

Uma Abordagem para Mineração de Argumentos em Redações do Português Brasileiro

Jonathan Nau
Universidade do Vale do Itajaí
Itajaí - Santa Catarina - Brasil
jonathan_nau@live.com

Aluizio Haendchen Filho
Universidade do Vale do Itajaí
Itajaí - Santa Catarina - Brasil
aluizio.h.filho@gmail.com

Rudimar Luís Scaranto Dazzi
Universidade do Vale do Itajaí
Itajaí - Santa Catarina - Brasil
rudimar@univali.br

ABSTRACT

Argument mining consists of extracting the argumentative structure of a text. The challenge in finding the argumentative structure lies in: identifying the components of the argument and the relationships that occur between them. Approaches that propose to solve both challenges of argument mining together are known as end-to-end methods. In the literature, some papers were found that perform argument mining in essays, but no papers were found for Brazilian Portuguese. Therefore, in this paper, we propose an end-to-end approach for argument mining of Brazilian Portuguese essays in the ENEM model.

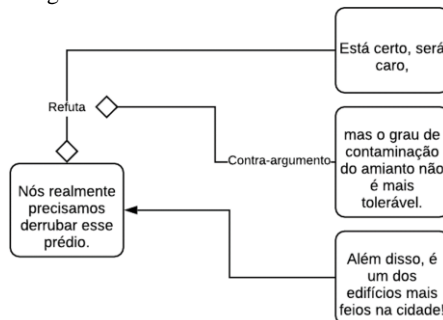
KEYWORDS

Argument Mining, Natural Language Processing, Discourse Analysis

1 Introdução

Conforme Stede e Schneider [6], a mineração de argumentos é uma técnica do Processamento de Linguagem Natural (PLN) útil para o entendimento do discurso. Ela visa identificar a estrutura argumentativa dos textos. É um trabalho complexo em vários aspectos, pois envolve problemas relacionados com coerência textual, coesão, e estrutura do argumento.

Segundo Walton [8], na literatura existem diferenças na teoria da argumentação sobre como definir um argumento. A definição mínima diz que, um argumento é um conjunto de proposições, composto de três partes: (i) uma conclusão; (ii) um conjunto de premissas; e (iii) uma inferência das premissas para a conclusão. Além disso, um determinado argumento pode ser suportado ou refutado por outros argumentos. A Figura 1 apresenta um exemplo da estrutura argumentativa de um texto.



O desafio em minerar argumentos está em identificar os componentes do argumento e as relações que ocorrem entre eles. As abordagens que propõem resolver os desafios da mineração de argumentos, de forma conjunta, são conhecidas como abordagens ponta a ponta. Normalmente, abordagens ponta a ponta são compostas por sete tarefas [6]: (i) identificar o texto argumentativo; (ii) segmentar o texto em unidades discursivas argumentativas; (iii) identificar a tese; (iv) identificar a função dos componentes argumentativos; (v) identificar as relações entre os componentes argumentativos; (vi) construir a representação estrutural geral; e (vii) identificar o tipo e a qualidade da argumentação.

O objetivo deste trabalho é apresentar uma abordagem ponta a ponta para a mineração de argumentos em redações dissertativas-argumentativas do português brasileiro no modelo ENEM [1]. Inicialmente, foram identificadas na literatura abordagens ponta a ponta para a mineração de argumentos em redações. Na segunda etapa, foi selecionada a abordagem com o melhor resultado. Por fim, a melhor abordagem foi adaptada para as redações dissertativas-argumentativas do português brasileiro.

2 Trabalhos Relacionados

Na literatura são encontrados alguns trabalhos para a mineração de argumentos [3, 7]. Entretanto, todos os trabalhos encontrados apresentam soluções apenas para a língua inglesa. O português brasileiro carece de corpora específicos de redações anotadas com a estrutura da argumentação no nível do discurso.

Para a abordagem ponta a ponta, foram selecionados dois trabalhos: (i) *End-to-End Argumentation Mining in Student Essays* [4]; e (ii) *Parsing Argumentation Structures in Persuasive Essays* [5]. Ambas utilizaram a métrica Valor-F para aferir o resultado dos classificadores. A métrica Valor-F utiliza a média harmônica entre os indicadores de precisão e recall, e é tanto melhor quanto mais próxima de 1.

Os autores Ng e Persing [4] apresentaram os primeiros resultados sobre argumentação ponta a ponta em redações persuasivas usando uma abordagem de pipeline. Nesta abordagem, a média das tarefas apresentou Valor-F de 0,30. Stab e Gurevych [5] apresentam uma abordagem que identifica os limites dos componentes, realiza a classificação, e relaciona os componentes. Esta abordagem apresentou Valor-F de 0,762 na média das tarefas.

3 Abordagem Proposta

A solução proposta neste trabalho apresenta uma abordagem ponta a ponta para a mineração de argumentos. A definição da extração de características e da modelagem da programação linear inteira será conforme desenvolvido por Stab e Gurevych [5], com adaptações para o português, caso necessário.

A metodologia compreende as seguintes tarefas: (i) separar as partes argumentativas das não argumentativas; (ii) classificar os tipos de componentes argumentativos; e (iii) identificar as relações entre os componentes.

A Figura 2 apresenta a estrutura da abordagem para mineração de argumentos. A redação é o arquivo de entrada para o processamento do texto. A redação é segmentada em sentenças, e a seguir, de cada sentença é extraído um conjunto de características. As características alimentam os dois classificadores, que realizam a classificação do componente e identificação a relação. O classificador de relação recebe ainda como entrada apenas os componentes que foram classificados como argumentativos. Por fim, a programação linear é aplicada na tentativa de otimizar os resultados. Para isto, são criados um conjunto de restrições na formação da estrutura argumentativa, como, por exemplo, um componente não pode ter duas saídas.

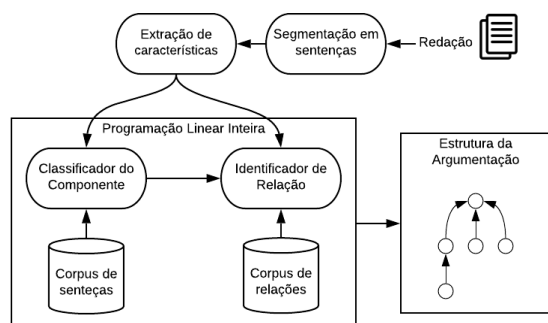


Figura 2: Abordagem para mineração de argumentos

Para o treinamento e validação da abordagem, foi criado um corpus com 50 redações, extraídas do portal Brasil Escola¹. As redações foram anotadas com a estrutura argumentativa por especialistas em linguística do Programa de Pós-Graduação em Educação (PPGE) da UNIVALI. A tarefa de anotação consistiu em dois especialistas anotarem de forma individual as redações, e um terceiro especialista julgar as divergências encontradas entre os dois primeiros especialistas. O índice de concordância para a anotação dos componentes foi medido pelo coeficiente de Krippendorff [2], e as anotações apresentaram índice de 0,92.

A tarefa de segmentação está sendo implementada com a biblioteca do NLTK². Para a extração de características. Foram utilizadas as bibliotecas do CoGrOO³ e spaCy⁴. Para implementar os dois classificadores, serão utilizados o Support Vector Machine

¹ <https://vestibular.brasilecola.uol.com.br/banco-de-redacoes>

² <http://www.nltk.org/>

³ <http://cogroo.org/>

⁴ <https://spacy.io/>

(SVM) com kernel polinomial. A programação linear será implementada com o algoritmo PuLP⁵.

4 Teste Preliminar

Um teste preliminar do classificador do componente foi executado com 659 sentenças, utilizando o algoritmo SVM. Destas sentenças, 80% (527) foram utilizadas para treinamento e os outros 20% (132) para validação. A Tabela 1 apresenta os resultados obtidos pelo algoritmo.

Tabela 1: Resultado do teste preliminar

Métrica	Valor
Valor-F	0,553
Precisão	0,601
Recall	0,568

5 Considerações Finais

Os resultados do teste preliminar mostram que houve um índice de acurácia que ficou com uma precisão intermediária entre os dois melhores resultados encontrados na literatura. Por outro lado, os resultados apontam que ainda existe um espaço para melhoria dos resultados.

As principais contribuições preliminares são: (i) criação de um corpus público de redações anotadas com a estrutura argumentativa para o português brasileiro; e (ii) uma abordagem para a mineração de argumentos ponto a ponto adaptada para o português brasileiro. Ambas contribuições são inéditas para a análise do entendimento do discurso com o foco em redações dissertativas-argumentativas.

REFERÊNCIAS

- [1] Brasil. Redação 2018: Cartilha do Participante: 2018. http://download.inep.gov.br/educacao_basica/enem/guia_participante/2018/manual_de_redacao_do_enem_2018.pdf. Acesso em: 2019-05-10.
- [2] Krippendorff, K. 2004. Measuring the Reliability of Qualitative Text Analysis Data. *Quality & Quantity*, 38, 6 (Dec. 2004), 787–800. DOI:<https://doi.org/10.1007/s11135-004-8107-7>.
- [3] Nguyen, H. and Litman, D. 2016. Improving Argument Mining in Student Essays by Learning and Exploiting Argument Indicators versus Essay Topics. *International Florida Artificial Intelligence Research Society Conference - FLAIRS*, 29., (Key Largo, 2016), 485–490.
- [4] Persing, I. and Ng, V. 2016. End-to-End Argumentation Mining in Student Essays. *Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - NAACL-HLT*, (San Diego, 2016), 1384–1394.
- [5] Stab, C. and Gurevych, I. 2017. Parsing Argumentation Structures in Persuasive Essays. *Computational Linguistics*, 43, 3 (Sep. 2017), 619–659. DOI:https://doi.org/10.1162/COLL_a_00295.
- [6] Stede, M. and Schneider, J. 2018. Argumentation Mining. *Synthesis Lectures on Human Language Technologies*, 11, 2 (2018), 1–191. DOI:<https://doi.org/10.2200/S00883ED1V01Y201811HLT040>.
- [7] Tahir, S.B. ud D. 2017. Comparative Analysis of Supervised Learning Approaches for Argument Identification. *IEEE International Multi-topic Conference - INMIC*, 20., (Lahore, 2017), 1–5.
- [8] Walton, D. 2009. *Argumentation in Artificial Intelligence*. *Argumentation in Artificial Intelligence*. G. Simari and I. Rahwan, eds. Springer US.

⁵ <https://pythonhosted.org/PuLP/>