

**UNIVERSIDADE DE SÃO PAULO**

Instituto de Ciências Matemáticas e de Computação

**ConSentiment: Análise de Sentimentos baseada em Aspectos  
a partir de Consenso de Modelos em Grafos**

**Guilherme dos Santos Marcon**

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências  
de Computação e Matemática Computacional (PPG-C<sup>2</sup>MC)



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Guilherme dos Santos Marcon**

## ConSentiment: Análise de Sentimentos baseada em Aspectos a partir de Consenso de Modelos em Grafos

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Ricardo Marcondes Marcacini

**USP – São Carlos**  
**Junho de 2024**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados inseridos pelo(a) autor(a)

M321c Marcon, Guilherme dos Santos  
ConSentiment: Análise de Sentimentos baseada em  
Aspectos a partir de Consenso de Modelos em Grafos  
/ Guilherme dos Santos Marcon; orientador Ricardo  
Marcondes Marcacini. -- São Carlos, 2024.  
69 p.

Dissertação (Mestrado - Programa de Pós-Graduação  
em Ciências de Computação e Matemática  
Computacional) -- Instituto de Ciências Matemáticas  
e de Computação, Universidade de São Paulo, 2024.

1. Análise de Sentimentos. 2. Comitê de Modelos.  
3. Grafos. 4. Mineração de Textos. I. Marcacini,  
Ricardo Marcondes, orient. II. Título.

**Guilherme dos Santos Marcon**

**ConSentiment: Consensus-based Sentiment Analysis with  
Graph Learning**

Dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP – in accordance with the requirements of the Computer and Mathematical Sciences Graduate Program, for the degree of Master in Science. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Ricardo Marcondes Marcacini

**USP – São Carlos**  
**June 2024**



*Dedico este trabalho a todos que, de alguma forma, contribuíram para o meu crescimento pessoal e acadêmico. Que este estudo possa servir como uma pequena expressão de gratidão por todo o apoio e incentivo recebidos ao longo desta jornada.*





# AGRADECIMENTOS

---

---

Gostaria de expressar meus sinceros agradecimentos a todas as pessoas e instituições que tornaram possível a realização deste trabalho de mestrado.

Primeiramente, quero expressar minha profunda gratidão ao meu orientador, Ricardo, pela paciência infinita e apoio constante demonstrados neste longo percurso acadêmico. Seu conhecimento e conselhos foram fundamentais para o desenvolvimento deste trabalho.

Agradeço também aos meus queridos familiares e amigos, que sempre me apoiaram e compreenderam as exigências e desafios deste período. Sua presença e apoio foram essenciais para minha jornada.

Não posso deixar de expressar minha sincera gratidão à CAPES por disponibilizar a bolsa que possibilitou minha dedicação integral aos estudos durante todo o curso. Sua contribuição foi fundamental para meu crescimento acadêmico e profissional.

Por fim, gostaria de agradecer a todos que estejam lendo este documento, mesmo que eu não os conheça. Obrigado.



*“As invenções são, sobretudo,  
o resultado de um trabalho de teimoso.”  
(Santos Dumont)*



# RESUMO

MARCON, G. S. **ConSentiment: Análise de Sentimentos baseada em Aspectos a partir de Consenso de Modelos em Grafos**. 2024. 69 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

Construir conjuntos de modelos de Análise de Sentimento Baseada em Aspectos (ABSA) pode ser um desafio. Cada modelo pode extrair aspectos distintos, onde alguns poderiam estar parcialmente corretos, e prever diferentes sentimentos associados a cada aspecto. A complexidade dos resultados é um dos principais problemas na construção de um consenso nesta configuração de aprendizagem. Além disso, a ABSA enfrenta desafios associados à eliminação de ambiguidades e ao discernimento de sentimentos em diversos domínios e contextos linguísticos. Este cenário é particularmente predominante na análise de produtos, serviços ou tópicos globalmente relevantes. Este projeto apresenta o ConSentiment (Análise de Sentimento Baseada em Consenso) como uma nova abordagem para enfrentar esses desafios. Propomos os princípios de aprendizagem do conjunto ABSA, denominados Grafo de Informação de Consenso e Redução de Incerteza, para encontrar um consenso entre os modelos básicos do ABSA. Em primeiro lugar, o ConSentiment sugere um grafo de consenso conectando revisões, aspectos e rótulos previstos semelhantes, enfatizando a concordância entre os modelos básicos. O grafo lida com a incerteza ao conectar também avaliações com aspectos semelhantes, mas sentimentos diferentes, gerenciando efetivamente a ambiguidade nas tarefas do ABSA. Em segundo lugar, o ConSentiment introduz uma estrutura de regularização ABSA baseada em grafos que prioriza informações consensuais, onde os nós com maior consenso compartilham mais fortemente as informações do rótulo com os vizinhos, reduzindo a incerteza onde os modelos básicos discordam sobre o sentimento. Avaliações experimentais em cenários multidomínios e multilíngues mostram que o ConSentiment supera os modelos individuais de base ABSA, incluindo os mais recentes baseados em BERT e GPT, bem como modelos de votação baseados em comitês.

**Palavras-chave:** Análise de Sentimentos, Redes Neurais Profundas, Modelos Multilíngues, Ensemble Learning, Regularização de Grafos.



# ABSTRACT

MARCON, G. S. **ConSentiment: Consensus-based Sentiment Analysis with Graph Learning**. 2024. 69 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

Building ensembles of Aspect-Based Sentiment Analysis (ABSA) models can be challenging. Each model might extract distinct aspects, where some could be partly correct, and predict different sentiments associated with each aspect. The output complexity is one of the core problems in building a consensus in this learning setup. Furthermore, ABSA confronts challenges associated with disambiguating aspects and discerning sentiments within diverse domains and linguistic contexts. This scenario is particularly prevalent in analyzing globally relevant products, services, or topics. This project introduces ConSentiment (Consensus-Based Sentiment Analysis) as a new approach to address these challenges. We propose ABSA ensemble learning principles, called Consensus Information Graph and Uncertainty Reduction, to find a consensus among ABSA base models. Firstly, ConSentiment suggests a consensus graph by connecting similar reviews, aspects, and predicted labels, emphasizing agreement among base models. The graph deals with uncertainty by also connecting reviews with similar aspects but different sentiments, effectively managing the ambiguity in ABSA tasks. Secondly, ConSentiment introduces a graph-based ABSA regularization framework that prioritizes consensual information, where nodes with higher consensus more strongly share label information with neighbors, reducing uncertainty where base models disagree on sentiment. Experimental evaluations across multidomain and multilingual scenarios show that ConSentiment outperforms individual ABSA base models, including more recent ones based on BERT and GPT, as well as committee-based voting models.

**Keywords:** Sentiment Analysis, Deep Neural Networks, Multilingual Models, Ensemble Learning, Graph Regularization.





# LISTA DE ILUSTRAÇÕES

---

---

Figura 1 – Tipos de abordagem para Análise de Sentimentos Baseada em Aspectos. . . . .	29
Figura 2 – Arquiteturas ABSA baseadas em <i>pipeline</i> com um modelo em cada tarefa. . . . .	30
Figura 3 – Arquiteturas de Análise de Sentimentos Baseada em Aspectos <i>End-to-End</i> . . . . .	35
Figura 4 – Arquitetura envolvendo múltiplas tarefas. . . . .	36
Figura 5 – Exemplo de uma arquitetura E2E-ABSA <i>Joint</i> . . . . .	37



# LISTA DE TABELAS

---

---

Tabela 1 – Pré-processamento de uma frase até sua representação inicial com o BERT. . . . .	33
Tabela 2 – Estrutura BIEOS unificada para E2E-ABSA. EA é a tarefa de Extração de Aspectos, AS de Análise de Sentimentos. . . . .	36
Tabela 3 – Tabela comparativa com trabalhos relacionados na literatura.	43
Tabela 4 – Características dos Conjuntos de Dados de ABSA . . . . .	53
Tabela 5 – Resultados experimentais (pontuações F1) com <i>reviews</i> de um único domínio e apenas em inglês . . . . .	56
Tabela 6 – Resultados experimentais (pontuações F1) com cenários multilíngues e entre domínios . . . . .	56
Tabela 7 – Resultados experimentais (pontuações F1) com a incorporação de LLMs nos modelos base e ConSentiment . . . . .	57



# SUMÁRIO

---

1	INTRODUÇÃO . . . . .	21
1.1	Contextualização . . . . .	21
1.2	Objetivos e Questões de Pesquisa . . . . .	23
1.3	Organização do Documento . . . . .	24
2	FUNDAMENTOS E TRABALHOS RELACIONADOS . . . . .	27
2.1	Mineração de Opiniões . . . . .	27
2.1.1	<i>Análise de Sentimentos Baseada em Aspectos</i> . . . . .	28
2.2	Representação Textual usando Word Embeddings . . . . .	30
2.2.1	<i>Word2Vec</i> . . . . .	31
2.2.2	<i>ELMO</i> . . . . .	31
2.2.3	<i>BERT</i> . . . . .	32
2.3	Arquiteturas ABSA <i>End-to-End</i> . . . . .	35
2.3.1	<i>Arquitetura Collapsed</i> . . . . .	35
2.3.2	<i>Arquitetura Joint</i> . . . . .	36
2.4	Comitê de Modelos . . . . .	37
2.5	Frameworks de Regularização . . . . .	38
2.6	Trabalhos Relacionados . . . . .	40
3	MÉTODO PROPOSTO: CONSENTIMENT . . . . .	45
3.1	Grafo de Consenso Baseado em ABSA . . . . .	45
3.2	Regularização Baseada em Consenso de ABSA . . . . .	47
3.2.1	<i>Complexidade do Algoritmo</i> . . . . .	49
3.2.2	<i>Ferramenta Computacional</i> . . . . .	50
4	AVALIAÇÃO EXPERIMENTAL . . . . .	53
4.1	Conjuntos de Dados . . . . .	53
4.2	Configuração Experimental . . . . .	54
4.3	Resultados e Discussão . . . . .	55
5	CONCLUSÃO . . . . .	59
	REFERÊNCIAS . . . . .	63



---

# INTRODUÇÃO

---

## 1.1 Contextualização

Com o avanço das redes sociais e sua integração na sociedade, as opiniões das pessoas tornaram-se cada vez mais acessíveis, desde opiniões em redes de amigos, análises de produtos e serviços até comentários sobre celebridades e políticos. Do ponto de vista corporativo, é interessante observar o que as pessoas falam sobre seus produtos e serviços. Por conta da imensidão de dados, técnicas de Mineração de Opiniões têm se tornado cada vez mais frequentes no dia-a-dia empresarial (LIU, 2020). Por meio dessas técnicas, as corporações podem entender a opinião de seus clientes sobre produtos, serviços e até mesmo sobre a empresa em si, por meio da análise de grandes volumes de textos opinativos (LIU; ZHANG, 2012; BIRJALI; KASRI; BENI-HSSANE, 2021).

Embora existam diferentes técnicas para Análise de Sentimentos, é possível organizá-las em três categorias (LIU, 2020) quando consideramos o nível de detalhamento dessa análise: em nível de documentos, em nível de sentenças ou em nível de aspectos. O cerne deste projeto é a Análise de Sentimentos Baseada em Aspectos (ABSA). Nesse nível, o objetivo é, a partir de um comentário sobre um produto, determinar o sentimento relacionado a certas partes do texto. Por exemplo, em "O processador é bom, mas a memória deixa a desejar", nós, como humanos, conseguimos identificar que "processador" tem uma conotação positiva, enquanto "memória" possui uma conotação negativa. Essa análise, caso presente em vários comentários, indica para a empresa que a memória do produto precisa de uma atenção maior. Ao contrário dos níveis baseados em documentos ou sentenças, que inferem o sentimento (por exemplo, positivo, negativo ou neutro) do texto de forma geral, no ABSA há um desafio extra que envolve identificar os aspectos dos textos opinativos, como

"processador" e "memória" no exemplo acima, e só então realizar a associação de aspecto com o sentimento considerando o contexto da sentença (LI *et al.*, 2019b).

Os métodos considerados estado-da-arte para ABSA são atualmente baseados em redes neurais profundas (DO *et al.*, 2019), conhecidas pela necessidade de um grande volume de dados para treinamento (DONG; WANG; ABBAS, 2021). No caso da Mineração de Opiniões, vale destacar também o uso de modelos pré-treinados para representação de textos, como o BERT (KENTON; TOUTANOVA, 2019) e o GPT (BROWN *et al.*, 2020), treinados por meio de uma tarefa de ajuste fino em análise de sentimentos baseada em aspectos, por exemplo. Outro fator importante é que a grande maioria dos métodos estado-da-arte para ABSA é dividida em duas tarefas: (1) Extração de Aspectos e (2) Análise de Sentimentos. Na prática, isso implica o uso de dois modelos de redes neurais profundas, um para cada tarefa. Vale destacar que, em geral, a tarefa de extração de aspectos não considera informações de rótulos de sentimentos. Uma vez extraídos os aspectos de um texto, as listas de aspectos e suas respectivas sentenças onde ocorrem são utilizadas como entrada para um modelo de análise de sentimentos (ZHANG *et al.*, 2021). Por outro lado, uma abordagem recente denominada modelos ABSA *End-to-End* ou E2E-ABSA (LI *et al.*, 2019b) foi proposta para combinar o processo de extração de aspectos e o processo de análise de sentimento, argumentando que as duas tarefas são complementares e podem melhorar o desempenho se consideradas em conjunto durante a etapa de treinamento.

O estudo apresentado por Li *et al.* (2019b) foi um dos pioneiros no contexto de E2E-ABSA e sua integração com modelos de representação textual mais recentes, como o BERT, para realizar a extração de aspectos e análise de sentimentos de forma unificada. Os resultados apresentados foram promissores em comparação com iniciativas anteriores, aumentando assim a atenção de pesquisadores para essas abordagens (LI *et al.*, 2019; HE *et al.*, 2019; MAO *et al.*, 2021; BIE; YANG, 2021). Consequentemente, diferentes arquiteturas de redes neurais para E2E-ABSA foram propostas, bem como novos experimentos, que são majoritariamente focados na língua inglesa. Já recursos multilíngues podem ser adicionados por meio do modelo de representação textual pré-treinado utilizado por essas arquiteturas. Um modelo popular é o Bert Multilíngue (mBERT), que foi pré-treinado considerando 104 línguas a partir de textos da Wikipedia. Identificar como realizar um ajuste fino desse modelo visando melhorar o desempenho na língua alvo com poucos dados anotados é um desafio de pesquisa.



Apesar do progresso alcançado, ainda há desafios a serem superados na área de ABSA. A diversidade de línguas, domínios e contextos torna a adaptação e generalização dos modelos uma tarefa complexa. Além disso, a falta de anotações detalhadas em aspectos e sentimentos para muitas línguas e domínios dificulta a criação de modelos eficazes para diferentes cenários. Nesse sentido, há uma necessidade crescente de avaliar quais estratégias têm obtido resultados mais satisfatórios, principalmente os modelos *End-to-End* e seu impacto em cenários multilíngues.

Uma estratégia promissora investigada neste projeto é o uso de comitê de modelos, ou *Ensemble Learning*, como uma alternativa para lidar com a variabilidade de desempenho entre diferentes modelos e cenários. A combinação de múltiplos modelos pode proporcionar uma visão mais abrangente e robusta da complexidade presente nos dados, resultando em previsões mais consistentes por meio de um consenso. Especificamente em Análise de Sentimentos Baseada em Aspectos (ABSA), a diversidade de línguas, domínios e expressões coloquiais pode levar a desafios significativos de generalização. A formação de comitês permite explorar a complementaridade entre diferentes modelos, aproveitando seus pontos fortes e compensando suas fraquezas. Portanto, a implementação de modelos comitê neste projeto visa não apenas otimizar a performance geral dos modelos de E2E-ABSA em cenários multilíngues, mas também fortalecer a capacidade de generalização para diferentes contextos.

## 1.2 Objetivos e Questões de Pesquisa

Neste projeto de mestrado, além do objetivo relacionado ao estudo e avaliação de diferentes modelos End-to-End ABSA para cenários multilíngues, há o objetivo de explorar estratégias de comitê para obter um resultado consensual a partir desses modelos. No entanto, métodos tradicionais de comitê, embora eficazes em muitos contextos, podem ser inadequados para abordar diretamente problemas específicos de ABSA. Essas abordagens frequentemente se baseiam em estratégias de votação, que podem não lidar de maneira ideal com situações em que os modelos concordam no aspecto, mas discordam no sentimento, ou quando a identificação do aspecto varia na quantidade de tokens para uma revisão. Diante disso, este projeto propõe uma nova estratégia de comitê baseada em grafos, integrada a um framework de regularização.

A estratégia de construção de grafos proposta é orientada por dois princípios fundamentais:

**Informação Consensual:** Quando vértices apresentam alta similaridade

entre revisão e aspecto, com previsões de sentimentos semelhantes, evidenciase um consenso entre os modelos-base. Esta abordagem visa explorar o consenso e disseminar informações com menor incerteza para áreas de maior ambiguidade no grafo.

**Redução de Incerteza:** Se dois vértices compartilham o mesmo aspecto identificado pelo modelo base, mas apresentam sentimentos diferentes, uma aresta é estabelecida entre eles. Essa conexão força o modelo de consenso a determinar o sentimento final do aspecto na presença de incerteza, utilizando a topologia da rede e a regularização. O consenso obtido na regra anterior é aproveitado para classificar o sentimento em situações de divergência entre os modelos-base.

Diante desse contexto, este projeto de mestrado formula as seguintes questões de pesquisa:

1. **Quais são os principais modelos End-to-End para ABSA e suas eficácias em cenários multilíngues?**

Foram investigados os principais modelos End-to-End baseados em BERT e técnicas de ajuste fino, incluindo a incorporação do uso de prompts para ABSA no contexto de Large Language Models (ChatGPT 3.5).

2. **Como combinar efetivamente os resultados de diferentes modelos End-to-End ABSA para obter um modelo consensual mais robusto?**

Foi proposto um framework de regularização para grafos representando o comitê de diferentes modelos. Este método permite utilizar instâncias de maior consenso para reduzir a incerteza em situações de divergência entre os modelos. Os resultados experimentais obtidos mostraram que o método proposto superou os modelos-base e o método de comitê baseado em votação.

## 1.3 Organização do Documento

Este texto está organizado da seguinte maneira:

O Capítulo 2 contextualiza o campo de Mineração de Opiniões, a tarefa de Análise de Sentimentos Baseada em Aspectos e os trabalhos relacionados. Ele apresenta os conceitos gerais da área e entra em mais detalhes nos elementos mais importantes para o projeto: BERT, arquiteturas E2E-ABSA e o comitê de modelos.

O Capítulo 3 discute mais a fundo sobre o método proposto, denominado ConSentiment, a metodologia utilizada, focando em como os grafos foram criados, modelos-base utilizados, análise de parâmetros e informações sobre a ferramenta criada.

O Capítulo 4 apresenta uma avaliação experimental envolvendo diferentes modelos ABSA, com detalhes sobre os conjuntos de dados, os critérios de avaliação e uma discussão detalhada dos resultados.

Por fim, o Capítulo 5 descreve as principais contribuições do trabalho, as limitações, além de possíveis trabalhos futuros.



---

## FUNDAMENTOS E TRABALHOS RELACIONADOS

---

### 2.1 Mineração de Opiniões

Mineração de Opiniões, ou Análise de Sentimentos, é um campo na área de Processamento de Linguagem Natural que busca analisar as opiniões das pessoas, suas atitudes e sentimentos em relação a algo, seja outra pessoa, um produto, serviço ou acontecimento (CORTIS; DAVIS, 2021).

Na indústria, Mineração de Opiniões processa textos de opinião em grande escala, suprimindo, por exemplo, a demanda das empresas de quererem saber o que as pessoas estão dizendo sobre seus produtos e serviços. E como a quantidade de opiniões é enorme, é inviável fazê-lo manualmente, tornando cada vez mais comuns métodos automáticos, especialmente métodos baseados em aprendizado de máquina (LIU, 2020). Dentre os tipos de dados opinativos, o escopo deste projeto envolve dados textuais, uma vez que com a imensa popularidade da internet, redes sociais e sites de compras, a disponibilidade de textos que expressam a opinião de alguém sobre algo se tornou ampla e acessível.

A Análise de Sentimentos pode ser realizada em três diferentes níveis (LIU, 2020): baseada em documentos, baseada em sentenças e baseada em aspectos. Nos níveis de análise baseada em documentos ou sentenças, o objetivo é identificar a polaridade do sentimento para o documento ou para uma sentença como um todo, respectivamente. No entanto, esses níveis não são adequados em situações em que o usuário apresenta comentários em diferentes características do alvo (produto ou serviço). Por exemplo, na frase “O processador é rápido, mas a memória deixa a desejar”, nós podemos identificar que o processador possui um sentimento positivo (por ser “rápido”) e que a

memória possui um sentimento negativo (por “deixar a desejar”). Observe que o alvo (notebook) é analisado considerando dois aspectos: processador e memória. Esse é o nível de análise baseada em aspectos (DO *et al.*, 2019; NAZIR *et al.*, 2020; TRISNA; JIE, 2022), que é foco deste projeto de mestrado.

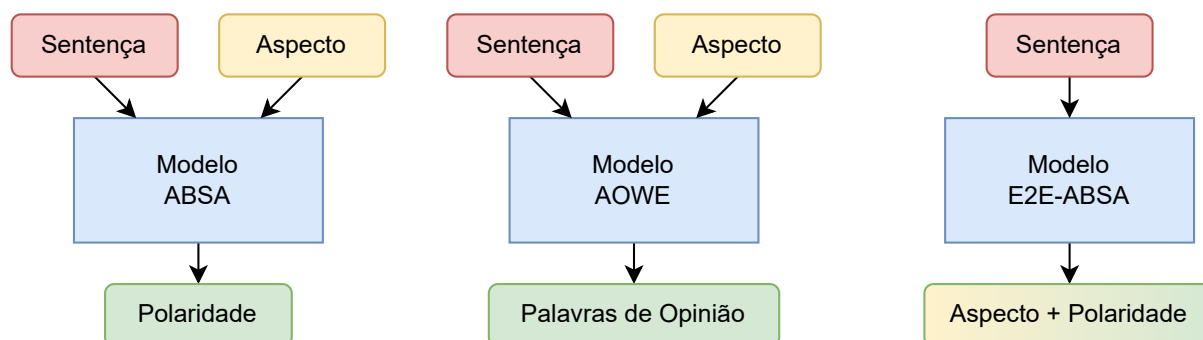
Essa Análise de Sentimentos Baseada em Aspectos (ABSA) geralmente requer a identificação prévia dos aspectos de um texto, que pode ser realizada por meio de modelos de Reconhecimento de Entidades Nomeadas (*Named Entity Recognition* - NER). Essa tarefa é responsável por identificar em textos quais partes indicam pessoas, organizações, tempo, etc. Yadav e Bethard (2018), Li *et al.* (2022) apresentaram *surveys* detalhados sobre a tarefa, mas um dos primeiros artigos na área foi sobre extração de nomes de empresas (RAU, 1991), com um sistema de heurísticas e regras manualmente criadas para esse problema em específico. A área então passou por uma evolução similar à outras áreas de aprendizado de máquina, utilizando técnicas de aprendizado não-supervisionado (COLLINS; SINGER, 1999; ZHANG; ELHADAD, 2013), de aprendizado supervisionado como Decision Trees (SZARVAS; FARKAS; KOC-SOR, 2006), Support Vector Machines (SVM) (MCNAMEE; MAYFIELD, 2002) e Conditional Random Field (CRF) (MCCALLUM; LI, 2003) e, mais recentemente, de aprendizado profundo como Convolutional Neural Networks (CNN) (COLLOBERT *et al.*, 2011), Long Short-Term Memory (LSTM) (KURU; CAN; YURET, 2016) e Bidirectional Encoder Representations from Transformers (BERT) (LI *et al.*, 2020; ALI *et al.*, 2022).

Voltando para ABSA, Liu (2020) define formalmente uma opinião como uma quintupla  $(e, a, s, h, t)$ , no qual  $e$  é a entidade alvo,  $a$  é o aspecto da entidade no qual a opinião foi dado,  $s$  é o sentimento da entidade em relação ao aspecto,  $h$  é o autor da opinião e  $t$  é o tempo que a opinião foi dada. No entanto, nem todos os conjuntos de dados contém a quintupla inteira, é normalmente omitido  $h$  e  $t$ , portanto definirei uma opinião apenas como  $(e, a, s)$ . Na próxima seção, são descritos mais detalhes sobre ABSA.

### 2.1.1 Análise de Sentimentos Baseada em Aspectos

No âmbito deste projeto, uma sentença será sinônimo de uma ou mais frases da qual existem aspectos e sentimentos/polaridades a serem extraídos. Um aspecto é uma ou mais palavras que compõem um conceito, podendo ou não possuir um sentimento relacionado na sentença. Portanto a quintupla de Liu (2020) para este projeto se dá por:  $e$  é uma sentença,  $a$  é o aspecto,  $s$  é um valor discreto positivo, neutro ou negativo. Os termos  $h$  e  $t$  são opcionais, podendo não estar presentes nos conjuntos de dados e, no escopo dos métodos

Figura 1 – Tipos de abordagem para Análise de Sentimentos Baseada em Aspectos.



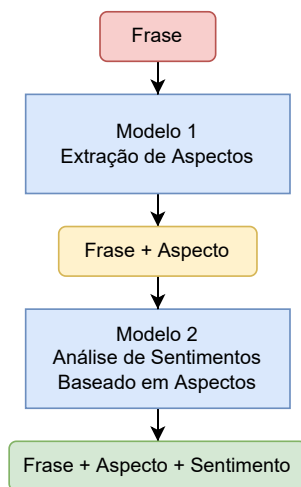
Fonte: Autor.

investigados neste projeto, não são utilizados para o desenvolvimento das técnicas e para a resolução do problema de ABSA.

De acordo com [Li et al. \(2019b\)](#), existem diferentes maneiras de lidar com ABSA, ilustradas na Figura 1. A estratégia tradicional de Análise de Sentimentos Baseada em Aspectos (ABSA) conforme proposta por [Mitchell et al. \(2013\)](#), [Zhang, Zhang e Vo \(2015\)](#), onde a partir de um texto e um aspecto, se extrai a polaridade da sentença em relação ao aspecto. Também há trabalhos que extraem palavras de opinião orientadas à aspectos conforme proposto por [Fan et al. \(2019\)](#), também chamado de AOWE (do inglês *Aspect Oriented Opinion Words Extraction*) que extrai palavras que demonstram a opinião do texto em relação ao aspecto. Por fim, a Análise de Sentimentos Baseada em Aspectos *End-to-End* (E2E-ABSA) ([MA; LI; WANG, 2018](#); [SCHMITT et al., 2018](#); [LI et al., 2019b](#); [WANG et al., 2021](#); [CHEN; YANG; HUANG, 2022](#)), em que a partir de uma sentença, se extrai tanto o aspecto quanto o sentimento relacionado a ele.

Na Figura 2 é apresentado um *pipeline* padrão para o ABSA e AOWE, em que a partir de cada frase do conjunto de dados, são extraídos seus possíveis aspectos com um modelo, depois a partir de cada tupla frase + aspecto, são inferidos os sentimentos da frase em relação aos aspectos.

Esse processo possui algumas limitações ([MA; LI; WANG, 2018](#); [SCHMITT et al., 2018](#); [LI et al., 2019b](#); [WANG et al., 2021](#); [CHEN; YANG; HUANG, 2022](#)), sendo as principais a necessidade de manter a manutenção de dois modelos diferentes, a propagação de erros de um modelo para o outro, bem como desconsiderar a complementaridade entre aspectos e sentimentos, lidando com a etapa de extração de aspectos sem considerar de forma mais direta os sentimentos associados. Por exemplo, se o modelo de aspectos extrair erroneamente, o modelo de sentimentos vai propagar esse erro, fazendo chegar ao usuário final um aspecto com sentimento errado, onde na verdade não faria

Figura 2 – Arquiteturas ABSA baseadas em *pipeline* com um modelo em cada tarefa.

Fonte: Autor.

nem sentido aquele aspecto ter sido extraído.

Já o E2E-ABSA condensa esses dois modelos em apenas um para mitigar essas limitações. Por outro lado, aumenta também a complexidade do problema (LI *et al.*, 2019b; WANG *et al.*, 2021; CHEN; YANG; HUANG, 2022), já que um modelo só precisa realizar o que outros dois desempenhavam. A Seção 2.3 discute de forma mais detalhada as arquiteturas específicas para Análise de Sentimentos Baseada em Aspectos *End-to-End*.

## 2.2 Representação Textual usando Word Embeddings

Como estamos lidando com modelos de redes neurais, precisamos de uma entrada estruturada, ao contrário do comportamento não-estruturado de textos. O comum é transformar cada palavra em vetores de características, processo chamado de Word Embedding, que tenta representar o texto em vetores de números reais (NASEEM *et al.*, 2021; GUPTA; AGRAWAL, 2022). Tal representação possibilita computar a similaridade entre palavras conforme a correlação entre seus respectivos vetores, bem como a similaridade entre sentenças e documentos.

Atualmente, métodos para representações de Word Embedding podem ser divididos em duas categorias (NASEEM *et al.*, 2021): sem contexto e com contexto. Uma embedding sem contexto, como Word2Vec (MIKOLOV *et al.*, 2013) e GloVe (PENNINGTON; SOCHER; MANNING, 2014), geram vetores em que uma palavra terá a mesma representação, após o treinamento, não importando como ela é utilizada na frase. Enquanto que a com contexto, como ELMo (PETERS *et al.*, 2018) e BERT (KENTON; TOUTANOVA, 2019), a representação



pode mudar dependendo do uso da palavra. Essa diferença ocorre na etapa de uso desses modelos, na qual sem contexto são consideradas representações estáticas e com contexto são representações dinâmicas (NASEEM *et al.*, 2021).

Usando como exemplo a palavra “banco” nas frases “sentei no banco”, “abri uma conta no banco” e “eu banco meu filho”, apesar de ser a mesma palavra, possuem sentidos diferentes em cada frase. Nós conseguimos intuitivamente saber a diferença, por já termos anos vivendo e aprendendo elas. Uma representação de Word Embedding sem contexto, após o treinamento, representaria “banco” da mesma maneira nas frases, enquanto que uma com contexto haveria diferença, sendo que as embeddings das palavras são calculadas de forma dinâmica ao apresentar novos textos.

Nas próximas seções é apresentada uma breve descrição dos principais métodos para word embedding, começando com Word2Vec e sua evolução até modelos baseados no BERT, que são utilizados neste projeto.

### 2.2.1 Word2Vec

O Word2Vec (MIKOLOV *et al.*, 2013) é um modelo para aprendizado de representações de palavras, em que as representações (*word vectors*) são estáticas após o treinamento. Existem variações de arquiteturas de rede neural para Word2Vec, como Skip-Gram e CBOW, mas ambas permitem computar *word vectors* de forma que seja possível realizar comparações de similaridade entre as palavras.

Uma palavra ainda possui a mesma representação com Word2Vec, sendo independente de contexto. Os autores do Word2Vec discutem que algumas operações podem ser empregadas com as *word vectors*. Por exemplo, o vetor de “Rainha” seria similar ao de “Rei” menos o de “Homem” mais o de “Mulher”.

Outro ponto importante discutido pelos autores é a escala de dados utilizados, enquanto que trabalhos anteriores não utilizavam mais do que poucos milhões de palavras, o Word2Vec aumentou para bilhões. Isso demonstra a popularidade do Word2Vec, uma vez que as técnicas de aprendizado de máquina da época conseguiam não só lidar com essa quantidade de dados maior, como também aprender mais com ela.

### 2.2.2 ELMO

Depois da grande popularidade do Word2Vec, a utilização de embeddings pré-treinadas se voltou para a criação de representações com contexto, resultando embeddings através de, por exemplo, RNNs bidirecionais (PETERS *et al.*,

2017) e LSTMs (MCCANN *et al.*, 2017), culminando no ELMo (PETERS *et al.*, 2018), que utiliza LSTMs bidirecionais.

A LSTM (Long Short-Term Memory) (HOCHREITER; SCHMIDHUBER, 1997) é uma RNN (*Recurrent Neural Network*) capaz de aprender a dependência de ordem nos dados de entrada, que são evidentes com palavras e sentenças. As RNNs possuem um estado interno que pode representar contexto, funcionando de maneira similar à uma memória. A LSTM foi criada com o objetivo de aumentar essa memória, conseqüentemente aumentando o contexto e influência de palavras mais distantes.

Essa estratégia da LSTM é direcional, ou seja, o contexto de uma palavra só depende das palavras anteriores. Por conta disso, o ELMo e muitas outras estratégias utilizam LSTMs bidirecionais (Bi-LSTM), que são duas camadas de LSTM, uma levando em consideração as palavras anteriores e a outra as posteriores. Os resultados das duas camadas são então concatenados, gerando as representações contextuais.

Entretanto, devido a esse comportamento sequencial e essa dependência rígida, é muito difícil paralelizar o treino de uma LSTM, o que limita a quantidade de dados que pode ser utilizado em tempo hábil.

### 2.2.3 BERT

Pouco tempo após o ELMo, o BERT (Bidirectional Encoder Representations from Transformers) (KENTON; TOUTANOVA, 2019) foi publicado e alcançou o estado-da-arte em 11 problemas diferentes em processamento de linguagem natural. Ele é um modelo baseado em Transformers (VASWANI *et al.*, 2017) pré-treinado em estrutura encoder-decoder, treinado em duas subtasks: Token Mascarado e Predição da Próxima Sentença.

Na entrada, o texto passa por um pré-processamento. Primeiro são extraídos o WordPiece (WU *et al.*, 2016) dos textos, na qual cada palavra (ou subpalavra) da sentença, existente num vocabulário predefinido, é processada de maneira a unificar espaços em branco, separar pontuações, radicais, marcadores de plural, posse, gênero, etc. Assim, as palavras do conjunto de dados são *tokenizadas* e um identificador para cada é associado. Em sequência são incluídos os tokens especiais auxiliares [CLS] e [SEP] respectivamente antes e depois dos tokens da primeira sentença. O token [CLS] será utilizado para representar a embedding das sentenças como um todo, enquanto o token [SEP] identificar separação de sentenças. O embedding final é composto pela soma das embeddings do tokens, as embeddings de tokens que representam qual

Tabela 1 – Pré-processamento de uma frase até sua representação inicial com o BERT.

Frase Inicial	John Johanson's house							
WordPiece		john	johan	##son	'	s	house	
Embedding dos Tokens		[john]	[johan]	[##son]	[']	[s]	[house]	
Tokens Auxiliares	[CLS]	[john]	[johan]	[##son]	[']	[s]	[house]	[SEP]
Embedding de Sentença	[A]	[A]	[A]	[A]	[A]	[A]	[A]	[B]
Embedding de Posição	[0]	[1]	[2]	[3]	[4]	[5]	[6]	[7]

Fonte: <https://github.com/google-research/bert#tokenization>

sentença ele pertence e as embeddings de posição dos tokens. Essa última, embedding de posição, possibilita identificar relações de ordem, porém de forma paralelizável. Esta característica torna o BERT interessante a respeito na quantidade de dados capaz de serem processadas, quando comparado aos modelos anteriores (LSTMs). A Tabela 1 ilustra esse processo com uma sentença.

Após essa etapa, o input passa pelas várias camadas de Transformers (VASWANI *et al.*, 2017), cujo funcionamento é baseado em auto-atenção. Cada embedding de token é aplicada três pesos diferentes, essas três derivações são então multiplicadas matricialmente e escaladas entre si, com o intuito de que o modelo aprenda qual parte daquela embedding é mais importante para resolver o problema. As Transformers envolvem várias dessas camadas de auto-atenção em paralelo, mas também mascaramento de token (AGGARWAL, 2018).

O modelo BERT é pré-treinado em duas tarefas diferentes: a do Token Mascarado e a da Predição da Próxima Sentença. Na do Token Mascarado, cerca de 15% dos tokens de entrada são trocados pelo fixo [MASK] (80% das vezes), um token aleatório (10%) ou não é trocado (10%). Depois, o objetivo do modelo é prever quais são os tokens mascarados, aprendendo relações entre os tokens. Já na tarefa de Predição da Próxima Sentença, são escolhidas duas sentenças A e B, onde 50% das vezes B é a que realmente sucede A, e a finalidade do modelo é prever as duas sentenças, aprendendo relações entre as sentenças.

Dessa maneira, o BERT visa encapsular um conhecimento geral de Processamento de Linguagem Natural na etapa de pré-treinamento. O BERT também é utilizado numa outra etapa conhecida como *fine-tuning*, na qual as embeddings podem ser refinadas de acordo com dados de uma tarefa específica. Por exemplo, na tarefa de análise de sentimentos, as embeddings podem ser refinadas considerando textos anotados como positivos, neutros ou negativos. Esta é uma técnica também conhecida como transferência de aprendizado (HOANG; BIHORAC; ROUCES, 2019; HAO *et al.*, 2020).

Apesar de não estar explícito no artigo original, o BERT também foi treinado em ambientes multilíngue<sup>1</sup>, permitindo seu uso em outros idiomas, não só o inglês. O conjunto de dados de treino multilíngue é composto pelas Wikipédias inteiras dos 100 idiomas mais presentes, com a adição de mais alguns específicos, chegando atualmente a 104 idiomas; eles são normalizados de forma a diminuir o impacto entre utilizar os idiomas com muito volume e ricos em recursos, como o Inglês, e as demais línguas com menos recursos (PIRES; SCHLINGER; GARRETTE, 2019).

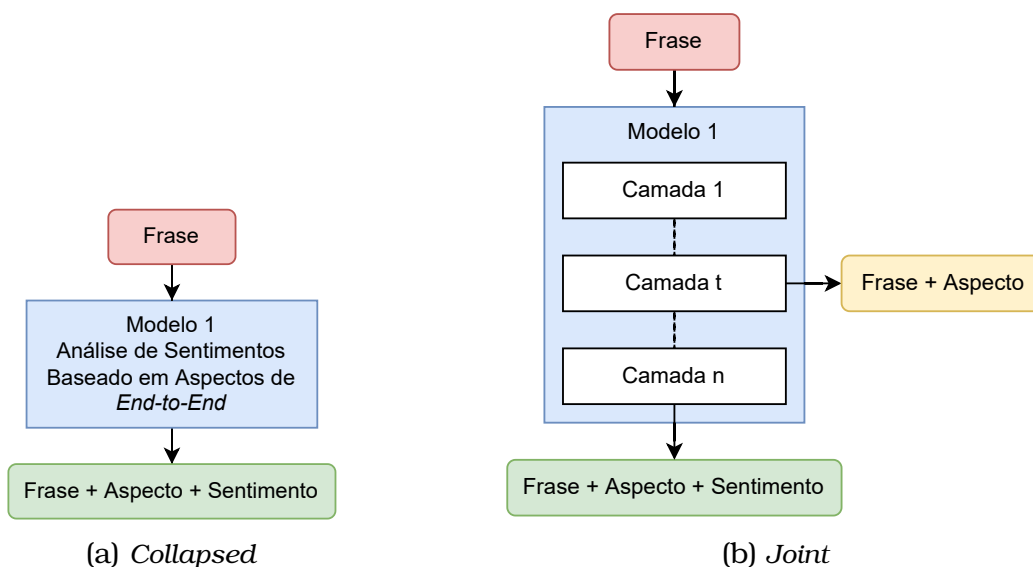
Ele é treinado da mesma forma: Token Mascarado e Predição da Próxima Sentença, conseguindo embeddings que prometem ser independentes de idioma. O BERT Multilíngue, ou m-BERT, foi investigado recentemente para algumas tarefas na língua portuguesa, com resultados promissores para a tarefa de similaridade de sentenças (FIALHO; COHEUR; QUARESMA, 2020). No trabalho de Rodrigues, Couto e Rodrigues (2019), foi apresentado um estudo descrevendo que um ajuste fino do m-BERT obtém resultados competitivos para similaridade entre sentenças, quando comparados com modelos apenas com dados em português. Também vale destacar o estudo apresentado por (LOPES; CORRÊA; FREITAS, 2021), no qual os resultados experimentais indicam que o m-BERT obtém resultados inferiores, mas competitivos, em relação ao BERT em Português denominado BERTimbau (SOUZA; NOGUEIRA; LOTUFO, 2020) para a tarefa de extração de aspectos.

Também vale destacar que vários estudos utilizam o BERT Multilíngue para ABSA em idiomas específicos. Alguns exemplos recentes incluem (AZHAR; KHODRA, 2020) para ABSA em indonésio, (JAFARIAN *et al.*, 2021) para ABSA em persa e (LOPES; CORRÊA; FREITAS, 2021) para NER em português. (PIRES; SCHLINGER; GARRETTE, 2019) investigou o modelo multilíngue em experimentos como NER, Marcação de Partes do Discurso (POS tagging), sobreposição de vocabulário, mudança de alfabeto, *etc.* Os autores concluíram que o BERT multilíngue mantém representações multilíngues, mas possui fraquezas específicas dependendo dos idiomas de baixo recurso e da aplicação.

Mais recentemente, estudos têm se concentrado em usar LLMs para tarefas de ABSA, incluindo cenários multidomínio e multilíngue. O estudo apresentado em (ZHANG *et al.*, 2023) examina as capacidades dos LLMs em abordar várias tarefas de análise de sentimentos, desde a análise de sentimentos convencional até tarefas mais complexas, como ABSA. Em particular, os LLMs foram comparados com modelos de linguagem tradicionais treinados em conjuntos de dados específicos do domínio. Um dos principais destaques

---

<sup>1</sup> <https://github.com/google-research/bert/blob/master/multilingual.md>

Figura 3 – Arquiteturas de Análise de Sentimentos Baseada em Aspectos *End-to-End*.

Fonte: Autor.

do estudo é que os modelos tradicionalmente ajustados em dados específicos do domínio permanecem altamente competitivos e até superam os LLMs. No entanto, os LLMs mostraram promessa para análise de sentimentos em um cenário de aprendizado sem supervisão (ou seja, na ausência de dados extensos de treinamento), um cenário mais frequente na indústria.

## 2.3 Arquiteturas ABSA *End-to-End*

Considerando os estudos apresentados por [Xiao et al. \(2021\)](#), podemos destacar duas arquiteturas *end-to-end*, cada uma com suas vantagens e desvantagens: a *collapsed* (também chamada de *unified*) e a *joint*. Essas arquiteturas estão ilustradas na Figura 3.

Nas próximas seções, cada arquitetura é descrita em mais detalhes.

### 2.3.1 Arquitetura *Collapsed*

Na arquitetura *Collapsed*, ao invés de usar um modelo para cada tarefa, busca-se resolver as duas tarefas, extração de aspectos e classificação do sentimento, com apenas um modelo.

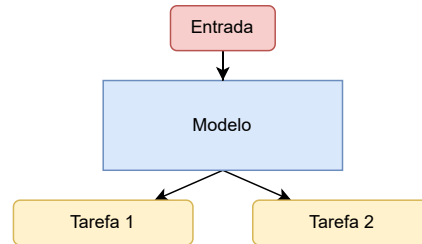
A Tabela 2 mostra o formato e diferenças de anotação dos tokens da arquitetura ABSA que usa dois modelos (EA para Extração de Aspectos e AS para Análise de Sentimentos) em relação à arquitetura *Collapsed*. Enquanto a tarefa de extração de aspectos utiliza um formato denominado BIEOS e a tarefa

Tabela 2 – Estrutura BIEOS unificada para E2E-ABSA. EA é a tarefa de Extração de Aspectos, AS de Análise de Sentimentos.

Tokens	The	AMD	Turin	Processor	seems	to	always	perform	better	than	Intel	.
EA	O	B	I	E	O	O	O	O	O	O	S	O
AS		POS	POS	POS							NEG	
<i>Collapsed</i>	O	B-POS	I-POS	E-POS	O	O	O	O	O	O	S-NEG	O

Fonte: (LI *et al.*, 2019b).

Figura 4 – Arquitetura envolvendo múltiplas tarefas.



Fonte: Autor.

de análise de sentimentos usa POS (positivo), NEU (neutro) e NEG (negativo), um modelo *Collapsed* as juntam, formando as tags: O, B-POS, I-POS, E-POS, S-POS, B-NEU, I-NEU, dentre outras variações.

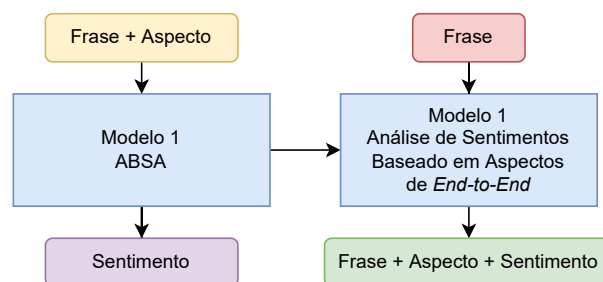
Esse formato BIEOS é uma expansão do BIO e possui o seguinte significado: B marca o começo de um aspecto (*beginning*); E marca o fim de um aspecto (*end*); I faz parte de um aspecto (*inside*), é uma marcação que existe entre B e E; O marca um token que não faz parte de um aspecto (*outside*) e S marca aspectos de tokens únicos (*single*).

Um desafio para essa arquitetura é o aumento da complexidade do problema, devido à quantidade de tags e a complexidade semântica envolvida. Mais especificamente, é a dificuldade de colocar em uma única função de perda da rede neural tanto a tarefa de *sequence labeling* comumente usada para extração de aspectos, quanto a tarefa de classificação que é a etapa de análise de sentimentos.

### 2.3.2 Arquitetura Joint

Outra arquitetura é a *Joint*, que também procura solucionar as duas tarefas em uma única estrutura, mas diferente da *Collapsed*, a *Joint* realiza um treino com múltiplas tarefas (ZHANG; YANG, 2021), conforme ilustrado na Figura 4.

No contexto de E2E-ABSA, a arquitetura *Joint* possui funções de perda diferentes durante o treinamento da rede neural, por exemplo, uma só para a extração de aspectos e outra para o E2E-ABSA na última camada. Durante o

Figura 5 – Exemplo de uma arquitetura E2E-ABSA *Joint*.

Fonte: Autor.

treinamento, cada tarefa pode ser treinada ao mesmo tempo, com compartilhamento de características entre as tarefas (e.g. embeddings) ou em forma de “cascata”, em que a saída de uma tarefa é utilizada como entrada para outra tarefa (CRAWSHAW, 2020; ZHANG; YANG, 2021).

Na Figura 5 é apresentado um exemplo da arquitetura *Joint* para E2E-ABSA. Primeiro um modelo é treinado apenas para a classificação de sentimento, sem considerar os aspectos. Essa etapa realiza um ajuste fino das embeddings do modelo BERT, por exemplo. Logo após, o processo de treinamento continua, mas alterando a função objetivo da rede neural para identificação de aspectos e sentimentos.

Assim, o intuito desta arquitetura é suprir o modelo de embeddings com informações de sentimentos gerais, para então para a etapa final de E2E-ABSA. Vale destacar que a ordem do treinamento pode ser alterada ou intercambiada durante várias épocas de treinamento.

## 2.4 Comitê de Modelos

Uma vez que neste projeto de mestrado diferentes arquiteturas de E2E-ABSA serão investigadas, a análise por meio de comitê é uma estratégia interessante (SAGI; ROKACH, 2018). De forma geral, o objetivo é verificar cenários em que diferentes arquiteturas obtêm resultados análogos ou não e, assim, estudar as características das arquiteturas considerando também as características do dataset.

Um exemplo para ilustrar um comitê é o conto de Sir Francis Galton (filósofo e estatístico inglês do século 19) sobre uma feira de gado, no qual todos os participantes deduziram o peso de um boi, mas nenhum acertou diretamente. No entanto, Galton observou que calculando a média entre os valores, ela chegava bem mais perto do peso real do animal que as outras estimativas.

Utilizando essa ideia como base, foram criadas várias alternativas para as estratégias de comitê (*ensemble learning*): a média, para valores contínuos; a votação, para valores discretos, na qual o valor mais comum é o escolhido como final; *bagging* (SAGI; ROKACH, 2018), onde cada modelo é treinado com uma parte aleatória diferente do conjunto de treino; *boosting* (FREUND; SCHAPIRE, 1997), onde cada modelo subsequente busca corrigir os erros do antecessor; e *stacking* (SARKAR, 2019), que utiliza um outro modelo para aprender direto com a saída dos modelos-base.

A princípio utilizaremos um sistema de votação, dado que estamos trabalhando com valores discretos (texto e os sentimentos positivo, neutro e negativo). Portanto, a resposta final do comitê será a mais frequente entre os modelos-base.

## 2.5 Frameworks de Regularização

Uma forma alternativa para lidar com comitê de modelos é por meio de frameworks de regularização, em que a resposta de diferentes modelos é organizada em uma rede de informação. Esta seção tem o objetivo de apresentar os fundamentos sobre esse assunto.

Uma rede de informação, denotada por  $\mathcal{N} = \langle \mathcal{O}, \mathcal{R}, \mathcal{W} \rangle$ , é composta por três elementos:

1.  $\mathcal{O}$ : O conjunto de objetos (ou elementos) da rede, onde  $\mathcal{O}$  pode conter diferentes tipos de entidades, representadas por nós ou vértices na rede. Matematicamente,  $\mathcal{O}$  é um conjunto finito de objetos.
2.  $\mathcal{R}$ : O conjunto de relações entre os objetos, representando as ligações ou conexões entre os elementos da rede. As relações podem ser direcionadas ou não direcionadas, ponderadas ou não ponderadas. Matematicamente,  $\mathcal{R}$  é um conjunto de pares ordenados que descrevem as conexões entre os objetos.
3.  $\mathcal{W}$ : O conjunto de pesos associados às relações da rede, representando a intensidade ou força das conexões entre os objetos. Os pesos podem ser usados para expressar a importância das relações em termos quantitativos. Matematicamente,  $\mathcal{W}$  é uma função que mapeia pares de objetos em números reais, onde  $w_{o_i, o_j}$  denota o peso da relação entre os objetos  $o_i$  e  $o_j$ .

Portanto, formalmente, uma rede de informação é uma tripla ordenada  $\mathcal{N} = \langle \mathcal{O}, \mathcal{R}, \mathcal{W} \rangle$ , que descreve as entidades (objetos), suas relações e a intensi-



dade dessas relações em um contexto específico. Esse modelo é frequentemente utilizado para representar informações complexas em diversas áreas, como ciência de dados, análise de redes sociais, processamento de linguagem natural, entre outros.

O framework de regularização geral em redes de informação pode ser expresso matematicamente através da seguinte equação:

$$Q(\mathbf{F}) = \frac{1}{2} \sum_{o_i, o_j \in \mathcal{O}} w_{o_i, o_j} \Omega(\mathbf{f}_{o_i}, \mathbf{f}_{o_j}) + \mu \sum_{o_i \in \mathcal{O}^L} \Omega'(\mathbf{f}_{o_i}, \mathbf{y}_{o_i}) \quad (2.1)$$

em que  $Q(\mathbf{F})$  representa a função objetivo a ser otimizada no processo de regularização,  $\mathbf{F}$  é um vetor que contém as informações de classe associadas aos nós da rede. Cada nó  $o_i$  possui um vetor  $\mathbf{f}_{o_i}$  que descreve as informações de classe. Já o  $w_{o_i, o_j}$  representa o peso da relação entre os nós  $o_i$  e  $o_j$  na rede. Esses pesos podem ser usados para expressar a intensidade das conexões entre os nós. O  $\Omega(\mathbf{f}_{o_i}, \mathbf{f}_{o_j})$ , primeiro termo da equação, mede a proximidade entre os vetores de informações de classe  $\mathbf{f}_{o_i}$  e  $\mathbf{f}_{o_j}$  de nós conectados na rede. Esta função pode ser formulada de várias maneiras, incluindo o uso de funções de distância ou dissimilaridade. O parâmetro  $\mu$  controla a importância das informações de classe reais durante a propagação dos rótulos. O parâmetro é aplicado em  $\Omega'(\mathbf{f}_{o_i}, \mathbf{y}_{o_i})$ , o segundo termo da equação, e mede a proximidade entre as informações de classe  $\mathbf{f}_{o_i}$  e os rótulos reais  $\mathbf{y}_{o_i}$  para os nós rotulados. Isso ajuda a manter a coerência com os rótulos conhecidos.

De forma geral, o objetivo do framework de regularização é encontrar uma distribuição de informações de classe  $\mathbf{F}$  que minimize a função objetivo  $Q(\mathbf{F})$ , de modo a alcançar uma propagação suave e consistente das informações de classe na rede, levando em consideração as relações na rede e os rótulos disponíveis.

Mais detalhes sobre a regularização no cenário de ABSA são apresentados no Capítulo 3.

## 2.6 Trabalhos Relacionados

Nesta seção é apresentada uma análise de trabalhos recentes para Análise de Sentimentos Baseada em Aspectos *End-to-End* (E2E-ABSA). Na Tabela 3 se encontra boa parte dos trabalhos mais relacionados com este projeto, com o tipo de arquitetura, a estratégia principal, inovações e limitações, além de ser ou não multilíngue e multi-domínio.

Em destaque, [Li et al. \(2019b\)](#) utilizaram o BERT, um modelo de rede neural genérico que revolucionou o contexto de NLP, avançando o estado-da-arte em vários problemas da área. Para o E2E-ABSA, o BERT também conseguiu melhores resultados que as estratégias anteriores e, por conta das características multilíngue do BERT, acredita-se neste projeto que ele consiga incorporar outros idiomas para o E2E-ABSA sem que se perca muita performance no idioma original dos dados de treinamento.

Os sete primeiros artigos na Tabela 3 também são discutidos no estudo apresentado por [Li et al. \(2019b\)](#). É possível observar que os tipos de modelos vão mudando com a melhoria da performance, indo de Bi-LSTM+RNN, para CNNs, Bi-LSTM+CRF, Bi-LSTM+CRF com embedding de caracteres e por fim o BERT+Camada, este último sendo a base inicial deste projeto.

Para a arquitetura *joint* com o BERT, [Xiao et al. \(2021\)](#) avaliaram entre as principais estratégias, uma arquitetura *joint* em que as saídas de cada camada do BERT são somadas e escaladas com diferentes pesos duas vezes, uma para a tarefa de classificação de sentimentos e outra para a de extração de entidades. Tal arquitetura obteve resultados levemente melhores, mostrando que para essa arquitetura, talvez seja mais interessante considerá-la realmente como um problema multi-task.

[Liang et al. \(2021\)](#) usam explicitamente a estrutura sintática nas embeddings de uma CNN, o que para inglês gerou bons resultados, porém para ambientes multilíngue pode dificultar, já que a estrutura sintática pode mudar bastante ao observar idiomas como inglês, português e japonês.

Seguindo o processo do BERT+Camada, [Lu et al. \(2022\)](#) criaram uma camada personalizada de atenção, na qual uma palavra prioriza o contexto de palavras no mesmo constituinte (palavras fortemente ligadas entre si). Os experimentos foram melhores em metade dos datasets, o que se dá a entender que é o BERT em si que faz o maior trabalho na predição, os tipos diferentes de camadas não parecem gerar uma diferença tão significativa.

Já no trabalho de [Tian et al. \(2021\)](#), o BERT foi utilizado para a geração de embeddings, que passam por uma Bi-LSTM+CRF. Eles exploram a ideia

de multidomínio, em que há a geração de parâmetros extras específicos para um domínio, que são concatenados à embedding, gerando uma representação única. De forma análoga, incorporar um BERT multilíngue nesta arquitetura pode ser interessante para comparação com o trabalho proposto, entretanto voltamos ao problema da LSTM não ser paralelizável, o que dificulta o treino e predição com muitos dados.

Chen e Qian (2022) propuseram o TransProto, onde eles substituem palavras em domínios desconhecidos com protótipos de mesma função semântica e sintática. Eles também encaram o problema como multi-task, onde o modelo possui diferentes camadas finais para diferentes tarefas. A grande limitação é a dependência explícita da sintática e semântica, o que não é ideal em um ambiente com vários idiomas.

No ambiente multilíngue, existem vários trabalhos que utilizam o m-BERT para NER ou ABSA em idiomas específicos. Alguns exemplos recentes são: (AZHAR; KHODRA, 2020) para ABSA Indonésio, (JAFARIAN *et al.*, 2021) para ABSA Persa e (LOPES; CORRÊA; FREITAS, 2021) para o NER Português. Entretanto ainda existe a lacuna de explorar a capacidade do modelo para E2E-ABSA.

Existem também trabalhos que analisam o próprio BERT multilíngue. Pires, Schlinger e Garrette (2019) averiguaram o modelo em experimentos como NER, part of speech tagging (POS tagging), sobreposição de vocabulário, mudança de alfabeto, etc. Os autores concluíram que o m-BERT aprende representações multilíngue, porém possui certas fraquezas dependendo dos idiomas utilizados e da aplicação. Além disso, Wu e Dredze (2020) investigaram a performance do modelo para os diferentes idiomas pré-treinados, alertando que m-BERT para línguas com poucos recursos exige um treinamento mais eficiente para o funcionamento, caso contrário obtém resultados inferiores a modelos sem pré-treinamento.

Recentemente, tem havido um crescente interesse em explorar Large Language Models (LLMs) para tarefas de Análise de Sentimento Baseada em Aspectos (ABSA). Um estudo recente de (ZHANG *et al.*, 2023) examinou as capacidades dos LLMs em lidar com uma variedade de tarefas de análise de sentimentos, desde a análise de sentimentos convencional até tarefas mais complexas como ABSA. Os resultados destacaram a capacidade dos LLMs de realizar análise de sentimentos em cenários de aprendizado sem supervisão, nos quais há escassez de dados de treinamento disponíveis, uma situação cada vez mais comum na indústria. Embora os modelos tradicionalmente ajustados em dados específicos do domínio ainda se mostrem altamente competitivos e,

em alguns casos, superiores aos LLMs, os resultados sugerem que os LLMs têm potencial para se destacar em ambientes com restrições de dados de treinamento.

A análise dos trabalhos da literatura selecionados, conforme a coluna de limitações da Tabela 3, indica que muitas estratégias foram propostas para E2E-ABSA, mas não há uma comparação experimental direta que permita um entendimento mais amplo de suas propriedades. Também indica poucos trabalhos com foco no multilíngue, algo cada vez mais necessário frente ao avanço de redes sociais e da globalização na comercialização de produtos e serviços. A língua portuguesa, por exemplo, tem sido pouco explorada para E2E-ABSA. Assim, no próximo capítulo é apresentado um plano de trabalho para o desenvolvimento deste projeto de mestrado focado em tais lacunas.

Tabela 3 – Tabela comparativa com trabalhos relacionados na literatura.

Paper	Arquitetura	Estratégia	Inovações	Limitações	Multilingue	Multidomínio
(LI <i>et al.</i> , 2019a)	<i>Joint</i>	Bi-LSTM+RNN	Propõe solucionar NER e ABSA como um só problema <i>end-to-end</i> .	LSTMs não são paralelizáveis, o que limita a quantidade de dados a serem utilizados.	Não	Não
(LUO <i>et al.</i> , 2019)	<i>Collapsed</i>	RNNs	Mesma premissa do paper acima.	Embeddings de palavras fora do vocabulário são iniciadas aleatoriamente, o que para ambientes multilingue não é ideal.	Não	Não
(HE <i>et al.</i> , 2019)	<i>Joint</i>	CNNs	Aprendizado simultâneo em nível de token e nível de documento.		Não	Não
(LAMPLE <i>et al.</i> , 2016)	<i>Collapsed</i>	Bi-LSTM+CRF	CNN para criar embeddings character-level.	LSTM	Não	Não
(MA; HOVY, 2016)	<i>Collapsed</i>	Word/character embeddings + Bi-LSTM+CRF	LSTM para criar embeddings character-level.	LSTM	Não	Não
(LIU <i>et al.</i> , 2018)	<i>Collapsed</i>	Word/character embeddings + Bi-LSTM+CRF		LSTM	Não	Não
(LI <i>et al.</i> , 2019b)	<i>Collapsed</i>	BERT+Layer	Paralelização (Transformers vs LSTM), permitindo o uso de uma quantidade maior de dados.	A camada extra não parece fazer uma diferença significativa, parece que o BERT é o responsável pelo melhor desempenho.	Não	Não
(XIAO <i>et al.</i> , 2021)	<i>Joint</i>	BERT+CRF	Utilização do BERT em uma arquitetura <i>joint</i> .	A arquitetura multi-task foi um pouco melhor que a <i>joint</i> tradicional.	Não	Não
(LIANG <i>et al.</i> , 2021)	<i>Joint</i>	BERT+DRECGN+CNN	Utilização da estrutura sintática como grafo.	Estrutura sintática de diferentes idiomas podem mudar drasticamente (como Português e Japonês, por exemplo).	Não	Não
(LU <i>et al.</i> , 2022)	<i>Collapsed</i>	BERT+Layer(C-ATT)	Criam uma variação da camada de self-attention, na qual a rede aprende a relacionar as palavras da sentença em "grupos", onde tokens no mesmo grupo se dão mais importância.	Resultados não são tão diferentes que os de (LI <i>et al.</i> , 2019b). No geral parece que é o BERT que está "puxando a carroça", a camada extra não parece fazer uma diferença significativa.	Não	Não
(TIAN <i>et al.</i> , 2021)	<i>Collapsed</i>	BERT+Bi-LSTM+CRF	Geração de diferentes parâmetros para diferentes domínios.	LSTM	Não	Sim
(CHEN; QIAN, 2022)	<i>Joint</i>	BERT+TransProto	Procura similaridades entre palavras do domínio origem e alvo através da sintaxe e semântica.	É voltado para a transferência de conhecimento de domínios com muito recurso para os com pouco. Dependem diretamente de sintaxe e semântica talvez não seja uma boa ideia para multilingue.	Não	Sim
(ZHANG <i>et al.</i> , 2023)	<i>Generativa</i>	LLM	Modelos pré-treinados para gerar textos orientados a ABSA.	Resultados muito dependentes do contexto ou <i>prompt</i> utilizado. Suscetível a alucinações ou respostas fora do padrão.	Sim	Sim



---

## MÉTODO PROPOSTO: CONSENTIMENT

---

Neste capítulo é apresentado o método “**ConSentiment: A Graph-Based Regularization Framework for Sentiment Analysis with Consensus and Uncertainty Reduction**” proposto para determinar o aspecto e o sentimento correspondente, considerando o consenso de diferentes modelos base de ABSA de ponta a ponta.

O método ConSentiment opera em duas etapas. Primeiro, envolve a representação unificada dos resultados de cada modelo base de ABSA em um grafo de consenso. Cada vértice inclui informações sobre a *review*, aspecto, sentimento e modelo base nesse contexto. A *review* está associada a um ou mais aspectos, e cada aspecto tem uma polaridade de sentimento (positiva, negativa ou neutra), que pode variar dependendo do modelo base usado. A construção do grafo também leva em consideração a similaridade semântica entre aspectos. Em segundo lugar, estendemos um mecanismo de regularização de grafo para lidar com nosso grafo de consenso baseado em ABSA. A ideia básica é propagar informações de rótulos ABSA de vértices com maior consenso para determinar o rótulo final de vértices com incerteza, estimulando assim a atualização dos rótulos inicialmente fornecidos pelos modelos base. Mais detalhes para cada etapa são apresentados nas subseções abaixo.

### 3.1 Grafo de Consenso Baseado em ABSA

Seja  $G(V, E)$  um grafo de consenso, onde um vértice  $v \in V$  é uma quádrupla  $(r, a, s, m)$  e  $E$  é o conjunto de arestas, onde:

- $r \in R$ : *review*  $r$  que pertence ao conjunto de revisões  $R$ ;

- $a \in A$ : aspecto  $a$  que pertence ao conjunto de aspectos  $A$ ;
- $s \in S$ : rótulo de sentimento  $s$  que pertence ao conjunto de polaridades de sentimento  $S$ ;
- $m \in M$ : identificador do modelo  $m$  que pertence ao conjunto de modelos base  $M$ .

No grafo de consenso baseado em ABSA, a conexão entre vértices é determinada pela similaridade dos embeddings de  $r$  e  $a$  e duas regras específicas do domínio ABSA propostas neste método:

1. **Informação Consensual:** quando os vértices têm alta similaridade entre a *review*  $r$  e o aspecto  $a$ , com o mesmo rótulo de sentimento previsto  $s$ , há evidências de que os modelos base  $m$  alcançaram um consenso na predição. Este cenário é relevante pois permite que o modelo explore o consenso e propague informações com menor incerteza para áreas de maior incerteza no grafo;
2. **Redução de Incerteza:** se dois vértices têm o mesmo aspecto  $a$  identificado pelo modelo base mas diferentes sentimentos associados  $s$ , uma aresta é criada entre eles. Essa conexão força o modelo de consenso a determinar o sentimento final do aspecto na presença de incerteza com base na topologia da rede. Na prática, o consenso obtido na primeira regra é esperado para ser usado para classificar o sentimento em situações em que há uma divergência entre os modelos base.

A regra de informação consensual determina que uma aresta seja criada entre dois vértices no grafo se a similaridade entre a *review*  $r$  e o aspecto  $a$  for suficientemente alta, dada um limiar, e o sentimento  $s$  previsto por um modelo base for o mesmo que outro modelo base. Para calcular a similaridade entre dois vértices  $v_i$  e  $v_j$  no grafo, considere os embeddings de revisões e aspectos, onde  $\sigma_r$  representa o embedding de revisões e  $\sigma_a$  o embedding de aspectos. Nosso método explora uma combinação linear ponderada das similaridades entre  $\sigma_r$  e  $\sigma_a$ , como definido na Equação 3.1.

$$\text{sim}(v_i, v_j) = \cos(w_r \cdot \sigma_r(r_i) + w_a \cdot \sigma_a(a_i), w_r \cdot \sigma_r(r_j) + w_a \cdot \sigma_a(a_j)) \quad (3.1)$$

onde  $\text{sim}(v_i, v_j)$  é a similaridade resultante da combinação linear dos embeddings de revisões  $r_i$  e  $r_j$  e seus respectivos aspectos  $a_i$  e  $a_j$ . Os termos  $w_r$  e  $w_a$  são pesos



que podem ser ajustados para controlar a contribuição relativa dos embeddings de *review* e aspecto.

Em relação às conexões para a regra de Redução de Incerteza, dados dois vértices com o mesmo aspecto  $a$ ,  $v_i = (r_i, a, s_i, m)$  e  $v_j = (r_j, a, s_j, m')$  para dois modelos base ABSA diferentes  $m \neq m'$ , a aresta  $E_{ij}$  é gerada se  $s_i \neq s_j$ , indicando que os sentimentos  $s_i$  e  $s_j$  associados aos vértices  $v_i$  e  $v_j$  são diferentes. Assim, também fortalecemos as relações na divergência entre modelos. O framework de regularização apresentado na próxima seção visa usar informações de vértices com consenso para aprender o sentimento final de vértices com mais divergência.

## 3.2 Regularização Baseada em Consenso de ABSA

Nosso framework de regularização proposto assume duas premissas fundamentais, similares ao framework de regularização geral proposto por (WANG, 2010). A primeira premissa considera que vértices vizinhos no grafo devem ter previsões de sentimento semelhantes, refletindo o conceito de que informações de consenso entre modelos devem ser preservadas. A segunda premissa postula que as previsões de sentimento de vértices com maior consenso devem permanecer estáveis ao longo do processo de regularização.

A função de regularização  $Q(\mathbf{F})$  é uma função objetivo a ser minimizada, combinando dois termos do método ConSentiment para atender às duas premissas, conforme definido na Equação 3.2:

$$Q(\mathbf{F}) = \frac{1}{2} \sum_{v_i, v_j \in V} e_{v_i, v_j} (\mathbf{f}_{v_i} - \mathbf{f}_{v_j})^2 + \mu \sum_{v_i \in V_L} (\mathbf{f}_{v_i} - \mathbf{g}_{v_i})^2 \quad (3.2)$$

onde

- $Q(\mathbf{F})$  é a função de regularização a ser minimizada em relação aos parâmetros  $F_{|V| \times 3}$ .  $F$  representa uma matriz de associação de rótulos de sentimento (positivo, negativo, neutro) para cada vértice  $v \in V$ .
- $e_{v_i, v_j}$  representa a presença (ou ausência) de uma relação entre dois vértices  $v_i$  e  $v_j$  no grafo. Em nosso caso, apenas usamos grafos não ponderados, mas o método proposto é flexível, pois pode ser facilmente adaptado para incorporar pesos de aresta.
- $\mathbf{f}_{v_i}$  são os valores de pertinência da previsão de sentimento para o vértice  $v_i$ . É um vetor com três dimensões, onde cada dimensão representa a força

do rótulo de sentimento entre 0 e 1. Quanto mais próximo de 1, maior a relevância para o respectivo sentimento.

- $\mu$  é um parâmetro que controla a importância das informações de consenso. O ConSentiment usa um valor entre 0 e 1, onde um valor maior representa dar mais importância aos vértices com maior consenso.
- $V_L$  é o conjunto de vértices com consenso de sentimento  $\mathbf{g}_{v_i}$ . O ConSentiment seleciona vértices com consenso da maioria dos modelos para incluir em  $V_L$ .
- $\mathbf{g}_{v_i}$  representa o vetor de informações de consenso de ABSA obtido dos modelos base. É uma média simples da confiança em cada classe, considerando as previsões dos modelos base.

Assim, a função  $Q(\mathbf{F})$  consiste em dois termos principais. O primeiro termo,  $\frac{1}{2} \sum_{v_i, v_j \in V} e_{v_i, v_j} (\mathbf{f}_{v_i} - \mathbf{f}_{v_j})^2$ , está relacionado à primeira premissa do método, implicando que vértices vizinhos devem ter previsões de sentimento semelhantes. Este termo penaliza diferenças significativas nas previsões de sentimento entre vértices conectados no grafo. O segundo termo,  $\mu \sum_{v_i \in V_L} (\mathbf{f}_{v_i} - \mathbf{g}_{v_i})^2$ , está relacionado à segunda premissa do método, sugerindo que os vértices devem manter suas previsões de sentimento, especialmente quando há forte consenso entre os modelos base. Este termo penaliza a diferença entre as previsões de sentimento dos vértices e o consenso ( $\mathbf{g}_{v_i}$ ).

A função  $Q(\mathbf{F})$  é usada como uma medida de custo que deve ser minimizada para encontrar as previsões de sentimento finais e reduzir a incerteza associada às previsões de sentimento em situações em que diferentes modelos base fornecem resultados divergentes. Minimizar a função  $Q(\mathbf{F})$  pode ser alcançado por meio de várias técnicas de otimização, como solucionadores e métodos de descida de gradiente, cada um com suas vantagens e considerações computacionais. O método ConSentiment usa uma abordagem iterativa, aproveitando sua eficiência para fornecer soluções aproximadas a um custo computacional acessível. Essa escolha é particularmente vantajosa no contexto de ABSA, onde lidar com conjuntos de dados extensos é comum.

O Algoritmo ConSentiment começa computando a matriz de grau diagonal, que captura a soma das relações entre os vértices no grafo (linha 1). Em seguida, o Laplaciano de grafo normalizado é calculado para modelar a interdependência dos vértices e suas previsões de sentimento; assim, o algoritmo suaviza as previsões iniciais com base nas informações de consenso dos vértices vizinhos (linha 2). Além disso, definimos as previsões iniciais de

**Algorithm 1:** ABSA Consensus-based Regularization (Iterative Method)

---

**Data:** Graph  $G = (V, E)$ , Initial sentiment predictions  $\mathbf{f}$ , Consensus information  $\mathbf{g}$ , Parameter  $\mu$

**Result:** Updated sentiment predictions  $\mathbf{f}$

- 1 Compute diagonal degree matrix  $D$ , where  $D_{ii} \leftarrow \sum_j E_{ij}$ ;
- 2 Compute normalized graph Laplacian  $\mathcal{L} = D^{-\frac{1}{2}}ED^{-\frac{1}{2}}$ ;  
// Initialize sentiment predictions and consensus information
- 3 Initialize  $\mathbf{f}^{(0)} \in \mathbf{F}$  for all vertices  $v \in V$  based on average confidence of base models;
- 4 Initialize  $\mathbf{g}^{(0)}$  for vertices  $v \in V_L$  with majority consensus information;
- 5 **while** not converged **do**
  - 6     // Consensus Uncertainty Reduction Step  
Update  $\mathbf{f}^{(t+1)}$  for all  $v \in V$  using  $(1 - \mu)\mathcal{L}\mathbf{f}^{(t)}$ ;
  - 7     Update  $\mathbf{f}^{(t+1)}$  for all  $v \in V_L$  using  $\mu\mathcal{L}\mathbf{f}^{(t)}$ ;
- 8 **return**  $\mathbf{F}$ ;

---

sentimento de cada vértice com base na confiança média dos modelos base e inicializamos as informações de consenso para vértices específicos com acordo da maioria (linhas 3-4). Conseqüentemente, o ConSentiment atualiza iterativamente as previsões de sentimento para todos os vértices, aproveitando o grafo normalizado do Laplaciano em um passo de Redução de Incerteza de Consenso (linhas 6-7). Esse processo visa minimizar a incerteza nas previsões de sentimento incorporando iterativamente informações dos vértices vizinhos e mantendo previsões estáveis para vértices com maior consenso. A convergência ocorre quando a diferença entre a matriz de previsões de sentimento, denotada como  $\mathbf{F}$ , torna-se menor que um limiar pré-definido de uma iteração para a próxima.

### 3.2.1 Complexidade do Algoritmo

O algoritmo pode ser dividido em duas partes principais: criação do grafo e regularização. Durante a fase de criação, a complexidade é  $O(t^2)$ , onde  $t$  representa o número total de tuplas (aspecto, sentimento) em todos os modelos. Essa complexidade surge porque o algoritmo precisa calcular a distância entre todos os nós para conectar os mais similares. Na fase de regularização, a complexidade é  $O(t \times n)$ , onde  $t$  permanece o mesmo e  $n$  representa o número médio de vizinhos para cada nó.

### 3.2.2 Ferramenta Computacional

O método proposto, ConSentiment, foi implementado por meio de uma ferramenta computacional com o objetivo de oferecer uma solução prática e eficiente para análise de sentimento baseada em consenso.

Essa ferramenta foi concebida com a intenção de facilitar a aplicação do método em diversos cenários e domínios, permitindo que pesquisadores e profissionais apliquem facilmente o ConSentiment, bastando adicionar novos modelos dentro da lista de modelos base. Ainda, a ferramenta permite a construção grafos de consenso baseados em ABSA, além de oferecer opções para ajuste de parâmetros e configurações personalizadas, que foram utilizadas na avaliação experimental.

A estruturação do repositório GitHub para a ferramenta ConSentiment é organizada em uma hierarquia de pastas e arquivos, cada um desempenhando um papel específico no processo de treinamento, avaliação e aplicação do método. O diretório “Data” contém os conjuntos de dados necessários para treinar e avaliar os modelos de análise de sentimento, subdivididos por domínio (Laptop, Restaurant, Hotel) e idioma (EN para inglês e PT para português). Os arquivos de treinamento (“train.xlsx”), validação (“dev.xlsx”) e teste (“test.xlsx”) estão disponíveis em cada subdiretório, juntamente com uma planilha que contém as *reviews*, anotações e previsões dos modelos base.

No diretório de exemplos, o fluxo de trabalho é iniciado com o notebook “Train\_Predict-E2E-ABSA.ipynb”, que permite treinar um novo modelo de análise de sentimento. Os notebooks “Committee-Graph.ipynb” e “Committee-Voting.ipynb” lidam com a criação e aplicação do comitê de modelos de análise de sentimento. No primeiro notebook, os resultados do teste são lidos, e um grupo de modelos é selecionado para construir o grafo de consenso e aplicar a regularização. Já no segundo notebook, os resultados do teste também são lidos, mas desta vez um grupo de modelos é escolhido para formar o comitê de votação, e os votos são calculados. Os resultados do comitê são então anexados ao arquivo “test.xlsx”.

Por fim, o notebook “Evaluate.ipynb” é utilizado para avaliar os resultados com a biblioteca Nervaluate, produzindo gráficos e tabelas com os resultados da análise de sentimento. Essa estruturação do repositório e a sequência de execução dos scripts facilitam o desenvolvimento, treinamento, avaliação e aplicação do método ConSentiment em uma variedade de cenários e conjuntos de dados.

Para transparência e reprodutibilidade, foi criado um GitHub com as

dados, a biblioteca e códigos utilizados, chamado [ConSentiment](#)<sup>1</sup>.

---

<sup>1</sup> <<https://github.com/GuilhermeMarcon/consentiment>>



## AVALIAÇÃO EXPERIMENTAL

### 4.1 Conjuntos de Dados

Três conjuntos de dados de referência amplamente utilizados em análise de sentimento foram selecionados e resumidos para a avaliação experimental na Tabela 4.

Conjunto de Dados	laptop14	rest14	hotel22
Tamanho Total	3.845	3.840	837
Tamanho do Treinamento	2.741	2.736	677
Tamanho da Validação	304	304	76
Tamanho do Teste	800	800	84
Total de Aspectos	2.933	4.706	3.084
Aspectos Positivos	1.327 (45%)	2.892 (61%)	2.120 (69%)
Aspectos Negativos	989 (33%)	993 (21%)	508 (16%)
Aspectos Neutros	617 (22%)	821 (18%)	456 (15%)

Tabela 4 – Características dos Conjuntos de Dados de ABSA

O conjunto de dados Laptop14 compreende dados de análise de sentimento no domínio de laptops e no idioma inglês. Foi introduzido no SemEval14 e obtido diretamente do repositório E2E-ABSA por (LI *et al.*, 2019b). O conjunto de dados Rest14 é semelhante ao anterior em inglês, mas se concentra no domínio de restaurantes. O conjunto de dados Hotel22 envolve análise de sentimento no domínio de hotéis em português, fornecido pelo desafio ABSAPT<sup>1</sup>. Esses conjuntos de dados fornecem vários domínios e suportam avaliação em diferentes idiomas, representando um cenário típico do mundo real.

<sup>1</sup> <<https://sites.google.com/view/iberlef2022>>

## 4.2 Configuração Experimental

Os seguintes cenários foram explorados:

- **Treinamento e Teste no Mesmo Domínio:** tanto os conjuntos de treinamento quanto de teste pertencem ao mesmo domínio, garantindo que os modelos estejam adaptados ao contexto específico;
- **Treinamento em um Domínio e Teste em Outro Domínio:** Também avaliamos cenários multidomínio em que o conjunto de treinamento é de um domínio, mas o conjunto de teste é de outro domínio. Isso nos permite entender a capacidade de generalização dos modelos em ambientes multidomínio ou quando há falta de dados rotulados no domínio de teste.
- **Aprendizado de Zero-Shot e Geração de Dados usando Grandes Modelos de Linguagem (LLMs):** Exploramos duas abordagens com LLMs. Primeiramente, o aprendizado de zero-shot foi empregado para ABSA de ponta a ponta, mostrando a capacidade do modelo de realizar análise de sentimento em um domínio sem dados de treinamento prévios. Em segundo lugar, usamos um LLM para gerar conjuntos de dados rotulados para treinar modelos ABSA de ponta a ponta baseados em BERT, demonstrando como modelos tradicionais podem ser integrados ao aprendizado de zero-shot e melhorar a diversidade em modelos baseados em comitê.

Exploramos as seguintes arquiteturas de modelos de ABSA:

- ABSA de ponta a ponta com ajuste fino do BERT (LI *et al.*, 2019b): diferentes camadas adicionais adicionadas sobre o modelo BERT para ajuste fino. As arquiteturas avaliadas incluíram Campo Aleatório Condicional (CRF), Unidade Recorrente Portão (GRU), Linear, Rede de Autoatenção (SAN) e Camada Transformadora (TFM).
- ABSA com LLM GPT 3.5 com aprendizado de prompt de zero-shot: um exemplo de prompt está disponível em <<https://chat.openai.com/share/4613b7a7-48ff-402b-bdb8-0ca5a3c53acd>>;
- Comitê com Voto Majoritário (Majority Voting Ensemble - MVE): um comitê de modelos (BERT e LLM) que realiza análise de sentimento por meio da votação majoritária dos modelos base;
- ConSentiment: o comitê proposto de modelos. Utilizamos  $w_a = 0.9$  e  $w_r = 0.1$  para controlar os pesos atribuídos aos componentes baseados em



aspectos e de regularização, respectivamente. Na construção do grafo de consenso, empregamos um limiar mínimo de similaridade de 0.7 para estabelecer conexões entre os nós. Além disso, definimos  $\mu$  como 0.9 para preservar nós com um alto nível de consenso durante o processo de regularização. Esses valores de hiperparâmetros foram determinados por meio de uma análise empírica, envolvendo validação cruzada em diferentes subconjuntos do conjunto de dados de treinamento. Importante destacar que o conjunto de teste não foi utilizado durante esse processo de seleção de hiperparâmetros, garantindo uma avaliação imparcial do desempenho de generalização do ConSentiment.

Para o ajuste fino do BERT em ABSA de ponta a ponta na tarefa de ABSA, foram utilizados dois modelos pré-treinados:

- BERT-base-uncased: Esta abordagem é particularmente relevante ao lidar com conjuntos de dados principalmente em inglês;
- BERT-base-multilingual-cased: O modelo BERT pré-treinado com suporte multilíngue foi usado para permitir a avaliação de conjuntos de dados em inglês e português, incluindo conjuntos de dados que apresentam uma mistura de ambos os idiomas.

O desempenho dos modelos é avaliado usando a pontuação F1 para Reconhecimento de Entidades Nomeadas (NER)<sup>2</sup>.

## 4.3 Resultados e Discussão

O primeiro cenário na avaliação experimental envolve *reviews* de um único domínio e apenas em inglês, conforme apresentado na Tabela 5. O ConSentiment demonstra desempenho superior em comparação com outros métodos, alcançando pontuações F1 de 0.701 no conjunto de dados Laptop14 e 0.806 no conjunto de dados Rest14. Esses resultados superam métodos de referência, incluindo ABSA-BERT-CRF, ABSA-BERT-GRU, ABSA-BERT-Linear, ABSA-BERT-SAN, ABSA-BERT-TFM e MVE.

Ao considerar o BERT multilíngue e cenários entre domínios, conforme apresentado na Tabela 6, o ConSentiment continua a exibir desempenho competitivo. Por exemplo, no conjunto de dados Hotel22, onde o treinamento

<sup>2</sup> <<https://pypi.org/project/nervaluate/>>

Modelo	Conjunto de Dados	
	Laptop14	Rest14
ABSA-BERT-CRF	0.600	0.724
ABSA-BERT-GRU	0.614	0.722
ABSA-BERT-Linear	0.603	0.735
ABSA-BERT-SAN	0.624	0.740
ABSA-BERT-TFM	0.625	0.736
MVE	0.606	0.710
ConSentiment	<b>0.701</b>	<b>0.806</b>

Tabela 5 – Resultados experimentais (pontuações F1) com *reviews* de um único domínio e apenas em inglês

envolve tanto um domínio quanto um idioma diferentes, o ConSentiment mostra uma melhoria média de aproximadamente 40% nas pontuações F1 em comparação com as pontuações F1 médias de outros modelos de referência. O ConSentiment apresenta resultados significativamente superiores, especialmente quando os modelos base enfrentam dificuldades para lidar com vários idiomas e domínios, exibindo consenso mínimo. Superando esses desafios, o ConSentiment aproveita o conceito de Redução de Incerteza, propagando informações de rótulos de forma mais precisa a partir de regiões de consenso.

Modelo	Conjunto de Dados		
	Hotel22	Laptop14	Rest14
ABSA-BERT-CRF	0.347	0.379	0.489
ABSA-BERT-GRU	0.358	0.353	0.499
ABSA-BERT-Linear	0.337	0.358	0.496
ABSA-BERT-SAN	0.381	0.388	0.524
ABSA-BERT-TFM	0.362	0.356	0.531
MVE	0.337	0.353	0.513
ConSentiment	<b>0.605</b>	<b>0.529</b>	<b>0.635</b>

Tabela 6 – Resultados experimentais (pontuações F1) com cenários multilíngues e entre domínios

Examinando o impacto da incorporação de grandes modelos de linguagem (GPT-3.5-turbo) nos modelos base e ConSentiment em um cenário multilíngue, avaliamos os modelos base em um cenário mais otimista, adquirindo dados de treinamento anotados por humanos para o conjunto de dados Hotel22 (coluna "Por Humanos" na Tabela 7). Além disso, avaliamos a capacidade do LLM de realizar tarefas de ABSA independentemente sem dados rotulados. Nesse caso, usamos o LLM para extrair aspectos e sentimentos de textos opinativos diretamente, alcançando uma pontuação F1 de 0.48. Além disso, exploramos os LLMs na anotação de dados por meio de aprendizado de zero-shot para treinar modelos ABSA-BERT (coluna "Por LLM" na Tabela

3). O ConSentiment teve um bom desempenho, alcançando uma pontuação F1 de 0.816 para dados anotados por humanos e 0.770 para rótulos gerados por um modelo de linguagem significativo (LLM). Esses resultados ilustram que a abordagem de comitê proposta pode obter resultados significativamente superiores em comparação com os modelos base.

É especialmente notável o cenário com dados rotulados limitados, onde um LLM que emprega aprendizado de zero-shot encontrou desafios na tarefa de ABSA. No entanto, o comitê ConSentiment, gerado por modelos base ABSA treinados em dados de LLM, superou com sucesso esses desafios. Isso se deve à capacidade dos modelos base ABSA-BERT de se especializarem para o domínio específico (primeiro cenário). Ao mesmo tempo, o ConSentiment captura as melhores características de cada modelo base durante a estratégia de redução de incerteza. Também é notável que os LLMs podem produzir respostas erradas ou sinônimos em várias situações, com aspectos semelhantes, mas ausentes na *review* original, impactando a avaliação. Por outro lado, em cenários com dados rotulados limitados, a integração de LLMs e aprendizado de zero-shot no processo junto com o ConSentiment tem sido promissor.

Modelo	Conjunto de Dados Hotel22	
	Por Humanos	Por LLM
ABSA-BERT-CRF	0.671	0.432
ABSA-BERT-GRU	0.705	0.439
ABSA-BERT-Linear	0.714	0.446
ABSA-BERT-SAN	0.719	0.435
ABSA-BERT-TFM	0.713	0.421
MVE	0.676	0.386
ConSentiment	<b>0.816</b>	<b>0.770</b>

Tabela 7 – Resultados experimentais (pontuações F1) com a incorporação de LLMs nos modelos base e ConSentiment

Demonstramos experimentalmente que modelos baseados em um comitê de votação majoritária dentro do comitê ABSA enfrentam limitações. Um obstáculo primário surge na obtenção de um consenso entre diversos modelos ABSA, já que cada modelo pode gerar aspectos distintos, alguns dos quais podem estar apenas parcialmente corretos, e diferentes sentimentos estão associados a cada aspecto. Essa variabilidade na saída introduz complexidade ao processo de construção de consenso. Além disso, a natureza inerente do ABSA desafia o manuseio de aspectos e sentimentos ambíguos, especialmente ao fazer a transição entre idiomas e domínios. Em todos os cenários, o MVE produziu resultados similares ou inferiores aos dos modelos base. Ao analisarmos mais de perto, observamos que o MVE se concentra principalmente em instâncias com maior consenso entre todos os modelos, muitas vezes negligenciando casos

com maior incerteza. É aqui que o ConSentiment entra em jogo, alavancando a similaridade de aspectos e os grafos de consenso para formar um modelo de comitê para tarefas de ABSA.

---

## CONCLUSÃO

---

Em vez de propor mais uma nova técnica única de ABSA, a abordagem ConSentiment explora uma estratégia baseada em consenso usando múltiplos modelos base de ABSA. Foram introduzidos dois princípios: *Informação Consensual* e *Redução de Incerteza*, que exploram similaridade semântica para vincular aspectos similares e focam em abordar a incerteza vinculando previsões de sentimento divergentes.

A avaliação experimental em conjuntos de dados do mundo real mostrou que o ConSentiment alcançou desempenho superior em comparação com os modelos base de ABSA e superou um comitê de votação majoritária, conforme abordado na segunda questão de pesquisa. A eficácia do modelo em cenários multilíngues e multidomínio, como discutido na primeira questão de pesquisa, é notável, especialmente considerando a crescente relevância desses contextos para lidar com uma variedade de tópicos, produtos e serviços em diferentes países e para diversos públicos consumidores. Além disso, os resultados mostraram que a integração de LLMs ao ConSentiment melhorou as pontuações F1 em comparação com os modelos base respectivos. Em resumo, as principais contribuições deste trabalho foram:

- Desenvolvimento do método ConSentiment, um framework robusto para ABSA que combina múltiplos modelos de forma eficaz, aproveitando o consenso entre os modelos para melhorar a precisão das previsões de sentimento.
- Proposição de um algoritmo de construção de grafos de consenso baseado em ABSA, que captura a similaridade semântica entre aspectos e utiliza regras específicas do domínio para determinar a conexão entre vértices.

- Introdução de um framework de regularização baseada em consenso, que utiliza o grafo de consenso para refinamento das previsões de sentimento, preservando o consenso entre os vértices e reduzindo a incerteza associada às previsões.
- Demonstração experimental do desempenho do ConSentiment em comparação com métodos de referência em uma variedade de cenários, incluindo avaliações em diferentes domínios, idiomas e com dados rotulados por humanos e gerados por modelos de linguagem.
- Submissão de um artigo para a International Joint Conference on Artificial Intelligence (IJCAI) 2024.

Embora o ConSentiment tenha mostrado resultados promissores, existem algumas limitações e desafios que podem ser abordados em trabalhos futuros:

- **Lidar com aspectos e sentimentos ambíguos:** A identificação e tratamento de aspectos e sentimentos ambíguos continuam sendo um desafio importante em ABSA. O ConSentiment mitiga este problema conectando reviews e aspectos similares por meio de suas embeddings, o que já permitiu obter resultados promissores. Na entanto, abordagens mais avançadas de modelagem de incerteza, usando os próprios LLMs para identificar esses cenários podem ser exploradas para lidar com essa ambiguidade de forma mais eficaz.
- **Adaptação a diferentes domínios e idiomas:** Embora o ConSentiment tenha mostrado capacidade de generalização em cenários multidomínio e multilíngue, ainda há espaço para melhorias na adaptação a novos domínios e idiomas. Novamente, estratégias de transferência de conhecimento e técnicas de pré-treinamento nas LLMs mais recentes podem ser exploradas para melhorar a adaptabilidade do ConSentiment.
- **Integração com outras fontes de informação:** Além das informações fornecidas pelos modelos de ABSA, o ConSentiment pode se beneficiar da integração de outras fontes de informação, como dados contextuais adicionais e conhecimento externo, para melhorar ainda mais a qualidade das previsões de sentimento. Por exemplo, na análise de sentimento baseado em aspectos de produtos eletrônicos, fichas sobre características técnicas dos produtos podem ser incorporados como informação no grafo.

- **Eficiência computacional:** Embora o ConSentiment seja eficaz em melhorar a precisão das previsões de sentimento, sua eficiência computacional pode ser aprimorada. Explorar técnicas de otimização e paralelização pode tornar o ConSentiment mais escalável para lidar com conjuntos de dados maiores e treinar modelos mais complexos. Como cada modelo base pode ser treinado de forma independente, há um espaço para melhoria na questão da escalabilidade.
- **Limitação de dados rotulados:** A relevância do ConSentiment em cenários com dados rotulados limitados foi explorada, principalmente em conjunto com LLMs e aprendizado de prompt sem rótulo. No entanto, trabalhos futuros podem explorar cenários adicionais de ABSA e abordar o impacto das arquiteturas de LLM em evolução. Especificamente, incorporar diferentes LLMs e técnicas de engenharia de prompt é um caminho natural para novas investigações do ConSentiment.

Para fins de reprodutibilidade no processo de revisão, neste trabalho é apresentado um repositório da biblioteca, código-fonte e exemplos do ConSentiment em <https://github.com/GuilhermeMarcon/consentiment>.





## REFERÊNCIAS

---

---

AGGARWAL, C. C. **Neural Networks and Deep Learning**: A textbook. Cham: Springer, 2018. 497 p. ISBN 978-3-319-94462-3. Citado na página 33.

ALI, S.; MASOOD, K.; RIAZ, A.; SAUD, A. Named entity recognition using deep learning: A review. In: IEEE. **2022 International Conference on Business Analytics for Technology and Security (ICBATS)**. [S.l.], 2022. p. 1–7. Citado na página 28.

AZHAR, A. N.; KHODRA, M. L. Fine-tuning pretrained multilingual bert model for indonesian aspect-based sentiment analysis. In: **2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA)**. [S.l.: s.n.], 2020. p. 1–6. Citado nas páginas 34 e 41.

BIE, Y.; YANG, Y. A multitask multiview neural network for end-to-end aspect-based sentiment analysis. **Big Data Mining and Analytics**, TUP, v. 4, n. 3, p. 195–207, 2021. Citado na página 22.

BIRJALI, M.; KASRI, M.; BENI-HSSANE, A. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. **Knowledge-Based Systems**, Elsevier, v. 226, p. 107134, 2021. Citado na página 21.

BROWN, T.; MANN, B.; RYDER, N.; SUBBIAH, M.; KAPLAN, J. D.; DHARIWAL, P.; NEELAKANTAN, A.; SHYAM, P.; SASTRY, G.; ASKELL, A. *et al.* Language models are few-shot learners. **Advances in neural information processing systems**, v. 33, p. 1877–1901, 2020. Citado na página 22.

CHEN, F.; YANG, Z.; HUANG, Y. A multi-task learning framework for end-to-end aspect sentiment triplet extraction. **Neurocomputing**, Elsevier, v. 479, p. 12–21, 2022. Citado nas páginas 29 e 30.

CHEN, Z.; QIAN, T. Retrieve-and-edit domain adaptation for end2end aspect based sentiment analysis. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, v. 30, p. 659–672, 2022. Citado nas páginas 41 e 43.

COLLINS, M.; SINGER, Y. Unsupervised models for named entity classification. In: **Conference on Empirical Methods in Natural Language Processing**. [S.l.: s.n.], 1999. Citado na página 28.

COLLOBERT, R.; WESTON, J.; BOTTOU, L.; KARLEN, M.; KAVUKCUOGLU, K.; KUKSA, P. Natural language processing (almost) from scratch. **Journal of machine learning research**, v. 12, n. ARTICLE, p. 2493–2537, 2011. Citado na página 28.

CORTIS, K.; DAVIS, B. Over a decade of social opinion mining: a systematic review. **Artificial intelligence review**, Springer, v. 54, n. 7, p. 4873–4965, 2021. Citado na página 27.

- CRAWSHAW, M. Multi-task learning with deep neural networks: A survey. **arXiv preprint arXiv:2009.09796**, 2020. Citado na página 37.
- DO, H. H.; PRASAD, P.; MAAG, A.; ALSADOON, A. Deep learning for aspect-based sentiment analysis: a comparative review. **Expert systems with applications**, Elsevier, v. 118, p. 272–299, 2019. Citado nas páginas 22 e 28.
- DONG, S.; WANG, P.; ABBAS, K. A survey on deep learning and its applications. **Computer Science Review**, Elsevier, v. 40, p. 100379, 2021. Citado na página 22.
- FAN, Z.; WU, Z.; DAI, X.-Y.; HUANG, S.; CHEN, J. Target-oriented opinion words extraction with target-fused neural sequence labeling. In: **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. p. 2509–2518. Disponível em: <<https://aclanthology.org/N19-1259>>. Citado na página 29.
- FIALHO, P.; COHEUR, L.; QUARESMA, P. Benchmarking natural language inference and semantic textual similarity for portuguese. **Information**, MDPI, v. 11, n. 10, p. 484, 2020. Citado na página 34.
- FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. **Journal of computer and system sciences**, Elsevier, v. 55, n. 1, p. 119–139, 1997. Citado na página 38.
- GUPTA, M.; AGRAWAL, P. Compression of deep learning models for text: A survey. **ACM Transactions on Knowledge Discovery from Data (TKDD)**, ACM New York, NY, v. 16, n. 4, p. 1–55, 2022. Citado na página 30.
- HAO, Y.; DONG, L.; WEI, F.; XU, K. Investigating learning dynamics of bert fine-tuning. In: **Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing**. [S.l.: s.n.], 2020. p. 87–92. Citado na página 33.
- HE, R.; LEE, W. S.; NG, H. T.; DAHLMEIER, D. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In: **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**. [S.l.: s.n.], 2019. p. 504–515. Citado nas páginas 22 e 43.
- HOANG, M.; BIHORAC, O. A.; ROUCES, J. Aspect-based sentiment analysis using bert. In: **Proceedings of the 22nd nordic conference on computational linguistics**. [S.l.: s.n.], 2019. p. 187–196. Citado na página 33.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural computation**, MIT Press, v. 9, n. 8, p. 1735–1780, 1997. Citado na página 32.
- JAFARIAN, H.; TAGHAVI, A. H.; JAVAHERI, A.; RAWASSIZADEH, R. Exploiting bert to improve aspect-based sentiment analysis performance on persian language. In: **2021 7th International Conference on Web Research (ICWR)**. [S.l.: s.n.], 2021. p. 5–8. Citado nas páginas 34 e 41.

KENTON, J. D. M.-W. C.; TOUTANOVA, L. K. Bert: Pre-training of deep bidirectional transformers for language understanding. In: **Proceedings of NAACL-HLT**. [S.l.: s.n.], 2019. p. 4171–4186. Citado nas páginas 22, 30 e 32.

KURU, O.; CAN, O. A.; YURET, D. CharNER: Character-level named entity recognition. In: **Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers**. Osaka, Japan: The COLING 2016 Organizing Committee, 2016. p. 911–921. Disponível em: <<https://aclanthology.org/C16-1087>>. Citado na página 28.

LAMPLE, G.; BALLESTEROS, M.; SUBRAMANIAN, S.; KAWAKAMI, K.; DYER, C. **Neural Architectures for Named Entity Recognition**. arXiv, 2016. Disponível em: <<https://arxiv.org/abs/1603.01360>>. Citado na página 43.

LI, J.; SUN, A.; HAN, J.; LI, C. A survey on deep learning for named entity recognition. **IEEE Transactions on Knowledge and Data Engineering**, v. 34, n. 1, p. 50–70, 2022. Citado na página 28.

LI, X.; BING, L.; LI, P.; LAM, W. A unified model for opinion target extraction and target sentiment prediction. In: **Proceedings of the AAAI conference on artificial intelligence**. [S.l.: s.n.], 2019. v. 33, n. 01, p. 6714–6721. Citado na página 43.

LI, X.; BING, L.; ZHANG, W.; LAM, W. Exploiting bert for end-to-end aspect-based sentiment analysis. In: **Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)**. [S.l.: s.n.], 2019. p. 34–41. Citado nas páginas 22, 29, 30, 36, 40, 43, 53 e 54.

LI, X.; FENG, J.; MENG, Y.; HAN, Q.; WU, F.; LI, J. A unified MRC framework for named entity recognition. In: **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**. Online: Association for Computational Linguistics, 2020. p. 5849–5859. Disponível em: <<https://aclanthology.org/2020.acl-main.519>>. Citado na página 28.

LI, Z.; LI, X.; WEI, Y.; BING, L.; ZHANG, Y.; YANG, Q. Transferable end-to-end aspect-based sentiment analysis with selective adversarial learning. In: **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**. [S.l.: s.n.], 2019. p. 4590–4600. Citado na página 22.

LIANG, Y.; MENG, F.; ZHANG, J.; CHEN, Y.; XU, J.; ZHOU, J. A dependency syntactic knowledge augmented interactive architecture for end-to-end aspect-based sentiment analysis. **Neurocomputing**, v. 454, p. 291–302, 2021. ISSN 0925-2312. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231221007815>>. Citado nas páginas 40 e 43.

LIU, B. **Sentiment Analysis: Mining Opinions, Sentiments, and Emotions**. 2. ed. [S.l.]: Cambridge University Press, 2020. (Studies in Natural Language Processing). Citado nas páginas 21, 27 e 28.

LIU, B.; ZHANG, L. A survey of opinion mining and sentiment analysis. In: \_\_\_\_\_. **Mining Text Data**. Boston, MA: Springer US, 2012. p. 415–463. ISBN 978-1-4614-3223-4. Disponível em: <[https://doi.org/10.1007/978-1-4614-3223-4\\_13](https://doi.org/10.1007/978-1-4614-3223-4_13)>. Citado na página 21.

LIU, L.; SHANG, J.; REN, X.; XU, F.; GUI, H.; PENG, J.; HAN, J. Empower sequence labeling with task-aware neural language model. In: **Proceedings of the AAAI Conference on Artificial Intelligence**. [S.l.: s.n.], 2018. v. 32, n. 1. Citado na página 43.

LOPES, E. U.; CORRÊA, U. B.; FREITAS, L. A. de. Exploring bert for aspect extraction in portuguese language. **The International FLAIRS Conference Proceedings**, 2021. Citado nas páginas 34 e 41.

LU, T.; XIANG, Y.; ZHANG, L.; ZHANG, J. Sentence constituent-aware attention mechanism for end-to-end aspect-based sentiment analysis. **Multimedia Tools and Applications**, v. 81, n. 11, p. 15333–15348, May 2022. ISSN 1573-7721. Disponível em: <<https://doi.org/10.1007/s11042-022-12487-x>>. Citado nas páginas 40 e 43.

LUO, H.; LI, T.; LIU, B.; ZHANG, J. Doer: Dual cross-shared rnn for aspect term-polarity co-extraction. In: **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**. [S.l.: s.n.], 2019. p. 591–601. Citado na página 43.

MA, D.; LI, S.; WANG, H. Joint learning for targeted sentiment analysis. In: **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**. Brussels, Belgium: Association for Computational Linguistics, 2018. p. 4737–4742. Disponível em: <<https://aclanthology.org/D18-1504>>. Citado na página 29.

MA, X.; HOVY, E. End-to-end sequence labeling via bi-directional lstm-cnns-crf. In: **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**. [S.l.: s.n.], 2016. p. 1064–1074. Citado na página 43.

MAO, Y.; SHEN, Y.; YU, C.; CAI, L. A joint training dual-mrc framework for aspect based sentiment analysis. In: **Proceedings of the AAAI Conference on Artificial Intelligence**. [S.l.: s.n.], 2021. v. 35, n. 15, p. 13543–13551. Citado na página 22.

MCCALLUM, A.; LI, W. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In: **Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003 - Volume 4**. USA: Association for Computational Linguistics, 2003. (CONLL '03), p. 188–191. Disponível em: <<https://doi.org/10.3115/1119176.1119206>>. Citado na página 28.

MCCANN, B.; BRADBURY, J.; XIONG, C.; SOCHER, R. Learned in translation: contextualized word vectors. In: **Proceedings of the 31st International Conference on Neural Information Processing Systems**. [S.l.: s.n.], 2017. p. 6297–6308. Citado na página 32.

MCNAMEE, P.; MAYFIELD, J. Entity extraction without language-specific resources. In: **Proceedings of the 6th Conference on Natural Language Learning - Volume 20**. USA: Association for Computational Linguistics, 2002. (COLING-02), p. 1–4. Disponível em: <<https://doi.org/10.3115/1118853.1118873>>. Citado na página 28.

MIKOLOV, T.; CHEN, K.; CORRADO, G.; DEAN, J. **Efficient Estimation of Word Representations in Vector Space**. arXiv, 2013. Disponível em: <<https://arxiv.org/abs/1301.3781>>. Citado nas páginas 30 e 31.

MITCHELL, M.; AGUILAR, J.; WILSON, T.; DURME, B. V. Open domain targeted sentiment. In: **Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing**. Seattle, Washington, USA: Association for Computational Linguistics, 2013. p. 1643–1654. Disponível em: <<https://aclanthology.org/D13-1171>>. Citado na página 29.

NASEEM, U.; RAZZAK, I.; KHAN, S. K.; PRASAD, M. A comprehensive survey on word representation models: From classical to state-of-the-art word representation language models. **Transactions on Asian and Low-Resource Language Information Processing**, ACM New York, NY, v. 20, n. 5, p. 1–35, 2021. Citado nas páginas 30 e 31.

NAZIR, A.; RAO, Y.; WU, L.; SUN, L. Issues and challenges of aspect-based sentiment analysis: a comprehensive survey. **IEEE Transactions on Affective Computing**, IEEE, 2020. Citado na página 28.

PENNINGTON, J.; SOCHER, R.; MANNING, C. D. Glove: Global vectors for word representation. In: **Empirical Methods in Natural Language Processing (EMNLP)**. [s.n.], 2014. p. 1532–1543. Disponível em: <<http://www.aclweb.org/anthology/D14-1162>>. Citado na página 30.

PETERS, M. E.; AMMAR, W.; BHAGAVATULA, C.; POWER, R. Semi-supervised sequence tagging with bidirectional language models. In: **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**. [S.l.: s.n.], 2017. p. 1756–1765. Citado na página 32.

PETERS, M. E.; NEUMANN, M.; IYYER, M.; GARDNER, M.; CLARK, C.; LEE, K.; ZETTLEMOYER, L. **Deep contextualized word representations**. arXiv, 2018. Disponível em: <<https://arxiv.org/abs/1802.05365>>. Citado nas páginas 30 e 32.

PIRES, T.; SCHLINGER, E.; GARRETTE, D. How multilingual is multilingual bert? In: **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**. [S.l.: s.n.], 2019. p. 4996–5001. Citado nas páginas 34 e 41.

RAU, L. Extracting company names from text. In: **[1991] Proceedings. The Seventh IEEE Conference on Artificial Intelligence Application**. [S.l.: s.n.], 1991. i, p. 29–32. Citado na página 28.

RODRIGUES, R.; COUTO, P.; RODRIGUES, I. Ipr: The semantic textual similarity and recognizing textual entailment systems. In: **ASSIN@ STIL**. [S.l.: s.n.], 2019. p. 39–48. Citado na página 34.

SAGI, O.; ROKACH, L. Ensemble learning: A survey. **WIREs Data Mining and Knowledge Discovery**, v. 8, n. 4, p. e1249, 2018. Disponível em: <<https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1249>>. Citado nas páginas 37 e 38.

SARKAR, K. A stacked ensemble approach to bengali sentiment analysis. In: SPRINGER. **International Conference on Intelligent Human Computer Interaction**. [S.l.], 2019. p. 102–111. Citado na página 38.

SCHMITT, M.; STEINHEBER, S.; SCHREIBER, K.; ROTH, B. Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks. In: **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**. Brussels, Belgium: Association for Computational Linguistics, 2018. p. 1109–1114. Disponível em: <<https://aclanthology.org/D18-1139>>. Citado na página 29.

SOUZA, F.; NOGUEIRA, R.; LOTUFO, R. Bertimbau: pretrained bert models for brazilian portuguese. In: SPRINGER. **Brazilian conference on intelligent systems**. [S.l.], 2020. p. 403–417. Citado na página 34.

SZARVAS, G.; FARKAS, R.; KOCSOR, A. A multilingual named entity recognition system using boosting and c4.5 decision tree learning algorithms. In: TODOROVSKI, L.; LAVRAČ, N.; JANTKE, K. P. (Ed.). **Discovery Science**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. p. 267–278. ISBN 978-3-540-46493-8. Citado na página 28.

TIAN, Y.; YANG, L.; SUN, Y.; LIU, D. Cross-domain end-to-end aspect-based sentiment analysis with domain-dependent embeddings. **Complexity**, Hindawi, v. 2021, p. 5529312, Mar 2021. ISSN 1076-2787. Disponível em: <<https://doi.org/10.1155/2021/5529312>>. Citado nas páginas 40 e 43.

TRISNA, K. W.; JIE, H. J. Deep learning approach for aspect-based sentiment classification: A comparative review. **Applied Artificial Intelligence**, Taylor & Francis, p. 1–37, 2022. Citado na página 28.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L. u.; POLOSUKHIN, I. Attention is all you need. In: GUYON, I.; LUXBURG, U. V.; BENGIO, S.; WALLACH, H.; FERGUS, R.; VISHWANATHAN, S.; GARNETT, R. (Ed.). **Advances in Neural Information Processing Systems 30**. Curran Associates, Inc., 2017. p. 5998–6008. Disponível em: <<http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf>>. Citado nas páginas 32 e 33.

WANG, F. A general learning framework using local and global regularization. **Pattern Recognition**, Elsevier, v. 43, n. 9, p. 3120–3129, 2010. Citado na página 47.

WANG, X.; XU, G.; ZHANG, Z.; JIN, L.; SUN, X. End-to-end aspect-based sentiment analysis with hierarchical multi-task learning. **Neurocomputing**, Elsevier, v. 455, p. 178–188, 2021. Citado nas páginas 29 e 30.

WU, S.; DREDZE, M. Are all languages created equal in multilingual bert? In: **Proceedings of the 5th Workshop on Representation Learning for NLP**. [S.l.: s.n.], 2020. p. 120–130. Citado na página 41.

WU, Y.; SCHUSTER, M.; CHEN, Z.; LE, Q. V.; NOROUZI, M.; MACHEREY, W.; KRIKUN, M.; CAO, Y.; GAO, Q.; MACHEREY, K.; KLINGNER, J.; SHAH, A.; JOHNSON, M.; LIU, X.; KAISER, L.; GOUWS, S.; KATO, Y.; KUDO, T.; KAZAWA, H.; STEVENS, K.; KURIAN, G.; PATIL, N.; WANG, W.; YOUNG, C.; SMITH, J.; RIESA, J.; RUDNICK, A.; VINYALS, O.; CORRADO, G.; HUGHES, M.; DEAN, J. Google’s neural machine translation system: Bridging the gap between human and machine translation. **CoRR**, abs/1609.08144, 2016. Disponível em: <<http://arxiv.org/abs/1609.08144>>. Citado na página 32.

XIAO, D.; REN, F.; PANG, X.; CAI, M.; WANG, Q.; HE, M.; PENG, J.; FU, H. A hierarchical and parallel framework for end-to-end aspect-based sentiment analysis. **Neurocomputing**, v. 465, p. 549–560, 2021. ISSN 0925-2312. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231221013795>>. Citado nas páginas 35, 40 e 43.

YADAV, V.; BETHARD, S. A survey on recent advances in named entity recognition from deep learning models. In: **Proceedings of the 27th International Conference on Computational Linguistics**. Santa Fe, New Mexico, USA: Association for Computational Linguistics, 2018. p. 2145–2158. Disponível em: <<https://aclanthology.org/C18-1182>>. Citado na página 28.

ZHANG, M.; ZHANG, Y.; VO, D.-T. Neural networks for open domain targeted sentiment. In: **EMNLP**. [S.l.: s.n.], 2015. Citado na página 29.

ZHANG, S.; ELHADAD, N. Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts. **Journal of Biomedical Informatics**, v. 46, n. 6, p. 1088–1098, 2013. ISSN 1532-0464. Special Section: Social Media Environments. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1532046413001196>>. Citado na página 28.

ZHANG, W.; DENG, Y.; LIU, B.; PAN, S. J.; BING, L. Sentiment analysis in the era of large language models: A reality check. **arXiv preprint arXiv:2305.15005**, 2023. Citado nas páginas 34, 41 e 43.

ZHANG, W.; LI, X.; DENG, Y.; BING, L.; LAM, W. Towards generative aspect-based sentiment analysis. In: **Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)**. [S.l.: s.n.], 2021. p. 504–510. Citado na página 22.

ZHANG, Y.; YANG, Q. A survey on multi-task learning. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, 2021. Citado nas páginas 36 e 37.

