



**UNIVERSIDADE
FEDERAL RURAL
DE PERNAMBUCO**

Guilherme Melo da Silva

Avaliação entre algoritmos de filtragem colaborativa baseada em vizinhança e transferência de conhecimento para CD-CARS

Recife

2019

Guilherme Melo da Silva

**Avaliação entre algoritmos de filtragem colaborativa
baseada em vizinhança e transferência de conhecimento
para CD-CARS**

Monografia apresentada ao Curso de Bacharelado em Ciência da Computação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Universidade Federal Rural de Pernambuco – UFRPE

Departamento de Computação

Curso de Bacharelado em Ciência da Computação

Orientador: Prof. Dr. Douglas Vêras e Silva

Recife

2019

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas da UFRPE
Biblioteca Central, Recife-PE, Brasil

S586a Silva, Guilherme Melo da
Avaliação entre algoritmos de filtragem colaborativa baseada em
vizinhança e transferência de conhecimento para CD-CARS
/ Guilherme Melo da Silva. – 2019.
60 f. : il.

Orientador: Douglas Vêras e Silva.

Trabalho de Conclusão de Curso (Graduação) – Universidade
Federal Rural de Pernambuco, Departamento de Ciência da
Computação, Recife, BR-PE, 2019.

Inclui referências.

1. Sistemas de Recomendação
2. Filtragem Colaborativa
3. Técnicas de Recomendação Sensíveis ao Contexto
4. CD-CARS
5. Tracer I. Silva, Douglas Vêras e, orient. II. Título

CDD 004

“A persistência é o caminho do êxito.”
(Charles Chaplin)

Agradecimentos

Agradeço à minha família, em especial aos meus pais Israel Luiz e Valquíria de Melo, por sempre estarem ao meu lado nos momentos felizes e difíceis da vida, pelos conselhos e direcionamentos em meu crescimento pessoal e profissional, enfim, por sempre me ajudarem a crescer na vida com educação e respeito ao próximo.

Aos meus professores pelos ensinamentos e experiências compartilhadas, em especial ao meu orientador Prof. Douglas Vêras, por confiar no meu trabalho, por todas as oportunidades e ensinamentos oferecidos, proporcionando meu crescimento profissional. Da mesma forma, agradeço à Prof. Érica Sousa, pela ajuda oferecida no projeto de iniciação científica da universidade.

Aos meus amigos, pelo apoio e fraternidade durante toda esta jornada, especialmente aos da turma de Bacharelado em Ciência da Computação de 2014.1. Agradeço a Universidade Federal Rural de Pernambuco, pelos excelentes serviços prestados a sociedade, formando não somente bons profissionais, mas cidadãos de respeito. Agradeço também a todos os funcionários da universidade, que zelam e fazem da universidade um local familiar e de crescimento humano.

Resumo

A realização de recomendações em cenários com a falta de preferências expressas por usuários é uma importante limitação para os Sistemas de Recomendação (SR). Devido a este problema pesquisas envolvendo SR de domínio cruzado (SRDC) vêm ganhando relevância, onde a filtragem colaborativa baseada (FC) é uma das técnicas mais exploradas nesta área. O sistema CD-CARS mostra que o uso de informações contextuais, disponíveis nas preferências dos usuários, pode otimizar algoritmos de FC baseada em vizinhança, técnica bastante difundida em FC para multidomínios. Embora apresentem recomendações precisas, alguns algoritmos de FC baseada em vizinhança, como o utilizado no CD-CARS, têm a limitação do uso de multidomínios apenas na ocorrência da sobreposição de usuários entre os domínios, cenário não trivial em bases de dados reais. Este trabalho apresenta uma análise comparativa entre diferentes algoritmos de recomendação envolvendo técnicas de filtragem colaborativa. Os algoritmos *NNUserNgb-transClosure* do CD-CARS (FC baseada em vizinhança) e *Tracer* (FC baseada em transferências de conhecimento), foram utilizados como base para os algoritmos de recomendação. Nos experimentos, os algoritmos de FC foram integrados às técnicas sensíveis ao contexto, abordadas no CD-CARS: Pré-Filtragem e Pós-Filtragem Contextual, sendo aplicados sobre dois conjuntos de dados, formados por dois domínios auxiliares e um alvo, com e sem sobreposição de usuários entre os domínios. As métricas de desempenho MAE e RMSE foram utilizadas para a avaliação dos algoritmos. Os resultados dos experimentos mostraram que o algoritmo *Tracer* apresentou melhores resultados, em relação ao algoritmo *NNUserNgb-transClosure*, em todos os experimentos envolvendo o cenário sem a sobreposição de usuários, com e sem o uso da Pré-Filtragem ou Pós-Filtragem Contextual.

Palavras-chave: Sistemas de Recomendação, Filtragem Colaborativa, Técnicas de Recomendação Sensíveis ao Contexto, CD-CARS, TRACER.

Abstract

Recommendations in scenarios with the lack of preferences expressed by users is an important limitation for Recommendation Systems (RS). Due to this problem, cross-domain RS (CDRS) searches have gained relevance, where collaborative filtering (CF) is one of the most exploited techniques in this area. The CD-CARS system shows that the use of contextual information, available in user preferences, can optimize CF neighborhood-based algorithms, a technique widely used in multidomain CF. Although they provide accurate recommendations, some neighborhood-based algorithms such as the one used in the CD-CARS have the limitation of the use of multi-domains only in the occurrence of user overlap between domains, a non-trivial scenario in real databases. This work presents a comparative analysis of different recommendation algorithms involving collaborative filtering techniques. The CD-CARS' NNUserNgbr-transClosure (CF neighborhood-based) and Tracer (CF transfer learning-based) algorithms, were used as the basis for the recommendation algorithms. In the experiments, the CF algorithms were integrated into the context-aware techniques, addressed in the CD-CARS: Contextual Pre-Filtering and Post-Filtering, being applied on two data sets, formed by two auxiliary domains and one target, with and without overlap between domains. The MAE and RMSE performance metrics were used to evaluate the algorithms. The results of the experiments showed that the Tracer algorithm presented better results concerning the NNUserNgbr-transClosure algorithm in all experiment scenarios without user overlap, with and without the use of the Contextual Pre-Filtering or Post-Filtering.

Keywords: Recommender Systems, Collaborative Filtering, Context-Aware Recommender Techniques, CD-CARS, TRACER.

Lista de ilustrações

Figura 1 – Domínio com escassez de preferências.	14
Figura 2 – Limitação do algoritmo <i>NNUserNgbr-transClosure</i> em multidomínios sem sobreposição de usuários.	15
Figura 3 – Sistemas de Recomendação de Domínio Cruzado.	18
Figura 4 – Definições de domínio.	19
Figura 5 – Cenários de recomendações em multidomínios.	19
Figura 6 – Filtragem Colaborativa Baseada em Usuários sobre Multidomínios.	20
Figura 7 – Agregação de Conhecimento.	21
Figura 8 – Filtragem Colaborativa <i>NNUserNgbr</i>	22
Figura 9 – Transferência de Conhecimento.	23
Figura 10 – Processo de clusterização e transferência de conhecimento do <i>Codebook-Transfer</i>	23
Figura 11 – Algoritmo <i>Tracer</i>	24
Figura 12 – Pré-Filtragem (a) e Pós-Filtragem (b) Contextual.	26
Figura 13 – Metodologia.	30
Figura 14 – Estrutura dos Dados.	31
Figura 15 – <i>NNUserNgbr-transClosure</i> com Pré-Filtragem Contextual.	34
Figura 16 – <i>Tracer</i> com Pré-Filtragem Contextual.	35
Figura 17 – <i>NNUserNgbr-transClosure</i> com Pós-Filtragem Contextual.	36
Figura 18 – <i>Tracer</i> com Pós-Filtragem Contextual.	37
Figura 19 – Funcionamento da Pré-Filtragem Contextual.	38
Figura 20 – Funcionamento da Pós-Filtragem Contextual.	39
Figura 21 – Distância Euclidiana.	39
Figura 22 – Métricas de Desempenho.	42
Figura 23 – Resultados dos algoritmos de Filtragem Colaborativa, MAE (a) e RMSE (b).	44
Figura 24 – Resultados (MAE e RMSE) boxplot, 20% e 0% de sobreposição.	44
Figura 25 – Pré-F. com algoritmos de Filtragem Colaborativa, MAE (a) e RMSE (b) (dimensão contextual temporal).	45
Figura 26 – Pós-F. com algoritmos de Filtragem Colaborativa, MAE (a) e RMSE (b) (dimensão contextual temporal).	46
Figura 27 – Resultados (MAE e RMSE) boxplot, 0% sobreposição (dimensão contextual temporal).	46
Figura 28 – Resultados (MAE e RMSE) boxplot, 20% sobreposição (dimensão contextual temporal).	47
Figura 29 – Pré-F. com algoritmos de Filtragem Colaborativa, MAE (a) e RMSE (b) (dimensão contextual de companhia).	48

Figura 30 – Pós-F. com algoritmos de Filtragem Colaborativa, MAE (a) e RMSE (b) (dimensão contextual de companhia).	49
Figura 31 – Resultados (MAE e RMSE) boxplot, 0% sobreposição (dimensão contextual de companhia).	49
Figura 32 – Resultados (MAE e RMSE) boxplot, 20% sobreposição (dimensão contextual de companhia).	50
Figura 33 – Pré-F. com algoritmos de Filtragem Colaborativa, MAE (a) e RMSE (b) (dimensão contextual de localização).	51
Figura 34 – Pós-F. com algoritmos de Filtragem Colaborativa, MAE (a) e RMSE (b) (dimensão contextual de localização).	52
Figura 35 – Resultados (MAE e RMSE) boxplot, 0% sobreposição (dimensão contextual de localização).	52
Figura 36 – Resultados (MAE e RMSE) boxplot, 20% sobreposição (dimensão contextual de localização).	53

Lista de tabelas

Tabela 1 – Classificação dos Trabalhos Relacionados.	29
Tabela 2 – Conjunto de Dados 1 (com sobreposição de usuários).	32
Tabela 3 – Nível de sobreposição de usuários (conjunto de dados 1).	32
Tabela 4 – Conjunto de Dados 2 (sem sobreposição de usuários).	33
Tabela 5 – Nível de sobreposição de usuários (conjunto de dados 2).	33
Tabela 6 – Cenários dos Experimentos.	41
Tabela 7 – Resultados (MAE / RMSE) com desvio padrão (dp) dos algoritmos de filtra- gem colaborativa.	43
Tabela 8 – Resultados (MAE / RMSE) com desvio padrão (dp), Pré-F. e Pós-F. com algoritmos de filtragem colaborativa (dimensão contextual temporal).	45
Tabela 9 – Resultados (MAE / RMSE) com desvio padrão (dp), Pré-F. e Pós-F. com algoritmos de filtragem colaborativa (dimensão contextual de companhia).	48
Tabela 10 – Resultados (MAE / RMSE) com desvio padrão (dp), Pré-F. e Pós-F. com algoritmos de filtragem colaborativa (dimensão contextual de localização).	51

Lista de abreviaturas e siglas

SR	Sistemas Recomendação
SRDC	Sistemas Recomendação de Domínio Cruzado
FC	Filtragem Colaborativa
FCV	Filtragem Colaborativa Baseada em Vizinhança
FCTC	Filtragem Colaborativa Baseada em Transferência de Conhecimento
TRSC	Técnicas de Recomendação Sensíveis ao Contexto
CD-CARS	Sistemas de Recomendação de Domínio Cruzado Sensíveis ao Contexto
AC	Agregação de Conhecimento
TC	Transferência de Conhecimento

Sumário

	Lista de ilustrações	6
1	INTRODUÇÃO	12
1.1	Problema de Pesquisa	14
1.2	Objetivos	16
1.2.1	Objetivo Geral	16
1.2.2	Objetivos Específicos	16
1.3	Organização do Trabalho	16
2	REFERENCIAL TEÓRICO	17
2.1	Sistemas de Recomendação	17
2.2	Sistemas de Recomendação de Domínio Cruzado	17
2.2.1	Definições de Domínio	18
2.2.2	Cenários de Recomendações em Multidomínios	19
2.3	Filtragem Colaborativa	20
2.3.1	Filtragem Colaborativa Baseada em Vizinhança	20
2.3.1.1	Algoritmo <i>NNUserNgbr-transClosure</i>	21
2.3.2	Filtragem Colaborativa Baseada em Transferência de Conhecimento	22
2.3.2.1	Algoritmo <i>TRACER</i>	24
2.4	Técnicas de Recomendação Sensíveis ao Contexto	25
2.4.1	Pré-Filtragem Contextual	25
2.4.2	Pós-Filtragem Contextual	25
2.5	CD-CARS	26
3	TRABALHOS RELACIONADOS	27
4	METODOLOGIA	30
4.1	Base de Dados	30
4.1.1	Descrição da Base de Dados	30
4.1.2	Preparação da Base de Dados	31
4.2	Algoritmos de Recomendação	33
4.2.1	Filtragem Colaborativa com Pré-Filtragem Contextual	34
4.2.2	Filtragem Colaborativa com Pós-Filtragem Contextual	35
4.2.3	Funcionamento dos Algoritmos Sensíveis ao Contexto	37
4.2.4	Configurações dos Algoritmos de Filtragem Colaborativa	39
4.3	Descrição dos Experimentos	40

4.3.1	Métricas de Desempenho	42
5	RESULTADOS	43
5.1	Resultados dos Algoritmos de Filtragem Colaborativa	43
5.2	Resultados na Dimensão Contextual Temporal	44
5.3	Resultados Dimensão Contextual de Companhia	47
5.4	Resultados na Dimensão Contextual de Localização	50
5.5	Discussão	53
6	CONCLUSÃO	55
6.1	Contribuições	56
6.2	Limitações	56
6.3	Trabalhos Futuros	57
	REFERÊNCIAS	58

1 Introdução

A grande quantidade e variedade de informações disponíveis hoje na Web, vem sobrecarregando a capacidade dos usuários em pesquisar e encontrar produtos ou serviços que satisfaçam suas necessidades, com o melhor custo-benefício. Devido a este problema os Sistemas de Recomendação (SR) têm ganhado cada vez mais importância, pela capacidade de sugerir os itens mais relevantes aos usuários (KUMAR; SHARMA, 2016; CANO; MORISIO, 2019).

Os SR têm importância para o mercado digital, gerando um grande retorno financeiro para as empresas deste setor, por exemplo, o mecanismo de recomendação da *Amazon*¹, gera cerca de 35% da receita da empresa (FORBES, 2018). Embora apresentem bons resultados a maioria dos sistemas de recomendações tradicionais apenas recomendam itens relacionados a um domínio específico (CANTADOR et al., 2015). A *Netflix*², por exemplo, baseia suas recomendações em filmes ou séries de TV. As recomendações baseadas em um único domínio podem ser imprecisas, quando há escassez de preferências por parte dos usuários, geralmente expressas por avaliações de satisfação (RICCI; ROKACH; SHAPIRA, 2015).

A escassez de informações relacionadas às preferências de usuários é atualmente uma das maiores limitações para a realização de recomendações (MURAD; IBRAHIM, 2017; CANO; MORISIO, 2019). Com o objetivo de solucionar este problema, alguns trabalhos na literatura (VERAS, 2016; ZHUANG et al., 2018; ZHANG et al., 2018), vêm adotando Sistemas de Recomendações de Domínio Cruzado (SRDC), onde preferências de usuários são exploradas em domínios auxiliares, para otimização de recomendações num domínio alvo.

A Filtragem Colaborativa Baseada em Vizinhaça (FCV) é uma das técnicas mais bem-sucedidas no desenvolvimento de SRDC, devido a fácil implementação e qualidade (KOREN; BELL, 2015; RAGHUWANSHI; PATERIYA, 2018). Esta técnica explora semelhanças entre usuários, segundo suas preferências em diferentes domínios, para a geração de recomendações. Por exemplo, a FCV considera que usuários que avaliaram bem algum livro; poderiam avaliar, semelhantemente, uma música ou filme (PORTUGAL; ALENCAR; COWAN, 2018).

Alguns trabalhos na literatura, abordam a integração de técnicas de recomendação sensíveis ao contexto (TRSC), como auxílio aos algoritmos de FC no processo de recomendação (VERAS et al., 2015; BEDI; BEDI, 2018; ABDULKAREM; ABOZAID; SOLIMAN, 2019). As TRSC possibilitam a seleção das preferências mais relevantes para a aplicação dos algoritmos de FC, segundo informações contextuais adquiridas no momento da avaliação dos itens, como a localização, ou companhia dos usuários (VERAS et al., 2015). Por exemplo, as preferências poderiam ser selecionadas com base nas avaliações realizadas no Brasil, no domingo ou na com-

¹ Amazon: <https://www.amazon.com>

² Netflix: <https://www.netflix.com>

panhia dos filhos do usuário. As TRSC também podem ser integradas aos SRDC como mostrado no sistema de recomendação CD-CARS proposto por [Veras \(2016\)](#), no qual utiliza TRSC para a otimização das recomendações geradas por algoritmos de FCV.

Embora seja uma boa técnica em relação a escassez de avaliações, alguns algoritmos de filtragem colaborativa baseada em vizinhança, apresentam uma limitação importante no processo de recomendação em multidomínios. Estes algoritmos são ineficientes em explorar informações de domínios auxiliares, sem a ocorrência da sobreposição de usuários entre os multidomínios, ou seja, os mesmos usuários devem avaliar itens tanto no domínio alvo da recomendação, quanto nos domínios auxiliares, um cenário não trivial em bases de dados reais ([FERNANDEZ et al., 2012](#); [CANTADOR et al., 2015](#); [VERAS, 2016](#)).

A Filtragem Colaborativa Baseada em Transferência de Conhecimento (FCTC), pode ser uma solução para o problema apresentado. A técnica de FCTC é um dos tópicos de pesquisas envolvendo SRDC ([PAN, 2016](#); [PORTUGAL](#); [ALENCAR](#); [COWAN, 2018](#)). Além de apresentar uma boa acurácia na realização de recomendações com escassez de preferências, alguns algoritmos de FCTC não necessitam da sobreposição de usuários para utilização de informações de domínios auxiliares no processo de recomendação, vantagem em relação a outras técnicas de filtragem colaborativa ([ZHANG et al., 2017](#); [ZHU et al., 2017](#); [ZHUANG et al., 2018](#)).

Este trabalho analisa o desempenho da integração de técnicas sensíveis ao contexto, aos algoritmos de filtragem colaborativa: baseada em vizinhança e transferência de conhecimento, quando aplicados à multidomínios sem a sobreposição de usuários. A realização dos experimentos tem o objetivo de comparar os algoritmos de recomendação, verificando se a FCTC pode ser mais eficiente que a FCV, com o auxílio de informações contextuais, em cenários com escassez de preferências e diferentes níveis sobreposição de usuários entre os domínios.

1.1 Problema de Pesquisa

Obter preferências de usuários não é uma tarefa trivial, muitas vezes os usuários possuem um tempo limitado, ou não tem interesse em explorar a grande variedade de opções disponíveis na Web. A escassez de preferências, expressas por usuários, é uma importante limitação para os sistemas de recomendação (RICCI; ROKACH; SHAPIRA, 2015; HUSSIEN; HELMY; HASAN, 2019). A Figura 1, mostra um exemplo de conjunto de dados com escassez de preferências no domínio relacionado a músicas, onde as avaliações podem variar de 1 a 5 e as interrogações representam itens não avaliados.

Figura 1 – Domínio com escassez de preferências.



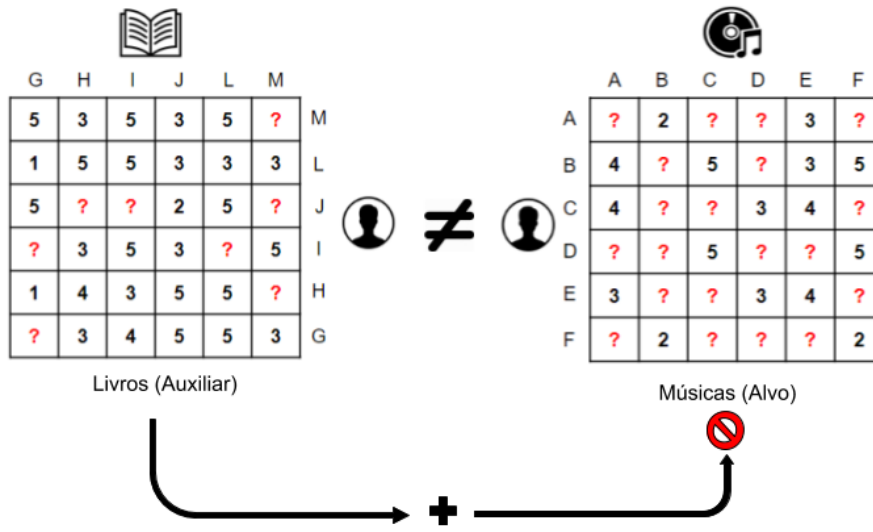

	A	B	C	D	E	F	G
A	?	2	?	?	1	?	?
B	5	?	5	?	3	3	?
C	4	?	?	1	4	?	3
D	?	?	5	?	?	5	?
F	1	?	?	5	5	?	?
G	?	3	?	?	?	3	1

Fonte: Li, Yang e Xue (2009)

A tese desenvolvida por Veras (2016), apresenta uma proposta de sistema de recomendação de domínio cruzado, sensível ao contexto (CD-CARS), para a resolução do problema relacionado a escassez de preferências. O CD-CARS integra algoritmos sensíveis ao contexto: Pré-Filtragem e Pós-Filtragem Contextual, ao algoritmo de filtragem colaborativa baseada em vizinhança (FCV), *NNUserNgbr-transClosure* (Seção 2.3.1.1). Os algoritmos de Pré-Filtragem e Pós-Filtragem Contextual têm a função de auxiliar o algoritmo de FCV, selecionando as preferências segundo o contexto das avaliações. O algoritmo *NNUserNgbr-transClosure* gera predições de avaliações, segundo a similaridade de preferências entre usuários, expressas em multidomínios, além de integrar a transitividade na construção das vizinhanças.

Embora o CD-CARS, proposto por Veras (2016), apresente bons resultados; também possui uma importante limitação. O algoritmo de FCV utilizado pelo CD-CARS (*NNUserNgbr-transClosure*), tem a limitação do uso de domínios auxiliares, apenas na ocorrência da sobreposição de usuários entre os domínios alvo e auxiliares. É importante também citar que o nível de sobreposição de usuários, pode afetar na contribuição dos domínios auxiliares no processo de recomendação, melhorando ou não as predições geradas pelo CD-CARS (VERAS, 2016). A Figura 2 mostra o algoritmo *NNUserNgbr-transClosure* aplicado a multidomínios sem sobreposição de usuários, cenário que impossibilita o uso de domínios auxiliares nas recomendações.

Figura 2 – Limitação do algoritmo *NNUserNgbr-transClosure* em multidomínios sem sobreposição de usuários.



Fonte: Próprio autor

Para a resolução do problema de pesquisa, exploramos o algoritmo *Tracer* (Seção 2.3.2.1), filtragem colaborativa baseada em transferência de conhecimento (FCTC), proposta por [Zhuang et al. \(2018\)](#). O algoritmo *Tracer* tem vantagem sobre o algoritmo *NNUserNgbr-transClosure* (FCV), por possibilitar a extração e transferência de padrões, referente as preferências dos usuários, dos domínios auxiliares para o domínio alvo da recomendação, não necessitando da sobreposição de usuários para a geração de recomendações baseadas em multidomínios ([ZHUANG et al., 2018](#)).

Com base no sistema CD-CARS, este trabalho apresenta um estudo comparativo entre os algoritmos de FCTC (*Tracer*) e FCV (*NNUserNgbr-transClosure*), integrados aos algoritmos sensíveis ao contexto: Pré-Filtragem e Pós-Filtragem Contextual. O estudo avalia a aplicação destes algoritmos de recomendação sobre conjuntos de dados, formados por multidomínios com e sem a sobreposição de usuários; sendo este estudo uma das linhas de pesquisa futuras propostas por [Veras \(2016\)](#). Neste trabalho foram construídos conjuntos de dados, baseados no conjunto de avaliações pertencente a *Amazon*, com destaque para a disponibilização de dois domínios auxiliares na realização dos experimentos.

1.2 Objetivos

1.2.1 Objetivo Geral

Avaliar o desempenho entre algoritmos de filtragem colaborativa baseada em vizinhança e transferência de conhecimento, com o auxílio de técnicas sensíveis ao contexto, na realização de recomendações em multidomínios com e sem sobreposição de usuários.

1.2.2 Objetivos Específicos

1. Otimizar as recomendações do sistema CD-CARS, com a integração do algoritmo *Tracer*.
2. Disponibilizar nova base de dados, com diferentes níveis de sobreposição de usuários e composta por dois domínios auxiliares.
3. Disponibilizar o código-fonte do trabalho, para pesquisas futuras.

1.3 Organização do Trabalho

Este trabalho está organizado com a seguinte estrutura: o Capítulo 2 apresenta o referencial teórico, com os fundamentos básicos de sistemas de recomendação, os possíveis cenários de recomendação em multidomínios, os tipos de filtragem colaborativa e as técnicas sensíveis ao contexto, explorados neste trabalho.

O Capítulo 3 contém a revisão da literatura, com trabalhos relacionados ao tema abordado. Estes trabalhos foram essenciais para o estudo das técnicas de recomendação utilizadas neste trabalho, para a definição da metodologia e métricas de desempenho.

O Capítulo 4 descreve a metodologia seguida para a realização deste trabalho. Como a descrição e construção da base de dados, os algoritmos de recomendação utilizados, a descrição dos experimentos e as métricas de desempenho para os experimentos.

No Capítulo 5 são apresentados os resultados dos experimentos. Por fim o Capítulo 6 contém a conclusão deste trabalho, com uma visão para trabalhos futuros.

2 Referencial Teórico

Este capítulo apresenta o referencial teórico do trabalho, com os conceitos básicos de sistemas de recomendação (Seções 2.1 e 2.2), as definições de domínio (Seção 2.2.1) e cenários de recomendação em multidomínios (Seção 2.2.2), os algoritmos de filtragem colaborativa (Seção 2.3) e técnicas sensíveis ao contexto (Seção 2.4), utilizados no trabalho.

2.1 Sistemas de Recomendação

Os Sistemas de Recomendação (SR) têm a função de filtrar da grande quantidade de informações disponíveis na Web, os conteúdos mais relevantes e úteis para os usuários, apresentando uma pequena seleção de itens que mais se adequem aos seus gostos, interesses e prioridades (RICCI; ROKACH; SHAPIRA, 2015). Desde o surgimento da área de recomendação em meados de 1990, uma grande quantidade de pesquisas têm sido publicadas nesta área, em diversos tipos de aplicações relacionadas a filmes, livros, músicas, programas de TV, entre outros (KUMAR; SHARMA, 2016). Os SR também vêm recebendo grandes investimentos por parte do comércio eletrônico, sendo utilizados em aplicações comerciais de grandes empresas do setor, como a *Amazon*, *Netflix*, entre outros (CANO; MORISIO, 2019).

Um dos maiores desafios desta área é a realização de recomendações onde há escassez de preferências por parte dos usuários, ou seja, avaliações de satisfação oferecidas aos itens; informações indispensáveis para os SR (RICCI; ROKACH; SHAPIRA, 2015; HUSSIEN; HELMY; HASAN, 2019). Alguns trabalhos na literatura vêm adotando técnicas que tentam solucionar este problema, como o uso de multidomínios para realização das recomendações (MURAD; IBRAHIM, 2017; PORTUGAL; ALENCAR; COWAN, 2018).

2.2 Sistemas de Recomendação de Domínio Cruzado

A maioria dos sistemas de recomendação tradicionais apenas recomendam itens pertencentes a um único domínio (CANTADOR et al., 2015). O *Spotify*¹, por exemplo, foca apenas na recomendação de músicas, já o *Netflix* recomenda filmes e séries de TV aos seus usuários. Apesar de ser amplamente utilizada, esta abordagem pode apresentar recomendações imprecisas quando há escassez de preferências, por parte dos usuários (FERNANDEZ et al., 2012; KUMAR; SHARMA, 2016).

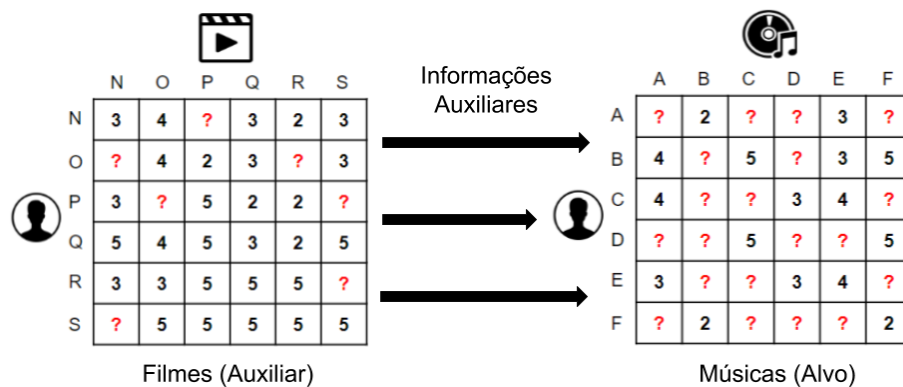
Por outro lado, grandes sites de comércio eletrônico, como a *Amazon* e o *Ebay*², utilizam sistemas de recomendação de domínio cruzado (SRDC), onde as recomendações são baseadas

¹ Spotify: <https://www.spotify.com>

² Ebay: <https://www.ebay.com>

em avaliações de itens que pertencem a múltiplos domínios, como: games, filmes e músicas. O uso de informações de diferentes domínios, possibilita a criação de perfis mais abrangentes para os usuários, gerando melhores recomendações (BEDI; BEDI, 2018; ZHANG et al., 2018). Ao invés de tratar cada domínio independentemente, os SRDC exploram informações em domínios auxiliares, com o objetivo de otimizar recomendações num domínio alvo, ou seja, é possível recomendar filmes aos usuários, baseado em suas preferências anteriores por livros ou games. Devido ao seu potencial em apresentar uma boa acurácia, mesmo com a escassez de preferências de usuários, os SRDC se tornaram um tópico de pesquisa emergente na área de sistemas de recomendação (CANTADOR et al., 2015; VERAS, 2016). A Figura 3, mostra o uso de informações através de um domínio auxiliar, por parte dos SRDC; onde as interrogações representam a falta de avaliações aos itens.

Figura 3 – Sistemas de Recomendação de Domínio Cruzado.



Fonte: Próprio autor com base em Cantador et al. (2015)

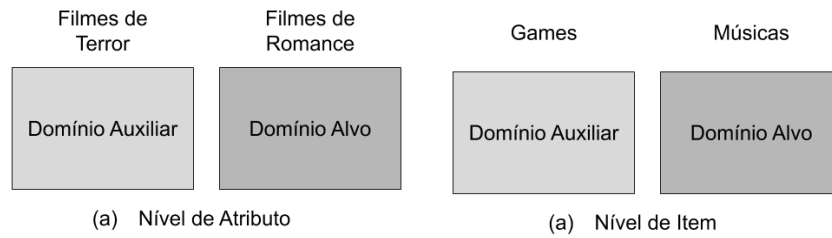
2.2.1 Definições de Domínio

Diferentes tipos de definições de domínio são encontradas na literatura. Algumas pesquisas consideram que livros e músicas não pertencem a um mesmo domínio, já outros trabalhos tratam diferentes estilos musicais (ex.: rock e pop), como domínios distintos. Dentre os diferentes tipos de definições podemos destacar algumas, segundo Cantador et al. (2015), são elas:

1. Nível de Atributo: Os itens recomendados apresentam o mesmo tipo, mas podem apresentar atributos com diferentes valores, como filmes de diferentes gêneros (ex.: terror, romance ou comédia).
2. Nível de Item: Os itens recomendados possuem diferentes tipos e atributos. Por exemplo, nesta definição, games e músicas pertencem a diferentes domínios.

Neste trabalho o domínio foi definido ao nível de item, onde foram utilizados domínios relacionados a livros, filmes e músicas. A Figura 4 detalha as duas diferentes definições de domínio, citadas anteriormente.

Figura 4 – Definições de domínio.



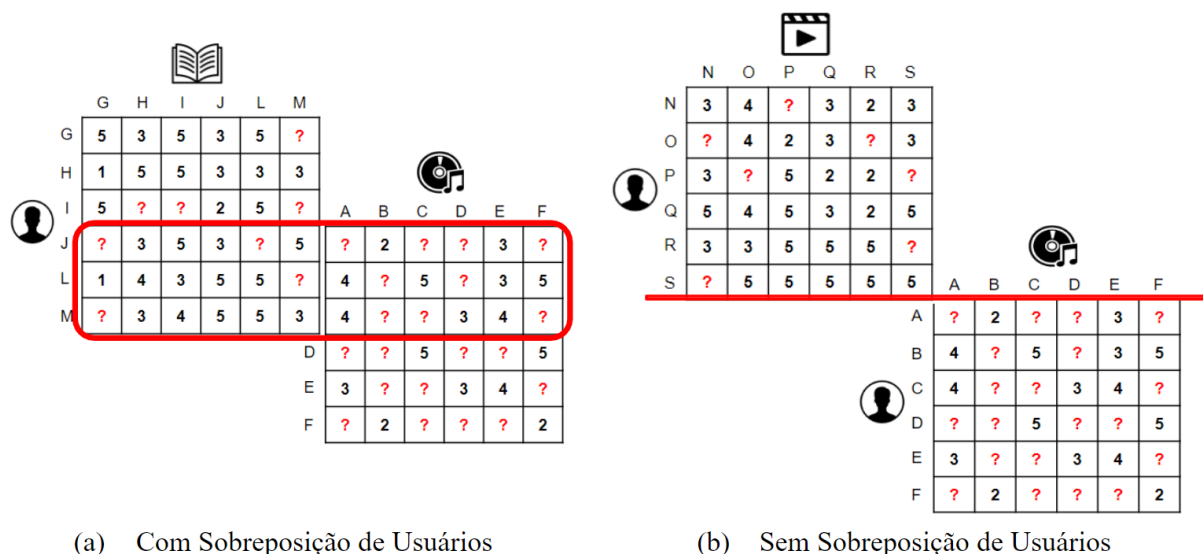
Fonte: Cantador et al. (2015)

2.2.2 Cenários de Recomendações em Multidomínios

As recomendações realizadas em multidomínios são baseadas em conjuntos de preferências de usuários, normalmente expressas por avaliações de itens. Os usuários podem avaliar itens de um único domínio ou itens de diferentes domínios, onde ocorre a sobreposição de usuários em diferentes níveis (ZHANG et al., 2018). Segundo Cremonesi, Tripodi e Turrin (2011), podemos ter dois cenários de recomendação, em relação a avaliações de usuários em multidomínios, são eles:

1. Com Sobreposição de Usuários: Quando um mesmo usuário apresenta avaliações em pelo menos dois domínios, por exemplo, neste caso poderiam existir comuns usuários que avaliaram filmes e livros numa mesma loja virtual (Figura 5a)
2. Sem Sobreposição de usuários: Quando não há avaliações de um mesmo usuário nos domínios utilizados para a recomendação, ou seja, cada usuário tem apenas preferências por itens de um mesmo domínio (Figura 5b).

Figura 5 – Cenários de recomendações em multidomínios.



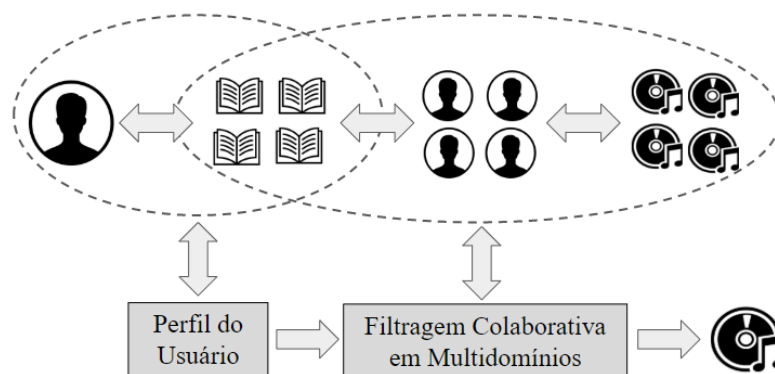
Fonte: Adaptado de Cremonesi, Tripodi e Turrin (2011)

2.3 Filtragem Colaborativa

Até o momento, a maioria das técnicas propostas para sistemas de recomendação, utilizam a Filtragem Colaborativa (FC) (CANO; MORISIO, 2019; PATEL et al., 2019), isto devido a sua fácil implementação, integração a diferentes domínios e sua qualidade que geralmente é superior a outras técnicas (KOREN; BELL, 2015). Esta técnica utiliza matrizes de avaliações de itens, expressas por usuários, para determinação de similaridades entre usuários ou itens. A FC explora apenas as avaliações dos itens, ignorando quaisquer informações relacionadas aos itens, como sua categoria ou público alvo (SHARMA; GOPALANI; MEENA, 2019). Esta técnica também não considera as informações contextuais das avaliações, como: localização ou dia da semana, no processo de recomendação (PATEL et al., 2019).

A Filtragem Colaborativa pode se classificada em FC baseada em usuários e FC baseada em itens. A FC baseada em usuários, gera recomendações com base na similaridade de preferências entre usuários, ou seja, considera que usuários com preferências similares no passado, também terão gostos semelhantes no futuro. A FC baseada em itens, calcula previsões baseado nos padrões de avaliações entre itens, ou seja, considera o conjunto de itens que o usuário avaliou, previamente, e computa quão similares são, em relação a um item alvo (SHARMA; GOPALANI; MEENA, 2019). Neste trabalho exploramos a FC baseada em usuários, para sistemas de recomendação de domínio cruzado, com foco em algoritmos de FC baseada em vizinhança e FC baseada em transferência de conhecimento. A Figura 6, mostra o funcionamento da FC baseada em usuários, aplicada a multidomínios.

Figura 6 – Filtragem Colaborativa Baseada em Usuários sobre Multidomínios.



Fonte: Adaptado de Veras (2016)

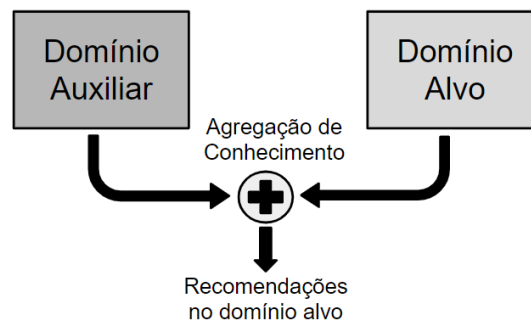
2.3.1 Filtragem Colaborativa Baseada em Vizinhança

Os algoritmos de filtragem colaborativa baseada em vizinhança (FCV), podem prever avaliações, com base na similaridade de preferências entre usuários (KOREN; BELL, 2015; SHARMA; GOPALANI; MEENA, 2019). Estes algoritmos conseguem criar perfis para usuários, baseado em suas preferências; comparando e relacionando os perfis que mais se assemelham, criando assim vizinhanças. Na FCV a similaridade entre dois usuários, no qual deter-

mina os vizinhos de um usuário, é normalmente obtida comparando a semelhança das avaliações feitas por esses usuários aos mesmos itens (NING; DESROSIERS; KARYPIS, 2015; RAGHUWANSHI; PATERIYA, 2018).

Alguns algoritmos de FCV utilizam a Agregação de Conhecimento (AC) dos domínios disponíveis, para a realização de recomendações em multidomínios. Esta abordagem agrega informações de um ou mais domínios auxiliares para a realização de recomendações em um domínio alvo (Figura 7), desta forma as preferências dos usuários, dispostas nos domínios auxiliares e alvo, são agrupadas em um único conjunto de dados, para a aplicação da FCV (CREMONESI; TRIPODI; TURRIN, 2011; VERAS, 2016).

Figura 7 – Agregação de Conhecimento.



Fonte: Cantador et al. (2015)

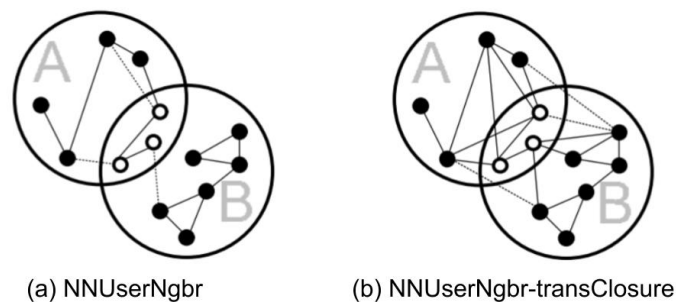
A Agregação de Conhecimento, em cenários onde ocorre a sobreposição de usuários entre os domínios (Seção 2.2.2), possibilita que algoritmos tradicionais de FCV baseados em um único domínio, também sejam utilizados em recomendações com uso de multidomínios (CREMONESI; TRIPODI; TURRIN, 2011; CANTADOR et al., 2015), dentre estes algoritmos está o "Baseado nos N Usuários Mais Próximos" ($NNUserNgbr$) (Figura 8a). Este algoritmo calcula a vizinhança de um determinado usuário, selecionando os N usuários mais próximos ou similares a este usuário, onde a similaridade entre os usuários é definida através de uma métrica de distância e os melhores valores para N são definidos através de experimentos. O algoritmo $NNUserNgbr$ gera previsões de avaliações de itens-alvo por um usuário-alvo, baseado na média ponderada das avaliações dos itens-alvo, feitas pelos usuários da vizinhança do usuário-alvo, onde cada usuário pertencente a vizinhança, contribui segundo o seu nível de similaridade entre o usuário-alvo (RICCI; ROKACH; SHAPIRA, 2015).

2.3.1.1 Algoritmo $NNUserNgbr-transClosure$

Neste trabalho utilizamos o algoritmo $NNUserNgbr-transClosure$, FCV proposta por Cremonesi, Tripodi e Turrin (2011). Este algoritmo é uma melhoria do $NNUserNgbr$, mostrado anteriormente, pois otimiza os cálculos de semelhanças entre usuários com o método de *Transitive-Closure*. Essa melhoria é alcançada pela descoberta indireta de relações entre os usuários, ou seja, o *Transitive-Closure* descobre todos os caminhos de similaridade de n -passos entre

qualquer par de usuários, estendendo sua vizinhança, como ilustrado na Figura 8b. Por exemplo, se houver a semelhança total entre dois usuários A e B: $(\text{Usuário A}) = (\text{Usuário B}) = 1$, e os usuários B e C também são iguais: $(\text{Usuário B}) = (\text{usuário C}) = 1$, o *Transitive-Closure* permite definir que os usuários A e C também são semelhantes, $(\text{Usuário A}) = (\text{Usuário C}) = 1$. [Cremonesi, Tripodi e Turrin \(2011\)](#), limita a transitividade para apenas dois passos, com base em experimentos onde mostram que a aplicação do algoritmo com mais de dois passos, não fornece melhoria na precisão das recomendações, aumentando requisitos computacionais.

Figura 8 – Filtragem Colaborativa NNUserNgbr.



Fonte: [Cremonesi, Tripodi e Turrin \(2011\)](#)

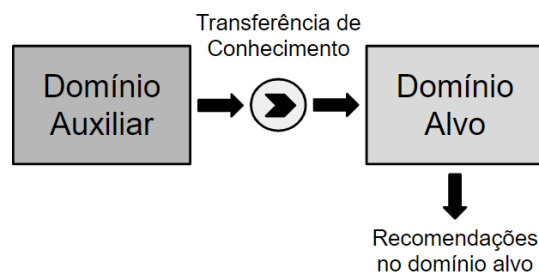
A Figura 8, mostra a melhoria da construção das vizinhanças, com a utilização do algoritmo *NNUserNgbr-transClosure*, em comparação ao *NNUserNgbr*, onde A e B representam os domínios; os pontos brancos na interseção, representam a sobreposição de usuários entre os domínios; os pontos pretos são os usuários com avaliações em apenas um domínio; e as linhas são os relacionamentos entre os usuários. É importante mencionar que o algoritmo *NNUserNgbr-transClosure*, necessita da sobreposição de usuários para uso de domínios auxiliares nas recomendações, uma importante limitação, pois este cenário não trivial em bases de avaliações reais. O nível de sobreposição de usuários também pode afetar o desempenho do algoritmo, podendo melhorar ou não a acurácia das predições ([CREMONESI; TRIPODI; TURRIN, 2011](#)). O processo de gerar predições do algoritmo *NNUserNgbr-transClosure* se assemelha ao *NNUserNgbr*, com a diferença de ser baseado em multidomínios, necessitando da agregação de conhecimento entre os domínios ([VERAS, 2016; CREMONESI; TRIPODI; TURRIN, 2011](#)).

2.3.2 Filtragem Colaborativa Baseada em Transferência de Conhecimento

A Transferência de Conhecimento (TC) é um tópico de pesquisa ativo na área de filtragem colaborativa em multidomínios ([PAN, 2016; PORTUGAL; ALENCAR; COWAN, 2018](#)). Esta técnica de recomendação (FCTC) consegue otimizar a tarefa de aprendizagem e recomendação em um domínio alvo específico, usando padrões de preferências transferidos de domínios auxiliares; estabelecendo assim o relacionamento de informações entre os domínios ([KOREN; BELL, 2015; IWATA; TAKEUCHI, 2015](#)).

As matrizes de avaliações podem compartilhar padrões de preferências semelhantes em diferentes domínios, assim é possível existir correlações entre preferências de grupos de usuários para grupos de itens, formando padrões de avaliações (QUADRANA; CREMONESI, 2014; HAO et al., 2017; JIANG et al., 2017). Alguns algoritmos de FCTC consideram que os padrões de avaliações podem agir como uma ponte de informações que relaciona os domínios auxiliares e alvo (Figura 9), sem a necessidade da sobreposição de usuários para o uso de domínios auxiliares no processo de recomendação (XU et al., 2016; ZHANG et al., 2017; ZHU et al., 2017).

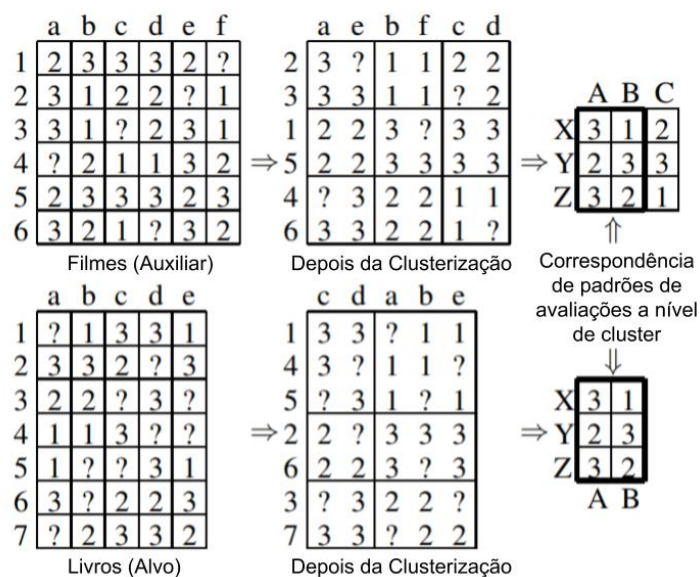
Figura 9 – Transferência de Conhecimento.



Fonte: Cantador et al. (2015)

O algoritmo *Codebook-Transfer* apresentado por Li, Yang e Xue (2009), é um dos algoritmos mais bem sucedidos em transferência de conhecimento, com foco em multidomínios sem a sobreposição de usuários (ZHANG et al., 2017; PAN, 2016). Este algoritmo extrai padrões de avaliações de domínios auxiliares, clusterizando simultaneamente usuários e itens, através da fatoração das matrizes de avaliações (Figura 10); então, o conhecimento é transferido através de uma matriz compacta ao nível de cluster usuário-item (*codebook*), auxiliando na predição de avaliações no domínio alvo das recomendações (LI; YANG; XUE, 2009).

Figura 10 – Processo de clusterização e transferência de conhecimento do *Codebook-Transfer*.



Fonte: Li, Yang e Xue (2009)

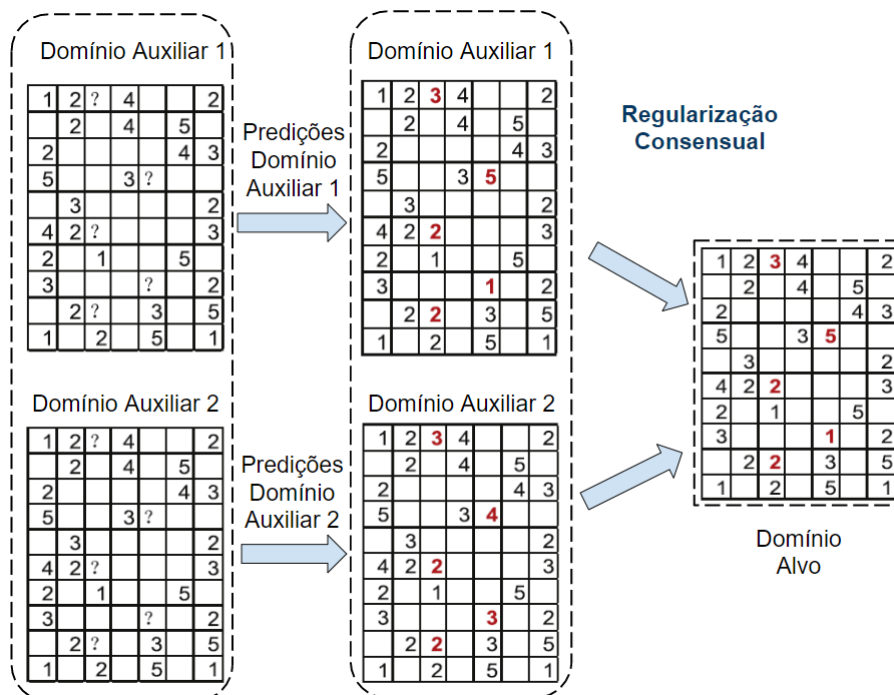
2.3.2.1 Algoritmo *TRACER*.

Neste trabalho foi abordado o algoritmo *Tracer* (*Transfer collaborative filtering via Consensus Regularization*), extensão otimizada do *Codebook-Transfer*, proposta por [Zhuang et al. \(2018\)](#). Através da fatoração de matrizes, o algoritmo constrói padrões de avaliações ao nível de cluster usuário-item em múltiplos domínios auxiliares, transferindo ao domínio alvo os padrões mais relevantes para a geração das predições ([ZHUANG et al., 2018](#)).

O algoritmo *Tracer* se destaca, por possibilitar o uso de dois ou mais domínios auxiliares no processo de recomendação, sem necessitar da sobreposição de usuários entre os domínios, ampliando assim a quantidade de informações disponíveis. Além disso, o algoritmo considera a aprendizagem nos domínios auxiliares e a transferência dos padrões de avaliações para o domínio alvo, como um único processo, onde o modelo criado pode ser ajustado iterativamente, otimizando assim a acurácia das predições ([ZHUANG et al., 2018](#)).

O *Tracer*, iterativamente, aprende e gera diferentes matrizes de predições de avaliações através de cada domínio auxiliar (Figura 11), por exemplo, para a predição de uma avaliação no domínio alvo, é possível obter diferentes resultados de predições, em cada domínio auxiliar (ex.: 5 e 4, dos domínios auxiliares 1 e 2, respectivamente). Em seguida a técnica de Regularização Consensual é aplicada, no qual força as predições dos domínios auxiliares a convergirem para o valor da maioria dos resultados, selecionando assim as melhores predições (ex.: 3 e 2, dos domínios 1 e 2), onde cada domínio auxiliar contribui, diferentemente, para cada predição de avaliação no domínio alvo das recomendações ([ZHUANG et al., 2018](#)).

Figura 11 – Algoritmo *Tracer*.



Fonte: [Zhuang et al. \(2018\)](#)

2.4 Técnicas de Recomendação Sensíveis ao Contexto

Nos últimos anos, pesquisadores e a indústria vêm desenvolvendo técnicas de recomendação sensíveis ao contexto (TRSC), aplicando-as a uma variedade de diferentes domínios, como: cinema, restaurante ou turismo ([ADOMAVICIUS; TUZHILIN, 2015](#); [ABDULKAREM; ABOZAID; SOLIMAN, 2019](#)). As TRSC utilizam diferentes informações contextuais de preferências (ex.: localização, companhia e tempo), como auxílio no processo de recomendação. Estas informações geralmente são adquiridas no momento em que os itens são avaliados pelos usuários. Por exemplo, a localização do usuário pode ser obtida, implicitamente, pelo GPS do seu smartphone; ou o sistema poderia obter a companhia do usuário, explicitamente, fazendo perguntas diretas sobre o contexto da avaliação ([VERAS, 2016](#)).

Algumas TRSC propostas para recomendações baseadas num único domínio, podem ser adaptadas para sistemas de recomendações de domínio cruzado (SRDC), como as técnicas sistemáticas de Pré-Filtragem e Pós-Filtragem Contextual, propostas por [Adomavicius e Tuzhilin \(2015\)](#). Neste trabalho foram utilizados dois algoritmos sensíveis ao contexto para SRDC, propostos por [Veras et al. \(2015\)](#). Estes algoritmos se baseiam na adaptação da Pré e Pós-Filtragem Contextual para utilização em multidomínios, podendo ser aplicados a diferentes tipos de domínios e algoritmos de recomendação de domínio cruzado, sendo adaptáveis, por exemplo, às diferentes técnicas de filtragem colaborativa citadas anteriormente (Seção 2.3). A pré e pós-filtragem contextual, podem auxiliar a aplicação de algoritmos de filtragem colaborativa em multidomínios, selecionando as preferências de usuários mais relevantes para as recomendações, segundo o contexto das avaliações ([VERAS et al., 2015](#)).

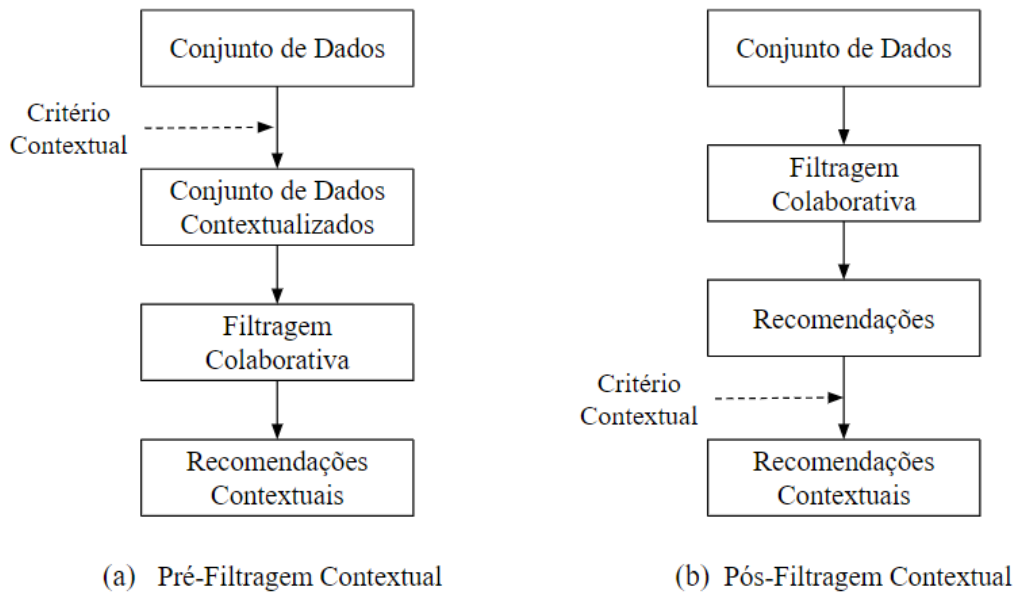
2.4.1 Pré-Filtragem Contextual

A Pré-Filtragem Contextual (Figura 12a), usa inicialmente um critério contextual, para selecionar relevantes avaliações de usuários de um domínio específico, gerando um conjunto de dados, segundo o contexto definido. Após isso, um algoritmo base de filtragem colaborativa é aplicado diretamente na base de dados gerada, produzindo as previsões de avaliações no domínio alvo da recomendação ([ADOMAVICIUS; TUZHILIN, 2015](#)).

2.4.2 Pós-Filtragem Contextual

Na Pós-Filtragem Contextual (Figura 12b), as informações contextuais são inicialmente ignoradas, e as previsões de avaliações são geradas no domínio alvo da recomendação, através da aplicação do algoritmo de filtragem colaborativa, em toda base de dados disponível. Em seguida, as avaliações aos itens produzidas pelo algoritmo de recomendação, são filtradas segundo critérios contextuais. Isto pode ser feito considerando os itens contidos no conjunto de categorias preferenciais do usuário num determinado contexto ([ADOMAVICIUS; TUZHILIN, 2015](#)).

Figura 12 – Pré-Filtragem (a) e Pós-Filtragem (b) Contextual.



Fonte: [Adomavicius e Tuzhilin \(2015\)](#)

2.5 CD-CARS

O CD-CARS (Cross domain context-aware recommender system) é um sistema de recomendação de domínio cruzado sensível ao contexto, proposto por [Veras \(2016\)](#). Este sistema foi baseado no motor de recomendações do *Mahout*³, *framework* distribuído pela *Apache Software Foundation*⁴, para implementação de algoritmos distribuídos ou escaláveis de aprendizagem de máquina, com foco nas áreas de filtragem colaborativa, clusterização e classificação ([OWEN et al., 2012](#)). O sistema de recomendação CD-CARS foi base para os algoritmos de recomendação e experimentos utilizados neste projeto.

O CD-CARS integra algoritmos sensíveis ao contexto, adaptados para multidomínios: Pré-Filtragem Contextual (Seção 2.4.1) e Pós-Filtragem Contextual (Seção 2.4.2), melhorando o desempenho do algoritmo de Filtragem Colaborativa Baseada em Vizinhaça, *NNUserNgbr-transClosure* (Seção 2.3.1.1). A Pré-Filtragem Contextual é utilizada sobre o domínio alvo das recomendações, selecionando as preferências segundo o contexto das avaliações, para a aplicação do algoritmo *NNUserNgbr-transClosure*. A Pós-Filtragem Contextual é aplicada sobre as predições geradas pelo algoritmo *NNUserNgbr-transClosure*, selecionando as predições segundo o contexto das avaliações e as categorias de itens preferenciais dos usuários.

³ Mahout: <https://mahout.apache.org>

⁴ Apache: <https://www.apache.org>

3 Trabalhos Relacionados

Na literatura é possível encontrar diversas técnicas e estudos envolvendo sistemas de recomendação, resultando no experimento de diversos algoritmos em diferentes contextos. Para a elaboração deste trabalho, uma pesquisa bibliográfica foi realizada, com o objetivo de identificar as técnicas mais atuais e efetivas na construção de sistemas de recomendação de domínio cruzado, com foco em algoritmos de filtragem colaborativa (FC) baseada em transferência de conhecimento e técnicas sensíveis ao contexto.

[Bedi e Bedi \(2018\)](#), apresentam uma técnica de FC baseada em vizinhança, paralela em multidomínios e sensível ao contexto. O algoritmo calcula, paralelamente, a similaridade entre os usuários presentes em cada domínio auxiliar. Após isso, o algoritmo computa a similaridade geral entre o usuário alvo e os usuários dos domínios auxiliares, verificando as possíveis vizinhanças e gerando uma lista dos *Top-N* melhores itens para o usuário alvo. O cálculo das similaridades é baseado nas avaliações contextuais expressas pelos usuários, sendo feito através da métrica *Pearson correlation coefficient*. A agregação de conhecimento e sobreposição de usuários entre domínios, se faz necessária para o cálculo da similaridade geral, onde um baixo nível de sobreposição pode afetar o desempenho das recomendações.

[Veras \(2016\)](#), em sua tese, apresenta o CD-CARS, sistema de recomendação de domínio cruzado que utiliza algoritmos sensíveis ao contexto: pré-filtragem e pós-filtragem contextual, como auxílio ao algoritmo de filtragem colaborativa baseado em vizinhança, *NNUserNgb-transColusure*, no processo de recomendação. O *NNUserNgb-transColusure* gera previsões no domínio alvo da recomendação, baseado na transitividade das vizinhanças entre os usuários dos multidomínios. Apesar da pré-filtragem e pós-filtragem contextual melhorarem o desempenho das previsões, o algoritmo *NNUserNgb-transColusure* tem a limitação de apenas utilizar informações do domínio alvo, quando aplicado em conjunto de dados sem a sobreposição de usuários, entre os domínios alvo e auxiliares. É importante também citar que o nível de sobreposição de usuários entre os domínios, pode afetar a precisão das previsões do CD-CARS.

Em [Jiang et al. \(2017\)](#), os autores mostram um algoritmo de FC baseado em transferência de conhecimento, específico para redes sociais. Os autores utilizam informações baseadas em “tags” postadas por usuários, para estimar a probabilidade de um usuário compartilhar ou não algum “post” relacionado a algum assunto. Como auxílio na estimativa das relações entre usuários e posts, métricas de similaridade são usadas na construção de relações internas em cada domínio (tags, posts e usuários). Também é calculado o nível de transferibilidade das tags, extraindo as tags mais significativas para o processo. O algoritmo *Random Walking Borda* utiliza a transferência de informações entre os domínios, para a estimativa das relações externas entre usuários, posts e tags. De acordo com o trabalho, o algoritmo proposto se restringe a mídias so-

ciais, impossibilitando outros tipos de informações, como avaliações de itens por usuários. Uma oportunidade a ser explorada seria a associação de informações contextuais relacionadas às tags e posts, para a melhoria da estimativa de relações entre os domínios.

Hao et al. (2017), propõem o algoritmo CTagCDR, no qual se baseia na transferência de informações presentes em tags. Este algoritmo define a semelhança entre usuários, com base nas categorias dos itens avaliados (tags dos itens). O CTagCDR cria clusters de itens com tags semelhantes e clusters de usuários com preferências similares. Através da fatoração de matrizes, são gerados padrões de avaliações de grupos de usuários à grupos de itens, para cada domínio auxiliar. Estes padrões servem como links de informações entre os multidomínios, auxiliando na geração de predições no domínio alvo. Embora o algoritmo não dependa da sobreposição de usuários, é necessário que os itens dos multidomínios possuam tags iguais ou semelhantes (sobreposição de tags), para o uso dos domínios auxiliares no processo de recomendação.

Em Zhu et al. (2017), os autores mostram um algoritmo de FC baseada em redes de relacionamentos entre itens de multidomínios. Primeiramente o algoritmo constrói uma rede de relacionamentos entre itens, interna e externa aos domínios, baseada em metadados disponíveis para cada item, como gênero, atores, diretores, entre outros. Após a criação dos relacionamentos entre os itens, o algoritmo utiliza a técnica de fatoração de matrizes *Low-Rank Matrix Factorization*, junto com as avaliações dos usuários, na geração de predições no domínio alvo da recomendação. Apesar da aplicação do algoritmo em multidomínios, os autores apenas focaram no uso de domínios relacionados a filmes, não realizando experimentos com outros tipos domínios, como músicas ou games.

Em Xu et al. (2016), os autores focam na geração de recomendações em cenários sem sobreposição de usuários e itens entre os domínios. Primeiro o algoritmo constrói relações de confiança entre usuários dos multidomínios, gerando predições grosseiras no domínio alvo. Este método considera que os gostos de um usuário são similares aos gostos de seus amigos de confiança (ex.: irmão, cônjuge). Através da métrica *Cosine similarity*, o algoritmo cria uma matriz de similaridade entre os itens dos multidomínios baseado no histórico de avaliações, onde os itens são relacionados segundo a semelhança de suas avaliações. Por fim, a FC baseada em item é aplicada, refinando a matriz com predições grosseiras através da matriz de similaridade de itens, gerando as predições definitivas. Os autores propõem a aplicação do algoritmo sobre dois ou mais domínios auxiliares em experimentos futuros, cenário que é abordado neste trabalho.

Zhang et al. (2018) apresentam o *KerKT* (kernel-induced knowledge transfer), no qual propõe a sobreposição de usuários e itens entre domínios, para otimização da FC baseada em transferência de conhecimento. Através da fatoração de matrizes, características de usuários e itens são extraídas de cada domínio. As características são ajustadas por um método de adaptação de domínios, baseado nas sobreposições. Estas informações são usadas para o cálculo da similaridade de usuários e itens. Através da técnica *Kernel-induced completion*, relações entre os domínios são construídas, garantindo a aprendizagem e transferência de informações ao domí-

nio alvo, onde as recomendações são feitas. Neste trabalho, os autores comparam o desempenho entre o *KerKT* e um algoritmo de FC baseada em vizinhança, em cenários com diferentes níveis de sobreposição. Com base em métricas de desempenho, o *KerKT* apresentou as melhores predições em todos os níveis de sobreposição abordados. Para melhoria das predições, informações contextuais das avaliações poderiam ser exploradas pelo *KerKT*, no processo de recomendação.

Em [Zhuang et al. \(2018\)](#), os autores propõem o algoritmo *Tracer* (Transfer collaborative filtering via Consensus Regularization), FC baseada na transferência de conhecimento de múltiplos domínios. O *Tracer* extrai padrões de avaliações de clusters de itens, feitas por clusters de usuários, através de dois domínios auxiliares. Estes padrões são transferidos ao domínio alvo através do processo iterativo de aprendizagem e transferência de conhecimento do algoritmo, auxiliando na geração das predições. O algoritmo gera predições de avaliações para cada domínio auxiliar e utiliza a técnica de Regularização Consensual na seleção das melhores predições para o domínio alvo. Uma proposta de melhoria para este trabalho seria a integração de técnicas sensíveis ao contexto, para auxílio do algoritmo *Tracer* na geração de predições mais precisas.

Diante das propostas dos trabalhos apresentados, este projeto se diferencia pela análise da integração de técnicas sensíveis ao contexto, aos algoritmos de FC baseada em transferência de conhecimento (FCTC) e baseada em vizinhança (FCV), quando aplicados em multidomínios sem a sobreposição de usuários. Para a realização desta análise, utilizamos o algoritmo *Tracer* (FCTC) proposto por [Zhuang et al. \(2018\)](#), por apresentar bons resultados, possibilitar o uso de dois ou mais domínios auxiliares no processo de recomendação, e a disponibilização do código-fonte do algoritmo. Também utilizamos o algoritmo *NNUserNgbr-transColusure* (FCV) abordado no sistema CD-CARS, proposto por [Veras \(2016\)](#) e os algoritmos sensíveis ao contexto: Pré-Filtragem e Pós-Filtragem Contextual, propostos para o CD-CARS.

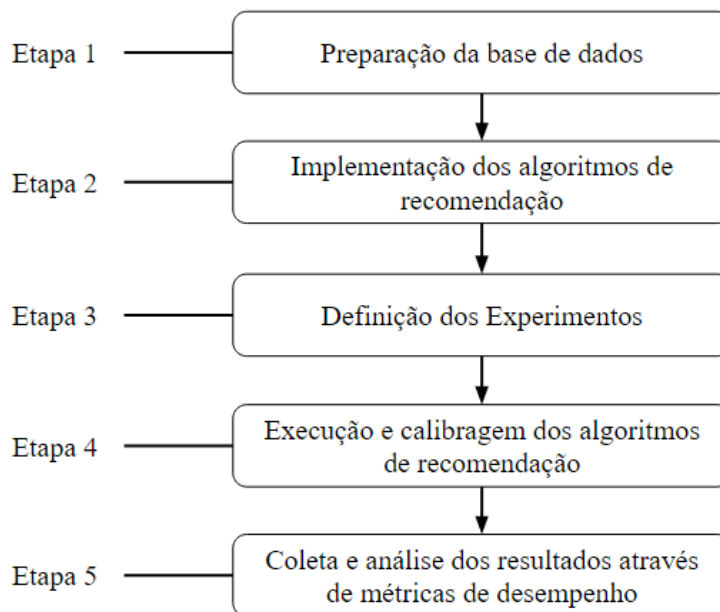
Tabela 1 – Classificação dos Trabalhos Relacionados.

Trabalhos	FCV	FCTC	Sobreposição de Usuários	Sensível ao Contexto
(BEDI; BEDI, 2018)	✓		✓	✓
(VERAS, 2016)	✓		✓	✓
(JIANG et al., 2017)		✓		
(HAO et al., 2017)		✓		
(ZHU et al., 2017)		✓		
(XU et al., 2016)		✓		
(ZHANG et al., 2018)		✓	✓	
(ZHUANG et al., 2018)		✓		
CD-CARS com TRACER		✓		✓

4 Metodologia

Este capítulo apresenta a metodologia utilizada neste trabalho. Dessa forma, com todas as etapas sendo ordenadas e especificadas como mostra a Figura 13, é possível replicar os resultados atingidos nos experimentos realizados.

Figura 13 – Metodologia.



Fonte: Próprio autor

4.1 Base de Dados

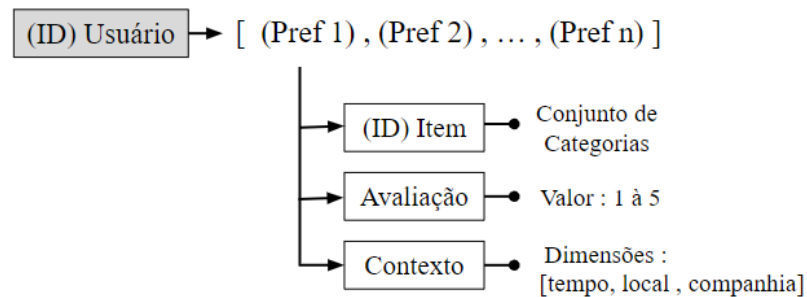
4.1.1 Descrição da Base de Dados

A base de dados utilizada neste trabalho foi construída através da base de informações da *Amazon*¹, disponibilizada por Veras (2016), contendo metadados e avaliações pertencentes a diferentes produtos da loja, relacionados aos domínios de livros, filmes e músicas. Como abordamos o problema da recomendação em multidomínios sob a perspectiva da filtragem colaborativa com técnicas sensíveis ao contexto, consideramos as avaliações dos usuários como uma função de três dimensões: (Usuário x Item x Contexto) → Avaliação Contextual. Assim, as avaliações contextuais podem ser armazenadas, representando um tensor de três dimensões, para cada domínio (livros, filmes e músicas).

¹ Base de dados original: <https://github.com/douglasveras/cd-cars-datasets>

Na estrutura das informações que compõem os domínios (Figura 14), cada usuário possui um ID e pode ter um conjunto de preferências, onde cada preferência (*Pref*) é composta por um item, a avaliação deste item e o contexto da avaliação. Cada item tem um ID e pode pertencer a uma ou mais categorias (ex.: drama, terror, comédia), as avaliações são valores definidos explicitamente pelos usuários, aos itens, numa escala de 1 a 5, já as informações contextuais das avaliações são representadas por diferentes atributos, referente as dimensões contextuais de tempo, localização e companhia.

Figura 14 – Estrutura dos Dados.



Fonte: Próprio autor

Informações contextuais de Tempo e Localização, foram obtidas implicitamente, enquanto as informações de Companhia foram inferidas. Na dimensão Temporal as informações foram extraídas através dos *timestamps* das avaliações, contendo informações como a hora, dia e mês da avaliação. Na dimensão de Localização, as informações foram obtidas com um rastreador Web, responsável por extrair localizações através de páginas Web pertencentes a usuários da *Amazon*, com base nos IDs *Amazon* destes usuários; o endereço completo do usuário foi fornecido pelo serviço *Google Maps Geocoding*², no qual recupera informações, como a cidade, estado ou país, onde ocorreu a avaliação. Na dimensão de Companhia, as informações foram inferidas, com a exploração de comentários avaliativos de usuários, através da mineração de texto. Da análise dos comentários são extraídas palavras-chave descrevendo a companhia do usuário no momento da avaliação (ex.: filhos, amigos).

4.1.2 Preparação da Base de Dados

Neste trabalho foram utilizados dois conjunto de dados distintos, apresentando os possíveis cenários de recomendação em multidomínios: com e sem sobreposição de usuários entre o domínio alvo da recomendação e os domínios auxiliares (Seção 2.2.2). Cada conjunto de dados foi composto por três domínios: Livros, Filmes e Músicas, onde os domínios relacionados a livros e filmes são auxiliares ao domínio de músicas, alvo das recomendações.

² Google Maps Geocoding: <https://developers.google.com/maps/documentation/geocoding/start>

Na construção do conjunto de dados com sobreposição de usuários, foram selecionados da base de informações da *Amazon*, 10000 usuários e 10000 itens com as maiores taxas de avaliações para cada domínio auxiliar (livros e filmes); em seguida foram selecionados 2000 usuários e 9986 itens do domínio alvo (músicas), onde estes 2000 usuários também avaliaram itens nos domínios auxiliares, totalizando 20% de sobreposição de usuários entre o domínio alvo e os domínios auxiliares. Neste conjunto de dados, 100% das avaliações têm informações com contexto temporal, enquanto aproximadamente, 65% possuem contexto de localização e 16% possuem contexto de companhia. As Tabelas 2 e 3, mostram a composição e o nível de sobreposição de usuários entre os domínios do primeiro conjunto de dados.

Tabela 2 – Conjunto de Dados 1 (com sobreposição de usuários).

Conjunto de Dados	Usuários	Itens	Avaliações	Avaliações por	
				Usuário	Item
Multidomínios (Total)	19386	29986	831442	42,89	27,73
Músicas (Alvo)	2000	9986	86357	43,18	8,65
Livros (Auxiliar)	10000	10000	366732	36,67	36,77
Filmes (Auxiliar)	10000	10000	378353	37,84	37,84
Contexto Tempo (Total)	19386	29986	831442	42,89	27,73
Contexto Local (Total)	13098	29542	536297	40,94	18,15
Contexto Companhia (Total)	15868	8091	129565	8,17	16,01

Tabela 3 – Nível de sobreposição de usuários (conjunto de dados 1).

Domínios	Usuários Sobrepostos	% Sobreposição
Músicas e Livros	1368	13,68
Músicas e Filmes	632	6,32
Livros e Filmes	614	6,14

O mesmo processo de construção do primeiro conjunto de dados, foi aplicado na construção do conjunto de dados sem a sobreposição de usuários, exceto na seleção dos usuários do domínio alvo, onde estes usuários apenas avaliaram itens pertencentes ao domínio alvo da recomendação. Neste conjunto de dados, 100% das avaliações possuem contexto temporal, enquanto aproximadamente, 64% possuem contexto de localização e 14% possuem contexto de

companhia. As Tabelas 4 e 5, mostram a composição e o nível de sobreposição de usuários entre os domínios do segundo conjunto de dados.

Tabela 4 – Conjunto de Dados 2 (sem sobreposição de usuários).

Conjunto de Dados	Usuários	Itens	Avaliações	Avaliações por	
				Usuário	Item
Multidomínios (Total)	16860	29952	869444	51,57	29,03
Músicas (Alvo)	2000	9952	104614	52,31	10,51
Livros (Auxiliar)	10000	10000	386477	38,65	38,65
Filmes (Auxiliar)	10000	10000	378353	37,84	37,84
Contexto Tempo (Total)	16860	29952	869444	51,57	29,03
Contexto Local (Total)	11047	29818	558865	50,59	18,74
Contexto Compainha (Total)	13691	8089	125990	9,20	15,58

Tabela 5 – Nível de sobreposição de usuários (conjunto de dados 2).

Domínios	Usuários Sobrepostos	% Sobreposição
Músicas e Livros	0	0
Músicas e Filmes	0	0
Livros e Filmes	5140	51,40

4.2 Algoritmos de Recomendação

Os algoritmos de recomendação utilizados neste trabalho, se basearam no sistema de recomendação de domínio cruzado CD-CARS (Seção 2.5), proposto por Veras (2016). É importante citar que toda a estrutura do sistema CD-CARS, como a arquitetura e o código-fonte do projeto, foram disponibilizados pelo autor para a realização dos experimentos deste trabalho.

Neste trabalho integramos o algoritmo de filtragem colaborativa baseada em transferência de conhecimento, *Tracer* (Seção 2.3.2.1), ao sistema de recomendação CD-CARS. Esta integração possibilitou a realização da análise comparativa entre os algoritmos de filtragem colaborativa: *NNUserNgbr-transClosure* abordado no CD-CARS (Seção 2.3.1.1) e *Tracer*, com o auxílio dos algoritmos de Pré-Filtragem e Pós-Filtragem Contextual, utilizados no CD-CARS (Seção 2.4). A análise comparativa também teve o objetivo de avaliar os benefícios da integração

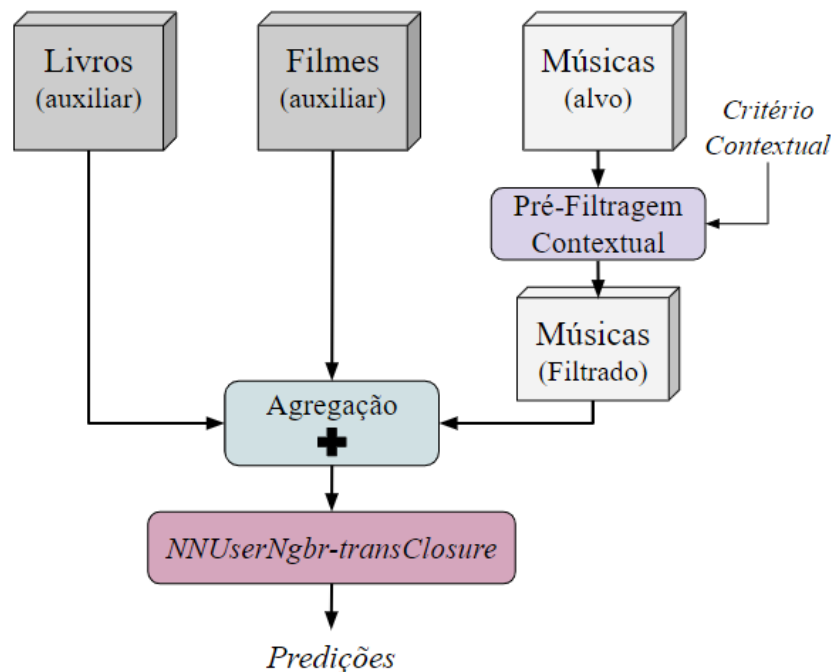
dos algoritmos de pré-filtragem e pós-filtragem contextual, na otimização das predições geradas pelo algoritmo *Tracer*.

Para a aplicação dos algoritmos de recomendação, temos como entrada três tensores, representando os domínios: livros, filmes (auxiliares) e músicas (alvo), formados por conjuntos de preferências, contendo avaliações contextuais (Seção 4.1.1). O contexto das avaliações são utilizados apenas pelos algoritmos de pré-filtragem e pós-filtragem contextual, visto que os algoritmos de filtragem colaborativa necessitam apenas das avaliações, presentes nas preferências de cada domínio: (usuário x item) (Seção 2.3).

4.2.1 Filtragem Colaborativa com Pré-Filtragem Contextual

O algoritmo de Pré-Filtragem Contextual inicializa, utilizando informações contextuais para filtrar as preferências do domínio alvo da recomendação, segundo um critério contextual definido. Em seguida as preferências filtradas do domínio alvo, juntamente com as preferências dos domínios auxiliares, são utilizadas por um algoritmo base de filtragem colaborativa: *Tracer* ou *NNUserNgbr-transClosure*.

Figura 15 – *NNUserNgbr-transClosure* com Pré-Filtragem Contextual.

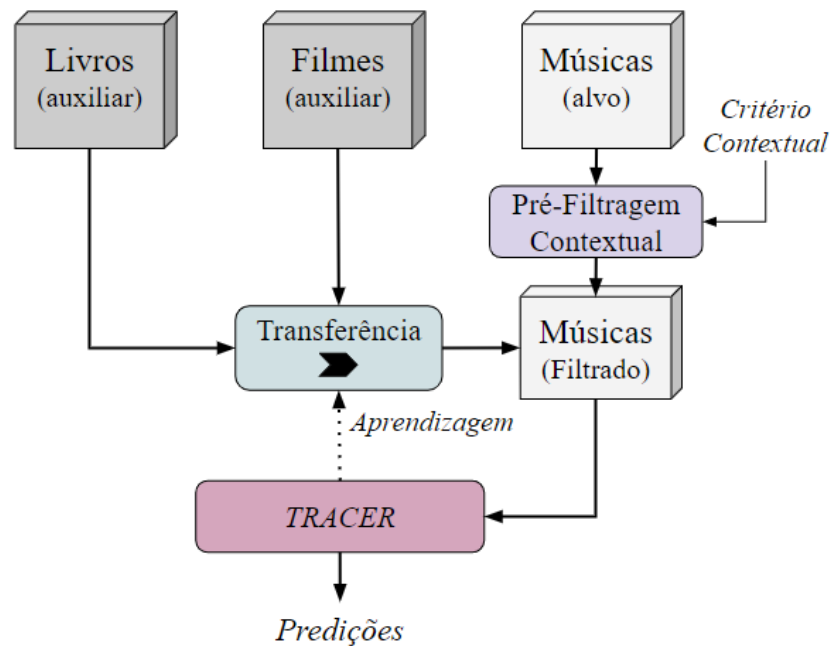


Fonte: Próprio autor com base em Veras (2016)

Para a aplicação do algoritmo *NNUserNgbr-transClosure*, as preferências pré-filtradas do domínio alvo, são agregadas às preferências dos domínios auxiliares, formando um conjunto unificado de preferências (Seção 2.3.1). Após a agregação de informações dos domínios, o algoritmo base *NNUserNgbr-transClosure* é aplicado ao conjunto de preferências, gerando predições de avaliações a itens pertencentes ao domínio alvo (Figura 15).

Na aplicação do algoritmo *Tracer*, após a pré-filtragem contextual, padrões de avaliações presentes nas preferências dos domínios auxiliares, são extraídos e transferidos para o domínio alvo pré-filtrado, através do processo iterativo de aprendizagem e transferência de conhecimento do algoritmo de recomendação. Após o processo de transferência, o algoritmo base *Tracer* gera previsões de avaliações a itens pertencentes ao domínio alvo (Figura 16).

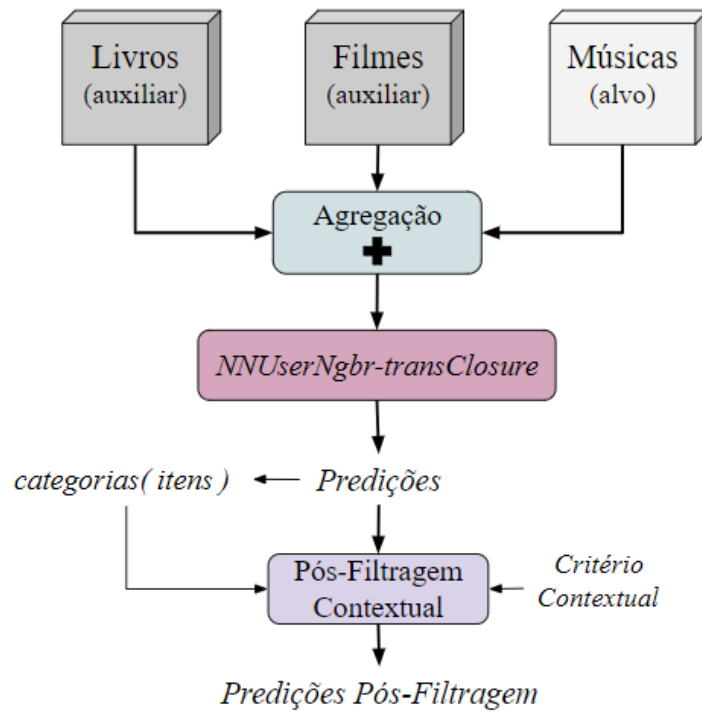
Figura 16 – *Tracer* com Pré-Filtragem Contextual.



Fonte: Próprio autor

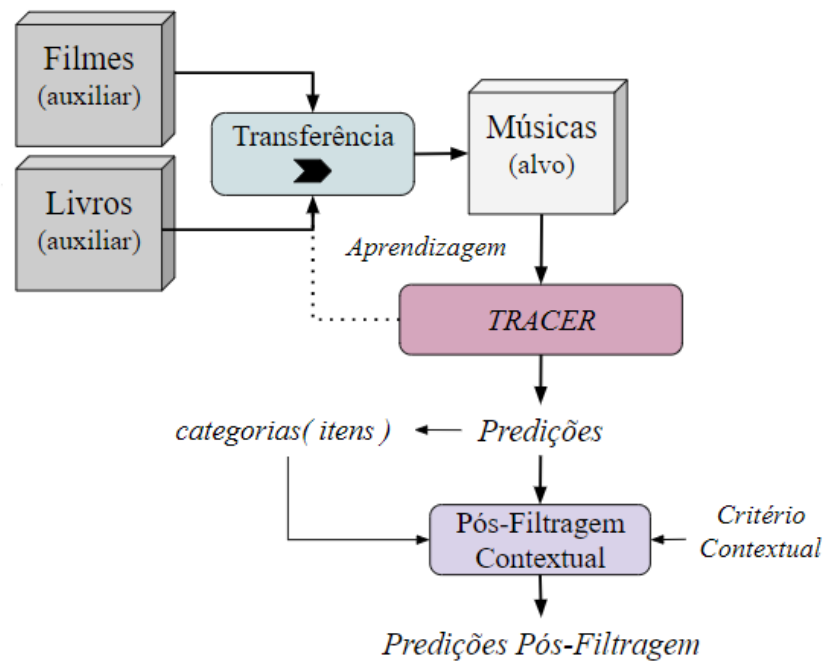
4.2.2 Filtragem Colaborativa com Pós-Filtragem Contextual

Primeiramente algum algoritmo de filtragem colaborativa base: *Tracer* ou *NNUserNgb-transClosure*, é aplicado sobre as avaliações presentes nos conjuntos de preferências de cada domínio. O algoritmo *NNUserNgb-transClosure* é aplicado ao conjunto de preferências formado pela agregação dos domínios auxiliares e alvo (Figura 17), por outro lado o algoritmo *Tracer* transfere padrões de avaliações dos domínios auxiliares para o domínio alvo, num processo iterativo de aprendizagem e transferência de conhecimento (Figura 18). Por fim, os dois algoritmos de filtragem colaborativa geram previsões de avaliações a itens pertencentes ao domínio alvo.

Figura 17 – *NNUserNgbr-transClosure* com Pós-Filtragem Contextual.

Fonte: Próprio autor com base em [Veras \(2016\)](#)

O algoritmo de Pós-Filtragem Contextual é aplicado nas predições de avaliações do domínio alvo, geradas pelos algoritmos de filtragem colaborativa. A pós-filtragem utiliza as informações contextuais das avaliações para a seleção das predições, segundo critério contextual definido. O algoritmo também considera os itens contidos no conjunto de categorias preferenciais dos usuários, para a seleção das melhores predições, gerando assim um conjunto de predições pós-filtragem (Figura 17 e 18).

Figura 18 – *Tracer* com Pós-Filtragem Contextual.

Fonte: Próprio autor

4.2.3 Funcionamento dos Algoritmos Sensíveis ao Contexto

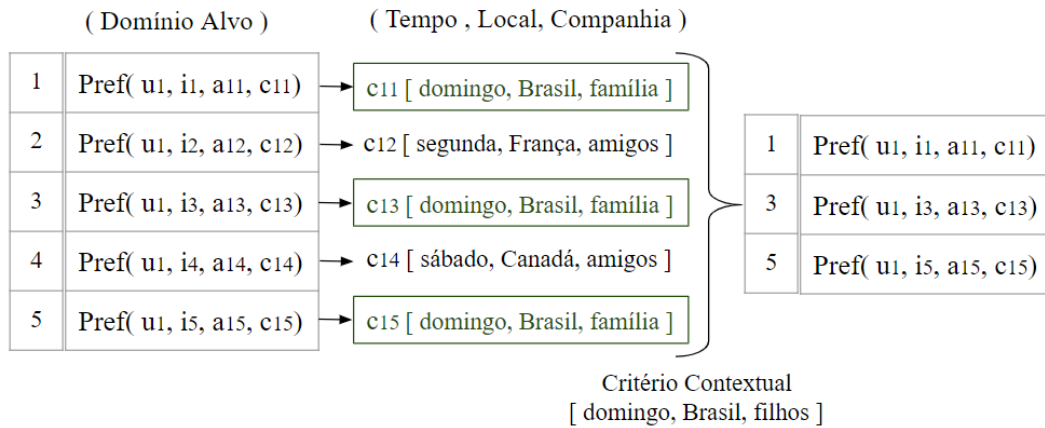
Como mencionado nas Seções 4.2.1 e 4.2.2, os algoritmos sensíveis ao contexto (pré e pós filtragem contextual), necessitam do critério contextual e dos contextos das avaliações para a seleção das preferências presentes nos domínios. Tanto os critérios contextuais, como os contextos das avaliações, possuem três dimensões contextuais: tempo, localização e campanha, onde as dimensões são formadas por atributos contextuais, podendo apresentar diferentes valores, como descrito abaixo:

- **Tempo:** Esta dimensão possui um atributo contextual relacionado a dias da semana, podendo apresentar 7 diferentes valores: sábado a domingo.
- **Localização:** Esta dimensão possui um atributo contextual relacionado a países, podendo apresentar 112 possíveis valores, onde cada valor representa um país diferente.
- **Companhia:** Esta dimensão possui um atributo contextual relacionado a diferentes tipos de companhia dos usuários no momento das avaliações, podendo apresentar 6 possíveis valores: sozinho, acompanhado, família, amigos, cônjuge e colegas.

Na Pré-Filtragem Contextual, as preferências do domínio alvo são filtradas segundo um critério contextual definido previamente, ou seja, as preferências apenas serão selecionadas se os atributos de seus contextos possuírem os mesmos valores dos atributos do critério contextual. Para o funcionamento do algoritmo de pré-filtragem, apenas foi necessário configurar os

possíveis valores dos atributos, pertencentes as dimensões do critério contextual. A Figura 19 apresenta uma possível configuração do critério contextual, onde as preferências (*Pref*) apenas serão selecionadas se as avaliações (*a*) dos itens (*i*), ocorreram no domingo, no Brasil e na companhia da família do usuário (*u*).

Figura 19 – Funcionamento da Pré-Filtragem Contextual.

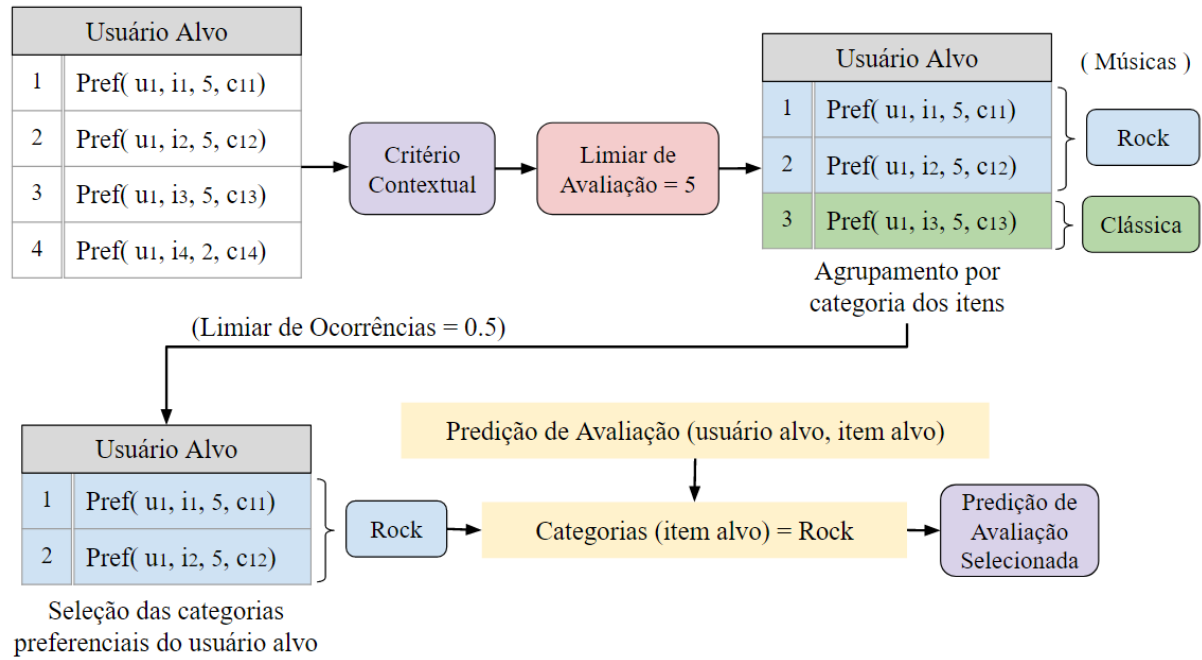


Fonte: Próprio autor com base em [Veras \(2016\)](#)

A Pós-Filtragem Contextual seleciona as predições de avaliações no domínio alvo, geradas pelos algoritmos de recomendação, segundo o critério contextual definido e as categorias dos itens preferenciais dos usuários, alvo das predições. Para a seleção das categorias preferências foram utilizados dois parâmetros: o *limiar de avaliação* e o *limiar de ocorrências*. O *limiar de avaliação* auxilia a pós-filtragem na seleção das preferências com itens avaliados igualmente ou superior a um determinado valor. O *limiar de ocorrências* é utilizado na seleção das categorias de itens preferenciais dos usuários, auxiliando no cálculo da quantidade média de itens por categoria. Os melhores valores para os parâmetros foram definidos através de experimentos preliminares, com os dois algoritmos de filtragem colaborativa, nas diferentes dimensões contextuais.

A Figura 20, apresenta um exemplo do algoritmo de Pós-Filtragem Contextual. Primeiramente as preferências do usuário, alvo da predição, são selecionadas segundo o critério contextual definido, em seguida apenas as preferências com itens avaliados igualmente ao *limiar de avaliação* são selecionados, onde definimos o *limiar de avaliação* com o valor 5 (avaliação mais alta), após isso as preferências são agrupadas com base nas categorias musicais dos itens (rock ou clássica), definindo em seguida as categorias preferenciais do usuário alvo (rock). A categoria pode ser considerada preferencial, se ao menos tiver a quantidade de itens, igual ou superior à quantidade média de itens por categoria, definida como: (*limiar de ocorrências* x *maior quantidade de itens em todas as categorias*), onde definimos o *limiar de ocorrências* com o valor 0,5. Por fim, como ao menos uma categoria musical do item, alvo da predição, está entre as categorias preferenciais do usuário alvo (rock), a predição da avaliação é selecionada.

Figura 20 – Funcionamento da Pós-Filtragem Contextual.



Fonte: Próprio autor com base em [Veras \(2016\)](#)

4.2.4 Configurações dos Algoritmos de Filtragem Colaborativa

Para o funcionamento do algoritmo *NNUserNgbr-transClosure* (Seção 2.3.1.1), foi necessário escolher uma métrica para o cálculo da similaridade entre os usuários, como também o tamanho da vizinhança para o algoritmo de recomendação, representado pelos “N” usuários mais próximos. Primeiramente o tamanho da vizinhança foi fixado num determinado valor (ex.: $N = 100$), em seguida algumas métricas de similaridade foram testadas no algoritmo *NNUserNgbr-transClosure*, com o uso de pré-filtragem e pós-filtragem contextual, como: *Cosine Similarity*, *Distância Euclidiana*, *Pearson Correlation* e *Spearman Correlation*, por fim a *Distância Euclidiana* foi definida como a métrica de similaridade, pois obteve o melhor êxito nas predições das avaliações, nos diferentes cenários de recomendação em multidomínios. A Figura 21, mostra a métrica de similaridade, onde au_i e av_i representam as avaliações de dois usuários a um mesmo conjunto de itens $I_{(u,v)}$. Após a escolha da métrica de similaridade, o tamanho da vizinhança foi inicializado com $N = 10$, sendo aumentado gradativamente de 5 em 5 em cada experimento, onde o algoritmo obteve os melhores resultados, quando o tamanho da vizinhança foi definido como os $N = 225$, vizinhos mais próximos.

Figura 21 – Distância Euclidiana.

$$\sqrt{(au_1 - av_1)^2 + (au_2 - av_2)^2 + \dots + (au_n - av_n)^2} = \sqrt{\sum_{i \in I(u,v)} (au_i - av_i)^2}$$

Fonte: [Owen et al. \(2012\)](#)

Para o funcionamento do algoritmo *Tracer* (Seção 2.3.2.1), foram definidos cinco parâmetros: o número de clusters de usuários (K), o número de clusters de itens (L), o coeficiente de importância da regularização consensual (α), o coeficiente de regularização de *over-fitting* (β) e o número máximo de iterações de treinamento (T). O número de clusters de usuários e itens (K e L), servem para a construção dos padrões de avaliações, transferidos dos domínios auxiliares ao domínio alvo das predições, já os coeficientes (α e β), servem como parâmetros de *trade-off*, para a regularização consensual na transferência de padrões. Os melhores valores para os parâmetros foram encontrados com base nos parâmetros utilizados originalmente no trabalho do algoritmo *Tracer* (ZHUANG et al., 2018). Testes foram executados com diferentes parâmetros, aplicando o algoritmo *Tracer* nos diferentes cenários de recomendação em multidomínios, com o uso da pré ou pós-filtragem contextual no processo de recomendação.

Enquanto diferentes números de clusters de usuários e itens (K e L) foram experimentados, os parâmetros de regularização (α e β) e o número de iterações de treinamento (T), foram fixados num determinado valor. Os valores de K e L foram ambos sendo aumentados gradativamente de 5 a 100, onde o *Tracer* obteve um melhor desempenho nas predições de avaliações, quando: $K = 15$ e $L = 15$. Em seguida, diferentes valores entre 0.00001 e 0.1 para o coeficiente beta (β) e valores entre 0.05 e 2 para o coeficiente alfa (α), foram testados, onde o algoritmo de recomendação obteve os melhores resultados com: $\alpha = 1$ e $\beta = 0.0001$. Por fim, o número de iterações de treinamento foi aumentado até as predições se estabilizarem, obtendo a melhor acurácia nas predições, quando definidas $T = 300$ iterações de treinamento, para cada experimento.

4.3 Descrição dos Experimentos

Neste trabalho foram realizados experimentos com o objetivo de avaliar e comparar diferentes algoritmos de recomendação, envolvendo filtragem colaborativa com técnicas sensíveis ao contexto (Seção 4.2.1 e 4.2.2), sendo avaliados segundo o desempenho apresentado na predição de avaliações. As análises comparativas de desempenho, realizadas nos experimentos, foram aplicadas nos seguintes algoritmos de recomendação:

- Apenas os algoritmos de filtragem colaborativa: baseada em transferência de conhecimento (*Tracer*) e baseada em vizinhança (*NNUserNgbr-transClosure*).
- Os algoritmos de filtragem colaborativa (*Tracer* ou *NNUserNgbr-transClosure*), com o algoritmo de Pré-Filtragem Contextual.
- Os algoritmos de filtragem colaborativa (*Tracer* ou *NNUserNgbr-transClosure*), com o algoritmo de Pós-Filtragem Contextual.

Os algoritmos de recomendação, descritos anteriormente, foram aplicadas sobre dois conjuntos de dados distintos, representando dois cenários de recomendação em multidomínios: com e sem sobreposição de usuários, onde cada conjunto de dados foi formado por três domínios: livros, filmes e músicas. Pelo fato do domínio relacionado a músicas apresentar uma densidade menor de avaliações em relação aos outros domínios (Seção 4.1.2), o mesmo foi definido como o domínio alvo das recomendações, enquanto os domínios restantes (livros e filmes), foram definidos como auxiliares ao domínio alvo. Em todos os experimentos, 100% das preferências disponíveis nos domínios auxiliares foram selecionadas para o treinamento dos algoritmos de recomendação, por outro lado, através do método de amostragem *Holdout*, foram selecionadas do domínio alvo, aleatoriamente, 80% das preferências para treinamento e 20% para teste dos algoritmos de recomendação. Para a realização dos experimentos utilizamos uma máquina com unidade de processamento central (CPU) Intel(R) Core(TM) i7-7500U 2.70GHz, 8GB de memória RAM e 4GB de GPU.

Os critérios contextuais para a pré e pós-filtragem contextual, foram formados por atributos de apenas uma dimensão contextual a cada experimento, ou seja, focamos na análise de uma dimensão contextual por vez (tempo, localização ou companhia), não combinando as dimensões nos experimentos. Os algoritmos de recomendação foram executados, para cada valor disponível nos atributos de cada dimensão contextual, ou seja, 7 vezes na dimensão temporal (sábado a domingo), 6 vezes na dimensão companhia (sozinho, acompanhado, família, amigos, cônjuge e colegas) e 112 vezes na dimensão localização (112 países).

Desta forma os experimentos exploraram a aplicação dos algoritmos de recomendação, em cada cenário de recomendação (com e sem sobreposição de usuários) e dimensão contextual disponível (tempo, localização, companhia). No geral, foram explorados 6 tipos de algoritmos de recomendação em 6 cenários distintos de recomendação (Tabela 6), onde cada experimento foi executado por 5 vezes consecutivas, para a coleta dos resultados e comparação dos algoritmos.

Tabela 6 – Cenários dos Experimentos.

Cenários dos Experimentos	Dimensões Contextuais	Conjunto de Dados
Cenário 1	Tempo	Com Sobreposição de Usuários
Cenário 2	Companhia	
Cenário 3	Localização	
Cenário 4	Tempo	Sem Sobreposição de Usuários
Cenário 5	Companhia	
Cenário 6	Localização	

4.3.1 Métricas de Desempenho

Para a avaliação dos algoritmos de recomendação utilizados neste projeto, foram utilizadas duas métricas probabilísticas de desempenho: *Mean Absolute Error* (MAE) e *Root Mean Square Error* (RMSE). Estas métricas são umas das mais efetivas e utilizadas em pesquisas, na avaliação da acurácia de predições, geradas por algoritmos de filtragem colaborativa (GUNAWARDANA; SHANI, 2015; RAGHUWANSHI; PATERIYA, 2018).

As Métricas MAE e RMSE, medem o desvio entre as avaliações reais, oferecidas aos itens do domínio alvo e as predições destas avaliações, geradas pelos algoritmos de filtragem colaborativa. A métrica MAE trata o erro absoluto médio entre cada par de avaliações-predições. A métrica RMSE computa a raiz quadrada do erro quadrado médio, entre as avaliações e as predições, punindo grandes erros nas medições (GUNAWARDANA; SHANI, 2015).

Figura 22 – Métricas de Desempenho.

$$\text{MAE} = \frac{\sum_{(u,i) \in R_{tgt}^{(test)}} |(R_{tgt} - \hat{R}_{tgt})_{(u,i)}|}{|Q_{tgt}^{(test)}|},$$

$$\text{RMSE} = \sqrt{\frac{\sum_{(u,i) \in R_{tgt}^{(test)}} [(R_{tgt} - \hat{R}_{tgt})_{(u,i)}]^2}{|Q_{tgt}^{(test)}|}}.$$

Fonte: Zhuang et al. (2018)

A Figura 22, mostra em detalhes as fórmulas das métricas de desempenho. Na métrica MAE, se computa o somatório do erro absoluto, entre as avaliações reais (R_{tgt}) e preditas ($\hat{R}_{tgt}$), para teste ($R_{tgt}^{(test)}$), dividido pelo tamanho do conjunto de avaliações para teste ($Q_{tgt}^{(test)}$). Na métrica RMSE, se computa a raiz quadrada do somatório do erro quadrado, entre as avaliações reais (R_{tgt}) e preditas ($\hat{R}_{tgt}$), para teste ($R_{tgt}^{(test)}$), dividido pelo tamanho do conjunto de avaliações para teste ($Q_{tgt}^{(test)}$).

5 Resultados

De acordo com os conjuntos de dados e a metodologia de avaliação descritos anteriormente, neste capítulo apresentamos e discutimos os resultados das avaliações dos algoritmos de recomendação formados pela integração dos algoritmos de Filtragem Colaborativa *NNUserNgbr-transClosure* e *Tracer*, aos algoritmos de Pré e Pós-Filtragem contextual. A Seção 5.1 mostra os resultados das avaliações sobre os algoritmos de Filtragem Colaborativa; as seções posteriores apresentam os resultados das avaliações sobre os algoritmos de Filtragem Colaborativa com Pré-Filtragem Contextual (PréF) ou Pós-Filtragem Contextual (PósF), considerando as dimensões contextuais: temporal (Seção 5.2), companhia (Seção 5.3) e localização (Seção 5.4).

5.1 Resultados dos Algoritmos de Filtragem Colaborativa

Nesta seção apresentamos os resultados da aplicação dos algoritmos de Filtragem Colaborativa, sobre os conjunto de dados com 20% e 0% de sobreposição de usuários, onde os valores das métricas de performance MAE e RMSE foram resultantes da média de cinco experimentos consecutivos, para cada algoritmo.

Tabela 7 – Resultados (MAE / RMSE) com desvio padrão (dp) dos algoritmos de filtragem colaborativa.

Abordagens de Recomendação	20% de Sobreposição		0% de Sobreposição	
	MAE $\pm$ dp	RMSE $\pm$ dp	MAE $\pm$ dp	RMSE $\pm$ dp
NNUserNgbr-transClosure	0,471 $\pm$ 0,006	0,822 $\pm$ 0,010	0,570 $\pm$ 0,004	0,890 $\pm$ 0,008
Tracer	0,493 $\pm$ 0,003	0,635 $\pm$ 0,003	0,501 $\pm$ 0,004	0,641 $\pm$ 0,003

Como pode ser observado na Tabela 7 e Figuras 23 e 24, nos experimentos realizados sobre o conjunto de dados com 0% de sobreposição de usuários, o algoritmo *Tracer* obteve melhores resultados que o algoritmo *NNUserNgbr-transColusure*, em todas as métricas de performance, cerca de $\approx 12\%$ (MAE) e $\approx 28\%$ (RMSE) de melhoria na acurácia das predições. Nos experimentos sobre o conjunto de dados com 20% de sobreposição o *Tracer* foi inferior em $\approx 5\%$ (MAE) e superior em $\approx 23\%$ (RMSE) nas predições em comparação ao *NNUserNgbr-transClosure*.

Figura 23 – Resultados dos algoritmos de Filtragem Colaborativa, MAE (a) e RMSE (b).

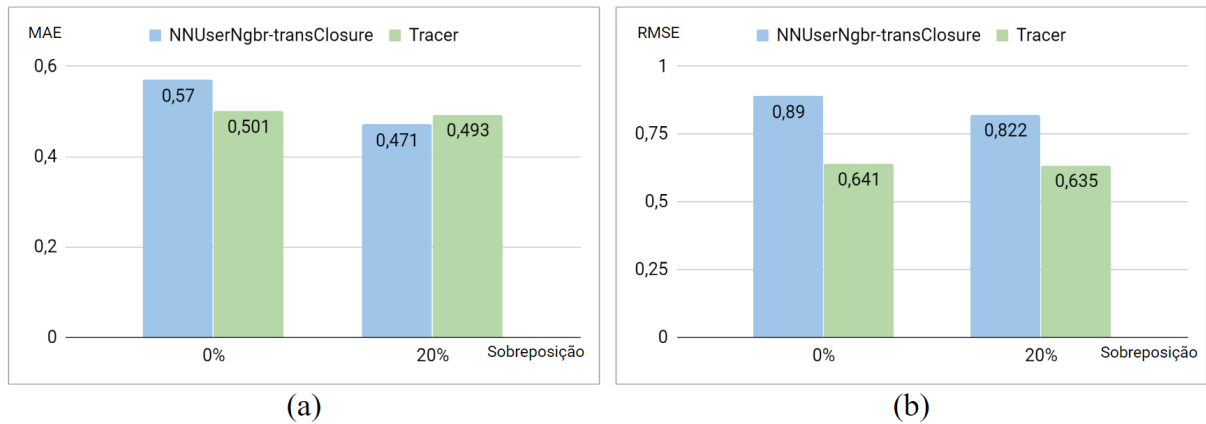
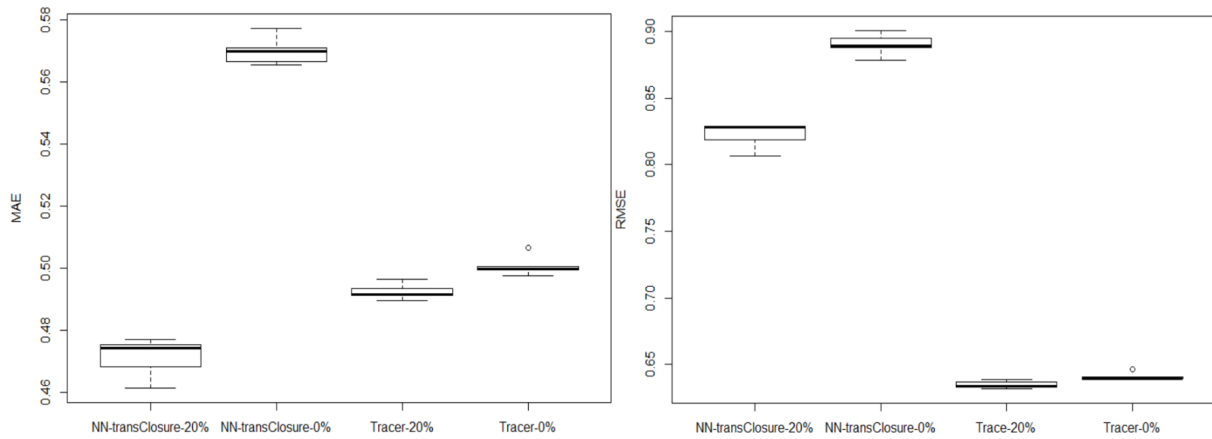


Figura 24 – Resultados (MAE e RMSE) boxplot, 20% e 0% de sobreposição.



5.2 Resultados na Dimensão Contextual Temporal

Nesta seção apresentamos os resultados da aplicação dos algoritmos de Filtragem Colaborativa com Pré e Pós-Filtragem Contextual (Pré-F e Pós-F), sobre os conjunto de dados com 20% e 0% de sobreposição de usuários, levando em consideração apenas a dimensão contextual temporal. Os resultados das métricas MAE e RMSE foram extraídos da realização de cinco experimentos consecutivos, para cada algoritmo de recomendação (Tabela 8).

Nos experimentos realizados sobre o conjunto de dados com 0% de sobreposição de usuários, o algoritmo *Tracer* com Pré-F foi superior em $\approx 27\%$ (MAE) e $\approx 42\%$ (RMSE) na acurácia das predições, em relação ao *NNUserNbr-transClosure* com Pré-F (Figura 25). O algoritmo *Tracer* com Pós-F também foi superior em $\approx 19\%$ (MAE) e $\approx 34\%$ (RMSE) nos resultados, em relação ao algoritmo *NNUserNbr-transClosure* com Pós-F (Figura 26).

Nos experimentos realizados sobre o conjunto de dados com 20% de sobreposição de usuários, o algoritmo *Tracer* com Pré-F foi inferior em $\approx 32\%$ (MAE) e superior em $\approx 13\%$ (RMSE) nas predições, em relação ao algoritmo *NNUserNgbr-transClosure* com Pré-F (Figura 25). O algoritmo *Tracer* com Pós-F foi superior em apenas $\approx 2\%$ (MAE) e superior em $\approx 29\%$ (RMSE) na acurácia das predições, em relação ao algoritmo *NNUserNgbr-transClosure* com Pós-F (Figura 26).

Tabela 8 – Resultados (MAE / RMSE) com desvio padrão (dp), Pré-F. e Pós-F. com algoritmos de filtragem colaborativa (dimensão contextual temporal).

Abordagens de Recomendação	20% de Sobreposição		0% de Sobreposição	
	MAE $\pm$ dp	RMSE $\pm$ dp	MAE $\pm$ dp	RMSE $\pm$ dp
NNUserNgbr-transClosure	0,471 $\pm$ 0,006	0,822 $\pm$ 0,010	0,570 $\pm$ 0,004	0,890 $\pm$ 0,008
Pré-F. com NNUserNgbr-transClosure	0,253 $\pm$ 0,013	0,600 $\pm$ 0,026	0,537 $\pm$ 0,015	0,914 $\pm$ 0,019
Pós-F. com NNUserNgbr-transClosure	0,438 $\pm$ 0,004	0,770 $\pm$ 0,006	0,529 $\pm$ 0,003	0,832 $\pm$ 0,003
Tracer	0,493 $\pm$ 0,003	0,635 $\pm$ 0,003	0,501 $\pm$ 0,004	0,641 $\pm$ 0,003
Pré-F. com Tracer	0,371 $\pm$ 0,002	0,520 $\pm$ 0,002	0,393 $\pm$ 0,005	0,533 $\pm$ 0,004
Pós-F. com Tracer	0,429 $\pm$ 0,001	0,550 $\pm$ 0,001	0,431 $\pm$ 0,001	0,551 $\pm$ 0,001

Figura 25 – Pré-F. com algoritmos de Filtragem Colaborativa, MAE (a) e RMSE (b) (dimensão contextual temporal).

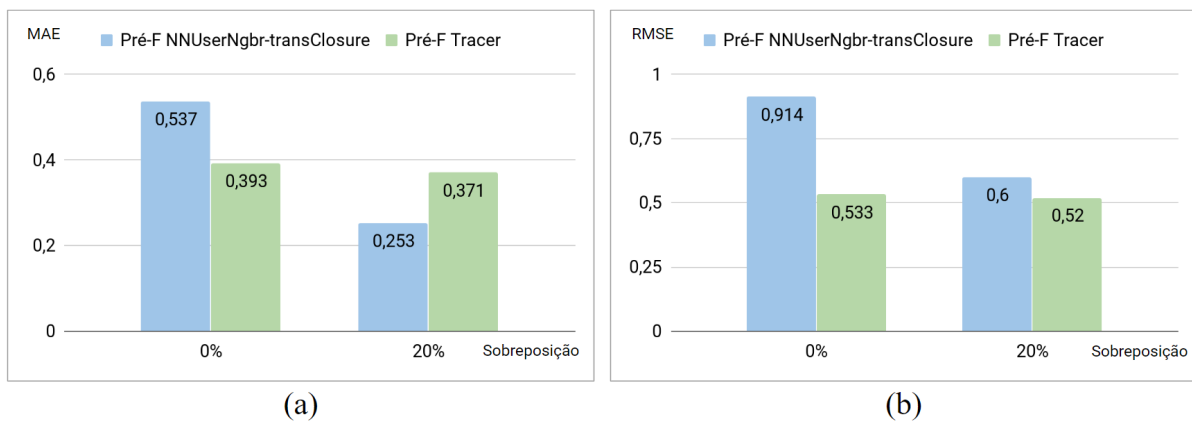
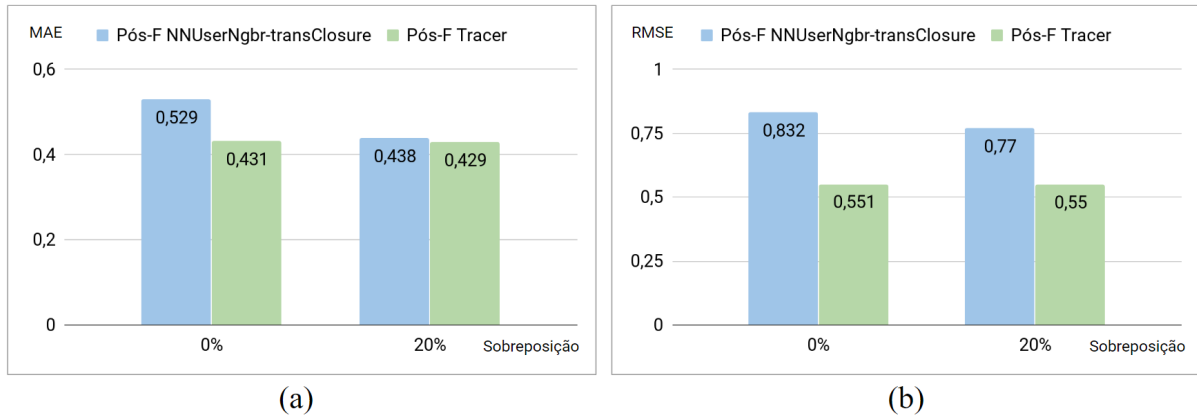


Figura 26 – Pós-F. com algoritmos de Filtragem Colaborativa, MAE (a) e RMSE (b) (dimensão contextual temporal).



Nos experimentos em 0% de sobreposição de usuários (Figura 27), o algoritmo *NNUserNbr-transClosure* com Pré-F obteve uma melhoria de $\approx 6\%$ (MAE) e uma piora de $\approx 3\%$ (RMSE) na acurácia das predições; com a Pós-F obteve melhorias de $\approx 7\%$ (MAE) e $\approx 6\%$ (RMSE). O algoritmo Tracer com Pré-F obteve melhorias de $\approx 21\%$ (MAE) e $\approx 15\%$ (RMSE) nas predições; com a Pós-F obteve melhorias de $\approx 14\%$ (MAE e RMSE).

Nos experimentos em 20% de sobreposição de usuários (Figura 28), o algoritmo *NNUserNbr-transClosure* com Pré-F obteve melhorias de $\approx 46\%$ (MAE) e $\approx 26\%$ (RMSE) na acurácia das predições; com a Pós-F obteve melhorias de $\approx 7\%$ (MAE e RMSE). O algoritmo Tracer com Pré-F obteve melhorias de $\approx 25\%$ (MAE) e $\approx 18\%$ (RMSE) nas predições; com a Pós-F obteve melhorias de $\approx 13\%$ (MAE e RMSE).

Figura 27 – Resultados (MAE e RMSE) boxplot, 0% sobreposição (dimensão contextual temporal).

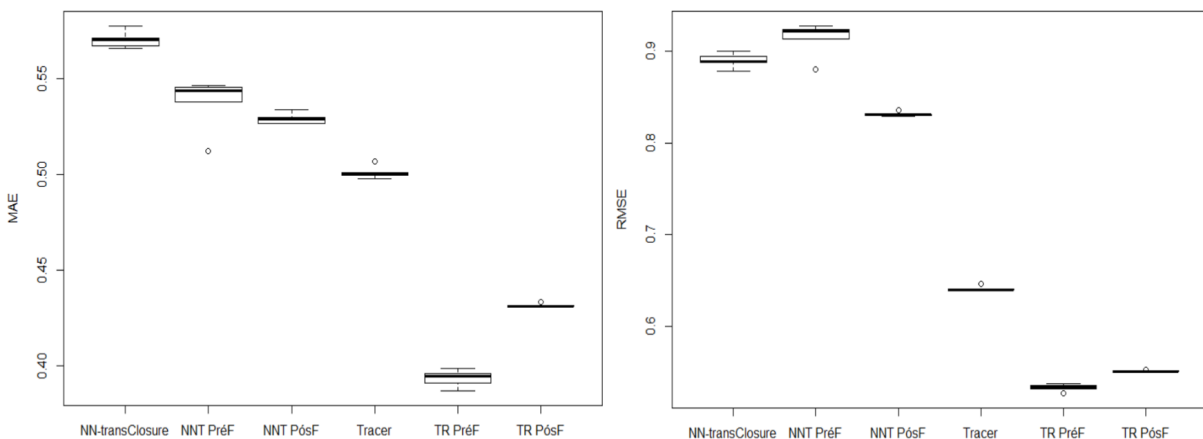
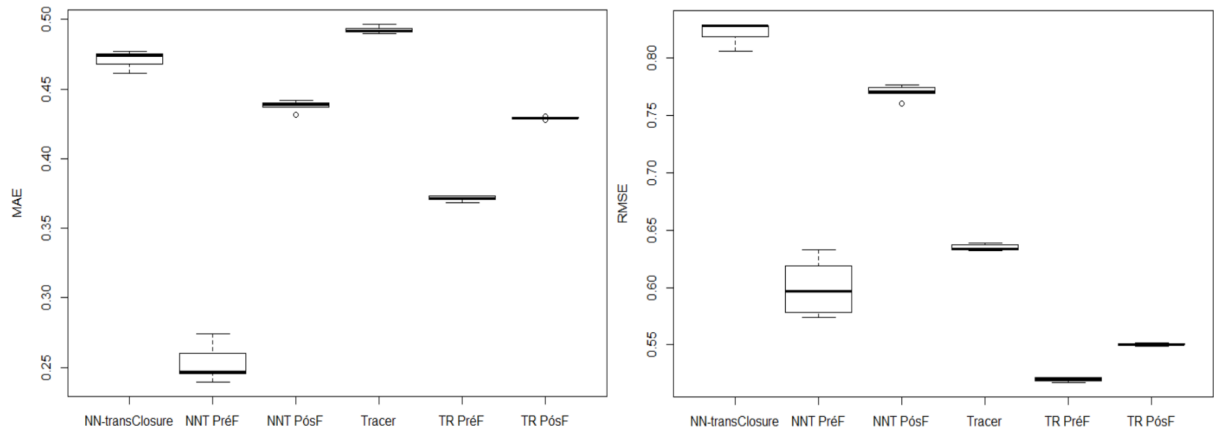


Figura 28 – Resultados (MAE e RMSE) boxplot, 20% sobreposição (dimensão contextual temporal).



5.3 Resultados Dimensão Contextual de Companhia

Nesta seção apresentamos os resultados da aplicação dos algoritmos de Filtragem Colaborativa com Pré e Pós-Filtragem Contextual (Pré-F e Pós-F), sobre os conjunto de dados com 20% e 0% de sobreposição de usuários, levando em consideração apenas a dimensão contextual de companhia. Os resultados das métricas MAE e RMSE foram extraídos da realização de cinco experimentos consecutivos, para cada algoritmo de recomendação (Tabela 9).

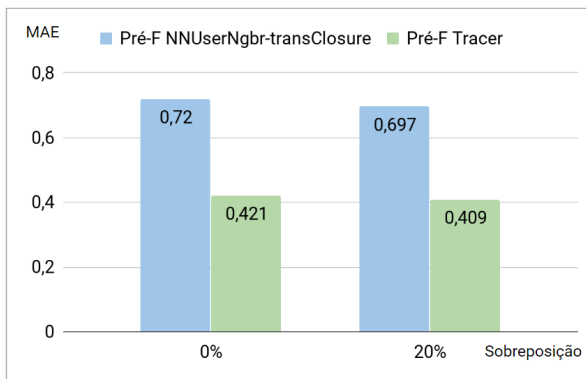
Nos experimentos realizados sobre o conjunto de dados com 0% de sobreposição de usuários, o algoritmo *Tracer* com Pré-F foi superior em $\approx 42\%$ (MAE) e $\approx 47\%$ (RMSE) na acurácia das predições, em relação ao *NNUserNgbr-transClousure* com Pré-F (Figura 29). O algoritmo *Tracer* com Pós-F também foi superior em $\approx 10\%$ (MAE) e $\approx 27\%$ (RMSE) nas predições, em relação ao algoritmo *NNUserNgbr-transClousure* com Pós-F (Figura 30).

Nos experimentos realizados sobre o conjunto de dados com 20% de sobreposição de usuários, o algoritmo *Tracer* com Pré-F foi superior em $\approx 41\%$ (MAE) e $\approx 46\%$ (RMSE) na acurácia das predições, em relação ao algoritmo *NNUserNgbr-transClousure* com Pré-F (Figura 29). O algoritmo *Tracer* com Pós-F foi inferior em $\approx 10\%$ (MAE) e superior em $\approx 20\%$ (RMSE) nas predições, em relação ao algoritmo *NNUserNgbr-transClousure* com Pós-F (Figura 30).

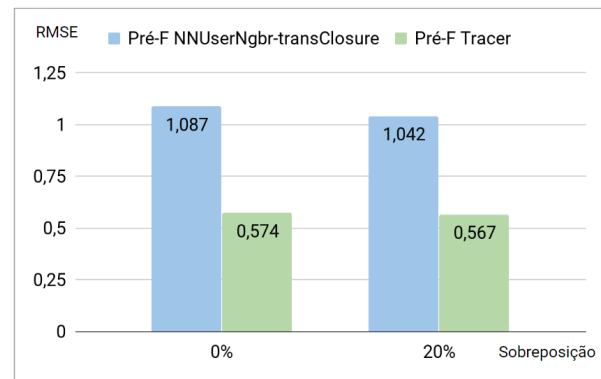
Tabela 9 – Resultados (MAE / RMSE) com desvio padrão (dp), Pré-F. e Pós-F. com algoritmos de filtragem colaborativa (dimensão contextual de companhia).

Abordagens de Recomendação	20% de Sobreposição		0% de Sobreposição	
	MAE $\pm$ dp	RMSE $\pm$ dp	MAE $\pm$ dp	RMSE $\pm$ dp
NNUserNgbr-transClosure	0,471 $\pm$ 0,006	0,822 $\pm$ 0,010	0,570 $\pm$ 0,004	0,890 $\pm$ 0,008
Pré-F. com NNUserNgbr-transClosure	0,697 $\pm$ 0,046	1,042 $\pm$ 0,082	0,720 $\pm$ 0,019	1,087 $\pm$ 0,019
Pós-F. com NNUserNgbr-transClosure	0,410 $\pm$ 0,012	0,729 $\pm$ 0,011	0,493 $\pm$ 0,004	0,787 $\pm$ 0,007
Tracer	0,493 $\pm$ 0,003	0,635 $\pm$ 0,003	0,501 $\pm$ 0,004	0,641 $\pm$ 0,003
Pré-F. com Tracer	0,409 $\pm$ 0,007	0,567 $\pm$ 0,007	0,421 $\pm$ 0,008	0,574 $\pm$ 0,010
Pós-F. com Tracer	0,456 $\pm$ 0,005	0,582 $\pm$ 0,005	0,447 $\pm$ 0,003	0,573 $\pm$ 0,003

Figura 29 – Pré-F. com algoritmos de Filtragem Colaborativa, MAE (a) e RMSE (b) (dimensão contextual de companhia).

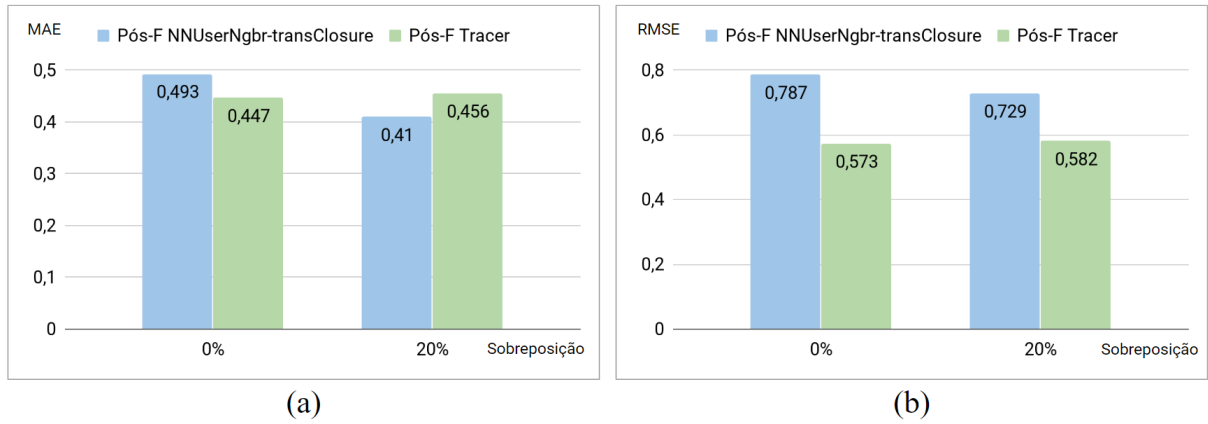


(a)



(b)

Figura 30 – Pós-F. com algoritmos de Filtragem Colaborativa, MAE (a) e RMSE (b) (dimensão contextual de companhia).



Nos experimentos em 0% de sobreposição de usuários (Figura 31), o algoritmo *NNUserNgbr-transClosure* com Pré-F obteve pioras de $\approx 21\%$ (MAE) e $\approx 18\%$ (RMSE) na acurácia das predições; com a Pós-F obteve melhorias de $\approx 14\%$ (MAE) e $\approx 12\%$ (RMSE). O algoritmo Tracer com Pré-F obteve melhorias de $\approx 16\%$ (MAE) e $\approx 10\%$ (RMSE) na acurácia das predições; com a Pós-F obteve melhorias de $\approx 11\%$ (MAE) e (RMSE).

Nos experimentos em 20% de sobreposição de usuários (Figura 32), o algoritmo *NNUserNgbr-transClosure* com Pré-F obteve pioras de $\approx 32\%$ (MAE) e $\approx 21\%$ (RMSE) na acurácia das predições; com a Pós-F obteve melhorias de $\approx 13\%$ (MAE) e $\approx 11\%$ (RMSE). O algoritmo Tracer com Pré-F obteve melhorias de $\approx 17\%$ (MAE) e $\approx 11\%$ (RMSE) na acurácia das predições; com a Pós-F obteve melhorias de $\approx 8\%$ (MAE) e (RMSE).

Figura 31 – Resultados (MAE e RMSE) boxplot, 0% sobreposição (dimensão contextual de companhia).

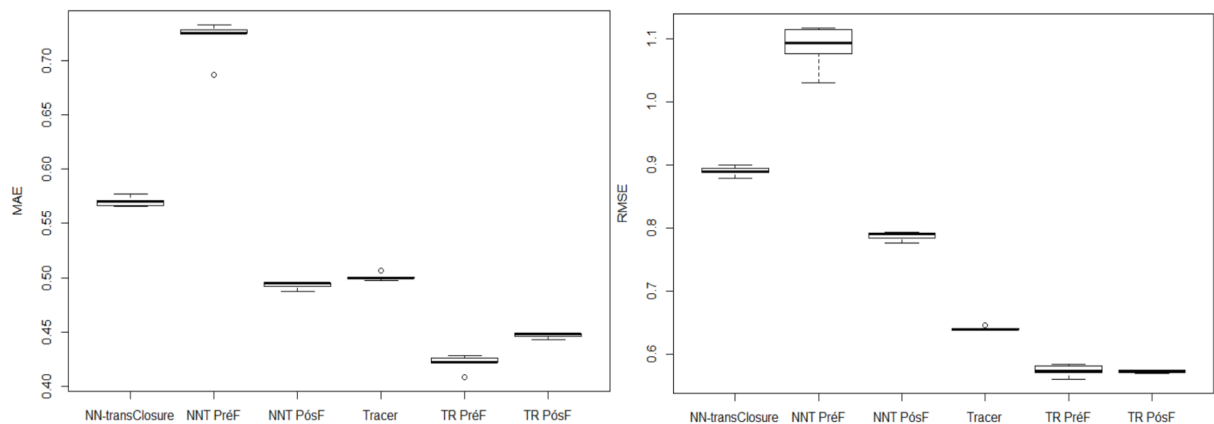
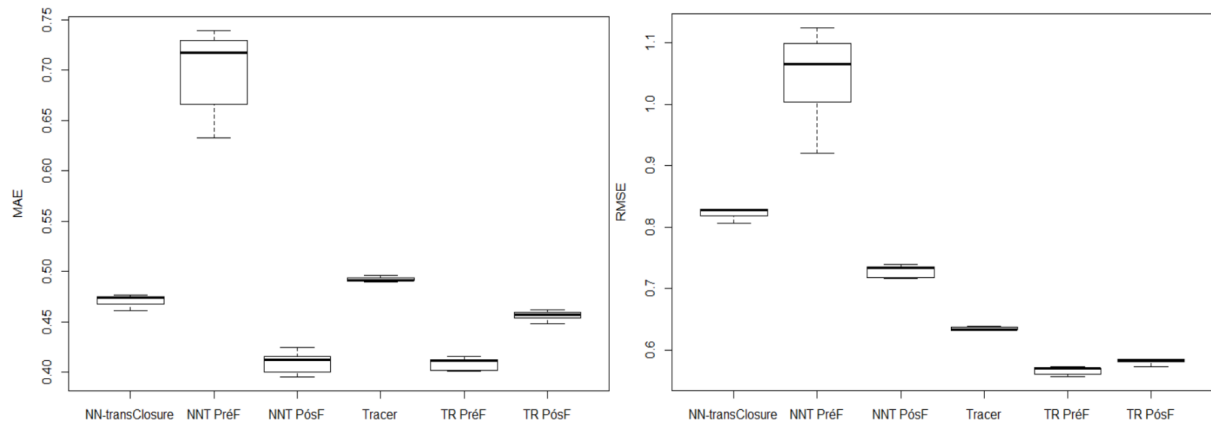


Figura 32 – Resultados (MAE e RMSE) boxplot, 20% sobreposição (dimensão contextual de companhia).



5.4 Resultados na Dimensão Contextual de Localização

Nesta seção apresentamos os resultados da aplicação dos algoritmos de Filtragem Colaborativa com Pré e Pós-Filtragem Contextual (Pré-F e Pós-F), sobre os conjunto de dados com 20% e 0% de sobreposição de usuários, levando em consideração apenas a dimensão contextual de localização. Os resultados das métricas MAE e RMSE foram extraídos da realização de cinco experimentos consecutivos, para cada algoritmo de recomendação (Tabela 10).

Nos experimentos realizados sobre o conjunto de dados com 0% de sobreposição de usuários, o algoritmo *Tracer* com Pré-F foi superior em $\approx 22\%$ (MAE) e $\approx 33\%$ (RMSE) na acurácia das predições em relação ao *NNUserNgbr-transClousure* com Pré-F (Figura 33). O algoritmo *Tracer* com Pós-F foi superior em $\approx 12\%$ (MAE) e $\approx 28\%$ (RMSE) nas predições, em relação ao algoritmo *NNUserNgbr-transClousure* com Pós-F (Figura 34).

Nos experimentos realizados sobre o conjunto de dados com 20% de sobreposição de usuários, o algoritmo *Tracer* com Pré-F foi praticamente igual (MAE) e superior em $\approx 27\%$ (RMSE) nas predições, em relação ao algoritmo *NNUserNgbr-transClousure* com Pré-F (Figura 33). O algoritmo *Tracer* com Pós-F foi inferior em $\approx 10\%$ (MAE) e superior em $\approx 20\%$ (RMSE) nas predições, em relação ao algoritmo *NNUserNgbr-transClousure* com Pós-F (Figura 34).

Tabela 10 – Resultados (MAE / RMSE) com desvio padrão (dp), Pré-F. e Pós-F. com algoritmos de filtragem colaborativa (dimensão contextual de localização).

Abordagens de Recomendação	20% de Sobreposição		0% de Sobreposição	
	MAE $\pm$ dp	RMSE $\pm$ dp	MAE $\pm$ dp	RMSE $\pm$ dp
NNUserNgbr-transClosure	0,471 $\pm$ 0,006	0,822 $\pm$ 0,010	0,570 $\pm$ 0,004	0,890 $\pm$ 0,008
Pré-F. com NNUserNgbr-transClosure	0,458 $\pm$ 0,014	0,797 $\pm$ 0,024	0,602 $\pm$ 0,008	0,923 $\pm$ 0,015
Pós-F. com NNUserNgbr-transClosure	0,432 $\pm$ 0,012	0,765 $\pm$ 0,015	0,554 $\pm$ 0,006	0,855 $\pm$ 0,008
Tracer	0,493 $\pm$ 0,003	0,635 $\pm$ 0,003	0,501 $\pm$ 0,004	0,641 $\pm$ 0,003
Pré-F. com Tracer	0,459 $\pm$ 0,004	0,608 $\pm$ 0,005	0,472 $\pm$ 0,003	0,615 $\pm$ 0,003
Pós-F. com Tracer	0,480 $\pm$ 0,005	0,612 $\pm$ 0,004	0,486 $\pm$ 0,002	0,620 $\pm$ 0,002

Figura 33 – Pré-F. com algoritmos de Filtragem Colaborativa, MAE (a) e RMSE (b) (dimensão contextual de localização).

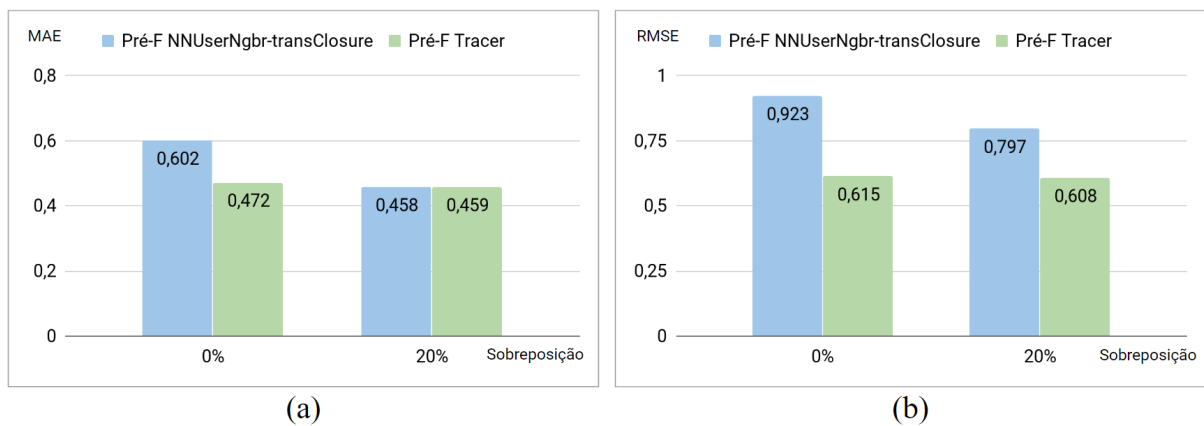
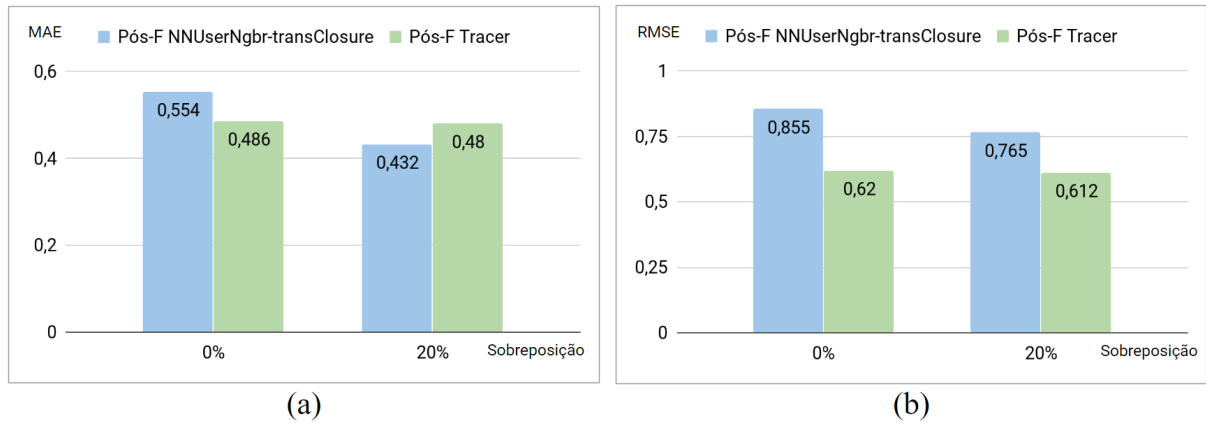


Figura 34 – Pós-F. com algoritmos de Filtragem Colaborativa, MAE (a) e RMSE (b) (dimensão contextual de localização).



Nos experimentos em 0% de sobreposição de usuários (Figura 35), o algoritmo *NNUserNgr-transClosure* com Pré-F obteve pioras de $\approx 5\%$ (MAE) e $\approx 4\%$ (RMSE) nos resultados das predições; com a Pós-F obteve melhorias de $\approx 3\%$ (MAE) e $\approx 4\%$ (RMSE). O algoritmo Tracer com Pré-F obteve melhorias de $\approx 6\%$ (MAE) e $\approx 4\%$ (RMSE) na acurácia das predições; com a Pós-F obteve melhorias de $\approx 3\%$ (MAE) e (RMSE).

Nos experimentos em 20% de sobreposição de usuários (Figura 36), o algoritmo *NNUserNgr-transClosure* com Pré-F obteve melhorias de $\approx 3\%$ (MAE) e (RMSE) na acurácia das predições; com a Pós-F obteve melhorias de $\approx 8\%$ (MAE) e $\approx 7\%$ (RMSE). O algoritmo Tracer com Pré-F obteve melhorias de $\approx 7\%$ (MAE) e $\approx 4\%$ (RMSE) nos resultados das predições; com a Pós-F obteve melhorias de $\approx 3\%$ (MAE) e $\approx 4\%$ (RMSE).

Figura 35 – Resultados (MAE e RMSE) boxplot, 0% sobreposição (dimensão contextual de localização).

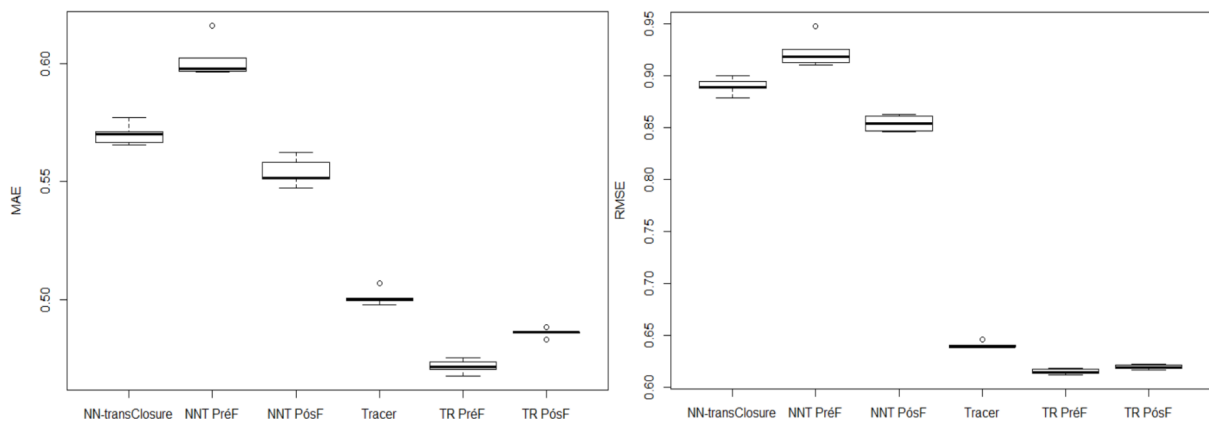
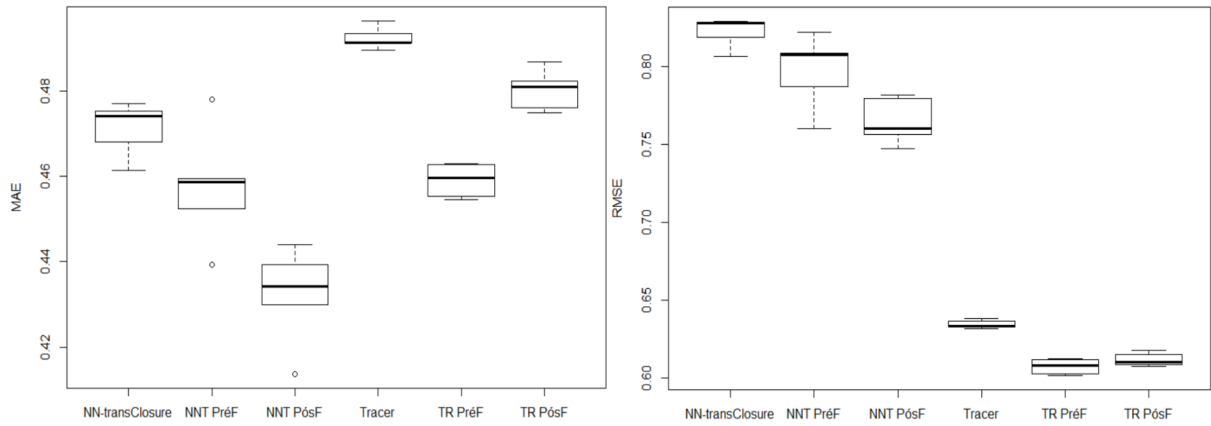


Figura 36 – Resultados (MAE e RMSE) boxplot, 20% sobreposição (dimensão contextual de localização).



5.5 Discussão

Após a coleta das métricas de precisão, foi possível observar que nos experimentos realizados sobre o conjunto de dados com 0% de sobreposição de usuários, o algoritmo *Tracer* apresentou melhores resultados que o algoritmo *NNUserNgbr-transClosure*, em todos os cenários dos experimentos (com e sem Pré-filtragem e Pós-filtragem Contextual e Dimensões Contextuais). Nos experimentos realizados sobre o conjunto de dados com 20% a sobreposição de usuários, o algoritmo *NNUserNgbr-transClosure* apresentou melhores resultados que o algoritmo *Tracer* (com e sem Pré e Pós-Filtragem Contextual), apenas com a métrica MAE; isso devido a uma maior variação nas predições apresentadas, em relação ao algoritmo *Tracer*.

Os resultados demonstram que a integração de algoritmos de filtragem colaborativa baseada em transferência de conhecimento, como o *Tracer*, pode melhorar o desempenho das predições geradas pelo sistema de recomendação CD-CARS, com a utilização de domínios auxiliares, mesmo sem a sobreposição de usuários entre os domínios; limitação encontrada em alguns algoritmos de filtragem colaborativa baseada em vizinhança, como o *NNUserNgbr-transClosure* abordado no CD-CARS.

A Pré-Filtragem Contextual apenas melhorou as predições do *NNUserNgbr-transClosure*, nas dimensões contextuais com maior densidade nas avaliações (tempo e localização). Na maioria dos casos a integração da Pós-Filtragem com o algoritmo *NNUserNgbr-transClosure* gerou melhores resultados em todos os níveis de sobreposição e dimensões contextuais. Também foi possível observar que a falta de sobreposição de usuários entre domínios, diminuiu a acurácia dos algoritmos de recomendação com o *NNUserNgbr-transClosure*. Estes resultados apresentaram o mesmo comportamento dos resultados encontrados no trabalho do CD-CARS.

Com base nos resultados apresentados, foi possível observar que os algoritmos de Pré-Filtragem e Pós-Filtragem Contextual melhoraram a acurácia das predições do algoritmo *Tracer* em todas as dimensões contextuais (tempo, companhia e localização) e conjuntos de dados (0% e 20% de sobreposição de usuários), onde os melhores resultados foram obtidos com a integração do algoritmo *Tracer* à Pré-Filtragem Contextual. Também foi identificado que o nível de sobreposição de usuários não influenciou significativamente as aplicações dos algoritmos de recomendação com o algoritmo *Tracer*, apresentando resultados semelhantes nos conjuntos de dados com 20% e 0% de sobreposição entre os domínios.

6 Conclusão

Este trabalho realizou um estudo sobre diferentes algoritmos de recomendação baseados em conjuntos de dados formados por multidomínios. Seguindo a tendência de pesquisa e motivado pelos ótimos resultados, foram adotados dois algoritmos de Filtragem Colaborativa (FC), base para as análises: o *Ngbr-transClosure* do sistema CD-CARS (FC baseada em vizinhança) e o *Tracer* (FC baseada em transferência de conhecimento), sendo integrados aos dois algoritmos sensíveis ao contexto, utilizados no CD-CARS: a Pré-Filtragem Contextual e a Pós-Filtragem Contextual. Os algoritmos de recomendação foram formados para a resolução do problema da geração de recomendações em multidomínios, sem a sobreposição de usuários, limitação encontrada pelos algoritmos de FC baseada em vizinhança.

Para que os algoritmos de recomendação fossem avaliados igualmente, os mesmos foram aplicados sobre dois conjuntos de dados distintos, formados por preferências contextuais de usuários, pertencentes a um domínio alvo (músicas) e dois auxiliares (livros e filmes). O primeiro conjunto de dados foi construído com 20% de sobreposição de usuários entre os domínios, já o segundo conjunto foi formado com 0% de sobreposição. Os experimentos realizados com a Pré-filtragem e Pós-filtragem Contextual foram focados em três dimensões contextuais, presentes nas preferências: temporal, localização e companhia, utilizadas de forma independente.

Os resultados dos experimentos, apresentados pelos algoritmos de recomendação, foram satisfatórios. Em termos gerais o algoritmo *Tracer* apresentou melhores resultados em relação ao algoritmo *Ngbr-transClosure*, na acurácia das predições realizadas sobre o conjunto de dados, sem a sobreposição de usuários. Nesta análise comparativa, o *Tracer* também apresentou melhores predições com o uso dos algoritmos de Pré-Filtragem e Pós-Filtragem Contextual, em todas as dimensões contextuais disponíveis. Estes resultados demonstram que o problema da falta de sobreposição de usuários nas predições em multidomínios, encontrado no CD-CARS, pode ser resolvido com algoritmos de FC baseada em transferência de conhecimento (*Tracer*).

Os resultados também demonstraram que em todas as dimensões contextuais e níveis de sobreposição de usuários, os algoritmos de Pré-Filtragem e Pós-Filtragem Contextual melhoraram o desempenho do algoritmo *Tracer*, onde os melhores resultados foram obtidos com a integração da Pré-Filtragem Contextual. Estes resultados demonstram que o uso de informações contextuais no processo de recomendação, podem melhorar o desempenho de algoritmos de FC baseada em transferência de conhecimento.

6.1 Contribuições

Durante este trabalho foram construídos dois conjuntos de dados, com e sem sobreposição de usuários entre os domínios, compostos por dois domínios auxiliares relacionados a livros e filmes, e um domínio alvo relacionado a músicas. Os conjuntos de dados¹ foram formados de modo que outros experimentos possam ser executados em trabalhos futuros. O código-fonte² com a estrutura dos algoritmos de recomendação utilizados neste trabalho está disponível para futuros experimentos.

Este trabalho possibilitou a aplicação da filtragem colaborativa baseada em vizinhança (*algoritmo Ngbr-transClosure*), com o auxílio das técnicas sensíveis ao contexto (Pré e Pós-Filtragem Contextual), adotados no CD-CARS, sobre conjuntos de dados compostos por multidomínios sem a sobreposição de usuários, contexto de avaliação ainda não explorado em trabalhos anteriores, até o melhor do nosso conhecimento.

Como contribuição, também foram adotados dois algoritmos de recomendação baseados em multidomínios, ainda não exploradas em pesquisas anteriores, com a integração dos algoritmos de pré-filtragem contextual e pós-filtragem contextual, ao algoritmo de filtragem colaborativa baseada em transferência de conhecimento (*algoritmo Tracer*), possibilitando o uso de informações disponíveis em domínios auxiliares, na geração de predições baseadas em multidomínios, sem a sobreposição de usuários.

6.2 Limitações

Este trabalho apresenta algumas limitações, dentre as principais estão:

- Ausência de um método nos algoritmos de recomendação, para exploração de avaliações com diferentes escalas e formas em diferentes domínios. Por exemplo, adaptar os algoritmos para uso de avaliações binárias, como a função "Curtir" das redes sociais. Técnicas de normalização de avaliações poderiam ser aplicadas, antes do processo de recomendação.
- Necessidade de dois ou mais domínios auxiliares para o uso do algoritmo *Tracer* no processo de recomendação. Devido a esta limitação, experimentos em cenários com apenas um domínio auxiliar não foram realizados, neste trabalho.

¹ Base de dados deste trabalho: <https://github.com/GuilhermeMelo18/dataset-cd-cars-tracer.git>

² Código-fonte deste trabalho: <https://github.com/GuilhermeMelo18/cd-cars-tracer.git>

6.3 Trabalhos Futuros

Para trabalhos futuros, outros algoritmos de filtragem colaborativa baseados em multi-domínios, podem ser analisados, integrando os mesmos aos algoritmos sensíveis ao contexto utilizados neste trabalho. Também é válido expandir a base de dados, com uma maior variação na densidade de avaliações e diferentes níveis de sobreposição de usuários, possibilitando experimentos mais diversificados. Outros tipos de domínios poderiam ser utilizados, como: games, alimentos, roupas; alternando esses domínios entre auxiliares e alvo, nos experimentos.

Para experimentos com dados mais próximos da realidade, os algoritmos de recomendação deste trabalho poderiam ser aplicados a sistemas online, possibilitando o uso de dados em tempo real. Em trabalhos futuros outras métricas de desempenho poderiam ser utilizadas na avaliação dos algoritmos de recomendação, como a métrica F-metric (utilizada para a avaliação do CD-CARS), onde os algoritmos são avaliados segundo as Top-N recomendações.

Em relação ao contexto das avaliações, outras dimensões contextuais poderiam ser utilizadas, como o humor do usuário no momento das avaliações. Outra possibilidade é a aplicação dos algoritmos sensíveis ao contexto, combinando diferentes dimensões contextuais nos experimentos, como: tempo e humor; ou companhia, tempo e localização. Em trabalhos futuros, os algoritmos de pré-filtragem e pós-filtragem contextual, poderiam ser integrados em conjunto a diferentes algoritmos de filtragem colaborativa, sendo possível avaliar as novas técnicas em diferentes contextos e conjuntos de dados.

Referências

- ABDULKAREM, H.; ABOZAID, G.; SOLIMAN, M. Context-aware recommender system frameworks, techniques, and applications: A survey. *International Conference on Innovative Trends in Computer Engineering (ITCE, 2019)*, IEEE, p. 180–185, Fevereiro 2019. Citado 2 vezes nas páginas 12 e 25.
- ADOMAVICIUS, G.; TUZHILIN, A. Context-aware recommender systems. In: RICCI, F.; ROKACH, L.; SHAPIRA, B. (Ed.). *Recommender Systems Handbook*. [S.l.]: Springer, 2015. cap. 6, p. 191–227. Citado 2 vezes nas páginas 25 e 26.
- BEDI, R.; BEDI, P. Parallel proactive cross domain context aware recommender system. *Journal of Intelligent and Fuzzy Systems*, 34, p. 205–227, 2018. Citado 3 vezes nas páginas 12, 18 e 27.
- CANO, E.; MORISIO, M. Hybrid recommender systems: A systematic literature review. *Intelligent Data Analysis*, vol.23, p. 1487–1524, Janeiro 2019. Citado 3 vezes nas páginas 12, 17 e 20.
- CANTADOR, I. et al. Cross-domain recommender systems. In: RICCI, F.; ROKACH, L.; SHAPIRA, B. (Ed.). *Recommender Systems Handbook*. [S.l.]: Springer, 2015. cap. 27, p. 919–961. Citado 7 vezes nas páginas 12, 13, 17, 18, 19, 21 e 23.
- CREMONESI, P.; TRIPODI, A.; TURRIN, R. Cross-domain recommender systems. *11th IEEE International Conference on Data Mining*, 2011. Citado 3 vezes nas páginas 19, 21 e 22.
- FERNANDEZ, I. et al. Cross-domain recommender systems: A survey of the state of the art. *Proceedings of the 2nd Spanish Conference on Information Retrieval*, Janeiro 2012. Citado 2 vezes nas páginas 13 e 17.
- FORBES. *How Amazon Has Reorganized Around Artificial Intelligence And Machine Learning*. 2018. Acessado em 02-04-2019. Disponível em: <<https://www.forbes.com/sites/blakemorgan/2018/07/16/how-amazon-has-re-organized-around-artificial-intelligence-and-machine-learning/#637b5a297361>>. Citado na página 12.
- GUNAWARDANA, A.; SHANI, G. Evaluating recommender systems. In: *Recommender Systems Handbook*. [S.l.]: Springer, 2015. cap. 8, p. 265–309. Citado na página 42.
- HAO, P. et al. Regularizing knowledge transfer in recommendation with tag-inferred correlation. *IEEE TRANSACTIONS ON CYBERNETICS*, IEEE, 2017. Citado 2 vezes nas páginas 23 e 28.
- HUSSIEN, M.; HELMY, M.; HASAN, M. Recommender systems challenges and solutions survey. *International Conference on Innovative Trends in Computer Engineering (ITCE, 2019)*, IEEE, Fevereiro 2019. Citado 2 vezes nas páginas 14 e 17.
- IWATA, T.; TAKEUCHI, K. Cross-domain recommendation without shared users or items by sharing latent vector distributions. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, Proceedings of Machine Learning Research, vol.38, Maio 2015. Citado na página 22.

- JIANG, M. et al. Social recommendation with cross-domain transferable knowledge. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, IEEE, 2017. Citado 2 vezes nas páginas 23 e 27.
- KOREN, Y.; BELL, R. Advances in collaborative filtering. In: *Recommender Systems Handbook*. [S.l.]: Springer, 2015. cap. 3, p. 79–119. Citado 3 vezes nas páginas 12, 20 e 22.
- KUMAR, B.; SHARMA, N. Approaches, issues and challenges in recommender systems: A systematic review. *Indian Journal of Science and Technology*, Vol.9, Dezembro 2016. Citado 2 vezes nas páginas 12 e 17.
- LI, B.; YANG, Q.; XUE, X. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction. *21st international joint conference on Artificial intelligence*, ACM, p. 2052–2057, Julho 2009. Citado 2 vezes nas páginas 14 e 23.
- MURAD, M.; IBRAHIM, R. Cross domain recommender systems: A systematic literature review. *ACM Computing Surveys (CSUR)*, vol.50, ACM, p. n.36, Outubro 2017. Citado 2 vezes nas páginas 12 e 17.
- NING, X.; DESROSIERS, C.; KARYPIS, G. A comprehensive survey of neighborhood-based recommendation methods. In: RICCI, F.; ROKACH, L.; SHAPIRA, B. (Ed.). *Recommender Systems Handbook*. [S.l.]: Springer, 2015. cap. 2, p. 37–77. Citado na página 21.
- OWEN, S. et al. Making recommendations. In: _____. *Mahout in Action*. [S.l.]: Manning Publications, 2012. cap. 4, p. 41–70. Citado 2 vezes nas páginas 26 e 39.
- PAN, W. A survey of transfer learning for collaborative recommendation with auxiliary data. *Neurocomputing*, Elsevier, 2016. Citado 3 vezes nas páginas 13, 22 e 23.
- PATEL, A. et al. Survey and evolution study focusing comparative analysis and future research direction in the field of recommendation system specific to collaborative filtering approach. *Information and Communication Technology for Intelligent Systems*, Springer, 2019. Citado na página 20.
- PORTUGAL, I.; ALENCAR, P.; COWAN, D. The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems With Applications*, Elsevier, p. 205–227, Maio 2018. Citado 4 vezes nas páginas 12, 13, 17 e 22.
- QUADRANA, M.; CREMONESI, P. Cross-domain recommendations without overlapping data: Myth or reality? *Proceedings of the 8th ACM Conference on Recommender systems*, ACM, p. 297–300, Outubro 2014. Citado na página 23.
- RAGHUWANSHI, S.; PATERIYA, R. Recommendation systems: Techniques, challenges, application, and evaluation. *Soft Computing for Problem Solving*, Spriger, p. 151–164, Outubro 2018. Citado 3 vezes nas páginas 12, 21 e 42.
- RICCI, F.; ROKACH, L.; SHAPIRA, B. Recommender systems: Introduction and challenges. In: *Recommender Systems Handbook*. [S.l.]: Springer, 2015. cap. 1, p. 1–37. Citado 4 vezes nas páginas 12, 14, 17 e 21.
- SHARMA, R.; GOPALANI, D.; MEENA, Y. Collaborative filtering–based recommender system: Approaches and research challenges. *3rd IEEE International Conference on "Computational Intelligence and Communication Technology" (IEEE-CICT 2017)*, IEEE, Janeiro 2019. Citado na página 20.

VERAS, D. *CD-CARS: Cross-Domain Context-Aware Recommender Systems. Tese (Doutorado em Computação) - Universidade Federal de Pernambuco*. 2016. Citado 18 vezes nas páginas 12, 13, 14, 15, 18, 20, 21, 22, 25, 26, 27, 29, 30, 33, 34, 36, 38 e 39.

VERAS, D. et al. Context-aware techniques for cross-domain recommender systems. *Brazilian Conference on Intelligent Systems*, Elsevier, p. 282–287, 2015. Citado 2 vezes nas páginas 12 e 25.

XU, Z. et al. Exploiting trust and usage context for cross-domain recommendation. *IEEE ACCESS*, 10, IEEE, Março 2016. Citado 2 vezes nas páginas 23 e 28.

ZHANG, Q. et al. A cross-domain recommender system with kernel-induced knowledge transfer for overlapping entities. *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, IEEE, p. 1–15, Novembro 2018. Citado 4 vezes nas páginas 12, 18, 19 e 28.

ZHANG, Q. et al. A cross-domain recommender system with consistent information transfer. *Decision Support Systems*, Elsevier, Outubro 2017. Citado 2 vezes nas páginas 13 e 23.

ZHU, J. et al. Chrs: Cold start recommendation across multiple heterogeneous information networks. *IEEE TRANSACTIONS ON CYBERNETICS*, IEEE, Agosto 2017. Citado 3 vezes nas páginas 13, 23 e 28.

ZHUANG, F. et al. Transfer collaborative filtering from multiple sources via consensus regularization. *Neural Network*, Elsevier, p. 287–295, Agosto 2018. Citado 7 vezes nas páginas 12, 13, 15, 24, 29, 40 e 42.