

**SISTEMAS DE RECOMENDAÇÃO
NÃO-PERSONALIZADOS PARA ATRAIR
USUÁRIOS NOVOS**

NÍCOLLAS DE CAMPOS SILVA

SISTEMAS DE RECOMENDAÇÃO
NÃO-PERSONALIZADOS PARA ATRAIR
USUÁRIOS NOVOS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais — Departamento de Ciência da Computação como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

ORIENTADOR: ADRIANO CÉSAR MACHADO PEREIRA
COORIENTADOR: LEONARDO CHAVES DUTRA DA ROCHA

Belo Horizonte

19 de fevereiro de 2018

© 2018, Nicollas de Campos Silva.
Todos os direitos reservados.

Silva, Nicollas de Campos

S586s Sistemas de recomendação não-personalizados para
atrair usuários novos / Nicollas de Campos Silva. —
Belo Horizonte, 2018
xxv, 96 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de
Minas Gerais — Departamento de Ciência da
Computação

Orientador: Adriano César Machado Pereira
Coorientador: Leonardo Chaves Dutra da Rocha

1. Computação — Teses. 2. Sistemas de
recomendação. I. Orientador II. Coorientador.
III. Título.

CDU 519.6*73(043)



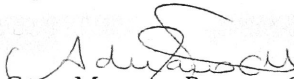
UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

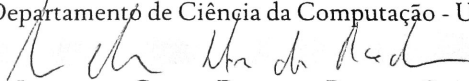
FOLHA DE APROVAÇÃO


Sistemas de recomendação não-personalizados para atrair usuários novos

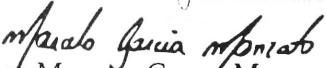
NÍCOLLAS DE CAMPOS SILVA

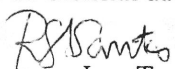
Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:


PROF. ADRIANO CÉSAR MACHADO PEREIRA - Orientador
Departamento de Ciência da Computação - UFMG


PROF. LEONARDO CHAVES DUTRA DA ROCHA - Coorientador
Departamento de Ciência da Computação - UFSJ


DR. FERNANDO HENRIQUE JESUS MOURAO
SEEK AI Labs - SEEK AI Labs


PROF. MARCELO GARCIA MANZATO
Departamento de Ciências da Computação - USP


PROF. RODRYGO LUIS TEODORO SANTOS
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 19 de Fevereiro de 2018.

Dedico este trabalho a todas as pessoas que fazem parte da minha vida.

Agradecimentos

Primeiramente a Deus, que permitiu que tudo isso acontecesse ao longo da minha vida, me ajudando a enfrentar todos os momentos de dificuldades dos quais confiei os meus medos e esperanças a Ele. Agradeço de maneira particular, a minha mãe Joana, heroína que me deu apoio, incentivo nas horas difíceis, de desânimo e cansaço, e que também me ensinou os valores e princípios da vida. Você que esteve sempre ao meu lado, guiando os meus passos e me dando a segurança necessária para eu vencer mais um objetivo de minha vida, minha eterna gratidão. Também de maneira muito especial, agradeço a minha namorada Bárbara Zanetti, por todo o incentivo, dedicação, amor e paciência. Nestes dois longos anos você foi a força necessária para eu continuar a lutar, para vencer os meus medos e conquistar a tão sonhada vitória. Com você pretendo viver toda a minha vida.

Ao meu pai Rogério, por sempre me fortalecer com palavras de consolo e motivação, me fazendo sempre sonhar mais alto. A minha irmã Maria Isabel, que veio ser um motivo para me alegrar nos dias difíceis. A minha avó Antônia, que sempre esteve ao meu lado sendo a força e o exemplo de fé que eu precisava. Ao meu tio Fernando, que me motivou, educou e ensinou a nunca desistir, me ensinando desde pequeno a seguir o caminho dos estudos. A minha madrinha Ana Paula, que gastou preciosas horas do seu Doutorado para me aconselhar, incentivar e me fazer lutar quando o desânimo era grande. Aos demais primos(as) e tios(as), pela contribuição valiosa. E a toda minha família, que sempre me deu forças e torceu pelo meu sucesso. Principalmente aos parentes que não são de sangue, mas que são parentes que Deus me deu a graça de ter.

Agradeço imensamente ao meu irmão de alma, Luís Felipe, por sempre me fazer sentir na presença de Deus. Suas palavras, seu carisma e sua força de vontade são os melhores exemplos que eu poderia ter tido em toda a minha vida. Agradeço também ao meu outro irmão, Guilherme Bassi, por ser um imenso tesouro que Deus me deu, sempre me ensinando as mais preciosas lições que a vida pode nos oferecer. As minhas cunhadas Amanda Assis e Izabel Vale, meus sinceros agradecimentos e orações por aguentarem os meus irmãos. Ao Padre Claudir, pastor guia e fiel, que sempre me aconselhou e ofereceu

uma inabalável amizade. Ao estimado Padre Ramiro, que pelo seu carisma nos ensina a viver o Céu. De maneira especial aos amigos Victor Argamin, Guilherme Carvalho e Jonas Martins pela preciosa companhia. Aos acólitos, meus verdadeiros amigos nascidos pela fé, os quais carregarei para sempre em meu coração. Aos meus amigos incríveis, Samuel Moreira, Rodrigo Agostinho e João Paulo, que acompanharam minha trajetória em Belo Horizonte. Ao Diego Carvalho, que de *pokémon* se tornou um grande amigo a ser levado para a vida toda. E a todos amigos e companheiros de trabalho do LabPi, que durante todos estes anos estiveram ao meu lado me incentivando e apoiando.

Não menos importantes, agradeço aos meus professores que fizeram parte deste trabalho. Ao orientador Adriano Pereira, por ter me acolhido nesta nova fase da minha vida e ensinado preciosos valores humanitários. Ao professor Leonardo Rocha, por desde o início se comprometer com o meu mestrado, lutando por uma bolsa e coorientando os meus passos. Ao professor e amigo Fernando Mourão, agradeço por toda paciência, dedicação e, principalmente, tempo nestes dois anos. Pretendo levar para a vida toda a amizade conquistada com vocês três durante estes anos de estudo. De maneira geral, também a todos os professores que me proporcionaram o conhecimento não apenas racional, mas a manifestação do caráter e afetividade da educação no processo de formação profissional. A palavra mestre nunca fará justiça a vocês, os quais terão os meus eternos agradecimentos. Abro espaço também para agradecer todas as agências de fomento que financiaram este trabalho com bolsas de estudo e reembolsos, como o Instituto Nacional de Ciência e Tecnologia para a Web, MASWeb, EUBra-BIGSEA, CAPES, CNPq e Fapemig.

Enfim, a todos vocês, que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigado! Somente Deus para lhes recompensar com uma vida repleta de graças e bênçãos. Que a Virgem Maria interceda sempre por vocês. Contem comigo para o que precisar!

*“De que serve ao homem conquistar
o mundo inteiro se perder a sua alma?”*

(Mc 8,36)

Resumo

O sucesso de aplicações Web, como sistemas de *e-commerce* e ferramentas de entretenimento, depende da capacidade de converter os usuários novos em usuários recorrentes. Este problema é conhecido como *Pure Cold-Start* e refere-se à capacidade dos Sistemas de Recomendação (SsR) para fornecer recomendações úteis aos usuários que não possuem nenhuma informação associada. Para atenuar esse problema, as estratégias estado-da-arte assumem que itens anteriormente relevantes para muitos usuários conhecidos (ou seja, itens populares) podem ser relevantes para a primeira interação dos usuários novos. Por outro lado, assumimos que itens não populares também são úteis para usuários novos. Por esse motivo, apresentamos duas novas abordagens para lidar com esse problema: Max-Coverage e Niche-Coverage. Para investigar as hipóteses associadas, realizamos testes experimentais *offline* sobre conjuntos de dados tradicionais, demonstrando que os SsR propostos complementam as estratégias não-personalizadas utilizadas na prática. Finalmente, avaliamos diferentes combinações de SsR não-personalizados para construir páginas de produtos em domínios reais. Um estudo *online* realizado com usuários reais apontou que SsR não-personalizados complementares nos permitem aprimorar as recomendações para o primeiro acesso de usuários novos.

Abstract

The success of Web-based applications, like e-commerce systems and entertainment tools, depends on their ability to convert first-time users into recurring ones. This problem is known as Pure Cold-Start and it refers to the capability of Recommender Systems (RSs) to provide useful recommendations to users without historical data. To address this problem, state-of-the-art strategies assume that items previously relevant to many known users (i.e., popular items) may be equally relevant to first-time users. Conversely, we assume that non-popular items are also useful to first-time users. For this reason, we introduce two new approaches to deal with this problem: Max-Coverage and Niche-Coverage. To investigate these hypotheses, we conducted offline experiments in traditional datasets demonstrating that the proposed methods complement recommendations provided by state-of-the-art non-personalized strategies. Finally, we evaluated distinct combinations of non-personalized RSs to construct product pages in real domains. An online study conducted with real users pointed out that mixing complementary non-personalized RSs allows us to enhance recommendations for first-time users.

Lista de Figuras

2.1	Exemplo de cenário de recomendação, com avaliações de 1 a 5.	10
2.2	Visão geral das classes de técnicas de Sistemas de Recomendação.	11
2.3	Processo de recomendação baseado em conteúdo.	12
2.4	Processo de recomendação das estratégias híbridas.	16
2.5	As três maneiras de representação do problema de <i>Cold-Start</i>	17
2.6	As três maneiras de representação do problema de <i>Pure Cold-Start</i>	18
3.1	Exemplo do sistema da Netflix que utiliza um Recomendador Interativo para ser capaz de gerar recomendações para seus usuários <i>novos</i>	22
3.2	Um exemplo de <i>login</i> com cadastros de redes sociais.	23
3.3	Um exemplo de Recomendadores não Personalizados utilizados na prática.	25
3.4	Exemplo de aplicação da estratégia de <i>Popularity</i> sobre o domínio.	26
3.5	Exemplo de aplicação da estratégia de <i>Best Rated</i> sobre o domínio.	26
3.6	Exemplo de aplicação da estratégia de <i>Recent Items</i> sobre o domínio.	27
4.1	As distribuições de popularidade dos itens seguem a <i>power-law</i> . Nos cenários de entretenimento, os itens mais populares são consumidos por mais da metade dos usuários, enquanto nos cenários de <i>e-commerce</i> poucos usuários são assíduos.	32
4.2	Porcentagem de itens populares consumidos por cada usuário, mostrando que existem: (1) usuários que optam por itens populares; (2) usuários que optam por itens não-populares; e (3) usuários que mesclam itens populares e não-populares.	33
4.3	Os nichos de usuários encontrados em cada cenário ressaltam a existência de grupos de interesse dos usuários que nem sempre estão relacionados à popularidade dos itens do domínio.	35

4.4	Exemplo prático da utilização da estratégia <i>Max-Coverage</i> . Pode-se notar que essa estratégia mescla suas recomendações com itens não-populares pois tenta cobrir a preferência do maior número de usuários possível.	38
5.1	Metodologia de avaliação proposta.	42
5.2	Corte temporal realizado em semanas específicas a fim de selecionar os usuários novos como os últimos 20% a ingressarem no sistema.	44
5.3	Similaridade dos primeiros 10 itens recomendados por cada estratégia. Em todos os cenários, mais de 40% dos itens são recomendados apenas pela estratégia de <i>Niche-Coverage</i> . Em ML-10M e Amazon, 50% e 20% dos itens são recomendados apenas por <i>Max-Coverage</i>	47
5.4	Nível de popularidade dos 10 primeiros itens recomendados por cada estratégia, apontando que as estratégias de <i>Max-Coverage</i> e <i>Niche-Coverage</i> são capazes de recomendarem itens menos populares que as estratégias tradicionais.	48
5.5	Principais resultados relacionados à utilidade dos recomendadores não-personalizados. <i>Max-Coverage</i> apresenta ganhos de acurácia em 5% para ML-10M e Amazon, e 47% para CiaoDVD. <i>Niche-Coverage</i> apresenta ganhos de 55% em acurácia para CiaoDVD com relação aos baselines.	50
5.6	Média harmônica dos valores de Precision e Recall obtidos por cada estratégia. Esses resultados contrastam as observações realizadas com a métrica de <i>Accuracy</i>	51
5.7	<i>Max-Coverage</i> apresenta ganhos de 39% para ML-10M e 5% para Amazon e CiaoDVD. <i>Niche-Coverage</i> apresenta ganhos de 5% para CiaoDVD, em relação aos <i>baselines</i>	52
5.8	Resultados da métrica de <i>Genre Coverage</i> , ressaltando a diversidade dos itens recomendados por cada estratégia em todos os domínios.	53
5.9	Frequência em que os gêneros dos itens aparecem nos 10 primeiros itens recomendados pelas principais estratégias. Existem diversos gêneros que são alcançados apenas pelas estratégias de <i>Max-Coverage</i> e <i>Niche-Coverage</i> . O cenário da Amazon não é mostrado nessa análise, pois existem 471 categorias associadas aos produtos.	54
5.10	Interseção dos usuários satisfeitos com pelo menos 1 item dentre os 10 recomendados pelas principais estratégias.	56
6.1	Esquema intuitivo do modelo atual utilizado pelas aplicações <i>e-commerces</i> e da nossa nova proposta para atrair mais usuários novos.	61

6.2	Metodologia de Avaliação <i>Online</i>	61
6.3	Corte temporal realizado no ano de 2012, a fim de selecionar as preferências dos usuários refletidas atualmente.	63
6.4	<i>Snapshots</i> da Interface Web.	65
6.4	Respostas dos usuários para o questionário da fase 2 da pesquisa.	69
6.5	Filmes apresentados pelos recomendadores.	70
6.6	Respostas dos usuários para o questionário pessoal da fase 6.	73

Lista de Tabelas

2.1	Classes de estratégias de recomendadores de filtragem colaborativa.	14
4.1	Descrição geral dos conjuntos de dados selecionados.	31
5.1	Porcentagem de usuários satisfeitos com pelo menos 1 item apresentado por cada lista de recomendação de tamanho k	55
6.1	Porcentagem de usuários novos satisfeitos pela combinação de estratégias, como feito em cenários reais como <i>Amazon</i> e <i>Netflix</i>	60
6.2	Porcentagem de usuários que selecionaram cada uma das opções durante a fase 1, especificando de qual recomendador o item foi escolhido.	67
6.3	Resultados da fase 3, especificando cada recomendador utilizado.	71
6.4	Resultados das fases 4 e 5, especificando cada recomendador utilizado.	72

Sumário

Agradecimentos	ix
Resumo	xiii
Abstract	xv
Lista de Figuras	xvii
Lista de Tabelas	xxi
1 Introdução	1
1.1 Contextualização & Motivação	1
1.2 Objetivos	4
1.3 Principais Contribuições	5
1.4 Organização do Documento	6
2 Referencial Teórico	9
2.1 O Problema de Recomendação	9
2.2 Métodos de Recomendação	11
2.2.1 Filtragem Baseada em Conteúdo	12
2.2.2 Filtragem Colaborativa	13
2.2.3 Métodos Híbridos	15
2.3 Desafios aos Métodos Clássicos	16
2.3.1 O Problema de Cold-Start	16
2.3.2 O Problema de Pure Cold-Start	18
2.4 Síntese do Capítulo	19
3 Estudo do Pure Cold-Start	21
3.1 Principais Abordagens Relacionadas	21
3.1.1 Recomendadores Interativos	22

3.1.2	Recomendadores Híbridos	23
3.1.3	Recomendadores não Personalizados	24
3.2	Estratégias não-Personalizadas Tradicionais	25
3.2.1	Popularity	26
3.2.2	Best Rated	26
3.2.3	Recent Items	27
3.3	Síntese do Capítulo	27
4	SsR não-Personalizados Baseados na Cobertura de Usuários	29
4.1	Estudo de Caso	29
4.1.1	Bases de Dados	31
4.1.2	Consumo de Itens Populares e Não-Populares	32
4.1.3	Nichos de Usuários	34
4.2	Estratégias não-Personalizadas Propostas	36
4.2.1	Estratégia de Max-Coverage	36
4.2.2	Estratégia de Niche-Coverage	38
4.3	Síntese do Capítulo	39
5	Avaliações Empíricas	41
5.1	Metodologia de Avaliação	41
5.2	Seleção de Usuários Novos	43
5.3	Requisitos de Qualidade	44
5.4	Estratégias Baselines	46
5.5	Resultados	46
5.5.1	Caracterização dos Itens Recomendados	47
5.5.2	Avaliação da Qualidade das Recomendações	49
5.5.3	Satisfação dos Usuários	55
5.6	Síntese do Capítulo	57
6	Estudos Direcionados aos Usuários	59
6.1	Combinação de Estratégias não-Personalizadas	59
6.2	Experimento Online	61
6.2.1	Metodologia de Avaliação	61
6.2.2	Seleção do Conjunto de Dados	63
6.2.3	Sistema Web de Avaliação	64
6.3	Análises dos Resultados	66
6.3.1	Análise Conjunta dos Recomendadores	66
6.3.2	Análise Individual dos Recomendadores	71

6.4 Síntese do Capítulo	74
7 Conclusão & Trabalhos Futuros	75
Referências Bibliográficas	79
Apêndice A Questionário 1	91
Apêndice B Questionário 2	93
Apêndice C Termo de Compromisso	95

Capítulo 1

Introdução

Este capítulo apresenta os principais objetivos deste trabalho, bem como as hipóteses associadas. Primeiramente, é feita uma contextualização do problema de recomendar itens para usuários novos, que não possuem nenhuma informação prévia. Além disso, são apresentados os principais fatores que tornam esse problema um dos mais desafiadores e relevantes para diversas aplicações *Web*. Em seguida, apresentamos o objetivo deste trabalho, que consiste em melhorar as recomendações geradas para usuários novos. De maneira complementar, consolidamos três hipóteses relevantes que derivam de estudos realizados sobre os principais cenários de recomendação. Nessa etapa, também são traçados os objetivos específicos para alcançar o nosso objetivo geral. Por fim, apresentamos as principais contribuições deste trabalho que derivam da realização dos objetivos específicos.

1.1 Contextualização & Motivação

Atualmente, com a enorme quantidade de informação disponível em uma variedade de aplicações Web, os usuários possuem mais opções do que podem lidar efetivamente [Schwartz, 2009; Ricci et al., 2015]. Ao disponibilizarem centenas de milhares de produtos distintos, involuntariamente, diversas aplicações de *e-commerce* dificultam os usuários a encontrarem itens potencialmente relevantes. Nesse contexto, ferramentas capazes de filtrar as informações disponíveis, mostrando apenas o que é de interesse do usuário, têm se tornado cada vez mais importantes [Bobadilla et al., 2013; Mourão, 2014]. Essas aplicações são denominadas Sistemas de Recomendação (SsR) e visam encontrar itens potencialmente relevantes para os usuários, com base em um conhecimento prévio sobre os comportamentos dos usuários e/ou características relevantes dos itens [Adomavicius & Tuzhilin, 2005].

Em sua maioria, os SsR analisam os padrões de consumo de cada usuário para fornecer uma recomendação personalizada que se adapte às preferências desses. Procura-se modelar o perfil de um usuário, baseando-se em avaliações feitas anteriormente pelo próprio usuário para outros itens, ou por usuários com um perfil similar àquele. De acordo com a abordagem utilizada, os SsR são classificados em três categorias: (1) filtragem baseada em conteúdo (*Content-Based* - CB); (2) filtragem colaborativa (*Collaborative Filtering* - CF); e (3) técnicas híbridas. CB assume que usuários possuem interesses em itens com características específicas, correlacionando os usuários com as características dos itens. Por outro lado, CF assume que usuários compartilham interesses em comum, correlacionando os usuários para recomendar itens em comum [Mourão, 2014]. Por sua vez, os métodos híbridos combinam os métodos de CB e CF, na tentativa de explorar as qualidades de ambas as abordagens. Existem também outros métodos híbridos que visam combinar as abordagens de CB e CF com informações externas aos sistemas, como as informações relacionadas aos usuários que são provenientes de redes sociais, localização demográfica e outras [Adomavicius & Tuzhilin, 2005].

Na prática, os SsR têm assumido uma importante função em diversas aplicações *Web*, por afetarem decisivamente distintas fases de negócio como a aquisição e retenção dos usuários [Boehmer et al., 2015; Klemm, 2016]. Contudo, apesar de todos os avanços na área, os SsR ainda se esforçam para fornecer recomendações potencialmente úteis aos usuários novos que acabaram de começar a interagir com o sistema [Bernardi et al., 2015]. Esse problema é chamado de *Pure Cold-Start* e ocorre em diversos cenários onde os usuários nunca se interagiram com o sistema ou não permitem a extração de informações dos *cookies*, como em uma navegação anônima [Hernando et al., 2017]. Por estes motivos, esse problema está relacionado a um dos maiores KPIs¹ do mercado, dado que a primeira impressão do usuário sobre o sistema é determinante para conquistar sua confiança. Apresentar itens pouco atrativos a usuários novos, faz com que esses acabem deixando o sistema sem proporcionar um retorno financeiro esperado [Pereira & Hruschka, 2015]. Além disso, a prosperidade de uma aplicação *e-commerce* está diretamente relacionada a sua capacidade de converter esses usuários novos em clientes ativos do sistema [Ramos, 2015]. Entretanto, de acordo com uma pesquisa da *Experian Hitwise*, a taxa média de conversão é baixa: somente 1,65 a cada 100 pessoas que entram em um *e-commerce* efetuam a compra [Reis, 2016]. Infelizmente, a maioria dos usuários saem do sistema e dificilmente retornam para realizar compras. A ausência de informações limita a efetividade dos SsR clássicos, por se basearem no histórico de consumo do usuário [Aharon et al., 2015].

¹Key Performance Indicator - indicadores de performance de um sistema.

A solução tradicionalmente adotada pela maioria dos domínios consiste em utilizar SsR não-personalizados por recomendarem os mesmo itens independente do perfil dos usuários. Essas estratégias se destacam devido à simplicidade, capacidade de generalização, independência do domínio e eficiência em diversos cenários [Bobadilla et al., 2013]. Entre os principais SsR não-personalizados utilizados em domínios reais, ressaltam-se as estratégias que exploram informações dos itens tais como: (1) popularidade; (2) avaliações obtidas; e o (3) período de consumo. A popularidade está relacionada aos usuários que assistiram, consumiram ou compraram um determinado item/produto. Itens mais populares (ou mesmo mais vendidos) são aqueles que interessaram a grande parte da população e podem ser relevantes para um novo usuário. Por outro lado, as avaliações obtidas são derivadas de *ratings* atribuídos por usuários aos itens, de modo que itens com maiores *ratings* tendem a ser preferidos pelos usuários. O período de consumo está relacionado ao momento (i.e., hora, dia, mês ou ano) em que o item foi consumido pelos usuários no sistema. Recomendar itens que estão na moda, ou seja, que foram consumidos recentemente, pode agradar usuários que estão interagindo com o sistema neste exato momento.

Apesar dessas estratégias explorarem informações distintas, são utilizados em conjunto sob a premissa: *itens enviados pela popularidade, recência e avaliações positivas tem o potencial de satisfazer os interesses de grande parte dos usuários novos*. Ainda que esse pressuposto seja válido para a maioria dos usuários, existem outros usuários com preferências específicas, que não se interessam apenas pelos itens consumidos por grande parte de uma população. Hoje em dia, a diversidade de usuários tornou-se uma característica marcante da web, que abriga homens e mulheres, jovens, adultos ou idosos, sem distinções entre faixa etária, sexo, classe social, religião ou qualquer outra [Hammond et al., 2017]. Dessa forma, nem sempre os itens que agradam a uma grande parcela da população serão capazes de satisfazer todas as preferências dos diversos usuários existentes. Apresentar itens pouco atrativos a usuários novos faz com que esses acabem deixando o sistema sem proporcionar um retorno financeiro esperado [Pereira & Hruschka, 2015]. Além disso, pensando a longo prazo, uma interação inicial não promissora entre o usuário e o sistema pode levar a dois outros problemas: (1) o usuário pode ter uma experiência desagradável, ficando insatisfeito com o sistema por um longo período; e (2) o sistema pode modelar um perfil incorreto sobre o usuário, por ter consumido itens que não lhe agradam, e assim afetar em futuras recomendações. Por essas razões, torna-se necessária a utilização de estratégias adequadas para atrair os consumidores que ainda estão nessa primeira fase de compra. O desafio nesse caso é selecionar um pequeno subconjunto de itens potencialmente úteis para agradar o interesse de usuários com perfis dissonantes.

1.2 Objetivos

O principal objetivo deste trabalho é melhorar as recomendações geradas para usuários novos, na tentativa de atenuar o problema de *Pure Cold-Start* em cenários reais de recomendação. Nesse sentido, analisamos e implementamos SsR não-personalizados estado-da-arte e propomos duas novas abordagens complementares, baseadas na premissa de que diversificar os itens recomendados também é útil para os usuários novos. Para avaliar a validade de nossa suposição, consolidamos três hipóteses relacionadas:

Hipótese 1: Uma parcela significativa da população de usuários não está interessada apenas nos itens populares em domínios reais.

Hipótese 2: SsR não-personalizados que exploram outros itens além dos populares são capazes de atenuar o problema de *Pure Cold-Start*.

Hipótese 3: Páginas de produtos compostas por SsR não-personalizados complementares satisfazem o interesse de um grande número de usuários novos distintos.

A motivação para consolidarmos a hipótese 1 provém da observação de que existem muitos usuários com preferências que não estão relacionadas a popularidade dos itens [Hammar et al., 2013; Ghazanfar & Prügel-Bennett, 2014]. Especificamente, caracterizamos diferentes cenários reais sobre dois aspectos distintos. Primeiro, avaliamos se existem usuários que optaram por consumir itens não-populares durante sua interação com o sistema. Em seguida, verificamos se os interesses dos usuários correspondem a dimensões que não são necessariamente relacionadas a popularidade dos itens. A veracidade dessas suposições motiva a consolidação da nossa segunda hipótese, que se preocupa em identificar e recomendar adequadamente itens não populares para os usuários novos. O desafio é justamente encontrar itens não populares que sejam relevantes dada a falta de informações relacionadas a esses itens. Para isso, assumimos que os SsR devem recomendar itens não populares: (1) relevantes para um grande número de usuários distintos; e (2) representativos para diferentes perfis de usuários. A primeira estratégia chamamos de *Max-Coverage*, pois deriva do problema de *Maximum k-Coverage*, avaliado recentemente em cenários de recomendação [Puthiya Parambath et al., 2016]. Por sua vez, a segunda estratégia é chamada de *Niche-Coverage* pois deriva de estudos comportamentais que modelam os usuários em grupos com interesses em comum (i.e., nichos de usuários) [Pereira & Hruschka, 2015]. De maneira complementar, a terceira hipótese consolida a nossa proposta para melhorar as recomendações geradas para usuários novos. Nesse caso, o desafio é encontrar as estratégias que alinhadas são capazes de conquistar um maior número de usuários distintos.

Portanto, para investigar nossas hipóteses e alcançar nosso objetivo, propomos sete objetivos específicos, descritos a seguir:

1. Definir o problema de *Pure Cold-Start*, distinguindo-o do *Cold-Start*.
2. Levantar e analisar as principais abordagens utilizadas atualmente para lidar com o problema de *Pure Cold-Start*.
3. Investigar e validar a primeira hipótese deste trabalho por meio de uma caracterização sobre os principais cenários de recomendação.
4. Propor dois novos SsR não-personalizados capazes de explorar itens não-populares para conquistar usuários novos.
5. Investigar e validar a segunda hipótese deste trabalho por meio de experimentos empíricos capazes de comparar os SsR propostos com as abordagens estado-da-arte.
6. Propor e avaliar diferentes cenários com SsR não-personalizados alinhados para compor páginas de produtos em sistemas reais.
7. Investigar e validar a terceira hipótese deste trabalho por meio de um estudo *online* direcionado aos usuários, no intuito de comparar a melhor proposta com as combinações tradicionalmente utilizadas.

1.3 Principais Contribuições

A principal contribuição deste trabalho é ser um guia prático para todos os proprietários de aplicações *e-commerce* ou de sistemas de entretenimento, que lidam constantemente com a tarefa de atrair e satisfazer novos usuários. Além disso, destaca-se também a avaliação de hipóteses contra-intuitivas que ainda não foram estudadas na literatura. Com essas hipóteses, este trabalho é capaz de apontar quais subconjuntos de recomendações maximizam a chance de converter usuários novos em clientes ativos, bem como os pontos fortes e fracos de cada SsR não-personalizado. Como outras contribuições deste trabalho, pode-se destacar:

1. A distinção prática dos problemas de *Pure Cold-Start* e *Cold-Start* que são apresentados de maneira confusa na literatura.
2. A consolidação de três categorias de estratégias capazes de serem aplicadas para o problema de *Pure Cold-Start*.

3. Duas novas estratégias de recomendação não-personalizadas que podem ser aplicadas em domínios reais de *e-commerce* e entretenimento.
4. Uma metodologia de avaliação empírica extensível a qualquer contexto onde deseja-se comparar em detalhes distintos SsR.
5. A verificação da primeira hipótese, mostrando que existem usuários que não são satisfeitos apenas com os itens que interessam a grande parte da população.
6. A verificação da segunda hipótese de que itens não-populares podem ser igualmente utilizados para atrair e conquistar usuários novos.
7. Propostas práticas para proprietários de sistemas *Web* utilizarem para compor as páginas de produtos afim de conquistar um maior número de usuários novos.
8. A verificação da terceira hipótese, por meio de um experimento *online* capaz de identificar a combinação ideal para compor as páginas de produtos e atrair mais usuários novos.

Cabe ainda salientar que não encontramos na literatura trabalhos que avaliem essas hipóteses contra-intuitivas, de maneira similar à proposta neste trabalho. Além disso, também não foram encontrados trabalhos que investiguem o problema de *Pure Cold-Start* no nível de detalhes aqui abordado.

1.4 Organização do Documento

Este trabalho é organizado como se segue. O Capítulo 2 apresenta um referencial teórico da tarefa de recomendação, bem como das principais estratégias clássicas que são utilizadas. Ainda neste, definimos o problema de *Pure Cold-Start* e mostramos porque as estratégias clássicas não são aplicáveis a esse problema. No Capítulo 3, investigamos as principais estratégias de recomendação que são capazes de atenuar o problema de *Pure Cold-Start*. Para essas, destacamos a simplicidade e eficiência das estratégias não-personalizadas que são tradicionalmente utilizadas na prática. No Capítulo 4, caracterizamos os principais cenários de recomendação a fim de consolidar a primeira hipótese deste trabalho. Nesse capítulo, visamos responder a duas principais questões levantadas pela hipótese 1, que se tornam fundamentais para as nossas estratégias propostas. Além disso, nesse capítulo propomos as estratégias de *Max-Coverage* e *Niche-Coverage* no intuito de encontrar os itens não-populares relevantes aos usuários de cada domínio. Por sua vez, no Capítulo 5 avaliamos de maneira *offline* os SsR

não-personalizados propostos frente aos SsR estado-da-arte. Nesse capítulo, definimos uma metodologia de avaliação para contrastar cada uma das estratégias avaliadas por meio dos principais requisitos de qualidade necessários para um recomendador. Essa metodologia nos permite, também nesse capítulo, consolidar a segunda hipótese deste trabalho ao verificar a eficácia dos itens não-populares recomendados. No Capítulo 6 é construído um sistema de avaliação *online* capaz de simular o cenário do *Pure Cold-Start* e avaliar as interações de usuários reais. Nesse capítulo, é consolidada a terceira hipótese deste trabalho, apresentando as estratégias que devem ser alinhadas para atrair mais usuários novos. Por fim, o Capítulo 7 apresenta as conclusões deste trabalho, resumindo os resultados obtidos em cada capítulo anterior. Nesse também são apresentados alguns possíveis trabalhos futuros a serem realizados.

Capítulo 2

Referencial Teórico

Este capítulo visa distinguir o problema de *Cold-Start* do *Pure Cold-Start*. Primeiramente, revisamos a definição formal da tarefa de recomendação enfatizando a tarefa de prever os itens de interesse do usuário. Em seguida, enumeramos os métodos clássicos de recomendação utilizados pela literatura para a maioria dos cenários: Filtragem Colaborativa (CF), Filtragem Baseada em Conteúdo (CB) e os métodos Híbridos. Logo após, destacamos o problema de recomendar itens para usuários com poucas avaliações, denominado como *Cold-Start*, e o problema de recomendar itens para usuários novos, denominado *Pure Cold-Start*.

2.1 O Problema de Recomendação

Atualmente, Sistemas de Recomendação têm se tornado um mecanismo fundamental para aplicações de *e-commerce*, pois auxiliam os usuários a encontrarem seus itens e serviços favoritos [Abbassi & Mirrokni, 2007; Ramos, 2015; Mavlanova et al., 2016]. Em geral, o principal objetivo da recomendação é definido como encontrar/predizer, entre um número potencialmente grande de itens, aqueles que melhor se adequam aos interesses individuais de cada usuário [Mourão, 2014]. Formalmente, seja U o conjunto de todos os usuários de um sistema e I o conjunto de possíveis itens a serem recomendados (i.e., livros, filmes ou músicas), o objetivo da tarefa de recomendação consiste em encontrar um subconjunto de itens $R_u \subset I$, de tamanho k (i.e., $|R_u| = k$), que maximizam a função de utilidade $f(u, i)$ para cada usuário $u \in U$ e $i \in R_u$ [Adomavicius & Tuzhilin, 2005], como mostra a equação 2.1.

$$R_u = k \left[\arg \max_{i \in I} f(u, i) \right] \quad (2.1)$$

Intuitivamente, essa estimativa dos melhores itens é feita com base nos *ratings* (i.e., avaliações) atribuídos no passado pelo usuário alvo para outros itens do domínio. Com isso, cada elemento do espaço U pode ser interpretado como um perfil que inclui várias características do próprio usuário. De maneira análoga, cada elemento do espaço I pode ser interpretado com um conjunto de características de cada item. O principal desafio dos SsR é que a utilidade $f(u, i)$ é normalmente definida apenas para itens pertencentes a um subconjunto restrito do espaço $U \times I$. Isso significa que $f(u, i)$ tem de ser extrapolada para todos os demais itens do domínio [Adomavicius & Tuzhilin, 2005]. A Figura 2.1 apresenta um exemplo de cenário em que é necessário gerar recomendações para um usuário alvo, tendo como base apenas suas preferências passadas.

Figura 2.1: Exemplo de cenário de recomendação, com avaliações de 1 a 5.

			1	3		2	
	1	2	5		4	1	
	4			3			
	?	3	?	?	5	4	?
		3	4		5		3

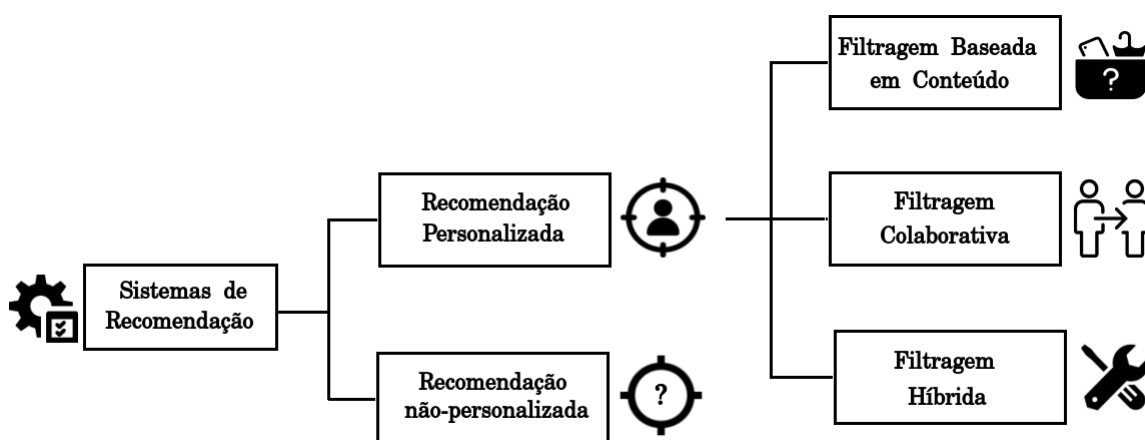
Em razão disso, o problema de recomendação é geralmente definido como o problema de estimar *ratings* para itens que ainda não foram *consumidos* pelos usuários. Esse processo consiste na tarefa de predição, onde o recomendador tenta prever *ratings* para os itens baseado em fatores relevantes dos usuários, como seu histórico de consumo ou mesmo suas características pessoais [Bobadilla et al., 2013]. Basicamente, o objetivo é minimizar a diferença entre o *rating* real, atribuído pelo próprio usuário, e o *rating* previsto, gerado pelo recomendador. Após essa etapa, a recomendação de um item para um usuário consiste na seleção de um grupo de itens com os melhores *ratings*. Em outras palavras, a recomendação pode ser feita selecionando os k melhores itens para um usuário, ou para um conjunto de usuários. Em alguns casos, não é possível estimar *ratings* para os itens, pois o recomendador utiliza outro critério para a seleção dos k melhores itens. Para esse tipo de recomendação dá-se o nome de tarefa de *ranking*, ou mesmo *Top-k*. Essa tarefa é recorrentemente abordada em diversas pesquisas e domínios comerciais, uma vez que o usuário nem sempre precisa visualizar o *rating* estimado [Hinz & Eckert, 2010]. Este trabalho centra-se especificamente nessa tarefa de *ranking*. Assim, por uma questão de notação e simplicidade, nas seguintes seções e capítulos, o termo recomendação refere-se especificamente à tarefa de *ranking*.

2.2 Métodos de Recomendação

Em SsR, diversas são as abordagens utilizadas para prever um conjunto de k itens para um usuário alvo [Jannach et al., 2010; Candillier et al., 2009; Ricci et al., 2011; Bobadilla et al., 2013]. De maneira geral, as principais abordagens são divididas em duas classes: personalizadas e não-personalizadas [Akshita & Smita, 2013]. Técnicas personalizadas são aquelas que levam em conta as informações relacionadas a um usuário, tentando sempre modelar um perfil para este, a fim de encontrar recomendações que melhor se adequem a este perfil. Por sua vez, técnicas não-personalizadas são simples, pois não consideram as informações existentes sobre o usuário alvo, recomendando apenas com base nas preferências dos usuários em geral. Repare que, por questões óbvias, não se pode comparar a aplicabilidade de técnicas não-personalizadas frente a técnicas personalizadas, sendo cada classe aplicada a distintos cenários.

Dentre as técnicas personalizadas, existem três subclasses que levam em conta o perfil do usuário alvo, e se distinguem devido às suas abordagens: Filtragem Baseada em Conteúdo (CB), Filtragem Colaborativa (CF) e Métodos Híbridos [Bobadilla et al., 2013; Beel et al., 2016]. Enquanto CB tenta recomendar com base nas características dos itens consumidos, CF tenta correlacionar usuários com preferências em comum para gerar as recomendações [Schafer et al., 2007]. Por sua vez, os métodos híbridos combinam ambas as abordagens a fim de melhorar as recomendações geradas para os usuários [Kawai & Nogami, 2016]. A Figura 2.2 ilustra toda essa divisão de classes dos SsR, que são apresentadas em mais detalhes nas subseções a seguir.

Figura 2.2: Visão geral das classes de técnicas de Sistemas de Recomendação.

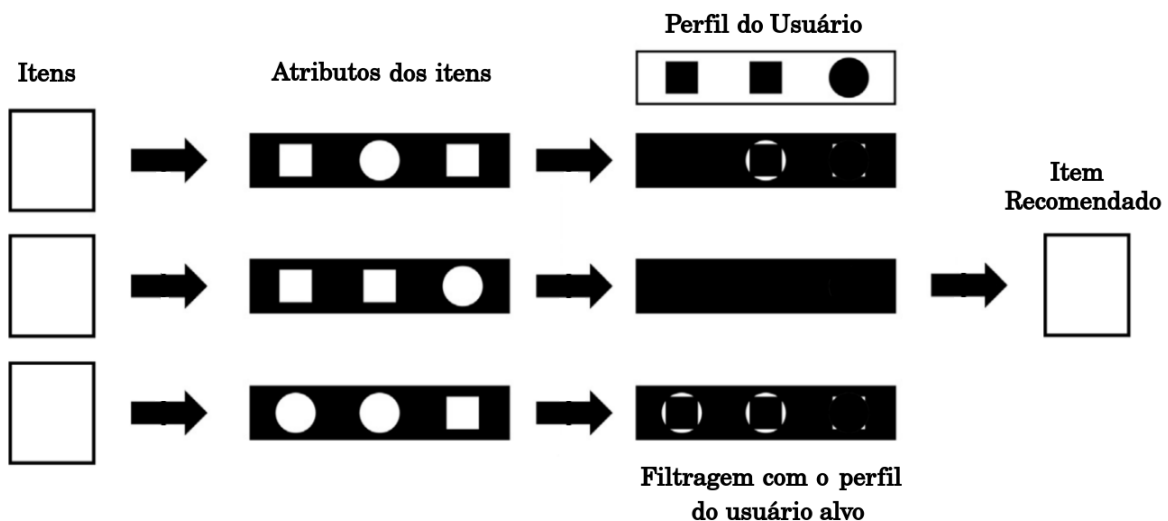


2.2.1 Filtragem Baseada em Conteúdo

Na sua definição formal, pode-se dizer que estratégias Baseadas em Conteúdo procuram estimar uma função de utilidade $f(u, i)$, do item i para o usuário u , por meio de uma função conhecida $f(u, j)$, que foi definida pelo mesmo usuário u a um item j similar a i [Van Meteren & Van Someren, 2000]. Em outras palavras, para estimar um *rating* que o usuário u atribuiria para o item i , esta estratégia consiste em encontrar itens similares a i e utilizar essas informações prévias como base para a recomendação. O diferencial dessa estratégia para as demais é que a similaridade de um item é estimada utilizando as características (i.e., atributos) relacionadas aos itens. A premissa dessa abordagem é que itens com atributos similares serão avaliados de maneira similar, pois cada usuário exibe uma preferência sistemática correlacionada com os atributos dos itens [Pazzani, 1999; Schafer et al., 2007; Ricci et al., 2011].

Nesses casos, um item i é descrito como um vetor $X_i = (x_1, x_2, \dots, x_n)$ de n atributos, que podem ser binários, nominais ou numéricos. Em um cenário de filmes, por exemplo, esses atributos podem ser informações como atores, diretores, gêneros, ano de lançamento, sinopse e outros. De maneira similar, o perfil de um usuário u é modelado com base nos atributos dos itens consumidos por u no passado, sendo formalmente descrito como um vetor $Y_u = (y_1, y_2, \dots, y_n)$, onde cada elemento é uma combinação dos atributos dos itens. Dessa forma, a recomendação consiste em filtrar todos os itens do domínio, na tentativa de encontrar o item que os atributos X combinem com o perfil Y do usuário. Esse processo é descrito visualmente em [Bobadilla et al., 2013], tal qual mostra a Figura 2.3.

Figura 2.3: Processo de recomendação baseado em conteúdo.



Conforme proposto por Bobadilla et al. [2013], o processo de recomendação baseada em conteúdo pode ser sumarizado em três passos:

1. **Extração dos atributos relacionados aos itens do domínio:** procura-se por informações capazes de descrever os itens do domínio, tais como gêneros, atores, ano de lançamento e outros, no caso de filmes.
2. **Comparação dos atributos dos itens com a preferência do usuário alvo:** procura-se por itens que se assemelhem ao perfil do usuário modelado com base nas características dos itens consumidos por esse.
3. **Recomendação dos itens:** procura-se apresentar os itens com características que interessem ao usuário alvo da recomendação.

Quando os atributos dos itens e os perfis dos usuários podem ser modelados, o desafio dessa abordagem é determinar o quanto um usuário se interessará por um item específico [Van Meteren & Van Someren, 2000]. Em geral, essa tarefa pode ser resolvida por estratégias heurísticas tradicionais [Baeza-Yates et al., 1999; Garcia & Amatriain, 2010] ou mesmo algoritmos de classificação como Regra de Indução [Cohen et al., 1995; Kim et al., 2001], Métodos dos Vizinhos Mais Próximos [Billsus et al., 2000], Algoritmo de Rocchio [Balabanović & Shohom, 1997], Classificadores Lineares [Joachims, 1998] e Métodos Probabilísticos [Park & Tuzhilin, 2008; De Gemmis et al., 2008]. Lops et al. [2011] apresentam detalhes dos métodos estado-da-arte para recomendação baseada em conteúdo. De certo modo, essas estratégias estão intimamente ligadas à área de Recuperação de Informação, uma vez que ambas estão relacionadas a avaliar as descrições dos itens por suas características [Jannach et al., 2010]. A suposição em comum é que os usuários são capazes de formular consultas que expressam seus interesses ou necessidades de informação em termos de atributos intrínsecos dos itens [Ramos et al., 2003].

2.2.2 Filtragem Colaborativa

Ao contrário das estratégias baseadas em conteúdo, Filtragem Colaborativa tenta prever a utilidade de um item i para um usuário u de acordo com o *feedback* atribuído pelo usuário a itens similares a i , ou o *feedback* atribuído por usuários similares a u [Ricci et al., 2011]. Mais formalmente, a função de utilidade $f(u, i)$, do item i para o usuário u , é estimada baseada na função $f(v, i)$ atribuída por um usuário v ao mesmo item i , desde que v tenha um limiar de similaridade com relação a u [Adomavicius & Tuzhilin, 2005]. Pode-se também levar em conta o valor obtido por $f(u, j)$, desde que j tenha um limiar de similaridade com relação a i [Bobadilla et al., 2013].

De acordo com Yang et al. [2014], as técnicas de CF podem ser separadas em dois grupos: *memory-based* e *model-based*. Estes dois grupos podem ser orientados segundo o usuário e segundo o item, tal qual sumarizado na tabela 2.1.

Tabela 2.1: Classes de estratégias de recomendadores de filtragem colaborativa.

Classes de Estratégia		
	Memory-based	Model-based
User-based	Combina as preferências dos k usuários mais similares, com características relacionadas.	Explora o histórico de preferências do usuário para treinar modelos de recomendação.
Item-based	Combina as avaliações dos k itens mais similares, considerando todos os usuários.	Explora as avaliações passadas do item para treinar modelos de recomendação.

Algoritmos *memory-based* são essencialmente heurísticas que fazem a predição de *ratings* baseando-se nos itens previamente avaliados pelos usuários [Adomavicius & Tuzhilin, 2005; Ricci et al., 2011]. Em geral, existem dois tipos de modelagens: *user-based* e *item-based*. As técnicas *user-based* procuram agregar os *ratings* atribuídos pelos N usuários mais similares ao mesmo usuário alvo da recomendação. Por outro lado, as técnicas *item-based* visam agregar os *ratings* recebidos pelos N itens mais similares ao item alvo da recomendação. Essa agregação de *ratings* pode ser feita por estratégias simples, como a média dos demais *ratings*, ou mesmo como estratégias ponderadas que levam em conta as funções de similaridade. Destacam-se a aplicação de algoritmos baseados em grafos [Huang et al., 2002; Lo & Lin, 2006; Silva et al., 2010] e abordagens relacionadas ao Vizinheiro Mais Próximo [Deshpande & Karypis, 2004; Dong et al., 2011], como o *UserKNN* e *ItemKNN* [Grčar et al., 2006; Campos et al., 2010].

Algoritmos *model-based* usam a coleção de *ratings* para aprender um modelo que é usado para fazer predições. Muitos desses métodos são inspirados em algoritmos de aprendizado de máquina, como os classificadores de redes neuronais [Billsus & Pazzani, 1998], a aprendizagem de regras de indução [Basu et al., 1998], as redes Bayesianas [Horvitz et al., 1998] e os modelos de fatores latentes [Bell & Koren, 2007]. Os métodos de fatores latentes representam uma das abordagens mais eficazes e populares de CF *model-based*, uma vez que são geralmente efetivos na estimativa da estrutura geral que relaciona, simultaneamente, a maioria ou todos os itens [Sarwar et al., 2000]. De maneira similar aos algoritmos *memory-based*, essas técnicas também pode ser avaliadas quanto a modelagens *user-based* ou *item-based*, sobre os mesmos pressupostos anteriores.

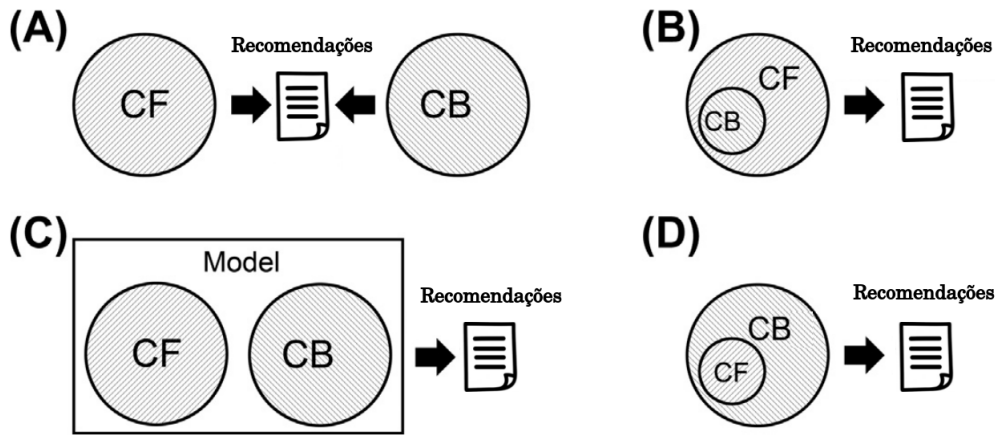
2.2.3 Métodos Híbridos

Em geral, as estratégias híbridas visam explorar as vantagens de cada abordagem clássica [Adomavicius & Tuzhilin, 2005]. A maioria das técnicas existentes podem ser classificadas conforme proposto por Martin et al. [2014], e ilustrado por Bobadilla et al. [2013] como na Figura 2.4:

- (A) **Combinando Recomendadores Separados:** implementam-se as estratégias de CF e CB separadamente e combinam-se suas previsões por meio de uma combinação linear dos *ratings* ou mesmo por meio de uma função de qualidade que determina, para um dado momento, qual é a melhor recomendação a se utilizar [Billsus & Pazzani, 2000; Kim et al., 2011].
- (B) **Adicionando CB em CF:** procuram-se incorporar algumas características de um CB em técnicas de CF, como por exemplo, integrando atributos de conteúdo no perfil do usuário (ou itens) no cálculo da função de similaridade entre usuários (ou itens) [Melville et al., 2002; Li & Kim, 2003; Hu & Pu, 2010].
- (C) **Modelo de Recomendação Unificado:** constrói-se um modelo que incorpora ambas as estratégias de CB e CF, como por exemplo, modelos capazes de englobar informações do conteúdo e também das preferências passadas dos usuários [Popescul et al., 2001; De Campos et al., 2010; Choi et al., 2012].
- (D) **Adicionando CF em CB:** procuram-se incorporar algumas características de uma CF em técnicas CB, como por exemplo, estratégias de redução de dimensionalidade aplicadas a grande coleção de conteúdos existentes [Mooney & Roy, 2000].

Diferentes estratégias de hibridização podem ser encontradas na literatura. Khrouf & Troncy [2013] apresentam um método híbrido para recomendações de eventos. Além de um sistema CB que supera a escassez de dados, esse método inclui um modelo de CF para modelar aspectos sociais. Por sua vez, Hussein et al. [2014] apresentam o *Hybreed*, um *software* para a construção de aplicações complexas de reconhecimento de contexto, juntamente com um conjunto de características de CF. Kouki et al. [2015] propõem o Hyper (Hybrid Probabilistic Extensible Recommender), um *framework* capaz de aprender automaticamente a equilibrar diferentes fontes de informação, como múltiplas medidas de similaridade entre usuários e itens, conteúdo e informações sociais, para prever *ratings*. Por fim, Kawai & Nogami [2016] apresentam o HCCF (*Hybrid Method of Collaborative and Content-Based Filtering*), que utiliza um modelo de *non-Negative Matrix Factorization* (NMF) associado com os atributos de um CB.

Figura 2.4: Processo de recomendação das estratégias híbridas.



2.3 Desafios aos Métodos Clássicos

Em sua maioria, os métodos de Sistema de Recomendação consistem de estratégias semi-supervisionadas que exploram o conhecimento prévio dos usuários e itens para gerar recomendações. Tanto os métodos de Filtragem Colaborativa, quanto os métodos de Filtragem Baseada em Conteúdo, utilizam o histórico de consumo (i.e., informações passadas) para modelar o perfil do usuário alvo. Por essa razão, os principais recomendadores clássicos sofrem com problemas relacionados à ausência de informações de um usuário ou item do domínio. Esses dois cenários prefiguram como problemas de destaque para os recomendadores atuais.

2.3.1 O Problema de Cold-Start

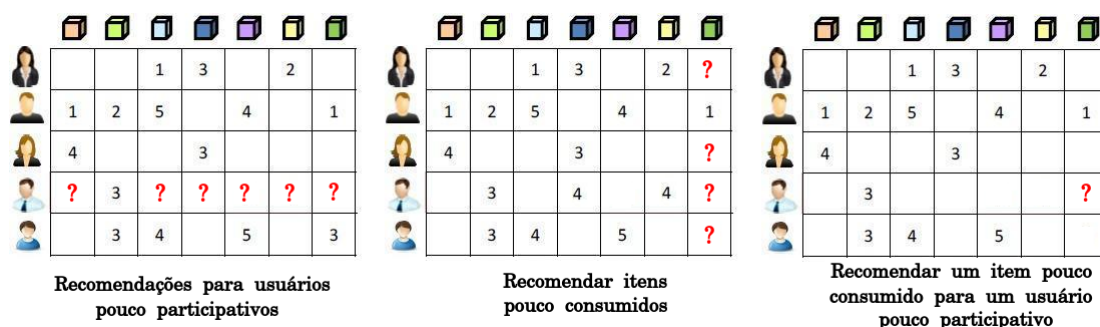
Um dos problemas mais comuns relacionados a falta de informação em Sistemas de Recomendação é o problema de *Cold-Start* [Schafer et al., 2007; Adomavicius & Tuzhilin, 2005]. Esse problema ocorre quando não é possível realizar recomendações potencialmente úteis devido a um baixo número de *ratings* relacionados aos usuários e/ou itens do domínio [Bobadilla et al., 2012]. Inicialmente, o termo *Cold-Start* foi utilizado para referenciar o problema de recomendação em cenários que não existe nenhum *rating* associado ao item ou usuário [Schein et al., 2002; Burke, 2002]. Entretanto, recentes trabalhos referem-se a *Cold-Start* como o problema de recomendar em cenários que existem poucas avaliações, i.e., *ratings* associados aos itens ou usuários do domínio [Bobadilla et al., 2012; Lika et al., 2014; Barjasteh et al., 2015; Sedhain et al., 2014; Wongchokprasitti et al., 2015; Hernando et al., 2017].

O problema de *Cold-Start* pode ser identificado sob três categorias: (1) gerar

recomendações para usuários pouco participativos; (2) recomendar itens que foram pouco consumidos; e (3) recomendar um item pouco consumido para um usuário pouco participativo [Lika et al., 2014]. Todos esses subproblemas prejudicam na tarefa de predição dos recomendadores e levam a maiores problemas para os sistemas. Usuários pouco participativos não conseguem receber recomendações úteis às suas preferências, ficando com a sensação de que o sistema não fornece o serviço esperado. Por outro lado, itens que foram pouco consumidos não são susceptíveis a serem recomendados e acabam ficando esquecidos no sistema, podendo causar prejuízos ao proprietário. A Figura 2.5 exemplifica essas três maneiras, mostrando a falta de informações sobre os usuários e/ou itens.

Na literatura, diversos trabalhos lidam com o problema de *Cold-Start*. Para o *Cold-Start* relacionado ao item, as primeiras propostas consistem em utilizar os métodos de CB, por não dependerem das informações fornecidas pelos usuários [Melville et al., 2002; Linden et al., 2003]. Entretanto, essa abordagem acaba levando ao problema de superespecialização, quando esses itens pouco consumidos são recomendados apenas a usuários que gostaram de itens similares [Pazzani & Billsus, 2007]. Por essa razão, diversos métodos híbridos, que combinam CF com CB, são propostos para esse problema [Schein et al., 2002; De Campos et al., 2010; Saveski & Mantrach, 2014]. De maneira similar, para o problema de *Cold-Start* relacionado ao usuário, são propostos métodos híbridos capazes de personalizar as recomendações para esses usuários pouco participativos [Basiri et al., 2010; Bobadilla et al., 2012]. Nesse caso, também é comum a utilização de modelos adaptados a outras informações dos usuários, a fim de aumentar o conhecimento desses [Sun et al., 2011; Saveski & Mantrach, 2014]. Por fim, para o *Cold-Start* relacionado ao item e usuário, poucas abordagens são propostas. Nesse caso, destaca-se um modelo de regressão baseado em características preditivas que exploram toda a informação disponível de usuários e itens [Park & Chu, 2009].

Figura 2.5: As três maneiras de representação do problema de *Cold-Start*.



2.3.2 O Problema de Pure Cold-Start

Na literatura, o problema de *Pure Cold-Start* é muitas vezes referenciado de maneira confusa com o problema de *Cold-Start*. Originalmente, o termo *Pure Cold-Start* referia-se ao problema de recomendar quando não existem informações suficientes sobre os itens ou usuários [Schein et al., 2002; Burke, 2002]. Entretanto, como recentes trabalhos [Barjasteh et al., 2015; Sedhain et al., 2014; Wongchokprasitti et al., 2015; Hernandez et al., 2017] relacionam *Cold-Start* ao problema de esparsidade dos dados, *Pure Cold-Start* passou a ser utilizado para descrever esse cenário inicial do sistema, quando não existe nenhuma informação disponível [Shah & Sahu, 2014; Shi, 2016]. Por essa razão, *Pure Cold-Start* pode ser visto como uma derivação do problema de *Cold-Start*.

Dessa forma, o problema de *Pure Cold-Start* também pode ser identificado de três maneiras, conforme ilustrado na Figura 2.6: (1) gerar recomendações para usuários novos, que não possuem informações; (2) recomendar itens novos, que nunca foram consumidos; e (3) recomendar um item novo para um usuário novo [Nguyen et al., 2007]. Quando usuários acessam pela primeira vez um sistema, os SsR não são capazes de prever as preferências desses, visto que não existe um histórico de consumo relacionados a eles. O mesmo pode ser dito para novos itens que são adicionados no catálogo de produtos. Por não existirem avaliações sobre esses novos produtos, os recomendadores não conseguem correlacioná-lo com as preferências dos usuários. Além disso, se torna ainda mais desafiador, quase improvável, recomendar um item novo a um usuário novo no domínio.

Contudo, apesar dessa relação com o problema de *Cold-Start*, as soluções aplicadas a esse não podem ser adaptadas para o *Pure Cold-Start* devido à inexistência de qualquer informação sobre os usuários e/ou itens. Por essa razão, torna-se necessário analisar em detalhes as soluções aplicáveis ao problema de *Pure Cold-Start*. Esse problema é o foco deste trabalho, que visa avaliar o cenário de usuários novos, por ser um dos desafios mais relevantes atualmente para sistemas *Web*.

Figura 2.6: As três maneiras de representação do problema de *Pure Cold-Start*.



2.4 Síntese do Capítulo

Este capítulo começa apresentando formalmente o problema geral de recomendação, bem como as tarefas inerentes a ele. Basicamente, o desafio é encontrar um conjunto de itens que maximizam a função utilidade relacionada a cada usuário. Além disso, restringimos o escopo desta dissertação para a tarefa de *ranking*, que visa recomendar k itens para cada usuário. Posteriormente, recordamos as classes de recomendadores tradicionalmente utilizadas como base das principais abordagens existentes. Nessa etapa, mostramos que existem os recomendadores focados em cada usuário individualmente (i.e., personalizados) e os que não se preocupam com apenas um usuário, gerando as mesmas recomendações para todos (i.e., não-personalizados). Dentre os personalizados, analisamos em detalhes as três abordagens clássicas que utilizam as informações passadas dos usuários para gerar recomendações.

Em seguida, vimos que todos os métodos personalizados sofrem com os problemas relacionados a ausência de informações passadas dos usuários. Assim, definimos em detalhes o problema de *Cold-Start* como a tarefa de recomendar itens aos usuários pouco participativos, que consumiram poucos itens do sistema. Desse problema geral, derivamos o problema de *Pure Cold-Start*, que considera esse mesmo problema em sua forma substancial. Nesse caso, a tarefa é recomendar itens a usuários novos, que não possuem nenhuma informação pessoal associada e nem mesmo consumiram sequer um item do sistema. Por essa razão, esse problema é ainda mais desafiador, sendo o foco deste trabalho. No próximo capítulo estudamos as principais estratégias utilizadas para atenuar esse problema.

Capítulo 3

Estudo do Pure Cold-Start

Este capítulo tem por objetivo investigar o problema de *Pure Cold-Start*, enfatizando a utilização de estratégias não-personalizadas em distintos cenários reais. Como mostrado no capítulo anterior, apesar do problema de *Pure Cold-Start* derivar do *Cold-Start*, as soluções aplicadas para o *Cold-Start* não podem ser adaptadas ao *Pure Cold-Start*. Por essa razão, primeiramente, levantamos as principais abordagens existentes na literatura para lidar com o problema em estudo. Devido à praticidade e eficácia, as estratégias não-personalizadas são comumente utilizadas. Por isso, definimos em detalhes os tradicionais recomendadores que são utilizados pelas aplicações de *e-commerce* para atenuar o problema de *Pure Cold-Start*, colocando-os como *baselines* deste trabalho.

3.1 Principais Abordagens Relacionadas

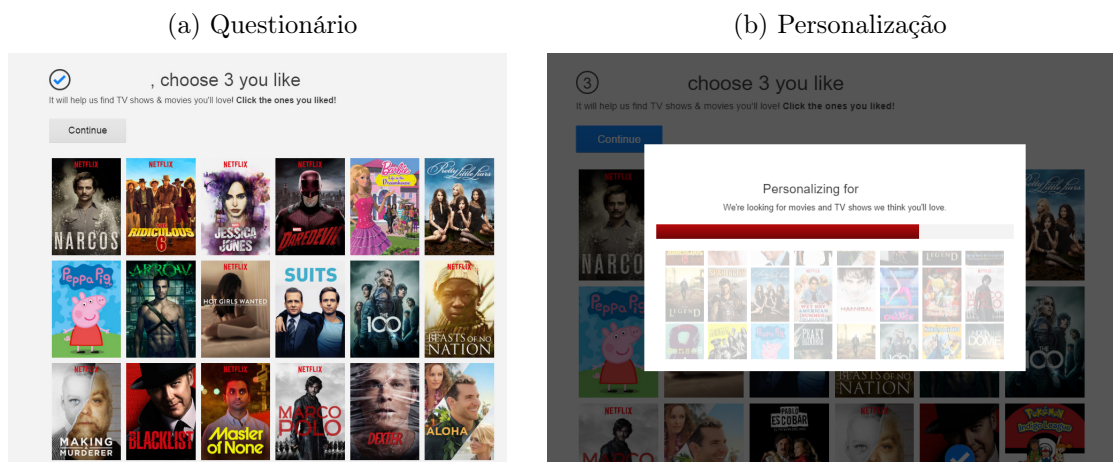
Neste trabalho, o termo *Pure Cold-Start* é adotado para denominar o problema de recomendar itens para usuários novos, i.e., usuários que ainda não avaliaram nenhum item do domínio e sequer forneceram informações pessoais. Como destacado no Capítulo 2, as estratégias clássicas de recomendação não são capazes de lidar com esse problema. Em geral, para atenuar esse problema, existem na literatura três principais categorias de SsR utilizadas: (1) SsR Interativos; (2) SsR Híbridos; e (3) SsR não-Personalizados. Em geral, as categorias (1) e (2) propõem utilizar informações externas relacionadas aos usuários, tais como informações demográficas ou mesmo pessoais como sexo, faixa etária, e outras. Basicamente, essas informações são agregadas às estratégias de Filtragem Colaborativa ou Baseada em Conteúdo. A categoria (3) propõe utilizar as informações globais do sistema para recomendar a usuários novos. Essa categoria diverge completamente das demais abordagens, por ignorarem o usuário alvo e recomendarem os mesmos itens para todos.

3.1.1 Recomendadores Interativos

Recomendadores Interativos foram propostos em [Aha & Breslow, 1997; Ardissono et al., 2003; Reilly et al., 2005] como estratégias que suportam interações humano-computador para personalizar a recomendação. Em geral, duas fases podem ser identificadas: (1) construção de um perfil provisório para o usuário e (2) geração de uma recomendação personalizada [Mahmood & Ricci, 2007]. O perfil provisório do usuário é adquirido por meio de interações prévias entre o usuário alvo e o sistema, durante a sessão de recomendação, como um *feedback* explícito. A partir desta interação, a fase de geração de recomendação consiste em identificar um subconjunto de itens que combine com o perfil provisório do usuário. Em geral, essa fase de interação usuário-sistema ocorre por meio de um questionário em que o sistema faz as perguntas e o usuário responde de acordo com sua preferência pessoal [Thompson et al., 2004].

No sistema da *Netflix*¹, por exemplo, essa estratégia é utilizada para lidar com usuários novos. A primeira interação do usuário com o sistema é para fornecer um conjunto de respostas a serem utilizadas como fonte de conhecimento inicial para o sistema. A Figura 3.1 esboça essas duas etapas características de Recomendadores Interativos, no caso da *Netflix*. Primeiro, o usuário deve especificar três filmes que representem a sua preferência pessoal, tal qual mostra a Figura 3.1a. Em seguida, o sistema modela essas respostas como um perfil provisório do usuário e gera as recomendações com base nesse *feedback*, tal qual mostra a Figura 3.1b. A partir dessa primeira interação, o sistema passa a utilizar os recomendadores clássicos para fidelizar esses usuários.

Figura 3.1: Exemplo do sistema da Netflix que utiliza um Recomendador Interativo para ser capaz de gerar recomendações para seus usuários *novos*.



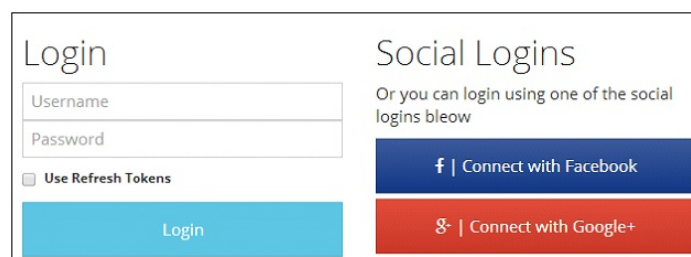
¹Disponível em: www.netflix.com

Na literatura, diversos trabalhos utilizam Recomendadores Interativos para lidar com o problema de *Pure Cold-Start* [Zhou et al., 2010; Mahmood & Ricci, 2007; Steck et al., 2015; He et al., 2016]. Zhou et al. [2010] propõe uma estratégia capaz de refinar o processo de Fatorização de Matrizes de acordo com as respostas fornecidas pelos usuários ao responderem alguns questionários. Em [Mahmood & Ricci, 2007] é apresentado um recomendador interativo capaz de aprender de maneira autônoma para auxiliar usuários em um sistema de viagens. Por sua vez, em [Steck et al., 2015], os autores apresentam um tutorial com vários aspectos fundamentais para um Recomendador Interativo ser capaz de atrair e simplificar o trabalho do usuário. Recentemente, He et al. [2016] apresentaram uma discussão dos principais Recomendadores Interativos, apontando os principais desafios relacionados a essas abordagens. Os autores argumentam que, assim como as máquinas de busca, a qualidade das recomendações depende da informação fornecida pelos usuários, que podem não definir claramente suas preferências e acabar tendo uma experiência negativa do sistema. Além disso, tais sistemas acabam não sendo as melhores opções para aplicações *e-commerce*, onde usuários não estão interessados em responder questões, mas apenas realizar uma compra.

3.1.2 Recomendadores Híbridos

Esta categoria de Recomendadores Híbridos se refere aos Métodos Híbridos revisados no Capítulo 2, que utilizam as abordagens clássicas de recomendação acrescidas de informações externas à aplicação, tal qual proposto em [Burke, 2002]. Essas informações externas estão relacionadas a qualquer outra informação sobre o usuário, como informações sociais, demográficas ou pessoais (e.g., sexo, idade, e outros). A premissa dessa categoria é que essas informações podem ajudar a identificar as preferências dos usuários novos, sem que eles nunca tenham interagido com o sistema. Em muitos cenários, os usuários são convidados a acessarem as aplicações de *e-commerce* por meio de suas contas nas redes sociais, como mostra a Figura 3.2, para que o sistema adquira um conhecimento inicial e modele um perfil prévio do usuário.

Figura 3.2: Um exemplo de *login* com cadastros de redes sociais.



The image shows a login interface with two main sections. On the left, under the heading "Login", there are two input fields for "Username" and "Password". Below these fields is a checkbox labeled "Use Refresh Tokens". A large blue button labeled "Login" is positioned at the bottom of this section. On the right, under the heading "Social Logins", there is a text prompt: "Or you can login using one of the social logins below". Below this text are two buttons: a blue button with the Facebook logo and the text "Connect with Facebook", and a red button with the Google+ logo and the text "Connect with Google+".

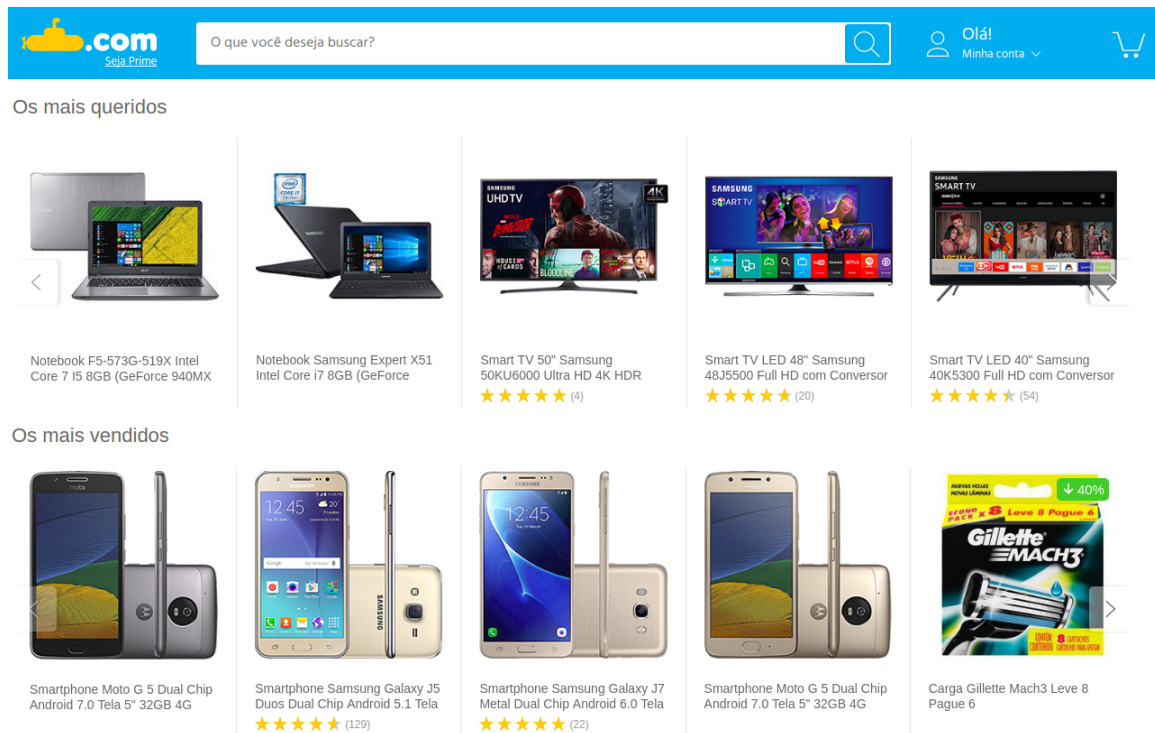
Essa categoria é chamada de Recomendadores Híbridos pois utiliza técnicas que combinam as estratégias clássicas de recomendação a essas informações externas [Zheng et al., 2017]. Em [Chikhaoui et al., 2011], os autores apresentam as vantagens de combinar as estratégias de Filtragem Colaborativa acrescida de informações demográficas com estratégias Baseadas em Conteúdos para atenuar o problema de *Pure Cold-Start*. Em [Safoury & Salah, 2013] também é apresentada uma estratégia híbrida *CF-demographic*, que correlaciona usuários novos com usuários ativos no sistema por meio de informações demográficas, tais como idade, sexo e ocupação profissional. Por outro lado, Carrer-Neto et al. [2012] apresentam uma estratégia híbrida *CF-social information*, que utiliza as informações obtidas pelos usuários através de redes sociais para melhorar a qualidade das recomendações. Em [Rosli et al., 2015], os autores usam comentários postados nas publicações do *Facebook* para modelar um perfil provisório a ser utilizado na etapa de recomendação. Entretanto, apesar das vantagens obtidas para o problema de *Pure Cold-Start* por essa categoria de estratégias, essas abordagens são limitadas à existência de informações externas. Em muitas aplicações de *e-commerce*, os usuários não estão interessados em vincular suas redes sociais a alguns sistemas por motivos de privacidade ou mesmo para não receber notícias e promoções. Além disso, existem também muitos usuários que não utilizam redes sociais ou mesmo possuem poucas informações pessoais cadastradas, dificultando o recomendador.

3.1.3 Recomendadores não Personalizados

A terceira categoria de estratégias relacionadas ao problema de *Pure Cold-Start* é dos Recomendadores não personalizados. Como o próprio nome sugere, estratégias não personalizadas geram recomendações que são independentes do usuário alvo, apresentando os mesmos itens para todos os usuários do domínio [Schafer et al., 1999]. Nessa categoria, as estratégias se baseiam nas informações globais dos itens ou usuários do sistema [Bobadilla et al., 2013]. Em uma analogia, pode-se enxergar as estratégias não personalizadas como a vitrine de uma loja, onde o comerciante coloca suas melhores opções não pensando apenas em uma pessoa, mas sim em atrair o maior número de pessoas possíveis. Esse pressuposto, associado à simplicidade, independência de domínio e eficiência dessas estratégias, faz com que diversas aplicações de *e-commerce* apostem em estratégias não personalizadas como mecanismo para lidar com o problema de *Pure Cold-Start*. A Figura 3.3, retirada da loja do *Submarino*², apresenta um exemplo de estratégias não personalizadas comumente utilizadas no mercado de vendas, apresentando os produtos mais preferidos e mais vendidos.

²Disponível em: www.submarino.com.br

Figura 3.3: Um exemplo de Recomendadores não Personalizados utilizados na prática.



Os mais queridos

- Notebook F5-573G-519X Intel Core 7 I5 8GB (GeForce 940MX)
- Notebook Samsung Expert X51 Intel Core i7 8GB (GeForce)
- Smart TV 50" Samsung 50KU6000 Ultra HD 4K HDR (4)
- Smart TV LED 48" Samsung 48J5500 Full HD com Conversor (20)
- Smart TV LED 40" Samsung 40K5300 Full HD com Conversor (54)

Os mais vendidos

- Smartphone Moto G 5 Dual Chip Android 7.0 Tela 5" 32GB 4G
- Smartphone Samsung Galaxy J5 Duos Dual Chip Android 5.1 Tela (129)
- Smartphone Samsung Galaxy J7 Metal Dual Chip Android 6.0 Tela (22)
- Smartphone Moto G 5 Dual Chip Android 7.0 Tela 5" 32GB 4G
- Carga Gillette Mach3 Leve 8 Pague 6 (40% off)

Ao contrário das demais categorias, os Recomendadores não-Personalizados não dependem de informações extras dos usuários novos, mas apenas das relações usuários-itens do próprio domínio. Por essa razão e devido à simplicidade, capacidade de generalização, independência do domínio e eficiência, essas estratégias são adotadas em diversos cenários [Bobadilla et al., 2013]. Este trabalho se restringe às estratégias não-personalizadas, uma vez que em muitos casos os usuários não estão dispostos a fornecer informações pessoais para a sua primeira compra.

3.2 Estratégias não-Personalizadas Tradicionais

A maioria das estratégias não-personalizadas utilizadas em aplicações *e-commerce* para usuários novos exploram três características tradicionais dos itens: (1) a popularidade; (2) os *ratings* recebidos; e (3) a recência de consumo [Poriya et al., 2014]. Ao recomendar tais itens, cada estratégia tem como premissa que esses itens são potencialmente relevantes para esse tipo de usuário. Em geral, o consumo é enviesado para essas dimensões onde SsR são aplicáveis, permitindo uma alta eficácia dessas abordagens. Tais estratégias são descritas ao longo desta seção, que utiliza a Figura 2.1 como modelo base para exemplificar cada abordagem.

3.2.1 Popularity

Popularity é a principal estratégia não-personalizada utilizada em cenários reais de aplicações *e-commerce* [Bobadilla et al., 2013]. Basicamente, essa estratégia visa recomendar os k itens mais populares do domínio, sobre o pressuposto de que itens que interessam a um grande número de usuários podem cobrir as distintas preferências existentes. A popularidade de um item i é estimada pela quantidade de usuários distintos que consumiram i . Essa estratégia, em geral, tende a apresentar uma alta eficácia prática por ser uma estratégia conservadora, focada em itens que tendem a produzir um maior número de acertos. A Figura 3.4 apresenta a popularidade de cada item do domínio da Figura 2.1, bem como o *ranking* gerado para se recomendar dois itens.

Figura 3.4: Exemplo de aplicação da estratégia de *Popularity* sobre o domínio.



3.2.2 Best Rated

Outra estratégia muito utilizada em cenários reais, *Best Rated* consiste na estratégia de recomendar os k itens mais bem avaliados do domínio. Basicamente, calcula-se a nota média de cada item considerando o total de avaliações recebidas pelo item e recomendam-se os k itens com maiores notas. O pressuposto dessa abordagem é que os itens mais bem avaliados tendem a interessar mais usuários distintos. Assim como a *Popularity*, trata-se de uma estratégia conservadora e também focada em acertos. A Figura 3.5 apresenta a nota média obtida por cada item do domínio da Figura 2.1, bem como o *ranking* gerado para se recomendar dois itens.

Figura 3.5: Exemplo de aplicação da estratégia de *Best Rated* sobre o domínio.



3.2.3 Recent Items

Recent Items consiste em recomendar os k últimos itens consumidos pelos usuários no domínio. O pressuposto dessa abordagem é que os usuários tendem a se interessar pelos itens que estão na moda, sendo consumidos atualmente. Em cenários como o de vendas de roupas, joias e afins, essa abordagem se mostra eficaz [Pathak et al., 2010]. Na prática, essa estratégia depende da informação da sessão de consumo dos itens, que indica quando cada item foi consumido. A Figura 3.6 apresenta os itens recomendados, considerando uma informação externa à matriz de *ratings* apresentada na Figura 2.1. Em geral, esse valor de consumo é apresentado em segundos e é conhecido como *timestamp* nas bases de dados tradicionais. Considerando que grande parte dos cenários reais possuem vários usuários, com distintas preferências, interagindo com o sistema simultaneamente, as recomendações tendem a ser mais diversificadas.

Figura 3.6: Exemplo de aplicação da estratégia de *Recent Items* sobre o domínio.



3.3 Síntese do Capítulo

Neste capítulo, apresentamos as principais abordagens capazes de atenuar o problema de *Pure Cold-Start*. Primeiramente, analisamos os SsR Interativos que buscam identificar o perfil de cada usuário novo por meio de questionários realizados. Em seguida, analisamos os SsR Híbridos que utilizam informações provenientes de redes sociais como *Facebook*, *Gmail* e outras. Por fim, analisamos os SsR não-personalizados que utilizam informações existentes no sistema para gerar recomendações independente dos usuários. A principal desvantagem das duas primeiras abordagens é que elas são dependentes dos usuários, que precisam responder as perguntas ou possuir uma rede social. Por essa razão, destacamos neste capítulo que SsR não-personalizados são os mais eficazes e utilizados na prática.

Nesse intuito, na última seção revisamos as estratégias não-personalizadas tradicionalmente utilizadas pelas principais aplicações *Web*. Entre elas destacamos as estratégias de *Popularity*, *Best-Rated* e *Recent Items*. *Popularity* assume que apre-

sentar os itens mais consumidos (i.e., os mais populares) é relevante para conquistar mais usuários novos. Por outro lado, *Best-Rated* assume que os itens preferidos pela maioria dos usuários (i.e., os mais bem avaliados) são potencialmente relevantes para atenuar o *Pure Cold-Start*. Por fim, *Recent Items* assume que os usuários se interessam pelos itens que estão na moda e recomenda os últimos itens a serem consumidos no sistema. No próximo capítulo estudamos as premissas assumidas por cada uma dessas estratégias, propondo estratégias complementares.

Capítulo 4

SsR não-Personalizados Baseados na Cobertura de Usuários

Este capítulo verifica a primeira hipótese deste trabalho: *“Uma parcela significativa da população de usuários não está interessada apenas nos itens populares em domínios reais”*. Nesse sentido, primeiramente, examinamos essa hipótese de forma a levantar duas questões pertinentes a serem respondidas. Em seguida, investigamos essas questões em cenários reais da literatura por meio de um estudo de caso. Por fim, apresentamos duas novas propostas de estratégias não-personalizadas, intituladas *Max-Coverage* e *Niche-Coverage*, que exploram essas questões de estudo.

4.1 Estudo de Caso

Como visto anteriormente, as estratégias não-personalizadas tradicionais utilizam informações globais sobre o interesse dos usuários no sistema. Entre essas, para atenuar o problema de *Pure Cold-Start*, destacam-se as estratégias que utilizam informações dos itens como popularidade, avaliações recebidas e período de consumo. Apesar dessas estratégias serem distintas, elas são utilizadas sob a mesma premissa: *itens enviados pela popularidade, recência e avaliações positivas tem o potencial de satisfazer os interesses de grande parte dos usuários novos*. Em outras palavras, essas estratégias assumem que qualquer novo usuário tende a gostar de itens relacionados à essas dimensões. Seja este interesse relacionado apenas a um período de tempo, com os itens mais recentes, ou mesmo relacionado ao período total de existência do sistema, com os itens mais populares e mais bem avaliados. De certo modo, essa premissa é eficaz pois se um item é preferido por distintos usuários, teoricamente, a probabilidade desse item agradar usuários com perfis desconhecidos é maior.

Em geral, os itens mais populares são preferidos por muitos usuários, mas não necessariamente por toda a população. Por essa razão, é necessário avaliar a primeira hipótese deste trabalho que sugere a existência de muitos usuários que não estão interessados apenas nos itens populares. Caso comprovada a veracidade dessa hipótese, a premissa assumida pelas estratégias tradicionais tornar-se-á válida apenas para satisfazer usuários interessados em itens que atraem a atenção da maioria da população. Para usuários que preferem itens específicos do sistema (e.g., itens distintos dos populares), por possuírem características peculiares, essas estratégias não serão adequadas para atraí-los. Nesse intuito, nossa primeira hipótese levanta duas questões pertinentes que nos orientam para confirmar sua veracidade:

H1 - Q1: Existem usuários que consomem mais itens diversos do que populares?

H1 - Q2: Existem outras dimensões de informação que atraem os usuários?

A primeira questão está relacionada à existência de usuários com preferências distintas aos itens populares, visando avaliar a preferência individual de cada usuário pelos itens do domínio. Por outro lado, a segunda questão levantada está relacionada à existência de outros fatores intrínsecos à preferência dos usuários. Nesse caso, o objetivo é avaliar a existência de grupos de usuários com interesses em comum, no intuito de mostrar fatores distintos ao conceito de popularidade. Sendo assim, avaliar em detalhes cada uma dessas questões torna-se tarefa essencial deste trabalho, uma vez que apresentar apenas itens populares a potenciais consumidores de itens específicos não representa uma boa estratégia para a aquisição desse tipo de usuário. Além disso, as respostas encontradas podem ser utilizadas para consolidar conceitos complementares aos utilizados pelas estratégias tradicionais. Tais conceitos podem nos permitir consolidar novas estratégias de recomendação que podem ser utilizadas para conquistar usuários novos.

Dessa forma, no intuito de validar as duas questões levantadas pela primeira hipótese, conduzimos um estudo de caso sobre domínios reais de recomendação. Devido à relevância desses domínios, optamos pelo cenário de entretenimento, que diz respeito ao consumo de filmes, séries, músicas e afins; e pelo cenário de *e-commerce*, que refere-se às principais aplicações *Web*, como vendas, trocas e outros. Para o primeiro cenário, selecionamos duas bases de filme do *MovieLens*. Para o segundo cenário, selecionamos as bases da *CiaoDVD*, relativa a venda de DVDs, e a da *Amazon*, relativa a venda de produtos relacionados a video games. Todas as coleções são detalhes na próxima seção. Vale ressaltar que em cada um dos cenários avaliamos as duas questões levantadas, extraindo informações relevantes para a análise.

4.1.1 Bases de Dados

Atualmente, serviços de entretenimento e *e-commerce* como *Netflix* e *Amazon* têm dominado o mercado mundial por oferecer qualidade e eficácia aos usuários [McDonald & Smith-Rowsey, 2016]. Constantemente, esses sistemas estão lidando com um número crescente de usuários que ingressam em busca de uma experiência pessoal positiva. Nos últimos 12 meses, por exemplo, mais de 20 milhões de novos assinantes foram registrados na *Netflix* [Portal, 2017]. Por essa razão, se torna crucial avaliar a aplicabilidade dos recomendadores não-personalizados para usuários *novos* nesses cenários. Dessa forma, selecionamos os seguintes conjunto de dados de cada contexto: (1) entretenimento: os cenários de filmes do *MovieLens 1M* e *MovieLens 10M*¹; e (2) *e-commerce*: os cenários de vendas do *CiaoDVD*² e *Amazon (Video Games)*³.

Os conjuntos de dados relativos a entretenimento contêm, respectivamente, 1M e 10M *ratings* atribuídos por usuários para filmes. Por sua vez, os conjuntos de dados de *e-commerce* possuem 73K e 157K *ratings* atribuídos por usuários para DVDs e produtos relacionados a *video games*, respectivamente. Todas essas bases de dados possuem *ratings* explícitos associados em um intervalo de 1 até 5, onde 1 é “muito ruim” e 5 significa “muito bom”. Para melhor análise do domínio da Amazon, os dados foram pré-processados deixando apenas usuários que compraram pelo menos 10 produtos. Assim, em todos os conjuntos de dados, cada usuário consumiu pelo menos 10 itens do domínio. Além disso, esses conjuntos de dados foram selecionados por possuírem duas informações fundamentais para este trabalho: (1) um *timestamp* relativo ao momento em que o usuário consumiu um item; e (2) gêneros associados aos itens do domínio. Cada item desses domínios possui pelo menos um gênero (e.g., ação, comédia, aventura, e outros) nos cenários de filmes e uma categoria (e.g., *playstation*, *xbox*, e outros) no cenário de *e-commerce*. Todas essas informações estão sumarizadas na Tabela 4.1 a seguir.

Tabela 4.1: Descrição geral dos conjuntos de dados selecionados.

Base de Dados	Usuários	Itens	Esparsidade	Gêneros
<i>MovieLens 1M</i>	6.040	3.952	95,82%	18
<i>MovieLens 10M</i>	69.878	10.283	98,60%	20
<i>CiaoDVD</i>	17.615	16.621	99,97%	17
<i>Amazon</i>	8.057	26.729	99,92%	471

¹Disponível em: <<http://www.grouplens.org/node/12>>

²Disponível em: <<https://www.librec.net/datasets.html>>

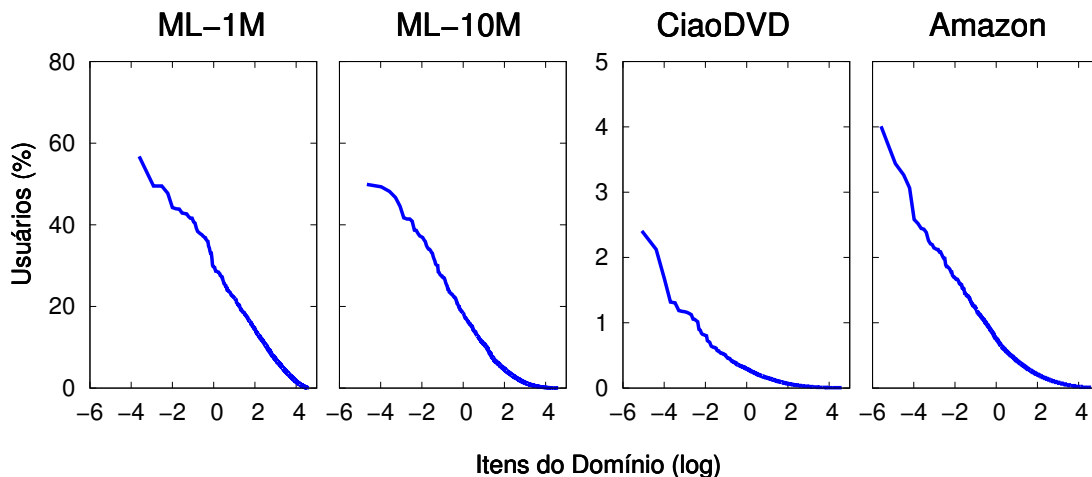
³Disponível em: <<http://jmcauley.ucsd.edu/data/amazon/links.html>>

4.1.2 Consumo de Itens Populares e Não-Populares

Freqüentemente, o consumo dos usuários é enviesado aos itens populares do domínio pois esses produtos são os mais *consumidos* (i.e., comprados, assistidos, etc.) e, assim, possuem a maior quantidade de informação disponível [Ho et al., 2014]. Esse fenômeno é descrito por grande parte dos economistas como o Princípio de Pareto (ou mesmo *power-law*) [Yin et al., 2012], onde uma pequena parcela de produtos (e.g., 20%) gera uma larga porção de consumo (e.g., 80%). Pode-se dizer que existe uma pequena parcela de itens que são muito consumidos, formando a *cabeça* da distribuição, e uma grande parcela de itens que são pouco consumidos, formando a *cauda* da distribuição [Anderson, 2006]. Por essa razão, neste trabalho, denominamos como itens populares aqueles que pertencem a *cabeça* da distribuição e como itens não-populares aqueles que pertencem a *cauda*. Essa divisão de *cabeça* e *cauda* é feita pelos estatísticos ao encontrar o *joelho* da distribuição (i.e., *elbow point*), por meio do cálculo da derivada segunda da distribuição [Fujimura & Zhu, 2008].

Nos domínios de recomendação selecionados, conforme mostra a Figura 4.1, a popularidade dos itens (i.e., número de usuários distintos que consumiram esse item) também segue uma *power-law*. Neste trabalho é definida uma porcentagem de 8%, 5%, 4% e 3% de itens como populares para as bases de ML-1M, ML-10M, CiaoDVD e Amazon, respectivamente. Como pode ser observado, o item mais popular foi consumido por aproximadamente 50% dos usuários, nos cenários de entretenimento, e por menos de 5% nos cenários de *e-commerce*. Por essa razão, se torna ainda mais pertinente a H1-Q1: *Existem usuários que consomem mais itens diversos do que populares?*

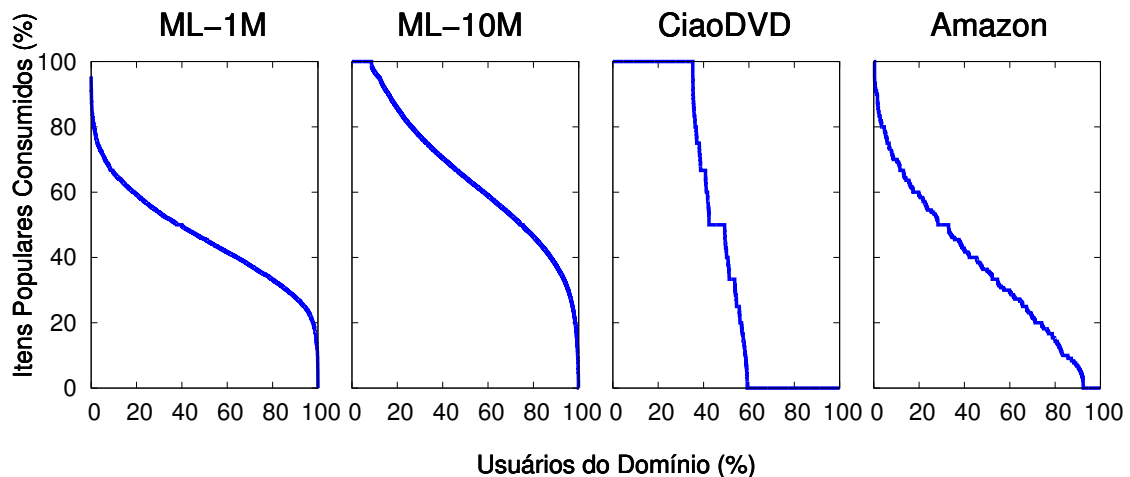
Figura 4.1: As distribuições de popularidade dos itens seguem a *power-law*. Nos cenários de entretenimento, os itens mais populares são consumidos por mais da metade dos usuários, enquanto nos cenários de *e-commerce* poucos usuários são assíduos.



Para responder a primeira questão levantada, avalia-se o *ranking* gerado pela porcentagem de itens populares existentes no histórico de consumo de cada usuário do domínio. Basicamente, como conhecemos os itens populares de cada cenário, basta contar o número de vezes que cada um desses aparece no histórico do usuário e normalizar pelo número de itens consumidos por esse. Valores próximos a 100% indicam que o consumo do usuário é, preferencialmente, formado por itens populares. Por outro lado, valores próximos a 0% mostram que o consumo do usuário é, preferencialmente, formado por itens não-populares. Os *rankings* gerados para cada conjunto de dados são ordenados decrescentemente e mostrados na Figura 4.2.

Com essa análise, pode-se notar que existem três tipos de comportamentos dos usuários relacionados à popularidade dos itens: (1) usuários que preferem itens populares, optando quase 100% das vezes por esses; (2) usuários que preferem itens não-populares, optando quase 0% das vezes por itens populares; e (3) usuários que escolhem itens populares e não-populares, optando entre 30% a 70% das vezes por itens populares. Com exceção do cenário da CiaoDVD, a maioria dos usuários mesclam por itens populares e não-populares durante seu consumo. Nos cenários de ML-1M, ML-10M e Amazon, poucos usuários possuem um comportamento totalmente linear aos tópicos (1) e (2), i.e., com valores no *ranking* equivalentes a 100% ou 0%. Apenas no cenário CiaoDVD que existe uma grande parcela de usuários (quase 40%) que consomem apenas populares, e um outra grande parcela de usuários (pouco mais de 40%) que consomem apenas não-populares. Esse caso distinto está relacionado à pequena quantidade de itens consumidos em média por cada usuário no cenário do CiaoDVD.

Figura 4.2: Porcentagem de itens populares consumidos por cada usuário, mostrando que existem: (1) usuários que optam por itens populares; (2) usuários que optam por itens não-populares; e (3) usuários que mesclam itens populares e não-populares.



Dessa forma, fica evidente que a premissa assumida pelas estratégias tradicionais, de que apenas os itens populares são capazes de satisfazer os usuários, nem sempre é verdadeira. Como visto, existem usuários que optam por itens com características específicas, que não são os populares. Esse resultado responde a primeira questão levantada por este trabalho, mostrando que a maioria dos usuários preferem mesclar entre itens populares e não-populares. Por essa razão, se torna evidente a necessidade de uma estratégia capaz de lidar com a maioria das preferências dos usuários. Afinal, satisfazer a cada uma dessas preferências é essencial para proporcionar uma melhor experiência do usuário com o sistema, principalmente no primeiro contato dos usuários novos.

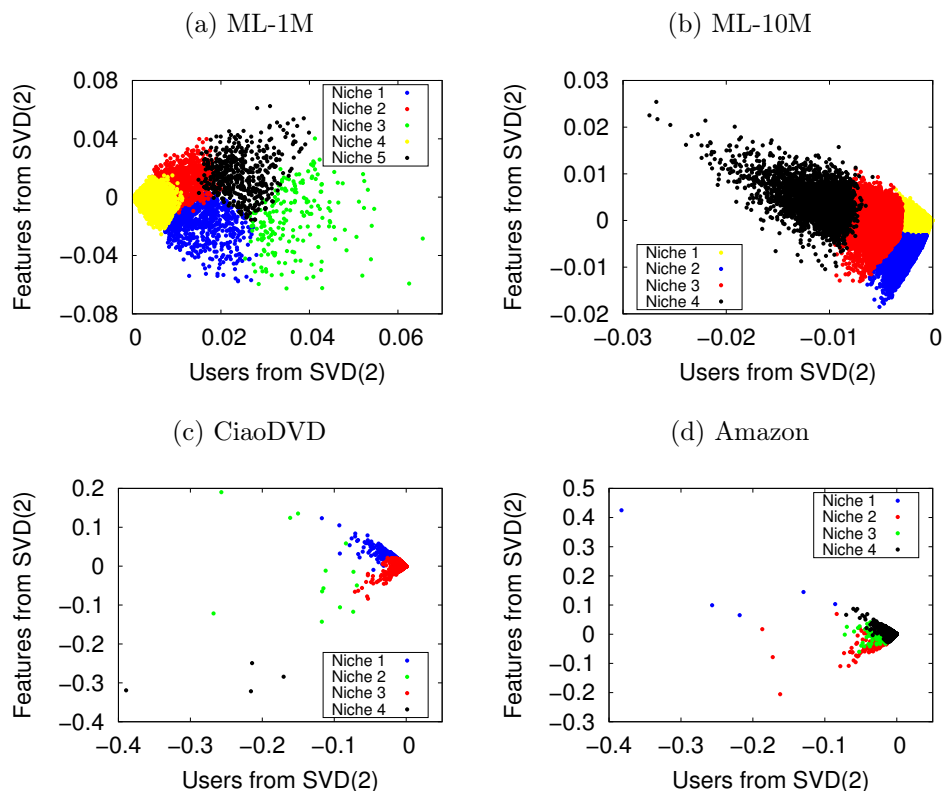
4.1.3 Nichos de Usuários

As principais estratégias de recomendação assumem que, mesmo que existam distintos perfis de usuários, esses compartilham interesses em comum no seu histórico de consumo [Bobadilla et al., 2013]. De fato, é possível dizer que grande parte dos itens compartilhados pelos usuários correspondem aos itens populares, pois esses itens são consumidos por grande parte da população. Entretanto, sabemos pela análise da seção anterior, que o consumo dos usuários não se retém apenas aos itens populares. Por essa razão, torna-se necessário avaliar a segunda questão levantada: *Existem outras dimensões de informação que atraem os usuários?*

Para responder a essa questão, este trabalho visa associar os usuários em grupos de interesses em comum, denominados pela literatura como nichos de usuários [Mobasher et al., 2005; Duarte Torres et al., 2012]. Em alguns casos esses nichos têm significado semântico, modelando perfis como crianças e adultos, mas, na grande maioria representam apenas grupos de usuários (i.e., *clusters*). Dessa forma, os nichos de usuários podem ser modelados por meio de qualquer estratégia de clusterização que utilize os *ratings* explícitos dos usuários como fonte de informação. Ghazanfar & Prügel-Bennett [2014] analisa diversos algoritmos de clusterização ao propor uma estratégia de recomendação, e observa que o algoritmo clássico do *k-means* é capaz de detectar esses nichos de usuários de maneira mais eficaz. Por essa razão, e também devido à simplicidade e eficiência da técnica, este trabalho utiliza o *k-means* para identificar nichos de usuários. A Figura 4.3 demonstra os nichos encontrados em cada domínio, por uma projeção 2D feita por meio do SVD⁴ aplicado sobre o histórico de consumo dos usuários. Nesses gráficos, cada ponto representa um usuário do domínio e as cores marcam os nichos (*clusters*) encontrados por meio do *k-means*.

⁴*Singular Value Decomposition* (SVD) é um método de decomposição de fatores latentes capaz de reduzir a dimensão de uma matriz representando-a em k autovetores e autovalores.

Figura 4.3: Os nichos de usuários encontrados em cada cenário ressaltam a existência de grupos de interesse dos usuários que nem sempre estão relacionados à popularidade dos itens do domínio.



Por meio dessa análise é possível identificar a existência de quatro a cinco nichos de usuários em cada domínio, demonstrando a existência de grupos de interesses distintos. Esse resultado responde a segunda questão levantada por este trabalho, concluindo que existem interesses intrínsecos ao comportamento dos usuários que vão além da popularidade dos itens. Para os cenários do ML-1M e ML-10M, é possível observar *clusters* bem distribuídos devido à grande interação dos usuários com os itens em um cenário de entretenimento. Por outro lado, nos cenários de *e-commerce*, como Amazon e CiaoDVD, o domínio é mais esparsa e os usuários são pouco assíduos, formando um *cluster* homogêneo com muitos usuários. Nesses cenários de *e-commerce*, os *outliers* são justamente os usuários mais assíduos do domínio.

Com esses resultados, se torna também evidente que uma estratégia capaz de satisfazer particularmente a cada um desses grupos de interesses pode ser efetiva para lidar com o perfil desconhecido de um usuário novo. Satisfazer todos esses nichos de usuários significa cobrir todos os perfis de usuários conhecidos e, assim, aumentar a probabilidade de atrair qualquer usuário novo, independente da sua preferência pessoal.

4.2 Estratégias não-Personalizadas Propostas

Com base neste estudo de caso realizado, pode-se confirmar a primeira hipótese deste trabalho mostrando que apenas o pressuposto assumido pelas estratégias tradicionais nem sempre é factível. Diferentemente da literatura, este trabalho propõe duas novas estratégias de recomendações não-personalizadas que consideram as respostas encontradas pelas questões levantadas. Propomos a utilização da estratégia de *Maximum k-Coverage* para o problema de *Pure Cold-Start*, no intuito de que essa estratégia seja capaz de extrair itens populares e não-populares relevantes aos usuários novos. Essa estratégia leva em conta a necessidade de lidar com nossa primeira premissa. Além disso, propomos uma nova estratégia intitulada *Niche-Coverage* que explora os itens mais representativos para cada grupo de usuários distintos. Por sua vez, essa outra estratégia leva em conta a segunda premissa destacada por este trabalho. Ambas estratégias consideram que maximizar a cobertura dos usuários encontrando os itens que satisfazem ao maior número de usuários ou grupos de usuários distintos é relevante para esse cenário de *Pure Cold-Start*. Essas estratégias são explicadas em detalhes nesta seção, que mescla teoria e prática para elaborar essas novas abordagens.

4.2.1 Estratégia de Max-Coverage

A estratégia de *Max-Coverage* consiste da aplicação da tradicional estratégia de *Maximum k-Coverage* [Hochbaum & Pathria, 1998] para gerar recomendações para usuários novos. Dado $U = \{u_1, \dots, u_m\}$ como os usuários existentes em um sistema e $I = \{i_1, \dots, i_n\}$ como os itens disponíveis em um catálogo de produtos, pode-se modelar um conjunto $F = \{S_1, \dots, S_n\}$ onde cada subconjunto S_i é relativo a um item existente. Nesta modelagem, o subconjunto S_i é formado pelos usuários que consumiram o item i , de forma que S_i se torna um subconjunto de usuários. Dessa forma, nosso objetivo é encontrar k subconjuntos S_i que contenham o maior número de usuários distintos possível, formando o conjunto $F^* = \{S_{i_1}, \dots, S_{i_k}\}$. Essa modelagem se assemelha ao problema de Otimização Combinatorial conhecido como *Maximum k-Coverage*, que pode ser definido formalmente como se segue:

MAXIMUM k -COVERAGE

Instância: Um universo de elementos $U = \{u_1, \dots, u_m\}$, um inteiro k e uma família de subconjuntos $F = \{S_1, \dots, S_n\}$, onde cada S_i é um subconjunto de U .

Objetivo: Encontrar uma subfamília $F^* \subseteq F$ dado que $|F^*| \leq k$ e o número de elementos cobertos $|\bigcup_{S \in F^*} S|$ seja maximizado, i.e. utilizando k subconjuntos, cobrir o maior número de elementos possível.

O problema de *Maximum k-Coverage* deriva do famoso problema de Cobertura de Vértices [Hochbaum, 1996; Alon et al., 2003; Buchbinder et al., 2009]. Infelizmente, esses problemas pertencem a classe de problemas NP-Difícil⁵, e não se conhece uma solução para tal em tempo polinomial. Entretanto, uma heurística gulosa simples, que a cada iteração maximiza o número de usuários ainda não cobertos, possui uma razão de aproximação de 63% da solução ótima, como mostrado em Chvatal [1979]. Na prática, essa heurística é implementada para a tarefa de recomendação como mostra o Algoritmo Guloso 1. Nesse algoritmo, são realizadas k iterações, em que todos os conjuntos S_i ainda não selecionados (i.e., $S \in F \setminus F^*$) são avaliados, procurando pelo conjunto que maximize a interseção $|S \cap R|$ de usuários, como mostra a linha 4. Repare que, a cada iteração, o conjunto R de usuários resultantes é atualizado, excluindo os usuários cobertos pelo conjunto S selecionado, conforme mostra a linha 6. O algoritmo termina quando os k itens são selecionados ou quando não existem mais usuários a serem cobertos. A complexidade deste algoritmo é de $O(kmn)$, onde k é o número de itens a serem recomendados, m é o número de usuários e n é o número de itens.

Algorithm 1 GREEDY-MAX-COVERAGE(U, k, F)

```

1:  $R \leftarrow U$ 
2:  $F^* \leftarrow \emptyset$ 
3: for  $itr$  from 1 to  $k$  do
4:    $S \leftarrow \max_{S \in (F \setminus F^*)} |S \cap R|$ 
5:    $F^* \leftarrow F^* \cup \{S\}$ 
6:    $R \leftarrow R \setminus S$ 
7:   if  $|R| = 0$  then
8:     break
9:   end if
10: end for
11: return  $F^*$ 

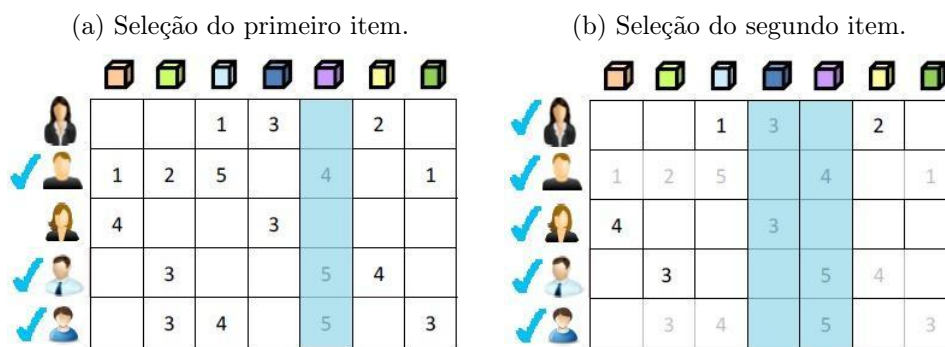
```

Uma análise preliminar da estratégia proposta pode levar a uma conclusão equivocada de que os itens recomendados por *Max-Coverage* são os itens mais populares do domínio, uma vez que o objetivo é cobrir o maior número de usuários. Entretanto, observando o comportamento do Algoritmo 1 sobre o exemplo proposto no Capítulo 2, na Figura 2.1, nota-se que essa estratégia diverge da tradicional que recomenda os itens populares. Devido ao algoritmo guloso utilizado, o primeiro item selecionado é sempre o mais popular (Figura 4.4a). Contudo, a partir desse item selecionado, desconsideram-se todas as avaliações dos usuários marcados com o \checkmark , pois estes já foram cobertos e removidos do conjunto de usuários resultantes, conforme mostra a linha 6 do algoritmo.

⁵Subconjunto de problemas que provavelmente não podem ser resolvidos em tempo polinomial.

Sendo assim, o novo item selecionado é aquele que cobre o maior número de usuários resultantes nessa iteração, e não o segundo mais popular (Figura 4.4b). A cada iteração, a estratégia de *Max-Coverage* tende a selecionar cada vez mais itens relacionados a uma preferência específica dos usuários. Por essa razão, pode-se dizer que a estratégia de *Max-Coverage* está alinhada a primeira premissa levantada anteriormente, uma vez que é capaz de recomendar itens populares e não-populares em conjunto.

Figura 4.4: Exemplo prático da utilização da estratégia *Max-Coverage*. Pode-se notar que essa estratégia mescla suas recomendações com itens não-populares pois tenta cobrir a preferência do maior número de usuários possível.



4.2.2 Estratégia de Niche-Coverage

A estratégia de *Niche-Coverage*, proposta por este trabalho, consiste em explorar os itens característicos de distintos nichos de usuários (i.e., grupos de usuários com interesses em comum). Definimos como item característico, o item mais provável de satisfazer a preferência individual da maioria dos usuários de um dado nicho. Dessa forma, *Niche-Coverage* consiste em uma estratégia que a cada etapa avalia um nicho distinto, procurando pelo item que o maior número de usuários consumiram e gostaram. Neste trabalho, definimos que um usuário u gosta de um item i se u consumiu i e forneceu um *rating* maior ou igual a sua média pessoal. Basicamente, a média pessoal de cada usuário é calculada sobre o histórico de avaliações do usuário no sistema, levando em conta apenas o número de *ratings* atribuídos por este. Por meio dessa abordagem conseguimos priorizar itens relevantes no processo de seleção das recomendações do *Niche-Coverage*. Essa estratégia deriva da segunda premissa feita anteriormente, pois propõe que encontrar os itens que satisfazem a esses nichos de usuários nos permite encontrar os itens que satisfazem a maioria das preferências existentes. Desse modo, esses itens são potencialmente relevantes para serem apresentados a usuários novos, nos quais o sistema desconhece o perfil desses.

Na prática, *Niche-Coverage* é implementado como um algoritmo guloso, conforme mostra o Algoritmo 2. Dado o conjunto de usuários $U = \{u_1, \dots, u_m\}$, o primeiro passo do algoritmo consiste em classificar esses usuários em x nichos distintos, por meio de qualquer algoritmo de clusterização que leve em conta as relações usuários-itens. Nesse caso, a família de subconjuntos F é modelada como $F = \{S^*_1, \dots, S^*_n\}$, onde o subconjunto S^*_i contém apenas os usuários que consumiram e gostaram do item i . O objetivo agora é encontrar uma subfamília $F^* = \{S_{i1}, \dots, S_{ik}\}$, contendo os k itens que satisfaçam a todos os nichos. Nesse algoritmo, cada nicho de usuários N_j é avaliado, procurando pelo item ainda não selecionado (i.e., $S^* \in F \setminus F^*$) que maximize os usuários cobertos (i.e., $|S^* \cap N_j|$), conforme mostra a linha 5. Repare que como S^*_i é modelado apenas com os usuários que consumiram e gostaram do item i , os itens selecionados são aqueles que satisfazem a grande parte dos usuários do nicho. A complexidade desse algoritmo também é $O(kmn)$.

Algorithm 2 GREEDY-NICHE-COVERAGE(U, k, F)

```

1:  $N \leftarrow \text{clustering}(U)$ 
2:  $itr \leftarrow 0$ 
3: while ( $itr < k$ ) do
4:   for  $j$  in  $N$  do
5:      $S^* \leftarrow \max_{S^* \in (F \setminus F^*)} |S^* \cap N_j|$ 
6:      $F^* \leftarrow F^* \cup \{S^*\}$ 
7:      $itr \leftarrow itr + 1$ 
8:   end for
9: end while
10: return  $F^*$ 

```

4.3 Síntese do Capítulo

Este capítulo se inicia levantando duas questões relacionadas à primeira hipótese deste trabalho. De maneira geral essas questões nos permitem analisar em detalhes as preferências dos usuários em cada domínio selecionado. Essa análise é feita por meio de um estudo de caso sobre dois dos principais domínios de recomendação: entretenimento e *e-commerce*. Os resultados da primeira questão confirmam que existem usuários que estão interessados por outros itens além dos populares. Por sua vez, os resultados da segunda questão mostram diversos nichos de usuários no domínio, apontando para a existência de outras preferências intrínsecas no perfil de cada usuário. Ambas as respostas nos permitem confirmar a primeira hipótese, mostrando que apenas as estratégias tradicionais não são suficientes para conquistar muitos usuários novos.

Dessa forma, na última seção, apresentamos duas novas propostas de SSR não-personalizados que se baseiam na cobertura de usuários. Basicamente, sabendo que existem usuários que preferem itens populares e não-populares, propomos *Max-Coverage* para mesclar esses itens e cobrir mais usuários distintos. De maneira complementar, sabendo que existem outros grupos de interesses, propomos *Niche-Coverage* para cobrir cada um dos nichos de usuários. Ambas as estratégias serão avaliadas em detalhes ao longo do próximo capítulo.

Capítulo 5

Avaliações Empíricas

Este capítulo tem por objetivo verificar a segunda hipótese deste trabalho: “SsR não-personalizados que exploram outros itens além dos populares são capazes de atenuar o problema de *Pure Cold-Start*”. Nesse intuito, primeiramente, destacamos as principais questões levantadas por essa hipótese que constitui nossa metodologia de avaliação. Nessa, propomos avaliar cada questão sobre os conjuntos de dados levantados no Capítulo 4. Para tal, primeiramente, selecionam-se um grupo de usuários do domínio para simularem os usuários novos, os quais todas as informações são removidas. Em seguida, são apresentados os requisitos necessários para definir a qualidade de um recomendador em satisfazer esses usuários novos. Por fim, são apresentados os resultados obtidos por cada métrica sobre os SsR tradicionais e os propostos.

5.1 Metodologia de Avaliação

A segunda hipótese deste trabalho sugere que estratégias complementares às tradicionais também podem ser utilizadas para atenuar o problema de *Pure Cold-Start*. Essa hipótese provém das observações realizadas no Capítulo 4, que propõe novas estratégias para serem utilizadas para satisfazer o interesse dos usuários por outros itens além dos populares. Dessa forma, surgem questões relevantes que estão relacionadas à aplicação dos SsR não-personalizados propostos, conforme enumeradas a seguir:

H2 - Q1: Qual a similaridade dos SsR propostos com os tradicionalmente utilizados?

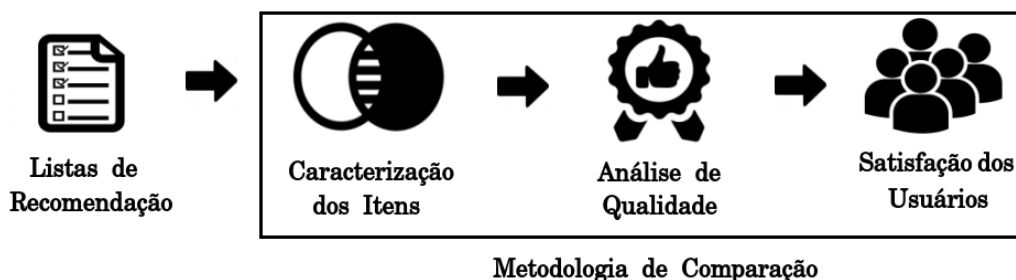
H2 - Q2: Qual a qualidade dos itens recomendados pelos SsR propostos?

H2 - Q3: Os itens apresentados pelos SsR propostos conquistam a atenção dos mesmos usuários ou subconjuntos complementares?

Nesse intuito, definimos nossa metodologia de avaliação sobre as três questões levantadas pela segunda hipótese deste trabalho. Especificamente, avaliamos cada uma das estratégias propostas sobre três dimensões distintas relacionadas a cada uma das questões apresentadas, tal qual mostra a Figura 5.1. Primeiramente, realiza-se uma distinção dos itens recomendados para responder a primeira questão levantada. Nessa etapa, avalia-se a semelhança dos itens a fim de identificar as estratégias que possuem um comportamento similar, bem como as que são complementares. Por sua vez, na segunda etapa, avalia-se as recomendações geradas sobre os principais requisitos de qualidade para responder a segunda questão. Por fim, avalia-se a satisfação dos usuários novos ao consumirem os itens fornecidos pelo sistema. Essa última questão visa avaliar se as estratégias recomendam itens que interessam aos mesmos usuários ou subconjuntos complementares. Cada etapa dessa metodologia é detalhada como se segue.

1. **Caracterização dos Itens Recomendados:** esta etapa tem por objetivo contrastar os itens recomendados por cada estratégia. Para tal, propomos avaliar a interseção dos itens recomendados por cada estratégia por meio do diagrama de *Venn*. Além disso, também avaliamos a popularidade dos itens recomendados por cada estratégia, deixando clara a diferença entre as principais abordagens existentes.
2. **Requisitos de Qualidade:** esta etapa consiste em avaliar a qualidade de cada recomendador sobre os critérios de utilidade e diversidade, mensurados pelas métricas selecionadas na Seção 5.3.
3. **Satisfação dos Usuários:** nesta etapa visamos mensurar quantos usuários são atraídos e satisfeitos por pelo menos um item recomendado por cada estratégia. Para este trabalho, nós consideramos que um item satisfaz um usuário se o *rating* atribuído por esse usuário ao item é maior ou igual à sua média pessoal de *ratings*. Após essa análise, avaliamos se os usuários alcançados por cada SsR são os mesmos ou subconjuntos complementares por meio do diagrama de *Venn*.

Figura 5.1: Metodologia de avaliação proposta.



Para essas análises, simulamos um cenário real de avaliação por meio dos conjuntos de dados selecionados no Capítulo 4. Especificamente, apresentamos listas de recomendações contendo 5, 10, 20 e 50 itens para um conjunto de usuários dos quais removemos todas as informações existentes. Essas recomendações são geradas pelas estratégias não-personalizadas propostas e pelos *baselines* da literatura. Para as etapas (1) e (3) simulamos um cenário onde apenas 10 itens são recomendados, e investigamos a efetividade de cada estratégia. Por sua vez, para a etapa (2) avaliamos as recomendações geradas sobre os principais requisitos de qualidade necessários.

Dessa forma, essa avaliação empírica está organizada conforme se segue. Primeiramente, selecionamos os usuários novos que serão utilizados como base para nossas avaliações. Para cada conjunto de dados selecionado, extraímos os últimos 20% dos usuários a acessarem o sistema e removemos todas as suas informações associadas. Em seguida, apresentamos os principais requisitos de qualidade necessários a um recomendador, que mesclam entre utilidade e diversidade. Por fim, destacamos os principais *baselines* selecionados da literatura, bem como a forma como foram implementados neste trabalho. Essas estratégias são utilizadas como base de nossas avaliações, visto que são amplamente utilizadas na prática.

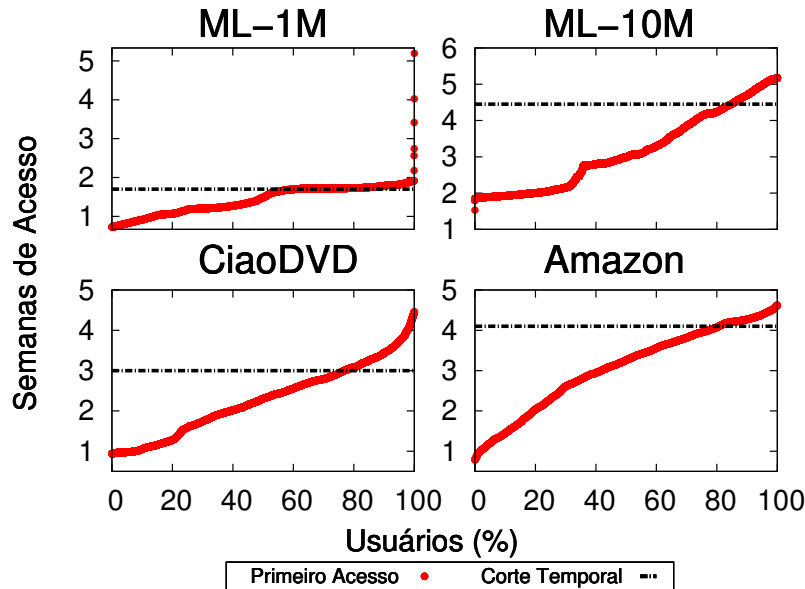
5.2 Seleção de Usuários Novos

A seleção de usuários novos é parte fundamental deste trabalho, dado que o objetivo consiste em avaliar a aplicabilidade das estratégias não-personalizadas para atenuar o problema de *Pure Cold-Start*. Neste trabalho, o conjunto de usuários novos é selecionado considerando uma análise temporal do consumo dos usuários em cada conjunto de dados. Basicamente, é realizado um corte temporal em uma semana específica do consumo de usuários, a fim de selecionar em torno de 20% dos últimos usuários a ingressarem em cada domínio. Para esses usuários selecionados são removidas todas as informações (i.e., dados pessoais e *ratings*) para simular o problema de *Pure Cold-Start*. A justificativa dessa abordagem é que desconsiderar os últimos usuários a ingressarem no sistema é a melhor aproximação *offline* que pode ser feita de um cenário real de usuários novos.

A Figura 5.2 esboça o corte temporal realizado nos conjuntos de dados levantados no Capítulo 4. Com essa estratégia são selecionados 1.277, 10.633, 3.523 e 1.612 usuários de ML-1M, ML-10M, CiaoDVD e Amazon, respectivamente, para serem usuários novos. Esses 20% de usuários selecionados são adicionadas ao conjunto de teste, enquanto o restante de 80% é utilizado como treino dos recomendadores. Repare

que nenhum dos recomendadores possui informações sobre os usuários presentes no conjunto teste, de forma a esboçar um cenário real como desejado.

Figura 5.2: Corte temporal realizado em semanas específicas a fim de selecionar os usuários novos como os últimos 20% a ingressarem no sistema.



5.3 Requisitos de Qualidade

A fim de atrair usuários novos com distintas preferências, não é suficiente que as estratégias foquem apenas na utilidade dos itens para os usuários [McNee et al., 2006; Lacerda & Ziviani, 2013; Ho et al., 2014]. Aspectos como diversidade, cobertura e surpresa são importantes para definir uma interface com os melhores itens do catálogo [Zhang, 2009; Shani & Gunawardana, 2011]. Por essa razão, as métricas selecionadas para essa avaliação *offline* visam avaliar utilidade e diversidade. A utilidade de uma lista de itens é avaliada pelas métricas de *accuracy*, *precision* e *recall* [Bobadilla et al., 2013]. Por outro lado, a diversidade dos itens recomendados é mensurada por meio das métricas de *Diversity* e *Genre Coverage* [Vargas & Castells, 2011; Puthiya Parambath et al., 2016]. Todas essas métricas são implementadas como se segue.

Accuracy - representa a quantidade de itens recomendados corretamente, que estão presentes na lista de recomendação R . Essa métrica é avaliada com base no conjunto de itens consumidos pelo usuário alvo, que estão presentes no conjunto de teste (I_u).

$$Accuracy(u) = |R_u \cap I_u| \quad (5.1)$$

Precision - representa a probabilidade de que um item selecionado seja relevante, sendo definida como a fração de itens relevantes selecionados (R^+) pelo número de itens presentes na lista de recomendação (R) [Herlocker et al., 2004].

$$Precision = \frac{|R^+|}{|R|} \quad (5.2)$$

Recall - representa a probabilidade de um item relevante ser selecionado, sendo definida como a fração de itens relevantes selecionados (R^+) pelo número de itens relevantes existentes no domínio (I_u^+) [Herlocker et al., 2004].

$$Recall = \frac{|R^+|}{|I_u^+|} \quad (5.3)$$

F-Measure - representa a média harmônica entre as métricas de *Precision* e *Recall*, permitindo avaliar de maneira global essas duas métricas.

$$F - Measure = 2 \cdot \frac{Prec \cdot Rec}{Prec + Rec} \quad (5.4)$$

Diversity - representa a diversidade dos itens baseada na distância média esperada (ILD) de um item para uma lista de itens. Basicamente, essa métrica é calculada como o complemento da similaridade de Pearson dos itens recomendados, como mostra a Equação 5.5, extraída do *framework* proposto por Vargas & Castells [2011].

$$div(R) = ILD = \frac{2}{|R|(|R| - 1)} \sum_{i_k \in R, l < k} d(i_k, i_l) \quad (5.5)$$

Genre Coverage - corresponde à proporção média de gêneros (i.e., características dos itens) recomendados para o usuário, a fim de mensurar a diversidade dos itens selecionados [Puthiya Parambath et al., 2016]. Essa métrica é definida para o conjunto de usuários como a razão dos gêneros recomendados sobre a quantidade de gêneros distintos existente, onde I_u é o conjunto de itens avaliados previamente por u ; R_u é a lista de recomendação para u ; e $genres(i)$ são os gêneros associados ao item i .

$$CG = \frac{1}{|U|} \sum_{u \in U} \frac{|\bigcup_{i \in R_u} genres(i)|}{|\bigcup_{i \in I_u} genres(i)|} \quad (5.6)$$

Genre Frequency - consiste em avaliar a frequência com que os gêneros aparecem em cada lista de recomendação apresentada. Basicamente, conta-se a quantidade de vezes que os gêneros aparecem e normaliza-se pela quantidade de itens recomendados.

5.4 Estratégias Baselines

Considerando a efetividade das estratégias não-personalizadas para atenuar o problema de *Pure Cold-Start*, este trabalho seleciona as estratégias estado-da-arte, destacadas no Capítulo 3, como *baselines*. Conforme descrito anteriormente, essas estratégias são:

- **Popularity (Pop):** visa recomendar os k itens mais populares do domínio, assumindo que os itens que interessam grande parte da população cobrem distintas preferências. Basicamente, a popularidade do item i é estimada pelo número de usuários distintos que consumiram o item i .
- **Best-Rated (BestR):** recomenda os k itens mais bem avaliados do domínio. Basicamente, calcula-se a média dos *ratings* recebidos por cada item, considerando todos os usuários do domínio, e recomendam-se os k com maiores médias.
- **Recent Items (RecItems):** consiste em recomendar os k últimos itens consumidos pelos usuários no domínio. A premissa dessa abordagem é que os usuários tendem a se interessar pelos itens que estão na moda e foram recentemente consumidos.
- **Random Popularity (RandP):** visa recomendar k itens aleatórios dentro do grupo de itens populares (i.e., que estão na *cabeça* da distribuição de popularidade). Para essa técnica, utilizou-se uma porcentagem de 8%, 5%, 4% e 3% de itens para as bases de ML-1M, ML-10M, Amazon e CiaoDVD, respectivamente. Essa estratégia não é utilizada na prática, sendo considerada apenas para determinar se as estratégias propostas por este trabalho são efetivas em selecionar itens potencialmente relevantes, ou se uma simples seleção aleatória seria suficiente.

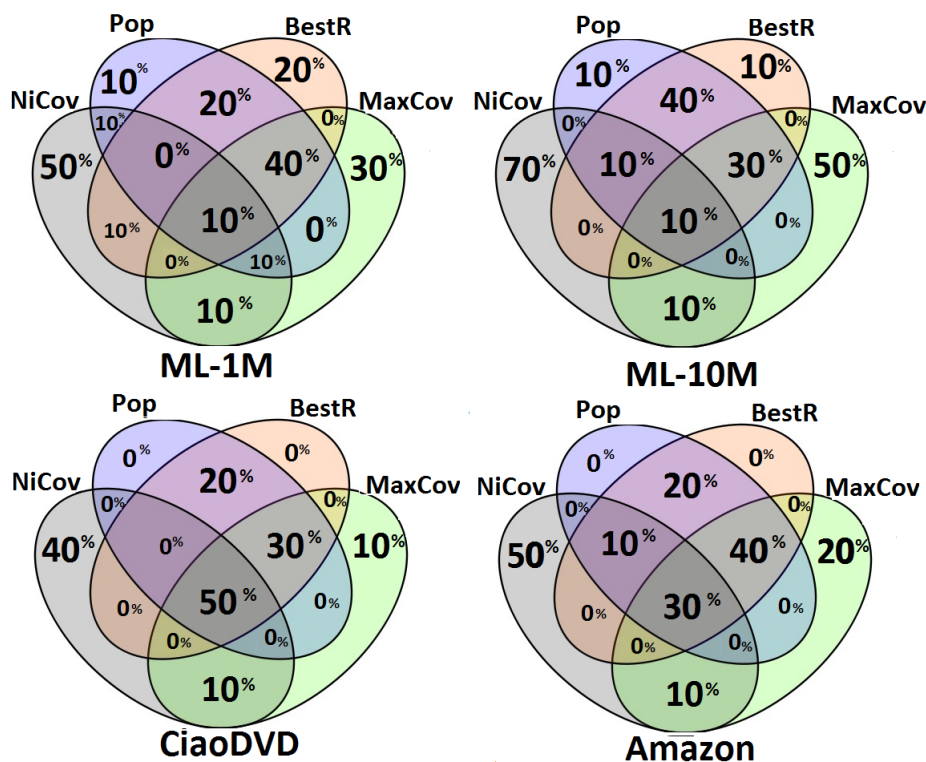
5.5 Resultados

Nesta etapa, analisamos os resultados da aplicação da metodologia de avaliação proposta sobre as novas estratégias propostas e os *baselines* selecionados. Para facilitar o entendimento, nossas análises são divididas em seções de acordo com cada etapa da metodologia. Primeiro, comparamos os itens recomendados por cada estratégia. Em seguida, avaliamos a qualidade das recomendações geradas sobre as métricas selecionadas. Por fim, avaliamos os usuários atraídos por cada recomendador quando aplicado sobre os conjuntos de dados selecionados.

5.5.1 Caracterização dos Itens Recomendados

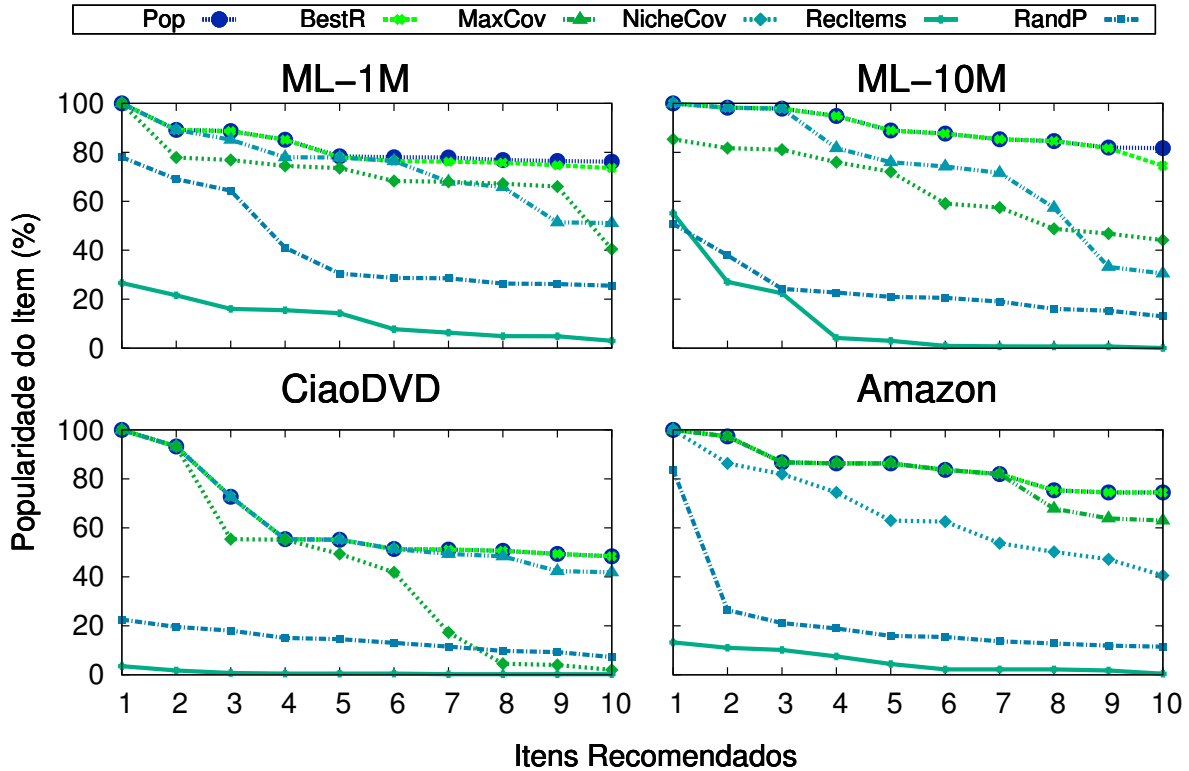
Nesta etapa avalia-se a interseção e a popularidade dos primeiros 10 itens recomendados por cada uma das estratégias não-personalizadas. A semelhança entre os itens recomendados é mostrada no diagrama de Venn da Figura 5.3. A interseção entre as estratégias *Pop* e *BestR*, que nos cenários somam 70% (ML-1M), 80% (ML-10M) e 100% (CiaoDVD e Amazon), confirma que ambas utilizam premissas similares na prática. Por outro lado, apesar de existirem itens em comum entre todas as estratégias, existem itens que são recomendados apenas por *MaxCov* e *NicheCov*. Nos cenários de ML-1M, ML-10M e *Amazon*, 70% e 50% dos itens (i.e., 7 e 5 itens) são recomendados apenas por *NicheCov*. Nesses mesmos cenários, 30%, 50% e 20% dos itens (i.e., 3, 5 e 2 itens) aparecem apenas nas listas de recomendação geradas por *MaxCov*. Além disso, em todos os cenários, existe um item que foi recomendado apenas pelas estratégias de *MaxCov* e *NicheCov*, aumentando ainda mais a relevância de nossas propostas. Essa análise nos permite consolidar que, de fato, as premissas dessas estratégias propostas divergem das tradicionais na prática.

Figura 5.3: Similaridade dos primeiros 10 itens recomendados por cada estratégia. Em todos os cenários, mais de 40% dos itens são recomendados apenas pela estratégia de *Niche-Coverage*. Em ML-10M e Amazon, 50% e 20% dos itens são recomendados apenas por *Max-Coverage*.



De maneira complementar, avalia-se a popularidade dos 10 primeiros itens recomendados por cada estratégia. Basicamente, para cada um dos recomendadores foi gerado um *ranking* ordenado decrescentemente, destacando a popularidade de cada um dos itens frente ao item mais popular do domínio. Dessa forma, a área sob a curva (AUC) da estratégia de *Pop* se torna o limite superior dessa análise, de forma que, quanto menor AUC, mais itens não-populares são recomendados. Claramente, com a Figura 5.4, pode-se observar que todas estratégias, com exceção de *BestR*, são capazes de recomendar itens menos populares. Como suposto, a cada iteração, a estratégia de *MaxCov* recomenda cada vez mais itens que são menos populares. *NicheCov* também diversifica seus itens das estratégias tradicionais apresentando itens menos populares. Como esperado, a performance da estratégia de *BestR* é similar a de *Pop*, dado que os itens recomendados são quase sempre idênticos. *RecItems* e *RandP*, que não possuem itens semelhantes às estratégias tradicionais, também diversificam as recomendações utilizando itens não-populares.

Figura 5.4: Nível de popularidade dos 10 primeiros itens recomendados por cada estratégia, apontando que as estratégias de *Max-Coverage* e *Niche-Coverage* são capazes de recomendarem itens menos populares que as estratégias tradicionais.



Essas duas análises apresentadas respondem a primeira questão levantada pela segunda hipótese, relacionada à similaridade dos itens recomendados, mostrando que: (1) as estratégias tradicionais são muito similares por recomendarem itens idênticos; (2) as novas estratégias são complementares As estratégias tradicionais pois, de fato, suas premissas recomendam itens distintos; e (3) os itens recomendados pelas novas estratégias são menos populares que os apresentados pelas tradicionais. Tais resultados ressaltam a validade das premissas feitas para *Max-Coverage* e *Niche-Coverage* e indicam que as estratégias avaliadas apresentam resultados práticos distintos, motivando uma análise detalhada da performance dessas abordagens para esses cenários.

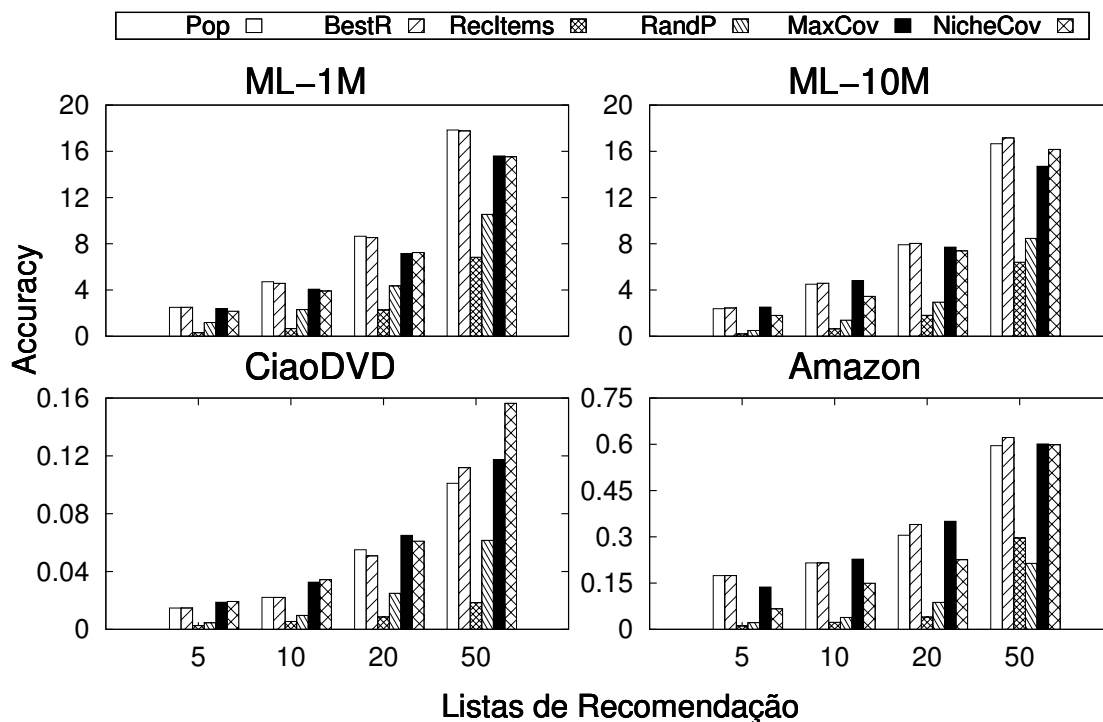
5.5.2 Avaliação da Qualidade das Recomendações

Nesta seção, avaliamos a performance das estratégias não-personalizadas sobre os principais requisitos de qualidade necessários para satisfazer usuários novos em SsR. Primeiramente avaliamos a utilidade de cada recomendador com as métricas de *Accuracy*, *Precision* e *Recall*. Nessa análise, resumimos *Precision* e *Recall* calculando a média harmônica dessas métricas, conforme a métrica de *F-measure*. Em seguida, avaliamos a diversidade dos itens apresentados com as métricas de *Diversity*, *Genre Coverage* e *Genre Frequency*. *Diversity* mensura a diversidade das listas utilizando a similaridade de *Pearson*. De maneira complementar, as outras duas métricas mensuram a diversidade com base nos gêneros (ou categorias) dos itens apresentados.

Calculamos a métrica de *Accuracy* para cada recomendador, variando o tamanho das listas geradas, e demonstramos a média dos valores encontrados como mostra a Figura 5.5. Devido à esparsidade dos dados nos cenários de *e-commerce* (i.e., CiaoDVD e Amazon), existe pouca informação dos usuários a ser utilizada pelos recomendadores, mesmo pelas estratégias estado-da-arte. Por essa razão, em uma análise preliminar dos resultados, pode-se notar que os valores de acurácia nesses domínios são muito inferiores aos valores encontrados no cenário de entretenimento (i.e., ML-1M e ML-10M). Além disso, pode-se notar uma superioridade das estratégias tradicionais (*Pop* e *BestR*) e das nossas estratégias propostas (*MaxCov* e *NicheCov*). Em outras palavras, pode-se dizer que as estratégias *RandP* e *RecItems* não conseguem apresentar itens potencialmente úteis aos usuários. A similaridade dos resultados obtidos pelas estratégias de *Pop* e *BestR* está relacionado à semelhança dos itens recomendados, como mostrado na seção anterior. Em geral, as estratégias tradicionais atingem maior acurácia nos domínios de entretenimento, em que os usuários gostam daquilo que está na moda, enquanto as nossas propostas são mais eficazes para os domínios de *e-commerce*, em que os usuários compram somente de acordo com suas preferências pessoais.

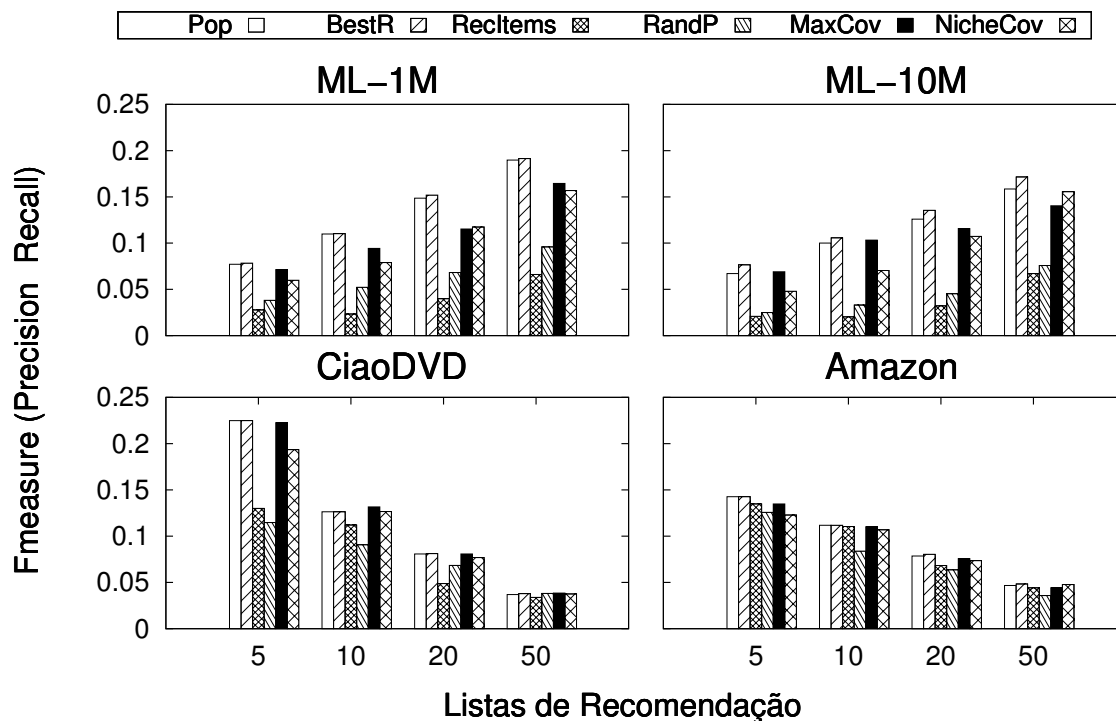
Além desse desempenho notável, se analisarmos em um cenário real de recomendação em que apenas 10 itens são apresentados, pode-se notar que *MaxCov* tem um desempenho superior aos baselines. Essa estratégia apresenta 5% de ganho para os cenários de ML-10M e Amazon, e 47% de ganho para o cenário de CiaoDVD. Também no cenário de CiaoDVD, a estratégia de *NicheCov* apresenta um ganho de 55% sobre os baselines. Para consolidar os resultados encontrados aplicamos um teste estatístico de *Wilcoxon* para distribuições não normais. Esse teste confirmou a superioridade de nossas estratégias propostas nesses cenários ressaltados com um *p-value* inferior a 0.001.

Figura 5.5: Principais resultados relacionados à utilidade dos recomendadores não-personalizados. *Max-Coverage* apresenta ganhos de acurácia em 5% para ML-10M e Amazon, e 47% para CiaoDVD. *Niche-Coverage* apresenta ganhos de 55% em acurácia para CiaoDVD com relação aos baselines.



Esses resultados de utilidade podem ser contrastados ao calcularmos a média harmônica de *Precision* e *Recall*. Basicamente, geramos um valor de *Precision* e *Recall* para cada usuário, e calculamos a média desses valores obtidos individualmente. Em seguida, calculamos a média dos valores obtidos, e sumarizamos esses resultados na Figura 5.6. Como esperado, as estratégias tradicionais conseguem ser relevantes para os usuários (i.e., apresentar itens que os usuários avaliaram positivamente). Além disso, os resultados obtidos destacam que as estratégias propostas também são capazes de apresentar itens potencialmente relevantes aos usuários.

Figura 5.6: Média harmônica dos valores de Precision e Recall obtidos por cada estratégia. Esses resultados contrastam as observações realizadas com a métrica de *Accuracy*.

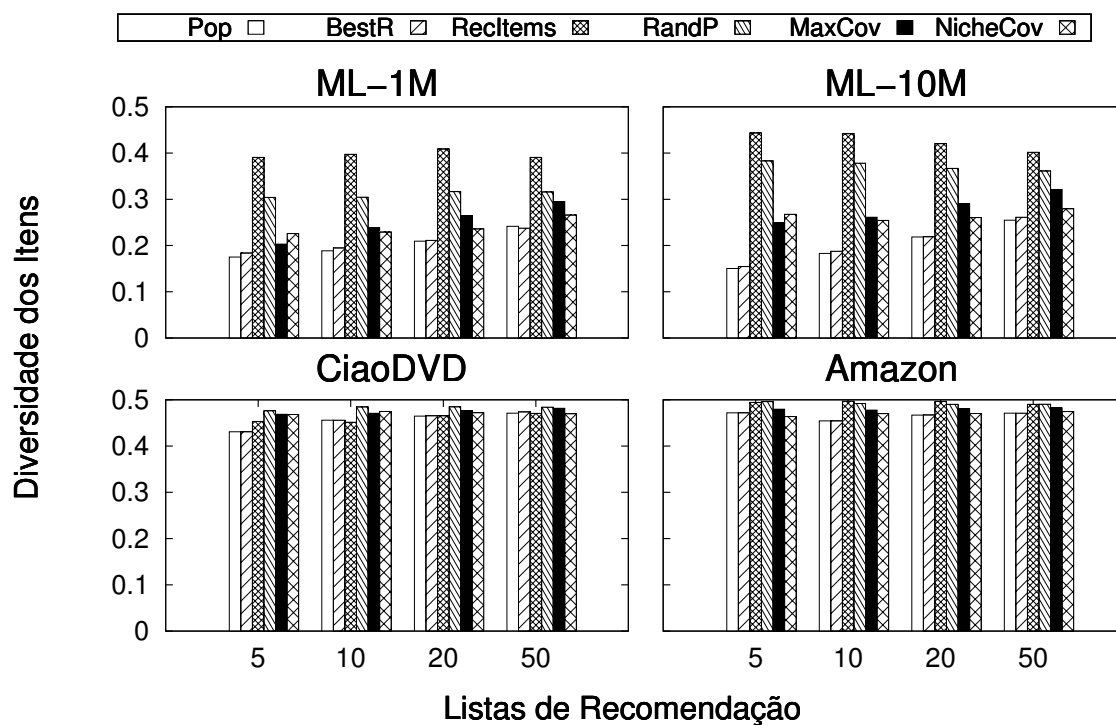


Em seguida, avalia-se a diversidade dos itens apresentados aos usuários, uma vez que o recomendador ideal para usuários novos deve apresentar itens úteis e também diversos para interessar a uma grande parcela de usuários. A Figura 5.7 sumariza a média dos resultados obtidos pela métrica de *Diversity*. Como pode-se notar, para quase todas as listas de recomendação geradas, as estratégias de *RecItems* e *RandP* apresentam mais diversidade que as demais. Entretanto, quando contrapomos essa diversidade com as análises de utilidade feitas anteriormente, pode-se notar que essas estratégias diversificam mais os itens justamente por selecionarem itens que não são relevantes aos usuários. Em outras palavras, apesar de apresentarem itens diversos, *RecItems* e *RandP* não apresentam itens úteis aos usuários. Por sua vez, as estratégias tradicionais de *Pop* e *BestR*, que apresentam maior utilidade aos usuários, não são capazes de apresentar itens muito diversos. Como essa métrica utiliza a similaridade de *Pearson*¹, pode-se dizer que esses itens estão relacionados aos mesmos usuários. Esse resultado enfatiza a hipótese levantada por este trabalho de que apenas essas estratégias tradicionais não são capazes de satisfazer todas as preferências dos diversos usuários existentes.

¹Coefficiente que mede a correlação entre dois vetores de pontos. Neste caso, os vetores de itens são formados pelos *ratings* atribuídos por usuários que consumiram o item.

As estratégias de *MaxCov* e *NicheCov* propostas atingem um nível de diversidade superior aos *baselines*. Nesse caso, o diferencial dessas estratégias é que, ao contrastarmos essa análise com as anteriores, pode-se dizer que essa diversidade é alcançada por meio de itens distintos potencialmente relevantes. Além de apresentarem itens potencialmente úteis, essas estratégias também são capazes de diversificar os itens recomendados. Especificamente, itens recomendados por *MaxCov* apresentam um ganho de *Diversity* relativo a 39% para ML-10M e 5% para Amazon e CiaoDVD. Por sua vez, *NicheCov* apresenta um ganho de 5% de diversidade no cenário de CiaoDVD, em relação aos *baselines*. Esses resultados de diversidade permitem entender um pouco mais de porque as estratégias propostas são mais eficazes em relação a acurácia nos domínios de *e-commerce*. Como dito anteriormente, nesses domínios os usuários preferem por itens específicos à sua preferência pessoal, que nem sempre são recomendados pelas estratégias tradicionais, pois essas não são capazes de diversificar. Vale destacar que todos esses ganhos de diversidade são comprovados estatisticamente, com um $p\text{-value} = 0.01$, usando o teste de *Wilcoxon* para distribuições não-normais.

Figura 5.7: *Max-Coverage* apresenta ganhos de 39% para ML-10M e 5% para Amazon e CiaoDVD. *Niche-Coverage* apresenta ganhos de 5% para CiaoDVD, em relação aos *baselines*.



Esses resultados podem ser contrastados pela métrica de *Genre Coverage*, conforme mostrado na Figura 5.8, que avalia a diversidade dos itens levando em conta as suas categorias associadas. Em uma primeira análise, pode-se notar que para os domínios relacionados aos filmes, em que as categorias são bem definidas como gêneros, todas as estratégias conseguem cobrir grande parte do domínio. Por outro lado, para o cenário da Amazon, onde existem muitas categorias, as estratégias não conseguem cobrir nem 10% dos tipos de conteúdo. Avaliando em detalhes, pode-se notar que a estratégia de *MaxCov* cobre um maior número de gêneros distintos ao apresentar apenas 10 itens na maioria dos cenários. Essa análise fica ainda mais clara quando analisamos a frequência em que os gêneros aparecem nos primeiros 10 itens recomendados pelos principais *baselines* e pelas duas novas abordagens. Conforme mostra a Figura 5.9, *MaxCov* e *NicheCov* conseguem recomendar itens com gêneros distintos aos recomendados pelos *baselines*, como “Children” em ML-1M, “Animation” e “Fantasy” em ML-10M, e “Westerns” em CiaoDVD. Esses resultados são um forte indício de que outros perfis de usuários podem ser satisfeitos. De maneira geral, essas análises de utilidade e diversidade respondem a segunda questão levantada, deixando claro que as duas novas estratégias propostas podem ser utilizadas na prática.

Figura 5.8: Resultados da métrica de *Genre Coverage*, ressaltando a diversidade dos itens recomendados por cada estratégia em todos os domínios.

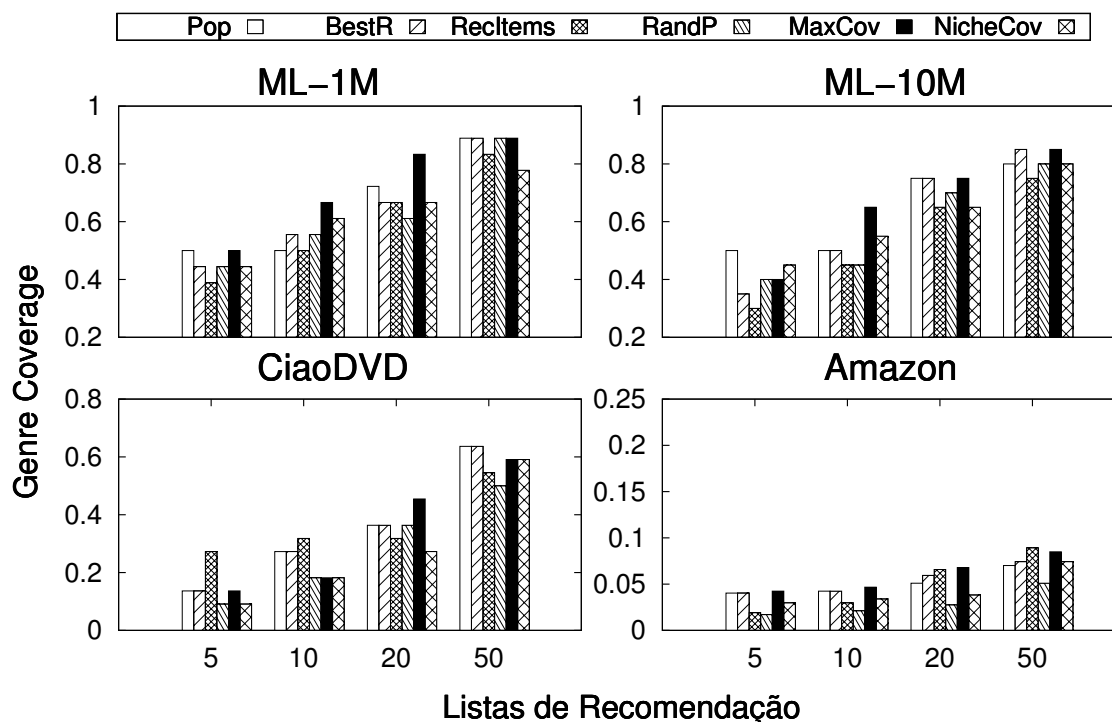
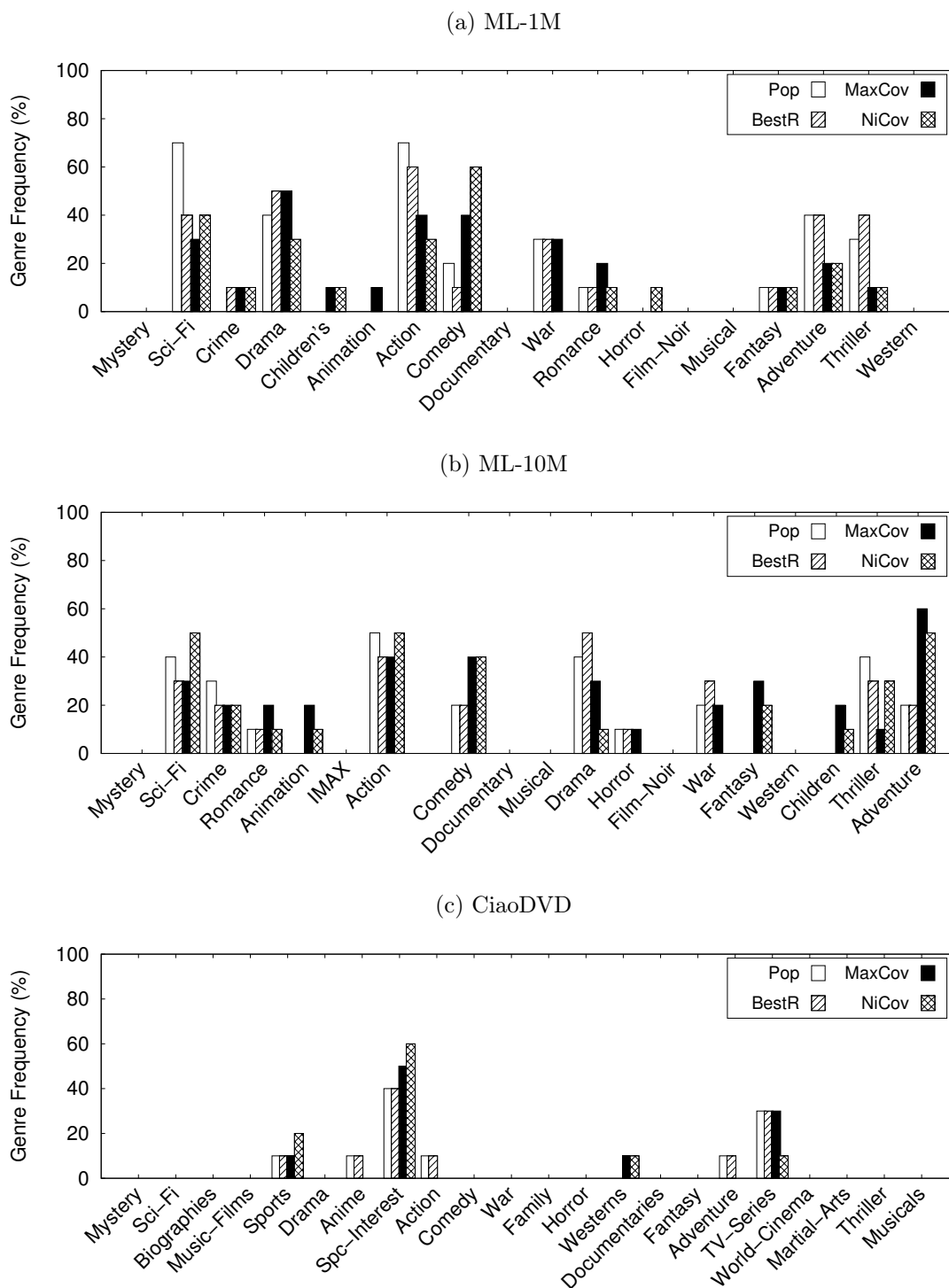


Figura 5.9: Frequência em que os gêneros dos itens aparecem nos 10 primeiros itens recomendados pelas principais estratégias. Existem diversos gêneros que são alcançados apenas pelas estratégias de *Max-Coverage* e *Niche-Coverage*. O cenário da Amazon não é mostrado nessa análise, pois existem 471 categorias associadas aos produtos.



5.5.3 Satisfação dos Usuários

A fim de mensurar a efetividade de cada recomendador em realizar a tarefa de atrair usuários novos, avalia-se a satisfação dos usuários em cada domínio. Neste trabalho, definimos que um usuário está satisfeito com um item, se e somente se, o usuário consumiu o item e atribuiu-o um *rating* maior ou igual a sua média pessoal de avaliações. Dessa forma, avaliam-se quantos usuários distintos estão satisfeitos com pelo menos 1 item recomendado por cada estratégia. O pressuposto dessa análise é que se o usuário gostou de pelo menos 1 item, o recomendador é capaz de atraí-lo para o sistema. De maneira prática, quanto mais usuários distintos uma estratégia é capaz de satisfazer, melhor são suas chances de ser eficaz para o cenário *Pure Cold-Start*. A Tabela 5.1 apresenta a porcentagem de usuários satisfeitos por cada estratégia, variando o número de itens recomendados. Os resultados marcados em negrito demonstram a maior quantidade de usuários novos satisfeitos por essas estratégias.

Tabela 5.1: Porcentagem de usuários satisfeitos com pelo menos 1 item apresentado por cada lista de recomendação de tamanho k .

	k	Pop	BestR	RItems	RandP	MaxCov	NiCov
ML-1M	5	77.9%	79.56%	11.28%	52.54%	80.1%	77.83%
	10	87%	88.72%	21.69%	68.20%	89.27%	86.84%
	20	95.3%	95.38%	60.37%	82.54%	95.85%	94.36%
	50	98.98%	98.9%	89.58%	93.81.6%	99.37%	98.27%
ML-10M	5	70.1%	75.2%	10.5%	17.8%	74.1%	55.8%
	10	80.9%	81.9%	19.9%	44.8%	82.2%	65.2%
	20	85.2%	85.5%	51.6%	66.4%	86.8%	83.8%
	50	90.5%	90.8%	81.0%	86.6%	92.9%	91.2%
CiaoDVD	5	1.24%	1.24%	0.22%	0.36%	1.70%	1.39%
	10	1.73%	1.73%	0.31%	0.82%	2.81%	2.72%
	20	4.39%	4.54%	0.51%	2.01%	5.30%	4.48%
	50	7.69%	8.43%	0.90%	4.14%	8.62%	10.78%
Amazon	5	12.59%	12.59%	0.49%	1.30%	10.42%	5.64%
	10	14.51%	14.51%	1.55%	2.35%	15.03%	10.04%
	20	17.99%	21.33%	2.29%	5.83%	22.27%	14.59%
	50	30.39%	31.88%	12.59%	13.02%	32.50%	31.82%

Em uma análise preliminar, é possível notar que mesmo recomendando 50 itens, nenhuma estratégia é capaz de satisfazer 100% dos usuários novos dos domínios. Também é possível notar que as estratégias de *RecItems* e *RandP* possuem resultados bem inferiores com relação às demais estratégias. Como foi dito anteriormente, essas estratégias não são capazes de apresentar itens relevantes aos usuários novos dos domínios.

Os resultados apresentados pelo Diagrama de Venn da Figura 5.10, mostram essas interseções dos usuários alcançados ao recomendar 10 itens. Em geral, pode-se notar que existe uma larga porção de usuários que são satisfeitos por todas as estratégias. Nos cenários de entretenimento, por exemplo, mais de 60% dos usuários satisfeitos agradaram com pelo menos 1 item de cada estratégia. Entretanto, os resultados mostram que existem usuários que são satisfeitos apenas pelas estratégias de *MaxCov* e *NicheCov*. No cenário do ML-10M, mais de 4% dos usuários satisfeitos, resultam somente da estratégia de *MaxCov*. O mesmo pode ser dito para o cenário da Amazon, onde mais de 5% dos usuários satisfeitos se interessaram apenas em itens do *NicheCov*. Esses resultados respondem a terceira questão levantada, evidenciando que essas estratégias satisfazem diferentes conjuntos de usuários e podem ser aplicadas em conjunto nos cenários reais.

5.6 Síntese do Capítulo

Este capítulo se inicia destacando três principais questões sobre a segunda hipótese deste trabalho. Por meio dessas questões, consolidamos uma metodologia de avaliação dividida em três dimensões distintas de análise, relacionadas a cada uma das questões levantadas. Além disso, definimos os principais passos a serem realizados ao longo deste capítulo para que possamos consolidar nossa segunda hipótese. Dessa forma, primeiro selecionamos um conjunto de usuários para simular o cenário de *Pure Cold-Start*. Nessa etapa, para cada um dos cenários avaliados, selecionamos em torno de 20% dos últimos usuários a ingressarem no sistema. Em seguida, destacamos os principais requisitos de qualidade necessários para um recomendador, relacionados à utilidade e diversidade dos itens. Logo após, destacamos as estratégias não-personalizadas tradicionais a serem utilizadas como *baselines* deste trabalho.

Finalmente, aplicamos a metodologia de avaliação sobre cada domínio, utilizando os usuários selecionados como o conjunto teste. Em síntese, a primeira dimensão de análise responde a primeira questão levantada neste capítulo, dado que: (1) as estratégias tradicionais são similares pois recomendam itens iguais; e (2) as novas estratégias são complementares às tradicionalmente utilizadas pois recomendam itens distintos. Também observamos que os novos SsR recomendam itens que não são classificados como populares, consolidando as premissas assumidas por cada estratégia. Por sua vez, a segunda dimensão de análise mostra que os itens distintos apresentados por esses novos SsR acabam adicionando diversidade às estratégias. Além disso, os resultados das métricas mostram que esses itens também são responsáveis por aumentar a acurácia das estratégias em alguns casos. Por fim, com a terceira dimensão vemos que esses SsR são

capazes de atrair mais usuários novos e subconjuntos complementares. Devido a esse comportamento complementar, esses resultados mostram que os SsR avaliados podem ser aplicados em conjunto nos cenários reais, como discutido no próximo capítulo.

Capítulo 6

Estudos Direcionados aos Usuários

O objetivo deste capítulo é verificar a terceira hipótese deste trabalho: “*Páginas de produtos compostas por SsR não-personalizados complementares satisfazem o interesse de um grande número de usuários novos distintos*”. Primeiramente, avaliamos quais estratégias que combinadas alcançam o maior número de usuários possível. Nesta etapa simulamos um cenário real de aplicação onde são apresentadas três listas de recomendação. Em seguida, avaliamos o desempenho de um sistema de entretenimento modelado com essa abordagem ressaltada. Para tal, construímos uma metodologia de avaliação *online* que visa caracterizar essa nova abordagem de acordo com as interações de usuários reais e mensurar métricas que não foram avaliadas de maneira *offline*. Os resultados encontrados nos permitem consolidar essa hipótese, bem como confirmar a aplicabilidade das estratégias propostas por este trabalho na prática.

6.1 Combinação de Estratégias não-Personalizadas

Nos principais domínios de recomendação, a maioria das páginas de vendas/entretenimento possuem diversos produtos selecionados por estratégias distintas. Em geral, esses sistemas optam por compor suas páginas com as estratégias de *Popularity*, *Best-Rated* e *Recent Items* a fim de apresentar diversos produtos. Entretanto, como visto no Capítulo 5, os itens populares tendem a ser os mais bem avaliados do sistema, de modo que o desempenho de *Best-Rated* se torna muito similar ao de *Popularity*. Além disso, vemos que os SsR propostos satisfazem subconjuntos complementares de usuários em relação aos SsR tradicionais. E, também, que existem usuários que são atraídos apenas pelos itens apresentados por *Max-Coverage* e *Niche-Coverage*. Dessa forma, os itens recomendados pelos novos SsR devem ser combinados com outras recomendações nas *interfaces* dos sistemas, como sugere nossa terceira hipótese.

Em uma simples análise para confirmar a veracidade dessa hipótese, avalia-se a satisfação dos usuários simulando um cenário real, como *netflix.com* ou *amazon.com*, onde as estratégias são apresentadas em conjunto. Para tal, geram-se cinco combinações distintas com as estratégias de recomendação, onde cada combinação possui três estratégias que recomendam 10 itens cada. Em seguida, avalia-se a quantidade de usuários satisfeitos com estes itens combinados. Os resultados obtidos estão sumarizados na Tabela 6.1.

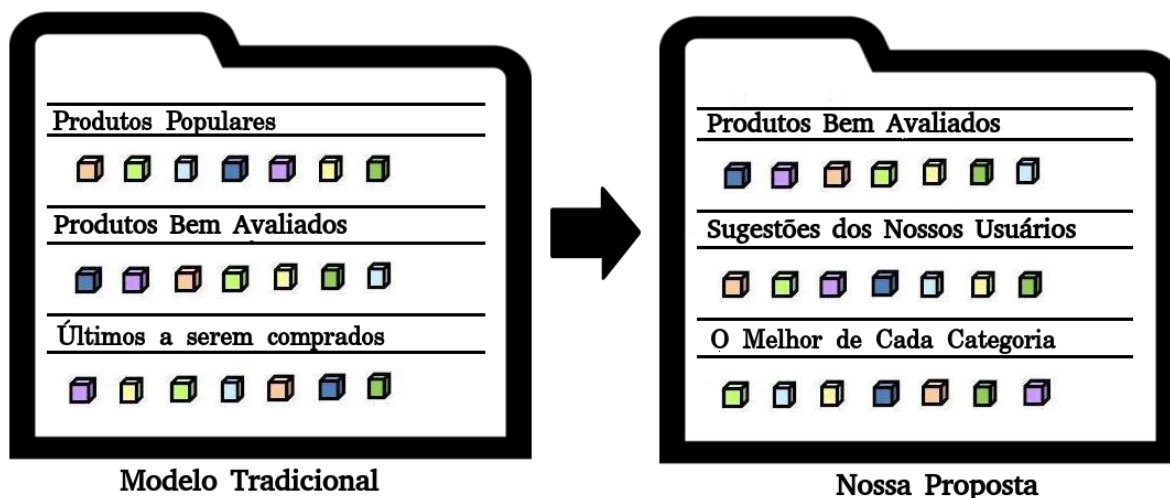
Tabela 6.1: Porcentagem de usuários novos satisfeitos pela combinação de estratégias, como feito em cenários reais como *Amazon* e *Netflix*.

Estratégias Combinadas	Usuários Satisfeitos			
	ML-1M	ML-10M	CiaoDVD	Amazon
Pop, BestR, RecItems	91.30%	83.38%	2.04%	15.38%
Pop, BestR, MaxCov	94.12%	87.78%	2.81%	16.19%
Pop, BestR, NiCov	94.04%	84.65%	3.20%	21.52%
Pop, MaxCov, NiCov	94.67%	88.05%	3.46%	22.02%
BestR, MaxCov, NiCov	95.45%	88.70%	3.46%	22.02%

Diferentemente das principais aplicações reais, onde as estratégias combinadas são $\langle \text{Popularity, Best-Rated, Recent Items} \rangle$ (cenário A), esse resultado mostra que combinar $\langle \text{Best-Rated, Max-Coverage, Niche-Coverage} \rangle$ (cenário B) é mais relevante para usuários novos. Enquanto a combinação tradicional (cenário A) atrai 1.166, 8.865, 71, e 247 usuários para ML-1M, ML-10M, CiaoDVD e Amazon, a combinação proposta (cenário B) atrai 1.219, 9.432, 121, e 355 usuários, respectivamente. Esses pequenos ganhos obtidos em cada domínio se tornam ainda mais significantes se levarmos em conta que essas estratégias vão lidar com milhões de usuários em cenários completos de utilização. Dessa forma, esses resultados apontam uma direção a ser seguida pelos proprietários de domínios de recomendação, no intuito de melhorar sua página de produtos para lidar com os usuários novos.

Na prática, nossa proposta consiste em alterar a forma com que os itens são apresentados pelas aplicações de *e-commerce*, conforme mostra a Figura 6.1. Como vimos, os proprietários de *e-commerce* optam por apresentar o cenário A, a fim de maximizar os seus lucros. Por sua vez, nossa proposta sugere que para melhor atrair usuários novos e conseqüentemente maximizar os lucros, devemos apresentar os itens mais bem avaliados (*BestR*), aqueles que os nossos usuários sugeriram (*MaxCov*) e os melhores de cada categoria (*NiCov*), tal qual o cenário B. Essa abordagem parte da premissa de que estratégias complementares são capazes de atrair mais usuários novos.

Figura 6.1: Esquema intuitivo do modelo atual utilizado pelas aplicações *e-commerces* e da nossa nova proposta para atrair mais usuários novos.



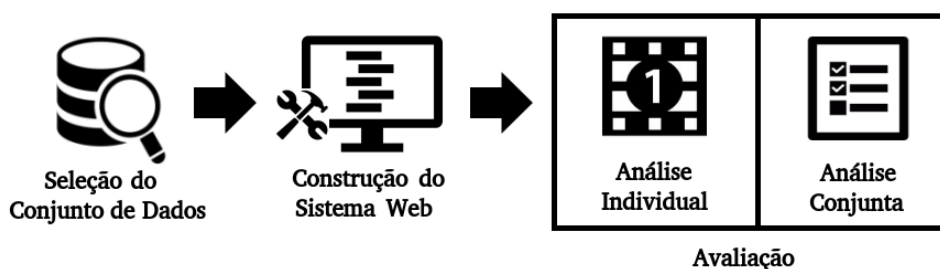
6.2 Experimento Online

Esse experimento *online* é conduzido a fim de confirmar a terceira hipótese deste trabalho ao estimar o real potencial dos SsR propostos aplicados em conjunto. Por meio dos experimentos *offline* do Capítulo 5, sabemos os prós e contras de cada SsR não-personalizado sobre os principais requisitos analisados. Por sua vez, com as avaliações feitas na seção anterior, vimos que a nossa proposta de estratégias alinhadas conquista o interesse de mais usuários do que o modelo tradicional. De maneira geral, ambos os resultados confirmam a necessidade de avaliar um *feedback* explícito dos usuários reais. Além disso, com esse experimento torna-se possível mensurarmos outras métricas de avaliação complementares às estudadas de maneira *offline*.

6.2.1 Metodologia de Avaliação

Com o intuito de avaliar cada SsR na prática e mensurar outras métricas de avaliação, definimos nossa metodologia em distintas etapas, como mostra a Figura 6.2.

Figura 6.2: Metodologia de Avaliação *Online*.



A primeira etapa consiste em selecionar um conjunto de dados relacionado ao cenário de filmes, que tem se tornado cada vez mais relevante para os usuários. O desafio nesse caso é encontrar uma base de dados com informações recentemente adicionadas, que se adequem aos interesses dos usuários. Em seguida, o desafio é construir um sistema de avaliação que seja capaz de simular um cenário real no qual os usuários novos lidam diariamente. Esse sistema *Web* deve também proporcionar a realização de todas as avaliações necessárias para esse experimento. Por essa razão, definimos o processo de avaliação em duas etapas: (1) *Avaliação Conjunta*: que visa comparar o novo modelo proposto com o modelo tradicionalmente utilizado; e (2) *Avaliação Individual*: que visa mensurar o interesse do usuário por cada recomendador, bem como outras métricas ainda não avaliadas. Na primeira etapa, apresentamos 10 filmes de cada estratégia visando realizar um teste A/B para comparar as combinações de estratégias. Basicamente, para cada usuário, apresentamos um cenário A com recomendações do modelo tradicional, ou um cenário B com filmes do modelo proposto. O objetivo dessa etapa é avaliar quantos usuários distintos foram atraídos pelo sistema, bem como qual item despertou o seu interesse. Além disso, nessa etapa também é realizado um questionário para entendermos com mais precisão as ações tomadas por cada usuário.

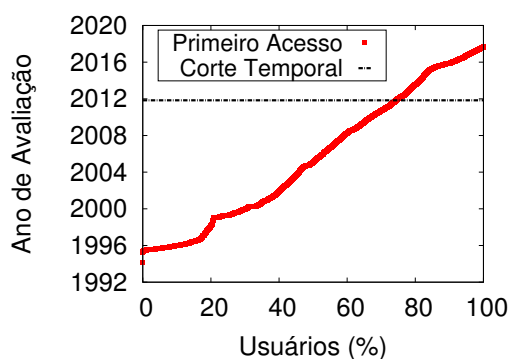
Na segunda etapa de avaliação, apresentamos para os usuários todos os 10 primeiros itens recomendados por cada uma das principais estratégias (i.e., *Popularity*, *Best-Rated*, *Max-Coverage* e *Niche-Coverage*). Nessa etapa, estamos interessados em avaliar o interesse dos usuários, bem como dois conceitos fundamentais em recomendação: *novidade* e *serendipidade*. A novidade de uma informação geralmente se refere a quão diferente essa informação é com relação a tudo que tinha sido previamente observado por um usuário específico [Vargas & Castells, 2011]. Por sua vez, o conceito de serendipidade está relacionado a uma forma de mensurar o quão surpreso o usuário ficou com o sucesso das recomendações [Zhang et al., 2012]. Repare que em avaliações *offline* esses conceitos são difíceis de avaliar, por não existirem informações sobre os usuários que nos permitam extrair esses conceitos. Para mensurar a utilidade das recomendações, primeiramente, perguntamos ao usuário quais filmes estão relacionados ao seu interesse pessoal. Em seguida, perguntamos quais filmes o usuário não conhecia ou nunca tinha ouvido falar, a fim de avaliarmos a novidade dos itens apresentados. Por fim, dos filmes novos para o usuário, perguntamos quais deles o usuário teria interesse em assistir para mensurarmos a *serendipidade* (i.e., algo inesperado e ao mesmo tempo útil ao usuário) de cada estratégia. Em suma, esperamos que esse experimento seja capaz de contrastar e complementar os resultados encontrados com a avaliação empírica *offline*. Além disso, vale a pena ressaltar que os participantes desconheciam a forma como geramos as recomendações dos itens.

6.2.2 Seleção do Conjunto de Dados

Um grande desafio da realização desse experimento *online* é a seleção de um conjunto de dados que contenha filmes recentes capazes de refletir nas preferências atuais dos usuários reais. Essa necessidade está relacionada às estratégias não-personalizadas que utilizam a premissa de popularidade, que é fortemente ligada a uma noção temporal. Um filme, por exemplo, pode ser popular apenas por um determinado período de tempo relacionado ao seu lançamento ou mesmo a sua divulgação. Desse modo, após uma vasta pesquisa na literatura, selecionamos o conjunto de dados de filmes da *MovieLens Latest*¹, atualizado em agosto de 2017. Originalmente, esse conjunto de dados possui 26M *ratings* atribuídos em uma escala de 1 a 5 por, aproximadamente, 270.000 usuários a 45.000 filmes disponíveis. Os filmes do domínio estão associados a 19 gêneros distintos, que podem ser comédia, ação, terror, drama e outros.

Esse conjunto de dados foi criado por usuários que interagiram com o sistema da *MovieLens* entre 09 de Janeiro de 1995 e 04 de Agosto de 2017. Por essa razão, mesmo esse conjunto de dados sendo atualizado em 2017, ao avaliarmos quando os usuários ingressaram no sistema (i.e., realizaram sua primeira avaliação), na Figura 6.3, notamos que essa base ainda é defasada. A maioria das avaliações relacionadas a esse conjunto de dados são referentes aos primeiros anos de coleta de dados. Dessa forma, para esse experimento *online* é necessário realizarmos um corte temporal nos dados obtidos. Para melhor esboçarmos as preferências dos usuários atualmente, utilizaremos apenas as informações relacionadas aos últimos 5 anos de coleta (i.e., de 2012 a 2017). Com esse corte temporal o conjunto de dados fica com, aproximadamente, 25% dos usuários da base (i.e., 69.295 usuários) e com quase todos os itens do domínio (i.e., 41.673 filmes).

Figura 6.3: Corte temporal realizado no ano de 2012, a fim de selecionar as preferências dos usuários refletidas atualmente.



¹Disponível em: <https://grouplens.org/datasets/movielens/>

6.2.3 Sistema Web de Avaliação

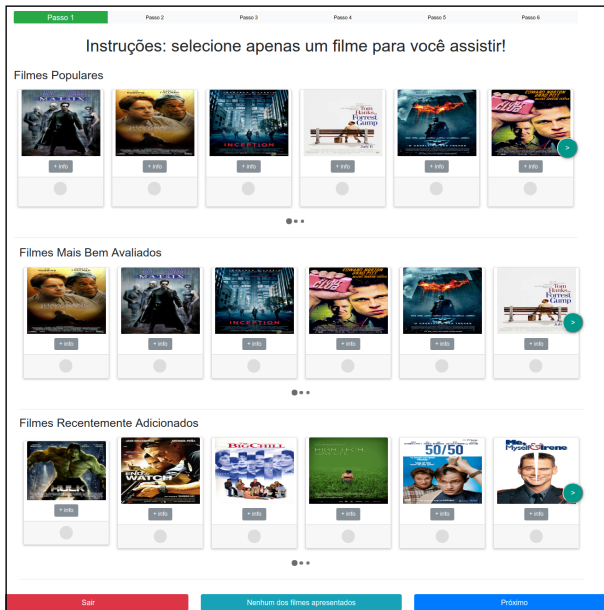
Para aplicar a metodologia de avaliação proposta, foi construído um sistema *Web* dividido em 6 fases distintas, conforme descrito a seguir e ilustrado na Figura 6.4.

1. **Seleção do Filme Ideal** - Para cada novo usuário é apresentada uma das combinações de três estratégias, contendo 10 filmes de cada um dos recomendadores. As estratégias selecionadas variam entre o conjunto $\langle \textit{Popularity}, \textit{Best-Rated}, \textit{Recent Items} \rangle$ e $\langle \textit{Best-Rated}, \textit{Max-Coverage}, \textit{Niche-Coverage} \rangle$. Nessa fase, pede-se ao usuário que selecione um filme para assistir ou a opção "Nenhum Filme Apresentado". Com os resultados é possível observar quais listas foram mais escolhidas pelos usuários, bem como qual combinação foi capaz de atrair mais usuários.
2. **Questionário** - Nesta fase pede-se aos usuários para responder seis perguntas sobre as recomendações apresentadas para ele/ela no passo anterior. O questionário completo está disponível no Apêndice A.
3. **Interesse Pessoal** - Para cada usuário são apresentados 10 itens de cada uma das estratégias de $\langle \textit{Popularity}, \textit{Max-Coverage}, \textit{Niche-Coverage}, \textit{Best-Rated} \rangle$. Em seguida pede-se aos usuários para selecionarem os filmes que eles/elas julgam como mais interessantes de acordo com suas preferências pessoais. Essa fase visa contrastar os resultados encontrados na avaliação *offline*, mostrando a relevância de cada um dos recomendadores.
4. **Novidade dos Itens** - Novamente, são apresentados aos usuários os mesmos 10 filmes do passo anterior. Porém, nessa fase, pede-se aos usuários para selecionarem os filmes que eles/elas não conhecem ou mesmo nunca ouviram falar. Com esses resultados é possível mensurar quanto de novidade cada recomendador apresenta.
5. **Serendipidade dos Itens** - Nessa fase são apresentados todos os filmes que foram selecionados pelo usuário no passo anterior. Contudo, pede-se aos usuários para selecionarem, dentre os filmes que são novos, aqueles que eles/elas teriam interesse em assistir. Com esses resultados pode-se extrair a utilidade dos itens novos apresentados aos usuários (i.e., o conceito de serendipidade).
6. **Perfil do Usuário** - Por fim, pede-se para cada participante responder um questionário sobre suas informações básicas, bem como o seu interesse pessoal por filmes. O questionário utilizado nessa fase está disponível no Apêndice B.

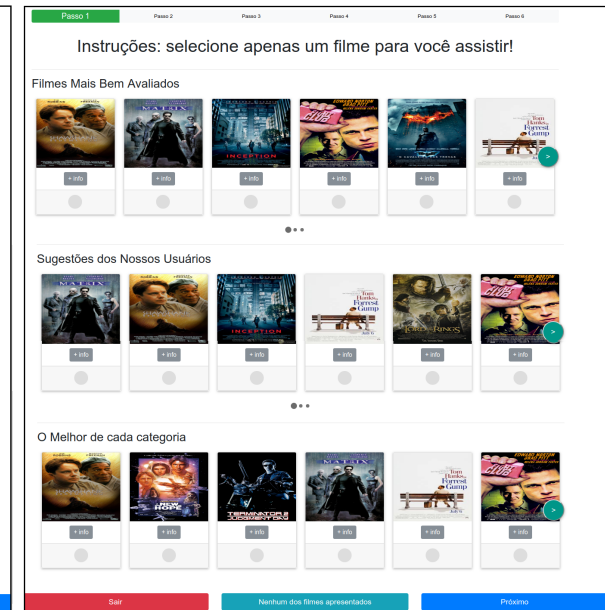
As fases 1 e 2 da pesquisa correspondem à etapa de *avaliação conjunta* de nossa metodologia, enquanto as demais fases correspondem a etapa de *avaliação individual*.

Figura 6.4: *Snapshots* da Interface Web.

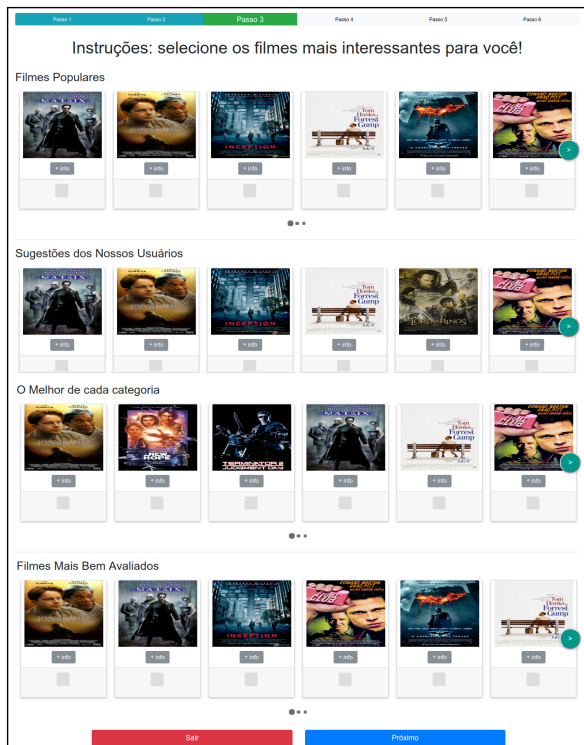
(a) Fase 1 - Cenário A



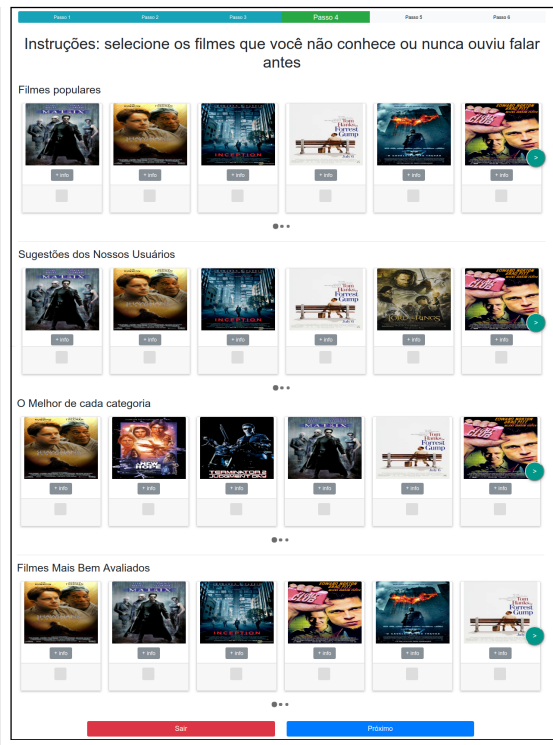
(b) Fase 1 - Cenário B



(c) Fase 3



(d) Fase 4



The image contains two screenshots of a web-based user interface for a study. The left screenshot, labeled '(e) Fase 5', shows a progress bar at the top with 'Passo 5' highlighted. Below it, the instruction reads: 'Instruções: Dentre os filmes que você não conhece, marque quais você teria interesse em assistir'. There are six movie posters in a row: 'The Matrix', 'Inception', 'The Bourne Supremacy', 'The Bourne Ultimatum', 'Pulp Fiction', and 'Schindler's List'. Each poster has a 'Marcar' button below it. At the bottom, there are 'Sair' and 'Próximo' buttons. The right screenshot, labeled '(f) Fase 6', shows a progress bar with 'Passo 6' highlighted. The instruction reads: 'Instruções: Fale um pouco de você'. It contains seven numbered questions with radio button options: 1. '1. Idade' (input field); 2. '2. Sexo' (male/female); 3. '3. Com que frequência você assiste filmes ou séries?' (daily/weekly/once a month/rarely/never); 4. '4. Qual a sua preferência por filmes?' (classic world taste/critically acclaimed taste/modern taste/specific tastes); 5. '5. Você é assinante de algum sistema de entretenimento, como Netflix, Spotify, etc que recomenda itens para o seu consumo?' (yes/no); 6. '6. Você se sente confuso (i.e., demora muito) para encontrar os itens de seu interesse (nó(s) sistema(s) que utiliza?)' (very confused/slightly confused/neutral/not me/definitely not); 7. '7. Você acha que os itens apresentados pelo(s) sistema(s) que utiliza são diversificados (i.e., apresentam muitos itens variados)?' (definitely yes/slightly yes/less or more/no/definitely not). Both screenshots have 'Sair' and 'Próximo' buttons at the bottom.

(e) Fase 5

(f) Fase 6

Vale a pena destacar que toda a interação do usuário com o sistema deu-se após a aceitação de um termo de compromisso que explica cada uma das fases, deixando claro os objetivos, riscos e benefícios da participação, como mostra o Apêndice C. Além disso, como pode-se notar na interface, era permitido ao usuário deixar o sistema quando quisesse ou se sentisse desconfortável. A divulgação da pesquisa foi realizada por meio de canais de comunicação como o *Facebook*, *Gmail*, *WhatsApp* e outros.

6.3 Análises dos Resultados

A pesquisa *online* foi divulgada durante dez dias, no período de 16 a 26 de Novembro de 2017, alcançando 236 usuários que participaram avaliando os filmes recomendados. Desses usuários alcançados, durante a etapa de avaliação conjunta dos recomendadores, 122 usuários interagiram com o cenário A tradicional e 114 usuários interagiram com o cenário B proposto. Vale ressaltar que nem todos os usuários responderam a todas as fases de pesquisa, pois era permitido pelo sistema que o usuário deixasse a pesquisa a qualquer momento.

6.3.1 Análise Conjunta dos Recomendadores

Nessa etapa avaliamos as fases 1 e 2 do nosso sistema *Web*, esboçando um teste A/B que visa comparar a combinação tradicional de recomendadores com a combinação proposta por este trabalho. Na fase 1, avaliamos se os usuários optaram por algum filme dentre os apresentados ou não escolheram nenhum para assistir. Além disso, avaliamos também quais os filmes foram selecionados pelos usuários durante a interação com o sistema. Os resultados dessa fase estão sumarizados na Tabela 6.2.

Lembrando que no cenário A os recomendadores disponíveis são *Popularity*, *Best-Rated* e *Recent Items*, respectivamente. Por sua vez, no cenário B os recomendadores relacionados são, respectivamente, *Best-Rated*, *Max-Coverage* e *Niche-Coverage*.

Tabela 6.2: Porcentagem de usuários que selecionaram cada uma das opções durante a fase 1, especificando de qual recomendador o item foi escolhido.

Opção Selecionada	Total de Usuários	
	Cenário A	Cenário B
Topo da Página	46,28%	51,36%
Meio da Página	14,88%	9,82%
Rodapé da Página	29,75%	34,82%
Nenhum filme	9%	4%

Em ambos os modelos avaliados, pode-se notar que a maioria dos usuários selecionaram itens da primeira lista de recomendação, referente a *Popularity* no cenário A e a *Best-Rated* no cenário B. Além da eficácia dessas duas estratégias, um dos motivos que podem ter influenciado na decisão dos usuários é a interface de avaliação que apresenta esses recomendadores antes dos demais. Em geral, em sistemas *Web* os usuários interagem de cima para baixo no sistema, de forma que visualizaram primeiro esses dois recomendadores [Shneiderman, 2010]. Outro ponto pertinente é que, apesar de ser o mesmo recomendador no meio da página do cenário A e no topo da página do cenário B, a quantidade de usuários que interagem com essas listas é diferente. Basicamente, esse resultado está relacionado à localização das estratégias e ao fato de que no cenário A as duas primeiras listas são muito similares, levando o usuário a escolher apenas da primeira. Por outro lado, nota-se também uma alta porcentagem de usuários que optaram por filmes presentes na terceira lista de recomendação, referente a *Recent Items* e *Niche-Coverage*, respectivamente. Nesse caso, um dos principais fatores relacionados a esse resultado é a eficácia dessas estratégias na prática. Novamente, nos dois modelos avaliados as duas primeiras listas de recomendação apresentam filmes muito similares entre si, enquanto a última lista diversifica o conteúdo.

De maneira complementar, nota-se que 9% dos usuários que interagiram com o cenário A não encontraram nenhum filme interessante para assistir. Por outro lado, apenas 4% dos usuários que interagiram com o cenário B optaram por não assistir nenhum filme dentre os apresentados. Apesar dessa pequena diferença, esses resultados apontam uma pequena melhoria em atrair novos usuários ao utilizar o modelo proposto, conforme sugerido por nossa hipótese. Entretanto, são necessárias análises mais extensivas, com um maior número de usuários, para comprovar essa hipótese.

Na fase 2 do sistema de avaliação, 182 usuários responderam a maioria das questões relativas aos modelos apresentados na fase anterior. Essas respostas estão sumarizadas na Figura 6.4 e são apresentadas em detalhes a seguir. Para esses resultados específicos, agrupamos as opções *concordo totalmente* e *concordo*, e as opções *discordo totalmente* e *discordo*.

Q1: Os filmes apresentados correspondem ao meu interesse por filmes.

A maioria dos usuários concordam com essa afirmação para ambos os modelos apresentados, com uma pequena superioridade de 3% para o cenário A em relação ao cenário B.

Q2: Os filmes recomendados para mim são famosos e bem conhecidos.

Um total de 93,02% dos usuários concordam que os filmes do cenário B são famosos e amplamente conhecidos, enquanto 84,38% dos usuários acham isso para os filmes do cenário A tradicional.

Q3: O sistema apresenta diversos gêneros de filmes distintos.

Em geral os usuários acharam o cenário B menos diverso que o cenário A. Enquanto no cenário A, 21,88% discordam que os filmes apresentam gêneros distintos, 30,23% discordam da afirmação no cenário B.

Q4: Essas recomendações me surpreenderam, pois eu não conhecia ou nunca tinha ouvido falar de vários dos filmes apresentados.

Em geral, os usuários não ficaram surpresos com essas recomendações geradas por nenhum dos dois modelos. Em torno de 85,42% e 88,37% dos usuários discordam dessa afirmação para os cenário A e B, respectivamente.

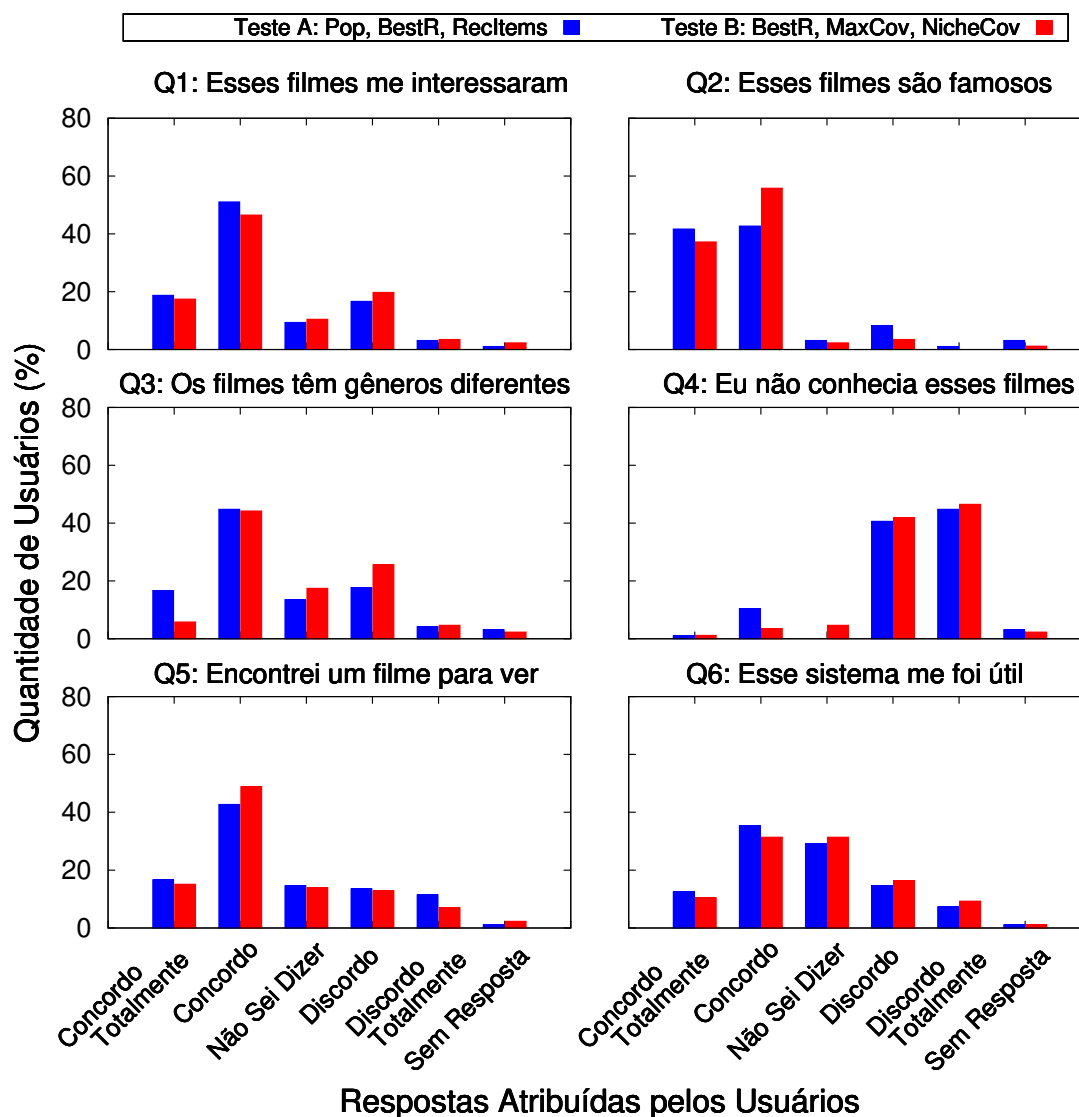
Q5: De maneira geral, eu estou satisfeito com as recomendações recebidas pois consegui encontrar um filme interessante para assistir.

A maioria dos usuários disseram que o cenário B é mais relevante, permitindo encontrarem um filme para assistir. Enquanto 59,38% dos usuários concordam com essa afirmação no cenário A, 63,96% dos usuários se sentiram satisfeitos com as recomendações do cenário B.

Q6: Eu utilizaria novamente essas listas de filmes para me auxiliar.

A maioria dos usuários concordaram ou não souberam opinar se utilizariam essas recomendações. Esses resultados eram esperados pois este não é um sistema real de consumo, e sim, apenas uma análise.

Figura 6.4: Respostas dos usuários para o questionário da fase 2 da pesquisa.

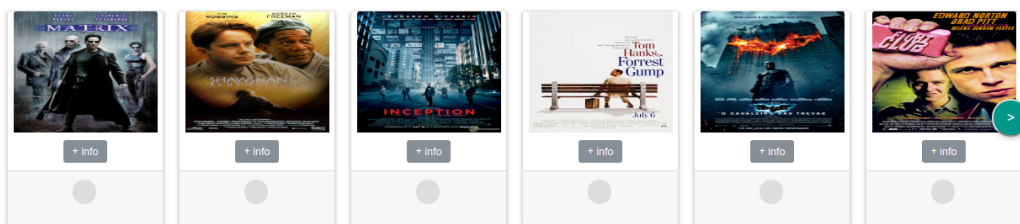


A princípio, essas respostas parecem contradizer alguns resultados das avaliações *offline*. Segundo os usuários as listas tradicionais do cenário A são mais diversas e surpreendentes do que as listas do cenário B, como mostram as questões Q3 e Q4. Além disso, os usuários disseram que os filmes presentes no cenário B são mais famosos do que os presentes no cenário A. Contudo, essas respostas refletem o fato de que a estratégia *Recent Items* está inserida no cenário A. Como mostrado nos resultados *offline*, essa estratégia diversifica muito os itens, apresentando filmes distintos das demais estratégias quando aplicada na prática. Sendo assim, o cenário A se torna mais diverso que o B, fazendo que os usuários apontem o cenário B como aquele capaz de apresentar pelo menos um item que os satisfaçam.

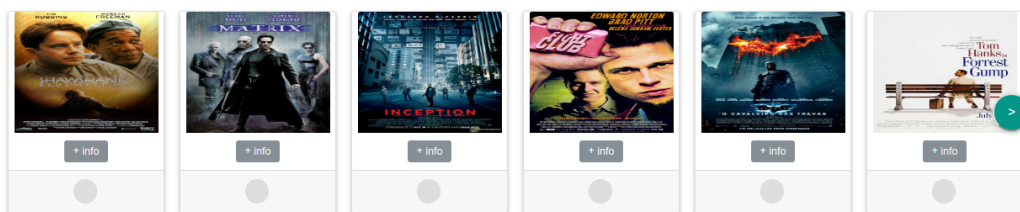
Para vermos as diferenças das estratégias, basta analisarmos os cinco filmes apresentados por cada SsR, conforme mostra a Figura 6.5. Note que os filmes recomendados por *Recent Items* (última lista) são totalmente distintos dos demais, não aparecendo em nenhuma das demais listas. Para os usuários reais, que estão interagindo com o sistema, tais itens tornam-se altamente relevantes e acabaram influenciando nas respostas de cada usuário nessa pesquisa.

Figura 6.5: Filmes apresentados pelos recomendadores.

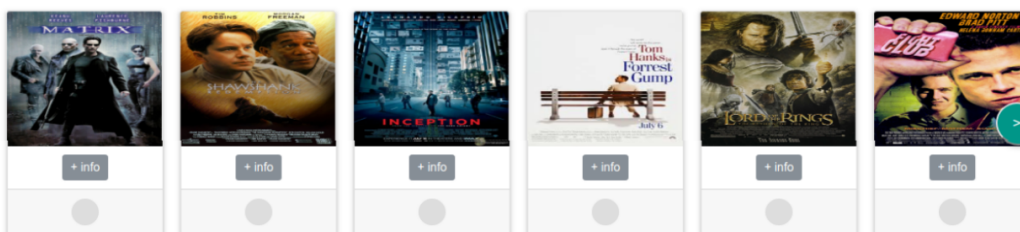
Filmes Populares



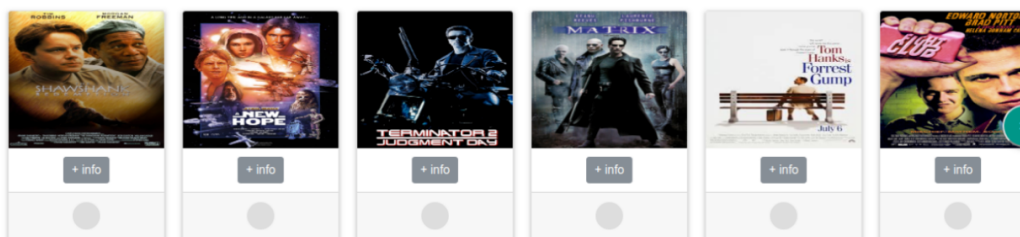
Filmes Mais Bem Avaliados



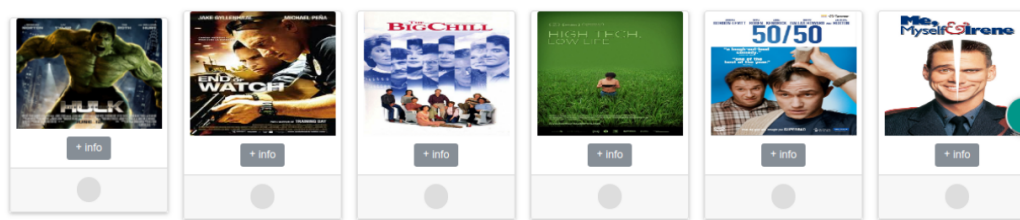
Sugestões dos Nossos Usuários



O Melhor de cada categoria



Filmes Recentemente Adicionados



Em suma, essas análises mostram o cenário proposto foi capaz de conquistar mais usuários que o cenário tradicional, uma vez que 96% dos usuários encontraram pelo menos um filme para assistir e 63,96% saíram satisfeitos com as recomendações. Esse resultado reforça a terceira hipótese, mostrando que SsR complementares satisfazem mais usuários novos distintos. Além disso, os resultados apontam a relevância da estratégia de *Recent Items* na prática, por ser capaz de diversificar mais os resultados. Logo, uma nova abordagem utilizando o *Recent Items* ao invés do *Best-Rated* pode ser analisada em um novo experimento com mais usuários. Vale a pena destacar que, infelizmente, apenas esse experimento não é suficiente para obter todas as respostas necessárias devido às restrições de usuários e domínio.

6.3.2 Análise Individual dos Recomendadores

Nessa etapa avaliamos as demais fases do nosso sistema *Web*, que visam mensurar métricas complementares às avaliadas de maneira *offline*: utilidade, novidade e serendipidade, respectivamente. Primeiramente, na fase 3, 152 pessoas avaliaram os filmes recomendados pelas listas de $\langle \textit{Popularity}, \textit{Max-Coverage}, \textit{Niche-Coverage}, \textit{Best-Rated} \rangle$ e marcaram aqueles que eram mais interessantes, de acordo com suas preferências pessoais. O intuito dessa análise foi para verificar se os filmes recomendados por cada estratégia são mesmo relevantes para os usuários reais. Para comparar as estratégias, construímos *rankings* contendo quantos filmes foram marcados como interessantes em cada lista de recomendação. Todas as informações foram sumarizadas na Tabela 6.3, que contém a área sob a curva (AUC), a média e desvio padrão de cada *ranking*. Assim como nos resultados *offline*, os itens populares e bem avaliados são os que despertam mais interesse nos usuários. Contudo, não encontramos uma diferença significativa entre as estratégias, enfatizando a relevância de cada uma das abordagens em recomendar itens relevantes aos usuários novos. *Max-Coverage* e *Niche-Coverage* se mostram igualmente relevantes às estratégias estado-da-arte sob a perspectiva dos usuários reais que participaram da pesquisa.

Tabela 6.3: Resultados da fase 3, especificando cada recomendador utilizado.

Recomendadores	AUC	Itens Selecionados	
		Média	Desvio Padrão
<i>Popularity</i>	0,38553	3,88	2,69
<i>Best-Rated</i>	0,37763	3,80	2,54
<i>Max-Coverage</i>	0,37463	3,40	1,99
<i>Niche-Coverage</i>	0,32632	3,29	2,26

Na fase 4 do sistema de avaliação, 133 usuários interagiram com os mesmos filmes recomendados na fase anterior a fim de selecionar os filmes que nunca assistiram ou ouviram falar. Nessa fase, visamos avaliar se os filmes recomendados por cada estratégia são novos para os usuários reais. Novamente, gera-se um *ranking* para cada estratégia contendo a quantidade de filmes selecionados por cada usuário. De maneira complementar, na fase 5 pedimos aos usuários para marcarem, dentre os filmes que são novos, aqueles que eles gostariam de assistir. Assim sendo, gera-se um *ranking* para cada estratégia esboçando a serendipidade de cada recomendador. Ambas essas fases de análise estão sumarizadas na Tabela 6.4, que indica a novidade e serendipidade de cada recomendador. Como consequência de apresentar filmes mais diversos que as demais estratégias, *Max-Coverage* alcança os maiores índices de novidade e serendipidade. Segundo os usuários, os filmes recomendados por essa estratégia são desconhecidos, mas também atraentes. Todavia, a estratégia de *Niche-Coverage* e as tradicionais também possuem um bom desempenho com relação a essas métricas. Devido aos altos valores de desvio padrão, não foram encontradas diferenças significativas entre as abordagens avaliadas, ressaltando a relevância de todas as estratégias.

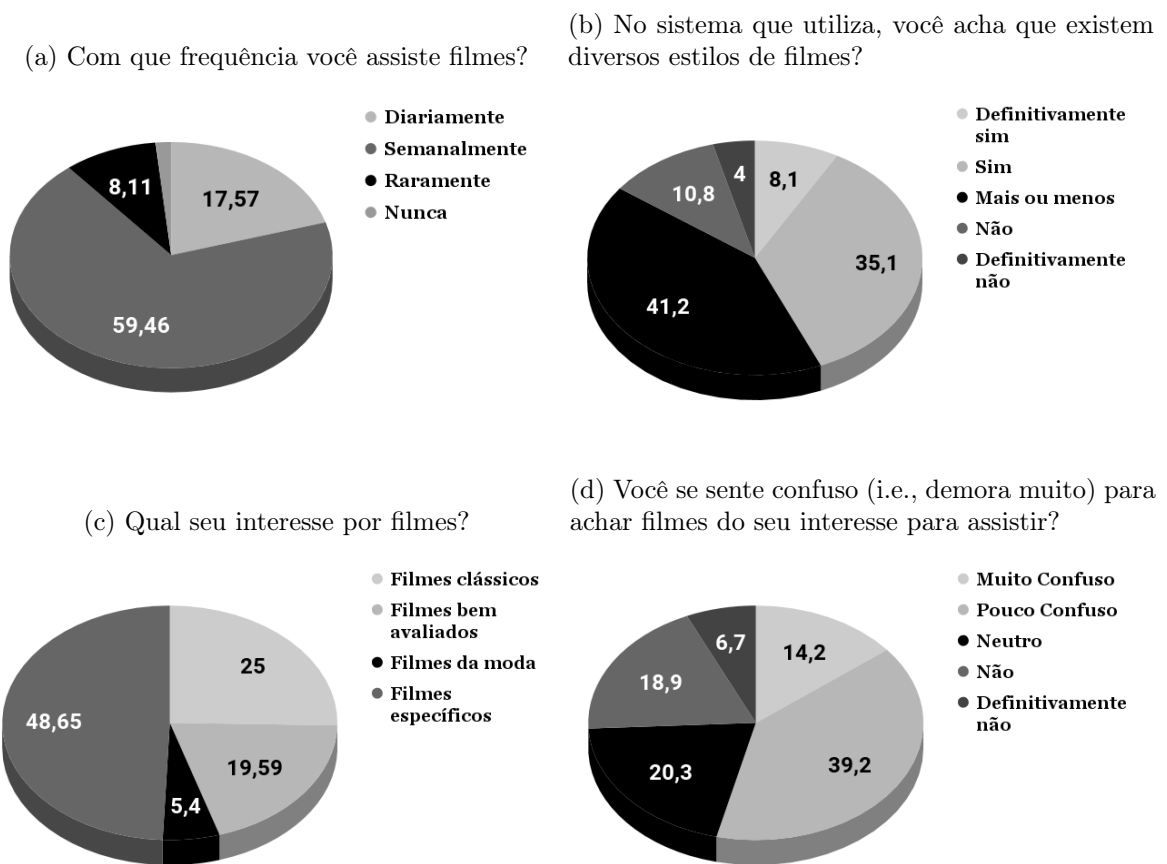
Tabela 6.4: Resultados das fases 4 e 5, especificando cada recomendador utilizado.

			Itens Selecionados		
		Recomendadores	AUC	Média	Desvio Padrão
Novidade		<i>Popularity</i>	0,30075	1,21	0,93
		<i>Best-Rated</i>	0,25789	1,30	1,06
		<i>Max-Coverage</i>	0,39323	1,98	1,04
		<i>Niche-Coverage</i>	0,20784	1,48	1,12
Serendipidade		<i>Popularity</i>	0,34756	0,70	0,59
		<i>Best-Rated</i>	0,24593	0,75	0,61
		<i>Max-Coverage</i>	0,33537	1,02	0,54
		<i>Niche-Coverage</i>	0,27439	0,84	0,68

Na fase 6 foi pedido aos usuários para deixarem algumas informações pessoais como idade e sexo, bem como suas opiniões pessoais sobre os recomendadores que utilizam diariamente, como *Netflix*, *Spotify* e outros. Com isso, descobrimos que 66,88% dos usuários são homens e 33,22% são mulheres, na faixa etária de 17 a 67 anos, com uma média de 27,41 anos. As demais avaliações estão sumarizadas na Figura 6.6, que apresenta a porcentagem de usuários que escolheram cada uma das alternativas. Vale destacar que essas porcentagens nem sempre somam 100% pois existem usuários que

deixaram essas questões em branco, sem marcar nenhuma resposta. Na Figura 6.6(a), vemos que a maioria dos usuários que participaram dessa pesquisa estavam contextualizados com o ambiente, pois gostam muito de filmes, tendo o hábito de assistir semanalmente ou mesmo diariamente. Entretanto, mesmo com os esforços para personalizar as recomendações, serviços como *Netflix* acabam apresentando centenas de filmes e confundindo o usuário, como mostram as Figuras 6.6(b) e 6.6(d). A maioria dos usuários ainda encontram dificuldades para selecionar o filme ideal para assistir.

Figura 6.6: Respostas dos usuários para o questionário pessoal da fase 6.



Por sua vez, o resultado exibido na Figura 6.6(c) corrobora nossa hipótese 1, mostrando que nem sempre os usuários vão ficar satisfeitos apenas com os itens recomendados pelas estratégias tradicionais. Com essa pesquisa vemos que 48,65% dos usuários preferem filmes peculiares, específicos às suas preferências. Apenas 25% e 19,59% dos usuários disseram gostar de filmes clássicos, que todo mundo gosta, ou de filmes aclamados pela crítica. Por essa razão e pelos resultados apresentados ao longo deste capítulo, vemos a necessidade de incorporar novas estratégias de recomendação às tradicionais em sistemas reais.

6.4 Síntese do Capítulo

Neste capítulo, avaliamos a última hipótese deste trabalho, que sugere a utilização de estratégias complementares para atenuar o problema de *Pure Cold-Start*. Com uma análise *offline*, vimos que a combinação de $\langle \textit{Best-Rated}, \textit{Max-Coverage}, \textit{Niche-Coverage} \rangle$ atrai mais usuários novos do que as demais estratégias. Por essa razão, consolidamos uma metodologia de avaliação que visa comparar essa combinação de estratégias proposta com o modelo tradicionalmente utilizado. Para tal, primeiramente selecionamos um conjunto de dados de filmes que foi atualizado pela última vez no ano de 2017. Apesar de existirem filmes recentes, foi necessário realizar um corte temporal visando selecionar apenas as informações após o ano de 2012, no intuito de refletir na popularidade real dos itens deste ano.

Em seguida, construímos uma interface de avaliação *Web* que se assemelha aos sistemas comumente utilizados por usuários novos. Essa interface nos permitiu realizar avaliações conjuntas e individuais sobre as estratégias. Com a avaliação conjunta é possível observar que novos SsR podem ser incorporados para lidar com esse cenário de usuários novos. Em geral, o cenário proposto por este trabalho permitiu que mais usuários encontrassem pelo menos um filme para assistir, alcançando a satisfação de 63,96% dos usuários que realizaram a pesquisa. De fato, as estratégias que são complementares em cada cenário (*Recent Items*, no cenário A, e *Max-Coverage* ou *Niche-Coverage*, no cenário B) alcançaram um alto interesse dos usuários, mostrando serem relevantes. Apesar de ser um experimento limitado, esses resultados apontam para a confirmação da terceira hipótese, sugerindo aos proprietários de sistemas *Web* alinhar recomendadores complementares para maximizar o interesse dos usuários. Sobretudo, com as análises individuais, os usuários mostraram que as listas de *Max-Coverage* e *Niche-Coverage* são igualmente relevantes aos modelos estado-da-arte. Por essas razões, este trabalho sugere fortemente a utilização de tais abordagens em cenários reais de aplicação.

Capítulo 7

Conclusão & Trabalhos Futuros

Este trabalho teve por objetivo melhorar as recomendações geradas para usuários novos, no intuito de atenuar o problema do *Pure Cold-Start*. Para alcançar esse objetivo, inicialmente, definimos o problema de *Pure Cold-Start*, definindo-o como um subproblema do *Cold-Start*. Basicamente, abordamos o *Cold-Start* como o problema de gerar recomendações para os usuários que possuem poucas informações associadas. Por outro lado, o *Pure Cold-Start* trata-se da versão primária do problema, onde é necessário gerar recomendações para os usuários que não possuem nenhuma informação prévia associada ao sistema. Além disso, nesse Capítulo 2, vimos que nenhuma das classes de estratégias personalizadas podem ser utilizadas nesse cenário, devido à ausência de informação a respeito dos usuários.

Por essa razão, no Capítulo 3, analisamos as principais abordagens utilizadas atualmente para lidar com o problema de *Pure Cold-Start*. Nessa etapa investigamos três principais categorias de estratégias, definidas como: (1) SsR Interativos; (2) SsR Híbridos; e (3) SsR não-Personalizados. Em geral, as categorias (1) e (2) propõem utilizar informações externas relacionadas aos usuários, como sexo, faixa etária, localização e outras. Por sua vez, a categoria (3) propõe utilizar as informações globais do sistema para recomendar a usuários novos, ignorando o perfil do usuário alvo. Devido à simplicidade, eficiência e independência de domínio, vimos que as estratégias não-personalizadas são as mais utilizadas nesse cenário. Entre elas, destacam-se as estratégias que exploram informações dos itens, como a popularidade, avaliações recebidas ou mesmo o momento de consumo.

Apesar dessas estratégias explorarem informações distintas, todas são utilizadas sob a mesma premissa: *itens enviados pela popularidade, recência e avaliações positivas tem o potencial de satisfazer os interesses de grande parte dos usuários novos*. Contudo, não necessariamente toda a população de usuários prefere esses itens

mais populares. Dessa forma, levantamos nossa primeira hipótese de que existe uma parcela significativa de usuários que não estão interessados apenas nos itens populares. De fato, no estudo de caso realizado no Capítulo 4, comprovamos essa hipótese sobre os principais domínios de recomendação. Avaliamos quatro cenários, relacionados as aplicações de *e-commerce* e entretenimento, e concluimos que existem: (1) usuários que preferem itens populares; (2) usuários que preferem itens não-populares; e (3) usuários que mesclam por itens populares e não-populares. Além disso, os resultados também mostram a existência de diversos nichos de usuários nos domínios, apontando para a existência de outras preferências intrínsecas no perfil de cada usuário. Com a comprovação da primeira hipótese, nesse mesmo Capítulo 4, propõe-se dois novos SsR não-personalizados que visam explorar as premissas destacadas pelo estudo de caso. Ambas as abordagens exploram a cobertura de usuários como ponto chave para alcançar os outros itens da preferência dos usuários.

Com a proposta de novos SsR não-personalizados, tornou-se possível propor uma segunda hipótese: *SsR não-personalizados que exploram outros itens além dos populares são capazes de atenuar o problema de Pure Cold-Start*. Dessa hipótese, três questões pertinentes foram levantadas e estudadas ao longo do Capítulo 5. Para tal, definimos uma metodologia de avaliação em três dimensões de análise que visam responder cada uma das perguntas destacadas. Como base de comparação, selecionamos os SsR não-personalizados comumente utilizados na literatura: *Popularity*, *Best-Rated* e *Recent Items*. Para simular o cenário do *Pure Cold-Start*, extraímos 20% dos últimos usuários de cada um dos domínios e removemos todas as informações associadas a esses. Na primeira dimensão de análise, verificamos a similaridade e a popularidade dos itens recomendados por cada estratégia. Em suma, os resultados mostram que as novas estratégias são complementares às tradicionalmente utilizadas pois recomendam itens distintos. Além disso, observamos que os novos SsR recomendam itens que não são classificados como populares, consolidando as premissas assumidas por cada estratégia.

Por sua vez, a segunda dimensão de análise avalia a qualidade das recomendações emitidas sobre os principais critérios de utilidade e diversidade. Os resultados mostram que os itens distintos apresentados por esses novos SsR acabam adicionando diversidade às estratégias. Além disso, também foi possível observar ganhos estatisticamente significantes em relação à acurácia das estratégias tradicionais. Por fim, com a terceira dimensão de análise avaliamos quais usuários são atraídos por cada estratégia de recomendação. Nessas análises vemos que a estratégia de *Max-Coverage* proposta é capaz de atrair mais usuários novos. Ademais, vemos que esses itens distintos permitem às novas estratégias alcançarem subconjuntos complementares de usuários, sendo possível pensar em um aumento da taxa de conversão de usuários novos.

Devido a esse comportamento complementar, esses resultados mostram que os SsR avaliados podem ser aplicados em conjunto nos cenários reais. Assim, foi possível consolidarmos nossa terceira hipótese no Capítulo 6: *Páginas de produtos compostas por SsR não-personalizados complementares satisfazem o interesse de um grande número de usuários novos distintos*. Basicamente, propomos que, para aumentar a taxa de conversão, os proprietários dos sistemas *Web* devem apresentar na sua *interface* de visualização itens das estratégias *Best-Rated*, *Max-Coverage* e *Niche-Coverage*. Para comprovarmos essa hipótese, construímos uma metodologia que visa avaliar os SsR apresentados em conjunto e individualmente. Dessa forma, selecionamos um conjunto de dados atualizado em 2017 e extraímos informações relativas aos últimos cinco anos, sob a justificativa de apresentar aos usuários filmes recentemente assistidos. Em seguida, construímos uma *interface* similar às que os usuários estão acostumados a lidar e realizamos um experimento *online* com 236 usuários.

Com a avaliação conjunta é possível observar o cenário proposto por este trabalho permitiu que mais usuários encontrassem pelo menos um filme para assistir, alcançando a satisfação de 63,96% dos usuários que realizaram a pesquisa. Em geral, 96% dos usuários que interagiram com o cenário proposto encontraram pelo menos um filme para assistirem. Além disso, os resultados apontam a relevância da estratégia de *Recent Items* na prática, por ser capaz de diversificar mais os resultados. Apesar de ser um experimento limitado, esses resultados reforçam a terceira hipótese, mostrando que SsR complementares satisfazem mais usuários novos distintos. Sobretudo, com a avaliação individual, os usuários mostraram que as listas de *Max-Coverage* e *Niche-Coverage* são igualmente relevantes aos modelos estado-da-arte.

Por todas essas conclusões, vemos que a forma para melhorar as recomendações geradas para usuários novos consiste em combinar estratégias complementares. Além disso, comprovamos a eficácia prática das estratégias de *Max-Coverage* e *Niche-Coverage* para usuários reais. Dessa forma, aconselhamos aos proprietários de aplicações de entretenimento ou *e-commerce* rever as estratégias utilizadas nessa primeira interação do usuário com o sistema. Por esse mesmo motivo, pretendemos como trabalhos futuros avaliar o desempenho dessas estratégias em sistemas reais de aplicação, tal como sugerido por nosso experimento *online*. Além disso, sabendo que os usuários existentes possuem esses comportamentos por terem iniciado suas interações com filmes populares, alterar o ponto de partida de cada usuário abre novas questões de pesquisa a serem respondidas, tais como: (1) *Como essas recomendações iniciais impactam no comportamento dos usuários?* (2) *Qual é o impacto dessa alteração na popularidade dos itens do domínio?* e (3) *Como serão a performance das estratégias atualmente consideradas estado-da-arte sobre esse novo cenário?*

Referências Bibliográficas

- Abbassi, Z. & Mirrokni, V. S. (2007). A recommender system based on local random walks and spectral methods. Em *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pp. 102--108. ACM.
- Adomavicius, G. & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734--749.
- Aha, D. W. & Breslow, L. A. (1997). Refining conversational case libraries. Em *International Conference on Case-Based Reasoning*, pp. 267--278. Springer.
- Aharon, M.; Anava, O.; Avigdor-Elgrabli, N.; Drachler-Cohen, D.; Golan, S. & Somekh, O. (2015). Excuseme: Asking users to help in item cold-start recommendations. Em *Proceedings of the 9th ACM Conference on Recommender Systems*, pp. 83--90. ACM.
- Akshita & Smita (2013). Recommender system: Review. *International Journal of Computer Applications*.
- Alon, N.; Awerbuch, B. & Azar, Y. (2003). The online set cover problem. Em *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pp. 100--105. ACM.
- Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. Hyperion.
- Ardissono, L.; Felfernig, A.; Friedrich, G.; Goy, A.; Jannach, D.; Petrone, G.; Schaffer, R. & Zanker, M. (2003). A framework for the development of personalized, distributed web-based configuration systems. *Ai Magazine*, 24(3):93.
- Baeza-Yates, R.; Ribeiro-Neto, B. et al. (1999). *Modern information retrieval*, volume 463. ACM press New York.

- Balabanović, M. & Shohom, Y. (1997). Content-based, collaborative recommendation. *Communications of the ACM*, 40(3).
- Barjasteh, I.; Forsati, R.; Masrour, F.; Esfahanian, A.-H. & Radha, H. (2015). Cold-start item and user recommendation with decoupled completion and transduction. Em *Proceedings of the 9th ACM Conference on Recommender Systems*, pp. 91--98. ACM.
- Basiri, J.; Shakery, A.; Moshiri, B. & Hayat, M. Z. (2010). Alleviating the cold-start problem of recommender systems using a new hybrid approach. Em *Telecommunications (IST), 2010 5th International Symposium on*, pp. 962--967. IEEE.
- Basu, C.; Hirsh, H.; Cohen, W. et al. (1998). Recommendation as classification: Using social and content-based information in recommendation. Em *Aaai/iaai*, pp. 714--720.
- Beel, J.; Gipp, B.; Langer, S. & Breitinger, C. (2016). paper recommender systems: a literature survey. *International Journal on Digital Libraries*, 17(4):305--338.
- Bell, R. M. & Koren, Y. (2007). Lessons from the netflix prize challenge. *ACM SIGKDD Explorations Newsletter*, 9(2):75--79.
- Bernardi, L.; Kamps, J.; Kiseleva, J. & Mueller, M. J. (2015). The continuous cold start problem in e-commerce recommender systems. *arXiv preprint arXiv:1508.01177*.
- Billsus, D. & Pazzani, M. J. (1998). Learning collaborative information filters. Em *Icml*, volume 98, pp. 46--54.
- Billsus, D. & Pazzani, M. J. (2000). User modeling for adaptive news access. *User modeling and user-adapted interaction*, 10(2-3):147--180.
- Billsus, D.; Pazzani, M. J. & Chen, J. (2000). A learning agent for wireless news access. Em *Proceedings of the 5th international conference on Intelligent user interfaces*, pp. 33--36. ACM.
- Bobadilla, J.; Ortega, F.; Hernando, A. & Bernal, J. (2012). A collaborative filtering approach to mitigate the new user cold start problem. *Knowledge-Based Systems*, 26:225--238.
- Bobadilla, J.; Ortega, F.; Hernando, A. & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46:109--132.

- Boehmer, J.; Jung, Y. & Wash, R. (2015). e-commerce recommender systems. *The International Encyclopedia of Digital Communication and Society*.
- Buchbinder, N.; Naor, J. S. et al. (2009). The design of competitive online algorithms via a primal–dual approach. *Foundations and Trends® in Theoretical Computer Science*, 3(2–3):93--263.
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331--370.
- Campos, P. G.; Bellogín, A.; Díez, F. & Chavarriaga, J. E. (2010). Simple time-biased knn-based recommendations. Em *Proceedings of the Workshop on Context-Aware Movie Recommendation*, pp. 20--23. ACM.
- Candillier, L.; Jack, K.; Fessant, F. & Meyer, F. (2009). State-of-the-art recommender systems. *Collaborative and Social Information Retrieval and Access Techniques for Improved User Modeling*.
- Carrer-Neto, W.; Hernández-Alcaraz, M. L.; Valencia-García, R. & García-Sánchez, F. (2012). Social knowledge-based recommender system. application to the movies domain. *Expert Systems with applications*, 39(12):10990--11000.
- Chikhaoui, B.; Chiazzaro, M. & Wang, S. (2011). An improved hybrid recommender system by combining predictions. Em *Advanced Information Networking and Applications (WAINA), 2011 IEEE Workshops of International Conference on*, pp. 644--649. IEEE.
- Choi, K.; Yoo, D.; Kim, G. & Suh, Y. (2012). A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis. *Electronic Commerce Research and Applications*, 11(4):309--317.
- Chvatal, V. (1979). A greedy heuristic for the set-covering problem. *Mathematics of operations research*, 4(3):233--235.
- Cohen, W.; Prieditis, A. & Russell, S. (1995). Proceedings of the twelfth international conference on machine learning.
- De Campos, L. M.; Fernández-Luna, J. M.; Huete, J. F. & Rueda-Morales, M. A. (2010). Combining content-based and collaborative recommendations: A hybrid approach based on bayesian networks. *International Journal of Approximate Reasoning*, 51(7):785--799.

- De Gemmis, M.; Lops, P.; Semeraro, G. & Basile, P. (2008). Integrating tags in a semantic content-based recommender. Em *Proceedings of the 2008 ACM conference on Recommender systems*, pp. 163--170. ACM.
- Deshpande, M. & Karypis, G. (2004). Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems (TOIS)*, 22(1):143--177.
- Dong, W.; Moses, C. & Li, K. (2011). Efficient k-nearest neighbor graph construction for generic similarity measures. Em *Proceedings of the 20th international conference on World wide web*, pp. 577--586. ACM.
- Duarte Torres, S.; Hiemstra, D.; Weber, I. & Serdyukov, P. (2012). Query recommendation for children. Em *Proceedings of the 21st ACM international conference on Information and knowledge management*, pp. 2010--2014. ACM.
- Fujimura, K. & Zhu, Y. (2008). Pose estimation based on critical point analysis. US Patent 7,317,836.
- Garcia, R. & Amatriain, X. (2010). Weighted content based methods for recommending connections in online social networks. Em *Workshop on Recommender Systems and the Social Web*, pp. 68--71.
- Ghazanfar, M. A. & Prügél-Bennett, A. (2014). Leveraging clustering approaches to solve the gray-sheep users problem in recommender systems. *Expert Systems with Applications*, 41(7):3261--3275.
- Grčar, M.; Fortuna, B.; Mladenič, D. & Grobelnik, M. (2006). knn versus svm in the collaborative filtering framework. *Data Science and Classification*, pp. 251--260.
- Hammar, M.; Karlsson, R. & Nilsson, B. J. (2013). Using maximum coverage to optimize recommendation systems in e-commerce. Em *Proceedings of the 7th ACM conference on Recommender systems*, pp. 265--272. ACM.
- Hammond, M. I.; Browne, K. M.; Estee, M. & Kliman-Silver, C. (2017). Multiple user interfaces of an artificial intelligence system to accommodate different types of users solving different types of problems with artificial intelligence. US Patent App. 15/417,033.
- He, C.; Parra, D. & Verbert, K. (2016). Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications*, 56:9--27.

- Herlocker, J. L.; Konstan, J. A.; Terveen, L. G. & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1):5--53.
- Hernando, A.; Bobadilla, J.; Ortega, F. & Gutiérrez, A. (2017). A probabilistic model for recommending to new cold-start non-registered users. *Information Sciences*, 376:216--232.
- Hinz, J. D. O. & Eckert, D.-K. J. (2010). The impact of search and recommendation systems on sales in electronic commerce. *Business & Information Systems Engineering*, 2(2):67--77.
- Ho, Y.-C.; Chiang, Y.-T. & Hsu, J. Y.-J. (2014). Who likes it more?: mining worth-recommending items from long tails by modeling relative preference. Em *Proceedings of the 7th ACM international conference on Web search and data mining*, pp. 253--262. ACM.
- Hochbaum, D. S. (1996). *Approximation algorithms for NP-hard problems*. PWS Publishing Co.
- Hochbaum, D. S. & Pathria, A. (1998). Analysis of the greedy approach in problems of maximum k-coverage. *Naval Research Logistics*, 45(6):615--627.
- Horvitz, E.; Breese, J.; Heckerman, D.; Hovel, D. & Rommelse, K. (1998). The lumiere project: Bayesian user modeling for inferring the goals and needs of software users. Em *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pp. 256--265. Morgan Kaufmann Publishers Inc.
- Hu, R. & Pu, P. (2010). Using personality information in collaborative filtering for new users. *Recommender Systems and the Social Web*, 17.
- Huang, Z.; Chung, W.; Ong, T.-H. & Chen, H. (2002). A graph-based recommender system for digital library. Em *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pp. 65--73. ACM.
- Hussein, T.; Linder, T.; Gaulke, W. & Ziegler, J. (2014). Hybreed: A software framework for developing context-aware hybrid recommender systems. *User Modeling and User-Adapted Interaction*, 24(1-2):121--174.
- Jannach, D.; Zanker, M.; Felfernig, A. & Friedrich, G. (2010). *Recommender systems: an introduction*. Cambridge University Press.

- Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *Machine learning: ECML-98*, pp. 137--142.
- Kawai, M. & Nogami, S. (2016). A hybrid recommender system of collaborative and content based filtering. *International Information Institute (Tokyo). Information*, 19(6B):2177.
- Khrouf, H. & Troncy, R. (2013). Hybrid event recommendation using linked data and user diversity. Em *Proceedings of the 7th ACM conference on Recommender systems*, pp. 185--192. ACM.
- Kim, H.-N.; Alkhalidi, A.; El Saddik, A. & Jo, G.-S. (2011). Collaborative user modeling with user-generated tags for social recommender systems. *Expert Systems with Applications*, 38(7):8488--8496.
- Kim, J. W.; Lee, B. H.; Shaw, M. J.; Chang, H.-L. & Nelson, M. (2001). Application of decision-tree induction techniques to personalized advertisements on internet storefronts. *International Journal of Electronic Commerce*, 5(3):45--62.
- Klemm, R. (2016). A platform approach to personalizing web applications with social network user profiles. Em *New Advances in Information Systems and Technologies*, pp. 807--816. Springer.
- Kouki, P.; Fakhraei, S.; Foulds, J.; Eirinaki, M. & Getoor, L. (2015). Hyper: A flexible and extensible probabilistic framework for hybrid recommender systems. Em *Proceedings of the 9th ACM Conference on Recommender Systems*, pp. 99--106. ACM.
- Lacerda, A. & Ziviani, N. (2013). Building user profiles to improve user experience in recommender systems. Em *Proceedings of the sixth ACM international conference on Web search and data mining*, pp. 759--764. ACM.
- Li, Q. & Kim, B. M. (2003). Clustering approach for hybrid recommender system. Em *Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on*, pp. 33--38. IEEE.
- Lika, B.; Kolomvatsos, K. & Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert Systems with Applications*, 41(4):2065--2073.
- Linden, G.; Smith, B. & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 7(1):76--80.

- Lo, S. & Lin, C. (2006). Wmr—a graph-based algorithm for friend recommendation. Em *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 121--128. IEEE Computer Society.
- Lops, P.; De Gemmis, M. & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. Em *Recommender systems handbook*, pp. 73--105. Springer.
- Mahmood, T. & Ricci, F. (2007). Learning and adaptivity in interactive recommender systems. Em *Proceedings of the ninth international conference on Electronic commerce*, pp. 75--84. ACM.
- Martin, R.; Walid, M.; Robert, W. & Thomas, Z. (2014). *Recommendation Systems in Software Engineering*. Springer.
- Mavlanova, T.; Benbunan-Fich, R. & Lang, G. (2016). The role of external and internal signals in e-commerce. *Decision Support Systems*, 87:59--68.
- McDonald, K. & Smith-Rowsey, D. (2016). *The Netflix effect: Technology and entertainment in the 21st century*. Bloomsbury Publishing USA.
- McNee, S. M.; Riedl, J. & Konstan, J. A. (2006). Being accurate is not enough: how accuracy metrics have hurt recommender systems. Em *CHI'06 extended abstracts on Human factors in computing systems*, pp. 1097--1101. ACM.
- Melville, P.; Mooney, R. J. & Nagarajan, R. (2002). Content-boosted collaborative filtering for improved recommendations. Em *Aaai/iaai*, pp. 187--192.
- Mobasher, B.; Burke, R.; Bhaumik, R. & Williams, C. (2005). Effective attack models for shilling item-based collaborative filtering systems. Em *Proceedings of the 2005 WebKDD Workshop, held in conjunction with ACM SIGKDD*, volume 2005.
- Mooney, R. J. & Roy, L. (2000). Content-based book recommending using learning for text categorization. Em *Proceedings of the fifth ACM conference on Digital libraries*, pp. 195--204. ACM.
- Mourão, F. H. d. J. (2014). A hybrid recommendation method that combines forgotten items and non-content attributes.
- Nguyen, A.-T.; Denos, N. & Berrut, C. (2007). Improving new user recommendations with rule-based induction on cold user data. Em *Proceedings of the 2007 ACM conference on Recommender systems*, pp. 121--128. ACM.

- Park, S.-T. & Chu, W. (2009). Pairwise preference regression for cold-start recommendation. Em *Proceedings of the third ACM conference on Recommender systems*, pp. 21--28. ACM.
- Park, Y.-J. & Tuzhilin, A. (2008). The long tail of recommender systems and how to leverage it. Em *Proceedings of the 2008 ACM conference on Recommender systems*, pp. 11--18. ACM.
- Pathak, B.; Garfinkel, R.; Gopal, R. D.; Venkatesan, R. & Yin, F. (2010). Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems*, 27(2):159--188.
- Pazzani, M. J. (1999). A framework for collaborative, content-based and demographic filtering. *Artificial intelligence review*, 13(5-6):393--408.
- Pazzani, M. J. & Billsus, D. (2007). Content-based recommendation systems. Em *The adaptive web*, pp. 325--341. Springer.
- Pereira, A. L. V. & Hruschka, E. R. (2015). Simultaneous co-clustering and learning to address the cold start problem in recommender systems. *Knowledge-Based Systems*, 82:11--19.
- Popescul, A.; Pennock, D. M. & Lawrence, S. (2001). Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. Em *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pp. 437--444. Morgan Kaufmann Publishers Inc.
- Poriya, A.; Bhagat, T.; Patel, N. & Sharma, R. (2014). Non-personalized recommender systems and user-based collaborative recommender systems. *Int. J. Appl. Inf. Syst*, 6(9):22--27.
- Portal, T. S. (2017). *Number of Netflix streaming subscribers worldwide from 3rd quarter 2011 to 2nd quarter 2017*.
- Puthiya Parambath, S. A.; Usunier, N. & Grandvalet, Y. (2016). A coverage-based approach to recommendation diversity on similarity graph. Em *Proceedings of the 10th ACM Conference on Recommender Systems*, pp. 15--22. ACM.
- Ramos, E. (2015). *E-commerce*. Editora FGV.
- Ramos, J. et al. (2003). Using tf-idf to determine word relevance in document queries. Em *Proceedings of the first instructional conference on machine learning*, volume 242, pp. 133--142.

- Reilly, J.; McCarthy, K.; McGinty, L. & Smyth, B. (2005). Incremental critiquing. *Knowledge-Based Systems*, 18(4):143--151.
- Reis, T. (2016). *5 dicas para atrair novos clientes com Inbound Commerce*.
- Ricci, F.; Rokach, L. & Shapira, B. (2011). *Introduction to recommender systems handbook*. Springer.
- Ricci, F.; Rokach, L. & Shapira, B. (2015). Recommender systems: introduction and challenges. Em *Recommender Systems Handbook*, pp. 1--34. Springer.
- Rosli, A. N.; You, T.; Ha, I.; Chung, K.-Y. & Jo, G.-S. (2015). Alleviating the cold-start problem by incorporating movies facebook pages. *Cluster Computing*, 18(1):187--197.
- Safoury, L. & Salah, A. (2013). Exploiting user demographic attributes for solving cold-start problem in recommender system. *Lecture Notes on Software Engineering*, 1(3):303.
- Sarwar, B.; Karypis, G.; Konstan, J. & Riedl, J. (2000). Application of dimensionality reduction in recommender system-a case study. Relatório técnico, DTIC Document.
- Saveski, M. & Mantrach, A. (2014). Item cold-start recommendations: learning local collective embeddings. Em *Proceedings of the 8th ACM Conference on Recommender systems*, pp. 89--96. ACM.
- Schafer, J. B.; Frankowski, D.; Herlocker, J. & Sen, S. (2007). Collaborative filtering recommender systems. Em *The adaptive web*, pp. 291--324. Springer.
- Schafer, J. B.; Konstan, J. & Riedl, J. (1999). Recommender systems in e-commerce. Em *Proceedings of the 1st ACM conference on Electronic commerce*, pp. 158--166. ACM.
- Schein, A. I.; Popescul, A.; Ungar, L. H. & Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. Em *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 253--260. ACM.
- Schwartz, B. (2009). *The Paradox of Choice: Why More Is Less, Revised Edition*. HarperCollins. ISBN 9780061748998.
- Sedhain, S.; Sanner, S.; Braziunas, D.; Xie, L. & Christensen, J. (2014). Social collaborative filtering for cold-start recommendations. Em *Proceedings of the 8th ACM Conference on Recommender systems*, pp. 345--348. ACM.

- Shah, J. & Sahu, L. (2014). A survey of various hybrid based recommendation method. *International Journal of Advanced Research in Computer Science and Software Engineering*.
- Shani, G. & Gunawardana, A. (2011). Evaluating recommendation systems. *Recommender systems handbook*, pp. 257--297.
- Shi, S. (2016). Real-time job recommendation engine based on college graduates' persona. *Journal of Residuals Science & Technology*, 13(7).
- Shneiderman, B. (2010). *Designing the user interface: strategies for effective human-computer interaction*. Pearson Education India.
- Silva, N. B.; Tsang, R.; Cavalcanti, G. D. & Tsang, J. (2010). A graph-based friend recommendation system using genetic algorithm. Em *Evolutionary Computation (CEC), 2010 IEEE Congress on*, pp. 1--7. IEEE.
- Steck, H.; van Zwol, R. & Johnson, C. (2015). Interactive recommender systems: Tutorial. Em *Proceedings of the 9th ACM Conference on Recommender Systems*, pp. 359--360. ACM.
- Sun, D.; Luo, Z. & Zhang, F. (2011). A novel approach for collaborative filtering to alleviate the new item cold-start problem. Em *Communications and Information Technologies (ISCIT), 2011 11th International Symposium on*, pp. 402--406. IEEE.
- Thompson, C. A.; Goker, M. H. & Langley, P. (2004). A personalized system for conversational recommendations. *Journal of Artificial Intelligence Research*, 21:393-428.
- Van Meteren, R. & Van Someren, M. (2000). Using content-based filtering for recommendation. Em *Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000 Workshop*, pp. 47--56.
- Vargas, S. & Castells, P. (2011). Rank and relevance in novelty and diversity metrics for recommender systems. Em *Proceedings of the fifth ACM conference on Recommender systems*, pp. 109--116. ACM.
- Wongchokprasitti, C.; Peltonen, J.; Ruotsalo, T.; Bandyopadhyay, P.; Jacucci, G. & Brusilovsky, P. (2015). User model in a box: Cross-system user model transfer for resolving cold start problems. Em *International Conference on User Modeling, Adaptation, and Personalization*, pp. 289--301. Springer.

- Yang, X.; Guo, Y.; Liu, Y. & Steck, H. (2014). A survey of collaborative filtering based social recommender systems. *Computer Communications*, 41:1--10.
- Yin, H.; Cui, B.; Li, J.; Yao, J. & Chen, C. (2012). Challenging the long tail recommendation. *Proceedings of the VLDB Endowment*, 5(9):896--907.
- Zhang, M. (2009). Enhancing diversity in top-n recommendation. Em *Proceedings of the third ACM conference on Recommender systems*, pp. 397--400. ACM.
- Zhang, Y. C.; Séaghdha, D. Ó.; Quercia, D. & Jambor, T. (2012). Auralist: introducing serendipity into music recommendation. Em *Proceedings of the fifth ACM international conference on Web search and data mining*, pp. 13--22. ACM.
- Zheng, L.; Noroozi, V. & Yu, P. S. (2017). Joint deep modeling of users and items using reviews for recommendation. Em *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 425--434. ACM.
- Zhou, T.; Kuscsik, Z.; Liu, J.-G.; Medo, M.; Wakeling, J. R. & Zhang, Y.-C. (2010). Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, 107(10):4511--4515.

Apêndice A

Questionário 1

1. Os filmes apresentados correspondem ao meu interesse por filmes.
 concordo totalmente concordo não sei discordo discordo totalmente
2. Os filmes recomendados para mim são famosos e bem conhecidos.
 concordo totalmente concordo não sei discordo discordo totalmente
3. O sistema apresenta diversos gêneros de filmes distintos (i.e., comédia, ação, romance).
 concordo totalmente concordo não sei discordo discordo totalmente
4. Essas recomendações me surpreenderam, pois eu não conhecia ou nunca tinha ouvido falar de vários dos filmes apresentados.
 concordo totalmente concordo não sei discordo discordo totalmente
5. De maneira geral, eu estou satisfeito com as recomendações recebidas pois consegui encontrar um filme interessante para assistir.
 concordo totalmente concordo não sei discordo discordo totalmente
6. Eu utilizaria novamente essas listas de filmes para me auxiliar.
 concordo totalmente concordo não sei discordo discordo totalmente

Apêndice B

Questionário 2

1. Idade
2. Sexo
3. Com que frequência você assiste filmes ou séries?
 diariamente semanalmente uma vez ao mês raramente nunca
4. Qual a sua preferência por filmes?
 gosto de filmes clássicos, que todo mundo gosta
 gosto de filmes aclamados pela crítica
 gosto de filmes que estão na moda
 gosto de filmes específicos, peculiares
 não gosto de filmes
5. Você é assinante de algum sistema de entretenimento, como Netflix, Spotify, etc que recomenda itens para o seu consumo?
 sim não
6. Você se sente confuso (i.e., demora muito) para encontrar os itens de seu interesse no(s) sistema(s) que utiliza?
 muito confuso um pouco confuso neutro não definitivamente não
7. Você acha que os itens apresentados pelo(s) sistema(s) que utiliza são diversificados (i.e., apresentam muitos itens variados)?
 definitivamente sim sim mais ou menos não definitivamente não

Apêndice C

Termo de Compromisso

I - INFORMAÇÕES GERAIS:

- **Título do trabalho experimental:** Pesquisa do comportamento dos usuários em Sistemas de Recomendação.
- **Pesquisador(es) responsável(is):** Nícollas Silva, Diego Carvalho, Adriano C. M. Pereira, Fernando Mourão, Leonardo Rocha.
- **Instituição/Departamento:** Universidade Federal de Minas Gerais (UFMG).
- **Local da coleta de dados:** Interface web.

Prezado(a) Senhor(a):

- Você está sendo convidado(a) a participar da pesquisa de forma totalmente voluntária.
- Antes de concordar em participar desta pesquisa, é muito importante que você compreenda as informações e instruções contidas neste documento.
- Os pesquisadores deverão responder todas as suas dúvidas antes que você se decida a participar.
- Para participar deste estudo você não terá nenhum custo, nem receberá qualquer vantagem financeira.
- Você tem o direito de desistir de participar da pesquisa a qualquer momento, sem nenhuma penalidade e sem perder os benefícios aos quais tenha direito, não acarretando qualquer penalidade ou modificação na forma em que é atendido pelo pesquisador.

As informações contidas neste termo visam firmar acordo por escrito, mediante o qual o entrevistado autoriza sua participação, com pleno conhecimento da natureza dos procedimentos e riscos a que se submeterá, com capacidade de livre arbítrio e sem qualquer coação.

II - OBJETIVOS

Esta pesquisa visa avaliar a interação do usuário com diversas listas de recomendação apresentadas por algoritmos determinísticos e não-enviesados.

III - JUSTIFICATIVA

Esta pesquisa online é realizada para corroborar as hipóteses levantadas no trabalho em estudo, dando um feedback das preferências dos usuários na prática.

IV - PROCEDIMENTOS DO EXPERIMENTO

AMOSTRA- É utilizada uma base de filmes recente da MovieLens, na qual são extraídas as principais informações sobre os filmes.

EXAMES- Serão realizadas perguntas e respostas referentes aos filmes apresentados aos usuários entrevistados.

V - RISCOS ESPERADOS

Não existem riscos.

VI - BENEFÍCIOS

O usuário poderá se distrair com a plataforma e conhecer um pouco mais sobre recomendadores.

VII - CRITÉRIOS PARA SUSPENDER OU ENCERRAR A PESQUISA

Caso qualquer irregularidade seja apurada, os pesquisadores responsáveis estão livres para suspender ou encerrar a pesquisa.