004) no treino e 83% (0.8369) no teste. Os melhores resultados obtidos foram utilizando o SNLI. O modelo utilizando MultiNLI matched a acurácia foi de 70% (0.7098) e com MultiNLI mismatched também 70% (0.7060).

Os resultados obtidos ao final da execução foram dispostos na tabela I

Table I
RESULTADOS

Época	Loss_train	Acc_train	Loss_test	Acc_test
1	0.6940	0.7031	0.4835	0.8081
2	0.5958	0.7543	0.4433	0.8238
3	0.5535	0.7740	0.4358	0.8309
4	0.5224	0.7888	0.4246	0.8380
5	0.4958	0.8004	0.4204	0.8369

V. CONCLUSÃO E TRABALHOS FUTUROS

Neste relatório apresentamos um método de aprendizado profundo como solução para a problemática de detecção de postura baseado no conjunto de dados, regras e métrica de avaliação do *FakeNews Challenge*(FNC-1). Para isto usufruímos das RNNs bidirecionais juntamente com mecanismos de atenção máxima e de concentração neural. Exploramos do uso de grandes conjuntos de dados propostos para a tarefa de inferência de linguagem natural para realização do processo de pré-treino da rede de codificação de frases.

Muitos dos conceitos aplicados neste projeto foram explanados em sala de aula na disciplina de *Deep Learning* (MO343) como redes neurais recorrentes e LSTM.

Apesar dos resultados satisfatórios, também existem muitas ideias possíveis para trabalhos futuros como forma de melhorias a fim de deixar a solução mais robusta. Como por

exemplo dentro do contexto do NLI propor métodos mais avançados para modelagem de sentenças como em [15], [16], ou métodos que, em vez de usar RNNs para codificação texto, considere apenas cálculos de *feed-forward* e abordagens de atenção como em [9].

REFERENCES

[1] Bispo, Thiago Dias. (2018). Arquitetura LSTM para classificação de discursos de ódio cross-lingual Inglês-PtBR. Dissertação de mestrado em Ciência da Computação – Universidade Federal de Sergipe Sergipe (UFS).

[2] Borges, Luís Martins, Bruno Calado, Pável. (2018). Combining Similarity Features and Deep Representation Learning for Stance Detection in the Context of Checking Fake News.

[3] Dhanya Sridhar, James R. Foulds, Bert Huang, Lise Getoor, and Marilyn A. Walker. 2015. Joint modelsof disagreement and stance in online debate. InProceedings of the 53rd Annual Meeting of the Associ-ation for Computational Linguistics and the 7th International Joint Conference on Natural LanguageProcessing of the Asian Federation of Natural Language Processing (ACL/IJCNLP). Beijing, China,pages 116–125.

[4] Goodfellow, Ian Bengio, Yoshua Courville, Aaron. (2016). Deep Learning. MIT Pres. http://www.deeplearningbook.org

[5] Hanselowski, Andreas PVS, Avinesh Schiller, Benjamin Caspelherr, Felix Chaudhuri, Debanjan Meyer, Christian Gurevych, Iryna. (2018). A Retrospective Analysis of the Fake News Challenge Stance Detection Task.

[6] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. SIGKDD Explor. Newsl. 19, 1 (September 2017), 22-36. DOI: https://doi.org/10.1145/3137597.3137600

[7] Lima, Nathan Viana Vazata, Pedro Antonio Guerra, Andreia Ostermann, Fernanda Cavalcanti, Cláudio. (2019). Educação em Ciências nos Tempos de Pós-Verdade: Reflexões Metafísicas a partir dos Estudos das Ciências de Bruno Latour. Revista Brasileira de Pesquisa em Educação em Ciências. 155-189. 10.28976/1984-2686rbpec2019u155189.

[8] Masood, Razan Aker, Ahmet. (2018). The Fake News Challenge: Stance Detection Using Traditional Machine Learning Approaches.

[9] Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, Jakob Uszkoreit, and ÅAukasz Kaiser. 2018. Universal Transformers.arXiv preprint-arXiv:1807.03819(2018)

[10] Recuero, Raquel, Gruzd, Anatoliy. (2019). Cascatas de Fake News Políticas: um estudo de caso no Twitter. Galáxia (São Paulo), (41), 31-47. Epub May 23, 2019.https://dx.doi.org/10.1590/1982-25542019239035

[11] Riedel, Benjamin Augenstein, Isabelle Spithourakis, Georgios Riedel, Sebastian. (2017). A simple but tough-to-beat baseline for the Fake News Challenge stance detection task.

[12] Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP).

[13] Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif Mohammad, Alan Ritter, and Veselin Stoy-anov. 2015. SemEval-2015 Task 10: Sentiment Analysis in Twitter. InProceedings of the 9th Inter-national Workshop on Semantic Evaluation (SemEval). Denver, CO, USA, pages 451–463.

[14] Vosoughi, Soroush Roy, Deb Aral, Sinan. (2018). The spread of true and false news online. Science. 359. 1146-1151. 10.1126/science.aap9559.

[15] Yi Tay, Luu Anh Tuan, and Siu Cheung Hui. 2018. A Compare-Propagate Architecture with Alignment Factorization for NaturalLanguage Inference. InProceedings of the Conference on Empirical Methods in Natural Language Processing.

[16] Yi Tay, Luu Anh Tuan, and Siu Cheung Hui. 2018. Co-Stack Residual Affinity Networks with Multi-level Attention Refinement forMatching Text Sequences. InProceedings of the Conference on Empirical Methods in Natural Language Processing.

[17] Yixin Nie and Mohit Bansal. 2017. Shortcut-Stacked Sentence Encoders for Multi-Domain Inference. InProceedings of the Workshop onEvaluating Vector Space Representations for NLP.

[18] Williams, Adina Nangia, Nikita Bowman, Samuel. (2018). A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. 1112-1122. 10.18653/v1/N18-1101.

[19] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alexander J Smola, and Eduard H Hovy. 2016. Hierarchical Attention Networks forDocument Classification. InProceedings of the Annual Conference of the North American Chapter of the Association for Computation-alLinguistics.