





ARTIGO DE PESQUISA/RESEARCH PAPER


Uma Abordagem Híbrida para Sistemas de Recomendação com Base em Avaliações Textuais e Filtragem Colaborativa


A Hybrid Approach for Recommender Systems Based on Textual Reviews and Collaborative Filtering

Naan Vasconcelos   [Universidade Federal de São João del Rei | naan.vasconcelos@aluno.ufsj.edu.br]

Yan Andrade  [Universidade Federal de Minas Gerais | yandandrade@ufmg.br]

Adriano Pereira  [Universidade Federal de Minas Gerais | adrianoc@dcc.ufmg.br]

Leonardo Rocha  [Universidade Federal de São João del Rei | lrocha@ufsj.edu.br]

 Departamento de Ciência da Computação, Universidade Federal de São João del-Rei (UFSJ), Av. Visconde do Rio Preto, s/n, Colônia do Marçal, São João del-Rei, MG, 36301-160, Brasil.

Resumo. Sistemas de recomendação (SsR) utilizam cada vez mais avaliações textuais para aprimorar a modelagem de preferências. Embora os SsR baseados em revisões (RARs) capturem nuances semânticas, apresentam dificuldades em lidar com padrões colaborativos. A filtragem colaborativa (FC) tradicional é robusta em relações estruturais, mas limitada na semântica. Propomos uma abordagem híbrida que integra representações semânticas derivadas de RARs em modelos de FC. Avaliada em três conjuntos de dados da Amazon, nossa proposta demonstra que os modelos híbridos superam os modelos de referência, alcançando um erro de previsão até 2,8 vezes menor e uma precisão 60% maior, destacando seu potencial para recomendações robustas e contextualizadas.

Abstract. Recommender systems increasingly leverage textual reviews to enhance preference modeling. While review-aware systems (RARs) capture semantic nuances, they often struggle with collaborative patterns. Traditional collaborative filtering (CF) is robust in structural relationships but limited in semantics. We propose a hybrid approach integrating RAR-derived semantic representations into CF models. Evaluated across three Amazon datasets, our results show that hybrid models outperform baselines, achieving up to 2.8 times lower prediction error and 60% higher precision. These findings highlight the potential of hybrid models for robust, context-aware recommendations.

Palavras-chave: Sistemas de recomendação, Filtragem colaborativa, Avaliações textuais, Processamento de Linguagem Natural

Keywords: Recommender systems, Collaborative filtering, Textual reviews, Natural Language Processing

Recebido/Received: 15 June 2026 • **Aceito/Accepted:** 15 June 2026 • **Publicado/Published:** 10 July 2026

1 Introdução

Com o crescimento exponencial da informação online, os Sistemas de Recomendação (SRs) consolidaram-se como ferramentas fundamentais para mitigar a sobrecarga informacional, inferindo interesses a partir de interações explícitas ou implícitas dos usuários [Jannach *et al.*, 2010; Ricci *et al.*, 2015]. Nos últimos anos, o uso de avaliações textuais, por meio dos *Review-Aware Recommender Systems* (RARs), ganhou grande destaque. A relevância e o crescimento acelerado desta abordagem são evidenciados por revisões recentes da literatura [Bittencourt *et al.*, 2025], que mapeiam como esses sistemas superam as notas numéricas ao capturar a subjetividade dos usuários. Essa abordagem parte da premissa de que textos capturam a subjetividade e as nuances de preferência de forma mais rica do que notas numéricas [Luiz *et al.*, 2018; Vasconcelos *et al.*, 2024]. Impulsionados por avanços em Processamento de Linguagem Natural (PLN) e arquiteturas de aprendizado profundo, os RARs têm demonstrado notável capacidade na extração de aspectos semânticos complexos.

Contudo, apesar desses avanços na modelagem semântica, estudos recentes apontam que os RARs frequentemente falham ao negligenciar relações colaborativas estruturais, como a coocorrência de consumo e as similaridades latentes

entre usuários. Por serem comumente derivados de tarefas supervisionadas, esses modelos tendem a apresentar baixa efetividade em cenários de alta esparsidade (o problema da cauda longa) e carecem de robustez frente a dados ruidosos ou ambíguos [Ferrari Dacrema *et al.*, 2019; Luo *et al.*, 2023]. Em contrapartida, as abordagens colaborativas clássicas [Amatriain and Basilico, 2015], embora limitadas na captura de nuances textuais, demonstram grande robustez na exploração de padrões coletivos de consumo e oferecem maior transparência na interpretação dos fatores de recomendação.

Para transpor as limitações isoladas de cada paradigma, este trabalho investiga o potencial de uma abordagem híbrida. A estratégia consiste em extrair representações semânticas de usuários e itens via RARs e incorporá-las como entrada em modelos colaborativos tradicionais. O objetivo é analisar se essa fusão entre a expressividade textual e a robustez estrutural resulta em recomendações mais eficazes e interpretáveis. Para guiar esta investigação, formulamos três perguntas de pesquisa principais:

- **PP1:** Métodos baseados em informações textuais capturam as preferências dos usuários de forma mais precisa do que os colaborativos?
- **PP2:** Essa precisão na modelagem se traduz em reco-

mendações práticas mais úteis e relevantes?

- **PP3:** Qual é o potencial da combinação híbrida entre métodos colaborativos e semântica textual?

Experimentos conduzidos com três bases da Amazon (*Musical Instruments, Digital Music e Office Products*), comparando modelos RARs (CARM, CARP, HRDR) com estratégias colaborativas clássicas, como a Decomposição em Valores Singulares (SVD), Fatoração de Matriz Não-Negativa (NNMF), *K-Nearest Neighbors* (KNN)), revelaram um contraste importante. Os resultados indicam que, embora os métodos textuais dominem a predição de preferência individual (PP1), eles falham em gerar listas de recomendação superiores aos modelos tradicionais (PP2). Contudo, a nossa proposta híbrida superou consistentemente ambas as abordagens isoladas, reduzindo o erro de predição em cerca de 3 vezes e elevando a precisão em até 60% (PP3). Adicionalmente, analisamos a diversidade das recomendações geradas, identificando um *trade-off* relevante entre os ganhos de precisão e a manutenção da variedade e surpresa nas listas sugeridas. Esse trabalho de iniciação científica resultou na publicação de um artigo no WebMedia 2025 (A3) [Vasconcelos et al., 2025].

2 Trabalhos Relacionados

Sistemas de Recomendação (SsR) são categorizados pela natureza das informações utilizadas. Enquanto abordagens tradicionais operam sobre registros históricos, métodos recentes incorporam conteúdo textual para capturar nuances do comportamento do usuário. Esta seção revisa os métodos colaborativos clássicos e os *Review-Aware Recommender Systems* (RARs).

2.1 Métodos Colaborativos Tradicionais

Métodos colaborativos constituem a base clássica dos SsR [Amatriain and Basilico, 2015], assumindo que usuários com comportamentos passados semelhantes compartilharão interesses futuros. Destacam-se as técnicas de fatoração de matrizes [Koren et al., 2009] como SVD [Zhou et al., 2015] e NNMF [Luo et al., 2014], que projetam usuários e itens em espaços latentes para capturar atributos como estilo e qualidade. Complementarmente, os métodos baseados em vizinhança estimam preferências via similaridade direta entre perfis (**UserKNN**) ou itens (**ItemKNN**) [Vargas and Castells, 2014]. Embora eficientes, esses métodos possuem limitações na captura de nuances subjetivas textuais.

2.2 Sistemas de Recomendação Review-Aware

Os RARs exploram revisões textuais para mitigar problemas de esparsidade e *cold start* [Srfi et al., 2020; Al-Ghuribi and Mohd Noah, 2019]. Uma taxonomia recente [Bittencourt et al., 2025] os organiza em quatro perspectivas: modelagem, extração, objetivo e arquitetura. Quanto à modelagem, diversas estratégias são exploradas, como *Document Modeling* [Zheng et al., 2017] e *Rating Aggregation* [Kim et al., 2016]. O modelo HRDR [Liu et al., 2020] destaca-se ao integrar *ratings* e textos via codificadores paralelos para robustez contra ruído.

A extração de informação identifica traços latentes globais [McAuley and Leskovec, 2013] ou locais. O

modelo CARP [Li et al., 2019] diferencia-se por vetorizar comentários via Redes Neurais Convolucionais (CNNs) para extrair características através de modelagem pareada. Quanto à arquitetura, os modelos evoluíram para redes profundas baseadas em atenção e grafos [Cai et al., 2022]. O modelo CARM [Li et al., 2021] alinha fatores latentes a contextos textuais via mecanismos de atenção. Apesar do progresso, os RARs falham em generalizar padrões colaborativos devido ao foco supervisionado em PLN [Ferrari Dacrema et al., 2019], motivando a abordagem híbrida proposta.

Diante do exposto, o avanço em relação ao estado da arte proposto neste trabalho reside justamente em preencher a lacuna existente entre os modelos puramente textuais e os colaborativos. Em vez de tratá-los como abordagens competidoras, propomos uma esteira híbrida onde a riqueza semântica supervisionada dos RARs atua como um pré-processamento enriquecedor para fortalecer a robustez consolidada dos métodos colaborativos clássicos.

3 Estratégia Proposta

Este trabalho propõe uma abordagem híbrida para Sistemas de Recomendação, com o objetivo de combinar a capacidade dos modelos *Review-Aware Recommender Systems* (RARs) de extrair informações semânticas ricas das avaliações textuais com a robustez dos métodos colaborativos tradicionais na modelagem estrutural. A estratégia divide-se em duas etapas principais, ilustradas na Figura 1: (i) a geração de *ratings* preditivos a partir de comentários textuais via RARs; e (ii) o uso desses valores como entrada para os algoritmos colaborativos, substituindo as interações explícitas originais.



Figura 1. Fluxo da abordagem híbrida proposta entre RARs e FC

Na primeira etapa, modelos RARs (como o CARP) são empregados para gerar estimativas contínuas de interesse. Cada instância de entrada é composta por uma quádrupla (u, i, r_{ui}, t_{ui}) , que representa a interação entre o usuário u e o item i , acompanhada do *rating* original atribuído r_{ui} e da revisão textual t_{ui} correspondente (Passo 1 da Figura 1). O objetivo é treinar arquiteturas de codificação semântica para aprender representações densas e prever um valor escalar \hat{r}_{ui} . Esse valor atua ponderando e refinando o *rating* original estritamente com base nas nuances do conteúdo textual (Passo 2).

Para ilustrar essa mecânica, considere um usuário u que avaliou um filme i com *rating* $r_{ui} = 5$ e escreveu a seguinte revisão: "O filme tem uma fotografia deslumbrante, mas a história é arrastada e os personagens são mal desenvolvidos". Apesar da nota numérica máxima, o conteúdo textual revela uma insatisfação significativa com aspectos centrais da obra. Ao processar essa entrada, o modelo RARs reavalia o real nível de satisfação do usuário, gerando um *rating* preditivo $\hat{r}_{ui} = 3.2$. Esse valor sintético captura as nuances negativas expressas na revisão que haviam sido omitidas pela nota original, resultando em uma nova coleção de interações (u, i, \hat{r}_{ui}) semanticamente ajustada (Passo 3).

Na segunda etapa da abordagem, essa coleção modificada de *ratings* refinados alimenta métodos de recomendação colaborativa clássicos, como fatoração matricial ou estra-

tégias baseadas em vizinhança (Passo 4). Diferentemente dos cenários tradicionais, os modelos colaborativos operam aqui sobre sinais semanticamente derivados. A motivação por trás dessa estratégia é investigar se a estrutura latente das relações usuário-item pode ser preservada e, sobretudo, potencializada quando os dados originais são substituídos por representações que carregam contexto.

Essa estruturação modular não apenas isola e avalia a contribuição das representações semânticas geradas por cada RAR em um ambiente colaborativo comum, mas também facilita a análise de complementaridade entre os modelos. Ao garantir que as duas etapas derivem de uma mesma fonte textual, a abordagem assegura que as recomendações finais sejam sensíveis tanto ao comportamento coletivo quanto à subjetividade individual.

4 Ambiente Experimental

Nesta seção, detalhamos o ambiente configurado para avaliar a proposta e responder às perguntas de pesquisa formuladas.

4.1 Bases de Dados e Métricas

Consideramos três coleções da Amazon (*Musical Instruments*, *Digital Music* e *Office Products*), amplamente utilizadas na literatura por apresentarem domínios e densidades de interação distintos [Vasconcelos et al., 2024; Bittencourt et al., 2023]. Como detalhado na Tabela 1, cada instância associa um *rating* numérico a uma avaliação textual. A avaliação fundamenta-se em dois grupos de métricas: (i) erro contínuo (MAE e MSE), para mensurar a precisão na modelagem das preferências; e (ii) precisão de ranking (*Hits*, *Precision* e *Recall* @ 10, 50 e 100), para aferir a utilidade das sugestões. A significância estatística das comparações é verificada pelo teste de Wilcoxon com nível de confiança de 5% [Vinagre et al., 2021].

Tabela 1. Visão geral das coleções utilizadas na avaliação.

Base de Dados	# Usuários	# Itens	Esparsidade
<i>Office Products</i>	4.905	2.420	99,98%
<i>Digital Music</i>	16.561	11.797	99,91%
<i>Musical Instruments</i>	27.528	10.620	99,92%

4.2 Modelos e Configuração Híbrida

Foram selecionados quatro modelos tradicionais (SVD, NNMF, UserKNN e ItemKNN) e três RARs (CARM, CARP e HRDR), visando diversidade metodológica e eficiência computacional [Bittencourt et al., 2023]. A implementação híbrida seguiu a modularidade descrita na Seção 3: os modelos RARs geraram predições de preferência a partir dos textos, que foram organizadas em uma matriz sintética de interações. Esta matriz serviu como entrada exclusiva para os algoritmos colaborativos, totalizando 12 combinações híbridas. A comparação entre esses híbridos e seus respectivos modelos base permite isolar o impacto das representações semânticas na eficácia das recomendações colaborativas.

5 Resultados e Discussões

Esta seção apresenta os resultados obtidos a partir dos experimentos, organizando a discussão em torno das três perguntas de pesquisa que guiam este trabalho. As análises empíricas detalhadas encontram-se reportadas nas Tabelas 2, 3 e 4.

5.1 PP1: Preferências com Informação Textual

A análise da **PP1** foca na capacidade dos modelos em capturar preferências individuais via métricas de erro contínuo (MAE e MSE - Bloco 1 das tabelas). Observamos uma superioridade consistente dos modelos RARs em todos os cenários avaliados. O modelo CARP obteve os menores erros nas bases *Musical Instruments* (MAE 0,3294 e MSE 0,2296) e *Digital Music* (MAE 0,5606), enquanto CARM e HRDR lideraram em *Office Products*. Esses resultados confirmam que as avaliações textuais são fontes semânticas ricas, permitindo uma modelagem de preferências significativamente mais precisa do que os métodos colaborativos puros, que dependem apenas de *ratings* numéricos.

Assim, a PP1 é respondida positivamente. Os métodos que exploram informações textuais capturam de forma mais precisa as preferências dos usuários. Contudo, essa capacidade descritiva precisa se traduzir em recomendações úteis, conforme avaliamos a seguir.

5.2 PP2: Utilidade com Informação Textual

Para responder à **PP2**, investigamos se a precisão na modelagem (PP1) se traduz em listas de recomendação relevantes, avaliando métricas de ranking (*Hits*, *Precision*, *Recall* - Bloco 2 das tabelas). Contrastando com os resultados de erro, os métodos colaborativos tradicionais superaram os RARs em quase todos os cenários. O SVD dominou as bases de música (ex: HIT10 de 0,0252 em *Musical Instruments*), enquanto o ItemKNN destacou-se em *Office Products*. A limitação dos RARs em gerar recomendações superiores evidencia uma falha em capturar padrões colaborativos e dinâmicas coletivas essenciais.

Assim, a PP2 é negativamente respondida. A riqueza semântica isolada não garante a utilidade prática das recomendações, o que abre uma possibilidade para emprego de estratégias integradas.

5.3 PP3: Potencialidades da Abordagem Híbrida

A **PP3** avalia se a integração entre semântica textual e aprendizado colaborativo supera as limitações individuais dos modelos. Os resultados mostram que as abordagens híbridas alcançaram o melhor desempenho global em ambos os conjuntos de métricas (Blocos 3 e 4 das tabelas). Em *Musical Instruments*, o modelo híbrido NNMF_{CARM} reduziu o erro em **quase 3x** em relação à sua versão base (MAE de 0,7163 para 0,2496). Na base *Digital Music*, o NNMF_{CARP} obteve um HIT10 **60% superior** à sua versão puramente colaborativa (de 0,0084 para 0,0136). Até mesmo modelos estruturalmente robustos, como o SVD, apresentaram ganhos generalizados ao operar sobre *ratings* refinados semanticamente.

Esses achados respondem positivamente à PP3, demonstrando que a combinação entre a expressividade textual e a robustez das relações colaborativas resulta em sistemas mais eficazes, adaptáveis e sensíveis ao contexto.

5.4 Limitações e Diversidade

Embora a estratégia híbrida apresente avanços significativos em erro e precisão, a análise da métrica de Diversidade Intra-Lista (ILD) (Tabela 5) revela um *trade-off* importante. De

Tabela 2. Avaliação dos algoritmos nas três coleções de dados. Os Blocos 1 e 3 comparam os modelos em métricas de erro, enquanto os Blocos 2 e 4 os avaliam em métricas de recomendação. Teste estatístico de Wilcoxon ($p = 0,05$): ▲ (ganho), ● (empate) e ▼ (perda).

Coleção			Musical Instruments								
Métrica	MAE	MSE	Hits			Precision			Recall		
T	-	-	10	50	100	10	50	100	10	50	100
Bloco 1			Bloco 2								
CARM	0,4732▲	0,5507▲	0,0141	0,0654	0,1100	0,0014	0,0013	0,0011	0,0168	0,0779	0,1323
CARP	0,3294▲	0,2296▲	0,0092	0,0403	0,0711	0,0009	0,0008	0,0007	0,0110	0,0480	0,0847
HRDR	0,4255▲	0,4906▲	0,0041	0,0157	0,0206	0,0004	0,0003	0,0002	0,0049	0,0187	0,0245
NNMF	0,7163	0,9747	0,0173▲	0,0754▲	0,1315▲	0,0017▲	0,0015▲	0,0013▲	0,0206▲	0,0899▲	0,1567▲
SVD	0,7152	0,9763	0,0252▲	0,1117▲	0,2006▲	0,0025▲	0,0022▲	0,0018▲	0,0300▲	0,1331▲	0,2391▲
ItemKNN	0,6059	1,1390	0,0153▲	0,0710▲	0,1325▲	0,0015▲	0,0014▲	0,0013▲	0,0182▲	0,0846▲	0,1579▲
UserKNN	0,6079	1,0951	0,0160▲	0,0726▲	0,1337▲	0,0016▲	0,0014▲	0,0013▲	0,0191▲	0,0865▲	0,1593▲
Bloco 3			Bloco 4								
NNMF _{CARM}	0,2496▲	0,1135▲	0,0183▲	0,0822▲	0,1513▲	0,0018▲	0,0016▲	0,0015▲	0,0218▲	0,0979▲	0,1803▲
SVD _{CARM}	0,2539▲	0,1174▲	0,0274▲	0,1270▲	0,2334▲	0,0027▲	0,0025▲	0,0023▲	0,0327▲	0,1513▲	0,2781▲
ItemKNN _{CARM}	0,2757▲	0,2510▲	0,0176▲	0,0807▲	0,1507▲	0,0017▲	0,0016▲	0,0015▲	0,0210▲	0,0962▲	0,1796▲
UserKNN _{CARM}	0,3787▲	0,2904▲	0,0172▲	0,0801▲	0,1500▲	0,0017▲	0,0016▲	0,0015▲	0,0205▲	0,0954▲	0,1788▲
NNMF _{CARP}	0,2941▲	0,1667▲	0,0180▲	0,0811▲	0,1521▲	0,0018▲	0,0016▲	0,0015▲	0,0215▲	0,0967▲	0,1813▲
SVD _{CARP}	0,3026▲	0,1762▲	0,0284▲	0,1256▲	0,2225▲	0,0028▲	0,0025▲	0,0022▲	0,0338▲	0,1497▲	0,2652▲
ItemKNN _{CARP}	0,3110▲	0,3223▼	0,0185▲	0,0853▲	0,1557▲	0,0018▲	0,0017▲	0,0015▲	0,0221▲	0,1017▲	0,1617▲
UserKNN _{CARP}	0,3980▼	0,3447▼	0,0182▲	0,0871▲	0,1626▲	0,0018▲	0,0017▲	0,0016▲	0,0217▲	0,1038▲	0,1938▲
NNMF _{HRDR}	0,2819▲	0,1394▲	0,0175▲	0,0756▲	0,1323▲	0,0017▲	0,0015▲	0,0013▲	0,0208▲	0,0901▲	0,1577▲
SVD _{HRDR}	0,2875▲	0,1451▲	0,0276▲	0,1227▲	0,2210▲	0,0027▲	0,0024▲	0,0022▲	0,0329▲	0,1462▲	0,2634▲
ItemKNN _{HRDR}	0,3291▲	0,3355▲	0,0173▲	0,0809▲	0,1509▲	0,0017▲	0,0016▲	0,0015▲	0,0206▲	0,0964▲	0,1798▲
UserKNN _{HRDR}	0,4419▼	0,3804▲	0,0187▲	0,0818▲	0,1515▲	0,0018▲	0,0016▲	0,0015▲	0,0223▲	0,0975▲	0,1806▲

Tabela 3. Continuação da Tabela 2. Resultados para a coleção *Digital Music*.

Coleção			Digital Music								
Métrica	MAE	MSE	Hits			Precision			Recall		
T	-	-	10	50	100	10	50	100	10	50	100
Bloco 1			Bloco 2								
CARM	0,5647▲	0,5680▲	0,0062	0,0206	0,0306	0,0006	0,0004	0,0003	0,0046	0,0153	0,0227
CARP	0,5606▲	0,5657▲	0,0079	0,0308	0,0435	0,0008	0,0006	0,0004	0,0059	0,0229	0,0323
HRDR	0,7898▲	0,7263▲	0,0082	0,0320	0,0518	0,0008	0,0006	0,0005	0,0061	0,0237	0,0385
NNMF	1,1743	0,8172	0,0084●	0,0336●	0,0544●	0,0008●	0,0006●	0,0005●	0,0062●	0,0249●	0,0403●
SVD	1,1807	0,8139	0,0112▲	0,0507▲	0,0686▲	0,0011▲	0,0009▲	0,0008▲	0,0083▲	0,0376▲	0,0510▲
ItemKNN	1,0037	1,2312	0,0108▲	0,0382▲	0,0672▲	0,0010▲	0,0007▲	0,0006▲	0,0076▲	0,0284▲	0,0499▲
UserKNN	0,9783	0,1186	0,0105▲	0,0362▲	0,0543●	0,0010▲	0,0007▲	0,0005●	0,0078▲	0,0269▲	0,0403●
Bloco 3			Bloco 4								
NNMF _{CARM}	0,2981▲	0,3978▲	0,0121▲	0,0567▲	0,1036▲	0,0012▲	0,0011▲	0,0010▲	0,0090▲	0,0421▲	0,0769▲
SVD _{CARM}	0,3079▲	0,4030▲	0,0152▲	0,0632▲	0,1102▲	0,0015▲	0,0012▲	0,0011▲	0,0113▲	0,0470▲	0,0819▲
ItemKNN _{CARM}	0,4127▲	0,4780▲	0,0123▲	0,0507▲	0,0807▲	0,0012▲	0,0010▲	0,0008▲	0,0091▲	0,0376▲	0,0599▲
UserKNN _{CARM}	0,4464▲	0,4899▲	0,0118▲	0,0454▲	0,0723▲	0,0011▲	0,0009▲	0,0007▲	0,0088▲	0,0337▲	0,0538▲
NNMF _{CARP}	0,3507▲	0,4302▲	0,0136▲	0,0517▲	0,0921▲	0,0012▲	0,0010▲	0,0009▲	0,0101▲	0,0384▲	0,0685▲
SVD _{CARP}	0,3587▲	0,4425▲	0,0162▲	0,0719▲	0,1333▲	0,0016▲	0,0014▲	0,0013▲	0,0120▲	0,0534▲	0,0992▲
ItemKNN _{CARP}	0,4328▲	0,4491▲	0,0120▲	0,0466▲	0,0824▲	0,0011▲	0,0009▲	0,0008▲	0,0089▲	0,0346▲	0,0612▲
UserKNN _{CARP}	0,4572▲	0,4786▲	0,0147▲	0,0481▲	0,0739▲	0,0012▲	0,0009▲	0,0007▲	0,0109▲	0,0357▲	0,0548▲
NNMF _{HRDR}	0,3888▲	0,4559▲	0,0117▲	0,0468▲	0,0846▲	0,0011▲	0,0009▲	0,0008▲	0,0087▲	0,0347▲	0,0629▲
SVD _{HRDR}	0,3901▲	0,4582▲	0,0173▲	0,0628▲	0,1028▲	0,0015▲	0,0012▲	0,0010▲	0,0129▲	0,0468▲	0,0764▲
ItemKNN _{HRDR}	0,4205▲	0,4482▲	0,0127▲	0,0468▲	0,0808▲	0,0012▲	0,0009▲	0,0008▲	0,0094▲	0,0347▲	0,0600▲
UserKNN _{HRDR}	0,4657▲	0,4860▲	0,0116▲	0,0406▲	0,0757▲	0,0011▲	0,0008▲	0,0007▲	0,0086▲	0,0301▲	0,0562▲

forma geral, as abordagens híbridas apresentaram diversidade inferior ou estatisticamente equivalente à dos seus respectivos modelos base. Em *Musical Instruments*, combinações como NNMF_{CARM} e NNMF_{CARP} perderam variedade nas listas recomendadas, um padrão que se repete de forma consistente nas coleções *Digital Music* e *Office Products*.

Esse comportamento sugere que o refinamento textual

tende a privilegiar padrões semânticos dominantes aprendidos das avaliações, reduzindo a descoberta de itens distintos e a variabilidade das sugestões. Tal limitação evidencia o *trade-off* entre acurácia e surpresa, reforçando a importância de investigar e incorporar, em trabalhos futuros, estratégias de controle de diversidade e re-ranqueamento [Zhao et al., 2025] para garantir listas equilibradas.

Tabela 4. Continuação da Tabela 2. Resultados para a coleção *Office Products*.

Coleção	Office Products										
	Métrica	MAE	MSE	Hits			Precision			Recall	
T	-	-	10	50	100	10	50	100	10	50	100
	Bloco 1			Bloco 2							
CARM	0,4251▲	0,4734▲	0,0010	0,0053	0,0101	0,0001	0,0001	0,0001	0,0009	0,0049	0,0093
CARP	0,5992▲	0,6626▲	0,0082	0,0307	0,0507	0,0008	0,0006	0,0005	0,0076	0,0283	0,0467
HRDR	0,4392▲	0,3598▲	0,0011	0,0061	0,0133	0,0001	0,0001	0,0001	0,0010	0,0056	0,0123
NNMF	0,7224	0,8640	0,0085●	0,0309●	0,0517●	0,0008●	0,0006●	0,0005●	0,0078●	0,0286●	0,0476●
SVD	0,7259	0,8725	0,0126▲	0,0526▲	0,0805▲	0,0012▲	0,0010▲	0,0008▲	0,0116▲	0,0485▲	0,0742▲
ItemKNN	0,6633	1,0773	0,0138▲	0,0557▲	0,1050▲	0,0013▲	0,0011▲	0,0010▲	0,0127▲	0,0513▲	0,0926▲
UserKNN	0,6317	0,8910	0,0135▲	0,0502▲	0,0921▲	0,0013▲	0,0010▲	0,0009▲	0,0124▲	0,0463▲	0,0849▲
	Bloco 3			Bloco 4							
NNMF _{CARM}	0,2433▲	0,1101▲	0,0123▲	0,0520▲	0,0917▲	0,0012▲	0,0010▲	0,0009▲	0,0113▲	0,0479▲	0,0845▲
SVD _{CARM}	0,2475▲	0,1143▲	0,0173▲	0,0752▲	0,1303▲	0,0017▲	0,0015▲	0,0013▲	0,0159▲	0,0693▲	0,1201▲
ItemKNN _{CARM}	0,2036▲	0,1616▲	0,0166▲	0,0741▲	0,1249▲	0,0016▲	0,0014▲	0,0012▲	0,0153▲	0,0683▲	0,1151▲
UserKNN _{CARM}	0,2316▲	0,1665▲	0,0161▲	0,0713▲	0,1207▲	0,0016▲	0,0014▲	0,0012▲	0,0148▲	0,0657▲	0,1112▲
NNMF _{CARP}	0,2920▲	0,1446▲	0,0133▲	0,0506▲	0,0913▲	0,0013▲	0,0010▲	0,0009▲	0,0123▲	0,0466▲	0,0841▲
SVD _{CARP}	0,2975▲	0,1527▲	0,0184▲	0,0811▲	0,1431▲	0,0018▲	0,0016▲	0,0014▲	0,0170▲	0,0747▲	0,1319▲
ItemKNN _{CARP}	0,2199▲	0,1726▲	0,0173▲	0,0704▲	0,1208▲	0,0017▲	0,0014▲	0,0012▲	0,0159▲	0,0649▲	0,1113▲
UserKNN _{CARP}	0,2435▲	0,1774▲	0,0161▲	0,0651▲	0,1108▲	0,0016▲	0,0013▲	0,0011▲	0,0148▲	0,0600▲	0,1021▲
NNMF _{HRDR}	0,2980▲	0,1564▲	0,0129▲	0,0511▲	0,0813▲	0,0012▲	0,0010▲	0,0008▲	0,0119▲	0,0471▲	0,0749▲
SVD _{HRDR}	0,3039▲	0,1626▲	0,0189▲	0,0759▲	0,1306▲	0,0018▲	0,0015▲	0,0013▲	0,0174▲	0,0699▲	0,1203▲
ItemKNN _{HRDR}	0,2270▲	0,1945▲	0,0167▲	0,0615▲	0,1119▲	0,0016▲	0,0012▲	0,0011▲	0,0154▲	0,0567▲	0,1031▲
UserKNN _{HRDR}	0,2625▲	0,2037▲	0,0164▲	0,0685▲	0,1115▲	0,0016▲	0,0013▲	0,0011▲	0,0151▲	0,0631▲	0,1027▲

Tabela 5. Avaliação dos modelos híbridos frente aos seus respectivos pares em relação à diversidade (ILD). Resultados são avaliados com teste de Wilcoxon ($p = 0,05$). Símbolos: ▲ (ganho significativo), ● (empate estatístico), ▼ (perda significativa).

Coleção	Musical Instruments			Digital Music			Office Products		
	Diversidade (ILD)			Diversidade (ILD)			Diversidade (ILD)		
T	10	50	100	10	50	100	10	50	100
CARM	0,1416	0,1388	0,1362	0,2423	0,2374	0,2358	0,1798	0,1681	0,1660
CARP	0,1415	0,1397	0,1384	0,2637	0,2580	0,2554	0,1436	0,1444	0,1443
HRDR	0,1397	0,1360	0,1330	0,2406	0,2377	0,2354	0,1650	0,1610	0,1605
NNMF	0,1323	0,1401	0,1423	0,2501	0,2486	0,2488	0,1970	0,2030	0,2057
SVD	0,1322	0,1414	0,1438	0,2480	0,2484	0,2485	0,2040	0,2030	0,2054
ItemKNN	0,1223	0,1098	0,1151	0,2246	0,2235	0,2213	0,1581	0,1646	0,1644
UserKNN	0,1223	0,1098	0,1151	0,2301	0,2214	0,2257	0,1523	0,1519	0,1500
NNMF _{CARM}	0,1320▼	0,1398●	0,1429●	0,2506●	0,2489●	0,2479●	0,1977●	0,2025●	0,2056●
SVD _{CARM}	0,1330▼	0,1411●	0,1425▼	0,2488●	0,2472▼	0,2473▼	0,2041●	0,2037●	0,2068▲
ItemKNN _{CARM}	0,1223▼	0,1098▼	0,1150▼	0,2234▼	0,2248▼	0,2234▼	0,1614▼	0,1618▼	0,1610▼
UserKNN _{CARM}	0,1223▼	0,1098▼	0,1151▼	0,2301▼	0,2214▼	0,2257▼	0,1523▼	0,1519▼	0,1500▼
NNMF _{CARP}	0,1320▼	0,1407●	0,1424●	0,2504▼	0,2497▼	0,2489▼	0,1989▲	0,2022●	0,2051●
SVD _{CARP}	0,1330▼	0,1411●	0,1428▼	0,2487▼	0,2482▼	0,2473▼	0,2041●	0,2037●	0,2055●
ItemKNN _{CARP}	0,1223▼	0,1098▼	0,1151▼	0,2301▼	0,2214▼	0,2257▼	0,1614▲	0,1618▼	0,1610▼
UserKNN _{CARP}	0,1223▼	0,1098▼	0,1151▼	0,2301▼	0,2214▼	0,2257▼	0,1523●	0,1519●	0,1500●
NNMF _{HRDR}	0,1336▼	0,1390▼	0,1424●	0,2498●	0,2487●	0,2480●	0,1986▲	0,2028●	0,2052●
SVD _{HRDR}	0,1337▼	0,1417●	0,1432●	0,2481●	0,2485●	0,2479●	0,2045●	0,2030●	0,2051●
ItemKNN _{HRDR}	0,1223▼	0,1098▼	0,1151▼	0,2234▼	0,2248▼	0,2234▼	0,1629▼	0,1618▼	0,1614●
UserKNN _{HRDR}	0,1223▼	0,1098▼	0,1151▼	0,2301▼	0,2214▼	0,2257▼	0,1523▼	0,1519●	0,1500●

6 Conclusão

Este trabalho investigou o potencial de abordagens híbridas que integram a riqueza semântica dos *Review-Aware Recommender Systems* (RARs) à robustez estrutural da filtragem colaborativa tradicional. Os resultados demonstram que, em-

bora os modelos RARs superem os métodos clássicos na modelagem de preferências individuais (erros significativamente menores), eles apresentam limitações na geração de recomendações úteis ao negligenciar padrões de consumo coletivo. A proposta híbrida resolve essa dicotomia ao utilizar representações textuais como entrada para algoritmos colaborativos,

superando sistematicamente os modelos isolados com redução de erro de até 3x e ganhos de precisão superiores a 60%.

Apesar da eficácia na acurácia, a análise de diversidade revelou um *trade-off* importante: a tendência de redução na variabilidade das sugestões devido ao reforço de padrões semânticos dominantes. Esse achado aponta para a necessidade de futuras investigações sobre mecanismos de controle de diversidade, como técnicas de re-ranqueamento [Carraro and Bridge, 2024], para garantir listas equilibradas e exploratórias.

Nossa meta é expandir esta abordagem em três frentes. Primeiro, planeja-se aplicar Grandes Modelos de Linguagem (LLMs) abertos, na etapa de refinamento textual, explorando sua capacidade de extrair nuances semânticas sem a necessidade de treinamento supervisionado. Segundo, conduziremos avaliações de custo computacional para compreender a viabilidade de implantação dessas arquiteturas híbridas em ambientes de alto volume de dados. Por fim, ampliaremos os experimentos para bases de diferentes domínios, bem como para paradigmas de recomendação em tempo real baseados em *Multi-Armed Bandits* (MAB) [Andrade et al., 2023; Vasconcelos et al., 2024], visando tornar o sistema mais sensível a contextos dinâmicos.

Declarações complementares

Financiamento

Esta pesquisa foi financiada por: CNPq, Capes, Fapemig, Fapesp, AWS, NVIDIA, CIIA-Saúde e Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional, Tratamento e Disseminação de Informação (INCT-TILD-IAR; 408490/2024-1).

Contribuições dos autores

N. Vasconcelos, Y. Andrade, A. Pereira e L. Rocha contribuíram para a concepção do estudo (Conceptualization) e análise formal dos resultados (Formal analysis). N. Vasconcelos foi o responsável pelo desenvolvimento do software (Software), investigação (Investigation) e escrita do rascunho original (Writing – original draft). L. Rocha atuou na supervisão da pesquisa (Supervision). Y. Andrade e A. Pereira colaboraram na validação da metodologia (Validation). Todos os autores participaram da revisão (Writing – review & editing) e aprovaram o manuscrito final.

Conflitos de interesse

Os autores declaram que não têm nenhum conflito de interesses

Disponibilidade de dados e materiais

Os conjuntos de dados (e/ou softwares) gerados e/ou analisados durante o estudo atual serão feitos mediante solicitação

Outras informações relevantes

Ferramentas de Inteligência Artificial Generativa foram utilizadas como suporte à revisão gramatical do texto. Os autores realizaram uma revisão completa do texto e assumem total responsabilidade pela integridade das informações apresentadas.

Referências

Al-Ghuribi, S. M. and Mohd Noah, S. A. (2019). Multi-criteria review-based recommender system—the state of the art. *IEEE Access*, 7:169446–169468. DOI: 10.1109/ACCESS.2019.2954861.

Amatriain, X. and Basilico, J. (2015). *Recommender Systems in Industry: A Netflix Case Study*, pages 385–419. DOI: 10.1007/978-1-4899-7637-6_11.

Andrade, Y., Silva, N., Silva, T., Pereira, A., Dias, D., Albergaria, E. T., and Rocha, L. (2023). A complete framework for offline and counterfactual evaluations of interactive recommendation systems. In *Proceedings of the 29th Brazilian Symposium on Multimedia and the Web, WebMedia '23*, page 193–197, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3617023.3617049.

Bittencourt, G., Fonseca, G., Andrade, Y., Silva, N., and Rocha, L. (2023). A survey on review-aware recommendation systems. In *Proceedings of the 29th Brazilian Symposium on Multimedia and the Web, WebMedia '23*, page 198–207, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3617023.3617050.

Bittencourt, G., Vasconcelos, N., Andrade, Y., Silva, N., Cunha, W., Colombo Dias, D. R., Gonçalves, M. A., and Rocha, L. (2025). Review-aware recommender systems (rars): Recent advances, experimental comparative analysis, discussions, and new directions. *ACM Comput. Surv.* Just Accepted. DOI: 10.1145/3744661.

Cai, Y., Wang, Y., Wang, W., and Chen, W. (2022). Rign: Review-aware interactive graph convolutional network for review-based item recommendation. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 475–484. DOI: 10.1109/BigData55660.2022.10020296.

Carraro, D. and Bridge, D. (2024). Enhancing recommendation diversity by re-ranking with large language models. *ACM Trans. Recomm. Syst.* DOI: 10.1145/3700604.

Ferrari Dacrema, M., Cremonesi, P., and Jannach, D. (2019). Are we really making much progress? a worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems, RecSys '19*, page 101–109, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3298689.3347058.

Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G. (2010). *Recommender systems: an introduction*. Cambridge University Press.

Kim, D., Park, C., Oh, J., Lee, S., and Yu, H. (2016). Convolutional matrix factorization for document context-aware recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16*, page 233–240, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/2959100.2959165.

Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37. DOI: 10.1109/MC.2009.263.

Li, C., Quan, C., Peng, L., Qi, Y., Deng, Y., and Wu, L. (2019). A capsule network for recommendation and explaining what you like and dislike. In *Proceedings of the 42nd International ACM SIGIR, SIGIR '19*, page 275–284. DOI: 10.1145/3331184.3331216.

Li, D., Liu, H., Zhang, Z., Lin, K., Fang, S., Li, Z., and Xiong, N. N. (2021). Carm: Confidence-aware recommender model via review representation learning and historical rating behavior in the online platforms. *Neurocomputing*, 455:283–296. DOI: 10.1016/j.neucom.2021.03.122.

- Liu, H., Wang, Y., Peng, Q., Wu, F., Gan, L., Pan, L., and Jiao, P. (2020). Hybrid neural recommendation with joint deep representation learning of ratings and reviews. *Neurocomputing*, 374:77–85. DOI: 10.1016/j.neucom.2019.09.052.
- Luiz, W., Viegas, F., Alencar, R., Mourão, F., Salles, T., Carvalho, D., Gonçalves, M. A., and Rocha, L. (2018). A feature-oriented sentiment rating for mobile app reviews. In *Proceedings of the 2018 World Wide Web Conference, WWW '18*, page 1909–1918, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee. DOI: 10.1145/3178876.3186168.
- Luo, S., Ma, C., Xiao, Y., and Song, L. (2023). Improving long-tail item recommendation with graph augmentation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM '23*, page 1707–1716. DOI: 10.1145/3583780.3614929.
- Luo, X., Zhou, M., Xia, Y., and Zhu, Q. (2014). An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. *IEEE Transactions on Industrial Informatics*, 10(2):1273–1284. DOI: 10.1109/TII.2014.2308433.
- McAuley, J. and Leskovec, J. (2013). Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13*, page 165–172, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/2507157.2507163.
- Ricci, F., Rokach, L., and Shapira, B. (2015). Recommender systems: introduction and challenges. *Recommender systems handbook*, pages 1–34.
- Srifi, M., Oussous, A., Ait Lahcen, A., and Mouline, S. (2020). Recommender systems based on collaborative filtering using review texts—a survey. *Information*, 11(6). DOI: 10.3390/info11060317.
- Vargas, S. and Castells, P. (2014). Improving sales diversity by recommending users to items. In *Proceedings of the 8th ACM Conference on Recommender Systems, RecSys '14*, page 145–152, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/2645710.2645744.
- Vasconcelos, N., Andrade, Y., Tuler, E., Pereira, A., and Rocha, L. (2025). Uma abordagem híbrida para sistemas de recomendação com base em avaliações textuais e filtragem colaborativa. In *Proceedings of the 31st Brazilian Symposium on Multimedia and the Web*, pages 455–463, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/webmedia.2025.14831.
- Vasconcelos, N., Reis, D., Silva, T., Silva, N., Cunha, W., Tuler, E., and Rocha, L. (2024). Integrando avaliações textuais de usuários em recomendação baseada em aprendizado por reforço. In *Proceedings of the 30th Brazilian Symposium on Multimedia and the Web*, pages 390–394, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/webmedia.2024.241405.
- Vinagre, J., Jorge, A. M., Rocha, C., and Gama, J. (2021). Statistically robust evaluation of stream-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering*, 33(7):2971–2982. DOI: 10.1109/TKDE.2019.2960216.
- Zhao, Y., Wang, Y., Liu, Y., Cheng, X., Aggarwal, C. C., and Derr, T. (2025). Fairness and diversity in recommender systems: A survey. *ACM Trans. Intell. Syst. Technol.*, 16(1). DOI: 10.1145/3664928.
- Zheng, L., Noroozi, V., and Yu, P. S. (2017). Joint deep modeling of users and items using reviews for recommendation. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM '17*, page 425–434, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3018661.3018665.
- Zhou, X., He, J., Huang, G., and Zhang, Y. (2015). Svd-based incremental approaches for recommender systems. *Journal of Computer and System Sciences*, 81(4):717–733. DOI: 10.1016/j.jcss.2014.11.016.