

Limitações e Barreiras na Implementação Prática de Sistemas de Cerca Geográfica Inteligentes

EMANUEL DA COSTA ANDRADE CARDOSO¹, FABRICIO BATISTA NARCIZO^{2,3}, MARIO T. SHIMANUKI⁴

¹ Cursando Tecnologia em Análise e Desenvolvimento de Sistemas. IFSP, Câmpus Caraguatatuba, emanuel.c@aluno.ifsp.edu.br

² Senior AI Research Scientist, GN Advanced Science, GN One, fbnarcizo@gn.com / ³ Part-time Lecturer, Computer Science Department, IT University of Copenhagen, København S, Denmark, narcizo@itu.dk

⁴ Professor no Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, IFSP, Campus Caraguatatuba, mario@ifspcaragua.net

Área de conhecimento (Tabela CNPq): Ciência da Computação – 1.03.00.00-7

RESUMO: Este estudo apresenta os resultados sobre os desafios na implementação de sistemas de cerca geográfica inteligentes integrados a algoritmos de aprendizado de máquina. O trabalho teve como objetivo analisar as limitações e as barreiras enfrentadas durante a reprodução de um modelo adaptativo de recomendação de Pontos de Interesse (POI), visando compreender os fatores que dificultam a reprodutibilidade e a eficácia desses sistemas. A pesquisa foi conduzida por meio da implementação do algoritmo de filtragem colaborativa adaptativa, utilizando bases de dados reais do *Foursquare* para as cidades de Nova Iorque e Tóquio. Foram aplicadas métricas de precisão e *recall*, cujos resultados demonstraram discrepâncias significativas em relação ao estudo de referência, evidenciando problemas na qualidade e consistência das bases de dados, além de lacunas na documentação dos parâmetros do modelo. Essas limitações impactaram diretamente o desempenho e a validação do sistema proposto. A análise revelou que a reprodutibilidade e a confiabilidade das soluções dependem fortemente da padronização dos dados e da transparência metodológica. Conclui-se que a integração entre cerca geográfica e aprendizado de máquina representa uma abordagem promissora, porém ainda desafiadora na prática. Recomenda-se que futuros estudos explorem bases de dados mais amplas, considerem variáveis contextuais como tempo de permanência e avaliações dos usuários, e investiguem o uso de técnicas de aprendizado profundo para aprimorar a precisão das recomendações e a escalabilidade dos sistemas.

PALAVRAS-CHAVE: Cerca Geográfica; Aprendizado de Máquina; Filtragem Colaborativa; Reprodutibilidade; Sistemas de Recomendação; POI.

Limitations and Barriers in the Practical Implementation of Intelligent Geofencing Systems

ABSTRACT: This study presents the results on the challenges involved in the implementation of intelligent geofencing systems integrated with machine learning algorithms. The objective was to analyze the limitations and barriers encountered during the reproduction of an adaptive Point of Interest (POI) recommendation model, aiming to understand the factors that hinder the reproducibility and effectiveness of such systems. The research was conducted through the implementation of the adaptive collaborative filtering algorithm, using real Foursquare datasets from New York and Tokyo. Precision and recall metrics were applied, and the results revealed significant discrepancies compared to the reference study, highlighting issues related to data quality and consistency, as well as gaps in the documentation of model parameters. These limitations directly impacted the system's performance and the validation of the proposed approach. The analysis indicated that the reproducibility and reliability of these solutions strongly depend on data standardization and methodological transparency. It is concluded that the integration of geofencing and machine learning represents a promising yet challenging approach in practice. Future studies are recommended to explore richer datasets, consider contextual variables such as user dwell time and place ratings, and investigate the use of deep learning techniques to improve recommendation accuracy and system scalability.

KEYWORDS: Geofencing; Machine Learning; Collaborative Filtering; Reproducibility; Recommender Systems; POI.

INTRODUÇÃO

O avanço das tecnologias de geolocalização e Internet das Coisas (IoT – Internet of Things), e a onipresença de dispositivos móveis, têm gerado um grande volume de dados sobre a mobilidade humana. Estes dados representam uma oportunidade única para o desenvolvimento de serviços personalizados, mas sua coleta contínua (*Geotracking*) acarreta preocupações significativas com a privacidade do usuário e o consumo de bateria [1, 2]. Nesse cenário, a cerca geográfica (Geofencing) surge como uma alternativa eficiente, acionando eventos baseados na entrada ou saída de perímetros virtuais pré-definidos, de forma menos intrusiva [2].

Contudo, a aplicação tradicional da cerca geográfica é majoritariamente reativa, limitando-se a disparar ações genéricas. Isso representa uma oportunidade desconsiderada de engajamento, pois as notificações ou ofertas não levam em conta o comportamento e as preferências individuais do usuário [2]. Este é o problema central que motiva a integração de cerca geográfica com algoritmos de machine learning (ML): a necessidade de sistemas proativos que utilizem dados de localização de forma inteligente e contextual.

Para superar essa limitação, propõe-se a utilização de cerca geográfica em conjunto com algoritmos de machine learning, em particular por meio da implementação de variações da filtragem colaborativa, como a proposta feita por Yueyu et al [3]. Essa integração transforma uma ferramenta reativa em preditiva, capaz de analisar o histórico de visitas e prever o comportamento futuro do usuário, permitindo a criação de sistemas de recomendação verdadeiramente proativos [4]. No entanto, a implementação prática desses sistemas revela desafios significativos que são frequentemente subestimados ou inadequadamente documentados na literatura científica.

O objetivo deste trabalho foi identificar, documentar e analisar os principais desafios encontrados na implementação prática de sistemas de cerca geográfica inteligentes. Utilizou-se como estudo de caso a reprodução do algoritmo de filtragem colaborativa adaptativa (*Adaptive-CF*) proposto por Yueyu et al [3]. Por meio desta pesquisa, foi identificado duas barreiras críticas à reprodutibilidade: 1) inconsistência e degradação de dados e 2) lacunas na documentação do algoritmo. Esses fatores impedem a replicação precisa dos resultados originais, sublinhando a necessidade de maior rigor e transparência na pesquisa para garantir a reprodutibilidade e a confiabilidade das soluções apresentadas.

METODOLOGIA

A metodologia tem como base uma pesquisa empírica por meio da implementação e avaliação de um sistema de recomendação de Pontos de Interesse (POIs) em busca de entender as lacunas existentes que dificultam a reprodutibilidade. Como estudo de caso, selecionou-se o trabalho de Yueyu et al [3], que propõe um algoritmo de filtragem colaborativa adaptativa para recomendação de POIs baseado em regiões de atividade. Esta escolha foi motivada por três fatores: 1) a relevância teórica para sistemas de recomendação de POI, 2) a disponibilidade de metodologia detalhada, e 3) a existência de uma base de dados pública mencionada pelos autores.

Os dados para o treinamento e avaliação do sistema foram obtidos a partir de uma base de dados de *check-ins* do Foursquare, mais precisamente Foursquare-Nova York e Foursquare-Tokyo, mencionado no trabalho de referência [5]. O pré-processamento seguiu as diretrizes do artigo original:

1. Limpeza de Dados: Remoção de registros duplicados ou anômalos e remoção de POI e usuário com menos de 10 *check-ins*;
2. Divisão Treino/Teste: 70% dos dados para treino, 30% para teste;
3. Mapeamento de ID: Conversão para índices numéricos para otimização.

Métricas de Avaliação

Foram utilizadas as métricas-padrão de sistemas de recomendação, amplamente aceitas na literatura [3, 4]:

Precisão@k (*Precision@k*): Mede a proporção de itens recomendados que foram efetivamente visitados pelo usuário no conjunto de teste.

$$Precision@k = \frac{(\text{Número de POI recomendados e visitados})}{k}$$

Onde k representa o número de recomendações apresentadas ao usuário.

Recall@k: Mede a proporção dos POI efetivamente visitados pelo usuário que foram corretamente recomendados pelo sistema.

$$Recall@k = \frac{(\text{Número de POI recomendados e visitados})}{\text{Total de POI visitados no teste}}$$

As avaliações foram realizadas para k = 5, 10, e 20, permitindo comparação direta com o trabalho de referência.

RELAÇÃO ENTRE O MODELO IMPLEMENTADO E A CERCA GEOGRÁFICA

O modelo de recomendação de POI se integra a um sistema de cerca geográfica inteligente em um ambiente de produção, atuando como o componente de inteligência que processa e analisa os dados coletados por meio de perímetros virtuais. A arquitetura conceitual de um sistema real é dividida em três camadas interconectadas. A primeira é a Camada de Captura de Dados, onde perímetros virtuais são definidos ao redor de pontos de interesse. Esta tecnologia é ativada apenas em locais específicos, registrando eventos de entrada e saída e permitindo o registro de *check-ins* que validam a presença física do usuário.

A segunda é a Camada de Processamento e Aprendizado, onde os *check-ins* validados pela cerca geográfica alimentam o algoritmo. O modelo analisa os padrões históricos de visita para identificar similaridades entre usuários e, assim, gerar recomendações personalizadas. Por fim, a terceira é a Camada de Ação Personalizada. Nesta etapa, o sistema não apenas gera recomendações, mas as transforma em ações customizadas, como o envio de notificações quando o usuário entra em uma região próxima a um POI com alta compatibilidade.

A integração dessas camadas oferece vantagens significativas, como a eficiência energética e a privacidade melhorada, pois a cerca geográfica processa dados de localização apenas em pontos de interesse específicos. Além disso, a combinação garante uma personalização inteligente baseada em padrões reais de comportamento e a validação de presença, assegurando que os dados de treinamento reflitam visitas autênticas.

RESULTADOS E DISCUSSÃO

As Tabelas 1 e 2 demonstram os resultados quantitativos obtidos para precisão e *recall*, respectivamente.

TABELA 1. Resultados Quantitativos da Implementação do Adaptive-CF - Precisão

K	@5	@10	@20
Foursquare-Nova York	0.0416	0.0323	0.0235
Foursquare-Tokyo	0.0976	0.0718	0.0521

TABELA 2. Resultados Quantitativos da Implementação do Adaptive-CF - Recall

K	@5	@10	@20
Foursquare-Nova York	0.0199	0.0328	0.0474
Foursquare-Tokyo	0.0299	0.0441	0.0623

Os resultados obtidos apresentam uma discrepância significativa quando comparados com o modelo proposto por Yueyu et al [3]. As Figuras 1a, 1b, 2a e 2b ilustram graficamente essas diferenças, onde o eixo Y representa a proporção de acertos (valores entre 0 e 1) e o eixo X representa o valor de k (número de recomendações).

FIGURA 1a: Comparação da Precisão - Nova York

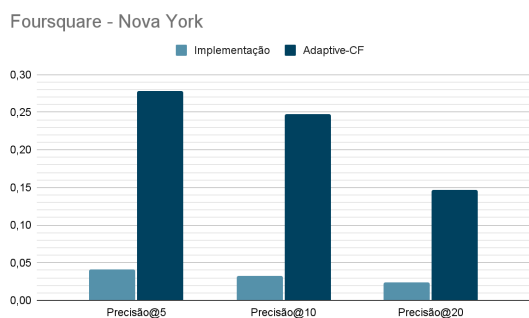
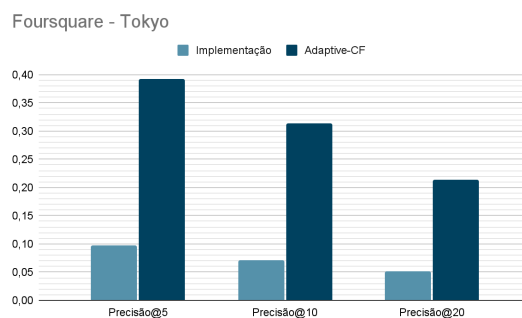


FIGURA 1b: Comparação da Precisão - Tokyo



Fonte: Autor (2025, p.4)

FIGURA 2a: Comparação do Recall – Nova York

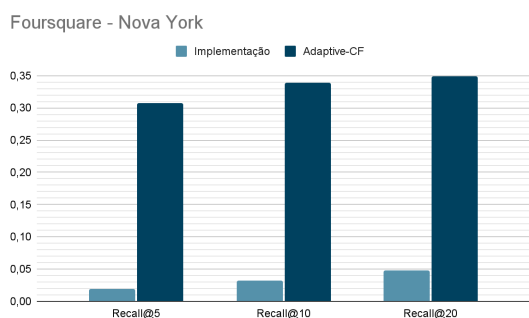
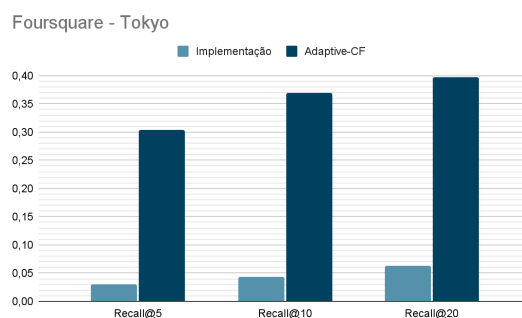


FIGURA 2b: Comparação do Recall – Tokyo



Fonte: Autor (2025, p.4)

Esta discrepância substancial motivou uma investigação aprofundada dos fatores causadores, revelando problemas fundamentais que podem ser categorizados em duas dimensões principais: 1) qualidade e consistência dos datasets, e 2) lacunas na documentação algorítmica.

Base de Dados: Limitações e Qualidade de Dados

As bases de dados utilizadas nesta implementação foram obtidas da mesma fonte citada no trabalho de referência. Contudo, ao realizar análise exploratória detalhada dos dados, foram identificadas inconsistências críticas em relação às estatísticas reportadas no estudo original.

Inicialmente, observou-se que as características estatísticas da base de dados de Tóquio obtidas neste estudo apresentavam maior semelhança com as características da base de dados de Nova Iorque reportadas na referência, e vice-versa. Esta evidência sugere uma possível inversão acidental das bases de dados durante a produção do trabalho de referência. Contudo, mesmo considerando essa possível troca, permanecem diferenças substanciais nos dados.

As Tabelas 4 e 5 apresentam comparações detalhadas entre as estatísticas das bases de dados utilizadas neste estudo e aquelas reportadas no trabalho de referência.

TABELA 4. Comparação entre estatísticas da base de dados Foursquare-Nova York

Datasets	Original	Pós pré-processamento	Referência
Quantidade de usuários	1083	1083	2293
Quantidade de POI	38333	5135	6017
Total de check-ins	227428	147729	307548

TABELA 5. Comparação entre estatísticas da base de dados Foursquare-Tokyo

Datasets	Original	Pós pré-processamento	Referência
Quantidade de usuários	2293	2293	1080
Quantidade de POI	61858	7873	3138
Total de check-ins	573703	447076	47753

Embora ambas bases de dados sejam altamente esparsas (característica comum em sistemas de recomendação), a discrepância nos dados pode afetar a qualidade da base de dados, impactando a distribuição e densidade relativa dos dados e, conseqüentemente, a capacidade do algoritmo de:

1. Identificar Padrões de Similaridade: Matrizes esparsas com distribuição diferente podem resultar em menor sobreposição de POI visitados entre usuários, dificultando o cálculo de similaridade.
2. Encontrar Vizinhos Relevantes: A etapa de seleção adaptativa de vizinhos depende de encontrar usuários com padrões de visitaçao similares. Dados com distribuição distinta reduzem o número de vizinhos úteis encontrados.
3. Gerar Recomendações Precisas: Com menos vizinhos similares e padrões menos claros, as predições de rating tornam-se menos confiáveis, impactando diretamente Precision@k e Recall@k.

Algoritmo: Dificuldades de Reprodutibilidade

Durante a implementação do algoritmo *Adaptive-CF*, foram identificadas lacunas críticas na documentação de parâmetros e procedimentos. Especificamente, o valor exato do limiar de frequência (α) não definido, que é crítico para determinar POIs frequentemente visitados e definir a região de atividade do usuário. Os critérios de convergência do LFM (Fatoração de Matriz), como o número de iterações, a taxa de erro mínima ou o delta de convergência, não foram detalhados. Houve também falta de clareza sobre os procedimentos de normalização, além da ausência de documentação explícita sobre o valor ou método de cálculo do limiar adaptativo de similaridade para a seleção de vizinhos. Essas lacunas resultaram em três problemas principais: a necessidade de inferência por tentativa e erro, que consumiu tempo e recursos computacionais significativos; a incerteza sobre a correção, pois a falta de especificação clara impede a garantia de que a implementação reproduz fielmente o algoritmo original; e a alta complexidade computacional, visto que, sem otimizações documentadas ou valores de referência para convergência, o tempo de execução se tornou proibitivo para a experimentação iterativa de parâmetros.

Conclusão

Por meio dos resultados obtidos foram identificadas duas barreiras reais à reprodutibilidade. A primeira delas é a degradação temporal das bases de dados públicas, a qual afeta a integridade e consistência dos dados, e a segunda é as lacunas na documentação dos algoritmos, dificultando a implementação fidedigna do modelo proposto [6, 7].

Dessa forma, os desafios documentados neste trabalho revelam uma necessidade urgente de mudança nas práticas de publicação científica. Especificamente, recomenda-se a adoção de protocolos rigorosos de documentação de algoritmos e gestão de base de dados. Tais protocolos devem incluir a especificação explícita de hiperparâmetros, o uso de pseudocódigo detalhado e a disponibilização do código-fonte. Em relação aos dados, é crucial o versionamento e a garantia da integridade via *hashes* e repositórios permanentes (DOI), a criação de metadados detalhados sobre estatísticas e pré-processamento, e a implementação de um controle de qualidade para validar a consistência [8].

Os desafios identificados também têm implicações diretas para o desenvolvimento de sistemas de cerca geográfica inteligentes. É fundamental a validação exaustiva dos sistemas de aprendizado de máquina antes da implantação, pois discrepâncias entre resultados experimentais e o desempenho real podem levar à má experiência do usuário e perda de confiança, sendo preferível a adoção de soluções mais simples e confiáveis. Além disso, este estudo reforça a necessidade de estabelecer expectativas realistas entre stakeholders e gestores, baseadas em validações empíricas rigorosas e provas de conceito (POC) em ambientes controlados, dada a dificuldade de reproduzir resultados reportados na literatura.

CONSIDERAÇÕES FINAIS

Este trabalho demonstra que o caminho entre uma ideia promissora na literatura científica e um sistema funcional em produção é repleto de desafios práticos que frequentemente são subestimados ou inadequadamente documentados. A reprodutibilidade científica não é apenas uma questão de rigor metodológico, mas uma necessidade prática para o avanço real da área.

Espera-se que esta pesquisa contribua para uma compreensão mais madura e realista dos desafios na implementação de sistemas de cerca geográfica inteligentes, estimulando a adoção de práticas mais rigorosas de documentação e compartilhamento de recursos científicos.

Futuros estudos sobre a integração entre cerca geográfica e aprendizado de máquina devem considerar a utilização de base de dados mais ricas e contextualizadas, incorporando variáveis como tempo de permanência, avaliações qualitativas de POI e fatores temporais e demográficos, explorar técnicas avançadas de aprendizado profundo, como redes neurais recorrentes (RNN/LSTM/GRU) para sequências temporais e *Graph Neural Networks* (GNN) para modelar relações complexas, além de aplicar *transfer learning* e *embeddings* para lidar com a esparsidade.

AGRADECIMENTOS

Os autores agradecem ao Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), por ter celebrado o Memorandum of Understanding (MOU) entre IFSP e GN Audio A/S, e ao Centro de Pesquisas CEPIN-CIS pela viabilização do estudo, bem como pelo suporte técnico e disponibilização de recursos essenciais para a realização deste trabalho, que viabilizou a realização deste projeto.

REFERÊNCIAS

- [1] ZHAO, P. et al. **Where to Go Next: A Spatio-Temporal Gated Network for Next POI Recommendation**. In: The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19), 2019. Disponível em: <https://cdn.aaai.org/ojs/4537/4537-13-7576-1-10-20190706.pdf>. Acesso em: 28 set. 2025.
- [2] SHEVCHENKO, Y.; REIPS, U.-D. **Geofencing in location-based behavioral research: Methodology, challenges, and implementation**. Behavior Research Methods, 2023. Disponível em: <https://doi.org/10.3758/s13428-023-02213-2>. Acesso em: 28 set. 2025.
- [3] YUEYU W.; LEYAN C.; JIANI C. **An Adaptive Collaborative Filtering Based Approach for Point-of-Interest Recommendations**. EAI, 2022. Disponível em: <http://dx.doi.org/10.4108/eai.17-6-2022.2322634>. Acesso em: 29 set. 2025.
- [4] CHEKOL, A. G.; FUFA, M. S. **A survey on next location prediction techniques, applications, and challenges**. J Wireless Com Network, v. 2022, n. 29, 2022. Disponível em: <https://doi.org/10.1186/s13638-022-02114-6>. Acesso em: 28 set. 2025.
- [5] Yang et al. **Modeling User Activity Preference by Leveraging User Spatial Temporal Characteristics in LBSNs**. IEEE Transactions on Systems, Man, and Cybernetics: Systems, v. 45, n. 1, 2015. Disponível em: <https://doi.org/10.1109/TSMC.2014.2327053>. Acesso em: 29 set. 2025.
- [6] Werner, L., Layaïda, N., Genevès, P., Euzenat, J., & Graux, D. **Reproduce, Replicate, Reevaluate. The Long but Safe Way to Extend Machine Learning Methods**. Proceedings of the AAAI Conference on Artificial Intelligence, 2024. Disponível em: <https://doi.org/10.1609/aaai.v38i14.29515>. Acesso em 9 nov 2025.
- [7] Semmelrock, Harald & Ross-Hellauer, Tony & Kopeinik, Simone & Theiler, Dieter & Haberl, Armin & Thalmann, Stefan & Kowald, Dominik. **Reproducibility in Machine Learning-based Research: Overview, Barriers and Drivers**. AI Magazine 46: e70002, 2025. Disponível em: <https://doi.org/10.1002/aaai.70002>. Acesso em: 9 nov 2025.
- [8] Alsudais, Abdulkareem. **Incorrect Data in the Widely Used Inside Airbnb Dataset**. Decision Support Systems, 2020. Disponível em: https://www.researchgate.net/publication/342763373_Incorrect_Data_in_the_Widely_Used_Inside_Airbnb_Dataset. Acesso em: 9 nov 2025