

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

GABRIEL MACHADO LUNARDI

**Diversificação multi-atributo para diminuir
o efeito “bolha de filtro” em sistemas de
recomendação de notícias**

Tese apresentada como requisito parcial para
a obtenção do grau de Doutor em Ciência da
Computação

Orientador: Prof. Dr. José Palazzo M. de Oliveira

Porto Alegre
2021

CIP — CATALOGAÇÃO NA PUBLICAÇÃO

Lunardi, Gabriel Machado

Diversificação multi-atributo para diminuir o efeito “bolha de filtro” em sistemas de recomendação de notícias / Gabriel Machado Lunardi. – Porto Alegre: PPGC da UFRGS, 2021.

133 f.: il.

Tese (doutorado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2021. Orientador: José Palazzo M. de Oliveira.

1. Diversificação de recomendação. 2. Sistemas de recomendação. 3. Bolhas de filtro. 4. Notícias. I. José Palazzo M. de Oliveira, . II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos André Bulhões

Vice-Reitora: Prof^ª. Patricia Pranke

Pró-Reitor de Pós-Graduação: Prof. Celso Giannetti Loureiro Chaves

Diretora do Instituto de Informática: Prof^ª. Carla Maria Dal Sasso Freitas

Coordenador do PPGC: Prof. Claudio Rosito Jung

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

RESUMO

Os Sistemas de Recomendação nasceram com o propósito de auxiliar as pessoas a encontrarem itens que satisfaçam suas preferências nos mais variados domínios, como músicas, notícias, vídeos, produtos, dentre outros. Entretanto, as abordagens de recomendação têm sido alvo de críticas ao longo do tempo por entregarem itens que são, muitas vezes, óbvios e redundantes, privando o usuário de outras possibilidades. Isso abre precedentes para o efeito da “bolha de filtro”, um termo cunhado em 2011, por Eli Pariser, que remete ao isolamento do usuário em relação a uma diversidade de conteúdos, o que limita suas possibilidades de novas experiências. Tal efeito se torna ainda mais crítico no domínio de notícias, uma vez que são fontes de informação e de formação de opinião, porque o usuário estando em uma “bolha” pode ficar sujeito à polarização de opiniões, à desinformação sobre diferentes assuntos, dentre outros malefícios. Nesse sentido, a diversificação de recomendação surge como uma estratégia para suavizar tal efeito, buscando entregar itens novos e diferentes aos usuários. Nesta tese, é apresentada uma abordagem para a diversificação de recomendação, baseada em múltiplos atributos, aplicada ao domínio de notícias eletrônicas. Além disso, é apresentada uma forma de medir o efeito de “bolhas de filtro” através do grau de homogeneidade dos itens consumidos e/ou entregues aos usuários. A avaliação compreende um estudo de usuário, utilizando um protótipo de recomendação de notícias sobre política. A partir desse, busca-se verificar se a estratégia de diversificação multi-atributo é capaz de suavizar o efeito de “bolha de filtro” quando comparada à recomendação convencional, empregando a metodologia de teste A/B e testes de diferença estatística entre os grupos de usuários. Os resultados para essa avaliação sugerem uma diminuição da homogeneidade para todos os casos, considerando apenas os itens entregues, isto é, sem qualquer tipo de interação. Além disso, a abordagem de diversificação multi-atributo foi comparada com uma abordagem de diversificação da literatura. Os resultados sugerem melhorias para alguns casos.

Palavras-chave: Diversificação de recomendação. Sistemas de recomendação. Bolhas de filtro. Notícias.

Multi-attribute diversification to diminish the “filter bubble” effect in news recommender systems

ABSTRACT

Recommender Systems were born with the purpose of helping people find items that satisfy their preferences in the most varied domains, such as music, news, videos, products, among others. However, recommendation approaches have been criticized over time for delivering items that are often obvious and redundant, depriving the user of other possibilities. This sets the stage for the “filter bubble” effect, a term coined in 2011 by Eli Pariser, which refers to the user’s isolation from a diversity of content, limiting their possibilities for new experiences. This effect becomes even more critical in the field of news, since they are sources of information and opinion formation, because the user being in a “bubble” may be subject to the polarization of opinions, misinformation on different subjects, among other evils. In this sense, recommendation diversification emerges as a strategy to soften this effect, seeking to deliver new and different items to users. In this thesis, an approach for recommendation diversification, based on multiple attributes, applied to the electronic news domain, is presented. In addition, a way to measure the effect of “filter bubbles” through the degree of homogeneity of items consumed and/or delivered to users is presented. The evaluation comprises a user study, using a prototype for recommending news about politics. From this, we seek to verify whether the multi-attribute diversification strategy is able to soften the filter bubble effect when compared to the conventional recommendation, using the A/B test methodology and statistical difference tests between user groups. The results for this evaluation suggest a decrease in homogeneity for all cases considering only the delivered items, i.e., without any kind of interaction. In addition, the multi-attribute diversification approach was compared with a diversification approach from the literature. The results suggest improvements for some cases.

Keywords: Diversification of recommendation. Recommender systems. Filter bubbles. News.

LISTA DE ABREVIATURAS E SIGLAS

ALS	<i>ALS - Alternating Least Squares</i>
ES	Esparsidade
GH	Grau de Homogeneidade
HTML	<i>HyperText Markup Language</i>
ILD	<i>Intra List Diversity</i>
KNN	<i>k-nearest neighbors</i>
LDA	<i>Latent Dirichlet Allocation</i>
MMR	<i>Maximal Marginal Relevance</i>
MF	<i>Matrix Factorization</i>
NER	<i>Named Entity Recognizer</i>
NRS	<i>News Recommender Systems</i>
PLN	Processamento de Linguagem Natural
PCA	<i>Principal Component Analysis</i>
RMSE	<i>Root Mean Squared Error</i>
REN	Reconhecimento de Entidades Nomeadas
SBRS	<i>Session-based Recommender Systems</i>
SGD	<i>SGD - Stochastic Gradient Descent</i>
SVD	<i>Single Value Decomposition</i>
SQL	<i>Standard Query Language</i>
SR	Sistema(s) de Recomendação
TD	<i>Topic Diversification</i>
URL	<i>Uniform Resource Locator</i>

LISTA DE FIGURAS

Figura 2.1	Processo genérico de recomendação.	15
Figura 2.2	Exemplo de fatores latentes.	26
Figura 2.3	Exemplo de fatoração da matriz de utilidade utilizando o algoritmo SVD... ..	27
Figura 2.4	Níveis de diversidade.	42
Figura 3.1	Esquema de revisão.	49
Figura 4.1	Trecho de duas notícias falsas e antagônicas sobre um mesmo assunto.	59
Figura 4.2	Arquitetura da abordagem de diversificação multi-atributo.	61
Figura 4.3	Exemplo de funcionamento do algoritmo LDA.	64
Figura 4.4	Valores de coerência do modelo para diferentes números de tópicos	66
Figura 4.5	Mapa de distância inter-tópico via escala multi-dimensional.	67
Figura 4.6	REN nas manchetes das duas notícias falsas motivacionais deste capítulo... ..	69
Figura 4.7	Gráfico de Pareto	70
Figura 4.8	As vinte entidades mais citadas no <i>corpus</i> de notícias.	71
Figura 4.9	Variabilidade dos valores de polaridade para cada um dos portais.	72
Figura 4.10	Variabilidade dos valores de subjetividade para cada um dos portais.	72
Figura 5.1	Primeira etapa do experimento.	76
Figura 5.2	Tela de feedback do processamento da recomendação.	77
Figura 5.3	Segunda etapa do experimento.	78
Figura 5.4	Interface móvel para a entrega das recomendações.	80
Figura 5.5	Tela de detalhe de uma notícia.	81
Figura 5.6	Tela das Notícias avaliadas.	82
Figura 5.8	Modelo lógico do banco de dados.	82
Figura 5.7	Tela com os resultados de interação.	83
Figura 5.9	Distribuição de cauda longa para o <i>dataset</i> de recomendações.	86
Figura 5.10	Acessos por país.	87
Figura 6.1	Relacionamento entre a homogeneidade (GH) e acurácia (inverso do RMSE).	96
Figura A.1	Os 30 termos mais relevantes para o tópico 0	123
Figura A.2	Os 30 termos mais relevantes para o tópico 1	123
Figura A.3	Os 30 termos mais relevantes para o tópico 2	124
Figura A.4	Os 30 termos mais relevantes para o tópico 3	124
Figura A.5	Os 30 termos mais relevantes para o tópico 4	125
Figura A.6	Os 30 termos mais relevantes para o tópico 5	125
Figura A.7	Os 30 termos mais relevantes para o tópico 6	126
Figura A.8	Os 30 termos mais relevantes para o tópico 7	126
Figura A.9	Os 30 termos mais relevantes para o tópico 8	127
Figura A.10	Os 30 termos mais relevantes para o tópico 9	127
Figura A.11	Os 30 termos mais relevantes para o tópico 10	128
Figura A.12	Os 30 termos mais relevantes para o tópico 11	128
Figura A.13	Os 30 termos mais relevantes para o tópico 12	129
Figura A.14	Os 30 termos mais relevantes para o tópico 13	129
Figura A.15	Os 30 termos mais relevantes para o tópico 14	130
Figura A.16	Os 30 termos mais relevantes para o tópico 15	130

LISTA DE TABELAS

Tabela 2.1	Base de dados de livros hipotética.....	16
Tabela 2.2	Matriz usuário <i>versus</i> item	20
Tabela 3.1	Comparação entre os trabalhos relacionados.	56
Tabela 4.1	Ranking de portais de notícias brasileiros.	62
Tabela 4.2	Quantidade de notícias coletadas de cada portal.	63
Tabela 4.3	Exemplo da Matriz de notícias <i>versus</i> tópicos.	67
Tabela 5.1	Algoritmos de recomendação e de diversificação implementados.....	76
Tabela 5.2	Característica geral do <i>dataset</i> de avaliações	85
Tabela 5.3	Característica do <i>dataset</i> de recomendações	86
Tabela 6.1	Estatísticas descritivas da variável GH, considerando itens recomenda- dos e interagidos, para cada um dos algoritmos.	93
Tabela 6.2	Comparação entre grupos da variável GH, considerando itens recomen- dados e interagidos.....	94
Tabela 6.3	Estatísticas descritivas do RMSE para cada um dos algoritmos.	95
Tabela 6.4	Comparação entre grupos da variável RMSE.....	96
Tabela 6.5	Resultados de interação com notícias falsas. R&L representa a porcen- tagem de notícias lidas que também receberam um “ <i>Like</i> ”, enquanto R&D é a porcentagem de notícias lidas que também receberam um “ <i>Dislike</i> ”.....	97
Tabela 6.6	Estatísticas descritivas da variável GH, considerando listas de recomen- dação, para cada um dos algoritmos.	101
Tabela 6.7	Comparação entre grupos da variável GH, considerando listas de reco- mendação.	102
Tabela 6.8	Estatísticas descritivas da variável GH, considerando listas de recomen- dação, para a comparação entre abordagens de diversificação multi-atributo e abordagens de diversificação <i>baseline</i>	103
Tabela 6.9	Comparação entre abordagens de diversificação multi-atributo e abor- dagens de diversificação <i>baseline</i>	104

SUMÁRIO

1 INTRODUÇÃO	10
1.1 Questão de Pesquisa.....	11
1.2 Objetivos	11
1.3 Método de Pesquisa.....	12
1.4 Organização do texto	13
2 FUNDAMENTAÇÃO CONCEITUAL	14
2.1 Sistemas de recomendação	14
2.1.1 Recomendação baseada em conteúdo.....	16
2.1.2 Recomendação baseada em filtragem colaborativa.....	19
2.1.3 Vantagens e desvantagens da filtragem colaborativa e da filtragem baseada em conteúdo	27
2.1.4 Recomendação híbrida.....	28
2.1.5 Principais Desafios e Oportunidades	29
2.1.6 Tipos de Experimentos.....	32
2.1.7 Métricas de Avaliação	35
2.1.8 Recomendação de notícias	36
2.2 Bolhas de filtro	39
2.3 Diversificação de recomendação	41
2.3.1 Máxima relevância marginal.....	42
2.3.2 Diversificação de tópicos	43
2.4 Análise de sentimentos	44
2.5 Reconhecimento de entidades nomeadas	46
2.6 Resumo do capítulo	47
3 TRABALHOS RELACIONADOS	48
3.1 Resumo do capítulo	57
4 DIVERSIFICAÇÃO MULTI-ATRIBUTO DE RECOMENDAÇÃO	58
4.1 Modelagem e implementação da abordagem para a diversificação multi- atributo de recomendação	59
4.1.1 Obtenção de notícias	61
4.1.2 Tópicos.....	63
4.1.3 Entidades nomeadas.....	68
4.1.4 Polaridade e subjetividade	70
4.2 Medição de bolha de filtro	73
4.3 Resumo do capítulo	73
5 NEWSREC: WEB APP PARA A RECOMENDAÇÃO DE NOTÍCIAS	75
5.1 Banco de dados	81
5.2 Coleta de dados	84
5.3 Resumo do capítulo	87
6 AVALIAÇÃO EXPERIMENTAL E RESULTADOS	88
6.1 Estudo de Usuário	88
6.2 Protocolo Experimental	89
6.2.1 Interação de usuários.....	90
6.2.2 Listas de recomendação	91
6.3 Resultados envolvendo interação de usuários	93
6.3.1 Formação de Bolha de Filtro.....	93
6.3.2 Interação com notícias falsas	97
6.3.3 Discussão	98

6.4 Resultados envolvendo listas de recomendação	100
6.4.1 Formação de Bolha de Filtro.....	101
6.4.2 Discussão	104
6.5 Resumo do Capítulo.....	105
7 CONCLUSÃO	106
7.1 Contribuições.....	107
7.2 Limitações, dificuldades e lições aprendidas	111
7.3 Trabalhos Futuros.....	112
REFERÊNCIAS.....	114
APÊNDICE A — PALAVRAS-CHAVE DO MODELO DE TÓPICOS LDA.....	123
APÊNDICE B — ENTIDADES NOMEADAS SELECIONADAS COMO PARTE DAS CARACTERÍSTICAS PARA A DIVERSIFICAÇÃO.....	131

1 INTRODUÇÃO

Sistemas de Recomendação (SR) auxiliam as pessoas a encontrarem itens que satisfaçam suas preferências (JANNACH et al., 2010). Atualmente, com um ambiente *web* cada vez mais veloz na geração e na disseminação de informação, o emprego desses sistemas traz vários benefícios como a diminuição da sobrecarga de informação sobre o usuário. Outras vantagens podem ser citadas: o suporte à tomada de decisão, a fidelização de clientes, o aumento de receita, dentre outras (JANNACH; ADOMAVICIUS, 2016).

Entretanto, as abordagens de recomendação têm acumulado críticas ao longo do tempo devido à sua ênfase demasiada em similaridade, produzindo recomendações que são muitas vezes óbvias e redundantes. Em uma plataforma de música, por exemplo, recomendar outras músicas do mesmo artista que o usuário está acostumado pode ser muito assertivo, porém de menor utilidade do que encontrar músicas de outros artistas para o usuário experimentar (KAPOOR et al., 2015).

Essa característica dos SR pode levar o usuário a experimentar, sem que perceba, o efeito de “bolha de filtro”. Bolhas de filtro são, segundo Pariser (2011) “universos invisíveis e pessoais de informações que podem prender os usuários a um paradoxo de relevância, confinando-os a vizinhanças de informações isoladas, impedindo-os de ver ou explorar uma vasta gama de outras possibilidades”. Dessa forma, as pessoas tendem a receber pontos de vista enviesados e, também, limitados a não experimentarem algo novo que possa estimular seu desenvolvimento.

Esse efeito é especialmente crítico no domínio de notícias já que elas são importantes fontes de informação e formação de opiniões. Da mesma forma que no exemplo anterior, no domínio de música, o *feed* de notícias de um usuário pode estar repleto de artigos sobre um artista popular e ignorar completamente artigos de outros tópicos como, por exemplo, eventos relacionados a outros artistas tal como o *Grammy*, dentre outros. Em uma situação ideal, ambas as notícias deveriam aparecer para o usuário (PARISER, 2011).

Entretanto, a recomendação no domínio de notícias envolve, ainda, outras dimensões, conforme aponta Karimi, Jannach e Jugovac (2018), como a popularidade de uma notícia (mais compartilhadas e mais comentadas), ou o quão recente ela é, o que pode corroborar ainda mais para a formação de bolhas de filtro que, por sua vez, contribuem para a polarização de opiniões e a disseminação de notícias falsas.

Na intenção de “romper” a bolha de filtro, autores como Pariser (2011) e Helber-

ger, Karppinen e D’Acunto (2018) sugerem que evidenciar o “diferente” pode ser uma solução, especialmente em relação a itens novos e que promovam uma surpresa boa ao usuário. A diversificação de recomendação surge, então, como factibilidade de solução já que visa a sugestão de novidades para o usuário. Para isso, entretanto, a diversificação no domínio de notícias envolve alguns desafios:

- O clássico dilema entre acurácia e diversidade. O aumento de um implica no decréscimo do outro (RICCI et al., 2015);
- Tornar perceptível, ao usuário, o efeito de bolha de forma que isso funcione como um fator de decisão para este escolher o quanto deseja ou não diversificar suas recomendações;
- Notícias são comumente expressas na forma de texto, um formato de dados desestruturado.

Ainda, as notícias costumam ser caracterizadas/organizadas por, pelo menos, um tópico/assunto. Utilizando apenas isso para caracterizá-las pode não ser suficiente, especialmente se estiverem sendo comparadas por algum método computacional como é necessário durante um processo de diversificação de itens. Por esse motivo, é essencial que o conteúdo textual seja analisado, buscando características que sejam realmente disruptivas. A partir disso, neste trabalho busca-se uma representação multi-atributo para os itens. Sendo assim, um algoritmo de diversificação seria capaz de apresentar notícias de diferentes formas.

1.1 Questão de Pesquisa

A utilização de múltiplos atributos em um algoritmo de diversificação de recomendação contribui para a diminuição do efeito de bolha de filtro?

1.2 Objetivos

Esta tese tem por **objetivo** desenvolver uma abordagem multi-atributo para a diversificação de recomendação de notícias, buscando diminuir o efeito de bolha de filtro.

Os **objetivos específicos** consistem em: introduzir uma métrica, junto de uma metodologia, para a mensuração do efeito de bolha de filtro em SR; e avaliar e validar tal

metodologia por meio da análise de experimentos.

1.3 Método de Pesquisa

Para responder à questão de pesquisa levantada e atingir os objetivos apresentados, foi necessário perpassar por um conjunto de etapas iterativas e interativas, ou seja, muitas vezes foi necessário dar um passo atrás, repensar, reimplementar, e agir novamente. Sendo assim, primeiramente, foi realizado um levantamento sistemático da literatura, visando identificar as oportunidades de pesquisa envolvendo diversificação de recomendação e bolhas de filtro, especialmente no domínio de notícias.

Após ampla pesquisa e estudo dos trabalhos obtidos, obteve-se um panorama das abordagens em relação ao tema proposto. Com isso, foram identificados os algoritmos, as ferramentas e os problemas em aberto. Desses, optou-se por aprofundar o problema da representação dos itens, pois, partindo do princípio de que todo algoritmo de diversificação de recomendação, baseado em pós-filtragem, opere adequadamente, pressupõe-se que o maior problema esteja na identificação de características que melhor representem os itens. Em outras palavras, se forem utilizadas características pouco representativas, o algoritmo possivelmente apresentará diversificações de pouco impacto, especialmente em relação à suavização do efeito de bolha de filtro.

Depois de definidos o problema e o escopo de pesquisa, foram estudadas diferentes possibilidades de representação de notícias, sendo, a primeira delas “tópicos”. A partir daí, vislumbrou-se a oportunidade de experimentar diferentes algoritmos de diversificação combinados com tópicos como características em um estudo de usuário. Esse estudo foi conduzido através de uma aplicação *web* de recomendação diversificada de notícias sobre as eleições presidenciais de 2018, cuja principal motivação foi a polarização e a disseminação de conteúdo duvidoso e controverso frente às notícias em redes sociais.

Posteriormente à utilização do sistema pelos usuários, e após mais estudos sobre representação de itens, foi concebida a representação multi-atributo composta de: entidades nomeadas, polaridade e subjetividade. Essa representação foi combinada com os algoritmos diversificadores de recomendação dando origem à abordagem de diversificação multi-atributo, que foi testada sobre os dados coletados com o estudo de usuário.

Por fim, foram analisados os resultados, empregando testes de significância estatística. A seguir, é apresentada a organização do texto desta tese.

1.4 Organização do texto

O restante deste texto está organizado da seguinte forma: O Capítulo 2 apresenta a fundamentação conceitual, compreendendo as áreas de Sistemas de Recomendação, Bolhas de Filtro, Diversificação de Recomendação, Análise de Sentimentos, e Reconhecimento de Entidades Nomeadas. O Capítulo 3 apresenta os trabalhos relacionados, evidenciando a oportunidade de pesquisa identificada e explorada nesta tese. O Capítulo 4 apresenta a modelagem, a definição e a implementação da abordagem desenvolvida para diversificar a recomendação de notícias com base em múltiplos atributos e, também, para a medição de bolhas de filtro. O Capítulo 5 apresenta a aplicação *web NewsREC* desenvolvida para recomendar notícias diversificadas e não diversificadas no intuito de coletar dados e com isso possibilitar a avaliação da abordagem de diversificação desenvolvida nesta tese. O Capítulo 6 apresenta a avaliação experimental da abordagem, considerando um estudo de usuário envolvendo a aplicação *NewsREC* cujos dados permitiram avaliar a abordagem através de testes de significância estatística. Por fim, o Capítulo 7 apresenta a conclusão, apontando para os próximos passos da pesquisa bem como pontuando os resultados obtidos e demais atividades realizadas durante o período de doutorado.

2 FUNDAMENTAÇÃO CONCEITUAL

A abordagem proposta nesta tese está diretamente relacionada às seguintes áreas de estudo: Sistemas de Recomendação e o domínio de itens de notícias, Bolhas de Filtro, análise de sentimentos e reconhecimento de entidades nomeadas. Por esse motivo, este capítulo apresenta os conceitos sobre cada uma delas, com ênfase nos elementos empregados ao longo da pesquisa, tal como formas de avaliação, métricas e algoritmos.

2.1 Sistemas de recomendação

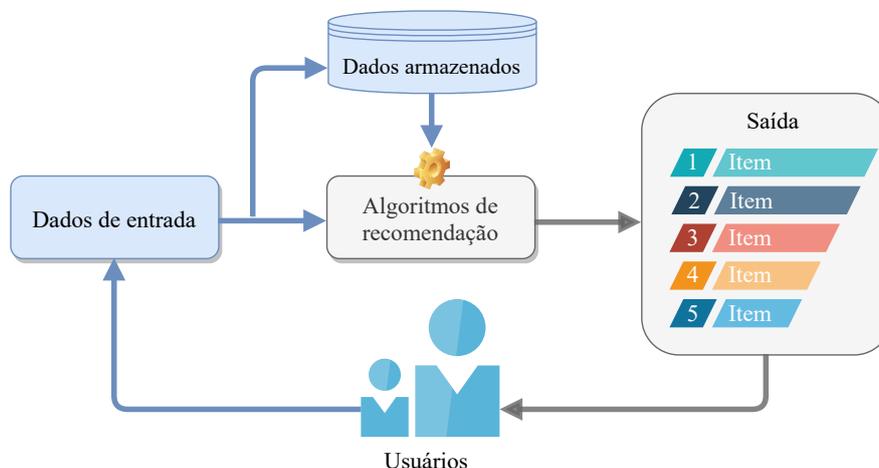
Os Sistemas de Recomendação (SR) configuram-se como uma técnica de filtragem de informação, utilizada para predizer os itens com os quais um usuário poderia demonstrar interesse (BOBADILLA et al., 2013). Grandes empresas investem no desenvolvimento desse tipo de sistema como a *Amazon* para a recomendação de livros, roupas e outros produtos; a *Netflix* para a recomendação de filmes e séries; o *YouTube* para a recomendação de vídeos; o *Facebook* para a recomendação de conteúdos diversos; o *BBC News* como representante dos portais que realizam recomendação de notícias.

Os SR exploram as possíveis relações existentes entre: os usuários e os itens; os usuários e outros usuários; ou os itens e outros itens. Com isso, o sistema busca as melhores recomendações de itens aos usuários. Uma determinada recomendação é dita relevante se o item oferecido ao usuário possuir alta probabilidade de ser utilizado. Por outro lado, para que sejam calculadas as recomendações, é essencial conhecer o usuário, no intuito de que as recomendações sejam relevantes para esse. As informações sobre o usuário são coletadas e, a partir delas, inicia-se o processo de recomendação (HERLOCKER et al., 2004; RICCI et al., 2015).

Burke (2002) descreve o processo de recomendação de forma genérica assim como os componentes envolvidos. A Figura 2.1 apresenta a visão geral desse processo. “Dados de entrada” correspondem a toda e qualquer informação fornecida pelo usuário para que o processo de recomendação possa ser iniciado; “Dados armazenados” representam o conjunto de itens a serem recomendados como, por exemplo, produtos, documentos, vídeos, filmes, notícias, etc. Além disso, considera-se o perfil do usuário como dado armazenado já que ele é essencial para o cálculo da recomendação; “Algoritmo de recomendação” representa o algoritmo responsável pelo cálculo da recomendação combinando os dados de entrada e os dados armazenados; por fim, a “Saída” é representada

como uma lista contendo os itens que o usuário possivelmente mais demonstraria interesse.

Figura 2.1: Processo genérico de recomendação.



Fonte: Adaptado de Burke (2002).

Burke (2002) destaca que os dados mais importantes nos SR são os dados de entrada e a saída obtida. De forma geral, quanto maior for a quantidade de dados de entrada, melhor será a recomendação. O perfil de cada usuário é preenchido a partir dos dados de entrada. Esses dados costumam ser coletados de forma explícita e implícita. De forma explícita, o usuário é incentivado a manualmente especificar as informações que descrevem os seus interesses, por meio de avaliações sobre os produtos. Por outro lado, com a forma de coleta implícita, o sistema infere os interesses do usuário analisando as interações usuário-itens. As informações são armazenadas em um histórico sempre que o usuário avalia um item, realiza uma consulta, uma compra ou efetua a leitura de algum conteúdo (BOBADILLA et al., 2013). O perfil do usuário costuma ser representado através de: vetores de características, modelos baseados em ontologia, matrizes de avaliações, dentre outros. Cada uma dessas estratégias de representação é empregada de acordo com o domínio de aplicação e os requisitos do sistema (Adomavicius; Tuzhilin, 2005).

Os algoritmos de recomendação distribuem-se em três classificações são mais proeminentes na literatura, a recomendação baseada em conteúdo, a recomendação baseada em filtragem colaborativa e a recomendação híbrida (JANNACH et al., 2010; RICCI et al., 2015). Cabe destacar que as abordagens serão aqui descritas em sua forma pura de concepção já que, atualmente, os sistemas não costumam empregar somente uma ou outra estratégia isoladamente mas, sim, uma combinação de estratégias a fim de minimizar as desvantagens de uma com as vantagens de outra.

2.1.1 Recomendação baseada em conteúdo

A recomendação baseada em conteúdo entrega itens similares àqueles que um usuário consumiu no passado. Para isso, é preciso que os itens possuam uma caracterização, isto é, atributos. Dessa forma, o sistema consegue estimar a utilidade de um item a um usuário. Por exemplo, para recomendar um livro um sistema buscaria medir a semelhança dele com os demais livros que o usuário avaliou no passado através de vários atributos como: gênero, autor(es), ano, dentre outros (Adomavicius; Tuzhilin, 2005).

A Tabela 2.1 apresenta um conjunto de livros armazenados hipoteticamente em um banco de dados relacional ligado a um sistema de recomendação de livros. Observe-se que esses itens são descritos pelo título, gênero, autor, tipo, preço e palavras-chave (derivadas automaticamente a partir dos textos dos livros). As preferências dos usuários são expressas, então, por esse mesmo conjunto de atributos. Para exemplificar, Maria é um usuário do sistema cujas preferências podem ser capturadas de diferentes maneiras. A maneira direta é solicitar explicitamente à Maria uma faixa de preço ou um conjunto de gêneros preferidos. A outra maneira é pedir à ela que avalie um conjunto de livros. Enquanto isso, o sistema pode derivar automaticamente um conjunto de palavras-chave daqueles livros que Maria gostou e, com isso, estimar as preferências com maior precisão (JANNACH et al., 2010).

Tabela 2.1: Base de dados de livros hipotética.

Título	Gênero	Autor	Tipo	Preço	keywords
Harry Potter	Aventura, fantasia	<i>J. K Rowling</i>	<i>e-book</i>	R\$ 50,00	aventura, magia, ação ...
O Código Da Vinci	Romance policial, suspense	Dan Brown	<i>hardcover</i>	R\$ 37,00	igreja, suspense, mistério ...
O Diário de Anne Frank	Drama bélico biográfico	Anne Frank	<i>e-book</i>	R\$ 41,00	biografia, Alemanha, Polônia, ...
...

Fonte: adaptado de Jannach et al. (2010).

Para gerar as recomendações, os sistemas baseados em conteúdo normalmente funcionam avaliando o potencial de semelhança que um item ainda não avaliado em re-

lação aos itens que o usuário ativo gostou no passado. A similaridade pode ser medida de diferentes maneiras. Dado um livro B ainda não avaliado, o sistema poderia verificar se o gênero do livro em questão está na lista de gêneros preferidos de Maria. A semelhança nesse caso é 0 ou 1. Outra opção é calcular a semelhança ou a sobreposição das palavras-chave envolvidas. Como uma métrica de similaridade típica que é adequada para características multi-valoradas, poder-se-ia adotar o coeficiente de *Dice* da seguinte forma: Se cada livro B_i é descrito por um conjunto de palavras-chave $keywords(B_i)$, o coeficiente de *Dice* mede a similaridade entre os livros b_i e b_j (JANNACH et al., 2010).

$$\frac{2 \times |keywords(b_i) \cap keywords(b_j)|}{|keywords(b_i)| + |keywords(b_j)|} \quad (2.1)$$

Entretanto, os atributos como autor ou gênero apresentados no exemplo anterior não representam o conteúdo de cada item, mas sim agregam conhecimento adicional a esses. Sendo assim, os sistemas baseados em conteúdo foram historicamente concebidos para filtrar itens baseados em texto como, por exemplo, e-mails e notícias. Dessa forma, essa abordagem centra-se em utilizar uma lista das palavras que aparecem dentro de um documento de texto, preferencialmente extraídas de forma automática (JANNACH et al., 2010). Nesse sentido, a seguir, são apresentadas algumas formas de representação de conteúdo de documentos.

2.1.1.1 Técnicas para a representação de conteúdo

Existem diferentes abordagens para a representação de documentos. Uma delas compreende criar uma lista de todas as palavras que aparecem em todos os documentos e descrever cada documento por um vetor binário, onde um 1 indica que uma palavra aparece em um documento e um 0 que a palavra não aparece. Se o perfil do usuário for descrito por uma lista semelhante (1 denotando interesse em uma palavra-chave), a correspondência de documento pode ser feita medindo a sobreposição de interesse e conteúdo do documento. A simplicidade dessa abordagem implica em alguns problemas, pois considera que todas as palavras possuem a mesma importância em um documento, embora pareça intuitivo que uma palavra mais frequente seja mais adequada para caracterizar o documento. Além disso, uma sobreposição maior do perfil do usuário e um documento serão encontrados quando os documentos forem mais longos. Como resultado, o recomendador tenderá a propor documentos longos (JANNACH et al., 2010).

Para solucionar esse problema da abordagem Booleana, surgiu o modelo TF-IDF

(*Term Frequency-Inverse Document Frequency*). TF-IDF é uma técnica muito conhecida na área de recuperação de informações e significa, respectivamente, **frequência do termo** e **inverso da frequência de termo**. Documentos de texto podem ser codificados em TF-IDF como vetores em um espaço euclidiano multidimensional. As dimensões do espaço correspondem às palavras-chave (também chamadas de termos) que aparecem nos documentos. As coordenadas de um determinado documento em cada dimensão (ou seja, para cada termo) são calculadas como um produto de duas submedidas: frequência do termo e frequência inversa do documento (JANNACH et al., 2010).

A **frequência do termo** descreve a frequência com que um determinado termo aparece em um documento (assumindo que palavras importantes aparecem com mais frequência). Para levar em consideração o comprimento do documento e evitar que documentos mais longos obtenham um peso de relevância maior, deve-se normalizar o comprimento do documento. Uma estratégia simples nesse sentido é relacionar o número de ocorrências de termos com a frequência máxima das outras palavras-chave do documento (JANNACH et al., 2010).

A **frequência inversa do documento** é combinada com a frequência do termo. O objetivo dessa métrica é reduzir o peso das palavras-chave que aparecem com frequência em todos os documentos. A ideia é que essas palavras frequentes não são muito úteis para discriminar um documento do outro e, portanto, deve-se dar mais peso às palavras que aparecem em apenas alguns documentos (JANNACH et al., 2010).

Uma vez definida a forma de representação de conteúdo dos itens, parte-se para o processo de recomendação. A seção seguinte apresenta a técnica mais comum de recomendação que se baseiam na representação de documentos espaço-vetorial.

2.1.1.2 *Técnicas para a recomendação de conteúdo*

Uma das técnicas para estimar o interesse que um determinado documento terá para um usuário consiste em verificar se esse gostou de documentos semelhantes no passado. Para tanto, duas informações são necessárias. A primeira delas é o histórico de avaliações do tipo “gostei/não gostei” fornecidas pelo usuário sobre os itens. A segunda, é uma medida que capture a semelhança de dois documentos (JANNACH et al., 2010).

A previsão para um item i ainda não avaliado é baseada nos k itens mais similares para os quais existe uma avaliação. Por exemplo, quatro de $k = 5$ dos itens mais similares foram avaliados pelo usuário-alvo (a quem se quer recomendar), o sistema estimará que a chance de i ser avaliado positivamente será relativamente alta. Além de variar o tama-

nho da vizinhança k , outras variações são possíveis, como binarização de classificações, usando um limite mínimo de similaridade ou ponderação dos votos com base no grau de similaridade. Essa técnica de recomendação é conhecida como kNN (*k-nearest-neighbors* - k vizinhos mais próximos) (JANNACH et al., 2010; RICCI et al., 2015).

Outra técnica de recomendação baseada no modelo de representação de documentos espaço-vetorial é o *feedback* de relevância. Essa técnica surgiu no contexto de um sistema pioneiro de recuperação de informações chamado de *SMART*. A ideia principal consistia em dividir os documentos já avaliados em dois grupos, D_+ e D_- , de documentos avaliados positivamente (interessantes/relevantes) e avaliados negativamente, e calcular um vetor protótipo (ou média) para esses dois grupos. Este vetor também pode ser visto como uma espécie de centroide de um *cluster* para conjuntos de documentos relevantes e não relevantes. A consulta atual Q_i , que é representada como um vetor de termos multidimensional assim como os documentos, é repetidamente refinada para Q_{i+1} por uma adição ponderada do vetor protótipo dos documentos relevantes e a subtração ponderada do vetor que representa os documentos não relevantes. Como efeito, o vetor de consulta deve se mover consistentemente em direção ao conjunto de documentos relevantes (JANNACH et al., 2010).

2.1.2 Recomendação baseada em filtragem colaborativa

A recomendação baseada em filtragem colaborativa foi desenvolvida a partir da constatação de que as pessoas tendem a adotar as recomendações dadas por outras pessoas. Alguém que esteja com a intenção de adquirir um determinado produto, por exemplo, procura por opiniões de outras pessoas que já adquiriram o mesmo produto antes de realizar a compra. Isso acontece com frequência no cotidiano das pessoas com as mais variadas situações: a escolha de um filme, de um livro, de uma peça de teatro, de um médico, dentre tantas outras (RICCI et al., 2015).

A filtragem colaborativa tem a ideia geral de entregar a um usuário alvo itens que outros usuários, possuindo preferências similares a ele/ela, gostaram no passado (Adomavicius; Tuzhilin, 2005). A proximidade das preferências costuma ser calculada com base na similaridade do histórico de avaliação/interação dos usuários. Um usuário alvo u avaliará um item i com uma nota possivelmente similar àquela de outro usuário v desde que u e v tenham avaliado itens no passado de forma semelhante (RICCI et al., 2015). Esse esquema de avaliação, isto é, através de notas, é conhecido como avaliação explícita

ou, em inglês, como *feedback* explícito, no qual o usuário explicitamente informa o seu grau de satisfação com determinado item. Por outro lado, ainda existe o esquema de *feedback* implícito no qual as informações são obtidas através de opções de compras passadas, histórico de sites visitados, *links* clicados, *cookies* do *browser* ou até mesmo localidade geográfica (BOBADILLA et al., 2013).

Essa abordagem leva em consideração a comunidade de usuários e é comumente empregada em várias aplicações comerciais como *sites* de *streaming* de filmes e músicas por exemplo. Algoritmos desta abordagem recebem uma matriz de usuários *versus* itens contendo as avaliações de cada usuário para cada item, tal como apresenta a Tabela 2.2. A partir disso, é realizada uma predição de avaliação para aqueles itens que um usuário não avaliou (células em branco da matriz) e que são ordenados de acordo com o valor de relevância predito, resultando numa lista de recomendações (KAPOOR et al., 2015).

Tabela 2.2: Matriz usuário *versus* item

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9	i_{10}
Mário	2	3	1							
José			4			3			1	
Maria	5		1		5		2	4		2
Mara	4				5					4
Joana	2	3	1							2
Jorge					3		5			3

Fonte: adaptado de Jannach et al. (2010).

Os algoritmos desta abordagem podem ser categorizados naqueles baseados em vizinhança, também conhecidos como baseados em memória, e naqueles baseados em modelo. A recomendação baseada em vizinhança pode ser baseada em usuário (*user-based*) e em itens (*item-based*). A primeira busca estimar a relevância de um item para um usuário com base na similaridade de avaliação de usuários. Já a segunda considera a similaridade de avaliação de itens.

Os algoritmos baseados em modelo, por outro lado, visam “aprender” o comportamento do usuário, a partir das suas avaliações (explícitas ou não), utilizando abordagens de aprendizado de máquina. O modelo derivado é utilizado, então, para prever a relevância dos itens. Uma das abordagens mais difundidas nesse meio é a fatoração de matrizes, um modelo capaz de decompor a matriz usuários *versus* itens, derivando um modelo de predição (RICCI et al., 2015).

A seguir são apresentados com mais detalhes alguns dos algoritmos mais conhecidos de cada uma dessas duas categorias.

2.1.2.1 Filtragem Colaborativa baseada em vizinhança

Uma das primeiras estratégias desta categoria é a recomendação **user-based**, proposta por Resnick et al. (1994), cuja ideia central é encontrar usuários similares a um usuário ativo, ou seja, a quem se quer recomendar itens.

Tendo como base a matriz de utilidade apresentada na Tabela 2.2 e que as notas dadas aos itens variam de 1 (menor satisfação) a 5 (maior satisfação), é exemplificado o cálculo das recomendações para o usuário Jorge, lançando mão da estratégia **user-based**. O primeiro passo envolve descobrir quais são os usuários que possuem preferências semelhantes às dele. Isso é possível através do cálculo da similaridade de Jorge em relação a outros usuários, utilizando as notas dadas aos itens em comum. Uma das medidas mais empregadas nesse cálculo é o coeficiente de correlação de *Pearson*, dado pela Equação (2.2) (JANNACH et al., 2010).

$$sim(u, v) = \frac{\sum_{i \in I_{u,v}} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{u,v}} (r_{ui} - \bar{r}_u)^2 \sum_{i \in I_{u,v}} (r_{vi} - \bar{r}_v)^2}} \quad (2.2)$$

Nessa equação, I_{uv} faz referência ao conjunto de itens que os usuários u e v avaliaram em comum; r_{ui} e r_{vi} são as notas dadas a cada item por cada usuário; \bar{r}_u e \bar{r}_v representam a média das notas fornecidas pelos usuários u e v aos itens em comum. O emprego de \bar{r}_u e \bar{r}_v serve para normalizar o comportamento dos usuários em relação à atribuição de notas. Por exemplo, alguns usuários podem nunca avaliar itens com notas muito baixas, mesmo que não gostem do item, já outros usuários podem se comportar de maneira totalmente oposta. Portanto, a subtração da nota em questão pela média de notas do usuário normaliza esse comportamento (JANNACH et al., 2010).

Ainda, o valor da similaridade obtido a partir do coeficiente de correlação de *Pearson* fica no intervalo $[-1, 1]$. Isso significa que valores próximos a 1 indicam que dois usuários concordam em relação à avaliação dos itens e valores próximos a -1 indicam o oposto. Já os valores iguais a 0 indicam que dois usuários não são correlacionados. Com isso em vista, os valores de similaridade entre o usuário Jorge e os demais usuários, considerando a matriz de utilidade apresentada anteriormente, são os seguintes:

- $sim(\text{Jorge}, \text{Joana}) = -0,08$
- $sim(\text{Jorge}, \text{Mara}) = 0,407$
- $sim(\text{Jorge}, \text{Maria}) = 0,274$
- $sim(\text{Jorge}, \text{José}) = -0,363$

- $sim(\text{Jorge}, \text{Mário}) = -0,374$

O próximo passo envolve prever a nota (\hat{r}_{ui}) para um determinado item que um usuário alvo ainda não avaliou. Para tanto, é utilizada a Equação (2.3).

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum_{v \in U, v \neq u} (r_{vi} - \bar{r}_v) \times sim(u, v)}{\sum_{v \in U, v \neq u} |sim(u, v)|} \quad (2.3)$$

Onde U é o conjunto de usuários selecionados como vizinhança, isto é, os vizinhos mais próximos do usuário alvo. Existem várias possibilidades para a escolha de tal vizinhança sendo, uma delas, selecionar apenas os usuários com similaridade positiva ou, então, definir um valor mínimo de similaridade. A escolha da vizinhança é crucial, pois selecionar muitos vizinhos pode incluir usuários que não são comparáveis, prejudicando a qualidade da predição. O mesmo vale para uma vizinhança muito pequena. Estudos como o de Herlocker et al. (2004) sugerem que um bom tamanho de vizinhança varia de 20 a 50 vizinhos, dependendo o tamanho do *dataset* em questão.

Já \bar{r}_u é a média de notas dada pelo usuário alvo, considerando os itens em comum com sua vizinhança; r_{vi} é a nota dada pelos vizinhos ao item cuja nota está sendo predita para o usuário alvo; \bar{r}_v é a média de notas do usuário v , considerando os itens em comum ao usuário alvo u ; $sim(u, v)$ corresponde ao valor de similaridade entre os dois usuários, sendo considerado seu módulo como divisor uma vez que valores negativos de similaridade podem existir, dependendo da heurística adotada na escolha da vizinhança.

Retomando o exemplo didático, Jorge é o usuário alvo u e almeja-se prever a nota que ele daria para o item i_1 . Primeiramente, define-se sua vizinhança U contendo os usuários Maria e Mara, cujos valores de similaridade são, respectivamente, 0,274 e 0,407. Observa-se que essa vizinhança é composta apenas por usuários que apresentam valores positivos de similaridade, seguindo uma das abordagens heurísticas comentadas anteriormente. Definida a vizinhança, calcula-se \bar{r}_u , considerando as notas dos itens i_5 , i_7 e i_{10} , pois eles foram avaliados em comum pelos três usuários, logo:

$$\bar{r}_u = \overline{r_{Jorge}} = \frac{(3 + 5 + 4)}{3} = 4 \quad (2.4)$$

Em seguida, selecionam-se as notas r_{vi} do item i_1 dadas por cada usuário da vizinhança: $r_{Maria, i_1} = 5$ $r_{Mara, i_1} = 4$. A média de notas de cada usuário vizinho \bar{r}_v também é necessária, considerando os itens em comum com o usuário alvo Jorge. Para o usuário Maria são considerados os itens i_5 , i_7 e i_{10} , para o usuário Mara, são considerados os itens

i_5 e i_{10} . Com isso, tem-se que:

$$\bar{r}_v = \bar{r}_{Maria} = \frac{(5 + 2 + 2)}{3} = 3 \quad (2.5)$$

$$\bar{r}_v = \bar{r}_{Mara} = \frac{(5 + 4)}{2} = 4,5 \quad (2.6)$$

A partir disso, tem-se todos os elementos necessários para o cálculo da predição do item i_1 ao usuário Jorge. A seguir, a Equação (2.7) apresenta a adequação das variáveis ao exemplo para um melhor entendimento.

$$\hat{r}_{Jorge,i_1} = \bar{r}_{Jorge} + \frac{((r_{Maria,i_1} - \bar{r}_{Maria}) \times sim(Jorge, Maria)) + ((r_{Mara,i_1} - \bar{r}_{Mara}) \times sim(Jorge, Mara))}{|sim(Jorge, Maria) + sim(Jorge, Mara)|} \quad (2.7)$$

Por fim, basta substituir os valores das variáveis de acordo com as explicações anteriores. A Equação (2.8) apresenta o valor final da nota predita para o item i_1 .

$$\hat{r}_{Jorge,i_1} = 4 + \frac{((5 - 3) \times 0,274) + ((4 - 4,5) \times 0,407)}{0,274 + 0,407} \approx 4,506 \quad (2.8)$$

Com isso, observa-se que Jorge daria, segundo o algoritmo baseado em usuários, uma nota de aproximadamente 4,5 estrelas ao item i_1 supondo esse o esquema de avaliações. Esse exemplo demonstrou a predição de nota para apenas um item. Entretanto, a mesma lógica é aplicada para cada item que se queira recomendar. Ao obter um conjunto de itens com suas respectivas notas preditas, costuma-se empregar algum critério de ordenação, geralmente da maior nota para a menor, produzindo uma lista de recomendação *top-K* em que k é um limiar que indica quantos itens devem aparecer na lista. Logo, no topo na lista aparecerão os itens potencialmente mais relevantes para o usuário alvo.

A abordagem *user-based* é considerada uma das precursoras na área de recomendação. Entretanto, ela sofre com o tempo elevado de computação da vizinhança quando essa é muito grande. A fim de suavizar tal problema e permitir maior agilidade na entrega das recomendações, Sarwar et al. (2001) propôs a abordagem *item-based*. Diferentemente da abordagem anterior que opera nas linhas da matriz de utilidade, isto é, nos usuários, um algoritmo *item-based* trabalha com as colunas dessa matriz, ou seja, nos itens. Com isso, obtém-se uma grande vantagem que é um menor tempo de computação

da vizinhança porque o relacionamento entre itens é relativamente mais estático quando comparado com o relacionamento entre usuários. Assim, é possível pré-computar as similaridades entre itens de forma *off-line*, economizando tempo, e recuperar esses valores quando uma predição for requerida em tempo real. De forma geral, o funcionamento das duas abordagens é similar, envolvendo as mesmas etapas: o cálculo de similaridade e o cálculo de predição.

O cálculo da similaridade entre dois itens i e j envolve, primeiramente, isolar os usuários que avaliaram esses itens e então aplicar alguma técnica para determinar a similaridade $s_{i,j}$. Sarwar et al. (2001) demonstrou que a técnica de similaridade de cossenos ajustada apresenta o menor erro médio no resultado final do processo de recomendação. A equação (2.9) apresenta essa técnica em U representa o conjunto de usuários que avaliaram os itens i e j ; e \overline{R}_u é a média das notas do n -ésimo usuário.

$$sim(i, j) = \frac{\sum_{u \in U} (R_{ui} - \overline{R}_u)(R_{uj} - \overline{R}_u)}{\sqrt{\sum_{u \in U} (R_{ui} - \overline{R}_u)^2 \sum_{u \in U} (R_{uj} - \overline{R}_u)^2}} \quad (2.9)$$

Uma vez isolado o conjunto de itens mais similares de acordo com a medida de similaridade, o próximo passo envolve olhar para as notas do usuário alvo, aplicando alguma técnica de predição. Uma delas é a soma ponderada cuja predição de nota de um item i para um usuário u é a soma das notas que u deu aos itens similares a i . Cada nota é ponderada de acordo com a similaridade $s_{i,j}$ entre os itens i e j , conforme apresenta a Equação (2.10)

$$\hat{r}_{u,i} = \frac{\sum_{\text{todosItensSimilares}, N} (s_{i,N} \times R_{u,N})}{\sum_{\text{todosItensSimilares}, N} (|s_{i,N}|)} \quad (2.10)$$

Essa técnica tenta endereçar como o usuário alvo avalia itens similares. A média aritmética é ponderada com a soma dos termos de similaridades para garantir que as predições estejam em limites pré-definidos (SARWAR et al., 2001).

2.1.2.2 Filtragem Colaborativa baseada em modelo

A segunda estratégia mais comum para a filtragem colaborativa é a baseada em modelo, a qual usa as avaliações dos usuários sobre os itens para “aprender” um modelo que é usado, posteriormente, para fazer as predições e as recomendações. Ao invés de utilizarem medidas de similaridade, os algoritmos se caracterizam pela criação de modelos. Pode-se dizer que os modelos são padrões a respeito do comportamento dos dados e

que os algoritmos que se baseiam neles analisam a estrutura da matriz de utilidade para encontrar relações entre os itens. O pressuposto dessa abordagem é que os usuários se interessam por itens similares aos bem avaliados por eles no passado (JANNACH et al., 2010; RICCI et al., 2015).

Outra característica é que os algoritmos baseados em modelos não necessitam calcular vizinhança de usuários ou itens similares que é, na verdade, um dos grandes limitadores de desempenho das abordagens baseadas em memória. Sendo assim, um algoritmo baseado em modelo costuma gerar recomendações mais rapidamente desde que os modelos já estejam treinados (SARWAR et al., 2001). Além disso, permitem adicionar informações extras no processo como *feedback* implícito, efeitos temporais e níveis de confiança (KOREN; BELL; VOLINSKY, 2009).

Diversas abordagens são possíveis para a obtenção de modelos, dentre elas: métodos de agrupamento, redes Bayesianas, regras de associação, modelos de fatoração de matrizes, dentre outros (Adomavicius; Tuzhilin, 2005). A fatoração de matrizes se ancora na ideia de projetar usuários e itens para um espaço de baixa dimensionalidade. Os algoritmos dessa técnica atrelam cada usuário u a um vetor de fatores latentes $p_u \in \mathbb{R}^f$; e cada item i a um vetor de fatores latentes $q_i \in \mathbb{R}^f$, em que f é o número de fatores latentes. A predição de nota \hat{r}_{ui} é obtida pelo produto escalar dos valores p_u e q_i , tal como apresenta a Equação (2.11).

$$\hat{r}_{ui} = p_u^T \times q_i \quad (2.11)$$

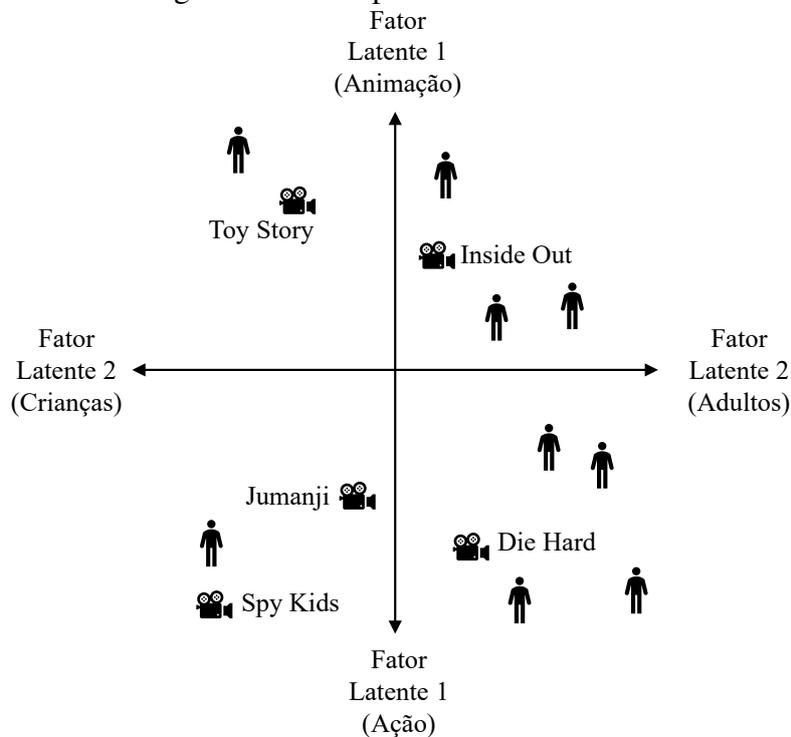
Este modelo tornou-se conhecido em SR após a solução dada pelo time *Bellkor's Pragmatic Chaos* ao prêmio Netflix em 2009, quando a técnica de fatoração de matrizes foi aplicada para fazer recomendações de filmes (KOREN, 2009).

As notas são influenciadas por um conjunto de fatores que são específicos de acordo com o domínio dos itens que estão sendo recomendados como, por exemplo, músicas, filmes, livros, notícias, dentre outros. Entre possíveis fatores latentes, relacionando usuários a filmes, por exemplo, poder-se-ia extrair características como gênero, faixa etária, profundidade no desenvolvimento dos personagens, ou mesmo características cujas denominações (ainda) não estão definidas. Em geral, os fatores não são óbvios, sendo difícil estimar o impacto de cada um nas notas. O espaço latente busca explicar as notas por meio da caracterização automática dos usuários e itens através dos fatores.

Para melhor exemplificar essa estratégia, é apresentado um exemplo ilustrativo de fatores latentes e de como esses permitem comparar usuários e itens. A Figura 2.2 repre-

seja um conjunto de dados distribuído em duas dimensões, os chamados fatores latentes que foram extraídos de uma matriz de utilidade hipotética de filmes. O primeiro fator caracteriza a semelhança dos itens/usuários de acordo com filmes de ação ou animação, ao passo que o segundo fator segue a mesma ideia, porém para filmes voltados para adultos ou crianças. Assim, tendo usuários e itens mapeados em um mesmo espaço no qual eles possam ser comparados, torna-se mais fácil fazer as recomendações. Por exemplo, para o usuário que se encontra no segundo quadrante, o filme mais próximo a ele no espaço de fatores latentes é *Toy Story*.

Figura 2.2: Exemplo de fatores latentes.



Fonte: Adaptado de Koren, Bell e Volinsky (2009)

Formalmente, para cada item i , os elementos de q_i medem em que grau o item possui os fatores latentes. Para cada usuário u , os elementos de p_u medem o grau de interesse que u tem nos itens com alta correspondência de fatores latentes. O aprendizado envolvido para a obtenção de $p_u \in P$, e $q_i \in Q$ é um problema de otimização em que dada uma matriz de utilidade esparsa R , busca-se encontrar as matrizes P e Q tal que o produto delas resulte numa matriz muito próxima de R , como ilustra a Figura 2.3 (KOREN; BELL; VOLINSKY, 2009), em que f_n representa os fatores; i_n representa os itens; e U_n representa os usuários.

Figura 2.3: Exemplo de fatoração da matriz de utilidade utilizando o algoritmo SVD

	i_1	i_2	i_3	i_4	i_5
U_1	4	2	3	1	4
U_2	2			5	
U_3	5	5	1		2

 $=$

	f_1	f_2	f_3	f_4	f_n
U_1					
U_2					
U_3					

 \times

	i_1	i_2	i_3	i_4	i_5
f_1					
\vdots					
f_n					

R = Matriz de utilidade
P
Q

Fonte: Adaptado de Koren e Bell (2015)

A matriz $P \times Q$ contém as previsões de nota. Entretanto, essa nova matriz, por ser uma aproximação, apresenta um nível de erro que deve ser minimizado, caso contrário, prejudica a qualidade das recomendações. Para tanto, existem duas técnicas amplamente utilizadas que são: descendente gradiente estocástico (*SGD - Stochastic Gradient Descent*); e mínimos quadrados alternados (*ALS - Alternating Least Squares*) (KOREN; BELL; VOLINSKY, 2009).

Além do algoritmo SVD, cuja ideia foi esboçada até aqui, existe uma variante dele chamada SVD++ que além da fatoração de matrizes, considera o *feedback* implícito dos usuários para realizar a previsão. Maiores detalhes sobre essa e outras variantes podem ser conferidos no trabalho de Koren e Bell (2015).

2.1.3 Vantagens e desvantagens da filtragem colaborativa e da filtragem baseada em conteúdo

De forma geral, a principal vantagem da filtragem colaborativa é a não necessidade de se conhecer os atributos do itens a serem recomendados já que sua sistemática reside nas interações dos usuários com os itens. Isso contribui para que se possa recomendar itens em diferentes domínios de aplicação (Adomavicius; Tuzhilin, 2005).

Por outro lado, essa abordagem sofre com a entrada de um novo usuário ou item, conhecido como problema de partida a frio (*cold-start*). Por se basear somente nas avaliações, essa abordagem não consegue estimar a relevância de um item quando não existe avaliação ou elas são poucas para um item. Outra desvantagem é a esparsidade de dados da matriz de avaliações, ou seja, os usuários fornecem avaliações para poucos itens o que acarreta em um grande número de valores faltantes. Poucas avaliações impactam a qualidade da previsão (Adomavicius; Tuzhilin, 2005; JANNACH et al., 2010).

Em relação à filtragem baseada em conteúdo, uma de suas grandes vantagens é que

ela não requer a comunidade para gerar as recomendações. Os itens podem ser imediatamente recomendados assim que entram no sistema, desde que possuam uma descrição suficiente, extraída automática ou manualmente. Por suficiente, Jannach et al. (2010) esclarece que as descrições sobre os itens devem possuir um número adequado de atributos e, também, que esses atributos possuam alta qualidade. Essa diretriz pode ser atingida por meio de estudos sobre os itens, principalmente se ancorando em áreas como a engenharia de características (*feature engineering*). Outra vantagem é que as recomendações podem ser geradas mesmo se existir apenas um único usuário (JANNACH et al., 2010).

Em contrapartida, prover descrições dos itens é uma tarefa custosa que enquadra-se como uma desvantagem desta abordagem. Por exemplo, o *Music Genome Project* é um esforço em que vários músicos participam anotando manualmente as músicas com informações sobre os tipos de instrumentos utilizados, influências de gênero, e outras centenas de atributos não fornecidos pelas gravadoras. Em ambientes práticos, as descrições de um item - como o gênero de um livro ou a lista de atores em um filme - já são parcialmente fornecidas pelo(s) criador(es). O que permanece desafiador, no entanto, é a aquisição de características subjetivas e qualitativas. Em relação ao gosto, por exemplo, os motivos pelos quais alguém gosta de algo nem sempre estão relacionados a certas características do produto e podem ser baseados em uma impressão superficial sobre o item (JANNACH et al., 2010).

Outra desvantagem refere-se à superespecialização, ou seja, o sistema tende a recomendar itens muito similares, “mais do mesmo” (JANNACH et al., 2010; Adomavicius; Tuzhilin, 2005). Por exemplo, um sistema baseado em conteúdo que recomenda músicas de um determinado artista e, a partir disso, o usuário marca como relevante esse tipo de conteúdo, então cada vez mais esse sistema entregará conteúdo daquele mesmo artista.

2.1.4 Recomendação híbrida

A recomendação híbrida surge como uma combinação das abordagens anteriores na intenção de cobrir as desvantagens de uma com as vantagens da outra (Adomavicius; Tuzhilin, 2005). Burke (2002) discute várias maneiras de combinar diferentes abordagens de recomendação em direção a uma abordagem híbrida, conforme consta a seguir:

- *Weighted*: as avaliações das abordagens de recomendação são combinadas para produzir uma recomendação única;

- *Switching*: o sistema escolhe a abordagem de acordo com a situação atual;
- *Mixed*: as recomendações de diferentes abordagens são apresentadas juntas;
- *Feature combination*: características de diferentes fontes de recomendação são combinadas em um único algoritmo;
- *Cascade*: um recomendador refina a lista gerada por outro e, assim, sucessivamente;
- *Feature augmentation*: a saída de uma técnica é utilizada como característica de entrada para outra abordagem.

Atualmente, grande parte das abordagens mercadológicas de recomendação, como um todo, não são puramente baseadas em conteúdo ou, então, baseadas em filtragem colaborativa. Elas são, na verdade, uma combinação dessas estratégias junto a outras como, por exemplo, baseadas em conhecimento, baseadas em contexto, dentre outras (JANNACH et al., 2010).

2.1.5 Principais Desafios e Oportunidades

Tradicionalmente, grande parte dos SR tem como principal objetivo melhorar cada vez mais a acurácia da predição de nota (KAPOOR et al., 2015). Entretanto, esforços da literatura têm apresentado um foco diferente de estudo, apontando que apenas melhorar a acurácia da predição de nota não necessariamente aumenta o nível de qualidade percebida pelo usuário nas recomendações (MCNEE; RIEDL; KONSTAN, 2006; JANNACH; ADOMAVICIUS, 2016). Portanto, uma nova tendência de pesquisa se mostra presente, cujo foco está no melhoramento de outros aspectos além da acurácia. Como exemplo, prover explicações do porquê das recomendações é um componente crítico que corrobora para a transparência dos algoritmos já que muitos deles operam sendo “caixas pretas”, ou seja, não se sabe ao certo como as recomendações são produzidas pelos mecanismos internos (ABDOLLAHI; NASRAOUI, 2016).

Outro aspecto que está em evidência na filosofia “além da acurácia” é a **diversidade** de itens em uma lista de recomendação. Defende-se que a diversidade pode aumentar o espectro de interesse de um usuário, fazendo com que ele descubra novos itens e melhore sua experiência com o sistema (ZHOU et al., 2010; Adomavicius; Kwon, 2012; ABISHEVA; GARCIA; SCHWEITZER, 2016). Esse é um dos focos desta pesquisa e, por isso, é tratada com maiores detalhes na Seção 2.3. Outro aspecto que também pode ser citado envolve questões de privacidade de dados. Os SR devem considerar a não coleta in-

discriminada de dados do usuário sem o seu conhecimento ou, então, ser intrusivo, fornecendo recomendações de forma impositiva e/ou fora de hora (KNIJNENBURG; KOBSA, 2013; ZHANG et al., 2019; Qi et al., 2020).

Além dos aspectos mencionados anteriormente, alguns domínios também se apresentam como desafios e temas de pesquisa relevantes. Um deles se refere às cidades inteligentes que são caracterizadas por conterem uma heterogeneidade de itens e dados muito grande, sendo, portanto, um desafio para os SR operarem. Vários subtópicos devem ser considerados tendo em vista sua estreita relação com cidades inteligentes como, por exemplo, internet das coisas, *web services* e dados abertos (Di Martino; ROSSI, 2016; Habibzadeh et al., 2018; Qi et al., 2020; QUIJANO-SÁNCHEZ et al., 2020). Outro domínio fortemente desafiador para os SR é o de notícias, pois esse tipo de item, além da sua característica informacional inerente, também possui um potencial de mudar opiniões para o bem e/ou para o mal, a depender da sua fonte geradora. Logo, um SR pode ou não amplificar esse efeito, dependendo de como foi implementado, especialmente em um ambiente de mídias sociais a partir dos quais muitas pessoas se mantêm informadas acerca dos fatos mundanos (BARILE et al., 2019; HARAMBAM et al., 2019; HU et al., 2020). Maiores detalhes sobre o domínio de recomendação de notícias podem ser encontrados na Seção 2.1.8, tendo em vista sua relação com este trabalho.

A seguir, são descritos alguns problemas considerados permanentes e transversais a qualquer SR. Esses problemas desafiam muitas pesquisas e sistemas comerciais que buscam, constantemente, maneiras de suavizá-los.

2.1.5.1 Esparsidade de dados

Segundo Bauer et al. (2012), a esparsidade (ES) mede o nível de valores faltantes da matriz e é calculada pela Equação (2.12), sendo $N_{\bar{0}}$ o número de elementos não-nulos; n e m o número de linhas e colunas da matriz.

$$ES = 1 - \frac{N_{\bar{0}}}{n \times m} \quad (2.12)$$

Quanto mais uma matriz estiver preenchida por valores não-nulos $ES \rightarrow 0$. Por outro lado, caso a matriz contenha muitos valores nulos $ES \rightarrow 1$. Nos SR, de forma geral, $ES \rightarrow 1$ porque os usuários não avaliam todos os itens do catálogo. Adomavicius e Tuzhilin (2005) sustentam essa afirmação dizendo que em qualquer SR o número de avaliações obtido é geralmente muito pequeno comparado ao número de avaliações

necessárias para realizar a predição. A filtragem colaborativa, por exemplo, depende da disponibilidade de uma quantidade considerável de usuários. Ao se recomendar filmes, por exemplo, podem existir muitos deles que foram avaliados por um pequeno grupo de usuários e que, por isso, raramente serão recomendados, ainda que as notas sejam altas. Uma forma de lidar com o problema é utilizar a técnica de decomposição de valores singulares, *SVD*, explicada anteriormente, para reduzir a dimensionalidade de uma matriz que apresente grande esparsidade (KOREN, 2009).

Da mesma forma, a abordagem baseada em conteúdo sofre com os poucos dados disponíveis. Se um usuário não interage com o sistema e não disponibiliza suas preferências de forma explícita ou implícita, dificilmente o SR conseguirá calcular boas recomendações (JANNACH et al., 2010).

2.1.5.2 *Partida a frio (cold-start)*

Esse é um desafio relacionado a usuários e itens novos. Um usuário que recém entrou na matriz de utilidade ou, então, que não possui o hábito de avaliar os itens, por exemplo, representa um grande problema, pois um SR colaborativo não conseguirá gerar recomendações personalizadas para esses usuários. Em uma situação ideal, desde as primeiras avaliações esse tipo de SR consegue recomendar, embora seja necessário um número razoável de avaliações a fim de se obter recomendações confiáveis (BOBADILLA et al., 2013; RICCI et al., 2015).

O mesmo efeito ocorre com os itens que, quando novos, não estão avaliados por nenhum usuário e, por isso, não podem ser recomendados, situação que os deixa fora do círculo da recomendação até que algum usuário os avalie. A teoria da cauda longa (*long tail*) está relacionada a esse problema, pois é observado nos SR, de forma geral, que alguns itens são muito populares ao passo que a grande maioria dos itens possuem poucas avaliações (JANNACH; ADOMAVICIUS, 2016).

O problema do item novo tem menos impacto sobre os SR quando comparado com o problema do usuário novo, pois pode-se criar políticas de propaganda para esses itens, a fim de promovê-los e, assim, suavizar o problema (BOBADILLA et al., 2013). A estratégia mais comum para suavizar os problemas de usuários e itens novos é buscar informações adicionais ao conjunto de avaliações de usuários sobre itens. Geralmente são utilizados algoritmos híbridos que combinam a filtragem colaborativa com outros tipos de recomendação (BURKE, 2002).

2.1.5.3 Superespecialização

A superespecialização é um desafio mais facilmente observável nos SR baseados em conteúdo. Isso porque, geralmente, esse tipo de abordagem não adota nenhum método para recomendar itens inesperados e, assim, apresenta um nível limitado de novidade na coleção. Um usuário, por exemplo, que avaliou positivamente livros de um determinado autor possivelmente receberá mais e mais livros desse mesmo autor (PAZZANI; BILLSUS, 2007). Entretanto, a filtragem colaborativa também sofre esse efeito, especialmente por conta do efeito da cauda longa, mencionado anteriormente. Isso faz com que o grande número de avaliações de determinados itens os promovam ainda mais, superespecializando as recomendações (ADAMOPOULOS; TUZHILIN, 2014).

2.1.5.4 Escalabilidade

A escalabilidade está ligada ao aumento demasiado de itens e usuários. Uma base de dados que contenha milhões de itens e usuários exigirá, de uma abordagem de filtragem colaborativa, o cálculo da similaridade do usuário ativo para a determinação de sua vizinhança, além da previsão de nota para milhões de itens que esse ainda não avaliou. Já uma abordagem baseada em conteúdo precisa ser escalável para o montante de características dos itens assim como para o montante de itens que precisam ser comparados ao perfil do usuário (TAKÁCS et al., 2009).

Portanto, os SR precisam ser escaláveis na intenção de reagir imediatamente às requisições, independente do tamanho da base de dados. Uma das principais alternativas para tratar esse desafio é o emprego de técnicas de redução de dimensionalidade. Algoritmos de filtragem colaborativa baseados em modelo lidam com o problema diminuindo o número de comparações necessárias. Com isso, tem-se uma otimização no tempo de processamento e, por consequência, maior responsividade na entrega das recomendações (GHAZANFAR; PRUGEL-BENNETT, 2010).

2.1.6 Tipos de Experimentos

Em diversas situações é necessário comparar diferentes SR a fim de decidir qual é mais adequado para determinada situação. Para tanto, são empregados diferentes tipos de experimentos sendo, os *offline*, estudos de usuário, e *online* os mais comuns em SR, conforme discutem Gunawardana e Shani (2015).

2.1.6.1 Offline

Este tipo de experimento é realizado utilizando um conjunto de dados coletado durante um intervalo de tempo. Os dados disponíveis, tipicamente as avaliações, são divididos em conjuntos de treinamento e de teste. As avaliações no conjunto de treino são usadas como entrada para o SR de forma que esse estabeleça um modelo de predição para as interações usuários-itens. A partir disso, o sistema calcula um *score*, isto é, prediz uma nota para cada item do conjunto de teste e que não esteja no conjunto de treinamento, associado ao seu respectivo usuário. As notas reais dos itens presentes no conjunto de teste são ocultadas momentaneamente e, após a predição, são comparadas com as notas preditas, permitindo o cálculo de várias métricas, dentre elas o erro, a acurácia, dentre outras (CAÑAMARES; CASTELLS; MOFFAT, 2020).

Assim, assume-se que o padrão de consumo dos usuários, quando os dados foram coletados, será similar o suficiente ao padrão no momento em que o SR entrar em produção. A principal vantagem desse tipo de experimento é que não há interação com os usuários no momento do treinamento, fazendo com que a complexidade de comparação de diferentes algoritmos seja reduzida. Dessa forma, o objetivo dos experimentos *offline* é filtrar algoritmos que demonstrem baixa acurácia, separando os algoritmos candidatos a testes utilizando métodos mais custosos como, por exemplo, experimentos *online* (GUNAWARDANA; SHANI, 2015).

Existem vários conjuntos de dados públicos disponíveis na *web* que podem ser utilizados em experimentos *offline*. Um exemplo deles é o MovieLens, coletado pelo *GroupLens*¹ *Research Project* da Universidade de Minnesota, cuja história é detalhada por Harper e Konstan (2016) e contém avaliações sobre filmes com notas de 1 a 5 dadas por usuários da plataforma de avaliação MovieLens. Outro exemplo é o Book-Crossing², criado por Ziegler et al. (2005), que contém avaliações de usuários sobre livros.

Tipicamente, esses *datasets* consistem de um ou mais arquivos de notas as quais são parte de uma tupla (u, i, r) , sendo u um usuário; i um item; e r uma nota. Em alguns casos, existe um *timestamp* atrelado também, assim como o conteúdo textual de cada item e a anonimização dos usuários (CAÑAMARES; CASTELLS; MOFFAT, 2020).

¹<https://grouplens.org/>

²<https://grouplens.org/datasets/book-crossing/>

2.1.6.2 Estudo de usuário

Um estudo de usuário é conduzido através do uso de um SR pensado para esse fim específico, no qual são selecionados usuários que interagem com o sistema. A partir disso, são registrados os comportamentos dos usuários, coletando quais tarefas foram ou não completadas, a acurácia das recomendações, o tempo de interação. Também é comum indagar o usuário acerca de questões qualitativas antes, durante e depois da tarefa, que permitem estimar parâmetros não observáveis quantitativamente como, por exemplo, a facilidade de uso do sistema, a acessibilidade da interface gráfica ou, então, a satisfação com as recomendações recebidas (GUNAWARDANA; SHANI, 2015).

Um estudo de usuário pode responder o maior número de questões quando comparado com as outras configurações experimentais apresentadas aqui, isto é, *offline* e *online*. Isso porque o estudo de usuário permite coletar dados qualitativos que podem ser cruciais para a interpretação de resultados quantitativos. Ainda, por ser especificamente concebido como um estudo, permite monitorar e coletar uma grande quantidade de dados quantitativos dos usuários de forma controlada (GUNAWARDANA; SHANI, 2015).

Por outro lado, de acordo com Kohavi et al. (2013), estudos de usuário são custosos de se conduzir. Afinal, recrutar um número expressivo de usuários que execute um número também expressivo de tarefas é custoso tanto em termos de tempo do usuário como quanto monetários, quando os usuários são pagos para participarem dos experimentos. Por isso, é importante que seja considerado um conjunto pequeno de usuários, um conjunto pequeno das tarefas mais importantes, e que não é possível testar todos os cenários possíveis (GUNAWARDANA; SHANI, 2015).

Tipicamente, um estudo de usuário é usado para comparar alguns sistemas candidatos sendo, cada um, testado com as mesmas tarefas. Os candidatos podem ser comparados entre sujeitos (*between subjects*), isto é, cada usuário é designado a cada sistema candidato. Essa abordagem é conhecida, também, como teste A/B (*All Between*) ou, ainda, teste e controle. O grupo de teste contém os usuários que testam o sistema cuja melhoria se quer medir e o grupo de controle contém os usuários que testam um sistema *baseline*, por exemplo (GUNAWARDANA; SHANI, 2015; KOHAVI, 2015).

2.1.6.3 Online

A definição de experimento *online* é muito similar à definição de um experimento *offline* dada anteriormente. A principal diferença é que um experimento *online* considera

uma aplicação de recomendação que não foi concebida especificamente para um estudo mas, sim, para ser utilizada de forma ampla. Por exemplo, o sistema de recomendação de filmes utilizado pela *Netflix* poderia ser utilizado pelo time de desenvolvimento da empresa para realizar um experimento *online*, comparando diferentes algoritmos a partir de tarefas executadas na plataforma por usuários reais.

Portanto, um experimento *online* é o que provê as evidências mais fortes no que se refere ao valor do sistema, uma vez que é realizado com usuários reais que realizam tarefas reais em um sistema real (GUNAWARDANA; SHANI, 2015). Por essa razão, vários sistemas de uso amplo empregam experimentos do tipo *online* (KOHAVI et al., 2009). Geralmente, esses sistemas direcionam uma pequena porcentagem do seu tráfego para diferentes algoritmos de recomendação, registrando as interações dos usuários com cada um deles (GUNAWARDANA; SHANI, 2015).

Avaliações *online* são as únicas que permitem a estimativa direta dos objetivos gerais do sistema, como lucro a longo prazo ou retenção de usuários, uma vez que o sistema é de uso amplo e, por isso, seu uso não é findado tão brevemente como acontece com estudos de usuário muito específicos.

2.1.7 Métricas de Avaliação

Um SR pode ser avaliado por meio de diferentes métricas que buscam mensurar aspectos diferentes como, por exemplo, a qualidade do *ranking*, a acurácia da predição de nota, a acurácia da predição de uso (precisão e revocação), dentre outras. A acurácia é o aspecto mais difundido em SR, principalmente porque grande parte dos recomendadores se baseiam em mecanismos de predição, seja de nota ou da probabilidade de uso dos itens, e porque existe a suposição de que um SR que possua maior acurácia nas predições será aquele preferido pelo usuário (GUNAWARDANA; SHANI, 2015).

Por isso, a seguir, são detalhadas duas métricas que foram adotadas nesta tese, uma para avaliar a acurácia da predição de nota como parte da análise de bolhas de filtro e, a outra, para avaliar estatisticamente a significância dos resultados obtidos entre abordagens de recomendação.

2.1.7.1 Acurácia da predição de nota

O RMSE (*Root Mean Squared Error*) mede o erro entre a nota predita (\widehat{r}_{ui}) para um item i e a nota dada (r) por um usuário u , de um conjunto de itens S , sendo r_{ui} conhecida. O RMSE é calculado de acordo com a Equação (2.13), sendo considerado como a métrica mais popular utilizada para avaliar a acurácia da predição de notas (GUNAWARDANA; SHANI, 2015).

$$RMSE = \sqrt{\frac{1}{|S|} \sum_{(u,i) \in S} (\widehat{r}_{ui} - r_{ui})^2} \quad (2.13)$$

2.1.7.2 Validação estatística

Todo e qualquer tipo de experimento envolvendo SR requer confiança nos resultados reportados. Só assim é possível tomar uma decisão mais segura de qual algoritmo deve ser adotado, dentre vários sendo comparados. Para reduzir as chances de erro, devem ser aplicados testes de significância nos resultados (GUNAWARDANA; SHANI, 2015).

O resultado de tais testes é o nível de significância, conhecido como p -valor, isto é, a probabilidade dos resultados obtidos ocorrerem ao acaso. Na prática, são adotados testes de significância paramétricos e/ou não paramétricos tais como teste t, teste de Wilcoxon, teste de Mann-Whitney, dentre outros, de acordo com a situação dos dados como, por exemplo, o tamanho da amostra disponível, a distribuição dos dados, ou o tipo de agrupamento dos sujeitos. Se o p -valor estiver abaixo de um limite, obtém-se confiança de que a hipótese nula não é verdadeira e, portanto, os resultados podem ser julgados significativos. Tradicionalmente, adota-se $p = 0,05$ como seu limite, o que indica 95% de confiança (BICKEL; DOKSUM, 2001).

2.1.8 Recomendação de notícias

Sistemas de Recomendação de Notícias (*News Recommender Systems - NRS*) são uma classe de SR onde os itens recomendados são notícias. Essas recomendações são entregues de maneira personalizada, seja em um portal de notícias ou, então, através de agregadores de notícias que aglomeram e recomendam a partir de várias fontes. Como exemplos pode-se citar o Google e o Yahoo! News. A entrega digital de conteúdo permite aos editores distribuir conteúdos novos e atualizados em tempo real, situação que resulta

em uma velocidade aumentada de publicação. Devido a essa velocidade, mais notícias são geradas em menos tempo, tornando a análise de tudo que foi publicado impraticável por um único usuário. Sendo assim, os NRS são essenciais para diminuir a sobrecarga de informação sobre o usuário (KARIMI; JANNACH; JUGOVAC, 2018).

De acordo com os mesmos autores, Karimi, Jannach e Jugovac (2018), recomendar notícias representa um domínio de aplicação tal como recomendar músicas, filmes, vídeos, dentre outros, nos quais várias das técnicas de recomendação conhecidas são aplicadas. Entretanto, o problema de recomendar notícias envolve certas características específicas que não se aplicam a outros domínios ou são menos evidentes. Uma delas é a relevância dos itens. No domínio de NRS a relevância de um item deve ser estimada não apenas através de seu conteúdo ou do quanto o usuário poderia gostar daquele tópico. Notícias têm tempo de vida, assim a relevância de um item pode mudar rapidamente (devido a eventos recentes, como uma copa do mundo ou um desastre natural). Outra característica é a rotatividade do item. De fato, como os sites de notícias são atualizados continuamente, alguns artigos podem ser substituídos por um artigo de “notícias de última hora” sobre o mesmo tópico várias vezes durante o mesmo dia, o que pode exigir atualizações constantes dos modelos de recomendação. Outra característica desafiadora é o dinamismo de mudança dos interesses dos usuários que depende de fatores contextuais como hora do dia, dispositivo utilizado ou localização.

Um dos desafios da utilização de uma estratégia de recomendação de notícias no ambiente acadêmico é a obtenção em tempo real de novos itens, já que muitos dos portais de notícias têm proteção de direitos autorais sobre seu conteúdo e exigem autorização prévia para utilização do mesmo. Outras fontes, como o *Twitter*, impõem um limite à coleta de novos *tweets*.

SR, em geral, são classificados em quatro principais áreas: filtragem colaborativa, filtragem baseada em conteúdo, técnicas baseadas em conhecimento, e abordagens híbridas (JANNACH et al., 2010). A filtragem colaborativa é a mais utilizada basicamente porque é independente de domínio, uma vez que não requer conhecimento das características sobre os itens recomendados. Entretanto, para a recomendação de notícias, a abordagem mais utilizada é a filtragem baseada em conteúdo já que os itens são documentos de texto representados com técnicas comumente conhecidas da área de Recuperação de Informações. Logo em seguida está a abordagem híbrida, combinando a filtragem colaborativa e a filtragem baseada em conteúdo (KARIMI; JANNACH; JUGOVAC, 2018).

Para que se possa personalizar as sugestões de leitura, os SR precisam manter per-

fis de usuário para capturar as preferências de leitura ao longo do tempo. O modo como esses perfis de usuário são coletados e quais informações são armazenadas geralmente estão atreladas ao paradigma de recomendação escolhido. Semelhante a outros domínios de aplicação, em geral, pode-se estimular os usuários a fornecerem informações de preferência explícitas, por exemplo, na forma de avaliações sobre os itens ou, então, monitorar e interpretar o comportamento passado do usuário, configurando-se como uma estratégia de *feedback* implícito. As características que são comumente utilizadas como *feedback* implícito são (KARIMI; JANNACH; JUGOVAC, 2018):

- a leitura de um artigo;
- o compartilhamento;
- a impressão ou o comentário;
- o tempo de permanência;
- o movimento do mouse como indicador de interesse.

Além disso, um dos fatores essenciais para NRS é o quão recente é um artigo, impactando diretamente o cálculo de relevância dos itens em uma função de *ranking*, por exemplo. Tecnicamente, a recência de um artigo no processo de recomendação é trabalhada em três etapas, sendo elas (KARIMI; JANNACH; JUGOVAC, 2018):

- Pré-filtragem: filtrar notícias supostamente desatualizadas antes de calcular previsões de relevância ou uma classificação de itens;
- Modelagem: incorporar o fator de recência dentro dos algoritmos de recomendação;
- Pós-filtragem: filtrar ou rebaixar artigos após o processo de classificação principal.

O principal objetivo de otimização dos pesquisadores na academia é prever com precisão a relevância de uma notícia para um usuário. No entanto, na prática, prever a relevância de um item é, em muitos casos, insuficiente. Se, por exemplo, um usuário está interessado em política e demonstrou grande interesse em artigos sobre uma eleição presidencial em andamento no passado, recomendar mais artigos sobre esse tópico provavelmente é uma boa escolha. No entanto, recomendar apenas artigos sobre a eleição, ou apenas sobre política, pode ser monótono demais para os usuários e provavelmente não levaria a um alto envolvimento no futuro. No caso de um site de agregação de notícias, também é importante que as notícias recomendadas não sejam muito semelhantes entre si. Apresentar três artigos de três fontes diferentes sobre, por exemplo, o mesmo acidente de avião pode ter pouco valor para os usuários. Portanto, um desafio adicional, além de

prever com precisão se um artigo é relevante para um usuário ou não, é levar em consideração fatores adicionais de qualidade que são apresentados brevemente abaixo com vistas ao domínio de notícias (KARIMI; JANNACH; JUGOVAC, 2018):

- diversidade: os usuários de sistemas de recomendação de notícias podem se interessar por vários tópicos. Um sistema de recomendação deve, portanto, ser capaz de abordar esses gostos variados e gerar listas de recomendações diversificadas;
- novidade: é definida em termos da não-obviedade das recomendações de itens. Também pode ser vista como aqueles itens que o usuário ainda não viu, mas que são relevantes para ele. O domínio de notícias tem uma forte influência de notícias mais lidas (populares), o que tende a prejudicar o fator de qualidade novidade;
- serendipidade: é uma visão mais forte de novidade, correspondendo a itens não apenas novos, mas também positivamente surpreendentes para o usuário.

A área de recomendação de notícias está estreitamente relacionada com o efeito de bolhas de filtro e, por este motivo, são discutidas na próxima seção.

2.2 Bolhas de filtro

Bolha de filtro (*filter-bubble*) é um termo cunhado em 2011 por Eli Pariser para descrever o potencial que a personalização *online* tem de isolar as pessoas de uma diversidade de pontos de vista ou conteúdos. Como resultado, o usuário é exposto apenas a informações de um determinado nicho ou a um tipo de informação que ele já conhece. Essa situação começou a ser observada a partir de 2009 quando plataformas como Google moveram esforços em direção à personalização das buscas fazendo com que diferentes usuários, com base em suas interações anteriores, preferências explícitas, localização, dentre vários outros fatores, obtivessem diferentes resultados (PARISER, 2011).

Tetlock (2005) encontrou indícios do efeito de bolha de filtro na área de pesquisa política. Durante um estudo, foi evidenciado que pessoas comuns, com diferentes opiniões e pontos de vista, fizeram previsões políticas e econômicas mais precisas quando comparadas a especialistas da área. A baixa precisão de previsão dos especialistas justificou-se pelo fato de que suas opiniões podem ser reforçadas e enviesadas após anos de estudo. Os sistemas de recomendação buscam exatamente uma situação parecida, isto é, entregar itens que sejam cada vez mais do interesse do usuário, fato que pode diminuir ligeiramente a diversidade de conteúdo que esses consomem, conforme apontado

no estudo de Nguyen et al. (2014).

Para entender como as bolhas de filtro são criadas, é importante entender como os algoritmos funcionam, porque, embora não sejam a única causa, desempenham um papel fundamental. Por exemplo, cada pessoa tende a consumir conteúdo com o qual concorda. Ninguém deseja que suas ideias sejam contestadas, ficando distantes de pessoas que não concordam. Em vez disso, as pessoas desejam estar na zona de conforto, convivendo com pessoas do círculo imediato que sejam fáceis de lidar. Essa premissa é utilizada na construção de vários algoritmos através da ênfase em funções de similaridade (PARISER, 2011). Passe, Drake e Mayger (2018) chamam tal efeito de homofilia, o comportamento humano no qual os indivíduos tendem a formar laços com outros indivíduos que compartilham suas crenças e interesses.

Vicario et al. (2016) afirmam que os usuários tendem a se agregar em comunidades de interesse, o que causa reforço e promove viés de confirmação, segregação e polarização. Esse é um fator que impulsiona a formação de bolhas de filtro e, quando as pessoas são polarizadas ao lado de uma discussão, é mais provável que aceitem falsidades como verdade.

Um dos maiores problemas das bolhas de filtro é que elas são invisíveis, e as pessoas não percebem que estão vendo algo diferente do que qualquer outra pessoa. Isso leva a acreditar que a opinião delas deve estar correta, porque só veem um lado, assumindo que todo mundo está vendo as mesmas coisas (PARISER, 2011). Por conta disso, essa pessoa pode ficar polarizada ao lado de um argumento específico. Quando indivíduos de ambos os lados de uma questão são polarizados e apenas veem seu lado de um argumento, e continuamente veem sua opinião reforçada, uma solução nunca será alcançada porque ambos os lados se recusarão a aceitar o argumento oposto. A maioria das pessoas obtém as notícias de uma fonte, e Panke e Stephens (2018) discutem isso dizendo que liberais e conservadores consistentes, por exemplo, geralmente vivem em mundos separados da mídia e mostram pouca sobreposição nas fontes em que confiam para notícias políticas.

As bolhas de filtro são um problema factível de solução, mas se nenhuma ação for tomada para conscientizar as pessoas sobre as bolhas de filtro e as pessoas não fizerem sua parte para se libertar delas, o resultado poderá ser altamente problemático. As bolhas de filtro têm muitos efeitos negativos que precisam ser reduzidos para que as pessoas se entendam e a democracia prospere. Vai ser difícil, porque essas bolhas de filtro estão enraizadas na natureza humana. Ao conscientizar as pessoas sobre as bolhas de filtro e como estamos sendo manipuladas por elas poderemos rompê-las, consumindo informa-

ções de vários sites confiáveis e buscando vários lados dos argumentos. É importante que as pessoas consumam conteúdos que desafiem seus pontos de vista (PARISER, 2011).

Tecnicamente falando, a literatura sugere algumas estratégias para lidar com o problema da bolha de filtro. A primeira delas é construir algoritmos que sejam orientados a diversidade. Helberger, Karppinen e D'Acunto (2018), por exemplo, sugerem a diversidade como um princípio de concepção para sistemas de recomendação na intenção de romper potenciais bolhas de filtro. A segunda estratégia está atrelada à construção de interfaces gráficas que promovam a exposição de diversidade ao usuário. Nagulendra e Vassileva (2014), Plettenberg et al. (2020) mostram que a visualização interativa aumenta a percepção do usuário da existência da bolha e o entendimento do mecanismo de filtragem.

2.3 Diversificação de recomendação

Os SR não geram recomendações diversas por natureza. Os motivos estão relacionados à heurística das técnicas de recomendação tradicionais e às métricas de avaliação atuais, que ignoram a diversidade e avaliam uma lista de recomendações como um agregado dos itens individuais. De forma geral, a diversificação de recomendação é um problema NP-Completo que envolve dois objetivos conflitantes (*trade-off*): a qualidade (em termos de acurácia) e a diversidade das recomendações (GALWAY, 2015). A intenção da diversificação é colocar o usuário frente a novos itens, que tenham um impacto positivo na sua satisfação, mantendo bons níveis de acurácia.

Galway (2015) define formalmente o problema da diversificação de recomendação como: a tarefa de selecionar um subconjunto de k itens pertencentes a um conjunto maior \mathcal{S} , em tempo polinomial, que maximize uma função objetivo que considera a qualidade e a diversidade dos k elementos. As abordagens de diversificação costumam utilizar algoritmos gulosos para otimizar uma função objetivo, utilizada para controlar o *trade-off* entre as dimensões consideradas.

A diversificação pode ser atingida em diferentes níveis. Ge, Gedikli e Jannach (2011) apresentam uma interessante classificação dos níveis de diversidade. A diversidade inerente é aquela calculada entre os itens de recomendação podendo ser classificada, ainda, em diversidade individual e diversidade agregada. A diversidade individual, também conhecida como diversidade intra-lista, é relacionada à diversidade da lista de recomendação de um usuário enquanto a diversidade agregada, também denominada di-

versidade inter-lista, se refere à diversidade considerando todos os usuários. Por outro lado, a diversidade percebida refere-se à diversidade experienciada pelo usuário e pode ser dividida em diversidade atual percebida e diversidade temporal percebida. A primeira refere-se à diversidade percebida por um usuário em um dado momento e, a segunda, à diversidade percebida pelo usuário em um período de tempo. A principal vantagem entre a diversidade percebida e a inerente é que a primeira pode capturar a opinião do usuário frente o fator de qualidade diversidade. A Figura 2.4 apresenta uma visão taxonômica dos níveis de diversidade explicados anteriormente.

Figura 2.4: Níveis de diversidade.



Fonte: O Autor

As abordagens de diversificação de recomendação costumam enquadrarem-se em duas categorias: aprimoramento de algoritmos tradicionais de recomendação; e pós-filtragem. As abordagens pertencentes à primeira categoria modificam algoritmos de recomendação conhecidos para que gerem listas de recomendação já diversificadas. As abordagens baseadas em pós-filtragem recebem como entrada uma lista de recomendação gerada por um algoritmo de recomendação convencional e geram uma lista final diversificada (GALWAY, 2015). As abordagens pertencentes a essa última categoria são as mais difundidas na literatura, tendo vantagens significativas como a independência do algoritmo de recomendação sendo utilizado (Adomavicius; Tuzhilin, 2005). Por esse motivo, são apresentadas a seguir duas abordagens de pós-filtragem.

2.3.1 Máxima relevância marginal

O algoritmo MMR (*Maximal Marginal Relevance*) foi inicialmente proposto para a diversificação de resultados de busca, na área de recuperação de informações, por Carbonell e Goldstein (1998). Mais tarde foi adaptado por Vargas (2012) para o contexto de recomendação. De acordo com Noia et al. (2017), o MMR diversifica implicitamente

uma lista considerando uma troca entre a relevância de um item e sua quantidade de novas informações fornecidas com relação aos itens selecionados anteriormente. Por implícito, entende-se que o algoritmo objetiva aumentar a distância média $((1 - sim(i, j)))$ entre os pares de itens na lista de recomendação. É um algoritmo de *reranking* guloso que consiste na **maximização** de uma função objetivo. O elemento de diversidade da função objetivo é determinado entre a combinação linear entre o *rating* e a similaridade máxima da nova lista, calculada utilizando as *features* compartilhadas dos itens. Formalmente, a função objetivo do MMR é definida como:

$$MMR(i) = \lambda \times r^*(i) + (1 - \lambda) \times avg_{j \in S} (1 - sim(i, j)) \quad (2.14)$$

Em que r^* representa o *rating* do item i e λ permite controlar o grau de diversidade da nova lista. Quando menor o valor de λ maior será a diversidade intra-lista. S representa o conjunto de itens já diversificados.

2.3.2 Diversificação de tópicos

O diversificador de tópicos (*Topic Diversification* - TD) foi um dos primeiros algoritmos de diversificação concebido para os Sistemas de Recomendação, apesar de ter sido inspirado no algoritmo MMR apresentado anteriormente. Ele foi proposto por Ziegler et al. (2005) e não considera somente a dissimilaridade dos atributos dos itens do usuário mas também a árvore taxonômica, a fim de expandir e generalizar o interesse do usuário em relação aos atributos dos itens recomendados. Iterativamente, depois de encontrar a dissimilaridade entre os itens e gerar uma lista de recomendações altamente diversificada, o algoritmo possui um método para mesclar a lista original com a nova lista diversificada a partir do fator de diversificação θ . A seleção do item que será adicionado à nova lista de recomendações consiste na **minimização** da seguinte função objetivo:

$$topic_div(i) = (1 - \theta) \times pos(rank_orig_i) + \theta \times pos(rank_inv_i) \quad (2.15)$$

Em que $pos(rank_orig_i)$ representa a posição do item na lista de recomendação original; e $pos(rank_inv_i)$ representa a posição do item na lista diversificada candidata que é criada em ordem inversa ao valor de similaridade dos itens $rank_inv$). Quanto maior o valor de θ , maior será a diversidade dos itens presentes na lista final *top-N*. Esta

é uma abordagem que trabalha a diversidade no nível intra-lista. Por isso, para avaliar o nível de diversidade da lista, os autores propuseram a métrica ILS (*Intra-List Similarity*) na qual maiores valores indicam menor diversidade.

2.4 Análise de sentimentos

A análise de sentimentos pode ser uma aliada poderosa aos SR como uma ferramenta para a descoberta de informações que contribuam para o processo de recomendação. Algumas dessas informações são sentimentos e opiniões expressas por usuários na forma de texto que podem expressar seu grau de satisfação ou insatisfação acerca de um determinado item. Exemplos disso compreendem comentários sobre itens como produtos, filmes, notícias, as próprias notícias em si, vídeos, dentre outros (HUNG, 2020).

O principal objetivo da análise de sentimentos é definir técnicas automáticas capazes de extrair informações subjetivas de textos em linguagem natural, como opiniões e sentimentos, na intenção de gerar conhecimento estruturado que possa ser utilizado por um sistema e que, por sua vez, venha auxiliar processos de tomada de decisão. A análise de sentimentos é uma das áreas de pesquisa em Processamento de Linguagem Natural (PLN) mais ativas desde o início dos anos 2000 (LIU, 2010).

Benevenuto, Ribeiro e Araújo (2015) salientam que a análise de sentimentos tem se desenvolvido cada vez mais, principalmente com o advento das redes sociais nas quais são expressas várias formas de sentimento. Isso traz uma série de conceitos que causam confusão a muitos pesquisadores. Por isso, os autores apresentam um esclarecimento dos conceitos principais envolvidos nesta área, conforme segue.

Polaridade: Representa o grau de positividade e negatividade de um texto. Alguns métodos tratam a polaridade como um resultado discreto binário (positivo ou negativo) ou ternário (positivo, negativo ou neutro). Por exemplo, a frase “Como você está bonita hoje” é positiva, a frase “Hoje é um péssimo dia” é negativa, e a frase “Hoje é 21 de Outubro” não possui polaridade e normalmente é classificada como neutra.

Sentimento: Indica um sentimento específico presente em uma mensagem (ex.: raiva, surpresa, felicidade, etc.). Alguns métodos apresentam abordagens capazes de identificar qual sentimento em específico uma sentença representa.

Força do sentimento: Representa a intensidade de um sentimento. Normalmente é um ponto flutuante entre (-1 e 1) ou até entre $-\infty$ e $+\infty$, muitas vezes tornando necessário o uso de um *threshold* para identificar a neutralidade de uma sentença.

Subjetividade e Objetividade: Uma sentença objetiva possui normalmente um fato ou uma informação, enquanto sentenças subjetivas expressam sentimentos pessoais e opiniões. Algumas técnicas utilizam a análise da objetividade para estimar se compensa realizar a análise de sentimentos. Portanto entender se um conjunto de dados possui mais sentenças objetivas ou subjetivas pode influenciar diretamente os resultados. Cabe ressaltar que textos informais por exemplo, coletados de redes sociais, tendem a ser mais subjetivos que textos formais como os de notícias.

Benevenuto, Ribeiro e Araújo (2015) ainda apresentam uma interessante divisão dos níveis de granularidade envolvidos nas tarefas de análise de sentimentos salientando que, quanto menor for a granularidade, mais específica se torna a classificação.

Nível de Documento: a classificação de sentimentos ocorre com a análise de todo um texto. Nesse nível, assume-se que todo o texto está relacionado a um único assunto que possui certa polaridade. Se no documento existirem várias entidades com opiniões diferentes, então seus sentimentos podem ser diferentes. Assim, é difícil assimilar um sentimento ao documento todo, mas um caso interessante em que a análise em nível de documento pode ser utilizado é em avaliações de produtos por exemplo Liu (2010).

Nível de Sentença: um único documento pode conter múltiplas opiniões ou mesmo entidades. Neste caso é assumido que o texto foi dividido em frases ou sentenças que possam conter uma opinião individualmente.

Nível de Palavra: os trabalhos focam neste nível buscam otimizar os Léxicos de sentimentos existentes na literatura. Não é a melhor maneira de se construir um dicionário de sentimentos. No entanto, existem diversos dicionários e suas principais diferenças são constituídas pelas palavras que os formam e às vezes na adição de gírias. A inclusão de diferentes termos é importante para alcançar melhor desempenho quando se trabalha com o foco em mídias sociais.

Nível de Aspecto: uma sentença pode ser julgada por várias entidades e pode conter múltiplos sentimentos associados a ela. Por exemplo, a sentença “Esse hotel, apesar de possuir um ótimo quarto, tem um atendimento péssimo!” possui duas diferentes polaridades associadas a “quarto” e “atendimento” para o mesmo hotel. Enquanto “quarto” é considerado positivo, “atendimento” pode ser analisado de forma negativa. Esta necessidade de avaliar a opinião para cada entidade é comum em avaliações de produtos ou em fóruns de discussão.

Os métodos de análise de sentimentos são categorizados, de forma geral, em duas classes: dicionário e aprendizado de máquina. A abordagem baseada em dicionário é

considerada a maneira mais simples de realizar análise de sentimentos. Dicionários como *wordNet*, *SentiwordNet* estão disponíveis publicamente. O dicionário pode ser criado com um conjunto de palavras de sentimento. A iteração do algoritmo termina quando não há novas palavras para adicionar ao dicionário (Abirami; Gayathri, 2017).

2.5 Reconhecimento de entidades nomeadas

O REN é considerado uma sub-tarefa pertencente a um conjunto de tarefas que servem para a extração de informação. O REN consiste em processar um texto e identificar as ocorrências de palavras ou expressões pertencentes à categorias de entidades nomeadas (MIKHEEV; MOENS; GROVER, 1999). As entidades nomeadas correspondem a todas as entidades que podem ser identificadas por um nome próprio, como pessoas, organizações, lugares, marcas, locais, produtos, datas, moedas, dentre outros (NADEAU; SEKINE, 2007). Por exemplo, a seguinte manchete de notícia: “Jair Messias Bolsonaro, do PSL, foi eleito presidente do Brasil em 2018” apresenta as seguintes entidades: “Jair Messias Bolsonaro” é uma pessoa; “PSL” é um partido político, uma organização; “Brasil” é um país, uma localização; “2018” é um marco temporal, neste caso, um ano.

Um SR poderia se beneficiar do REN para a extração de outras dimensões de informação para caracterizar os itens. Por exemplo, as entidades nomeadas em um livro, ao invés de seu conteúdo na íntegra, podem fornecer uma visão mais objetiva da representação do conteúdo. Um usuário lendo um texto biográfico possivelmente estaria interessado nos protagonistas do livro, ou seja, um ou mais personagens principais que poderiam ser do interesse do leitor. Pode-se considerar que o leitor também pode estar interessado em ler outros livros nos quais algumas das pessoas no livro atual se destacam (SARIKI; G et al., 2018). Essa ideia também é adequada para notícias nas quais muitas entidades são citadas e podem ser exploradas para representar as preferências de um usuário em relação à essas entidades (KARIMI; JANNACH; JUGOVAC, 2018).

Uma das técnicas para o REN envolve a utilização de regras escritas manualmente que indicam a existência de uma entidade dado um determinado contexto. Entretanto, a escrita manual de regras tem a desvantagem de ser tediosa, pois necessita da escrita de muitas regras que precisam cobrir os mais diversos cenários possíveis, sendo praticamente impossíveis de serem descobertos manualmente. A partir disso, surgiram as técnicas de aprendizado automatizado. Essas técnicas são divididas duas categorias: aprendizado supervisionado e aprendizado não supervisionado (SEKINE; RANCHHOD, 2009).

O aprendizado supervisionado tem sido a técnica mais utilizada no REN. A ideia dessa técnica consiste em extrair de uma coleção de treinamento as características necessárias para a correta classificação das entidades. Inicialmente um ou mais especialistas avaliam um conjunto de textos e, para cada texto, os especialistas anotam manualmente as suas entidades nomeadas de forma a criar exemplos. Em seguida, o conjunto de textos anotados é dividido em um conjunto de treino e de teste. O conjunto de treino é utilizado para transferir o conhecimento do especialista para o sistema de aprendizado. O conjunto de teste é utilizado para validar a generalização do aprendizado obtido pelo sistema por meio do conjunto de treino. Alguns dos algoritmos que são empregados no aprendizado supervisionado são: *Hidden Markov Models* (HMM), árvores de decisão, Modelos de máxima entropia, *Support Vector Machines* (SVM), e *Conditional Random Fields* (CRF) (NADEAU; SEKINE, 2007).

No aprendizado não supervisionado não existe um conjunto de exemplos para treinamento. Ao invés disso, a abordagem comumente empregada nesse tipo de aprendizado envolve a utilização de algoritmos de agrupamento (NADEAU; SEKINE, 2007). Por exemplo, Alfonseca e Manandhar (2002) rotularam uma palavra de entrada com um tipo de entidade apropriado. Os tipos de entidade são retirados do *WordNet* (por exemplo, local = país, animado = pessoa, animado = animal, etc.). A abordagem atribui uma assinatura da categoria listando palavras que frequentemente co-ocorrem com ela. Em seguida, dada uma palavra de entrada em um determinado documento, o contexto da palavra (palavras que aparecem em uma janela de tamanho fixo ao redor da palavra de entrada) é comparado às assinaturas de tipo e classificado na mais semelhante.

2.6 Resumo do capítulo

Neste capítulo, foram apresentados os principais fundamentos das áreas que escoram a abordagem desenvolvida nesta tese, sendo elas: Sistemas de Recomendação e o domínio de itens de notícias, Bolhas de Filtro, análise de sentimentos e reconhecimento de entidades nomeadas. A diversificação baseada em pós-filtragem é ser explorada através da investigação de múltiplos atributos a fim de potencializar a qualidade de seus resultados e, assim, possibilitar uma suavização do efeito de bolha de filtro. A extração e a estruturação desses multi-atributos são o elo entre a área de recomendação e as áreas de reconhecimento de entidades nomeadas (REN) e de análise de sentimentos. É através dessas duas áreas que a abordagem desta tese é fundamentada e implementada.

3 TRABALHOS RELACIONADOS

O levantamento dos trabalhos relacionados a esta pesquisa foi realizado a partir de uma adaptação do processo de revisão sistemática abordado por Kitchenham e Brereton (2013). De forma geral, tal processo envolve a definição do escopo da pesquisa seguida de uma “*string*” de busca, motores de busca, critérios de inclusão e exclusão, e extração de informações dos artigos de forma estruturada. Com isso, tem-se menos chances de introduzir viés do pesquisador durante as análises.

Nesse sentido, considerando a questão de pesquisa: “Uma estratégia de diversificação de recomendação é capaz de romper a Bolha de Filtro?”, o escopo de pesquisa foi definido a partir das seguintes palavras-chave: Recomendação, Bolhas de Filtro e Diversificação. A partir daí, formulou-se a seguinte “*string*” de busca (*recommend* AND bubble AND divers**), submetida a três motores de busca: *Scopus*, *Web of Science* e *ACM DL*. A pesquisa foi realizada em janeiro de 2021, considerando os seguintes metadados: título, resumo e palavras-chave. Como resultado, obteve-se, respectivamente, 83, 71 e 26 artigos de *journals* e conferências importantes. Dos 180 artigos retornados, foram removidas duplicatas e artigos fora da área de computação, o que resultou em 107 artigos. Cabe destacar que outras bases não foram consideradas pelo fato dos motores de busca aqui adotados também as indexarem.

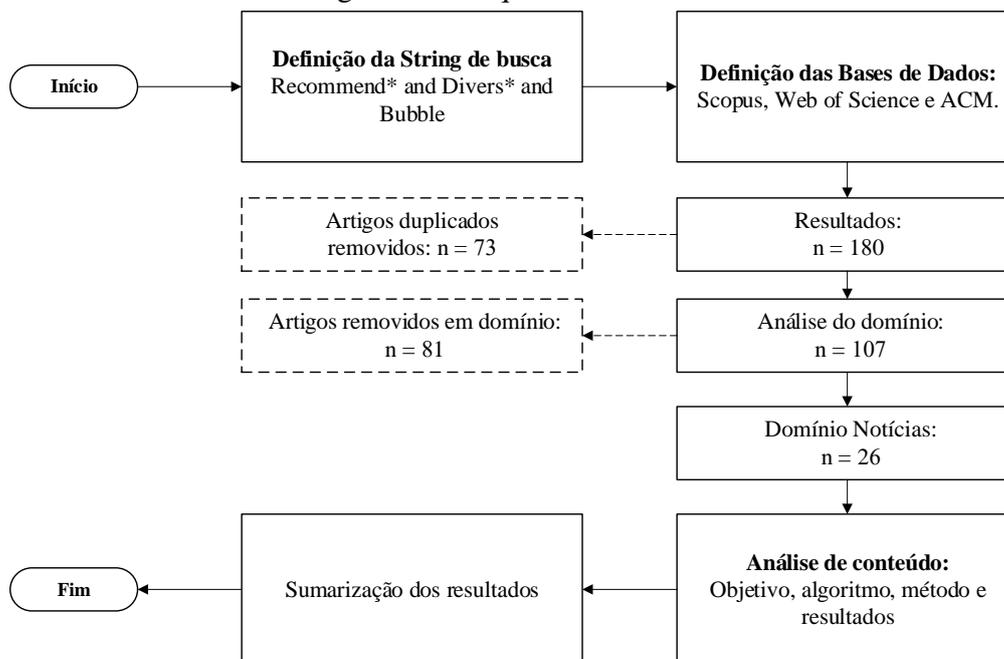
Dentre os 107 artigos restantes, foi aplicado um critério de exclusão referente ao não pertencimento ao domínio de aplicação, notícias, por meio da leitura do título, resumo e palavras-chave. Com isso, restaram 26 artigos levando em consideração o problema de bolha de filtro. O intervalo anual desse conjunto varia de janeiro/2013 a dezembro/2020, sendo o maior número de trabalhos distribuído em 2018 e 2020. Por esse motivo, não foram aplicados critérios de exclusão referentes ao ano de publicação. Vale destacar, ainda, que para o ano de 2020, foi retornado o trabalho do autor desta tese, Lunardi et al. (2020), no qual foi publicada a métrica para medição do efeito de bolha de filtro. Detalhes sobre esse trabalho podem ser conferidos na Seção 6.3.

Por fim, a análise de conteúdo dos 26 artigos foi realizada a partir da leitura, na íntegra, do objetivo, método e resultados, utilizando formulários para o registro qualitativo das informações. Após essa leitura aprofundada, 12 trabalhos foram considerados relevantes para esta pesquisa. Os demais 14 trabalhos, apesar de satisfazerem os critérios de busca, foram desconsiderados por diferentes razões como, por exemplo, a superficialidade das informações apresentadas. A Figura 3.1 ilustra o fluxo do trabalho de seleção

dos trabalhos relacionados.

A apresentação da análise qualitativa dos 12 trabalhos considerados relevantes foi realizada destacando as palavras: objetivo, algoritmo e métrica. Também, é apresentada uma análise comparativa de cada trabalho com a abordagem desenvolvida nesta tese.

Figura 3.1: Esquema de revisão.



Fonte: O Autor

Möller et al. (2018) **objetivam** entender os efeitos que diferentes estratégias de recomendação tem na diversidade de conteúdo. Para avaliar a diversidade dos itens, utilizam diferentes características: tópicos (LDA; categoria e tag; tom (polaridade e pela subjetividade); e taxa de relevância do conteúdo político. A partir disso, é utilizada a entropia de *Shannon* como *métrica* para avaliar a diversidade de conteúdo. A **avaliação** se dá através da implementação dos **algoritmos**: Filtragem colaborativa e semântica baseadas em itens (não personalizados) e levando em consideração o histórico do usuário (personalizados); e um *baseline* aleatório que considera que cada artigo de notícia tem igual probabilidade de ser recomendado. Ao todo são sete algoritmos que geram recomendações de notícias de uma única fonte (*Valkstrant* - jornal Holandês de posição política central). Os dados de recomendação são simulados, isto é, são gerados utilizando cada uma das características propostas separadamente e sem considerar que usuários de fato interagiram com elas. Em suma, os resultados mostram que as recomendações personalizadas não mostraram redução na diversidade frente às recomendações não personalizadas; pelo contrário, a

filtragem colaborativa personalizada produziu a maior quantidade de diversidade de tópicos nesse ambiente controlado. **Comparação:** Este artigo não trata diretamente de uma abordagem de diversificação de recomendação, mas sim da análise da diversidade de conteúdo. Mesmo assim, foi considerado por discutir indícios importantes em relação às características dos itens. O *dataset* possui conteúdo na língua alemã e os dados de recomendação são simulados, ao passo que, nesta tese, os dados são reais. Bolha de filtro não é abordada e as notícias consideradas pertencem a um único portal.

Gao, Do e Fu (2018) **objetivam** aprimorar a conscientização de diversas opiniões sociais, fornecendo dicas visuais e recomendações de opiniões (por exemplo, notícias e comentários) de diferentes lados com diferentes indicadores (emoções). Um sistema é **avaliado** utilizando artigos de notícias da CNN e FoxNews sobre a questão da revogação de Obamacare e seus comentários de usuários correspondentes do Facebook. Os resultados demonstram que o sistema pode aumentar a conscientização das pessoas sobre suas posições e preferências de seleção de opinião, o que diminui a exposição seletiva e, portanto, leva a uma percepção mais equilibrada das opiniões sociais. **Comparação:** Este é um trabalho que não tem como foco a recomendação nem a diversificação. Entretanto, explora recursos de interface para promover a entrega de conteúdo de lados opostos e a percepção do usuário da polarização de opiniões. A bolha de filtro está implícita na discussão, mas não é medida explicitamente.

Bountouridis et al. (2019) **propõem** um ferramenta de simulação chamada SIREN para avaliar os efeitos dos sistemas de recomendação no domínio de notícias, permitindo que os provedores de conteúdo (editores de jornais) selecionem e parametrizem diferentes algoritmos de recomendação para visualizar seus efeitos de acordo com duas métricas de diversidade. O efeito de bolha de filtro procura-se medir através de uma **métrica** de diversidade em relação aos itens mais ao final das listas de recomendação. A ferramenta é **avaliada** fazendo uso de notícias do portal *BBC News*, com uma taxa de 100 notícias por dia e com diferentes parâmetros para diferentes **algoritmos**: filtragem colaborativa (*User-KNN*, *Item-KNN*), *Random*, *MostPopular* e *WeightedBPRMF*. Os resultados sugerem que os algoritmos não levam à falta de diversidade de conteúdo, mas que isso se deve mais à mudança das preferências dos usuários ao longo do tempo. **Comparação:** Não abordam a diversificação de recomendação, tão pouco fazem menção em relação a caracterização de conteúdo de notícias. A medição de bolha fica implícita à métrica de diversidade, significando que baixa diversidade corresponde ao surgimento do efeito. Os dados utilizados são simulados, portanto é sensível generalizar os resultados apontados.

Sonoda, Seki e Toriumi (2019) **analisam** os fatores que afetam as mudanças comportamentais de usuários em serviços de notícias. Para isso, basearam as análises em dados do portal Gunosy, verificando a interação dos usuários com as notícias (comportamento), mais especificamente as categorias de artigos consumidos. A mudança na diversidade das categorias de um artigo é avaliada com o comportamento do usuário pela diversidade de categorias às quais o artigo que está sendo lido está atribuído. Os artigos pesquisados foram avaliados de cada usuário com base no grupo ao qual o artigo pertence e a diversidade do comportamento de navegação com base no grau de concentração das categorias de artigos lidos, calculado usando a entropia **métrica** de informações. A mudança na diversidade é avaliada com os dois indicadores da entropia da categoria e da entropia da categoria de *ranking* para diferentes períodos de tempo. De forma geral, os resultados confirmam o decréscimo da diversidade de conteúdo ao longo do tempo. **Comparação:** Este é um artigo que concentra seus esforços na análise de diversidade de conteúdo, e não em algoritmos de recomendação e/ou diversificação. Apesar disso, fornece noções interessantes sobre medição de diversidade de conteúdo que podem ser exploradas e aproveitadas. A noção de bolha de filtro não é explicitamente abordada.

Nechushtai e Lewis (2019) **analisam** o quanto os sistemas de recomendação de notícias contribuem para a formação de bolhas de filtro. Para isso, recrutam uma série de pessoas diferentes, utilizando seus perfis pessoais Google, para pesquisarem notícias no *Google News* durante as eleições presidenciais americanas de 2016 sobre *Hilary Clinton e Donald Trump*. Cada participante reportou as 5 primeiras notícias de cada candidato. Usuários com diferentes tendências políticas e de diferentes estados receberam notícias muito semelhantes, reafirmando a suposição de que os algoritmos incentivam as bolhas de filtro. Também foi evidenciado alto grau de homogeneidade e concentração nas recomendações de notícias e que o *Google News* replica mais as estruturas tradicionais da indústria do que as “rompe”. **Comparação:** Apesar de reportar uma análise empírica interessante, não abordam diversificação de recomendação. Ainda, não propõem uma forma de medir bolha de filtro, assumindo a existência de duas bolhas de acordo com cada um dos candidatos políticos. Portanto, ainda que seja um trabalho que menciona recomendação, não contempla o tema como foco de estudo.

Abbar et al. (2013) **objetivam** recomendar, em tempo real, os mais próximos e diversos artigos de notícias em relação a um artigo que o usuário está lendo no momento. Para isso, utilizam os comentários postados por usuários nas notícias para extrair entidades nomeadas e sentimentos que são utilizados como atributos para caracterizar as

notícias. Os autores utilizam isso como base para atingir um balanço entre relevância e diversidade. Como **métricas** para as duas grandezas, definem relevância como a menor distância a partir de um artigo atualmente sendo lido, que é utilizada para encontrar um conjunto de artigos relacionados; e diversidade como a distância par a par entre dois artigos de acordo com o conteúdo gerado pelo usuário na forma de entidades nomeadas e sentimentos contidos nos comentários. A partir disso, tem-se um problema de alta complexidade, especialmente quando a quantidade de artigos relacionados é muito alta, e que afeta a performance da recomendação em tempo real. Os autores propõem uma adaptação das técnicas de *hashing* sensíveis a localidade (LSH), para encontrar os vizinhos mais próximos de um artigo de interesse, para implementar o **algoritmo** dLSH que, por sua vez, busca maximizar a diversidade de um subconjunto de artigos dentro do primeiro conjunto (relevância). A abordagem é **avaliada** por meio de um estudo de usuário utilizando notícias reais dos portais *Al Jazeera English* e *Reuters*. É utilizada plataforma *Amazon Mechanical Turk* em que os usuários avaliam a qualidade das listas de recomendação diversificadas com o uso de entidades nomeadas e sentimentos dos comentários com as diversificadas com base no conteúdo da notícia. Os resultados se mostraram superiores com a utilização do algoritmo e do conjunto de *features* propostos, quando comparado ao diversificador *baseline* MMR. **Comparação:** O conjunto de *features* apresentado é similar ao desta tese para caracterizar cada notícia. Entretanto, os esforços são concentrados nos comentários das notícias e não no seu conteúdo. Apesar de terem sido revelados resultados superiores, destaca-se que o conteúdo de cada notícia pode não ter sido explorado o suficiente o que pode justificar o resultado inferior. Outra diferença significativa é com relação ao efeito de bolha de filtro. Apesar de citá-lo, os autores não se preocupam em medi-lo, especialmente em um domínio tão crítico como o de notícias.

Desarkar e Shinde (2014) **objetivam** recomendar e diversificar notícias para um usuário que não possui um perfil definido sendo, portanto, uma abordagem não personalizada. As recomendações são geradas a partir do tempo de vida e a popularidade da notícia. A diversificação é atingida utilizando entidades nomeadas e tópicos (como características) extraídos de notícias, através de LDA e o Alchemy API, respectivamente. O **algoritmo** diversificador utiliza tais características (aspectos de diversificação) em conjunto com a combinação de duas **métricas** de distância: euclidiana para tópicos e a raiz quadrada de Jaccard para entidades nomeadas. A diversificação é tratada como um problema de otimização bi-critério, isto é, máxima utilidade dos itens recomendados e máxima diferença entre eles. A **avaliação** consiste em gerar recomendação diversificadas e

não diversificadas a partir de um *dataset* Alemão de notícias. A qualidade das recomendações foi medida através do número de cliques em cada notícia e do número de usuários satisfeitos (que leram uma notícia). Um dos resultados é que a diversificação utilizando entidades nomeadas satisfaz melhor os usuários quando comparada com a utilização de tópicos. No geral, recomendações diversificadas mostraram satisfazer melhor os usuários, entretanto, com o aumento do valor da variável de controle da diversificação λ , a satisfação tende a diminuir. **Comparação:** Utilizam entidades nomeadas e tópicos para caracterizar as notícias tal como no trabalho aqui proposto, porém, não consideram sentimentos nem definem outras características. Utilizam essas características como aspectos de diversificação e apresentam resultados interessantes quanto a isso. Não abordam o aspecto de bolha de filtro.

Nguyen et al. (2014) **objetivam** investigar o impacto que as abordagens de recomendação baseadas em filtragem colaborativa tem sobre os usuários em relação ao efeito de bolha de filtro em termos da diversidade de conteúdo no nível individual. Os autores definem duas questões que balizam a metodologia e a sumarização dos resultados: Sistemas de Recomendação expõem os usuários a itens limitados ao longo do tempo? Como a experiência dos usuários que seguem as recomendações diferem daqueles que regularmente não as seguem? Para responder tais questões, os autores propõem uma **avaliação** que separa os usuários de acordo com a frequência esses consomem o conteúdo recomendado. Essa separação permite analisar o grupo de usuários onde é possível uma bolha de filtro e compará-los com um grupo de controle que usa o mesmo sistema, mas não segue regularmente as recomendações. Outra parte da avaliação consiste na proposição de uma **métrica** para explorar mudanças na diversidade dos itens consumidos (avaliados) ao longo do tempo. Essa métrica de diversidade é baseada na distância euclidiana entre os itens que são caracterizados por *genome tags*. O *dataset* utilizado é o *MovieLens*, contendo dados longitudinais sobre filmes como, por exemplo, nota recebida, *timestamp*, e as *genome tags* que os caracterizam. De forma geral, os resultados apontam que a recomendação baseada em filtragem colaborativa limita o espectro do usuário, de uma forma mais lenta quando comparada com o grupo de usuários que não seguiu as recomendações. Entretanto, ao analisar os itens consumidos pelos usuários, constatou-se haver maior diversidade entre eles, o que representa um risco menor da presença de bolha de filtro. **Comparação:** A primeira diferença significativa é com relação ao objetivo que, aqui, não visa diversificar as recomendações, embora os autores sugiram nas discussões como uma estratégia para diminuir a limitação do espectro de itens e, conseqüentemente, mini-

mizar as chances de bolha de filtro. Ainda, não é proposto nenhum **algoritmo**, apenas é realizada uma análise do que os usuários consumiram ou não das recomendações. Essas, por sua vez, foram geradas através de um algoritmo de filtragem colaborativa baseado em vizinhança de itens. Outra diferença que se faz necessário destacar é o domínio de aplicação, filmes, que pode apresentar influências significativas nos resultados apresentados.

Chaney, Stewart e Engelhardt (2018) analisam o impacto dos sistemas de recomendação, especialmente em relação ao *feedback loop*, na homogeneização do comportamento dos usuários. O *feedback loop* acontece quando um modelo é retreinado inúmeras vezes considerando novos dados de *feedback* dos usuários perante os itens a fim de gerar recomendações mais acuradas. Os autores defendem que esse fenômeno tende a degradar a utilidade das recomendações bem como homogeneizar o comportamento dos usuários, levando o usuário para dentro de bolha de filtro. É definida uma **métrica** para medir a homogeneização dos usuários como o índice Jaccard entre os dois conjuntos de itens consumidos pelos dois usuários mais similares, definidos através do pareamento de todos com todos utilizando a similaridade de cossenos. Os autores utilizaram dados simulados para gerar as recomendações com base nas seguintes abordagens (**algoritmo**): conteúdo, fatoração de matrizes, popularidade, aleatório e social. A **avaliação** se deu, primeiramente, gerando recomendação com um único treinamento de cada algoritmo e, em seguida, com treinamentos repetidos. Os resultados mostraram que o *feedback loop* amplia a homogeneização do comportamento do usuário sem ganhos correspondentes em utilidade. **Comparação:** Esse é um trabalho que não aborda a diversificação de recomendação, porém, apresenta importantes direcionamentos para a medição de bolha de filtro. Os autores utilizam dados simulados ao invés de dados reais o que poderia ter levado a resultados diferentes. A bolha de filtro não é explicitamente avaliada ou, então, uma métrica definida para sua medição.

Em Tintarev et al. (2018), o **objetivo** dos autores é diversificar a recomendação baseando-se em vários atributos de notícias. Para isso, caracterizam um conjunto de notícias através dos seguintes atributos: fonte geradora, tópico (extraído com LDA), emoção, gravidade e complexidade do tópico, dentre outras. Utilizam isso, então, junto ao **algoritmo** de diversificação MMR, propondo uma **métrica** baseada em distância euclidiana combinando tais características. A abordagem é **avaliada** por meio de testes *offline* e estudos de usuário. O teste *offline* buscou evidenciar a melhor combinação de valores de parâmetros para cada uma das características e, também, do fator de diversificação λ . Já o estudo de usuário foi realizado com 15 pessoas. Cada uma delas foi exposta a artigos de

um recomendador comum e a artigos de um recomendador diversificado. Os resultados mostraram que os usuários preferiram mais as recomendações do recomendador comum. **Comparação:** Não medem bolha de filtro tão pouco propõem uma forma de verificar a suavização, ou não, de bolha de filtro, apesar de especularem ideias nesse sentido. O estudo de usuário envolve uma amostra muito pequena e que parece não envolver o uso de um sistema, o que leva a certa dúvida em relação aos resultados apresentados. Além disso, apenas um algoritmo de diversificação de recomendação é testado e o conjunto de características não parece ser implementado conforme ideias apresentadas, pois não demonstram os resultados dessa implementação. De forma geral, pode-se dizer que se trata de um artigo especulativo.

Lu, Dumitrache e Graus (2020) têm por **objetivo** avaliar a relação entre recomendações de notícias geradas por algoritmos e por curadoria, isto é, por editores de portais de notícias. O **algoritmo** desenvolvido é orientado a aspectos “além da acurácia”, incorporando **métricas** como a cobertura, a serendipidade, e a diversidade. Essa última, por sua vez, é utilizada como uma estratégia de *re-ranking* para trazer maior dinamismo às diversas listas de recomendação geradas por um usuário ao longo do tempo. A estratégia é **avaliada** através de estudos de usuário, cujos resultados sugerem que as recomendações geradas por algoritmos têm maior diversidade quando comparadas às recomendações editoriais. **Comparação:** Os autores não medem o efeito de bolha de filtro, apesar de o citarem como efeito colateral aos leitores que possivelmente sigam as recomendações dos editores ao invés das recomendações geradas via algoritmo; Não mencionam quais foram os algoritmos de recomendação adotados; Não utilizam características como sentimentos e/ou entidades nomeadas extraídas das notícias.

Gharahighehi e Vens (2020) **objetivam** introduzir diversidade em sistemas de recomendação de notícias baseados em sessão - *Session-based Recommender Systems* (SBRS). Nesse tipo de sistema, os usuários são anônimos, não possuindo um perfil explícito e as recomendações são geradas enquanto a sessão do usuário durar, através das sequências de escolhas que o usuário faz. Os autores adotaram o **algoritmo** SKNN, uma versão do algoritmo KNN para SBRS, que é flexível e simples para introduzir a característica de diversidade por meio do conteúdo de notícias. A abordagem é **avaliada** utilizando três *datasets* de notícias (Globo.com; Roularta, Adressa) a partir dos quais foram extraídas representações de conteúdo, na forma vetorial, para os cálculos de similaridade realizados pelas funções de diversificação inseridas no algoritmo SKNN. As **métricas** empregadas para avaliar as listas de recomendação geradas foram: precisão

top-10 e a diversidade intra-lista (ILD) *top-10*. **Comparação:** Os autores representam o conteúdo das notícias por meio de *word-embeddings*, mas não consideram outras possibilidades. Entretanto, sugerem, como trabalhos futuros, caracterizar as notícias por outros meios como, por exemplo, seus textos, *tags*, sentimento e polaridade, características que muito se aproximam da abordagem desenvolvida nesta tese. O efeito de bolha de filtro é utilizado como motivador do estudo, mas não é medido diretamente, assim como o seu aumento ou diminuição. A avaliação considera apenas dados *offline*.

A Tabela 3.1 apresenta a comparação entre os doze trabalhos considerados relevantes para esta pesquisa. Os trabalhos são comparados utilizando cinco aspectos:

- DV: Diversificação de recomendação;
- BF: Bolha de filtro;
- MT: Métrica para a medição do efeito de bolha de filtro;
- DR: Dados reais utilizados na avaliação;
- MA: Multi-atributos como aspectos de diversificação e/ou como forma de representação de conteúdo dos itens.

Cada um desses aspectos é marcado com o sinal + para quando o trabalho apresenta o aspecto como seu diferencial de pesquisa. O sinal – indica que o trabalho apresenta o aspecto parcialmente. Por exemplo, para o aspecto MA considera-se parcial quando um trabalho não utiliza todos os atributos que são desenvolvidos nesta tese, isto é, tópicos, entidades nomeadas, polaridades, e subjetividades. Quando não há marcação, significa que o trabalho não possui determinado aspecto como diferencial de pesquisa.

Tabela 3.1: Comparação entre os trabalhos relacionados.

Trabalho	DV	BF	MT	DR	MA
Abbar et al. (2013)	+	–		+	–
Desarkar e Shinde (2014)	+			+	–
Nguyen et al. (2014)	+	–		+	
Möller et al. (2018)					–
Gao, Do e Fu (2018)		–		+	–
Chaney, Stewart e Engelhardt (2018)	–	–			
Tintarev et al. (2018)	+	–		+	–
Bountouridis et al. (2019)	–	–			–
Sonoda, Seki e Toriumi (2019)	–			+	
Nechushtai e Lewis (2019)		–		+	
Lu, Dumitrache e Graus (2020)	+	–		+	
Gharahighehi e Vens (2020)	+	–		+	–
Esta tese	+	+	+	+	+

Fonte: O Autor.

Detalhes sobre a parcialidade (–) em relação ao aspecto MA podem ser conferidos na descrição de cada trabalho pela palavra-chave “**comparação**”. Assim, é possível identificar quais foram os atributos que cada trabalho, considerado parcial nesse aspecto, utilizou.

3.1 Resumo do capítulo

Neste capítulo, foram apresentados 12 trabalhos que foram selecionados a partir de um levantamento sistemático da literatura. Ao longo da pesquisa, esse processo foi realizado diversas vezes a fim de manter a consciência em relação às soluções mais atuais e, com isso, contribuir para a evolução do estado atual da literatura.

Nesse sentido, esta tese contribui desenvolvendo um conjunto multi-atributo para caracterizar notícias, aliado a algoritmos de diversificação de recomendação, baseados em pós-filtragem, para suavizar o efeito de bolha de filtro. Assim, um dos diferenciais deste trabalho é a medição do efeito de bolha de filtro. Os trabalhos relacionados avaliam a diversificação por meio de métricas de diversidade como, por exemplo, o grau de entropia de *Shannon* (MÖLLER et al., 2018) ou a métrica *ILD (Intra List Diversity)* (VARGAS, 2014), dentre outras. Entretanto, nesta tese, a diversificação é avaliada através da homogeneização do conjunto de itens do usuário e, com isso, tem-se a vantagem de medir diretamente a formação da bolha de filtro. Essa é uma inovação que pode corroborar com a concepção de ferramentas que auxiliem os usuários na percepção da sua “bolha”, conforme apontado por Helberger, Karppinen e D’Acunto (2018).

Em relação ao conjunto de características multi-atributo proposto e desenvolvido nesta tese, os trabalhos relacionados não são unânimes. Em outras palavras, nenhum trabalho considera tópicos, entidades nomeadas, polaridade, subjetividade e fonte em relação aos textos das notícias. Alguns trabalhos propõem algo parecido, entretanto, para comentários postados junto às notícias. Portanto, esse é outro diferencial central desta tese que pode ser explorado não só para a diversificação de recomendação, ou para a caracterização de itens em uma abordagem baseada em conteúdo mas, também, por profissionais de diferentes áreas que desejarem representar e sumarizar grandes volumes textuais, permitindo extrair informações relevantes de forma automatizada.

4 DIVERSIFICAÇÃO MULTI-ATRIBUTO DE RECOMENDAÇÃO

Um sistema que entrega somente notícias de um determinado assunto/tópico, e que não leva em consideração aspectos como o nível de controvérsia ou sentimento, pode ter consequências como, por exemplo, a polarização de opiniões, a não contestação de ideias, e a desinformação a respeito de diferentes assuntos de interesse.

Uma das apostas para diminuir (romper) o efeito da bolha de filtro é diversificar a recomendação, isto é, promover a dissimilaridade entre os itens da lista e, ao mesmo tempo, manter bons níveis de acurácia. Entretanto, para isso, é necessário conhecer as características dos itens que, sendo notícias, comumente são utilizados tópicos. Todavia, essas características podem não ser suficientes para promover uma diversificação que de fato suavize o efeito de bolha de filtro.

No intuito de exemplificar tal problema, foram selecionados dois itens do *corpus* de notícias coletadas para esta tese (ver Seção 4.1.1). Essas notícias são falsas e possuem conteúdos antagônicos, porém compartilham um mesmo assunto: o candidato Jair Messias Bolsonaro. O fato de duas ou mais notícias compartilharem um mesmo assunto/tópico pode limitar as potencialidades de uma possível diversificação. Entretanto, analisando o conteúdo de cada uma das duas notícias é possível perceber que elas possuem posicionamentos diferentes. A notícia da Figura 4.1(a) apresenta posicionamento positivo em relação ao assunto tratado, ao passo que a notícia da Figura 4.1(b) manifesta um posicionamento negativo em relação ao mesmo tópico.

Portanto, é preciso considerar a adição de outras dimensões de informação ao processo de diversificação de recomendação. Essa ideia norteia, então, a concepção de uma abordagem que inclui a extração de múltiplos atributos, conforme formalização apresentada na Seção 4.1. Vale ressaltar que a busca por uma abordagem de diversificação multi-atributo está alinhada com o tratamento do problema da bolha de filtro apontada por Pariser (2011): “consumir conteúdos que desafiem nossos pontos de vista”. Por isso, espera-se que esta abordagem contribua para a suavização do efeito de bolhas de filtro.

Destaca-se ainda que as notícias são utilizadas neste trabalho como um domínio de aplicação somente, não levando em consideração todos os aspectos da área de recomendação de notícias tais como vida útil da notícia, escolha de notícias do dia colocadas em destaque pelos editores, comentários, etc. O motivo dessa escolha são explicados por dois argumentos. O primeiro, porque permite a realização de um experimento mais simples, que não precisa se preocupar em construir um *feed* de notícias alimentado em tempo

Figura 4.1: Trecho de duas notícias falsas e antagônicas sobre um mesmo assunto.



Fonte: O Autor

real, e que tem a capacidade de avaliar o papel dos algoritmos tradicionais na formação de bolhas. O segundo motivo é que permite a aplicação da abordagem em qualquer outro domínio. Isso abre precedentes para a aplicação da abordagem em qualquer outro domínio em que os itens possuam conteúdo textual.

4.1 Modelagem e implementação da abordagem para a diversificação multi-atributo de recomendação

Define-se \mathcal{N} como sendo o conjunto total de itens de notícias em que cada $n \in \mathcal{N}$ é caracterizada a partir de quatro dimensões vetoriais de informação, a partir de seu conteúdo. A primeira delas é representada por \vec{t} que contém um ou mais tópicos do conjunto \mathcal{T} de tópicos extraídos a partir de \mathcal{N} . A segunda dimensão de informação, denotada por \vec{e} , diz respeito a um conjunto de entidades nomeadas \mathcal{E} , por exemplo, pessoas, localizações, organizações, dentre outras. A terceira dimensão considerada é um valor de polaridade $p \in [-1, 1]$ que define se o conteúdo de n expressa uma opinião positiva ($p > 0$), negativa ($p < 0$) ou neutra ($p = 0$). A quarta dimensão tem a ver com a subjetividade de um texto, expressa através de um valor $s \in [0, 1]$. Quanto mais próximo do valor 1, mais subjetivo é considerado o texto. Por outro lado, quanto mais próximo do valor 0, mais objetivo, significando que apresenta pouca ou nenhuma opinião pessoal, e informações factuais.

Dessa forma, duas notícias sobre uma eleição presidencial do Brasil, por exemplo, podem ser semelhantes ao compartilharem o mesmo assunto (Política), local (Brasil), e pessoa (candidato X), mas ainda diferentes quando considerado a um valor de polaridade e/ou subjetividade. A Equação (4.1) sumariza a caracterização de cada notícia n .

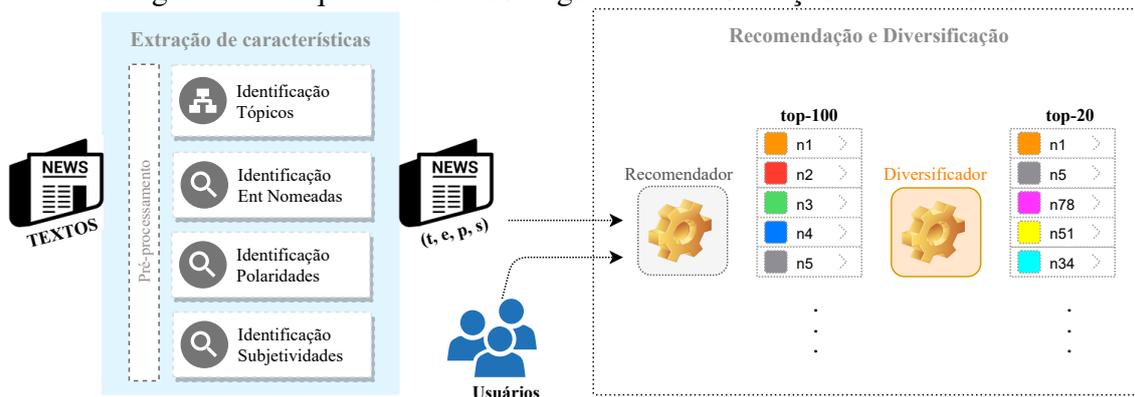
$$n = [\vec{t}, \vec{e}, p, s] \quad (4.1)$$

A abordagem de diversificação de recomendação desta tese envolve dois componentes principais. O primeiro deles é o conjunto de múltiplos atributos que são necessários para a operação do segundo componente: o algoritmo de diversificação de recomendação baseado em pós-filtragem. Nesse sentido, dois algoritmos da literatura foram adotados, o MMR e o TD, apresentados na Capítulo 2, tendo em vista que são considerados seminais e amplamente aceitos pela comunidade de sistemas de recomendação. A combinação entre o conjunto de múltiplos atributos e cada um dos algoritmos adotados foi extensamente testada conforme apresentado no próximo capítulo.

Esses diversificadores funcionam de forma semelhante (VARGAS, 2014), porém divergem pelas suas funções objetivo, responsáveis por selecionar os itens a serem incluídos na lista final diversificada de acordo com o fator de diversificação θ (TD) ou λ (MMR). De forma geral, eles recebem uma lista de recomendação como entrada $\mathcal{S} = \{\vec{n}_1, \vec{n}_2, \dots, \vec{n}_n\}$, gerada por uma abordagem comum de recomendação como, por exemplo, baseada em conteúdo, filtragem colaborativa, dentre outras, ordenada de acordo com um valor estimado de relevância. A partir dela, então, constrói-se uma lista *top-K* diversificada $\mathcal{S}_* \subset \mathcal{S}$, contendo itens menos similares entre si. Por esse motivo, \mathcal{S} deve ser, pelo menos, cinco vezes maior que o tamanho pretendido k para \mathcal{S}_* (ZIEGLER et al., 2005). A nova lista de recomendação \mathcal{S}_* é inicializada com o primeiro elemento de \mathcal{S} com o objetivo de manter um bom nível de acurácia para a nova lista e de obter a base para o cálculo da similaridade entre o primeiro item da lista com os itens candidatos \mathcal{C} advindos da lista de recomendações base \mathcal{S} .

A Figura 4.2 apresenta uma visão arquitetural da abordagem, considerando todas as etapas, isto é, desde o processamento dos textos das notícias, em busca da representação formalizada pela Equação (4.1), até a geração das recomendações diversificadas. Na etapa de “Recomendação e Diversificação” fica ilustrada a ideia da diversificação em que, a partir de uma lista ligeiramente maior (*top-100*) gerada por algum algoritmo comum de recomendação, um diversificador baseado em pós-filtragem constrói uma lista diversificada menor (*top-20*).

Figura 4.2: Arquitetura da abordagem de diversificação multi-atributo.



Fonte: O Autor

A fim de atingir tal representação, são apresentados a seguir uma série de processos, apoiando-se na área de PLN e em uma base de dados de notícias. Essa representação resultou em um conjunto de múltiplos atributos que foram empregados na condução de um estudo de usuário, envolvendo uma aplicação *web* de recomendação diversificada de notícias. A próximas quatro subseções se apresentam de forma técnica, buscando evidenciar os detalhes e escolhas de cada etapa a fim de corroborar para a reprodutibilidade da pesquisa em outros domínios em relação à extração de múltiplos atributos.

4.1.1 Obtenção de notícias

O conjunto de notícias utilizado neste experimento é referente ao período das eleições presidenciais brasileiras de 2018, sendo composto de notícias verdadeiras e falsas. Essa escolha teve como principal motivação a ampla disseminação de notícias falsas e de conteúdo controverso, especialmente através das redes sociais durante o referido período eleitoral. A construção desse conjunto ocorreu a partir seleção de alguns portais de notícias que mantivessem atividade no Facebook e que possuíssem maior número de seguidores (mais populares). A Tabela 4.1 mostra o ranking dos portais de acordo com o critério mencionado a partir da qual é possível perceber que G1¹ e R7² são os portais mais famosos e, portanto, foram os escolhidos como fonte para a obtenção de notícias verdadeiras.

Os demais portais, apesar de interessantes pelo fato de apresentarem vieses políticos diferentes, não puderam ser considerados já que o Facebook no início do ano de 2018

¹<https://g1.globo.com/>

²<https://www.r7.com/>

veiculou uma notícia divulgando que mudaria a sistemática de entrega de conteúdo dos seus algoritmos. Essa, por sua vez, priorizaria conteúdos de amigos e não mais de páginas tais como as mantidas pelos portais de notícias (CUSTÓDIO, 2018). Sendo assim, outros portais populares como Quebrando o Tabu, UOL, Estadão, BBC News Brasil, Yahoo Brazil e MSN Brasil deixaram de postar conteúdo.

Como fonte de notícias falsas verificadas foi escolhida a Agência LUPA³, em função de: (i) ter sido o primeiro portal de checagem de fatos (*fact-checking*) do Brasil; (ii) ter alta popularidade; e (iii) ter concedido permissão para uso do conteúdo. Vale destacar que a Agência LUPA, apesar de pertencer ao grupo UOL, manteve sua página ativa no Facebook dada a sua importância.

Tabela 4.1: Ranking de portais de notícias brasileiros.

Portal	Seguidores
R7	13,5 Milhões
G1	10,6 Milhões
Quebrando o Tabu	9,9 Milhões
UOL	8,2 Milhões
Estadão	3,7 Milhões
BBC News Brasil	3,1 Milhões
Yahoo Brasil	2,9 Milhões
MSN Brasil	2,7 Milhões

Fonte: O Autor.

Os *scrappers* foram desenvolvidos em linguagem Python com o auxílio da biblioteca *BeautifulSoup*⁴ para percorrer cada uma das páginas web dos portais de notícias selecionados. De forma geral, foram coletadas notícias que contivessem as seguintes informações: título, imagem, e conteúdo (HTML). Da Agência LUPA ainda foi coletado o título original, isto é, o título da notícia falsa. O título, no caso desse portal, refere-se à chamada da notícia desmentida, essencial para não perpetuar conteúdo falso através do experimento.

A coleta compreendeu o início do mês de julho até o fim do mês de outubro de 2018 quando, de fato, ocorreram as votações incluindo o 2º turno. Analisou-se a organização da paginação das notícias de cada portal através da URL. No portal R7, por exemplo, uma URL base⁵ foi utilizada e acrescida do intervalo de páginas (7 a 56) que compreendia o período de meses buscado. Na Agência LUPA foi utilizada a mesma sistemática

³<https://piaui.folha.uol.com.br/lupa/>

⁴<https://www.crummy.com/software/BeautifulSoup/>

⁵https://noticias.r7.com/eleicoes-2018/noticias?mobile_cookie=true&page=

(páginas 6 a 14), porém com uma URL base⁶ específica que possuía somente conteúdo veiculado no Facebook, já que esse portal verifica conteúdo não só do Facebook mas também de diversas outras fontes.

Com o portal G1 a mesma estratégia foi adotada, entretanto, com maiores dificuldades já que esse portal utiliza *lazy loading* nas suas páginas, isto é, o conteúdo de cada notícia só é carregado se o acesso for realizado através do navegador. Para contornar esse problema, foi utilizado o *Selenium webdriver* para o navegador *Chromium*, de código aberto, que possibilitou ao *script* interagir com as páginas como se fosse um usuário navegando. Dessa forma, foi possível, através da URL base⁷ (páginas 11 a 201) mais a extensão ".ghtml", perfazer a coleta.

Durante o processo de coleta, cada *scraper* foi responsável por criar um arquivo de comandos *inserts* SQL, contendo todas as notícias coletadas, e de acordo com a estrutura do banco de dados desenvolvido, detalhado no Capítulo 5. Com isso, foi possível popular o banco de dados com as notícias coletadas de forma automática e facilitada. Ao todo, foram coletadas 1938 notícias. A Tabela 4.2 sumariza o total de notícias coletadas em cada um dos portais.

Tabela 4.2: Quantidade de notícias coletadas de cada portal.

Portal	Veracidade	Quantidade
R7	verdadeira	959
G1	verdadeira	901
Agência LUPA	falsa	78

Fonte: O Autor.

4.1.2 Tópicos

A *web* está cada vez mais repleta de todo tipo de informação - na forma de notícias, blogs, páginas da web, artigos científicos, livros, imagens, som, vídeo e redes sociais - e, com isso, torna-se mais difícil encontrar o que se procura. Assim, surge a necessidade de se empregar ferramentas computacionais para ajudar a organizar, pesquisar e compreender todas as informações (M Blei, 2011).

Uma dessas ferramentas é o algoritmo LDA (*Latent Dirichlet Allocation*), proposto por Blei, Ng e Jordan (2003), para a modelagem automática de tópicos latentes,

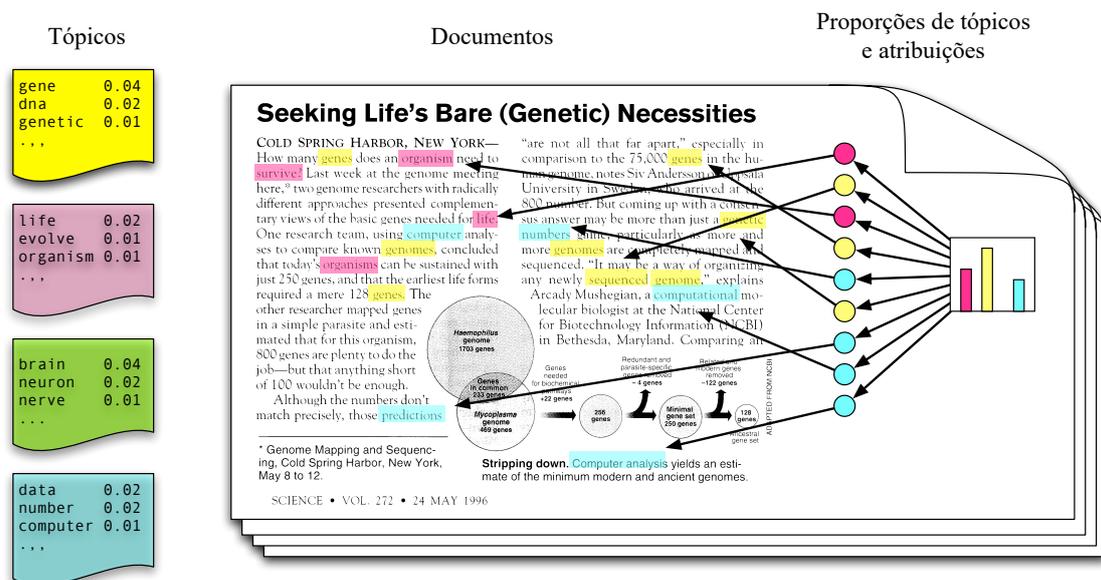
⁶<https://piaui.folha.uol.com.br/lupa/category/no-facebook/page/>

⁷<https://g1.globo.com/politica/eleicoes/2018/index/feed/pagina->

isto é, que não são explicitamente observáveis. O algoritmo trabalha com a suposição de que uma coleção de documentos é representada por vários tópicos. Para exemplificar essa suposição, M Blei (2011) utilizou a Figura 4.3 que apresenta um artigo científico tratando sobre a utilização da análise de dados para determinar o número de genes que um organismo precisa para sobreviver. De forma manual, o autor destacou as diferentes palavras usadas no artigo. Palavras envolvendo “análise de dados”, tais como *computer* e *prediction* foram destacadas na cor azul; palavras sobre “biologia” como *life* e *organism* foram destacadas em rosa; palavras sobre “genética” como *sequenced* e *genes* foram destacadas em amarelo. Seguindo essa sistemática para todas as palavras do artigo, excluindo as *stopwords*, seria possível observar que o artigo combina genética, análise de dados e biologia evolutiva com proporções diferentes.

Sendo assim, é pressuposto um número qualquer de tópicos, que são distribuições sobre palavras, existentes para toda a coleção (quadros coloridos à esquerda). Cada documento é tratado da seguinte maneira. Primeiro, escolhe-se uma distribuição sobre os tópicos (o histograma à direita); em seguida, para cada palavra, escolhe-se a qual tópico ela pertence (os círculos coloridos). Essa é a característica distintiva do LDA, todos os documentos da coleção compartilham o mesmo conjunto de tópicos, mas cada documento exibe esses tópicos com proporções diferentes (M Blei, 2011). Essa é apenas uma visão ilustrativa do funcionamento do algoritmo LDA e maiores detalhes podem ser conferidos no trabalho de Blei, Ng e Jordan (2003).

Figura 4.3: Exemplo de funcionamento do algoritmo LDA.



Fonte: Adaptado de M Blei (2011).

Considerando que um tópico é um conjunto de palavras importantes (RöDER; BOTH; HINNEBURG, 2015) e trazendo a ideia do exemplo acima para o contexto desta pesquisa, isto é, notícias, e supondo que as top-5 palavras de um tópico fossem: {"bolsonaro"; "atentado"; "dizer"; "adelio"; e "transtorno"}, poder-se-ia dizer que esse tópico trataria sobre o atentado contra um candidato à presidência, Jair Bolsonaro, durante o período eleitoral de 2018.

Antes da geração do modelo de tópicos, foi necessária uma etapa de pré processamento dos textos das notícias. Primeiramente foi realizada uma limpeza através de expressões regulares de forma a remover os seguintes itens de conteúdo: caracteres de nova linha, aspas, pontuação, e palavras recorrentes sem importância, por exemplo, Foto, Autor(a), dentre outras. Vale destacar que todo o processo de limpeza foi executado inúmeras vezes, sempre analisando os textos a cada iteração. Em seguida, os textos foram separados em palavras (*tokens*) o que permitiu remover *stopwords* e criar bigramas. Com isso, foi possível remover as flexões das palavras com a utilização do lematizador, em Python, da biblioteca spaCy⁸. Por fim, foi criado o *corpus* estruturado como um dicionário composto de termos, identificadores e frequências de palavras, utilizando o modelo *Bag Of Words (BoW)*.

A implementação do modelo de tópicos ocorreu com o apoio da biblioteca Python GenSim⁹. O algoritmo LDA da biblioteca gera um modelo que permite a inferência de tópicos em um determinado documento. Os seus principais parâmetros de entrada são o *corpus* e um número de tópicos a serem extraídos. Portanto, o próximo desafio foi encontrar um número ideal de tópicos que melhor representasse o conjunto de notícias.

Nesse sentido, foram gerados 102 modelos, cada um possuindo um número crescente, com incrementos de cinco unidades, de tópicos a serem extraídos. Com isso, foi possível estimar a qualidade de cada modelo a partir do índice de coerência que mede o quão interpretáveis e corretos são os tópicos extraídos por um modelo LDA (RöDER; BOTH; HINNEBURG, 2015). A Figura 4.4 mostra um gráfico que relaciona a quantidade de tópicos de cada modelo com o índice de coerência alcançado. Observa-se que o maior valor de coerência (0,4939) é atingido quando todas as notícias são representadas por 16 tópicos.

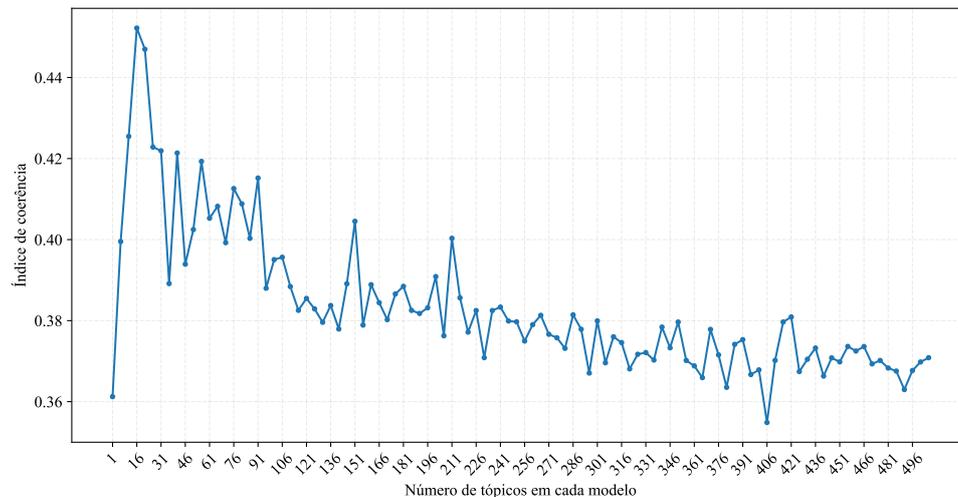
A Figura 4.5 apresenta os 16 tópicos identificados, utilizando um mapa de distância inter-tópicos, obtido através da ferramenta LDAVis¹⁰. Os círculos representam cada

⁸<https://spacy.io/>

⁹<https://radimrehurek.com/gensim/>

¹⁰Fornece uma visão global dos tópicos (e como eles diferem uns dos outros), enquanto ao mesmo tempo

Figura 4.4: Valores de coerência do modelo para diferentes números de tópicos



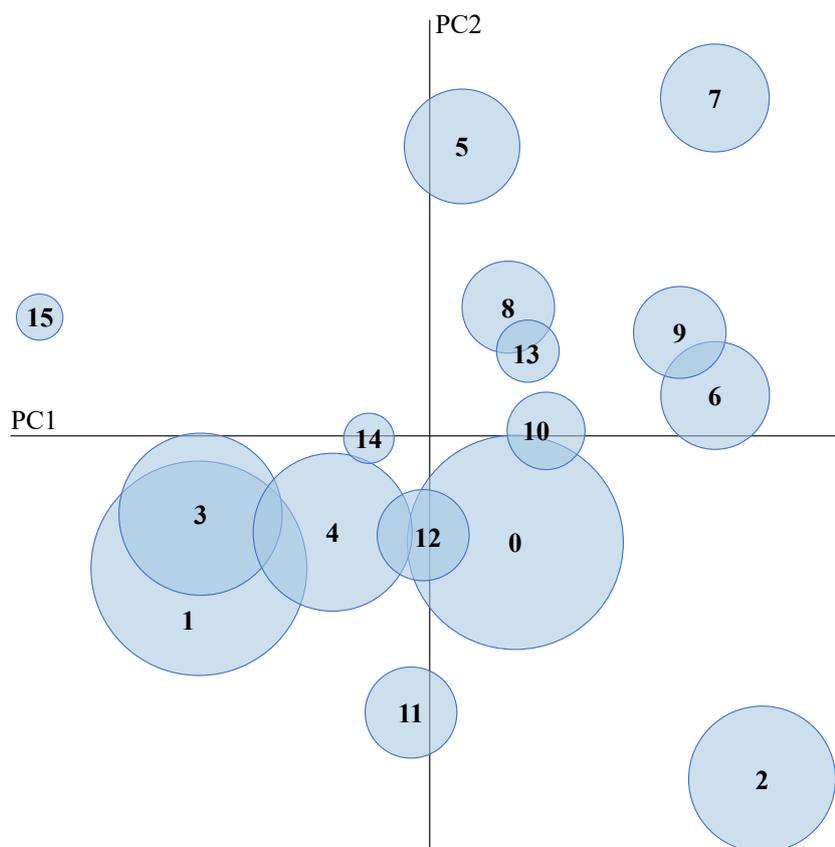
Fonte: O Autor

um dos 16 tópicos plotados sobre um plano bidimensional cujos centros são determinados calculando a distância entre os tópicos. Os tópicos são exibidos de acordo com os eixos dos componentes principais PC1 e PC2. Essa é uma técnica estatística conhecida como PCA (*Principal Component Analysis*) que emprega uma transformação ortogonal para converter um conjunto de variáveis com alguma relação em um conjunto de variáveis não relacionadas denominadas componentes principais - *Principal Components* (PC). A área de cada círculo representa a prevalência do tópico, sendo essa métrica utilizada para ordenar os tópicos em ordem decrescente (SIEVERT; SHIRLEY, 2014). Com isso, observa-se que o tópico 0 é o mais prevalente em relação a todo o conjunto de notícias, o tópico 1 possui uma prevalência menor e, assim, sucessivamente. A sobreposição de um círculo ao outro indica que os tópicos tratam de assuntos próximos. Sievert e Shirley (2014) sugerem que quanto menos sobreposições existirem, mais distintivos serão os tópicos.

Após selecionar 16 como o número ideal de tópicos, o modelo foi utilizado para inferir os tópicos presentes em cada notícia. O resultado desta inferência é uma matriz binária cujo cabeçalho contém os tópicos e, as linhas, o identificador do banco de dados de cada notícia, conforme mostrado pela Tabela 4.3. Dessa forma, obteve-se um mapeamento de quais tópicos caracterizavam cada notícia. Essa matriz foi armazenada em um

permite uma inspeção dos termos mais relevantes a cada tópico. **Esses termos podem ser conferidos no Apêndice A para cada um dos 16 tópicos identificados.** Cada tópico é ilustrado por meio de um gráfico de barras horizontais que representam os termos mais úteis para interpretar o significado do tópico. Um par de barras sobrepostas representa a frequência de todo o *corpus* para um determinado termo, bem como a frequência do termo no tópico em específico.

Figura 4.5: Mapa de distância inter-tópico via escala multi-dimensional.



Fonte: O Autor

arquivo csv que, por sua vez, foi utilizado pelos algoritmos diversificadores toda vez que uma recomendação diversificada era requerida por um usuário. Essa matriz foi utilizada pelas funções de similaridade dentro de cada um dos dois algoritmos.

Tabela 4.3: Exemplo da Matriz de notícias versus tópicos.

Id	t0	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	t11	t12	t13	t14	t15
1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
3	0	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0
...
1938	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0

Fonte: O Autor.

A construção dessa matriz ocorreu a partir do modelo de tópicos aprendido. Para tanto, um método disponível no próprio modelo pela biblioteca GenSim, retorna uma lista de dominância dos tópicos, em valores de porcentagem, para cada uma das notícias passadas como parâmetro. Logo, obteve-se uma matriz com a mesma estrutura apresentada anteriormente pela Tabela 4.3, porém, com valores de porcentagem em suas células

referentes à probabilidade de pertencimento de uma dada notícia a um dado tópico.

Para a forma final, foi implementada uma função de binarização baseada no valor de porcentagem. Em outras palavras, diz-se que uma notícia trata de um tópico caso a probabilidade deste seja maior ou igual a zero, conforme a heurística da Equação (4.2). Essa estratégia foi adotada porque haviam poucos valores de probabilidade (tópicos) em cada notícia. Com poucos valores, a aplicação de um *threshold*, por exemplo, para a atribuição do valor 1, diminuiria ainda mais a quantidade de tópicos a ponto de erradicar a presença de todos para algumas notícias, situação que prejudicaria o processo de diversificação.

$$f(x) = \begin{cases} 1 & \text{se } x \leq 1.0 \wedge x \neq 0 \\ 0 & \text{Outro} \end{cases} \quad (4.2)$$

4.1.3 Entidades nomeadas

O Reconhecimento de Entidades Nomeadas (REN) é uma sub-tarefa da extração de informações e tem como objetivo localizar e classificar elementos do texto em categorias pré-definidas como nome de pessoas, organizações, lugares, datas, moedas e outras classes (CARVALHO, 2012). As entidades nomeadas foram extraídas a partir da biblioteca *Python spaCy*, utilizando um conjunto de treinamento de notícias da própria biblioteca¹¹ em português, com o módulo NER (*Named Entity Recognizer*). Esse conjunto é denominado *pt_core_news_sm*¹² e contém notícias manualmente anotadas por milhares de colaboradores da biblioteca.

A Figura 4.6 apresenta a saída referente à utilização da biblioteca *spaCy* para o REN. Como exemplo, foram utilizadas as manchetes das duas notícias falsas utilizadas como motivação na abertura deste capítulo. A marcação “PER” indica que a entidade em questão é uma pessoa e, a marcação “LOC”, indica que a entidade é uma localização. Além dessas marcações, o modelo treinado ainda conta com outras duas: “ORG” (organizações) e “MISC” (miscelânea).

Todos os textos das notícias passaram pelo procedimento ilustrado no exemplo anterior para a extração das entidades nomeadas. Ao todo, foram extraídas 18.383 entidades únicas, isto é, após um processamento para a remoção de duplicatas. Mesmo assim, foi possível perceber, ainda, a presença de algumas ambiguidades em função da divergência

¹¹<https://spacy.io/models/pt>

¹²<https://tinyurl.com/1alup5xk>

Figura 4.6: REN nas manchetes das duas notícias falsas motivacionais deste capítulo.

Roberto Carlos PER manifesta publicamente seu apoio a Jair Messias Bolsonaro PER
 Pablo Vittar PER : se Bolsonaro PER ganhar em 2018, eu declaro toda a minha
 carreira encerrada no Brasil LOC . Não tenho como viver no mesmo país que esse homem

Fonte: O Autor

de nomes como, por exemplo, “Jair Bolsonaro” e ‘Bolsonaro’. Esse tipo de ambiguidade foi mantida porque está ligada à acurácia do algoritmo utilizado para a extração de entidades.

Esse número elevado de entidades trouxe à tona um problema comum em aprendizado de máquina que é o elevado número de características, que acarreta em tempo elevado de processamento e, muitas vezes, até impraticável. Neste caso, cada entidade passou a ser representada como um vetor binário acrescentado à matriz de tópicos, conforme exemplo apresentado na Tabela 4.3 anteriormente. Por isso, foi empregado um método para a seleção de características que melhor representassem o conjunto original.

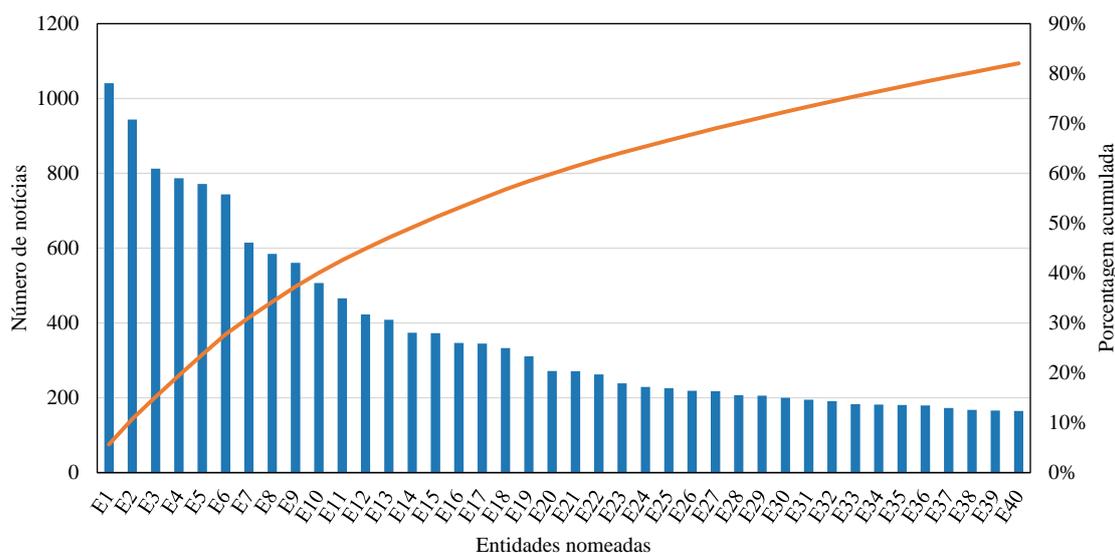
Nesse sentido, foi utilizado o princípio de Pareto em função de sua simplicidade e eficácia. Essa análise consiste no princípio da contagem das ocorrências de cada categoria e sua representatividade frente ao conjunto total. Em termos mais formais, o princípio de Pareto, que também é conhecido como regra do 80/20, lei dos poucos vitais ou princípio de escassez do fator, consiste em afirmar que aproximadamente 80% dos efeitos surgem de 20% das causas (DUNFORD; SU; TAMANG, 2014).

A Figura 4.7 apresenta um gráfico de Pareto com as *top* 40 entidades mais citadas nos textos das notícias bem como a porcentagem acumulada de representação que cada entidade tem em relação ao conjunto original de entidades extraídas, isto é, 18.383. No gráfico, é possível observar que as 40 entidades mais citadas atingem 82,05% de representação do conjunto original de entidades.

Entretanto, visando uma representatividade maior, foram selecionadas 65 entidades que, juntas, totalizam uma porcentagem acumulada de 99%. O leitor interessado pode verificar a correspondência de cada entidade E_n presente no gráfico Figura 4.7 e além das demais 25 entidades selecionadas no Apêndice B. Portanto, a matriz de tópicos passou a contar com 65 colunas a mais além das 16 referentes a cada tópico.

Figura 4.8 apresenta uma visão informacional ao leitor acerca das 20 entidades extraídas e mais citadas de acordo com o número de notícias que mencionaram cada uma.

Figura 4.7: Gráfico de Pareto



Fonte: O Autor

4.1.4 Polaridade e subjetividade

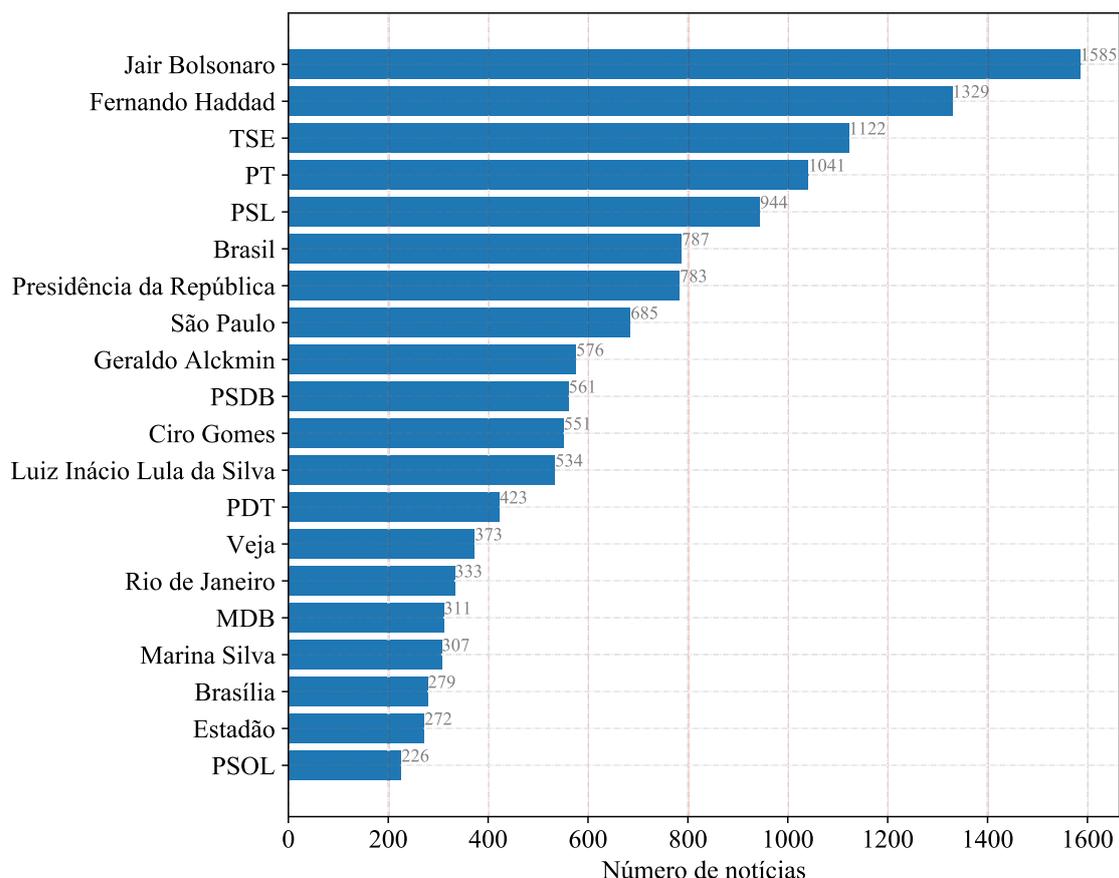
A estratégia empregada para a extração dos sentimentos dos textos das notícias baseou-se na biblioteca *TextBlob*¹³. O algoritmo retorna uma tupla, contendo a polaridade e a subjetividade de um documento. A polaridade varia entre $[-1,1]$, sendo -1 uma polaridade mais negativa, e 1 uma polaridade mais positiva. O algoritmo também possui rótulos semânticos que ajudam na análise detalhada. Por exemplo - *emoticons*, ponto de exclamação, emojis, etc. A subjetividade varia entre $[0,1]$, quantificando o montante de opiniões pessoais e informações factuais contidas no texto. Valores maiores de subjetividade indicam que o texto contém opinião pessoal ao invés de informações factuais.

Entretanto, a biblioteca *TextBlob* exige que os textos submetidos para análise estejam em língua inglesa. Como os textos das notícias se referiam a conteúdos em português, foi necessário traduzir cada uma das notícias para a língua inglesa. Benevenuto, Ribeiro e Araújo (2015) salientam que essa é uma forma simples e eficiente para realizar a análise de sentimentos em diferentes idiomas, pois combina o poder dos métodos já existentes em inglês com a eficiência dos tradutores automáticos tal como *Google Translate*, *DeepL*¹⁴, dentre vários outros. Uma simples tradução do texto de entrada a partir dessas ferramentas para o idioma que o método análise opera pode trazer bons resultados.

Assim, a biblioteca calcula os valores de polaridade e subjetividade com base no

¹³<https://textblob.readthedocs.io/>

¹⁴<https://www.deepl.com/translator>

Figura 4.8: As vinte entidades mais citadas no *corpus* de notícias.

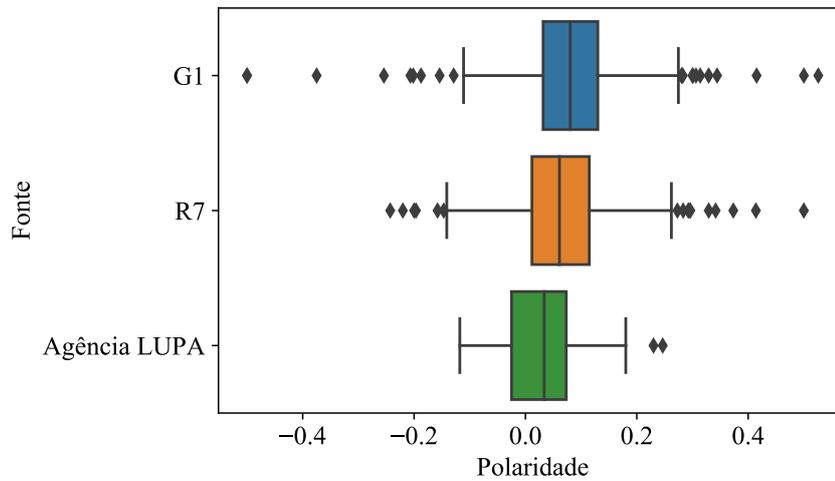
Fonte: O Autor

algoritmo *NaiveBayes*, estendido da biblioteca NLTK¹⁵, o conjunto de bibliotecas *Python* mais amplamente utilizado para o processamento de PLN. O algoritmo foi treinado a partir de 278 manchetes de notícias, sendo 78 da Agência LUPA, 100 do portal R7, e 100 do portal G1. Essas manchetes foram rotuladas manualmente pelo autor.

A Figura 4.9 apresenta um gráfico do tipo *boxplot* com os valores referentes às polaridades identificadas nos textos das notícias de cada um dos portais, isto é, G1, R7 e Agência LUPA. É possível observar que a Agência LUPA é o portal que mais se aproxima da neutralidade, apresentando uma mediana próxima de zero. Já o portal R7 se mostra com textos com um tom mais positivo. O portal G1 tende para uma positividade mediana, em seus textos, levemente mais acentuada quando comparada ao portal R7.

¹⁵<https://www.nltk.org/>

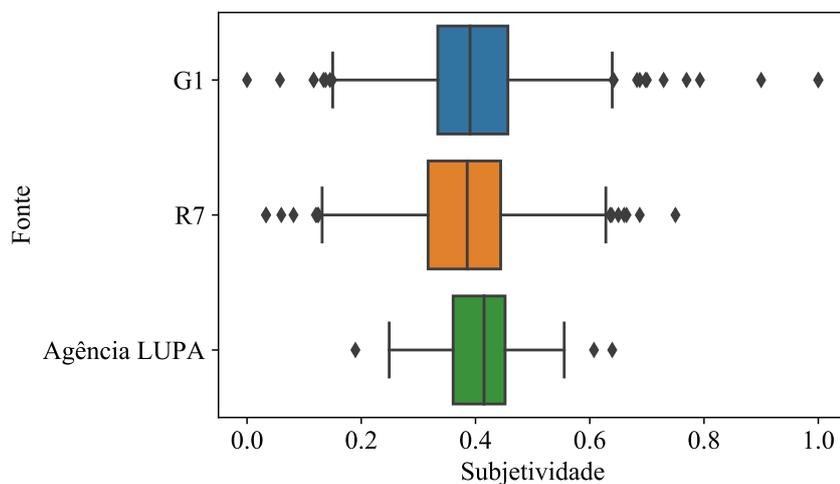
Figura 4.9: Variabilidade dos valores de polaridade para cada um dos portais.



Fonte: O Autor

A Figura 4.10 apresenta um gráfico do tipo *boxplot* com os valores referentes às subjetividades identificadas nos textos das notícias de cada um dos portais. É possível observar que os três portais coadunam com uma subjetividade mediana próxima de 0,4. Entretanto, os portais G1 e R7 se mostram mais objetivos, com mediana levemente inferior a 0,4. Apesar disso, G1 e R7 apresentam valores de máximo e mínimo mais distantes e, também, maior quantidade de *outliers*.

Figura 4.10: Variabilidade dos valores de subjetividade para cada um dos portais.



Fonte: O Autor

4.2 Medição de bolha de filtro

Uma bolha de filtro é aqui definida como a homogeneidade presente no conjunto \mathcal{R} que representa os itens recomendados. Assim, pode-se avaliar a “bolha” a partir do conjunto de itens com os quais cada usuário u demonstrou interesse ou, então, a partir das listas de recomendação que ainda não receberam interação. Com esse último caso, é possível avaliar o potencial que um lista possui de homogeneizar as possíveis interações dos usuários. Quanto maior for a homogeneidade de itens mais considera-se que u está em uma bolha e, conseqüentemente, maior é a tendência desse estar enviesado por um conjunto de notícias.

O efeito de bolha de filtro é medido através do grau de homogeneidade gh que é calculado como a média de valores pareados de alguma medida de similaridade sim entre duas notícias \vec{n}_i e $\vec{n}_j \in \mathcal{R}$, conforme mostra a Equação (4.3). A medida de similaridade a ser empregada depende do contexto de medição que inclui: o domínio de aplicação e, conseqüentemente, as características dos itens e os tipos de dados envolvidos. Durante os experimentos desta tese, adotou-se o índice de similaridade de Jaccard e a similaridade de cossenos que são melhor detalhados no próximo capítulo.

$$gh(u) = \frac{\sum_{n=1}^{|\mathcal{R}|} sim(\vec{n}_i, \vec{n}_j)}{|\mathcal{R}|} \quad (4.3)$$

Uma forma complementar de análise do efeito de bolha de filtro, quando há disponibilidade da nota predita e da nota real fornecida pelo usuário, é através da métrica RMSE, que mede o erro entre elas. Quanto menor for esse erro, maior é a acurácia do sistema que pode ser um indicativo da formação de bolha de filtro, desde que essa acurácia esteja relacionada com a métrica GH.

4.3 Resumo do capítulo

Neste capítulo, foram apresentados: a abordagem para a diversificação multi-atributo de recomendação no domínio de notícias, composta por: um conjunto multi-atributo para a representação de notícias; e algoritmos de diversificação baseados em pós-filtragem. Além disso, foi apresentada uma metodologia para a medição do efeito de bolhas de filtro em dois níveis: (i) nos itens recomendados com os quais um usuário interagiu, empregando a métrica GH e o RMSE; e (ii) nos itens presentes em uma lista

de recomendação que ainda não tenha recebido interações, empregando apenas a métrica GH.

Nos próximos capítulos, a abordagem de diversificação aqui descrita é experimentada, assim como a metodologia para a medição de bolhas de filtro, nos dois níveis mencionados, por meio de um estudo de usuário. Esse estudo de usuário compreende dados coletados a partir de uma aplicação *web* para a recomendação diversificada de notícias cujos detalhes de implementação são descrito no próximo capítulo.

5 NEWSREC: WEB APP PARA A RECOMENDAÇÃO DE NOTÍCIAS

Tendo em vista a dificuldade de acesso aos dados de recomendação, isto é, que algoritmo gerou determinada recomendação, gerados por grandes meios de comunicação tais como *Facebook*, portais de notícias, dentre outros, foi desenvolvida uma aplicação *web* denominada *NewsREC*¹²³ em linguagem Python, utilizando o *framework* Django, banco de dados PostgreSQL, e recursos de interface como Bootstrap, CSS e Javascript.

Para acessá-la, o participante precisa se registrar e, neste momento é designado a ele, de forma transparente e automática, um dos algoritmos de recomendação de filtragem colaborativa implementados utilizando a biblioteca Python Surpr!se⁴, sendo: KNN baseado em usuário (*knnu*); KNN baseado em item (*knni*); e fatoração de matrizes com SVD (*svd*). Além desses, dois algoritmos de diversificação foram adotados, baseados em pós-filtragem, e implementados em linguagem Python pura, sendo eles: o MMR (*mmr*), proposto por Carbonell e Goldstein (1998); e o diversificador de tópicos (*td*), proposto por Ziegler et al. (2005). Esses algoritmos de diversificação foram escolhidos, pois estão entre os mais populares quando se aborda a diversificação de recomendação.

Combinando essas estratégias, obteve-se um total de nove algoritmos, conforme mostra a Tabela 5.1, que são distribuídos aos usuários de uma forma *Round-Robin*, isto é, o primeiro usuário recebe o algoritmo *knnu*, o segundo usuário recebe o algoritmo *knnu td* e assim sucessivamente. Ao esgotarem-se os algoritmos, o processo inicia novamente. Com isso, buscou-se uma distribuição que fosse justa e que facilitasse o trabalho de análise e comparação de resultados, separando os usuários em grupos de teste (recomendações diversificadas) e de controle (recomendações não diversificadas).

O sistema possui duas interfaces gráficas principais. A primeira delas é utilizada para coletar as preferências iniciais de um usuário. A coleta de preferências iniciais é a primeira etapa do experimento e ocorre sempre que um usuário é novo no sistema. A Figura 5.1 apresenta duas imagens da primeira etapa do experimento na qual o usuário recebe 25 itens de notícias por página. O usuário recebe algumas instruções a respeito do que fazer nesta etapa, conforme consta na lateral esquerda da Figura 5.1(a), sendo: “(i) analise as imagens e leia os títulos”; “(ii) abra a notícia para ler seu conteúdo se desejar”; e “(iii) forneça uma avaliação através de um *Like* ou *Dislike*”. Ao final da

¹Disponível em: <https://sisrec.inf.ufrgs.br/news-rec/>

²Software registrado no Instituto Nacional de Propriedade Intelectual (INPI) sob o número BR512020000372-9.

³Código-fonte disponível em: https://github.com/gmlunardi/news_rec

⁴Disponível em: <http://surpriselib.com/>

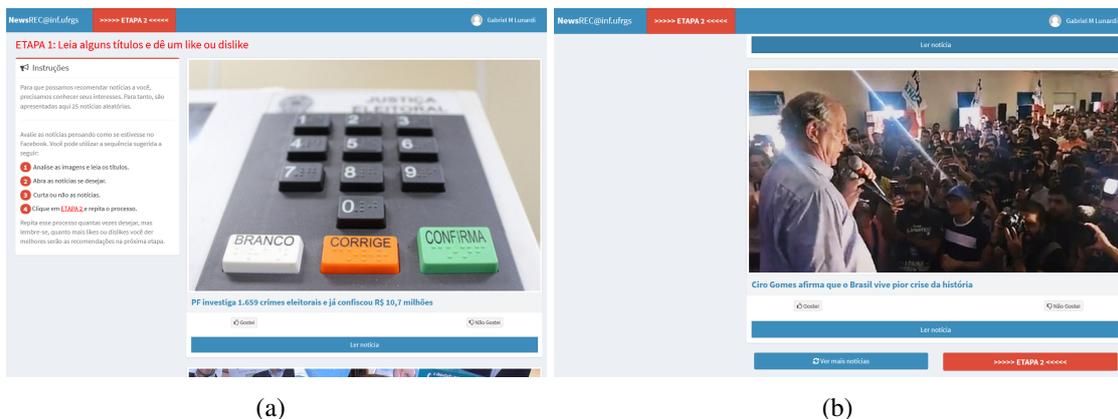
Tabela 5.1: Algoritmos de recomendação e de diversificação implementados.

Acrônimo	Algoritmo	Grupo
(i) knnu	<i>KNN user-based</i>	controle
(ii) knnu td	<i>KNN user-based com TD</i>	teste
(iii) knnu mmr	<i>KNN user-based com MMR</i>	
(iv) knni	<i>KNN item-based</i>	controle
(v) knni td	<i>KNN item-based com TD</i>	teste
(vi) knni mmr	<i>KNN item-based com MMR</i>	
(vii) svd	<i>MF with SVD</i>	controle
(viii) svd td	<i>MF with SVD com TD</i>	teste
(ix) svd mmr	<i>MF with SVD com MMR</i>	

Fonte: O Autor.

página, o usuário tem duas opções: recarregar uma nova página com outras 25 notícias aleatórias ou, então, prosseguir para a próxima etapa. É possível verificar essas opções na Figura 5.1(b) como dois botões na parte inferior da interface, “Ver mais notícias” e “ETAPA 2”, respectivamente.

Figura 5.1: Primeira etapa do experimento.



Fonte: O Autor

Sendo assim, o usuário não fica limitado a 25 notícias, podendo avaliar quantas notícias desejar durante a coleta inicial do seu perfil. Quanto mais o usuário interagir com as notícias nessa etapa, isto é, abrir e/ou ler as notícias, maior será a qualidade do seu perfil que será utilizado para calcular as recomendações mais tarde. Um maior número de avaliações colabora para a suavização do problema de partida a frio (*cold-start*). Não foi estabelecido um número mínimo de itens que um usuário deveria avaliar porque isso poderia inserir algum tipo de viés no seu comportamento. Então, foi sugerido ao usuário que quanto mais interagisse com os itens, melhor.

Dentro das 25 notícias aleatórias, 8,5% são falsas sendo este número definido a partir de um estudo referente às últimas eleições presidenciais dos EUA o qual revela que

8,5% das pessoas compartilhava pelo menos uma notícia falsa na rede social (GUESS; NAGLER; TUCKER, 2019). O que justifica a escolha desse número é a inexistência de uma estatística brasileira a respeito da proporção de conteúdo falso e verdadeiro disseminado em redes sociais.

Antes de seguir para a próxima etapa, é preciso calcular a recomendação de acordo com o algoritmo atribuído. Durante esta fase, o usuário recebe o *feedback* de uma tela que o informa o *status* do processamento, conforme mostra a Figura 5.2.

Figura 5.2: Tela de feedback do processamento da recomendação.

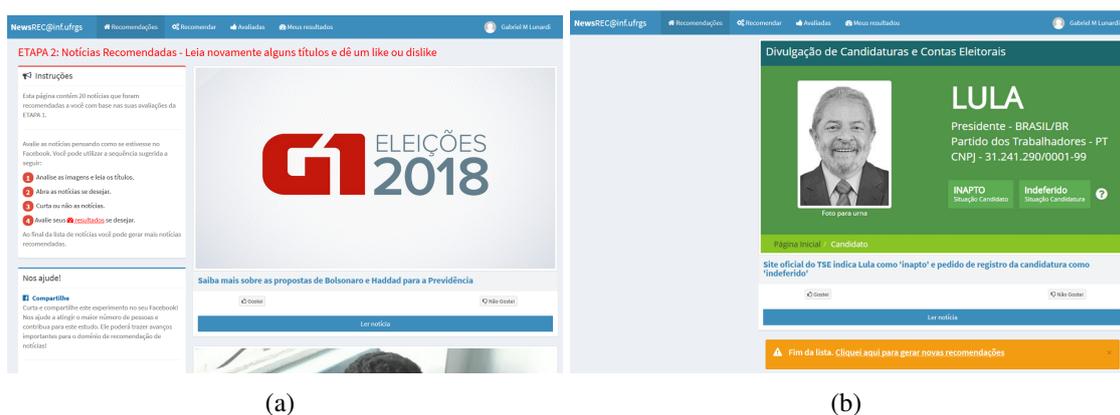


Fonte: O Autor

Na segunda etapa do experimento, o usuário interage com as recomendações de notícias geradas pelos algoritmos de recomendação, fazendo uso da interface gráfica apresentada na Figura 5.3. A partir dela, destaca-se a semelhança com os *feeds* comumente encontrados em várias redes sociais atuais. Da mesma forma que foi sugerido ao usuário na primeira etapa, esse deve avaliar cada uma das notícias de acordo com os mesmos três critérios, conforme apresentado na lateral esquerda da Figura 5.3(a). O usuário recebe *top-20* notícias por página e, ao final de cada uma, existe a opção de gerar uma nova lista de recomendação com outras *top-20* notícias, através de um botão conforme mostra a Figura 5.3(b). Portanto, o usuário pode gerar quantas listas de recomendação desejar e, por consequência, não há um limite de interações usuário-notícia recomendada.

O processamento da recomendação é disparado toda a vez que um usuário solicita uma nova lista de recomendação. O processamento consiste em uma pré-filtragem para selecionar todas as notícias que tenham recebido pelo menos um *rating*. Esse pré-requisito é necessário porque foram adotados algoritmos de filtragem colaborativa com *feedback* explícito e, por isso, requerem alguns *ratings* sobre os itens. A pré-filtragem também objetiva garantir que a proporção de 91,5% de notícias verdadeiras e 8,5% de notícias falsas seja respeitada no conjunto de notícias passadas para o treinamento de cada algoritmo. Esta pré-filtragem também garante que cada algoritmo de recomendação re-

Figura 5.3: Segunda etapa do experimento.



Fonte: O Autor

ceba a mesma quantidade de conteúdo falso e verdadeiro, isolando os efeitos da interação exclusivamente ao algoritmo e ao comportamento do usuário, criando uma semelhança suficiente com uma rede social real. Com isso, espera-se evitar viés nos resultados já que a exposição maior ou menor a conteúdo falso tende o usuário a interagir mais ou menos com isso.

Em seguida, cada um dos três algoritmos de filtragem colaborativa é treinado. Para isso, cada um deles recebe uma matriz de *ratings* (usuários x notícias), contendo todas as notícias avaliadas por todos os usuários, respeitando a pré-filtragem explicada anteriormente. Como o tamanho dessa matriz é variável de acordo com o número de usuários se registrando no sistema ao longo do tempo, foi necessário garantir a pré-filtragem de forma dinâmica. Em relação ao treinamento, os algoritmos baseados em memória, isto é, *knnu* e *knni* foram parametrizados com um tamanho de vizinhança (k) igual a 5 e a medida de similaridade adotada foi o coeficiente de correlação de Pearson junto a um parâmetro de encolhimento⁵ a fim de evitar *overfitting* quando poucos *ratings* estão disponíveis. Já o algoritmo baseado em modelo, *svd*, não requer parametrizações.

O treinamento de cada algoritmo ocorre a partir da construção de um conjunto de treino baseado em toda a matriz de *ratings* informada como parâmetro. Na sequência, é gerado um conjunto de teste que contém todas notícias que cada usuário não deu *rating* possibilitando, assim, fazer as previsões de *rating* e gerando todas as recomendações possíveis. A partir disso, é construída a lista de recomendação, ordenada pelo *rating* predito de cada notícia, contendo 20 notícias (*top-k*).

Para fins de diversificação das recomendações, foram implementados dois algoritmos: o MMR e o diversificador de tópicos. Ambos foram parametrizados com um fator

⁵maiores detalhes podem ser encontrados em: <https://surprise.readthedocs.io/en/stable/similarities.html>

de diversificação (λ) de 40%, seguindo a sugestão de Ziegler et al. (2005) que mencionam ser um valor ideal a fim de não prejudicar a experiência do usuário. Afinal, uma lista de recomendação muito diversificada pode prejudicar o usuário no sentido de fornecer itens fora do seu espectro de interesses.

A construção da lista de recomendação diversificada, por ser baseada em pós-filtragem, é realizada a partir de uma lista de recomendação com top-100 notícias. Isso se justifica pelo fato de que é necessário, conforme apontado por Ziegler et al. (2005), pelo menos cinco vezes o tamanho da lista pretendida (diversificada), no caso, top-20 para ficar de acordo com as listas não diversificadas. Por exemplo, para uma lista diversificada pretendida com 20 itens é necessário uma lista de entrada (não diversificada) de pelo menos 100 itens.

Uma das premissas para o funcionamento dos algoritmos de diversificação adotados é que sejam identificadas características (*features*) relativas ao conteúdo dos itens. No caso das notícias coletadas nenhuma delas possui atributos extras como, por exemplo, o tópico tratado (reforma da previdência, pandemia, eleições, etc.). Há somente o texto na íntegra. Com base nisso, foi adotada uma estratégia para identificar características latentes nos textos, isto é, características que não estão explicitamente aparentes. Neste caso, foi adotada a técnica LDA, que permite extrair tópicos de forma automática, conforme detalhes apresentados no Capítulo 4. O resultado é uma matriz binária, também apresentada no referido capítulo, que mapeia o pertencimento ou não de cada notícia aos tópicos. Assim, essa matriz binária serve de base para o funcionamento das funções de similaridade utilizadas pelos diversificadores.

Por fim, o usuário recebe a recomendação de 20 notícias, diversificadas ou não, com base no algoritmo atribuído, e a depender da plataforma sendo utilizada. Durante a concepção do experimento foi levado em consideração o cuidado de desenvolver as interfaces responsivas de modo que nenhum usuário tivesse a experiência prejudicada em função da plataforma utilizada para o acesso. A Figura 5.4 mostra a interface de recomendações caso o usuário esteja utilizando um dispositivo móvel.

Adicionalmente, o sistema conta com outras interfaces gráficas. A Figura 5.5 mostra o conteúdo de uma notícia assim que o usuário clica sobre a imagem ou, então, sobre o título desta. A figura mostra, como exemplo, uma notícia falsa cujo conteúdo pode ser verificado como falso através da leitura. Essa estratégia foi adotada para evitar a disseminação de conteúdo falso através do experimento.

O usuário pode consultar todas as notícias com as quais interagiu, tanto na fase de

Figura 5.4: Interface móvel para a entrega das recomendações.



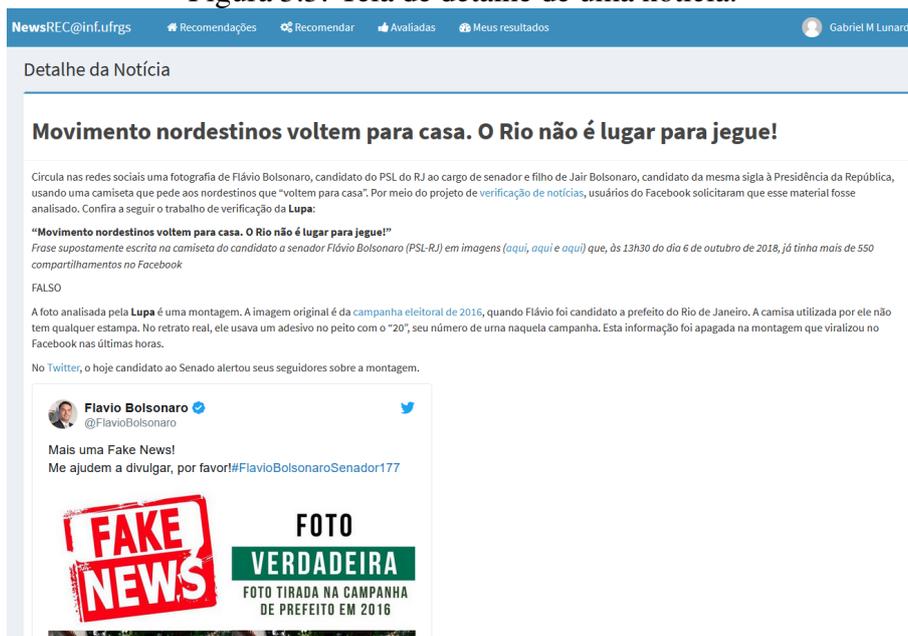
Fonte: O Autor

coleta do perfil como na fase de recomendações como mostra a Figura 5.6. Na referida tela, existe a possibilidade de remover a avaliação de uma notícia, assim como nas demais telas apresentadas anteriormente. Tal ação faz com que a notícia deixe de fazer parte do perfil do usuário, o que influencia o cálculo da recomendação. Entretanto, o *log* de interação é mantido para fins de análise de comportamento do usuário.

Outra possibilidade que o usuário tem à sua disposição é a de verificar como foi a sua interação com as notícias recomendadas por meio de um conjunto de informações, em um *dashboard*, tal como mostra a Figura 5.7, sendo elas:

- número total de notícias avaliadas;
- número de notícias verdadeiras curtidas e não curtidas;
- número de notícias falsas curtidas e não curtidas;
- percentual de interação com notícias falsas e verdadeiras;
- percentual de interação considerando todos os usuários.

Figura 5.5: Tela de detalhe de uma notícia.



Fonte: O Autor

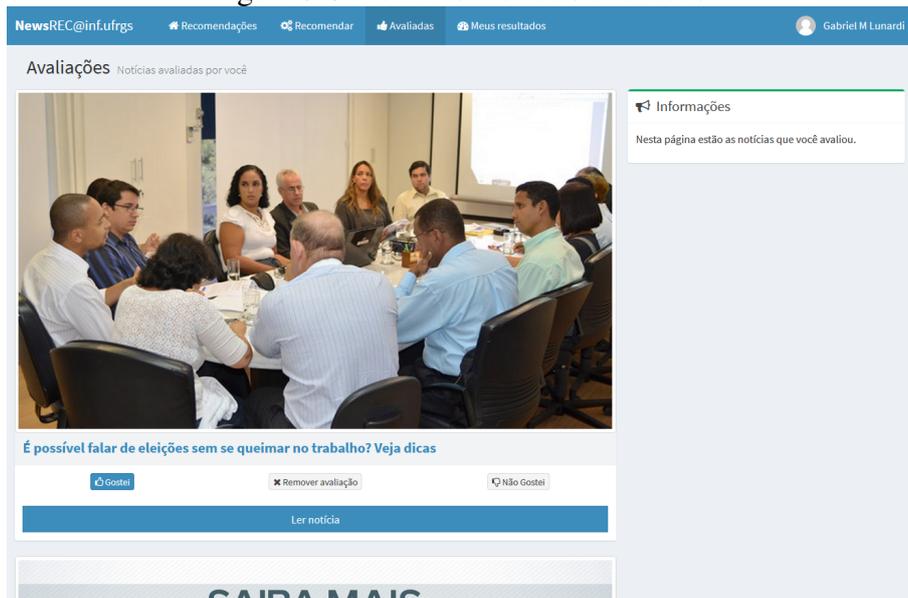
Esse espaço no sistema foi desenvolvido como uma estratégia de estímulo à participação do experimento durante a divulgação do mesmo. Isso pode ser visto, também, como uma recompensa ao usuário após sua participação.

5.1 Banco de dados

O banco de dados desenvolvido para suportar a aplicação descrita até aqui possui a estrutura apresentada pela Figura 5.8. A tabela `News` armazena os dados coletados de cada notícia. Merecem destaque os atributos:

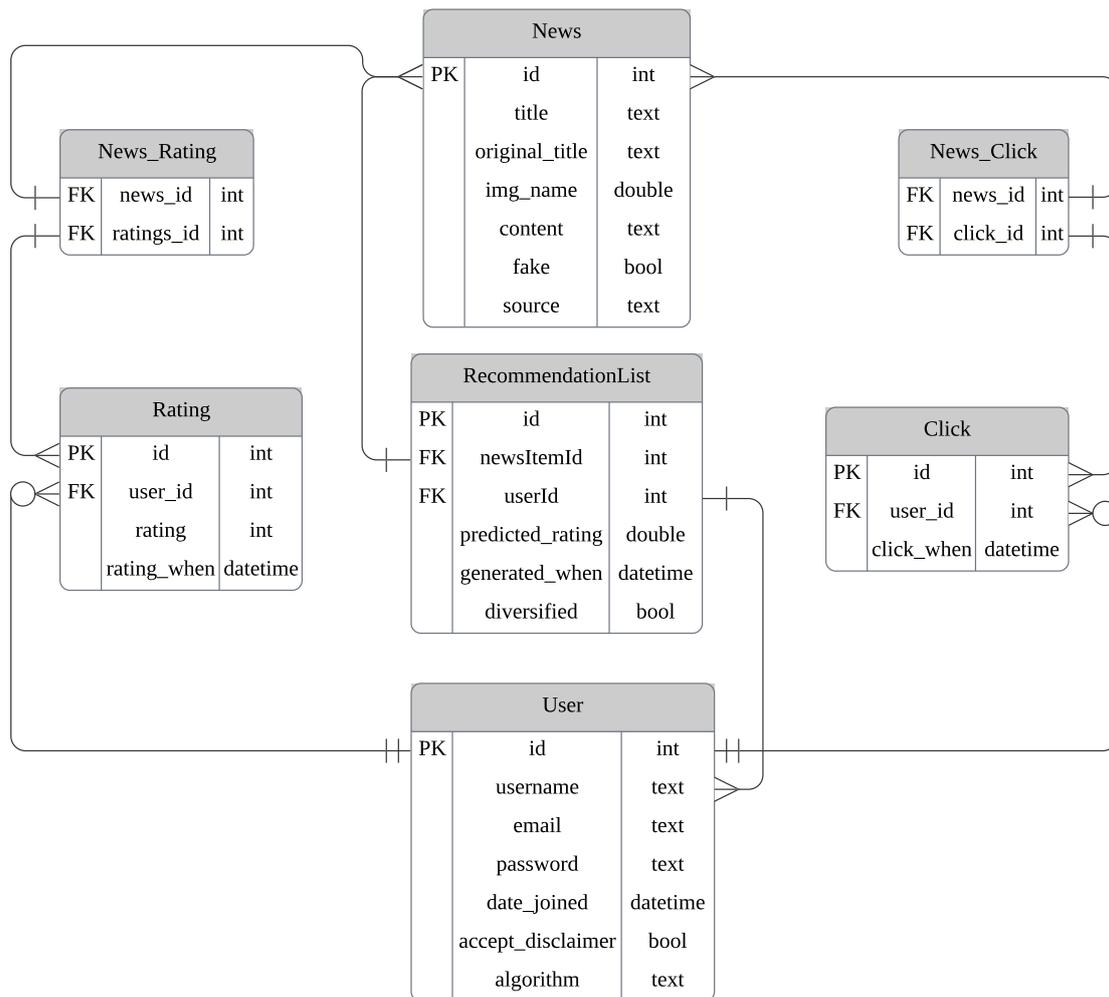
- `titulo_original` armazena o título da notícia falsa;
- `titulo` armazena o título após verificação dos fatos no caso das notícias falsas e o título padrão no caso de uma notícia verdadeira;
- `img_name` armazena o nome da imagem destaque cujo arquivo é guardado em um diretório;
- `conteudo` guarda o texto da notícia em formato HTML;
- `fake` possui valor booleano indicando se a notícia é falsa ou não;
- `fonte` armazena a fonte de cada notícia dentre três possibilidades: R7, G1 e Agência LUPA.

Figura 5.6: Tela das Notícias avaliadas.



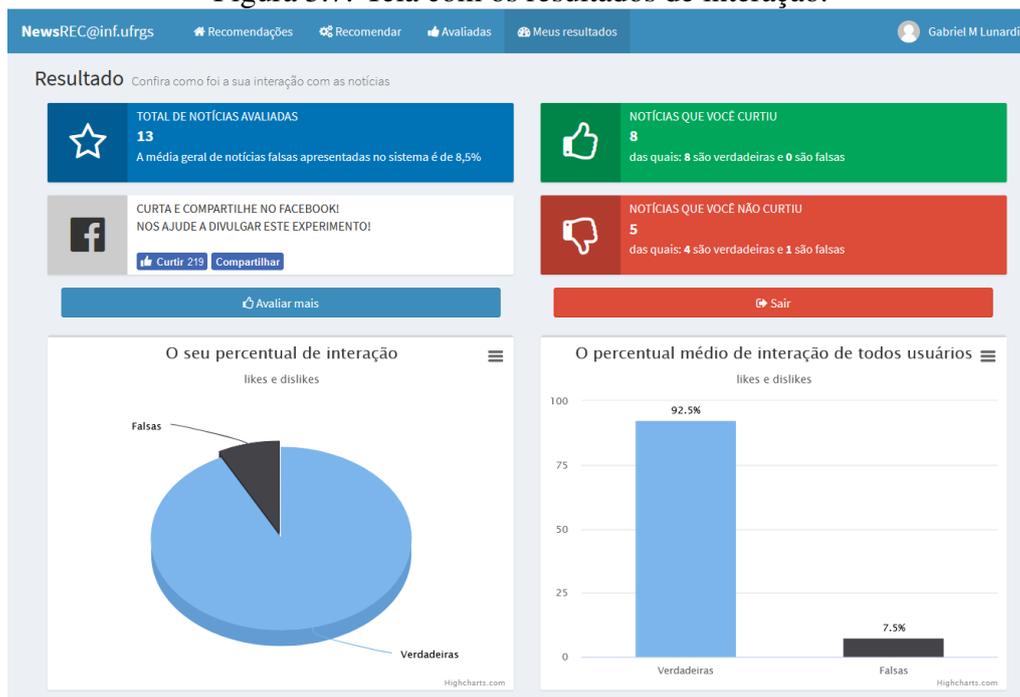
Fonte: O Autor

Figura 5.8: Modelo lógico do banco de dados.



Fonte: O Autor

Figura 5.7: Tela com os resultados de interação.



Fonte: O Autor

A tabela `Ratings` guarda as notas valor 1 (curtir) ou -1 (não curtir) no atributo `rating` de um determinado usuário para uma determinada notícia em um dado momento por meio do atributo `rating_when`. O relacionamento entre as tabelas `Ratings` e `News` é do tipo muitos para muitos, o que permite o armazenamento completo do comportamento das notas do usuário. A nota considerada para o cálculo das recomendações é sempre a última cuja recuperação é realizada a partir do valor de `timestamp` do atributo `rating_when`. A mesma lógica se aplica para a tabela `Clicks` com o atributo `click_when`. Toda vez que o usuário abre uma notícia para a leitura em um dado momento uma entrada é registrada sendo, também, um relacionamento muitos para muitos entre `Ratings` e `News`.

Sobre a tabela `User`, faz-se necessário o detalhamento dos seguintes atributos: (i) `algorithm` armazena o nome do algoritmo atribuído durante o registro do usuário no sistema; (ii) `accept_disclaimer` valor booleano para controlar o acesso ao experimento condicionado à aceitação de ciência do usuário aos termos de uso das notícias concedidas pela Agência LUPA.

A tabela `RecommendationList` é a responsável por guardar as listas de recomendação geradas por cada um dos algoritmos. Para tanto, foram definidos os seguintes atributos: `predicted_rating` que contém a nota predita para uma notícia; `generated_when` é um `timestamp` cuja função é separar as listas tantas quanto forem

as vezes que o usuário solicitar recomendações e, além disso, para fins de recuperação e apresentação da última lista gerada; e *diversified* é um valor booleano cuja função é diferenciar a lista diversificada (20 itens), que é apresentada ao usuário, da lista não diversificada (100 itens) que é utilizada para a construção da primeira. Essa lista é armazenada para fins de log e documentação de tempo de execução do sistema.

5.2 Coleta de dados

A coleta de dados ocorreu a partir do uso da aplicação pelos usuários e, também, através da ferramenta de rastreamento *Google Analytics* dada a sua facilidade de utilização para a visualização de diferentes tipos de informação como, por exemplo, número de acessos por país, número de *clicks*, origem de tráfego, taxa de rejeição, dentre várias outras. A aplicação do experimento ocorreu *on-line*, conforme já descrito em seções anteriores. Inicialmente, foi realizado um teste piloto com um grupo pequeno de usuários para fins de testes de estabilidade e correção de possíveis *bugs* do sistema. Em seguida, foi realizada uma campanha de divulgação nas redes sociais, compreendendo os meses de maio a julho de 2019.

Um total de 317 usuários se registraram no sistema, entretanto, 296 avaliaram pelo menos uma notícia através de *likes* (valor 1) e *dislikes* (valor -1) sendo, portanto, esse o número de usuários válidos. Ao todo, foram coletadas 12.363 notas (*ratings*) sendo a esparsidade total da matriz de utilidade (“usuários \times itens”) de 97,84%.

A Tabela 5.2 apresenta a caracterização geral dos dados coletados, incluindo tanto as notícias que foram avaliadas em listas de recomendação como as notícias que foram avaliadas na fase de coleta do perfil de usuário. Das 1938 notícias cadastradas, 1864 são verdadeiras e 74 são falsas, diferentemente do número de notícias falsas coletadas inicialmente, apresentado na Seção 4.1.1, isto é, 78. Isso é explicado por ter sido constatado, mais tarde, a presença de notícias sem indicação de falsidade em seu conteúdo e, por isso, foram removidas. A média de notícias avaliadas por usuário é de 38,58 (desvio padrão de 35,69); O número médio de avaliações que cada notícia recebeu é de 6,73. Esse *dataset* permite que outros pesquisadores façam experimentos com seus algoritmos utilizando, por exemplo, validação cruzada em testes *offline* da mesma forma que poderiam fazer utilizando *datasets* famosos na área de recomendação conforme foi apresentado na Seção 2.1.6.1.

Seguindo a mesma diretriz dos *datasets* famosos, isto é, de anonimização e dis-

Tabela 5.2: Característica geral do *dataset* de avaliações

Atributo	Valor
Escala de <i>rating</i>	-1 a 1
Notas coletadas (<i>Ratings</i>)	12.363
Esparsidade total	97,84%
Usuários que se registraram no sistema	317
Usuários que avaliaram notícias	296
Notícias cadastradas	1938
Notícias verdadeiras cadastradas	1864
Notícias falsas cadastradas	74
Notícias avaliadas	1832
Notícias verdadeiras avaliadas	1758
Notícias falsas avaliadas	74
Média de notícias avaliadas por usuário	38,58
Média de avaliações por notícia	6,73

Fonte: O Autor.

ponibilização, o *dataset* desta tese está disponível no repositório *Zenodo*⁶. O pesquisador interessado pode ter acesso a esse conjunto de dados em Lunardi e Oliveira (2021), citando a referência ao *dataset* e o seguinte artigo: Lunardi et al. (2020). Detalhes sobre os dados coletados podem ser verificados na explicação da Figura 5.8.

A Tabela 5.3 apresenta a caracterização de uma porção menor do *dataset* de avaliações, compreendendo somente as notícias que foram avaliadas em listas de recomendação, isto é, na segunda etapa do experimento. Ao todo, 221 usuários interagiram com, pelo menos, um item da(s) lista(s) de recomendação através de *likes* e *dislikes*. Das 74 notícias falsas cadastradas, 40 foram recomendadas, enquanto que das 1864 notícias verdadeiras cadastradas 697 foram recomendadas. Isso se deve, em parte, à aleatoriedade empregada durante o oferecimento de notícias na fase de coleta de perfil do usuário. A média de notícias avaliadas por usuário foi de 24,11 (desvio padrão 32,77) e o número médio de listas de recomendação geradas foi de 3. Já a média de avaliações por notícia foi de 11,94 (desvio padrão 13,95). Destaca-se, ainda, que esses foram os dados utilizados na primeira etapa de avaliação de resultados (Seção 6.3) apresentados nesta tese, isto é, considerando a interação dos usuários com os itens recomendados.

A Figura 5.9 apresenta um gráfico com a distribuição de cauda longa para o *dataset* de recomendações. A partir dela, é possível observar a popularidade das 697 notícias a partir do número de avaliações (*ratings*) que cada uma recebeu. Apenas 67 itens foram considerados populares e, por isso, estão concentrados na “cabeça” da distribuição,

⁶<https://zenodo.org/>

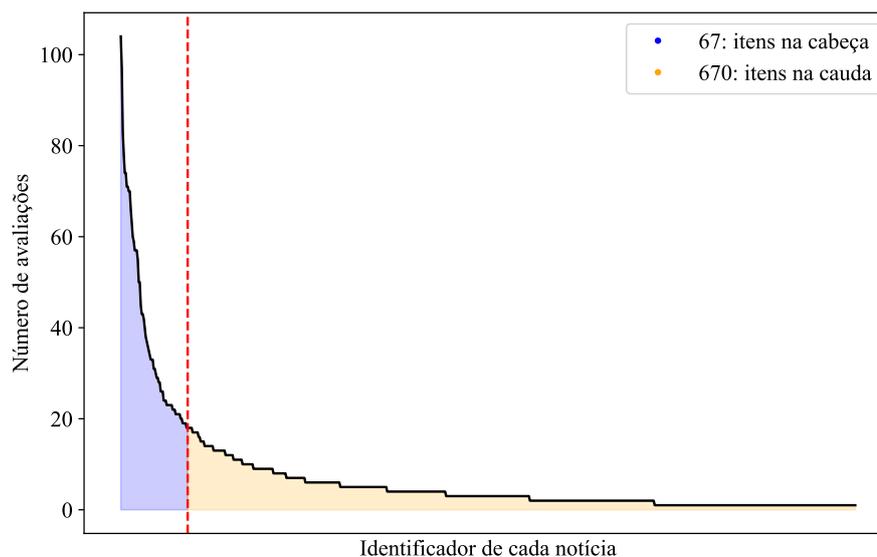
Tabela 5.3: Característica do *dataset* de recomendações

Atributo	Valor
Usuários que finalizaram o experimento	221
Notícias falsas	40
Notícias verdadeiras	697
Média de notícias avaliadas por usuário	24,11
Média de avaliações por notícia	11,94
Número médio de listas de recomendação por usuário	3

Fonte: O Autor.

possuindo o mínimo aproximado de 20 avaliações. Já na cauda, separada pela linha vermelha, se encontram os itens menos populares. Esse gráfico, portanto, revela a presença do efeito de cauda longa nas recomendações geradas.

Figura 5.9: Distribuição de cauda longa para o *dataset* de recomendações.



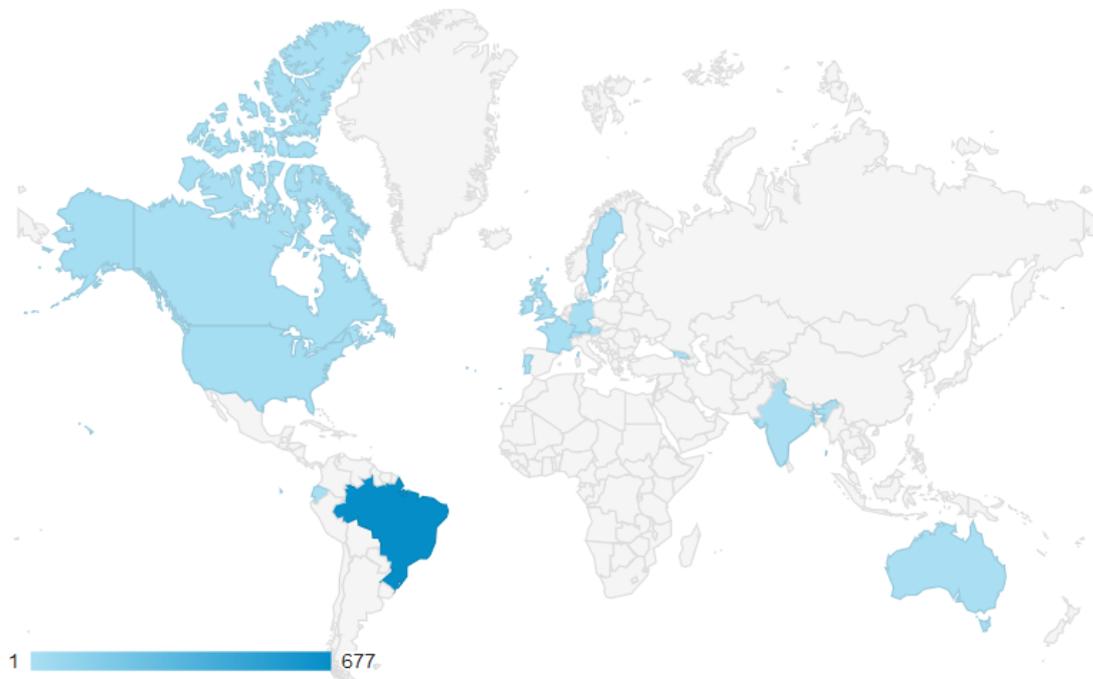
Fonte: O Autor.

Para a segunda etapa, ou seja, a avaliação das listas de recomendação (Seção 6.4) sem levar em conta a interação usuário-notícia, foram considerados 260 usuários, mais do que a quantidade considerada na etapa anterior. Isso se justifica porque foram incluídos os usuários que geraram listas de recomendação mas não avaliaram seus itens. Das 1938 notícias cadastradas no banco de dados, 1085 estão presentes em listas de recomendação sendo, 1015 verdadeiras e 70 falsas.

Apesar do experimento possuir exclusivamente conteúdo em português do Brasil, verificou-se uma certa diversidade demográfica na origem do tráfego, conforme mostra a

Figura 5.10. Os países coloridos em azul tiveram usuários acessando a aplicação e quanto mais escuro o azul, maior é a quantidade de acessos. Os top-5 países com mais sessões foram: Brasil com 677, Estados Unidos com 20, Alemanha com 5, Canadá com 4, e França com 4.

Figura 5.10: Acessos por país.



Fonte: O Autor

5.3 Resumo do capítulo

Neste capítulo foi apresentada a aplicação *web* para recomendação diversificada de notícias *NewsREC*, incluindo todos os detalhes de *design* e implementação de algoritmos, de interface gráfica e dos dados coletados. Os dados coletados estão disponíveis publicamente na plataforma *Zenodo* para outros pesquisadores interessados em testar suas abordagens relacionadas ao tema desta tese. Esses dados são utilizados para a avaliação da abordagem desenvolvida nesta tese reportada no próximo capítulo.

6 AVALIAÇÃO EXPERIMENTAL E RESULTADOS

Neste capítulo é apresentada a avaliação experimental da abordagem para a medição de bolhas de filtro bem como da abordagem para a diversificação multi-atributo de recomendação para a suavização do referido efeito. Nesse sentido, um experimento do tipo estudo de usuário foi empregado com duas intenções de avaliação. A primeira delas concentra-se na interação usuário-notícia, inclusive de notícias falsas, utilizando, para isso, dados coletados da aplicação *NewsREC*. A diversificação baseou-se em tópicos como atributos, um método já empregado na literatura servindo assim como um *baseline*.

A segunda intenção de avaliação possui dois objetivos: estimar o quanto a abordagem de diversificação multi-atributo é capaz de diminuir a homogeneidade das listas de recomendação quando comparada às recomendações comuns; e estimar o quanto a abordagem de diversificação multi-atributo é capaz de diminuir a homogeneidade das listas de recomendação quando comparada à diversificação de tópicos. Os resultados são reportados de forma fundamentada por diferentes testes de significância estatística, considerando o valor GH médio ou mediano dos usuários que utilizaram cada uma das diferentes abordagens de recomendação.

6.1 Estudo de Usuário

De forma geral, com este experimento, é avaliado o papel que os algoritmos de recomendação mais comuns têm na formação de bolhas de filtro. Também é verificado se uma estratégia de diversificação, seja ela multi-atributo ou de tópicos, pode fornecer recomendações menos homogêneas através do emprego da métrica GH, introduzida no Capítulo 4. De forma complementar, foi avaliado o potencial dos algoritmos de recomendação em colocar o usuário em contato maior ou menor com notícias falsas.

As eleições presidenciais brasileiras do ano de 2018 foram caracterizadas por um intenso uso de redes sociais e opiniões polarizadas (BROOKS; BOADLE, 2018). Nos últimos anos, a comunidade de sistemas de recomendação tem chamado a atenção para o papel que esses sistemas desempenham em tais cenários. Quando esses sistemas são usados para recomendar notícias, uma câmara de eco pode ser criada, ou seja, uma situação em que um usuário em uma bolha de filtro tem uma opinião reforçada pela comunidade ou pelas notícias sendo compartilhadas, mesmo que essa opinião não seja baseada na verdade, cooperando para a disseminação de notícias falsas.

Este cenário se mostrou potencial para a máxima participação das pessoas no experimento apresentado nesta seção, uma vez que o tema era algo atual e do interesse comum, principalmente por envolver notícias falsas e redes sociais, em especial o Facebook, elementos recorrentes durante as eleições de 2018. Para tanto, foi desenvolvida a aplicação *web NewsREC* com o intuito de simular um ambiente de rede social no qual os participantes foram convidados a ler e avaliar (através de *Likes* e *Dislikes*) as notícias recomendadas, sendo essas verdadeiras e/ou falsas, porém sem o conhecimento do participante.

É importante frisar que a aplicação desenvolvida não é uma rede social já que os usuários não interagem entre si, mas que simula tal. Isso porque os algoritmos de recomendação escolhidos são baseados em filtragem colaborativa, embora seja mais comum utilizar uma estratégia baseada em conteúdo, no domínio de notícias (KARIMI; JANNACH; JUGOVAC, 2018). Os motivos dessa escolha são: (i) isolar o papel que os algoritmos podem ter na criação das bolhas e levar os usuários a interagir com conteúdo falso; (ii) a filtragem colaborativa é a técnica mais popular nos sistemas de recomendação (RICCI et al., 2015); e (iii) seus algoritmos dependem da comunidade de usuários, parecendo mais apropriados para reproduzir o comportamento de uma rede social.

6.2 Protocolo Experimental

Neste experimento, os usuários foram separados em grupos de teste e controle, conforme já introduzido na seção anterior. Esse método é também conhecido como teste *A/B* (*All between*) no qual cada um dos grupos de sujeitos interage com experiências diferentes no intuito comparar resultados (GUNAWARDANA; SHANI, 2015). Neste estudo, o grupo de controle corresponde aos usuários que interagiram com as recomendações não diversificadas e, o grupo de teste, corresponde aos usuários que interagiram com as recomendações diversificadas.

A separação dos grupos foi realizada através de uma diretriz, mencionada por Gunawardana e Shani (2015), que é a definição de variáveis de controle. As variáveis de controle são importantes porque permitem rastrear os acontecimentos no sistema, por exemplo, se o conjunto de algoritmos não diversificados promoveu mais ou menos o efeito de bolha de filtro. Nesse sentido, a variável de controle *algorithm*, vista como um atributo do banco de dados apresentado no capítulo anterior, permite esse controle. A distribuição dos algoritmos em cada um desses grupos ocorre no momento do registro do

usuário no sistema.

Outra diretriz importante na condução deste tipo de teste se refere à interface gráfica que deve ser a mesma para ambos os grupos. De forma geral, vale ressaltar ainda que essa estratégia de avaliação foi adotada dada a sua ampla utilização e aceitabilidade na avaliação de sistemas de recomendação (GUNAWARDANA; SHANI, 2015).

Os esclarecimentos apresentados nos três parágrafos anteriores servem como protocolo geral para as duas análises indicativas da presença, ou não, de bolhas de filtro. Cada uma delas conta com especificidades de protocolo que foram empregadas de acordo com os objetos de análise em questão. O primeiro objeto envolve os itens recomendados e que receberam avaliações dos usuários, e o segundo objeto envolve as listas de recomendação sem considerar as interações. Os motivos dessa escolha estão melhor detalhados em cada um dos protocolos a seguir.

6.2.1 Interação de usuários

Para a análise do efeito de bolha de filtro gerado pelos algoritmos de recomendação, foram selecionados usuários dos grupos de teste e controle de cada algoritmo bem como todos os itens recomendados que receberam avaliação (notas). Com isso, foi aplicada a métrica para a medição de bolhas de filtro apresentada no Capítulo 4, GH, que mede a homogeneidade dos itens recomendados com os quais um usuário interagiu. O componente da métrica, responsável pelo cômputo da similaridade, foi o coeficiente de Jaccard que é adequado para quando os dados sendo computados são binários¹ (VERMA; AGGARWAL, 2020). Logo, isso se aplica neste caso, no qual os tópicos eram representados por uma matriz binária que indicava a presença ou não de um determinado tópico.

Nesta análise também foi considerada a acurácia das recomendações a fim de relacioná-la com o efeito de bolha de filtro, servindo como um indicador complementar. Nesse sentido, foi empregado o RMSE que mede o erro entre a nota predita pelo algoritmo para um item e a nota realmente dada por um usuário ao mesmo item. Valores maiores indicam menor acurácia enquanto valores menores indicam maior acurácia (HERLOCKER et al., 2004).

A fim de traçar conclusões fundamentadas da comparação de grupos, foram empregados testes paramétricos e não paramétricos de significância estatística de acordo com os seguintes critérios: distribuição de dados de cada grupo (normal, binomial, qui qua-

¹<https://ibm.co/3aywjQK>

drado, dentre outras); tipo de variável (contínua ou discreta); e natureza do grupo (pareado ou não pareado). Os grupos, nesta análise, se enquadram como grupos não pareados visto que os usuários foram atribuídos uma única vez a um algoritmo ou a outro.

De forma complementar, para avaliar se uma estratégia de diversificação seria capaz de diminuir a interação dos usuários com notícias falsas, foi adotado um cálculo de porcentagem de interação dos usuários com as notícias falsas. Para tanto, verificou-se o número de *likes*, *dislikes*, leituras, leituras + *likes*, e leituras + *dislikes*.

6.2.2 Listas de recomendação

Nesta análise, as listas de recomendação correspondem ao objeto para o estudo do efeito de bolha de filtro. Diferentemente da análise anterior, aqui não são considerados apenas os itens recomendados com os quais um usuário interagiu mas, sim, todos os itens presentes na última lista de recomendação *top-20* gerada por cada usuário. Isso foi possível graças ao atributo *generated_when*, do banco de dados, que armazenava o *timestamp* no momento que o usuário solicitava uma nova lista de recomendação. Essa escolha se justifica por conta da última lista possuir maior acurácia teórica em função do *feedback loop* (CHANEY; STEWART; ENGELHARDT, 2018; JIANG et al., 2019; MANSOURY et al., 2020). Por isso, teoricamente, essas listas possuem maior homogeneidade, o que pode contribuir para a formação de bolhas de filtro.

A abordagem de diversificação multi-atributo é experimentada nesta análise nos mesmos moldes metodológicos apresentados no protocolo anterior, isto é, os algoritmos não diversificados são comparados com os algoritmos diversificados. Entretanto, foram geradas novas listas diversificadas (abordagem multi-atributo) em relação àquelas existentes no banco de dados (abordagem de tópicos). A diversificação multi-atributo foi realizada utilizando a mesma lista não diversificada empregada pela diversificação de tópicos, ou seja, a lista *top-100* de entrada, gerada por algum dos algoritmos recomendação sem diversificação, requerida por algoritmos de diversificação de pós-filtragem. Isso só foi possível graças ao armazenamento dessa lista na tabela *RecommendationLists* no banco de dados a cada vez que um usuário diversificado solicitava uma nova lista de recomendação. A lista diversificada final *top-20* mostrada ao usuário também era armazenada na tabela, porém sinalizada com a *flag diversified*, o que permitiu fazer a distinção das listas para a recuperação, apenas, da lista *top-100*.

Dado o exposto, cada usuário foi avaliado de acordo com uma lista de recomen-

dação *top-20* cada um. Para os usuários não diversificados foram recuperadas as suas últimas listas *top-20* e, para os usuários diversificados, foram geradas novas listas diversificadas *top-20*, seguindo os mesmos fatores de diversificação, θ e λ , empregados pela diversificação de tópicos nos algoritmos TD e MMR. A partir disso, a métrica de homogeneidade GH foi empregada novamente, porém utilizando a similaridade de cossenos como componente de cálculo da similaridade. Essa escolha justifica-se em função da matriz de características não conter apenas dados binários (tópicos) mas também dados contínuos, especialmente em relação aos valores de polaridade e subjetividade das notícias. A similaridade de cossenos também se justifica por ser amplamente utilizada e aceita pela comunidade de sistemas de recomendação.

Nesta análise não foi possível empregar o RMSE para mostrar a relação entre o efeito de bolha de filtro e a acurácia. Isso porque não houve nova interação de usuários provendo avaliações às notícias, principalmente às novas listas diversificadas. Apesar de alguns itens em algumas das listas possuírem a nota real do usuário advindas do experimento anterior, não foi possível garantir a mesma situação para todas as listas. Logo, por questões de padronização, decidiu-se não avaliar a acurácia das recomendações.

A fim de traçar conclusões fundamentadas da comparação entre abordagens diversificadas e não diversificadas, foram empregados testes paramétricos e não paramétricos de significância estatística de acordo com os mesmos critérios empregados na análise anterior, isto é: distribuição dos dados, tipo de variável, e natureza de grupamento (neste caso grupos não pareados). Ainda, foi verificada a homogeneidade das listas de recomendação diversificadas pela abordagem baseada em tópicos como atributos servindo, essa, como abordagem *baseline*. Assim, foi possível comparar a influência, maior ou menor, no grau de homogeneidade da abordagem desta tese com a abordagem da literatura. Para isso, foram empregados testes de significância estatística pareados, isto é, os mesmo usuários que foram avaliados com a abordagem *baseline* também foram avaliados com a abordagem multi-atributo.

É válido pontuar que o algoritmo de diversificação “*Topic diversification - TD*”, apesar do nome fazer alusão ao conjunto de atributos empregado, isto é, tópicos, não o obriga a utilizar apenas esses (ZIEGLER et al., 2005). Inclusive, essa possibilidade é uma das responsáveis por permitir os vários testes realizados com esse algoritmo durante a condução desta tese.

6.3 Resultados envolvendo interação de usuários

Nesta seção, são reportados os resultados sobre a homogeneização e, consequentemente, a possibilidade de formação de bolha de filtro, considerando a interação dos usuários com as notícias recomendadas, verdadeiras e falsas. Detalhes sobre os procedimentos experimentais podem ser conferidos na Seção 6.2.1.

6.3.1 Formação de Bolha de Filtro

A Tabela 6.1 apresenta as estatísticas descritivas com relação à variável GH para cada algoritmo. Os valores de desvio padrão são baixos o que significa valores de GH próximos da média, com um mínimo de 0,72 e um máximo de 0,76. A variação pode parecer pequena à primeira vista, mas elas são examinadas com maior cuidado através dos testes de diferença estatística. Para a escolha adequada de cada teste, verificou-se o tipo de distribuição dos dados através do teste de normalidade *Shapiro-Wilk*, que resultou no p-valor. Levando em conta um nível de significância α de 0,05, os únicos algoritmos que não seguiram uma distribuição normal foram o “knnu mmr” e o “knni mmr” cujos valores de p-valor são menores que 0,05, em negrito na Tabela 6.1.

Tabela 6.1: Estatísticas descritivas da variável GH, considerando itens recomendados e interagidos, para cada um dos algoritmos.

Algoritmo	Usuários	Média	Mediana	Desvio Padrão	p-valor
knnu	23	0,744	0,751	0,0361	0,113
knnu td	24	0,753	0,755	0,0359	0,621
knnu mmr	20	0,723	0,738	0,0505	0,026
knni	33	0,760	0,761	0,0277	0,451
knni td	27	0,754	0,749	0,0246	0,861
knni mmr	22	0,727	0,729	0,0267	0,02
svd	25	0,737	0,732	0,0263	0,595
svd td	23	0,744	0,747	0,0333	0,424
svd mmr	20	0,736	0,737	0,0305	0,418

Fonte: o Autor.

É importante destacar que o número de usuários em cada algoritmo é diferente porque alguns usuários se registraram no sistema, porém não interagiram com as recomendações. Além disso, o leitor mais atento verificará que o número total de usuários não é igual a 221, conforme reportado na caracterização do *dataset* de recomendações an-

teriormente, mas sim igual a 217. Isso porque, para calcular a homogeneidade, é preciso que cada usuário tenha avaliado, pelo menos, duas notícias. Os usuários que avaliaram apenas uma notícia foram desconsiderados para não enviesarem os resultados caso fosse considerada uma homogeneidade máxima por terem interagido com apenas um item.

A Tabela 6.2 apresenta a comparação entre grupos de teste e controle, levando em conta a variável GH e o teste de significância estatística aplicado. O teste não-paramétrico *Mann-Whitney* foi adotado apenas quando os algoritmos “knnu mmr” e “knni mmr” estavam envolvidos, visto que seus dados não seguiam uma distribuição normal. O único algoritmo que mostrou potencial para diminuir a homogeneidade em relação aos itens com os quais um usuário interagiu foi o “knni” associado ao diversificador “MMR”, conforme destacado na tabela. O p-valor $< 0,001$ permite aceitar a hipótese de que existe diferença estatística significativa entre medianas, já que o p-valor é inferior à margem de erro estabelecida de 0,05. Retomando a Tabela 6.1, observa-se que a diferença se refere a um decréscimo da mediana de GH de 0,760 para 0,729 do algoritmo “knni” para o “knni mmr”, respectivamente. Com isso, deduz-se que há uma menor chance do algoritmo diversificado “knni mmr” colocar o usuário em uma bolha de filtro quando comparado com aqueles usuários que utilizaram o recomendador comum “knni”. Essa conclusão foi determinada a partir da análise das medianas porque é a medida mais adequada a se considerar quando os dados não seguem uma distribuição normal, isto é, ao redor da média, e um teste não paramétrico tal como o *Mann-Whitney* (SIEGEL; JR, 1975).

Tabela 6.2: Comparação entre grupos da variável GH, considerando itens recomendados e interagidos.

Controle	Teste	Teste estatístico	p-valor
knnu	knnu td	teste t	0,381
knnu	knnu mmr	<i>Mann-Whitney</i>	0,223
knni	knni td	teste t	0,388
knni	knni mmr	<i>Mann-Whitney</i>	0,001
svd	svd td	teste t	0,407
svd	svd mmr	teste t	0,908

Fonte: o Autor.

A mesma estratégia de testes foi aplicada entre diversificadores na intenção de avaliar suas potencialidades de homogeneização quando comparados entre si. O “knnu mmr” provou ser melhor que o “knnu td” com um p-valor = 0,04, diminuindo o GH médio de 0,753 para 0,723. O “knni mmr” foi melhor que o “knni td” com um p-valor $< 0,001$, diminuindo o GH médio de 0,754 para 0,727. Outros diversificadores não apresentaram

uma diferença significativa nas suas médias de GH, quando comparados entre si, e, por isso, não mereceram destaque.

A Tabela 6.3 apresenta as estatísticas descritivas para a variável RMSE. Essa variável assume valores entre 0 e 2 já que está relacionada às notas que os usuários deram a cada notícia na forma de *likes* (1) e *dislikes* (-1). A média é maior em todos os algoritmos diversificados quando comparada com suas versões não diversificadas. Esse era um comportamento esperado já que as recomendações diversificadas tendem a uma menor precisão (maior erro na predição de nota), tendo em vista a possível introdução de itens fora o espectro de interesse do usuário. Esse fato é conhecido na literatura como o clássico dilema entre acurácia e diversidade (MCNEE; RIEDL; KONSTAN, 2006). Para a variável RMSE, os algoritmos “knnu td”, “knnu mmr”, “knni”, e “knni td” não seguem uma distribuição normal dado que seus p-valores são inferiores a 0,05. O número de usuários foi suprimido desta tabela porque corresponde ao mesmo número apresentado na tabela referente à variável GH anteriormente.

Tabela 6.3: Estatísticas descritivas do RMSE para cada um dos algoritmos.

Algoritmo	Min.	Máx.	Média	Mediana	Desvio padrão	p-valor
knnu	0,087	1,55	0,911	0,894	0,428	0,245
knnu td	0	1,63	0,989	1,095	0,503	0,029
knnu mmr	0,247	1,418	1	1,080	0,326	0,042
knni	0	1,48	0,676	0,632	0,465	0,028
knni td	0	1,414	0,967	1,003	0,358	0,032
knni mmr	0	1,43	0,859	0,925	0,359	0,587
svd	0	1,07	0,705	0,719	0,291	0,175
svd td	0	1,51	0,734	0,712	0,329	0,927
svd mmr	0,13	1,36	0,761	0,820	0,316	0,293

Fonte: o Autor.

Na Tabela 6.4, são apresentadas as comparações entre grupos de teste e controle para a variável RMSE, utilizando testes de significância estatística de acordo com a distribuição de dados de cada algoritmo. O “knni” e o “knni td” apresentaram uma diferença estatística significativa entre as medianas do RMSE. Isso se confirma pelo p-valor = 0,014 < α . De acordo com Tabela 6.3, observa-se um aumento do RMSE, através da mediana em função do teste não paramétrico aplicado, de 0,632 (“knni”) para 1,003 (“knni td”), o que permite afirmar que “knni td” erra mais que sua versão não diversificada ‘knni’. Ao comparar as abordagens de diversificação entre si, não foram observadas diferenças estatísticas significativas. Outro resultado interessante se refere ao algoritmo “knni mmr” que foi capaz de diminuir a homogeneidade e, ao mesmo tempo, não apresentou um aumento

significativo do RMSE quando comparado ao “knni”.

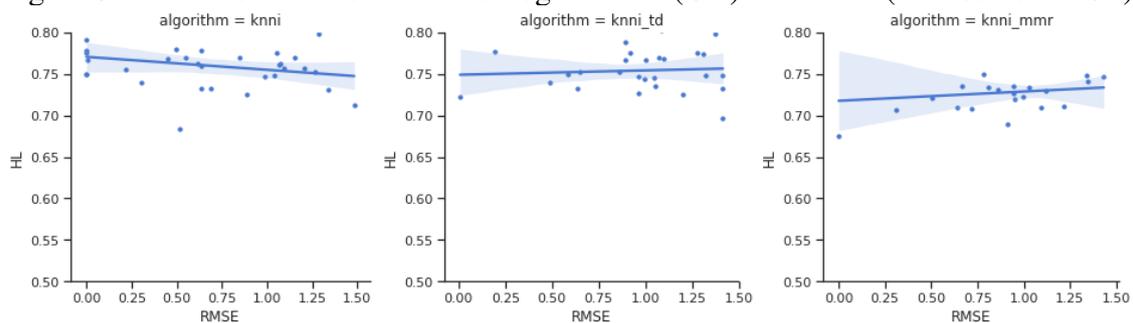
Tabela 6.4: Comparação entre grupos da variável RMSE.

Controle	Teste	Teste estatístico	p-valor
knnu	knnu td	<i>Mann-Whitney</i>	0,317
knnu	knnu mmr	<i>Mann-Whitney</i>	0,336
knni	knni td	<i>Mann-Whitney</i>	0,014
knni	knni mmr	teste t	0,126
svd	svd td	teste t	0,749
svd	svd mmr	teste t	0,542

Fonte: o Autor.

Após apresentar as comparações individuais das variáveis GH e RMSE, é apresentada, a seguir, a combinação dessas duas variáveis em um gráfico de dispersão no qual cada ponto é um usuário. O objetivo desse gráfico é servir como uma ferramenta complementar à homogeneidade para identificar a formação de bolhas de filtro quando a métrica RMSE estiver disponível. Para tanto, uma linha é definida para mostrar a tendência dos dados. A Figura 6.1 apresenta o gráfico de dispersão para os algoritmos que envolvem o “knni” apenas, dado que o “knni mmr” apresentou uma queda estatisticamente significativa da mediana de GH quando comparado ao “knni”. O gráfico referente ao algoritmo “knni” demonstra que o aumento da acurácia do algoritmo, isto é, o inverso ao RMSE, implica no aumento da homogeneidade dentro do conjunto de itens recomendados com os quais os usuários interagiram.

Figura 6.1: Relacionamento entre a homogeneidade (GH) e acurácia (inverso do RMSE).



Fonte: O Autor.

Por outro lado, os algoritmos “knni td” e “knni mmr” foram os únicos que não apresentaram uma tendência de dados positiva entre o aumento da acurácia e o aumento da homogeneidade. Neles, é possível perceber uma inversão no comportamento desses algoritmos quando comparados com sua versão não diversificada, ou seja, o conjunto de itens se torna mais homogêneo quando o algoritmo é menos preciso (valores mais altos

de RMSE). O “knni mmr” foi o algoritmo que apresentou os menores valores de GH para todos os seus usuários. A partir disso e, de forma geral, pode-se dizer que quando a acurácia é alta, o GH é menor de acordo com a tendência dos dados, não criando uma bolha. Com isso, tem-se um algoritmo que consegue produzir recomendações diversificadas e, ainda, com tendência de acurácia elevada.

6.3.2 Interação com notícias falsas

Os resultados referentes à interação com notícias falsas são apresentados pela Tabela 6.5 que sumariza as estatísticas de interação agrupadas por algoritmo. Os valores são apresentados em porcentagem em relação ao número total de interações registradas. As interações computadas na tabela são as seguintes: curtir - “Like” (L); não curtir - “Dislike” (D); um clique para abrir uma notícia que possivelmente indica uma leitura - “Reading” (R); a leitura R e um curtir L (R&L); uma leitura R e um não curtir D (R&D). Como um “curtir” em uma notícia pode representar uma maior probabilidade de alguém compartilhá-la, então valores mais baixos envolvendo essa estatística foram considerados como um bom efeito.

Tabela 6.5: Resultados de interação com notícias falsas. R&L representa a porcentagem de notícias lidas que também receberam um “Like”, enquanto R&D é a porcentagem de notícias lidas que também receberam um “Dislike”.

Grupo	Algoritmo	Like	Dislike	Read	R&L	R&D
Controle	knnu	64,71%	35,29%	18%	67%	33%
	knnu td	27,59%	72,41%	3%	0%	100%
	knnu mmr	37,50%	62,50%	6%	0%	100%
Controle	knni	40,82%	59,18%	6%	100%	0%
	knni td	73,33%	26,67%	33%	80%	20%
	knni mmr	33,33%	66,67%	33%	25%	75%
Controle	svd	73,68%	26,32%	16%	0%	100%
	svd td	83,33%	16,67%	0%	0%	0%
	svd mmr	100%	0%	0%	0%	0%

Fonte: O Autor.

Ao observar os valores na coluna “Like”, três abordagens de recomendação apresentaram valores menores comparadas às suas versões não diversificadas: “knnu td”, “knnu mmr”, e “knni mmr”. Esse último merece destaque, pois a abordagem que apresentou diferença estatística significativa nos valores de GH comparado ao “knni”. Em outras palavras, ele apresentou menor homogeneização, ou seja, menor tendência à formação

de bolha de filtro, fato que pode contribuir para diminuir a interação com notícias falsas. Em contrapartida, as outras três abordagens diversificadoras, isto é, “knni td”, “svd td”, e “svd mmr”, apresentaram maior porcentagem de curtidas quando comparadas às suas versões não diversificadas. A coluna “*Dislike*” apresenta os valores de porcentagem complementares à coluna “*Like*”.

A coluna “*Read*” contém a porcentagem do número total de possíveis leituras independente se o usuário curtiu (coluna “*Like*”) ou não curtiu (coluna “*Dislike*”). São observadas porcentagens baixas, o que é ruim, pois indica que os usuários curtiram e não curtiram várias notícias sem lerem o conteúdo. Isso fica ainda mais evidente ao analisar as versões diversificadas do algoritmo “svd”. Ambas aumentaram o número de curtidas, possivelmente devido a não leitura (0%) das notícias recomendadas. Quando um determinado usuário curte uma notícia apenas lendo o título e/ou imagem sem verificar o conteúdo, a possibilidade de espalhar possíveis notícias falsas é maior.

Por outro lado, o “knni mmr” apresentou a maior porcentagem de leituras “*Read*” (33%) dos quais 75% leram e não curtiram (R&D) as notícias falsas. Portanto, acredita-se que isso também contribuiu para que o algoritmo fosse hábil em diminuir a quantidade de curtidas (coluna “*Like*”). Afinal, quando o usuário abria a notícia era possível identificar sua falsidade e, assim, possivelmente avaliaria o item com uma não curtida.

Portanto, esses dados mostram que o efeito de bolha de filtro pode influenciar diretamente um aumento da interação de curtidas nas notícias falsas. Os usuários do único algoritmo que não apresentou tendência de formação de bolha, isto é, “knni mmr”, apresentaram menores níveis de curtidas em notícias falsas.

6.3.3 Discussão

A hipótese de que recomendar notícias de vários tópicos a fim de diminuir o efeito de bolha do filtro, ou seja, menor homogeneidade, foi parcialmente aceita. Isso porque apenas a estratégia de recomendação diversificada “knni mmr” foi capaz de diminuir a métrica GH (valores de média e mediana mais elevados) de forma estatisticamente significativa quando comparada com a sua versão não diversificada “knni”. De outra forma, era esperado que todas as abordagens de diversificação diminuíssem a homogeneidade, mas os testes de significância estatística não forneceram essa confirmação para todos os casos. Assim, esta seção de discussão foi pensada a fim apresentar algumas razões pelas quais acredita-se que isso aconteceu.

A primeira razão se refere à essência heurística de cada algoritmo diversificador, o MMR e o TD. Esses algoritmos contam com um fator de diversificação estático que varia de 0 a 1 cuja responsabilidade é controlar o nível de diversificação inserido nas lista de recomendações. Enquanto a acurácia aumenta, em função do *feedback loop* ou mesmo para um perfil inicial de usuário mais bem coletado, o fator de diversificação permanece estático, podendo resultar em um processo de diversificação ineficiente, especialmente em conjuntos de itens mais homogêneos.

A segunda razão depende das características dos itens, pois elas têm um papel central no processo de diversificação pós-filtragem. No caso dos resultados apresentados anteriormente, as características são tópicos, representados em uma matriz binária de itens (notícias) vs. tópicos. Muitas notícias apresentaram uma sobreposição de tópicos, ou seja, eram semelhantes entre si, embora fossem diferentes ao considerar seu conteúdo de texto. Assim, os diversificadores não poderiam promover a diversidade de notícias porque algumas delas eram muito semelhantes do ponto de vista de tópicos. Neste estudo, os tópicos foram extraídos automaticamente e talvez um método diferente, por exemplo, classificação humana, pudesse fornecer resultados diferentes. As características precisam ser de alta qualidade e distintivas para caracterizar um conjunto de dados, o que expande a discussão para uma área mais ampla, de *feature engineering*, que está fora do escopo desta tese.

Para exemplificar essa situação, foram analisados dois itens de notícias que compartilhavam o mesmo tópico. Identificou-se que a primeira notícia falava exclusivamente sobre um candidato presidencial de ideologia política de esquerda. Em contraste, a segunda notícia incluía outros candidatos com uma ideologia política oposta, de direita. Do ponto de vista do tópico, essas duas notícias são semelhantes, uma vez que compartilham o mesmo tópico na representação vetorial. No entanto, com um olhar mais atento aos seus conteúdos, foi identificada uma diferença clara: polaridades em relação ao tópico. Ao mesmo tempo que uma notícia apresenta um negativismo para um tópico, ela também apresenta um positivismo para outro. Em outras palavras, as notícias deixam de ter semelhanças e passam a apresentar uma diferença ao se considerar uma polaridade, o que poderia potencializar a possibilidade de colocar o usuário em contato com dois pontos de vista distintos sobre um mesmo assunto.

Em relação aos resultados de interação com notícias falsas, ressalta-se que eles são complementares à análise do efeito de bolha de filtro, não sendo o foco central de discussão deste trabalho. Apesar disso, obtiveram-se alguns resultados interessantes porque

a única estratégia que não apresentou tendência ao efeito bolha (GH significativamente inferior) “knni mmr” também apresentou uma diminuição na interação positiva do usuário (curtidas) com notícias falsas. Isso vai ao encontro das probabilidades e preocupações sobre uma bolha de filtro ser um ambiente onde notícias falsas podem se espalhar mais rapidamente.

Os algoritmos de diversificação que apresentaram maior GH, quando comparados com suas versões não diversificadas, apresentaram resultados divergentes. Por exemplo, a abordagem diversificada “svd td” aumentou a interação positiva com notícias falsas de 73,68% para 83,33% quando comparada com sua abordagem não diversificada “svd”. De outra forma, o “knnu td” diminuiu a mesma interação de 64,71% para 27,59%, o que significaria diversificação hábil em diminuir a interação positiva com notícias falsas, mesmo apresentando tendência à formação de bolha de filtro pela análise da variável GH. Essa divergência pode ser explicada pela capacidade de cada usuário em identificar ou não e, partir disso, curtir ou não curtir uma dada notícia falsa. Como esses algoritmos não apresentaram diferenças significativas no valor de GH, então, não é possível afirmar, para eles, que a diversificação foi capaz de diminuir a interação com notícias falsas.

O último ponto de discussão acerca dos resultados apresentados até aqui novamente tem a ver com as características de medição de homogeneidade. Acredita-se que outras características devem ser consideradas além de tópicos, especialmente no domínio de notícias no qual são abordados inúmeros assuntos. Com isso, a existência de uma abordagem que possibilite maior distinção entre os itens poderia potencializar as capacidades de diversificação. Nesse sentido, são apresentados os resultados a seguir.

6.4 Resultados envolvendo listas de recomendação

Esta seção apresenta os resultados sobre a homogeneização e, conseqüentemente, a possibilidade de formação de bolha de filtro, considerando a aplicação da abordagem desta tese considerando todos os itens de uma lista de recomendação, mesmo aqueles que não foram avaliados, isto é, sem interação com o usuário. Além disso, os valores de GH foram calculados de acordo com a similaridade de cossenos ao invés do índice de Jaccard. Por esses motivos, os resultados aqui apresentados não podem ser comparados diretamente com os resultados apresentados na seção anterior.

Os resultados estão relacionados com a análise de bolha de filtro através da métrica GH, a partir da qual são comparadas listas de recomendação com e sem diversificação

e, também, as listas diversificadas pela abordagem desta tese com as listas diversificadas pela abordagem *baseline*. Todos os detalhes sobre os procedimentos experimentais podem ser conferidos na Seção 6.2.2.

6.4.1 Formação de Bolha de Filtro

A Tabela 6.6 apresenta as estatísticas descritivas com relação à variável GH para cada algoritmo. Para a escolha adequada de cada teste, verificou-se o tipo de distribuição dos dados através do teste de normalidade *Shapiro-Wilk*, que resultou no p-valor. Levando em conta um nível de significância α de 0,05, todos os algoritmos seguiram uma distribuição normal, apresentando valores para o p-valor maiores que 0,05. É importante destacar que o número de usuários em cada algoritmo é diferente porque nem todos finalizaram o experimento. Além disso, o número total de usuários é igual a 260 porque foram incluídos aqueles que geraram listas de recomendação, mas não avaliaram seus itens. Lembrando que esses usuários haviam sido excluídos na análise da Seção 6.3.

Tabela 6.6: Estatísticas descritivas da variável GH, considerando listas de recomendação, para cada um dos algoritmos.

Algoritmo	Usuários	Média	Mediana	Desvio Padrão	p-valor
knnu	26	0,233	0,233	0,0260	0,980
knnu td	32	0,222	0,215	0,0300	0,168
knnu mmr	25	0,125	0,125	0,0134	0,635
knni	34	0,245	0,246	0,0343	0,152
knni td	32	0,235	0,238	0,0290	0,470
knni mmr	29	0,124	0,126	0,0134	0,142
svd	26	0,234	0,238	0,0252	0,111
svd td	29	0,235	0,230	0,0265	0,390
svd mmr	27	0,130	0,129	0,0133	0,932

Fonte: o Autor.

A Tabela 6.7 apresenta a comparação entre grupos de teste e controle, levando em conta a variável GH e o teste de significância estatística aplicado. Observa-se que o teste não-paramétrico *Mann-Whitney* foi adotado em três das seis comparações, mesmo todos os algoritmos tendo apresentado distribuições normais, conforme mencionado anteriormente. Isso pode ser explicado pelo p-valor de alguns algoritmos estarem mais próximos do $\alpha = 5\%$, estabelecido para o teste de normalidade individual. Assim, quando esses algoritmos estavam envolvidos em um teste de comparação, a junção dos dados convergia

para uma distribuição que não seguia uma curva Gaussiana.

Três abordagens de diversificação mostraram potencial para diminuir a homogeneidade em relação às listas de recomendação, sendo o “knnu”, o “knni”, e o “svd” associados ao diversificador “MMR”, conforme destacado na tabela. O p-valor $< 0,001$ permite aceitar a hipótese de que existe diferença estatística significativa entre medianas, já que o p-valor é inferior à margem de erro estabelecida de 0,05. Retomando a Tabela 6.6, observa-se que a diferença se refere a um decréscimo da mediana de GH de: 0,233 para 0,125 do algoritmo “knnu” para o “knnu mmr”; de 0,246 para 0,126 do algoritmo “knni” para o “knni mmr”; e de 0,238 para 0,129 do algoritmo “svd” para o “svd mmr”. Com isso, deduz-se que há uma menor chance das abordagens de diversificação multi-atributo colocarem o usuário em uma bolha de filtro quando comparadas com aquelas abordagens dos usuários que utilizaram o recomendador comum. Essa conclusão foi determinada a partir da análise das medianas porque é a medida mais adequada a se considerar quando os dados não seguem uma distribuição normal, isto é, ao redor da média, e um teste não paramétrico tal como o *Mann-Whitney* (SIEGEL; JR, 1975).

Tabela 6.7: Comparação entre grupos da variável GH, considerando listas de recomendação.

Controle	Teste	Teste estatístico	p-valor
knnu	knnu td	teste t	0,134
knnu	knnu mmr	<i>Mann-Whitney</i>	< 0,001
knni	knni td	teste t	0,218
knni	knni mmr	<i>Mann-Whitney</i>	< 0,001
svd	svd td	teste t	0,952
svd	svd mmr	<i>Mann-Whitney</i>	< 0,001

Fonte: o Autor.

Na Tabela 6.8, são apresentadas as estatísticas descritivas para a variável GH que servem de suporte, considerando listas de recomendação, para a comparação entre abordagens de diversificação multi-atributo e abordagens de diversificação *baseline*, comparação, essa, que será apresentada logo a seguir através dos testes de significância estatística. A palavra “tópicos” identifica a abordagem *baseline* enquanto a palavra “multi” identifica a abordagem desta tese. Para a escolha adequada de cada teste, verificou-se o tipo de distribuição dos dados através do teste de normalidade *Shapiro-Wilk*, que resultou no p-valor. Levando em conta um nível de significância α de 0,05, os únicos algoritmos que não seguiram uma distribuição normal foram o “knni td tópicos”, o “knnu mmr tópicos”, e o “knni mmr tópicos” cujos valores de p-valor são menores que 0,05. Além disso, a

igualdade do número de usuários entre abordagens sinaliza a adoção de testes estatísticos pareados, isto é, os mesmos usuários que utilizaram a abordagem *baseline* também foram submetidos à abordagem desta tese.

Tabela 6.8: Estatísticas descritivas da variável GH, considerando listas de recomendação, para a comparação entre abordagens de diversificação multi-atributo e abordagens de diversificação *baseline*.

Diversificador	Usuários	Média	Mediana	Desvio Padrão	p-valor
knnu td tópicos	32	0,170	0,166	0,0303	0,545
knnu td multi	32	0,221	0,215	0,0308	0,316
knni td tópicos	32	0,171	0,165	0,0371	0,004
knni td multi	32	0,235	0,239	0,0304	0,313
svd td tópicos	29	0,163	0,156	0,0214	0,508
svd td multi	29	0,237	0,231	0,0270	0,565
knnu mmr tópicos	25	0,140	0,134	0,0282	< 0,001
knnu mmr multi	25	0,125	0,125	0,0115	0,607
knni mmr tópicos	29	0,131	0,130	0,0134	0,041
knni mmr multi	29	0,125	0,127	0,0133	0,223
svd mmr tópicos	27	0,137	0,135	0,0166	0,215
svd mmr multi	27	0,131	0,129	0,0131	0,987

Fonte: o Autor.

A Tabela 6.9 apresenta a comparação entre a abordagem desta tese (coluna multi) e a abordagem empregada na literatura para a diversificação de recomendação (coluna *baseline*). Para tanto, foram computados os valores de GH para a abordagem *baseline*, considerando as listas de recomendação *top-20* bem como o emprego da similaridade de cossenos para que a comparação fosse justa, já que, na análise da Seção 6.3, foi empregado o coeficiente de Jaccard. A coluna “Teste estatístico” indica o teste de significância empregado de acordo com a distribuição dos dados das abordagens sendo comparadas. O “teste t pareado” sinaliza uma distribuição normal ao passo que o teste de “Wilcoxon” aponta o oposto. Esse último é equivalente ao teste “Mann-Whitney”, porém, para grupos pareados. As diferenças de GH são evidenciadas através da média (teste t pareado) ou da mediana (teste de “Wilcoxon”), apresentadas Tabela 6.8 (SIEGEL; JR, 1975).

Os resultados da Tabela 6.9 mostram que através do p-valor < 0,05 existe diferença significativa entre todas as abordagens, exceto entre “knni mmr tópicos” e “knni mmr multi”. Apesar da diferença não ser significativa, a abordagem “knni mmr multi” ainda foi capaz de diminuir a média GH de 0,131 para 0,125 em relação à abordagem “knni mmr tópicos”. Destaca-se o p-valor = 0,157 que ficou próximo do $\alpha = 0,05$, ou seja, próximo de uma diferença estatística significativa. Da mesma forma, porém com di-

Tabela 6.9: Comparação entre abordagens de diversificação multi-atributo e abordagens de diversificação *baseline*.

<i>Baseline</i>	Multi-atributo	Teste estatístico	p-valor
knnu td tópicos	knnu td multi	teste t pareado	< 0,001
knni td tópicos	knni td multi	teste t pareado	< 0.001
svd td tópicos	svd td multi	teste t pareado	< 0.001
knnu mmr tópicos	knnu mmr multi	<i>Wilcoxon</i>	0,027
knni mmr tópicos	knni mmr multi	teste t pareado	0,157
svd mmr tópicos	svd mmr multi	teste t pareado	0,042

Fonte: o Autor.

ferença estatística significativa, a abordagem “knnu mmr multi” diminuiu o GH mediano de 0,140 para 0,125. A mesma situação ocorreu para a abordagem “svd mmr multi” que diminuiu o GH médio de 0,137 para 0,131 em relação à “svd mmr tópicos”. Por outro lado, as três abordagens *baseline*, que empregaram o algoritmo TD, se mostraram superiores às suas respectivas abordagens multi-atributo no que se refere à diminuição do GH médio.

6.4.2 Discussão

A hipótese de que recomendar notícias diversificadas a partir de múltiplos atributos na intenção de diminuir o efeito de bolha do filtro, ou seja, menor homogeneidade, foi totalmente aceita. Isso porque todas as estratégias de recomendação diversificada foram capazes de diminuir os valores de média e/ou mediana da métrica GH, de forma estatisticamente significativa, quando comparadas com a suas versões não diversificadas. Esse resultado permitiu concluir que mesmo as listas de recomendação oferecendo itens mais diversos, os conjuntos de itens avaliados pelos usuários se mostraram mais homogêneos. Isso pode ser conferido na Tabela 6.1 da Seção 6.3 na qual constata-se que alguns dos conjuntos de itens diversificados apresentaram maior homogeneidade devido às escolhas de interação dos usuários. Em outras palavras, sugere-se que, independente da diversidade existente em uma lista de recomendação, a homogeneização está intimamente ligada ao desejo do usuário em permanecer na zona de conforto de sua bolha.

Em relação aos resultados comparativos entre abordagens de diversificação *baseline* e multi-atributo, era esperado que todas as abordagens de diversificação multi-atributo diminuíssem mais a homogeneidade, isto é, menores chances de formação de bolha de filtro. Entretanto, os testes de significância estatística não forneceram essa confirmação para

todos os casos. A principal razão pela qual acredita-se que isso aconteceu está relacionada às características das notícias. Sugere-se que apenas os tópicos já teriam sido suficientes para caracterizar de forma distinta os itens, tendo em vista que foram extraídos de forma automática e fundamentada pelo índice de coerência do algoritmo LDA. Por outro lado, as entidades nomeadas e sentimentos foram extraídas sem relação direta com o modelo de tópicos e isso pode ter contribuído para a não melhoria de todos os casos.

Ainda, uma extração de sentimentos a nível de aspecto, ao invés da extração a nível de documento, empregada nesta tese, também poderia ter contribuído para uma maior distinção entre notícias. Essa seria uma estratégia para lidar com a neutralidade existente textos jornalísticos envolvendo diferentes assuntos. Por fim, pontua-se a possibilidade de inclusão de outros atributos à abordagem que possam agregar vantagens ao processo de diversificação.

6.5 Resumo do Capítulo

Neste capítulo, foi apresentada a avaliação experimental de diferentes abordagens de diversificação, através da aplicação da métrica GH para e medição de bolhas de filtro. Nesse sentido, foi desenvolvido um sistema de recomendações de notícias, denominado *NewsREC*, envolvendo as eleições presidenciais brasileiras de 2018. Tal sistema fez parte de um estudo de usuário que possuiu duas intenções de avaliação. A primeira delas concentrou-se na interação usuário-notícia, inclusive de notícias falsas, diversificadas e não diversificadas. A diversificação baseou-se em tópicos como atributos, um método já empregado na literatura servindo, assim, como um *baseline*.

A segunda intenção de avaliação teve dois objetivos: estimar o quanto a abordagem de diversificação multi-atributo foi capaz de diminuir a homogeneidade das listas de recomendação quando comparada às recomendações comuns; e estimar o quanto a abordagem de diversificação multi-atributo foi capaz de diminuir a homogeneidade das listas de recomendação quando comparada à diversificação de tópicos. Os resultados do primeiro objetivo mostraram que houve uma diminuição dos valores de GH de forma significativa para todos os casos. Já os resultados do segundo objetivo mostraram que houve uma diminuição de GH, porém, não para todos os casos. Discussões foram estabelecidas de forma a destacar possíveis porquês acerca desses resultados.

7 CONCLUSÃO

Levando em consideração o potencial que os SR tem de entregar recomendações homogêneas, isto é, itens que são muito similares entre si, surge a problemática do efeito de bolha de filtro, especialmente crítico no domínio de notícias. Um ambiente altamente suscetível à polarização de ideias e opiniões e que pode ter consequências sociais danosas. Nesse sentido, esta tese apresentou uma abordagem para a diversificação de recomendação com o intuito de suavizar o efeito de bolha de filtro no domínio de notícias. Essa abordagem é composta de: um algoritmo de diversificação baseado em pós-filtragem; e um conjunto de múltiplos atributos envolvendo tópicos, polaridades, subjetividades e entidades nomeadas, todos extraídos dos textos das notícias. Para alcançar essa abordagem, primeiramente foi realizado um levantamento dos trabalhos relacionados e, na sequência, foi conduzido um estudo de usuário.

O estudo de usuário envolveu a construção de um sistema de recomendação de notícias, denominado *NewsREC*, referentes às eleições presidenciais brasileiras de 2018, cujo conteúdo foi coletado a partir de portais de notícias que mantinham postagens de seus conteúdos no *Facebook*. As recomendações geradas nesse sistema foram separadas em abordagens diversificadas e não diversificadas, sendo os usuários alocados em cada uma como grupos de teste e controle para fins de comparação de resultados.

Nesse sentido, os resultados mostraram que sim, a utilização de múltiplos atributos em algoritmos de diversificação de recomendação, diminui o efeito de bolha de filtro, respondendo a questão de pesquisa desta tese. Isso pôde ser verificado graças à medição da homogeneidade, através da métrica GH, em listas de recomendação ao comparar abordagens comuns de recomendação (algoritmos “knnu”, “knni”, e “svd”) com suas respectivas versões diversificadas, por meio de testes de significância estatística. Entretanto, ao estimar o quanto a abordagem de diversificação multi-atributo foi capaz de diminuir a homogeneidade das listas de recomendação quando comparada à diversificação de tópicos (*baseline*), não constatou-se uma diminuição significativa de GH para todos os casos, apenas para alguns.

Apesar da abordagem desta tese ser aplicada e testada no domínio de notícias, os resultados fornecem algumas pistas significativas que podem ajudar outros domínios de aplicação. Afirma-se isso porque os algoritmos de recomendação escolhidos são baseados em filtragem colaborativa, não sendo, portanto, dependentes do item sendo recomendado, o que significa que funcionam com qualquer tipo de item. Assim, o efeito de bolha de

filtro é suscetível de surgir em qualquer domínio de aplicação. Os resultados reportados neste trabalho são úteis para indicar a importância de se avaliar as recomendações por outros meios além da acurácia, conforme mencionado Kapoor et al. (2015) e realizado nesta tese, através da medição de bolha de filtro.

7.1 Contribuições

A seguir são destacadas as contribuições desta tese em duas grandes frentes. Na primeira, são pontuadas as contribuições diretamente relacionadas à abordagem desenvolvida e sua avaliação. Na segunda, são apresentadas as contribuições no âmbito acadêmico, isto é, as publicações e demais atividades realizadas que demonstram a extensão desta tese, seus relacionamentos, sua aplicabilidade, bem como o trabalho já desenvolvido e a experiência adquirida.

- Uma métrica, baseada em homogeneidade de itens, para a medição do efeito de bolha de filtro em sistemas de recomendação;
- O processo de extração de múltiplos atributos de notícias que permite sumarizar seus conteúdos;
- Uma abordagem multi-atributo para a diversificação de recomendação de notícias;
- Um sistema para a recomendação diversificada de notícias denominado *NewsREC* cujos detalhes e ideias de implementação apresentadas no Capítulo 5 podem auxiliar futuros pesquisadores interessados em desenvolver uma aplicação para a avaliação de suas abordagens;
- A realização de experimentos que avaliam a capacidade de diversificação da abordagem e a comparam, por meio de testes de significância estatística, com abordagens diversificadoras *baseline*, demonstrando melhorias significativas parciais em relação à não formação de bolhas de filtro;
- Um conjunto de dados que contém avaliações de usuários em notícias recomendadas, incluindo referência ao algoritmo responsável pela recomendação, seja ela diversificada ou não.
- Um conjunto de resultados de interação com as recomendações, com atenção às notícias falsas que foram obtidas através de uma parceria entre a UFRGS e a Agência LUPA, uma das maiores agências de checagem de fatos do Brasil. Recentemente, Cristina Tardáguila, fundadora da Agência LUPA, junto da rede internacional de

checadores de fatos foram indicados ao prêmio Nobel da Paz de 2021¹. Por isso, esta tese tem um impacto social significativo já que corrobora com resultados nesse sentido.

A seguir são apresentados os artefatos de pesquisa² já publicados, submetidos à publicação e/ou em processo de redação durante o período de doutorado, em ordem cronológica, do mais atual até o mais antigo. Ao todo, são dez publicações sendo, três em conferências e sete em *Journals*. Abaixo, são apresentadas as publicações, como primeiro autor, diretamente relacionadas à esta tese:

1. **A metric for Filter Bubble measurement in recommender algorithms considering the news domain** - este artigo trata de um estudo de usuário utilizando notícias veiculadas no Facebook utilizando as estratégias mais comuns de recomendação, as baseadas em filtragem colaborativa. A ideia central é avaliar e confirmar se, de fato, tais algoritmos tem influência na formação de bolhas de filtro e, como análise periférica, o quanto contribuem para a interação dos usuários com notícias falsas. Em outras palavras, busca-se analisar o comportamento dos usuários frente às recomendações e como as bolhas são formadas. (LUNARDI et al., 2020). Artigo publicado na *special issue* “Soft Computing for Recommender Systems and Sentiment Analysis” do *Journal Elsevier Applied Soft Computing*. QUALIS A1.
2. **Representing the Filter Bubble: Towards a Model to Diversification in News** - este artigo apresenta uma arquitetura e uma ideia de abordagem para diversificação de recomendação de notícias. A ideia central busca “romper” a bolha de filtro, utilizando de uma abordagem multi-atributo para a diversificação (LUNARDI, 2019). Foi publicado no Simpósio de Doutorado da *International Conference on Conceptual Modeling* (ER 2019).
3. **Diversificação de Recomendação em Cidades Inteligentes: Estudo e Estrutura de Abordagem** - este artigo apresenta um estudo de revisão preliminar de conceitos para diversificação de recomendação bem como uma abordagem inicial para diversificar itens para o domínio de cidades inteligentes, buscando auxiliar os cidadãos a consumirem itens mais diversos na intenção de fomentar o desenvolvimento próprio e da cidade em si. A partir disso, prospectou-se o domínio de notícias como uma fonte informacional permanente aos cidadãos a ser considerada como delimi-

¹<https://piaui.folha.uol.com.br/lupa/2021/01/21/ifcn-nobel-da-paz/>

²As publicações bem como número de citações podem ser conferidas no perfil do Google Scholar do autor: <<https://bit.ly/33ULD70>>

tação do escopo de pesquisa (LUNARDI et al., 2019). Foi publicado no volume 10, número 1 dos Cadernos de Informática do Instituto de Informática da UFRGS: Modelos e Sistemas para Cidades Inteligentes (2019).

A seguir, são apresentadas as publicações como primeiro autor, desenvolvidas durante o período de doutorado, porém não diretamente relacionadas ao tema desta tese.

1. **Probabilistic Ontology Reasoning in Ambient Assistance: Predicting Human Actions** - este artigo apresenta uma estratégia de recomendação de ações humanas com base em uma para auxiliar pessoas com declínio cognitivo no domínio de *Smart Homes*. Este foi um trabalho desenvolvido com cooperação internacional com a Universidade de Klagenfurt, na Áustria (LUNARDI et al., 2018b). Foi publicado na trilha principal da IEEE International Conference on Advanced Information Networking and Applications (2018). QUALIS A2;
2. **IoT-based human action prediction and support** - este artigo apresenta uma visão mais ampla da abordagem aplicada e apresentada no artigo anterior, mostrando desde a revisão de conceitos até os mecanismos de inferência. (LUNARDI et al., 2018). Foi publicado no Journal da Elsevier Internet of Things; Engineering Cyber Physical Human Systems (2018).
3. **Um Modelo Ontológico Probabilístico para Assistir Pessoas com Declínio Cognitivo** - este artigo apresenta a ontologia probabilística adotando um discussão semântica conceitual (LUNARDI et al., 2018a). Foi publicado na trilha principal do Seminário Brasileiro de Pesquisa em Ontologias - Ontobras (2018). QUALIS B3.

Por fim, são apresentadas as publicações concebidas em coautoria e colaboração com os trabalhos de outros colegas.

1. **Recommender Systems in IoT-driven Smart Cities: A systematic Mapping Towards a Service-based Architecture** - este artigo apresenta um mapeamento sistemático da literatura, buscando evidenciar o potencial de concepção de uma arquitetura baseada em serviços para prover dados a sistemas de recomendação em cidades inteligentes. Um agente de software controla as inscrições e solicitações de diferentes sistemas de recomendação e de diferentes provedores de dados. Um desses provedores é um agregador de notícias, responsável por entregar notícias representadas pelo modelo desta tese a um sistema de recomendação de notícias específico. Trabalho submetido ao Journal Elsevier Future Generation Computer Systems. QUALIS A1.

2. **AwARE: a Framework for Adaptive Recommendation of Educational rEsources** - este é um artigo escrito em colaboração com outros colegas, fruto da tese de Guilherme Medeiros Machado e atualmente professor da Universidade Lorraine, em Nancy, na França, que trata de uma abordagem de recomendação adaptativa de recursos educacionais (MACHADO et al., 2021). Trabalho publicado no Journal Springer Computing. QUALIS B1.
3. **Explicação de Recomendações com Diversificação: uma Revisão Bibliográfica** - este artigo faz um levantamento dos principais trabalhos que combinam dois temas de pesquisa em evidência em sistemas de recomendação: a explicação e a diversificação de recomendação. Ao mesmo tempo que se busca por itens mais diversos, também se busca entender os motivos que levaram o sistema a entregar aqueles itens. Esse é um grande desafio visto que grande parte dos algoritmos de recomendação, principalmente os baseados em aprendizado de máquina, são “caixas pretas” cujo funcionamento não é claro (DIAS et al., 2019). Trabalho publicado no volume 10, número 1 dos Cadernos de Informática do Instituto de Informática da UFRGS: Modelos e Sistemas para Cidades Inteligentes (2019).

As produções e os desdobramentos desta tese estão intimamente ligados aos projetos de pesquisa aos quais o autor está atrelado. O primeiro deles é o projeto CNPq, chamada universal, submetido pelo prof. orientador e, aprovado em 2017, denominado: Recomendação adaptativa para cidades inteligentes. Tem como objetivo recomendar recursos informacionais adaptados de acordo com o nível de conhecimentos e contexto do usuário. Acredita-se que a recomendação de recursos informacionais de tipos diversos pode estimular usuários de dispositivos móveis a tornarem-se mais inseridos socialmente em suas cidades, contribuindo assim para o aperfeiçoamento do nível de participação cidadã da população.

Em 2019, outro projeto de pesquisa, também sob chamada universal CNPq foi aprovado: Middleware de Gerenciamento de Dados para Aplicações Pervasivas em Campus Universitários Inteligentes. Este projeto foi submetido pelo prof. Vinícius Maran, egresso de doutorado, e colaborador com o qual mantém-se estreita colaboração.

7.2 Limitações, dificuldades e lições aprendidas

Durante a condução deste trabalho, várias dúvidas e dificuldades foram identificadas e estão aqui descritas para auxiliar futuras pesquisas na forma de lições aprendidas. Tal situação levou à tomada de algumas decisões que trouxeram limitações ao trabalho mas, ao mesmo tempo, abriram oportunidades para o desenvolvimento de trabalhos futuros, apresentados na próxima seção.

Uma das primeiras dificuldades encontradas foi a não disponibilização de dados do *Facebook*, especialmente dados referentes às recomendações do *feed* de notícias da rede social para o estudo de bolha de filtro. Para contornar a situação, foi desenvolvido, então, um sistema de recomendação, dentro de um estudo de usuário, para coletar tais dados, isto é, o item recomendado e o algoritmo responsável pela recomendação. Entretanto, por não ser um sistema de uso amplo, como o *Facebook*, o número de usuários não foi tão elevado como se gostaria, principalmente por não envolver notícias de última hora.

Antes da coleta/extração de qualquer dado da *web*, é preciso verificar questões de permissão de uso e redistribuição de dados. Essa é uma séria lição aprendida a partir da coleta das notícias que foram recomendadas durante o estudo de usuário. As notícias falsas, coletadas da Agência LUPA, por exemplo, exigiam a permissão explícita dos responsáveis do grupo UOL ao qual a agência pertence. Com isso, uma negociação foi estabelecida a fim de obter uma permissão por tempo determinado sob a condição de que o experimento deveria incluir um *disclaimer* avisando os usuários da não possibilidade de reusarem o conteúdo sem permissão prévia.

Determinar a polaridade das notícias pode ser citada como uma lição aprendida. Isso porque, diferentemente do esperado pelo senso comum, um mecanismo tradutor não prejudica a identificação de sentimentos. O estudo de Araújo, Pereira e Benevenuto (2020) sustenta tal afirmação, apresentando uma série de experimentos, sugerindo que a tradução automática do texto de entrada da língua específica para o inglês e, em seguida, a utilização dos métodos estado-da-arte criados para o inglês pode ser melhor que os métodos existentes desenvolvidos para uma língua específica.

O fato de não ter sido avaliada a opinião dos usuários sobre os itens recomendados, tanto os diversificados como os não diversificados, através de um formulário de satisfação, se configura como uma limitação do trabalho. Afinal, sem isso, é difícil estimar o quão úteis foram as recomendações para o usuário. Assim, uma lição aprendida é a importância de se avaliar a satisfação do usuário com as recomendações. Entretanto,

a decisão de não avaliar a satisfação foi tomada porque poderia aumentar o tempo que o usuário dispenderia com o experimento e, com isso, poderia prejudicar o engajamento dos usuários com o sistema.

A abordagem de diversificação multi-atributo não foi experimentada por usuários assim como foi a abordagem de diversificação *baseline*. Isso porque, a abordagem de diversificação multi-atributo foi concebida após a aplicação do experimento com usuários, utilizando a abordagem de diversificação *baseline*. A abordagem multi-atributo não foi aplicada a usuários pelo fato das notícias serem desatualizadas e, por isso, possivelmente haveria um baixo interesse dos usuários no experimento. Notícias mais atuais poderiam ter sido coletadas para solucionar esse problema, porém, isso implicaria em um custo elevado de tempo envolvendo: a coleta de novas notícias; a extração de novos atributos para a diversificação, a adequação da plataforma de recomendação; e a divulgação e o uso do sistema aos usuários.

7.3 Trabalhos Futuros

Como trabalhos futuros, aponta-se, como primeira possibilidade, um estudo para verificar o impacto que o uso de longo prazo das recomendações teria na formação de bolha de filtro de um usuário. Sugere-se, para tanto, o emprego de um sistema de recomendação com diversificação que fornecesse algum serviço de uso diário por um usuário como, por exemplo, um agregador de notícias. Especula-se que quanto mais o usuário provê *feedback* ao sistema mais esse “aprende” suas preferências e, possivelmente, homogeniza as recomendações, corroborando para o efeito de bolha de filtro. Esse fenômeno é conhecido como *feedback loop* e maiores detalhes sobre ele podem ser encontrados no trabalho de Chaney, Stewart e Engelhardt (2018).

Outra possibilidade de trabalho é aplicar a abordagem de diversificação multi-atributo em um sistema, podendo ser inclusive um sistema de longo prazo sugerido, no qual os usuários interajam com os itens recomendados, assim obter-se-ia uma análise de bolha de filtro em relação à interação dos usuários com os itens recomendados tal como apresentada para a abordagem *baseline*. Além disso, seria de extrema valia coletar a satisfação dos usuários em relação às recomendações, diversificadas e não diversificadas, estratégia que permitiria estimar e comparar a diversidade percebida e a utilidade das recomendações.

Seria interessante a concepção de um mecanismo capaz de detectar a formação

das bolhas de forma dinâmica. Uma variável poderia ser utilizada para ajustar o fator de diversificação conforme a bolha surge. Nesse sentido, o fator de diversificação dos algoritmos TD e MMR se mostram como essa possível variável.

Outra frente de trabalho identificada durante a condução desta pesquisa se refere às características dos itens, especialmente na sua identificação contínua, interativa e semi ou totalmente automatizada, especialmente no domínio de notícias. Em busca disso, são elencados, a seguir, alguns pontos de pesquisa que podem servir de alicerce para uma abordagem nessa direção, com o apontamento de alguns trabalhos norteadores, intimamente ligados a área de mineração de textos e/ou sistemas de recomendação.

- Empregar técnicas de enriquecimento semântico utilizando grafos baseados em conhecimento (*Knowledge graphs*), tais como a DBpedia e a Wikipedia, a fim de agregar informações externas às entidades nomeadas. A compreensão delas ajuda no entendimento do conteúdo das notícias de maneira mais rápida e fácil (LIU et al., 2020). Esse arcabouço pode ser empregado na desambiguação de entidades nomeadas (AL-MOSLMI et al., 2020). A ambiguidade é algo que pode prejudicar um algoritmo diversificador de recomendação, por exemplo, quando existem entidades distintas, mas que se referem a uma mesma entidade em uma base de conhecimento;
- Escalar e automatizar, dentro do possível, o processo de identificação de características. No domínio de notícias última hora, por exemplo, haverá uma exigência muito maior desse organismo computacional já que várias e várias notícias são publicadas em um pequeno espaço de tempo ao longo do dia (KARIMI; JANNACH; JUGOVAC, 2018; DIAKOPOULOS, 2019; SANCHEZ et al., 2020). Com isso, mais e mais características acabam surgindo. Outros domínios são mais estáticos como, por exemplo, filmes, cujas descrições não possuem grande rotatividade.

A fim de auxiliar o possível leitor interessado em seguir em alguma das frentes de trabalho atreladas à área de sistemas de recomendação, destaca-se, aqui, a Seção 2.1.5, apresentada na fundamentação conceitual, que discorre sobre oportunidades e desafios constantes e atuais.

REFERÊNCIAS

ABBAR, S. et al. Real-time recommendation of diverse related articles. **WWW 2013 - Proceedings of the 22nd International Conference on World Wide Web**, p. 1–11, 2013.

ABDOLLAHI, B.; NASRAOUI, O. Explainable matrix factorization for collaborative filtering. In: **Proceedings of the 25th International Conference Companion on World Wide Web**. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2016. (WWW '16 Companion), p. 5–6. ISBN 9781450341448.

Abirami, A. M.; Gayathri, V. A survey on sentiment analysis methods and approach. In: **2016 Eighth International Conference on Advanced Computing (ICoAC)**. [S.l.: s.n.], 2017. p. 72–76.

ABISHEVA, A.; GARCIA, D.; SCHWEITZER, F. When the filter bubble bursts: Collective evaluation dynamics in online communities. In: **Proceedings of the 8th ACM Conference on Web Science**. New York, NY, USA: Association for Computing Machinery, 2016. (WebSci '16), p. 307–308. ISBN 9781450342087.

ADAMOPOULOS, P.; TUZHILIN, A. On over-specialization and concentration bias of recommendations: Probabilistic neighborhood selection in collaborative filtering systems. In: **Proceedings of the 8th ACM Conference on Recommender Systems**. New York, NY, USA: Association for Computing Machinery, 2014. (RecSys '14), p. 153–160. ISBN 9781450326681.

Adomavicius, G.; Kwon, Y. Improving aggregate recommendation diversity using ranking-based techniques. **IEEE Transactions on Knowledge and Data Engineering**, v. 24, n. 5, p. 896–911, 2012.

Adomavicius, G.; Tuzhilin, A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. **IEEE Transactions on Knowledge and Data Engineering**, v. 17, n. 6, p. 734–749, June 2005. ISSN 2326-3865.

AL-MOSLMI, T. et al. Named entity extraction for knowledge graphs: A literature overview. **IEEE Access**, IEEE, v. 8, p. 32862–32881, 2020.

ALFONSECA, E.; MANANDHAR, S. An unsupervised method for general named entity recognition and automated concept discovery. In: **In: Proceedings of the 1st International Conference on General WordNet**. [S.l.: s.n.], 2002.

ARAÚJO, M.; PEREIRA, A.; BENEVENUTO, F. A comparative study of machine translation for multilingual sentence-level sentiment analysis. **Information Sciences**, Elsevier, v. 512, p. 1078–1102, 2020.

BARILE, F. et al. A news recommender system for media monitoring. In: **IEEE/WIC/ACM International Conference on Web Intelligence**. New York, NY, USA: Association for Computing Machinery, 2019. (WI '19), p. 132–140. ISBN 9781450369343.

BAUER, F. et al. **Handbook for Automatic Computation: Volume II: Linear Algebra**. Springer Berlin Heidelberg, 2012. (Grundlehren der mathematischen Wissenschaften). ISBN 9783642869402. Available from Internet: <<https://books.google.com.br/books?id=ebmrCAAQBAJ>>.

BENEVENUTO, F.; RIBEIRO, F.; ARAÚJO, M. Métodos para análise de sentimentos em mídias sociais. **Sociedade Brasileira de Computação**, 2015.

BICKEL, P.; DOKSUM, K. **Mathematical Statistics: Basic Ideas and Selected Topics**. Prentice Hall, 2001. (Mathematical Statistics: Basic Ideas and Selected Topics, v. 1). ISBN 9780138503635. Available from Internet: <<https://books.google.com.br/books?id=8poZAQAIAAJ>>.

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent dirichlet allocation. **J. Mach. Learn. Res.**, JMLR.org, v. 3, p. 993–1022, mar. 2003. ISSN 1532-4435.

BOBADILLA, J. et al. Recommender systems survey. **Knowledge-Based Systems**, v. 46, p. 109–132, jul 2013. ISSN 09507051.

BOUNTOURIDIS, D. et al. SIREN: A Simulation Framework for Understanding the Effects of Recommender Systems in Online News Environments. In: **Proceedings of the Conference on Fairness, Accountability, and Transparency**. New York, NY, USA: ACM, 2019. (FAT* '19), p. 150–159. ISBN 978-1-4503-6125-5.

BROOKS, B.; BOADLE, A. **Divisive Brazil election careens into 'dangerous' polarization**. 2018. Available from Internet: <<https://www.reuters.com/article/us-brazil-election/divisive-brazil-election-careens-into-dangerous-polarization-idUSKCN1LZ23S>>.

BURKE, R. Hybrid recommender systems: Survey and experiments. **User Modeling and User-Adapted Interaction**, v. 12, n. 4, p. 331–370, Nov 2002. ISSN 1573-1391.

CAÑAMARES, R.; CASTELLS, P.; MOFFAT, A. Offline evaluation options for recommender systems. **Information Retrieval Journal**, v. 23, n. 4, p. 387–410, 2020. ISSN 15737659.

CARBONELL, J.; GOLDSTEIN, J. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. In: **ACM SIGIR Forum**. [S.l.: s.n.], 1998. v. 51, n. 2, p. 209–210. ISSN 01635840.

CARVALHO, W. S. **Reconhecimento de entidades mencionadas em português utilizando aprendizado de máquina**. Thesis (PhD) — Universidade de São Paulo, 2012.

CHANEY, A. J. B.; STEWART, B. M.; ENGELHARDT, B. E. How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In: **Proceedings of the 12th ACM Conference on Recommender Systems**. New York, NY, USA: ACM, 2018. (RecSys '18), p. 224–232. ISBN 978-1-4503-5901-6.

CUSTÓDIO, M. **Facebook anuncia mudança no algoritmo: menos relevância para fanpages e mais para amigos e familiares**. 2018. Available from Internet: <<https://resultadosdigitais.com.br/blog/facebook-mudanca-algoritmo/>>.

DESARKAR, M. S.; SHINDE, N. Diversification in news recommendation for privacy concerned users. In: **2014 International Conference on Data Science and Advanced Analytics (DSAA)**. [S.l.]: IEEE, 2014. p. 135–141. ISBN 978-1-4799-6991-3.

Di Martino, S.; ROSSI, S. An architecture for a mobility recommender system in smart cities. **Procedia Computer Science**, v. 98, p. 425 – 430, 2016. ISSN 1877-0509.

DIAKOPOULOS, N. **Automating the News: How Algorithms Are Rewriting the Media**. Harvard University Press, 2019. ISBN 9780674976986. Available from Internet: <<https://books.google.com.br/books?id=mlWQDwAAQBAJ>>.

DIAS, A. S. et al. Explicação de recomendações com diversificação: uma revisão bibliográfica. **Cadernos de Informática – Modelos e Sistemas para Cidades Inteligentes**, v. 10, n. 1, p. 17, 2019.

DUNFORD, R.; SU, Q.; TAMANG, E. The pareto principle. University of Plymouth, 2014.

GALWAY, N. U. I. **XPLODIV: DIVERSIFICATION APPROACH FOR RECOMMENDER SYSTEMS**. [S.l.], 2015. v. 1, 133–150 p. Available from Internet: <<https://aran.library.nuigalway.ie/handle/10379/5081>>.

GAO, M.; DO, H. J.; FU, W.-T. Burst Your Bubble! An Intelligent System for Improving Awareness of Diverse Social Opinions. In: **Proceedings of the 2018 Conference on Human Information Interaction & Retrieval - IUI 18**. New York, New York, USA: ACM Press, 2018. (IUI '18), p. 371–383. ISBN 9781450349451.

GE, M.; GEDIKLI, F.; JANNACH, D. Placing High-Diversity Items in Top-N Recommendation Lists. In: CITESEER. **ITWP@ IJCAI**. [S.l.], 2011.

GHARAHIGHEHI, A.; VENS, C. Making session-based news recommenders diversity-aware. In: **CEUR Workshop Proceedings**. [S.l.: s.n.], 2020. v. 2758, p. 60–66. ISSN 16130073.

GHAZANFAR, M. A.; PRUGEL-BENNETT, A. A scalable, accurate hybrid recommender system. In: IEEE. **2010 Third International Conference on Knowledge Discovery and Data Mining**. [S.l.], 2010. p. 94–98.

GUESS, A.; NAGLER, J.; TUCKER, J. Less than you think: Prevalence and predictors of fake news dissemination on facebook. **Science Advances**, American Association for the Advancement of Science, v. 5, n. 1, 2019.

GUNAWARDANA, A.; SHANI, G. Evaluating recommender systems. In: _____. **Recommender Systems Handbook**. Boston, MA: Springer US, 2015. p. 265–308. ISBN 978-1-4899-7637-6.

Habibzadeh, H. et al. Soft sensing in smart cities: Handling 3vs using recommender systems, machine intelligence, and data analytics. **IEEE Communications Magazine**, v. 56, n. 2, p. 78–86, 2018.

HARAMBAM, J. et al. Designing for the better by taking users into account: A qualitative evaluation of user control mechanisms in (news) recommender systems. In:

Proceedings of the 13th ACM Conference on Recommender Systems. New York, NY, USA: Association for Computing Machinery, 2019. (RecSys '19), p. 69–77. ISBN 9781450362436.

HARPER, F. M.; KONSTAN, J. A. The MovieLens Datasets. **ACM Transactions on Interactive Intelligent Systems**, v. 5, n. 4, p. 1–19, jan 2016. ISSN 2160-6455.

HELBERGER, N.; KARPPINEN, K.; D'ACUNTO, L. Exposure diversity as a design principle for recommender systems. **Information Communication and Society**, Taylor & Francis, v. 21, n. 2, p. 191–207, 2018. ISSN 14684462.

HERLOCKER, J. L. et al. Evaluating collaborative filtering recommender systems. **ACM Trans. Inf. Syst.**, Association for Computing Machinery, New York, NY, USA, v. 22, n. 1, p. 5–53, jan. 2004. ISSN 1046-8188.

HU, L. et al. Graph neural news recommendation with long-term and short-term interest modeling. **Information Processing & Management**, v. 57, n. 2, p. 102142, 2020. ISSN 0306-4573.

HUNG, B. T. Integrating sentiment analysis in recommender systems. In: _____. **Reliability and Statistical Computing: Modeling, Methods and Applications.** Cham: Springer International Publishing, 2020. p. 127–137. ISBN 978-3-030-43412-0. Available from Internet: <https://doi.org/10.1007/978-3-030-43412-0_8>.

JANNACH, D.; ADOMAVICIUS, G. Recommendations with a Purpose. In: **Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16.** New York, New York, USA: ACM Press, 2016. p. 7–10. ISBN 9781450340359.

JANNACH, D. et al. **Recommender Systems.** Cambridge: Cambridge University Press, 2010. 1–335 p. ISBN 9780511763113.

JIANG, R. et al. Degenerate feedback loops in recommender systems. In: **Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society.** [S.l.: s.n.], 2019. p. 383–390.

KAPOOR, K. et al. "i like to explore sometimes". In: **Proceedings of the 9th ACM Conference on Recommender Systems - RecSys '15.** New York, New York, USA: ACM Press, 2015. p. 19–26. ISBN 9781450336925.

KARIMI, M.; JANNACH, D.; JUGOVAC, M. News recommender systems – Survey and roads ahead. **Information Processing & Management**, v. 54, n. 6, p. 1203–1227, 2018. ISSN 0306-4573.

KITCHENHAM, B.; BRERETON, P. A systematic review of systematic review process research in software engineering. **Information and Software Technology**, v. 55, n. 12, p. 2049 – 2075, 2013. ISSN 0950-5849.

KNIJNENBURG, B. P.; KOBSA, A. Making decisions about privacy: Information disclosure in context-aware recommender systems. **ACM Trans. Interact. Intell. Syst.**, Association for Computing Machinery, New York, NY, USA, v. 3, n. 3, oct. 2013. ISSN 2160-6455.

- KOHAVI, R. Online controlled experiments: Lessons from running a/b/n tests for 12 years. In: **Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2015. (KDD '15), p. 1. ISBN 9781450336642.
- KOHAVI, R. et al. Online controlled experiments at large scale. In: **Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2013. (KDD '13), p. 1168–1176. ISBN 9781450321747.
- KOHAVI, R. et al. Controlled experiments on the web: survey and practical guide. **Data Mining and Knowledge Discovery**, v. 18, n. 1, p. 140–181, feb. 2009. ISSN 1573-756X.
- KOREN, Y. The bellkor solution to the netflix grand prize. In: . [S.l.: s.n.], 2009.
- KOREN, Y.; BELL, R. Advances in Collaborative Filtering. In: _____. **Recommender Systems Handbook**. Boston, MA: Springer US, 2015. p. 77–118. ISBN 978-1-4899-7637-6.
- KOREN, Y.; BELL, R.; VOLINSKY, C. Matrix Factorization Techniques for Recommender Systems. **Computer**, v. 42, n. 8, p. 30–37, aug 2009. ISSN 0018-9162.
- LIU, B. Sentiment analysis and subjectivity. In: **Handbook of Natural Language Processing, Second Edition**. Taylor and Francis Group, Boca. [S.l.: s.n.], 2010.
- LIU, D. et al. KRED: Knowledge-Aware Document Representation for News Recommendations. p. 200–209, 2020.
- LU, F.; DUMITRACHE, A.; GRAUS, D. Beyond Optimizing for Clicks: Incorporating Editorial Values in News Recommendation. In: **Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization**. New York, NY, USA: ACM, 2020. p. 145–153. ISBN 9781450368612. ISSN 23318422.
- LUNARDI, G. M. Representing the filter bubble: Towards a model to diversification in news. In: SPRINGER. **International Conference on Conceptual Modeling**. [S.l.], 2019. p. 239–246.
- LUNARDI, G. M. et al. Um modelo ontológico probabilístico para assistir pessoas com declínio cognitivo. In: **Seminário Brasileiro de Pesquisa em Ontologias**. [S.l.: s.n.], 2018. v. 2228, n. 10.
- LUNARDI, G. M. et al. Probabilistic ontology reasoning in ambient assistance: Predicting human actions. In: **2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA)**. [S.l.: s.n.], 2018. p. 593–600.
- LUNARDI, G. M. et al. A metric for filter bubble measurement in recommender algorithms considering the news domain. **Applied Soft Computing**, p. 106771, 2020. ISSN 1568-4946.
- LUNARDI, G. M. et al. Iot-based human action prediction and support. **Internet of Things**, Elsevier, v. 3, p. 52–68, 2018.

LUNARDI, G. M. et al. Diversificação de recomendação em cidades inteligentes: Estudo e estrutura de abordagem. **Cadernos de Informática – Modelos e Sistemas para Cidades Inteligentes**, v. 10, n. 1, p. 17, 2019.

LUNARDI, G. M.; OLIVEIRA, J. P. M. de. **NewsREC dataset: news recommendation and diversification**. Zenodo, 2021. Available from Internet: <<https://doi.org/10.5281/zenodo.4604008>>.

M Blei, D. Introduction to Probabilistic Topic Models. **Communications of the ACM**, v. 55, n. 4, p. 77–84, 2011.

MACHADO, G. M. et al. AwARE: a framework for adaptive recommendation of educational resources. **Computing**, 2021. ISSN 1436-5057.

MANSOURY, M. et al. Feedback loop and bias amplification in recommender systems. In: **Proceedings of the 29th ACM International Conference on Information & Knowledge Management**. [S.l.: s.n.], 2020. p. 2145–2148.

MCNEE, S. M.; RIEDL, J.; KONSTAN, J. A. Being accurate is not enough: How accuracy metrics have hurt recommender systems. In: **CHI '06 Extended Abstracts on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2006. (CHI EA '06), p. 1097–1101. ISBN 1595932984.

MIKHEEV, A.; MOENS, M.; GROVER, C. Named entity recognition without gazetteers. In: **Ninth Conference of the European Chapter of the Association for Computational Linguistics**. Bergen, Norway: Association for Computational Linguistics, 1999. p. 1–8. Available from Internet: <<https://www.aclweb.org/anthology/E99-1001>>.

MÖLLER, J. et al. Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. **Information Communication and Society**, Routledge, v. 21, n. 7, p. 959–977, 2018. ISSN 1369118X.

NADEAU, D.; SEKINE, S. A survey of named entity recognition and classification. **Linguisticae Investigationes**, John Benjamins, v. 30, n. 1, p. 3–26, 2007. ISSN 0378-4169. Available from Internet: <<https://www.jbe-platform.com/content/journals/10.1075/li.30.1.03nad>>.

NAGULENDRA, S.; VASSILEVA, J. Understanding and controlling the filter bubble through interactive visualization: A user study. In: **Proceedings of the 25th ACM Conference on Hypertext and Social Media**. New York, NY, USA: ACM, 2014. (HT '14), p. 107–115. ISBN 978-1-4503-2954-5.

NECHUSHTAI, E.; LEWIS, S. C. What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. **Computers in Human Behavior**, v. 90, p. 298–307, jan 2019. ISSN 07475632.

NGUYEN, T. T. et al. Exploring the filter bubble: The effect of using recommender systems on content diversity. In: **Proceedings of the 23rd International Conference on World Wide Web**. New York, NY, USA: ACM, 2014. (WWW '14), p. 677–686. ISBN 978-1-4503-2744-2.

- NOIA, T. D. et al. Adaptive multi-attribute diversity for recommender systems. **Information Sciences**, v. 382-383, p. 234 – 253, 2017. ISSN 0020-0255.
- PANKE, S.; STEPHENS, J. Beyond the echo chamber: Pedagogical tools for civic engagement discourse and reflection. **Educational Technology & Society**, v. 21, p. 248–263, 01 2018.
- PARISER, E. **The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think**. [S.l.]: Penguin Publishing Group, 2011. ISBN 9781101515129.
- PASSE, J.; DRAKE, C.; MAYGER, L. Homophily, echo chambers, & selective exposure in social networks: What should civic educators do? **The Journal of Social Studies Research**, v. 42, n. 3, p. 261 – 271, 2018. ISSN 0885-985X.
- PAZZANI, M. J.; BILLSUS, D. Content-based recommendation systems. In: _____. **The Adaptive Web: Methods and Strategies of Web Personalization**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 325–341. ISBN 978-3-540-72079-9.
- PLETTENBERG, N. et al. User behavior and awareness of filter bubbles in social media. In: SPRINGER. **International Conference on Human-Computer Interaction**. [S.l.], 2020. p. 81–92.
- Qi, L. et al. Privacy-aware cross-platform service recommendation based on enhanced locality-sensitive hashing. **IEEE Transactions on Network Science and Engineering**, p. 1–1, 2020.
- QUIJANO-SÁNCHEZ, L. et al. Recommender systems for smart cities. **Information Systems**, v. 92, p. 101545, 2020. ISSN 0306-4379.
- RESNICK, P. et al. GroupLens: An open architecture for collaborative filtering of netnews. In: **Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work**. New York, NY, USA: Association for Computing Machinery, 1994. (CSCW '94), p. 175–186. ISBN 0897916891.
- RICCI, F. et al. **Recommender Systems Handbook**. Boston, MA: Springer US, 2015. 9–16 p. ISSN 16130073. ISBN 978-1-4899-7636-9.
- RÖDER, M.; BOTH, A.; HINNEBURG, A. Exploring the space of topic coherence measures. In: **Proceedings of the Eighth ACM International Conference on Web Search and Data Mining**. New York, NY, USA: ACM, 2015. (WSDM '15), p. 399–408. ISBN 978-1-4503-3317-7.
- SANCHEZ, L. et al. Easing legal news monitoring with learning to rank and bert. In: JOSE, J. M. et al. (Ed.). **Advances in Information Retrieval**. Cham: Springer International Publishing, 2020. p. 336–343. ISBN 978-3-030-45442-5.
- SARIKI, T. P.; G, B. K. et al. A book recommendation system based on named entities. **Annals of Library and Information Studies (ALIS)**, v. 65, n. 1, p. 77–82, 2018.
- SARWAR, B. et al. Item-based collaborative filtering recommendation algorithms. In: **Proceedings of the 10th International Conference on World Wide Web**. New York, NY, USA: Association for Computing Machinery, 2001. (WWW '01), p. 285–295. ISBN 1581133480.

SEKINE, S.; RANCHHOD, E. **Named Entities: Recognition, classification and use**. John Benjamins Publishing Company, 2009. (Benjamins Current Topics). ISBN 9789027289223. Available from Internet: <<https://books.google.com.br/books?id=FVI6AAAAQBAJ>>.

SIEGEL, S.; JR, N. J. C. **Estatística não-paramétrica para ciências do comportamento**. [S.l.]: Artmed Editora, 1975.

SIEVERT, C.; SHIRLEY, K. LDAvis: A method for visualizing and interpreting topics. In: **Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces**. Baltimore, Maryland, USA: Association for Computational Linguistics, 2014. p. 63–70.

SONODA, A.; SEKI, Y.; TORIUMI, F. Analysis of Factors That Affect Users' Behavioral Changes in News Service. In: **IEEE/WIC/ACM International Conference on Web Intelligence - Companion Volume**. New York, NY, USA: ACM, 2019. (WI '19 Companion), p. 35–42. ISBN 978-1-4503-6988-6.

TAKÁCS, G. et al. Scalable collaborative filtering approaches for large recommender systems. **The Journal of Machine Learning Research**, JMLR. org, v. 10, p. 623–656, 2009.

TETLOCK, P. E. **Expert Political Judgment: How Good Is It? How Can We Know?** Stu - student edition. [S.l.]: Princeton University Press, 2005. ISBN 9780691128719.

TINTAREV, N. et al. Same, Same, but Different: Algorithmic Diversification of Viewpoints in News. In: **Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization**. New York, NY, USA: ACM, 2018. (UMAP '18), p. 7–13. ISBN 978-1-4503-5784-5.

VARGAS, S. Novelty and diversity enhancement and evaluation in recommender systems and information retrieval. In: **Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval**. New York, NY, USA: ACM, 2014. (SIGIR '14), p. 1281–1281. ISBN 978-1-4503-2257-7.

VARGAS, S. S. **Novelty and Diversity Evaluation and Enhancement in Recommender Systems**. 5–8 p. Thesis (PhD) — Universidad Autónoma de Madrid, 2012.

VERMA, V.; AGGARWAL, R. K. A comparative analysis of similarity measures akin to the jaccard index in collaborative recommendations: empirical and theoretical perspective. **Social Network Analysis and Mining**, Springer, v. 10, p. 1–16, 2020.

VICARIO, M. D. et al. The spreading of misinformation online. **Proceedings of the National Academy of Sciences**, National Academy of Sciences, v. 113, n. 3, p. 554–559, 2016. ISSN 0027-8424.

ZHANG, F. et al. Privacy-aware smart city: A case study in collaborative filtering recommender systems. **Journal of Parallel and Distributed Computing**, v. 127, p. 145 – 159, 2019. ISSN 0743-7315.

ZHOU, T. et al. Solving the apparent diversity-accuracy dilemma of recommender systems. **Proceedings of the National Academy of Sciences**, National Academy of Sciences, v. 107, n. 10, p. 4511–4515, 2010. ISSN 0027-8424.

ZIEGLER, C.-N. et al. Improving recommendation lists through topic diversification. In: **Proceedings of the 14th international conference on World Wide Web - WWW '05**. New York, New York, USA: ACM Press, 2005. p. 22. ISBN 1595930469.

APÊNDICE A — PALAVRAS-CHAVE DO MODELO DE TÓPICOS LDA

Figura A.1: Os 30 termos mais relevantes para o **tópico 0**

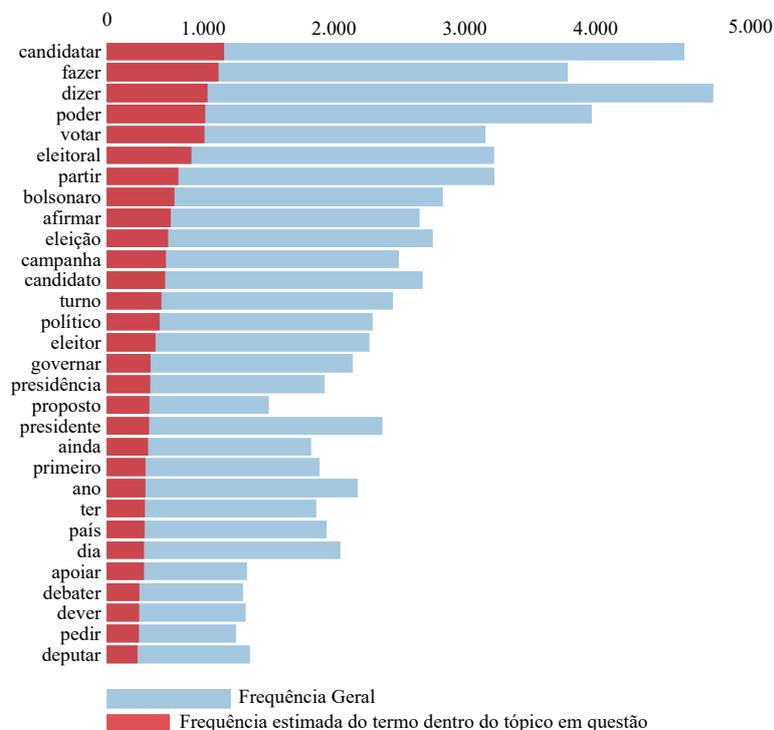


Figura A.2: Os 30 termos mais relevantes para o **tópico 1**

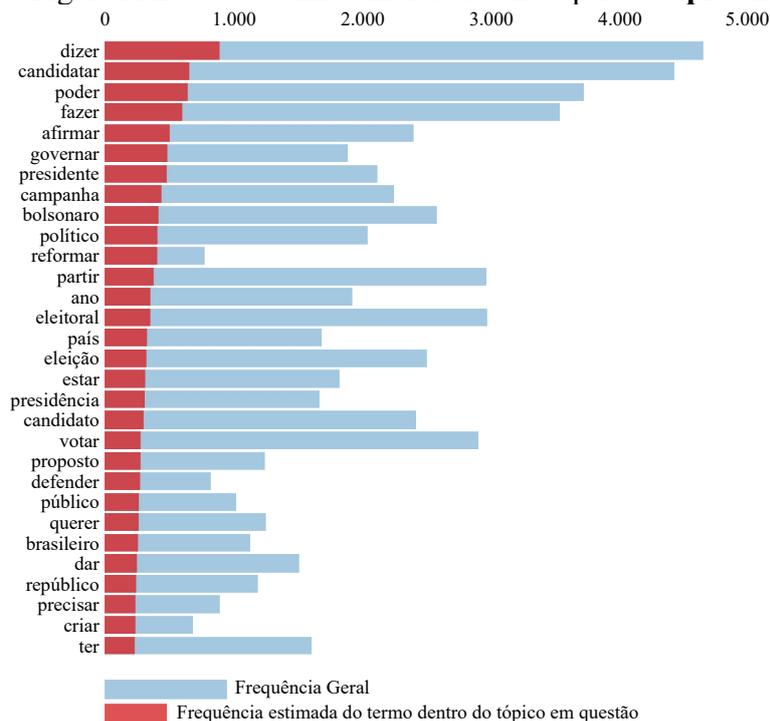


Figura A.3: Os 30 termos mais relevantes para o **tópico 2**

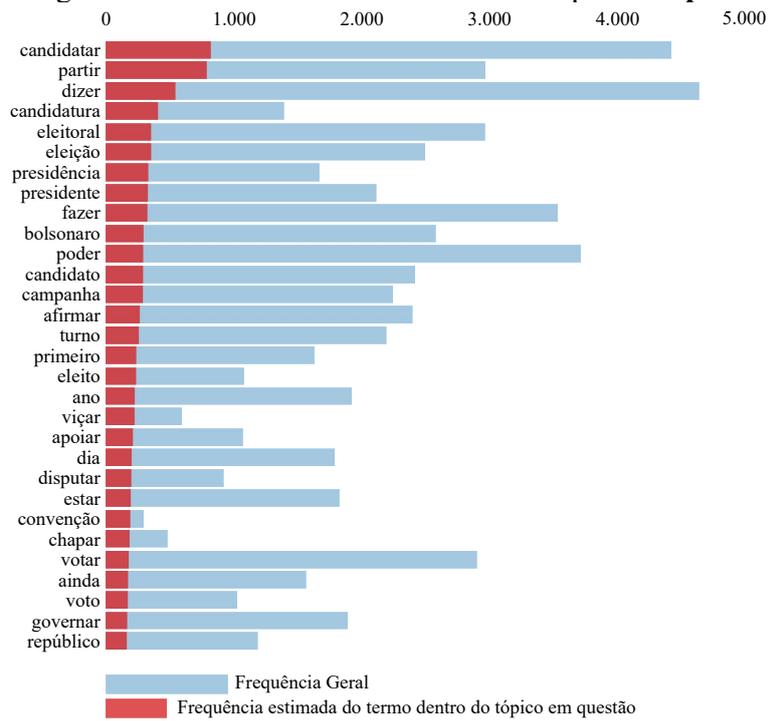


Figura A.4: Os 30 termos mais relevantes para o **tópico 3**

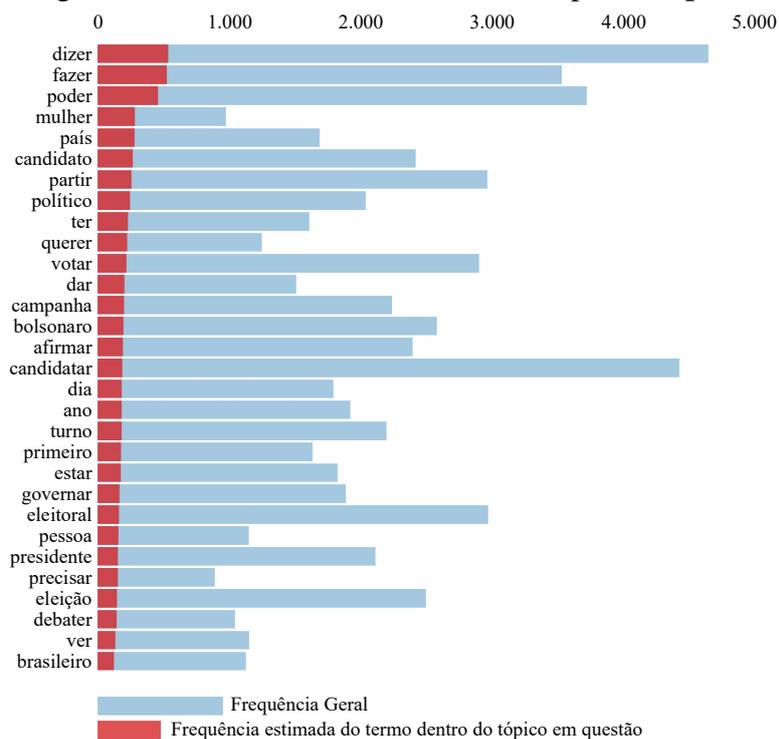


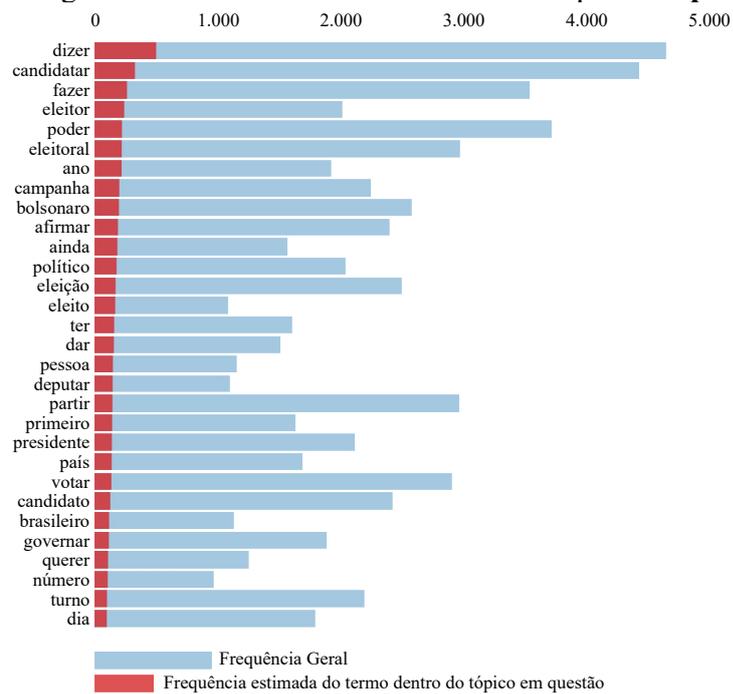
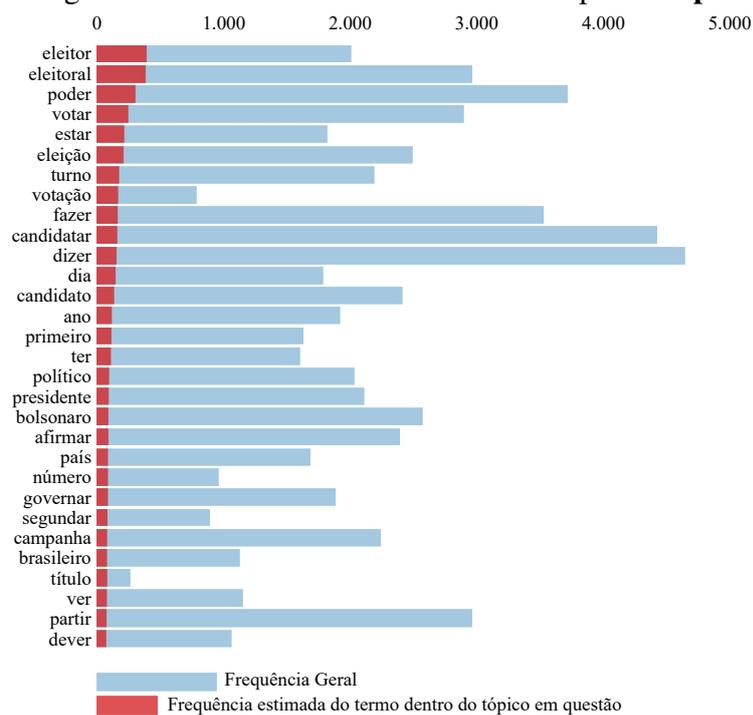
Figura A.5: Os 30 termos mais relevantes para o **tópico 4**Figura A.6: Os 30 termos mais relevantes para o **tópico 5**

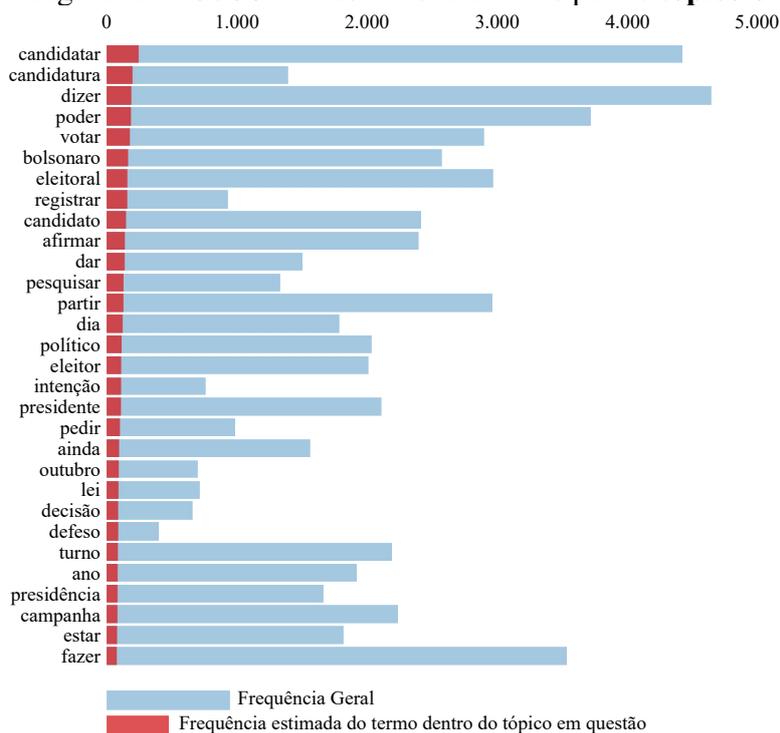
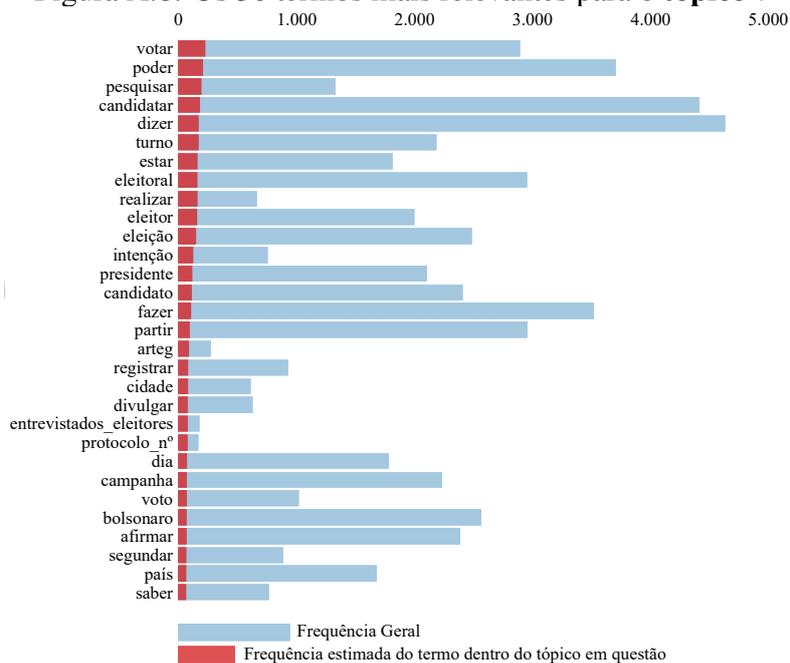
Figura A.7: Os 30 termos mais relevantes para o **tópico 6**Figura A.8: Os 30 termos mais relevantes para o **tópico 7**

Figura A.9: Os 30 termos mais relevantes para o **tópico 8**

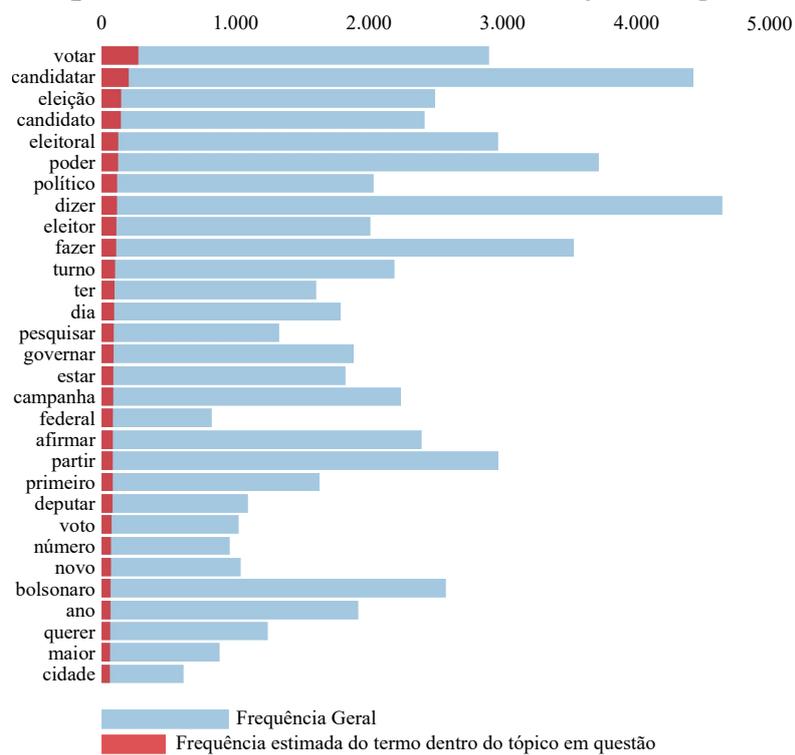


Figura A.10: Os 30 termos mais relevantes para o **tópico 9**

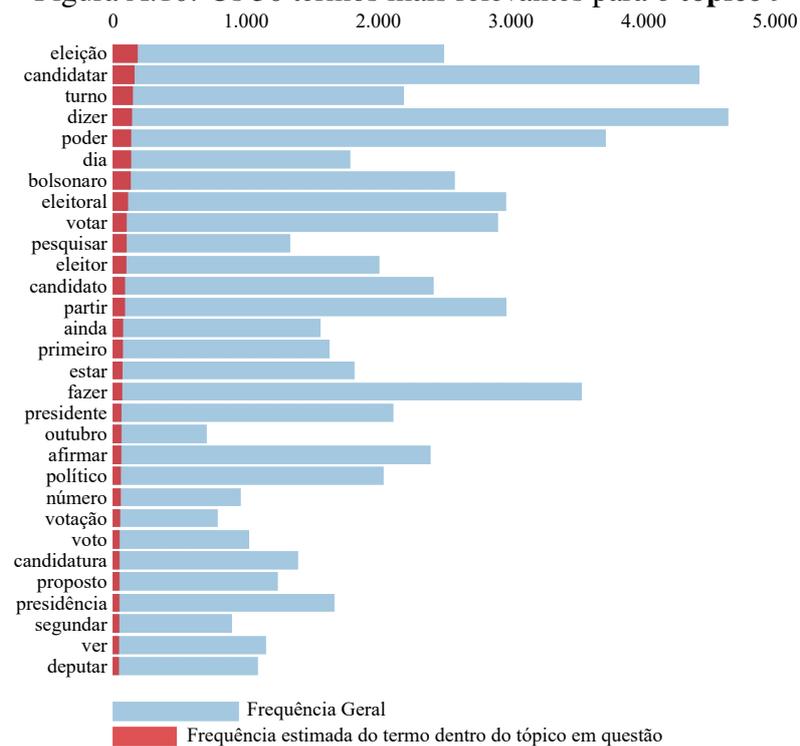


Figura A.11: Os 30 termos mais relevantes para o **tópico 10**

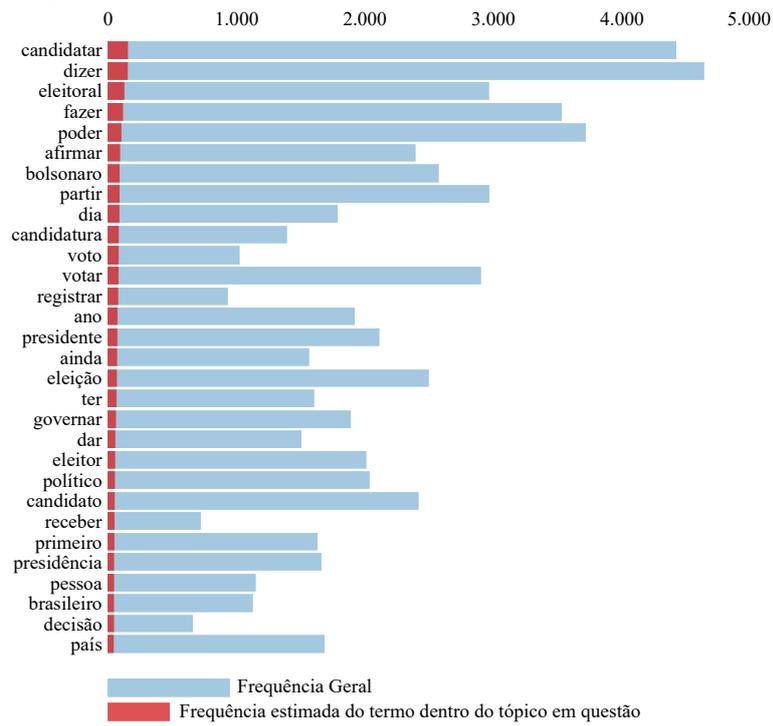


Figura A.12: Os 30 termos mais relevantes para o **tópico 11**

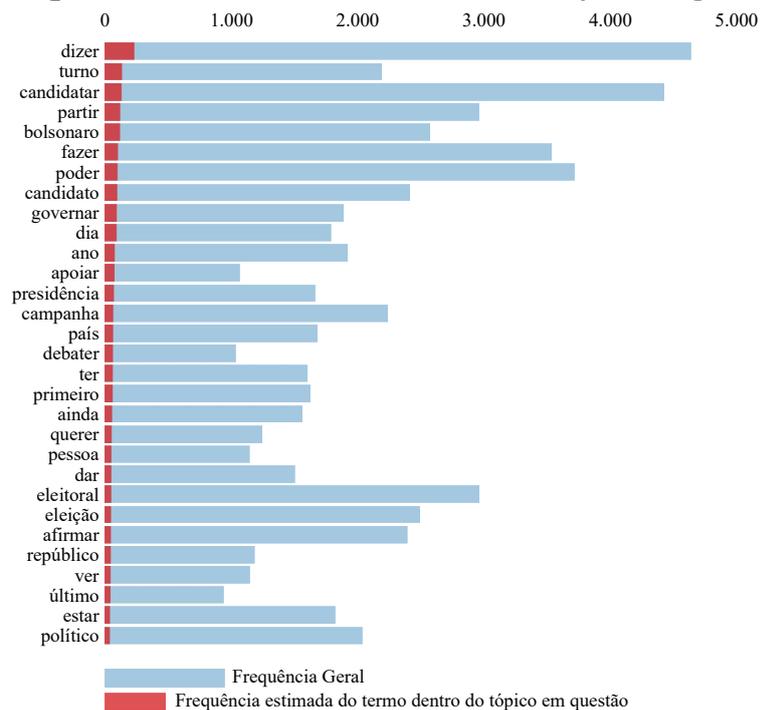


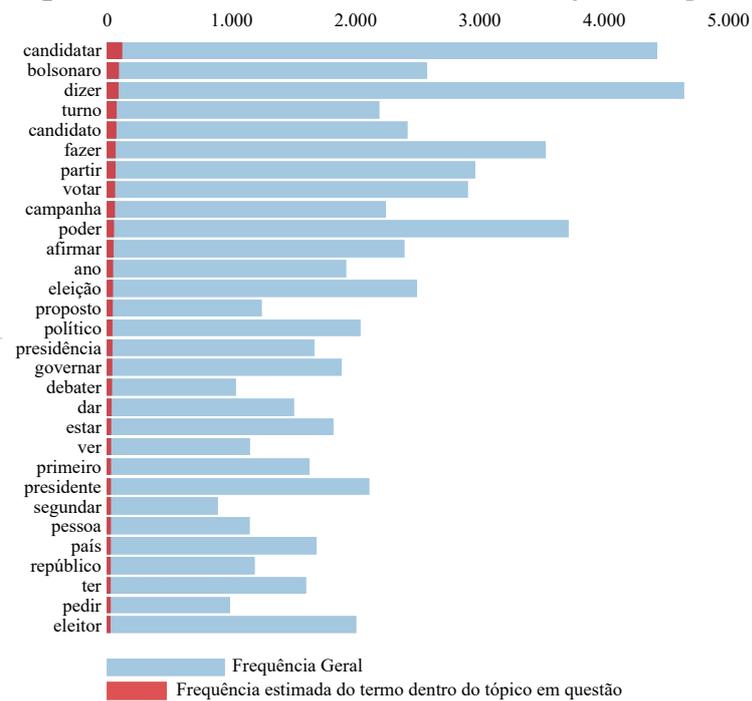
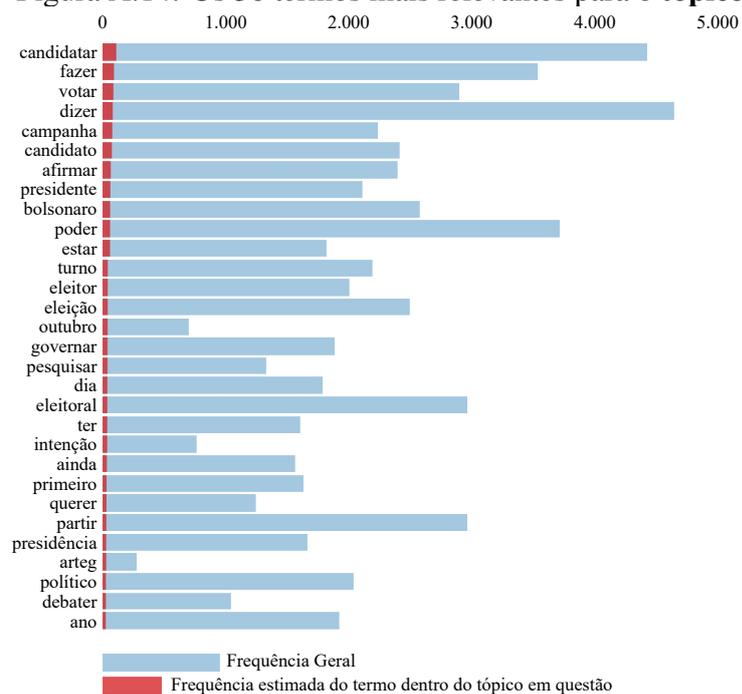
Figura A.13: Os 30 termos mais relevantes para o **tópico 12**Figura A.14: Os 30 termos mais relevantes para o **tópico 13**

Figura A.15: Os 30 termos mais relevantes para o **tópico 14**

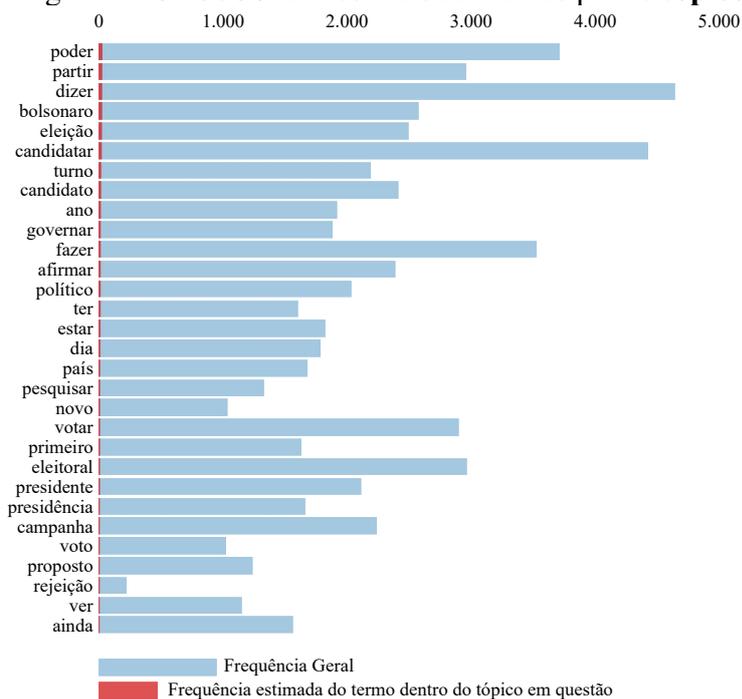
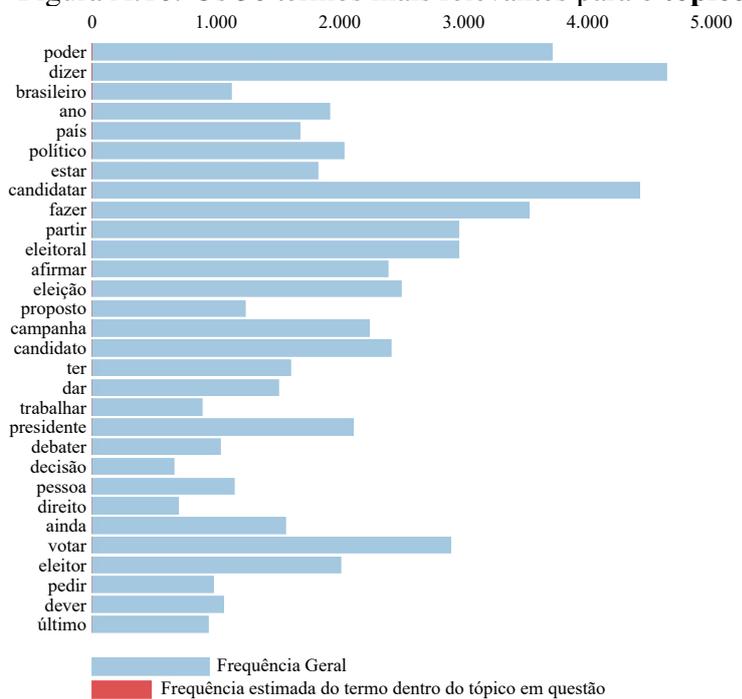


Figura A.16: Os 30 termos mais relevantes para o **tópico 15**



**APÊNDICE B — ENTIDADES NOMEADAS SELECIONADAS COMO PARTE
DAS CARACTERÍSTICAS PARA A DIVERSIFICAÇÃO**

Identificador	Entidade
E1	PT
E2	PSL
E3	Bolsonaro
E4	Brasil
E5	Jair Bolsonaro
E6	Fernando Haddad
E7	TSE
E8	Haddad
E9	PSDB
E10	Tribunal Superior Eleitoral
E11	São Paulo
E12	PDT
E13	Presidência da República
E14	Presidência
E15	Veja
E16	Geraldo Alckmin
E17	Ciro Gomes
E18	Rio de Janeiro
E19	MDB
E20	Estadão
E21	Lula
E22	Luiz Inácio Lula da Silva
E23	Estado
E24	Alckmin
E25	PSOL
E26	SP
E27	Marina Silva
E28	Brasília
E29	Ciro

E30	Rede
E31	Copyright
E32	STF
E33	PSB
E34	Ibope
E35	R\$
E36	Supremo Tribunal Federal
E37	Justiça Eleitoral
E38	Minas Gerais
E39	Facebook
E40	Polícia Federal
E41	República
E42	Michel Temer
E43	Vamos
E44	G1
E45	Henrique Meirelles
E46	DEM
E47	Confira
E48	Folha de São Paulo
E49	Justiça
E50	MG
E51	Marina
E52	Podemos
E53	Dilma Rousseff
E54	Curitiba
E55	TV Globo
E56	Juiz de Fora
E57	Câmara
E58	Novo
E59	PSTU
E60	Distrito Federal
E61	Guilherme Boulos
E62	Nordeste
E63	Senado

E64	Rosa Weber
E65	Temer
