

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

Yan Andrade¹, Naan Vasconcelos², Elisa Tuler², Adriano Pereira¹, Leonardo Rocha²

¹Universidade Federal de Minas Gerais, ²Universidade Federal de São João del-Rei

yan.ribeiro@ddc.ufmg.br;naan.vasconcelos@aluno.ufsj.edu.br;etuler@ufsj.edu.br;adrianoc@dcc.ufmg.br;lcrocha@ufsj.edu.br

ABSTRACT

Recommender systems are increasingly leveraging textual reviews to enhance user preference modeling. Although review-aware recommender systems (RARs) have shown advances in capturing semantic nuances, they often fall short in representing collaborative patterns among users and items. In contrast, traditional collaborative filtering methods, though limited in semantic modeling, are robust in exploiting structural relationships. In this study, we propose a hybrid approach that integrates semantic representations derived from RARs into traditional collaborative models. We evaluated this strategy across three Amazon datasets, comparing RARs, collaborative, and hybrid models in terms of preference modeling accuracy (MSE and MAE), and recommendation relevance (Hits, Precision, Recall). Our results show that hybrid models outperform both baselines, achieving up to 2.8 times lower prediction error and 60% higher precision. These findings highlight the potential of hybrid models to deliver more robust and context-aware recommendations.

KEYWORDS

Review Aware, Reccomedation System, Colaborative Filtering

1 INTRODUÇÃO

Com o crescimento exponencial da informação disponível online, os Sistemas de Recomendação (SRs) consolidaram-se como ferramentas fundamentais para auxiliar na tomada de decisão e aprimorar a experiência dos usuários. Presentes em diversas esferas do cotidiano digital, como na sugestão de músicas [10], filmes [16] e conexões sociais [12], esses sistemas tornaram-se onipresentes em plataformas e serviços online. Seu funcionamento baseia-se na inferência dos interesses e preferências dos usuários, a partir de interações explícitas ou de dados implícitos extraídos de seus padrões de uso [22, 43]. A personalização resultante desse processo permite ajustar conteúdos e sugestões de forma mais precisa às necessidades individuais. No entanto, a qualidade dessas recomendações depende diretamente da capacidade do sistema de capturar, modelar e representar adequadamente as preferências do usuário [42, 43, 61].

Nos últimos anos, a comunidade científica tem voltado sua atenção para o aproveitamento de avaliações explícitas, frequentemente manifestadas por meio de comentários e resenhas textuais, na modelagem do perfil dos usuários e na geração de recomendações. Esses sistemas, conhecidos como *Review-Aware Recommender Systems* (RARs) [1, 49], partem da premissa de que as opiniões textuais dos usuários refletem de forma mais fidedigna suas percepções reais

sobre a relevância de um item [26, 34, 55]. Embora não constituam uma abordagem recente [21, 52], os RARs têm ganhado novo fôlego com os avanços em técnicas de Processamento de Linguagem Natural (PLN), especialmente aquelas baseadas em arquiteturas de redes neurais profundas aplicadas à classificação automática, análise de sentimentos e extração de aspectos semânticos dos textos [23, 40]. Diversas abordagens têm sido propostas nesse campo, explorando modelos baseados em mecanismos de atenção [14, 53], redes convolucionais [24, 66], grafos [32, 59], entre outros.

Apesar dos avanços significativos trazidos por RARs na modelagem semântica das preferências dos usuários, estudos recentes apontam limitações importantes desses modelos quanto à sua capacidade de capturar relações colaborativas entre usuários e itens [19, 58]. Por serem comumente derivados de arquiteturas originalmente concebidas para tarefas supervisionadas de classificação com dados independentes e identicamente distribuídos, esses modelos podem negligenciar padrões estruturais fundamentais do comportamento coletivo, como a co-ocorrência de consumo entre itens, similaridades latentes entre usuários e correlações item-usuário [39, 44]. RARs também têm apresentado falhas ao tentar capturar relações implícitas e estruturas latentes complexas, essenciais à efetividade da filtragem colaborativa [19, 58]. Arquiteturas neurais recentes, como AutoEnconders, que apresentam resultados excelentes em diversos cenários [27], tendem a apresentar uma efetividade limitada em contextos de alta esparsidade, no qual maioria absoluta dos itens é pouco avaliada/consumida (i.e. problema da cauda longa [35]). Além disso, os modelos resultantes dessas arquiteturas são determinísticos, uma vez que não realizam a modelagem probabilística do ruído e da incerteza, o que pode comprometer a robustez desses modelos frente a entradas incompletas ou ambíguas, reduzindo sua generalização em contextos variáveis de recomendação [60].

Por outro lado, sistemas de recomendação baseados em abordagens colaborativas tradicionais [2, 45, 47], apesar de não modelarem os perfis dos usuários tão bem quanto os RARs, demonstram maior robustez na captura de relações latentes entre usuários e itens, especialmente no que se refere à exploração de padrões coletivos de consumo. Essas técnicas tendem a ser mais simples em termos computacionais e oferecem maior transparência na interpretação dos fatores que influenciam as recomendações, facilitando tanto a análise dos resultados quanto a explicação para os usuários [34].

Neste trabalho, investigamos o potencial de uma abordagem híbrida para sistemas de recomendação, que combina RARS com métodos colaborativos tradicionais. Utilizamos estratégias de RARs baseadas em aprendizado profundo para extrair representações semânticas de usuários e itens a partir de comentários textuais, visando capturar nuances contextuais e emocionais não refletidas nas notas (ratings) atribuídas. Em seguida, essas representações são

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2025). Rio de Janeiro, Brazil. Porto Alegre: Brazilian Computer Society, 2025.
© 2025 SBC – Brazilian Computing Society.

ISSN 2966-2753

incorporadas como entrada em modelos tradicionais de recomendação, buscando preservar as propriedades colaborativas durante a etapa de recomendação propriamente dita. Ao integrar essas duas camadas — semântica e colaborativa —, analisamos se a combinação entre modelagens textual e estrutural resulta em recomendações mais eficazes, robustas e interpretáveis. Para conduzir esta investigação, formulamos três perguntas de pesquisa que orientam a análise comparativa e fundamentam as escolhas metodológicas:

- **PP1:** Métodos baseados em informações textuais são capazes de capturar as preferências dos usuários de forma mais precisa que métodos colaborativos?
- **PP2:** Métodos baseados em informações textuais geram recomendações mais relevantes e precisas do que métodos colaborativos?
- **PP3:** Qual é o potencial da combinação entre métodos colaborativos e informações textuais?

Para responder a essas perguntas, conduzimos experimentos com três conjuntos de dados referentes à plataforma Amazon, amplamente utilizados na literatura: *Musical Instruments*, *Digital Music* e *Office Products*[5, 55], relacionados a diferentes domínios e volumes de interações. Como métodos baseados em informações textuais, consideramos os modelos RARs CARM[29], CARP [28] e HRDR [31], os quais apresentaram os melhores resultados em um benchmark recente [6]. Como modelos de filtragem colaborativa, adotamos métodos tradicionais associados às estratégias mais clássicas da literatura: SVD [57], NNMF [36], UserKNN [18] e ItemKNN [54].

Para responder à **PP1**, realizamos uma comparação entre estratégias baseadas em avaliações textuais (RARs) e colaborativas tradicionais sob a perspectiva da modelagem das preferências dos usuários. Para isso, utilizamos métricas de erro contínuo (MSE e MAE), que medem o quanto bons são os modelos na predição das avaliações dos usuários. Quanto menores os erros, mais fiel é a representação das preferências. Os resultados obtidos indicam que os modelos baseados em avaliações textuais apresentam desempenho superior na modelagem das preferências dos usuários, evidenciado por menores erros de predição em todos os conjuntos de dados analisados, respondendo positivamente à primeira pergunta de pesquisa.

Na segunda pergunta (**PP2**), o foco está em avaliar a capacidade de cada conjunto de modelos em gerar listas de recomendação úteis. Novamente, comparamos os modelos RARs aos tradicionais, mas agora avaliando com métricas de precisão (i.e., *Hits*, *Precision* e *Recall*), que capturam a relevância dos itens sugeridos para os usuários. Os resultados revelaram um contraste importante: apesar de representarem mais adequadamente as preferências dos usuários, os modelos RARs não conseguiram superar a precisão das recomendações dos modelos colaborativos, os quais mantiveram desempenho superior ou estatisticamente equivalente em todos os cenários analisados, respondendo negativamente à segunda pergunta de pesquisa.

Por fim, para responder à **PP3**, avaliamos nossa proposta híbrida, na qual a representação do usuário extraída das avaliações textuais pelos modelos RARs é incorporada a algoritmos colaborativos, sob duas perspectivas: modelagem das preferências dos usuários (métricas de erro) e utilidade das predições (métricas de precisão). Comparamos nossa proposta com ambos os conjuntos de abordagens: RARs e colaborativas tradicionais. **Os resultados mostram que a abordagem híbrida supera consistentemente as demais**,

com erros quase 3x menores e precisão até 60% maiores. A fusão dos RARs com modelos colaborativos demonstrou-se eficaz em melhorar a qualidade das recomendações e refinar a modelagem das preferências, com ganhos observados mesmo quando comparados com modelos colaborativos que já apresentavam bom desempenho isoladamente. Esses achados respondem à PP3, evidenciando o potencial das abordagens híbridas como um caminho promissor para sistemas de recomendação mais robustos e sensíveis ao contexto.

Para complementar a análise, investigamos também a diversidade das recomendações geradas, uma dimensão frequentemente negligenciada, mas essencial para garantir variedade e promover a descoberta de novos itens pelos usuários [48]. Os resultados indicam que, apesar dos ganhos nas métricas de erro e precisão, houve uma redução na diversidade, evidenciando o *trade-off* inerente entre predição e diversidade. Considerando que os modelos híbridos já alcançam alto desempenho preditivo, surge uma oportunidade concreta para explorar, em trabalhos futuros, estratégias que ampliem a diversidade sem comprometer a precisão das recomendações.

O restante desse artigo está organizado da seguinte forma. Na Seção 2 apresentamos as principais abordagens de recomendação tradicionais colaborativas e os principais modelos RARs. Na Seção 3 detalhamos nossa proposta de abordagem híbrida com combinação dos pontos fortes de métodos baseados em avaliações textuais com a robustez estrutural de algoritmos colaborativos. Na Seção 4 apresentamos o ambiente experimental utilizado para avaliar as abordagens. Na Seção 5 discutimos os resultados obtidos de acordo com as perguntas de pesquisa. Concluímos nosso trabalho na Seção 6.

2 TRABALHOS RELACIONADOS

Sistemas de Recomendação (SsR) podem ser categorizados com base na natureza das informações utilizadas para inferir as preferências dos usuários. Enquanto abordagens tradicionais operam, em geral, sobre registros históricos de interações entre usuários e itens, métodos mais recentes incorporam dados complementares, como conteúdo textual, metadados e sinais contextuais, com o objetivo de capturar aspectos adicionais do comportamento do usuário. Esta seção revisa duas vertentes relevantes para este trabalho: os métodos colaborativos tradicionais e os *Review-Aware Recommender Systems*.

2.1 Métodos Colaborativos Tradicionais

Métodos colaborativos tradicionais constituem a base clássica dos SsR e têm sido amplamente adotados em aplicações do mundo real [2, 47]. Esses métodos inferem preferências futuras a partir de padrões observados nas interações passadas entre usuários e itens, sem depender de informações de conteúdo. A hipótese subjacente é que usuários que avaliaram itens de maneira semelhante tendem a compartilhar gostos e comportamentos similares [45].

Entre as abordagens mais representativas estão os modelos baseados em fatoração de matrizes [25], como o *Singular Value Decomposition* (SVD)[67] e o *Non-negative Matrix Factorization* (NNMF)[36]. Esses métodos identificam padrões comuns em grandes volumes de dados ao projetar usuários e itens em um espaço de características latentes. Nessa representação, usuários e itens são descritos por vetores de baixa dimensionalidade, que capturam aspectos implícitos como estilo, qualidade ou popularidade, mesmo que esses atributos não estejam explicitamente rotulados. O SVD decompõe a matriz de

interações em componentes lineares, permitindo prever avaliações futuras com base na proximidade entre representações no espaço latente. Essa abordagem apresenta boa capacidade preditiva e desempenho robusto, mesmo em cenários com alta esparsidade. O NNMF impõe a restrição de não negatividade nos vetores gerados, favorecendo uma interpretação mais direta das dimensões latentes, pois cada fator representa contribuições positivas em diferentes aspectos de preferência. Essa característica torna o NNMF especialmente útil em aplicações que exigem maior transparência ou explicabilidade.

Além dos modelos de fatoração, outra classe amplamente empregada é a de métodos baseados em vizinhança [18]. Esses métodos estimam a preferência de um usuário com base nas avaliações de usuários semelhantes (**UserKNN**) ou em itens similares ao item de interesse (**ItemKNN**) [54]. A similaridade costuma ser calculada por métricas como correlação de Pearson ou similaridade de cosseno, aplicadas aos vetores de interação. O UserKNN parte do princípio de que usuários tendem a gostar de itens bem avaliados por perfis semelhantes, enquanto o ItemKNN assume que usuários apreciarão itens similares aos que avaliaram positivamente. Por se basearem em relações diretas de similaridade, essas abordagens são intuitivas, simples de implementar e apresentam boa interpretabilidade, sobretudo em sistemas com dados de interação relativamente densos.

As duas abordagens apresentam vantagens significativas em termos de eficiência computacional, escalabilidade e simplicidade de implementação. No entanto, por se basearem exclusivamente em padrões observados de interação, são limitadas na captura de nuances subjetivas frequentes em informações textuais.

2.2 Sistemas de Recomendação Review-Aware

Em contraste, os *Review-Aware Recommender Systems* (RARs) exploram o conteúdo textual das interações dos usuários para enriquecer a representação de suas preferências [1, 49]. A motivação central é que comentários e opiniões contêm informações complementares – ou mesmo substitutas – aos *ratings* explícitos, ajudando a mitigar problemas clássicos como esparsidade e *cold start*. Um trabalho recente [6] propôs uma taxonomia abrangente para os RARs, estruturada em quatro perspectivas: modelagem dos dados, extração de informação, objetivo do modelo e arquitetura de aprendizado.

A modelagem dos dados descreve como os textos são processados e representados, com três estratégias principais: *Document Modeling*, que concatena todos os comentários de um usuário ou item em um único documento (ex.: DeepCoNN [66], ANR [15]); *Sentence Modeling*, que trata sentenças individualmente com atenção para identificar palavras-chave (ex.: NARRE [11], TransNet [9]); e *Rating Aggregation*, que associa diretamente avaliações numéricas aos textos (ex.: ConvMF [24], TARMF [33]). O **HRDR** [31] combina *ratings* com modelagem textual via codificadores paralelos, oferecendo uma solução robusta para lidar com esparsidade e ruído textual.

A extração de informação abrange técnicas para capturar preferências e características a partir dos textos. Essa perspectiva se divide em: *Sentiment Extraction*, que aplica análise de sentimentos (ex.: AGTR [63], NRPA [62]); e *Aspect Extraction*, que identifica aspectos latentes, com três subtipos: *Global* (ex.: HFT [37], baseado em modelagem de tópicos); *Local*, com foco em preferências individuais (ex.: A3NCF [13], HRDR [31]); e *Pairwise*, que modela interações entre usuários e itens por meio de grafos (ex.: SENG [46], CARP [28]).

O modelo **CARP** [28] se destaca por vetorizar os comentários de usuários e itens os quais são processados por camadas CNN para capturar o contexto e extrair características importantes.

A perspectiva do objetivo do modelo classifica os RARs quanto ao uso das informações extraídas. Modelos *Predictive* preveem avaliações ou rankings diretamente (ex.: RBLT [51], DeepCoNN [66]); já os *Hybrid* combinam predição e explicações, adotando estratégias como: *Item Description* (ex.: ESCOFILT [41]); *User Description*, baseada em traços de personalidade (ex.: JRL [64]); e *Opinion Generation*, que utiliza modelos gerativos para fornecer justificativas personalizadas (ex.: U-ARM [50], ExpansionNet [38], CARP [28]).

A arquitetura de aprendizado diz respeito às técnicas de modelagem utilizadas. Modelos *Non-Neural* adotam abordagens clássicas, como modelagem de tópicos e regressão (ex.: HFT [37], SUML [4]). Já os *Neural* empregam arquiteturas profundas, como: *Redes Convolucionais* (DeepCoNN [66], ConvMF [24], HRDR [31]); *Mecanismos de Atenção* (NARRE [11], DAML [30], CARM [29]); *Redes Recorrentes* (DRR [17]); e *Redes Neurais em Grafos* (RI-GCN [7], SSG [20]). O modelo **CARM** [29] se destaca por integrar aspectos semânticos à filtragem colaborativa, usando um mecanismo de atenção para alinhar fatores latentes aos contextos extraídos dos textos.

Embora os RARs ofereçam avanços significativos na modelagem semântica das preferências dos usuários, como evidenciado pela sua capacidade de capturar nuances textuais e semânticas, esses modelos ainda enfrentam limitações na generalização de padrões colaborativos complexos. Isso se deve, em parte, ao serem derivados de arquiteturas supervisionadas, geralmente focadas em tarefas de classificação com dados independentes e identicamente distribuídos, os RARs tendem a negligenciar relações colaborativas fundamentais, como a co-ocorrência de consumo entre itens e as similaridades latentes entre usuários [19, 58]. Esse foco supervisionado, comum em tarefas de PLN, tende a negligenciar estruturas relacionais entre usuários e itens. [39, 44]. Para superar essas limitações, propomos uma abordagem híbrida que combina a riqueza semântica dos textos com a robustez estrutural dos métodos colaborativos, explorando a complementaridade entre ambos.

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 das relações latentes entre usuários e itens. A estratégia consiste em duas partes principais: (i) a geração de *ratings* preditivos a partir de comentários textuais por meio de RARs; e (ii) o uso desses *ratings* como entrada dos algoritmos colaborativos, substituindo os *ratings* explicitamente assinalados. Ilustramos nossa proposta na Figura 1, detalhando-a a seguir.

A primeira parte dessa abordagem emprega modelo RARs, como CARP, para gerar estimativas contínuas de interesse do usuário por um item. Cada instância de entrada é composta por uma quadrupla (u, i, r_{ui}, t_{ui}) , que representa a interação entre o usuário u e o item i , acompanhada do *rating* atribuído r_{ui} e da revisão textual t_{ui} correspondente, passo 1 da Figura 1. O modelo é treinado para aprender representações densas dos usuários e itens a partir das avaliações textuais, utilizando arquiteturas de codificação semântica. O objetivo é prever um valor escalar \hat{r}_{ui} , que represente o grau

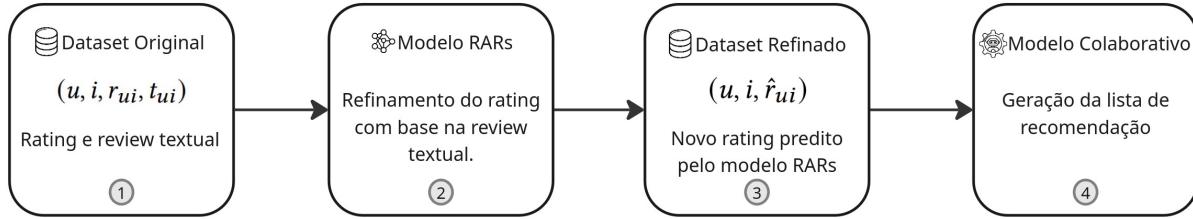


Figura 1: Fluxo da abordagem híbrida proposta entre RARs e filtragem colaborativa

de interesse estimado com base nas nuances semânticas da revisão, ponderando e refinando o *rating* original com base em conteúdo textual, ilustrado no passo 2 da Figura 1.

Por exemplo, 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 máxima, o conteúdo textual revela uma insatisfação significativa com aspectos centrais da obra. Ao processar essa entrada, o modelo RARs infere o real nível de satisfação do usuário, gerando um *rating* preditivo $\hat{r}_{ui} = 3.2$. Esse valor mais baixo reflete melhor as nuances negativas expressas na revisão, que por sua vez não foram capturadas pelo *rating* numérico original. Após processar todas as interações do conjunto de dados, o resultado é uma nova coleção (u, i, \hat{r}_{ui}) no qual cada interação usuário–item contém um *rating* sintético obtido à partir da análise semântica das revisões, como ilustrado no passo 3 da Figura 1.

Na segunda parte de nossa abordagem, essa coleção modificada de *ratings* refinados é utilizada como entrada para métodos de recomendação colaborativa, como fatoração matricial ou métodos baseados em vizinhança, conforme passo 4 na Figura 1. Diferentemente dos cenários tradicionais, em que se parte de interações históricas reais, aqui os métodos colaborativos operam sobre os *ratings* preditivos derivados dos RARs. A motivação por trás dessa estratégia é investigar se a estrutura latente dos modelos colaborativos pode ser preservada e explorada mesmo quando os dados originais são substituídos por sinais semanticamente derivados.

Essa estruturação modular de nossa abordagem permite isolar e avaliar a contribuição das representações semânticas geradas por cada RARs em um ambiente colaborativo clássico. Além disso, facilita a comparação direta entre modelos e a análise de complementaridades, uma vez que os dados utilizados em ambas as etapas são derivados de uma mesma fonte textual, conforme veremos a seguir.

4 AMBIENTE EXPERIMENTAL

Com o intuito de avaliar nossa proposta, bem como responder às três perguntas de pesquisa deste trabalho, nesta seção detalhamos o ambiente experimental considerado em nossos experimentos.

4.1 Bases de Dados

Consideramos três coleções de dados amplamente utilizadas na avaliação de Sistemas de Recomendação (SsR) [5, 55], oriundas da plataforma Amazon: *Musical Instruments*, *Digital Music* e *Office Products*. Esses conjuntos foram selecionados por apresentarem características complementares, como domínios distintos, diferentes densidades de interação e variações na extensão média dos comentários. Cada instância inclui tanto uma nota numérica (*rating*) quanto

sua respectiva avaliação textual, tornando essas bases adequadas para a avaliação de modelos tradicionais e RARs. A Tabela 1 apresenta detalhes sobre essas coleções.

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%

Tabela 1: Visão geral das coleções utilizadas na avaliação.

4.2 Métricas e Validação Estatística

A avaliação dos modelos foi conduzida com base em dois grupos de métricas. O primeiro grupo, voltado à análise da modelagem das preferências, compreende métricas de erro contínuo [5]: MAE (*Mean Absolute Error*) e MSE (*Mean Squared Error*). Tais métricas quantificam a discrepância entre as notas previstas pelos modelos e as avaliações reais atribuídas pelos usuários, avaliando a capacidade dos modelos de inferir as preferências dos usuários. O segundo grupo avalia a utilidade das recomendações por meio de métricas de precisão [3]: *Hits*, *Precision* e *Recall*, que capturam a relevância dos itens sugeridos para os usuários. Avaliamos listas de tamanho 10, 50 e 100 itens recomendados. Para assegurar a robustez das comparações, aplicamos o teste estatístico não paramétrico de Wilcoxon [56], com nível de significância de 5%, a fim de verificar se as diferenças observadas entre os modelos são estatisticamente significativas.

4.3 Modelos de Recomendação

Os modelos tradicionais escolhidos para execução neste trabalho foram o **SVD**, **NNMF**, **UserKNN** e **ItemKNN**, todos descritos na Seção 2.1. A escolha desses modelos se deu por sua simplicidade e baixo custo computacional. Os modelos RARs selecionados foram o **CARM**, **CARP** e **HRDR**, detalhados na Seção 2.2. A escolha desses modelos se deve à diversidade metodológica, à representatividade nas diferentes categorias da taxonomia e ao desempenho destacado nos estudos comparativos da literatura [5, 55].

A implementação da abordagem híbrida no ambiente experimental seguiu a estrutura descrita na Seção 3. Na primeira etapa, os modelos RARs foram utilizados para gerar previsões de preferência com base nas avaliações textuais dos usuários. Essas previsões foram organizadas em uma nova matriz de interações, que serviu como entrada para a segunda etapa. Nessa fase, os algoritmos colaborativos tradicionais operaram sobre a matriz preditiva derivada semanticamente, em vez das interações históricas originais. Considerando que cada modelo RAR foi combinado com um modelo tradicional, avaliamos um total de 12 novas estratégias híbridas.

Comparamos os modelos híbridos implementados com seus respectivos modelos de base. Essa configuração experimental permitiu analisar de forma isolada o impacto das representações textuais no desempenho dos modelos colaborativos, possibilitando uma avaliação sistemática da efetividade da combinação entre semântica e estrutura colaborativa, conforme detalharemos na seção a seguir.

5 RESULTADOS E DISCUSSÕES

Esta seção apresenta os resultados obtidos a partir dos experimentos conduzidos no ambiente descrito na Seção 4. Os resultados das avaliações experimentais encontram-se reportados na Tabela 2. As análises são apresentadas nas subseções a seguir, organizadas de acordo com nossas três perguntas de pesquisa.

5.1 PP1: Preferências com Informação Textual

Para responder à **PP1**, compararamos o desempenho dos modelos RARs (CARM, CARP e HRDR) com os métodos colaborativos tradicionais (SVD, NNMF, UserKNN e ItemKNN), com foco em sua capacidade de modelar as preferências dos usuários. Para isso, consideramos as métricas de erro, cujos resultados são apresentados no **Bloco 1** da Tabela 2, para cada uma das coleções avaliadas.

De forma geral, observamos uma superioridade consistente dos modelos RARs em todos os conjuntos de dados. O modelo CARP destacou-se ao apresentar os menores valores de erro nas bases *Musical Instruments* (MAE de 0,3294 e MSE de 0,2296) e *Digital Music* (MAE de 0,5606 e MSE de 0,5657), enquanto o CARM e o HRDR obtiveram os melhores desempenhos, respectivamente em MAE (0,4251) e MSE (0,3598), na base *Office Products*. Esses resultados demonstram que os modelos RARs são mais eficazes na captura das nuances subjetivas e contextuais expressas nas avaliações textuais, resultando em uma modelagem mais precisa das preferências dos usuários. A confirmação da superioridade dos RARs na modelagem das preferências individuais evidencia o valor das informações textuais como fonte rica e semântica, capaz de captar aspectos subjetivos e contextuais que os métodos tradicionais, centrados apenas em *ratings* numéricos, não conseguem alcançar.

Dessa forma, é possível afirmar que a PP1 foi respondida positivamente. Os métodos que exploram informações textuais são capazes de capturar de forma mais precisa as preferências individuais dos usuários em comparação aos métodos colaborativos. Embora essa capacidade descritiva seja essencial para compreender o comportamento do usuário, ela precisa ser traduzida em recomendações úteis, conforme avaliamos a seguir.

5.2 PP2: Utilidade com Informação Textual

A próxima análise visa responder à **PP2**, investigando se a modelagem mais precisa das preferências dos usuários, observada nos modelos RARs, resulta em listas de recomendação mais relevantes e úteis. Para isso, novamente comparamos os modelos RARs (i.e., CARM, CARP e HRDR) com os métodos colaborativos (i.e., SVD, NNMF, UserKNN e ItemKNN), agora considerando as métricas de precisão, cujos resultados são apresentados no **Bloco 2** da Tabela 2.

Observa-se que os modelos colaborativos tradicionais apresentaram desempenho superior aos RARs em todos os cenários avaliados (com poucas exceções). O SVD foi consistentemente o modelo mais eficaz nas bases *Musical Instruments* (e.g., HIT10 de 0,0252) e *Digital Music* (e.g., HIT50 de 0,0507), superando os demais em todas as

métricas de precisão. Já na base *Office Products*, o destaque foi o modelo ItemKNN, que obteve os melhores resultados em todas as métricas (e.g., HIT100 de 0,1050). Embora o HRDR tenha demonstrado competitividade pontual, frequentemente empatando com modelos como NNMF e UserKNN, ele não apresentou superioridade estatisticamente significativa. Essa limitação dos RARs evidencia uma dissociação entre a qualidade da modelagem das preferências e a efetividade prática das recomendações.

Assim, a PP2 é respondida negativamente. Embora os métodos que exploram informações textuais capturem nuances semânticas mais refinadas, eles não exploram de forma eficiente os padrões colaborativos subjacentes ao comportamento coletivo dos usuários, reforçando a necessidade de estratégias híbridas, como as propostas neste trabalho e avaliadas a seguir.

5.3 PP3: Potencialidades da Abordagem Híbrida

Nesta seção, apresentamos as análises que visam responder à terceira pergunta de pesquisa, **PP3**. Nossa objetivo é investigar se a integração entre informações textuais e aprendizado colaborativo pode superar as limitações observadas isoladamente em cada conjunto de modelos. Conforme detalhado na Seção 3, construímos abordagens híbridas nas quais a modelagem semântica dos usuários, extraída por modelos RARs, é incorporada como entrada para os algoritmos colaborativos tradicionais. Em nossa análise, compararamos os 12 modelos híbridos gerados a partir da combinação dos três modelos RARs (CARM, CARP e HRDR) com os quatro métodos colaborativos (SVD, NNMF, UserKNN e ItemKNN) e suas respectivas versões base. Avaliamos tanto a capacidade desses métodos híbridos em modelar preferências (métricas de erro), quanto a utilidade das recomendações (métricas de precisão). Os resultados são apresentados, respectivamente, nos **Blocos 3 e 4** da Tabela 2.

Observando nos resultados que todos os modelos híbridos apresentaram desempenho consistente, alcançando os melhores resultados tanto sob a perspectiva da modelagem de preferências quanto da utilidade das recomendações. Na coleção *Musical Instruments*, destaca-se o modelo híbrido NNMF_{CARM}, que passou de pior desempenho em termos de erro (no Bloco 1), com MAE de 0,7163, para o melhor entre todos no **Bloco 3**, com MAE de 0,2496 – ou seja, **quase 3x menor**. Trata-se de um indicativo claro do ganho expressivo obtido com a fusão de informações textuais com estratégias colaborativas. Esse padrão de melhora repete-se de forma consistente em outras combinações, com exceções pontuais como o ItemKNN_{CARP} no MSE e os modelos UserKNN_{CARP} e UserKNN_{HRDR} em MAE e MSE. Ainda assim, a tendência geral é positiva. Analisando as métricas de precisão no **Bloco 4**, observa-se uma melhora generalizada com a abordagem híbrida. Todos os modelos apresentaram ganhos, incluindo o SVD, que já era superior em sua versão base e ainda assim conseguiu aprimorar seu desempenho.

Na base *Digital Music*, esse comportamento se repete. Novamente, o NNMF_{CARM} se destaca nas métricas de erro, e os modelos híbridos, de maneira geral, superam suas versões base, tanto em MAE e MSE quanto em *Hits*, *Precision* e *Recall*. Um destaque é o NNMF_{CARP}, que obteve um **Hit 60% superior ao da versão base** (HIT10 de 0,0084 passou para 0,0136). Os resultados obtidos na coleção *Office Products* também confirmam a robustez das estratégias híbridas, com melhorias sistemáticas em ambos os conjuntos de métricas, reforçando sua consistência em diferentes contextos.

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	0,3291▲	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▲			
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▲			
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▲</												

Coleção	Musical Instruments			Digital Music			Office Products		
	Diversidade (ILD)			Diversidade (ILD)			Diversidade (ILD)		
Métrica	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*	0,1425▼	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*

Tabela 3: Avaliação dos modelos híbridos frente aos seus respectivos pares, em relação a diversidade. Resultados são avaliados com teste de Wilcoxon ($p = 0,05$). Símbolos: ▲ (ganho significativo), * (empate estatístico), ▼ (perda significativa).

Em suma, as abordagens híbridas mostraram-se capazes de aproveitar o melhor de ambos os mundos — a expressividade semântica das avaliações textuais e a robustez das relações colaborativas —, produzindo ganhos consistentes tanto na modelagem das preferências quanto na geração de recomendações. Os ganhos obtidos em todas as coleções e em ambos os grupos de métricas respondem à PP3 de forma positiva, indicando que a combinação entre métodos colaborativos e informações textuais resulta em sistemas de recomendação mais robustos, eficazes, adaptáveis e sensíveis ao contexto.

5.4 Limitações

Embora os resultados apresentados nas seções anteriores demonstrem avanços significativos na modelagem das preferências e na precisão das recomendações obtidas por meio da estratégia híbrida proposta, é importante destacar que a avaliação experimental concentrou-se em duas dimensões tradicionais: erro e precisão. No entanto, outras dimensões também são fundamentais para avaliar a efetividade de um sistema de recomendação, como a diversidade. Nesta seção, exploramos a diversidade como dimensão adicional para analisar criticamente o desempenho da abordagem híbrida. Mais especificamente, consideramos a métrica ILD [65], que quantifica o grau de dissimilaridade entre os itens recomendados. Os resultados dessa análise estão apresentados na Tabela 3.

De forma geral, nota-se que as abordagens híbridas apresentaram desempenho inferior ou empates estatísticos com seus respectivos modelos base em relação à diversidade das recomendações. Apesar dos valores absolutos serem bastante próximos, a maioria das combinações híbridas apresentou diversidade estatisticamente inferior — especialmente os modelos baseados em vizinhança (ItemKNN e UserKNN), que já tendem a gerar listas menos diversas. Essa tendência é consistente nas três coleções avaliadas. Em *Musical Instruments*, os modelos híbridos NNMF_{CARM}, NNMF_{CARP} e NNMF_{HRDR} empataram ou perderam diversidade em relação aos seus modelos base. Situação semelhante ocorre nas coleções *Digital Music* e *Office Products*, onde, embora alguns empates e ganhos pontuais sejam observados — como NNMF_{HRDR} em *Office Products* —, o padrão dominante é de desempenho igual ou inferior.

Esses resultados sugerem que, embora a estratégia híbrida proposta seja eficaz na modelagem das preferências e na precisão das recomendações, ela tende a privilegiar padrões dominantes aprendidos a partir das representações textuais, reduzindo a variabilidade dos itens recomendados. Isso indica um possível efeito de reforço de popularidade ou de compressão semântica em torno de tópicos recorrentes nas avaliações, ao ancorar o espaço latente em sinais textuais frequentes, o módulo de representação tende a aproximar itens lexicalmente semelhantes e a atenuar nuances menos prevalentes, o que reduz artificialmente as distâncias entre itens. Em seguida, o componente colaborativo opera sobre esse espaço já concentrado, amplificando coocorrências em torno de itens dominantes e, na prática, penalizando itens de cauda longa cujo sinal semântico é mais esparsos. O resultado combinado é uma menor dispersão no conjunto recomendado e, portanto, queda de ILD mesmo quando a precisão de ponto aumenta. Portanto, a análise de diversidade evidencia um importante *trade-off* entre precisão e diversidade. Os ganhos obtidos em precisão indicam que há espaço para explorar mecanismos que promovam maior diversidade nas recomendações, sem comprometer significativamente a acurácia. Isso reforça a importância de investigar e incorporar estratégias de controle de diversidade, como discutiremos na próxima seção.

6 CONCLUSÃO E TRABALHOS FUTUROS

Neste trabalho, investigamos o potencial de abordagens híbridas em sistemas de recomendação, combinando os pontos fortes de métodos baseados em avaliações textuais (*Review-Aware Recommender Systems - RARs*) com a robustez estrutural de algoritmos colaborativos tradicionais. Avaliamos, de forma sistemática, o desempenho de diferentes estratégias na modelagem das preferências dos usuários e na geração de recomendações úteis. Os resultados obtidos indicam que os modelos RARs são superiores na modelagem das preferências individuais dos usuários. Essa superioridade foi evidenciada por erros significativamente menores nas tarefas de predição, o que reforça o valor semântico contido nas avaliações textuais como fonte rica para capturar nuances subjetivas e contextuais. No entanto, ao serem avaliados quanto à utilidade prática das recomendações, os

modelos baseados em texto apresentaram desempenho inferior às abordagens colaborativas tradicionais, demonstrando limitações na exploração dos padrões de consumo coletivo.

A proposta híbrida, que integra representações textuais geradas por modelos RARs como entrada para algoritmos colaborativos, demonstrou-se eficaz ao combinar as vantagens de ambas as abordagens. Os modelos híbridos superaram sistematicamente seus equivalentes tradicionais e textuais, com erros até três vezes menores e melhorias expressivas nas métricas de precisão, em alguns casos superando em mais de 60% os resultados das abordagens isoladas. Esses achados evidenciam o potencial das abordagens híbridas na construção de sistemas de recomendação mais eficazes, sensíveis ao contexto e adaptáveis a diferentes domínios.

Apesar dos avanços observados, a análise adicional de diversidade revelou uma limitação relevante: os modelos híbridos tendem a reduzir a variabilidade das recomendações, possivelmente em decorrência do reforço de padrões semânticos dominantes. Esse *trade-off* entre precisão e diversidade sugere a necessidade de futuras investigações voltadas à incorporação de mecanismos de controle de diversidade, tais como re-rankeamento [8], para garantir recomendações não apenas precisas, mas também diversas e exploratórias. Além disso, pretendemos expandir a aplicação da abordagem híbrida proposta para outros paradigmas de sistemas de recomendação, com destaque para modelos baseados em *Multi-Armed Bandits* (MAB) [3, 55], que são particularmente relevantes em cenários de recomendação online e em tempo real. Investigaremos como incorporar representações semânticas derivadas de avaliações textuais em algoritmos MAB, de modo a equilibrar o dilema *Exploration-exploitation* com maior sensibilidade às preferências contextuais dos usuários.

AGRADECIMENTOS:

Financiado por CNPq, CAPES, INCT-TILD-IAR, Fapemig e AWS.

REFERÊNCIAS

- [1] Sumaia Mohammed Al-Ghuribi and Shahrul Azman Mohd Noah. 2019. Multi-criteria review-based recommender system—the state of the art. *IEEE Access* 7 (2019), 169446–169468.
- [2] Xavier Amatriain and Justin Basilico. 2015. *Recommender Systems in Industry: A Netflix Case Study*. 385–419. doi:10.1007/978-1-4899-7637-6_11
- [3] Yan Andrade, Nicollas Silva, Thiago Silva, Adriano Pereira, Diego Dias, Elisa T. Albergaria, and Leonardo Rocha. 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* (Ribeirão Preto, Brazil) (*WebMedia* '23). Association for Computing Machinery, New York, NY, USA, 193–197. doi:10.1145/3617023.3617049
- [4] Konstantin Bauman, Bing Liu, and Alexander Tuzhilin. 2017. Aspect Based Recommendations: Recommending Items with the Most Valuable Aspects Based on User Reviews. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Halifax, NS, Canada) (KDD '17). Association for Computing Machinery, New York, NY, USA, 717–725. doi:10.1145/3097983.3098170
- [5] Guilherme Bittencourt, Guilherme Fonseca, Yan Andrade, Nicollas Silva, and Leonardo Rocha. 2023. A Survey on Review - Aware Recommendation Systems. In *Proceedings of the 29th Brazilian Symposium on Multimedia and the Web* (Ribeirão Preto, Brazil) (*WebMedia* '23). Association for Computing Machinery, New York, NY, USA, 198–207. doi:10.1145/3617023.3617050
- [6] Guilherme Bittencourt, Naan Vasconcelos, Yan Andrade, Nicollas Silva, Washington Cunha, Diego Roberto Colombo Dias, Marcos André Gonçalves, and Leonardo Rocha. 2025. Review-Aware Recommender Systems (RARSs): Recent Advances, Experimental Comparative Analysis, Discussions, and New Directions. *ACM Comput. Surv.* (June 2025). doi:10.1145/3744661 Just Accepted.
- [7] Yijin Cai, Yilei Wang, Weijin Wang, and Wenting Chen. 2022. RI-GCN: Review-aware Interactive Graph Convolutional Network for Review-based Item Recommendation. In *2022 IEEE International Conference on Big Data (Big Data)*. 475–484. doi:10.1109/BigData55660.2022.10020296
- [8] Diego Carraro and Derek Bridge. 2024. Enhancing Recommendation Diversity by Re-ranking with Large Language Models. *ACM Trans. Recomm. Syst.* (Oct. 2024). doi:10.1145/3700604
- [9] Rose Catherine and William Cohen. 2017. TransNets: Learning to Transform for Recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (Como, Italy) (RecSys '17). Association for Computing Machinery, New York, NY, USA, 288–296. doi:10.1145/3109859.3109878
- [10] Oscar Celma and Paul Lamere. 2011. Music recommendation and discovery revisited. In *Proceedings of the fifth ACM conference on Recommender systems*. 7–8.
- [11] Chong Chen, Min Zhang, Yiqun Liu, and Shaoping Ma. 2018. Neural attentional rating regression with review-level explanations. In *Proceedings of the 2018 world wide web conference*. 1583–1592.
- [12] Chien Chin Chen, Shun-Yuan Shih, and Meng Lee. 2016. Who should you follow? Combining learning to rank with social influence for informative friend recommendation. *Decision Support Systems* 90 (2016), 33–45.
- [13] Zhiyong Cheng, Ying Ding, Xiangnan He, Lei Zhu, Xuemeng Song, and Mohan Kankanhalli. 2018. A3NCF: an adaptive aspect attention model for rating prediction. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence* (Stockholm, Sweden) (IJCAI'18). AAAI Press, 3748–3754.
- [14] Zhiyong Cheng, Ying Ding, Lei Zhu, and Mohan Kankanhalli. 2018. Aspect-Aware Latent Factor Model: Rating Prediction with Ratings and Reviews. In *Proceedings of the 2018 World Wide Web Conference* (Lyon, France) (WWW '18). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 639–648. doi:10.1145/3178876.3186145
- [15] Jin Yao Chin, Kaiqi Zhao, Shafiq Joty, and Gao Cong. 2018. ANR: Aspect-based Neural Recommender. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (Torino, Italy) (CIKM '18). Association for Computing Machinery, New York, NY, USA, 147–156. doi:10.1145/3269206.3271810
- [16] Dan Cosley, Shyong K Lam, Istvan Albert, Joseph A Konstan, and John Riedl. 2003. Is seeing believing? How recommender system interfaces affect users' opinions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 585–592.
- [17] Kostadin Cvejoski, Ramsés J. Sánchez, Christian Bauckhage, and César Ojeda. 2022. Dynamic Review-based Recommenders. In *Data Science – Analytics and Applications*, Peter Haber, Thomas J. Lampoltshammer, Helmut Leopold, and Manfred Mayr (Eds.). Springer Fachmedien Wiesbaden, Wiesbaden, 66–71.
- [18] Christian Desrosiers and George Karypis. 2011. *A Comprehensive Survey of Neighborhood-based Recommendation Methods*. Springer US, Boston, MA, 107–144. doi:10.1007/978-0-387-85820-3_4
- [19] Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. 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* (Copenhagen, Denmark) (RecSys '19). Association for Computing Machinery, New York, NY, USA, 101–109. doi:10.1145/3298689.3347058
- [20] Jingyue Gao, Yang Lin, Yasha Wang, Xiting Wang, Zhao Yang, Yuanduo He, and Xu Chu. 2020. Set-Sequence-Graph: A Multi-View Approach Towards Exploiting Reviews for Recommendation. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (Virtual Event, Ireland) (CIKM '20). Association for Computing Machinery, New York, NY, USA, 395–404. doi:10.1145/3340531.3411939
- [21] Negar Hariri, Bamshad Mobasher, Robin Burke, and Yong Zheng. 2011. Context-Aware Recommendation Based On Review Mining. In *Proceedings of the 9th Workshop on Intelligent Techniques for Web Personalization & Recommender Systems, ITWP@IJCAI 2011, Barcelona, Spain, July 16, 2011 (CEUR Workshop Proceedings, Vol. 756)*. Sarabjot Singh Anand and Dietmar Jannach (Eds.). CEUR-WS.org. <https://ceur-ws.org/Vol-756/paper04.pdf>
- [22] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. 2010. *Recommender systems: an introduction*. Cambridge University Press.
- [23] Armand Joulin, Edouard Grave, and Piotr Bojanowski Tomas Mikolov. 2017. Bag of Tricks for Efficient Text Classification. *EACL 2017* (2017), 427.
- [24] Donghyun Kim, Chanyoung Park, Jinoh Oh, Sungyoung Lee, and Hwanjo Yu. 2016. Convolutional Matrix Factorization for Document Context-Aware Recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) (RecSys '16). Association for Computing Machinery, New York, NY, USA, 233–240. doi:10.1145/2959100.2959165
- [25] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. *Computer* 42, 8 (2009), 30–37. doi:10.1109/MC.2009.263
- [26] Balraj Kumar and Neeraj Sharma. 2016. Approaches, Issues and Challenges in Recommender Systems: A Systematic Review. *Indian Journal of Science and Technology* 9 (2016), 47.
- [27] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*. 7871–7880.

- [28] Chenliang Li, Cong Quan, Li Peng, Yunwei Qi, Yuming Deng, and Libing Wu. 2019. A Capsule Network for Recommendation and Explaining What You Like and Dislike. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) (SIGIR'19). Association for Computing Machinery, New York, NY, USA, 275–284. doi:10.1145/3331184.3331216
- [29] Duantengchuan Li, Hai Liu, Zhaoli Zhang, Ke Lin, Shuai Fang, Zhifei Li, and Neal N Xiong. 2021. CARM: Confidence-aware recommender model via review representation learning and historical rating behavior in the online platforms. *Neurocomputing* 455 (2021), 283–296.
- [30] Donghua Liu, Jing Li, Bo Du, Jun Chang, and Rong Gao. 2019. DAML: Dual Attention Mutual Learning between Ratings and Reviews for Item Recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Anchorage, AK, USA) (KDD '19). Association for Computing Machinery, New York, NY, USA, 344–352. doi:10.1145/3292500.3330906
- [31] Hongtao Liu, Yian Wang, Qiyao Peng, Fangzhao Wu, Lin Gan, Lin Pan, and Pengfei Jiao. 2020. Hybrid neural recommendation with joint deep representation learning of ratings and reviews. *Neurocomputing* 374 (2020), 77–85.
- [32] Yun Liu and Jun Miyazaki. 2022. Knowledge-aware attentional neural network for review-based movie recommendation with explanations. *Neural Comput. Appl.* 35, 3 (Sept. 2022), 2717–2735. doi:10.1007/s00521-022-07689-1
- [33] Yichao Lu, Ruihai Dong, and Barry Smyth. 2018. Coevolutionary Recommendation Model: Mutual Learning between Ratings and Reviews. In *Proceedings of the 2018 World Wide Web Conference* (Lyon, France) (WWW '18). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 773–782. doi:10.1145/3178876.3186158
- [34] Washington Luiz, Felipe Viegas, Rafael Alencar, Fernando Mourão, Thiago Salles, Dárlinton Carvalho, Marcos Andre Gonçalves, and Leonardo Rocha. 2018. A feature-oriented sentiment rating for mobile app reviews. In *Proceedings of the 2018 world wide web conference*. 1909–1918.
- [35] Sichun Luo, Chen Ma, Yuanzhang Xiao, and Linqi Song. 2023. Improving Long-Tail Item Recommendation with Graph Augmentation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* (CIKM '23). 1707–1716. doi:10.1145/3583780.3614929
- [36] Xin Luo, Mengchu Zhou, Yunni Xia, and Qingsheng Zhu. 2014. An Efficient Non-Negative Matrix-Factorization-Based Approach to Collaborative Filtering for Recommender Systems. *IEEE Transactions on Industrial Informatics* 10, 2 (2014), 1273–1284. doi:10.1109/TII.2014.2308433
- [37] Julian McAuley and Jure Leskovec. 2013. Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems*. 165–172.
- [38] Jianmo Ni and Julian McAuley. 2018. Personalized Review Generation By Expanding Phrases and Attending on Aspect-Aware Representations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Iryna Gurevych and Yusuke Miyao (Eds.). Association for Computational Linguistics, Melbourne, Australia, 706–711. doi:10.18653/v1/P18-2112
- [39] Aäron van den Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep content-based music recommendation. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2* (Lake Tahoe, Nevada) (NIPS'13). Curran Associates Inc., Red Hook, NY, USA, 2643–2651.
- [40] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.
- [41] Reinold Adrian Pugoy and Hung-Yu Kao. 2021. Unsupervised Extractive Summarization-Based Representations for Accurate and Explainable Collaborative Filtering. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 2981–2990. doi:10.18653/v1/2021.acl-long.232
- [42] Massimo Quadrana, Paolo Cremonesi, and Dietmar Jannach. 2018. Sequence-aware recommender systems. *ACM computing surveys (CSUR)* 51, 4 (2018), 1–36.
- [43] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender systems: introduction and challenges. *Recommender systems handbook* (2015), 1–34.
- [44] Tara N. Sainath, Brian Kingsbury, Vikas Sindhwani, Ebru Arisoy, and Bhuvana Ramabhadran. 2013. Low-rank matrix factorization for Deep Neural Network training with high-dimensional output targets. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. doi:10.1109/ICASSP.2013.6638949
- [45] J. Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. 2007. *Collaborative Filtering Recommender Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg, 291–324. doi:10.1007/978-3-540-72079-9_9
- [46] Liye Shi, Wen Wu, Wang Guo, Wenxin Hu, Jiayi Chen, Wei Zheng, and Liang He. 2022. SENGR: Sentiment-Enhanced Neural Graph Recommender. *Information Sciences* 589 (2022), 655–669. doi:10.1016/j.ins.2021.12.120
- [47] Alex Shtoff, Michael Viderman, Naama Haramaty-Krasne, Oren Somekh, Ariel Raviv, and Tularam Ban. 2024. Low Rank Field-Weighted Factorization Machines for Low Latency Item Recommendation. In *Proceedings of the 18th ACM Conference on Recommender Systems* (Bari, Italy) (RecSys '24). Association for Computing Machinery, New York, NY, USA, 238–246. doi:10.1145/3640457.3688097
- [48] Nicollas Silva, Heitor Werneck, Thiago Silva, Adriano C. M. Pereira, and Leonardo Rocha. 2021. A contextual approach to improve the user's experience in interactive recommendation systems. In *Proceedings of the Brazilian Symposium on Multimedia and the Web* (Belo Horizonte, Minas Gerais, Brazil) (Web-Media '21). Association for Computing Machinery, New York, NY, USA, 89–96. doi:10.1145/3470482.3479621
- [49] Mehdi Sifi, Ahmed Oussous, Ayoub Ait Lahcen, and Salma Mouline. 2020. Recommender systems based on collaborative filtering using review texts—a survey. *Information* 11, 6 (2020), 317.
- [50] Peijie Sun, Le Wu, Kun Zhang, Yu Su, and Meng Wang. 2021. An Unsupervised Aspect-Aware Recommendation Model with Explanation Text Generation. *ACM Trans. Inf. Syst.* 40, 3, Article 63 (Nov. 2021), 29 pages. doi:10.1145/3483611
- [51] Yunzhi Tan, Min Zhang, Yiqun Liu, and Shaoping Ma. 2016. Rating-boosted latent topics: understanding users and items with ratings and reviews. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (New York, New York, USA) (IJCAI '16). AAAI Press, 2640–2646.
- [52] Jiliang Tang, Huiji Gao, Xia Hu, and Huan Liu. 2013. Context-aware review helpfulness rating prediction. In *Proceedings of the 7th ACM Conference on Recommender Systems* (Hong Kong, China) (RecSys '13). 1–8.
- [53] Yi Tay, Anh Tuan Luu, and Siu Cheung Hui. 2018. Multi-Pointer Co-Attention Networks for Recommendation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (London, United Kingdom) (KDD '18). Association for Computing Machinery, New York, NY, USA, 2309–2318. doi:10.1145/3219819.3220086
- [54] Saúl Vargas and Pablo Castells. 2014. Improving sales diversity by recommending users to items. In *Proceedings of the 8th ACM Conference on Recommender Systems* (Foster City, Silicon Valley, California, USA) (RecSys '14). Association for Computing Machinery, New York, NY, USA, 145–152. doi:10.1145/2645710.2645744
- [55] Naan Vasconcelos, Davi Reis, Thiago Silva, Nicollas Silva, Washington Cunha, Elisa Tuler, and Leonardo Rocha. 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* (Juiz de Fora/MG). SBC, Porto Alegre, RS, Brasil, 390–394. doi:10.5753/webmedia.2024.241405
- [56] Joao Vinagre, Alípio Mário Jorge, Conceição Rocha, and Joao Gama. 2019. Statistically robust evaluation of stream-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering* 33, 7 (2019), 2971–2982.
- [57] Manolis Vozalis, Angelos Markos, and Konstantinos Margaritis. 2009. Evaluation of standard SVD-based techniques for Collaborative Filtering. *HERCMA'09 Proceedings* (2009).
- [58] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. 2015. Collaborative Deep Learning for Recommender Systems. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Sydney, NSW, Australia) (KDD '15). Association for Computing Machinery, New York, NY, USA, 1235–1244. doi:10.1145/2783258.2783273
- [59] Kr Wang, Yanmin Zhu, Haobing Liu, Tianzi Zang, and Chunyang Wang. 2023. Learning Aspect-Aware High-Order Representations from Ratings and Reviews for Recommendation. *ACM Trans. Knowl. Discov. Data* 17, 1, Article 3 (Feb. 2023), 22 pages. doi:10.1145/3532188
- [60] Xinxin Wang and Ye Wang. 2014. Improving Content-based and Hybrid Music Recommendation using Deep Learning. In *Proceedings of the 22nd ACM International Conference on Multimedia* (Orlando, Florida, USA) (MM '14). Association for Computing Machinery, New York, NY, USA, 627–636. doi:10.1145/2647868.2654940
- [61] Heitor Werneck, Nicollas Silva, Matheus Carvalho Viana, Fernando Mourão, Adriano CM Pereira, and Leonardo Rocha. 2020. A survey on point-of-interest recommendation in location-based social networks. In *Proceedings of the Brazilian Symposium on Multimedia and the Web*. 185–192.
- [62] Chuhuan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. NPA: Neural News Recommendation with Personalized Attention. In *Proceedings of the 25th ACM SIGKDD* (Anchorage, AK, USA) (KDD '19). Association for Computing Machinery, 9 pages. doi:10.1145/3292500.3330665
- [63] HuiLin Yu, Tieyun Qian, Yile Liang, and Bing Liu. 2020. Adversarial Generation of Target Review for Rating Prediction. In *Database Systems for Advanced Applications: 25th International Conference, DASFAA 2020, Jeju, South Korea, September 24–27, 2020, Proceedings, Part II* (Jeju, Korea (Republic of)). Springer-Verlag, Berlin, Heidelberg, 73–89. doi:10.1007/978-3-030-59416-9_5
- [64] Yongfeng Zhang and Xu Chen. 2020. Explainable Recommendation: A Survey and New Perspectives. (01 2020). doi:10.1561/9781680836592
- [65] Yuying Zhao, Yu Wang, Yunchao Liu, Xueqi Cheng, Charu C. Aggarwal, and Tyler Derr. 2025. Fairness and Diversity in Recommender Systems: A Survey. *ACM Trans. Intell. Syst. Technol.* 16, 1, Article 2 (2025), 28 pages. doi:10.1145/3664928
- [66] Lei Zheng, Vahid Noroozi, and Philip S. Yu. 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* (Cambridge, United Kingdom) (WSDM '17). doi:10.1145/3018661.3018665
- [67] Xun Zhou, Jing He, Guangyan Huang, and Yanchun Zhang. 2015. SVD-based incremental approaches for recommender systems. *J. Comput. System Sci.* 81, 4 (2015), 717–733. doi:10.1016/j.jcss.2014.11.016