

Análise de Sentimento Baseada em Aspectos no Domínio de Acomodações Utilizando o modelo BERTimbau

Franco Noronha Pereira e Larissa Astrogildo de Freitas e Ulisses Brisolara Corrêa

Universidade Federal de Pelotas, Pelotas, RS, Brasil

{fnpereira, larissa, ulisses}@inf.ufpel.edu.br

Resumo

Este trabalho investiga a aplicação do modelo monolíngue BERTimbau para a Análise de Sentimentos Baseada em Aspectos (ABSA) em português, visando estabelecer um *baseline* robusto para o domínio hoteleiro. São comparadas duas estratégias via *fine-tuning*: uma abordagem *pipeline* (extração seguida de classificação) e uma abordagem *end-to-end* (multitarefa com esquema de *tags* colapsadas). Avaliadas no conjunto de dados da competição ABSAPT 2024, os resultados evidenciam um *trade-off* arquitetural: o *pipeline* favorece a revocação na extração de aspectos (F1: 0,840), enquanto o *end-to-end* prioriza a precisão, mas sofre com a dispersão de classes. A análise composta demonstra desempenho competitivo (Medida-F $\approx 0,72$ para ambos), oferecendo um ponto de partida para futuras investigações em arquiteturas híbridas e generativas para o português.

1 Introdução

O crescimento acelerado da produção de conteúdo textual em plataformas digitais, como redes sociais e sistemas de avaliação de serviços, criou um ambiente rico para a extração de informações opinativas. A tarefa de Análise de Sentimentos (AS) busca identificar a polaridade emocional expressa em textos, podendo ser aplicada em diferentes níveis de granularidade, como documento, sentença ou aspecto. A variante Análise de Sentimento Baseada em Aspectos (do inglês, *Aspect-Based Sentiment Analysis* - ABSA) oferece uma análise mais refinada ao identificar tanto os aspectos mencionados quanto os sentimentos associados a cada um deles.

Nos últimos anos, modelos baseados na arquitetura *Transformer* revolucionaram a área de Processamento da Linguagem Natural (PLN), permitindo o processamento eficiente de dependências contextuais longas e obtendo resultados superiores em diversas tarefas. O BERTimbau, versão monolíngue

pré-treinada para o português do modelo *Bidirectional Encoder Representations from Transformers* (BERT), é um modelo do tipo *encoder*, com foco em gerar representações textuais com relações de contexto bidirecional, e tem se destacado em diversas tarefas de PLN. Este trabalho implementa e avalia duas estratégias de ABSA em português baseadas no modelo BERTimbau, utilizando o conjunto de dados disponibilizado na competição ABSAPT 2024 (Bender et al., 2024).

2 Fundamentação Teórica

Esta seção está organizada da seguinte forma: a Subseção 2.1 apresenta as definições de AS e suas especializações na vertente ABSA; a Subseção 2.2 discute os modelos de linguagem BERT e sua versão para o português, o BERTimbau; a Subseção 2.3 detalha o conjunto de dados da competição ABSAPT 2024, destacando suas características e desafios, como o desbalanceamento de classes.

2.1 AS e ABSA

A AS consiste na identificação da polaridade opinativa presente em textos. Segundo (Liu, 2012), uma opinião pode ser estruturada como uma quádrupla (alvo, sentimento, autor, tempo). Na ABSA, o objetivo é identificar aspectos (entidades ou atributos mencionados) e determinar a polaridade dos aspectos, como positivos, negativos ou neutros.

A ABSA é tradicionalmente dividida em duas subtarefas principais, que podem ser realizadas separadamente ou de forma conjunta:

- **Extração de Aspectos (do inglês, *Aspect Extraction* - AE):** Na primeira subtarefa, o objetivo é identificar quais partes do texto representam os aspectos mencionados pelo usuário. Considere, por exemplo, a seguinte avaliação do setor hoteleiro:

“A **localização** é excelente e a **cama**

é muito confortável, porém o **ar-condicionado** fazia muito barulho.”

Neste caso, os termos destacados (**localização**, **cama**, **ar-condicionado**) seriam os aspectos extraídos pelo modelo.

- **Classificação de Sentimento por Aspecto (do inglês, *Aspect Sentiment Classification - ASC*):** Na segunda sub tarefa, o objetivo é determinar a polaridade dos aspectos identificados. Diferente da AS tradicional, que atribui uma única polaridade ao documento inteiro, a ASC analisa cada aspecto de forma independente. Assim, o mesmo texto pode expressar opiniões positivas e negativas simultaneamente, dependendo do aspecto considerado.

Utilizando o mesmo exemplo da etapa anterior:

“A **localização** é excelente e a **cama** é muito confortável, porém o **ar-condicionado** fazia muito barulho.”

Neste caso, o modelo deve classificar a polaridade de cada aspecto:

- **localização** → **Positivo**;
- **cama** → **Positivo**;
- **ar-condicionado** → **Negativo**.

2.2 BERT e BERTimbau

O BERT, proposto por Devlin et al. (2019), é um marco no avanço do PLN ao utilizar exclusivamente o *encoder* da arquitetura *Transformer*. Seu mecanismo de atenção permite capturar relações contextuais bidirecionais entre palavras, produzindo representações profundas e sensíveis ao contexto.

O pré-treinamento do BERT envolve duas tarefas: *Masked Language Modeling* (MLM), em que parte dos *tokens* é ocultada para que o modelo aprenda a prever palavras com base no contexto; e *Next Sentence Prediction* (NSP), usada para ensinar o modelo a capturar relações entre sentenças. Após essa etapa, o modelo pode ser refinado via *fine-tuning* para diferentes tarefas de PLN, como classificação de texto, análise de sentimentos e sistemas de perguntas e respostas.

O BERTimbau é uma adaptação do BERT voltada ao português brasileiro. Criado para superar as limitações de modelos multilíngues — que dividem vocabulário e capacidade entre diversas

línguas —, o BERTimbau foi treinado exclusivamente em textos escritos em português, seguindo o mesmo procedimento de pré-treinamento do BERT, baseado nas tarefas de MLM e NSP.

2.3 Conjunto de Dados Disponibilizado na Competição ABSAPT 2024

O corpus utilizado neste trabalho é o conjunto de dados disponibilizado na competição ABSAPT 2024 (Bender et al., 2024). Ele reúne avaliações de viajantes sobre hospedagens da plataforma TripAdvisor¹, anotados em nível de aspecto. Os dados são divididos em três partes: um conjunto de treino público e dois conjuntos de teste privados (um para AE e um para ASC). A Tabela 1 apresenta o tamanho do corpus ABSAPT 2024.

Conjunto	Nº de amostras	Nº de avaliações
Treino	4828	1322
Teste AE	283	283
Teste ASC	1176	282

Tabela 1: Tamanho do corpus ABSAPT 2024.

Um desafio relevante do corpus é o desbalanceamento das classes em ambas as tarefas. Na AE, alguns aspectos aparecem com mais frequência que outros; na ASC, a polaridade predominante é a positiva. A Figura 1 ilustra essa distribuição desigual, que pode introduzir viés nos modelos e dificultar o aprendizado de classes menos frequentes, neste caso, negativa e neutra.

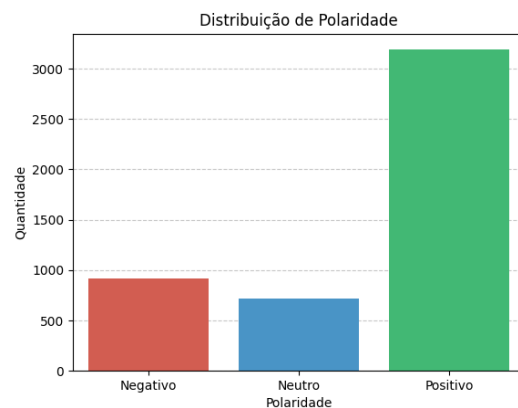


Figura 1: Distribuição de polaridade do *dataset* de treinamento da tarefa de SA.

¹<https://www.tripadvisor.com.br/>

3 Trabalhos Relacionados

O trabalho de [de Freitas \(2015\)](#) apresenta uma abordagem de ABSA para o português baseada em ontologias e regras linguísticas. Na sub tarefa de AE é utilizada uma ontologia de domínio para identificar aspectos explícitos e implícitos, enquanto na sub tarefa de ASC são utilizados léxicos de sentimento combinados com regras que tratam negações e relações sintáticas para determinar a polaridade dos aspectos identificados.

[Corrêa \(2021\)](#) desenvolveu o corpus TripAdvisorBR-2020, baseado no protocolo de anotação do trabalho de [de Freitas \(2015\)](#), e avaliou duas arquiteturas convolucionais com entrada em nível de caractere: uma abordagem *end-to-end* (E2E) e outra que separa ABSA em sub tarefas. O corpus TripAdvisorBR-2020 serviu de base para a construção dos conjuntos de dados disponibilizados nas competições ABSAPT 2022([da Silva et al., 2022](#)) e 2024 ([Bender et al., 2024](#)).

Na competição ABSAPT 2024, o trabalho de [Dal Ponte et al. \(2024\)](#) investigou o impacto do uso de dados sintéticos na sub tarefa de ASC. Utilizando o modelo BERT e gerando 1.281 novas instâncias com o auxílio do ChatGPT, os autores observaram um aumento significativo tanto na acurácia quanto na Medida-F após o aumento de dados no conjunto de treino.

A equipe ABCD ([Van Thin et al., 2024](#)) comparou modelos de classificação (como: BERTimbau, mDeBERTa_v3 e XLM-R) e modelos generativos (como: mT5 e mT0) nas sub tarefas AE e ASC. Os resultados indicam vantagem das abordagens de classificação na sub tarefa de AE e competitividade — por vezes superioridade — das abordagens generativas na sub tarefa de ASC.

Avanços recentes exploram arquiteturas de extração conjunta que mitigam problemas de *tags* colapsadas, como o framework DOER ([Luo et al., 2019](#)), que utiliza atenção cruzada entre tarefas. No contexto do português, abordagens generativas vêm ganhando destaque. O trabalho de [Gomes et al. \(2023\)](#) utilizou *ensembles* de modelos Transformer para extração e o modelo generativo PTT5 para classificação e logrou a primeira colocação do desafio ABSAPT 2022 ([da Silva et al., 2022](#)).

[Lopes et al. \(2022\)](#) usaram o BERTimbau na sub tarefa de ASC convertendo ABSA em um problema de classificação de pares de sentenças por meio de uma formulação em estilo

Pergunta-Resposta (do inglês, *Question Answering - QA*). Um pós-treinamento no domínio específico mostrou-se essencial para melhorar o desempenho da abordagem proposta pelos autores.

O estudo de [da Rocha Junqueira et al. \(2023\)](#) compara o modelo Albertina ao BERTimbau em quatro tarefas (AE, Detecção de Discurso de Ódio, Detecção de Ironia e QA). O modelo Albertina obteve melhores resultados na Detecção de Ironia e QA, enquanto o modelo BERTimbau obteve melhores resultados na sub tarefa de AE e Detecção de Discurso de Ódio.

[Gomes \(2024\)](#) demonstra que modelos baseados em Redes de Atenção em Grafos alcançam resultados competitivos na sub tarefa de ASC (Acurácia Balanceada de 74%) e oferecem maior leveza e interpretabilidade, sendo adequados a cenários com restrições computacionais.

Outra vertente promissora é o uso de aumento de dados (Data Augmentation) e adaptação de domínio para lidar com o desbalanceamento de classes, conforme explorado em trabalhos recentes que utilizam LLMs (do inglês, *Large Language Models*) para gerar exemplos sintéticos ([Zhong et al., 2024](#)). Diferente destes trabalhos, nossa investigação foca em isolar e quantificar o desempenho de um modelo *encoder* monolíngue (BERTimbau-Large) em configurações padrão, estabelecendo um *baseline* claro para arquiteturas não-generativas.

4 Metodologia

Esta seção está organizada da seguinte forma: a Subseção 4.1 descreve o tratamento dos dados e o esquema de anotação BIO; as Subseções 4.2 e 4.3 detalham, respectivamente, as abordagens *pipeline* e *end-to-end* implementadas; a Subseção 4.4 define as métricas de avaliação para as tarefas de extração e classificação; e, finalmente, a Subseção 4.5 apresenta os detalhes da configuração e treinamento dos modelos.

4.1 Pré-processamento dos dados

O pré-processamento realizado concentra-se, principalmente, na organização dos dados e na formatação adequada dos dados para o treinamento dos modelos. Como o BERTimbau é do tipo *cased*, o texto não foi convertido para letras minúsculas. Essa decisão se justifica pelo fato de que modelos *cased* são treinados para levar em consideração a distinção entre maiúsculas e minúsculas, extraindo

possíveis diferenças de significado ou função a partir do uso de caixa alta no texto.

No *dataset* de treino do ABSAPT 2024 (Bender et al., 2024), cada linha representa um aspecto identificado em uma avaliação de hotel. Textos com múltiplos aspectos aparecem repetidamente, com apenas um aspecto por linha. Para melhor aprendizado dos modelos, foi necessário realizar o agrupamento de aspectos por avaliações de hotéis. Esse agrupamento consiste em consolidar todas as instâncias de uma mesma avaliação em uma única entrada, contendo todos os aspectos associados àquela avaliação.

Para o treinamento dos modelos de classificação de *tokens*, as avaliações de hotéis passam por um processo de rotulagem no formato BIO, que identifica quais *tokens* correspondem a determinado aspecto. Esse formato é amplamente utilizado em tarefas de segmentação de entidades em PLN, onde cada palavra é classificada em uma das três *tags*:

- B (*beginning*): o *token* inicia uma entidade de interesse;
- I (*inside*): o *token* está dentro de uma entidade de interesse;
- O (*outside*): o *token* não pertence a nenhuma entidade de interesse.

Em muitos cenários — incluindo este trabalho — essas *tags* podem ser combinadas com diferentes classes, como aspecto e polaridade (por exemplo, B-POS, I-NEG), permitindo distinguir múltiplas categorias dentro da mesma sequência.

A marcação das *tags* BIO ocorre após a tokenização do texto. Ao utilizar o parâmetro `return_offsets_mapping` na função de tokenização, é possível obter, para cada *token*, as posições de início e fim no texto original, o que permite alinhar os *tokens* aos trechos anotados como aspectos.

Na Figura 2 é ilustrado o resultado final da marcação das *tags* BIO no texto original: “Funcionários muito atenciosos, porém o café da manhã é muito fraco.”. Abaixo estão enumerados os passos que foram realizados durante a marcação das *tags* BIO.

1. O texto original é tokenizado utilizando o parâmetro `return_offsets_mapping`, que retorna as posições de início e fim de cada *token* no texto original.

2. Inicialmente, todos os *tokens* recebem a *tag* 0, indicando que não pertencem a nenhum aspecto.
3. *Tokens* especiais, como [CLS], [SEP] e [PAD], recebem a *label* -100, de forma que sejam ignorados durante o treinamento do modelo.
4. Para cada aspecto anotado no texto original, o mapeamento de *offsets* é percorrido para identificar os *tokens* cujas posições de início e fim estejam contidas nos limites “start” e “end” do aspecto. O primeiro *token* correspondente recebe a *tag* “B”, enquanto os *tokens* subsequentes recebem a *tag* “I”, ambos com o sufixo da sua polaridade associada.
5. *Tokens* que representam sub-palavras — identificados por compartilharem o mesmo `word_id` com o *token* anterior — recebem também a *label* -100, para que não sejam considerados na avaliação e no treinamento do modelo.

Texto original: Funcionários muito atenciosos, porém o café da manhã é muito fraco.
 Tokens: [[CLS], 'Func.', '##onários', 'muito', 'aten', '##ciosos', '.', 'porém', 'o', 'café', 'da', 'manhã', 'é', 'muito', 'fraco', ',', '[SEP]', [PAD], [PAD], [PAD], ...]
 Labels: [-100, 1, -100, 0, 0, -100, 0, 0, 0, 3, 4, 4, 0, 0, 0, -100, -100, -100, -100, ...]
 BIO tags: [O, B-POS, O, O, O, O, O, O, O, B-NEG, I-NEG, I-NEG, O, O, O, O, O, O, ...]

Figura 2: Exemplo de marcação das *tags* BIO.

4.2 Abordagem Pipeline

A abordagem *pipeline* para ABSA estrutura a tarefa em duas subtarefas distintas: AE e ASC. Primeiramente, a AE identifica todos os aspectos que são alvos de opinião em uma avaliação de hotel fornecida. Em seguida, a ASC determina a polaridade dos aspectos identificados — seja ela positiva, negativa ou neutra — para cada aspecto que foi previamente extraído do texto. Este processo é ilustrado na Figura 3.

4.2.1 Extração de Aspectos

O modelo pré-treinado é submetido a um processo de *fine-tuning* para classificação de *tokens*, no qual os textos originais tokenizados são usados como entrada e as *tags* BIO correspondentes funcionam como rótulos. O objetivo é que a cabeça do modelo seja treinada para prever a *tag* BIO de cada *token*.

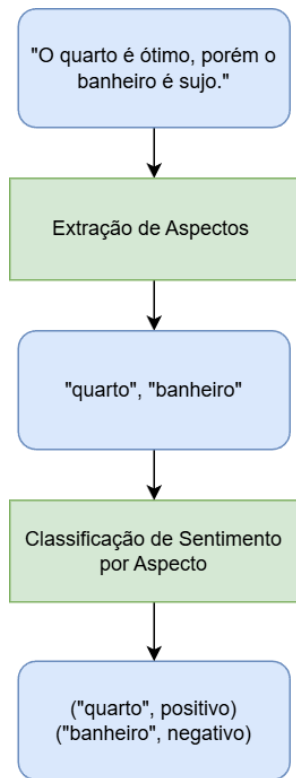


Figura 3: Fluxograma da abordagem *pipeline*.

4.2.2 Classificação de Sentimento por Aspecto

Essa tarefa é realizada por meio da Classificação de Pares de Sequência, que consiste em atribuir um rótulo que descreve a relação entre duas sequências de texto, sendo comumente utilizada para medir similaridade ou identificar relações semânticas entre sentenças. No contexto de ABSA, essa tarefa é adaptada para prever a polaridade expressa em um texto opinativo com relação a um aspecto específico mencionado neste texto.

Para isso, cada par (*texto opinativo*, *aspecto*) é transformado em uma sequência única, na qual as duas partes são separadas por um *token* especial [SEP]. Em seguida, o modelo passa por um processo de *fine-tuning*, no qual é ajustado com base em um conjunto de dados anotado com as polaridades correspondentes (por exemplo, *positivo*, *negativo* ou *neutro*). O modelo aprende, assim, a prever a polaridade correta para cada par, tratando a tarefa como uma classificação de rótulo único.

Ressalta-se que, para evitar vazamento de dados durante a avaliação do *pipeline* na sub tarefa ASC isolada, o treinamento e teste utilizaram os aspectos *gold-standard*. No entanto, para a **Avaliação Composta**, o módulo ASC recebeu estritamente os aspectos preditos pelo módulo AE, propagando-se os erros de extração conforme um cenário real de

uso.

4.3 Abordagem *End-to-End*

A abordagem E2E para ABSA visa integrar, em um único modelo, as subtarefas de AE e ASC. Diferente da abordagem *pipeline*, onde os aspectos são inicialmente extraídos e, em seguida, classificados quanto à polaridade, na abordagem E2E o modelo é treinado para realizar as duas subtarefas simultaneamente. Em um cenário típico, o modelo recebe um texto opinativo como entrada e, como saída, retorna pares (*aspecto*, *polaridade*), esse processo é ilustrado na Figura 4.

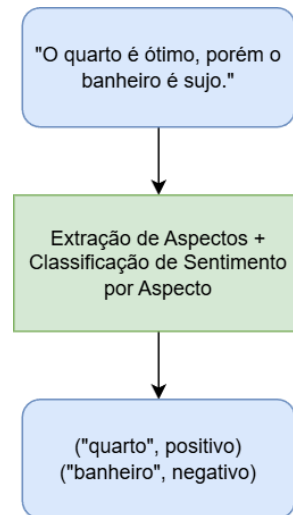


Figura 4: Fluxograma da abordagem *end-to-end*.

Para que o modelo seja capaz de aprender simultaneamente a identificar os aspectos e a determinar a polaridade dos aspectos identificados, os dados de entrada são anotados utilizando o esquema BIO, com a adição de classes que representam cada polaridade. Assim, cada palavra relevante pode ser rotulada com uma combinação de posição e polaridade, como B-POS, I-POS, B-NEG, I-NEG, B-NEU ou I-NEU. Dessa forma, o modelo aprende não apenas a segmentar os aspectos na sequência, mas também a associá-los diretamente à sua respectiva polaridade.

4.4 Métricas para AE, ASC e Composta

A sub tarefa AE avalia a correspondência exata entre os aspectos identificados e os aspectos de referência, conforme adotado na competição ABSAPT 2024 (Bender et al., 2024).

Para garantir a consistência com a ontologia de referência, aplicou-se uma normalização de aspectos baseada em regras léxicas (lematização e correção ortográfica, ex: “apto” → “apartamento”,

“café” → “café da manhã”). A lista de normalização está disponível publicamente no repositório do projeto² para garantir reprodutibilidade.

Dado um conjunto de aspectos inferidos e outro de referência, definem-se:

- **VP:** aspectos corretamente identificados;
- **FP:** aspectos identificados que não pertencem ao conjunto de referência;
- **FN:** aspectos do conjunto de referência que não foram identificados.

A partir deles, calcula-se:

$$\text{Precisão} = \frac{VP}{VP + FP}$$

$$\text{Revocação} = \frac{VP}{VP + FN}$$

$$\text{Medida-F} = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}}$$

A subtarefa ASC é tratada como uma classificação ternária. Ela avalia as mesmas métricas da AE, acrescidas da acurácia balanceada (Bacc), adequada para cenários desbalanceados. A Bacc corresponde à média da revocação por classe:

$$\text{Bacc} = \frac{1}{C} \sum_{i=1}^C \frac{VP_i}{VP_i + FN_i},$$

onde C é o número de classes.

A avaliação composta mede o desempenho completo da tarefa de ABSA, considerando simultaneamente as subtarefas AE e ASC. As métricas (Precisão – P, Revocação – R e Medida-F – F1) são calculadas sobre tuplas (*aspecto, polaridade*), de modo que VP, VN, FN e FP passam a refletir correspondência exata entre tuplas identificadas e tuplas de referência. Essa métrica é aplicada na abordagem *pipeline* e E2E.

4.5 Treinamento

O treinamento dos modelos foi realizado utilizando a classe Trainer da biblioteca Hugging Face Transformers³, que abstrai todo o ciclo de treinamento, desde o preparo dos dados até a avaliação.

²omitteddueblindreview

³<https://github.com/huggingface/transformers>

Foram executadas até 6 épocas, com o uso de *early stopping* (paciência de 1 época e *threshold* de 0,01) para reduzir o risco de sobreajuste. O modelo utilizado foi a versão Large do BERTimbau.

Para garantir uma avaliação robusta da capacidade de generalização, aplicou-se validação cruzada com *k-fold* igual a 5. Em cada iteração, um subconjunto é usado como teste e os demais para treinamento. A métrica final é obtida pela média dos cinco testes. Os hiperparâmetros utilizados estão apresentados na Tabela 2.

Hiperparâmetro	Valor
<i>Train Epochs</i>	Máx. 6
<i>Batch Size</i>	16
<i>Learning Rate</i>	$5 * 10^{-5}$
<i>Weight Decay</i>	0,01
<i>Patience</i>	1
<i>Threshold</i>	0,01
K	5

Tabela 2: Hiperparâmetros utilizados.

5 Resultados Obtidos

Este capítulo apresenta e discute os resultados experimentais obtidos pelas abordagens propostas (*end-to-end* e *pipeline*). Inicialmente, é apresentada a análise de estabilidade dos modelos por meio da validação cruzada. Na sequência, o desempenho é detalhado separadamente para as subtarefas de AE e ASC, incluindo comparações com os resultados da competição ASBAPT 2024. Por fim, realiza-se uma avaliação composta para verificar a eficácia global dos sistemas ao integrarem ambas as subtarefas.

A Tabela 3 resume os resultados obtidos durante o treinamento com validação cruzada com *k-fold* igual a 5. A métrica de avaliação da abordagem E2E e *pipeline* na tarefa de AE é a Medida-F, e a métrica de avaliação da abordagem *pipeline* na tarefa ASC é a Bacc. Todas as abordagens demonstraram um baixo desvio padrão da média da métrica de avaliação, isso significa que o desempenho foi consistente entre as diferentes amostras de dados da validação cruzada. Essa estabilidade minimiza a possibilidade de que os resultados positivos sejam um acaso ou fruto de um sobreajuste a uma amostra particular de treinamento. Isso é um indicativo de que as abordagens são capazes de generalizar para dados inéditos.

A Tabela 4 exhibe os resultados obtidos no con-

Abordagem	Média	Desvio Padrão
E2E*	0,733	0,014
<i>Pipeline</i> AE*	0,844	0,013
<i>Pipeline</i> ASC†	0,760	0,016

* Métrica de avaliação: Medida-F.

† Métrica de avaliação: Bacc (Acurácia Balanceada).

Tabela 3: Resultados do treinamento com validação cruzada ($k = 5$).

junto de teste para a sub tarefa AE. Na mesma tabela, também são apresentados os resultados obtidos pelos competidores da ASBAPT 2024. Uma análise dos dados indica que o aumento no número de classes eleva significativamente a complexidade da tarefa. Essa observação é sustentada pela comparação de desempenho: a abordagem *pipeline* AE, que opera com um esquema de 3 classes (O, B-ASP, I-ASP), superou a abordagem E2E de 3 classes, a qual exige a classificação de *tokens* em 7 classes distintas para incluir também a polaridade (O, B-POS, I-POS, B-NEG, I-NEG, B-NEU, I-NEU).

Abordagem	F1	P	R
E2E	0,774	0,957	0,650
<i>Pipeline</i> AE	0,840	0,978	0,737
ABCD ¹	0,637	0,855	0,730

Tabela 4: Resultados obtidos na sub tarefa de AE. ¹Resultado extraído do artigo que relata a competição ABSAPT 2024 (Bender et al., 2024).

A Tabela 5 exibe os resultados obtidos no conjunto de teste para a sub tarefa ASC. Na mesma tabela também são apresentados os resultados obtidos pelos competidores da ASBAPT 2024.

Abordagem	Bacc	F1	P	R
<i>Pipeline</i> ASC	0,766	0,762	0,759	0,766
UFPR ¹	0,652	0,653	0,656	0,652
ABCD ¹	0,571	0,568	0,569	0,571

Tabela 5: Resultados obtidos na sub tarefa de ASC. ¹Resultado extraído do artigo que relata a competição ABSAPT 2024 (Bender et al., 2024).

A Tabela 6 apresenta os resultados obtidos da avaliação composta. Observa-se que as abordagens E2E e *pipeline* atingiram valores de Medida-F muito semelhantes (0,725 e 0,721, respectivamente), evidenciando um claro *trade-off*. A abordagem E2E obteve maior Precisão, enquanto a abordagem *pipeline* obteve maior Revocação. Essa paridade no resultado final é particularmente notável

ao recordar que, na sub tarefa de AE avaliada isoladamente, a abordagem *pipeline* demonstrou um desempenho consideravelmente superior. Pode-se inferir, portanto, que a abordagem E2E compensa sua menor capacidade de AE com uma acurácia balanceada maior na sub tarefa de ASC.

Abordagem	F1	P	R
E2E	0,725	0,754	0,698
<i>Pipeline</i>	0,721	0,702	0,741

Tabela 6: Resultados obtidos da avaliação composta.

6 Considerações Finais e Limitações

Este trabalho se propôs a desenvolver e avaliar um sistema de ABSA para a língua portuguesa, utilizando a arquitetura *Transformer* por meio da técnica de *transfer learning*. O objetivo foi adaptar modelos pré-treinados para as sub tarefas AE e ASC, explorando a viabilidade das abordagens *pipeline* e E2E em um contexto no qual ainda há poucos recursos, que é o caso da língua portuguesa.

Para alcançar esse objetivo, foram implementadas duas abordagens distintas: uma em formato *pipeline*, que executa as sub tarefas AE e ASC de forma sequencial, e outra no formato E2E, que realiza as duas sub tarefas simultaneamente. Os resultados obtidos utilizando o modelo pré-treinado BERTimbau (Souza et al., 2020) e o conjunto de dados disponibilizado na competição ABSAPT 2024 (Bender et al., 2024) (composto por avaliações do setor hoteleiro) demonstraram que a aplicação de modelos *Transformer* para ABSA em português é uma estratégia promissora. Os resultados foram competitivos em relação aos apresentados na competição ABSAPT 2024, indicando robustez e boa capacidade de generalização.

Apesar dos resultados positivos, o estudo apresenta limitações. A principal é o foco em um domínio específico. Além disso, o trabalho aborda apenas opiniões diretas e aspectos explícitos. Não são abordadas opiniões comparativas e aspectos implícitos, frequentemente presentes em textos opinativos.

Como desdobramentos futuros, diversas direções podem ser exploradas. Dentre elas podemos citar: a investigação de técnicas de aumento de dados, incluindo estratégias baseadas em grandes modelos de linguagem, para mitigar o desbalanceamento de classes; a avaliação do desempenho de outros modelos de linguagem pré-treinados, tanto

monolíngues quanto multilíngues; o efeito de um pré-treinamento adicional com um *corpus* específico do domínio antes do *fine-tuning*; o desenvolvimento de métodos para AE e ASC em opiniões comparativas e aspectos implícitos, os quais representam desafios importantes na área.

Referências

- Alexandre Thurow Bender, Gabriel A Gomes, Emerson P Lopes, Ricardo M Araujo, Larissa A de Freitas, e Ulisses B Corrêa. 2024. Overview of absapt at iberlef 2024: Overview of the task on aspect-based sentiment analysis in portuguese. *Procesamiento del Lenguaje Natural*, 73:315–322.
- Ulisses Brisolará Corrêa. 2021. *Análise de Sentimento Baseada em Aspectos Usando Aprendizado Profundo: uma Proposta Aplicada à Língua Portuguesa*. Tese (doutorado em ciência da computação), Universidade Federal de Pelotas, Porto Alegre. Disponível em <http://guaiaca.ufpel.edu.br/handle/prefix/7433>. Acesso em: 2025-08-29.
- Júlia da Rocha Junqueira, Claudio Luis Junior, Félix Leonel V Silva, Ulisses Brisolará Corrêa, e Larissa A de Freitas. 2023. Albertina in action: an investigation of its abilities in aspect extraction, hate speech detection, irony detection, and question-answering. Em *Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL)*, páginas 146–155, Belo Horizonte. SBC, SBC.
- Felix LV da Silva, Guilherme da S Xavier, Heliks M Mensenburg, Rodrigo F Rodrigues, Leonardo P dos Santos, Ricardo M Araújo, Ulisses B Corrêa, e Larissa A de Freitas. 2022. Absapt 2022 at iberlef: Overview of the task on aspect-based sentiment analysis in portuguese.
- Bruno Dal Pontte, Anderson Aparecido do Carmo Frasnão, Marcus Vinícius Reisdöfer Pereira, Tiago Heinrich, e Vinicius Fulber-Garcia. 2024. Teamufpr at absapt 2024: Improving sentiment polarity classification with data augmentation using chatgpt. Em *Proceedings of the 6th Iberian Languages Evaluation Forum (IberLEF 2024)*, CEUR Workshop Proceedings, Valladolid, Spain. CEUR-WS.org.
- Larissa Astrogildo de Freitas. 2015. *Feature-level sentiment analysis applied to brazilian portuguese reviews*. Tese (doutorado em ciência da computação), Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre. Disponível em <https://tede2.pucrs.br/tede2/handle/tede/6031>. Acesso em: 2025-08-24.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, e Kristina Toutanova. 2019. *Bert: Pre-training of deep bidirectional transformers for language understanding*. *Preprint*, arXiv:1810.04805.
- Gabriel Almeida Gomes. 2024. *Avaliando a utilização de redes de grafos com atenção para a tarefa de análise de sentimentos em nível de aspecto em línguas de baixo recurso*. Tese (mestrado em ciência da computação), Universidade Federal de Pelotas, Pelotas, Novembro. Disponível em <https://guaiaca.ufpel.edu.br/handle/prefix/15627>. Acesso em: 2025-08-24.
- Juliana Resplande Santanna Gomes, Eduardo Augusto Santos Garcia, Adalberto Ferreira Barbosa Junior, Ruan Chaves Rodrigues, Diogo Fernandes Costa Silva, Dyonatan Ferreira Maia, Nádia Félix Felipe da Silva, Anderson da Silva Soares, and 1 others. 2023. Deep learning brasil at absapt 2022: Portuguese transformer ensemble approaches. *arXiv preprint arXiv:2311.05051*.
- Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, San Rafael.
- Émerson Philippe Lopes, Larissa Freitas, Gabriel Gomes, Gerônimo Lemos, Luiz Otávio Hammes, e Ulisses B Corrêa. 2022. Exploring bert for aspect-based sentiment analysis in portuguese language. Em *The International FLAIRS Conference Proceedings*, volume 35, Daytona Beach, FL. Florida Online Journals.
- Huaishao Luo, Tianrui Li, Bing Liu, e Junbo Zhang. 2019. *DOER: Dual cross-shared RNN for aspect term-polarity co-extraction*. Em *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, páginas 591–601, Florence, Italy. Association for Computational Linguistics.
- Fábio Souza, Rodrigo Nogueira, e Roberto Lotufo. 2020. *Bertimbau: Pretrained bert models for brazilian portuguese*. Em *Intelligent Systems: 9th Brazilian Conference, BRACIS 2020, Rio Grande, Brazil, October 20–23, 2020, Proceedings, Part I*, página 403–417, Berlin, Heidelberg. Springer-Verlag.
- Dang Van Thin, Nguyen Tuan Kiet, Duong Ngoc Hao, e Ngan Luu-Thuy Nguyen. 2024. Abcd team at absapt 2024: Classification-based versus generation-based approach for aspect-based sentiment analysis in portuguese. Em *Proceedings of the 6th Iberian Languages Evaluation Forum (IberLEF 2024)*, CEUR Workshop Proceedings, Valladolid, Spain. CEUR-WS.org.
- Qihuang Zhong, Haiyun Li, Luyao Zhuang, Juhua Liu, e Bo Du. 2024. Iterative data generation with large language models for aspect-based sentiment analysis. *arXiv preprint arXiv:2407.00341*.