

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Informática

Dissertação de Mestrado

Um estudo comparativo da acurácia de algoritmos de
recomendação em sistemas de Compras Coletivas

Filipe da Silva Costa

Campina Grande - Paraíba
Novembro de 2014

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Informática

Um estudo comparativo da acurácia de algoritmos de
recomendação em sistemas de Compras Coletivas

Filipe da Silva Costa

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande – Campus I como parte dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação

Linha de Pesquisa: Metodologia e Técnicas da Computação

Dr. Leandro Balby Marinho

(Orientador)

Dr. Nazareno Ferreira de Andrade

(Orientador)

Campina Grande, Paraíba, Brasil

© Filipe da Silva Costa, 27/11/2014

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

C837e Costa, Filipe da Silva.

Um estudo comparativo da acurácia de algoritmos de recomendação em sistemas de compras coletivas / Filipe da Silva Costa. – Campina Grande, 2014.

95 f. : il. color.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2014.

"Orientação: Prof. Dr. Leandro Balby Marinho, Prof. Dr. Nazareno Ferreira de Andrade".

Referências.

1. Compras Coletivas. 2. Sistema de Recomendação. 3. Algoritmos de Recomendação. I. Marinho, Leandro Balby. II. Andrade, Nazareno Ferreira de. III. Título.

CDU 004.738.5(043)

**"UM ESTUDO COMPARATIVO DA ACURÁCIA DE ALGORITMOS DE
RECOMENDAÇÃO EM SISTEMAS DE COMPRAS COLETIVAS"**

FILIPPE DA SILVA COSTA

DISSERTAÇÃO APROVADA EM 19/11/2014


NAZARENO FERREIRA DE ANDRADE, D.Sc, UFCG
Orientador(a)


LEANDRO BALBY MARINHO, Dr., UFCG
Orientador(a)


CARLOS EDUARDO SANTOS PIRES, Dr., UFCG
Examinador(a)


MARCELO JOSÉ SIQUEIRA COUTINHO DE ALMEIDA, D.Sc., IFPB
Examinador(a)

CAMPINA GRANDE - PB

Resumo

Sites de Compras Coletivas (SCC) correspondem a um tipo específico de site que tem como objetivo divulgar ofertas (produtos, serviços ou viagens) com alto valor de desconto, por um tempo determinado, o que proporciona ao vendedor das ofertas um maior número de negociações em um curto período de tempo. Para atingir esse objetivo, os SCC utilizam estratégias de *marketing* que vão desde a divulgação das ofertas em redes sociais até o envio de e-mails diários para os usuários cadastrados. No entanto, a divulgação das ofertas para a maioria dos SCC é realizada de forma não personalizada, de maneira que todos os usuários recebem diariamente o mesmo conjunto de ofertas divulgadas. Logo, por falta de personalização dessa divulgação, os usuários acabam por receber uma grande quantidade de ofertas irrelevantes ou desinteressantes. Nesse sentido, propomos o estudo de um Sistema de Recomendação que leve em consideração aspectos importantes para os usuários de SCC, aspectos estes definidos a partir da análise da base de dados real de uma empresa que atua no domínio de compras coletivas. Neste estudo, avaliamos quatro algoritmos aplicados aos dados desse domínio, dois desses algoritmos considerados estado da arte em recomendação. Discutimos os resultados obtidos a partir dos experimentos, indicando qual dos algoritmos apresenta maior eficácia no domínio estudado, de acordo com as métricas definidas neste trabalho. A avaliação do trabalho foi realizada por meio experimental em parceria com a empresa *QueroDois*, sediada em Ribeirão Preto - São Paulo.

Abstract

Daily Deals Sites (*DDSs*) correspond to a specific website type designed to advertise offers (products, services or travel) at a significantly reduced prices, for a certain time, so the seller of the offers can make a large number of negotiations in a short period of time. To achieve this goal, the *DDSs* use marketing strategies ranging from advertising the offers on social network until to sending daily e-mails to registered users. However, the disclosure of the offers for most *DDSs* is not performed in a personalized manner, so all users receive the same set of daily offers. Thus, because of the lack of customization of this disclosure, users end up receiving a large amount of irrelevant or uninteresting offerings. Accordingly, we propose the study of a Recommender System that takes into account important aspects of users of *DDSs*. This aspects are defined by analyzing the real database of a company engaged in the group of buying domain. We evaluated four algorithms applied to data in this domain, two of these considered state of the art on recommendation, and discuss the results obtained from the experiments, also indicating which of those algorithms presents better efficacy, according to the metrics defined in this work. The evaluation of this work was performed by experimental means in partnership with company *QueroDois*, based in Ribeirão Preto - São Paulo.

Agradecimentos

Agradeço primeiramente a Deus, por ter me abençoado com saúde e, acima de tudo, sabedoria para concluir mais uma etapa da minha jornada: a conclusão do mestrado. Por ter me cercado de pessoas especiais e importantes, que sempre me apoiaram de diversas maneiras.

Aos meus pais e irmã por toda motivação e todo amor, além de tudo compreensão durante o período em que me ausentei do âmbito familiar para me dedicar aos estudos.

A minha namorada por toda a paciência, amor, companheirismo e incentivo nos momentos mais difíceis.

Aos meus orientadores Leandro Balby e Nazareno Andrade por todos os puxões de orelha, broncas e paciência, além de todos os *feedbacks*, pela motivação e empenho nas correções e orientação.

A todos aos amigos da OFF Thread e demais colegas que me incentivaram, direta ou indiretamente, durante essa trajetória, seja a partir do apoio moral ou por ouvir meus anseios.

Agradeço ainda à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Conteúdo

1. Introdução	1
1.1. Motivação.....	2
1.2. Objetivo.....	3
1.3. Relevância.....	3
1.4. Estrutura da Dissertação	4
2. Fundamentação Teórica	6
2.1. Sistemas de Recomendação	6
2.2. Abordagens de Algoritmos de Recomendação.....	11
2.2.1. Recomendação não personalizada baseada em sumarização dos dados	11
2.2.1.1. Mais Popular.....	12
2.2.2. Personalizada.....	13
2.2.2.1. Baseado em Conteúdo (Query & Search).....	16
2.2.2.2. Baseado em filtragem colaborativa.....	17
2.2.2.3. Baseado em Redução de Dimensionalidade (Fatoração de Matrizes)	20
2.3. Considerações Finais.....	23
3. Trabalhos Relacionados	24
3.1. Trabalhos relacionados ao modelo de negócio	24
3.2. Sistemas relacionados à recomendação de ofertas.....	26
3.3. Contribuições deste trabalho	29
4. Metodologia	32

4.1.	Base de Dados.....	32
4.2.	Fatores e Design do Experimento	37
4.3.	Objetivos	39
4.4.	Execução do experimento	39
4.4.1.	Escolha dos dias amostrais.....	39
4.4.2.	Montagem dos arquivos de treino, teste e arquivo auxiliar	39
4.4.3.	MyMediaLite: algoritmos, parâmetros e obtenção das métricas.....	41
4.4.4.	Lucene e implementação de algoritmo de recomendação baseado em conteúdo 42	
4.5.	Considerações Finais.....	42
5.	Resultados	44
5.1.	Experimento inicial: ofertas válidas por 24 horas	44
5.1.1.	Análise para o fator "algoritmo"	44
5.1.2.	Análise para o fator "hora de corte"	48
5.1.3.	Análise para o fator "número de itens a ser recomendado"	51
5.1.4.	Análise para o fator "número mínimo de compras por usuário"	55
5.2.	Experimento Adicional: Agrupamento do histórico de compras por quatro dias consecutivos.....	60
5.2.1.	Análise para o fator "algoritmo"	61
5.2.2.	Análise para o fator "hora de corte"	63
5.2.3.	Análise para o fator "número de itens a ser recomendado"	67
5.2.4.	Análise para o fator "número mínimo de compras por usuário"	70
5.3.	Discussão.....	74
5.4.	Considerações Finais.....	78
6.	Considerações Finais da Dissertação.....	80
6.1.	Conclusões	80
6.2.	Limitações e Trabalhos Futuros	83

7. Referências Bibliográficas	84
Apêndice A	88
Apêndice B	91

Lista de Figuras

Figura 1 - Predição de avaliação pelo sistema MovieLens.	7
Figura 2 - Recomendação de itens na loja <i>Amazon</i>	8
Figura 3 - Recomendação de vídeos no <i>YouTube</i>	8
Figura 4 - Exemplo de busca "Query & Search".....	15
Figura 5 - Representação da filtragem colaborativa baseada em itens, Item kNN[42].	18
Figura 6 - Alguns resultados da técnica de similaridade baseada no cosseno.....	19
Figura 7 - Distribuição de compras por cidade.....	33
Figura 8 - Ofertas mais compradas.....	34
Figura 9 - Distribuição das aquisições das ofertas por hora.	35
Figura 10 - Intervalo de confiança considerando ofertas e horários de pico para usuários com 5 ou mais compras	36
Figura 11 - Intervalo de confiança considerando ofertas e horários de pico para usuários com 10 ou mais compras	37
Figura 12 - Esquema do Banco de Dados.....	40
Figura 13 - Divisão da base em treino, teste e arquivo auxiliar.	41
Figura 14 - Intervalo de confiança para a métrica <i>MAP</i> observada.....	45
Figura 15 - Intervalos de confiança para as métricas precisão e <i>recall</i>	46
Figura 16 - Comparação entre os algoritmos em relação a precisão e <i>recall</i>	47
Figura 17 - Intervalos de confiança, considerando os níveis do fator "hora de corte", para a métrica <i>MAP</i>	49
Figura 18 - Intervalos de confiança, considerando os níveis do fator "hora de corte", para a métrica precisão.	50
Figura 19 - Intervalos de confiança, considerando os níveis do fator "hora de corte", para a métrica <i>recall</i>	51
Figura 20 - Intervalos de confiança, considerando os níveis do fator "número de itens a ser recomendado", para a métrica <i>MAP</i>	52

Figura 21 - Intervalos de confiança, considerando os níveis do fator "número de itens a ser recomendado", para a métrica precisão.	53
Figura 22 - Intervalos de confiança, considerando os níveis do fator "número de itens a ser recomendado", para a métrica <i>recall</i>	54
Figura 23 - Intervalos de confiança, considerando os níveis do fator "número mínimo de compras por usuário", para a métrica <i>MAP</i>	56
Figura 24 - Intervalos de confiança, considerando os níveis do fator "número mínimo de compras por usuário", para a métrica precisão.	58
Figura 25 - Intervalos de confiança, considerando os níveis do fator "número mínimo de compras por usuário", para a métrica <i>recall</i>	59
Figura 26 - Ofertas de 4 dias consecutivos agrupadas para simulação de experimento.	61
Figura 27 - Intervalo de confiança para a métrica <i>MAP</i> observada.	62
Figura 28 - Intervalos de confiança para a métrica precisão.	62
Figura 29 - Intervalos de confiança para a métrica <i>recall</i>	63
Figura 30 - Intervalos de confiança, considerando os níveis do fator "hora de corte", para a métrica <i>MAP</i>	64
Figura 31 - Intervalos de confiança, considerando os níveis do fator "hora de corte", para a métrica precisão.	65
Figura 32 - Intervalos de confiança, considerando os níveis do fator "hora de corte", para a métrica <i>recall</i>	66
Figura 33 - Intervalos de confiança, considerando os níveis do fator "número de itens a ser recomendado", para a métrica <i>MAP</i>	68
Figura 34 - Intervalos de confiança, considerando os níveis do fator "número de itens a ser recomendado", para a métrica precisão.	69
Figura 35 - Intervalos de confiança, considerando os níveis do fator "número de itens a ser recomendado", para a métrica <i>recall</i>	70
Figura 36 - Intervalos de confiança, considerando os níveis do fator "número mínimo de compras por usuário", para a métrica <i>MAP</i>	71
Figura 37 - Intervalos de confiança, considerando os níveis do fator "número mínimo de compras por usuário", para a métrica precisão.	73
Figura 38 - Intervalos de confiança, considerando os níveis do fator "número mínimo de compras por usuário", para a métrica <i>recall</i>	74
Figura 39 - Página referente ao Site de Compras Coletivas.	88

Figura 40 - Atributos de uma oferta em Site de Compras Coletivas 89

Lista de Tabelas

Tabela 1 - Técnicas de Recomendação.....	14
Tabela 2 - Diferenças entre dados de Sites de Compras Coletivas e de outros domínios	30
Tabela 3 - Teste de Wilcoxon.....	47
Tabela 4 - Conjunto de ensaios do experimento.....	93

Lista de Equações

Equação 1 - Predição de Avaliação de Item.	8
Equação 2 - Função Utilidade.	9
Equação 3 - Utilidade da importância de conteúdo.	17
Equação 4 - Similaridade baseada no cosseno.	19
Equação 5 - Similaridade baseada na correlação de Pearson.	19
Equação 6 - Similaridade baseada no cosseno ajustado.	20
Equação 7 - Cálculo da preferência do usuário sobre pares de itens.	22
Equação 8 - Cálculo simplificado da LPP.	22
Equação 9 - Cálculo do BPR para todos os usuários.	22

Lista de Acrônimos

SCC - Site de Compras Coletivas.

BPRMF - Bayesian Personalized Ranking Matrix Factorization.

ItemkNN - K Vizinhos mais próximos a um item.

SVD - Singular Value Decomposition.

LSI - Latent Semantic Indexing.

LPP - Likelihood of Pairwise Preference.

MAP - Mean Average Precision.

ROC - Receiver Operating Characteristic.

AUC - Area under the ROC Curve.

Capítulos da Dissertação

1. Introdução

Sites de Compras Coletivas (SCC) são sistemas compostos por três principais autores: o próprio domínio de Compras Coletivas, os usuários cadastrados nesses sistemas e também as entidades ou empresas que desejam anunciar itens nesses domínios. Os usuários têm intuito de acessar ofertas com descontos expressivos, devido ao alto volume de compras, sendo essas ofertas previamente acordadas entre um estabelecimento e o site. O objetivo desses sistemas é negociar com as empresas ofertas que serão propostas e divulgadas para seus usuários, construindo uma estratégia de marketing diferenciada [38]. No ano de 2012, o Brasil contava com cerca de 800 SCC ativos. Juntos, estes sites tiveram no primeiro trimestre desse mesmo ano um faturamento de, aproximadamente, R\$731,7 milhões, e o mercado ainda se encontrava em expansão e amadurecimento [39].

SCC possuem diferentes regras de negócios para divulgar suas ofertas e atrair os consumidores. Uma dessas regras, por exemplo, impõe um limiar mínimo de compradores, num intervalo de tempo, para que a compra seja concretizada (entende-se por compra concretizada como a efetivação e emissão do cupom referente ao item ofertado) [9]. Outra regra consiste na divulgação da compra na rede social e somente há concretização da oferta para um usuário quando ao menos três amigos comprarem também aquela oferta [7] (no Apêndice A são apresentados mais detalhes sobre o funcionamento de um SCC).

Com o crescente interesse das empresas em ofertar seus produtos para atrair cada vez mais clientes, os SCC passaram a utilizar diferentes estratégias de divulgação, como, por exemplo, anúncios publicitários, redes sociais e também ferramentas de recomendação para contatar os usuários cadastrados no sistema via e-mail. Tais ferramentas de recomendação estão ajustadas apenas para sugerir as ofertas que estarão, ou estão atualmente, divulgadas nos sites e ainda possuem disponibilidade para compra, ou seja, não atingiram o tempo limite da oferta e também não atingiram o limite máximo de compras. Portanto, SCC tipicamente divulgam toda e qualquer oferta sem a preocupação de realizar recomendações personalizadas de acordo com o perfil de compras do usuário, mesmo quando há métodos computacionais capazes de analisar cada perfil e realizar a recomendação individualizada, utilizando Sistemas de Recomendação.

1.1. Motivação

Os Sistemas de Recomendação são capazes de prever, de modo geral, a preferência de um indivíduo por um item, podendo utilizar duas abordagens básicas.

Na primeira abordagem, personaliza-se uma lista de itens que mais se adequam ao histórico do próprio usuário ou às preferências dos usuários com perfis semelhantes ao dele. A esta abordagem se dá o nome de predição de itens. Na segunda abordagem, tenta-se prever a avaliação do usuário, também de acordo com seu histórico de compras ou preferência de outros usuários. A esta abordagem se dá o nome de predição de notas. Logo, a partir da utilização de Sistemas de Recomendação, os SCC poderiam se tornar mais poderosos e eficientes quanto à predição de ofertas (consideradas itens) recomendadas para os usuários.

Apesar da importância de recomendar ofertas que realmente são atrativas para o usuário, há poucas informações na literatura quanto à análise dos dados de SCC, com o intuito de gerar uma ferramenta que possua capacidade para recomendar ofertas relevantes aos usuários. Esses dados são extraídos apenas quanto à realização de compra das ofertas por parte dos usuários (histórico de compras), portanto não possuem ranqueamento e são classificados como dados de *feedback* implícito [5]. Logo, a predição para a recomendação de ofertas é mais complexa, pois não é possível realizar uma classificação destas a partir da preferência do usuário. Além desse fato, falta entendimento sobre a mecânica de recomendação para esses usuários, tendo em vista que há fatores que podem ser relevantes para melhor recomendação nesse contexto e ainda não foram utilizados, como o melhor momento do dia para realizar a recomendação e o número mínimo de ofertas presentes no histórico do usuário, por exemplo. Tais fatores são importantes pois interferem diretamente na qualidade da recomendação que o usuário recebe.

O problema da recomendação em SCC, no sentido de melhorar a recomendação de ofertas diárias, vem sendo alvo de interesse por grandes empresas internacionais (GroupOn, LivingSocial) [35] e também nacionais (Peixe Urbano, ClickOn, Quero2), tendo em vista o grande sucesso de Sistemas de Recomendação para áreas afins, como, por exemplo, filmes (*Netflix*), músicas (*Last.fm*) e produtos (*Amazon*).

De forma geral, o problema que motiva este trabalho é a falta de conhecimento dos fatores importantes para que haja recomendação personalizada para os usuários no domínio de compras coletivas. A análise desses fatores importantes, seja da perspectiva temporal, comportamental, número mínimo de compras e etc., auxilia no perfilamento do usuário e,

consequentemente, garante melhor recomendação de ofertas, aliviando a sobrecarga de ofertas enviadas diariamente aos usuários.

Portanto, a análise dos dados de SCC e a compreensão de fatores importantes irão auxiliar no estudo do mecanismo de recomendação para esse domínio, avaliando uma ferramenta especialista em recomendação para SCC, considerando algoritmos que já são utilizados em outros contextos.

1.2. Objetivo

O objetivo principal deste trabalho é estudar e avaliar o comportamento de algoritmos de recomendação, que considerem também o cenário de compras coletivas. Esses algoritmos incluem fatores predominantes em cada contexto de estudo (produtos, filmes, músicas, livros) e são utilizados em outras áreas com sucesso, porém ainda sem aplicação alguma no domínio deste trabalho. Esses objetivos estão subdivididos nos seguintes objetivos específicos:

- Compreender a composição das ofertas em SCC, observando e analisando padrões de vendas, de acordo com o número de ofertas diárias e número de compras por oferta, verificando a relevância do estudo de algoritmos de recomendação que considere o perfil de cada cliente;
- Analisar as características específicas dos dados de SCC, com o intuito de extrair informações importantes para o Sistema de Recomendação, como, por exemplo, o número de ofertas por dia, o número de usuários que comprou ofertas e a distribuição do número de compras de ofertas por tempo (em horas);
- Avaliar a importância da perspectiva de tempo - em que momento é indicado para que se haja recomendação? - e conteúdo - a descrição da oferta é relevante para o perfil do usuário? - das ofertas nos SCC;
- Comparar, a partir de métricas relevantes para Sistemas de Recomendação, diferentes algoritmos estado da arte em recomendação, verificando se há melhorias na abordagem de recomendação atual (ofertas mais populares).

1.3. Relevância

Apesar do mercado de compras coletivas atualmente se encontrar em processo de depressão [35], alguns domínios ainda seguem, diariamente, com a divulgação de ofertas de produtos e serviços, além de diárias em hotéis, viagens e ingressos. A partir do presente estudo, há a possibilidade dos remanescentes nesse mercado compararem sua estratégia de divulgação,

bem como a análise dos dados, e verificar padrões em comum para possível melhoria de seu sistema.

Aliando ao ferramental de *marketing* um Sistema de Recomendação que personalize as ofertas para cada usuário, os SCC tendem a melhorar o relacionamento com o usuário, que recebe apenas ofertas de seu interesse, como também ter a opção de verificar as ofertas que são de maior interesse por parte dos usuários em geral, a partir da avaliação do *ranking* que o Sistema de Recomendação gera individualmente. Vale ainda a ressalva de que a expansão do acesso à Internet contribui para o crescimento acelerado de usuários e itens nos mais diversos domínios, como, por exemplo, nas redes sociais e sites de vendas de produtos. Portanto, personalizar informações relevantes ao usuário, analisando suas preferências, acentua o problema de recomendação em SCC.

Logo, é possível perceber que a literatura de recomendação deve ser expandida para contemplar também o cenário no qual as ofertas são válidas durante um certo período, há um número máximo de usuários compradores das ofertas e outros fatores que só fazem sentido quando SCC são analisados, auxiliando na escolha de algoritmos capazes de suprir a necessidade de recomendação nesse domínio, como também mostrar fraquezas que os algoritmos utilizados como *baseline* e estado da arte em recomendação possuem quando esse domínio é avaliado.

1.4. Estrutura da Dissertação

Os demais capítulos deste documento estão estruturados da seguinte forma: no *Capítulo 2* é apresentado o embasamento teórico acerca dos Sistemas de Recomendação, destacando seus tipos, limitações e aplicações e quatro algoritmos são demonstrados. Além disso, o capítulo expõe a definição de SCC e os cenários observados nesses.

No *Capítulo 3*, apresentamos a revisão da literatura, na qual são discutidos trabalhos desenvolvidos com o intuito de melhorar a recomendação em SCC.

Já no *Capítulo 4*, apresentamos a metodologia do presente trabalho, a qual inclui a explicação da estratégia de investigação, definição dos fatores importantes e dos experimentos realizados.

O *Capítulo 5* é composto pelos resultados obtidos com a realização dos experimentos. Além disso, discutimos o que foi observado em relação à execução dos experimentos e indicação do melhor nível de cada fator estudado.

Finalmente, no *Capítulo 6*, apresenta-se a conclusão do estudo acerca dos dados de SCC, bem como suas limitações e possíveis trabalhos futuros.

2. Fundamentação Teórica

O uso de Sistemas de Recomendação tem proporcionado resultados positivos na interação dos usuários com os diversos sistemas *online*. A grande disponibilidade de itens e a dificuldade de mapeamento entre as preferências do usuário e os itens à disposição, aliados à popularização da Internet fizeram com que a tarefa de achar rapidamente o que se deseja se tornasse custosa. Logo, surgiram diversos estudos acerca das recomendações de livros, músicas, filmes e produtos a partir da década de 90.

Hoje, sabe-se que há a possibilidade de recomendação personalizada de itens, de maneira a aproximar o usuário dos itens que são mais relevantes para seu perfil. Porém, diversos estudos nas áreas de psicologia, *marketing*, engenharia e computação ainda são desenvolvidos para aprimorar cada vez mais as recomendações em relação ao perfil do usuário, aprimorando as técnicas e modelos que são considerados estado da arte nessa área, facilitando a escolha dos usuários no momento de tomadas de decisão.

Neste capítulo são apresentados os fundamentos teóricos que utilizamos no decorrer da dissertação, considerando os principais conceitos da área de Sistemas de Recomendação.

2.1. Sistemas de Recomendação

Com o aumento do conteúdo disponível e a popularização da internet, aliados ao crescimento emergente do comércio eletrônico (*e-commerce*), instaurou-se um custoso e difícil processo para que os usuários consigam escolher produtos e serviços, dentre um leque de variedades e alternativas que lhe são apresentadas. Pensando nesta dificuldade, o desenvolvimento de sistemas de recomendação se tornou indispensável [19].

Um Sistema de Recomendação é uma combinação de técnicas computacionais que, a partir das preferências dos usuários e do contexto no qual estão inseridos, seleciona automaticamente itens para oferecer ao usuário. Itens, em geral, podem ser representados por livros, músicas, filmes, fotos, notícias, vídeos, programas de TV, anúncios, páginas na internet, produtos e serviços em uma loja virtual, etc. Grandes empresas como *Amazon*, *Netflix*, *Google* e *Facebook* utilizam intensamente Sistemas de Recomendação para obter vantagem competitiva em seus serviços.

Nos Sistemas de Recomendação, há duas possibilidades de computação de resultados: a predição da avaliação de itens ou a recomendação de itens. Quando há predição da

avaliação, o sistema de recomendação geralmente utiliza uma determinada escala para prever o resultado da avaliação de um determinado item. Por exemplo, em um sistema de recomendação de filmes, um usuário avalia explicitamente os filmes que assistiu, ranqueando-os em estrelas, no qual o nível mais baixo indica a insatisfação do usuário com o filme e o nível mais alto indica a total satisfação do usuário com o filme. Nesses sistemas, geralmente, há a indicação da avaliação, em quantidade de estrelas, dos filmes não assistidos pelo usuário, levando em consideração os filmes que foram previamente assistidos. Logo, o Sistema de Recomendação está predizendo que, caso o usuário assista aquele filme, possivelmente a avaliação a ser atribuída será a indicada. O Sistema de Recomendação de filmes *MovieLens* é um exemplo de sistema que utiliza, dentre as suas computações de resultados, a predição de itens (como pode ser visto na Figura 1).

Prediction or Rating	Your Rating	Movie Information	Wish List
★★★★	Not seen	Incendies (2010) DVD info imdb flag Movie Tuner Drama, Mystery, War - French, Arabic, English Popular tags: disturbing twist ending nonlinear	<input type="checkbox"/>
★★★★	Not seen	Ip Man (2008) DVD info imdb flag Movie Tuner Action, Drama, War - Cantonese, Mandarin, Japanese Popular tags: action kung fu martial arts	<input type="checkbox"/>
★★★★	Not seen	X-Men: First Class (2011) DVD info imdb flag Movie Tuner Action, Adventure, Sci-Fi, Thriller, War - English, Russian, German, French, Spanish Popular tags: fantasy comic book superhero	<input type="checkbox"/>
★★★★	Not seen	Lincoln (2012) info imdb flag Movie Tuner Drama, War Popular tags: politics Steven Spielberg Daniel Day-Lewis	<input type="checkbox"/>
★★★★	Not seen	Restrepo (2010) DVD info imdb flag Movie Tuner Documentary, War Popular tags: afghanistan documentary military	<input type="checkbox"/>
★★★★	Not seen	Beneath Hill 60 (2010) info imdb flag Movie Tuner Drama, War - English, German Popular tags: WWI World War I war	<input type="checkbox"/>
★★★★	Not seen	Patience Stone, The (2012) info imdb flag Movie Tuner Drama, War - Persian Popular tags: women's rights Persian Afghanistan	<input type="checkbox"/>

Figura 1 - Predição de avaliação pelo sistema MovieLens.

Já quando há recomendação de itens, o Sistema de Recomendação gera sugestões de itens que o usuário gostaria ou que se encaixam na atividade atual do usuário, geralmente representadas na forma de ranques *top-n*. Por exemplo, em um sistema de busca, um usuário escreve a busca e o motor de busca recomenda as páginas ou assuntos que mais se aproximam dos termos utilizados, levando em consideração também outros aspectos, tais como a localidade do usuário e a língua. Nesses sistemas, uma lista ordenada de maneira decrescente considerando similaridade do conteúdo das páginas com a busca é recomendada para o usuário, indicando as páginas com conteúdo mais relevantes. Diversos sistemas *online* utilizam esta abordagem, principalmente sistemas de comércio eletrônico, como *Amazon*, *WallMart*, *eBay* e outros. Também empresas como *Netflix*, na área de filmes, *Last.fm* na área

de música e *YouTube* na área de vídeos utilizam a recomendação de itens ao invés da predição de itens. Alguns exemplos podem ser observados nas Figura 2 e Figura 3.

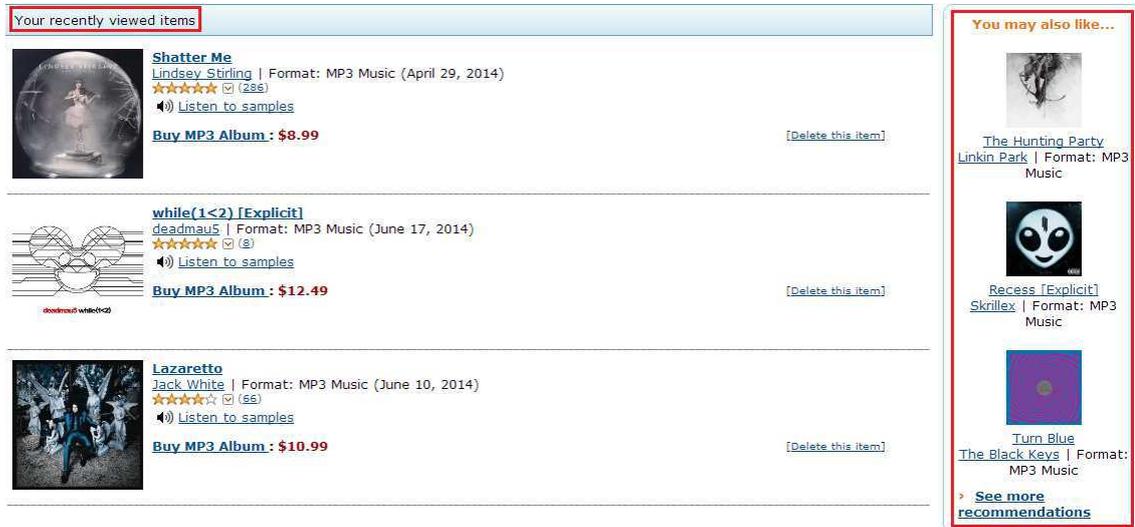


Figura 2 - Recomendação de itens na loja Amazon



Figura 3 - Recomendação de vídeos no YouTube

Formalmente, um sistema de recomendação baseado na predição de avaliações de itens pode ser definido da seguinte maneira: seja U , um conjunto de usuários do sistema e I , um conjunto de todos os possíveis itens do sistema, passíveis de recomendação. Dado um usuário $a \in U$ e um item $i \in I$, a predição de avaliação ρ do usuário a com relação ao item i pode ser definida, em uma de suas diversas formas, como:

Equação 1 - Predição de Avaliação de Item.

$$\rho_{a,i} = (\bar{r}_a + \sum_{u=1}^n (r_{u,i} - \bar{r}_u)) / n$$

onde $(\sum_{U=1}^n r_{u,i} - \bar{r}_u)$ representa a média normalizada das avaliações recebidas pelo item i , $r_{u,i}$ representa a avaliação de usuário com relação ao item i e \bar{r}_u representa a média de avaliações do usuário u . Em geral, temos por definição que a predição da avaliação de um usuário u sobre um item i depende da sua média de avaliações somada à média normalizada das avaliações que o item i recebeu. Variações foram estudadas na literatura para ajustar a predição dos itens, como nos trabalhos [21] e [27].

Já um Sistema de Recomendação de itens pode ser definido matematicamente da seguinte maneira: seja U o conjunto de usuários do sistema e I o conjunto de todos os possíveis itens do sistema, passíveis de recomendação. Seja μ a função utilidade que determina quão útil um determinado item i é para um determinado usuário u , isto é, $\mu: U \times I \rightarrow \mathbb{R}$, sendo \mathbb{R} o conjunto dos números reais. Logo, para cada usuário $u \in U$, há a necessidade de se determinar um item $i' \in I$ tal que a utilidade seja maximizada. Estendendo para todos os usuários do sistema, temos que

Equação 2 - Função Utilidade.

$$\forall u \in U, i'_u = \operatorname{argmax}_{i \in I} \mu(u, i).$$

Ou seja, para todo usuário u , um item i' representa o maior valor calculado por meio da função utilidade, que tem como parâmetros o próprio usuário e todos os itens de I .

A função de utilidade μ possui diferentes representações nos mais diversos ambientes em que os Sistemas de Recomendação estão inseridos, sendo esta geralmente representada pela avaliação que indica a preferência de um usuário por um determinado item (o usuário "gosta" de um item). A problemática central dos Sistemas de Recomendação consiste justamente nesse ponto, avaliar a utilidade de cada item para cada usuário não é conhecida, ou seja, a utilidade não está definida em todo o espaço $U \times I$, mas em apenas uma parte deste. Em outras palavras, as avaliações estão definidas em apenas um subconjunto do espaço, subconjunto este representado pelos itens que foram previamente avaliados pelo usuário. Logo, o algoritmo de recomendação deve ser capaz de prever, a partir dos itens avaliados previamente, quais seriam as avaliações dos itens que ainda não foram avaliados - evento mais conhecido como **extrapolação** - e realizar as recomendações de acordo com a predição [23]. Esta predição é realizada com auxílio dos métodos de teorias de aproximações, as mais diversas heurísticas e também pelos métodos de aprendizagem de máquina.

A técnica de extrapolação tem suma importância para que as avaliações desconhecidas possam ser aproximadas. Pode ocorrer de diferentes formas, porém há duas abordagens geralmente utilizadas: a estimativa da função de utilidade, que verifica a otimização de algum critério de desempenho; e a especificação de heurísticas para definir e validar a função de utilidade.

Para cada tipo de abordagem, há utilização dos mesmos dados, porém sob uma perspectiva de obtenção diferente. Esses dados podem ser obtidos por meio das interações dos usuários com o sistema, sendo estas interações ações classificadas como explícitas ou implícitas. As interações são classificadas como explícitas quando o usuário transpõe sua opinião acerca de determinados itens, por meio de votações, notas ou mesmo opiniões. Exemplos clássicos de dados explícitos são: as votações com estrelas, nas quais os usuários avaliam um determinado item escolhendo um número de estrelas que varia de um a cinco; e *likes* e *thumbs*, em que o usuário vota pela preferência de gostar ou não de um item.

Já a classificação implícita ocorre quando o usuário expressa sua preferência a partir de ações, como páginas visitadas e compras realizadas. Exemplos clássicos de dados implícitos são: tempo gasto pelo usuário na leitura ou próximo clique em *link* em uma determinada página, compras feitas pelo usuário em um determinado momento do tempo.

Sistemas de Recomendação ainda possuem estratégias de recomendação para aumentar a fidelidade dos seus usuários e a lucratividade das empresas. Diferentes estratégias são utilizadas para apresentar recomendações personalizadas, já que cada estratégia envolve um grau de complexidade variado e distinto no tratamento da informação coletada. As estratégias geralmente utilizadas são: reputação do produto, recomendações por associações, associações por conteúdo e análise na sequência de ações [30].

Recomendar itens relevantes ou úteis, em meio à grande quantidade de itens que fazem parte dos mais diversos sistemas de *e-commerce*, é uma tarefa difícil porque a utilidade dos itens para determinados usuários depende de fatores que muitas vezes são desconhecidos. Seja a necessidade representada na forma de uma viagem, um filme, uma música, um novo produto para cabelos, um brinquedo infantil, uma fantasia para determinada época do ano ou mesmo artigos tecnológicos, a recomendação deve aproximar cada vez mais os itens relevantes dos usuários para que somente um subconjunto dos itens sejam apresentados como resultados. Porém, esse é um problema complexo e, por isso, algumas abordagens vêm sendo

estudadas na área de Sistemas de Recomendação, das quais estão: a recomendação não personalizada baseada em sumarização dos dados, a filtragem de itens baseada em conteúdo, a filtragem colaborativa e abordagens híbridas, que geralmente se beneficiam dos pontos fortes de cada uma dessas abordagens.

2.2. Abordagens de Algoritmos de Recomendação

2.2.1. Recomendação não personalizada baseada em sumarização dos dados

Um algoritmo de recomendação é dito não personalizado quando todos os usuários recebem a mesma recomendação, seja ela definida por um critério ou não. Por critério, define-se como uma regra que serve de base para comparação dos itens a serem recomendados e julgamento de quais itens obedecem a regra definida.

Quando não há critérios definidos previamente, a recomendação é randômica, ou seja, escolhe-se aleatoriamente um subconjunto de itens do conjunto total e se recomenda para todos os usuários do sistema. Já quando há definição de critérios, há sumarização dos dados anteriormente à recomendação e, posteriormente, a recomendação de determinados itens, escolhendo um subconjunto pré-definido, de acordo com o critério, e se recomenda para todos os usuários do sistema. A sumarização dos dados ocorre como forma de produzir as regras que serão impostas à recomendação, como, por exemplo, "recomendar apenas itens que possuem valor acima de R\$100,00" ou mesmo "recomendar apenas itens que precisam ter um número mínimo de compradores igual a 50".

Supondo que o critério seja definido como "os itens mais acessados pelos usuários", temos uma sumarização prévia de quais itens foram mais acessados durante um determinado tempo e recomendamos os itens que possuam o maior número de visualizações. Definimos este tipo de critério, no escopo deste trabalho, como "Mais Vistos". Quando o critério definido é "os itens que foram melhor avaliados pelos usuários", temos uma sumarização prévia de quais itens tiveram a melhor avaliação dos usuários, seja esta uma avaliação qualitativa - bom, muito bom, ótimo, ruim, péssimo - ou uma avaliação quantitativa - cinco estrelas, nota dez. Este tipo de critério é definido como "Melhor avaliados".

Alguns critérios podem ser definidos por uma função que indica a pré-disposição de um item ser recomendado com relação a outros itens, numa recomendação conjunta, como por exemplo "Compraram Junto".

Schafer *et al.* [3] mostraram em seu trabalho uma taxonomia que define técnicas de recomendação, sendo a recomendação não personalizada classificada como efêmera - não permite customização da recomendação, pois o sistema de recomendação não faz distinção entre os usuários do sistema, recomendando os mesmos itens para usuários em sessões diferentes - e automática - há pouco ou nenhum esforço da parte do usuário para que haja recomendação.

Para o estudo da recomendação em SCC conduzido nesta dissertação, houve a análise do algoritmo "Mais Popular" como exemplo de algoritmo que utiliza recomendação não personalizada. Este algoritmo é bastante simples e, muitas vezes, eficiente em alguns dos mais diversos cenários da recomendação.

2.2.1.1. Mais Popular

Como parte do processo de recomendação utilizando técnicas que permitem recomendar itens de maneira não personalizada, o algoritmo "Mais Popular" se destaca em diversos sistemas mais robustos.

Empresas como *Amazon*, *Netflix* e *eBay* no mercado internacional e, no mercado nacional, a *Netshoes*, fazem a utilização desse algoritmo, sendo este também utilizado na literatura como um algoritmo de *baseline*. Por *baseline*, entende-se por um algoritmo que servirá como base para comparar a eficiência de novos algoritmos ou de novos casos em que não houve aplicação de sistemas de recomendação e utilizado em sistemas reais, como os citados anteriormente, fazendo parte da recomendação não-personalizada (por exemplo, a recomendação de produtos, filmes, músicas e outros quando o usuário ainda não possui um perfil no sistema).

Esse algoritmo é utilizado, também como *baseline*, na recomendação de marcação (*tags*) [14], recomendação de filmes [15], músicas, notícias [18], vídeos, ferramentas, aplicativos e outros. A ideia principal é que os itens mais populares sejam recomendados a todos os usuários, pois como premissa dessa abordagem têm-se que os itens que possuem alta popularidade entre os usuários são interessantes para a população em geral.

Como ponto forte da utilização desse algoritmo, é possível destacar que é de fácil entendimento e execução, simples e eficaz, tendo em vista que o efeito da popularização de

um determinado item é bem aceito pela maior parte da população por produzir um efeito de familiarização dos usuários com um determinado item, aumentando a confiança dos mesmos na tomada de decisões importantes [LUHMANN, N. Trust and power. 1982]. Como pontos fracos, podemos destacar a falta de personalização da recomendação para os usuários, ou seja, todos recebem a mesma recomendação. Ainda há problemas com a descoberta de novidades (*novelty*) e a impossibilidade de diversidade (*diversity*) [28] já que, por definição, os itens a serem recomendados são os que mais foram difundidos na população. Logo, não há possibilidade de itens de diversas categorias serem recomendados, para o problema da diversidade, e de que itens que surpreendam o usuário sejam recomendados, para o problema de descoberta de novidades - esta medida de forma qualitativa, determinada a partir do quanto o usuário se surpreendeu com a recomendação - e aquela medida de forma quantitativa - determinada a partir do quanto os itens recomendados diferem entre si, a partir de um critério.

2.2.2. Personalizada

O algoritmo de recomendação é dito personalizado quando cada usuário recebe uma recomendação sob medida, sendo obrigatoriamente esta definida por um critério que, ao ser aplicado, recomenda itens de acordo com o histórico de interações entre um usuário específico e o sistema. *Burke* [25] define um Sistema de Recomendação personalizado como: "*qualquer sistema que produz recomendação personalizada como saída, ou tem o efeito de guiar o usuário, de maneira personalizada, a itens de seu interesse ou itens úteis em um grande espaço de possibilidades*". O algoritmo estabelece o ranqueamento ou a predição de um subconjunto de itens pré-definidos de acordo com um critério e, de forma personalizada, realiza recomendação de um subconjunto específico de itens para cada usuário do sistema, que é próximo ao seu interesse.

Os critérios dependem diretamente do conjunto de dados. Esses podem ser referentes às compras efetuadas, páginas visitadas, questionários respondidos, notícias lidas ou qualquer outro tipo de dado nos quais seja possível obter, explícita ou implicitamente, o interesse do usuário por determinados itens. Supondo que o critério seja definido como "os itens que possuem a mesma categoria de itens que o usuário já visitou", temos uma análise prévia de quais itens foram visitados pelo usuário em um determinado momento e recomendamos os itens que possuem semelhança na categoria do item que foi previamente visitado. Este tipo de critério é definido como " Similares a". Alguns critérios podem ser definidos por meio de uma função que indica a pré-disposição de um item ser recomendado com relação a outros

itens, como quando o critério definido é "os itens que são semelhantes aos itens melhor avaliados pelo usuário". Temos uma análise prévia de quais itens tiveram a melhor avaliação do usuário, seja esta uma avaliação qualitativa - bom, muito bom, ótimo, ruim, péssimo - ou uma avaliação quantitativa - cinco estrelas, nota dez; para que haja a escolha de um subconjunto de itens melhor avaliados e posterior comparação com outros itens que são semelhantes a estes. Este tipo de critério é definido como "Similares aos melhor classificados" no escopo desta dissertação.

Os Sistemas de Recomendação que utilizam algoritmos personalizados podem ser divididos em: colaborativos, demográficos, baseados em conteúdos, baseados na utilidade e baseados no conhecimento. As técnicas de recomendação estão na Tabela 1.

Tabela 1 - Técnicas de Recomendação

Técnica	O que é explorado	Entrada	Processo
Filtragem Colaborativa	Avaliações de U em itens I .	Avaliações de u em itens I .	Identifica usuários em U similares a u , extrapolando deles a avaliação de i .
Conteúdo	Características dos itens I .	Avaliações de u em itens I .	Gera um classificador que se ajusta ao comportamento de avaliação de u e utiliza em i .
Demográfico	Informação demográfica sobre U e suas avaliações dos itens em I .	Informações demográficas sobre u .	Identifica usuários que são demograficamente similares a u , extrapolando deles a avaliação de i .
Baseado em Utilidade	Características dos itens I .	Uma função utilidade sobre os itens em I que descrevem as preferências de u .	Aplica a função ao item e determina o <i>rank</i> de i .
Baseado em Conhecimento	Características dos itens I . Conhecimento em como esses itens atendem às necessidades	Descrição dos interesses ou preferências de u .	Infere um casamento entre i e a necessidade de u .

A Tabela 1 explica, de maneira resumida, as técnicas utilizadas nos algoritmos de recomendação, a entrada que deve ser fornecida e o processo realizado para determinar os itens que serão recomendados, de forma a exemplificar cada técnica utilizada nos algoritmos de recomendação.

Dentre as possíveis escolhas de algoritmos que realizam recomendação personalizada, três deles foram estudados no presente trabalho, com o intuito de verificar a efetividade da recomendação - observando as métricas definidas no Capítulo de Metodologia. O primeiro algoritmo pertence à classe dos algoritmos de recomendação baseados em conteúdo, sendo a abordagem de análise de conteúdo realizada por intermédio da técnica de "Query & Search", similar a ferramentas de busca, como *Google* e *Bing*. Essa técnica, em sua forma mais simples, analisa a similaridade entre os termos utilizados na busca e os principais termos dos documentos, estes podendo ser arquivos, páginas e outros, como pode ser visto na Figura 4, como exemplo de busca na ferramenta de busca do *Google*. Utilizou-se essa abordagem considerando que cada *query* é representada pela descrição dos itens que estão sendo ofertados no SCC e como conjunto de *search* as compras efetuadas previamente pelo usuário.

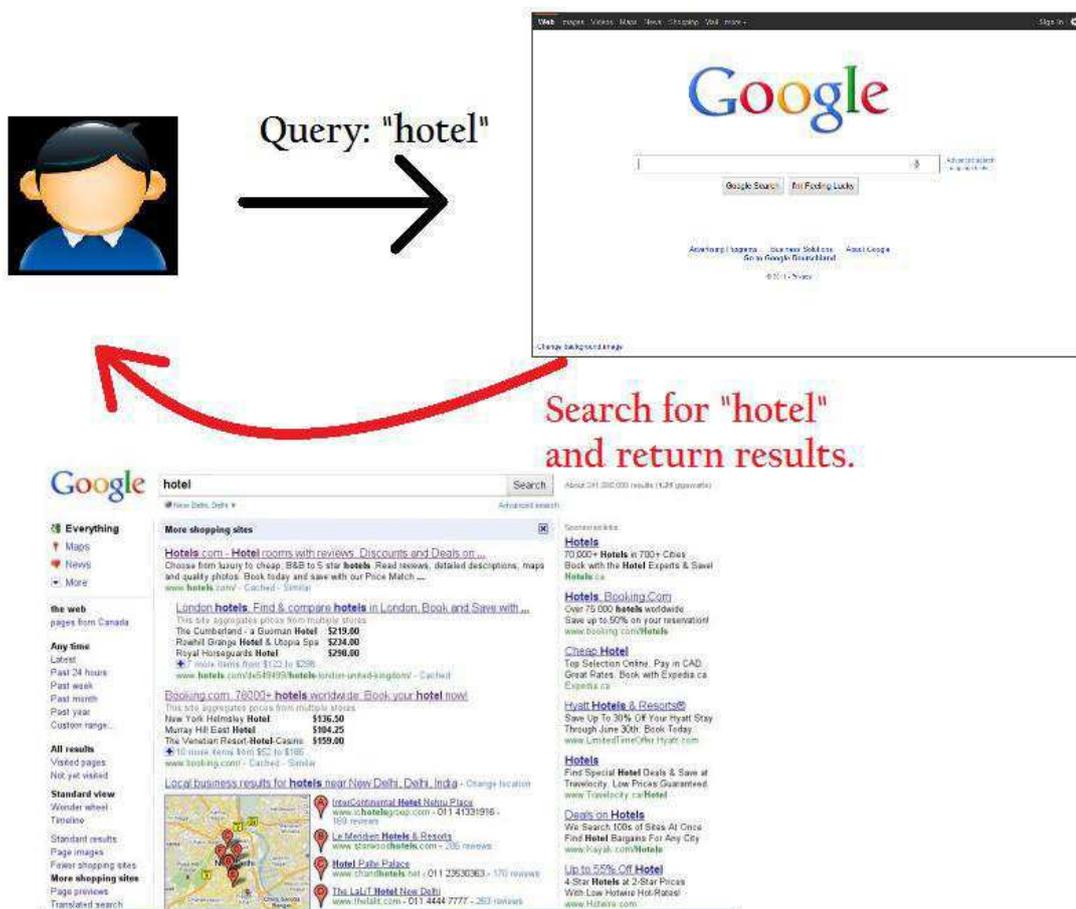


Figura 4 - Exemplo de busca "Query & Search".

Já o segundo algoritmo utilizado neste estudo pertence à classe dos algoritmos de recomendação colaborativa baseada em itens. Essa técnica, em sua forma mais simples, analisa a similaridade entre pares de itens (utilizando técnicas de similaridade, como por

exemplo correlação de *Pearson*) e realiza a predição de itens para os usuários. Apresenta vantagens quando o número de usuários é maior que o número de itens, o que ocorre em SCC.

O terceiro algoritmo também pertence à classe dos algoritmos de recomendação que utilizam filtragem colaborativa, na qual a abordagem de redução de dimensionalidade "Singular Value Decomposition"- SVD - foi estudada. Essa técnica, em sua forma mais simples, gera um subespaço de preferências dos usuários de forma mais compacta e robusta, em que há redução de dimensionalidade em relação a matriz total de usuários x itens. Apresenta vantagens na obtenção dos "vizinhos" - itens ou usuários que possuem maior similaridade de acordo com um determinado critério - pois utiliza modelo mais compacto que restringe o espaço de busca do total de itens ou usuários mais próximos.

2.2.2.1. Baseado em Conteúdo (Query & Search)

Na abordagem baseada em conteúdo, itens são recomendados para um usuário a partir da verificação do conteúdo dos demais itens que foram acessados previamente. Por exemplo, se filmes do gênero "Comédia" forem bem avaliados por um usuário em um Sistema de Recomendação de filmes, outros filmes do gênero "Comédia" ou mesmo "Comédia-Romance" serão recomendados para este usuário, pois há maior similaridade desse gênero com o perfil do usuário do que filmes do gênero "Ação" ou "Aventura".

Cada item, em um Sistema de Recomendação que utiliza um algoritmo baseado em conteúdo, possui atributos descritores que os caracterizam e fazem parte do seu perfil. Esse perfil geralmente é extraído do conteúdo do item, seja esse a cor, o valor, a descrição ou qualquer outro conteúdo significativo para a recomendação. Filmes podem ser caracterizados pelos autores, diretor, nome, gênero e ano, já músicas pelos seus autores, ritmo, estilo musical e ano. Ofertas de SCC podem ser caracterizadas pela sua descrição, valor, desconto, validade, regras da oferta e categoria.

Existem duas classificações correspondentes às técnicas utilizadas na recomendação baseada em conteúdo: consulta e busca (*query & search*) e as técnicas de elicitación e estimación (*elicitation and estimation techniques*) [29]. Para o escopo desta dissertação, consideramos apenas a primeira técnica, pela facilidade de implementação e de análise dos resultados, além da novidade, já que esse algoritmo não foi estudado com entrada de dados de SCC, de acordo com a literatura.

Da área de Recuperação da Informação (RI), podemos estruturar o problema de recomendação contextual utilizando a técnica de consulta e busca da seguinte maneira: Seja u

um usuário e $Perfil(u)$ um vetor que representa as preferências do usuário (classificação de filmes, estilos de músicas, etc.), obtidas por meio dos itens que foram anteriormente acessados por u , utilizando as palavras-chaves contidas no conteúdo dos itens. Sendo i um item qualquer e $Conteúdo(i)$ um vetor que indicam os atributos individuais que caracterizam i , podemos definir a função utilidade do item i para o usuário u como

Equação 3 - Utilidade da importância de conteúdo.

$$Utilidade(u,i) = Avaliação(Perfil(u), Conteúdo(i))$$

na qual essa avaliação consiste na medição da importância do conteúdo de i para o perfil de u , ou seja, quão próximo está o conteúdo do item ao perfil do usuário.

Como pontos fortes da utilização desse algoritmo, é possível destacar a facilidade de entendimento e execução, simplicidade e tratamento do problema de *cold start*. Por *cold start*, define-se a situação em que um novo usuário ou item é inserido no sistema e não há avaliações suficientes para prever quais são os usuários/itens que possuem similaridade com este, fazendo com que não haja como calcular a similaridade entre usuários ou itens. Logo, há melhores resultados na recomendação, em casos típicos de *cold start*, quando se utiliza puramente o algoritmo baseado em conteúdo [22] ou mesmo em abordagens híbridas, segundo Gunawardana *et al.* [24].

Como pontos fracos, há a necessidade de estudo aprofundado das características dos itens que estão sendo utilizados nos algoritmos de recomendação para extração dos atributos, pois a inclusão/exclusão dos atributos impacta diretamente na recomendação, diminuindo a acurácia da mesma.

2.2.2.2. Baseado em filtragem colaborativa

A filtragem colaborativa é classificada de acordo com duas abordagens básicas: baseada em memória (*user-based*) e baseada em modelo (*item-based*).

A abordagem baseada em memória tem pilares no conceito de vizinhança, esta representada por usuários que possuem ou possuíram, historicamente, o mesmo "gosto" nos mais diversos itens da base de dados de um sistema. A partir de então, itens são recomendados a partir da proximidade de um usuário com os que estão na sua vizinhança.

Amplamente utilizada em sua concepção, atualmente esta abordagem caiu em desuso por três principais motivos [23]: a grande quantidade de itens a serem avaliados, em que usuários ativos no sistema avaliam poucos itens, gerando recomendações com baixa acurácia, sendo este problema denominado **esparsidade** da matriz de avaliação. O segundo motivo é o

problema da **escalabilidade**, pois cada vez mais cresce o número de usuários e itens em um sistema, fazendo com que se torne cada vez mais custoso avaliar e escolher os itens que, supostamente, fazem parte do perfil do usuário. O terceiro motivo é a possibilidade de rápida mutabilidade das preferências do usuário, impossibilitando da computação em tempo real da recomendação.

Já a abordagem baseada em modelo observa o conjunto de itens avaliados por um determinado usuário e a similaridade entre estes itens e novos itens-alvo - itens a serem recomendados ou não para o usuário. Basicamente, computar a similaridade entre os itens i e j consiste em, inicialmente, isolar os usuários que avaliaram os dois itens e, posteriormente, aplicar uma técnica computacional de cálculo de similaridade entre eles, determinada por $S_{i,j}$. Esta abordagem está representada na Figura 5.

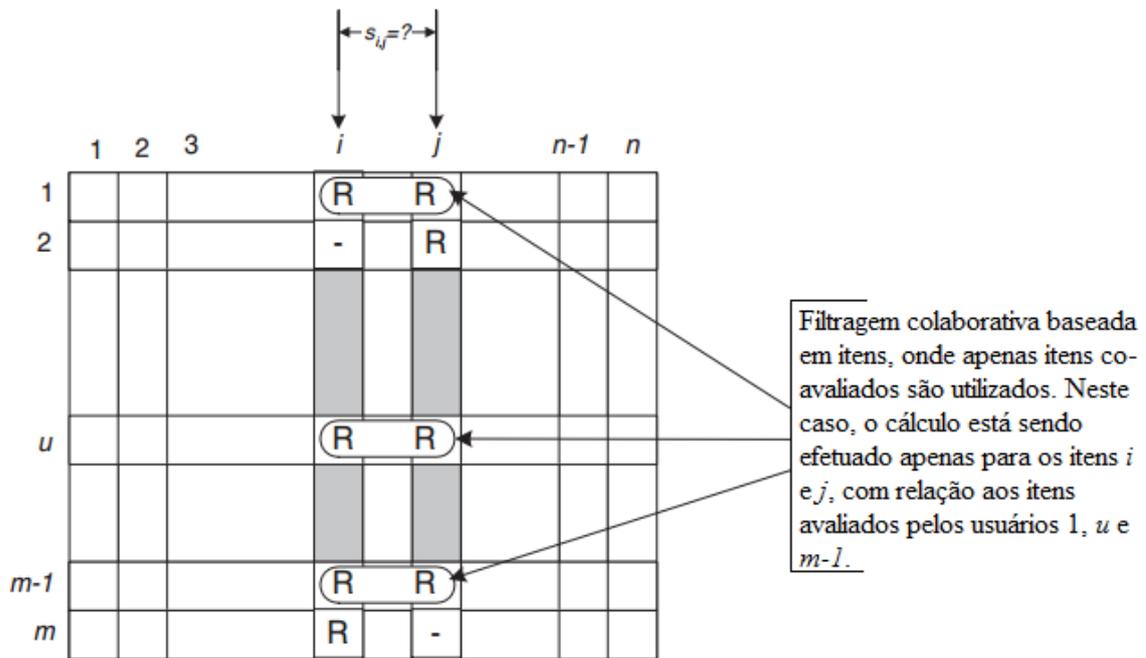


Figura 5 - Representação da filtragem colaborativa baseada em itens, Item kNN[42].

O algoritmo de recomendação calcula a proximidade entre os itens-alvo e os itens avaliados por intermédio de técnicas de similaridade, nas quais as mais aplicadas na literatura correspondem à similaridade baseada no cosseno, baseada no cosseno ajustado e baseada na correlação.

Quando se utiliza a similaridade baseada no cosseno, dois itens são representados por vetores no espaço m dimensional de usuários. A similaridade então é calculada a partir do cosseno do ângulo entre estes dois vetores, em que podemos definir formalmente o cálculo de similaridade entre os itens i e j , apresentado na Figura 5, como

Equação 4 - Similaridade baseada no cosseno.

$$sim(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\| * \|\vec{j}\|}$$

em que o divisor da fração indica o produto escalar de dois vetores e o dividendo da fração indica a multiplicação da norma dos vetores. Alguns dos possíveis resultados podem ser visualizados na Figura 6.

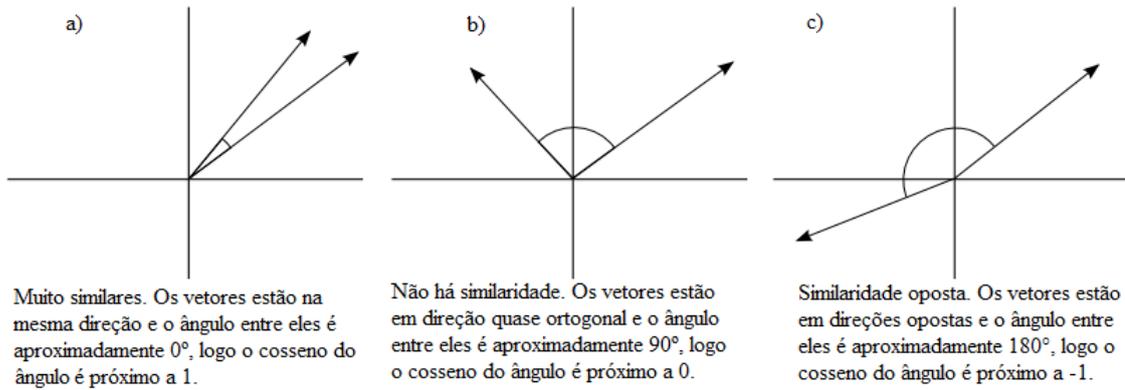


Figura 6 - Alguns resultados da técnica de similaridade baseada no cosseno.

Já quando se utiliza a similaridade baseada na correlação, a similaridade de dois itens i e j é calculada a partir da correlação de *Pearson*, que mede a força e a direção da relação linear entre dois vetores. Formalmente, o cálculo da correlação entre os itens i e j pode ser realizado da seguinte maneira:

Equação 5 - Similaridade baseada na correlação de Pearson.

$$sim(i, j) = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_{u \in U} (r_{u,j} - \bar{r}_j)^2}}$$

em que $r_{u,i}$ e $r_{u,j}$ representam a avaliação do usuário u sobre os itens i e j , \bar{r}_i e \bar{r}_j representam a média das avaliações dos itens i e j e U representa o conjunto de usuários.

A similaridade baseada no cosseno ajustado utiliza a pré-condição de que os usuários personalizam seus padrões para avaliação de itens. Para um mesmo filme, dois usuários que gostem muito do filme podem ranqueá-lo com 4 e 5 estrelas, respectivamente. Já para um filme que gostem pouco, podem ranqueá-lo com 1 e 3 estrelas, respectivamente. Logo, cada usuário tem seu padrão para avaliação e isto pode gerar ruídos na recomendação. Esta técnica busca diminuir essa desvantagem, subtraindo a média de avaliações de cada usuário de cada

avaliação dada por este ao par de itens i e j . Podemos, formalmente, calcular a similaridade baseada no cosseno ajustado da seguinte forma:

Equação 6 - Similaridade baseada no cosseno ajustado.

$$sim(i, j) = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u)(r_{u,j} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{u,j} - \bar{r}_u)^2}}$$

em que \bar{r}_u representa a média de avaliações do usuário u .

A utilização de filtragem colaborativa baseada em modelo, para gerar recomendações, gerou diversos métodos para buscas associativas, dentre eles o Item kNN . A denominação kNN surge a partir do momento em que deseja-se ranquear os k vizinhos mais próximos, ou *k-nearest neighbor*, sendo esse o modelo mais popular utilizado no algoritmo de filtragem colaborativa.

Em diversos estudos, cientistas provaram a eficácia dessa abordagem, no qual o foco pode estar tanto na interação entre os usuários (baseada em usuários) como na interação entre os itens (baseado em itens) [10][11].

Como pontos fortes da utilização desses algoritmos, é possível destacar a qualidade de recomendação para aplicações de *e-commerce* e a alta desempenho, dado que o número de itens é praticamente estático em relação ao número de usuários e a similaridade entre os produtos pode ser pré-computada e utilizada posteriormente como peso no modelo para a recomendação. Além disso, é bastante flexível, podendo ser aplicada tanto ao carrinho de compras do usuário quanto ao perfil, por exemplo.

2.2.2.3. Baseado em Redução de Dimensionalidade (Fatoração de Matrizes)

Single Value Decomposition (SVD) é uma técnica de fatoração de matrizes comumente utilizada para produzir aproximações de baixo *ranking* [12]. Esta técnica foi proposta pela comunidade de Recuperação da Informação, devido à impossibilidade de recuperação da informação a partir de consultas que possuíam o mesmo valor semântico. Como exemplo para o problema, podemos realizar a consulta "*Frameworks de Compiladores*" e, no conjunto de documentos, possuir um documento intitulado "*Sistemas para transformação de linguagens de programação em código de máquina*". Logo, como não há identificação do valor semântico da consulta com os dados da base de documentos, não há possibilidade de acesso à informação desejada, muito embora o documento seja de suma importância para a consulta. Em outras

palavras, se não há a utilização das palavras exatas, há a impossibilidade de recuperação da informação.

A partir do problema destacado, cientistas buscaram nas técnicas matemáticas uma solução, elaborando a técnica de indexação semântica latente (**LSI**) [16], decompondo a matriz de informações (supondo documentos \times termos) em várias dimensões, para que se compute a aproximação do *ranking* de cada termo, e cuja multiplicação dos elementos dos vetores formados pela decomposição da matriz original esteja o mais próximo possível da matriz original.

Os pontos positivos da redução de dimensionalidade utilizando a técnica de **SVD** resultam em modelos mais rápidos e menores, com conexões de vizinhanças mais eficazes - a redução da dimensão elimina os vizinhos que não estão no mesmo espaço de preferência. Apesar dos pontos positivos, a redução de dimensionalidade apresenta desafios, tais como o preenchimento de valores ausentes (a falta de preferência de um usuário por um item), a complexidade computacional e a dificuldade de transparência (baixa *explainability*) dos resultados.

A técnica de *Ranking* Personalizado Bayesiano (**BPR**) para aproximação da classificação de todos os itens é descrita por Rendle *et al.* [5]. Para aplicação da técnica, faz-se necessário, inicialmente, definir formalmente o problema. Utilizando U e I como o conjunto de usuários e itens, respectivamente, e para cada usuário $a \in U$, temos o conjunto de itens $I_u \subseteq I$ cujo usuário expressou *feedback* positivo, ou seja, adquiriu o item. Nosso objetivo é, então, recomendar uma lista personalizada de itens de I/I_u , ou seja, um conjunto de itens que seja personalizado em relação aos itens que o usuário já adquiriu, sendo este problema nominado ranqueamento colaborativo a partir de *feedback* implícito.

Para representar a preferência de um usuário acerca de dois itens i e j , Rendle *et al.* [5] utilizaram uma variável randômica binária $\delta((u, i) \succ (u, j))$ para inferir a preferência do usuário u sobre o item i e não o item j . Esta função $\delta(z)$ é binária e possui valor igual a 1 caso z seja positivo e 0 caso z seja negativo, tendo essa notação, usualmente, a denominação de preferência do usuário sobre pares de itens. Em um Sistema de Recomendação que utilize essa abordagem, é necessário calcular a preferência do usuário sobre todos os pares de itens do sistema para gerar a lista de recomendação. Esse processo é conhecido como o cálculo da LPP (*likelihood of pairwise preference*), que utiliza a distribuição de *Bernoulli* em conjunto com a função de preferência do usuário sobre os pares de itens para realizar a computação das

preferências do usuário sobre todos os pares de itens do sistema. Formalmente, Rendle *et al.* definiram o cálculo da LPP como:

Equação 7 - Calculo da preferência do usuário sobre pares de itens.

$$\begin{aligned} LPP(u) &= \prod_{i,j \in I} \Pr(\hat{r}_{ui} > \hat{r}_{uj})^{\delta((u,i) > (u,j))} \\ &\times [1 - \Pr(\hat{r}_{ui} > \hat{r}_{uj})]^{[1-\delta((u,i) > (u,j))]} \\ &= \prod_{(u,i) > (u,j)} \Pr(\hat{r}_{ui} > \hat{r}_{uj}) [1 - \Pr(\hat{r}_{uj} > \hat{r}_{ui})] \end{aligned}$$

, onde $(u, i) > (u, j)$ denota a preferência do usuário u no item i ao invés do item j .

Utilizando o modelo **BPR** apresentado anteriormente, assume-se que duas premissas fundamentais serão adotadas. A primeira premissa corresponde à preferência individual de um usuário u sobre os itens i e j , em que u prefere i se o par (u, i) for observado e o par (u, j) não for. A partir desta premissa, o cálculo do $LPP(u)$ pode ser simplificado por

Equação 8 - Calculo simplificado da LPP.

$$BPR(u) = \prod_{i \in I_u} \prod_{j \in I/I_u} \Pr(\hat{r}_{ui} > \hat{r}_{uj}) [1 - \Pr(\hat{r}_{uj} > \hat{r}_{ui})]$$

onde $i \in I_u$ significa que o par usuário-item (u,i) é observado e $j \in I/I_u$ significa que o par usuário-item (u,j) não foi observado.

Já a segunda premissa corresponde à independência entre os usuários, quando se assume que a junção de preferência de pares de dois usuários u e w pode ser decomposta como $BPR(u, w) = BPR(u) BPR(w)$, significando que a $LPP(u)$ independe da $LPP(w)$. A partir desta premissa, podemos estender a LPP para todos os usuários do sistema como

Equação 9 - Calculo do BPR para todos os usuários

$$BPR = \prod_{u \in U} BPR(u).$$

Com a utilização deste método genérico e à aplicação do mesmo em modelos de fatoração de matrizes, verificou-se que os modelos de aprendizagem utilizando **BPR** são melhores do que outros modelos na qualidade de predição, utilizando a métrica de **AUC** (*area under the ROC curve*), que corresponde à habilidade de classificar corretamente itens que são ou não preferência do usuário em dados binários.

2.3. Considerações Finais

Neste capítulo foram apresentados os principais conceitos relacionados aos Sistemas de Recomendação e Sites de Compras Coletivas, bem como a aplicação de técnicas de recomendações em diversos contextos em que o usuário está inserido, seja ele na recomendação de músicas, filmes ou produtos. Estes conceitos e aplicações são essenciais para compreensão do presente trabalho. No próximo capítulo serão apresentados alguns trabalhos recentes do estado da arte, que utilizam técnicas computacionais de recomendação para melhorar a eficiência dos Sistemas de Recomendação em diferentes áreas de aplicação.

3. Trabalhos Relacionados

Este capítulo é destinado à revisão da literatura, no contexto de Sistemas de Recomendação e está dividido em duas partes. A primeira parte apresenta a revisão da literatura de modelagem do negócio em SCC, apresentando os estudos realizados para maximizar o número de compras dos usuários do sistema e alternativas para atração da população. Já a segunda parte apresenta a revisão da literatura do segmento computacional da divulgação das ofertas, que corresponde ao estudo dos Sistemas de Recomendação. Nessa parte há tanto estudos realizados para aumentar a eficácia da recomendação de ofertas para os usuários, analisando a melhoria de se utilizar recomendações personalizadas ao invés de não-personalizadas, como também o desenvolvimento de novos algoritmos que são específicos para recomendar ofertas em SCC.

3.1. Trabalhos relacionados ao modelo de negócio

SCC atraem diversos tipos de usuários, que podem ser categorizados a partir de inúmeros fatores. O número de compras de ofertas que o usuário realizou, indicando a assiduidade deste (usuário muito assíduo, pouco assíduo), a sua idade, região geográfica, renda familiar e outros fatores foram alvos de estudo da comunidade científica que não está relacionada diretamente à computação, mas sim ao modelo de negócios que o SCC representa [6].

Dholakia e Kimes [6] analisaram, a partir da perspectiva do usuário, as ofertas divulgadas em SCC e o motivo da atratividade deste tipo de negócio. Essa atração, segundo os autores, corresponde à percepção de economia no valor das ofertas que já são de seu interesse, o entusiasmo em checar e-mails diariamente com ofertas coletivas e poucos problemas ao utilizar as ofertas antes da data de expiração, bem como o aumento da frequência de compras de usuários que já conhecem o mercado. Os autores ainda acreditavam no crescimento do mercado para os anos posteriores, dado que, na época, o mercado ainda não era consolidado e muitas dúvidas quanto à segurança do negócio, dos dados e a credibilidade da empresa faziam parte do cotidiano dos usuários.

Utilizando dados obtidos por meio do monitoramento de dois dos maiores SCC dos Estados Unidos, Byers *et al.* [9] estudam as características das compras de ofertas, resumindo as cidades que mais tiveram compras de ofertas, os valores, em média, gastos por compras, a disposição das ofertas quanto ao número de dias em que estavam disponíveis, a categoria das ofertas que foram mais adquiridas e outras análises pontuais.

Além disso, os mesmos autores realizaram a modelagem de um algoritmo que prevê o número de aquisições de ofertas, a partir de fatores observados durante a coleta de dados, como: o valor da oferta, o valor do desconto, o número mínimo de compras para que a oferta seja ativada (número mínimo de compras de uma oferta, para que se tenha o desconto promovido), o número máximo de itens da oferta que serão disponibilizados para os usuários e se as ofertas irão ser divulgadas durante múltiplos dias ou não. A partir desse modelo, os pesquisadores conseguiram observar que há alta correlação entre o preço da oferta original e o desconto que é aplicado sobre ela ao realizar a compra em SCC e que o melhor nível do fator que corresponde ao número de dias em que a oferta está disponível é de até dois dias de divulgação. Em suas conclusões, os autores afirmam superestimar as ofertas mais fracas, ou seja, as ofertas que são pouco adquiridas pelos usuários; e acabam por subestimar as ofertas que são muito adquiridas pelos usuários.

Ainda no trabalho de Byers *et al.*, os autores afirmaram que há importância na propagação das ofertas por intermédio das redes sociais, pois há maior aquisição de ofertas quando divulgadas em redes sociais, como o *Facebook*. Seguindo essa linha, o modelo de negócios, envolvendo as estratégias de *marketing* para atrair o público-alvo, foi analisado e até mesmo incorporado ao mercado de SCC por duas das maiores empresas mundiais do ramo, o *GroupOn* e o *Living Social* [7].

Enquanto o *GroupOn* utilizou a estratégia de divulgação de suas ofertas por e-mails e propagandas para que a oferta atinja um número mínimo de compras, a fim de ser ativada, a rede *Living Social* apostou na divulgação das ofertas em redes sociais, em que o usuário adquire a oferta com desconto desde que três amigos comprem também a mesma oferta, a partir de um *link* de página que valida a compra. Ye *et al.* [7] tentaram prever a ativação da oferta, utilizando as duas diferentes estratégias de *marketing*, pela observação comportamental dos usuários que utilizam os dois sites e do tempo em que uma oferta leva para obter sucesso (ser ativada, nesse contexto). Para tanto, utilizaram dados do *GroupOn* obtendo informações significantes quanto às ofertas, como: descrição, desconto, data de início de divulgação, número máximo de compras e número mínimo de compras para que a oferta fosse ativada. Já para obter os dados do *Living Social*, realizaram o desenvolvimento de um *crawler* que visita a página periodicamente com o intuito de observar as ofertas e o crescimento do número de compras com o tempo.

Além disso, os autores buscaram prever um modelo de atenção coletiva para o comportamento de compras. Esse modelo consiste em validar a atração dos consumidores em adquirir ofertas, utilizando a mescla das duas estratégias apresentadas anteriormente. Para isso, verificaram qual das abordagens possui maior eficácia quanto à rapidez em atingir o número mínimo de compras, para que a oferta seja ativada. Em suas conclusões, os autores apontaram que não há diferenças significativas ao utilizar as duas estratégias e que, apesar de as abordagens divergirem em sua ideia principal, convergem para o melhor entendimento do mercado de SCC quanto ao tratamento de ofertar itens e somente permitir a ativação após um número mínimo de compras.

O presente trabalho se beneficia, observando os modelos de negócio propostos anteriormente, por possuir uma base de dados real com o histórico de aquisições, tornando a análise mais fidedigna ao que acontece na realidade, validando alguns dos diversos cenários apresentados, como:

- a previsão do número de compras de uma oferta, que pode ser medida pela métrica *recall*, avaliada no estudo, em que há possibilidade de verificar a quantidade esperada de compras quando há recomendação de ofertas;
- a análise de dados históricos de compras para verificar a relação entre a divulgação das ofertas e a sede na qual a empresa de SCC está instalada;
- a análise da categoria das ofertas que possuem maior popularidade entre os usuários, fazendo com que haja maior confiança de que a oferta será ativada.

Observando ainda a estratégia de divulgação por e-mails, a partir do presente trabalho é possível personalizar esse tipo de divulgação com ofertas que realmente são de interesse do usuário, ao invés de utilizar recomendação não-personalizada, como o próprio SCC *QueroDois*, fornecedor da base de dados, utiliza.

3.2. Sistemas relacionados à recomendação de ofertas

Com o objetivo de auxiliar o usuário a encontrar informações de seu interesse, surgem os Sistemas de Recomendação. Nestes sistemas, ao menos três passos são necessários para que o objetivo seja atingido:

- I. Usuários provêm informações com exemplos de seus interesses ou preferências, que podem ser coletados de maneira direta ou explícita - que ocorre quando o usuário

- avalia os itens, por exemplo - ou de maneira indireta/implícita - que ocorre quando o usuário visita uma página, por exemplo;
- II. Tais exemplos são utilizados para gerar a representação das preferências do usuário, sendo esta representação o perfil do usuário;
 - III. O sistema analisa o perfil do usuário e provê recomendações para este usuário.

Para Resnick *et al.* [20], Sistemas de Recomendação são definidos como sistemas que auxiliam e aumentam a eficácia do processo natural de recomendação "verbal" ou "boca-a-boca" de revistas, músicas, livros, filmes, restaurantes e quaisquer outros tipos de itens que possam ser recomendados, estando incluídas as ofertas em SCC. Nesses sistemas, há possibilidade de recomendar itens de determinadas categorias aos usuários, tomando por base o conhecimento e as utilizações prévias (histórico de navegação, mensagens, **histórico de compras** etc.) de outros usuários do sistema. Em outras palavras, um Sistema de Recomendação é uma tecnologia de filtragem de informação personalizada, utilizada para predição de quais ou quantos itens o usuário possivelmente demonstraria interesse.

Nos últimos anos, Sistemas de Recomendação têm sido usados em diversas aplicações, como mencionadas no Capítulo 2 e, com o advento dos SCC no comércio eletrônico, alguns trabalhos foram difundidos na literatura com o objetivo de analisar e propor novos algoritmos que realizassem a recomendação das ofertas que são divulgadas nos SCC, melhorando o relacionamento e a confiança do usuário nesses sistemas.

Ebrahimi [33] estudou um Sistema de Recomendação sensível ao contexto para a recomendação de ofertas diárias, intitulado *SmarterDeals*. Este sistema utiliza um *framework* especializado em gerência de conteúdo, monitorando as interações do usuário com entidades da web para obter informações relevantes [17]. Ou seja, utilizam um sistema que realiza o monitoramento das páginas que o usuário acessou para realizar a recomendação das ofertas, por meio de análise do conteúdo das ofertas que o usuário previamente adquiriu ou teve intenção de adquirir (inserindo no carrinho de compras, por exemplo), mesclando dados referentes ao passado (histórico de compras) e ao presente (intenção de compras). A partir dos dados obtidos por essa ferramenta, o *SmarterDeals* realiza recomendação de itens utilizando abordagens de recomendação tais como Filtragem Colaborativa e informações pessoais contextuais. O autor utilizou a base de dados real do Yelp Academic Data Set [34] para validar sua proposta, porém transformando a avaliação (*feedback* explícito) de produtos e serviços num modelo de predição de categorias de ofertas. Posteriormente, obtendo dados do

GroupOn, o autor identifica a categoria das ofertas do SCC e recomenda, para cada usuário, ofertas que casam com este perfil.

O presente trabalho se diferencia do trabalho apresentado por conter informações reais dos usuários, por meio de conteúdo fornecido por uma empresa atuante no interior de São Paulo. Ademais, utilizamos apenas os dados que já são disponibilizados pelo usuário, a partir de sua interação com o sistema, sem a necessidade de um *framework* especializado que necessita de autorização do usuário para realizar o monitoramento. De acordo com estudos apresentados anteriormente, a confiança do usuário no fornecimento de dados cadastrais em SCC é um dos fatores primordiais que impedem o crescimento da abordagem de comércio eletrônico, portanto utilizar um *framework* que necessita da autorização do usuário para realizar o monitoramento de atividades pode inviabilizar e enviesar os estudos e conclusões apresentados por esse autor.

Ainda no presente trabalho, utilizamos algoritmos que necessitam apenas dos dados das ofertas diárias e do histórico de compras do usuário, dados estes anônimos e pertencentes à empresa, o que viabiliza o estudo de recomendação por meio do conteúdo das ofertas divulgadas diariamente e o conteúdo do histórico de compras do usuário. Encontramos apenas um trabalho na literatura que realize tal estudo e análise, porém aplicando algoritmos e técnicas mais sofisticadas, sem nenhuma experimentação de outros algoritmos *baseline* em recomendação sensível ao contexto.

Seguindo a linha de recomendação em SCC, o trabalho realizado por Lacerda [32] apresenta o estudo e proposição de um algoritmo baseado em uma sequência de passos para tomada de decisão em cenários de incerteza, denominados *multi-armed bandit*.

Nesses algoritmos, deve-se tomar decisões a cada tempo t , como no cenário em que deve haver decisão entre mandar ou não uma recomendação por e-mail para um usuário. Esses algoritmos se diferenciam de outras abordagens baseadas em aprendizagem por reforço [31].

Os autores propuseram uma forma de obter dados implícitos, a partir do envio de e-mails diários para usuários específicos, sendo esta fase chamada de *exploration*, em que o usuário, ao acessar as ofertas do e-mail, registra dados de acesso naquele momento. Posteriormente, os dados obtidos são utilizados como base para recomendar ofertas personalizadas aos outros usuários que tem perfil próximo àquele na fase de *exploration*, sendo esta etapa chamada de *exploitation*. Apresenta ainda o *trade-off* entre utilizar um grande número de usuários na fase de exploração vs usuários beneficiados com a

recomendação, já que precisam decidir a quantidade de usuários para formar os centros dos *clusters* de similaridade com os outros usuários. Lacerda *et al.* utiliza a base de dados da maior empresa brasileira do segmento de SCC, *Peixe Urbano*, com dados de dois meses do histórico de compras dos usuários.

Assim como o presente trabalho, os autores utilizaram uma base de dados real, porém com menor histórico de compras. A base utilizada no presente trabalho é representada por cerca de 80% dos dados totais da empresa *QueroDois*, do final de 2010 à meados de 2012, possibilitando o acompanhamento do crescimento do mercado junto ao histórico dos usuários, utilizando um conjunto mais complexo de dados, fazendo com que haja complementação do trabalho anterior, que possui apenas um *snapshot* da base da empresa.

Outrossim, a pesquisa atual busca estudar, compreender e comparar os algoritmos estado da arte em recomendação, para que seja possível a avaliação do cenário de recomendação utilizando dados de SCC. Os autores citados anteriormente buscaram realizar a incorporação de novos algoritmos, utilizando novas abordagens de recomendação apesar de, na atual literatura, ainda faltarem estudos mais aprofundados sobre a eficácia e limitações dos algoritmos estado da arte em recomendação no atual domínio. Logo, o presente trabalho difere dos anteriores por complementar os estudos realizados na área de Sistemas de Recomendação para SCC, acrescentando conhecimento acerca dos algoritmos estado da arte em recomendação e também analisando um conjunto de dados diferente dos anteriores.

3.3. Contribuições deste trabalho

Neste capítulo, apresentamos alguns trabalhos na área de Negócios e Sistemas de Recomendação voltados aos SCC que tratam da melhoria dos métodos de recomendação a partir do uso das características dos itens, dos usuários ou mesmo da aplicação de técnicas matemáticas para superação dos problemas oriundos de técnicas já consolidadas na área.

Nesses trabalhos, é notório o esforço para superar técnicas que são consideradas *baseline*, analisando possibilidades para melhoria nas métricas consideradas chave na recomendação. Entretanto, não há estudos apenas de Sistemas de Recomendação, que envolvem matemática, inteligência artificial e recuperação da informação. É possível observar que há estudos para a análise das características comportamentais dos usuários em relação às ofertas (preço, desconto, localidade, regras de utilização, número mínimo de compras) e concernimentos acerca da continuidade do mercado de compras coletivas e também do

público alvo que estivera utilizando os SCC e, mais ainda, do público que ainda poderia ser atraído pelos SCC.

O presente trabalho estuda alguns algoritmos de recomendação considerados *baseline*, aplicando-os ao domínio de compras coletivas. Esse domínio possui diferenças cruciais quando comparados aos outros domínios, como produtos, filmes, livros e músicas, principalmente no tocante à disponibilidade das ofertas no tempo. As principais características encontradas nos SCC estão listadas na Tabela 2, na qual é possível verificar as diferenças entre os dados de SCC e de outros domínios.

Tabela 2 - Diferenças entre dados de Sites de Compras Coletivas e de outros domínios

Quesito de avaliação	Dados de compras coletivas	Dados de outras áreas (filmes, músicas, produtos)
Tipo de dado	<i>Positive-only</i> (indicando somente se a oferta foi ou não comprada)	Dados de feedback explícito (<i>e.g.</i> , avaliações em escalas de 5 estrelas) e <i>feedback</i> implícito (<i>e.g.</i> compra de uma oferta)
Obtenção dos dados	Somente histórico de compras	Histórico de compras, <i>clicks</i> , preferência prévia, <i>tags</i>
Fator temporal	Sim, ofertas são cadastradas e possuem tempo de expiração. Não se pode recomendar ofertas inativas.	Não englobam, pois todos os itens são de livre escolha
Composição dos itens	Ofertas são compostas não só por produtos, mas também serviços. Em alguns casos, possuem mais de um item por oferta.	Itens concretos e/ou digitais, geralmente categorizados.
Regras de compra	Possuem tempo de expiração do cupom; localidade em que deve ser utilizado obrigatoriamente	Itens sem restrições de localidade; não possuem validade
<i>Timing</i> regressivo	A oferta está disponível somente por algum tempo	Os itens estão disponíveis por muito tempo (ou desde o seu lançamento, no caso de músicas e filmes)
Conteúdo dos itens	Poucas informações, como preço, descrição, categoria	Informações completas acerca dos itens.
Esparsidade	Muito alta, devido ao fato dos itens estarem disponíveis por apenas um período	Média/Alta, pois itens estão disponíveis por tempo indeterminado, podendo ser avaliados a qualquer momento

A análise da base real de um SCC é uma contribuição muito importante deste trabalho. A partir do histórico de compras de dois anos da empresa *QueroDois*, foi possível observar alguns comportamentos dos usuários que são importantes para a recomendação e serão mostrados no Capítulo de Metodologia. Essa análise proporcionou a escolha das métricas e fatores a serem estudados, além de observar o motivo pelo qual alguns algoritmos apresentam melhores métricas que outros e como seria possível aumentar a qualidade dessas métricas. Apesar da análise de apenas uma base de dados, é possível estender o presente trabalho para outros SCC, pois a mesma dinâmica de funcionamento nesses domínios é observada, como por exemplo, a divulgação diária de ofertas, a limitação de tempo e quantidade de usuários mínimo que compram ofertas e outros.

Ainda que trabalhos como o de Lacerda [32] e Ebrahimi [33] aprofundem o conhecimento acerca dos SCC e Sistemas de Recomendação, há poucos trabalhos que foquem nos Sistemas de Recomendação, na análise de algoritmos *baseline*, predição de compras a partir do perfil do usuário e análise dos dados dos SCC, identificando possíveis carências dos algoritmos estado da arte em recomendação aplicados nesses sistemas. Além disso, poucos trabalhos possuem base de dados de um longo período em um sistema real, limitando a generalização das conclusões.

4. Metodologia

Este capítulo descreve a metodologia utilizada no presente trabalho, que é dividida em duas partes. A primeira consiste na análise dos dados de SCC, bem como a motivação em relação à recomendação de ofertas e os fatores e níveis utilizados na seção seguinte. A segunda parte consiste na descrição geral da metodologia aplicada, bem como na definição dos algoritmos, fatores, níveis e hipóteses a serem estudadas neste trabalho.

A estratégia de investigação é uma pesquisa do tipo experimental, meio pelo qual testaremos hipóteses que dizem respeito à acurácia da recomendação, utilizando as métricas de precisão, recuperação e *mean average precision (MAP)* como variáveis de resposta.

A partir da definição de quatro fatores e motivados pela análise da interação dos usuários e ofertas em SCC, definimos a metodologia a ser utilizada, as métricas e ferramentas que nos ajudam a melhor compreender os SCC e suas especificidades.

O experimento foi executado uma vez para cada nível dos fatores definidos nas próximas seções. Os resultados são baseados nos valores de precisão, recuperação e *MAP*, que foram coletados e armazenados, para posterior análise durante o este trabalho.

4.1. Base de Dados

A base de dados utilizada no presente trabalho foi cedida pela empresa de compras coletivas *QueroDois*, situada na cidade de Ribeirão Preto - São Paulo, em parceria com a Universidade Federal de Campina Grande, considerando o período dos dados registrados entre meados de 2010 e meados de 2012.

A base de dados é formada por dois conjuntos. O primeiro corresponde às características das ofertas que estavam em divulgação. Nesses dados, as ofertas são descritas por um identificador único da oferta (valor numérico), a descrição textual da oferta, a cidade em que a oferta pode ser utilizada, o preço da oferta e a data em que houve o cadastro da oferta no sistema. Já o segundo conjunto corresponde aos pedidos que foram efetuados, sendo estes caracterizados por um identificador único de venda (valor numérico), identificador único de usuário (valor numérico), identificador único da oferta (valor numérico), a data em que a aquisição foi realizada e a cidade de cadastro do usuário.

Para o período considerado, a base possui cerca de 117 mil vendas, tal que 43 mil usuários realizaram compras no site e ainda um total de 1.537 ofertas estavam à disposição do usuário no site, com *esparsidade* de 95,80%. A alta *esparsidade* indica que grande quantidade valores nulos da matriz de associação entre usuários e ofertas. A partir da análise desses dados foi possível identificar alguns padrões de comportamento da interação de compra entre os usuários e as ofertas.

Quanto aos usuários, verificou-se que, do total de usuários, cerca de 6 mil realizaram compras de 5 ou mais ofertas, enquanto 1.700 usuários realizaram compras de 10 ou mais.

Considerando a interação dos usuários e ofertas, verificou-se que, apesar do SCC divulgar ofertas nacionais, a maior parte das aquisições partiu de ofertas divulgadas na região de Ribeirão Preto, Franca e Presidente Prudente, devido ao fator de localização física da empresa estar diretamente interligado com os empresários que cadastram seus produtos e serviços no site, como pode ser verificado na Figura 7.

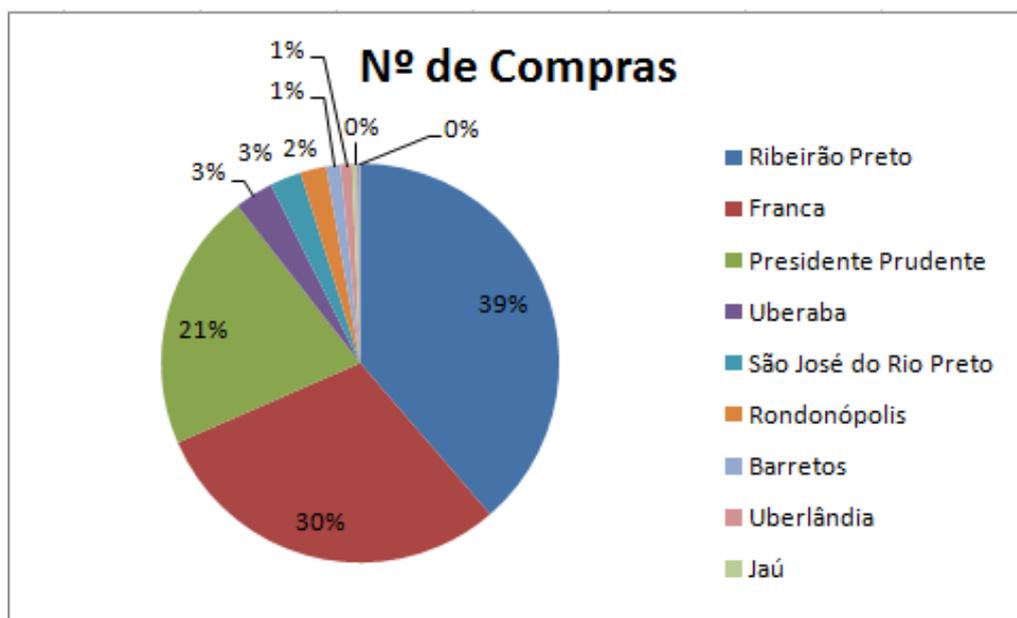
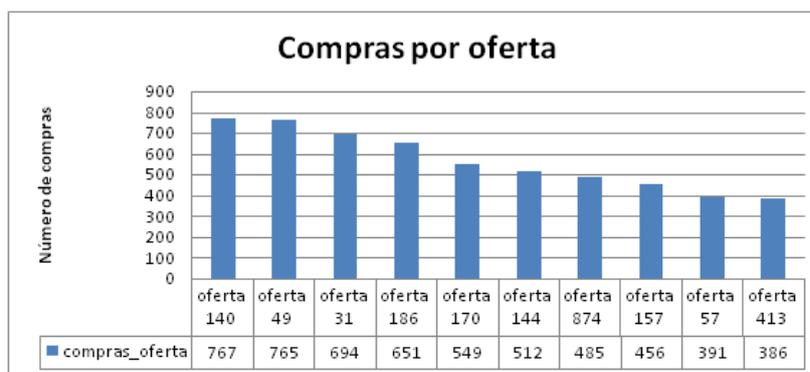


Figura 7 - Distribuição de compras por cidade.

Ainda foi possível observar que o número de compras de ofertas da categoria "Restaurante / Gastronomia" é predominante. Na análise das 10 ofertas que mais foram compradas (Figura 8), 9 são referentes a esta e apenas 1 referente a "Serviços Automotivos". Quando aumentamos o número de ofertas que mais foram compradas para as 100, 85 são da categoria "Restaurante / Gastronomia".

Quanto às ofertas, é possível observar que o número médio de divulgações diárias é de 23 ofertas. O número médio de compras é de 300 aquisições diárias. Foi possível ainda extrair o número de ofertas que foram anunciadas no site de compras coletivas e estas variam entre 10 e 33 anúncios de produtos ou serviços divulgados diariamente.



Oferta 31 - (24 hs) A famosa Bloomin' Onion do Outback 50% off! De R\$ 24,75 por R\$ 12,35!

Oferta 49 - Happy Hour Com Classe! 80% OFF em Porção de Calabresa Acebolada + 3 Chopes Heineken no

Classe A Gourmet da João Fiusa! De R\$ 30,60 por R\$ 5,95!

Oferta 57 - (Limite 500 cupons) Seu Carro Novinho! 60% OFF em Lavagem Completa (Interna, Externa, Por Baixo, Aspiração e Motor) + Cera 3M + Tornador (Limpeza de Painéis, Partes Plásticas e Trilhos dos Bancos)

Oferta 140 - O Passado Foi em Dobro, Agora é Quase de Graça! 77% OFF no Double Cheese Burger com Batata Palha no Betts Burger.

Oferta 144 - We Love Pizza! 50% OFF em Entrada com Porção de Casquetas + Pizza Grande (escolher entre 10 sabores) + 1 Jarra de Suco

OU 1 Cerveja OU 2 Refrigerantes Lata na Pappagone Pizzaria!

Oferta 157 - A Sofisticação Árabe! 50% OFF em Combinado Executivo com Kafta + Arroz com Lentilha + Esfirra de Carne Pequena + Mix de Folhas Verdes + Porção de Coalhada Seca, Pão Árabe e Torradas no Mabruk!

Oferta 170 - Ela está de volta! 50% OFF na Inigualável e Deliciosa Bloomin' Onion do Outback!

Oferta 186 - A Água Doce faz 15 Anos e Quem Ganha o Presente é Você! 60% OFF no Mais Famoso e Copiado Escondidinho de Carne Seca Grande + Saquerita de Frutas

Oferta 413 - Aproveite e Faça a Festa com 50% OFF em 100 Salgadinhos para Festa, Sendo 30 Coxinhas, 40 Bolinhas de Queijo e 30 Kibes Fritos

Oferta 874 - Mané Gula Apresenta: 60% OFF em Rodízio de Pizza + Rodízio Gastronômico + Coca Cola, Guaraná Antarctica

Figura 8 - Ofertas mais compradas.

Ainda foi possível verificar que, para a base de dados utilizada, há um padrão de horário em que as compras são realizadas. Com a divulgação das ofertas iniciando as 5 horas da manhã, há dois horários em que há picos de aquisições: as 8 da manhã - 3 horas após o início da divulgação - e às 16 horas - 9 horas após o início da divulgação. A Figura 9 mostra o agrupamento de todas as aquisições realizadas no site, agrupando os dados pelo horário de compra.

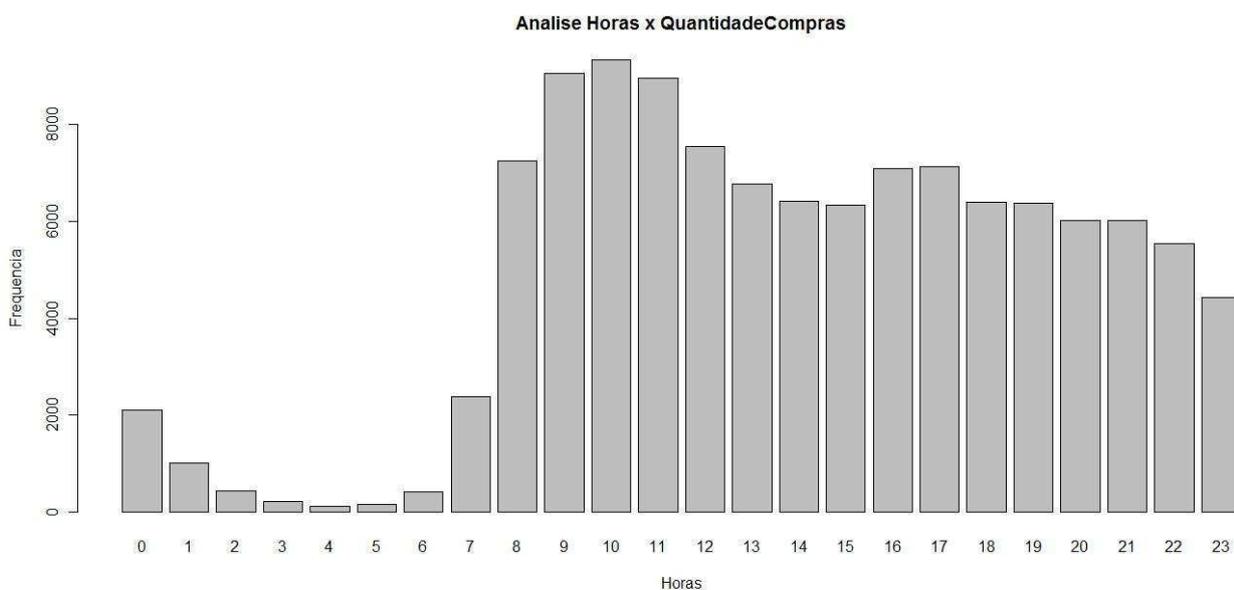


Figura 9 - Distribuição das aquisições das ofertas por hora.

A partir dos horários de pico de aquisição das ofertas, verificou-se que a quantidade de usuários que efetuam compras no primeiro horário de pico - entre 7 e 12 da manhã - representa um valor bastante semelhante se comparado ao segundo horário de pico - entre 15 e 17 horas. Também foi possível verificar que, para a base de dados da empresa, é bastante comum que haja concentração de compras no horário da tarde (entre as 13 e 18 horas)/noite (entre as 19 e 24 horas) em detrimento ao horário da manhã (entre as 7 e 12 horas), como pode ser visto no próprio gráfico.

A partir dos intervalos de confiança verificamos que o percentual de aquisições entre 8 horas e 15 horas é próximo ao percentual de aquisições de 15 horas até as 23 horas e que esses dois horários são extremamente superiores ao intervalo da manhã. Os intervalos de confiança foram calculados a partir do $\alpha = 0.1$, ou seja, podemos afirmar que os valores das **médias de aquisições diárias** são os mesmos observados nos períodos das Figuras Figura 10 e Figura 11, com 90% de certeza. O eixo X corresponde ao intervalo de horas durante um dia, observando que a divulgação se inicia às 5 horas da manhã, diariamente. Logo, consideramos os períodos até 3 horas de divulgação - compras realizadas até as 8 horas da manhã -, até 10 horas de divulgação - compras realizadas até as 15 horas da tarde - e após 10 horas de divulgação - compras realizadas após as 15 horas, considerando apenas usuários que efetuaram 5 ou mais aquisições e apenas usuários que efetuaram 10 ou mais aquisições. Já o eixo Y indica, em porcentagem, a quantidade de compras que foram realizadas durante os dias da amostra analisada, valor correspondente ao conjunto de dias amostrais de divulgação, escolhidos

aleatoriamente, com cerca de 5.600 eventos de compras de ofertas, em que apenas usuários que realizaram 5 ou mais compras, e por volta de 2.700 eventos de compras de ofertas considerando usuários com 10 ou mais compras no site.

Logo, a partir das Figuras Figura 10 e Figura 11, podemos concluir que há horários de picos do número de compras ao longo do dia e este atinge seu ápice no período de considerando até 10 horas do início da divulgação das ofertas. Além do mais, é possível observar que após às 10 horas de início de divulgação das ofertas, o número de usuários que compraram ofertas é bem próximo ao número de usuários que efetuaram compras até 10 horas de início de divulgação das ofertas, sendo este um diferencial importante na recomendação em SCC, pois a oferta é válida apenas durante um dia e é necessário um histórico de aquisições para que a recomendação personalizada seja eficaz.

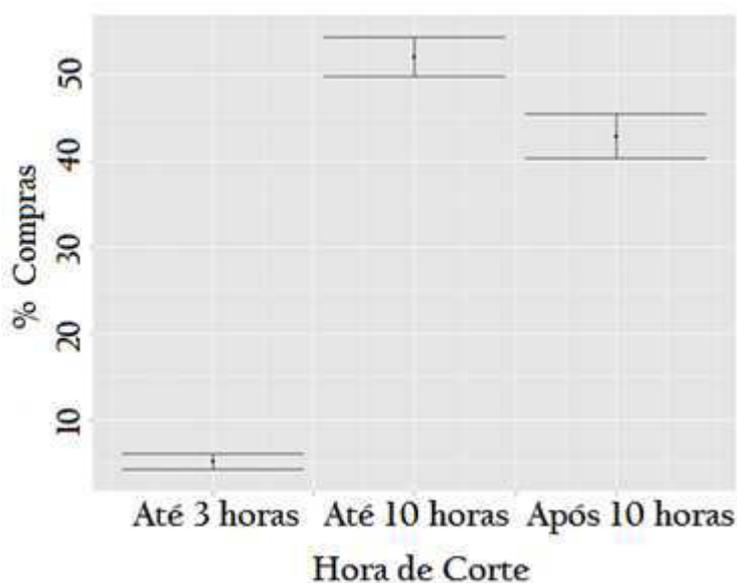


Figura 10 - Intervalo de confiança considerando ofertas e horários de pico para usuários com 5 ou mais compras

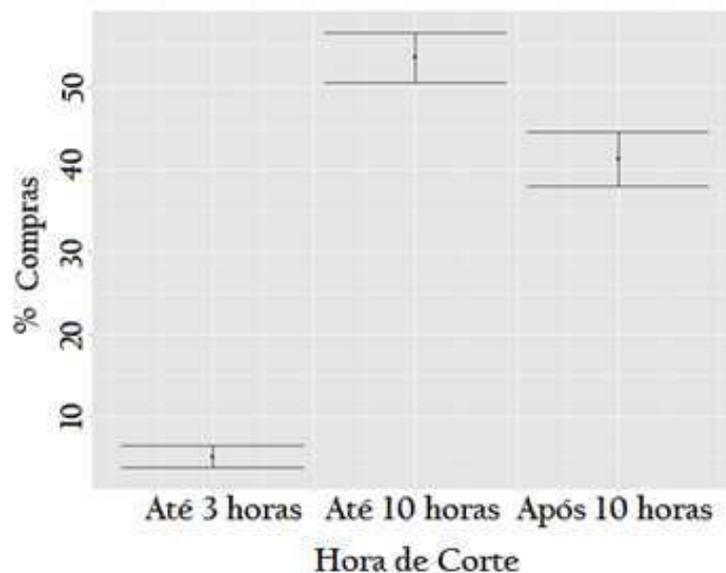


Figura 11 - Intervalo de confiança considerando ofertas e horários de pico para usuários com 10 ou mais compras

Portanto, a partir da análise descritiva apresentada, definimos que o horário de recomendação e o número de compras mínimo realizado pelo usuário são fatores relevantes para nosso experimento. O primeiro é relevante pois avaliamos que há horários de picos na atividade de compras de ofertas. O segundo porque há uma grande quantidade de usuários pouco ativos (com poucas compras realizadas) e para os quais provavelmente há pouca informação para realizar as recomendações.

4.2. Fatores e Projeto do Experimento

Para realizar o experimento, inicialmente houve a definição de quatro fatores essenciais, sendo estes:

1. Fator algoritmo: possui quatro níveis, indicando os algoritmos utilizados para realizar a recomendação.
 - 1.1. BPRMF;
 - 1.2. ItemkNN;
 - 1.3. Mais popular;
 - 1.4. Query & Search;
2. Fator hora de corte: possui três níveis, indicando o horário em que a recomendação seria efetuada a partir do Sistema de Recomendação. É importante destacar que as ofertas são válidas apenas durante o dia em que foi anunciada e não há possibilidade de aquisição após o término do dia. Além disso, para o algoritmo que utiliza recomendação baseada em

conteúdo (**Query & Search**) não iremos avaliar este fator, pois de acordo com a mecânica do algoritmo utilizado, não existe nenhuma variação nos resultados quando os níveis dos fatores são variados*. Os níveis são:

2.1. 2 horas após o início da divulgação da oferta;

2.2. 3 horas após o início da divulgação da oferta;

2.3. 10 horas após o início da divulgação da oferta.

3. Fator número de itens a ser recomendados: possui três níveis, indicando a quantidade de itens, no máximo, que serão recomendados aos usuários.

3.1. 3 itens recomendados;

3.2. 4 itens recomendados;

3.3. 5 itens recomendados.

4. Fator número mínimo de compras por usuário: possui dois níveis, indicando a quantidade mínima de compras que o usuário deve ter adquirido antes da recomendação para que os itens possam ser recomendados.

4.1. Ao menos 5 compras efetuadas pelo usuário;

4.2. Ao menos 10 compras efetuadas pelo usuário.

Como variáveis de resposta, que correspondem às variáveis nas quais os efeitos são observados no processo de experimentação, escolheu-se as métricas de precisão, *recall* e *MAP*, utilizadas para avaliar o desempenho dos algoritmos de recomendação em um determinado contexto quando há necessidade de recomendação de itens, ao invés de predição de notas. Com a precisão, espera-se mensurar a quantidade de ofertas relevantes que são recomendadas para os usuários, em detrimento a quantidade de ofertas irrelevantes. Já a partir do *recall*, espera-se mensurar a quantidade, partindo da maioria das ofertas relevantes, de ofertas que são recomendadas. Por último, a partir da métrica *MAP*, espera-se que, em média, a recomendação de um item na posição *i* da lista de recomendação de cada usuário represente a ordem de relevância da oferta para o mesmo.

O design do experimento é fatorial completo, em que cada nível de um fator é combinado a todos os níveis dos demais fatores, possibilitando a análise dos efeitos nas variáveis de resposta. Portanto, 60 repetições para cada um dos 50 dias amostrais escolhidos foram realizadas. O valor de dias amostrais foi escolhido por incluir a maior variabilidade da amostra com relação ao número de ofertas divulgadas em um dia específico, e o número de compras efetuadas. Mais detalhes sobre a combinação dos níveis dos fatores podem ser observados no Apêndice B.

4.3. Objetivos

Baseando-se no problema de recomendação utilizando dados de SCC, foco da análise, algumas perguntas seguem abaixo, considerando as métricas apontadas anteriormente.

Pergunta 1: Variando cada nível do fator algoritmo definido anteriormente, é possível afirmar que há diferença significativa nos valores das métricas observadas nos experimentos, quando utilizamos dados de compras coletivas?

Pergunta 2: Variando o fator hora de corte definido anteriormente, é possível afirmar que há diferença significativa nos valores das métricas observadas nos experimentos, quando utilizamos dados de compras coletivas?

Pergunta 3: Variando o fator número de itens a ser recomendados definido anteriormente, é possível afirmar que há diferença significativa nos valores das métricas observadas nos experimentos, quando utilizamos dados de compras coletivas?

Pergunta 4: Variando o fator número mínimo de compras, é possível afirmar que há diferença significativa nos valores das métricas observadas nos experimentos, quando utilizamos dados de compras coletivas?

A partir das perguntas formuladas, nortearmos a pesquisa em analisar e responder cada uma delas de acordo com os experimentos na seção seguinte.

4.4. Execução do experimento

4.4.1. Escolha dos dias amostrais

De acordo com a base de dados utilizada neste trabalho, cerca de 500 dias formam a população total considerando o período entre 2010 e 2012. Para o presente estudo, escolhemos apenas 10% para amostragem. As datas foram escolhidas de forma aleatória, excluindo o ano de 2010, pois de outra forma não haveria dados suficientes após realizar a filtragem de usuários a partir do número mínimo de aquisições.

4.4.2. Montagem dos arquivos de treino, teste e arquivo auxiliar

Utilizando a linguagem de programação *Java* e o IDE *eclipse*, além do banco de dados *MySQL*, os arquivos de treino e teste foram formados a partir da população total (arquivo completo do base de dados), filtrando e isolando os usuários que realizaram 5 ou mais

aquisições de ofertas e os usuários que realizaram 10 ou mais aquisições, a partir da tupla de aquisição <id_usuario, id_oferta>, como exemplificado na figura abaixo.



Figura 12 - Esquema do Banco de Dados

Os dois valores foram escolhidos por possuírem quantidade de dados significativa para a recomendação, tanto em histórico de aquisição de ofertas quanto em número de usuários utilizados para validação dos algoritmos de filtragem colaborativa. De outra forma, os resultados seriam pouco significativos, pois o menor número de compra de ofertas acarretaria em maior esparsidade e maior número de compra de ofertas acarretaria em poucos usuários para treinamento dos algoritmos de filtragem colaborativa. Ainda de acordo com esse fator, Dholakia *et al.* [6] definem usuários com esta característica como usuários experientes, ou seja, usuários que utilizam os SCC com frequência, portanto focamos a recomendação para esses usuários

Após esse processo, houve a real divisão dos arquivos de treino e teste utilizados no trabalho, a partir da filtragem das compras realizadas em um determinado dia e anterior ao horário de corte definido. Logo, a partir de uma data específica, as tuplas de compra foram separadas em dados de treino (se ocorrerem na data definida e anterior ou igual ao horário de corte) e teste (se ocorrerem na data definida e posterior ao horário de corte). Essa configuração para os dados de testes foi utilizada pois a maioria dos itens de uma data não estão presentes nas datas posteriores. Portanto, apenas a data atual deve ser levada em consideração ao realizar os testes.

Utilizamos ainda um arquivo auxiliar para conter apenas as ofertas que são válidas no dia definido, para que houvesse recomendação considerando apenas as ofertas que estavam presentes nesse. O arquivo é composto apenas por um valor numérico, <id_oferta_válida>. Chamamos esse arquivo de arquivo de itens candidatos. A Figura 13 exemplifica esse cenário.



Figura 13 - Divisão da base em treino, teste e arquivo auxiliar.

Para os arquivos de treino e teste, a partir do conteúdo das ofertas (a descrição da oferta propriamente dita), utilizamos o mesmo processo, porém com o auxílio da base de dados para identificar as descrições das ofertas válidas no dia atual e a descrição das ofertas que os usuários adquiriram previamente.

4.4.3. MyMediaLite: algoritmos, parâmetros e obtenção das métricas

A biblioteca do *MyMediaLite* possui diversos algoritmos para formação de uma ferramenta de recomendação, apesar de não realizá-la, gerando apenas as saídas da aplicação dos algoritmos em determinada base de dados. Utilizamos a biblioteca, com foco na recomendação de itens, para gerar os arquivos de saída que contêm o identificador de usuário ou item (dependente do algoritmo utilizado) e uma lista de usuários ou itens que formam a lista de recomendações (também dependente do algoritmo).

No presente trabalho, utilizamos o algoritmo **Mais Popular**, chamado na ferramenta como *MostPopular*, adicionado aos parâmetros de número de predição de itens (*--predict-items-number*) igual a 3, 4 ou 5 (de acordo com os níveis do fator definido anteriormente), o arquivo de itens candidatos e também a lista das métricas que devem ser expurgadas ao final do processamento (*--measures*).

Já no algoritmo de fatoração de matrizes **BPRMF**, utilizamos os mesmos parâmetros do algoritmo **Mais Popular**, adicionado apenas os parâmetros opcionais de recomendação (--*recommender-options*), com valor igual a 30, indicando o número de fatores latentes por usuário, e o número de iterações igual a 60. Estes valores foram escolhidos via tentativa e erro, nos quais os resultados foram piores com valores menores e foram semelhantes ou piores com valores maiores.

No algoritmo **ItemkNN**, os mesmos parâmetros do algoritmo **BPRMF** foram utilizados, porém com alteração apenas nos parâmetros opcionais da recomendação (--*recommender-options*), no qual foi informado o valor $k=30$, indicando o número máximo de vizinhos a serem conectados. Conforme exposto anteriormente, esse valor foi escolhido por apresentar melhores resultados, já que utilizando valores menores, apresentou piores resultados e com os valores maiores, apresentou resultados semelhantes ou piores.

4.4.4. Lucene e implementação de algoritmo de recomendação baseado em conteúdo

Utilizando a biblioteca *Lucene*, houve o desenvolvimento de um Sistema de Recomendação simples, baseado no conteúdo das ofertas. Cada oferta possui um campo de texto denominado *descrição*, que corresponde a descrição completa da oferta (exemplo no Apêndice A).

A partir das aquisições de cada usuário e utilizando as entidades de um Sistema de Recuperação da Informação, foi possível implementar o sistema. Considerando que cada oferta possui uma descrição textual, modelamos o usuário como um documento que possui todas as descrições das ofertas adquiridas até um certo intervalo de tempo (considerando a data em que há recomendação e a hora de corte). Logo, uma oferta será recomendada para um usuário se a mesma possuir similaridade com o conjunto das ofertas previamente adquiridas, ou seja, se há similaridade entre o documento completo e a oferta. A similaridade utilizada é baseada no cosseno, conforme definida no capítulo de Fundamentação Teórica.

4.5. Considerações Finais

Neste capítulo definimos a metodologia do trabalho - avaliação de algoritmos de recomendação a partir de dados de SCC. Inicialmente, apresentamos uma visão geral da base de dados utilizada. Na sequência, descrevemos os fatores e níveis que serão levados em consideração neste trabalho e o design de experimento formulado para o mesmo; definimos as hipóteses e os objetivos e, finalmente, descrevemos como se dá a execução do experimento, a partir da escolha dos dias amostrais, a montagem dos arquivos de treino, teste e arquivo de

itens candidatos, a ferramenta utilizada e os parâmetros de execução e o Sistema de Recomendação baseado no conteúdo, desenvolvido para avaliação juntamente com os demais.

5. Resultados

Neste capítulo apresentaremos os resultados dos experimentos realizados, de acordo com a abordagem proposta no Capítulo de Metodologia. Os experimentos foram realizados com o objetivo de mensurar a eficiência dos algoritmos definidos anteriormente, levando em consideração os dados de SCC e analisando as métricas precisão, *recall* e *MAP*, identificando os pontos positivos e negativos de cada um dos níveis dos fatores propostos.

Realizamos um experimento adicional para verificar se há melhoria nas métricas quando agrupamos 4 dias de divulgação consecutivos, ao invés de considerar apenas um dia de divulgação das ofertas, simulando o cenário que foi observado no trabalho de Lacerda [32].

5.1. Experimento inicial: ofertas válidas por 24 horas

Conforme definido no Capítulo de Metodologia, foram realizadas 50 execuções (correspondentes aos 50 dias escolhidos aleatoriamente) para cada algoritmo avaliado, considerando todos os níveis dos fatores hora de corte, número de itens a ser recomendados e número mínimo de compras por usuário. As execuções representam diferentes dias de divulgação das ofertas, de acordo com a base de dados fornecida para este trabalho. O conjunto de experimentos podem ser vistos no Apêndice B, no qual se encontra a tabela com a combinação dos fatores, bem como os dias e o número de ofertas divulgadas que foram escolhidas para cada execução.

5.1.1. Análise para o fator "algoritmo"

A avaliação experimental mostrou que, fixando-se o fator algoritmo, a métrica *MAP* apresentou os melhores resultados para o algoritmo **BPRMF**. De acordo com a Figura 14, o intervalo de confiança com $\alpha = 0,1$ (ou seja, 90% de confiança que o intervalo contém o verdadeiro valor da média populacional) mostra que a métrica *MAP* é superior, em todos os casos, no algoritmo **BPRMF**, enquanto nada podemos afirmar quando são comparados os algoritmos **Mais Popular** e **ItemkNN**. Os piores resultados são observados no algoritmo baseado em conteúdo **Query & Search**. Os valores observados estão no intervalo de 0 a 1, no qual a proximidade de 0 significa que a recomendação de um item i na posição k do ranking não é relevante para o usuário e a proximidade de 1 significa que a recomendação de um item i na posição k do ranking é muito relevante.

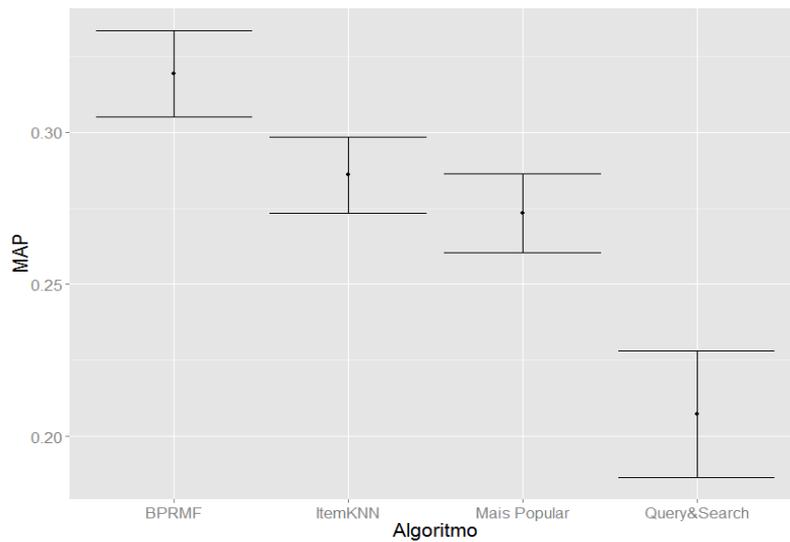


Figura 14 - Intervalo de confiança para a métrica *MAP* observada.

Quanto às métricas de **precisão** e *recall*, não foi possível identificar um melhor algoritmo, porém foi possível identificar que o algoritmo baseado em conteúdo apresenta os piores resultados, de acordo com o intervalo de confiança com nível de confiança igual a 90%, como é possível observar na Figura 15. Os valores observados estão no intervalo de 0 a 1. Para a precisão, quanto mais próximo de 0 o valor estiver, menor será a **quantidade** de ofertas relevantes recomendadas do conjunto de ofertas divulgadas; e quanto mais próximo de 1 for o valor, maior será a **quantidade** de ofertas relevantes recomendadas, **em detrimento às ofertas irrelevantes**. Já para o *recall*, o valor mais próximo a 1 indica que a maioria das ofertas relevantes será recomendada para o usuário, dentre as ofertas que realmente são relevantes para este; caso contrário, a minoria das ofertas relevantes será recomendada.

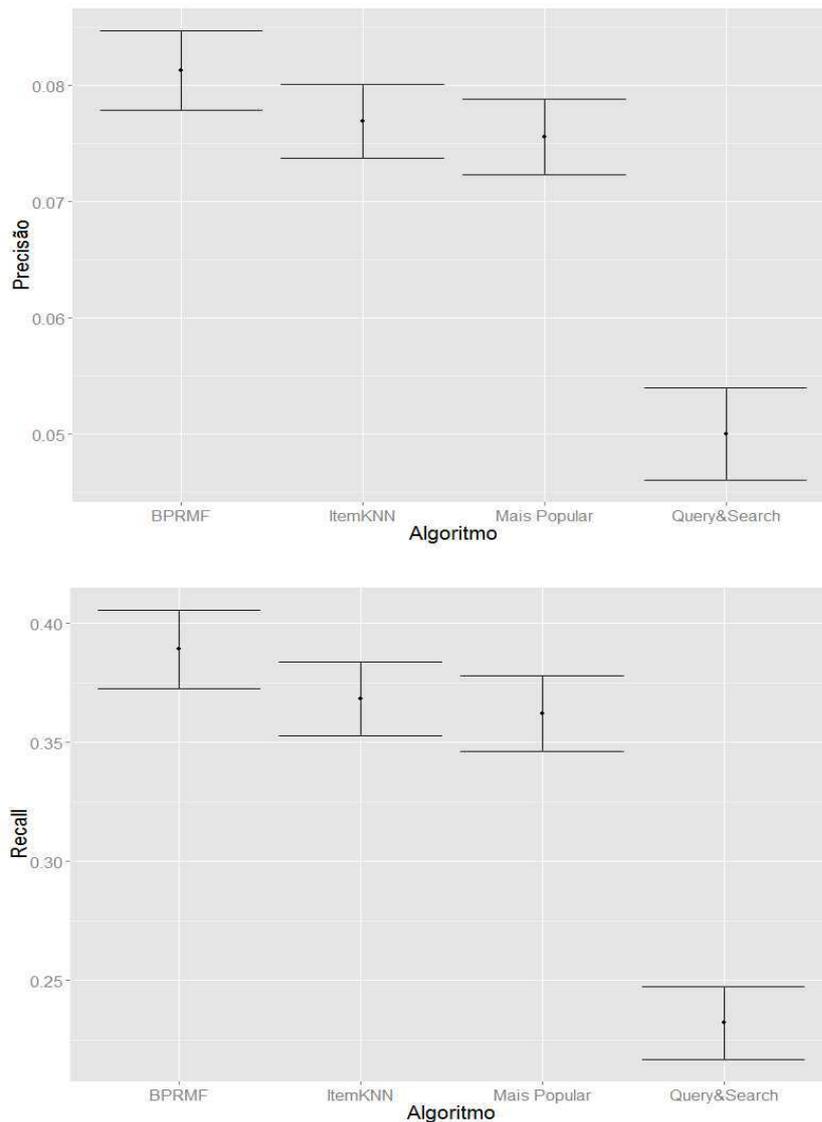


Figura 15 - Intervalos de confiança para as métricas precisão e recall.

Com o propósito de verificar qual dos algoritmos possui o melhor indicador das métricas de precisão e *recall*, fez-se necessário a realização de um testes estatísticos para determinar se há diferença nos resultados observados. Para tal propósito, utilizaremos o teste de **Wilcoxon**, de acordo com a Tabela 3. Utilizamos esse teste em específico pois é um teste não-paramétrico que calcula a diferença entre todos os valores obtidos, onde há possibilidade de observação de três condições: aumento (+), diminuição (-) ou igualdade (=) de valores. É considerado um teste não-paramétrico dos mais poderosos e populares, por considerar o valor das diferenças entre os pares ao invés de considerar apenas o sinal das diferenças entre os pares [40].

Tabela 3 - Teste de Wilcoxon.

Teste	p-value	Tipo de teste
Wilcoxon	Se $> 0,05$, aceitar H_0 ; caso contrário, aceitar H_1 .	Não-paramétrico com avaliação de equivalência entre níveis de fatores

O teste de **Wilcoxon** responde às seguintes hipóteses:

H_0 : A métrica em questão é, em média, igual para todos os algoritmos analisados;

H_1 : A métrica em questão não é, em média, igual para todos os algoritmos analisados.

Utilizando o teste estatístico definido na Tabela 3, verificou-se que, para as métricas de **precisão** e **recall**, os algoritmos **BPRMF**, **ItemKNN** e **Mais Popular** apresentam valores semelhantes e não há confirmação estatística que determine a diferença entre esses. Não há verificação para o algoritmo baseado em conteúdo pois este não possui intervalo de confiança sobreposto aos demais.

Para a precisão, o *p-valor* foi de 0,1486. Já para o **recall**, o *p-valor* foi de 0,1625. Em ambos os casos, aceitamos a hipótese nula e confirmamos que, a nível estatístico, não há diferença na precisão e **recall** para os algoritmos considerados.

Avaliando o número de vezes em que cada algoritmo foi **superior a todos os outros**, por comparações pareadas, podemos verificar que o **BPRMF** apresentou, em cerca de 55% das vezes o melhor valor para as métricas de **precisão** e **recall**, como pode ser visto na Figura 16.

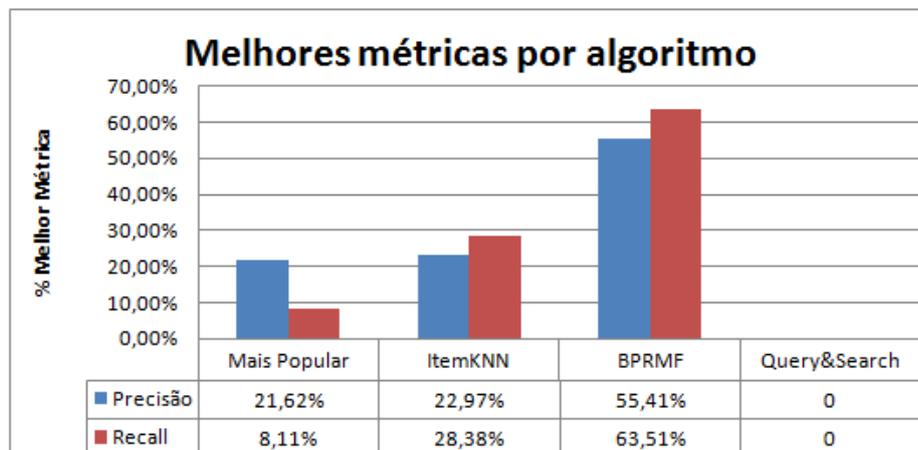


Figura 16 - Comparação entre os algoritmos em relação a precisão e **recall**.

Logo, ao se utilizar a média amostral, nada se pode afirmar quanto à diferença estatística dos resultados.

5.1.2. Análise para o fator "hora de corte"

A avaliação experimental mostrou que, fixando-se o fator hora de corte, as métricas de **MAP**, **precisão** e **recall** apresentaram os melhores resultados quando consideramos o nível "10 horas após o início da divulgação das ofertas". De acordo com a Figura 17, o intervalo de confiança com $\alpha = 0,1$ mostra que a métrica **MAP** é superior, considerando todos os algoritmos, nos casos em que o nível do fator é o terceiro nível. As Figura 18 e Figura 19 mostram a mesma situação, porém considerando as métricas de **precisão** e **recall**, respectivamente. Os piores resultados são observados no primeiro nível, no qual há tentativa de recomendação 2 horas após o início de divulgação das ofertas.

O algoritmo baseado em conteúdo **Query & Search** não foi avaliado para os níveis desse fator, porque de acordo com a mecânica do algoritmo, não existe nenhuma alteração quando os níveis do fator são alterados. Na mecânica do algoritmo, apenas o perfil do usuário é avaliado, de acordo com a descrição das ofertas que foram adquiridas até o momento de corte. Logo, a variação do histórico de compras de outros usuários é irrelevante para um usuário específico.

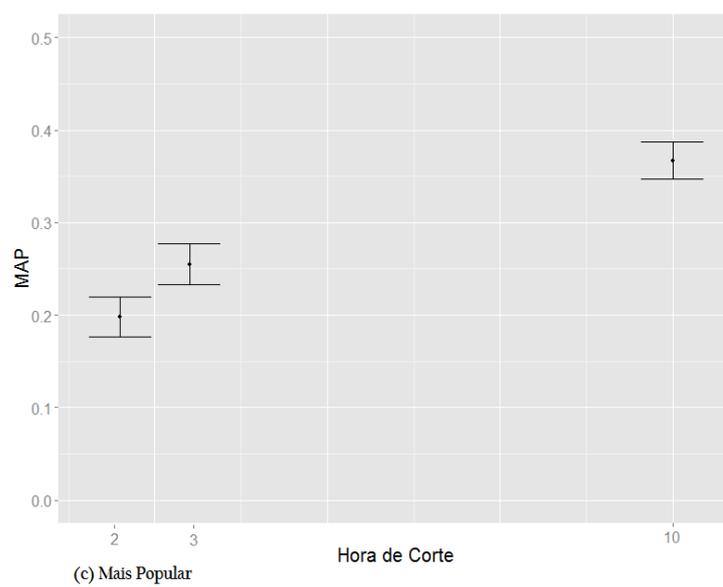
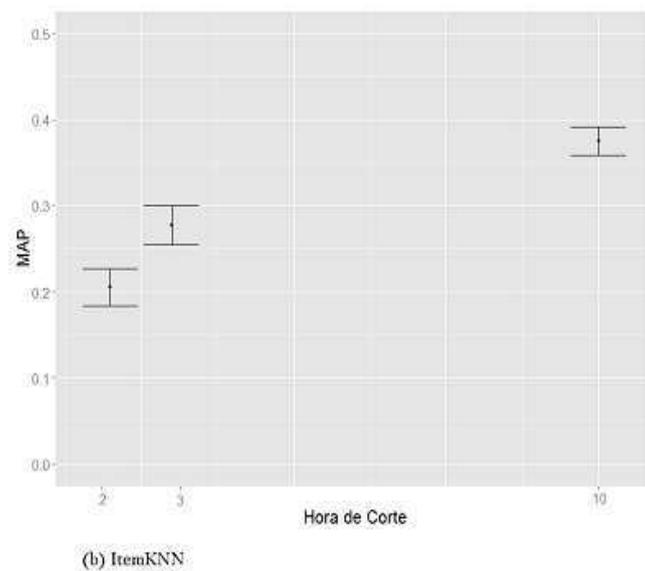
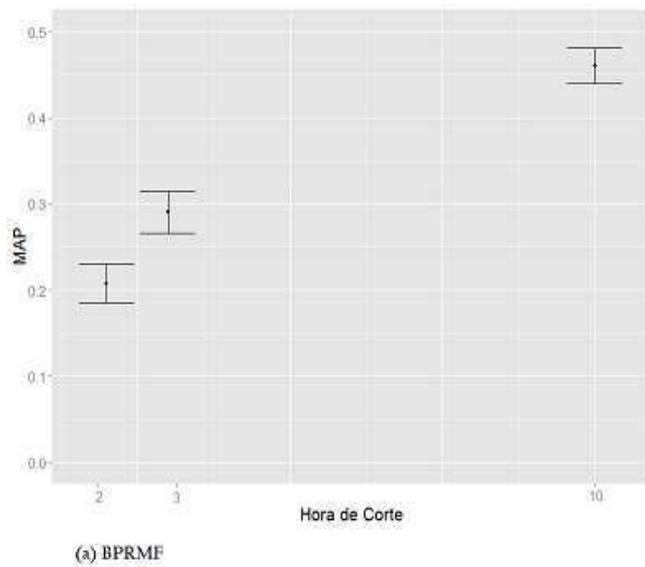
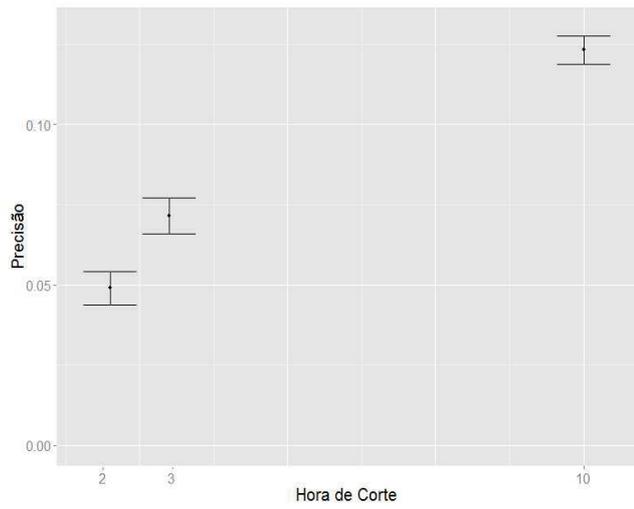
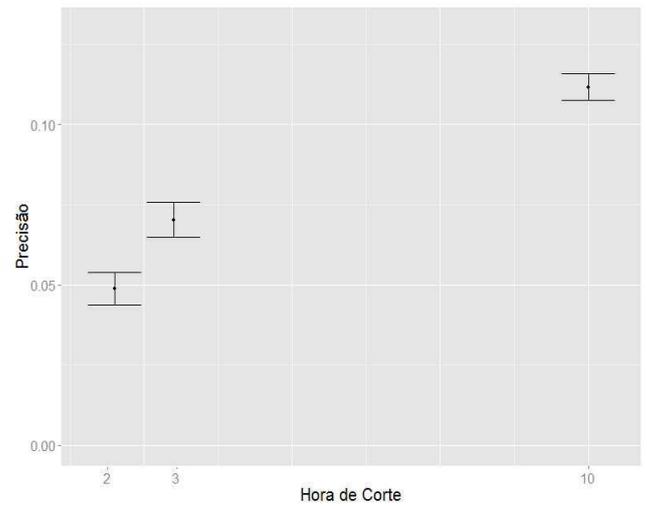


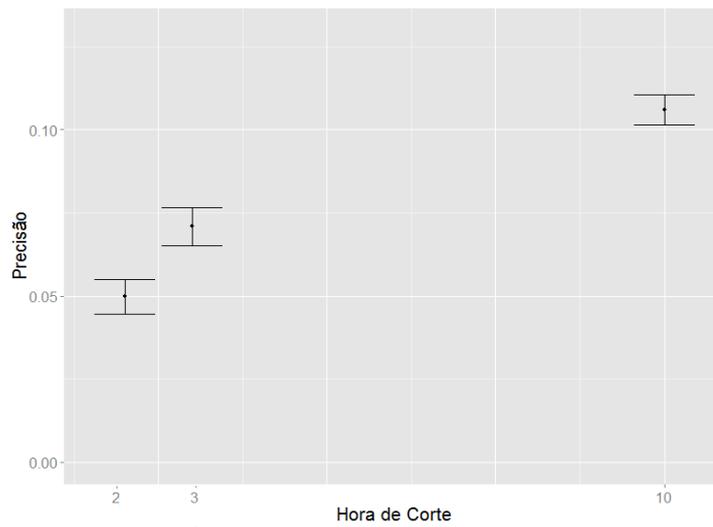
Figura 17 - Intervalos de confiança, considerando os níveis do fator "hora de corte", para a métrica *MAP*.



(a) BPRMF



(b) ItemKNN



(c) Mais Popular

Figura 18 - Intervalos de confiança, considerando os níveis do fator "hora de corte", para a métrica precisão.

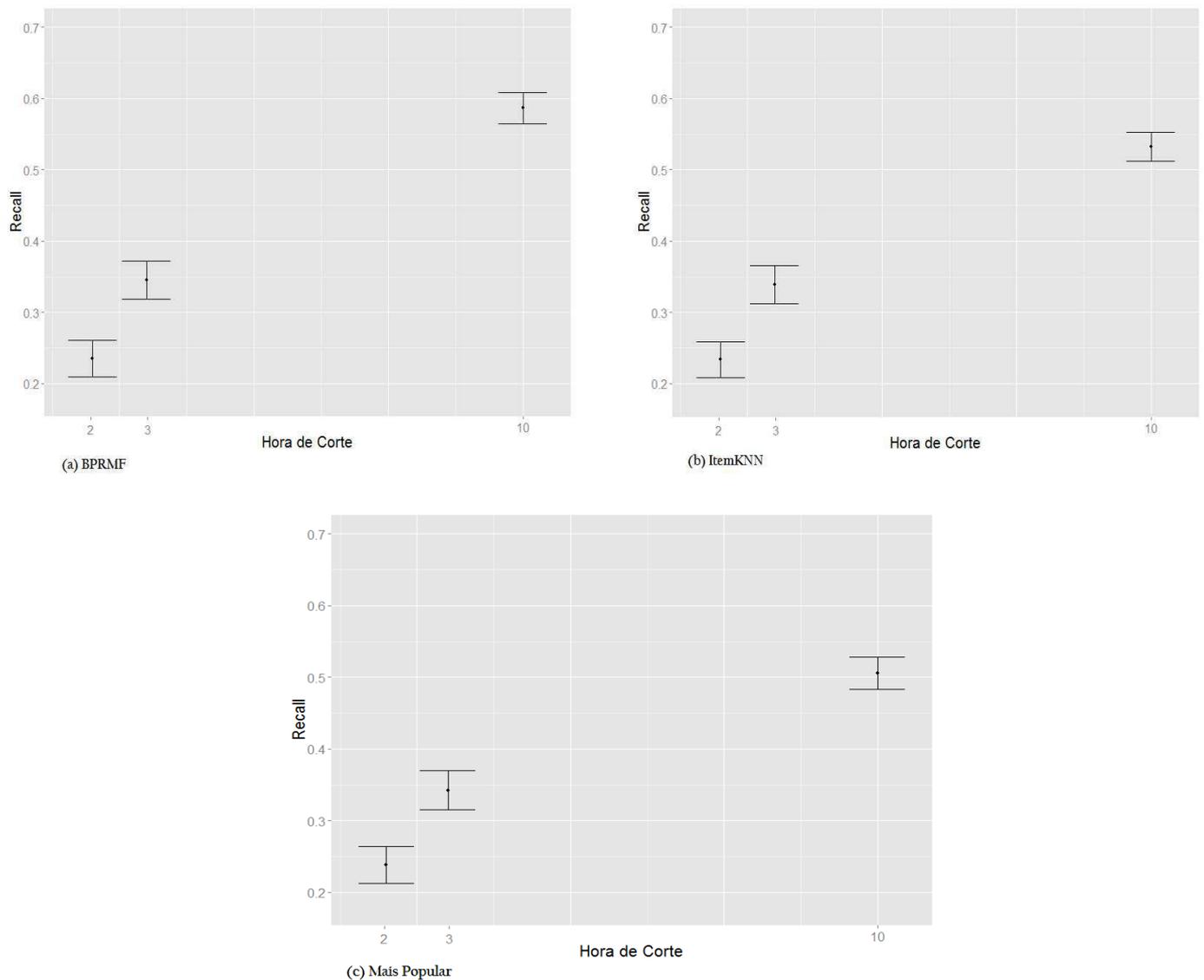
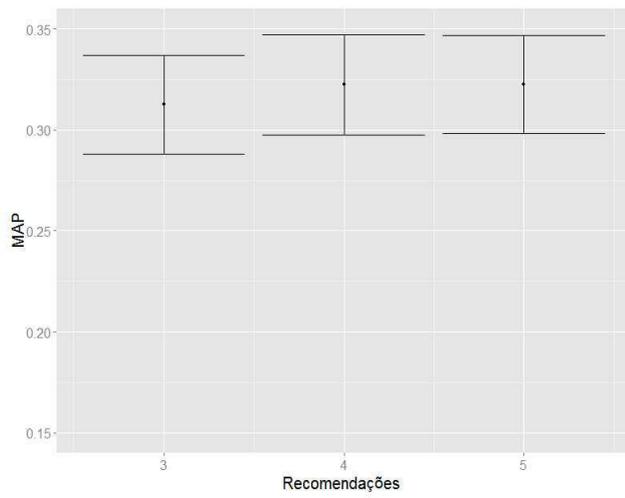


Figura 19 - Intervalos de confiança, considerando os níveis do fator "hora de corte", para a métrica *recall*.

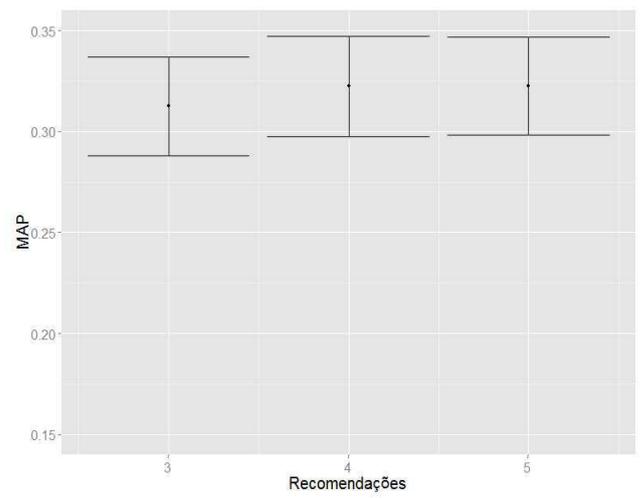
Logo, para o fator hora de corte, os melhores resultados para a recomendação das ofertas divulgadas durante um dia são obtidos após as 10 horas de divulgação inicial da oferta.

5.1.3. Análise para o fator "número de itens a ser recomendado"

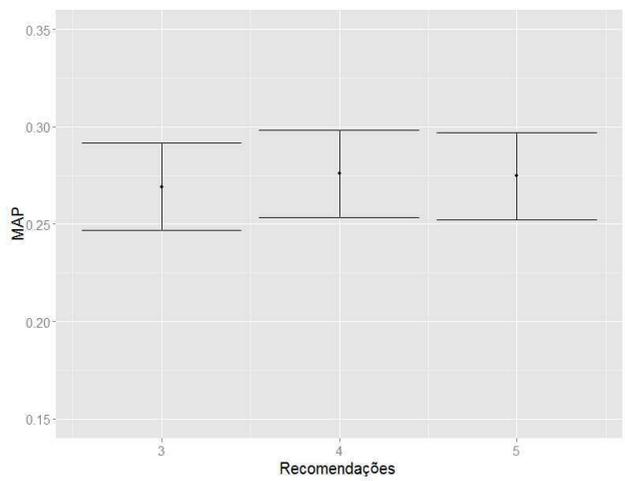
Semelhantemente ao que foi analisado nos fatores anteriores, verificou-se que variando os níveis do fator número de itens a ser recomendado, as métricas de *MAP*, *precisão* e *recall* apresentaram o mesmo comportamento, em que a variação do intervalo de confiança não permite que haja uma conclusão de qual o melhor nível para este fator. De acordo com as Figura 20, Figura 21 e Figura 22, o intervalo de confiança com $\alpha = 0,1$ mostra-se que o comportamento da alteração do nível do número de recomendações a ser efetuadas pouco influenciam as métricas, mantendo-se equiparadas em quase todo o intervalo.



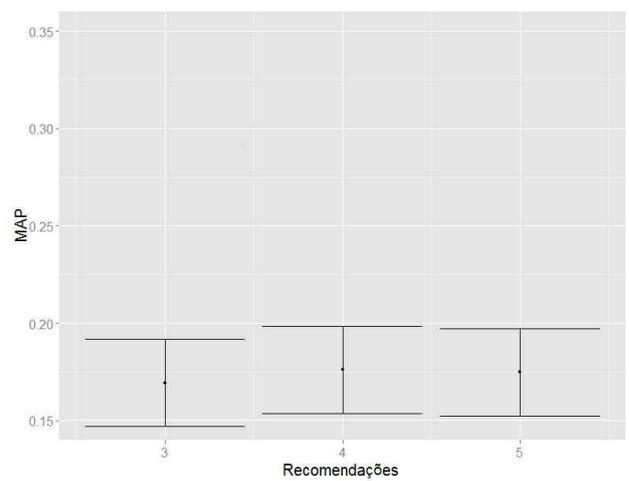
(a) BPRMF



(b) ItemKNN

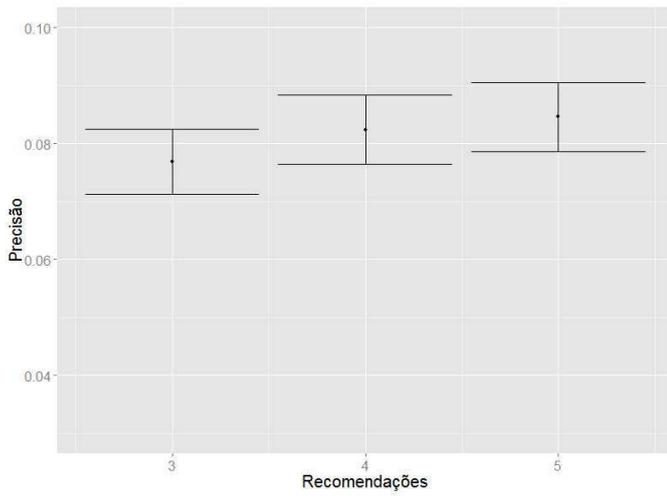


(c) Mais Popular

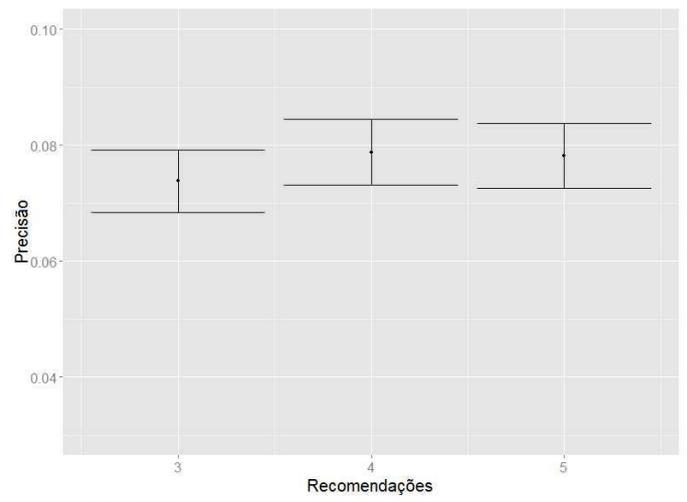


(d) Query&Search

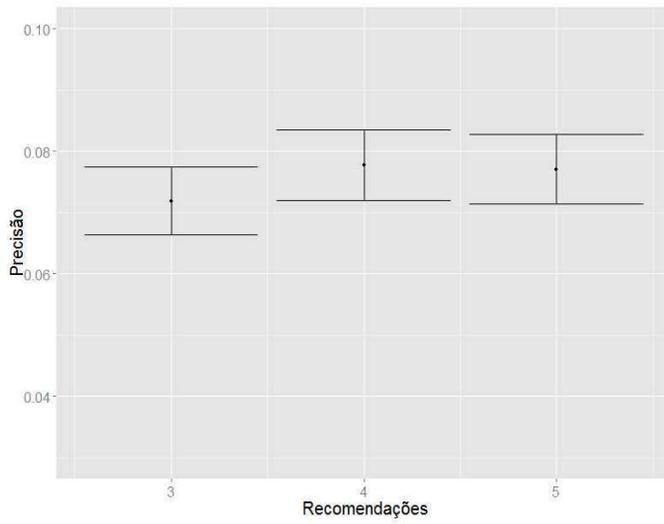
Figura 20 - Intervalos de confiança, considerando os níveis do fator "número de itens a ser recomendado", para a métrica *MAP*.



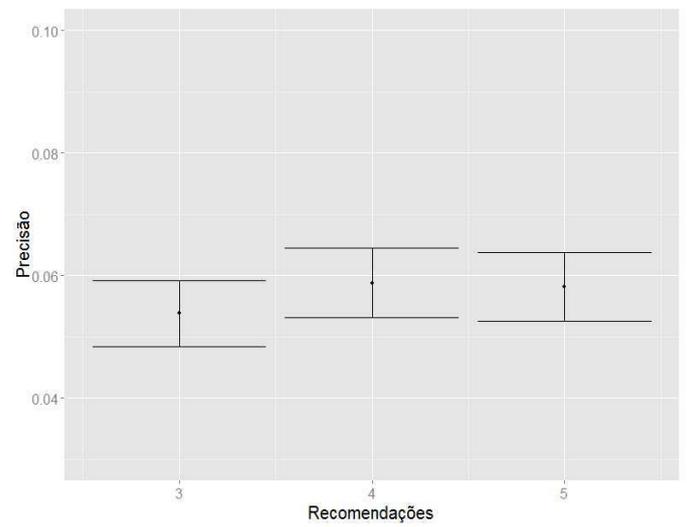
(a) BPRFM



(b) ItemKNN

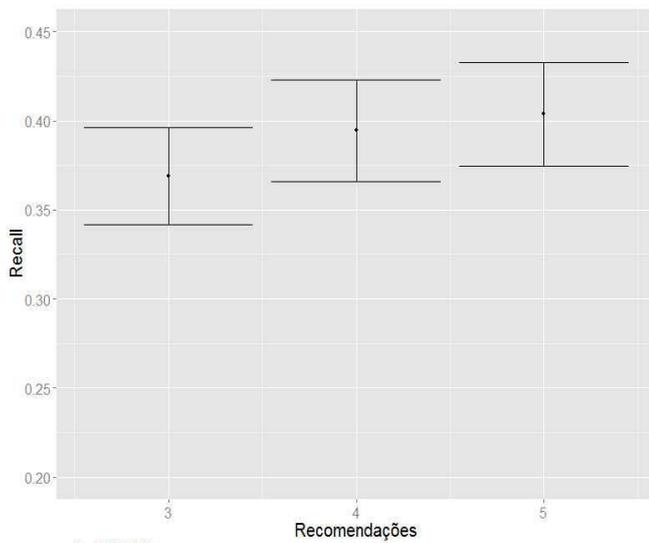


(c) Mais Popular

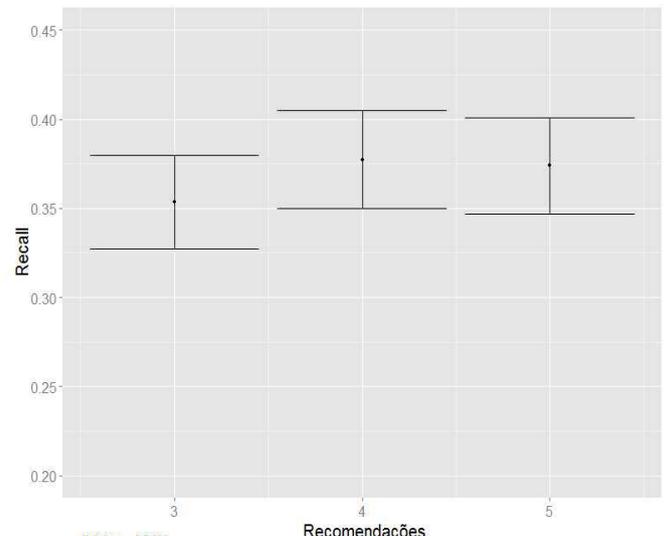


(d) Query&Search

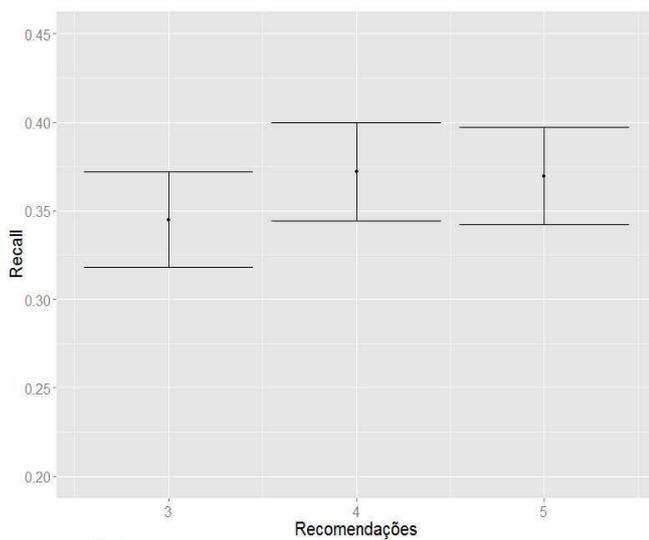
Figura 21 - Intervalos de confiança, considerando os níveis do fator "número de itens a ser recomendado", para a métrica precisão.



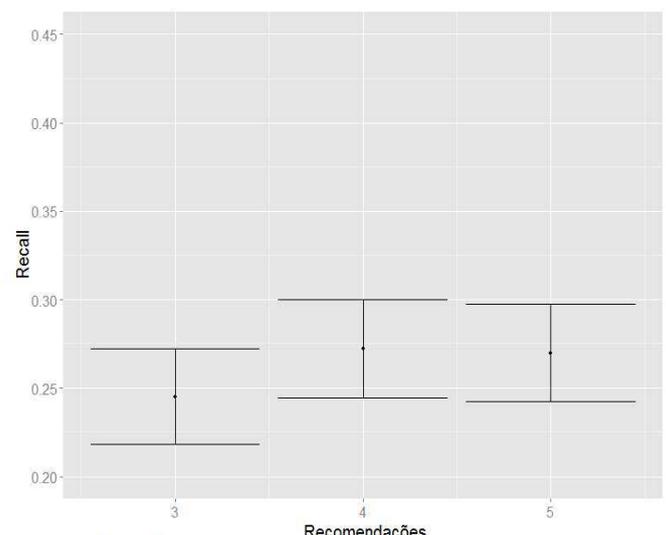
(a) BPRMF



(b) ItemKNN



(c) Mais Popular



(d) Query&Search

Figura 22 - Intervalos de confiança, considerando os níveis do fator "número de itens a ser recomendado", para a métrica *recall*.

O mesmo teste estatístico de **Wilcoxon** foi realizado para verificar se há destaque em um dos níveis, considerando exclusivamente um único algoritmo, a partir das seguintes hipóteses:

H0: A métrica em questão é, em média, igual para todos os níveis do fator analisado, de acordo com o algoritmo utilizado;

H1: A métrica em questão não é, em média, igual para todos os níveis do fator analisado, de acordo com o algoritmo utilizado.

Utilizando o teste estatístico, verificou-se que, para as métricas de **MAP**, **precisão** e **recall**, os algoritmos **BPRMF**, **ItemkNN** e **Mais Popular e Query & Search** apresentam valores semelhantes para as métricas, mesmo quando há variação do número de itens a ser recomendados, não havendo confirmação estatística que determine a diferença entre estes. Para todas as métricas, variando-se o algoritmo e o número de recomendações, o *p-valor* foi igual 1, fazendo com que a hipótese nula fosse aceita.

5.1.4. Análise para o fator "número mínimo de compras por usuário"

Analogamente ao que foi analisado nos fatores anteriores, verificou-se que fixando o fator número mínimo de compras por usuário, a métrica **MAP** apresenta diferenças substanciais quando há variação do número de compras nos algoritmos. Para os algoritmos **BPRMF** e **ItemkNN**, o comportamento é muito semelhante, nos quais os valores da métrica em questão são muito próximos, mesmo na ocorrência da variação do nível do fator, como pode ser visto na Figura 23.

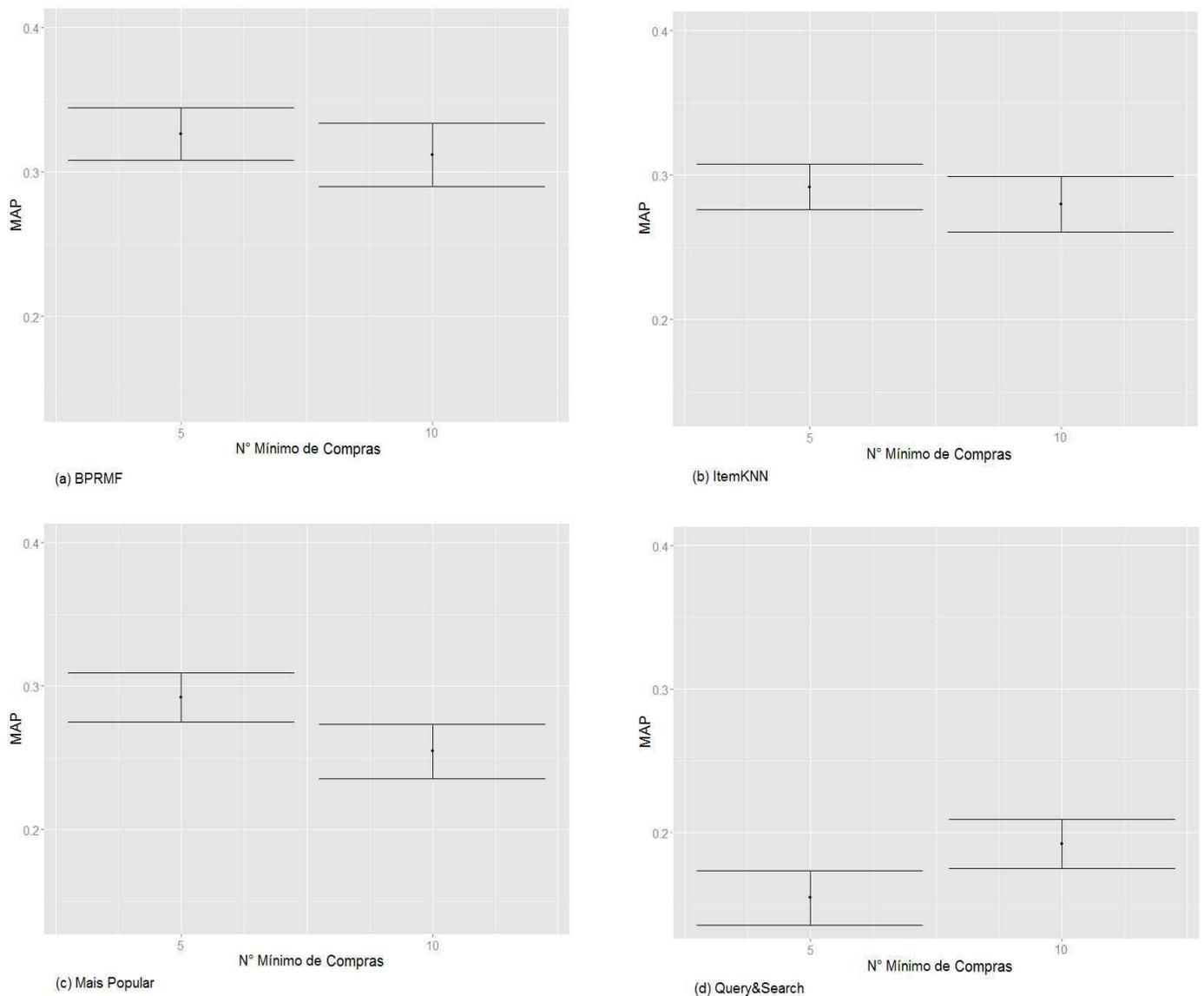


Figura 23 - Intervalos de confiança, considerando os níveis do fator "número mínimo de compras por usuário", para a métrica MAP.

Analisando estatisticamente, utilizando o teste de **Wilcoxon**, podemos formular as seguintes hipóteses:

H0: A métrica **MAP** é, em média, igual para todos os níveis do fator analisado;

H1: A métrica **MAP** não é, em média, igual para todos os níveis do fator analisado.

E, aplicando-se o teste estatístico em que o *p-valor* foi igual a 0,1671 para o algoritmo **BPRMF** e igual a 0,1468 para o algoritmo **ItemkNN**, verificou-se que não há diferença na métrica, a partir dos valores medidos quando aplicamos tais algoritmos. Já o algoritmo **Mais Popular** apresenta os melhores valores quando o nível do fator está definido como "ao menos 5 compras efetuadas pelo usuário", enquanto o contrário ocorre para o algoritmo **Query &**

Search, que apresenta melhores resultados quando o nível do fator está definido como "ao menos 10 compras efetuadas pelo usuário".

Quando analisamos a métrica **precisão**, verificamos que os algoritmos **BPRMF** e **ItemkNN** apresentam comportamento semelhante ao apresentado na métrica **MAP**, em que não há diferenciação na métrica em questão quando há variação do nível do fator, como pode ser visto na Figura 24. Isto pode ser comprovado pelo teste de **Wilcoxon**, no qual podemos formular as seguintes hipóteses:

H0: A métrica **precisão** é, em média, igual para todos os níveis do fator analisado;

H1: A métrica **precisão** não é, em média, igual para todos os níveis do fator analisado.

Aplicado o teste estatístico, verificou-se que, com *p-valor* igual a 0,0633 para o algoritmo **BPRMF** e 0,1801 para o algoritmo **ItemkNN**, não há evidências estatísticas para afirmar que há diferença nos valores apresentados para a métrica em questão quando variamos os níveis, para os algoritmos **BPRMF** e **ItemkNN**, respectivamente.

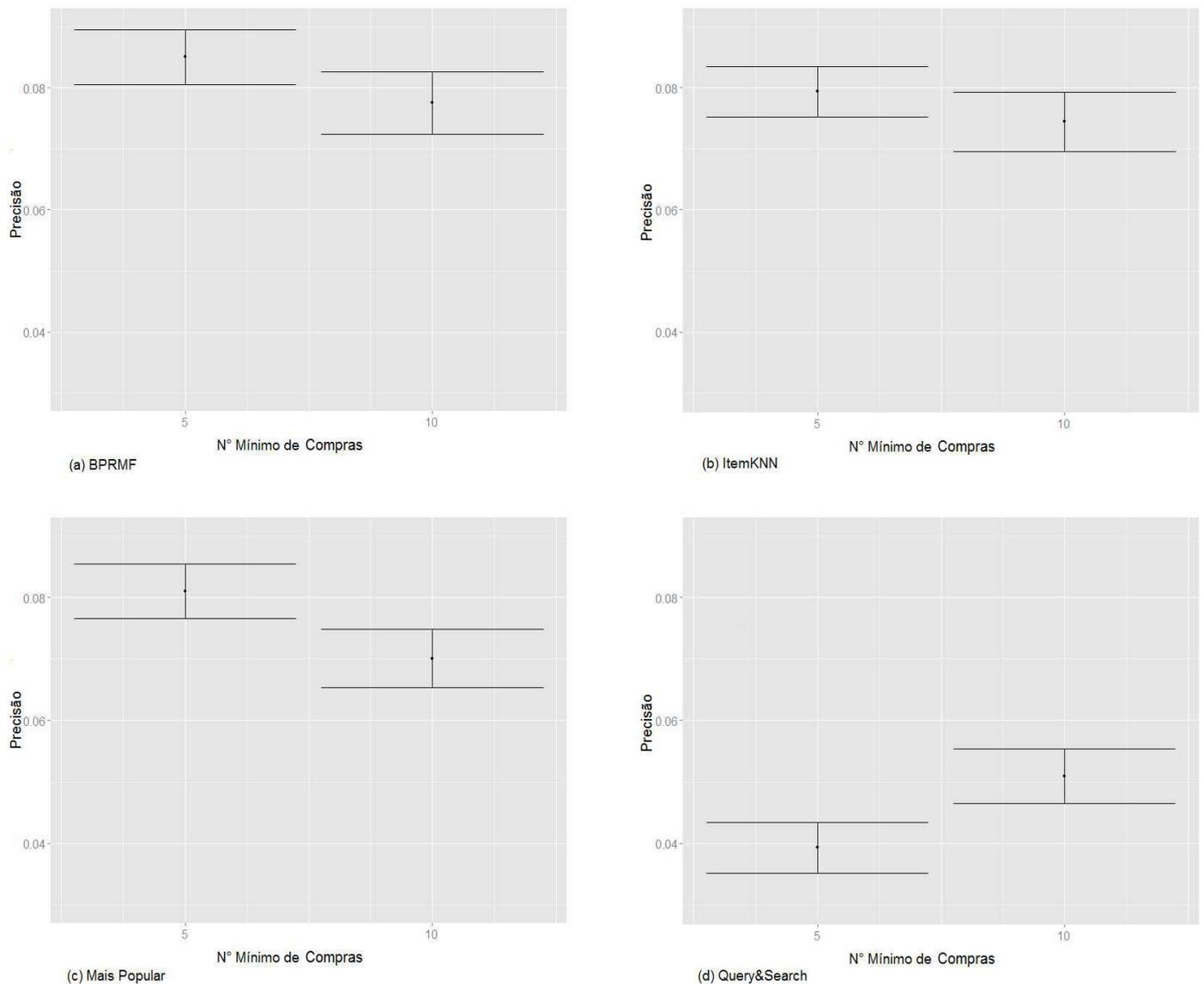


Figura 24 - Intervalos de confiança, considerando os níveis do fator "número mínimo de compras por usuário", para a métrica precisão.

Já os algoritmos **Mais Popular** e **Query & Search** apresentam comportamento semelhante ao analisado na métrica **MAP**. O primeiro apresenta melhores valores quando no primeiro nível, "ao menos 5 compras efetuadas pelo usuário", e o segundo quando no segundo nível, "ao menos 10 compras efetuadas pelo usuário".

Quando analisamos a métrica **recall**, verificamos que o algoritmo **BPRMF** apresenta comportamento diferente ao apresentado quando analisado a partir das métricas **MAP** e **precisão**, apresentando diferenciação na métrica em questão quando há variação do nível do fator, como pode ser visto na Figura 25. Isto pode ser comprovado pelo teste de **Wilcoxon**, no qual é possível formular as seguintes hipóteses:

H0: A métrica *recall* é, em média, igual para todos os níveis do fator analisado;

H1: A métrica *recall* não é, em média, igual para todos os níveis do fator analisado.

Aplicado o teste estatístico, verificou-se que, com *p-valor* igual a 0,00002 há estatisticamente evidências para afirmar que há diferença nos valores apresentados para a métrica em questão quando variamos o nível, para o algoritmo **BPRMF**, no qual o nível que apresenta os melhores resultados é o primeiro, "ao menos 5 compras efetuadas pelo usuário".

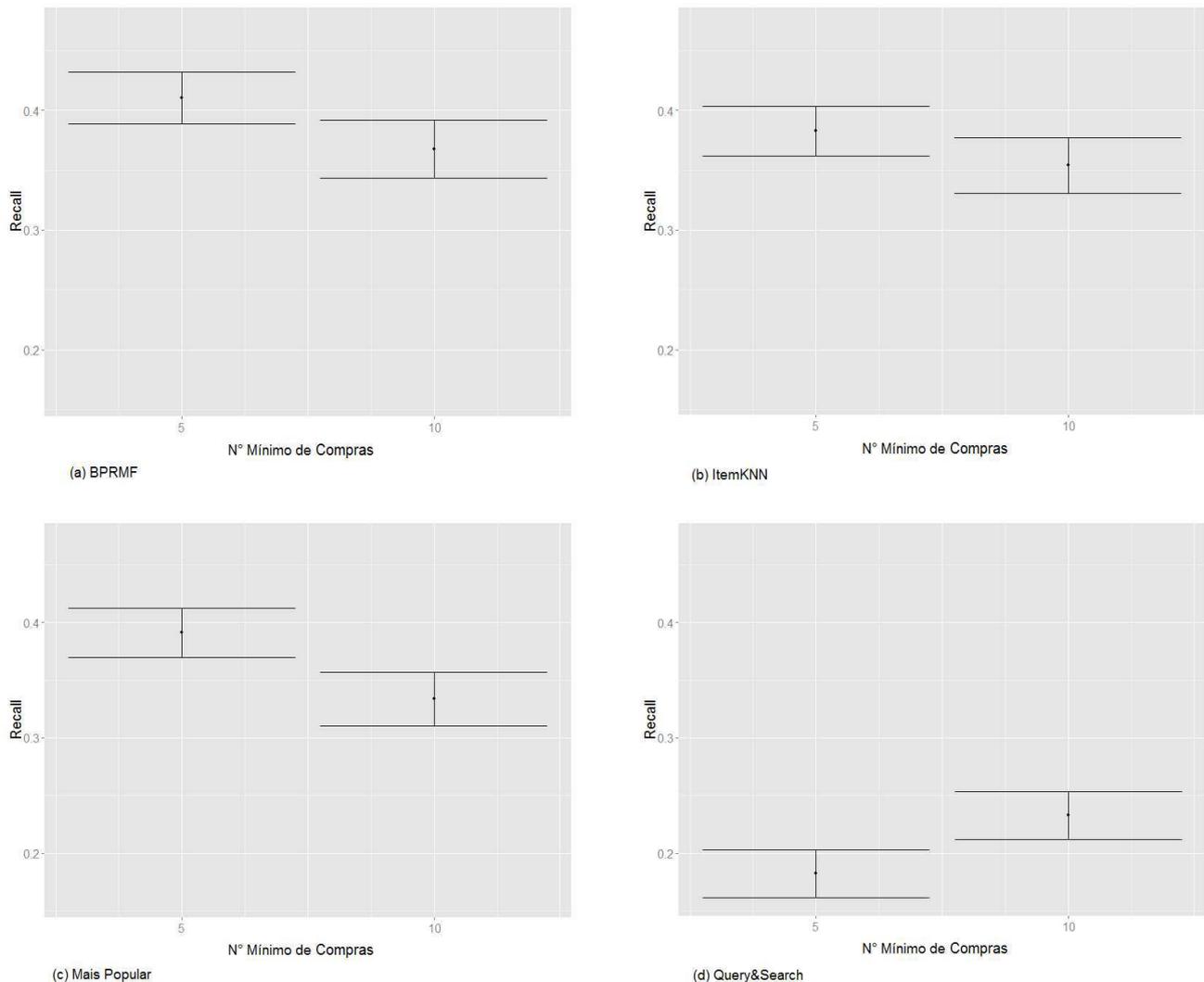


Figura 25 - Intervalos de confiança, considerando os níveis do fator "número mínimo de compras por usuário", para a métrica *recall*.

Ainda analisando a métrica *recall*, verificamos que o algoritmo **ItemKNN** apresenta comportamento semelhante ao apresentado quando analisado a partir das métricas **MAP** e **precisão**, no qual não há diferenciação na métrica em questão quando há variação do nível do fator, como pode ser visto na Figura 25. Isto pode ser comprovado pelo teste de **Wilcoxon**, no qual é possível formular as seguintes hipóteses:

H0: A métrica *recall* é, em média, igual para todos os níveis do fator analisado;

H1: A métrica *recall* não é, em média, igual para todos os níveis do fator analisado.

Aplicado o teste estatístico, verificou-se que, com *p-valor* igual a 0,0912, não há evidências estatísticas para afirmar que há diferença nos valores apresentados na métrica em questão, quando variamos o nível, para o algoritmo em questão.

Já os algoritmos **Mais Popular** e **Query & Search** apresentam comportamento semelhante ao analisado na métrica *MAP* e *precisão*. O primeiro apresenta melhores valores quando no primeiro nível, "ao menos 5 compras efetuadas pelo usuário", e o segundo quando no segundo nível, "ao menos 10 compras efetuadas pelo usuário".

5.2. Experimento Adicional: Agrupamento do histórico de compras por quatro dias consecutivos

No capítulo de Trabalhos Relacionados, o trabalho de Lacerda [32] menciona que, de acordo com a base de dados da empresa Peixe Urbano, as ofertas são válidas por mais de um dia. Além disso, em sua grande maioria, a média de dias em que as ofertas estão divulgadas é de 4 dias. Como os algoritmos de filtragem colaborativa funcionam melhor de acordo com a disponibilidade de dados, por possuírem mais dados para a etapa de treinamento, a recomendação que é gerada por esses algoritmos possivelmente superam as do experimento anterior.

Portanto, para fins de análise dos algoritmos de recomendação sob a perspectiva na qual os itens estão disponíveis por mais dias, essa mesma configuração foi repetida em um novo experimento, agrupando as ofertas de 4 dias consecutivos e realizando a recomendação. Este agrupamento se deu de maneira a acumular por 4 dias consecutivos os itens que foram divulgados e separá-los em dados de treino e teste apenas de acordo com o horário da compra. O conjunto de ofertas que antes era formado por apenas ofertas de um único dia agora corresponde ao conjunto de ofertas de 4 dias consecutivos. Um breve exemplo pode ser visto na Figura 26.

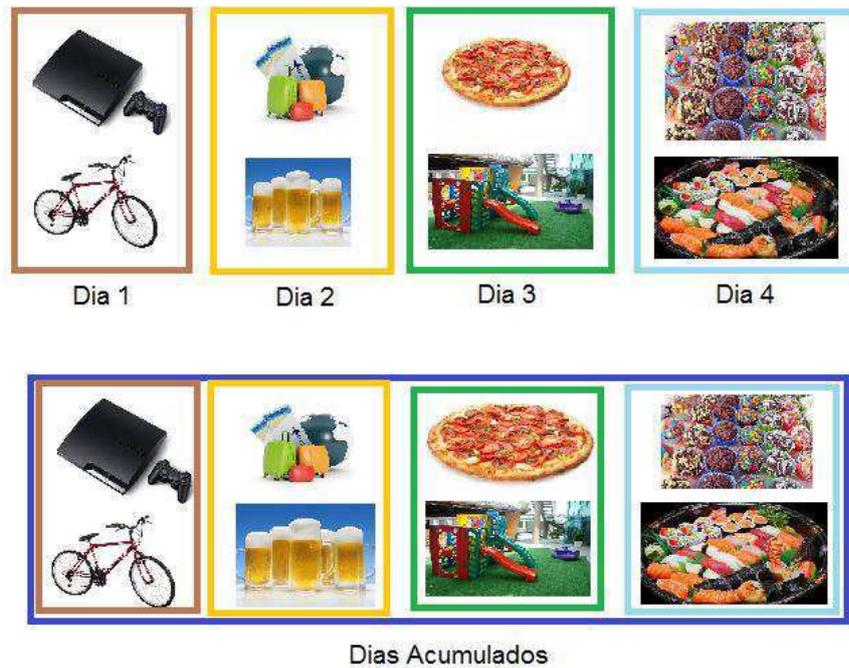


Figura 26 - Ofertas de 4 dias consecutivos agrupadas para simulação de experimento.

5.2.1. Análise para o fator "algoritmo"

A avaliação experimental mostrou que, fixando-se o fator algoritmo, as métricas de *MAP*, *precisão* e *recall* apresentaram os melhores resultados para o algoritmo **BPRMF**. De acordo com as Figura 27, Figura 28 e Figura 29, e intervalos de confiança com $\alpha = 0,1$, mostram que as métricas *MAP*, *precisão* e *recall* são superiores, em todos os casos, no algoritmo **BPRMF**, seguido pelos algoritmos **ItemkNN** e **Mais Popular**. Os piores resultados são observados no algoritmo baseado em conteúdo **Query & Search**. Pela análise estatística, podemos concluir que o algoritmo **BPRMF** apresenta os melhores resultados para essa métrica.

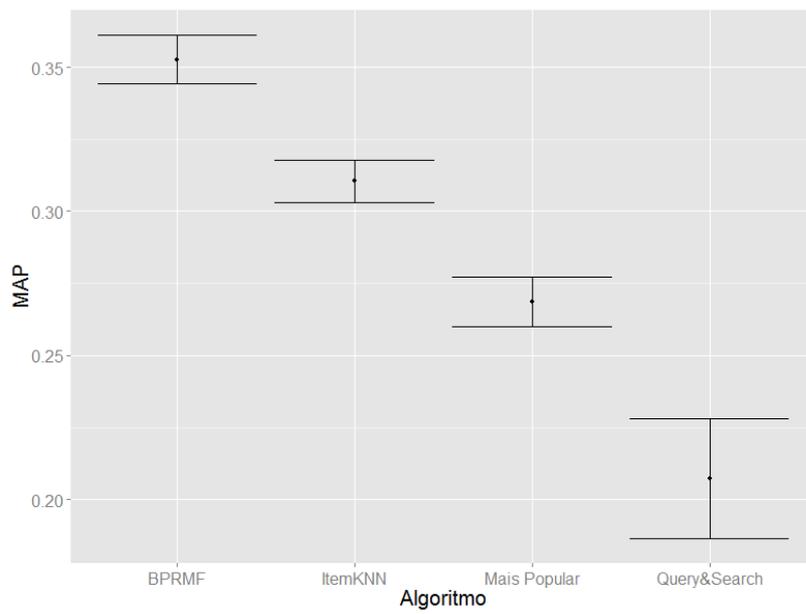


Figura 27 - Intervalo de confiança para a métrica MAP observada.

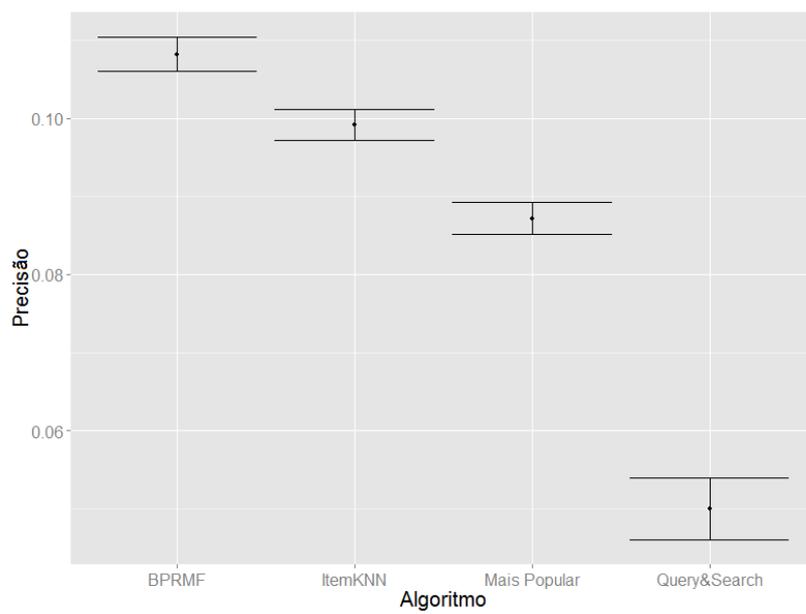


Figura 28 - Intervalos de confiança para a métrica precisão.

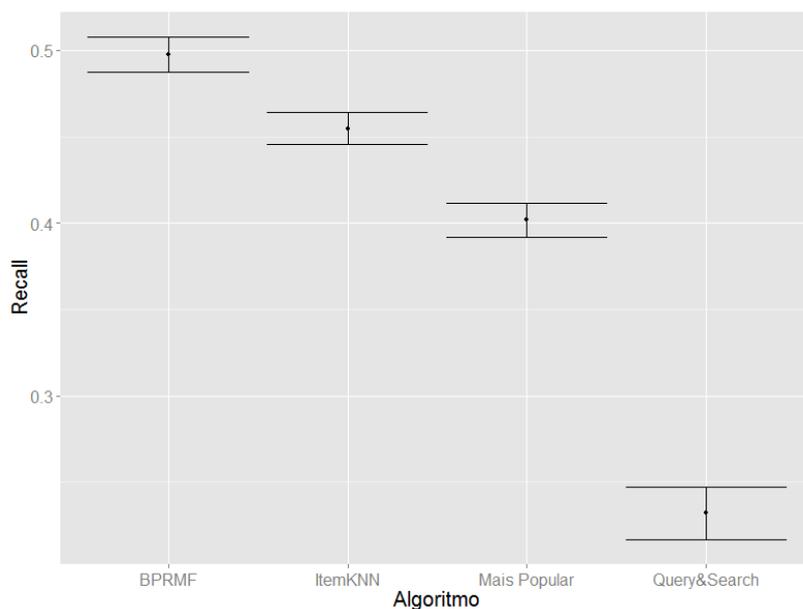


Figura 29 - Intervalos de confiança para a métrica *recall*.

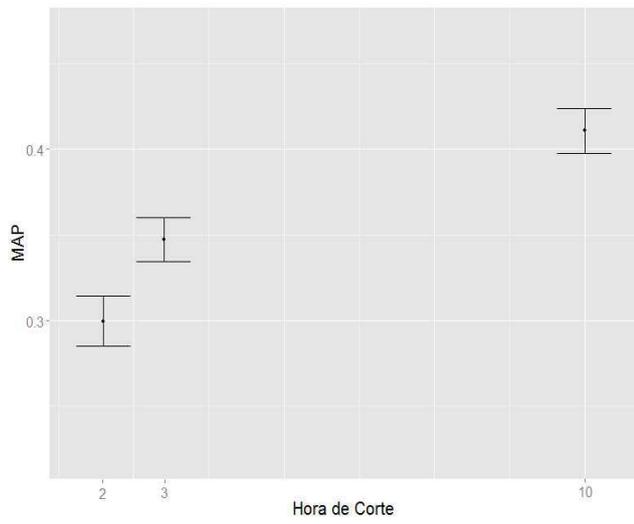
5.2.2. Análise para o fator "hora de corte"

A avaliação experimental mostrou que, fixando-se o fator hora de corte, as métricas de **MAP**, **precisão** e **recall** apresentaram os melhores resultados quando consideramos o nível "10 horas após o início da divulgação das ofertas", em dois dos três algoritmos analisados.

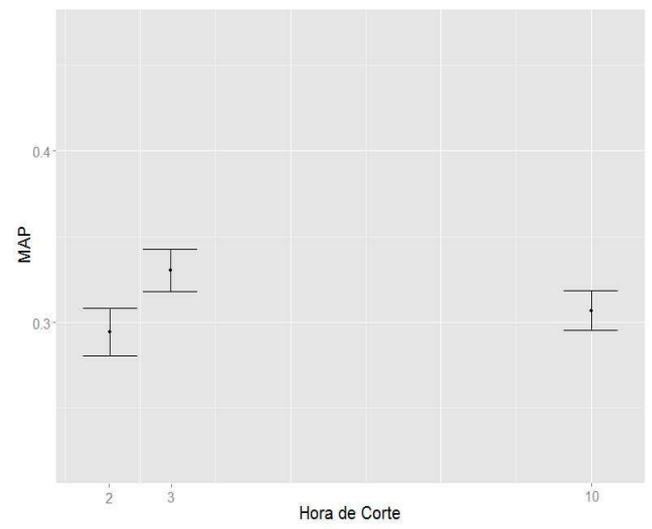
De acordo com a Figura 30, o intervalo de confiança com $\alpha = 0,1$ mostra que a métrica **MAP** é superior, considerando os algoritmos **BPRMF** e **Mais Popular**, nos casos em que o nível do fator é o terceiro nível. As Figura 31 e Figura 32 mostram a mesma situação, porém considerando as métricas de **precisão** e **recall**, respectivamente. Os piores resultados são observados no primeiro nível, quando há tentativa de recomendação 2 horas após o início de divulgação das ofertas.

A mesma justificativa demonstrada na seção 5.1.2 é válida para a variação desse nível, logo não analisaremos a variação dos níveis do fator para o algoritmo de recomendação baseado em conteúdo **Query & Search**.

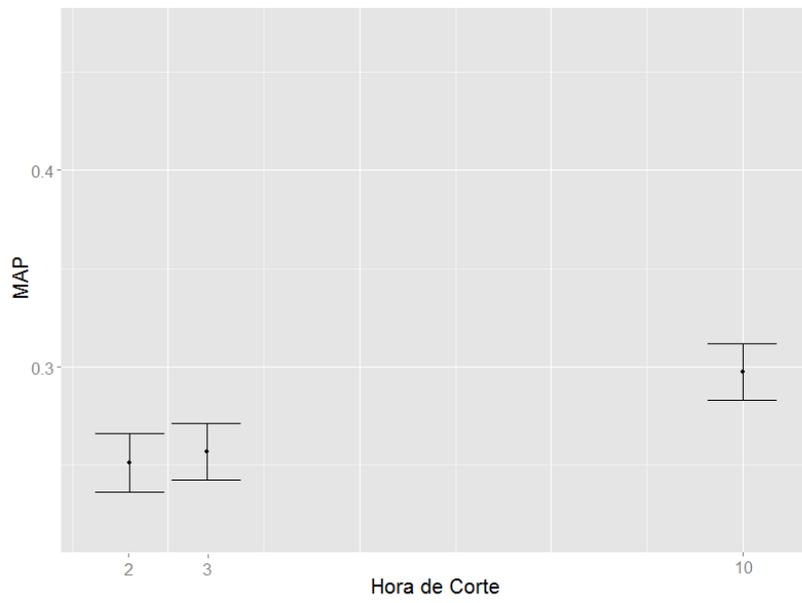
Quanto ao algoritmo **ItemkNN**, este apresenta resultados divergentes quando comparado aos outros. Para a métrica **MAP** é possível identificar que o segundo nível, "após 3 horas do início da recomendação", apresenta os melhores resultados. Já para as métricas **precisão** e **recall**, este resultado é inconclusivo, pelo teste estatístico de **Wilcoxon**.



(a) BPRMF

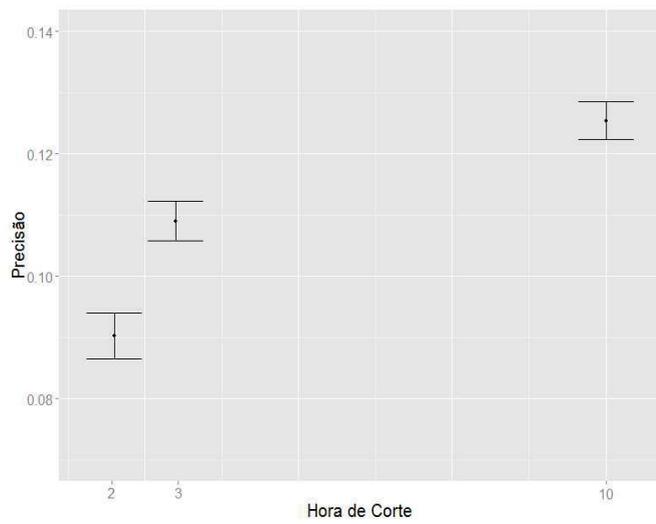


(b) ItemKNN

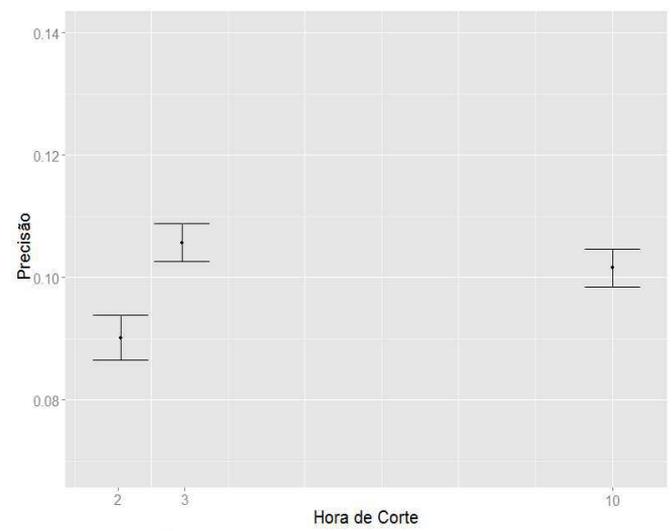


(c) Mais Popular

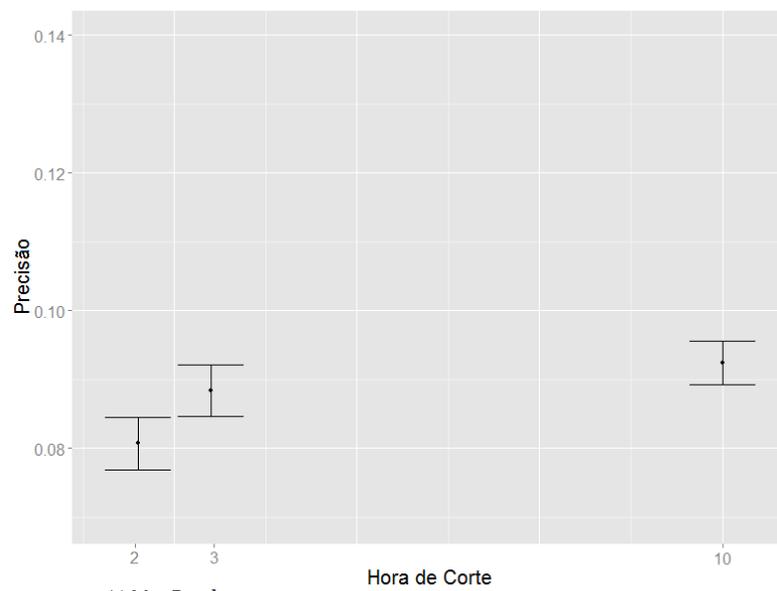
Figura 30 - Intervalos de confiança, considerando os níveis do fator "hora de corte", para a métrica MAP.



(a) BPRMF



(b) ItemKNN



(c) Mais Popular

Figura 31 - Intervalos de confiança, considerando os níveis do fator "hora de corte", para a métrica precisão.

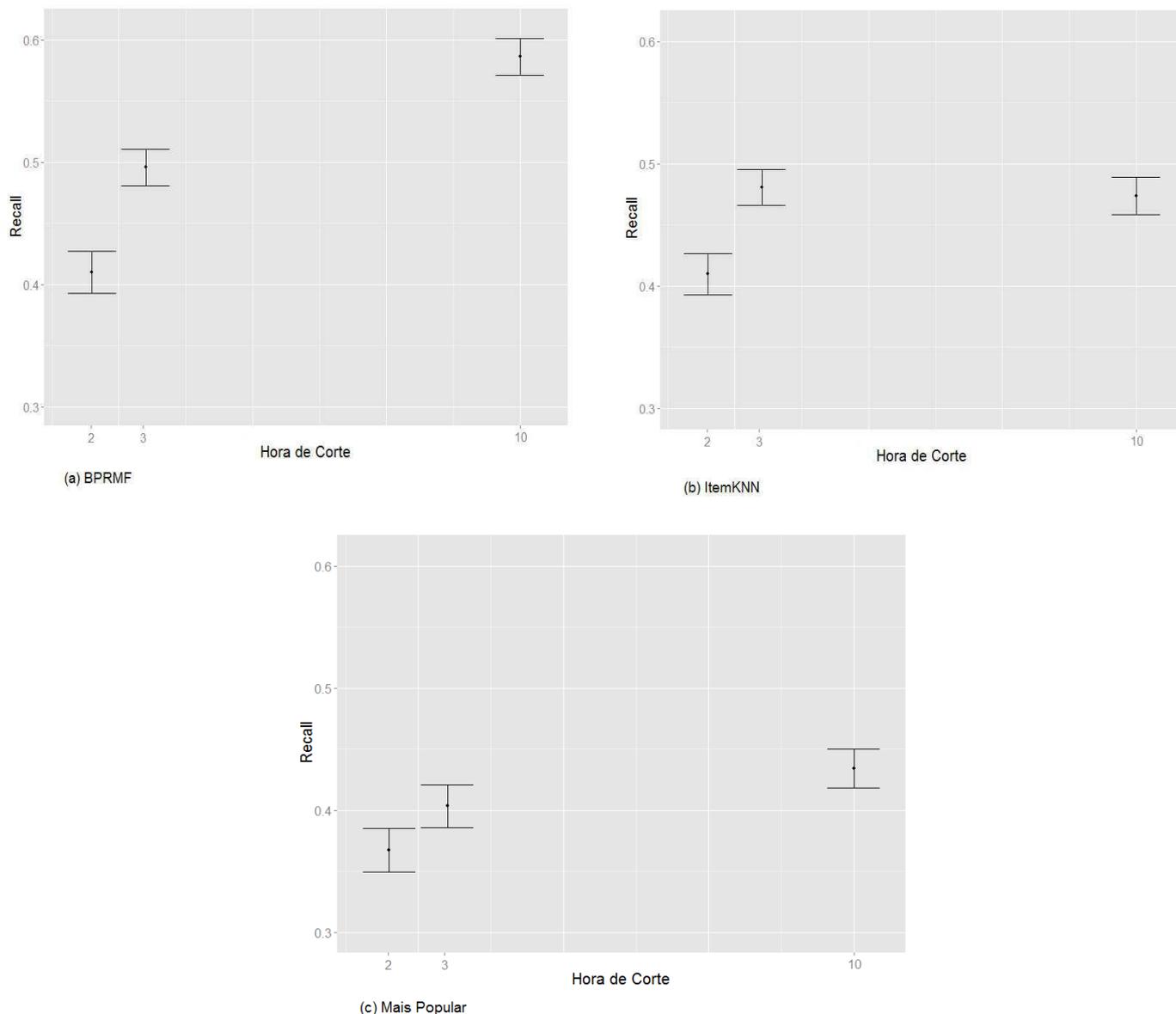


Figura 32 - Intervalos de confiança, considerando os níveis do fator "hora de corte", para a métrica *recall*.

Logo, para o fator "hora de corte" os melhores resultados foram observados quando há maior quantidade de dados, correspondendo ao terceiro nível deste fator, nos algoritmos de recomendação **BPRMF** e **Mais Popular**. Tal fato é observado pois os dados de treinamento já foram obtidos pelas compras realizadas até o momento, para o primeiro, e as ofertas mais adquiridas durante o dia já estão definidas para o segundo.

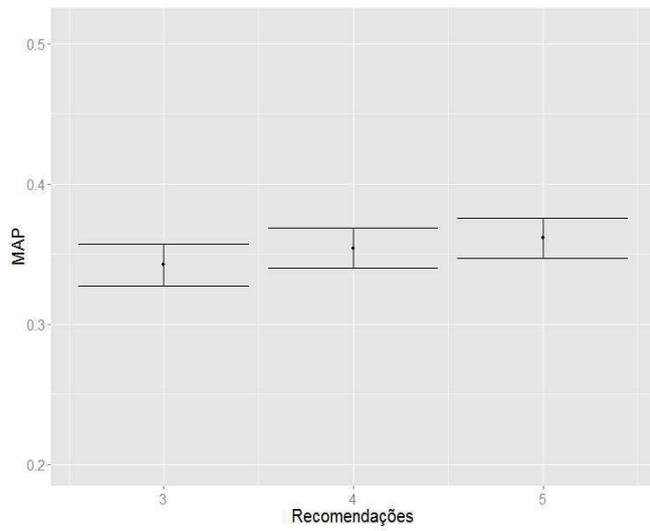
Já o algoritmo de recomendação baseado em memória **ItemkNN** apresenta os melhores resultados no segundo nível em geral, pois neste é possível definir claramente quais são os vizinhos mais próximos dos usuário, fazendo com que a recomendação seja mais precisa. Com a inclusão de mais histórico de compras, no terceiro nível do fator, o algoritmo apresenta resultados iguais ou piores pois há dificuldade em definir os vizinhos mais

próximos, quando há agrupamento de 4 dias consecutivos de divulgação. Porém, deve-se observar que a variação é pequena nos níveis, podendo não influenciar diretamente na recomendação das ofertas.

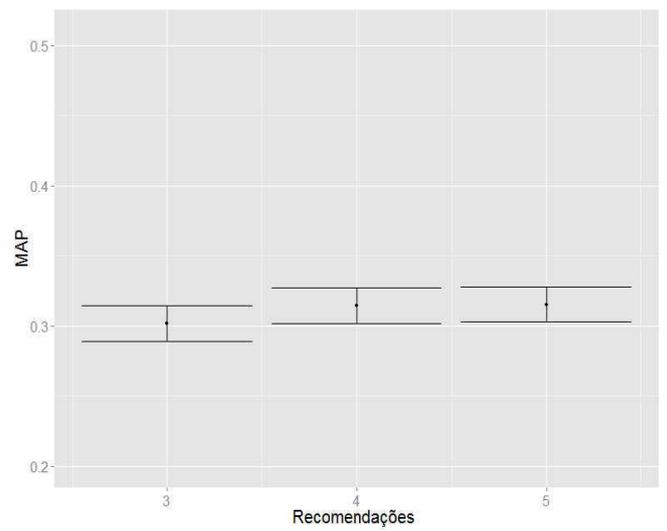
5.2.3. Análise para o fator "número de itens a ser recomendado"

De maneira semelhante ao que foi analisado nos fatores anteriores, verificou-se que fixando o fator "número de itens a ser recomendado", as métricas de *MAP*, *precisão* e *recall* apresentaram o mesmo comportamento, em que a variação do intervalo de confiança não permite que haja uma conclusão de qual o melhor nível para este fator, para os algoritmos de recomendação **ItemkNN**, **Mais Popular** e **Query & Search**. De acordo com as Figura 33, Figura 34 e Figura 35, com intervalos de confiança com $\alpha = 0,1$, mostram que o comportamento da alteração do nível do número de recomendações a ser efetuada pouco influenciam as métricas, mantendo-se equiparadas em quase todo o intervalo.

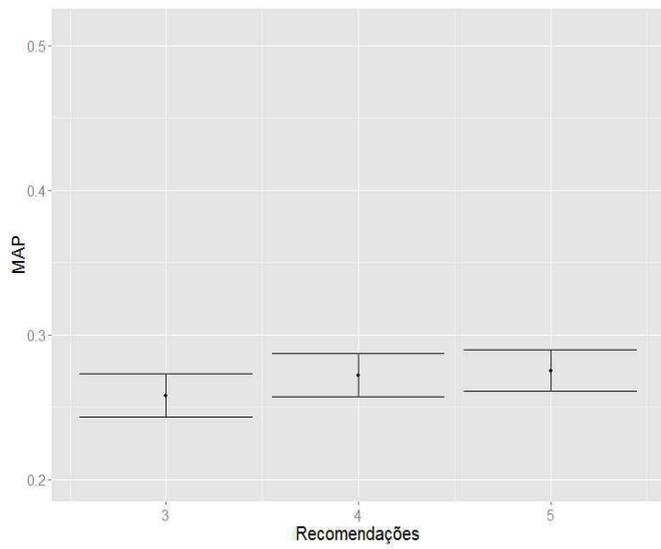
Quando analisamos o algoritmo de recomendação **BPRMF**, vimos que as métricas analisadas apresentam os melhores resultados no terceiro nível deste fator, ou seja, recomendando 5 ofertas aos usuários. Tal fato pode ser explicado pois as ofertas de interesse podem estar ranqueadas em posições mais a frente, ou seja, as três primeiras opções podem não conter exatamente as ofertas que são de interesse para o usuário. Por esta variação ser pequena, os resultados podem não ter diferenças quanto à qualidade da recomendação, pois os ganhos são mínimos, de acordo com os valores medidos.



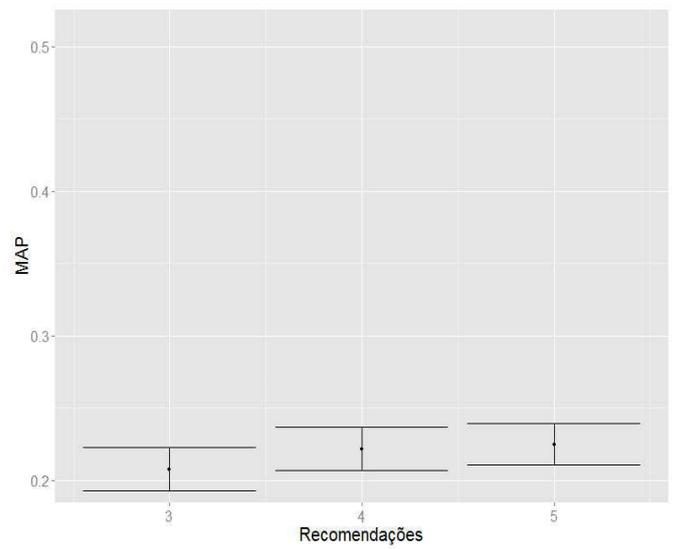
(a) BPRMF



(b) ItemKNN

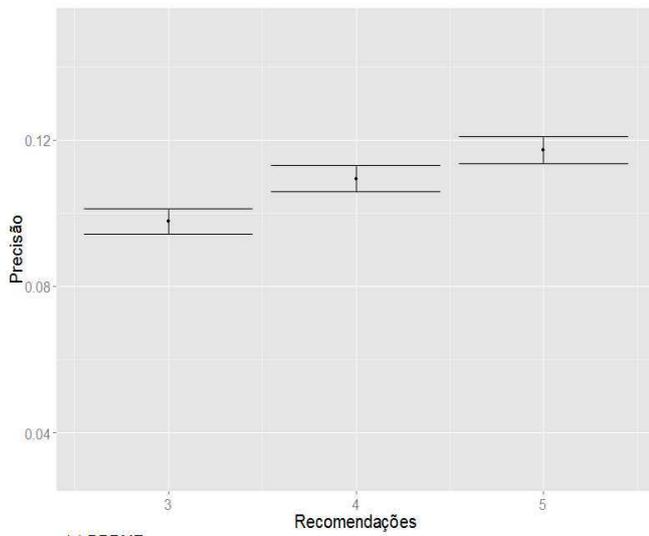


(c) Mais Popular

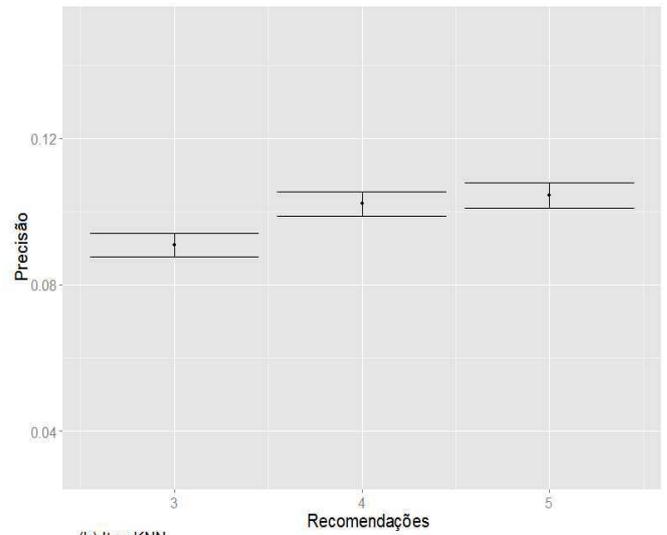


(d) Query&Search

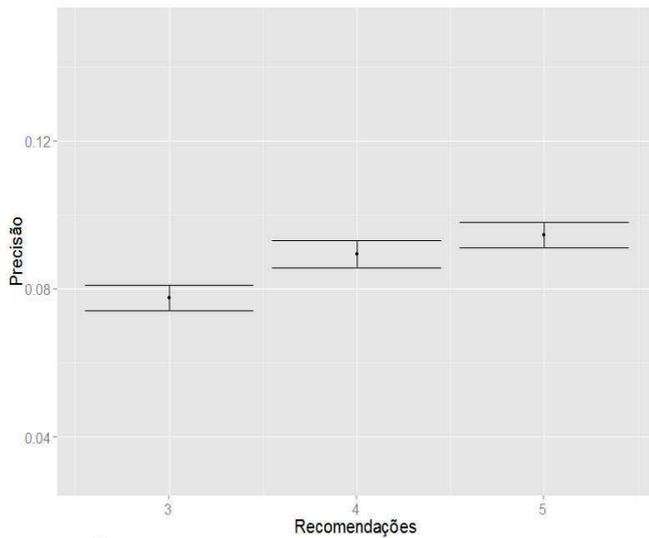
Figura 33 - Intervalos de confiança, considerando os níveis do fator "número de itens a ser recomendado", para a métrica *MAP*.



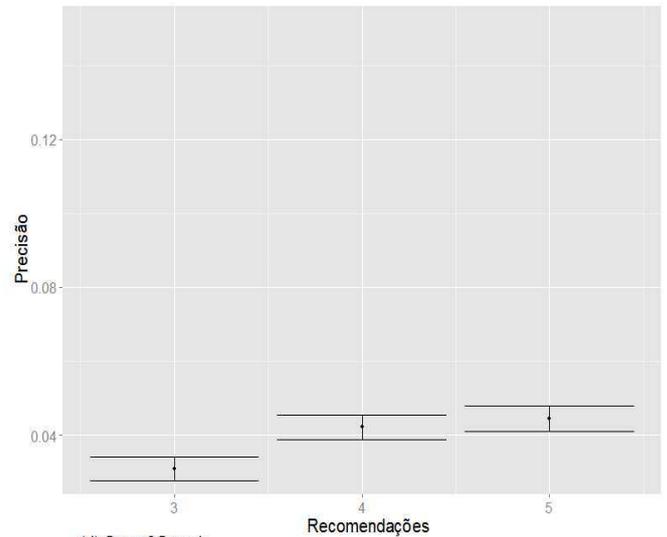
(a) BPRMF



(b) ItemKNN



(c) Mais Popular



(d) Query&Search

Figura 34 - Intervalos de confiança, considerando os níveis do fator "número de itens a ser recomendado", para a métrica precisão.

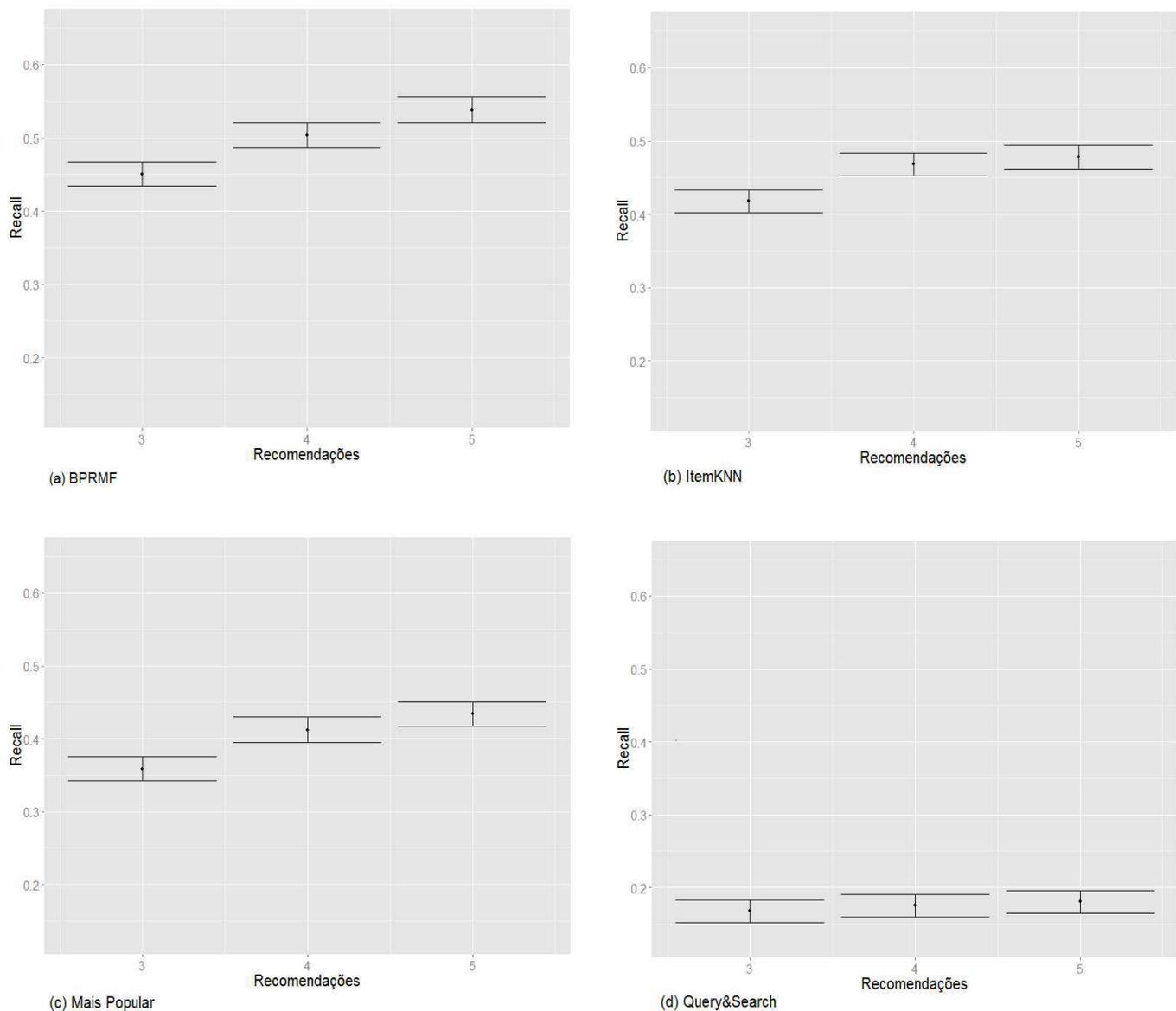


Figura 35 - Intervalos de confiança, considerando os níveis do fator "número de itens a ser recomendado", para a métrica *recall*.

Logo, para o fator número de itens a ser recomendado, nada se pode afirmar sobre sua superioridade em relação aos outros níveis, apesar de, em alguns casos do algoritmo **BPRMF**, os melhores resultados são observados no terceiro nível do fator. Porém, os ganhos praticamente imperceptíveis, de acordo com os valores medidos.

5.2.4. Análise para o fator "número mínimo de compras por usuário"

Analogamente ao que foi analisado nos fatores anteriores, verificou-se que fixando o fator número mínimo de compras por usuário, a métrica **MAP** apresenta diferenças quando há variação do número de compras nos algoritmos. Para os algoritmos **BPRMF** e **ItemkNN**, o

comportamento é muito semelhante, em que os melhores resultados podem ser observados no segundo nível do fator, ou seja, a partir do histórico de compras de usuários com, no mínimo, 10 aquisições de ofertas, como pode ser visto na Figura 36.

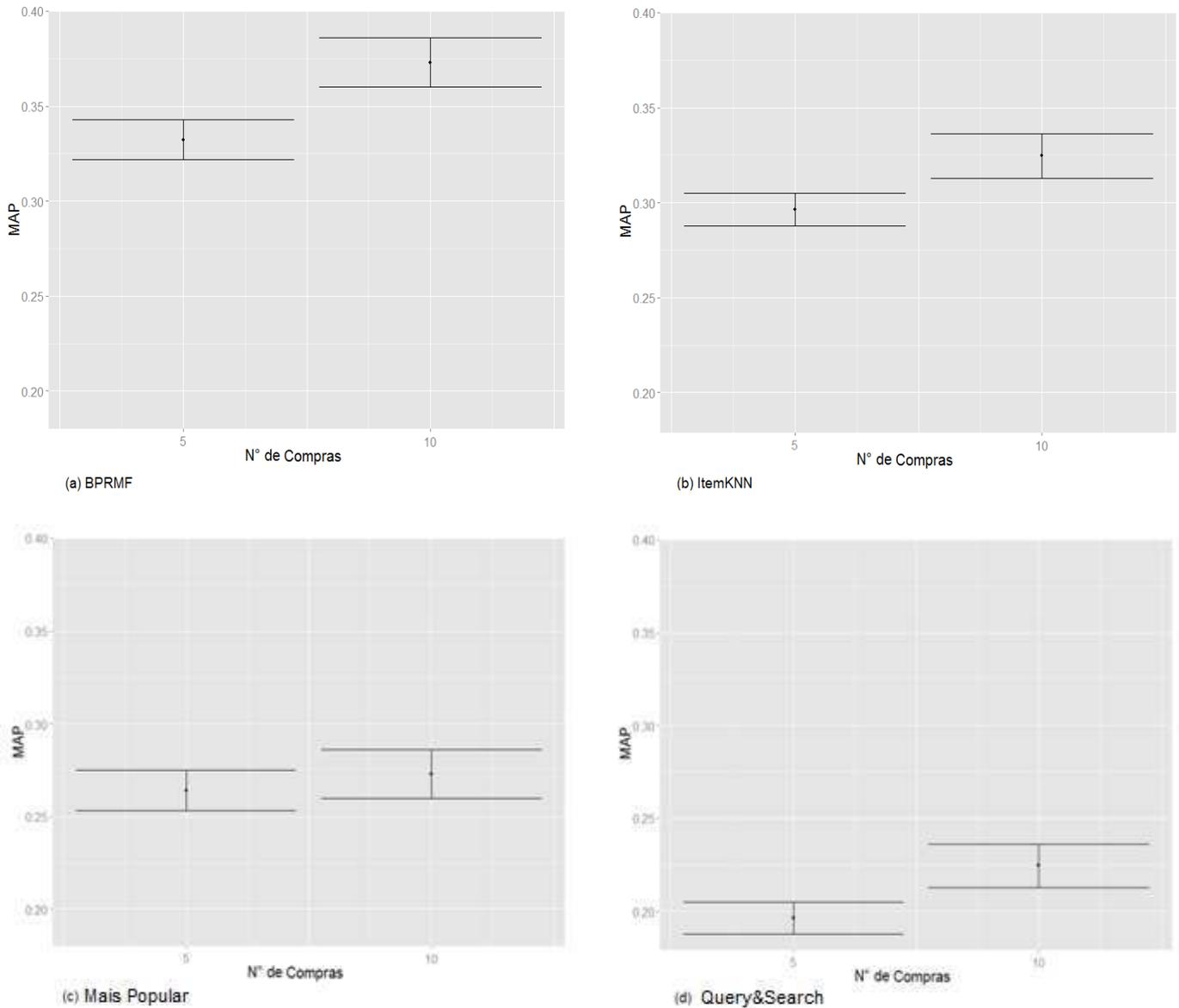


Figura 36 - Intervalos de confiança, considerando os níveis do fator "número mínimo de compras por usuário", para a métrica MAP.

Analisando estatisticamente, utilizando o teste de **Wilcoxon**, podemos formular as seguintes hipóteses:

H0: A métrica **MAP** é, em média, igual para todos os níveis do fator analisado;

H1: A métrica **MAP** não é, em média, igual para todos os níveis do fator analisado.

E, aplicando-se o teste estatístico em que o *p-valor* foi igual a 0,6574 para o algoritmo **Mais Popular**, verificou-se que não há diferença na métrica, a partir dos valores medidos quando aplicamos tais algoritmos.

Já o algoritmo **Query & Search**, apresenta melhores resultados quando o nível do fator está definido como "ao menos 10 compras efetuadas pelo usuário".

Quando analisamos a métrica **precisão**, verificamos que os algoritmos **BPRMF** e **ItemkNN** apresentam comportamento semelhante ao apresentado pela métrica **MAP**, no qual há melhores resultados a partir do segundo nível do fator, como pode ser visto na Figura 37. Ainda quanto à precisão, o algoritmo **Mais Popular** não possui definição de melhores resultados com a variação dos níveis do fator, podendo isto ser comprovado pelo teste de **Wilcoxon**, no qual é possível formular as seguintes hipóteses:

H0: A métrica **precisão** é, em média, igual para todos os níveis do fator analisado;

H1: A métrica **precisão** não é, em média, igual para todos os níveis do fator analisado.

Aplicado o teste estatístico, verificou-se que, com *p-valor* igual a 0,3653, não há estatisticamente evidências para afirmar que há diferença nos valores apresentados para a métrica em questão quando variamos o nível, para o algoritmo **Mais Popular**.

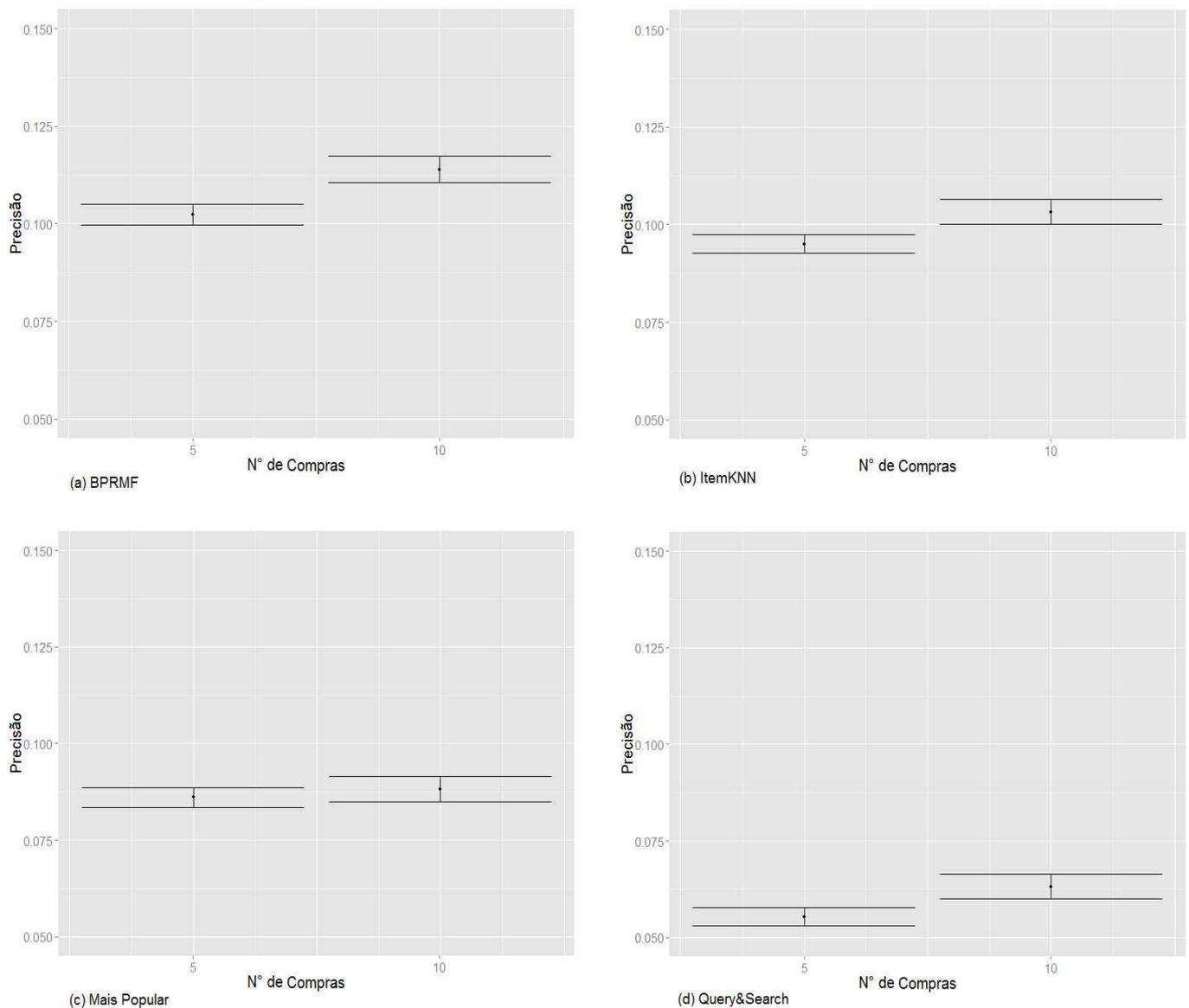


Figura 37 - Intervalos de confiança, considerando os níveis do fator "número mínimo de compras por usuário", para a métrica precisão.

Já o algoritmo **Query & Search** apresenta melhores resultados quando o segundo nível, "ao menos 10 compras efetuadas pelo usuário" é fixado.

Quando analisamos a métrica *recall*, verificamos que os algoritmos **BPRMF** e **ItemkNN** apresentam comportamentos semelhantes ao apresentado quando analisado a partir das métricas *MAP* e *precisão*, como pode ser visto na Figura 38. Da mesma forma, o algoritmo **Mais Popular** possui comportamento semelhante, como pode ser comprovado pelo teste de **Wilcoxon**, no qual é possível formular as seguintes hipóteses:

H0: A métrica *recall* é, em média, igual para todos os níveis do fator analisado;

H1: A métrica *recall* não é, em média, igual para todos os níveis do fator analisado.

Aplicado o teste estatístico, verificou-se que, com p -valor igual a 0,8042 não há estatisticamente evidências para afirmar que há diferença nos valores apresentados para a métrica em questão quando variamos o nível.

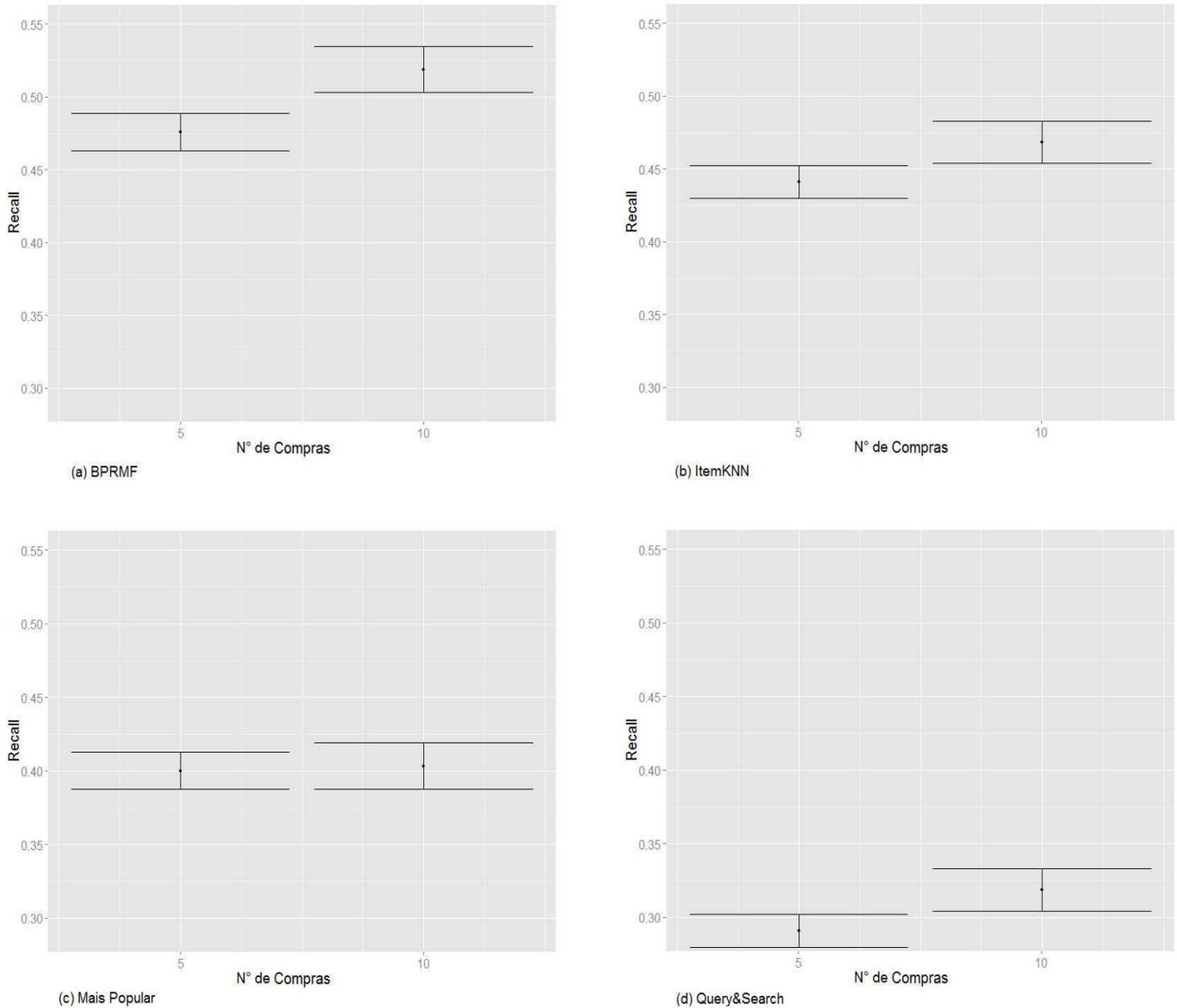


Figura 38 - Intervalos de confiança, considerando os níveis do fator "número mínimo de compras por usuário", para a métrica recall.

Já o algoritmo **Query & Search** apresenta comportamento semelhante ao analisado a partir das métricas **MAP** e **precisão**, apresentando melhores valores quando o segundo nível, "ao menos 10 compras efetuadas pelo usuário", é fixado.

5.3. Discussão

Dentre os algoritmos de recomendação utilizados neste trabalho, o algoritmo mais apropriado para recomendação em sites de compras coletivas é o algoritmo **BPRMF**, nos dois

experimentos realizados. Esse algoritmo apresenta os melhores resultados para a métrica **MAP** e, na maioria das vezes, os melhores resultados para a **precisão e recall**, muito embora nenhuma afirmação estatística possa ser feita quanto ao algoritmo que apresenta os melhores resultados.

Quando analisamos os melhores níveis dos fatores escolhidos, vemos que o comportamento dos algoritmos **BPRFM** e **ItemkNN** é bastante similar, com relação aos fatores hora de corte, número de itens a ser recomendado e número mínimo de compras por usuário. Apesar de pertencerem a mesma classe de algoritmos de recomendação - a classe dos algoritmos de recomendação personalizados baseados em filtragem colaborativa - esses algoritmos utilizam diferentes estratégias de recomendação, e apresentaram resultados bastante semelhantes quando há restrições de tempo e ofertas para a recomendação. Porém, a experimentação realizada na seção 2 indica que o algoritmo **BPRMF** é superior ao algoritmo **ItemkNN** quando há agrupamento de ofertas em 4 dias consecutivos.

Utilizando o algoritmo **BPRMF**, podemos concluir que:

- 1) **MAP**: Representa a recomendação de um item na posição i da lista de recomendações representa a ordem de relevância da oferta para o usuário. Logo, para esta métrica, apenas 32% e 35,5% das ofertas recomendadas, seguindo a ordem de recomendação do algoritmo, indicam a preferência do usuário, quando levamos em consideração o primeiro e segundo experimentos, respectivamente.
 - a) Quando há variação no fator **número de itens a ser recomendado**, há pouco ou nenhum ganho significativo na métrica, em ambos os experimentos;
 - b) Para o fator **número mínimo de compras efetuadas por usuário**, o primeiro experimento mostra equivalência dos níveis deste fator. Já no segundo, há ganho de aproximadamente 5% quando há agrupamento de ofertas em 4 dias consecutivos;
 - c) Já o melhor nível para o fator **hora de corte** é o terceiro nível, indicando que há melhor recomendação após 10 horas do início de divulgação da oferta;
- 2) **Precisão**: Representa a quantidade de ofertas relevantes que são recomendadas para o usuário, em detrimento as ofertas que não são relevantes. Logo, para esta métrica, apenas 8% e 11% das ofertas recomendadas, no primeiro e segundo experimentos, são de interesse do usuário.
 - a) Quando há variação no fator **número de itens a ser recomendado**, há pouco ganho na métrica, em média, em ambos os experimentos;

- b) Os experimentos apresentam comportamentos diferentes para o fator **número mínimo de compras efetuadas por usuário**, nos quais o primeiro experimento apresenta melhores resultados quando há filtro de usuários que contém 5 ou mais ofertas adquiridas. Já o segundo apresenta melhores resultados quando há filtro de usuários que contém 10 ou mais ofertas adquiridas;
 - c) O melhor nível para o fator **hora de corte** é o terceiro nível, indicando que há melhor recomendação após 10 horas do início de divulgação da oferta.
- 3) **Recall**: Representa a quantidade de ofertas relevantes recomendadas dentre as ofertas relevantes. Logo, para esta métrica, apenas 39% e 49% das ofertas recomendadas são relevantes para o usuário, apenas considerando o conjunto das ofertas que são relevantes, para o primeiro e segundo experimento, respectivamente.
- a) Quando há variação no fator **número de itens a ser recomendado**, há pouco ou nenhum ganho na métrica em ambos os experimentos, apesar do segundo experimento apresentar melhores resultados no terceiro nível desse fator. Porém, a diferença é pouco significativa;
 - b) Os experimentos apresentam comportamentos diferentes para o fator **número mínimo de compras efetuadas por usuário**, semelhantemente ao que foi observado na seção 2.b;
 - c) Já o melhor nível para o fator **hora de corte** é o terceiro nível, semelhantemente ao que foi observado na seção 2.c;

Utilizando o algoritmo **ItemkNN**, podemos concluir que:

- 1) **MAP**: Para essa métrica, apenas 28% e 31% dos itens recomendados, na ordem de recomendação do algoritmo, indicam a preferência o usuário, no primeiro e segundo experimentos, respectivamente;
 - a) Quando há variação no fator **número de itens a ser recomendado**, há pouco ou nenhum ganho na métrica, em ambos os experimentos;
 - b) Quando o fator **número mínimo de compras efetuadas por usuário** é variado, o primeiro experimento mostra que o primeiro nível apresenta uma ligeira vantagem em relação ao segundo. Já quando analisamos o segundo experimento, o segundo nível é superior ao primeiro. Porém, em ambos os casos, a variação é pouco significativa;Já o fator **hora de corte** apresenta divergências de acordo com o experimento. O terceiro nível apresenta os melhores resultados, no primeiro experimento, indicando que há melhor recomendação após 10 horas do início de divulgação das ofertas. Já o

segundo experimento apresenta o segundo nível, após 3 horas do início de divulgação das ofertas, com melhores resultados.

- 2) **Precisão:** Para essa métrica, apenas 7,7% e 10% das ofertas recomendadas são relevantes para o usuário. Isto indica que, dentre todas as ofertas, poucas ofertas recomendadas são de interesse do usuário, a partir do primeiro e segundo experimentos, respectivamente;
 - a) Quando há variação para o fator **número de itens a ser recomendado**, há pouco ganho na métrica, em média, para ambos os experimentos;
 - b) A mesma observação da seção 1.b pode ser estendida para esta seção, na qual há divergência nos resultados dos experimentos, de acordo com o fator **número mínimo de compras efetuadas por usuário**;
 - c) A mesma observação da seção 1.c pode ser estendida para esta seção, na qual há divergência nos resultados dos experimentos, de acordo com o fator **hora de corte**.
- 3) **Recall:** Para essa métrica, os resultados apresentaram média de 37,5% e 45% para o primeiro e segundo experimento, respectivamente, indicando que há relevância média entre as ofertas que foram recomendadas e as ofertas que são relevantes para o usuário;
 - a) Quando há variação para o fator **número de itens a ser recomendado**, há pouco ou nenhum ganho na métrica, para ambos os experimentos;
 - b) Quando analisamos os níveis para o **número mínimo de compras efetuadas por usuário**, observamos o mesmo comportamento apresentado na seção 2.b;
 - c) A mesma observação pode ser estendida para o fator **hora de corte**, observando o mesmo comportamento da seção 2.c.

Utilizando o algoritmo **Mais Popular**, notamos que há quase o mesmo poder de recomendação quando comparado aos dois algoritmos anteriores, considerando o primeiro experimento. Muito se deve ao fato de, diariamente, existirem ofertas que são mais adquiridas, de acordo com análise dos dados. Essa peculiaridade motivou o experimento adicional.

De acordo com o segundo experimento, observamos que há pouco ou nenhum ganho quando há agrupamento das ofertas em 4 dias consecutivos. Muito se deve ao fato de que, devido ao agrupamento, muitas ofertas são consideradas populares e o algoritmo por vezes acerta as ofertas que são relevantes, recomendando 3, 4 ou 5 ofertas, e por vezes acaba errando, por não conseguir ranquear as ofertas mais populares em casos de igualdade, por exemplo, para recomendar aos usuários.

Em muitos casos, pudemos observar que há empate nos resultados, porque realmente os usuários possuem preferências semelhantes, seja pela atratividade da oferta em relação

desconto aplicado no valor final ou mesmo por essa oferta estar em destaque no site. É possível concluir que, pela variabilidade dos produtos/serviços divulgados no site, há sempre uma oferta que é mais popular entre os usuários e geralmente essa oferta é da categoria "Restaurante".

Quando há agrupamento de 4 dias consecutivos de divulgação de ofertas, observamos que o algoritmo **Mais Popular** ainda consegue ter a mesma taxa de acerto do experimento sem agrupamento de dias consecutivos. Porém, em todas as rodadas do segundo experimento, os algoritmos **BPRMF** e **ItemkNN** superam, em todas as métricas estudadas, esse algoritmo. Logo, para o cenário em que há empate de itens que são populares, o algoritmo **Mais Popular** não é tão eficaz e preciso como nos casos em que há destaque de apenas uma oferta como a mais popular.

Finalmente, utilizando o algoritmo **Query & Search**, notamos que esse apresenta o menor poder de recomendação quando comparado aos três algoritmos anteriores. Muito se deve ao fato de não utilizarmos uma abordagem mais sofisticada e esse algoritmo ter pouca influência dos fatores escolhidos nesta pesquisa, já que a recomendação baseada no conteúdo da descrição das ofertas considera apenas a similaridade da oferta divulgada e as ofertas que foram compradas pelo usuário antes do período de corte.

Utilizando apenas o algoritmo que verifica a similaridade entre o conteúdo das ofertas, notamos a insuficiência para superar os demais algoritmos, porque a descrição das ofertas pode ser feita de diversas, para ofertas de mesma categoria ou semelhantes, como por exemplo "Venha saborear deliciosas panquecas no Restaurante Sul Do Lar" e "Delicie-se com deliciosas massas finas recheadas com carne no S&R". As duas ofertas podem conter o mesmo assunto, porém descrições diferentes, fazendo com que a similaridade do conteúdo não seja suficiente para aproximá-las e não haja recomendação.

5.4. Considerações Finais

Neste capítulo mostramos os resultados obtidos a partir da metodologia definida no Capítulo 4, com o intuito de avaliar os algoritmos estado da arte em recomendação e, a partir das métricas, explicar os resultados obtidos.

Os resultados mostram que, no primeiro experimento, para a métrica **MAP**, o algoritmo que mais se destaca é o **BPRMF**, porém a variação não influencia totalmente na recomendação, pois é pequena quando comparadas aos algoritmos **ItemkNN** e **Mais Popular**. O algoritmo de recomendação baseado em conteúdo **Query & Search** apresentou os piores

resultados. Já no segundo experimento, o algoritmo **BPRMF** apresenta resultados superiores aos demais, seguido por **ItemkNN** e **Mais Popular**.

O mesmo comportamento é observado para as métricas de **precisão** e **recall**. Muito se deve ao fato da recomendação ser restritiva, ou seja, não há possibilidade de recomendação após o dia de divulgação da oferta. Logo, os algoritmos dependem do horário em que a divulgação das ofertas é iniciada e também do horário em que há dados disponíveis para que haja treinamento dos mesmos, com exceção o **Mais Popular**, que não tem influência de treino e teste, pois se baseia apenas no total de ofertas adquiridas até o momento e recomenda de acordo com este critério.

As métricas de **precisão** e **recall** medidas foram satisfatórias, em geral, devido a esparsidade dos dados. Considerando que, para usuários que efetuaram 5 ou mais compras, temos 6.450 usuários e 1.449 ofertas, a possibilidade de 9.346.050 compras de ofertas e a base de dados contém apenas 56.963 compras, observamos que a esparsidade chega a 99,4%. Como a análise é baseada no *feedback* implícito, em que a preferência do usuário é baseada nas compras que realizou, esse valor indica que há muitos usuários que não efetuaram aquisições de várias ofertas. Quando consideramos os usuários que efetuaram 10 ou mais compras, temos 1.753 usuários e 1.313 ofertas, com possibilidade de 2.301.689 compras, e a base de dados contém apenas 27.027 destas, observamos que a esparsidade chega a 99,9%, favorecendo, por vezes, o algoritmo **BPRMF** e desfavorecendo algoritmos como **ItemkNN** e **Query & Search** ainda mais, observando as métricas definidas anteriormente.

Logo, de acordo com a análise realizada neste capítulo, há ganhos justificáveis para que o algoritmo **BPRMF** seja utilizado na recomendação das ofertas, por apresentar resultados semelhantes ou, por vezes, melhores do que os outros algoritmos considerados. Além disso, o agrupamento das ofertas em 4 dias consecutivos beneficia esse, fazendo com que haja melhora nas métricas, porém com mais significância na métrica *recall*. Também considerar os usuários com 10 ou mais aquisições de ofertas e recomendar após as 10 horas de início de divulgação da oferta auxiliam na eficácia desse algoritmo.

6. Considerações Finais da Dissertação

Neste capítulo vamos expor as considerações finais referentes ao estudo dos algoritmos de recomendação, quando aplicados aos dados de SCC, a partir das métricas de precisão, *recall* e *MAP*. Em seguida, apresentam-se as limitações do atual trabalho e sugestões para trabalhos futuros, a fim de promover a continuidade da pesquisa documentada.

6.1. Conclusões

Neste trabalho estudamos algoritmos de recomendação a serem aplicados aos dados de SCC. Estes dados contem cerca de 117 mil vendas realizadas no período de 2010 a meados de 2012 e um total aproximado de 43 mil usuários que realizaram compras de ofertas, cuja divulgação por parte do SCC excedeu 1.500 ofertas locais, regionais e nacionais.

A partir da análise dos dados, foi possível verificar que existe um padrão de compras por parte dos usuários em relação as ofertas. De acordo com nossa análise, usuários estão mais propensos a adquirir ofertas referentes à categoria "Restaurante/Gastronomia", responsável por 80% das vendas do site.

Quanto à distribuição diária de vendas, verificamos que existem dois grandes picos: o primeiro tendo período situado entre as 8 e 10 horas e o segundo tendo período entre as 15 e 17 horas. Apesar desses picos isolados, ainda foi possível verificar que o número de compras possui pouca variação com o decorrer das horas, após o primeiro horário de pico. Constatamos ainda que o número de compras na parte da manhã é bastante inferior em relação ao número de compras durante o período da tarde/noite. Esse detalhe serviu de motivação para que pudéssemos incluir o fator hora de corte como um dos fatores a serem estudados na presente pesquisa. Esse fator simboliza o momento em que deve haver separação dos dados em treino e dados de testes dos algoritmos de recomendação, etapas essenciais para que pudéssemos avaliar os algoritmos, a partir das métricas de **precisão**, *recall* e *MAP*. Foram definidos três níveis para este fator: 2 horas, 3 horas e 10 horas após o início de divulgação da oferta. Esses níveis foram definidos nessas horas por representarem os momentos pré-picos de compras, no qual é possível obter dados de compras efetuadas por outros usuários, sendo esses dados utilizados como entrada nos algoritmos de recomendação.

Mostramos que as métricas **precisão**, *recall* e *MAP* apresentam os mesmos resultados: após as 10 horas do início de divulgação da oferta é o melhor nível desse fator, em todos os

algoritmos analisados, nos experimentos em que consideramos a divulgação da oferta coletiva por apenas um dia e agrupando 4 dias consecutivos de divulgação de ofertas. Muito se explica pelo fato da quantidade disponível de dados ser grande para treino, logo os algoritmos conseguem recomendar com mais qualidade.

Outro dado que nos ajuda a entender a melhoria das métricas no terceiro nível é que a variação do primeiro e segundo nível, quanto ao tamanho do conjunto de dados disponível para treino, é bastante pequena, variando entre 4 e 12%. Logo, essa variação é insignificante para os algoritmos de recomendação. Quando verificamos a variação do segundo para o terceiro nível do fator, o número de compras varia entre 40 a 70%, o que aumenta significativamente a base de treino para os algoritmos, fazendo com que esses sejam mais precisos e eficazes.

Além disso, ainda de acordo com a análise efetuada, verificamos que há um limiar mínimo de compras que os usuários efetuaram que influenciam também na recomendação, em que cerca de 6 mil usuários realizaram mais de 5 compras e cerca de 1.700 usuários efetuaram mais de 10 compras. Escolhemos esse fator para estudo pois é importante recomendar itens para usuários que possuem uma quantidade mínima de compras, para que possamos analisar o perfil do usuário e gerar recomendação para esses. Usuários com número de compras inferior a 5 dificultam a tarefa de recomendação, pois seu perfil não é bem definido, portanto estes não são considerados na presente pesquisa. Como alternativa para estes usuários, há a possibilidade de continuar com a recomendação dos itens mais populares, recomendação já padrão para o SCC avaliado. Poucos usuários possuem um número de compras superior a 10, o que faria com que o número de usuários atingidos pela recomendação fosse bastante reduzido.

Avaliando os níveis do fator número mínimo de compras por usuário, mostrou-se que, para a métrica *MAP* e algoritmos **BPRMF** e **ItemkNN**, não há, estatisticamente, diferenças na métrica quando alteramos os níveis do fator. Já para o algoritmo **Mais Popular**, o melhor nível para essa métrica ocorre quando o número mínimo de compras é igual a 5, pois o histórico de compras dos usuários apresenta maior quantidade de dados, logo o algoritmo se beneficia pelo destaque de uma oferta mais adquirida no dia e recomenda para todos os usuários. Já o algoritmo *Query & Search* apresenta os melhores resultados quando o número mínimo de compras é igual a 10, logo o algoritmo se beneficia pela quantidade de ofertas que

o usuário já adquiriu, facilitando a definição do perfil do usuário a partir da descrição das ofertas que foram adquiridas por este.

Analogamente ao observado na métrica *MAP*, os algoritmos **BPRMF** e **ItemkNN** apresentam resultados semelhantes quando analisamos a métrica *precisão*, sem comprovação estatística de que a alteração no nível do fator influência na métrica. Os algoritmos **Mais Popular** e **Query & Search** também possuem comportamento semelhante ao apresentado pela métrica *MAP*. A concordância quanto ao comportamento acontece pois a métrica *MAP* utiliza a *precisão* como parâmetro, logo o resultado era esperado.

Analisando a métrica *recall*, o algoritmo **BPRMF** apresentou melhor resultado quando o nível do número mínimo de compras foi igual a 5, pois esta métrica é influenciada pela quantidade de dados disponíveis para os testes. Os outros algoritmos apresentaram comportamento semelhante ao analisado nas métricas de *MAP* e *precisão*.

Já o fator número de itens a ser recomendado não influenciou nenhuma das métricas estudadas. Para os níveis 3, 4 ou 5 recomendações, as métricas *MAP*, *precisão* e *recall* apresentaram os mesmos resultados, não havendo predominância de nenhum dos níveis, no primeiro experimento. Quando analisamos o segundo experimento, divergências ocorreram, beneficiando por vezes o terceiro nível deste fator, porém com variação muito pequena para a recomendação (variação de, no máximo, 2% em relação aos outros níveis). Logo, podemos afirmar que o comportamento dos algoritmos é o mesmo, independente de quantas recomendações forem efetuadas.

Analisando apenas o fator algoritmo, percebeu-se que o algoritmo **Mais Popular**, que utiliza abordagem de recomendação não-personalizada, se destaca para a métrica *MAP*. Muito se deve ao fato de que há ofertas que são mais adquiridas em determinados dias, ou seja, na maioria das vezes, uma oferta tem preferência sobre as outras, sendo essa oferta recomendada para os usuários. Para as métricas *precisão* e *recall*, não há evidências estatisticamente significativas de que os algoritmos **BPRMF**, **ItemkNN** e **Mais Popular** se destaquem. O algoritmo **Query & Search** apresentou os piores resultados para as métricas estudadas.

Conforme mencionado no Capítulo de Resultados, há ganhos justificáveis para que o algoritmo **BPRMF** seja utilizado na recomendação das ofertas, considerando o agrupamento das ofertas em 4 dias consecutivos que proporciona melhoria nas métricas estudadas, porém com mais significância na métrica *recall*. Também considerar os usuários com 10 ou mais

aquisições de ofertas e recomendar após as 10 horas de início de divulgação da oferta auxiliam na eficácia desse algoritmo.

6.2. Limitações e Trabalhos Futuros

As principais limitações do trabalho são:

- Utilização de apenas alguns algoritmos de recomendação, representando a classe dos algoritmos personalizados que são baseados em Filtragem Colaborativa (baseado em modelo **ItemkNN** e baseado em memória **BPRMF**), baseados em Conteúdo ou Sensíveis ao Contexto (**Query & Search**) e Não-Personalizado (**Mais Popular**);
- Falta de validação cruzada para estimação dos parâmetros utilizados no *framework* MyMediaLite;
- Falta de validação cruzada para estimação dos níveis dos fatores: "número mínimo de compras por usuário", "hora de corte" e "número de itens a ser recomendado";
- A utilização de outros algoritmos de cálculo de similaridade de conteúdo para a recomendação por conteúdo ou sensível ao contexto (**Query & Search**).

Como trabalhos futuros, têm-se:

- Investigar algoritmos que se propõem a resolver problemas de *cold start* para a recomendação baseada em conteúdo;
- Investigar algoritmos que se propõem a resolver o problema de esparsidade nos dados de *feedback* implícito, como nos SCC;
- Propor e avaliar algoritmos que utilizam abordagens híbridas, para que se tenha recomendação baseada no conteúdo das ofertas e também utilizando o histórico de compras do usuário, bem como recomendações baseadas em contexto (localização do usuário, idade, sexo, se possui família), pois estas informações podem ser relevantes na obtenção do perfil do usuário com relação as ofertas divulgadas diariamente.

7. Referências Bibliográficas

- [1]. LINDER, G., SMITH, B., YORK, J. 2003. Amazon.com recommendations: item-to-item collaborative filtering. In *Internet Computing, IEEE*, Volume 7 Issue 1, 76-80.
- [2]. ECK, D., LAMERE, P., BERTIN-MAHIEUX, T., GREEN, S. Automatic Generation of Social Tags for Music Recommendation. 2007. In *Advance in neural information processing systems*, 1-8.
- [3]. SCHAFER, J. Ben, KONSTAN, J., RIEDL, J. Recommender Systems in E-Commerce. 1999. In *EC '99 Proceedings of the 1st ACM conference on Electronic commerce*, 158-166.
- [4]. TAKÁCS, G., PILÁSZY, I., NÉMETH, B., TIKK, D. Scalable Collaborative Filtering Approaches for Large Recommender Systems. 2009. In *Journal of Machine Learning Research, Volume 10*, 623-656.
- [5]. RENDLE, C., FREUDENTHALER, C., GANTNER, Z., SCHMIDT-THIEME, L. BPR: Bayesian personalized ranking from implicit feedback. 2009. In *UAI '09 Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, 452-461.
- [6]. DHOLAKIA, U., KIMES, S. Daily Deal Fatigue or Unabated Enthusiasm? A Study of Consumer Perceptions of Daily Deal Promotions. 2011. In *Social Science Research Network*.
- [7]. MAO, Y., SANDHOLM, T., WANG, C., APERJIS, C., HUBERMAN, B. Collective attention and the dynamics of group deals. 2012. In *Proceedings of the 21st international conference companion on World Wide Web (WWW '12 Companion)*, 1205-1212.
- [8]. DESROSIERS, C., KARYPIS, G. A comprehensive Survey of Neighborhood-based Recommendation Methods. 2009. In *Handbook on Recommender Systems*, Kantor, Ricci, Rokach et Shapira éditeurs, Springer.
- [9]. BYERS, J., MITZENMACHER, M., ZERVAS, G. Daily Deals: Prediction, Social Diffusion, and Reputation Ramifications. 2012. In *Web Search and Data Mining 2012*. 543-552.
- [10]. KONSTAN, J., MILLER, B., MALTZ, D., HERLOCKER, J., GORDON, L. GroupLens: Applying Collaborative Filtering to Usenet News. 1997. In *Communications of the ACM Volume 40 Issue 3*, 77-87.
- [11]. RESNICK, P., IACOVOU, N., SUCHAK, M., BERGSTROM, P., RIEDL, J. GroupLens: an open architecture for collaborative filtering of netnews. 1994. In *CSCW '94*

- Proceedings of the 1994 ACM conference on Computer supported cooperative work, 175-186.
- [12]. SARWAR, B., KARYPIS, G., KONSTAN, J., RIEDL, J. Incremental Singular Value Decomposition Algorithms for Highly Scalable Recommender Systems. 2002. In *Fifth International Conference on Computer and Information Science*, 27-28.
- [13]. VAN MEETEREN, R., VAN SOMEREN, M. Using Content-based Filtering for Recommendation. 2000. In *ECML/MLNET Workshop on Machine Learning and the New Information Age*, 47-56.
- [14]. JÄSCHKE, R., MARINHO, L., HOTHO, A., SCHMIDT-THIEME, L., STUMME, G. Tag Recommendations in Folksonomies. 2007. In *PKDD 2007 Proceedings of the 11th European conference on Principles and Practice of Knowledge Discovery in Databases*, 506-514.
- [15]. CREMONESI, P., KOREN, Y., TURRIN, R. Performance of Recommender Algorithms on Top-N Recommendation Task. 2010. In *RecSys '10 Proceedings of the fourth ACM conference on Recommender systems*, 39-46.
- [16]. DEERWESTER, S., DUMAIS, S., FURNAS, G., LANDAUER, T., HARSHMAN, R. Indexing by latent semantic analysis. 1990. In *Journal Of The American Society For Information Science, Volume 41, Issue 6*, 391-407.
- [17]. VILLEGAS, N., MÜLLER, H., MUÑOZ, J., LAU, A. A dynamic context management infrastructure for supporting user-driven web integration in the personal web. 2011. In *CASCON '11 Proceedings of the 2011 Conference of the Center for Advanced Studies on Collaborative Research*, 200-214.
- [18]. BILLSUS, D. A Hybrid User Model for News Story Classification. 1999. In *UM '99 Proceedings of the seventh international conference on User modeling*, 99-108.
- [19]. RESNICK, P., VARIAN, H. R. Recommender Systems. 1997. In *Communications of the ACM*, 56-58.
- [20]. RESNICK, P., KUWABARA, K., ZECKHAUSER, R., FRIEDMAN, E. Reputations Systems. 2000. In *Communications of the ACM*, 45-48.
- [21]. PAPAGELIS, M., PLEXOUSAKIS, D. Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents. 2005. In *Engineering Applications of Artificial Intelligence*, 781-789.

- [22]. POIRIER, D., FESSANT, F., TELLIER, I. Reducing the Cold-start Problem in Content-Recommendation Through Opinion Classification. 2010. In *ACM International Conference Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 204-207.
- [23]. SARWAR, B., KARYPIS, G., KONSTAN, J., REIDL, J. Item-based collaborative filtering recommendation algorithms. 2001. In *Proceedings of the 10th international conference on World Wide Web*, 285-295.
- [24]. GUNAWARDANA, A., MEEK, C. A Unified Approach to Building Hybrid Recommender Systems. 2009. In *RecSys '09 Proceedings of the third ACM conference on Recommender systems*, 117-124.
- [25]. BURKE, R. Hybrid Recommender Systems: Survey and Experiments. 2002. In *User Modeling and User-Adapted Interaction*, 331-370.
- [26]. BREESE, J., HECKERMAN, D., KADIE, C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. 1998. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, 43-52.
- [27]. WICKRAMARATNA, K., KUBAT, M., PREMARATNE, K. Predicting Missing Items in Shopping Carts. 2009. In *IEEE Transactions on Knowledge & Data Engineering*, 985-998.
- [28]. VARGAS, S., CASTELLS, P. Rank and relevance in novelty and diversity metrics for recommender systems. 2011. In *RecSys '11 Proceedings of the fifth ACM conference on Recommender systems*, 109-116.
- [29]. RICCI, F., ROKACH, L., SHAPIRA, B., KANTOR, P.B. 2009. In *Handbook on Recommender Systems*, Kantor, Ricci, Rokach et Shapira éditeurs, Springer.
- [30]. CAZELLA, C., NUNES, Maria Augusta S. N., REATEGUI, Eliseo Berni. A Ciência da Opinião: Estado da arte em Sistemas de Recomendação. 2010. In *Congresso da Sociedade de Computação: Computação Verde – desafios científicos e tecnológicos*, 52.
- [31]. SUTTON, S., BARTO, A. 1998. In *Reinforcement Learning: An Introduction*, Bradford Book.
- [32]. LACERDA, Anisio M. Revenue optimization and customer targeting in daily-deals sites. 2013. 97 f. Tese (Doutorado em Ciência da Computação) - Universidade Federal de Minas Gerais, 2013.
- [33]. EBRAHIMI, Sahar. SmarterDeals: A Context-aware Deal Recommendation System based on the SmarterContext Engine. 2012. 68 f. Dissertação (Mestrado em Ciência da Computação) - Tarbiat Moallem University, 2012.

- [34]. Yelp's Academic Data Set. Disponível em <https://www.yelp.com.br/academic_dataset>. Acessado em 12 jul. 2014.
- [35]. VERA, L., LADEIRA, R., OLIVEIRA, R. Sites de Compras Coletivas e o Processo de Inovação no Mercado de Vendas pela Internet. 2013. In *Simpósio de Excelência em Gestão e Tecnologia*.
- [36]. Facebook entra no mercado de compras coletivas. Disponível em <<http://www1.folha.uol.com.br/mercado/907659-facebook-entra-no-mercado-de-compras-coletivas.shtml>>. Acessado em 26 nov. 2014.
- [37]. QueroDois Compras Coletivas. Disponível em <<http://www.quero2.com.br/>>. Acessado em 26 nov. 2014.
- [38]. Compras Coletivas. In *Wikipédia*. Disponível em <http://pt.wikipedia.org/wiki/Compras_coletivas>. Acessado em 26 nov. 2014.
- [39]. UOL: Donos de sites de compras coletivas mudam estratégia para crescer. In *UOL*. Disponível em <<http://economia.uol.com.br/ultimas-noticias/redacao/2012/08/29/ex-donos-de-sites-de-compras-coletivas-mudam-estrategia-para-crescer.jhtm>>. Acessado em 26 nov. 2014.
- [40]. Teste U de Mann-Whitney. In *Wikipédia*. Disponível em <http://pt.wikipedia.org/wiki/Teste_U_de_Mann-Whitney>. Acessado em 26 nov. 2014.
- [41]. Paraná Online. Sites de Compras Coletivas com novidades para recuperar público. Disponível em <<http://www.parana-online.com.br/editoria/cidades/news/673898/?noticia=SITES+DE+COMPRAS+COLETIVAS+COM+NOVIDADES+PARA+RECUPERAR+PUBLICO>>. Acessado em 26 nov. 2014.
- [42]. Recommender Systems. Item-based collaborative filtering. Disponível em <http://www.cs.carleton.edu/cs_comps/0607/recommend/recommender/itembased.html>. Acessado em 26 nov. 2014.

Apêndice A

Funcionamento de um Sites de Compras Coletivas

Os SCC seguem o modelo comum de sites de anúncios. Na Figura 39, temos uma página típica de ofertas, com seus atributos, de acordo com o site *QueroDois*, parceiro que forneceu os dados para estudo. Consideram-se atributos da oferta a descrição da mesma, o valor da compra, o tempo de validade e o número de compradores. Tais atributos podem ser vistos na Figura 40.



Figura 39 - Página referente ao Site de Compras Coletivas.



Figura 40 - Atributos de uma oferta em Site de Compras Coletivas.

As etapas comuns no processo de um usuário comprar uma oferta são as seguintes:

- O usuário efetua *login* na página, caso já possua cadastro. Caso contrário, efetua um cadastro com algumas informações básicas (endereço, telefone, CPF, cidade, *login* e senha);
- Escolhe, dentre as ofertas que ainda não expiraram, a(s) oferta(s) desejada(s);
- Visualiza as regras de compra das ofertas. Tais regras geralmente são básicas, indicando apenas a validade do uso da oferta, as cidades que estão contempladas, o endereço do local, os dias de funcionamento e/ou dias de validade do item;
- Efetua o pagamento via cartão de crédito, obrigatoriamente;
- Recebe um cupom, correspondente à sua compra. Este cupom geralmente deve ser apresentado no estabelecimento que oferece o item e/ou serviço;
- É agregada à página pessoal do usuário a informação da oferta que foi comprada, bem como as informações para impressão do cupom, valor economizado, data de validade da oferta e outras;

Por fim, o usuário pode marcar a oferta como utilizada e fornecer algumas informações acerca do estabelecimento/serviço prestado, enviando um *feedback* para a empresa que hospeda o SCC.

Ainda há algumas opções nos SCC que procuram personalizar o serviço, tais como:

- Lista de cidades que possuem parceria com o SCC, sendo possível listar apenas ofertas para uma localidade em específico;
- Cadastro de e-mail para que as ofertas sejam enviadas, assim que cadastradas;
- Verificar quais foram as ofertas recentes, juntamente com seus atributos.

Para um novo usuário no SCC, apenas as seguintes funcionalidades são oferecidas:

- Cadastro de e-mail para recebimento de oferta;
- Visualização das ofertas de todas as localidades disponíveis;
- Visualização de oferta individualmente e seus atributos, cadastro de conta para efetuar compras.

Como bonificação, os SCC incentivam a indicação de ofertas aos amigos, via cadastramento do e-mail e contas nas redes sociais em geral (*Facebook, Twitter* e outras).

Com base nesta descrição de funcionamento de um SCC, vemos que os usuários não possuem nenhum tipo de recomendação personalizada de acordo com as compras que efetuou recentemente. Esta recomendação de ofertas poderia tanto ser feita na página principal (como a indicação de um filme no *Netflix*, ou produto na *Amazon*) ou mesmo no envio de ofertas diárias por e-mail, dando maior ênfase nos produtos que tiverem mais afinidade com o perfil de compras do usuário.

Apêndice B

Experimentos executados

De acordo com a metodologia definida no Capítulo 4 - Metodologia, definimos o espaço amostral com 50 dias de ofertas, variando o mês e o ano, aleatoriamente. Além disso, definimos ainda como rodada de experimento um conjunto de 10 dias. Estas rodadas, juntamente com os dias e o número de anúncios que estavam ativos neste dia estão listado a seguir.

1. Rodada 1

- a. 2012-06-06 (23 anúncios)
- b. 2012-06-04 (24 anúncios)
- c. 2012-05-26 (15 anúncios)
- d. 2012-05-04 (19 anúncios)
- e. 2012-04-13 (21 anúncios)
- f. 2012-03-30 (28 anúncios)
- g. 2012-03-16 (28 anúncios)
- h. 2012-03-09 (25 anúncios)
- i. 2012-02-24 (21 anúncios)
- j. 2012-01-30 (32 anúncios)

2. Rodada 2

- a. 2012-06-10 (16 anúncios)
- b. 2012-06-02 (13 anúncios)
- c. 2012-05-26 (15 anúncios)
- d. 2012-05-13 (14 anúncios)
- e. 2012-05-06 (11 anúncios)
- f. 2012-03-11 (16 anúncios)
- g. 2012-02-05 (13 anúncios)
- h. 2011-12-31 (10 anúncios)
- i. 2011-12-17 (17 anúncios)
- j. 2012-01-12 (24 anúncios)

3. Rodada 3

- a. 2011-10-27 (33 anúncios)
- b. 2011-12-16 (33 anúncios)
- c. 2012-05-18 (25 anúncios)
- d. 2011-06-10 (20 anúncios)
- e. 2011-08-24 (18 anúncios)
- f. 2012-04-22 (18 anúncios)
- g. 2011-07-04 (15 anúncios)
- h. 2012-06-01 (22 anúncios)
- i. 2011-09-10 (23 anúncios)
- j. 2011-12-27 (24 anúncios)

4. Rodada 4

- a. 2012-03-08 (30 anúncios)
- b. 2012-03-15 (29 anúncios)
- c. 2011-12-14 (28 anúncios)
- d. 2011-10-03 (22 anúncios)
- e. 2011-10-21 (22 anúncios)
- f. 2012-05-01 (18 anúncios)
- g. 2011-11-12 (16 anúncios)
- h. 2012-04-15 (12 anúncios)
- i. 2012-02-22 (21 anúncios)
- j. 2011-11-14 (23 anúncios)

5. Rodada 5

- a. 2012-02-15 (32 anúncios)
- b. 2012-04-04 (30 anúncios)
- c. 2011-10-31 (30 anúncios)
- d. 2012-03-28 (26 anúncios)
- e. 2012-03-13 (26 anúncios)
- f. 2012-04-02 (24 anúncios)
- g. 2011-11-21 (20 anúncios)
- h. 2012-01-03 (18 anúncios)
- i. 2011-09-19 (16 anúncios)

j. 2011-09-21 (15 anúncios)

Definimos ainda um conjunto de ensaios a serem executados, a partir da combinação dos fatores escolhidos para o presente estudo. Este conjunto pode ser visualizado na Tabela 4. Vale a ressalva de que este conjunto de experimentos foi executado para cada dia definido anteriormente.

Tabela 4 - Conjunto de ensaios do experimento.

N° do Experimento	Algoritmo	Hora de Corte	N° de Itens Recomendados	N° Mínimo de compras
1	<i>ItemkNN</i>	2	3	5
2	<i>ItemkNN</i>	2	3	10
3	<i>ItemkNN</i>	2	4	5
4	<i>ItemkNN</i>	2	4	10
5	<i>ItemkNN</i>	2	5	5
6	<i>ItemkNN</i>	2	5	10
7	<i>ItemkNN</i>	3	3	5
8	<i>ItemkNN</i>	3	3	10
9	<i>ItemkNN</i>	3	4	5
10	<i>ItemkNN</i>	3	4	10
11	<i>ItemkNN</i>	3	5	5
12	<i>ItemkNN</i>	3	5	10
13	<i>ItemkNN</i>	10	3	5
14	<i>ItemkNN</i>	10	3	10
15	<i>ItemkNN</i>	10	4	5
16	<i>ItemkNN</i>	10	4	10
17	<i>ItemkNN</i>	10	5	5
18	<i>ItemkNN</i>	10	5	10
19	BPRMF	2	3	5
20	BPRMF	2	3	10

21	BPRMF	2	4	5
22	BPRMF	2	4	10
23	BPRMF	2	5	5
24	BPRMF	2	5	10
25	BPRMF	3	3	5
26	BPRMF	3	3	10
27	BPRMF	3	4	5
28	BPRMF	3	4	10
29	BPRMF	3	5	5
30	BPRMF	3	5	10
31	BPRMF	10	3	5
32	BPRMF	10	3	10
33	BPRMF	10	4	5
34	BPRMF	10	4	10
35	BPRMF	10	5	5
36	BPRMF	10	5	10
37	<i>Query & Search</i>	N/A	3	5
38	<i>Query & Search</i>	N/A	3	10
39	<i>Query & Search</i>	N/A	4	5
40	<i>Query & Search</i>	N/A	4	10
41	<i>Query & Search</i>	N/A	5	5
42	<i>Query & Search</i>	N/A	5	10
43	Mais Popular	2	3	5

44	Mais Popular	2	3	10
45	Mais Popular	2	4	5
46	Mais Popular	2	4	10
47	Mais Popular	2	5	5
48	Mais Popular	2	5	10
49	Mais Popular	3	3	5
50	Mais Popular	3	3	10
51	Mais Popular	3	4	5
52	Mais Popular	3	4	10
53	Mais Popular	3	5	5
54	Mais Popular	3	5	10
55	Mais Popular	10	3	5
56	Mais Popular	10	3	10
57	Mais Popular	10	4	5
58	Mais Popular	10	4	10
59	Mais Popular	10	5	5
60	Mais Popular	10	5	10