

Uma Abordagem para Avaliar Aderência ao Tema e Estrutura Dissertativa-Argumentativa de Redações

Jonathan Nau, Vinicius Cavaco, Aluizio Haendchen Filho

Núcleo de Inteligência Artificial e Sistemas Inteligentes (NIASI)
Centro Universitário de Brusque (UNIFEBE) – Brusque, SC – Brasil

{jonathan.nau, vinicavaco3, aluizio.h.filho}@gmail.com

Abstract. *This paper presents an approach for automatic evaluation of the argumentative structure and adherence to the theme of essays. The approach can be integrated into a solution that addresses the other competencies involved in the correction of essays. Promising results were achieved in the initial phase of the work on indicators of mean error and Pearson correlation.*

1. Introdução

Dados recentes do ENEM de 2015, mostram que mais de 10 mil corretores foram contratados e o custo estimado da correção por redação foi R\$15,88. Esse valor engloba aparatos físicos e tecnológicos, capacitação de corretores, serviço de correção e outros custos relacionados ao processo de correção [Globo.com 2016]. Considerando o total de 6,54 milhões de redações submetidas em 2015 [Globo.com 2016], pode-se estimar um custo total de mais de 100 milhões de reais com a correção de redações.

O texto dissertativo-argumentativo se organiza na defesa de um ponto de vista sobre determinado assunto. “Daí sua dupla natureza: é argumentativo porque defende uma tese, uma opinião, e é dissertativo porque são utilizadas explicações para justificá-la” [INEP 2017]. Além disso, é avaliada a aderência do texto ao tema. O objetivo deste trabalho é apresentar uma proposta para automatizar a correção da estrutura dissertativa-argumentativa e pertinência ao tema de redações. A abordagem trata a Competência 2 do ENEM, podendo ser integrada a soluções que tratem as outras competências.

2. Abordagem Proposta

A solução proposta foi desenvolvida de forma iterativa, em cinco etapas:

1. **Criação do corpus:** utilizou-se o corpus de redações obtido do banco de redações do site UOL Educação [UOL Educação 207]. É um corpus público com 2123 redações comentadas e nota atribuída por especialistas em correção do ENEM.
2. **Criação e seleção de *features*:** o sistema incorpora *features* que medem aspectos que codificam conhecimento linguístico relevante para a tarefa em questão, e se mostraram importantes durante a fase de treinamento.
3. **Treinamento do algoritmo:** para a atribuição das notas, utilizamos o algoritmo de regressão linear múltipla, cuja performance apontou resultados melhores do que algoritmos de classificação e redes neurais testados.
4. **Análise dos resultados:** utilizamos duas métricas para medir os resultados: (i) erro absoluto médio; e (ii) correlação de Pearson. Para medir a correlação, usamos a validação cruzada, mais especificamente o método *k-fold*, onde *k* é igual a 10.

5. **Ajustes:** para melhorar os resultados, são realizados experimentos na ferramenta RapidMiner, manipulando parâmetros e adicionado operadores para gerar pesos.

3. Criação e Seleção das *Features*

Para o treinamento do algoritmo, foram consideradas 68 *features*. Durante a fase de treinamento, dessas 68 *features*, foram selecionadas 24 que, combinadas alcançaram os melhores resultados de erro médio e correlação. Elas foram agregadas em macro-*features*, e alinhadas com aspectos específicos de construção de escrita, nos seguintes grupos:

(1) Estatísticas: comumente usadas para avaliar textos dissertativos [Amorin e Veloso, 2017] [Burrows et al 2015] [Dikli 2006]: quantidade de caracteres, quantidade de palavras, tamanho médio dos parágrafos; quantidade de palavras únicas; repetição média de palavras, tamanho médio das palavras, quantidade de vírgulas, quantidade de pontos, tamanho médio das sentenças, tamanho médio dos parágrafos, quantidade de erros ortográficos e gramaticais, quantidade de substantivos e quantidade de adjetivos.

(2) Dissertativas-Argumentativas: obtidas com a formação de pacotes de palavras que representam ideias de criação de hipóteses e argumentação, extraídas do Dicionário Analógico da Língua Portuguesa [Azevedo 2010]: relações abstratas de causa e efeito, relações abstratas de comparação, relações abstratas quantitativas simples, e marcadores discursivos relacionados a objetivos.

(3) Pertinência ao tema: obtidas com *features* que medem o grau de similaridade semântica; o principal objetivo é verificar se o texto está ou não aderente ao tema. Para avaliar o índice de similaridade, foi utilizado o algoritmo LSA (*Latent Semantic Analysis*), proposto por Landauer e Dutnais (1997) e que consiste em uma técnica estatístico-matemática de abstração de conhecimento a partir de um corpus de textos. As seguintes *features* foram consideradas: grau de similaridade entre a introdução e o texto motivador, grau de similaridade entre o(s) parágrafo(s) intermediário(s) e o texto motivador, e quantidade de quebras de linearidade relacionadas com o tamanho do texto.

(4) Bag de palavras: derivadas da análise do corpus de redações, o bag foi constituído através de análise de n-gramas. Foram considerados os seguintes pacotes de palavras: (i) unigramas e bigramas de palavras cujas ocorrências máximas comportam-se de forma crescente ou decrescente no intervalo de notas de 0 até 2 (mesmo critério ou escala do ENEM); (ii) unigramas e bigramas de palavras que ocorrem em alta escala nas redações que obtiveram nota mínima (zero); e (iii) unigramas e bigramas de palavras que ocorreram em grande número nas redações que obtiveram nota máxima (dois).

Foram utilizados algoritmos de Redes Neurais (*Deep Learning*) com 10 camadas ocultas e Regressão Linear Múltipla com *kernel* não linear. Os resultados finais obtidos foram similares, com pequena vantagem para o método de Regressão Linear Múltipla.

4. Resultados

A Tabela 1 apresenta os resultados obtidos por cada grupo de *features*, e o resultado global, considerando todos. É possível observar que os resultados melhoram após os ajustes efetuados com a ferramenta RapidMiner [Hofmann e Klinkenberg 2013] com o operador *Generate Weight* para *Local Polynomial Regression*. O erro médio de 0,3270 alcançado num intervalo de notas que varia de 0 a 2 corresponde a 16, 35% de diferença entre as notas atribuídas pelo avaliador humano e àquela calculadas pelo computador.

Atende, assim, aos critérios do ENEM que só encaminha a correção para um terceiro avaliador se a diferença entre as notas atribuídas for maior do que 20%.

Tabela 1. Resultados de erro médio e correlação obtidos com as *features*

Conjunto	Erro médio	Correlação
<i>Features</i> estatísticas	0.3547	0.3756
<i>Features</i> dissertativas-argumentativas	0.3657	0.2332
<i>Features</i> de pertinência ao tema	0.3748	0.2223
<i>Features bag</i> de palavras	0.3566	0.3576
Resultado com todas	0.3518	0.3963
Ajustes com a RapidMiner	0,3270	0.4070

Um dos problemas reside no fato dos dados serem esparsos para que o aprendizado seja eficaz. Dados esparsos apresentam pouca redundância, que é a condição necessária para que técnicas de Aprendizado de Máquina tenham um bom desempenho [Pardo 2005]. Outra característica que dificulta o aprendizado é a alta concentração de notas com conceito 1 (um). Os algoritmos tendem a aproximar o aprendizado ao *baseline*.

5. Conclusão

Ainda em fase preliminar de desenvolvimento, consideramos que a solução apresentada mostrou ser eficaz para realizar correção automática da Competência 2 do ENEM. Nossos esforços estão direcionados para aprimorar a solução, criando novas *features* que possam reduzir o erro médio e aumentar o índice de correlação de Pearson.

Referências

- Amorim, E. C. F e Veloso, A. (2017). A Multi-aspect Analysis of Automatic Essay Scoring for Brazilian Portuguese. 15th Conference of the European Chapter of the Association for Computational Linguistics, pages 94–102, Valencia, Spain.
- Azevedo, F. F. S. (2010). Dicionário analógico da língua portuguesa, Lexikon Editora Digital, 2ª edição.
- Burrows, S., Gurevych, I., & Stein, B. (2015). The eras and trends of automatic short answer grading. *International Journal of Artificial Intelligence in Education*, 25(1).
- Dikli, S. (2006). An Overview of Automated Scoring of Essays. *Journal of Technology, Learning, and Assessment*, 5(1).
- Globo.com G1 (2016). Corretores de redação do Enem avaliam em média 74 textos por dia. <https://g1.globo.com/educacao/enem/2016/noticia/>.
- Hofmann M., Klinkenberg R. (2013) RapidMiner: Data Mining Use Cases and Business Analytics Applications. CRC Press, October 25, 2013.
- INEP (2017). “Redação no ENEM 2017 Cartilha do Participante”.
- Landauer, T. K. e Dutnais, S. T. (1997). A Solution to Plato’s Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 104(2), 211–240.
- Pardo, T. A. S. (2005). Métodos para análise discursiva automática. Tese de Doutorado, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo.
- UOL Educação (2017) “Banco de Redações”.