

Em Direção à Pontuação Multimodal de Redações Narrativas

Hyan H. N. Batista¹, Gabriel Augusto Barbosa¹

¹Departamento de computação – Universidade Federal Rural do Pernambuco (UFRPE)
Recife – PE – Brasil

hyan.batista@ufrpe.br

gabriel.augusto@ufrpe.br

Abstract.

Resumo.

1. Introdução

2. Trabalhos Relacionados

A presente seção apresenta uma análise dos trabalhos realizados na área. Inicialmente, são discutidas as pesquisas relacionadas à fusão de *features* multimodais no contexto do processamento de documentos digitais. Posteriormente, são abordados os estudos sobre AES (*Automated Essay Scoring*) e os métodos aplicados, com ênfase nas abordagens voltadas para a língua portuguesa.

O uso de técnicas de fusão de *features* multimodais, principalmente no contexto específico de construção de *Vision Language Models*, não é algo novo na literatura. [Jain and Wigington 2019], por exemplo, explorou o uso de um modelo multimodal *early fusion* para classificar automaticamente documentos de imagem. A abordagem demonstrou desempenho superior à métodos que faziam uso de *Transfer Learning*. [Su et al. 2023], por sua vez, desenvolveram um sistema de AES para avaliar redações manualmente escritas de estudantes chineses usando um modelo que fundia *features* multimodais de linguagem e imagem.

No que se refere ao estudo de sistemas AES para Língua Portuguesa, existem uma série de estudos abordando o uso de métodos de *Machine Learning* na detecção automática de elementos essenciais para a avaliação de textos narrativos. [Batista et al. 2022], por exemplo, aplicou métodos de *Machine Learning* para identificar computacionalmente a presença de clímax em textos narrativos de estudantes do ensino fundamental. Na mesma linha, utilizou técnicas de *Generative AI* e outros algoritmos de inteligência computacional para detectar categorias e elementos narrativos.

Os trabalhos citados acima focam em elementos específicos de um texto que são cruciais no processo de *assesment* dos estudantes. Entretanto, existem trabalhos que apresentam sistemas de AES completos. Se baseando no conjunto de dados introduzido por [Mello et al. 2024], por exemplo, [da Silva Filho et al. 2023] combinou diferentes algoritmos de *Machine Learning* com um conjunto de *features* linguísticas construídas manualmente. O sistema de AES alcançou níveis de concordância semelhante ao de dois anotadores humanos. Usando esse mesmo conjunto de dados, [Ribeiro et al. 2024] explorou o uso de modelos pré-treinados baseados na arquitetura *Transformer*.

Na literatura, portanto, há muitos trabalhos explorando o uso de métodos de *Deep Learning*, *Machine Learning* e *Multimodal Feature Fusion* para processar documentos de imagem e redações. Entretanto, há uma escassez de trabalhos explorando a aplicação de métodos de *Deep Learning* empregando fusão de *features* multimodais para o desenvolvimento de sistemas de AES para o *assessment* de textos de alunos do ensino fundamental em Língua Portuguesa.

3. Metodologia

3.1. Dados

O dataset utilizado é uma extensão do dataset do KAGGLE, um dataset de avaliação de redações criado a partir de redações do ensino fundamental do 5º ao 9º ano. Este dataset é composto por imagens das redações, textos anotados manualmente, e avaliações destes textos. Cada texto é avaliado em quatro competências, criadas a partir de um guia de correção feito pela UFAL. Cada uma das competências avalia um aspecto único do texto. São consideradas competências:

- Registro Formal: Avalia o nível da escrita do aluno em relação a norma culta da língua portuguesa, aspectos como erros ortográficos e de pontuações são avaliados nesse quesito.
- Coerência Temática: Avalia o nível de adesão ao tema proposto para a escrita da redação, o aluno deve atender ao tema proposto acrescentando informações, mas sem fugir ao tema.
- Tipologia Textual: O tipo textual proposto no dataset é uma breve narrativa, o aluno deve criar uma história que tenha estrutura narrativa, com introdução, desenvolvimento, clímax e um desfecho. Este aspecto avalia a completude dos elementos narrativos na redação.
- Coesão: Avalia o nível de coesão do texto, entre cada sentença, o aluno deve manter uma sequência lógica de parágrafos.

Como mostra a figura 1, a distribuição dos dados de cada competência é bem diferente. Por exemplo, a competência de Tipologia Textual tem seu pico de frequência no nível 4, enquanto as competências de Coesão, Coerência Temática e Registro Formal têm mais exemplos no nível 3. Já o nível 5 quase não possui exemplares de redações, mostrando que falta, nessa base, redações de nota máxima.

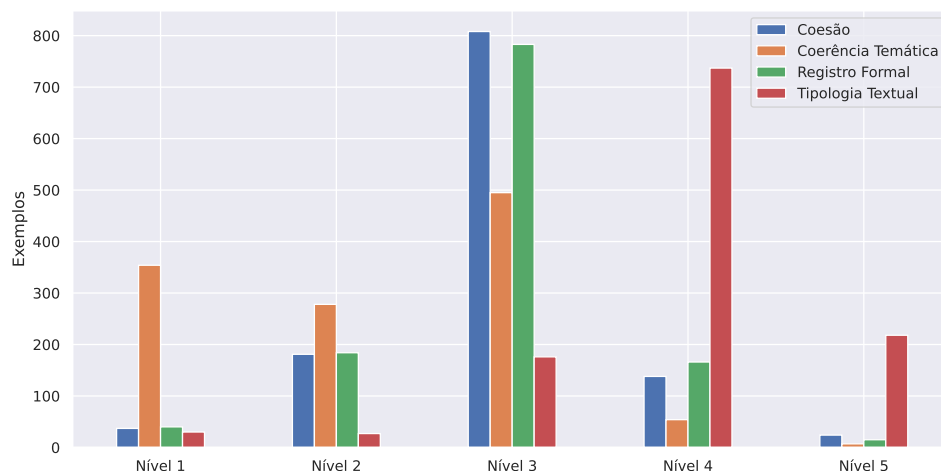


Figure 1. Redações Agrupadas por Nível

	Encoder	Acc	Prec _m	Prec _w	Recall _m	Recall _w	F _{1m}	F _{1w}	Kappa
SVC	TF-IDF	0.680	0.170	0.463	0.250	0.680	0.202	0.551	0.000
SVC	BERT	0.680	0.170	0.463	0.250	0.680	0.202	0.551	0.000
SVC	LBP	0.680	0.170	0.463	0.250	0.680	0.202	0.551	0.000
RF	TF-IDF	0.660	0.168	0.458	0.243	0.660	0.199	0.540	-0.022
RF	BERT	0.690	0.275	0.526	0.288	0.690	0.263	0.584	0.081
RF	LBP	0.620	0.169	0.459	0.228	0.620	0.193	0.527	-0.031
DT	TF-IDF	0.560	0.349	0.588	0.329	0.560	0.328	0.561	0.140
DT	BERT	0.500	0.315	0.578	0.324	0.500	0.295	0.509	0.116
DT	LBP	0.480	0.226	0.482	0.229	0.480	0.223	0.472	0.004
ET	TF-IDF	0.660	0.168	0.458	0.243	0.660	0.199	0.540	-0.022
ET	BERT	0.690	0.279	0.536	0.288	0.690	0.265	0.591	0.110
ET	LBP	0.620	0.169	0.461	0.228	0.620	0.194	0.527	-0.036
XGB	TF-IDF	0.640	0.267	0.539	0.345	0.640	0.296	0.580	0.161
XGB	BERT	0.700	0.335	0.574	0.360	0.700	0.338	0.624	0.198
XGB	LBP	0.550	0.188	0.486	0.215	0.550	0.199	0.513	0.009
MLP	TF-IDF	0.660	0.168	0.458	0.243	0.660	0.199	0.540	-0.022
MLP	BERT	0.610	0.386	0.592	0.356	0.610	0.361	0.594	0.160
MLP	LBP	0.680	0.170	0.463	0.250	0.680	0.202	0.551	0.000

Table 1. Pontuações nas métricas dos vários classificadores e *encoders* para a competência de coesão.

4. Discussão

5. Conclusão

References

Batista, H. H., Barbosa, G. A., Miranda, P., Santos, J., Isotani, S., Cordeiro, T., Bittencourt, I. I., and Mello, R. F. (2022). Detecção automática de clímax em produções de

	Encoder	Acc	Prec _m	Prec _w	Recall _m	Recall _w	F _{1m}	F _{1w}	Kappa
SVC	TF-IDF	0.570	0.308	0.452	0.396	0.570	0.338	0.492	0.336
SVC	BERT	0.620	0.324	0.469	0.428	0.620	0.364	0.526	0.405
SVC	LBP	0.380	0.095	0.145	0.250	0.380	0.138	0.210	0.000
RF	TF-IDF	0.610	0.321	0.468	0.421	0.610	0.359	0.522	0.394
RF	BERT	0.610	0.366	0.496	0.425	0.610	0.375	0.530	0.394
RF	LBP	0.400	0.292	0.374	0.295	0.400	0.273	0.368	0.100
DT	TF-IDF	0.460	0.342	0.470	0.373	0.460	0.341	0.457	0.225
DT	BERT	0.480	0.383	0.520	0.365	0.480	0.360	0.482	0.256
DT	LBP	0.240	0.182	0.251	0.177	0.240	0.173	0.237	-0.098
ET	TF-IDF	0.600	0.322	0.470	0.415	0.600	0.356	0.518	0.380
ET	BERT	0.560	0.310	0.451	0.383	0.560	0.334	0.488	0.321
ET	LBP	0.360	0.247	0.320	0.269	0.360	0.247	0.327	0.045
XGB	TF-IDF	0.580	0.391	0.525	0.416	0.580	0.387	0.537	0.369
XGB	BERT	0.480	0.342	0.470	0.356	0.480	0.339	0.464	0.240
XGB	LBP	0.390	0.320	0.403	0.290	0.390	0.281	0.372	0.084
MLP	TF-IDF	0.530	0.406	0.531	0.397	0.530	0.381	0.504	0.290
MLP	BERT	0.510	0.337	0.478	0.361	0.510	0.341	0.482	0.264
MLP	LBP	0.330	0.110	0.161	0.226	0.330	0.144	0.211	-0.057

Table 2. Pontuações nas métricas dos vários classificadores e *encoders* para a competência de coerência temática.

	Encoder	Acc	Prec _m	Prec _w	Recall _m	Recall _w	F _{1m}	F _{1w}	Kappa
SVC	TF-IDF	0.650	0.156	0.423	0.240	0.650	0.189	0.512	0.000
SVC	BERT	0.650	0.158	0.427	0.240	0.650	0.190	0.515	0.013
SVC	LBP	0.650	0.156	0.423	0.240	0.650	0.189	0.512	0.000
RF	TF-IDF	0.640	0.149	0.420	0.227	0.640	0.180	0.507	-0.009
RF	BERT	0.690	0.335	0.580	0.315	0.690	0.306	0.612	0.228
RF	LBP	0.590	0.171	0.458	0.216	0.590	0.190	0.514	0.036
DT	TF-IDF	0.530	0.204	0.465	0.230	0.530	0.214	0.490	0.079
DT	BERT	0.670	0.386	0.639	0.405	0.670	0.384	0.645	0.331
DT	LBP	0.490	0.202	0.467	0.223	0.490	0.208	0.474	0.016
ET	TF-IDF	0.640	0.149	0.420	0.227	0.640	0.180	0.507	-0.009
ET	BERT	0.660	0.240	0.518	0.260	0.660	0.236	0.570	0.152
ET	LBP	0.590	0.168	0.452	0.217	0.590	0.188	0.510	0.027
XGB	TF-IDF	0.620	0.277	0.541	0.313	0.620	0.286	0.566	0.184
XGB	BERT	0.740	0.480	0.688	0.406	0.740	0.419	0.694	0.413
XGB	LBP	0.510	0.187	0.459	0.210	0.510	0.197	0.482	0.021
MLP	TF-IDF	0.630	0.151	0.427	0.223	0.630	0.180	0.509	0.002
MLP	BERT	0.640	0.323	0.574	0.312	0.640	0.307	0.595	0.211
MLP	LBP	0.650	0.156	0.423	0.240	0.650	0.189	0.512	0.000

Table 3. Pontuações nas métricas dos vários classificadores e *encoders* para a competência de registro formal.

textos narrativos. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 932–943. SBC.

- da Silva Filho, M. W., Nascimento, A. C., Miranda, P., Rodrigues, L., Cordeiro, T., Isotani, S., Bittencourt, I. I., and Mello, R. F. (2023). Automated formal register scoring of student narrative essays written in portuguese. In *Anais do II Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil*, pages 1–11. SBC.
- Jain, R. and Wigington, C. (2019). Multimodal document image classification. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 71–77. IEEE.
- Mello, R. F., Oliveira, H., Wenceslau, M., Batista, H., Cordeiro, T., Bittencourt, I. I., and Isotani, S. (2024). Propor’24 competition on automatic essay scoring of portuguese narrative essays. In *Proceedings of the 16th International Conference on Computational Processing of Portuguese-Vol. 2*, pages 1–5.
- Ribeiro, E., Mamede, N., and Baptista, J. (2024). Exploring the automated scoring of narrative essays in brazilian portuguese using transformer models. In *Proceedings of the 16th International Conference on Computational Processing of Portuguese-Vol. 2*, pages 14–17.
- Su, T., Wang, J., You, H., and Wang, Z. (2023). Multimodal scoring model for handwritten chinese essay. In *International Conference on Document Analysis and Recognition*, pages 505–519. Springer.