Inferência de Avaliações em Cenários de *Item Cold Start* a Partir de Grafos de Conhecimento de Produtos Enriquecidos com Opiniões

Kirk Sahdo[1], **Tiago de Melo**[1]

[1]Escola Superior de Tecnologia – Universidade do Estado do Amazonas (UEA)
Manaus – AM – Brasil

{kmis.eng21,tmelo}@uea.edu.br

Abstract. *This paper addresses the cold start problem in recommender systems, which occurs when there is little or no information about new products. We propose the construction of Product Knowledge Graphs (PKGs) enriched with opinions extracted from user evaluations using Large Language Models (Sabiá-3.1), integrating product specifications with subjective information. The enriched PKG is processed for rating inference in item cold start scenarios: for each aspect category, the inferred polarity is computed as the arithmetic mean of the polarities of opinion tuples from similar products that share the same technical attribute in the PKG. Results show that this approach achieves 88.9% agreement in polarity classification against direct extraction from the target product's own reviews.*

Resumo. *Este trabalho aborda o problema de cold start em sistemas de recomendação, que ocorre quando há pouca ou nenhuma informação sobre novos produtos. Propõe-se a construção de Product Knowledge Graphs (PKGs) enriquecidos com opiniões extraídas das avaliações dos usuários utilizando Large Language Models (Sabiá-3.1), integrando as especificações dos produtos com informações subjetivas. O grafo enriquecido é processado para inferência de avaliações em cenários de item cold start: para cada categoria de aspecto, a polaridade inferida é a média aritmética das polaridades das tuplas de opinião de produtos similares que compartilham o mesmo atributo técnico no PKG. Os resultados mostram concordância de 88,9% na classificação de polaridade em relação à extração direta das avaliações do produto alvo.*

1. Introdução

Os sistemas de recomendação são amplamente empregados em plataformas de *e-commerce*, como a Amazon e a Taobao. Eles utilizam a análise de dados e algoritmos de aprendizado de máquina (*machine learning*) para recomendar produtos com base em interações prévias e nas preferências dos usuários.

O problema de *cold start* consiste na dificuldade de gerar recomendações precisas pela falta de histórico de interações e se manifesta em dois cenários principais [Yuan and Hernandez 2023]: (i) *item cold start*, quando um novo produto é adicionado à plataforma sem avaliações; e (ii) *user cold start*, quando um novo usuário cria sua conta sem histórico.

Além dos métodos tradicionais, pode-se citar a estruturação dos produtos a partir de *Product Knowledge Graphs* (PKGs). Os *PKGs* são representações estruturadas que capturam relações entre produtos, suas marcas, categorias e atributos, além de conexões entre produtos similares [Xu et al. 2020]. Uma abordagem inovadora para enriquecê-los e atenuar o *cold start* é o método *Product Graph enriched with Opinions* (PGOpi) proposto por Moreira et al. [Moreira et al. 2022], que mapeia opiniões extraídas de avaliações de usuários aos nós dos PKGs via aprendizado profundo. Neste trabalho, adaptou-se o PGOpi substituindo a extração de opiniões por uma abordagem baseada em *Large Language Models* (LLMs), especificamente o Sabiá-3.1 [Abonizio et al. 2024], que oferece maior precisão contextual, pode reduzir a dependência de anotação manual extensiva e de heurísticas de *distant supervision* [Simmering and Huoviala 2023].

O objetivo é avaliar o método para inferir avaliações de *smartphones* a partir de avaliações de produtos semelhantes, mitigando o problema *item cold start*. As principais contribuições são: (i) adaptação do PGOpi com LLM para extração de opiniões em português sem anotação manual; (ii) construção e enriquecimento de PKG para inferência em cenários de *item cold start*; e (iii) análise comparativa com concordância de 88,9% e MAE de 0,167.

2. Trabalhos Relacionados

Os *Knowledge Graphs* (KGs) são estruturas que representam informações como redes de entidades interconectadas [Hogan et al. 2021]. Nos sistemas de recomendação, tornaram-se fundamentais em cenários de *cold start*, em que as relações entre produtos compensam a falta de histórico [Wang et al. 2019, Hogan et al. 2021]. Xu et al. [Xu et al. 2020] propuseram o *Product Knowledge Graph Embedding*, e grandes empresas como Amazon [Dong 2018] e Walmart [Xu et al. 2020] já utilizam KGs para organizar catálogos e aprimorar recomendações.

O artigo-base deste estudo, Moreira et al. [Moreira et al. 2022], apresenta o PGOpi, que usa aprendizado profundo e supervisão à distância para mapear opiniões de avaliações aos atributos no grafo, permitindo que novos produtos recebam avaliações estimadas a partir de produtos similares já avaliados. Outros trabalhos, como Opinion Link [de Melo et al. 2019], também enriquecem catálogos com opiniões, mas dependem de dados rotulados manualmente; o PGOpi diferencia-se pela rotulação automática.

Para mitigar o *cold start*, outras abordagens incluem a decomposição de matrizes com atributos [Guo et al. 2019], que é limitada por tratar avaliações apenas como números, e o MetaKG [Du et al. 2023], que combina meta-aprendizado com KGs, mas não incorpora dados subjetivos. O LLM-PKGs [Wang et al. 2024] usa LLMs para a construção do grafo, mas é mais suscetível a alucinações do que o PGOpi, por gerar conhecimento via LLM. Na extração de opiniões, LLMs [Brown et al. 2020] superam métodos baseados em padrões linguísticos pela compreensão contextual e aprendizado com poucos exemplos [Simmering and Huoviala 2023]; no domínio de avaliações de aplicativos, Kotsifas et al. [Kotsifas et al. 2025] aplicam técnicas de PLN, incluindo análise de sentimento baseada em aspectos.

3. Metodologia

3.1. Conjunto de Dados

Escolheu-se a categoria de *smartphone* por sua abundante quantidade de avaliações e especificações técnicas. A lista dos itens foi retirada do catálogo do Mercado Livre, e as informações técnicas foram extraídas do GSMArena[1] via *scraping*. O *smartphone* alvo é o Samsung Galaxy A35 5G, escolhido por não ser topo de linha, o que favorece a existência de dispositivos similares, e por sua alta disponibilidade de avaliações. Cinco *smartphones* similares foram selecionados com base no compartilhamento de especificações técnicas e no maior número de avaliações disponíveis no Mercado Livre. A Tabela 1 apresenta as especificações dos seis dispositivos. Os dados foram armazenados em banco de dados SQLite para facilitar a integração com o código de construção do grafo.

Tabela 1. Smartphones selecionados e suas especificações técnicas

Especificação	Galaxy A35 5G *(alvo)*	Galaxy A54 5G	Moto G84	Redmi Note 13 Pro 5G	Infinix Note 40 5G	Poco X6 Pro
Processador	Exynos 1380	Exynos 1380	Snapdragon 695	Snapdragon 7s G2	Dimensity 7020	Dimensity 8300U
RAM (GB)	8	8	8	8	8	12
Armazen. (GB)	128/256	128/256	128/256	128/256	256	256/512
Tela (pol., tipo)	6,6”, S-AMOLED	6,4”, S-AMOLED	6,55”, P-OLED	6,67”, AMOLED	6,78”, AMOLED	6,67”, AMOLED
Câm. Traseira (MP)	50+8+5	50+12+5	50+8	200+8+2	108+2+2	64+8+2
Câm. Frontal (MP)	13	32	16	16	32	16
Bateria (mAh)	5000	5000	5000	5100	5000	5000
S.O.	Android 14	Android 13	Android 13	Android 13	Android 14	Android 14

3.2. Representação dos Produtos

Para a representação dos produtos, foi usado o *Product Knowledge Graph*. Segundo Moreira et al. [Moreira et al. 2022], o *PKG* é um grafo direcionado $G = \langle V, E, L_V, L_E \rangle$, em que V é o conjunto de nós (produtos ou atributos), $E \subseteq V \times V$ é o conjunto de arestas direcionadas, L_V é o conjunto de rótulos dos nós e L_E é o conjunto de rótulos das arestas.

Para simular o cenário de *item cold start*, as avaliações do *smartphone* alvo foram excluídas da fase de enriquecimento do *PKG*, sendo utilizadas apenas como referência de validação. A inferência de polaridade para cada categoria de aspecto é calculada como a **média aritmética** das polaridades das tuplas de opinião dos cinco produtos similares que compartilham o mesmo atributo técnico no *PKG*. Isso permite comparar diretamente a polaridade inferida com a polaridade real extraída das avaliações do Samsung Galaxy A35 5G.

3.3. Coleta de Avaliações

A coleta das avaliações foi feita por meio da API do Mercado Livre. O processo resultou em um *dataset* contendo 6.166 avaliações distribuídas pelos seis *smartphones*. Do total, 5.630 (91,3%) contêm comentários textuais válidos.

3.4. Extração de Opinião Baseada em Aspectos

A metodologia de extração de opiniões segue uma abordagem baseada em aspectos. Nove categorias foram definidas: armazenamento, processador, memória RAM, tela, resolução, câmera frontal, câmera traseira, bateria e sistema operacional. Cada opinião identificada pelo modelo foi representada como uma tupla: `<aspecto, polaridade,`

[1] https://www.gsmarena.com

`evidência, categoria, sentença>`, em que a polaridade varia de -1 (muito negativo) a $+1$ (muito positivo).

Para a extração, foi utilizado o Sabiá-3.1, desenvolvido pela Maritaca AI [Abonizio et al. 2024]. Um *prompt* de instrução hierárquica foi criado para guiar a LLM na extração, incluindo: mensagem do sistema, descrição da atividade, formato de saída em JSON e a entrada com o aspecto e comentário a ser analisado. A estrutura hierárquica segue o *prompting* passo a passo de Li et al. [Li et al. 2025], que melhora a precisão na extração da tupla de opinião. A qualidade da extração foi verificada por inspeção manual qualitativa de uma amostra aleatória de 50 tuplas, observando-se a coerência entre aspecto, polaridade, evidência e a sentença original. Essa etapa foi utilizada como checagem exploratória, sem cálculo formal de concordância entre os anotadores.

4. Experimentos e Resultados

4.1. Construção e Enriquecimento do PKG

Utilizou-se a biblioteca PyVis [Perrone et al. 2020] com Python para a construção do *PKG*. O grafo resultante representa os seis *smartphones* e suas especificações técnicas na forma de nós interconectados. A partir da extração das opiniões, o grafo foi enriquecido: um nó contendo o aspecto e a evidência da opinião é criado e se liga ao seu respectivo atributo técnico por meio de uma aresta com o rótulo "opiniao". A Figura 1 ilustra esse processo. Os dados e o código utilizados neste trabalho podem ser disponibilizados mediante solicitação aos autores.

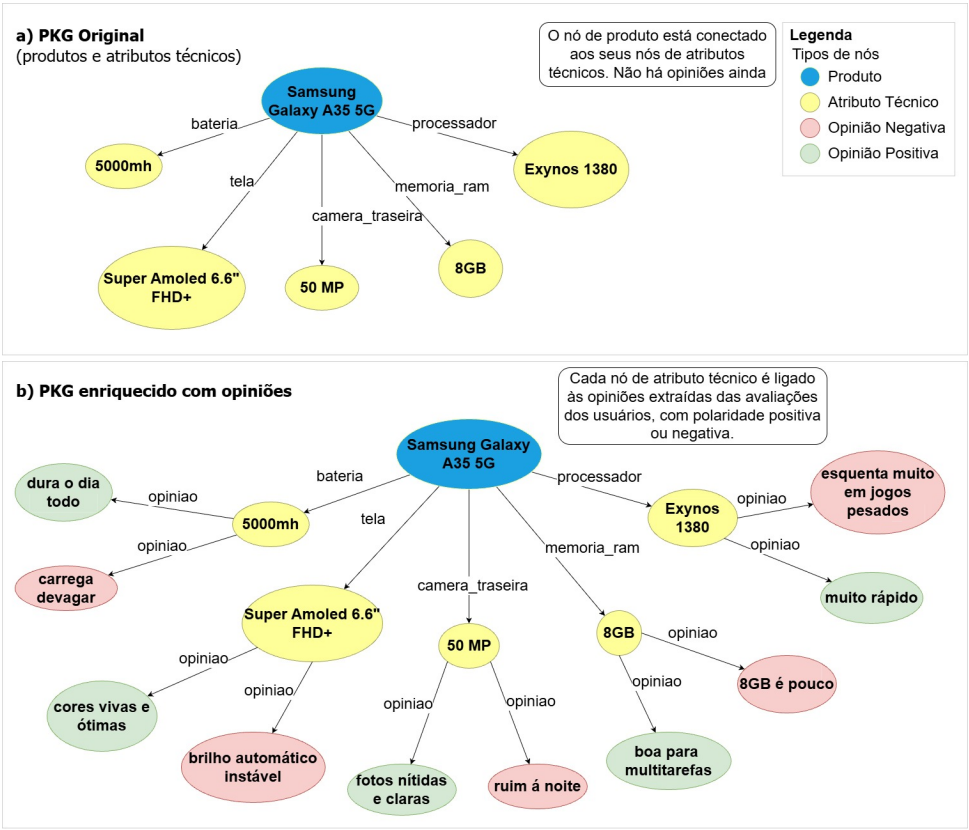

Figura 1. Exemplo ilustrativo do *PKG* enriquecido com opiniões

4.2. Resultados da Extração

O processo de extração resultou em 14.389 tuplas de opinião das 5.630 avaliações analisadas, com uma média de 2,56 opiniões por avaliação. Do total, 61,8% são positivas, 37,0% negativas e 1,2% neutras. Neste conjunto de dados, observou-se a dominância de sentimentos fortes ($|\text{polaridade}| \geq 0{,}8$, correspondendo a 84,9% das tuplas), em que respostas emocionais extremas motivam a escrita de avaliações.

4.3. Análise Comparativa

A avaliação compara duas abordagens para o Samsung Galaxy A35 5G: (1) **Extração Direta**, que calcula as polaridades médias a partir das avaliações do próprio produto alvo (2.685 opiniões); e (2) **PKG enriquecido**, que infere as polaridades agregando, por média aritmética, as opiniões de produtos similares no *PKG* (11.704 opiniões). A Tabela 2 apresenta os resultados por categoria.

	Extração Direta		**Enriquecimento PKG**			
Categoria	**N**	**Pol. Média**	**N**	**Pol. Média**	**Dif. Abs.**	**Concordância**
Armazenamento	88	+0,524	334	+0,532	0,008	Sim
Bateria	646	−0,290	2.267	−0,164	0,126	Sim
Câmera Frontal	27	+0,656	155	+0,322	0,334	Sim
Câmera Traseira	402	+0,672	1.872	+0,271	0,400	Sim
Memória RAM	67	+0,578	305	+0,676	0,099	Sim
Processador	904	+0,397	4.262	+0,406	0,009	Sim
Resolução	17	+0,818	102	+0,512	0,306	Sim
Sist. Operacional	121	+0,053	732	−0,121	0,174	Não
Tela	413	+0,538	1.675	+0,486	0,052	Sim
Total	**2.685**		**11.704**		**0,167**	**8/9**

Tabela 2. Comparação por categoria: quantidade (N), polaridade média e diferença absoluta (Dif. Abs.) nos dois métodos, com concordância de polaridade. Na linha Total, as colunas N trazem a soma das opiniões e a coluna Dif. Abs. traz o erro absoluto médio (MAE) das nove categorias.

A concordância na classificação de polaridade alcançou 88,9% (8 de 9 categorias), com erro absoluto médio (MAE) de 0,167 (média não ponderada das diferenças absolutas das nove categorias). Como ambos os lados da comparação são produzidos pelo próprio Sabiá-3.1, essas métricas medem a *consistência* com a leitura direta do modelo, não a acurácia frente a julgamento humano. A única discordância ocorreu em Sistema Operacional, em que a extração direta indicou sentimento levemente positivo (+0,053) enquanto o enriquecimento produziu sentimento levemente negativo (−0,121); essa divergência pode ser atribuída ao tamanho amostral reduzido no método direto (121 opiniões contra 732 no enriquecimento). Categorias com maior volume amostral, como processador (904 opiniões) e bateria (646 opiniões), apresentaram as menores diferenças absolutas (0,009 e 0,126, respectivamente). Observa-se ainda que o enriquecimento tende a *moderar* polaridades extremas, como câmera traseira (+0,672 → +0,271), ao agregar opiniões de múltiplos produtos com perfis de sentimento distintos.

5. Conclusões

Este estudo aborda a inferência de avaliações de *smartphones* em *item cold start* usando *PKGs* enriquecidos com opiniões extraídas pelo Sabiá-3.1, integrando dados objetivos (especificações técnicas) e subjetivos (opiniões) para inferir polaridades de novos produtos a partir de similares.

A análise revelou alinhamento promissor (concordância de 88,9%, MAE de 0,167). Por medirem consistência com a leitura direta do modelo, e não acurácia frente a anotação humana, esses valores indicam que a inferência por produtos similares pode servir de *fallback* no *cold start*, embora a acurácia real ainda exija validação humana. Entre as limitações, a similaridade foi avaliada qualitativamente e os resultados restringem-se a uma categoria (*smartphones*); ganhos de experiência do usuário e descoberta de produtos permanecem como hipótese. Trabalhos futuros devem explorar métricas formais de similaridade, rotulação humana para medir acurácia, modelos multilíngues e validação em outras plataformas e categorias.

Agradecimentos

Este trabalho foi parcialmente financiado pelo Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional, Tratamento e Disseminação de Informação (INCT-TILD-IAR), financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo nº 408490/2024-1. Esta pesquisa é parcialmente apoiada pela Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM), por meio do Projeto NeuralBond (UNIVERSAL 2023, Proc. 01.02.016301.04300/2023-04).

Declaração de Uso de IA Generativa

Neste trabalho, foram utilizadas ferramentas de IA generativa (LLMs) para as seguintes finalidades: (1) o modelo Sabiá-3.1 foi utilizado como parte da metodologia de pesquisa para extração de opiniões baseada em aspectos a partir das avaliações dos produtos; (2) ferramentas de IA generativa foram utilizadas para auxílio na revisão linguística e na estruturação do texto do artigo. A responsabilidade integral pelo conteúdo permanece com os autores humanos, incluindo a verificação de todas as referências bibliográficas.

Origem do Trabalho

Este trabalho é resultado de um Trabalho de Conclusão de Curso (TCC) do curso de Engenharia da Computação da Escola Superior de Tecnologia da Universidade do Estado do Amazonas (UEA).

Referências

Abonizio, H., Almeida, T. S., Laitz, T., Malaquias Junior, R., Bonás, G. K., Nogueira, R., and Pires, R. (2024). Sabiá-3 technical report. Technical report, Maritaca AI.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901.

de Melo, T., da Silva, A. S., de Moura, E. S., and Calado, P. (2019). Opinionlink: Leveraging user opinions for product catalog enrichment. *Information Processing & Management*, 56(3):823–843.

Dong, X. L. (2018). Challenges and innovations in building a product knowledge graph. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, page 2869.

Du, Y., Zhu, X., Chen, L., Fang, Z., and Gao, Y. (2023). Metakg: Meta-learning on knowledge graph for cold-start recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 35(10).

Guo, X., Yin, S.-C., Zhang, Y., Li, W., and He, Q. (2019). Cold start recommendation based on attribute-fused singular value decomposition. *IEEE Access*, 7:11349–11359.

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G. d., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., et al. (2021). Knowledge graphs. *ACM Computing Surveys*, 54(4):1–37.

Kotsifas, M. F. A., Lüders, R., and Silva, T. H. (2025). How culture shapes customers: A cross-continent analysis of apps reviews using NLP techniques. In *Anais do XL Simpósio Brasileiro de Banco de Dados (SBBD)*, pages 760–766.

Li, X., Zhang, K., and Han, D. (2025). Duality-driven aspect sentiment triplet extraction with llm and iterative reinforcement. *Symmetry*, 17(5):642.

Moreira, J., de Melo, T., Barbosa, L., and da Silva, A. (2022). A distantly supervised approach for enriching product graphs with user opinions. *Journal of Intelligent Information Systems*, 59:435–454.

Perrone, G., Unpingco, J., and Lu, H.-m. (2020). Network visualizations with pyvis and visjs. *arXiv preprint arXiv:2006.04951*.

Simmering, P. F. and Huoviala, P. (2023). Large language models for aspect-based sentiment analysis. *arXiv preprint arXiv:2310.18025*.

Wang, M., Guo, Y., Zhang, D., Jin, J., Li, M., Schonfeld, D., and Zhou, S. (2024). Enabling explainable recommendation in e-commerce with llm-powered product knowledge graph. *arXiv preprint arXiv:2412.01837*.

Wang, X., He, X., Cao, Y., Liu, M., and Chua, T.-S. (2019). Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 950–958.

Xu, D., Ruan, C., Korpeoglu, E., Kumar, S., and Achan, K. (2020). Product knowledge graph embedding for e-commerce. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 672–680.

Yuan, H. and Hernandez, A. A. (2023). User cold start problem in recommendation systems: A systematic review. *IEEE Access*, 11:136958–136977.