

Aprimorando a Personalização de Sistemas de Recomendação por Calibração Multicategoria

Paul Atauchi¹, André Levi Zanon², Leonardo Rocha³, Marcelo G. Manzato¹

¹Universidade de São Paulo, ²University College Cork, ³Universidade Federal de São João del-Rei, paul.atauchi@usp.br; andre.zanon@insight-centre.org; manzato@icmc.usp.br; lcrocha@ufsj.edu.br

ABSTRACT

Recommender systems (RSs) are widely used to personalize suggestions based on users' historical data. Despite recent advances, many RSs still prioritize accuracy metrics, often generating non-personalized lists that fail to reflect users' actual interests. Calibration strategies have been proposed to address this misalignment by tailoring recommendations to users' profiles. However, most approaches consider only a single category (e.g., genre), limiting their ability to capture the multifaceted nature of user preferences. This paper proposes a multi-category calibration method that simultaneously incorporates multiple categories extracted from knowledge graphs. The approach aims to balance recommendation relevance with alignment across diverse user profile dimensions. We evaluate the method on two well-known datasets, MovieLens and LastFM, using two contrasting recommendation algorithms: NCF (neural) and BPR-MF (non-neural). Results show that multi-category calibration can improve alignment with multifaceted user preferences while maintaining or enhancing accuracy, fostering fairer, more personalized, and more relevant recommendations.

KEYWORDS

Recomendação, Calibração multicategoria, Personalização

1 INTRODUÇÃO

Sistemas de Recomendação (SR) geram sugestões personalizadas a partir das interações prévias dos usuários [11]. Apesar dos avanços, muitos SR ainda priorizam métricas como acurácia, o que frequentemente leva a divergências entre a lista recomendada e a distribuição real das preferências do usuário. No domínio de filmes, por exemplo, se um usuário interagiu com 30% de dramas, 45% de filmes de ação e 25% de animações, espera-se que as recomendações refletissem essas proporções [4, 14, 15]. No entanto, na prática, observa-se que os sistemas podem não reproduzir fielmente esses padrões, possivelmente devido a vieses, à natureza do algoritmo adotado ou aos objetivos do sistema de recomendação [1, 9, 16].

Para enfrentar esse problema, Steck [15] introduziu o conceito de calibração ao contexto de sistemas de recomendação, visando alinhar a distribuição das categorias das recomendações às preferências do perfil do usuário. Contudo, sua abordagem realiza essa calibração considerando apenas uma única categoria (por exemplo, gênero), limitando a capacidade de capturar a complexidade das preferências reais.

Partimos da premissa de que os interesses dos usuários são multifacetados (Figura 1), ou seja, moldados por múltiplas categorias simultaneamente [2]. Assim, além do gênero, outros fatores, tais como prêmios recebidos ou país de origem, podem influenciar as escolhas. Por exemplo, um usuário pode preferir não apenas filmes de ação, mas filmes de ação que venceram o Oscar de Melhor Filme.

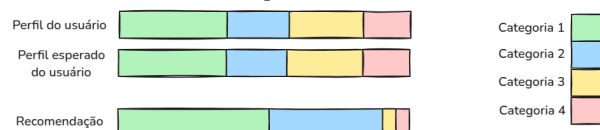


Figura 1: Exemplo ilustrativo de desalinhamento entre o perfil do usuário, o perfil esperado e a lista de recomendação. As cores representam diferentes categorias de interesse.

Neste artigo, propomos um método de calibração multicategoria que estende a abordagem de Steck [15], incorporando múltiplas categorias extraídas de um Grafo de Conhecimento (GC). Nossa proposta adiciona restrições explícitas de calibração ao processo de recomendação, garantindo que as listas geradas representem de forma mais fiel os interesses dos usuários. Formulamos a seguinte Questão de Pesquisa **QP: A calibração multicategoria melhora o alinhamento das listas de recomendação com os interesses multifacetados dos usuários?**

Para responder a essa Questão de Pesquisa (QP), adaptamos o método de Steck [15] para considerar três categorias (i.e. gênero, prêmio recebido e país de origem), em dois conjuntos de dados, utilizando um GC. Avaliamos acurácia e alinhamento, comparando abordagens de categoria única com a proposta multicategoria, e observamos ganhos consistentes em ambos os aspectos.

2 TRABALHOS RELACIONADOS

A equidade em SR pode estar associada à igualdade de oportunidade, exposição balanceada e representação das preferências, tornando-se um tema cada vez mais relevante [9]. Entre as diversas técnicas voltadas à equidade propostas na literatura, a calibração destaca-se por buscar alinhar a lista de recomendações recebida pelo usuário ao seu histórico de interesses [15]. Wang et al. [16] apresentaram uma taxonomia abrangente sobre equidade em SR, situando a calibração em um grupo que considera os sujeitos envolvidos (usuários ou itens), a granularidade (individual ou em grupo) e os objetivos de otimização. Na literatura, as técnicas de calibração geralmente são implementadas por meio de métodos algorítmicos de pós-processamento ou reordenação.

Steck [15] introduziu a primeira definição de calibração em SR, propondo o reordenamento da lista de recomendações em uma etapa de pós-processamento para alinhar a distribuição de gêneros dos itens recomendados às preferências do usuário, avaliando a calibração por meio da divergência de Kullback-Leibler (KL).



Figura 2: Pipeline do processo de calibração para categoria única e multicategoria em sistemas de recomendação.

Entretanto, abordagens iniciais de calibração não consideravam vieses estruturais. Abdollahpouri et al. [1] evidenciaram a influência do viés de popularidade, mostrando que usuários que consomem itens mais populares tendem a receber recomendações descalibradas, pois os sistemas favorecem esses itens.

Para mitigar tais limitações, Sacilotti et al. [12] desenvolveram um *framework* de calibração que integra informações de gênero e popularidade. Seu modelo híbrido ajusta as recomendações de forma a equilibrar a equidade com a exposição a itens de cauda longa. De forma semelhante, Da Silva et al. [5] generalizaram a calibração ao propor um *framework* de pós-processamento agnóstico ao SR, capaz de suportar múltiplas métricas de divergência e valores de *trade-off* personalizados por usuário, permitindo controlar o equilíbrio entre equidade e acurácia. Naghiaei et al. [9] expandiram o conceito tradicional ao introduzir um algoritmo de calibração que incorpora dimensões como novidade, serendipidade e diversidade. Seu método agrupa usuários e aplica um algoritmo de reordenação multiobjetivo, ampliando o escopo da calibração para além do alinhamento de categorias. Ainda assim, a maioria dos métodos de calibração, conforme inicialmente proposto por Steck [15], restringe-se a uma única categoria, no caso de Steck, o gênero cinematográfico, para alinhar as recomendações ao perfil do usuário.

3 CALIBRAÇÃO MULTICATEGORIA

A calibração em SR tem como objetivo ajustar a lista recomendada de forma que a distribuição de categorias dos itens apresentados reflita, de maneira equilibrada, as preferências e interesses do usuário. SR que não aplicam calibração podem apresentar listas enviesadas para categorias dominantes, o que pode comprometer outras métricas além da precisão [2, 9, 15].

Neste trabalho, propomos uma extensão da estratégia de calibração unidimensional apresentada por [15], introduzindo uma abordagem de *calibração multicategoria*. Essa abordagem ajusta, de forma sistemática, a distribuição das diferentes categorias na lista recomendada, incorporando informações sobre a correlação entre o histórico do usuário e os itens recomendados em cada categoria. O objetivo é garantir que a lista final não apenas maximize a relevância predita, mas também mantenha um alinhamento proporcional e equilibrado com os múltiplos interesses do usuário.

A Figura 2 ilustra a *pipeline* do processo de calibração, contemplando tanto a abordagem de categoria única quanto a multicategoria. O fluxo é executado da esquerda para a direita. Inicialmente, o sistema de recomendação recebe como entrada o conjunto de dados e o algoritmo selecionado, gerando uma lista recomendada inicial.

Em seguida, essa lista, juntamente com o histórico de interações do usuário e as categorias que compõem seu perfil, é utilizada como entrada no módulo de calibração. Nesse módulo, o ajuste é realizado considerando ou uma única categoria ou múltiplas categorias.

A Equação (1) define a função objetivo como um problema de otimização que busca maximizar a relevância das recomendações e minimizar a divergência entre as categorias da lista recomendada e o perfil do usuário.

O termo $s(I_u)$ quantifica a relevância das recomendações, enquanto $D_{KL}^C(p, q(I_u))$ representa a divergência entre a distribuição de categorias observada na lista gerada e o perfil de preferências do usuário p .

O parâmetro $\lambda \in [0, 1]$ atua como um fator de regularização, controlando o equilíbrio (*trade-off*) entre relevância e alinhamento categórico.

A formulação abrange tanto o caso de categoria única ($|C| = 1$), em que $\omega(c) = 1$, quanto o caso multicategoria ($|C| > 1$), no qual cada categoria é ponderada por $(1 - \rho(c))$, conforme estabelecido na Equação (3).

$$I_u^* = \arg \max_{I_u} \left[(1 - \lambda) s(I_u) - \lambda D_{KL}^C(p, q(I_u)) \right], \quad (1)$$

$$D_{KL}^C(p, q(I_u)) = \sum_{c \in C} \omega(c) D_{KL}(p^c, q(I_u)^c), \quad (2)$$

com $D_{KL}(\cdot, \cdot)$ sendo a divergência de Kullback–Leibler, C o conjunto de categorias consideradas, e $\omega(c)$ um coeficiente de ponderação definido por

$$\omega(c) = \begin{cases} 1 - \rho(c), & \text{se } |C| > 1 \quad (\text{caso multicategoria}), \\ 1, & \text{se } |C| = 1 \quad (\text{caso de categoria única}). \end{cases} \quad (3)$$

Sendo $\rho(c) \in [0, 1]$ a *correlação categórica* entre o histórico do usuário e a lista recomendada com alinhamento na categoria c . Os pesos $\rho(c)$ podem ser estimados com base na similaridade entre o perfil do usuário e a lista recomendada, derivada de seu histórico de interações. Essa formulação permite incorporar simultaneamente múltiplas categorias, minimizando divergências específicas sem comprometer a relevância global da lista. Além disso, o termo de ponderação $\rho(c)$ mitiga o risco de sobreajuste em categorias com baixa correlação, assegurando maior robustez e equilíbrio na calibração das recomendações.

Por fim, a formulação proposta será avaliada em cenários de recomendação com múltiplas categorias de interesse, analisando seu

impacto nas métricas de relevância e alinhamento das categorias ao perfil do usuário, em comparação à calibração por categoria única.

4 MATERIAIS E MÉTODOS

Para responder a QP e analisar o alinhamento da listas de recomendação com o perfil do usuário, por meio do método proposto e descrito na Seção 3, utilizamos duas bases de dados: *MovieLens-Latest-Small* (MovieLens) [7] e LastFM [3]. Para gerar recomendações para os usuários dessas bases de dados executamos dois algoritmos de recomendação: o *Bayesian Personalized Ranking Matrix Factorization* (BPR-MF) [10] e o Neural Collaborative Filtering (NCF) [8], que representam um algoritmo não neural e neural, respectivamente.

Diferentemente de trabalhos anteriores em somente uma categoria foi utilizada para o alinhamento entre perfil do usuário e as recomendações, em uma etapa de pós-processamento, extraímos um GC da Wikidata¹ a fim de obter várias categorias sobre os itens das bases de dados. Para a obtenção do GC para o MovieLens, utilizamos a informação do identificador *imdbId* comum a base de dados e a Wikidata para extrair dados multicategoria sobre itens. Para extrair as informações sobre artistas para a base de dados LastFM, utilizamos o nome dos artistas para encontrar suas respectivas informações. Um GC conecta itens a atributos por meio de categorias, criando triplas (*item, categoria, atributo*). No GC extraído para o MovieLens encontramos 9535 itens, 23 categorias e 69487 atributos em 295787 triplas. Já para o GC da base de dados do LastFM encontramos 11646 itens, 33 categorias diferentes e 23001 atributos em 134197 triplas.

Como etapa de pré-processamento nas bases de dados excluímos interações com itens que não foram encontrados na Wikidata. Isso resultou em uma base de dados com 9724 itens, 610 usuários e 100836 interações para a base do MovieLens. Já para a LastFM resultou em uma base de dados com 11641 itens, 1875 usuários e 13883 interações. A nossa base de dados resultante se encontra no nosso repositório de código² Em seguida, com as bases de dados processadas, executamos os dois algoritmos de recomendação BPR-MF e NCF e então o algoritmo de calibração multicategoria. Para avaliação, utilizamos as métricas de *Mean Average Precision* (MAP) para medir a acurácia antes e depois da calibração e *Mean Rank Miscalibration* (MRMC) [6] para medir o alinhamento do perfil do usuário também antes e após a calibração.

5 RESULTADOS

Para responder à QP, avaliamos o impacto da calibração multicategoria sobre duas métricas: a precisão das listas recomendadas, medida pelo MAP, que considera a quantidade e a ordenação dos itens relevantes, e o erro de calibração, quantificado pelo MRMC.

Para cada conjunto de dados (MovieLens e *LastFM-2k* (LastFM)), treinamos dois algoritmos de recomendação (BPR-MF e NCF). Em seguida, aplicamos o processo de calibração considerando:

- (1) Categoria única: *gênero*, *prêmios recebidos* e *país de origem*.
- (2) Multicategoria: combinação das três categorias anteriores.

No domínio de filmes (Figura 3), o MAP (gráfico à esquerda) apresenta variações ao longo de diferentes valores de *trade-off*. As calibrações por gênero e multicategoria demonstram desempenho superior em relação às calibrações por prêmios recebidos e país de

origem. Entretanto, a calibração por gênero supera ligeiramente a calibração multicategoria em praticamente todos os pontos de *trade-off*. Quanto ao MRMC (gráfico à direita), observa-se que as calibrações por categoria única apresentam reduções consistentes na divergência, enquanto a calibração multicategoria também reduz o erro, mas de forma mais gradual. Para valores de *trade-off* próximos de 1.0, ocorre queda acentuada no MAP tanto para a calibração por categoria única quanto para a multicategoria, indicando que o excesso de alinhamento com a distribuição das categorias do perfil do usuário compromete a precisão da lista recomendada. Assim, valores intermediários de *trade-off* oferecem o melhor equilíbrio entre precisão e calibração, assegurando maior justiça na distribuição das recomendações com perdas mínimas (ou até ganhos) de acurácia.

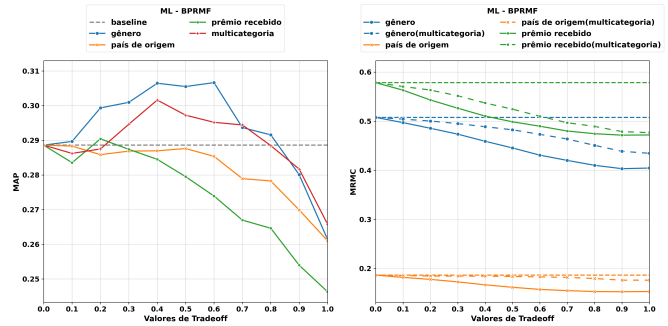


Figura 3: Resultados do MAP e MRMC para o algoritmo de recomendação BPR-MF no conjunto de dados MovieLens.

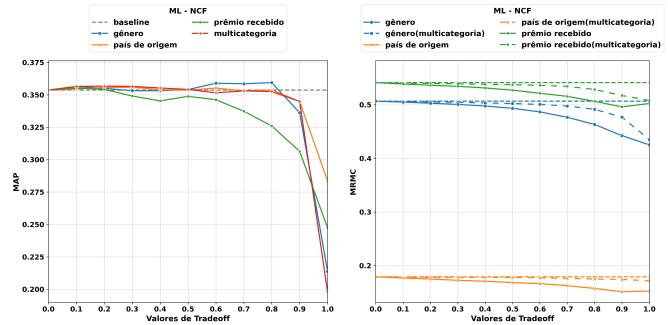


Figura 4: Resultados do MAP e MRMC para o algoritmo de recomendação NCF no conjunto de dados MovieLens.

Na Figura 4, o MAP (gráfico à esquerda) apresenta valores homogêneos próximos ao *baseline* para ambas as calibrações, categoria única e multicategoria, com exceção da categoria prêmios recebidos, que apresenta queda em relação às demais. As calibrações por gênero mantêm desempenho estável, com pequeno ganho em torno dos *trade-offs* 0.5 e 0.8. Após esse ponto, todas as curvas apresentam queda acentuada, especialmente quando o *trade-off* se aproxima de 1.0, reforçando que priorizar excessivamente o alinhamento com as distribuições das preferências do usuário resulta em perda de precisão. No MRMC (gráfico à direita), as calibrações por categoria única apresentam reduções graduais da divergência, enquanto a calibração multicategoria mostra redução mais acentuada à medida que o *trade-off* se aproxima de 1.0. Valores de *trade-off* entre 0 e aproximadamente 0.8 oferecem o melhor equilíbrio entre acurácia

¹<https://www.wikidata.org/>

²<https://gitlab.com/paul.atauchi/multi-calibration>

e alinhamento com o perfil do usuário, sobretudo nas calibrações por gênero, país de origem e multicategoria.

No domínio musical (LastFM), o comportamento difere do observado nos dados de filmes. Na Figura 5 (gráfico à esquerda), para a métrica MAP com o algoritmo BPR-MF, observa-se crescimento acentuado nas calibrações por gênero (entre *trade-offs* 0 e aproximadamente 0.45) e multicategoria (entre *trade-offs* 0 e aproximadamente 0.55). Nota-se também que, na maior parte da faixa analisada (0 até cerca de 0.9), a calibração multicategoria supera as calibrações por categorias únicas. Já as categorias país de origem (até cerca de 0.2) e prêmios recebidos (até cerca de 0.15) apresentam crescimento mais modesto. No gráfico à direita (MRMC), tanto as calibrações por categoria única quanto a multicategoria reduzem a divergência em relação ao perfil do usuário. As curvas de categoria única mantêm queda consistente até o final da faixa de *trade-offs*, enquanto a calibração multicategoria também diminui, mas com declive mais suave, mantendo-se sempre abaixo do *baseline*. Assim como no MovieLens com BPR-MF, no LastFM valores intermediários de *trade-off* promovem simultaneamente ganhos de precisão e melhor alinhamento às preferências do usuário.

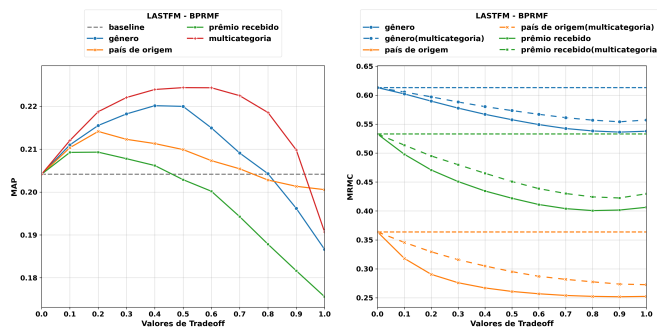


Figura 5: Resultados do MAP e MRMC para o algoritmo de recomendação BPR-MF no conjunto de dados LastFM.

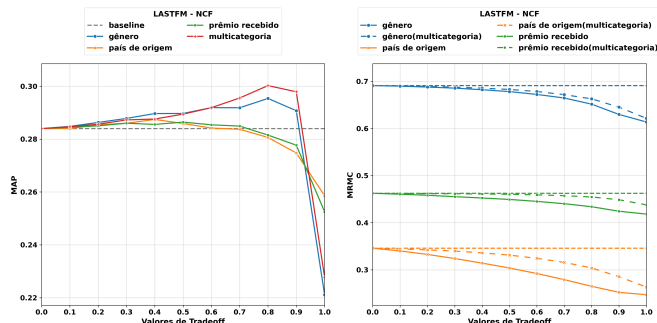


Figura 6: Resultados do MAP e MRMC para o algoritmo de recomendação NCF no conjunto de dados LastFM.

Na Figura 6, para o conjunto LastFM com o algoritmo NCF, o MAP (gráfico à esquerda) apresenta crescimento contínuo nas calibrações por gênero e multicategoria, entre *trade-offs* 0 e aproximadamente 0.8. Em diversos pontos dessa faixa, a calibração multicategoria supera as calibrações por categorias únicas. Tanto as configurações de categoria única quanto multicategoria permanecem acima do *baseline* entre *trade-offs* 0 e aproximadamente

0.7. No gráfico à direita (MRMC), todas as configurações reduzem suavemente a divergência em relação ao perfil do usuário, com exceção da categoria país de origem, que apresenta redução mais pronunciada. Observa-se que a redução de MRMC é mais expressiva nas calibrações por categoria única do que na multicategoria, embora nesta última a recomendação seja ajustada simultaneamente às preferências do usuário em todas as categorias.

Com base nos resultados obtidos, verifica-se que a calibração multicategoria, considerando simultaneamente gênero, país de origem e prêmios recebidos, tende a promover um equilíbrio entre a precisão (MAP) e o alinhamento às preferências multifacetadas dos usuários (MRMC). No MovieLens, valores intermediários de *trade-off* (no BPR-MF) e valores entre 0 e aproximadamente 0.8 (no NCF) mostraram-se mais favoráveis, especialmente para as calibrações por gênero e multicategoria, enquanto valores próximos de 1.0 prejudicaram significativamente a precisão. No LastFM, o efeito variou conforme o algoritmo: com BPR-MF, a maior precisão ocorreu nas calibrações por gênero e multicategoria, enquanto valores próximos de 1.0 reduziram tanto a precisão quanto o alinhamento. Já com NCF, valores de *trade-off* até aproximadamente 0.9 proporcionaram o melhor equilíbrio, com ganhos iniciais de precisão e redução consistente da divergência. Esses resultados indicam que a calibração multicategoria pode contribuir para recomendações mais precisas e melhor alinhadas aos interesses multifacetados dos usuários, desde que aplicado um *trade-off* moderado.

Respondendo à QP, **a calibração multicategoria melhora o alinhamento das listas de recomendação com os interesses multifacetados dos usuários, promovendo redução consistente do erro de calibração em todos os cenários avaliados.**

6 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho, apresentamos uma abordagem de calibração multicategoria para sistemas de recomendação, com o objetivo de alinhar as listas recomendadas aos interesses multifacetados dos usuários. Adaptamos o método de Steck [15] de forma a integrar múltiplas facetas extraídas de grafos de conhecimento, gênero, país de origem e prêmios recebidos, e avaliamos seu impacto tanto na acurácia das recomendações quanto no alinhamento com o perfil do usuário.

Nossos resultados indicam que a calibração multicategoria melhora o alinhamento das recomendações com as preferências multifacetadas dos usuários (MRMC), sem comprometer a acurácia (MAP). Nos domínios LastFM e MovieLens, utilizando os algoritmos BPR-MF e NCF, a calibração multicategoria apresentou ganhos de precisão em diferentes faixas de *trade-off* quando comparada às calibrações por categorias únicas. Como principal contribuição, demonstramos que incorporar múltiplas categorias no processo de calibração não apenas é viável, mas também traz benefícios sob diferentes perspectivas: proporciona alinhamento mais fiel ao perfil do usuário e preserva, ou até melhora, a acurácia das recomendações.

Como trabalhos futuros, propomos investigar estratégias adaptativas de calibração personalizadas por usuário, avaliar o impacto da calibração multicategoria na qualidade e na transparência das explicações [13, 17] geradas pelos sistemas de recomendação e ampliar a análise para estudos com usuários em cenários reais.

AGRADECIMENTOS

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

REFERÊNCIAS

- [1] Himan Abdollahpour, Masoud Mansoury, Robin Burke, and Bamshad Mobasher. 2020. The connection between popularity bias, calibration, and fairness in recommendation. In *Proceedings of the 14th ACM conference on recommender systems*. 726–731.
- [2] Himan Abdollahpour, Zahra Nazari, Alex Gain, Clay Gibson, Maria Dimakopoulou, Jesse Anderton, Benjamin Carterette, Mounia Lalmas, and Tony Jebara. 2023. Calibrated recommendations as a minimum-cost flow problem. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*. 571–579.
- [3] Iván Cantador, Peter Brusilovsky, and Tsvi Kuflik. 2011. 2nd Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011). In *Proceedings of the 5th ACM conference on Recommender systems* (Chicago, IL, USA) (*RecSys 2011*). ACM, New York, NY, USA.
- [4] Diego Corrêa da Silva and Frederico Araújo Durão. 2025. Benchmarking fairness measures for calibrated recommendation systems on movies domain. *Expert Systems with Applications* 269 (2025), 126380.
- [5] Diego Corrêa da Silva, Marcelo Garcia Manzato, and Frederico Araújo Durão. 2021. Exploiting personalized calibration and metrics for fairness recommendation. *Expert Systems with Applications* 181 (2021), 115112.
- [6] Diego Corrêa da Silva, Marcelo Garcia Manzato, and Frederico Araújo Durão. 2021. Exploiting personalized calibration and metrics for fairness recommendation. *Expert Systems with Applications* 181 (2021), 115112. <https://doi.org/10.1016/j.eswa.2021.115112>
- [7] F Maxwell Harper and Joseph A Konstan. 2015. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)* 5, 4 (2015), 1–19.
- [8] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *Proceedings of the 26th International Conference on World Wide Web* (Perth, Australia) (*WWW '17*). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 173–182. <https://doi.org/10.1145/3038912.3052569>
- [9] Mohammadmehdi Naghiaei, Mahdi Dehghan, Hossein A Rahmani, Javad Azizi, and Mohammad Aliannejadi. 2024. Personalized beyond-accuracy calibration in recommendation. In *Proceedings of the 2024 ACM SIGIR International Conference on Theory of Information Retrieval*. 107–116.
- [10] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence* (Montreal, Quebec, Canada) (*UAI '09*). AUAI Press, Arlington, Virginia, USA, 452–461.
- [11] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2022. *Recommender Systems: Techniques, Applications, and Challenges*. Springer US, New York, NY, 1–35. https://doi.org/10.1007/978-1-0716-2197-4_1
- [12] Andre Sacilotti, Rodrigo Ferrari de Souza, and Marcelo Garcia Manzato. 2023. Counteracting popularity-bias and improving diversity through calibrated recommendations. In *Proceedings*.
- [13] Diego Silva and Frederico Durão. 2022. Explorando Justiça em Sistemas de Recomendação. In *Anais Estendidos do XXVIII Simpósio Brasileiro de Sistemas Multimídia e Web* (Curitiba). SBC, Porto Alegre, RS, Brasil, 11–14. https://doi.org/10.5753/webmedia_estendido.2022.225303
- [14] Rodrigo Souza and Marcelo Manzato. 2024. Explorando Formas de Calibração e Redução do Viés de Popularidade em Sistemas de Recomendação. In *Anais Estendidos do XXX Simpósio Brasileiro de Sistemas Multimídia e Web* (Juiz de Fora/MG). SBC, Porto Alegre, RS, Brasil, 9–10. https://doi.org/10.5753/webmedia_estendido.2024.244380
- [15] Harald Steck. 2018. Calibrated recommendations. In *Proceedings of the 12th ACM conference on recommender systems*. 154–162.
- [16] Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. 2023. A survey on the fairness of recommender systems. *ACM Transactions on Information Systems* 41, 3 (2023), 1–43.
- [17] André Zanon, Leonardo Rocha, and Marcelo Manzato. 2024. O Impacto de Estratégias de Embeddings de Grafos na Explicabilidade de Sistemas de Recomendação. In *Proceedings of the 30th Brazilian Symposium on Multimedia and the Web* (Juiz de Fora/MG). SBC, Porto Alegre, RS, Brasil, 231–239. <https://doi.org/10.5753/webmedia.2024.241857>