EM DIREÇÃO À AVALIAÇÃO MULTIMODAL DE TEXTOS NARRATIVOS

Hyan H. N. Batista, Gabriel A. Barbosa

INTRODUÇÃO

A **avaliação de textos narrativos em português** é uma tarefa complexa e que consome muito tempo. Embora, trabalhos anteriores tenham explorado o uso de métodos de ***deep learning***, há uma escassez de trabalhos sobre o uso ***features* multimodais** nesse contexto.

OBJETIVO

O objetivo deste trabalho é desenvolver um sistema de **avaliação multimodal automática de textos narrativos**.

METODOLOGIA

Para este estudo usou-se o *Brazilian Portuguese Narrative Essays Dataset*. O *dataset* possui **1235** amostras, mas após passar pelas etapas de *data cleansing* ficou com **1163**. Os textos e imagens passaram por etapas de redimensionamento, remoção de tags, caracteres desnecessários e normalização. A Figura 1 apresenta a distribuição de suas classes.

Os métodos de extração de *features* de imagens e textos *baseline* empregados foram **LBP**, **ViT**, **BERT** e **TF-IDF** e para a classificação, **SVM**, ***Random Forest***, ***Decision Tree***, ***Extra Trees***, **XGBoost** e **MLP**. Para avaliação, selecionou-se as métricas **precisão**, ***recall*** e ***f-score*** em suas versões de média ponderada e macro. Além disso, inclui-se **acurácia** e o **Cohen's kappa score**. Na Tabela 1, pode-se observar os resultados. A ideia da abordagem proposta (**BERT + ViT + BiLSTM + MLP**) pode ser entendida pela Figura 2.

RESULTADOS

Approach	Cohesion		Formal Register		Text Typology		Thematic Coherence	
	Kappa	Weighted F1	Kappa	Weighted F1	Kappa	Weighted F1	Kappa	Weighted F1
SVC + TF-IDF	-0.012	0.546	-0.009	0.507	**0.176**	0.548	0.398	0.542
SVC + BERT	**0.274**	0.636	0.397	0.666	0.065	0.461	0.374	0.531
SVC + LBP	0.128	0.400	0.086	0.204	-0.030	0.053	0.010	0.113
SVC + ViT	-0.067	0.512	0.054	0.528	0.000	0.460	-0.078	0.258
RF + TF-IDF	0.049	0.562	0.065	0.541	0.161	0.539	0.380	0.532
RF + BERT	0.253	**0.639**	**0.410**	**0.682**	0.016	0.478	0.376	0.535
RF + LBP	0.139	0.572	0.104	0.523	-0.083	0.398	0.060	0.347
RF + ViT	-0.012	0.546	-0.008	0.505	0.159	0.541	-0.041	0.284
DT + TF-IDF	0.088	0.542	0.029	0.450	0.031	0.441	0.225	0.456
DT + BERT	0.136	0.538	0.229	0.562	-0.018	0.427	0.085	0.358
DT + LBP	0.043	0.523	0.027	0.471	0.077	0.451	-0.033	0.244
DT + ViT	0.004	0.498	0.085	0.492	-0.014	0.403	-0.021	0.262
ET + TF-IDF	-0.012	0.546	-0.009	0.507	0.176	0.548	0.407	0.540
ET + BERT	0.000	0.551	0.029	0.517	-0.042	0.448	**0.458**	**0.560**
ET + LBP	0.109	0.578	0.087	0.527	0.020	0.461	0.083	0.369
ET + ViT	-0.012	0.546	-0.009	0.507	**0.176**	0.548	-0.060	0.278
XGB + TF-IDF	0.205	0.603	0.256	0.605	0.128	0.511	0.308	0.515
XGB + BERT	0.234	0.623	0.269	0.607	0.020	0.464	0.248	0.458
XGB + LBP	0.165	0.584	0.097	0.511	0.009	0.426	0.142	0.400
XGB + ViT	0.040	0.555	0.030	0.505	0.091	0.512	-0.109	0.247
MLP + TF-IDF	-0.002	0.542	0.027	0.520	0.287	**0.599**	0.338	0.541
MLP + BERT	0.205	0.604	0.348	0.652	-0.012	0.443	0.322	0.515
MLP + LBP	0.104	0.484	0.160	0.424	0.017	0.097	-0.070	0.110
MLP + ViT	-0.027	0.504	-0.025	0.481	0.127	0.521	-0.062	0.276

Tabela 1. Comparação das diferentes abordagens ao longo das métricas selecionadas.

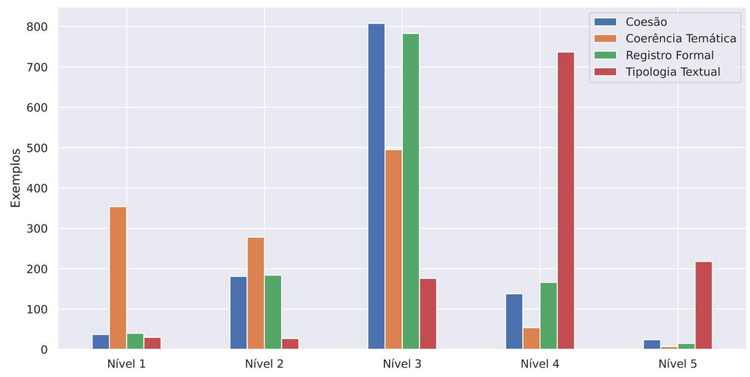

Figura 1. Distribuição das classes do *Brazilian Portuguese Narrative Essays Dataset*.

DISCUSSÃO

Observando a Tabela 1, é possível chegar as seguintes conclusões:

- Para coesão, registro formal e coerência temática é essencial o entendimento contextual e semântico do texto, desse modo, faz sentido que técnicas de *word embedding* como o **BERT** entreguem melhores resultados;
- Os melhores resultados para tipologia textual, entretanto, não seguem essa lógica, tendo o **ViT** e **TF-IDF** como melhores *encoders*. Isso pode indica que **vocabulário específico** e a **disposição do texto na folha** são bons preditores para essa competência.

CONCLUSÃO

Este trabalho teve como objetivo implementar e avaliar um sistema multimodal fim-a-fim para avaliação de textos narrativos. Entre suas principais contribuições está uma análise comparativa de métodos de NLP, processamento de imagens e *machine learning* no contexto de AES para textos narrativos em português, o estabelecimento de uma base para o estudo de abordagens multimodais como um método de avaliação de redações desse tipo e, também, a construção de um modelo multimodal fim-a-fim para AES. Para o futuros trabalhos, pretende-se explorar o uso de outras abordagens multimodais tendo em foco este problema, como **VisualBERT**, **ViBERT** e ***Text Vision Dual Encoders***.

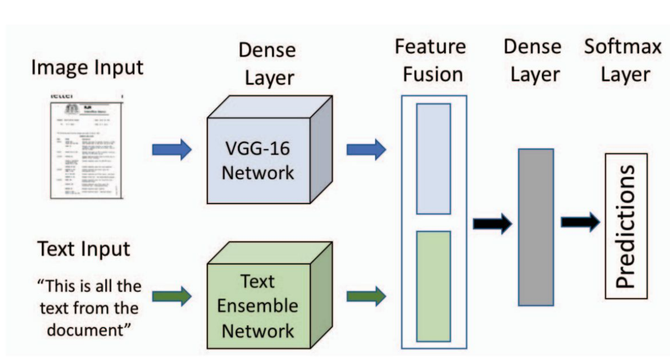

Figura 2. *Multimodal early feature fusion*.