

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

**AGENTE DE INTERFACE PERSUASIVO BASEADO EM
CLASSIFICADORES BAYESIANOS E TEORIA DA ESCOLHA
SOCIAL PARA O PROBLEMA DE COLD-START EM
SISTEMAS DE RECOMENDAÇÃO**

por

Marcus Vinicius Freire Guimaraes

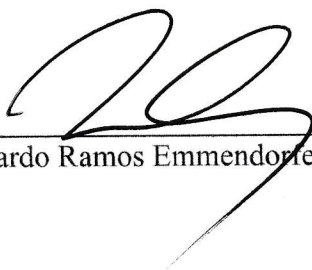
Rio Grande, Agosto de 2021.

“AGENTE DE INTERFACE PERSUASIVO BASEADO EM CLASSIFICADORES BAYESIANOS E TEORIA DA ESCOLHA SOCIAL PARA O PROBLEMA DE COLD-START EM SISTEMAS DE RECOMENDAÇÃO”

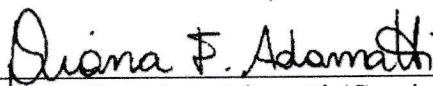
Tese apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal do Rio Grande - FURG, como requisito parcial para obtenção do Grau de Doutor. Área concentração: Modelagem Computacional.

Aprovado em:

BANCA EXAMINADORA



Prof. Dr. Leonardo Ramos Emmendorfer (Orientador – FURG)



Profa. Dra. Diana Francisca Adamatti (Coorientadora – FURG)



Profa. Dra. Viviane Leite Dias de Mattos (FURG)



Prof. Dr. Paul Gerhard Kinas (FURG)



Prof. Dr. Jomi Fred Hübner (UFSC)

Ficha Catalográfica

G963a Guimaráes, Marcus Vinicius Freire.

Agente de interface persuasivo baseado em classificadores Bayesianos e Teoria da Escolha Social para o problema de *Cold-start* em sistemas de recomendação / Marcus Vinicius Freire Guimaráes. – 2021.

73 f.

Tese (doutorado) – Universidade Federal do Rio Grande – FURG, Programa de Pós-Graduação em Modelagem Computacional, Rio Grande/RS, 2021.

Orientador: Dr. Leonardo Ramos Emmendorfer.

Coorientadora: Dra. Diana Francisca Adamatti.

1. Sistemas de Recomendação 2. Tecnologia Persuasiva
3. Classificador Bayesiano 4. Teoria Escolha Social 5. *E-commerce*
6. *Cold-start* I. Emmendorfer, Leonardo Ramos II. Adamatti, Diana Francisca III. Título.

CDU 004

Catálogo na Fonte: Bibliotecário José Paulo dos Santos CRB 10/2344

ATA DE DEFESA DE TESE

ATA Nº 05/2021

Ao décimo terceiro dia do mês de agosto, através de videoconferência, realizou-se a quinta sessão de defesa de tese do ano de dois mil e vinte um do Programa de Pós-Graduação em Modelagem Computacional. O doutorando, **Marcus Vinicius Freire Guimarães**, apresentou e defendeu a tese **“AGENTE DE INTERFACE PERSUASIVO BASEADO EM CLASSIFICADORES BAYESIANOS E TEORIA DA ESCOLHA SOCIAL PARA O PROBLEMA DE COLD-START EM SISTEMAS DE RECOMENDAÇÃO”** como requisito parcial para a obtenção do grau de Doutor em Modelagem Computacional. A sessão foi aberta às quinze horas pelo Prof. Dr. Leonardo Ramos Emmendorfer, orientador da tese e presidente da Comissão de Avaliação. A referida Comissão esteve integrada, além do presidente e orientador, pela Profa. Dra. Diana Francisca Adamatti (coorientadora) da FURG, Profa. Dra. Viviane Leite Dias de Mattos da FURG, Prof. Dr. Paul Gerhard Kinas da FURG e Prof. Dr. Jomi Fred Hübner da UFSC. Após cada integrante da Banca ter arguido o doutorando, e este ter tido a oportunidade de responder a cada um, a Comissão reuniu-se para deliberar sobre o conceito a ser atribuído ao trabalho. A Comissão de Avaliação considerou o candidato aprovado neste requisito parcial para a obtenção do grau de Doutor em Modelagem Computacional. Por fim, o presidente publicou e encerrou a sessão da qual eu, Emanuel da Silva Diaz Estrada, na qualidade de Coordenador do Programa de Pós-Graduação em Modelagem Computacional, lavrei a presente ata que vai assinada por todos os membros da Comissão de Avaliação. Rio Grande, 13 de agosto de 2021.

Documento assinado digitalmente
 Leonardo Ramos Emmendorfer
Data: 23/09/2021 11:52:20-0300
Verifique em <https://verificador.iti.br>

Prof. Dr. Leonardo Ramos Emmendorfer
Orientador

Documento assinado digitalmente
 Diana Francisca Adamatti
Data: 29/09/2021 08:17:47-0300
Verifique em <https://verificador.iti.br>

Profa. Dra. Diana Francisca Adamatti
Coorientadora

Documento assinado digitalmente
 Viviane Leite Dias de Mattos
Data: 24/09/2021 20:14:50-0300
Verifique em <https://verificador.iti.br>

Profa. Dra. Viviane Leite Dias de Mattos
Membro Interno

Documento assinado digitalmente
 PAUL GERHARD KINAS
Data: 24/09/2021 15:56:24-0300
Verifique em <https://verificador.iti.br>

Prof. Dr. Paul Gerhard Kinas
Membro Interno

Documento assinado digitalmente
 Jomi Fred Hubner
Data: 23/09/2021 13:15:02-0300
CPF: 596.334.769-68
Verifique as assinaturas em <https://v.ufsc.br>

Prof. Dr. Jomi Fred Hübner
Membro Externo

Prof. Dr. Emanuel da Silva Diaz Estrada
Coordenador do Programa de Pós-Graduação
Em Modelagem Computacional

RESUMO

Esta tese visa apresentar uma abordagem para solucionar o problema de *Cold-Start* em sistemas de recomendação de conteúdo. Este problema caracteriza-se por tentar responder o seguinte questionamento: como e o que recomendar para usuários novos em uma plataforma digital? Para isso, diversas abordagens já foram propostas na literatura, contudo, ainda sem um consenso para solucionar o problema. Entretanto, para que um sistema de recomendação seja desenvolvido, são necessários dados. Porém, estas informações não são facilmente acessíveis ou disponibilizadas por empresas. As principais tecnologias de recomendação de conteúdo são proprietárias e não encontradas em domínio público. Para contornar este cenário, esta tese apresenta uma contribuição onde um algoritmo, baseado em interações de agentes autônomos, é capaz de simular um conjunto de dados com propósito de atuação em situações de comércio eletrônico. Consequentemente, após a validação das informações geradas, um novo passo para contribuir com o estado da arte na área de sistemas persuasivos e de recomendação foi a elaboração de um agente de interface baseado em classificadores Bayesianos e Teoria da Escolha Social para processar, refinar e recomendar conteúdos personalizados. Tais processos atuarão como filtragem dos dados em duas etapas. A primeira com o uso de classificadores Bayesianos, visa a refinamento das informações em tempo real do coletivo próximo às características do usuário. Em um segundo momento, ocorre o ranqueamento de preferências individuais em busca de uma regra coletiva. Este segundo filtro será obtido com o uso da Teoria da Escolha Social através do método de contagem de Borda. Os resultados encontrados demonstram que o algoritmo gerador de dados assemelha-se à alguns bancos de dados disponibilizados para fins de pesquisa assim como o agente de interface é capaz de recomendar conteúdos personalizados de acordo com as características do usuário que estiver acessando a plataforma naquele exato momento.

Palavras-chaves: Sistemas de Recomendação, Tecnologia Persuasiva, Classificador Bayesiano, Teoria Escolha Social, E-commerce, Cold-start

ÍNDICE

1	Introdução	6
1.1	Relevância e Contribuições	9
1.2	Hipótese científica	9
1.3	Objetivos	10
1.3.1	Objetivo Geral	10
1.3.2	Objetivos Específicos	10
1.4	Organização do trabalho	13
2	Revisão sistemática	14
3	Revisão teórica	21
3.1	Sistemas de Recomendação	21
3.1.1	Tipos de sistemas de recomendação	22
3.1.2	Filtros colaborativos	22
3.1.3	Filtros baseados em conteúdo	24
3.1.4	O problema de cold-start	25
3.1.5	Privacidade em sistemas de recomendação	25
3.2	Classificadores Bayesianos	27
3.3	Teoria da Escolha Social	28
3.3.1	Método de contagem de Borda	30
3.4	Simulação e Sistemas multiagente	30
4	Métodos	32
4.1	Processo de aquisição de dados	32
4.2	Protocolo de privacidade do usuário	35
4.2.1	Processo gerador de dados baseado em agentes	36
4.2.2	Domínios e características avaliadas	41
4.3	Classificadores Bayesianos como filtro colaborativo em problemas de <i>cold-start</i>	41
4.4	Teoria da Escolha Social como filtro baseado em conteúdo	43
5	Resultados e Discussão	45
5.1	Análise dos dados sintéticos	45
5.1.1	Obtenção dos dados	45
5.1.2	Análise descritiva dos dados	46
5.2	Aplicação modelo	55
5.3	Processamento dos filtros: Agente de Interface	56
6	Considerações finais	62
6.1	Limitações e Trabalhos futuros	62

LISTA DE FIGURAS

Figura 1:	Abstração dos processos para recomendação de conteúdo com problema de <i>cold-start</i> (Fonte: Autor)	8
Figura 2:	Áreas do conhecimento abordadas para solução do problema estudado. (Fonte: Autor)	12
Figura 3:	Abstração de um sistema de recomendação, (Fonte: Autor)	22
Figura 4:	Abstração do processo de filtro colaborativo entre usuários.	23
Figura 5:	Abstração do processo de filtro baseado em conteúdo.	24
Figura 6:	Mapa de calor para análise de usabilidade, (Fonte: [82])	26
Figura 7:	Interação Usuário/Plataforma com atuação de agente de interface. (Fonte: [9] Adaptado)	31
Figura 8:	Elementos acessados no protocolo de identificação do usuário.	36
Figura 9:	Fluxograma do processo gerador de dados sintéticos baseado na interação entre agentes.	38
Figura 10:	Estrutura dos processos no algoritmo de simulação baseada em agentes.	40
Figura 11:	Estrutura gráfica dos modelos avaliados.	43
Figura 12:	Distribuição do acesso aos departamentos.	47
Figura 13:	Estudo da correlação entre as simulações e acesso aos departamentos	49
Figura 14:	Amostra da distribuição do acesso aos itens contidos em departamento.	50
Figura 15:	Avaliação da distribuição de acesso referente às horas e dias da semana	51
Figura 16:	Comparação entre simulação e literatura referente aos navegadores	52
Figura 17:	Comparação entre simulação e literatura referente aos tipos de dispositivos	53
Figura 18:	Distribuição dos reviews da Amazon	54
Figura 19:	Comparação da distribuição dos resultados de reviews.	54
Figura 20:	Estrutura de processamento de informações e recomendação de conteúdo	55
Figura 21:	Análise do agente de interface e processamento Bayesiano. Exemplo 1.	57
Figura 22:	Análise do agente de interface e processamento Bayesiano. Exemplo 2.	58
Figura 23:	Análise do agente de interface e processamento Bayesiano. Exemplo 3.	58
Figura 24:	Análise do agente de interface e raqueamento de itens	60
Figura 25:	Análise do agente de interface e raqueamento de itens comparado com outros usuários	60

LISTA DE TABELAS

1	Revisão sistemática em bases científicas	15
Tabela 2:	Contagem de Borda para n=3 itens	30
3	Desempenho no processo gerador dos dados	45
4	Análise incremental e desempenho computacional	46
5	Dados de acesso coletadas pelo protocolo	57

LISTA DE ABREVIATURAS

FBM	Fogg Behavior Model
SCT	Social Choice Theory

1 INTRODUÇÃO

É seguro afirmar que o atual desenvolvimento computacional proporcionou o surgimento de uma característica peculiar: o grande volume de dados produzidos e armazenados. Assim como a quantidade de informação está em uma crescente constante, o processamento de tanta informação também pode resultar em uma tarefa trabalhosa.

Devida esta situação, algoritmos cada vez mais elaborados devem ser implementados com a intenção de refinar a pesquisa do usuário na busca por algum conhecimento. Este processo de desenvolver algoritmos e sistemas capazes de garimpar a melhor informação com base em informações prévias do usuário a fim de transformar toda informação armazenada em conhecimento útil é denominado *sistemas de recomendação* [80].

Diversos serviços digitais apresentam mecanismos para recomendação de conteúdo e estas ações passam silenciosamente pelos usuários. Facebook, por exemplo, constantemente recomenda novas pessoas, com o intuito de aumentar sua rede de contatos e através de seus algoritmos recalibrar o sistema e continuar o processo de recomendação. Por outro lado, o Youtube, a plataforma de serviços de vídeo mais acessada no mundo, recomenda vídeos relacionados ao interesse do usuário. Spotify recomenda músicas de acordo com o gosto ou procura prévia, Netflix recomenda novos filmes e séries e assim como praticamente todo e qualquer serviço digital.

O uso de plataformas digitais registram as "pegadas" que o usuário deixa durante o processo de busca por conteúdo [45],[81]. Seja o tempo de acesso, palavras-chave, assuntos relacionados, etc. Com essa informação o sistema é capaz de recalibrar as preferências do usuário e continuar o processo de recomendação em tempo real.

Entretanto, esta tarefa de recomendar conteúdo de acordo com os interesses implícitos do usuário não é algo trivial. É necessário utilizar diversos processos para refinar as intenções do usuário e assegurar a correta identificação de padrões comportamentais e características pessoais, para então adotar o melhor método de recomendação. Estes processos de identificação comportamental podem ser implementados de diversas maneiras [64],[92],[58],[17],[105].

Com isso, uma quantidade suficiente de informação do usuário deve estar disponível para o sistema processar tais conteúdos. Esta etapa é crucial para o desenvolvimento de sistemas de recomendação de conteúdo. Tais sistemas, utilizam procedimentos capazes de estimular seu uso e consequentemente recomendar itens que possam apresentar maior interesse do usuário.

Contudo, o problema não se finaliza na quantidade de informação do usuário. A veracidade e qualidade das informações coletadas são, muitas vezes, desconhecidas, visto que grandes empresas não tornam seus bancos de dados públicos para que processos de recomendação de conteúdo sejam avaliados e implementados. Isso é compreensível, dado o ponto de vista jurídico, e preservando assim a privacidade do usuário. Contudo, isso é um fator limitante no processo de implementação de sistemas de recomendação. Com isso, a falta de banco de

dados se torna um dos principais problemas.

Para contornar essa situação será abordado nessa tese o desenvolvimento de um algoritmo baseado em agentes que simule as intenções de compra de usuários de uma plataforma de comércio eletrônico e produza um banco de dados capaz de atender a demanda de desenvolvimento e implementações de processos persuasivos para elaboração de sistemas de recomendação de conteúdos em tempo real.

Neste contexto, uma abordagem através de agentes fornece uma dinâmica no processo de geração de informações. Um sistema multiagente (MAS, *Multiagent System*) consiste em uma coleção de entidades autônomas, denominadas *agentes* que interagem em um ambiente compartilhado [95]. Neste trabalho, o ambiente em comum é uma comunidade arbitrária de utilizadores de serviços de comércio eletrônico. Dessa forma, tais entidades autônomas serão identificadas como usuários da plataforma e poderão interagir, influenciar e ser influenciada de acordo com seus interesses.

Consequentemente, toda informação produzida pelo sistema baseado em agentes será utilizada como base para o desenvolvimento de processos com intenção em solucionar o problema de recomendação de conteúdo para usuários novos ou desconhecidos, denominado como problema de cold-start. Mais além, novos processos persuasivos podem explorar esta base de dados obtida, visto que representa informações coletadas de usuários (agentes) autônomos contendo intenções de compra, revisão de itens e registros de compartilhamento de informação entre seus pares. Tudo simulado através de interações com propriedades não convergentes¹.

Dados são ativos importantes, e a grande maioria das vezes não são de fácil acesso. Grandes empresas sabem disso, do fato de quanto mais informações forem coletadas de seus usuários, serviços mais persuasivos e personalizados podem ser entregues. Contudo, com esta abordagem proposta, o conjunto de dados sintéticos obtido pode atuar como um primeiro passo para modelagem de sistemas baseados em recomendação de conteúdo.

Diante disto, a aplicação de métodos estatísticos e computacionais que auxiliem na filtragem da informação é a etapa seguinte a ser adotada. Para isto, será utilizado o processo de classificação bayesiana como filtro colaborativo, ou seja, detectar quais departamentos dentro de um bando de dados modelo com características de comércio eletrônico apresentam maior probabilidade de compra dado características em comum dos usuários avaliados.

Posteriormente, o ranqueamento de informação será realizado através de processos matemáticos de acordo com a Teoria da Escolha Social e o método de contagem de Borda . Esta etapa visa a elaboração de um filtro baseado em conteúdo, onde os itens contidos nos departamentos serão ordenados de acordo com a preferência dos usuários.

O produto final entregue nesta tese é um algoritmo capaz de emular o comportamento de

¹Devido as características de não convergência, as simulações não apresentam uma tendência definida, ou seja, não convergem para um resultado específico. O sistema continua sendo atualizado com os interesses e com isso permanece ativo e dinâmico.

usuários de uma plataforma de comércio eletrônico, assim como todo processo de tomada de decisão. Este algoritmo poderá ser utilizado sempre que dados reais sejam inviáveis de obter. Em seguida, uma abordagem única para o processo de recomendação de conteúdo em tempo real combinando duas metodologias: classificação bayeasiana no filtro colaborativo e a teoria da escolha social como filtro baseado em conteúdo concluindo assim uma contribuição para o estado da arte com uma abordagem híbrida.

A Figura 1 demonstra de forma abstrata todo processo desta tese, que será detalhadamente explicada posteriormente, mas, em resumo, a abstração apresenta uma aplicação baseada em processos persuasivos e com conteúdos de interesse dos usuários que serão continuamente acessados. Com isso, um banco de dados é a base para a implementação de sistema capaz de recomendar conteúdo à quem estiver interessado. Para isso, o protocolo coletando informações do usuário é ativado. Para o processo de recomendação, como as preferências do usuário são desconhecidas pela plataforma, o problema de *cold-start* é caracterizado, e consequentemente, dois filtros são utilizados para refinar a informação mais adequada.

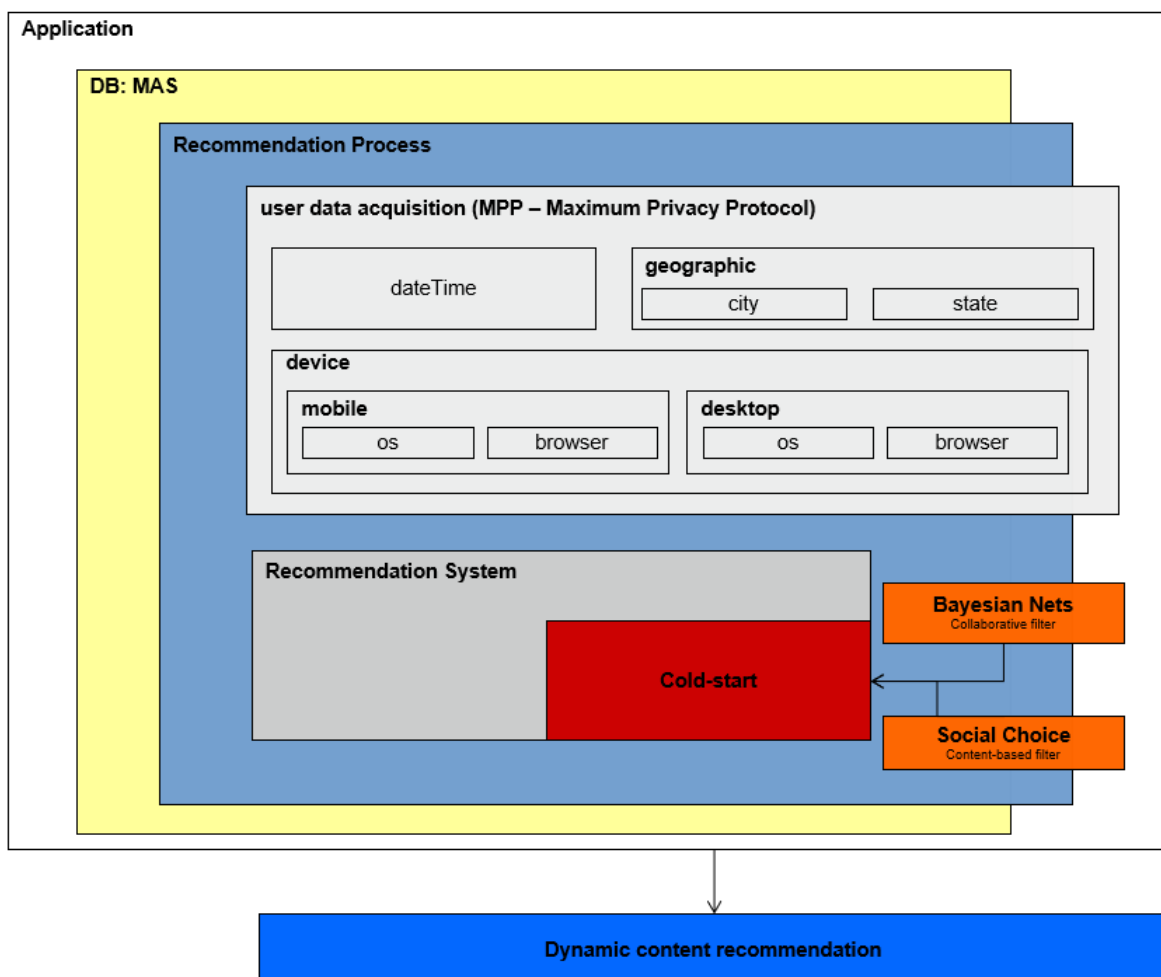


Figura 1: Abstração dos processos para recomendação de conteúdo com problema de *cold-start* (Fonte: Autor)

O conteúdo será então processado e obtido através de Classificadores Bayesianos e pela Teoria da Escolha Social. Uma vez todo procedimento executado, o sistema é capaz de recomendar um conteúdo em tempo real cujo resultado, para o usuário mesmo desconhecido pelo ambiente, apresenta a maior probabilidade de ser de seu interesse naquele momento.

1.1 Relevância e Contribuições

A necessidade em fornecer mecanismos capazes de filtrar e classificar informações disponibilizadas em uma plataforma digital faz com que processos persuasivos sejam implementados a fim de proporcionar uma melhor experiência ao usuário final [80].

Dessa forma, implementações computacionais que estabeleçam relações diretas de recomendação de conteúdo tornam-se uma necessidade devido a vasta quantidade de informação disponível.

Com isso, de acordo com o comportamento do usuário, procedimentos computacionais especificamente desenvolvidos para esse fim asseguram a operacionalidade do processo de apresentação de conteúdos personalizados e, de acordo com as preferências do usuário, promovam uma maior retenção na plataforma.

Contudo, alguns problemas ainda específicos persistem nestas implementações persuasivas. Um problema muito recorrente é a questão de o que recomendar para novos usuários? Esse novo agente operador de um sistema computacional não apresenta muitas informações relevantes para que a plataforma possa analisá-lo. Até o momento, esse novo usuário é um completo desconhecido para a plataforma. Esse problema é denominado na literatura como *Problema de Cold-Start*.

Essa situação particular é um problema aberto a ser explorado e de grande interesse comercial. Dessa forma, diversos estudos vem sendo conduzidos a fim de propor abordagens para contornar esse problema. Diante desta necessidade, tanto acadêmica, preenchendo uma lacuna teórica no conhecimento, quanto comercial, dado o interesse de aplicação, neste trabalho será apresentado uma abordagem inovadora propondo uma solução para esse problema.

Através de um banco de dados, gerado e avaliado em tempo real por processos baseados em gerenciamento de informações com sistemas multiagente e utilizando a combinação de metodologias distintas de base estatística computacional como processos de Classificação Bayesiana aliada a Teoria da Escolha social, esta tese visa apresentar uma solução para o problema assim como assegurando um tempo de execução computacional satisfatório para o desenvolvimento de futuras aplicações comerciais.

1.2 Hipótese científica

Processos distintos vem sendo implementados para abordar de frente esse problema de *Cold-Start* em sistemas de recomendação de conteúdo. Contudo, uma lacuna ainda não

explorada na literatura será abordada neste presente trabalho. O uso de um ambiente baseado em agentes para filtragem automática de conteúdo através de processos em duas etapas: a primeira etapa consiste em um filtro baseado no comportamento do coletivo, de acordo com as características explícitas dos usuários da plataforma. Em um segundo momento, um filtro baseado em conteúdo para refinamento da informação e apresentação em tempo real ao usuário da plataforma.

Entende-se por características explícitas toda informação registrada de maneira espontânea pelo usuário no momento do cadastro e acesso à plataforma. Nenhuma outra informação em segundo plano será coletada sem o consentimento do utilizador.

É esperado que essa abordagem apresente tempo computacional viável para implementação em tempo real, visto que o modelo é utilizado em ambientes dinâmicos. Adotar abordagens que utilizem complexos procedimentos matemáticos e inviabilizem a execução em tempo real da análise do usuário não é de interesse para esse trabalho.

1.3 Objetivos

Para que uma solução seja apresentada de maneira satisfatória, esse trabalho pretende atender os seguintes objetivos:

1.3.1 Objetivo Geral

- Apresentar uma solução para o problema de *cold-start* em sistema de recomendação de conteúdo através de uma abordagem baseada em agentes e utilizando Classificação Bayesianas e Teoria da Escolha Social para processamento e filtragem em tempo real de informação.

1.3.2 Objetivos Específicos

- Desenvolver um algoritmo capaz de gerar um conjunto de dados sintéticos baseados em processos de tomada de decisão em um ambiente de aplicação comercial digital.
- Elaborar um sistema multiagente para explorar o conjunto de dados em tempo real.
- Elaborar um protocolo direcionado à privacidade do usuário, para acesso de dados pessoais em processos baseados em implementações persuasivas.
- Apresentar um filtro colaborativo utilizando Classificadores Bayesianos.
- Desenvolver um filtro baseado em conteúdo utilizando a Teoria da Escolha Social.
- Testar e validar o sistema multiagente gerado em banco de dados multidomínios.

- Implementar uma plataforma modelo de comércio eletrônico para validação do processo e agente desenvolvido.
- Validar em tempo real o processo de recomendação de conteúdo através do sistema multiagente e filtros desenvolvidos.

Para apresentar uma solução robusta para o problema de *cold-start* este trabalho transita entre diversos processos interdisciplinares, conforme demonstra a Figura 2.

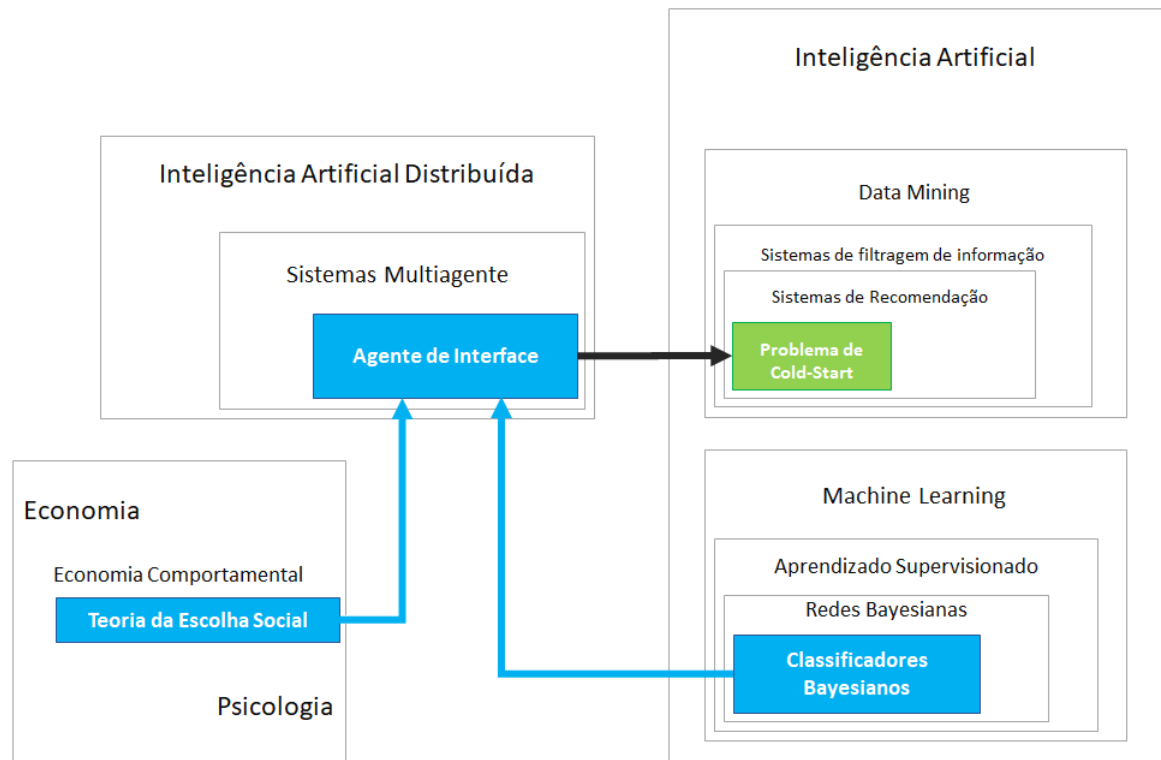


Figura 2: Áreas do conhecimento abordadas para solução do problema estudado. (Fonte: Autor)

Portanto, para apresentar uma solução para o problema de *cold-start* esta tese utilizará abordagens baseadas em Sistemas Multiagente, que é uma sub-área de conhecido da Inteligência Artificial Distribuída. Combinada com Classificadores Bayesianos, uma sub-área de conhecido relacionado à técnicas de aprendizado de máquina visa estabelecer filtros que refinem a informação disponível. Por fim, a intersecção entre a área de Economia e Psicologia da origem a uma sub-área denominada Economia Comportamental. Nessa área, uma teoria latente é a Teoria da Escolha Social. Com essa área de estudo, será proposto o refinamento da informação disponível para então recomendar o melhor conteúdo em tempo real.

Dessa forma, estas três áreas do conhecimento atuarão em conjunto para apresentar uma abordagem para um problema da área de Mineração de Dados atuando em sistemas de filtragem de informação e recomendação.

A estrutura é complexa, porém representa uma lacuna importante na literatura que deve ser avaliada a fim de propor novas teorias e recursos computacionais que viabilizem o estudo de um problema em aberto e de forte apelo comercial.

1.4 Organização do trabalho

A seção de *Revisão Sistemática* apresenta ao leitor os principais resultados de pesquisas desenvolvidos nos últimos cinco anos e coletados através da plataforma portal periódicos CAPES. Os critérios para refinamento dos resultados são apresentados. Além disso, uma tabela com resumos dos principais resultados, métricas e observações gerais dos trabalhos que auxiliaram o desenvolvimento dessa tese.

Em seguida, uma *Revisão Teórica* dos principais temas abordados computacionalmente, e estatisticamente serão apresentados assim como as principais contribuições na área de sistemas de recomendação informando seus métodos e filtros baseados tanto em forma colaborativa quanto de conteúdo. Por fim, o problema de *Cold-start* é formalmente demonstrado.

Posteriormente, toda metodologia de desenvolvimento desse trabalho é detalhadamente apresentada de acordo com os resultados obtidos através do processamento das informações e algoritmos implementados.

2 REVISÃO SISTEMÁTICA

O processo de revisão sistemática para esse trabalho consiste em obter uma base referencial capaz de assegurar que pesquisas relacionadas com esse trabalho possam contribuir com a solidificação do conhecimento adquirido.

A revisão sistemática foi realizada entre outubro/2017 e setembro/2020, com periodicidade de conferência e atualização semanal. Para isso foram realizadas pesquisas adotando uma plataforma institucional² vinculada à mais de vinte bases científicas, considerando periódicos revisados por pares e recursos online disponíveis.

Os critérios utilizados para o refinamento desta pesquisa consiste em avaliar palavras-chave no corpo dos conteúdos disponíveis nas bases científicas, assim como em resumo e título, respeitando critérios de exatidão, similaridade e unicidade. Os seguintes termos foram procurados:

- cold-start (cold start)
- multi-agent (multi agent)
- persuasive
- recommender
- recommendation
- bayesian classifiers
- social choice theory

Os materiais selecionados foram avaliados, inicialmente, de acordo com o conteúdo apresentado em seu resumo. Quando o material apresentasse alguma contribuição de interesse, era então estudado por completo. A Tabela 1 reflete os resultados obtidos durante o período avaliado.

²Portal de periódicos CAPES/MEC disponível em periodicos.capes.gov.br

Tabela 1: Revisão sistemática em bases científicas

					Resultados		
Ind.	Posição	Critério	Adc.	Pesquisa	Geral	Ult.5 anos	Triagem
1.a	qualquer	contém		cold-start recommender	1711	1148	
2.a	título	contém		cold-start recommender	30	21	
3.a	qualquer	contém		cold-start recommendation	2400	1414	
4.a	título	contém		cold-start recommendation	61	39	16
<i>triagem</i>	<i>selecionado ultimos 5 anos e lido título e resumo para avaliar quais materiais podem contribuir com a tese.</i>						
<i>Selecionados</i>	[66], [91], [2], [43], [106], [65], [36], [102], [98], [103], [100], [25], [47], [22], [108], [75]						
5.a	qualquer	contém	AND	cold-start	139	107	
	qualquer	contém		bayesian classifiers			
6.a	qualquer	contém		cold-start bayesian classifiers	142	107	
7.a	qualquer	contém	AND	cold-start recommender	22	12	
	qualquer	exato		bayesian classifier			
8.a	qualquer	contém	AND	cold-start recommendation	22	13	
	qualquer	exato		bayesian classifier			
9.a	qualquer	contém	AND	cold-start recommender	22	12	
	qualquer	exato		bayesian classifiers			
10.a	qualquer	contém	AND	cold-start recommendation	22	13	
	qualquer	exato		bayesian classifiers			
11.a	título	contém		cold-start bayesian	2	1	2
<i>triagem</i>	<i>selecionados todos trabalhos encontrados e lido título e resumo para avaliar quais materiais podem contribuir com a tese.</i>						

Tabela 1 continuação da página anterior

					Resultados		
Ind.	Posição	Critério	Adc.	Pesquisa	Geral	Ult.5 anos	Triagem
<i>Selecionados</i>	[104], [57]						
12.a	qualquer qualquer	contém exato	AND	cold-start recommender social choice	12	10	
13.a	qualquer qualquer	contém exato	AND	cold-start recommendation social choice	12	10	7
<i>triagem</i>	<i>selecionados todos trabalhos encontrados e lido título e resumo para avaliar quais materiais podem contribuir com a tese. pesquisa 13.a apresenta os mesmos resultados da pesquisa 12.a</i>						
<i>Selecionados</i>	[85], [79], [41], [21], [59], [40], [5]						
14.a	título	contém		cold-start social choice	0	0	
15.a	qualquer qualquer	contém exato	AND	cold-start recommender social choice theory	8	6	
<i>obs:</i>	<i>todos trabalhos encontrados na pesquisa 15.a estão contidos na pesquisa 13.a</i>						
16.a	qualquer qualquer	contém exato	AND	cold-start recommendation social choice theory	8	6	
<i>obs:</i>	<i>todos trabalhos encontrados na pesquisa 16.a estão contidos na pesquisa 13.a. pesquisa 16.a retorna os mesmos resultados da pesquisa 15.a.</i>						
17.a	qualquer	contém		cold-start recommendation persuasive	151	58	
17.b	qualquer qualquer	contém exato	AND	cold-start recommendation persuasive bayesian	6	4	4
<i>triagem</i>	<i>selecionados todos trabalhos encontrados e lido título e resumo para avaliar quais materiais podem contribuir com a tese.</i>						

Tabela 1 continuação da página anterior

					Resultados		
Ind.	Posição	Critério	Adc.	Pesquisa	Geral	Ult.5 anos	Triagem
<i>Selecionados:</i>	[96], [51], [72], [56]						
17.c	qualquer	contém	AND	cold-start recommendation persuasive	1	1	1
	qualquer	exato		social choice			
<i>triagem</i>	<i>selecionados todos trabalhos encontrados e lido título e resumo para avaliar quais materiais podem contribuir com a tese.</i>						
<i>Selecionados:</i>	[79]						
18.a	qualquer	contém		cold-start recommendation multiagent	97	72	
18.b	qualquer	contém	AND	cold-start recommendation multiagent	16	12	8
	qualquer	exato		bayesian			
<i>triagem</i>	<i>selecionados todos trabalhos encontrados e lido título e resumo para avaliar quais materiais podem contribuir com a tese.</i>						
<i>Selecionados:</i>	[18], [19], [83], [69], [97], [88], [14], [7]						
18.c	qualquer	contém	AND	cold-start recommendation multiagent	0	0	
	qualquer	exato		social choice theory			
19.a	qualquer	contém		cold-start recommendation	128	77	
	qualquer	exato		multi agent			
19.b	qualquer	contém	AND	cold-start recommendation bayesian	42	29	
	qualquer	exato		multi agent			
19.c	qualquer	contém	AND	cold-start recommendation social choice	66	39	
	qualquer	exato		multi agent			
<i>obs:</i>	<i>os resultados da pesquisa 19.c foram desconsiderados pois o termo social apresentava relações com o termo social networks</i>						
20.a	qualquer	contém		cold-start recommendation multi-agent	130	79	
20.b	qualquer	contém	AND	cold-start recommendation multi-agent	4	3	2

Tabela 1 continuação da página anterior

					Resultados		
Ind.	Posição	Critério	Adc.	Pesquisa	Geral	Ult.5 anos	Triagem
	qualquer	exato		bayesian classifiers			
<i>obs:</i>	<i>a pesquisa 20.b esta contida no resultado encontrado em 19.b visto que o sistema da CAPES desconsidera o hífen em multi agente e multi-agent. Foi considerado um refinamento no termo para bayesian para bayesian networks</i>						
<i>triagem</i>	<i>selecionado ultimos 5 anos e lido título e resumo para avaliar quais materiais podem contribuir com a tese.</i>						
<i>Selecionados:</i>	[14], [39]						
20.c	qualquer	contém	AND	cold-start recommendation multi-agent	3	3	3
	qualquer	exato		social choice theory			
<i>triagem</i>	<i>selecionados todos trabalhos encontrados e lido título e resumo para avaliar quais materiais podem contribuir com a tese.</i>						
<i>Selecionados:</i>	[21], [59], [5]						

A triagem realizada possibilitou um refinamento do material bibliográfico encontrado. Dessa forma, os materiais selecionados que estão relacionados aos objetivos dessa tese foram estudados e após identificação adequada de seus objetivos e principais resultados, fazem parte da bibliografia do presente trabalho conforme é demonstrado abaixo:

Referência: [75]

Objetivo: Fatoração matricial para recomendação de itens para novos usuarios.

Resultados: Experimental results show that it can effectively solve each of the three cold-start problems and outperforms several matrix factorization

Referência: [66]

Objetivo: Fatoração matricial baseada em agrupamentos em filtros colaborativos.

Resultados: Recomendação de itens do banco de dados da Amazon.

Referência: [91]

Objetivo: Filtro colaborativo para recomendação de itens baseados em avaliações de compras através de redes neurais.

Resultados: Explorar uma abordagem híbrida com outros métodos aliados com o filtro colaborativo

Referência: [2]

Objetivo: Fatoração matricial ponderada com tensores de informação.

Resultados: Aprimoramento no tempo computacional na recomendação direta de itens para novos usuários.

Referência: [106]

Objetivo: Filtro colaborativo com processo de tomada de decisão baseado em modelos de Markov.

Resultados: Utilização do banco de dados de Movie Lens 100K e avaliação de desempenho computacional em proposta futura de escalabilidade.

Referência: [64]

Objetivo: Analisar o problema de Cold-start com regras de decisão.

Resultados: Resultados mostraram desempenho computacional ao propor escalabilidade dos dados.

Referência: [36]

Objetivo: Avaliar o problema de cold-start com correlações e informações geográficas.

Resultados: Proposta de análise geográfica e desempenho computacional

Referência: [43]

Objetivo: Avaliação de matriz de fatoração com raqueamento Bayesiano

Resultados: A abordagem Bayesiana promove desempenho computacional e um critério probabilístico para a tomada de decisão.

Referência: [100]

Objetivo: Analisar o problema de cold-start com redes neurais

Resultados: Utilização de redes neurais em filtros colaborativos promove um desempenho computacional melhor ao escalar os dados

Referência: [40]

Objetivo: Filtro colaborativo baseado em análise de componentes principais

Resultados: Ganho de tempo computacional e precisão na recomendação

Referência: [5]

Objetivo: Ranqueamento de informações e recomendação através de algoritmo de PageRank

Resultados: Ganho de tempo computacional e precisão na recomendação

Referência: [21]

Objetivo: Recomendação de itens diretos através de análise de agrupamento e algoritmo K-Means

Resultados: Escalabilidade e possibilidade de novas modelagens no agrupamento de informações

Referência: [69]

Objetivo: Ambiente multiagente baseado em grafos para tomada de decisão

Resultados: Recomendação de itens avaliadas de maneira qualitativa necessitando de uma análise mais quantitativa

Através dessas avaliações, foi identificado uma lacuna na literatura que esta tese visa preencher.

A utilização de um processo híbrido na abordagem do problema de cold-start. A utilização de filtragem em duas etapas com um processo probabilístico, através de classificadores Bayesianos e um refinamento do conteúdo utilizando a Teoria da Escolha Social através do método de contagem de Borda, metodologia esta ainda não explorada na literatura. Assim, este trabalho contribui com o estado da arte fornecendo uma abordagem híbrida para o problema e com o uso de uma metodologias ainda não aplicadas ao problema analisado.

3 REVISÃO TEÓRICA

Esta seção visa instruir brevemente o leitor aos tópicos abordados para realização desse trabalho. Noções fundamentais serão apresentados a fim de que possa ser plenamente compreendida a metodologia utilizada para obtenção dos resultados presentes nessa tese.

3.1 Sistemas de Recomendação

A crescente quantidade de informação apresentada em qualquer meio impõe uma necessidade de filtrar e selecionar conteúdo e conseqüentemente essa tarefa pode tornar-se um tanto trabalhosa. O volume de informação cresce de maneira quase que exponencial e com isso, a utilização de metodologias e processos computacionais que visem garantir uma otimização de tempo, qualidade do conteúdo e a transformação de informações em conhecimento torna-se uma necessidade para a sociedade atual.

A utilização de recomendação de conteúdo ocorre constantemente e passa quase sempre despercebida. Essa prática é facilmente reconhecida em nossas atividades cotidianas onde, por exemplo, um amigo recomenda alguma coisa para outro, e conseqüentemente, essa analogia é potencializada para aplicações comerciais atualmente presentes em toda e qualquer plataforma digital. O sistema tenta então de alguma forma apresentar conteúdo mais adequado aos interesses do usuário final.

Exemplos de serviços do dia a dia que apresentam essa abordagem são plataformas baseadas em interações sociais, tais como diversas redes sociais. Esse tipo de serviço em algum determinado momento está recomendando novas pessoas para que seu círculo e amizade ou rede de contato social aumente. Não obstante, serviços de fornecimento de audiovisuais tais como Netflix, Spotify e o YouTube apresentam de maneira direta recomendação de seus conteúdos, sejam esses filmes, séries musicais ou qualquer outro vídeo que seus algoritmos avaliam ser de interesse do usuário.

Sendo assim, entende-se por *sistema de recomendação* toda e qualquer implementação computacional que vise de alguma forma explorar conjunto de dados, filtrar toda aquela informação e apresentá-la ao usuário [60] de maneira eficiente, associada diretamente aos seus interesses. A principal motivação dessa abordagem é a retenção do usuário na plataforma e otimizar o processo para transformar toda informação disponível em conhecimento, de tal modo que seja útil ao usuário tomador de decisão. A Figura 3 informa uma relação abstrata de como funciona o fluxo de acesso, filtragem, processamento e apresentação do conteúdo personalizado em sistemas de recomendação.

Dessa forma, independentemente do volume de informação que a plataforma disponha, robustos e dinâmicos algoritmos estarão filtrando todo o conjunto de dados a fim de apresentar soluções personalizadas e matematicamente relacionadas às intenções do usuário.

Contudo, certos protocolos e boas práticas devem ser adotadas em implementações que vi-

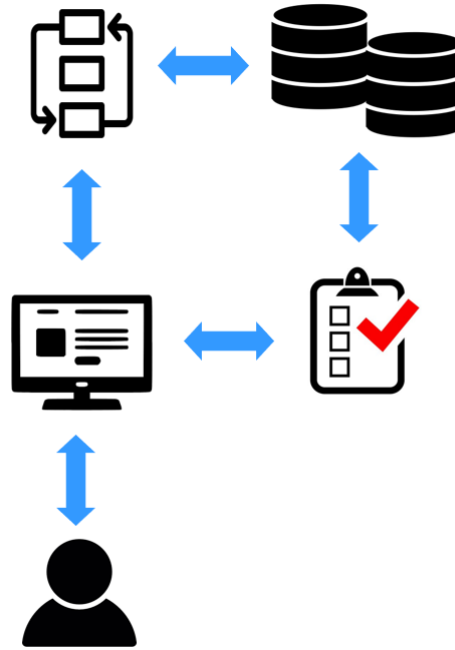


Figura 3: Abstração de um sistema de recomendação, (Fonte: Autor)

sem essa abordagem. De acordo com as características do serviço, sistemas de recomendação tendem a se comportar de maneiras específicas, adotando métricas e metodologias que buscam aperfeiçoar a experiência do usuário. Os diferentes tipos de sistemas de recomendação mais explorados na literatura serão apresentados a seguir.

3.1.1 Tipos de sistemas de recomendação

Diversas maneiras podem ser utilizadas para abordar o problema de recomendação de conteúdo. Entretanto, existem generalizações que podem ser assumidas em tais práticas. Dessa forma, é reconhecido na literatura que existem aproximadamente cinco tipos ou categorias básicas de sistemas de recomendação [15]: colaborativa, baseada em conteúdo, demográfica, utilitária e baseada em conhecimento. É também assumido uma abordagem híbrida, de forma que para satisfazer a melhor recomendação de acordo com a aplicação proposta, combinações desses métodos podem ser realizadas.

Entretanto, para esse trabalho, uma abordagem híbrida e com duas metodologias ainda não exploradas em conjunto para esses processos serão abordadas: Processos de recomendações baseados em filtros colaborativos e filtros de conteúdo.

3.1.2 Filtros colaborativos

Uma abordagem baseada em filtragem colaborativa visa estabelecer critérios de similaridade entre os agentes envolvidos [55]. Dessa forma, esse método caracteriza-se por ser uma abordagem que examina itens avaliados por outros usuários e consequentemente esta-

belece métricas para sua classificação e ponderação quanto à níveis de relevância. A Figura 4 exemplifica o funcionamento desta abordagem.

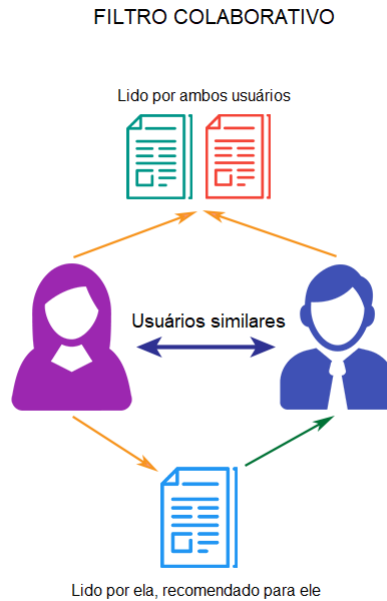


Figura 4: Abstração do processo de filtro colaborativo entre usuários.

Conforme visto na revisão da literatura diversos procedimentos matemáticos são implementados a fim de realizar essa filtragem. Neste trabalho, será utilizado um processo baseado em classificação bayesiana para verificar o comportamento do coletivo e dessa forma refinar as informações coletadas afim de restringir o vasto volume de dados para uma quantidade significativa, com alta relevância e interesse ao usuário.

Entretanto, tais processos ainda podem ser subclassificados em abordagens que seguem critérios de memória ou de modelo. Entende-se por um filtro colaborativo baseado em memória, situações que utilizem critérios matemáticos e estocásticos que de alguma forma possam armazenar informações do usuário conforme seu uso na plataforma.

Esse cenário visa criar documentos que registram as pegadas do usuário e dessa forma um histórico de uso da plataforma seja gerado. Através de algoritmos e critérios matemáticos e estatísticos [13] o sistema de recomendação é capaz de analisar esse histórico e consequentemente tomar decisão no processo de recomendação. Destaca-se na literatura que de acordo com a implementação e demanda do serviço, o vasto volume de informação armazenada tende a tornar essa abordagem uma tarefa extremamente lenta sob o ponto de vista de processamento computacional [37],[46].

Para contornar esse cenário, outra abordagem pode ser implementada: uma filtragem de informação baseada em modelo. Esse critério consiste em realizar técnicas de filtragem através do processamento dinâmico de um modelo matemático. Essa abordagem, quando modelada corretamente, tende a produzir melhores desempenhos computacionais. Contudo,

o problema de escalabilidade de informações é uma constante preocupação para abordagens desse tipo [84].

3.1.3 Filtros baseados em conteúdo

Outra maneira de filtrar o conteúdo de um serviço digital é apresentá-lo de uma forma refinada ao usuário, consiste em estabelecer métricas capazes de organizar a informação em critérios relacionados ao conteúdo. Entende-se com filtro baseado em conteúdo algoritmos desenvolvidos para analisar grandes volumes de dados através de seleção de descrições [94],[77], avaliações [42] e ranqueamento de informação [93] de modo que seja adequada às necessidades e interesse do usuário [47].

A utilização de filtros baseado em conteúdos apresenta diversas vantagens, não apenas sob o ponto de vista de correlação entre ocorrência de itens, mas também apresenta a capacidade de recomendar itens que apresentem características referente ao problema de *cold-start*, ou seja, itens novos e não ainda avaliados por outros usuários [47]. A Figura 5 exemplifica essa abordagem.

FILTRO BASEADO EM CONTEÚDO



Figura 5: Abstração do processo de filtro baseado em conteúdo.

Portanto, o uso de filtros baseados em conteúdo fornece uma abordagem mais que visa relacionar as preferências do usuário com atributos que possam ser acessados de uma maneira facilitada [47]

Contudo, os critérios julgados como fundamentais para realização do processo de recomendação podem conter informações coletadas de maneira implícita ou explícita [76],[50],[44]. Dessa forma, é necessário respeitar a privacidade do usuário e aconselha-se adotar metodologias capazes de assegurar que a privacidade deste agente utilizador seja preservada.

3.1.4 O problema de cold-start

Identificar as necessidades e interesse de um usuário de determinado serviço digital, sabendo suas pesquisas prévias e hábitos de consumo de informação facilita muito o processo de indicação de material. Por outro lado, caso o usuário seja um completo desconhecido para a plataforma, como e o que recomendar?

Essa tarefa torna-se, portanto, nada trivial. Serviços digitais de forma recorrente tentam coletar pelo menos alguma informação referente as intenções do usuário já no momento de cadastro no serviço. Essas informações, coletadas de maneira explícitas, ou seja, o próprio usuário está informando-as, facilita o trabalho dos algoritmos de recomendação de conteúdo.

Por outro lado, caso o usuário não tenha acessado a área de cliente na plataforma ou o cadastro não realiza tal coleta prévia, esse é considerado desconhecido. Assim, como e o que recomendar para esse usuário é o denominado Problema de *Cold-Start* em sistemas de recomendação de conteúdo.

Dessa forma, o problema de *cold-start* que se refere à falta de informações iniciais tanto de itens, ou outro atributo avaliado, para que o sistema seja capaz de avaliar as necessidades do utilizador e apresentar um conteúdo refinado que melhor atenda às suas necessidades de pesquisa [68],[105].

Dependendo do método abordado para atacar esse problema diversas situações de desempenho computacional poderão surgir [94],[54]. Existem algumas abordagens híbridas [53],[24], [83] para contornar essa situação. Algumas abordagens adotam critérios baseado em informação [109],[1], outras em relevância social [68],[86],[27],[30] outras ainda focam em contextos da informação [20]. Contudo o problema está longe de ser resolvido, especialmente recomendando novos itens para novos usuários. Recentemente alguns métodos [10],[99] entre plataformas foram propostos para aliviar o grande volume de dados que necessita ser analisado de acordo com a metodologia implementada.

Assim, este trabalho foca em combinar duas metodologias que transitam entre os domínios estatísticos, computacionais e matemáticos para apresentar uma solução para o problema de *Cold start* em tempo computacional de execução viável. Através da combinação de filtros baseados em classificação Bayesiana e a teoria da Escolha social uma nova métrica será apresentada.

3.1.5 Privacidade em sistemas de recomendação

Privacidade de usuário é um tema sempre delicado. Diversos processos e implementações persuasivas acessam dados do usuário com base em seus termos de uso [34]. Esse consentimento, muitas vezes negligenciado pelo usuário, através da não leitura dos termos de uso do serviço, faz com que as empresas aumentem seu poder de avaliação e consequentemente desenvolvam algoritmos mais sofisticados e com mais eficiência persuasiva [50]. Contudo, esse acesso a informação do utilizador da plataforma ocorre essencialmente de duas formas:

uma implícita e outra explícita.

Entende-se por coleta implícita de dados todo acesso a informação que não seja diretamente mencionada ao utilizador [50]. Dessa forma, "pegadas digitais" do usuário ficam registradas e são dinamicamente catalogadas e analisadas pelas empresas [45],[81]. A Figura 6 mostra um sistema de *HeatMap*, ou seja, um mapa de calor que registra as regiões da plataforma onde o usuário passa mais tempo com o mouse posicionado, ou regiões com maiores cliques no uso diário [67].

Vale ressaltar que, pelo lado do cliente, o usuário não faz ideia de que esse registro está sendo coletado de maneira contínua e em tempo real. Outros sistemas também podem ser implementados em segundo plano, tais como EyeTracking ³ [67], ScrollMap ⁴ [62], Mobile Pressure Analysis ⁵ [73]

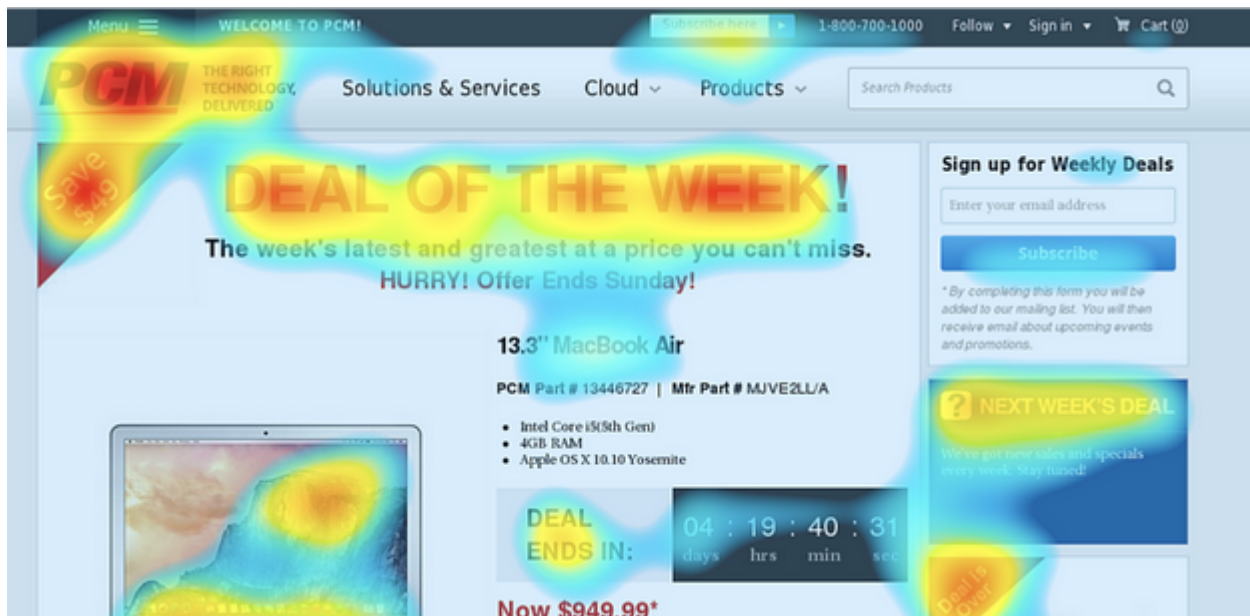


Figura 6: Mapa de calor para análise de usabilidade, (Fonte: [82])

Esse processamento ocorre em segundo plano, sem o consentimento do usuário. Existe um debate contínuo na literatura para estabelecer padrões [50] dos limites de coleta de dados implícitos do usuário. É um debate moral, que está longe de chegar a um consenso, visto que os resultados são visíveis para ambas as partes.

Pelo lado comercial, a plataforma é beneficiada por poder recomendar conteúdo de acordo com um comportamento indireto do usuário. Por outro lado, o próprio cliente também usufrui de um serviço mais personalizado, e automaticamente gerado com base exclusivamente na dinâmica de sua utilização.

³Análise do fluxo visual do usuário na página

⁴Avaliação da rolagem e permanência no setor de conteúdo

⁵Análise pressão aplicada na tela do dispositivo

A segunda abordagem de coleta de informação decorrente do acesso do usuário ocorre de maneira explícita [32]. Através desse procedimento, o usuário está plenamente ciente e de acordo com os dados acessados pela plataforma. Vale destacar que a grande maioria dos usuários não lê os termos de uso do serviço de deseja contratar. Dessa maneira, diversas informações são coletadas, mesmo sem a ciência por parte do usuário, mas resguardado com seu consentimento, visto que a leitura dos termos de uso, e sua aceitação, é um procedimento obrigatório para utilização em todos serviços digitais [22].

Atualmente com a rápida expansão de uso de dispositivos móveis, a instalação de aplicativos em celulares ocorre quase que de maneira automática. Mas, vale ressaltar que antes de qualquer instalação, a aplicação solicita autorização do usuário para coletar diversos dados, tais como posição geográfica, contatos, galeria de imagens, entre outros. O usuário, quase sempre, autoriza a coleta dessas informações de maneira espontânea, sem a devida leitura de seus termos de uso.

Uma vez de posse dessas informações, o serviço pode então estabelecer métricas sofisticadas capazes de analisar o comportamento do usuário e conseqüentemente elaborar mecanismos adequados para a recomendação de conteúdo.

Neste trabalho é apresentado um protocolo para filtragem de informação para o usuário de maneira híbrida, combinando abordagem implícita e explícita mas com apenas dados não invasivos e legalmente disponibilizados pelo usuário no momento de acesso à plataforma digital.

3.2 Classificadores Bayesianos

Classificação consiste em relacionar classes à objetos. Essencialmente existem duas formas de estabelecer critérios para esse fim: supervisionada e não supervisionada [11].

Processos de classificação supervisionados assumem que as classes são conhecidas à priori, e dessa maneira, o problema central consiste em estabelecer uma função que relacione classes às regras, ou classificadores, implementadas para classificação.

Por outro lado, processos de classificação não supervisionados assumem que as classes não são conhecidas de antemão. Dessa forma, o problema inicial consiste em identificar as classes presentes no conjunto de dados para então, em um segundo momento, iniciar o processo de classificação [11],[29].

Com isso, um classificador é uma função $f : \Omega_X \rightarrow \Omega_C$ que associa uma determinada classe $c \in \Omega_C = \{c_1, \dots, c_m\}$ a objetos descritos por um conjunto de atributos $X = \{X_1, X_2, \dots, X_n \in \Omega_X\}$. Com isso, os atributos estabelecidos são considerados variáveis aleatórias, e assim, as ocorrências ou não de tais características fica relacionada a ocasionalidade amostral [29], [49].

A eficiência computacional em utilizar processos de classificação está vinculada à avaliação probabilística que deve ser realizada. Com isso, ao invés de mapear todo conjunto, ob-

servando todo espaço amostral e suas características de interesse da forma $X \rightarrow C$, é possível mapear a ocorrência de determinado evento e probabilisticamente classificá-lo quanto ao objeto como $X \rightarrow P(C_j|X)$. Sendo assim, a hipótese de que determinada observação pertence a um conjunto objeto é definida por:

$$c^* = h_c(x) = \arg \max_{j=1..m} P(c_j|x) \quad (3.1)$$

Portanto, classificadores Bayesianos são funções que relacionam objetos com determinadas características às classes [107]. O termo Bayesiano surge pois a classe analisada c^* associada à eventos aleatórios X é determinada pela equação 3.2 através do Teorema de Bayes.

$$P(C|X) = \frac{P(C)P(X|C)}{P(X)} \quad (3.2)$$

Dessa maneira, a classificação do objeto analisado referente à classe de interesse é associado de acordo com a maior distribuição posterior obtida. Logo, as equações 3.1 e 3.2 podem ser reescritas da forma:

$$c^* = h_{Bayes}(x) = \arg \max_{j=1..m} P(c_j)P(x|c_j) \quad (3.3)$$

A Eq. 3.3 é definida como um Classificador Bayesiano, ou seja, uma função matemática capaz relacionar variáveis aleatórias às determinadas classes ou conjunto de informações através de avaliações baseadas em critérios probabilísticos de acordo com o teorema de Bayes [101], [107].

Consequentemente, o resultado obtido pelo método avaliado consiste em um valor probabilístico. Este característico numérico pode ser associado com classes definidas à priori assim como estabelecer um critério hierárquico de acordo com o valor da probabilidade calculada[4].

3.3 Teoria da Escolha Social

A Teoria da Escolha Social é o estudo de processos e decisões individuais que resultem em um padrão de escolha coletiva [12], ou seja, é o processo de como agregar as preferências individuais em uma única escolha coletiva de maneira racional. Para isso, métricas capazes de agregar parâmetros individuais são adotadas de tal modo que um padrão comportamental coletivo seja obtido.

Os pilares teóricos da teoria estão nas abordagens iniciais por Nicolas de Condorcet e Kenneth Arrow. O resultado geral de Condorcet, conhecido também como o paradoxo de Condorcet é a observação de que preferências de uma maioria podem ser irracionais mesmo que preferências individuais sejam racionais.

Supondo que um terço de um grupo faça uma escolha racional e ordenada de elementos (x ,

y, z). Um segundo grupo escolha (y, z, x) e um terceiro grupo escolha (z, x, y) . Racionalmente é notável que x tem prioridade de escolha sobre y (veja os pares x, y). Entretanto, nenhum par é considerado um "vencedor" visto que não existe uma alternativa que represente a maioria quando comparado em pares.

Esse paradoxo se mantém na sociedade atual, e é considerado um dos pilares da Teoria da Escolha Social: uma regra da maioria é um método plausível de tomada de decisão coletiva e, mesmo assim, extremamente subjetivo [71].

Por outro lado, Arrow considerou um método baseado em classes de possíveis agregações denominada funções de bem-estar social (welfare functions) no qual devem satisfazer determinadas regras. O resultado dessa abordagem é conhecido como o Teorema da impossibilidade de Arrow [61] informando que a agregação de ordens de preferência individuais não produz uma preferência coletiva.

As regras estabelecidas por Arrow para a generalização desse problema partem de três premissas: as preferências são ranqueadas, existem três ou mais alternativas e as escolhas individuais são transitivas.

Para isso, algumas propriedades foram definidas:

Unanimidade: Se todos os participantes preferem uma mesma determinada alternativa, então essa é a preferência geral do coletivo.

Escolha não ditatorial: O resultado não deve refletir sempre a preferência de um único indivíduo.

Independência das alternativas irrelevantes: Dado um ranqueamento formado por $a > b > c$, se $a > b$ então essa mesma preferência não pode ser alternada com a possibilidade de c .

A importância em estabelecer critérios exatos para um sistema de escolhas é fundamental em diversas áreas da sociedade. De acordo com Arrow: *"A maior parte dos sistemas não funcionará mal o tempo todo. Tudo que fiz foi provar que eles podem funcionar mal as vezes"*.

Diante disso, um dos métodos matemáticos utilizados na Teoria da Escolha Social é a contagem de Borda. Essa abordagem apresenta uma implementação computacional com grande desempenho computacional dado sua abordagem matemática direta e será implementada para ranqueamento das preferências dos usuários no banco de dados obtido através da simulação baseada em agentes.

3.3.1 Método de contagem de Borda

Proposto por Jean-Charles Borda em 1770, o método consiste em obter um ranqueamento para as escolhas disponíveis em um conjunto de possibilidade apresentados ao usuário⁶.

Dessa forma, os itens são ordenados de acordo com as preferências de cada usuário.

Na contagem, cada posição na ordenação é atribuída uma pontuação de acordo com:

$$\forall x_i \in X | x_n = 1, x_{n-1} = 2, x_{n-2} = 3 \dots \quad (3.4)$$

A distância entre cada preferência será de um ponto absoluto. Ao final, os pontos são somados para decidir qual a opção que vencedora. Esta abordagem considera, além da primeira escolha, todas as outras opções disponíveis ao usuário. Com isso, o método garante que nem sempre o item mais vezes utilizados como primeiro lugar será o vencedor.

A Tabela 2 ilustra o sistema de contagem Borda para uma aplicação de seleção de itens em um departamento arbitrários de um sistema de comércio eletrônico.

Tabela 2: Contagem de Borda para n=3 itens

Rank	Item	Pontos
1	A	3
2	B	2
3	C	1

A implementação desse método foi realizada para fins de ranqueamento dentro de um departamento com maior probabilidade de compra, maximizando assim o poder persuasivo de um sistema de recomendação. Os elementos selecionados correspondem aos itens que tiveram seu processo de aquisição finalizado por parte do usuário.

Dessa forma, a utilização de método proposto através da Teoria da Escolha Social favorece a obtenção de uma preferência coletiva com base em critérios individuais. Assim, a escolha latente reflete os interesses ponderados do coletivo e assim assegurando uma representatividade e mais adequada apresentação de recomendações de itens.

3.4 Simulação e Sistemas multiagente

Um sistema multiagente consiste em múltiplas entidades denominadas *agentes* que interagem em um ambiente compartilhado com o interesse em satisfazer algum objetivo individual ou coletivo. Por outro lado, uma simulação consiste na avaliação de um modelo operacional ao longo do tempo [95]. A utilização de métodos para simulação baseada em agentes tem

⁶O algoritmo original refere-se à aplicações em eleições e sistemas de votação, e trata as respectivas escolhas oriundas por decisões de eleitores. Para manter a associação com o presente trabalho, será substituído o termo eleitores por usuários.

como objetivo avaliar o comportamento de um conjunto de operações em função do tempo, caracterizadas por um domínio de aplicação dinâmico.

O processo de modelagem via agentes foi adotado para emular um ambiente capaz de satisfazer as condições para implementação de interações sociais com intencionalidade comercial, de acordo com o uso e decisões proferidas no processo de compra e venda em plataformas de comercio digital, e dessa maneira criando laços e interações sociais simuladas [89].

Consequentemente, outra abordagem adotada com estas características consiste em estabelecer uma relação de assistência ao usuários[23]. Esta prática consiste no envolvimento de entidades denominadas agentes de interface[16]. Agentes desse tipo são recursos computacionais baseados em técnicas de inteligência artificial que atuam a fim de fornecer um auxílio ao usuário, lidando dessa maneira como um facilitador no processo de interação entre o usuário e a plataforma[9]. Analogamente, pode ser considerado como um assistente pessoal que colabora com o usuário interagindo, recomendando e direcionando-o durante sua experiência de utilização do sistema. A Figura 7 demonstra este relacionamento entre os envolvidos no processo de aquisição de informações.

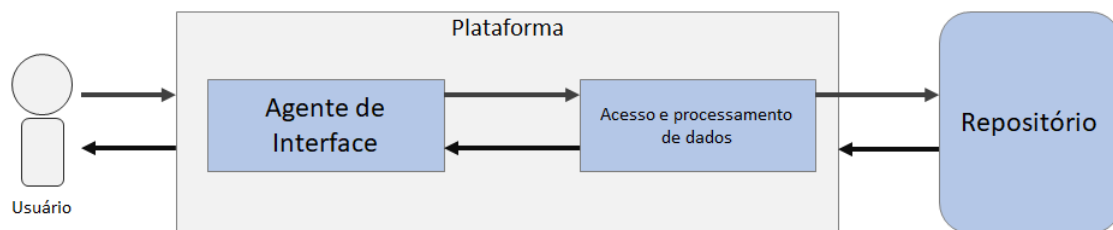


Figura 7: Interação Usuário/Plataforma com atuação de agente de interface. (Fonte: [9] Adaptado)

Com esta abordagem, é possível implementar ferramentas que atuam em uma camada entre o usuário e o banco de dados, processando a informação, filtrando e enviando ao usuário conteúdos previamente avaliados como sendo de maior probabilidade de interesse à quem estiver acessando o sistema naquele momento.

4 MÉTODOS

4.1 Processo de aquisição de dados

Para o planejamento e desenvolvimento de um sistema para recomendação de conteúdo é fundamental que uma base de dados seja disponibilizada e adotada como referencial.

O domínio das variáveis disponíveis no conjunto de dados, características na coleta, posterior análise e processamento de dados na tomada de decisão é fundamental para selecionar, classificar e apresentar o conteúdo personalizado ao usuário.

Como o problema de *cold-start* em sistemas de recomendação é uma característica independente de domínio, ou seja, não importa o setor ou segmento fornecido pelo serviço, o problema ainda estará presente. Com isso, acredita-se que uma solução geral possa atender necessidades distintas em bancos de dados com configurações e modelagens diferentes.

Neste trabalho alguns banco de dados disponibilizados foram inicialmente avaliados. Com isso, o banco de dados modelo disponibilizado pela empresa Amazon⁷ foi analisado. Esse conjunto de dados representa aproximadamente 100 milhões de registros de compra realizados em suas plataformas. A estrutura do banco de dados é apresentada abaixo:

```
{
  "asin": "0000031852",
  "title": "Girls Ballet Tutu Zebra Hot Pink",
  "price": 3.17,
  "imUrl": "http://ecx.images-amazon.com/images/I/51fAmVkTbyL._SY300_.jpg",
  "related":
  {
    "also_bought": ["B00JHONN1S", "B002BZX8Z6", ...],
    "also_viewed": ["B002BZX8Z6", "B00JHONN1S", ...],
    "bought_together": ["B002BZX8Z6"]
  },
  "salesRank": {"Toys & Games": 211836},
  "brand": "Coxlures",
  "categories": ["Sports & Outdoors", "Other Sports", "Dance"]
}
```

onde

asin - Número identificação do produto

⁷Disponível pela University of California San Diego para pesquisas acadêmicas em jmcauley.ucsd.edu/ata/amazon/index.2014.html

title - nome do produto
 price - preço
 imageUrl - url da imagem do produto
 related - produtos relacionados
 salesRank - organização referente categoria e vendas
 brand - marca
 categories - outras categorias que o produto esta presente

Com isso, neste momento o setor de comércio eletrônico, baseado em consumo de itens no varejo estaria muito bem assegurado com o bando de dados disponibilizado para testes. Visto que a Amazon representa em média 42% de marketshare no segmento e-commerce nos EUA e 10% no total em relevante participação global⁸.

Outro segmento latente de uso diário da sociedade são plataformas de streaming de vídeo. Entende-se por streaming de vídeo todo serviço que apresente seu conteúdo de maneira áudio visual em formatos de vídeos curtos, didáticos, filmes e séries através de transmissões pela internet[48]. Empresas representativas desse setor são Netflix e Youtube. Contudo, o acesso à informações dessas empresas é uma tarefa praticamente inviável.

Para contornar esse obstáculo, foi adotado o conjunto de dados do plataforma MovieLens⁹. Em essência, embora essa plataforma não transmita ou hospede conteúdos dessa natureza, ela caracteriza-se por ser um repositório de grande poder informativo, visto que seu catalogo contém um acervo expressivo de opiniões (*Reviews*) de usuários.

Em tese, essa informação é a base de análise para os algoritmos semelhantes aos contidos no Youtube e Netflix. Uma grande quantidade de conteúdo, ranqueado e avaliado por usuários distintos. Essa avaliação é ponderada através de processos matemáticos, estatísticos e computacionais para então entrar no mérito de recomendação de conteúdo personalizado ao usuário.

Dessa forma, devido a similaridade conceito da aplicação em destaque, foi adotado esse conjunto de dados para avaliação. A estrutura das informações coletadas são baseadas em avaliações de usuários em itens.

Assim, apresenta uma relação direta entre $Usuário \rightarrow Item \rightarrow Avaliação$.

Com aproximadamente 20 milhões de registros de mais de 135.000 usuários em 27.000 itens, estas informações resguardam este domínio de aplicação com expressiva relevância e abordando conteúdos baseados em características áudio visuais e com apelo ao entretenimento.

Por fim, um outro segmento muito utilizado é o de músicas. Destaca-se como pioneira

⁸De acordo com relatórios analíticos: Statista (<https://www.statista.com/statistics/788109/amazon-retail-market-share-usa/>) e Bloomberg (<https://www.bloomberg.com/news/articles/2019-06-13/emarketer-cuts-estimate-of-amazon-s-u-s-online-market-share>)

⁹Disponível em grouplens.org/datasets/movielens/

nesse setor a aplicação Spotify. Novamente, por questões operacionais, o acesso a informações dessa plataforma não é imediata. Para contornar esse cenário, foi adotado um conjunto de dados modelo disponibilizado com informações da empresa Last.fm¹⁰. Essa empresa apresenta um conjunto de dados público e é uma referência como plataforma de rádio online. Ambas são similares em propósito visto que o usuário busca apenas um tipo de conteúdo: músicas.

Os recursos dessa plataforma são baseados em estações de rádio on-line. Contudo, como cada "rádio" pode ser criada na forma de lista de reprodução, é comum encontrar rádios com temas específicos. Neste momento os recursos do Spotify se sobrepõem ao da Last.fm, visto que o vasto acervo do Spotify é apresentado ao consumidor no formato de *playlist*.

Assim, o conjunto de dados modelo disponibilizado pelo serviço Last.fm foi utilizado para avaliação. A estrutura dos dados é apresentada na forma:

```
{
  user-id:
  timestamp:
  artist-mbid:
  artist-name:
  song-mbid:
  song-title:
}
```

onde,

```
user-id: Número de identificação do usuário,
timestamp: Data/hora de acesso,
artist-mbid: Identificação do artista no banco de dados
artist-name: Nome do artista
song-mbid: Identificação da música no banco de dados
song-title: Nome da música
```

Devido as características de cada conjunto de dados, certas informações de usuário não são disponibilizadas por questões legais completamente aceitáveis. No entanto, dados de acesso como data e hora, foram omitidas de alguns conjuntos de dados mas outras informações como sistema de origem, plataforma e localização geográfica não estão disponíveis para

¹⁰<http://ocelma.net/MusicRecommendationDataset/> Esse conjunto de dados foi coletado para fins de pesquisa e recomendação de músicas pelo pesquisador independente Oscar Celma (ocelma.net)

consulta tornando assim os banco de dados analisados inviáveis para atuar como base no desenvolvimento dos filtros propostos.

Para isso, um processo gerador de dados baseado em agentes foi desenvolvido. O algoritmo proposto visa criar um banco de dados com as características ausentes nos bancos já analisados.

Com isso, será definido alguns elementos básicos que serão utilizados na filtragem de conteúdo e incorporados à um protocolo de acesso, porém respeitando a privacidade do usuário. A modelagem para o problema de *cold-start* está sustentada justamente nesse pilar. Dados geográficos e tempo de acesso do usuário na plataforma. O principal motivo em escolher essas variáveis está no fato de que, com essa modelagem, é o único território em comum que todo e qualquer banco de dados pode facilmente coletar e analisar em tempo real e não viola qualquer prerrogativa de invasão de privacidade.

Novamente, por questões legais esses valores não foram publicamente divulgados nos banco de dados acima analisados. Por isso a necessidade em simular estas informações. Vale destacar que, embora sejam coletadas de maneira implícita, nenhuma violação à privacidade do usuário é cometida visto que tais informações são facilmente acessadas pelo próprio navegador, assim como dados geográficos são obtidos devido o registro do protocolo TCP/IP do cliente[63].

Para isso, será necessário estabelecer métricas que envolvam essas variáveis de tal modo que não seja caracterizada como abusiva. Dessa forma, um protocolo de dados do utilizador focado em sua máxima privacidade será apresentado.

Com essa particularidade, foi necessário estabelecer critérios objetivos para avaliar determinadas características do usuário. Isso resultou no desenvolvimento de um protocolo específico que será apresentado a seguir.

4.2 Protocolo de privacidade do usuário

Existem certos dados em comum que podem ser instantaneamente coletados quando o usuário de um serviço online acessa determinada plataforma. No entanto, o limite de coleta de informação pode definir noções de responsabilidade caracterizando práticas abusivas e invasivas à privacidade do cliente.

Contudo, certas informações diretas serão adotadas como base para a implementação da solução proposta nesse trabalho. A Figura 8 destaca uma abstração modelada especificamente para essa abordagem, de tal modo que possa ser facilmente adaptada para qualquer aplicação persuasiva baseada em recomendação de conteúdo.

Tais informações são disponibilizadas em tempo real pelo usuário através da plataforma de acesso. Nenhum critério invasivo ou que ocorra violação da privacidade do usuário é adotado.

O primeiro registro importante é data e hora de acesso. Neste critério informações como

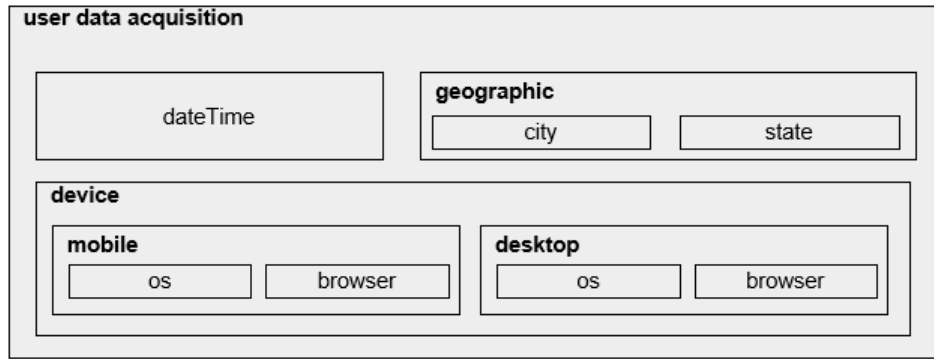


Figura 8: Elementos acessados no protocolo de identificação do usuário.

dia da semana, turno e horários de acesso à plataforma são registrados pelo sistemas.

Outro ponto coletado consiste nas informações geográficas. Cidade e consequentemente estado do usuário já garantem um refinamento de toda informação disponível.

Por fim, o terceiro ponto é o tipo de dispositivo que está sendo utilizado. Duas plataformas são avaliadas: dispositivos móveis e desktop¹¹. Com essa informação é possível ainda acessar qual tipo de sistema operacional e qual navegador esta sendo utilizado.

Com estes elementos é possível fazer um refinamento criterioso de dados de compra em uma plataforma baseada em comércio eletrônico. Contudo, vale destacar que estas informações são apresentadas em tempo real, sem que a plataforma saiba se é um usuário recorrente ou não, que, neste caso, não caracterizaria mais um problema de *cold-start*.

4.2.1 Processo gerador de dados baseado em agentes

Os conjuntos de dados coletados e disponibilizados na internet para fins acadêmicos e pesquisa em desenvolvimento de sistemas de recomendação não apresentam uma característica fundamental para a abordagem proposta desta tese: informações do usuário. Para contornar essa exigência, foi necessário desenvolver um sistema autônomo, que seja capaz de simular interações sociais em processos baseados em comércio eletrônico.

Para que esta abordagem apresente relevância, é necessário que esta simulação seja capaz de identificar as seguintes características:

- Explorar diversos departamentos e itens;
- Detectar influência da opinião de outras pessoas;
- Identificar a intenção de compra do cliente (hedônica/racional);
- Verificar o potencial persuasivo (potencializar a compra);

¹¹Demais dispositivos como consoles de vídeo game e televisores não foram considerados nesse protocolo visto que representam pouco acesso à plataformas de comercio eletrônico. GSStatcounter, 2019. Disponível em <https://gs.statcounter.com/vendor-market-share/mobile/worldwide>

- Registrar avaliações dos usuários (review).

Não apenas emular a compra e venda, mas também criar mecanismos capazes de analisar a opinião de outros consumidores, sua dinâmica e interação com demais usuários e, através de um julgamento e percepção interna, tomar a decisão de adquirir ou não determinado item do catálogo modelo de uma aplicação digital.

Para isto, foi modelado através de um ambiente multiagente, os cenários possíveis para que entidades autônomas possam interagir e escolher os itens que desejam consumir. Cada passo é registrado no conjunto de dados a fim de recriar um modelo capaz de reproduzir um banco de informações sintéticas e que possam ser comparadas às disponibilizadas para fins de pesquisa em diversos repositórios com esta finalidade.

A Figura 9 demonstra todo processo gerador através de uma abordagem baseada em agentes.

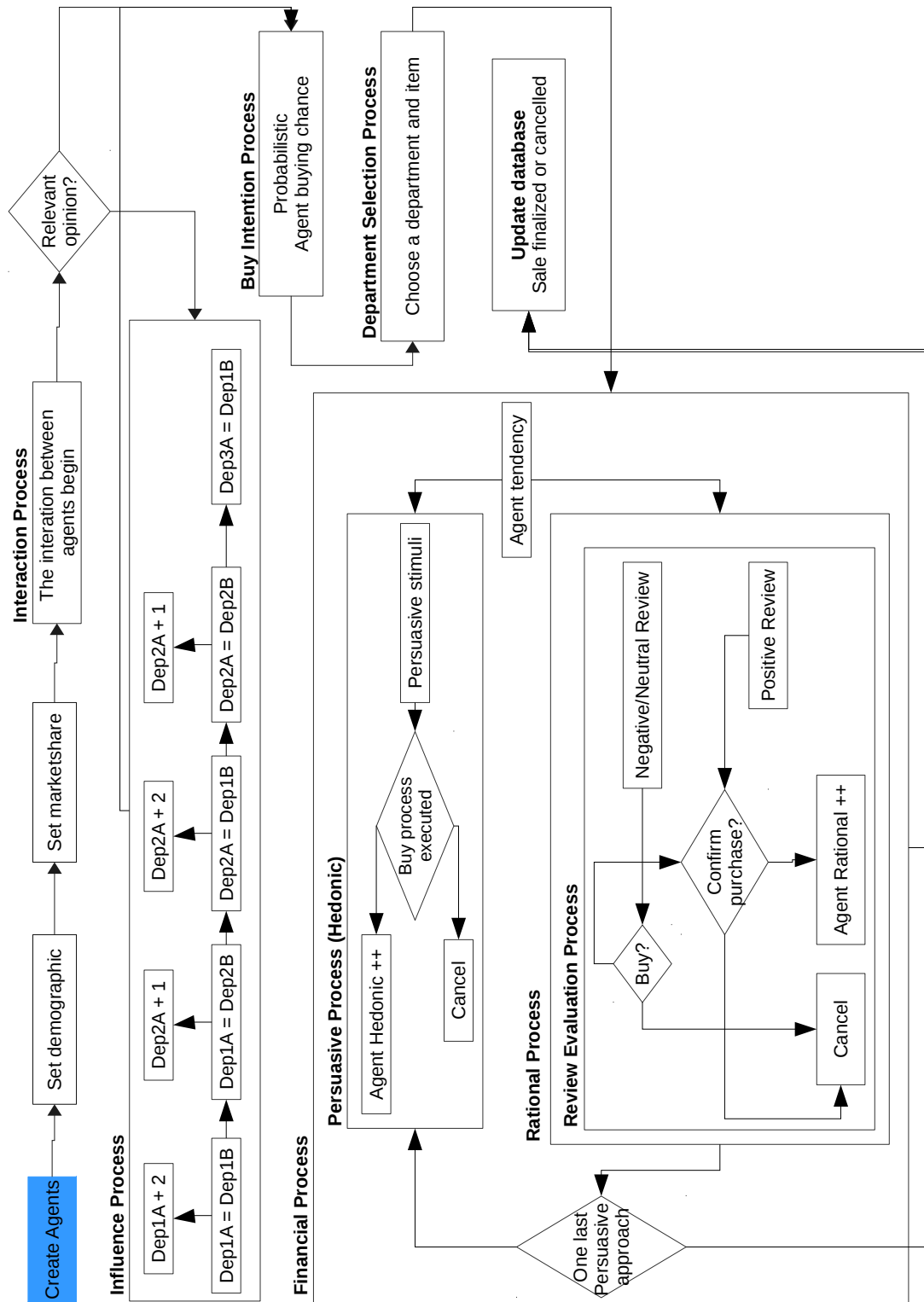


Figura 9: Fluxograma do processo gerador de dados sintéticos baseado na interação entre agentes.

O fluxograma apresentado é executado com base em processos distintos que garantem a dinâmica da simulação vinculada às características sociais e psicológicas das entidades autônomas envolvidas.

Para obtenção de um conjunto de dados que represente o comportamento de usuários em sistemas de compra e venda em ambiente digital, foi modelado uma estrutura baseada em processos (ações) que precisam ser satisfeitos sequencialmente. Os agentes iniciam a interação conjunta para obter as ações e médias hedônicas e racionais de cada entidade participante no ambiente implementado. Os valores iniciais foram probabilisticamente atribuídos de acordo com uma distribuição uniforme de probabilidade. Essa abordagem visa não produzir uma tendência nos dados para alguma centralidade assim como arregurar que não é imposta nenhuma variabilidade à priori.

Os critérios hedônicos e racionais são necessários pois, implementações persuasivas exploram os pontos emocionais do usuário. Por outro lado, para assegurar o equilíbrio no processo de tomada de decisão, o agente também pode ser racional e escolher determinado item de acordo com seus valores coletados em avaliações prévias.

Os processos de influência e interação são fundamentais e análogos ao envolvimento causado por uma rede de contatos, visto que a opinião de outros podem contribuir com a decisão do próprio agente.

No *Processo de influência* o contato obtido por agentes próximos, denominados vizinhos, podem impactar no processo de tomada de decisão. Essa modelagem visa estabelecer a mesma relação social causada por interação pessoal entre pessoas próximas. Exemplo, a opinião de um amigo, ou encontrar alguém que comente ou recomende algo sobre determinado item pode aumentar ou diminuir o interesse em consumir determinado bem.

O *Processo de Intenção* de compra é inicializado e verificado probabilisticamente se a ação deve ser executada ou não. Cada agente possui uma probabilidade intrínseca em realizar a ação ou não. Caso seja realizada, a modelagem continua seu fluxo de execução, caso contrário, o agente seguirá sua caminhada e interações com as demais entidades do ambiente.

No *Processo de escolha* o agente seleciona, autonomamente, qual setor, subsetor e item de compra tem intenção de adquirir. Cada agente apresenta setores de preferência, assim como qualquer pessoa tem. Esta modelagem assemelha-se às preferências individuais e não necessariamente influências causadas pela opinião de algum coletivo próximo.

Uma vez satisfeitos todos requisitos até o momento, o agente é capaz de iniciar o *Processo Financeiro*. Nesta etapa alguns elementos serão considerados para a efetivação do procedimento de compra. Inicialmente é verificado se o agente apresenta tendência à estímulos persuasivos ou tende a atuar de uma maneira mais racional.

Esta etapa é crucial, pois simula o cenário de pessoas consumistas, impulsivas e que adquirem determinados itens basicamente por estímulos persuasivos. Por outro lado, outras focam na utilidade[74] do bem caracterizando-se por ser usuários mais racionais, no qual estímulos diversos podem não apresentar resultados expressivos. Este tipo de usuário tende

a tomar decisão com base em outras avaliações (reviews).

Sendo um agente com grande disposição à estímulos emocionais, o *Processo Persuasivo* é ativado. Neste um estímulo é enviado e verificado se o usuário registra a compra ou não. Por outro lado, o *Processo racional* apresenta as avaliações prévias realizadas por outros usuários no item selecionado. Esta etapa visa assegurar que tanto usuários com apelo ao emocional quando os mais racionais serão avaliados no processo. Um estímulo final é empregado, visto que embora o agente tenha predisposição à tomada de decisão baseada em critérios técnicos, um estímulo final pode ser o gatilho para ele concretizar a aquisição.

Por fim, o usuário pode deixar ou não uma avaliação (*review*) do item pesquisado, efetivando a compra ou não, para que essa métrica sirva como critério no processo de tomada de decisão para eventos e interações futuras. Este recurso ficará disponível para todos usuários, inclusive para quem registrou a avaliação do item. A Figura 10 demonstra de forma abstrata todos os processos do algoritmo proposto.

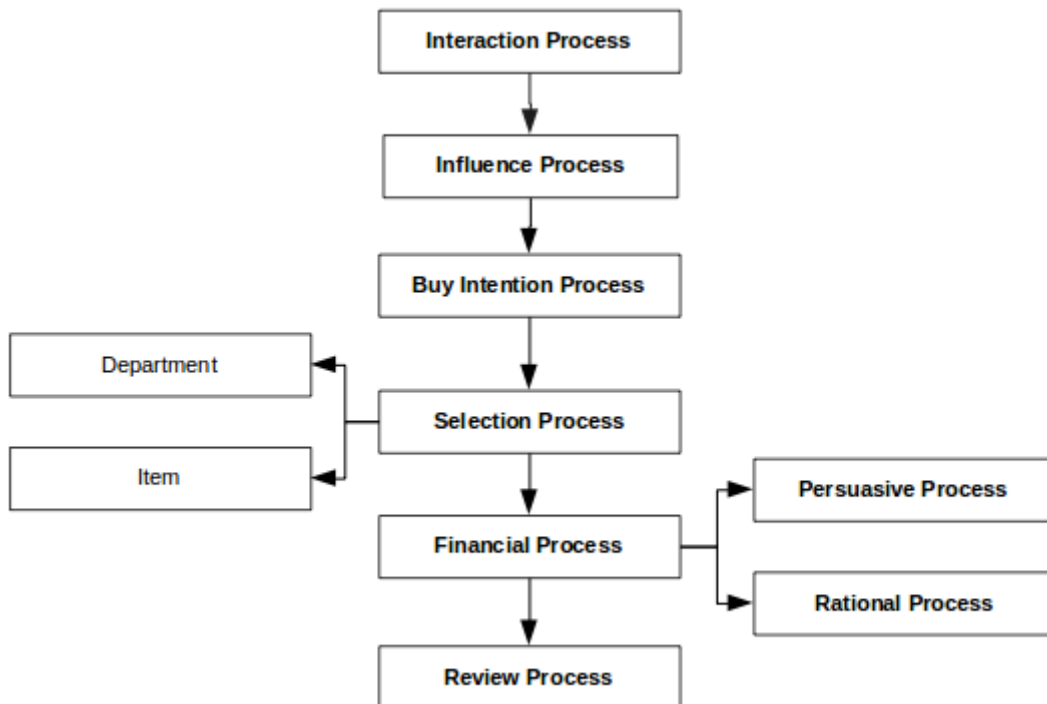


Figura 10: Estrutura dos processos no algoritmo de simulação baseada em agentes.

Essa estrutura é a base para execução do modelo persuasivo baseado em agentes para criar um banco de dados sintético capaz de apresentar características contidas em aplicações de comércio eletrônico. Todo dado produzido por estes processos serão utilizados na sequência para o desenvolvimento de uma metodologia que visa apresentar uma solução para o problema de *cold-start* em sistemas de recomendação.

4.2.2 Domínios e características avaliadas

Por questões de método, a utilização de um cenário baseado em comércio eletrônico engloba as principais características comuns em qualquer domínio de aplicação:

- existem diversos usuário novos, caracterizando assim um problema de *cold-start*;
- os usuários escolhem livremente itens/setores;
- informações sobre os usuários: data de acesso, itens consumidos, localização, etc;
- interações com outros usuários influenciam o processo de tomada de decisão;
- existe informação de *review* (opinião de outros usuários);
- registro de compra finalizada, análogo à vídeo/música reproduzida;
- registro de cancelamento/desistência na aquisição do item;
- agentes possuem preferências por certos setores;
- suas opiniões podem ser alteradas ao longo do tempo;
- a opinião do coletivo pode influenciar a decisão do agente.

Assim, tais pontos refletem características em comum presente em diversos domínios de aplicação comercial. Dessa forma, o sistema modelado pode ser utilizado em cenários que necessitam um conjunto de dados sintético, para fins de avaliação do comportamento e dinâmica de seus usuários e que satisfaçam os tópicos acima mencionados.

4.3 Classificadores Bayesianos como filtro colaborativo em problemas de *cold-start*

A utilização de classificadores Bayesianos visa estabelecer um critério probabilístico no processo de tomada de decisão por parte do sistema em ponderar suas recomendações.

Será utilizada essencialmente para responder a seguinte pergunta: Dentro do sistema de comércio eletrônico do modelo desenvolvido, qual o departamento que apresenta maior probabilidade de compra dadas as características do usuário que está acessando?

Desse questionamento emerge imediatamente uma questão probabilística condicional, ou seja, a probabilidade de um evento ocorrer dado a presença de determinados critérios.

O processo de classificação Bayesiano é um caso particular de Redes Bayesianas [52]. A modelagem do problema desta tese foi refinada de tal forma que, de acordo com o protocolo definido e as características intrínsecas de sistemas de comércio eletrônico, a avaliação probabilística modelada não envolvem necessariamente expressivos relacionamentos entre banco

de dados a ponto que seja relevante a criação de vínculos de dependência. Dessa forma, tal caso particular é o candidato perfeito para o problema proposto.

Primeiro, porque é computacionalmente rápido de ser executado. Segundo, e principal motivo, é que devido ao protocolo de privacidade desenvolvido e apresentado anteriormente, essa abordagem satisfaz todas exigências impostas no processo de modelagem.

Para obter o departamento com maior probabilidade de compra é necessário solucionar a seguinte equação:

$$P(Co|X) = \frac{P(X|Co) \times P(Co)}{P(X)} \quad (4.1)$$

Onde $P(Co|X)$ é a distribuição posterior, ou seja, a resposta com a probabilidade de cada departamento no conjunto de dados do comércio eletrônico modelo. Além de $P(X|Co)$ ser a verossimilhança, ou seja, a probabilidade observada pelo critério, o cálculo relaciona também toda informação à priori disponível no conjunto de dados referente à classe, sem influência dos critérios dado por $P(Co)$ e por fim essa expressão é ponderada pela probabilidade de evidência dado por $P(X)$ ou seja, o que já se sabe sobre os critérios avaliados no banco de dados.

Diante das informações coletadas e processadas pela abordagem Bayesiana, é possível reescrever a Eq. 4.1 de acordo com o protocolo de máxima privacidade desenvolvido, que resulta na seguinte equação:

$$P(Co_j|E, C, D, T, De, O, B) = \frac{P(E, C, D, T, De, O, B|Co_j) \times P(Co_j)}{P(E, C, D, T, De, O, B)} \quad (4.2)$$

Onde Co_j é a probabilidade de Compra do departamento j, que está condicionada às variáveis: Estado (E), Cidade (C), Data (D), Turno de acesso (T), Dispositivo (De), Sistema Operacional (O) e Navegador (B).

Como um classificador Bayesiano é um caso particular de um Rede Bayesiana[52], diferentes estruturas[78] foram avaliadas para obter o modelo que satisfaça a obtenção coerente do departamento com maior probabilidade de compra.

Esta escolha ocorreu simplificando a estrutura no processo de tomada de decisão do algoritmo, visto que um requisito para utilização de classificadores Bayesianos é a independência entre as variáveis[70]. Dessa forma, a principal dependência ocorre entre as variáveis Cidade (C) e Estado (E).

Com isso, os modelos gráficos são apresentados na Figura 11:

O modelo ajustado foi útil visto que a dependência criada pelas variáveis De , OS e B neste momento pode não influenciar diretamente a análise. Assim, é seguro manter a variável geral dispositivo De , visto que a dependência das demais condicionantes é imediata.

O modelo ajustado também agrupa os nós Estado (E) e Cidade (C) em um valor processado anteriormente e resultante com o nó posterior Geografico (G). Esta abordagem visa minimizar o fenômeno de cidades cidades homônimas. Atualmente o Brasil possui cerca de

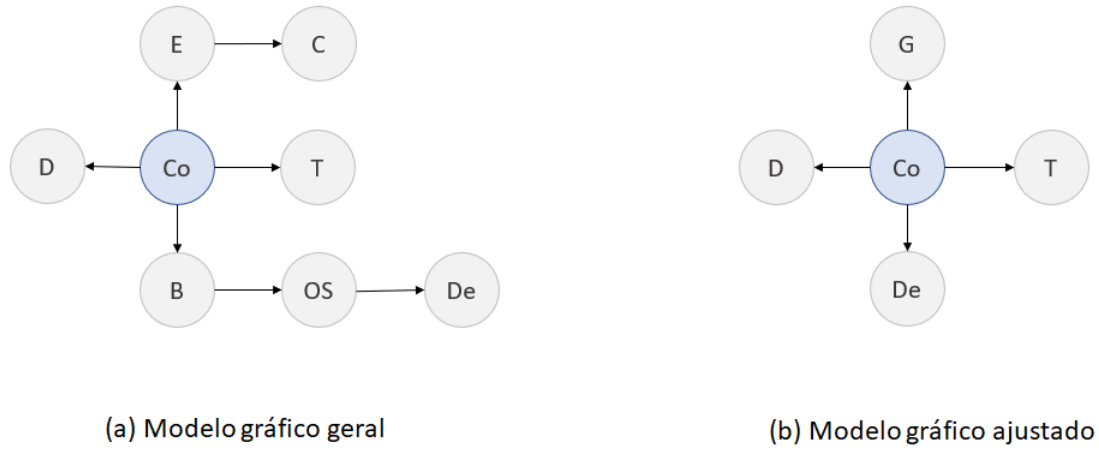


Figura 11: Estrutura gráfica dos modelos avaliados.

5.570 municípios sendo 505 homônimos¹² o que corresponde a 9,06% dos municípios. Sendo assim, para assegurar a integridade da avaliação este elo será mantido. Logo, a Equação 4.2 pode ser reescrita de acordo com o modelo ajustado dado por:

$$P(Co_j|G, D, T, De) = \frac{P(G, D, T, De|Co_j) \times P(Co_j)}{P(G, D, T, De)} \quad (4.3)$$

Diferentemente de modelos baseados em memória[3],[55],[90], esta abordagem garante um modelo em tempo real que filtra o conjunto de dados e calcula a probabilidade associado aos departamentos de acordo com critérios que estão sendo constantemente atualizados no banco de dados. Com isso, uma abordagem em tempo real é apresentada em excelente tempo computacional. Ressalta-se que o algoritmo gerador de dados desse trabalho utiliza o modelo geral apresentado na Fig.11 para obter as informações de usuários.

4.4 Teoria da Escolha Social como filtro baseado em conteúdo

A implementação de um segundo filtro visa estabelecer um refinamento no sistema de ranqueamento e preferências. O sistema processado pelo filtro anterior garante que um determinado setor apresenta maior probabilidade de compra. Entretanto, esse grupo apresenta diversos elementos contidos.

Para ordenar seus elementos, será implementado um sistema de ranqueamento baseado na Teoria da Escolha Social denominado Método de Contagem de Borda, explicado na seção 4.5.1.

Visando a obtenção do melhor processamento computacional, uma abordagem através desse método favorece o desempenho geral do modelo, uma vez que seu processamento direto

¹²Disponível em <https://www.embrapa.br/manual-de-referenciacao/anexo-cidades-homonimas>

e não envolve aplicações matriciais.

Além de ser um método clássico e pioneiro, o uso desse procedimento neste trabalho está no fato de sua relevante contribuição em diversas áreas do conhecimento, tais como biologia [6], medicina [31] e direito internacional [33].

5 RESULTADOS E DISCUSSÃO

5.1 Análise dos dados sintéticos

5.1.1 Obtenção dos dados

O processo gerador de dados foi implementado utilizando a ferramenta para modelagem baseada em agentes Netlogo¹³. É reconhecido que um dos principais problemas em sistemas de recomendação é o fenômeno de escalabilidade [38],[8],[35], ou seja, conforme o conjunto de dados aumenta, suas características tornam o processo de classificação cada vez mais difícil.

Como o passo principal dessa etapa é criar um sistema capaz de gerar um conjunto de dados com determinadas características de uma plataforma digital de conteúdos, o ferramenta Netlogo apresentou algumas limitações. Para contornar essa situação, o algoritmo proposto na Figura 9 foi implementado completamente em linguagem Python¹⁴. O desempenho computacional foi melhorado e o problema de escalabilidade contornado no processo de geração de dados. A Tabela 3 informa os resultados desse procedimento de migração de tecnologia para obtenção dos dados em tempo computacional aceitável.

Tabela 3: Desempenho no processo gerador dos dados

	Netlogo	Python
Ambiente	GUI	Console
Agentes	100	100.000
Interações	8M	+10M
Tempo	168h	15h

O desempenho computacional foi expressivo quando a tecnologia foi alterada. Nota-se que além de aumentar a quantidade de entidades interagindo (1000 vezes maior), o tempo computacional demonstrou ser 91,07% mais eficiente. A Tabela 4 informa o procedimento incremental até obtenção do banco completo.

¹³Linguagem de programação e um ambiente de desenvolvimento integrado baseada em agentes, disponível em <https://ccl.northwestern.edu/netlogo/>

¹⁴Linguagem de programação de alto nível, interpretada, orientada a objetos, de tipagem dinâmica e forte, disponível em python.org

Tabela 4: Análise incremental e desempenho computacional

Ambiente	Agentes	Interações	Tempo
NetLogo	100	1.000	1.29s
Python	100	1.000	1.16s
NetLogo	100	10.000	15.6s
Python	100	10.000	14.2s
NetLogo	100	100.000	5.9m
Python	100	100.000	3.8m
NetLogo	100	1.000.000	3.2h
Python	100	1.000.000	45m
NetLogo	100	3.000.000	26.3h
Python	100	3.000.000	3.2h
NetLogo	100	8.000.000	168h
Python	100.000	10.000.000	15h

Dessa forma, o conjunto de dados sintético foi obtido com a modelagem proposta. Até o momento não identificado na literatura específica uma ferramenta capaz de produzir um conjunto que simule as características independentes de intenções de compra em ambiente relacionado à comércio eletrônico por entidades que tomam decisões autônomas de acordo com critérios técnicos, racionais ou emocionais.

5.1.2 Análise descritiva dos dados

Uma avaliação descritiva dos dados visa obter confirmação de como o processo de simulação das interações entre os agentes ocorreu, e se esta pode ser utilizada de maneira satisfatória como base de informação para implementações futuras de métodos para recomendação de conteúdo.

A primeira análise consiste em verificar como as entidades acessam e demonstram seus interesse referente aos diferentes departamentos no cenário de comércio eletrônico modelado. Existem departamentos abstratos possíveis para seleção. A metodologia para obtenção desses resultados consiste em nove simulações com um milhão de interações. O resultado apresentado na Figura 12 mostra como a seleção dos departamentos ocorreu durante as simulações realizadas.

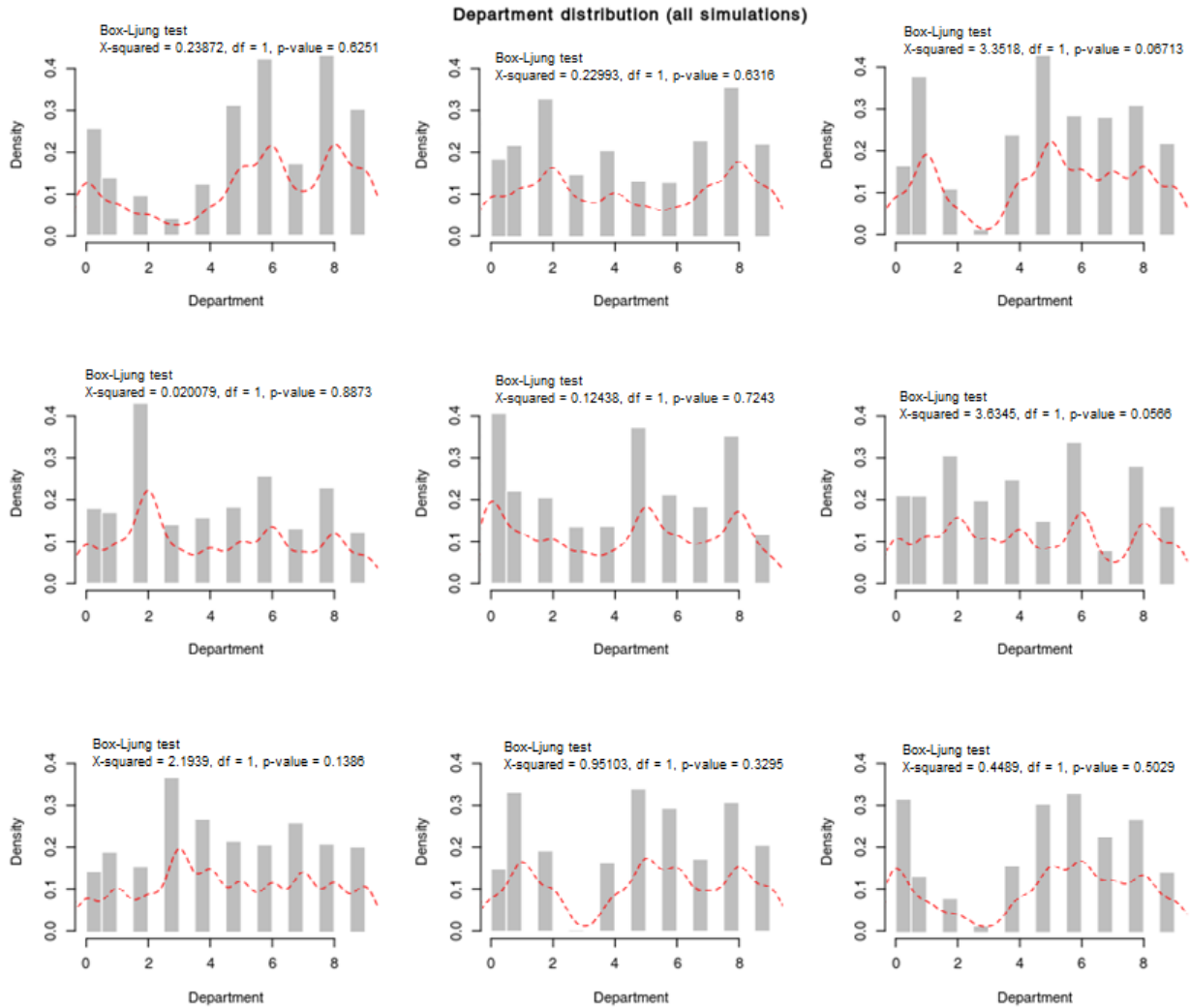


Figura 12: Distribuição do acesso aos departamentos.

Os resultados mostraram que não existe uma relação que estabeleça alguma pré disposição à seleção de algum departamento, tornando assim o processo de seleção completamente aleatório. É possível notar que diferentes simulações resultam em diferentes departamentos escolhidos pelos agentes. Além disso, é possível perceber que, de acordo com a estatística Ljung-Box, não foram encontradas evidências para rejeitar a hipótese de que as escolhas ao longo do tempo são de fato aleatórias e independentes.

Para assegurar a integridade do resultado obtido, alguns testes estatísticos foram realizados. Para verificar a homogeneidade das variâncias entre cada simulação foi conduzido o teste de Fligner-Killeen (Fligner-Killeen K-squared = 8.4362, df = 8, p-value = 0.3921). Com isso, não é possível rejeitar a hipótese de que variância entre as simulações foram praticamente iguais ($p > 0.05$). Este resultado foi condizente com o esperado pois, embora a simulação produza valores distintos, é previsto que a dispersão entre as seleções de preferência entre os agentes envolvidos não seja desproporcionalmente apresentada.

Contudo, é necessário ainda assegurar que as simulações obtidas sejam estatisticamente

semelhantes. Esta premissa visa garantir a integridade dos dados gerados e que, embora sejam dinamicamente produzidos, toda nova simulação resulte em características adequadas ao comportamento modelado. Para isso, o teste de Kruskal-Wallis ¹⁵ foi realizado (Kruskal-Wallis chi-squared = 1.0899, df = 8, p-value = 0.6976) e, com base nas simulações realizadas, existem evidências estatísticas entre as populações obtidas ($p > 0.05$) de que as simulações são semelhantes. Esse resultado é importante pois garante que todo processo simulado é capaz de produzir resultados coerentes, porém distintos.

Por fim, os dados devem ser estatisticamente não correlacionados, e isso deve ser confirmado com um teste de correlação de Spearman ¹⁶ para toda matriz da distribuição dos departamentos acessados na simulação. O intuito em verificar esta característica está no fato de que cada novo processo gerado deve ser independente e não produzir resultados pseudo-aleatórios. Se existir alguma correlação forte, é possível que em determinado momento o código possa estar induzindo a simulação à algum comportamento. A Figura 13 demonstra as correlações entre as simulações obtidas.

Com os resultados obtidos, percebe-se que apenas três simulações apresentaram um grau de correlação maior do que 0.7, o que poderia indicar alguma correlação forte. Contudo, como outras simulações apresentaram valores próximos de zero, é possível afirmar que correlações mais elevadas, quando ocorrer, podem ser derivadas de fatores associadas à erros aleatórios produzidos e não necessariamente implica alguma tendência no algoritmo gerador de dados.

¹⁵O teste de Kruskal-Wallis (KW) é um teste não paramétrico utilizado para avaliar a hipótese nula de que todas as populações possuem funções de distribuição iguais contra a hipótese alternativa de que ao menos duas das populações possuem funções de distribuição diferentes.

¹⁶O coeficiente de correlação de Spearman é uma medida não paramétrica da correlação de postos (dependência estatística do ranking entre duas variáveis) e usado principalmente para medir a força e direção da associação entre duas variáveis classificadas

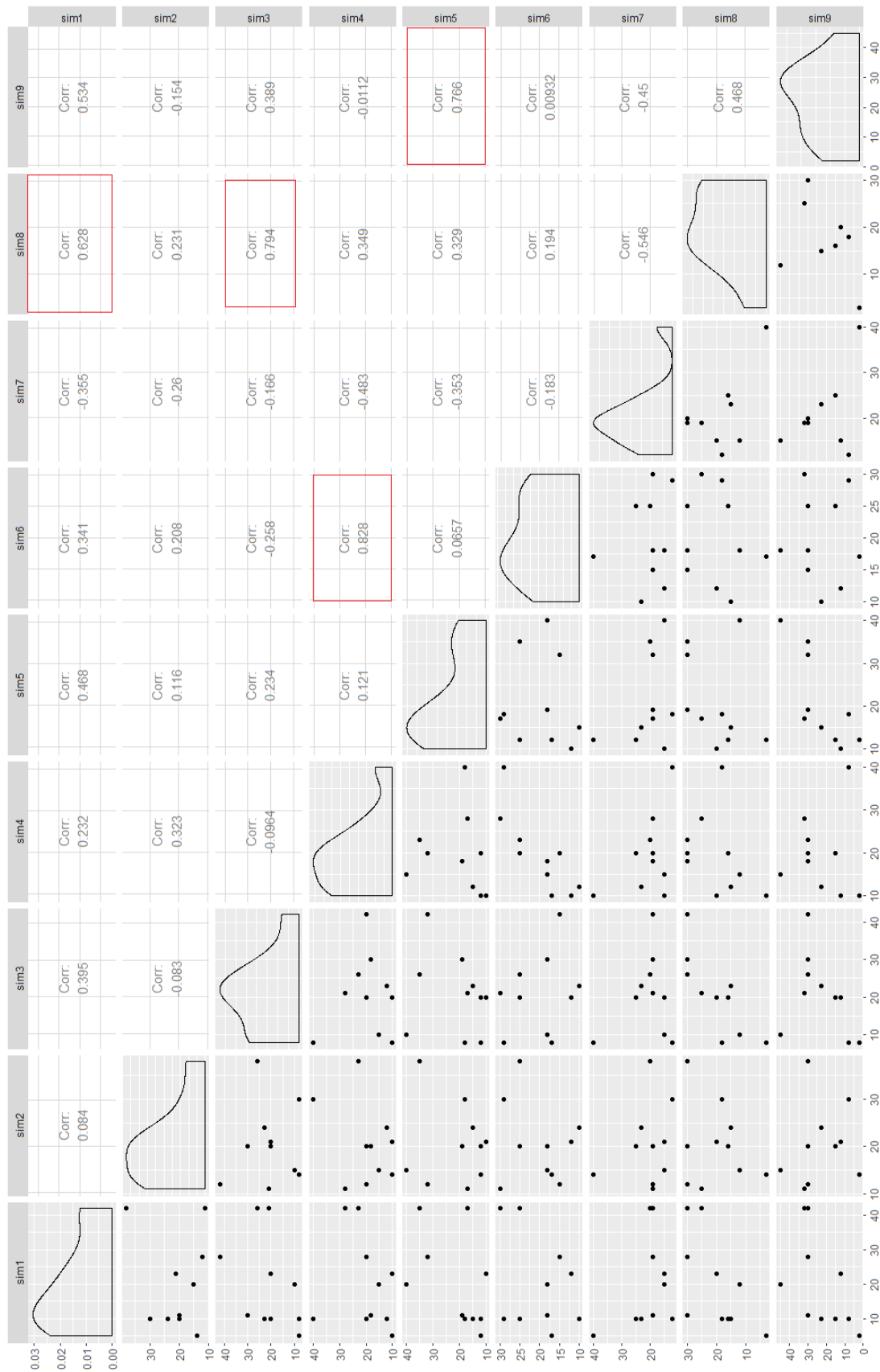


Figura 13: Estudo da correlação entre as simulações e acesso aos departamentos

De acordo com os resultados obtidos e a avaliação estatística, é possível afirmar que as

simulações apresentam consistência nos resultados e relevante para o andamento posterior e objetivo da tese, visto que a aleatoriedade constatada assegura que o algoritmo é capaz de criar entidades realmente autônomas e que avaliam de forma independente qual departamento acessar. Vale destacar que os setores preferenciais dos agentes pode ser alterado durante a execução, de acordo com avaliações prévias ou quantidade de itens adquiridos. Essa característica reforça ainda mais o comportamento aleatório e não convergente de todo processo simulado e, portanto, não apresentam tendência involuntária que caracteriza algum erro no processo de seleção.

Esta análise pode ser expandida para os demais processos realizados pelo algoritmo gerador. Dessa maneira, a etapa seguinte consiste em verificar a integridade e distribuição de acesso aos itens contidos em cada departamento selecionado, conforme demonstra a Figura 14.

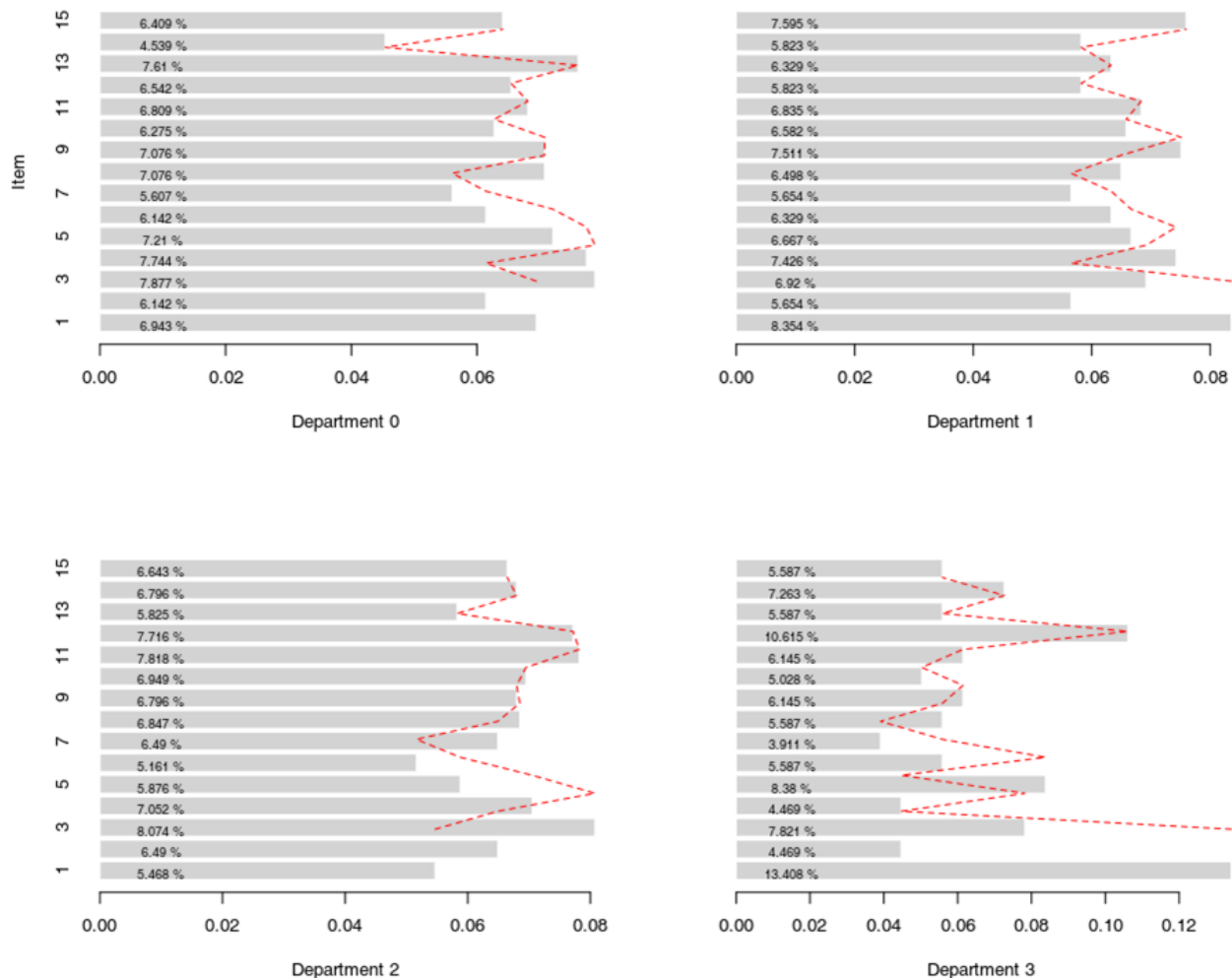


Figura 14: Amostra da distribuição do acesso aos itens contidos em departamento.

Da mesma forma que ocorreu com a seleção dos departamentos, o acesso a cada item

contido no departamento, arbitrariamente selecionado, demonstra que também foi realizado de maneira aleatória porém com evidências de uma mesma população (Fligner-Killeen K -squared = 9.3415, $df = 14$, p -value = 0.8086, Kruskal-Wallis χ^2 -squared = 14.893, $df = 14$, p -value = 0.3855). Essa aleatoriedade se faz presente devido ao *Processo de Influência* e *Processo de escolha* modelados no algoritmo para geração de dados.

Esta etapa demonstra que o algoritmo gerador não tem tendência na elaboração e acesso aos itens, assim não favorecendo determinado item sobre os demais. Em nenhum cenário avaliado, o sistema converge. Essa característica é extremamente forte para o bom andamento do processo e assegura a integridade, aleatoriedade e dinâmica na obtenção de um conjunto de dados capaz de simular aplicações de comércio eletrônico.

Analogamente, foi avaliada a distribuição de acesso e interesse de consumo durante as horas do dia e os dias da semana. Percebe-se, de acordo com a Figura 15, que a distribuição para ambos casos é quase homogênea. Devido a grande quantidade de dados produzidos, nenhuma presença latente que sinalize alguma preferência, ou tendência erroneamente gerada, de horário de acesso ou dias da semana foi detectada.

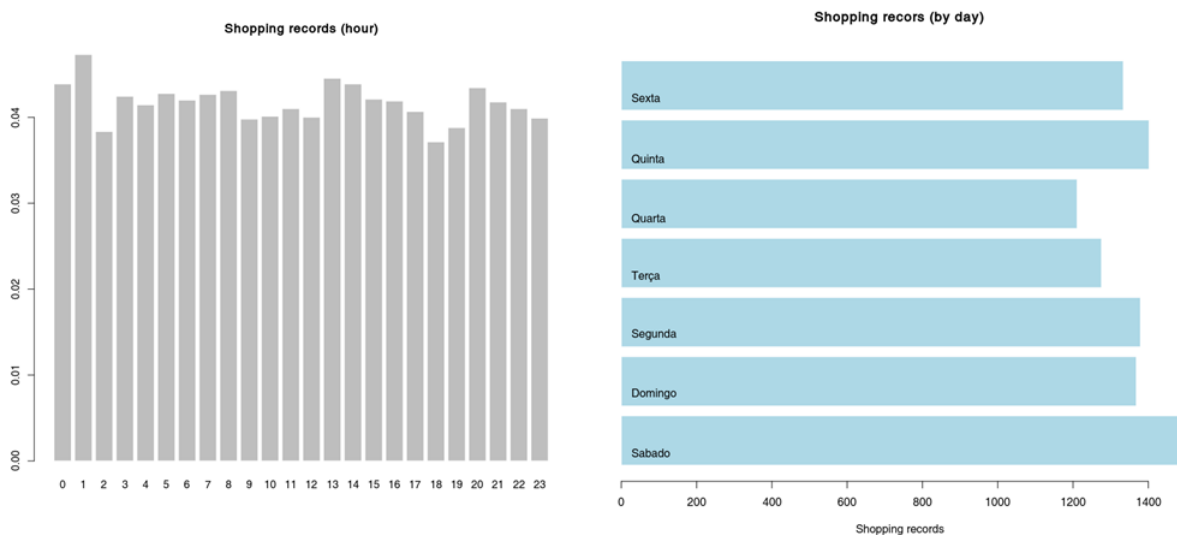


Figura 15: Avaliação da distribuição de acesso referente às horas e dias da semana

Este critério é importante pois será a base de informação para a obtenção do turno de acesso identificado no protocolo desenvolvimento.

Com informações de marketshare ¹⁷ o algoritmo foi calibrado visando identificar, novamente, se algum agente tende a estabelecer alguma preferência sobre as demais entidades que estão interagindo no ambiente de simulação.

¹⁷Dados coletados dos Institutos NetMarketShare e StatCounter Global Stats, disponível em netmarketshare.com e gs.statcounter.com

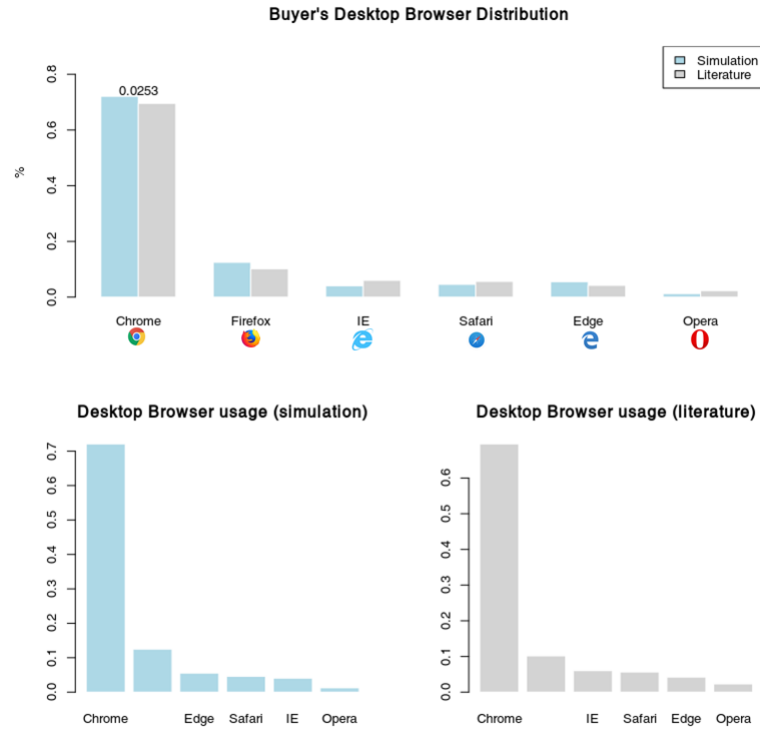


Figura 16: Comparação entre simulação e literatura referente aos navegadores

Nota-se que o processo de tomada de decisão dos agentes envolvidos na simulação não apresenta tendência. Isto é mais uma vez garantido com os dados obtidos via relatórios de participação no mercado. Uma vez o sistema calibrado, a premissa de que, caso o algoritmo gerados apresentasse algum erro que implicasse em tendência, os valores de marketshare deveriam estar destoados da realidade obtida. Com isso, os dados gerados foram coerentes e robustos ao apresentar resultados aleatórios.

A Figura 16 informa ainda um fenômeno de seleção. A distribuição presente no processo de seleção referente aos navegadores de acesso foram semelhantes, contudo a utilização dos navegadores IE e Edge apresentam colocações distintas. O ocorrência desta característica reforma o argumento de que, embora o sistema seja calibrado com dados de marketshares, o processo de simulação continua randômico e produzindo resultados condizentes com a realidade porém não forçados à tendências arbitrárias ou intencionais.

Avaliação análoga pode ser realizada para valores obtidos referentes ao tipo de dispositivo utilizado pelo agente no processo de tomada de decisão conforme demonstra a Figura 17.

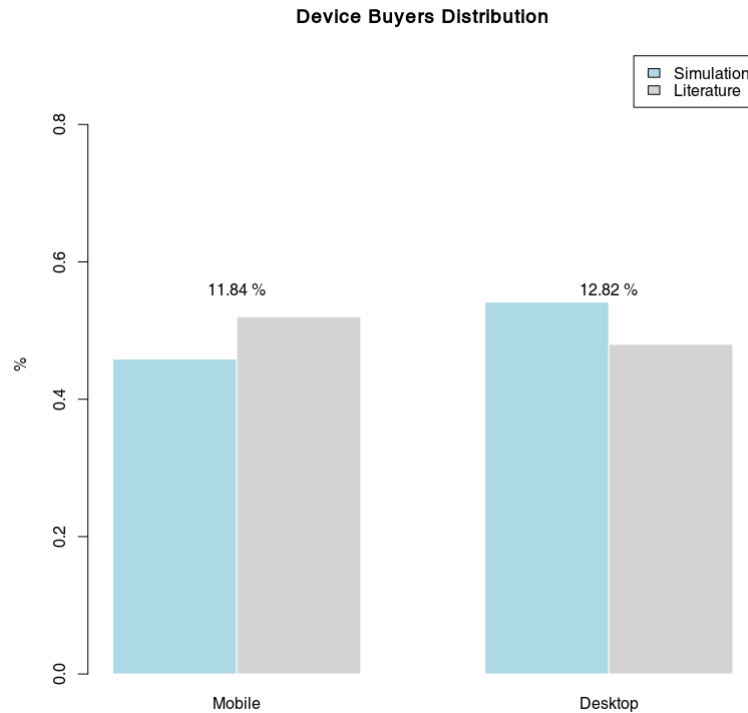


Figura 17: Comparação entre simulação e literatura referente aos tipos de dispositivos

Os valores obtidos nos processos de simulação apresentam um erro médio, de acordo com a distribuição amostral das médias das diferenças, correspondendo à 11.84% para dispositivos do tipo Mobile (dispositivos móveis) e 12.82% para dispositivos Desktop. Esse resultado representa um importante passo no processo de simulação assegurando a aleatoriedade e dinâmica de tomada de decisão dos agente no ambiente, visto que, embora seja ponderado os valores disponibilizados de participação no mercado, o sistema comporta-se de maneira independente, porém coerente com a realizada apresentada nos relatórios de pesquisas institucionais com este propósito¹⁸

Por fim, o processo de obtenção de reviews, ou seja, a opinião deixada pelos usuários nos itens selecionados foi avaliada. Esta etapa é relevante na simulação visto que é o cerne no processo de tomada de decisão em abordagens racionais. Quando um agente decide ponderar sua escolha e efetivar ou não a aquisição do item, são estes valores que serão observados. A Figura 18 mostra o comportamento dos valores de avaliação de itens de acordo com o banco de dados disponibilizado para pesquisa de registro da Amazon.

¹⁸Conforme dados obtidos dos Institutos NetMarketShare e GS StatCounter.

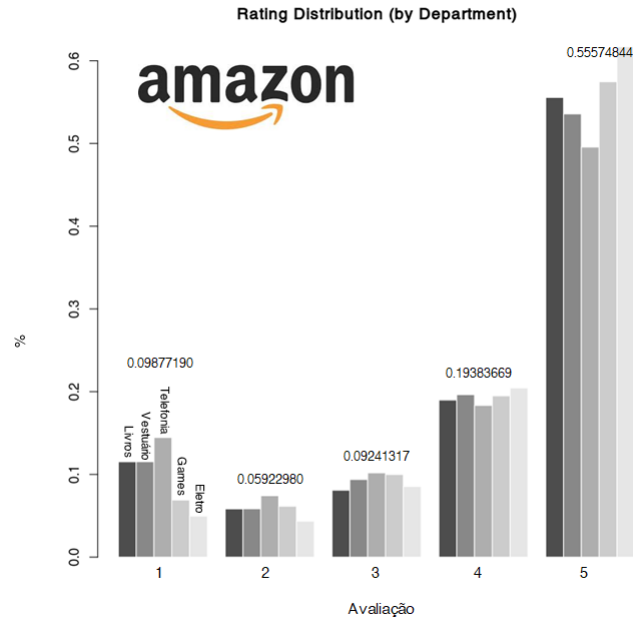


Figura 18: Distribuição dos reviews da Amazon

Percebe-se que diversos departamentos apresentam um comportamento similar: avaliações com nota máxima é latente para vários departamentos. Assim, de acordo com as simulações realizadas, foi coletado a média das distribuições e valores de avaliações e comparado com os dados obtidos na literatura. A Figura 19 mostra este resultado.

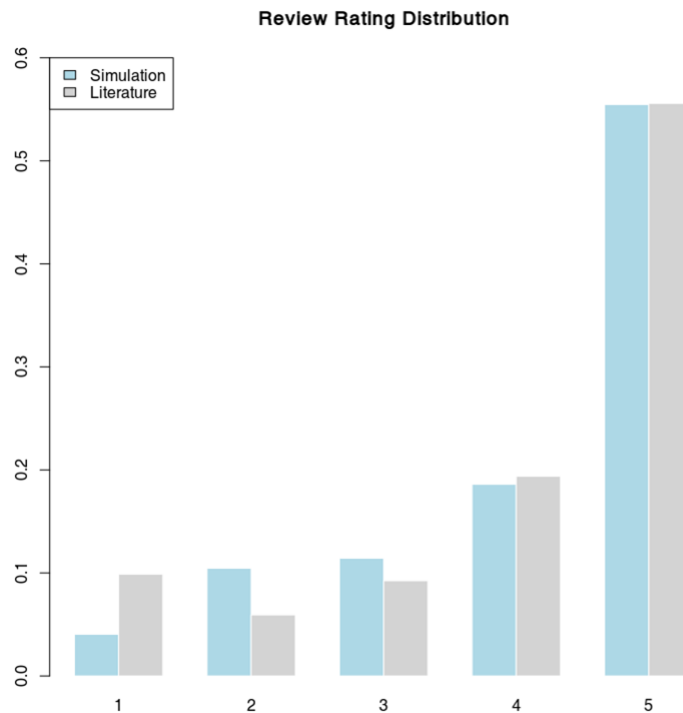


Figura 19: Comparação da distribuição dos resultados de reviews.

Os valores obtidos no processo de geração de dados sintéticos foram extremamente satisfatórios e, através de inspeção visual direta, é reconhecido que representam, incluindo todo fator aleatório da simulação, resultados condizentes com dados coletados e disponibilizados por uma empresa real. Com isso, é assegurado que o algoritmo gerador de dados sintéticos é capaz de produzir informações simuladas de usuários em um sistema de comércio eletrônico registrando seu comportamento e todo processo de tomada de decisão.

5.2 Aplicação modelo

A etapa final consiste em apresentar uma proposta de aplicação do conceito demonstrado nesse trabalho. A elaboração de uma estrutura capaz de coletar informações não invasivas do usuário, filtrar esses dados e apresentar um conteúdo personalizado que reflita o interesse do coletivo e tenha a maior probabilidade de que seja consumido.

Para isso, a Figura 20 ilustra uma sequência de passos que devem ser executados para que o conteúdo seja recomendado ao usuário desconhecido. Essa abordagem deve ser inicial e executada para cada usuário não reconhecido pela plataforma, caracterizando assim a solução proposta para contornar o problema de *cold-start* em sistemas de recomendação.

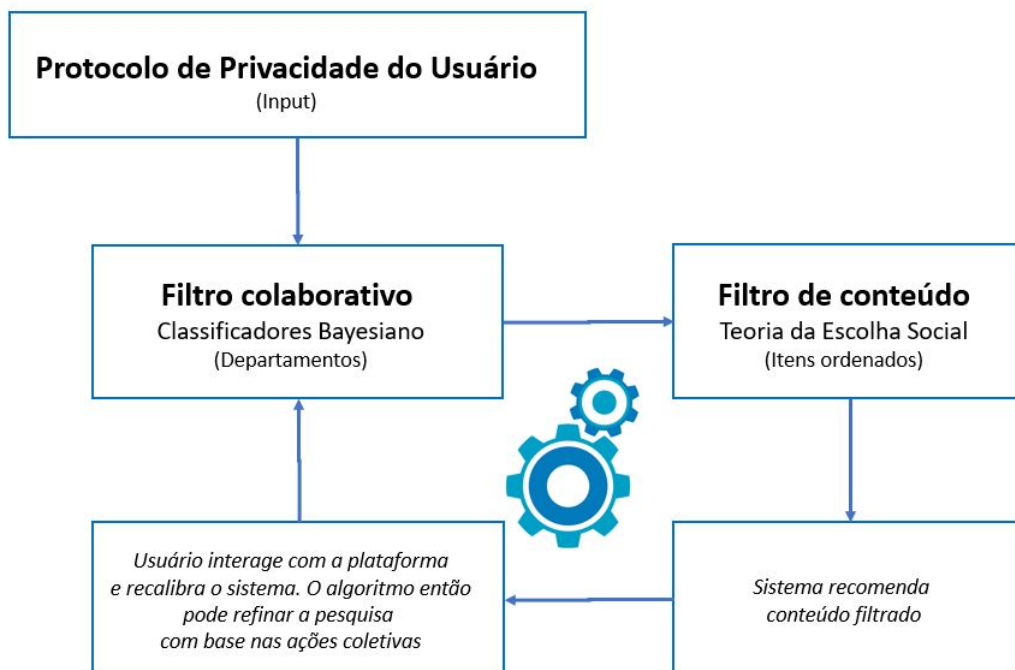


Figura 20: Estrutura de processamento de informações e recomendação de conteúdo

Uma vez atendida a estrutura proposta, o sistema implementado terá condições de coletar as informações iniciais dos usuários desconhecidos e, no banco de dados da aplicação, realizar o processamento dos filtros.

5.3 Processamento dos filtros: Agente de Interface

A abordagem proposta, através de classificadores Bayesianos como filtro colaborativo faz com que o resultado produzido seja um característico numérico que representa o departamento com maior probabilidade de compra dado os critérios intrínsecos do usuário que está acessando a plataforma de comércio eletrônico.

Com isso, é possível obter a probabilidade para todos departamentos cadastrados no ambiente modelo e organizá-los em ordem decrescente de acordo com a probabilidade obtida. Para agilizar todo processo, o sistema faz um refinamento onde apenas o primeiro departamento é de interesse da aplicação. Sendo assim, dentre todos departamentos cadastrados, o sistema é capaz de, em tempo real, selecionar aquele que apresenta maior probabilidade de compra dado os critérios do usuário que estiver acessando o sistema naquele exato momento.

Inicialmente, é simulado o acesso por um usuário arbitrário coletando as características definidas no protocolo e solucionando. Apenas para fins de demonstração do equacionamento, por ser uma expansão extensa para o modelo geral, será utilizado o modelo gráfico ajustado da Fig.11.

$$P(Co_j|G, D, T, De) = \frac{P(G, D, T, De|Co) \times P(Co_j)}{P(G, D, T, De)} \quad (5.1)$$

O sistema então pega as informações desse usuário e processa em tempo real no conjunto de dados. Logo a Equação 4.3 deve ser avaliada conforme o classificador Bayesiano prevê. Assim, o calculo abaixo deve ser satisfeito:

$$P(Co_j|G, D, T, De) = \frac{P(G, D, T, De|Co_j) \times P(Co_j)}{P(G, D, T, De)} = \quad (5.2)$$

$$= \frac{P(G|Co_j) \times P(D|Co_j) \times P(T|Co_j) \times P(De|Co_j) \times P(Co_j)}{P(G|Co_j)P(D|Co_j)P(T|Co_j)P(De|Co_j) + P(G|\neg Co_j)P(D|\neg Co_j)P(T|\neg Co_j)P(De|\neg Co_j)} \quad (5.3)$$

O denominador da equação 5.3 é um fator normalizador. Logo a utilização deste fator não influencia[87],[26] o resultado do classificador visto que o interesse é encontrar a maior distribuição posterior[28] conforme demonstrado na Equação 3.3. Implementações de Classificador Bayesian usualmente desconsideram esse elemento visto que não apresenta relação com a classe avaliada[26]. Com isso, este fator foi removido do processo para ganho de tempo em desempenho computacional e assim retornando apenas o valor da verossimilhança relativa, e não uma probabilidade propriamente explícita dado a não utilização do fator normalizador.

Portanto, o sistema está interessado em encontrar o departamento que apresente maior verossimilhança relativa de compra definido de acordo com a equação dada por:

$$P(Co_j|G, D, T, De) \propto P(G, D, T, De|Co_j) \times P(Co_j) = \quad (5.4)$$

$$= P(G|Co_j) \times P(D|Co_j) \times P(T|Co_j) \times P(De|Co_j) \times P(Co_j) \quad (5.5)$$

Assim, a Equação 5.6 define qual departamento (j) deve ser escolhido como sendo aquele que apresenta a maior verossimilhança relativa de compra dado as características do usuário.

$$P(Co_j|X) \propto \left(\prod_{i=1}^n P(X_i|Co_j) \right) P(Co_j) \quad (5.6)$$

Uma vez implementado, o agente de interface passa a filtrar os dados do conjunto gerado em tempo real quando os dados do usuário são informados. A Figura 21 mostra um exemplo de usuário acessando a plataforma retornando o departamento com a maior verossimilhança relativa de compra dado as características do usuário.

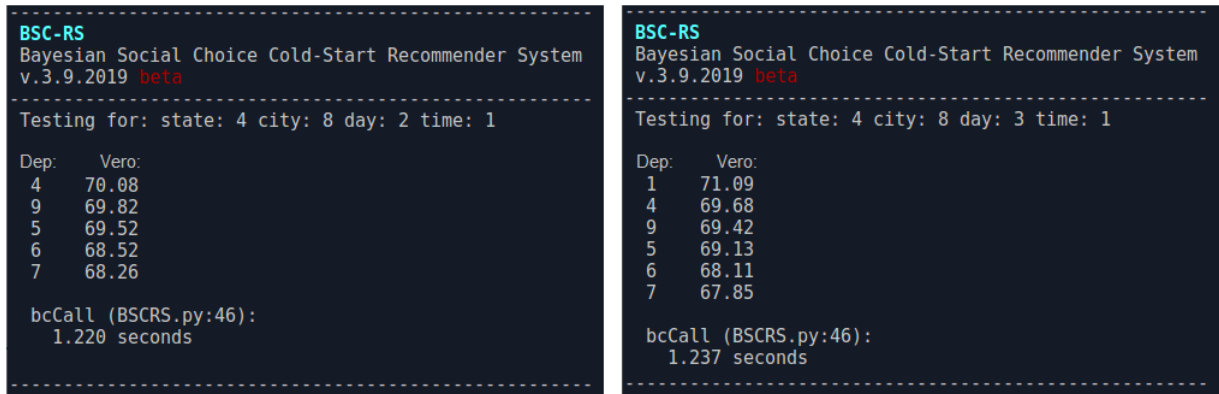


Figura 21: Análise do agente de interface e processamento Bayesiano. Exemplo 1.

Um usuário arbitrário executa dois acessos ao sistema. A Tabela 5 demonstra as informações coletadas do usuário.

Tabela 5: Dados de acesso coletadas pelo protocolo

Acesso	Estado	Cidade	Dia da semana	Turno	Dispositivo
#1	4	8	2	1	0
#2	4	8	3	1	0

Com estas informações, o sistema é capaz de processar em tempo real e filtrar o departamento adequado para o cliente. Note que no primeiro acesso, ocorre no dia descrito como 2. Para fins de avaliação, seja 2 considerado Segunda-feira.

Com base nos dados, o usuário da Cidade 8 e Estado 4, quando acessa o sistema no turno da manhã (Turno = 1) o departamento com maior verossimilhança relativa de compra

é identificado como sendo o Departamento 4, com uma verossimilhança relativa de 70,08.

Contudo, quando este mesmo usuário acessa a plataforma no dia seguinte (Dia = 3) no mesmo turno, o sistema recomenda um departamento completamente diferente, sendo agora o Departamento 1 com uma verossimilhança relativa de 71,09.

```

BSC-RS
Bayesian Social Choice Cold-Start Recommender System
v.3.9.2019 beta
-----
Testing for: state: 4 city: 8 day: 3 time: 3

Dep:    Vero:
2       67.77
9       67.59
7       64.93
6       63.63

bcCall (BSCRS.py:46):
1.648 seconds

```

Figura 22: Análise do agente de interface e processamento Bayesiano. Exemplo 2.

Entretanto, quando outro usuário acessa a plataforma no mesmo dia e turno (Dia = 2, Turno = 1) o sistema recomenda departamentos completamente conforme demonstra a Figura 23.

```

BSC-RS
Bayesian Social Choice Cold-Start Recommender System
v.3.9.2019 beta
-----
Testing for: state: 0 city: 1 day: 2 time: 3

Dep:    Vero:
9       71.25
0       70.32
4       69.84
1       69.59
5       69.29
8       68.85
6       68.28
7       68.02
2       67.77

bcCall (BSCRS.py:46):
1.584 seconds

```

Figura 23: Análise do agente de interface e processamento Bayesiano. Exemplo 3.

Este resultado é importante visto que o processo de recomendação de conteúdo personalizado de acordo com as características de utilizador reflete exatamente o esperado: usuários distintos, conteúdos distintos.

Com base nos dados obtidos, o tipo de dispositivo não apresentou diferença expressiva quanto à recomendação de departamento. Contudo, dentre os recomendados, alguns, excluindo o primeiro, apresentaram ordenação distintas.

Uma vez obtido o departamento com maior probabilidade de compra, é necessário ordenar os itens contidos nesse departamento de acordo com a preferência dos usuários. Será utilizado como critério os itens no qual o processo de compra tenha sido finalizado.

Do ponto de vista prático, considerar o item que de fato foi adquirido é relevante para uma aplicação comercial, visto que existe um potencial de comercialização para o bem destacado e dessa forma potencializando sua receita.

O filtro anterior retorna um característico numérico correspondente ao índice do departamento. Nesta etapa, o sistema deverá pesquisar todo conjunto de dados e ordenar seus itens. Para isso, será utilizado o método para obter um indicativo de preferência coletiva com base em decisões individuais, denominado contagem de Borda.

A vantagem nessa abordagem reside no fato de que, nem sempre o item mais vendido, será necessariamente o preferido pelos usuários. Existe um ordenamento no qual deve ser observados suas transitividades no processo de escolha.

O sistema então, dentro do departamento alvo, filtrará todas as compras realizadas por usuários que pertençam à amostragem realizada de acordo com os critérios de análise (protocolo de privacidade e departamento alvo).

Com isso, seja X_u uma amostra de n itens adquiridos por usuários que satisfaçam os critérios resultantes do filtro Bayesiano faça:

$$\forall X_u^j \Rightarrow \forall x_i \in X_u^j | x_n = 1, x_{n-1} = 2, x_{n-2} = 3 \dots \quad (5.7)$$

Assim, seja x_α os itens indexados contidos no departamento alvo, então:

$$\forall x_\alpha \in X \sum x_\alpha \in X_u^j \quad (5.8)$$

Dessa forma, $\exists x_\alpha \in X_u^j$ tal que esse item é o de maior preferência entre os usuários selecionados.

A Figura 24 demonstra o algoritmo para processamento do filtro de conteúdo de forma completa.

O algoritmo identifica em tempo real o as informações do usuário e processa o primeiro filtro (F1). O resultado desse é então utilizando como parâmetro para o filtro baseado em conteúdo (F2) através da Teoria da Escolha Social.

O resultado retorna uma lista de itens ordenados de acordo com a preferência dos usuários, ou seja, uma preferência coletiva é recomendada ao usuário através de uma classificação de preferências individuais.

A Figura 25 demonstra um resultado de maneira mais completa.

A Figura 24 demonstra as informações e recomendação para um usuário arbitrário. O departamento selecionado foi o 5 e a lista de itens ordenados é dada pelo filtro F2. Entretanto, caso outro usuário acesse a plataforma no mesmo dia (Weekday=3) e mesmo turno (DateTime=3), Figura 25.b, o sistema ainda retorna como departamento com maior verossimilhança relativa de compra: o departamento 5. Contudo, a lista ordenada de itens é completamente diferente.

Esse resultado demonstra que o sistema é dinâmico, e que o comportamento dos usuários,

```

-----
BSC-RS
Bayesian Social Choice Cold-Start Recommender System
v.22.1.2021 beta
-----

User information:
State: 3
City: 5
Weekday: 3
DateTime:3
-----

F1 : Department
Dep:      Vero:
5         70.49
1         70.23
0         69.98
9         68.97
4         67.70
8         67.67
6         66.80
2         64.92

bcCall (bscrs.py:64):
0.364 seconds

Department selected: 5
-----

F2 : Item ID Rank
[2, 6, 10, 4, 11, 7, 9]
-----

sctbc (bscrs.py:178):
0.056 seconds

```

Figura 24: Análise do agente de interface e raqueamento de itens

```

-----
BSC-RS
Bayesian Social Choice Cold-Start Recommender System
v.22.1.2021 beta
-----

User information:
State: 3
City: 5
Weekday: 3
DateTime:3
-----

F1 : Department
Dep:      Vero:
5         70.49
1         70.23
0         69.98
9         68.97
4         67.70
8         67.67
6         66.80
2         64.92

bcCall (bscrs.py:64):
0.364 seconds

Department selected: 5
-----

F2 : Item ID Rank
[2, 6, 10, 4, 11, 7, 9]
-----

sctbc (bscrs.py:178):
0.056 seconds

```

(A)

```

-----
BSC-RS
Bayesian Social Choice Cold-Start Recommender System
v.22.1.2021 beta
-----

User information:
State: 0
City: 1
Weekday: 3
DateTime:3
-----

F1 : Department
Dep:      Vero:
5         69.38
1         69.11
9         67.83
4         66.53
8         66.50
6         65.62
7         65.56
2         63.71

bcCall (bscrs.py:64):
0.332 seconds

Department selected: 5
-----

F2 : Item ID Rank
[3, 8, 14, 10, 6, 11, 5]
-----

sctbc (bscrs.py:178):
0.050 seconds

```

(B)

```

-----
BSC-RS
Bayesian Social Choice Cold-Start Recommender System
v.22.1.2021 beta
-----

User information:
State: 0
City: 1
Weekday: 3
DateTime:1
-----

F1 : Department
Dep:      Vero:
9         67.80
4         66.51
6         65.60
7         65.53

bcCall (bscrs.py:64):
0.223 seconds

Department selected: 9
-----

F2 : Item ID Rank
[1, 7, 3, 10, 2, 11, 9, 15]
-----

sctbc (bscrs.py:178):
0.057 seconds

```

(C)

Figura 25: Análise do agente de interface e raqueamento de itens comparado com outros usuários

de um mesmo local, pode implicar em recomendações complemente distintas dado o perfil de consumo de outras variáveis, como o dia de semana e turno de acesso.

Consequentemente caso um usuário de mesmo estado e cidade acesse o ambiente no mesmo dia, entretanto em turnos distintos, conforme visto na Figura 25.b e Figura 25.c, o

padrão comportamental dos demais agentes no banco faz que o resultado novamente seja distinto. Essa capacidade de filtro em tempo real das características e ordenação de preferências faz com que o algoritmos seja capaz de detectar mínimas variações.

A vantagem do método de Teoria da Escolha Social adotado reforça essa abordagem, uma vez que o item que apresenta diretamente mais vendas não significa, necessariamente o mais preferido pelos usuários.

6 CONSIDERAÇÕES FINAIS

Toda abordagem realizada neste trabalho direcionará futuras pesquisas na área de tecnologias persuasivas e sistemas de recomendação, pois, através do algoritmo gerador de dados sintéticos novos processos persuasivos e de recomendação de conteúdo podem ser explorados e implementados. Visto que um banco de dados completo com características, intenções e processo de tomada de decisão de agentes autônomos são registrados em tempo real pode ser gerado dinamicamente, esse novo recurso poderá facilitar a aquisição de dados e consequentemente agilizar o processo de desenvolvimento de novas metodologias.

O sistema multagente foi implementado para atuar como base no processo de aquisição de dados. Posteriormente, o agente de interface é responsável pelo processamento dos filtros em dois estágios: por explorar um filtro colaborativo e refinando a sugestão com filtros baseados em conteúdo.

Outro ponto de destaque é de que a privacidade do usuário é preservada, visto que, embora seja um usuário desconhecido para a plataforma, dados implicitamente coletados podem auxiliar no processamento da recomendação. Com isso, com dados públicos estabelecidos nos protocolos de comunicação digital, o usuário começa a receber dados personalizados.

O tempo computacional obtido no processamento com Classificadores Bayesianos atuando como filtro colaborativo e a Teoria da Escolha Social como filtro baseado em conteúdo foram extremamente satisfatórios, visto que todo algoritmo foi desenvolvido e executado em um ambiente computacional não profissional.

Uma vez implementada essa solução em computadores de alto desempenho e serviços comerciais, embora também aumente significativamente o volume de dados, os algoritmos tendem a apresentar um melhor desempenho computacional, visto que em ambientes de simulação os dados gerados foram na grandeza de aplicações comerciais reais.

Com isso, contorna-se também um problema de escalabilidade, sendo esse um efeito colateral de implementações persuasivas e em sistemas de recomendação de conteúdo.

A aplicabilidade do algoritmo gerador de dados pode ir além de processos de análise em situações de *cold-start*. Qualquer futuro trabalho que necessite de dados resultantes de processos interativos entre usuários poderá utilizar o método implementado nesse trabalho.

6.1 Limitações e Trabalhos futuros

Este trabalho apresentou uma solução para o problema de *cold-start* utilizando uma combinação de metodologias ainda não exploradas na literatura. A principal limitação foi a obtenção de dados, mas foi contornada com o desenvolvimento de um algoritmo baseado em agentes capaz de emular padrões comportamentais de usuários de um ambiente de comércio eletrônico.

Os objetivos dessa tese foram atingidos com o desenvolvimento do algoritmo gerador, e

com a adequada modelagem através de sistemas multiagente. A elaboração dos filtros colaborativos através de Classificadores Bayesianos aliados ao processamento secundário para refinamento dos dados com procedimentos da Teoria da Escolha Social proporcionaram apresentar uma nova abordagem para solucionar o problema de *cold-start* em sistemas de recomendação. Além disso, a criação de uma metodologia capaz de ser implementada em diversos domínios de interesse da sociedade.

Com isso, pretende-se explorar a implementação dessa abordagem para o problema de *cold-start* em trabalhos futuros com dados reais, em ambiente comercial digital assim como averiguar o comportamento da metodologia proposta em outros domínios.

Referências

- [1]
- [2] Learning informative priors from heterogeneous domains to improve recommendation in cold-start user domains. *ACM Transactions on Information Systems*, 35(2):1–37, 2016.
- [3] P. H. Aditya, I. Budi, and Q. Munajat. A comparative analysis of memory-based and model-based collaborative filtering on the implementation of recommender system for e-commerce in indonesia: A case study pt x. In *2016 International Conference on Advanced Computer Science and Information Systems (ICACSYS)*, pages 303–308, 2016.
- [4] C. Aggarwal. *Data Classification: Algorithms and Applications*. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. CRC Press, 2015.
- [5] M. Albanese, A. D’Acerno, V. Moscato, F. Persia, and A. Picariello. A multimedia recommender system. *ACM Transactions on Internet Technology*, 13(1):1–32, 2013.
- [6] M. Alsharif, A. Thrwat, I. Amin, A. E. Hassanien, and H. Hefny. Enzyme function classification based on sequence alignment. *Advances in Intelligent Systems and Computing*, 340:409–418, 01 2015.
- [7] F. Alyari and N. Jafari Navimipour. Recommender systems: A systematic review of the state of the art literature and suggestions for future research. *Kybernetes*, 47(5):985–1017, 2018.
- [8] A. Baldominos, Y. Saez, E. Albacete, and I. Marrero. An efficient and scalable recommender system for the smart web. In *2015 11th International Conference on Innovations in Information Technology (IIT)*, pages 296–301, 2015.
- [9] M. Barley and N. Kasabov. *Intelligent Agents and Multi-Agent Systems: 7th Pacific Rim International Workshop on Multi-Agents, PRIMA 2004, Auckland, New Zealand, August 8-13, 2004, Revised Selected Papers*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2005.
- [10] A. Bellogin, J. Parapar, and P. Castells. Probabilistic collaborative filtering with negative cross entropy. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys ’13, page 387–390, New York, NY, USA, 2013. Association for Computing Machinery.

- [11] M. Berry, A. Mohamed, and B. Yap. *Supervised and Unsupervised Learning for Data Science*. Unsupervised and Semi-Supervised Learning. Springer International Publishing, 2019.
- [12] C. Blackorby, W. Bossert, and D. Donaldson. *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*. Econometric Society Monographs. Cambridge University Press, 2005.
- [13] R. Boisvert and E. Houstis. *Computational Science, Mathematics, and Software: Proceedings of the International Symposium on Computational Science in Celebration of the 65th Birthday of John R. Rice, West Lafayette, Indiana, USA, 22-26 May, 1999*. Number v. 1 in Computational Science, Mathematics, and Software: Proceedings of the International Symposium on Computational Science in Celebration of the 65th Birthday of John R. Rice, West Lafayette, Indiana, USA, 22-26 May, 1999. Purdue University Press, 2002.
- [14] J. Borràs, A. Moreno, and A. Valls. Intelligent tourism recommender systems: A survey. *Expert Systems with Applications*, 41(16):7370–7389, 2014.
- [15] R. Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370, Nov 2002.
- [16] D. Buse and Q. Wu. *IP Network-based Multi-agent Systems for Industrial Automation: Information Management, Condition Monitoring and Control of Power Systems*. Power Systems. Springer London, 2007.
- [17] C. L. C. E. Calle, J. Statistical user model supported by r-tree structure. *Appl Intell*, 39:545–563, 2013.
- [18] J. Calle, L. Castaño, E. Castro, and D. Cuadra. Statistical user model supported by R-Tree structure. *Applied Intelligence*, 39(3):545–563, 2013.
- [19] F. Cena, S. Likavec, and A. Rapp. Real World User Model: Evolution of User Modeling Triggered by Advances in Wearable and Ubiquitous Computing: State of the Art and Future Directions. *Information Systems Frontiers*, 21(5):1085–1110, 2019.
- [20] S. Chaiken. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, page 752–766, 1998.
- [21] J. H. Chang, C. F. Lai, and M. S. Wang. A fair scheduler using cloud computing for digital TV program recommendation system. *Telecommunication Systems*, 60(1):55–66, 2015.

- [22] C. C. Chen, Y. H. Wan, M. C. Chung, and Y. C. Sun. An effective recommendation method for cold start new users using trust and distrust networks. *Information Sciences*, 224:19–36, 2013.
- [23] A. Cheung. *A Collaborative Interface Agent for Lotus ESuite Mail*. Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 1998.
- [24] S.-B. Cho and J.-M. Yu. Hierarchical modular bayesian networks for low-power context-aware smartphone. *Neurocomputing*, 326-327:100 – 109, 2019.
- [25] M. Colhon and A. Iftene. Recommendation Technique for the ”Cold-Start” Problem. *Romanian Journal of Human - Computer Interaction*, 9(3):255–268, 2016.
- [26] J. Cordeiro, B. Shishkov, A. Ranchordas, and M. Helfert. *Software and Data Technologies: Third International Conference, ICSOFT 2008, Porto, Portugal, July 22-24, 2008*. Communications in Computer and Information Science. Springer Berlin Heidelberg, 2009.
- [27] R. Cordier, R. Speyer, Y.-W. Chen, S. Wilkes-Gillan, T. Brown, H. Bourke-Taylor, K. Doma, and A. Leicht. Evaluating the psychometric quality of social skills measures: A systematic review. *PLOS ONE*, 10(7):1–32, 07 2015.
- [28] G. Dougherty. *Pattern Recognition and Classification: An Introduction*. SpringerLink : Bücher. Springer New York, 2012.
- [29] J. Drugowitsch. *Design and Analysis of Learning Classifier Systems: A Probabilistic Approach*. Studies in Computational Intelligence. Springer Berlin Heidelberg, 2008.
- [30] X. Du, H. Liu, and L. Jing. Additive co-clustering with social influence for recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, RecSys ’17, page 193–200, New York, NY, USA, 2017. Association for Computing Machinery.
- [31] M. Erp and L. Schomaker. Variants of the borda count method for combining ranked classifier hypotheses. 01 2000.
- [32] S. Fazeli, B. Loni, A. Bellogin, H. Drachsler, and P. Sloep. Implicit vs. explicit trust in social matrix factorization. In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys ’14, page 317–320, New York, NY, USA, 2014. Association for Computing Machinery.
- [33] J. Fraenkel and B. Grofman. The borda count and its real-world alternatives: Comparing scoring rules in nauru and slovenia. *Australian Journal of Political Science*, 49, 04 2014.

- [34] C. Fuchs. *Internet and Society: Social Theory in the Information Age*. Routledge Research in Information Technology and Society. Taylor & Francis, 2007.
- [35] e. a. Gabor T., Istvan P. Scalable collaborative filtering approaches for large recommender systems.
- [36] H. Gao, J. Tang, and H. Liu. Addressing the cold-start problem in location recommendation using geo-social correlations. *Data Mining and Knowledge Discovery*, 29(2):299–323, 2015.
- [37] F. Gedikli. *Recommender Systems and the Social Web: Leveraging Tagging Data for Recommender Systems*. SpringerLink : Bücher. Springer Fachmedien Wiesbaden, 2013.
- [38] O. Georgiou and N. Tsapatsoulis. Improving the scalability of recommender systems by clustering using genetic algorithms. In K. Diamantaras, W. Duch, and L. S. Iliadis, editors, *Artificial Neural Networks – ICANN 2010*, pages 442–449, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.
- [39] M. Gil, E. Serral, P. Valderas, and V. Pelechano. Designing for user attention: A method for supporting unobtrusive routine tasks. *Science of Computer Programming*, 78(10):1987–2008, 2013.
- [40] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins. Eigentaste: A Constant Time Collaborative Filtering Algorithm. *Information Retrieval*, 4(2):133–151, 2001.
- [41] J. Goldsmith, N. Mattei, and R. H. Sloan. Who Is Watching you eat? *AAAI Workshop - Technical Report*, WS-14-10(4):43–48, 2014.
- [42] N. Houlsby, J. M. Hernández-Lobato, and Z. Ghahramani. Cold-start active learning with robust ordinal matrix factorization. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*, ICML’14, page II–766–II–774. JMLR.org, 2014.
- [43] G. N. Hu and X. Y. Dai. Integrating reviews into personalized ranking for cold start recommendation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10235 LNAI:708–720, 2017.
- [44] L. Hu, L. Cao, J. Cao, Z. Gu, G. Xu, and D. Yang. Learning informative priors from heterogeneous domains to improve recommendation in cold-start user domains. *ACM Trans. Inf. Syst.*, 35(2), Dec. 2016.
- [45] Q. Huang. Mining online footprints to predict user’s next location. *International Journal of Geographical Information Science*, 31.3:523–541, 2017.

- [46] D. Hwang, J. Jung, and N. Nguyen. *Computational Collective Intelligence – Technologies and Applications: 6th International Conference, ICCCI 2014, Seoul, Korea, September 24-26, 2014, Proceedings*. Lecture Notes in Computer Science. Springer International Publishing, 2014.
- [47] K. Ji and H. Shen. Jointly modeling content, social network and ratings for explainable and cold-start recommendation. *Neurocomputing*, 218:1–12, 2016.
- [48] D. Jin. *Globalization and Media in the Digital Platform Age*. Taylor & Francis, 2019.
- [49] D. Jin and S. Lin. *Advances in Computer Science, Intelligent Systems and Environment*. Number v. 2 in Advances in Intelligent and Soft Computing. Springer Berlin Heidelberg, 2011.
- [50] M. Kaptein, P. Markopoulos, B. de Ruyter, and E. Aarts. Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human-Computer Studies*, 77:38 – 51, 2015.
- [51] J. A. Konstan and J. Riedl. Recommender systems: From algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22(1-2):101–123, 2012.
- [52] D. L, K. B, and O. J. *Information, Statistics And Induction In Science - Proceedings Of The Conference, Isis '96*. World Scientific Publishing Company, 1996.
- [53] J. Laskar and P. Robutel. High order symplectic integrators for perturbed hamiltonian systems, 2000.
- [54] J. Lee, M. Sun, and G. Lebanon. A comparative study of collaborative filtering algorithms. 05 2012.
- [55] C. Levinas, M. Llorente, and F. d’Informàtica de Barcelona. *An Analysis of Memory Based Collaborative Filtering Recommender Systems with Improvement Proposals*. Universitat Politècnica de Catalunya. Facultat d’Informàtica de Barcelona, 2014 (Master in Artificial Intelligence - MAI), 2014.
- [56] R. Liang, W. Guo, and D. Yang. Mining product problems from online feedback of Chinese users. *Kybernetes*, 46(3):572–586, 2017.
- [57] T. Lillegraven. Design of a Bayesian Recommender System for Tourists Presenting a Solution to the Cold-Start User Problem. *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Industry track*, (June):1681–1688, 2010.
- [58] J. Lu, D. Wu, M. Mao, and W. Wang. Recommender system application developments: A survey. *Decision Support Systems*, 74:12–32, 2015.

- [59] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang. Recommender system application developments: A survey. *Decision Support Systems*, 74:12–32, 2015.
- [60] M. Ma. *Communication Systems and Information Technology: Selected Papers from the 2011 International Conference on Electric and Electronics (EEIC 2011) in Nanchang, China on June 20-22, 2011*. Number v. 4 in Lecture Notes in Electrical Engineering. Springer Berlin Heidelberg, 2011.
- [61] E. Maskin, A. Sen, K. Arrow, P. Dasgupta, P. Pattanaik, and J. Stiglitz. *The Arrow Impossibility Theorem*. Kenneth J. Arrow Lecture Series. Columbia University Press, 2014.
- [62] A. Maslen. *Persuasive Copywriting: Using Psychology to Engage, Influence and Sell*. Kogan Page, 2015.
- [63] P. Miller. *TCP/IP: The Ultimate Protocol Guide*. Number v. 2 in TCP/IP: The Ultimate Protocol Guide : Complete 2 Volume Set. BrownWalker Press, 2009.
- [64] F. Min and W. Zhu. Cold-start recommendation through granular association rules. abs/1305.1372, 2013.
- [65] F. Min and W. Zhu. Cold-start recommendation through granular association rules. 2013.
- [66] N. Mirbakhsh and C. X. Ling. Improving top-N recommendation for cold-start users via cross-domain information. *ACM Transactions on Knowledge Discovery from Data*, 9(4):1–19, 2015.
- [67] N. Nahai. *Webs of Influence: The Psychology of Online Persuasion: The Psychology of Online Persuasion*. Pearson Education Limited, 2017.
- [68] W. Neto and A. Nowe. Insights on social recommender system. *CEUR Workshop Proceedings*, 910:33–38, 01 2012.
- [69] B. Neville, M. Fasli, and J. Pitt. Utilising social recommendation for decision-making in distributed multi-agent systems. *Expert Systems with Applications*, 42(6):2884–2906, 2015.
- [70] T. Nielsen and F. JENSEN. *Bayesian Networks and Decision Graphs*. Information Science and Statistics. Springer New York, 2007.
- [71] S. Nitzan. *Collective Preference and Choice*. Cambridge University Press, 2010.
- [72] I. Nunes and D. Jannach. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27(3-5):393–444, 2017.

- [73] P. Olla. *Mobile Health Solutions for Biomedical Applications*. Advances in healthcare information systems and informatics (AHISI) series. IGI Global, 2009.
- [74] Y. Pavlov. *Decision Control, Management, and Support in Adaptive and Complex Systems: Quantitative Models: Quantitative Models*. Premier reference source. Information Science Reference, 2013.
- [75] F. Peng, J. Lu, Y. Wang, R. Y. D. Xu, C. Ma, and J. Yang. N-dimensional Markov random field prior for cold-start recommendation. *Neurocomputing*, 191:187–199, 2016.
- [76] L. Peska and P. Vojtáš. Evaluating various implicit factors in e-commerce. *CEUR Workshop Proceedings*, 910, 01 2012.
- [77] A. Popescul, D. M. Pennock, and S. Lawrence. Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI’01, page 437–444, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.
- [78] O. Pourret, P. Naïm, and B. Marcot. *Bayesian Networks: A Practical Guide to Applications*. Statistics in Practice. Wiley, 2008.
- [79] L. Quijano-Sanchez, C. Sauer, J. A. Recio-Garcia, and B. Diaz-Agudo. Make it personal: A social explanation system applied to group recommendations. *Expert Systems with Applications*, 76:36–48, 2017.
- [80] P. Resnick and H. R. Varian. Recommender systems. *Commun. ACM*, 40:56–58, 1997.
- [81] F. Ricci, K. Bontcheva, O. Conlan, and S. Lawless. *User Modeling, Adaptation and Personalization: 23rd International Conference, UMAP 2015, Dublin, Ireland, June 29 – July 3, 2015. Proceedings*. Lecture Notes in Computer Science. Springer International Publishing, 2015.
- [82] J. Rowe. How to use heatmaps to improve your website. <https://medium.com/quicksixty/how-to-use-heatmaps-to-improve-your-website-1f844543a0f5>, 2018.
- [83] G. M. Sarnè. A novel hybrid approach improving effectiveness of recommender systems. *Journal of Intelligent Information Systems*, 44(3):397–414, jun 2015.
- [84] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, WWW ’01, pages 285–295, New York, NY, USA, 2001. ACM.
- [85] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock. CROC: A new evaluation criterion for recommender systems. *Electronic Commerce Research*, 5(1):51–74, 2005.

- [86] S. Sedhain, S. Sanner, D. Braziunas, L. Xie, and J. Christensen. Social collaborative filtering for cold-start recommendations. In *Proceedings of the 8th ACM Conference on Recommender Systems*, page 345–348, New York, NY, USA, 2014. Association for Computing Machinery.
- [87] E. Services. *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. Wiley, 2014.
- [88] A. Seth, J. Zhang, and R. Cohen. A personalized credibility model for recommending messages in social participatory media environments. *World Wide Web*, 18(1):111–137, 2015.
- [89] R. Shannon. Simulation: A survey with research suggestions. *AIIE Transactions*, 7(3):289–301, 1975.
- [90] B. Shlomo, C. Ivan, and T. Domonkos. *Collaborative Recommendations: Algorithms, Practical Challenges And Applications*. World Scientific Publishing Company, 2018.
- [91] M. Sun, F. Li, and J. Zhang. A multi-modality deep network for cold-start recommendation. *Big Data and Cognitive Computing*, 2(1):1–15, 2018.
- [92] M. Sun, F. Li, and J. Zhang. A multi-modality deep network for cold-start recommendation. *Big Data and Cognitive Computing*, 2(1.7), 2018.
- [93] S. S. Suvash S., Aditya K. Low-rank linear cold-start recommendation from social data. pages 1502–1508. AAAI Conference on Artificial Intelligence, 2017.
- [94] M. Taghavi, K. Bakhtiyari, and E. Scavino. Agent-based computational investing recommender system. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys ’13, page 455–458, New York, NY, USA, 2013. Association for Computing Machinery.
- [95] A. Uhrmacher and D. Weyns. *Multi-Agent Systems: Simulation and Applications*. Computational Analysis, Synthesis, and Design of Dynamic Systems. CRC Press, 2009.
- [96] J. Urner, D. Bucher, J. Yang, and D. Jonietz. Assessing the influence of spatio-temporal context for next place prediction using different machine learning approaches. *ISPRS International Journal of Geo-Information*, 7(5), 2018.
- [97] B. Veloso, B. Malheiro, and J. C. Burguillo. A Multi-Agent Brokerage Platform for Media Content Recommendation. *International Journal of Applied Mathematics and Computer Science*, 25(3):513–527, 2015.

- [98] X. Wang, Z. Peng, S. Wang, P. S. Yu, W. Fu, and X. Hong. Cross-domain recommendation for cold-start users via neighborhood based feature mapping. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10827 LNCS:158–165, 2018.
- [99] X. Wang, Z. Peng, S. Wang, P. S. Yu, W. Fu, and X. Hong. Cross-domain recommendation for cold-start users via neighborhood based feature mapping, 2018.
- [100] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang. Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Systems with Applications*, 69:1339–1351, 2017.
- [101] I. Witten, E. Frank, and M. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science, 2011.
- [102] C. Yang, Y. Zhou, L. Chen, X. Zhang, and D. M. Chiu. Social-group-based ranking algorithms for cold-start video recommendation. *International Journal of Data Science and Analytics*, 1(3-4):165–175, 2016.
- [103] W. X. Zhao, S. Li, Y. He, E. Y. Chang, J. R. Wen, and X. Li. Connecting Social Media to E-Commerce: Cold-Start Product Recommendation Using Microblogging Information. *IEEE Transactions on Knowledge and Data Engineering*, 28(5):1147–1159, 2016.
- [104] X. Zhao and P. I. Frazier. A Markov Decision Process Analysis of the Cold Start Problem in Bayesian Information Filtering. pages 1–12, 2014.
- [105] X. Zhao and P. I. Frazier. A markov decision process analysis of the cold start problem in bayesian information filtering. *Computing Research Repository*, abs/1410.7852, 2014.
- [106] X. Zhao and J. Wang. A theoretical analysis of two-stage recommendation for cold-start collaborative filtering. *ICTIR 2015 - Proceedings of the 2015 ACM SIGIR International Conference on the Theory of Information Retrieval*, pages 71–80, 2015.
- [107] N. Zheng and J. Xue. *Statistical Learning and Pattern Analysis for Image and Video Processing*. Advances in Computer Vision and Pattern Recognition. Springer London, 2009.
- [108] J. Zhu, J. Zhang, C. Zhang, Q. Wu, Y. Jia, B. Zhou, and P. S. Yu. CHRS: Cold Start Recommendation Across Multiple Heterogeneous Information Networks. *IEEE Access*, 5:15283–15299, 2017.

- [109] J. Zhu, J. Zhang, C. Zhang, Q. Wu, Y. Jia, B. Zhou, and P. S. Yu. Chrs: Cold start recommendation across multiple heterogeneous information networks. *IEEE Access*, 5:15283–15299, 2017.