

Identificação de aspectos explícitos e implícitos em críticas gastronômicas em português: avaliando o potencial dos LLMs

Luiz H. N. Silva¹, Eloize R. M. Seno¹, Rozane R. Rebechi², Helena M. Caseli³
Fabiano M. Rocha Júnior¹, Guilherme A. Faller²

¹Instituto Federal de São Paulo (IFSP) – São Carlos, SP

² Depto de Letras – Universidade Federal do Rio Grande do SUL – Porto Alegre, RS

³ Depto de Ciência da Computação – Universidade Federal de São Carlos – São Carlos, SP

{nascimento.henrique1, fabiano.j}@aluno.ifsp.edu.br, guilfaller@gmail.com
eloize@ifsp.edu.br, rozanereb@gmail.com, helenacaseli@ufscar.br

Abstract. *Aspect identification is a fundamental step in Aspect-Based Sentiment Analysis (ABSA), which involves detecting the opinion target aspects in product or service reviews published on social media. Although there are many works developed for detecting aspects in English, there are few studies in this area for Portuguese, and LLMs have been little explored. Given this context, this research investigated the potential use of LLMs for aspect identification in culinary reviews in Portuguese.*

Resumo. *A identificação de aspectos é uma etapa fundamental da Análise de Sentimentos Baseada em Aspectos (ASBA) que consiste em detectar os aspectos alvos de opinião em avaliações de produtos ou serviços publicadas nas mídias sociais. Enquanto existem vários estudos focados na detecção de aspectos na língua inglesa, para o português há poucos trabalhos na área e os LLMs praticamente não têm sido explorados. Dado esse contexto, esta pesquisa investigou o potencial de uso de LLMs na identificação de aspectos em críticas gastronômicas em português.*

1. Introdução

A análise de sentimentos baseada em aspectos (ASBA) é uma subárea da Análise de Sentimentos (AS) que busca identificar e analisar opiniões e sentimentos relacionados a aspectos ou atributos específicos de uma entidade, produto ou serviço. Em uma avaliação de um restaurante, por exemplo, aspectos como “comida”, “serviço” e “preço” podem ser analisados individualmente, permitindo uma compreensão mais detalhada das opiniões dos clientes sobre cada um deles.

A ASBA representa o nível mais complexo da análise automática, devido à dificuldade de se modelar as conexões semânticas entre um determinado aspecto (termo) e as palavras que fazem parte do seu contexto [Zhang et al. 2018]. Uma etapa fundamental da ASBA consiste na identificação de aspectos, os quais podem ser explícitos ou implícitos, de acordo com a literatura [Schouten and Frasinca 2016, Soni and Rambola 2022]. Enquanto o aspecto explícito ocorre diretamente no texto, o aspecto implícito não é mencionado explicitamente, mas pode ser inferido pelo contexto. Por exemplo, na avaliação de um restaurante “A comida estava deliciosa, mas demorou muito para chegar.”, temos um aspecto explícito “comida” com sentimento positivo e um aspecto implícito “serviço” com sentimento negativo.

Enquanto para o inglês há uma vasta literatura relacionada à detecção de aspectos [Schouten and Frasinicar 2016, Zhang et al. 2018, Soni and Rambola 2022], para o português as pesquisas ainda são emergentes [Pereira 2021]. Além disso, os trabalhos existentes se baseiam principalmente no uso de regras, léxicos e em algoritmos de aprendizado de máquina, sendo que o uso de modelos de linguagem em larga escala (*Large Language Model – LLM*, no inglês) tem sido pouco explorado. À medida que o interesse público por modelos generativos pré-treinados, como os modelos da OpenAI, continua a crescer, espera-se que a utilidade desses modelos em resolver tarefas de PLN seja investigada. E nesse sentido, algumas iniciativas recentes têm surgido [Oliveira et al. 2023, Santos and Paraboni 2023].

Dado esse contexto, este estudo investigou a potencialidade de cinco LLMs na identificação de aspectos explícitos e implícitos em críticas gastronômicas em português. Críticas gastronômicas são textos escritos por críticos profissionais da gastronomia com experiência em avaliar restaurantes, pratos e experiências culinárias. A escolha desse domínio se justifica pelo fato de que as críticas gastronômicas, até onde se sabe, ainda não foram exploradas no contexto da ASBA em português.

2. Trabalhos Relacionados

Os trabalhos de identificação de aspectos para o português se baseiam, principalmente, no uso de léxicos [Costa and Pardo 2020], regras de linguagem [Vargas and Pardo 2020, Machado et al. 2021], em algoritmos de aprendizado de máquina tradicionais [Balage Filho 2017, Vargas and Pardo 2018] e no uso de *deep learning* [Lopes et al. 2021, Assi et al. 2022, Machado and Pardo 2022, Resplande et al. 2022]). Em [Resplande et al. 2022], por exemplo, os autores avaliaram o uso de modelos baseados em *Transformers* na extração de aspectos em avaliações de hotéis. Os aspectos extraídos foram classificados, posteriormente, como positivos, negativos ou neutros usando o LLM GPT-3. Em um trabalho anterior [Seno et al. 2024], o GPT-3.5 Turbo foi empregado na tarefa de detecção de aspectos e classificação de polaridade em comentários do domínio político. Em [Machado 2023], os autores compararam o uso de LLMs – GPT-3.5, Maritaca e Llama – com um modelo BERT e com vários classificadores tradicionais na identificação de aspectos em revisões de produtos eletrônicos, livros e hotéis. Nos experimentos, os melhores resultados para os aspectos explícitos foram obtidos pelo classificador *CRF* (o melhor *F-score* foi 81% para revisões de hotéis). Porém, para os aspectos implícitos o melhor resultado, em termos de porcentagem de acerto, foi obtido com o Llama 7B (52%).

De forma similar, este estudo também explorou o uso do GPT-3.5 e dos modelos da família Maritaca na detecção de aspectos em críticas gastronômicas. Porém, os modelos investigados aqui são variações mais recentes das versões usadas por [Machado 2023].

3. Identificação de aspectos em Críticas Gastronômicas usando LLMs

Para a identificação de aspectos em críticas gastronômicas foram explorados alguns dos LLMs mais populares da atualidade como o GPT-3.5 Turbo, o GPT-4o e GPT-4o mini¹. Segundo a OpenAI², o GPT-4o é o seu modelo mais avançado e inteligente para tarefas

¹<https://platform.openai.com/docs/api-reference/introduction>

²<https://platform.openai.com/docs/models>

Tabela 1. Prompts usados na anotação de aspectos explícitos e implícitos.

| |
|---|
| <p>Aspectos explícitos: Dada a sentença <i>EXEMPLO</i> com os alvos de opiniões explícitos, identifique os alvos de opinião explícitos na sentença (se houver) no formato [e - alvo1], se não houver nenhum alvo, indique com um '-'. <i>EXEMPLO</i>: “A pizza estava gostosa. E a sobremesa também.”. Saída: [e - pizza] [e - sobremesa]</p> |
| <p>Aspectos implícitos: Dada a sentença <i>EXEMPLO</i> com os alvos de opiniões implícitos, identifique os alvos de opinião implícitos (se houver) no formato [i - alvo], se não houver nenhum alvo, indique com um '-'. <i>EXEMPLO</i>: “A pizza estava gostosa, mas era muito cara. Além disso, estava fria”. Saída: [i - preço] [i - temperatura]</p> |

mais complexas. O GPT-4o mini é o modelo mais avançado na categoria de modelos pequenos, que também inclui o GPT-3.5 Turbo. Além desses LLMs, também foram investigados dois modelos monolíngues treinados para o português, o Sabiá-2-medium e o Sabiá-3³. Em experimentos reportados por [Almeida et al. 2024], o Sabiá-2-medium é comparado a vários outros LLMs, alcançando desempenho igual ou melhor que GPT-3.5 Turbo em várias análises. O Sabiá-3, por sua vez, lançado em julho de 2024, até o momento da escrita deste artigo não se tinha informações sobre o seu desempenho.

Todos os LLMs são modelos generativos baseados em *prompt*, que recebem como entrada um texto (*prompt*) contendo a descrição da tarefa a ser realizada e geram as saídas conforme solicitado. O grande desafio em lidar com esses modelos consiste em definir um *prompt* que gere as saídas exatamente como se espera para a tarefa. Vários *prompts* diferentes foram testados para a identificação de aspectos explícitos e implícitos no *corpus*. Foram experimentados *prompts* específicos para cada tipo de aspecto usando exemplos de anotação humana (i.e. abordagem *few-shot*) e sem o uso de exemplos de anotação (i.e. abordagem *zero-shot*). Contudo, percebeu-se uma facilidade maior dos modelos ao usar a abordagem *few-shot*. Assim, na anotação do *corpus* foram adotados os *prompts* apresentados na Tabela 1. Em todos os LLMs investigados a temperatura foi ajustada em zero, a fim de obter modelos mais determinísticos, conforme apontado por outros trabalhos da literatura [Oliveira et al. 2023, Santos and Paraboni 2023].

4. Corpus

Para os experimentos foi usado um conjunto de 1005 sentenças extraídas do *corpus* de críticas gastronômicas de [Rebecchi et al. 2021]. Cada sentença foi anotada por 5 anotadores humanos, todos pesquisadores da área de PLN, em duas etapas. Primeiramente os anotadores classificaram as sentenças em opinativa ou factual. Em seguida, aspectos explícitos e implícitos foram anotados, em dupla/trio, para as 374 (37,2%) sentenças consideradas opinativas pelos anotadores. Para estas, 432 aspectos foram identificados, sendo 88,6% explícitos e 11,4% implícitos. A Tabela 2 apresenta exemplos de sentenças com anotação de aspectos explícitos (em negrito) e implícitos.

Dado o fato de que não é possível determinar todos os aspectos possíveis para o *corpus*, não foi possível calcular o coeficiente *Kappa* para estimar a concordância entre os anotadores. Embora não se tenha obtido uma estimativa da concordância na anotação do *corpus*, a busca pelo consenso, seguida da clara convergência dos anotadores, permite assegurar que os aspectos identificados reproduzem de forma bastante fiel os aspectos que geralmente são considerados na avaliação de uma experiência gastronômica.

³Disponíveis por meio da MariTalk API como um *chatbot*.

Tabela 2. Exemplos de anotação de aspectos explícitos (em negrito) e implícitos.

| Sentença | Implícito |
|---|---------------------------|
| Se estiver sozinho, desista de tentar o omakassê (sequência de iguarias decididas e enviadas aos poucos pelo chef) — ele é gigante (para uma pessoa) e caro (42 itens, R\$ 390). | tamanho; preço |
| Não é demais lembrar: a casa só aceita dinheiro ou cheque – costume fora de moda, também trazido de outros tempos. | forma de pagamento |
| Carta de vinhos: Excelente, com muitas opções argentinas para todos os bolsos. | variedade (vinhos); preço |

5. Experimentos e Resultados

A Tabela 3 apresenta os resultados obtidos por cada LLM na detecção de aspectos explícitos e implícitos. O Sabiá-medium-2 obteve o melhor *F-score* (48,37%) para os aspectos explícitos, alcançando também a maior cobertura (77,75%). Contudo, a maior precisão (40,90%) foi obtida pelo GPT-4o mini. Já no que se refere aos aspectos implícitos, os resultados mostram uma grande dificuldade dos LLMs em identificar esse tipo de aspecto. Vale mencionar que essa dificuldade também foi relatada pelos humanos na anotação do *corpus*. Como os aspectos implícitos são inferidos pelo contexto, nem sempre é trivial perceber qual é o alvo de opinião. Em alguns casos, essa inferência exige um conhecimento mais especializado como no exemplo “Na boca, é equilibrado, com taninos firmes e boa estrutura.”, que se refere ao aspecto “vinho”. Para esse caso específico, apenas o modelo GPT-3.5 Turbo conseguiu identificar o aspecto implícito.

Tabela 3. Resultados obtidos para aspectos explícitos e implícitos.

| LLM | Explícitos | | | Implícitos | | |
|----------------|---------------|---------------|---------------|--------------|---------------|--------------|
| | Precisão | Cobertura | F-score | Precisão | Cobertura | F-score |
| Sabiá-2-medium | 35,11% | 77,75% | 48,37% | 1,60% | 26,00% | 3,01% |
| GPT-3.5 turbo | 31,53% | 76,18% | 44,60% | 1,95% | 32,00% | 3,68% |
| Sabiá-3 | 33,21% | 67,80% | 44,58% | 2,21% | 32,00% | 4,13% |
| GPT-4o | 21,51% | 76,18% | 33,54% | 3,90% | 34,00% | 7,00% |
| GPT-4o mini | 40,90% | 15,82% | 22,81% | 2,23% | 20,00% | 4,01% |

6. Conclusões

Este estudo investigou o uso de LLMs na detecção de aspectos em críticas gastronômicas. Nos experimentos, o LLM monolíngue Sabiá-2-medium mostrou um potencial maior na detecção de aspectos explícitos, do que os modelos multilíngues analisados. Enquanto que o Sabiá-3, também monolíngue, mostrou-se equivalente ao GPT-3.5 Turbo, superando o GPT-4o e o GPT-4o mini. Além de apresentarem desempenho superior ou equivalente aos obtidos pelos modelos multilíngues, os modelos monolíngues são bem mais acessíveis⁴. Com relação aos aspectos implícitos, todos os LLMs tiveram bastante dificuldade em identificar esse tipo de aspecto. O melhor desempenho foi obtido pelo GPT-4o (7% de *F-score*).

Como trabalhos futuros, pretende-se investigar a combinação de LLMs para a tarefa de identificação de aspectos, bem como a utilização de conhecimento do domínio de críticas gastronômicas para enriquecer os *prompts*.

⁴Os valores podem ser consultados em <https://openai.com/api/pricing/> e <https://www.maritaca.ai/>

Referências

- Almeida, T. S., Abonizio, H., Nogueira, R., and Pires, R. (2024). Sabiá-2: A new generation of portuguese large language models. *ArXiv*, abs/2403.09887.
- Assi, F. M., Candido, G. B., dos Santos Silva, L. N., Silva, D. F., and Caseli, H. M. (2022). Ufscar’s team at ABSAPT 2022: using syntax, semantics and context for solving the tasks. In Montes-y-Gómez, M. and et al., editors, *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*, volume 3202 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Balage Filho, P. P. (2017). *Aspect extraction in sentiment analysis for portuguese language*. PhD thesis, São Carlos - SP.
- Costa, R. and Pardo, T. (2020). Métodos baseados em léxico para extração de aspectos de opiniões em português. In *Anais do IX Brazilian Workshop on Social Network Analysis and Mining*, pages 61–72, Porto Alegre, RS, Brasil. SBC.
- Lopes, E., Correa, U., and Freitas, L. (2021). Exploring BERT for aspect extraction in portuguese language. *The International FLAIRS Conference Proceedings*, 34.
- Machado, M., Pardo, T., Ruiz, E., and Felippo, A. (2021). Learning rules for automatic identification of implicit aspects in portuguese. In *Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 82–91, Porto Alegre, RS, Brasil. SBC.
- Machado, M. and Pardo, T. A. S. (2022). Evaluating methods for extraction of aspect terms in opinion texts in Portuguese - the challenges of implicit aspects. In Calzolari, N., Béchet, F., Blache, P., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Odijk, J., and Piperidis, S., editors, *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3819–3828, Marseille, France. European Language Resources Association.
- Machado, M. T. (2023). *Methods for identifying aspects in opinion texts in Portuguese: the case of implicit aspects and their typological analysis*. PhD thesis, São Carlos - SP.
- Oliveira, A., Cecote, T., Silva, P., Gertrudes, J., Freitas, V., and Luz, E. (2023). How good is ChatGPT for detecting hate speech in portuguese? In *Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 94–103, Porto Alegre, RS, Brasil. SBC.
- Pereira, D. A. (2021). A survey of sentiment analysis in the portuguese language. *Artificial Intelligence Review*, 54(2):1087–1115.
- Rebechi, R. R., Nunes, R. R., Munhoz, L. R., and Marcon, N. O. (2021). Restaurant reviews in Brazil and the USA: A feast of cultural differences and their impact on translation. *Mutatis Mutandis. Revista Latinoamericana de Traducción*, 14:372–396.
- Resplande, J., Garcia, E., Junior, A., Rodrigues, R., Silva, D., Maia, D., Da Silva, N., Filho, A., and Soares, A. (2022). Deep learning Brasil at ABSAPT 2022: Portuguese transformer ensemble approaches. In Montes-y-Gómez, M. and et al., editors, *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*, volume 3202 of *CEUR Workshop Proceedings*. CEUR-WS.org.

- Santos, W. and Paraboni, I. (2023). Predição de transtorno depressivo em redes sociais: Bert supervisionado ou ChatGPT zero-shot? In *Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 11–21, Porto Alegre, RS, Brasil. SBC.
- Schouten, K. and Frasincar, F. (2016). Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(3):813–830.
- Seno, E., Silva, L., Anno, F., Rocha, F., and Caseli, H. (2024). Aspect-based sentiment analysis in comments on political debates in Portuguese: evaluating the potential of ChatGPT. In Gamallo, P., Claro, D., Teixeira, A., Real, L., Garcia, M., Oliveira, H. G., and Amaro, R., editors, *Computational Processing of the Portuguese Language: 16th Conference, PROPOR 2024*, pages 312–320, Santiago de Compostela, Galicia/Spain. Association for Computational Linguistics.
- Soni, P. K. and Rambola, R. (2022). A survey on implicit aspect detection for sentiment analysis: Terminology, issues, and scope. *IEEE Access*, 10:63932–63957.
- Vargas, F. A. and Pardo, T. A. S. (2018). Aspect clustering methods for sentiment analysis. In *Computational Processing of the Portuguese Language: 13th International Conference, PROPOR 2018, Canela, Brazil, September 24–26, 2018, Proceedings*, page 365–374, Berlin, Heidelberg. Springer-Verlag.
- Vargas, F. A. and Pardo, T. A. S. (2020). Linguistic rules for fine-grained opinion extraction. proceedings of the 14th International AAAI Conference on Web and Social Media, 2020.
- Zhang, L., Wang, S., and Liu, B. (2018). Deep learning for sentiment analysis : A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8.