

Explorando Redes Neurais Profundas para Tarefa de Aceitabilidade Linguística

Henrique Santos¹, Késia P. Alencar², Rogério F. de Sousa², Rafael T. Anchiêta²

¹Universidade Federal do Piauí – Picos (UFPI)

²Instituto Federal do Piauí – Picos (IFPI)

Laboratório de Inteligência Artificial, Robótica e Automação – (LIARA)

{henriques.santod, kesiapereira2003}@gmail.com

{rogerio.sousa, rta}@ifpi.edu.br

Abstract. *Linguistic acceptability is the task of determining whether a sentence is grammatically correct. Although some grammar correction tools exist for Portuguese, they are based on manually defined rules, which is a laborious task. This work investigated deep neural networks and language models for the linguistic acceptability task to develop more robust tools/methods for Portuguese to obtain better results than the existing tools. Recurrent networks, convolutional networks, and the BERTimbau and Albertina language models were explored. These models were trained on a corpus translated from English to Portuguese and evaluated on the Probi corpus. The recurrent and convolutional networks achieved the best results (0.37 f1), being competitive with the LanguageTool tool (0.40 f1).*

Resumo. *Aceitabilidade linguística é a tarefa de determinar se uma sentença está gramaticalmente correta. Apesar de existir algumas ferramentas de correção gramatical para o português, elas são baseadas em regras manualmente definidas, o que é uma tarefa laboriosa. Neste trabalho, investigaram-se redes neurais profundas e modelos língua para a tarefa de aceitabilidade linguística a fim de desenvolver ferramentas/métodos mais robustos para o português que obtenham resultados melhores do que as ferramentas existentes. Explorou-se redes recorrentes, redes convolucionais e os modelos de língua BERTimbau e Albertina. Esses modelos foram treinados em um corpus traduzido do inglês para o português e avaliados no corpus Probi. As redes recorrentes e convolucionais atingiram os melhores resultados (0,37 f1), sendo competitivas com a ferramenta LanguageTool (0,40 f1).*

1. Introdução

Aceitabilidade linguística é a tarefa de determinar se uma sentença está gramaticalmente correta. Essa tarefa vem do campo da linguística generativa [Klezi et al. 2022] que se baseia em julgamentos intuitivos de falantes nativos sobre se uma sentença é aceitável ou não [T Schütze 2016]. Essa tarefa possui diversas aplicações na área de Processamento de Línguas Naturais (PLN), por exemplo: analisar a robustez de modelos de língua [Yin et al. 2020] e verificar se tais modelos adquirem conhecimentos gramaticais [Zhang et al. 2021, Choshen et al. 2022]. Além disso, aceitabilidade linguística tem

sido utilizada para avaliar a correção gramatical e a semântica de métodos de geração automática de textos [Bakshi et al. 2021, Batra et al. 2021].

Embora aceitabilidade linguística seja uma área de estudo com aplicações recentes focadas, principalmente, em modelos de língua, ela é pouco explorada na língua portuguesa. Acredita-se que seja devido à falta de um *corpus* manualmente anotado. Apesar da falta desse recurso, existem algumas ferramentas que fazem correção ortográfica e gramatical na língua portuguesa, por exemplo: CoGrOO [Kinoshita et al. 2007] e LanguageTool¹. Essas ferramentas, fazem as correções em um texto através de regras linguísticas manualmente projetadas. Métodos baseados em regras, embora apresentem bons resultados, são dependentes de especialistas para a construção de regras. Ademais, a criação de regras denota muito esforço.

Com o objetivo de diminuir o esforço na criação de regras linguísticas, neste trabalho exploraram-se métodos de aprendizado profundo (*deep learning*) e modelos de língua para a identificação automática de sentenças aceitáveis. Esses métodos mapeiam uma entrada para a saída desejada de maneira automática através de uma grande quantidade de dados sem a necessidade de desenvolver regras manuais.

A fim de treinar os modelos para a tarefa de aceitabilidade linguística, utilizou-se a versão traduzida do *corpus* CoLA (*Corpus of Linguistic Acceptability*) [Warstadt et al. 2019] para o português. O *corpus* CoLA é um conjunto de 1.657 sentenças em inglês extraídas da literatura linguística e rotuladas como gramatical ou agramatical. A versão portuguesa² do *corpus* CoLA foi traduzida automaticamente utilizando ferramenta *Google Cloud Translation*. Após o treinamento dos modelos, eles foram avaliados no *corpus* Probi [Martins 2002] que é um recurso composto por 11.625 sentenças anotadas manualmente. Do total de sentenças, 2.616 são anotadas como agramatical, ou seja, possuem algum tipo de erro. Obteve-se o melhor resultado com uma rede convolucional, atingido 0,37 de medida F1 que é um valor competitivo com a ferramenta LanguageTool que atingiu 0,40 na mesma métrica.

O restante do artigo está organizado da seguinte forma. Seção 2 introduz brevemente os trabalhos relacionados. Na Seção 3, é detalhado os métodos desenvolvidos. Seção 4 apresenta os resultados e discussões. Por fim, Seção 5 conclui o artigo indicando futuras direções.

2. Trabalhos Relacionados

O ReGra foi o primeiro revisor gramatical para a língua portuguesa [Nunes and Jr. 2000]. Ele foi desenvolvido através de regras manualmente definidas. Posteriormente, o revisor foi incorporado ao MS Word e ao processador de texto REDADOR da Itaotec.

CoGrOO (Corretor Gramatical para OpenOffice) [Kinoshita et al. 2007] é uma ferramenta de correção gramatical para a língua portuguesa. Assim como o ReGra, o corretor gramatical foi desenvolvido através de regras gramaticais manualmente definidas.

LanguageTool é ferramenta de correção gramatical disponível para vários idiomas, como: inglês, alemão, espanhol, holandês, português, entre outros. Da mesma forma

¹<https://languagetool.org/>

²<https://github.com/ju-resplande/PLUE>

que as ferramentas anteriores, o módulo de correção gramatical para língua portuguesa é baseado em regras.

Percebe-se que as principais ferramentas de correção gramatical para o português são baseados em regras linguísticas. Nossa abordagem investiga métodos baseados em redes neurais profundas e modelos de língua como uma alternativa aos métodos tradicionais baseados em regras. No que segue, detalha-se os métodos desenvolvidos.

3. Métodos Desenvolvidos

A fim de desenvolver um método de aceitabilidade linguística para a língua portuguesa, investigaram-se as Redes Neurais Convolucionais (CNN), Redes Neurais Recorrentes (RNN) e os modelos de língua BERTimbau [Souza et al. 2020] e Albertina [Rodrigues et al. 2023]. Esses modelos foram treinados ou ajustados no *corpus* CoLA traduzido para a língua portuguesa e testados no *corpus* Probi que foi criado para avaliar o revisor gramatical ReGra. Analisou-se diversas configurações para os modelos supracitados. Os melhores parâmetros para a RNN e CNN são apresentados nas Tabelas 1 e 2, respectivamente.

Em ambas as redes neurais, utilizou-se as embeddings pré-treinadas do GloVe de 300 dimensões [Hartmann et al. 2017]. Para a RNN, adotou-se uma rede *Bidirectional Long Short-Term Memory* (BiLSTM). Para a CNN, utilizou-se três camadas de convolução e cem filtros com dimensões de três, quatro e cinco. As final da RNN e CNN, implementou-se uma camada linear para realizar a classificação.

Tabela 1. Parâmetros rede RNN.

| Parâmetro | Valor |
|--------------|-------|
| Embeddings | 300 |
| Dropout | 0,5 |
| Bidirecional | sim |
| Otimizador | Adam |
| Erro | BCE |
| Épocas | 2 |

Tabela 2. Parâmetros rede CNN.

| Parâmetro | Valor |
|---------------------|-----------|
| Embeddings | 300 |
| Dropout | 0,1 |
| Convolução | 3 camadas |
| Filtros# | 100 |
| Tamanho dos filtros | [3, 4, 5] |
| Otimizador | Adam |
| Erro | BCE |
| Épocas | 2 |

Para o modelo de língua Albertina, investigou-se algumas variações de parâmetros. No entanto, os resultados sempre foram iguais. Na Tabela 3, são apresentados os parâmetros desse modelo. Diferente do Albertina, os resultados do BERTimbau variaram de acordo com o parâmetro definido. Na Tabela 4, são apresentados os parâmetros que apresentaram melhores resultados.

O código fonte para o treinamento e ajuste fino dos modelos está disponível em <https://github.com/liara-ifpi/aceitabilidade>. Na seção seguinte, os resultados dos modelos desenvolvidos são apresentados.

4. Resultados e Discussão

A partir dos modelos implementados, realizou-se uma comparação entre eles. Além disso, também comparou-se os resultados com as ferramentas CoGrOO e LanguageTool. Na

Tabela 3. Parâmetros Albertina.

| Parâmetro | Valor |
|--------------|--------------------|
| Dropout | 0,5 |
| Batch size | 8 |
| Otimizador | AdamW |
| Aprendizagem | 4×10^{-5} |
| Erro | CrosEntropy |
| Épocas | 6 |

Tabela 4. Parâmetros BERTimbau.

| Parâmetro | Valor |
|--------------|--------------------|
| Batch size | 8 |
| Otimizador | AdamW |
| Aprendizagem | 4×10^{-5} |
| Erro | CrossEntropy |
| Épocas | 2 |

Tabela 5, são apresentados os resultados obtidos.

Tabela 5. Melhores resultados alcançados.

| Modelo/Ferramenta | Precisão | Cobertura | F1 |
|-------------------|-------------|-------------|-------------|
| CoGrOO | 0,59 | 0,19 | 0,28 |
| LanguageTool | 0,56 | 0,32 | 0,40 |
| Albertina | 0,00 | 0,00 | 0,00 |
| BERTimbau | 0,36 | 0,16 | 0,22 |
| RNN | 0,24 | 0,78 | 0,37 |
| CNN | 0,23 | 1,00 | 0,37 |

A partir dos resultados, pode-se ver que o modelo Albertina teve o pior resultado. O Albertina é um modelo de língua para o português recente, que nas avaliações realizadas por seus desenvolvedores atingiu resultados melhores que o BERTimbau. Em trabalhos futuros será investigado porque nesta tarefa os resultados foram muito inferiores ao BERTimbau. O CoGrOO e LanguageTool atingiram os melhores resultados em precisão e f1, respectivamente. Nossos melhores modelos foram a RNN e a CNN, ambas alcançaram 0,37 na medida f1. Além disso, a CNN atingiu 1,00 na cobertura.

Os resultados alcançados tanto pela RNN quanto pela CNN são, de certa forma, promissores, pois os modelos foram treinados em um *corpus* que foi traduzido de forma automática para o português. Isso pode justificar a baixa precisão dos modelos, ou seja, uma alta taxa de falsos positivos. Um trabalho futuro é investigar a qualidade da tradução automática nesse *corpus*.

5. Conclusão

Neste trabalho, explorou-se redes neurais profundas e modelos de língua para a tarefa de aceitabilidade linguística. Apesar de existir algumas ferramentas de correção gramatical para o português, elas são baseadas de regras manualmente definidas. Este artigo buscou uma forma alternativa para identificar se uma sentença é aceitável, ou seja, se está gramaticalmente correta. Os melhores resultados obtidos foram com uma rede BiLSTM CNN, atingindo um resultado competitivo com a ferramenta LanguageTool e superando a ferramenta CoGrOO.

Como trabalho futuro, além de explorar mais o modelo de língua Albertina que obteve os piores resultados e investigar o *corpus* CoLA traduzido para o português, pretende-

se realizar uma detalhada análise de erros a fim de descobrir onde os modelos erraram, desenvolver um método que possa capturar os pontos fortes do BERTimbau e da CNN para melhorar os resultados e uma ferramenta que online para uso dos modelos desenvolvidos.

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq pelo apoio financeiro.

Referências

- Bakshi, S., Batra, S., Heidari, P., Arun, A., Jain, S., and White, M. (2021). Structure-to-text generation with self-training, acceptability classifiers and context-conditioning for the GEM shared task. In *Proceedings of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2021)*, pages 136–147, Online. Association for Computational Linguistics.
- Batra, S., Jain, S., Heidari, P., Arun, A., Youngs, C., Li, X., Donmez, P., Mei, S., Kuo, S., Bhardwaj, V., Kumar, A., and White, M. (2021). Building adaptive acceptability classifiers for neural NLG. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 682–697, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Choshen, L., Hacohen, G., Weinshall, D., and Abend, O. (2022). The grammar-learning trajectories of neural language models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8281–8297, Dublin, Ireland. Association for Computational Linguistics.
- Hartmann, N., Fonseca, E., Shulby, C., Treviso, M., Silva, J., and Aluísio, S. (2017). Portuguese word embeddings: Evaluating on word analogies and natural language tasks. In *Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology*, pages 122–131, Uberlândia, Brazil. Sociedade Brasileira de Computação.
- Kinoshita, J., Salvador, L., Menezes, C., and Silva, W. (2007). Cogroo - an openoffice grammar checker. In *Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007)*, pages 525–530, Rio de Janeiro, Brazil. IEEE.
- Klezl, J., Mohammed, Y. A., and Volodina, E. (2022). Exploring linguistic acceptability in Swedish learners’ language. In *Proceedings of the 11th Workshop on NLP for Computer Assisted Language Learning*, pages 84–94, Louvain-la-Neuve, Belgium. LiU Electronic Press.
- Martins, R. (2002). Probi: um corpus de teste para o revisor gramatical regra. Technical report, ICMC-USP.
- Nunes, M. d. G. V. and Jr., O. (2000). O processo de desenvolvimento do revisor gramatical regra. In *XXVII Seminário Integrado de Software e Hardware*, pages 1–15, Curitiba, Brasil. SBC.
- Rodrigues, J., Gomes, L., Silva, J., Branco, A., Santos, R., Cardoso, H. L., and Osório, T. (2023). Advancing neural encoding of portuguese with transformer albertina pt-*.

- Souza, F., Nogueira, R., and Lotufo, R. (2020). Bertimbau: pretrained bert models for brazilian portuguese. In *9th Brazilian Conference on Intelligent Systems*, pages 403–417, Rio Grande, Brazil. Springer.
- T Schütze, C. (2016). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Language Science Press.
- Warstadt, A., Singh, A., and Bowman, S. R. (2019). Neural network acceptability judgments. *Transactions of the Association for Computational Linguistics*, 7:625–641.
- Yin, F., Long, Q., Meng, T., and Chang, K.-W. (2020). On the robustness of language encoders against grammatical errors. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3386–3403, Online. Association for Computational Linguistics.
- Zhang, Y., Warstadt, A., Li, X., and Bowman, S. R. (2021). When do you need billions of words of pretraining data? In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1112–1125, Online. Association for Computational Linguistics.