CENTRO UNIVERSITÁRIO FEI

MILLER HORVATH

ESTUDO DA APLICAÇÃO DE SISTEMAS DE RECOMENDAÇÃO PARA APOIO À DECISÃO CLÍNICA NO TRATAMENTO DE PACIENTES PORTADORES DE CARDIOPATIAS CONGÊNITAS

São Bernardo do Campo

2020

MILLER HORVATH

ESTUDO DA APLICAÇÃO DE SISTEMAS DE RECOMENDAÇÃO PARA APOIO À DECISÃO CLÍNICA NO TRATAMENTO DE PACIENTES PORTADORES DE CARDIOPATIAS CONGÊNITAS

Dissertação de Mestrado apresentada ao Centro Universitário FEI para obtenção do título de Mestre em Engenharia Elétrica. Orientado pelo Prof. Dr. Paulo Sérgio Silva Rodrigues.

São Bernardo do Campo

2020

Horvath, Miller.
Estudo da Aplicação de Sistemas de Recomendação para Apoio à Decisão Clínica no Tratamento de Pacientes Portadores de Cardiopatias Congênitas / Miller Horvath. São Bernardo do Campo, 2020.
112 f. : il.

Dissertação - Centro Universitário FEI.
Orientador: Prof. Dr. Paulo Sérgio Silva Rodrigues.

1. Sistemas de Recomendação para Saúde. 2. Sistema de Apoio à Decisão Clinica. 3. Históricos Médicos Pessoais. 4. Recomendação de Tratamento. I. Silva Rodrigues, Paulo Sérgio, orient. II. Título.

Elaborada pelo sistema de geração automática de ficha catalográfica da FEI com os dados fornecidos pelo(a) autor(a).

centro universitário FEI

APRESENTAÇÃO DE DISSERTAÇÃO
ATA DA BANCA EXAMINADORA

Programa de Pós-Graduação Stricto Sensu em Engenharia Elétrica

Mestrado

PGE-10

Aluno: Miller Horvath **Matrícula:** 117202-2

Título do Trabalho: Estudo da aplicação de sistemas de recomendação para apoio à decisão clínica no tratamento de pacientes portadores de cardiopatias congênitas.

Área de Concentração: Processamento de Sinais e Imagens

Orientador: Prof. Dr. Paulo Sérgio Silva Rodrigues

Data da realização da defesa: 19/08/2020

ORIGINAL ASSINADA

Avaliação da Banca Examinadora:

_____

_____

_____

_____

_____

São Bernardo do Campo, / / .

MEMBROS DA BANCA EXAMINADORA	
Prof. Dr. Paulo Sérgio Silva Rodrigues	Ass.: _____
Prof. Dr. Flavio Tonidandel	Ass.: _____
Prof. Dr. Jones Oliveira de Albuquerque	Ass.: _____

A Banca Julgadora acima-assinada atribuiu ao aluno o seguinte resultado:

APROVADO ☒ REPROVADO ☐

VERSÃO FINAL DA DISSERTAÇÃO

APROVO A VERSÃO FINAL DA DISSERTAÇÃO EM QUE FORAM INCLUÍDAS AS RECOMENDAÇÕES DA BANCA EXAMINADORA

_____

Aprovação do Coordenador do Programa de Pós-graduação

_____

Prof. Dr. Carlos Eduardo Thomaz

RESUMO

Sistema de Recomendação (SR) são técnicas já consolidadas, desenvolvidas para auxiliar usuários na tarefa de encontrar conteúdo de interesse em meio a um grande conjunto de opções, que pode ser uma tarefa exaustiva quando feita manualmente. Estes sistemas têm sido extensivamente explorados pela comunidade acadêmica, mostrando serem eficazes para lidar com dados esparsos e prover recomendações relevantes, principalmente em cenários de comércio e entretenimento *online*. Trabalhos de pesquisa recentes passaram a investigar Sistemas de Recomendação em cenários relacionados à saúde, discutindo desafios e oportunidades para apoiar profissionais da saúde na tomada de decisão, ou até mesmo para gerar recomendações diretamente para pacientes. Este trabalho apresenta uma extensa revisão da área denominada Sistema de Recomendação para Saúde (SRS), com o intuito de pontuar as principais contribuições e desafios atuais. Alguns dos principais desafios pontuados no levantamento bibliográfico são: a escassez de trabalhos publicados que avaliam SRSs empiricamente, a adoção dos chamados Históricos Médicos Eletrônicos (HMEs) para a construção de SRSs e a modelagem do perfil dos pacientes em cenários de recomendações na área médica. Portanto, este trabalho apresenta uma avaliação e comparação das principais técnicas de SRs aplicadas em um cenário de recomendação na área médica, com o objetivo de recomendar procedimentos para apoiar profissionais da saúde no processo de definição de tratamento de pacientes que apresentam cardiopatias congênitas. Os experimentos foram realizados utilizando uma base de dados fornecida por um hospital de São Paulo, extraída de um sistema de Históricos Médicos Eletrônicos. Os resultados mostraram que os modelos de Filtragem Colaborativa (FC) superaram os modelos de Filtragem Baseada em Conteúdo (FBC) em todas as comparações realizadas. Além disso, os Sistemas de Recomendação Sensíveis ao Contexto (SRSCs) proporcionaram um aumento do desempenho dos SRSs, uma vez que os melhores resultados, tanto dos modelos de FC como dos modelos de FBC, foram obtidos com a utilização de informações contextuais na definição das recomendações. Por fim, os Sistemas de Recomendação Híbridos superaram todos os modelos individualmente, atingindo desempenho de 0,78 em relação ao $nDCG$ e de 0,87 em relação à Revocação.

Palavras-chave: Sistemas de Recomendação para Saúde. Sistema de Apoio à Decisão Clinica. Históricos Médicos Pessoais. Recomendação de Tratamento.

ABSTRACT

Recommender Systems are well-known techniques designed to support users in the overwhelming task of finding content of interest among a vast set of options. These systems have been extensively explored by the academia, demonstrating the Recommender Systems capabilities to efficiently handle sparse data and to provide valuable recommendations, especially in e-commerce and online entertainment scenarios. Recently, researchers started to investigate Recommender Systems into health-related scenarios, discussing challenges and opportunities of generating recommendations to support health professional decisions or even to recommend content directly for patients. This work presents an extensive review of the research field called Health Recommender Systems, aiming at showing the main contributions and challenges discussed in this field. Among these challenges, the lack of empirical experiments evaluating Health Recommender Systems, the adoption of the so-called Personal Health Records to build Recommender Systems, and the patient profile modeling are noteworthy. Therefore, this work assesses and compares the most relevant Recommender Systems techniques into a health recommendation scenario, focusing on producing procedures recommendations in order to assist health professionals to build the treatment plan of patients born with cardiac diseases. The experiments was performed using a dataset provided by a hospital in São Paulo, Brazil, built from a Personal Health Records system. The results show that Collaborative Filtering outperformed Content-Based Filtering in all evaluations. Moreover, Context-Aware Recommender Systems improved the models performance, since the best results for both Collaborative Filtering and Content-Based Filtering were achieved using context information to compute recommendations. Finally, the Hybrid Recommender Systems outperformed all other models, achieving a $nDCG$ of 0.78 and a Recall of 0.87.

Keywords: Health Recommender Systems, Clinical Decision Support Systems, Personal Health Records, Treatment Planning.

LISTA DE ILUSTRAÇÕES

Ilustração 1 – Ilustração que exemplifica o funcionamento de um modelo de Filtragem Colaborativa Baseada em Usuários em um cenário de recomendações de notícias. 40
Ilustração 2 – Ilustração que exemplifica o funcionamento de um modelo de Filtragem Colaborativa Baseada em Itens em um cenário de recomendações de notícias. 41
Ilustração 3 – Arquitetura em alto nível apresentando as etapas que compõem um Sistema de Recomendação por Filtragem Baseada em Conteúdo 43
Ilustração 4 – Fluxograma que ilustram os paradigmas de incorporação de informações contextuais em Sistemas de Recomendação. 46
Ilustração 5 – Exemplo do funcionamento do modelo kNN para diferentes valores de k. Sendo a observação objetivo, representada por $\star$, classificada como $\hexagon$ para $k = 3$ e como $\square$ para $k = 5$. 56
Ilustração 6 – Ilustra as etapas da metodologia desenvolvida. 64
Ilustração 7 – Diagrama que ilustra o cálculo das recomendações através da Filtragem Colaborativa, com e sem a adoção de dados contextuais, dentro da metodologia proposta. 72
Ilustração 8 – Diagrama que ilustra o cálculo das recomendações através da Filtragem Baseada em Conteúdo dentro da metodologia proposta. 74
Ilustração 9 – Avaliação, através do nDCG, do impacto da quantidade de pacientes similares (k) adotada pelos modelos de Filtragem Colaborativa (FC) em relação a cada uma das métricas de similaridade consideradas. 80
Ilustração 10 – Avaliação dos modelos de recomendação de procedimentos cirúrgicos através da métrica nDCG. 82
Ilustração 11 – Sobreposição das curvas suavizadas de distribuição suavizadas de similaridade entre o Perfil de Contexto Completo de cada paciente em relação aos demais através da Distância Euclidiana. 84
Ilustração 12 – Sobreposição das curvas suavizadas de distribuição de similaridade entre o Perfil de Contexto Completo de cada paciente em relação aos demais através da Distância dos Cossenos. 84
Ilustração 13 – Avaliação, através da métrica nDCG, do impacto dos perfis de contexto específicos na recomendação de procedimentos cirúrgicos através dos modelos de Filtragem Baseada em Conteúdo. 86
Ilustração 14 – Avaliação, através da métrica nDCG, do impacto dos perfis de contexto específicos na recomendação de procedimentos cirúrgicos através dos modelos de Filtragem Baseada em Conteúdo. 87
Ilustração 15 – Avaliação dos modelos de recomendação de procedimentos cirúrgicos através da Curva de Precisão-Revocação. 89

Ilustração 16 – Avaliação dos modelos de recomendação de procedimentos cirúrgicos através do *F-Score*. 91
Ilustração 17 – Avaliação dos modelos de recomendação de procedimentos cirúrgicos através da Curva ROC. 92
Ilustração 18 – Avaliação comparativa entre os sistemas híbridos que utilizam diferentes métricas para cálculo de similaridade. As avaliações são realizadas através do $nDCG$, F-*Score*, Curva de Precisão-Revocação e Curva ROC. 94
Ilustração 19 – Avaliação dos sistemas híbridos de recomendação de procedimentos cirúrgicos através do $nDCG$. 96
Ilustração 20 – Avaliação dos sistemas híbridos de recomendação de procedimentos cirúrgicos através do F-*Score*. 97
Ilustração 21 – Avaliação dos sistemas híbridos de recomendação de procedimentos cirúrgicos através da Curva de Precisão-Revocação. 98
Ilustração 22 – Avaliação dos sistemas híbridos de recomendação de procedimentos cirúrgicos através da Curva ROC. 99

LISTA DE TABELAS

Tabela 1 – Matriz de Confusão que relaciona os itens recomendados pelo Sistema de Recomendação com os itens "úteis" aos usuários, ou seja, os itens que os usuários de fato consumiram, se interessaram e/ou avaliaram positivamente. . 53
Tabela 2 – Exemplo da transformação *one-hot encoding* aplicado a uma variável categórica que representa o tipo sanguíneo, sendo (a) a variável original e (b) o conjunto de variáveis resultantes. 58
Tabela 3 – Aproximação com 5 casas decimais dos pesos de cada modelos de recomendação que compõem os modelos híbridos, separados pela métrica adotada para o cálculo de similaridade entre os perfis. 94
Tabela 4 – Tabela que apresenta os melhores resultados dos Sistemas de Recomendação Híbridos em relação ao $nDCG$ e à Revocação na primeira e na última posição da lista de recomendações. 103

LISTA DE ALGORITMOS

Algoritmo 1 – Descrição do cálculo de pesos para cada um dos modelos que compõem o Sistema de Recomendação Híbrido. A função *RMSE* representa a raíz do erro quadrado médio, descrito pela Equação (14), da predição de adequações $p_{cp}(u, i)$ em relação à adequação observada $a(u, i)$, para todos os pacientes u e procedimentos i contidos na base de dados. A constante λ é utilizada para limitar a porção do peso do modelo a ser transferida em cada iteração; neste trabalho, foi adotado o valor de $\lambda = 0.3$. 76

LISTA DE ABREVIATURAS

AUC	*Area Under Curve*.
DCG	*Discounted Cumulative Gain*.
FBC	Filtragem Baseada em Conteúdo.
FC	Filtragem Colaborativa.
HME	Histórico Médico Eletrônico.
HMP	Histórico Médico Pessoal.
kNN	*k-Nearest Neighbors*.
*n*DCG	*Normalized Discounted Cumulative Gain*.
PLN	Processamento de Linguagens Naturais.
ROC	*Receiver Operating Characteristic*.
SADC	Sistema de Apoio à Decisão Clinica.
SR	Sistema de Recomendação.
SRBC	Sistema de Recomendação Baseados em Conhecimento.
SRS	Sistema de Recomendação para Saúde.
SRSC	Sistema de Recomendação Sensível ao Contexto.
SVM	*Support Vector Machine*.
TFP	Taxa de Falsos Positivos.
UCI	*University of California, Irvine*.

SUMÁRIO

1 Introdução . . . 13
1.1 Objetivo . . . 14
1.2 Contribuições . . . 14
1.3 Estrutura do Trabalho . . . 15
2 Trabalhos Relacionados . . . 16
2.1 Sistemas de Recomendação para Saúde (SRS) . . . 16
2.1.1 Artigos de Revisão e Discussão em Sistemas de Recomendação para Saúde 16
2.1.2 Sistemas de Recomendação para Profissionais da Saúde . . . 23
2.1.3 Sistemas de Recomendação para Pacientes . . . 26
2.1.4 Sistemas de Recomendação Sensíveis à Saúde . . . 29
2.2 Sistemas de Apoio à Decisão Clinica (SADCs) . . . 30
2.2.1 Trabalhos de Revisão e Discussão em Sistemas de Apoio à Decisão Clinica 30
2.2.2 Aplicações em Sistemas de Apoio à Decisão Clinica . . . 33
3 Conceitos Fundamentais . . . 38
3.1 Sistemas de Recomendação . . . 38
3.1.1 Transação (Avaliação) . . . 39
3.1.2 Filtragem Colaborativa (FC) . . . 39
3.1.2.1 Filtragem Colaborativa Baseada em Usuários . . . 40
3.1.2.2 Filtragem Colaborativa Baseada em Itens . . . 41
3.1.3 Filtragem Baseada em Conteúdo (FBC) . . . 42
3.1.4 Sistemas de Recomendação Baseados em Conhecimento (SRBC) . . . 44
3.1.5 Sistemas de Recomendação Sensíveis ao Contexto (SRSC) . . . 45
3.1.6 Sistemas de Recomendação Híbridos (*Ensemble*) . . . 47
3.1.7 Métricas de Similaridade . . . 48
3.1.8 Avaliação . . . 50
3.1.8.1 Acurácia Preditiva . . . 52
3.1.8.1.1 Precisão ao Prever a Avaliação dos Usuários . . . 52
3.1.8.1.2 Precisão ao Prever a Adoção dos Usuários . . . 53
3.1.8.1.3 Precisão no Ranqueamento de Itens . . . 55
3.2 *k*-Nearest Neighbors (*k*NN) . . . 55
3.3 Pré-Processamento de Dados . . . 57
3.3.1 Validação Cruzada *k-Fold* . . . 57
3.3.2 Transformação de Dados . . . 57
3.3.2.1 Transformação One-Hot Encoding . . . 57
3.3.2.2 Normalização Min-Max . . . 58
3.3.2.3 Normalização Logarítmica . . . 58
3.4 Base de Dados . . . 59

4 Metodologia 63
4.1 Pré-Processamento de Dados 63
4.1.1 Limpeza de Dados 65
4.1.2 Modelagem dos Dados 65
4.1.3 Amostragem de Dados 67
4.2 Treinamento dos Modelos 68
4.2.1 Perfil de Contexto 69
4.2.2 Perfil de Paciente 70
4.2.3 Pré-Filtragem Contextual 70
4.2.4 Perfil de Procedimento Cirúrgico 70
4.2.5 Filtros Colaborativos 71
4.2.6 Filtros Baseados em Conteúdo 73
4.3 Hibridização 75
4.4 Avaliação dos Modelos 77
4.4.1 Descrição Experimental 78
5 Resultados 79
5.1 Quantidade De Pacientes Mais Similares 79
5.2 Desempenho Através do nDCG 81
5.2.1 Impacto das Métricas de Similaridade nos Modelos de Filtragem Baseada em Conteúdo Sensíveis ao Contexto 83
5.2.2 Impacto dos Perfis de Contexto 85
5.3 Desempenho Através da Curva de Precisão-Revocação 88
5.4 Desempenho Através do *F-Score* 90
5.5 Desempenho Através da Curva ROC 93
5.6 Desempenho dos Sistemas Híbridos 93
6 Conclusão 101
REFERÊNCIAS 104

1 INTRODUÇÃO

Nas últimas décadas, a área de Sistemas de Recomendação (SRs) tem sido extensivamente explorada pela comunidade acadêmica. Atualmente, pode-se considerar que esta é uma área de pesquisa consolidada, com diversos trabalhos publicados que comprovam a eficácia de SRs, principalmente quando aplicados em serviços de mídia e entretenimento e em comércio de produtos *online* (RICCI; ROKACH; SHAPIRA, 2015). Outro indício da consolidação de SRs é a adoção e o desenvolvimento destes sistemas por grandes empresas que oferecem serviços globais, como Amazon®, YouTube®, Netflix®, dentre muitos outros (RICCI; ROKACH; SHAPIRA, 2015; SCHAFER; KONSTAN; RIEDL, 1999).

Recentemente, estudos que visam explorar Sistemas de Recomendação para abordar problemas na área médica têm se intensificado, dando origem aos chamados Sistemas de Recomendação para Saúde (SRSs) (WIESNER; PFEIFER, 2014). O foco desta área de pesquisa é dividido na literatura em três vertentes principais, que são:

a) Sistemas de Recomendação para Pacientes — Focados em gerar recomendações no domínio da saúde diretamente aos pacientes. Alguns exemplos de conteúdos recomendados são: diagnósticos, medicamentos e tratamentos, literatura médica de fácil entendimento, médicos e hospitais.
b) Sistemas de Recomendação para Profissionais da Saúde — Voltados na produção de recomendações, também no domínio da saúde, para profissionais da área médica, com o intuito de auxiliá-los em desafios como: diagnosticar e tratar pacientes, gerenciar recursos, entre outros.
c) Sistemas de Recomendação Sensíveis à Saúde — Apesar de terem a finalidade de recomendar conteúdos não diretamente relacionados com a saúde, modelos desta natureza buscam incluir informações do âmbito da saúde no processo de recomendação. Por exemplo, considerar condições clínicas como hipertensão, diabetes, alergias ou até mesmo gestação, para gerar recomendações de comida.

Em relação aos SRSs para Profissionais da Saúde, o levantamento bibliográfico, apresentado no Capítulo 2, aponta a recomendação de diagnósticos e de tratamento como os principais cenários de discussões da área. Dentre os desafios ressaltados na literatura, pode-se destacar:

a) A falta de publicações focadas em análises experimentais da aplicação de modelos de SRs aplicados à saúde (HORS-FRAILE et al., 2017).
b) A adoção de tecnologias de Históricos Médicos Eletrônicos (HMEs), que se baseiam na aplicação de conceitos de *Big Data* para padronização e integração de dados históricos de pacientes (WIESNER; PFEIFER, 2014; HORS-FRAILE et al., 2017).
c) Escassez de bases de dados médicas públicas disponíveis para pesquisa (VALDEZ et al., 2016)
d) A construção do perfil dos pacientes utilizado para gerar as recomendações (SCHÄFER et al., 2017; HORS-FRAILE et al., 2017)

Portanto, o trabalho aqui proposto tem como objetivo apresentar uma avaliação e comparação empírica de algumas das principais técnicas de Sistemas de Recomendação tradicionais aplicadas à saúde. As técnicas de SR adotadas neste estudo são: Filtragem Colaborativa (FC), Filtragem Baseada em Conteúdo (FBC) e Sistemas de Recomendação Sensíveis ao Contexto (SRSCs).

Os estudos foram desenvolvidos em um cenário de recomendação de procedimentos de tratamento para pacientes que apresentam cardiopatias congênitas, servindo como um Sistemas de Apoio à Decisão Clinica para auxiliar médicos na construção dos planos de tratamento de seus pacientes. Estes experimentos foram conduzidos utilizando uma base de dados médica real, composta por informações sobre o processo de tratamento de pacientes, fornecida por um hospital de São Paulo.

Os resultados apontaram que os modelos de FC obtiveram resultados melhores em relação aos modelos de FBC em todas as comparações realizadas. No melhor caso, os modelos de FBC atingiram 0,6976 de $nDCG$, enquanto os modelos de FC obtiveram 0,7480. Além disso, os melhores resultados de ambos os modelos foram alcançados em conjunto com a integração de dados contextuais na construção das recomendações, através de SRSCs.

Por fim, a construção de Sistemas de Recomendação Híbridos através da combinação das recomendações dos demais modelos resultou em modelos de melhor desempenho, recomendando 87,54% dos procedimentos que foram de fato adotados no tratamento dos pacientes da base e atingindo 0,7834 de $nDCG$.

1.1 OBJETIVO

Desenvolver uma análise experimental de modelos de Sistemas de Recomendação tradicionais aplicados em um cenário de recomendação para saúde.

1.2 CONTRIBUIÇÕES

As principais contribuições apresentadas neste trabalho são:

a) Uma avaliação experimental da aplicação de alguns dos principais modelos de Sistemas de Recomendação em um cenário de recomendação de procedimentos de tratamento para pacientes que sofrem de cardiopatias congênitas.
b) Uma revisão da área de Sistemas de Recomendação para Saúde (SRSs) e áreas correlatas, levantando as principais contribuições e desafios relacionados ao tema, assim como os estudos publicados recentemente pela comunidade acadêmica.
c) Um conjunto de possibilidades de novos estudos que podem ser desenvolvidos a partir deste trabalho.

1.3 ESTRUTURA DO TRABALHO

O restante deste trabalho será apresentado da seguinte maneira:

a) Capítulo 2 — são descritos os trabalhos relacionados disponíveis na literatura, com o objetivo de apresentar uma visão completa do estado-da-arte em Sistemas de Recomendação para Saúde atualmente.
b) Capítulo 3 — são descritos os conceitos fundamentais necessários para a compreensão dos experimentos e discussões aqui propostos.
c) Capítulo 4 — foi definida a metodologia utilizada para a análise empírica dos modelos de Sistemas de Recomendação aplicados no cenário proposto.
d) Capítulo 5 — São apresentados os resultados e discussões obtidos através da aplicação da metodologia proposta e da revisão da área apresentada.
e) Capítulo 6 — São pontuadas as conclusões finais deste trabalho

2 TRABALHOS RELACIONADOS

2.1 SISTEMAS DE RECOMENDAÇÃO PARA SAÚDE (SRS)

De uma maneira geral, Sistemas de Recomendação para Saúde é um campo de pesquisa que explora a aplicação de técnicas de Sistemas de Recomendação tradicionais para abordar problemas inerentes à área da saúde. Com o objetivo de retratar o estado da arte em SRSs, a apresentação dos trabalhos na área foi dividido em quatro categorias.

Primeiramente, apesar de ser uma área de pesquisa relativamente nova, alguns trabalhos de revisão em SRSs já foram publicados. Além do mais, alguns trabalhos focam em discutir e propor aspectos para o desenvolvimento de SRSs, principalmente em comparação com SRs tradicionais. Os pontos principais destes trabalhos de revisão e discussão são apresentados na Seção 2.1.1.

Em seguida, as Seções 2.1.2 e 2.1.3 abordam a revisão bibliográfica de trabalhos enquadrados na classificação proposta em (WIESNER; PFEIFER, 2014), onde SRSs são divididos em: Sistemas de Recomendação para Profissionais da Saúde; e Sistemas de Recomendação para Pacientes.

Por fim, na Seção 2.1.4, são abordados trabalhos que, apesar de não serem aplicados diretamente na área médica, levam em consideração informações relacionadas à saúde para gerar recomendações em outros domínios de aplicação. Esta categoria de SRSs é referenciada neste trabalho como "Sistemas de Recomendação Sensíveis à Saúde".

2.1.1 Artigos de Revisão e Discussão em Sistemas de Recomendação para Saúde

Em (SODSEE; KOMKHAO, 2013), foi apresentada uma revisão em SRS e em Telemedicina, com foco na aplicação destas técnicas em áreas remotas da Tailândia. Os autores apresentam problemas na qualidade dos serviços médicos prestados nessas regiões da Tailândia, causados por diversos fatores, como falta de infraestrutura e quantidade insuficiente de médicos qualificados, além de sugerir a aplicação de Telemedicina e SRS para mitigar estes problemas (SODSEE; KOMKHAO, 2013). Apesar de restringir-se a regiões remotas na Tailândia, o trabalho proposto em (SODSEE; KOMKHAO, 2013) mostra-se relevante pela possibilidade de sua replicação para atender outras áreas que apresentem condições semelhantes. Sendo assim, foi proposta a técnica chamada "Sistema de Recomendação Baseado em Evidencias" (SODSEE; KOMKHAO, 2013), que é definida basicamente pela aplicação de algoritmos, como Rede Neurais, técnicas de Mineração de Dados, entre outras, em conjunto com técnicas de filtragem híbridas, que combinam Filtros Colaborativos, Filtros Baseados em Conteúdo e Filtros Baseados em Conhecimento, para realizar recomendações a partir de três fontes de dados, que são: histórico médico de pacientes, estudos experimentais e opiniões e relatórios de especialistas.

Por fim, em (SODSEE; KOMKHAO, 2013), um estudo de caso para um trabalho futuro foi proposto, no qual os autores sugerem utilizar a técnica proposta para auxiliar médicos não-especialistas que atuam em áreas remotas da Tailândia no diagnóstico de desproporção céfalo-pélvica em mulheres grávidas, com o objetivo de determinar se cesarianas devem ou não ser realizadas.

Segundo Sezgin (SEZGIN; OZKAN, 2013), a área de SRSs era considerada nova, pois poucos trabalhos relevantes haviam sido publicados até então. Estes trabalhos foram divididos em duas categorias, que são: SRSs para auxílio aos médicos na definição de diagnósticos e SRSs para realizar sugestões médicas personalizadas para usuários. A revisão bibliográfica em (SEZGIN; OZKAN, 2013) foi realizada metodologicamente, onde um total de 310 artigos foram selecionados em bases de trabalhos acadêmicos. Em seguida, uma análise de adequação ao tema foi realizada com base no título e no resumo dos trabalhos, onde o número de trabalhos foi reduzido para 35. Por fim, 8 destes trabalhos foram adotados por suprirem os critérios de qualidade definidos pelos autores, que são: a confiabilidade da fonte, a integridade do conteúdo e a apresentação de estudos aplicáveis.

Os autores em (SEZGIN; OZKAN, 2013) sugerem os seguintes desafios na área:

a) SRSs podem ser alvos de ataques cibernéticos devido a sensibilidade dos dados que são gerenciados nestes sistemas, sendo uma oportunidade de desenvolvimento em pesquisas na área de segurança da informação e privacidade de dados;
b) SRs tradicionais são baseados na popularidade de conteúdos, onde os conteúdos mais frequentemente consumidos ou bem avaliados pelos usuários tendem a ter maior chance de serem recomendados com maior relevância em comparação com conteúdos menos populares. Em contrapartida, essa característica pode gerar resultados ilusoriamente positivos em SRSs. Por exemplo, em casos onde o procedimento de tratamento mais adequado pra uma determinada condição clínica é pouco adotado, SRs tradicionais tendem a dar mais relevância para procedimentos mais populares ao gerar as recomendações. Esta constatação sugere a necessidade de uma adequação dos algoritmos de SRs para abordarem problemas na área médica;
c) Ausência de integração de dados entre aplicações *online* voltadas para saúde; e
d) Problemas em relação a questões éticas em relação ao uso das informações dos pacientes, como etnia, gênero, orientação sexual, entre outras.

Além disso, em (SEZGIN; OZKAN, 2013), as seguintes oportunidades são destacadas:

a) O desenvolvimento de SRSs não requer grande expertise em medicina, já que estes sistemas operam autonomamente através de abordagens colaborativas para detecção de padrões latentes de dados;
b) A integração de dados históricos dos pacientes com SRSs pode melhorar o poder preditivo destes sistemas e também ajudar a lidar com novos pacientes.
c) Soluções colaborativas podem permitir a obtenção de padrões e o desenvolvimento conhecimentos ainda não constatados na área médica, servindo para fins educacionais e de pesquisa na saúde;

d) Estruturas autônomas podem melhorar recomendações em relação à consistência e melhorar o conhecimento obtido sobre os usuários.

Dentre as técnicas mais utilizadas em SRSs, os autores em (SEZGIN; OZKAN, 2013) destacam: filtragem colaborativa e por conteúdo, SRs híbridos, semântica computacional e arquitetura orientada a software.

Em (WIESNER; PFEIFER, 2014), os autores apresentam uma classificação de SRSs baseada no cenário no qual são aplicados, podendo ter profissionais da saúde ou pacientes como usuários finais dos SRSs, sendo que o foco principal está direcionado ao cenário onde as recomendações são entregues aos pacientes. O trabalho em (WIESNER; PFEIFER, 2014) também aborda a exploração dos chamados Históricos Médicos Pessoais (HMPs), do inglês *Personal Health Records*, também referenciado na literatura como Históricos Médicos Eletrônicos (HMEs), ou *Electronic Health Records*, para construir SRSs mais eficientes; principalmente quando aplicados aos SRSs que visam melhor informar os pacientes sobre suas condições médicas, recomendando conteúdos médicos confiáveis, corretos e de fácil entendimento para leigos em medicina. Esse conceito é chamado na literatura de "empoderamento do paciente" (*patient empowerment*), que permite que pacientes exerçam um papel mais ativo no seu processo de tratamento e recuperação (WIESNER; PFEIFER, 2014). Ademais, são apresentados alguns desafios e requerimentos necessários para a integração de HMPs em SRSs, que são os seguintes:

a) Tratar erros de digitação e termos coloquiais ou imprecisos que podem constar nos HMPs;
b) Lidar com vocabulário técnico utilizado por médicos e outros profissionais de saúde;
c) Detectar quando condições clínicas aparecem negadas nos HMPs;
d) Preservar a confidenciabilidade das informações dos usuários contidas nos HMPs;
e) Possibilitar ajustes automáticos nos dados devido a mudanças de localização do sistema, como troca do idioma do usuário.

Por fim, uma proposta de implementação inicial de SRSs é apresentada em (WIESNER; PFEIFER, 2014), fazendo uso do chamado *Health Graph G*, uma estrutura de dados baseada em grafo construída a partir de informações médicas extraídas do *Wikipédia*[1].

O trabalho de (SCHÄFER et al., 2017), publicado em 2017, também tem como foco central o empoderamento de pacientes. Primeiramente, os autores apresentam um resumo de SRs tradicionais. Em seguida, é realizado um levantamento dos principais tópicos em SRSs abordados por trabalhos científicos até então, que são:

a) Construção de perfil de usuários;
b) Personalização;
c) Persuasão, empoderamento e confiança;
d) Avaliação médica;
e) Intervenção no estilo de vida;

[1] https://www.wikipedia.org/

f) Satisfação dos pacientes.

Por fim, os desafios em SRSs são apresentados, divididos em 3 perspectivas: do paciente; da recomendação; e da avaliação (SCHÄFER et al., 2017). Na perspectiva do paciente, os principais desafios discutidos são:

a) Construção do perfil dos pacientes;
b) identificação das necessidades dos pacientes.

Na perspectiva da recomendação, os desafios apresentados em (SCHÄFER et al., 2017) são:

a) Personalização e modelos de mudança comportamental;
b) Explicações simples e concisas;
c) Privacidade;
d) Aprovação do uso de medicamentos e medicina complementar.

Finalmente, os desafios apresentados em (SCHÄFER et al., 2017) na perspectiva de avaliação são:

a) Satisfação do usuário;
b) Avaliação em aplicações reais;
c) Consequências das recomendações.

Ainda com o foco no empoderamento do paciente, uma revisão de trabalhos que exploram as chamadas *redes sociais centradas em pacientes* é apresentada em (BISSOYI; MISHRA; PATRA, 2016). Tais serviços oferecem a oportunidade de pacientes compartilharem suas experiências médicas com outros usuários e receberem ajuda, sugestões e recomendações de outros usuários parecidos (BISSOYI; MISHRA; PATRA, 2016). Também é apresentada uma discussão sobre os principais tópicos em SRSs aplicados neste contexto, tais como:

a) Construção de perfil do usuário;
b) Similaridade entre dados de pacientes;
c) Similaridade entre pacientes a partir de suas condições médicas e tratamento;
d) Correspondência de postagens.

Em (EKSTRAND; EKSTRAND, 2016), são discutidos conceitos de como SRSs, desenvolvidos para ajudar usuários, podem ser prejudiciais aos mesmos, baseando-se no princípio bioético da não-maleficência. Esta discussão foi desenvolvida considerando um estudo de caso em nutrição, onde os autores apresentam situações onde SRSs podem ser aplicados visando recomendar alimentos para incentivar uma alimentação mais saudável, levantando discussões de contextos nos quais estes sistemas podem produzir recomendações inapropriadas (EKSTRAND; EKSTRAND, 2016). As principais discussões levantadas em (EKSTRAND; EKSTRAND, 2016) são:

a) Recomendações podem ser frustrantes para os usuários quando são difíceis de serem executadas. Por exemplo, recomendar receitas caras pode frustrar usuários com condições socio-econômicas menos favorecidas; receitas complicadas de serem executadas podem frustrar usuários que não possuam habilidades para realizá-las; ou recomendar receitas que contenham ingredientes difíceis de serem obtidos;

b) Condições clínicas específicas não consideradas podem ocasionar recomendações inapropriadas para pacientes sob essas condições. Por exemplo, cada tipo de diabetes tem uma variação em seu tratamento e cuidados. Todo paciente que possuí diabetes deve controlar a quantidade de carboidratos que são consumidos. Obesidade é uma condição comumente relacionada a diabetes, onde pacientes também buscam perder peso. Em contrapartida, mulheres que apresentam diabetes gestacional temporária devem controlar o nível de açúcar no sangue enquanto consomem os nutrientes apropriados para a gestação, ao invés de buscarem perda de peso;

c) Questões relacionadas ao projeto e *design* de sistemas podem causar estresse na busca por dietas apropriadas. Por exemplo, sistemas que não permitem que usuários indiquem alergia a alimentos específicos.

d) SRSs podem encorajar comportamentos não-saudáveis. Por exemplo, ao realizar recomendação de alimentos para pessoas com disfunção alimentar, podendo incentivar comportamentos destrutivos.

Em (VALDEZ et al., 2016), os conceitos tradicionais de Sistemas de Recomendação, como algoritmos, bases de dados e métricas, são abordados mais amplamente. Devido ao foco em Interação Humano-Computador, os autores também apresentam um resumo de interfaces e interação em SRs. Além disso, é proposto um ciclo iterativo de desenvolvimento de SRSs, que é dividido em 3 etapas: (1) compreensão do domínio de aplicação do SRS; (2) métodos para avaliação do SRS; e (3) métodos de implementação do SRS. Cada uma dessas etapas é definidas por 5 tópicos que devem ser analisados. Na primeira etapa, os tópicos levantados são:

a) Compreensão dos itens a serem recomendados — Exemplos de categorias são brevemente apresentados, como nutrição e alimentos, exercícios físicos e esportes, diagnósticos, terapia e medicação. Sendo que uma série de trabalhos na área de nutrição/alimentação foram citados, sugerindo uma tendência em pesquisas com este foco;

b) Quem são os usuários do sistema? — Além dos usuários considerados óbvios pelos autores, que são: pacientes, médicos especialistas e enfermeiros; também são apontados usuários alternativos, como farmacêuticos, pesquisadores e administradores, que poderiam, por exemplo, usufruir de SRSs com o objetivo de reduzir custos de serviços de assistência médica.

c) Contexto de uso — O contexto de uso é apresentado no âmbito de analisar os objetivos do sistema e como os fatores contextuais podem interferir nas recomendações.

d) Disponibilidade de dados — A falta de disponibilidade de bases de dados públicas para SRSs é apontada como uma das dificuldades da área, que dificulta a avaliação experimental de técnicas, além das questões relacionadas à ética e privacidade inerentes deste tipo de base de dados.

e) Outros interessados — O trabalho sugere que deve-se analisar outros interessados no sistema, além do usuário final, que possam se beneficiar de alguma funcionali-

dade do sistema. Por exemplo, um familiar que seja responsável por acompanhar a recuperação de um determinado paciente.

Na segunda etapa, os tópicos levantados em (VALDEZ et al., 2016) são:

a) Critérios relevantes para avaliar recomendações — Analisar métodos de avaliar quantitativamente as recomendações, como precisão, revocação, acurácia, cobertura, entre outros.
b) Avaliações centradas no usuário — Analisar SRSs em relação à satisfação do usuário, considerando aspectos como confiança e privacidade. Os autores destacam como o SRS lida com incertezas, assim como elas são comunicadas ao usuário, sugerindo como um fator essencial para SRSs que busquem efeitos a longo-prazo. Um exemplo adotado é a comparação de SRs tradicionais com SRSs. Por exemplo, em um SR de filmes, a recomendação de um filme ruim ao usuário custará em torno de 90 minutos do seu dia; em contrapartida, em um SRS, a recomendação de um tratamento inadequado ao paciente pode causar consequências com danos muito mais graves.
c) Avaliação comportamental — Avaliação comportamental refere-se a métodos para verificar se usuários estão efetuando as recomendações propostas. Por exemplo, no contexto de recomendação de atividades físicas, um SRS móvel poderia usar os sensores do *smartphone* para tentar supervisionar a realização das recomendações. Por outro lado, seria mais difícil desenvolver um método de supervisão ativo para um SRS que visa ajudar usuários a pararem de fumar, dependendo apenas do *feedback* do usuário.
d) Avaliação dos impactos na saúde — Averiguar os impactos do uso do SRS, por exemplo, através de exames clínicos ou análise de parâmetros vitais.
e) Implicações éticas — Os autores também fazem referência ao princípio bioético da não-maleficência citado anteriormente, dando importância aos riscos que um SRS pode oferecer a determinados usuários.

Na terceira etapa, os tópicos levantados em (VALDEZ et al., 2016) são:

a) Personalização — São levantadas considerações sobre a importância da modelagem de SRSs focada no usuário final, dando destaques ao uso da modelagem participativa, onde o sistema fornece artifícios para que os usuários possam contribuir ativamente na modelagem do sistema para melhor adequá-lo às suas necessidades específicas.
b) Privacidade diferencial — Privacidade diferencial refere-se à manipulação de dados sem expor a identidade dos indivíduos. Os autores destacam a importância de comunicar os riscos de privacidade aos usuários e levantam considerações de como usuários consideram a privacidade diferentemente conforme suas necessidades.
c) Metodologia de estudo de modelagem — É discutida a importância de se adotar visualizações efetivas e eficientes ao exibir informações em um SRS.

d) Riscos e incertezas compreensíveis — É destacada a importância de comunicação de riscos e incertezas relacionados ao SRS de forma bidirecional. Por exemplo, usuários devem ser capazes de comunicar ao SRS quando alguma informação de entrada é imprecisa ou desconhecida, assim como o sistema deve apresentar os riscos e incertezas adequadamente.

e) Aproveitamento das capacidades computacionais e humanas — Os autores destacam o paradigma chamado *Doctor-in-the-loop*, que coloca o médico não apenas como um usuário final do SRS, mas também como alguém que pode manipular diretamente ferramentas e algoritmos interativamente, aproveitando ao máximo as capacidades do médico e do SRS no contexto de uso.

Em (HORS-FRAILE et al., 2017), uma revisão metodológica em SRSs é apresentada. Esta revisão tem como objetivo analisar os trabalhos seguindo uma taxonomia multidisciplinar, proposta por eles. Os principais desafios da área apontados em (HORS-FRAILE et al., 2017) são:

a) Em relação ao domínio — A maioria dos trabalhos focam apenas em pacientes adultos e são desenvolvidos para promoção de saúde genérica e nutrição. Além do mais, foram conduzidas análises em um número restrito de países.

b) Aspectos metodológicos e procedurais — Há uma falta de exposição de resultados e análises de custo-benefício, casos de teste limitados e considerando poucos pacientes, além de alguns trabalhos não consideram HMEs.

c) Fatores teóricos em promoção de saúde e mudança de comportamento — Não foi possível avaliar a efetividade do uso destas teorias devido à completa ausência de informações em relação à como as mensagens recomendadas foram construídas.

d) Aspectos técnicos — Estudos não reportam informações completas sobre os sistemas, há sistemas rotulados incorretamente como SRs e a maioria dos sistemas tem limitações em relação à geração de perfil do usuário, adaptação desses perfis às mudanças e obtenção de *feedback* dos pacientes.

Em (BATEJA; DUBEY; BHATT, 2018), são discutidos os benefícios que podem ser obtidos através da aplicação de *frameworks*, mais precisamente os que oferecem *MapReduce* (DEAN; GHEMAWAT, 2008), para o desenvolvimento de SRSs. *MapReduce* é um modelo de programação que provê artifícios para facilitar o processamento de grandes massas de dados, dividindo esses dados em sub-grupos menores, permitindo processamento em paralelo com tolerância a falhas e flexibilidade (DEAN; GHEMAWAT, 2008). Um exemplo de proposta de *framework* para SRSs pode ser observado em (MACEDO et al., 2016), onde o *Health Surveillance Software Framework (HSSF)* é proposto. O HSSF é um *framework* que visa dar suporte ao desenvolvimento de sistemas em biomedicina ao oferecer artifícios computacionais, como Processamento de Linguagem Naturais para processamento de artigos científicos e históricos médicos, Ontologias, modelos classificadores, entre outros.

O levantamento bibliográfico desenvolvido neste trabalho apresenta ao menos 8 trabalhos de revisão em SRSs, publicados entre 2013 e 2018. Este volume de trabalhos de revisão em

SRSs desenvolvidos nos últimos anos sugere um grande interesse da comunidade acadêmica na área, explorando diversas vertentes da mesma, tais como: SRSs para pacientes, SRSs para profissionais da saúde, telemedicina, interação humano-computador, modelagem de software em SRSs, *frameworks* para desenvolvimento de SRSs, ética, privacidade, redes sociais centradas nos pacientes e possíveis riscos que SRSs podem apresentar aos pacientes.

Devido a quantidade de trabalhos com este foco, pode-se observar que SRSs para pacientes é a principal abordagem da área tendo como foco principal o chamado empoderamento do paciente. Em contrapartida, tendo em vista que este trabalho foca em SRSs para profissionais da saúde, as principais oportunidades e desafios constatados neste contexto pelos trabalhos de revisão e discussão apresentados nesta seção, são:

a) O uso dos chamados HMPs em SRSs é um recurso que oferece um grande potencial para melhoria da qualidade dos SRS.
b) A construção do perfil dos pacientes é um dos pontos principais para o desenvolvimento de SRSs adequados.
c) A importância de se analisar os possíveis riscos que o SRS pode oferecer aos pacientes devido a recomendações inapropriadas.
d) Poucos trabalhos apresentam resultados experimentais.
e) Há uma escassez de base de dados públicas disponíveis na área da saúde, dificultando a realização de experimentos e a comparação de resultados com outros trabalhos.

2.1.2 Sistemas de Recomendação para Profissionais da Saúde

SRSs para profissionais da saúde tem como objetivo auxiliar estes profissionais na tomada de decisão e realização de suas tarefas. O principal foco de SRSs para profissionais da saúde é a recomendação automática de diagnósticos médicos, sendo que Filtragem Colaborativa, frequentemente combinada com outros métodos computacionais, é uma técnica popular em estudos com esta finalidade.

Em (HASSAN; SYED, 2010), Hassan e Syed ajustam técnicas de Filtros Colaborativos para modelar estratificação de risco em pacientes cardíacos. O modelo proposto supera as técnicas de Regressão Logística e Máquinas de Vetores de Suporte, ou SVM, ao prever morte por parada cardíaca súbita e reincidência de infarto do miocárdio com maior precisão, a partir da análise da área sob a curva ROC (HASSAN; SYED, 2010).

Métodos de clusterização são amplamente adotados em conjunto com Filtros Colaborativos.

Em (DAVIS et al., 2008), é proposto o sistema *CARE*, onde Filtros Colaborativos e Clusterização são combinados para analisar o histórico médico de pacientes e prever o risco dos mesmos apresentarem reincidência de doenças ou outras doenças relacionadas no futuro, possibilitando que intervenções médicas sejam menos reativas, como usualmente são, e passem a ser mais preventivas. Além do mais, uma versão interativa do *CARE* também é apresentada,

chamada *ICARE* (DAVIS et al., 2008). Dado que Filtros Colaborativos tendem a ser predominantemente afetados por itens populares, o *ICARE* visa mitigar o domínio da recomendação de doenças mais frequentes na base de dados, clusterizando as amostras de treinamento pelas doenças e aplicando Filtros Colaborativos individualmente em cada cluster (DAVIS et al., 2008). A metodologia proposta foi avaliada através das métricas de cobertura, *average rank* e *half-life accuracy* (DAVIS et al., 2008).

Similarmente, Filtros Colaborativos e clusterização também são combinados em (HUSSEIN et al., 2014). Este trabalho propõe um SRS para lidar com pacientes que sofrem de doenças crônicas, utilizando Filtros Colaborativos e diversas técnicas, clássicas e modernas, de clusterização e classificação, com o objetivo de avaliar o impacto de cada uma destas técnicas no SRS (HUSSEIN et al., 2014). As avaliações dos métodos adotados foram realizadas considerando a precisão, erro médio absoluto e outros método estatísticos (P, R, *F-measure*) (HUSSEIN et al., 2014). Além da recomendação de diagnósticos, esse sistema recomenda sugestões médicas para melhorar as condições clínicas, melhorar a qualidade de vida e ajudar nas adaptações ao estilo de vida de pacientes com diabetes (HUSSEIN et al., 2014). Consequentemente, diminuindo o custo das instituições de saúde para tratar estes pacientes (HUSSEIN et al., 2014).

Filtros Colaborativos também são frequentemente combinados com métodos *Fuzzy*. Em alguns casos, ambos são também combinados com métodos de clusterização.

Primeiramente, Thong et al. (THONG et al., 2015b) combinam Conjuntos *Fuzzy* Intuicionistas com Filtros Colaborativos para analisar os registros médicos dos pacientes e prever diagnósticos. Essa combinação proporcionou uma diminuição do erro médio absoluto ao prever diagnósticos quando comparada à aplicação individual das técnicas utilizadas (THONG et al., 2015b). Em seu trabalho posterior (THONG et al., 2015a), foram integradas técnicas de Clusterização *Fuzzy* para diminuir ainda mais o erro na recomendação de diagnósticos. Em ambos os trabalhos (THONG et al., 2015b,a), foi utilizada uma base de dados da *UCI Machine Learning Repository* (DHEERU; KARRA TANISKIDOU, 2017), chamada *HEARTH*.

Conjuntos *Fuzzy* e Filtros Colaborativos também são combinados em (KOMKHAO; LU; ZHANG, 2012), com o objetivo de prever casos em que o parto por cesárea deve ser realizado devido à desproporção céfalo-pélvica. Posteriormente, foi publicada uma extensão deste trabalho (KOMKHAO; HALANG, 2013), onde Filtros Colaborativos, Filtros Baseados em Conhecimento e Aprendizado por Reforço são utilizados de modo a aumentar a acurácia das previsões realizadas no contexto abordado, em comparação com Filtros Colaborativos tradicionais.

Ali et al. (ALI; THANH; VAN MINH et al., 2017) utiliza uma generalização dos Conjuntos *Fuzzy* Intuicionistas, chamada de Conjuntos Neutrosóficos, para criar um Sistema de Recomendação Neutrosófico. Em (THANH; ALI et al., 2017), Thanh, Ali e outros colaboradores propõem um algoritmo de clusterização para aprimorar esse sistema. Apesar do alto custo computacional, o método proposto foi comparado com outros métodos do estado-da-arte até então, superando-os em relação à precisão na recomendação de diagnósticos (THANH; ALI et al., 2017).

O sistema Ebinformatics (OLUWAGBEMI; OLUWAGBEMI; ABIMBOLA, 2016) utiliza Lógica *Fuzzy* para auxiliar no diagnóstico de doença por vírus Ebola no oeste africano, assim como recomendar sugestões de prevenção, tratamento e contensão de disseminação da doença. A avaliação do sistema foi realizada através de questionários aplicados para analisar o conhecimento dos usuários sobre fatores de transmissão do vírus Ebola, assim como para avaliar a usabilidade do sistema proposto.

Pode-se observar a adesão de técnicas mais modernas em SRs para recomendação de diagnósticos. Os autores de (MONTEIRO et al., 2015) propõem a utilização de Sistemas de Recomendação Baseados em Contexto, em conjunto com técnicas de Processamento de Linguagens Naturais (PLN) e Recuperação de Imagens por Conteúdo (*Content-Based Image Retrieval*), para recomendar textos e artigos científicos que auxiliem médicos a realizar diagnósticos de exames de imagem.

A recomendação de tratamentos pode ser apontada como outra demanda de SRSs para profissionais da área médica.

Em (DUAN; STREET; XU, 2011), técnicas estatísticas clássicas são combinadas com Filtros Colaborativos para gerar planos de cuidados de enfermagem baseados nos diagnósticos e intervenções realizadas, além do histórico de desfecho dos pacientes. Foi proposta também uma nova métrica para avaliação de SRSs, chamada *average ranking position* (DUAN; STREET; XU, 2011).

Em (GRÄSSER et al., 2017), é proposta uma metodologia que também faz uso dos benefícios oriundos de Filtros Colaborativos, principalmente em relação à personalização e recomendação baseada nos dados, em conjunto com Sistemas de Apoio à Decisão Clinica (SADCs), que oferecem decisões confiáveis com base em regras. Essa metodologia é aplicada para recomendar terapias para pacientes que sofrem de psoríase, que é uma doença auto-imune que afeta a pele (GRÄSSER et al., 2017). A avaliação experimental foi realizada com base na precisão e raiz do erro médio quadrado (GRÄSSER et al., 2017).

O algoritmo Leap (ZHANG et al., 2017) combina SRs Baseados em Conhecimento e em Conteúdo, juntamente com Aprendizado por Reforço e técnicas aplicadas em Problemas de Tomada de Decisão Sequenciais, para recomendar medicação para pacientes que apresentam multimorbidade (mais de uma patologia crônica); reduzindo o risco de interações medicamentosas causarem reações adversas. A avaliação do sistema foi realizada considerando as seguintes métricas: *Mean Jaccard Coefficient*, precisão das predições, métricas relacionadas à interação medicamentosa e avaliação por profissionais médicos (ZHANG et al., 2017).

Sun e os demais autores em (SUN et al., 2016) apresentam duas contribuições para serem aplicadas à recomendação de tratamentos médicos, sendo uma nova métrica de similaridade entre tratamentos e um novo algoritmo de clusterização. Seus experimentos mostram que é possível aumentar a taxa de cura de pacientes que sofreram infarto cerebral de 19,89% para 21,28%, e a taxa de efetividade dos tratamentos de 97,25% para 98,29% (SUN et al., 2016).

Em (YANG et al., 2017), é proposta uma metodologia de clusterização de procedimentos de modo a extrair protótipos de procedimentos, dado um conjunto de informações contextuais. Apesar da proposta não ter sido desenvolvida especificamente para o contexto da saúde, podendo ser replicada em outros domínios, a metodologia foi avaliada com bases de dados médicas, mais especificamente em intubação endotraqueal, ressuscitação traumatológica e emergência; podendo ser utilizada para apoio à tomada de decisão (YANG et al., 2017). Foram utilizadas as métricas de similaridade média, *F1-score* e *G-mean* (YANG et al., 2017).

Em (NACHAWATI et al., 2012), é proposta uma metodologia para recomendar tratamento, dicas de prevenção, terapias para doenças, e até mesmo médicos e hospitais, baseado no conteúdo compartilhado pelos usuários em redes sociais.

A recomendação de diagnósticos aparece como o principal foco de SRSs para profissionais da saúde. Para este fim, Filtros Colaborativos e métodos de clusterização são as técnicas mais amplamente adotadas; técnicas *Fuzzy* também são frequentemente adotadas neste contexto. Entretanto, outras técnicas clássicas de SRs, como Filtros Baseados em Conteúdo (LOPS; DE GEMMIS; SEMERARO, 2011) e Filtros Baseados em Conhecimento (FELFERNIG et al., 2011), ainda não foram amplamente exploradas na recomendação de diagnósticos.

Outro foco de SRSs para profissionais da saúde, sendo também o foco deste trabalho, é a recomendação de tratamentos. Os trabalhos apresentados neste contexto mostram divergência, tanto nas métricas de avaliação adotadas, como nos domínios clínicos em que são aplicados. Esta divergência sugere que SRSs para recomendação de tratamento é um campo de pesquisa que ainda demanda grande investigação. Além disso, nenhum dos trabalhos apresentados é aplicado em pacientes que apresentam cardiopatias congênitas, que é o contexto clínico abordado neste trabalho.

2.1.3 Sistemas de Recomendação para Pacientes

SRSs para pacientes são sistemas que geram recomendação diretamente aos pacientes com o foco em auxiliá-los em objetivos clínicos específicos, como perder peso ou encontrar o médico mais adequado para sua condição clínica. A área mais explorada em recomendação voltada aos pacientes visa recomendar conteúdos informativos apropriados, corretos e de fácil entendimento para que os usuários possam ter um maior conhecimento sobre suas condições médicas. Esse aumento de conhecimento dos pacientes é conhecido como *Empoderamento do Paciente*, que tem como consequência aumentar a capacidade dos pacientes em participar mais ativamente do seu processo de recuperação, e até mesmo aumentar o senso crítico dos mesmos em relação as suas alternativas de tratamento, medicação, e assim por diante. Para este fim, muitos trabalhos fazem uso de técnicas de PLN (do inglês *Natural Language Processing*) como base de SRs com o intuito de extrair informações de documentos de texto.

Em (CHO et al., 2014), é proposto um agente automático que recomenda links para postagens em fóruns que discutem problemas médicos similares aos reportados pelo usuário através

de Filtros Baseados em Conteúdo e PLN. Técnicas do estado-da-arte em Recuperação de Informação foram comparadas utilizando as métricas MAP (*Mean Average Precision*) e *Precision@k*.

Lage et al. (LAGE et al., 2011) propõem utilizar técnicas de Análise Semântica, oriundas de sistemas de Bio-Supervisão tradicionais, para recomendar sinais que inferem doenças com base em sintomas, assim como documentos que possam ser pertinentes às doenças inferidas. Nesse trabalho, a base de treinamento foi construída a partir de postagens do Twitter[2].

Em (KONDYLAKIS et al., 2015), Kondylakis e seus colegas propõem utilizar Análise Semântica de Textos em um SRS que ajude usuários a encontrar documentos na internet que possam muni-los de conhecimento relevante sobre suas condições médicas.

Similarmente, em (KONDYLAKIS et al., 2014), é proposta a utilização de um modelo estatístico de PLN, chamado *Latent Dirichlet Allocation*, para filtrar documentos relacionados com as condições médicas e preferências dos usuários.

Além disso, outros trabalhos buscam combinar técnicas de PLN com modelos de Representação de Conhecimento na área médica, mais especificamente Ontologias. Ontologias são utilizadas para representar formalmente os elementos de um domínio específico, assim como a relação entre estes elementos.

Wiesner e Pfeifer (WIESNER; PFEIFER, 2010) extraem termos médicos dos registros médicos dos pacientes e de páginas de medicina do Wikipedia[3] para construir uma Rede Semântica que represente os conceitos extraídos como um modelo de Representação de Conhecimento. Esta rede foi usada em seu trabalho posterior (WIESNER; PFEIFER, 2014), juntamente com técnicas de Recuperação de Informação, expansão de consultas (*queries*) e um conceito de negação de termos, para apresentar uma arquitetura básica de um SRS para recomendar conteúdo médico com base em uma *query* de entrada. Utilizando uma base de dados fornecida pelo *Heidelberg University Hospital* , a arquitetura proposta foi avaliada estatisticamente, utilizando precisão e o teste t de Student, com base em um padrão ouro obtido através de um levantamento com especialistas.

No trabalho de Bayyapu e Dolog (BAYYAPU; DOLOG, 2010), rótulos são explorados através de técnicas de similaridade tradicionais e semânticas para ajudar usuários a encontrar textos médicos adequados.

Em (BOCANEGRA et al., 2017), conceitos provenientes da Web Semântica são aplicados para relacionar páginas educacionais de saúde com *tags* extraídas de vídeos relacionados à saúde do YouTube[4].

Em (MARTÍNEZ-ROMERO et al., 2017), um SR de Ontologias biomédicas é proposto, onde a recomendação é realizada com base num texto de entrada, levando em consideração alguns critérios, como a extensão do texto no qual a ontologia abrange, a aceitação da ontologia pela comunidade biomédica, o nível de detalhamento da ontologia e a especialização da ontologia no domínio dos dados de entrada.

[2]https://twitter.com/
[3]https://www.wikipedia.org/
[4]https://www.youtube.com/

Outros trabalhos focam em recomendar médicos e hospitais adequados para determinadas condições do usuário.

Em (JIANG; XU, 2014), utiliza-se técnicas de PLN para relacionar as condições médicas (sintomas, doença, histórico, etc) e preferencias do usuário com o perfil dos médicos, que também são avaliados pelos usuários; recomendando, por fim, os médicos mais adequados para cada usuário.

Alvarez et al. (ÁLVAREZ; SALAZAR; OVALLE, 2013) propõem um SR baseado em Ontologias que permite que usuários encontrem médicos adequados, não só às suas condições médicas, mas também considerando outras questões, como custo, localização, disponibilidade de agendamento, etc.

Em (MBANDU et al., 2015), é proposto um sistema onde usuários possam verificar a qualidade dos servidores de saúde *a priori*. Podendo determinar o local mais adequado para realizar o procedimento médico necessário com base nas medidas baseadas em reputação propostas pelos autores, chamadas de "pontuação de reputação" (*reputation score*) (MBANDU et al., 2015).

Por fim, outros trabalhos utilizam SRs aplicados a problemas específicos.

Hors-Fraile et al. (HORS-FRAILE et al., 2016) propõem um SRs híbrido de mensagens, motivacionais e/ou informativas, de para auxiliar fumantes em seu processo de cessar o consumo de cigarros. Sua proposta determina quais mensagens são mais adequadas para cada usuário, assim como qual o intervalo ideal entre o envio dessas mensagens (HORS-FRAILE et al., 2016). O SR híbrido combina Filtragem Colaborativa, Filtragem Baseada em Conteúdo e algoritmos baseados em funções de utilidade.

Em (HAMED et al., 2012), Árvores de Decisão são treinadas com base em mensagens do Twitter[5], manualmente selecionadas e classificadas por profissionais médicos a partir dos sintomas relatados nas mensagens. Utilizando a Árvore de Decisão para classificar novas mensagens que relatem sintomas em seus determinados grupos e aplicar SRs para recomendar medicamentos que tratem estes sintomas (HAMED et al., 2012).

Em (LI; JONES, 2015), é proposta uma combinação de técnicas em Recuperação de Informação com Filtros Baseados em Conteúdo com o objetivo de auxiliar pacientes e profissionais da saúde na realização de buscas por informações médicas.

Assim como foi apontado pelos trabalhos de revisão e discussão na Seção 2.1.1, o principal foco dos trabalhos em SRSs para pacientes está relacionado com o empoderamento do paciente. As técnicas de PLN são as mais utilizadas neste contexto, dada a necessidade de se analisar textos médicos diversos, assim como os HMPs. Curiosamente, estas técnicas de PLN não aparecem com frequência em SRSs para profissionais da saúde. Dado que SRSs para profissionais da saúde também podem tirar proveito do processamento deste textos, a aplicação de PLN pode ser um possível caminho de pesquisa. Ontologias também são frequentemente apli-

[5]https://twitter.com/

cadas em SRSs para paciente e pouco em SRSs para profissionais, podendo também ser um caminho de pesquisa dada a sua capacidade de representação de domínios.

2.1.4 Sistemas de Recomendação Sensíveis à Saúde

Apesar de não terem foco direto na área médica, alguns trabalhos buscam explorar questões relacionadas a saúde para aprimorar seus sistemas de recomendação.

López-Nores e os demais autores (LÓPEZ-NORES et al., 2012) propõem um SR que analisa as condições médicas dos usuários para determinar os produtos que serão recomendados. Desta forma, evitando que produtos que possam ser prejudiciais ao usuário, devido às suas condições específicas, sejam recomendados e refinando a recomendação com produtos mais adequados (LÓPEZ-NORES et al., 2012). Seus experimentos apontam um aumento na precisão das recomendações para usuários que apresentam problemas médicos específicos (LÓPEZ-NORES et al., 2012).

Em (AGU; CLAYPOOL, 2016), é proposto um Sistema de Recomendação de "*exergames*", jogos que exigem algum tipo de atividade física como ferramenta de interação com o jogo, para dispositivos móveis. Com o intuito de manter os usuários praticando atividades físicas constantemente, o sistema utiliza os sensores do celular e outros fatores de interação do usuário com os jogos para inferir os tipos de *exergames* que o usuário gosta (AGU; CLAYPOOL, 2016). Assim, quando o usuário perde o interesse por um determinado *exergame*, o sistema é capaz de recomendar novos jogos que se enquadrem nas preferencias do usuário (AGU; CLAYPOOL, 2016).

Alguns trabalhos recentes tem focado na recomendação de alimentos.

Em (GE et al., 2015), os autores propõem um SR de alimentos personalizado como um aplicativo para dispositivos móveis, apresentando todo o estudo de Interfaces do mesmo. Em (GE; RICCI; MASSIMO, 2015), o sistema proposto em (GE et al., 2015) é expandido, de modo aos usuários poderem determinar se desejam recomendações de alimentos baseados no sabor ou na saúde. Permitindo assim que os usuários possam adequar seus hábitos alimentares com recomendações saudáveis e equilibradas com os seus paladares (GE; RICCI; MASSIMO, 2015).

Em (ACHANANUPARP; WEBER, 2016), é proposto um sistema que identifica e recomenda alimentos substitutos, ou seja, alimentos que possam ser trocados um pelo outro em uma determinada refeição. Sua hipotese baseia-se num conceito de PLN de que, assim como palavras utilizadas em contextos semelhantes tendem a ter significados semelhantes, alimentos consumidos em contextos semelhantes tendem a ser possíveis substitutos entre si (ACHANANUPARP; WEBER, 2016).

2.2 SISTEMAS DE APOIO À DECISÃO CLINICA (SADCS)

Outro campo de pesquisa que visa utilizar técnicas computacionais para abordar problemas na área médica é conhecido como Sistemas de Apoio à Decisão Clinica (SADCs).

Até onde foi possível constatar, os trabalhos de Sistemas de Recomendação para Saúde raramente fazem referência aos trabalhos de SADCs, apesar de ser um campo de pesquisa mais antigo do que o campo de SRSs, levando em consideração as datas de publicação dos principais trabalhos de ambas as áreas. Ademais, como será discutido nas seções seguintes, trabalhos de SADCs abordam pontos em comum com SRSs, tais como personalização, construção de perfil dos pacientes, filtragem e recomendação. Estas constatações sugerem que o intercâmbio de conhecimento entre estes campos de pesquisa deveria ser considerado para trabalhos futuros, podendo beneficiar-se do progresso, até então individual, de ambas as áreas.

Sendo assim, a apresentação da revisão bibliográfica em SADCs foi dividida em duas categorias de trabalhos, que são: trabalhos de revisão e discussão em SADCs (Seção 2.2.1), que foca em mostrar um pouco da evolução da área como um todo, assim como listar os seus principais desafios e oportunidades; e aplicações em SADCs (Seção 2.2.2), onde são levantadas as principais vertentes de trabalhos, assim como as principais técnicas computacionais adotadas por elas e os seus resultados.

2.2.1 Trabalhos de Revisão e Discussão em Sistemas de Apoio à Decisão Clinica

Muitos trabalhos de revisão buscaram investigar os benefícios de SADCs através do estudo da aplicação destes sistemas em testes clínicos.

O estudo desenvolvido em (HUNT et al., 1998) aponta que 68% dos trabalhos considerados apresentam uma melhoria no desempenho dos profissionais da saúde quando apoiados por um SADC. Segundo (HUNT et al., 1998), a quantidade de publicações na área e a qualidade destes trabalhos estava aumentando rapidamente. Foi possível observar benefícios para prescrição de medicamentos, cuidados preventivos, entre outros (HUNT et al., 1998). Por outro lado, SADCs não ofereciam benefícios significativos para diagnósticos até então (HUNT et al., 1998).

Em (GARG et al., 2005), foi conduzida uma revisão metodológica de trabalhos que avaliam os efeitos de SADCs aplicados em testes clínicos controlados. Em (GARG et al., 2005), os SADCs foram divididos em quatro categorias, que são: sistemas de diagnóstico, sistemas de lembrete para prevenção, sistemas de tratamento e sistemas de dosagem e prescrição de medicamentos. Segundo (GARG et al., 2005), a quantidade e qualidade metodológica dos estudos melhoraram com o tempo. Além disso, 64% dos SADCs avaliados melhoram o desempenho dos profissionais médicos (GARG et al., 2005). Entretanto, o efeito dos SADCs no desfecho dos pacientes ainda era indeterminado devido a falta e inconsistência dos trabalhos realizados até então (GARG et al., 2005).

Similarmente, o trabalho em (KAWAMOTO et al., 2005) também conduz uma revisão para identificar características de SADCs que melhorem a prática clínica através da análise estatística da aplicação dos mesmos em testes clínicos. Foi constatado que 68% dos trabalhos apontaram uma melhoria significativa na prática clínica devido a utilização de SADCs (KAWAMOTO et al., 2005). Ao todo, 15 características foram analisadas, divididas em 4 categorias, que são: gerais do sistema, interação clínico-sistema, conteúdo de comunicação e características auxiliares (KAWAMOTO et al., 2005). Através de regressões logísticas, 4 características foram identificadas como preditores independentes de melhoria de prática clínica (KAWAMOTO et al., 2005), que são:

a) Provisão automática de apoio à decisão como parte do fluxo de trabalho do profissional clínico — Por exemplo, recomendações de cuidados de diabetes impressas e anexadas no prontuário dos pacientes, para que o médico não tenha que buscar a recomendação no SADC.
b) Prover uma recomendação, não apenas uma analise — Por exemplo, o sistema recomendar a prescrição de antidepressivos a um paciente ao invés de apenas identificar que o mesmo apresenta depressão.
c) Fornecer apoio a decisão no momento e local da tomada de decisão — Por exemplo, ao invés de gerar relatórios mensais listando os pacientes que necessitam de cuidados preventivos, imprimir a recomendação de tais cuidados na ficha de atendimento destes pacientes.
d) Uso de computação para gerar apoio a tomada de decisão — Por exemplo, ao invés de realizar auditoria manuais de prontuários para identificar pacientes em atraso para rastreamento de câncer no colo do útero, identificar estes pacientes através da análise de banco de dados.

Dentre os sistemas analisados que integram todas as 4 características, 98% deles apresentaram melhoria na prática clínica (KAWAMOTO et al., 2005).

Em (KUPERMAN et al., 2007), foi realizada uma revisão em SADCs para medicação. Estes SADCs são divididos em duas categorias, que são: SADCs básicos e avançados (KUPERMAN et al., 2007). Os SADCs básicos lidam com os seguintes desafios: checagem de alergia medicamentosa, orientação básica de dosagem, suporte a decisão de formulários, checagem de terapias duplicadas e checagem de interação medicamentosa (KUPERMAN et al., 2007). Em contrapartida, os SADCs avançados lidam com: suporte de dosagem para pacientes geriátricos e pacientes que sofrem de insuficiência renal, orientação para testes laboratoriais relacionados a medicação, checagem de medicação para gestantes e checagem de contraindicação medicamentosa em doenças (KUPERMAN et al., 2007).

O trabalho em (SITTIG et al., 2008) fez um levantamento que apontou os 10 principais desafios em Sistemas de Apoio à Decisão Clinica (SADCs) até então. Estes desafios foram divididos em 3 categorias (SITTIG et al., 2008), listados a seguir:

a) Melhorar a eficácia das intervenções dos SADCs:
   – Melhorar interface humano-computador.

- Resumir informações dos usuários.
- Ranquear e filtrar recomendações para o usuário.
- Combinar recomendações para pacientes que apresentam comorbidade.
- Utilizar informações de texto para conduzir SADCs.

b) Criar novos SADCs:
- Priorizar o desenvolvimento e implementação de SADCs.
- Aplicar mineração de dados em grandes bases de dados médicas para criar novos SADCs.

c) Disseminar o conhecimento e intervenção dos SADCs já conhecidos:
- Disseminar boas práticas para projetar, desenvolver e implementar SADCs
- Criar arquiteturas para compartilhar módulos e serviços em SADCs.
- Criar repositórios online para SADCs

Interessante como, já em 2008, estudos em SADCs destacavam desafios como o ranqueamento e filtragem de recomendações aos usuários, além da aplicação de mineração de dados em grandes bases de dados, que são características fundamentais da área de SRs. Além do mais, a utilização de informações textuais, através da aplicação de técnicas de PLN, também é uma abordagem popular em SRSs.

Em (JASPERS et al., 2011), foi realizada uma revisão para avaliar os efeitos dos SADCs no desempenho dos profissionais da saúde e no desfecho dos pacientes. Foi constatado que 57% dos trabalhos indicam que o SADC impactou positivamente o desempenho dos profissionais da saúde, e apenas 30% dos trabalhos indicam benefícios no desfecho dos pacientes devido ao uso de SADCs (JASPERS et al., 2011). Os autores também destacam os estudos de SADCs que oferecem prescrição de medicamentos (com exceção da prescrição de anticoagulantes) e lembretes de tratamento preventivo como sendo os que mais oferecem evidências dos benefícios dos SADCs (JASPERS et al., 2011). Entretanto, estudos em SADCs para diagnóstico apresentam resultados menos positivos (JASPERS et al., 2011).

Em (VAGHELA; BHATT; MISTRY, 2015), foi apresentada uma revisão curta que visa comparar diversas técnicas de classificação, tais como Rede Bayesiana, Rede Neural, Árvore de Decisão, *Naive Bayes*, SVM, Conjuntos *Fuzzy*, Algoritmo Genético, *Rough Set* e kNN; aplicadas para realizar diagnóstico em diferentes contextos clínicos, como câncer de mama, doenças de coração e diabetes.

Em (ZUCCOTTI et al., 2014), foi feito um estudo junto à uma companhia de seguros e sete hospitais em Boston, Estados Unidos, para identificar oportunidades de evitar negligências médicas com o uso de SADC. O estudo aponta que mais da metade dos casos de indenização devido a negligência médica que foram considerados, totalizando mais de 40 milhões de dólares em indenizações, poderiam ser evitados com o uso de SADCs (ZUCCOTTI et al., 2014).

Em (MARCO-RUIZ et al., 2016), é levantada uma discussão de como serviços da web semântica poderiam auxiliar na generalização, difusão, reutilização de funcionalidades e compartilhamento de Sistemas de Apoio à Decisão Clinica (SADCs) entre instituições de saúde.

Esta investigação foca no desenvolvimento de métodos para definição da semântica em SADCs, explorar serviços baseados nestes métodos e estudar formas de interagir com artifícios, como metadados, funcionalidades e modelos de dados, baseado-se em suas descrições semânticas (MARCO-RUIZ et al., 2016).

2.2.2 Aplicações em Sistemas de Apoio à Decisão Clinica

Grande parte das publicações em SADCs visam apoiar profissionais médicos na realização de diagnósticos. Neste contexto, Redes Neurais Artificiais é a técnica mais frequentemente adotada.

Os autores em (GUDADHE; WANKHADE; DONGRE, 2010) propõem dois SADC, um baseado em SVM e outro em Redes Neurais Artificiais, para diagnóstico de doenças no coração. A avaliação desta proposta foi realizada utilizando uma base de dados do *UCI Machine Learning Repository*, chamada *Heart Disease Data Set* (DHEERU; KARRA TANISKIDOU, 2017). Os SADC baseados em SVM e redes neurais atingiram, respectivamente, 80,41% e 97,5% no diagnóstico.

Em (SUGANTHI; MADHESWARAN, 2012), foi proposto um SADC para realizar o diagnóstico de câncer de mama analisando imagens obtidas através de mamografias. As imagens selecionadas passam por um processo de melhoria, segmentação e extração de características, onde um sub-conjunto de características ótimas é obtido pela aplicação de um algoritmo genético (SUGANTHI; MADHESWARAN, 2012). A etapa de classificação do tumor é realizada através de uma rede neural com retro-propagação otimizada por algoritmos de colonia de formigas e enxame de partículas (SUGANTHI; MADHESWARAN, 2012). O SADC proposto obteve 99,5% de acurácia na classificação dos tumores, aplicado à uma base de dados chamada *Digital Database for Screening Mammography (DDSM)*.

Um SADC baseado em *Deep Learning* foi proposto em (HU et al., 2016), para diagnosticar doença de Alzheimer a partir de imagens de ressonância magnética. Essas imagens são convertidas em uma matriz de série temporal, que representa 90 regiões do cérebro (HU et al., 2016). Em seguida, uma matriz de correlação entre os pares de regiões do cérebro é calculada e usada para classificação através de uma rede neural artificial, chamada *autoencoder* (HU et al., 2016). A avaliação experimental foi realizada utilizando uma base de dados do *Alzheimer's Disease Neuroimaging Initiative*, onde o método proposto foi comparado com regressão logística e SVM, comparando também o impacto da utilização da matriz de correlação proposta como entrada dos classificadores em relação a matriz de série temporal (HU et al., 2016). Os resultados apontaram que as técnicas de regressão logística, SVM e a rede neural *autoencoder* atingiram, respectivamente, uma acurácia de 59,7%, 62,9% e 67,5% utilizando a matriz de série temporal como entrada, e uma acurácia de 83,5%, 82,1% e 87,5% utilizando a matriz de correlação como entrada (HU et al., 2016).

Ademais, pode-se observar uma diversidade de técnicas adotadas por SADCs para diagnósticos, tais como: classificador *Nayve Bayes*, *Rough Sets* e Lógica *Fussy*.

Em (YAO; AZAM, 2015), foi proposto um SADC baseado em *Game-theoretic Rough Sets*. Essa técnica é baseada em uma aproximação de conjuntos, onde os seus limiares são determinados através de um processo de competição e cooperação baseado na teoria dos jogos (HERBERT; YAO, 2011). Esta proposta foi avaliada utilizando 4 bases de dados disponibilizadas pelo *UCI Machine Learning Repository* (DHEERU; KARRA TANISKIDOU, 2017), chamadas *Contraceptive Method Choice*, *Haberman's Survival*, *Thyroid Disease* e *Pima Indian Diabetes*, que abrangem os seguintes contextos clínicos: métodos contraceptivos, mortalidade pós cirurgia de câncer de mama, diagnósticos de tiroide e diagnóstico de diabetes, respectivamente. Para as duas primeiras bases, os resultados apontaram uma acurácia condizente com as apresentadas por outros trabalhos que utilizaram as mesmas bases; por outro lado, para as bases restantes, os resultados foram considerados promissores por superarem outros trabalhos em relação também a acurácia (YAO; AZAM, 2015).

Um SADC foi proposto em (LIU et al., 2016), utilizando o método de classificação *Naive Bayes* para realizar previsão de diagnósticos com o foco na preservação da privacidade das informações dos pacientes. Este SADC foi testado com uma base de dados real fornecida pelo *UCI Machine Learning Repository* (DHEERU; KARRA TANISKIDOU, 2017), chamada *Acute Inflammations Dataset*, que abrange 2 doenças do sistema urinário: inflamação na bexiga e pielonefrite. O SADC proposto em (LIU et al., 2016) não apresentou nenhum falso negativo e uma taxa de 26,23% e 14,29% de falso positivos para inflamação na bexiga e pielonefrite, respectivamente.

Um SADC também baseado em um classificador *Naive Bayes* é proposto em (SUBBALAKSHMI; RAMESH; RAO, 2011) com o objetivo de indicar se um paciente apresenta doença cardíaca ou não com base em uma série de características contidas na base de dados *Heart Disease Data Set*, do *UCI Machine Learning Repository* (DHEERU; KARRA TANISKIDOU, 2017).

O trabalho em (DOUALI et al., 2015) propõem explorar o modelo de *Fuzzy Cognitive Maps* juntamente com técnicas da Web Semântica para modelar diretrizes e informações médicas para gerar diagnóstico em UTI. O trabalho parte de uma árvore de decisão, criada manualmente e previamente em trabalhos anteriores, que passa por etapas de melhorias, extração de conceitos clínicos e diagnósticos, normalização através de ontologia e criação da regras até a construção do *Fuzzy Cognitive Maps* (DOUALI et al., 2015). A metodologia proposta foi construída utilizando diretrizes para infecção do trato urinário, sendo que foram considerados 70 pacientes anônimos de diversos hospitais europeus (DOUALI et al., 2015). Os resultados mostram que 92% dos diagnósticos gerados pelo *Fuzzy Cognitive Maps* estavam integralmente alinhados com as diretrizes consideradas *Fuzzy Cognitive Maps*. Além disso, 25 conceitos clínicos e 13 diagnósticos foram identificados para representar o problema de diagnóstico de infecção do trato urinário *Fuzzy Cognitive Maps*.

Em (CHEN; GOH; CHONG, 2013), é proposto um SADC para diagnósticos fundamentado em medicina baseada em evidências. Este sistema gera sistematicamente uma lista de evidências externas, como artigos em medicina publicados em jornais científicos, para cada tipo de doença possível, determinadas pelo resultado de hemograma (CHEN; GOH; CHONG, 2013). Foi realizada uma avaliação qualitativa do sistema, onde um grupo de médicos e estudantes de medicina avaliaram a usabilidade do SADC (CHEN; GOH; CHONG, 2013).

Outro foco de SADCs é sugerir e apoiar a construção de estratégias de tratamento médico.

Segundo (MAGO et al., 2012), a literatura aponta que há uma falta de consistência entre os dentistas na definição do plano de tratamento dentário. Sendo assim, foi proposto um SADC baseado em Lógica *Fuzzy*, que sugere os tratamentos possíveis a partir dos sinais e sintomas do paciente (MAGO et al., 2012). O teste de qui-quadrado foi utilizado para averiguar a uniformidade entre os tratamentos propostos pelo SADC e pelos dentistas (MAGO et al., 2012). Os autores em (MAGO et al., 2012) também destacam que os dentistas tendem a prescrever tratamentos excessivos em comparação com o SADC, sugerindo que o SADC pode servir como uma segunda opinião para dentistas, com o objetivo de evitar a execução de procedimentos desnecessários e excesso de medicação.

Em (O'CONNOR et al., 2011), é proposto avaliar o impacto de um SADC no controle da taxa de hemoglobina glicada, pressão sanguínea, colesterol LDL (*low-density lipoprotein*, ou lipoproteína de baixa densidade) em pacientes que apresentam diabetes tipo 2, através de um teste clínico aleatorizado. Detalhes sobre o desenvolvimento do SADC não são expostos em (O'CONNOR et al., 2011). Os resultados apontam melhorias significativas no controle de glicose e da pressão sistólica, uma melhoria pequena no controle da pressão diastólica e nenhum benefício no controle do colesterol LDL (O'CONNOR et al., 2011).

Em (BÖTTIGER et al., 2009), é apresentada uma base de dados, chamada SFINX, de interações entre pares de medicamentos projetada para SADC. A base SFINX possuí mais de oito mil interações entre pares de medicamentos, classificadas pela relevância clínica, ou seja, o impacto de uma determinada interação medicamentosa no tratamento, e o nível de documentação, que representa o embasamento teórico e prático utilizado para descrever uma determinada interação medicamentosa (BÖTTIGER et al., 2009). A relevância clínica das interações medicamentosas é classificada entre quatro categorias (BÖTTIGER et al., 2009), entre "A" e "D", que são as seguintes:

a) "A" para interações sem impacto clínico;
b) "B" para interações que possuem efeitos incertos e/ou variados;
c) "C" para interações clinicamente relevantes que podem ser contornadas, por exemplo, através de ajustes na dosagem; e
d) "D" para interações clinicamente relevantes que devem ser evitadas.

Além disso, o nível de documentação das interações é classificada em cinco níveis (BÖTTIGER et al., 2009), que vão de "0" a "4", que são as seguintes:

a) "0" para interações derivadas de estimativas baseadas em estudos com medicamentos semelhantes;

b) "1" para interações derivadas de estudos *in vitro* e/ou estudos de caso incompletos;
c) "2" para interações derivadas de estudos de caso devidamente documentados;
d) "3" para interações derivadas de estudos em voluntários saudáveis e/ou estudos iniciais em pacientes; e
e) "4" para interações derivadas de estudos controlados em uma populações relevantes de pacientes.

Em (HUSSAIN; ABIDI; ABIDI, 2007), é proposto um SADC que utiliza *Web* Semântica para inferir recomendações baseadas em diretrizes clinicas para auxiliar médicos no acompanhamento de pacientes pós-tratamento de câncer de mama. Este sistema integra múltiplas Ontologias que representam o domínio clínico, os pacientes e as diretrizes de práticas clinicas(HUSSAIN; ABIDI; ABIDI, 2007).

Alguns trabalhos em SADC propõem soluções capazes de apoiar tanto o diagnóstico como o tratamento médico, normalmente através da utilização de lógica derivada de bases de conhecimento.

Em (KARIM; BAJWA, 2011), é proposto um SADC baseado em telemedicina para operar em regiões remotas, onde não haja um especialista. Este SADC utiliza técnicas de PLN para processar um relatório das condições do paciente e detectar sinais e sintomas (KARIM; BAJWA, 2011). Em seguida, é utilizado para definir o diagnóstico e prescrever a medicação; caso o SADC não seja capaz de realizar o diagnóstico, um e-mail é enviado para um médico especialista e a sua resposta é usada para atualizar o SADC para futuras intervenções semelhantes (KARIM; BAJWA, 2011). Foi utilizado um estudo de caso para analisar o SADC proposto, onde o sistema apresentou 92,2% e 89,6% de acurácia na detecção dos sintomas e realização de diagnóstico, respectivamente (KARIM; BAJWA, 2011).

Em (ALHARBI; BERRI; EL-MASRI, 2015), é proposto um SADC para diagnóstico e tratamento de diabete em estágio inicial. Para este fim, foi construída uma ontologia com base em diretrizes de prática de medicina e conhecimento de especialistas, em conjunto com a construção de um base de conhecimento baseada na Web Semântica (ALHARBI; BERRI; EL-MASRI, 2015).

Uma técnica para identificar quais as características dos pacientes que são mais relevantes para prever corretamente o diagnóstico e para recomendar o tratamento mais adequado para cada paciente, chamada *discovery engine*, é apresentada em (YOON; DAVTYAN; SCHAAR, 2017). Essas características são determinadas a partir da análise dos HMEs e informações da literatura médica (YOON; DAVTYAN; SCHAAR, 2017). A técnica proposta foi comparada com 10 técnicas do estado-da-arte que são amplamente adotadas por trabalhos em SADCs (YOON; DAVTYAN; SCHAAR, 2017), que são: CFS, all contextual learning, regressão logística multivariada, regressão linear, SVM multivariada com *radial basis function* como kernel, SVM com seleção de características, *adaptative boosting*, árvore de classificação, regressão logística multivariada regularizada utilizando laço e regressão linear regularizada utilizando laço. A avaliação da proposta em (YOON; DAVTYAN; SCHAAR, 2017) foi realizada utilizando

uma base de dados disponível no *UCI Machine Learning Repository* (DHEERU; KARRA TANISKIDOU, 2017), chamada *Diagnostic Wisconsin Breast Cancer Database*. Os experimentos apontam melhoria de 16,6%, através do coeficiente *Cohen's kappa*, na recomendação de tratamentos personalizados de quimioterapia para pacientes com câncer de mama, e 2,18% e 4,20% de melhoria em relação à taxa de erro e à taxa de falso positivos, respectivamente, para previsão de diagnóstico de câncer de mama (YOON; DAVTYAN; SCHAAR, 2017). Ademais, o desempenho da técnica proposta se mostrou robusto em relação à ausência de informações (YOON; DAVTYAN; SCHAAR, 2017).

SADCs também são aplicados para realização de prognósticos ou determinar o desfecho de pacientes que são submetidos a determinados procedimentos.

Em (FERNANDES et al., 2010), é proposto um SADC para prognóstico de pacientes com câncer de mama com base na estratificação de risco de mortalidade dos mesmos. Este SADC utiliza três metodologias para prognósticos: *Nottingham prognostic index* (NPI), *Cox regression modelling* e *partial logistic artificial neural network with automatic relevance determination* (PLANN-ARD) (FERNANDES et al., 2010).

3 CONCEITOS FUNDAMENTAIS

3.1 SISTEMAS DE RECOMENDAÇÃO

A evolução do campo de pesquisa em Sistemas de Recomendação está fortemente relacionado com o crescimento da quantidade de conteúdo disponível na internet. Serviços de mídia, entretenimento e comércio digital oferecem uma quantidade abundante de opções aos usuários. Em meio a tantas opções, a busca manual pelo conteúdo de interesse se torna uma tarefa exaustiva, podendo causar a frustração, o desinteresse e até mesmo o abandono destes serviços pelos usuários. Dado este contexto, SRs é uma sub-área de aprendizado de máquina que tem como objetivo reduzir o volume de conteúdo apresentado ao usuário, realizando recomendações personalizadas de modo a discriminar o conteúdo no qual o usuário tende a ter maior interesse, com base nos padrões de consumo de conteúdo no passado.

O conceito de Sistemas de Recomendação é proveniente do trabalho em (GOLDBERG et al., 1992), referenciado na época como Filtragem Colaborativa. Posteriormente, em (RESNICK; VARIAN, 1997), foi proposta uma definição mais abrangente de Sistemas de Recomendação, como é conhecida hoje, devido à constatação de que outras estratégias, diferentes da Filtragem Colaborativa, poderiam ser adotadas para realizar recomendações personalizadas.

Em outubro de 2006, ocorreu um grande marco na área de SRs, o chamado "*Netflix Prize*" (BENNETT; LANNING et al., 2007), onde a empresa de *streaming* de video Netflix® disponibilizou uma base de dados anonimizada, contendo um conjunto de avaliações de filmes realizadas por usuários de sua plataforma, desafiando a comunidade acadêmica a desenvolver um sistema de recomendação que superasse a precisão do sistema que eles utilizavam na época. Este desafio impulsionou o crescimento de pesquisas na área de SRs.

Uma estratégia popular em Sistemas de Recomendação é tentar prever o nível de satisfação que os usuários teriam caso consumissem novos itens, ainda não consumidos pelos mesmos, para gerar as recomendações. Em um Sistema de Recomendação, os elementos que são recomendados pelo sistema são chamados de itens. Por exemplo, os itens no contexto do *Netflix Prize* são os filmes e séries contidos na base de dados disponibilizada.

Frequentemente, essa predição baseia-se na similaridade entre usuários e/ou itens, calculada através de métricas matemáticas. As principais métricas utilizadas para o cálculo de similaridade em SRs serão listadas e descritas na Seção 3.1.7.

O cálculo de similaridade pode ser focado nos padrões de como os itens são consumidos e/ou avaliados pelos usuários, ou focado nas características dos itens que o usuário consome. Outra estratégia de SRs foca na exploração de conhecimentos aprofundados sobre o domínio no qual a recomendação é realizada e nas necessidades específicas dos usuários, onde recomendações são construídas baseando-se em regras. Além disso, SRs podem ser focados em identificar o contexto no qual o usuário está interagindo com o sistema para identificar itens mais adequados para situações específicas.

Estas variações de estratégias compõem as classes de Sistemas de Recomendação tradicionais, já consolidadas na literatura. Baseando-se na taxonomia proposta em (BURKE, 2007), SRs podem ser divididos em quatro classes de técnicas principais: (1) Filtragem Colaborativa (*Collaborative Filtering*); (2) Filtragem Baseada em Conteúdo (*Content-Based Filtering*); (3) Filtragem Baseada em Conhecimento (*Knowledge-Based Filtering*); e (4) Filtragem Baseada em Contexto (*Context-Based Filtering*); que serão descritos respectivamente nas Seções 3.1.2, 3.1.3, 3.1.4 e 3.1.5.

3.1.1 Transação (Avaliação)

Transação é uma informação a respeito da interação do usuário com o Sistema de Recomendação, que é armazenada por ser relevante para o algoritmo responsável por gerar as recomendações (RICCI; ROKACH; SHAPIRA, 2015). Uma transação contem informações explícitas ou latentes relacionadas ao nível de satisfação do usuário que consumiu um determinado item, popularmente referenciadas na literatura como *ratings*, ou avaliações. Nas avaliações explícitas, é necessário que os usuários forneçam o seu parecer quanto a sua satisfação ao consumir um item, respeitando uma determinada escala de avaliação, que pode ser:

a) Avaliação numérica, tal como uma nota de 1 a 5 estrelas utilizada para avaliar uma compra realizada na Amazon.com [1];
b) Avaliação ordinal, como um questionário sobre um item que adota respostas em escala Likert (LIKERT, 1932);
c) Avaliação binária, onde usuários informam se gostaram ou não de um determinado item. Por exemplo, como o sistema de avaliação de vídeos no YouTube [2]; e
d) Avaliação unária, como por exemplo, quais músicas um determinado usuário escutou no Spotify [3].

Outro tipo de transação explicita popular em SRs são os rótulos, ou *tags*. Onde usuários podem atribuir rótulos que descrevam a sua experiência ao consumir um determinado item.

Um exemplo de transação com informações latentes pode ser quando um usuário seleciona um filme no catálogo para ler a sinopse, pois pode-se considerar que o mesmo se interessou por este filme de alguma forma.

3.1.2 Filtragem Colaborativa (FC)

Filtragem Colaborativa foi o primeiro modelo de SR a ser proposto, sendo publicado primeiramente em (GOLDBERG et al., 1992) e formalmente estruturado em (SCHAFER et al., 2007), sendo até hoje um dos modelos mais populares, devido a sua simplicidade e facilidade de implementação.

[1] https://www.amazon.com.br/
[2] https://www.youtube.com/
[3] https://www.spotify.com/

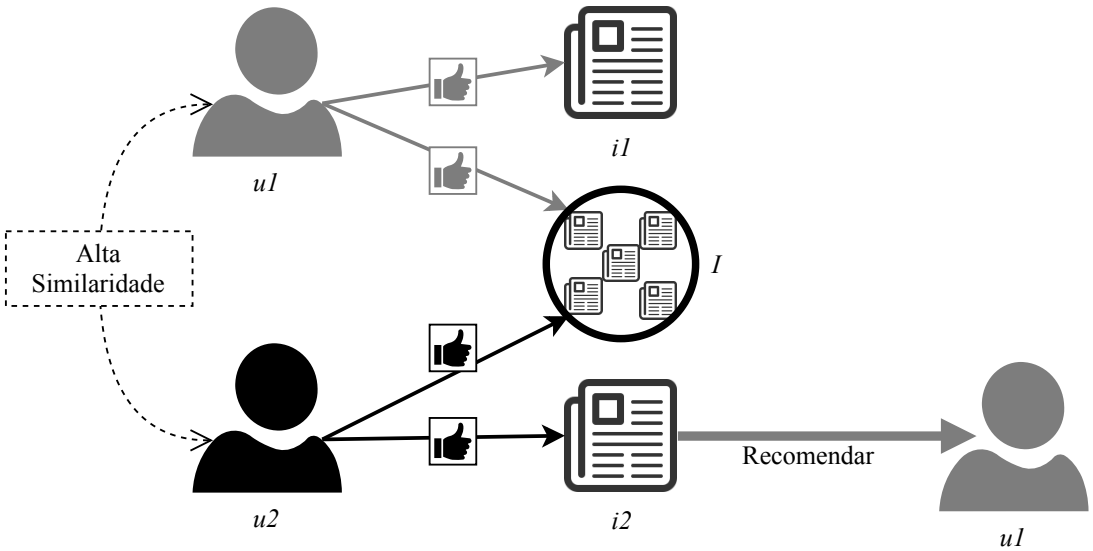
Nesta classe de SR, a recomendação é baseada exclusivamente nas avaliações que os usuários atribuem aos itens que eles consumiram (RICCI; ROKACH; SHAPIRA, 2015). Desta forma, itens diferentes que são avaliados de maneira semelhante, assim como usuários diferentes que avaliam os mesmos itens de forma parecida, são relacionados neste tipo de modelo. Como resultado, há duas subclasses de modelos de Filtragem Colaborativa, que são: (1) Filtragem Baseada em Usuários (*User-Based* Collaborative Filtering) e (2) Filtragem Baseada em Itens (*Item-Based* Collaborative Filtering); que serão descritos respectivamente nas Seções 3.1.2.1 e 3.1.2.2.

3.1.2.1 Filtragem Colaborativa Baseada em Usuários

Modelos de Filtragens Colaborativas Baseada em Usuários parte da premissa que usuários semelhantes tendem a se interessar por conteúdos semelhantes (NING; DESROSIERS; KARYPIS, 2015). Desta forma, esta classe de modelos de recomendação foca em identificar a similaridade entre usuários e utilizar as informações daqueles que tem alta similaridade entre si para gerar recomendações.

Por exemplo, vamos considerar um cenário de recomendação de notícias, ilustrado na Figura 1, onde as recomendações devem ser feitas ao usuário u_1. Considerando que o usuário u_2 é altamente similar ao usuário u_1, pelo fato de ambos gostarem de notícias em comum, representadas pelo conjunto de notícias I, um modelo de Filtragem Colaborativa Baseada em Usuários recomendaria as notícias que o usuário u_2 gostou que ainda não foram lidas pelo usuário u_1, neste caso, a notícia i_2 seria recomendada.

Figura 1 – Ilustração que exemplifica o funcionamento de um modelo de Filtragem Colaborativa Baseada em Usuários em um cenário de recomendações de notícias.



Fonte: Autor.

A estimativa de quais itens um determinado usuário tende a ter preferência é realizada através da predição da avaliação que este usuário realizaria aos itens que ele ainda não consumiu. Em Filtragem Colaborativa (FC) Baseada em Usuários, esta predição é calculada com base nas avaliações que usuários semelhantes realizaram à estes itens.

$$p(u,i) = \frac{\sum_{n \in N} sim(u,n) * a(n,i)}{\sum_{n \in N} sim(u,n)} \tag{1}$$

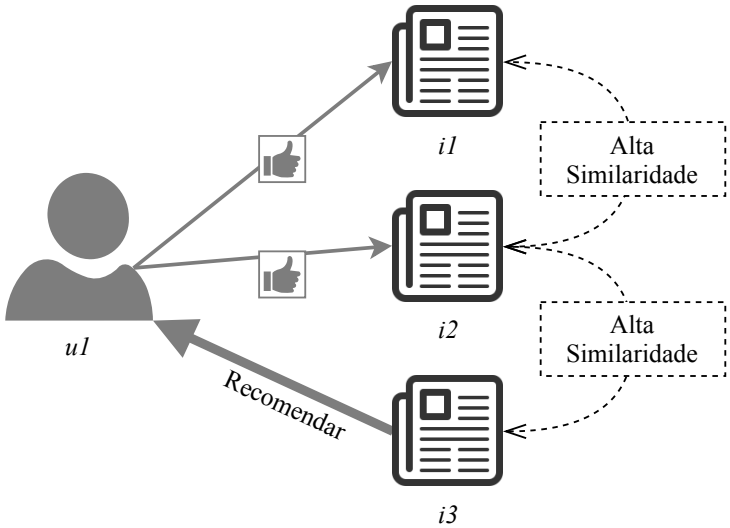
Sendo assim, $p(u,i)$, que é a predição da avaliação que um usuário u faria a um item i, é dada pela Equação (1), sendo N um conjunto de usuários semelhantes à u, $sim(u,n)$ a similaridade entre os usuários u e n e $a(n,i)$ é a avaliação que o usuário n atribuiu ao item i.

3.1.2.2 Filtragem Colaborativa Baseada em Itens

Em contrapartida à Filtragem Colaborativa Baseada em Usuários, a Filtragem Colaborativa Baseada em Itens parte da premissa que usuários tendem a se interessar por itens semelhantes aos que eles já se interessaram no passado (NING; DESROSIERS; KARYPIS, 2015). Sendo assim, o foco nesta classe de FC é identificar a similaridade entre os itens para produzir recomendações.

Seguindo o mesmo cenário de recomendação de notícias adotado na Seção 3.1.2.1, a Figura 2 apresenta um exemplo de um modelo de Filtragem Colaborativa Baseada em Itens. Neste cenário, um usuário u se interessou pelas notícias i_1 e i_2. Levando em conta que a notícia i_3 é similar às notícias i_1 e i_2 e ainda não foi lida por u, um modelo de FC Baseada em Itens recomendaria a notícia i_3 ao usuário u.

Figura 2 – Ilustração que exemplifica o funcionamento de um modelo de Filtragem Colaborativa Baseada em Itens em um cenário de recomendações de notícias.



Fonte: Autor

Nesta classe de FC, a predição da avaliação que usuários fariam aos itens que eles ainda não consumiram é realizada com base na Equação (2); onde $p(u,i)$ é a predição da avaliação que um usuário u faria a um item i, M é o conjunto de itens semelhantes à i, $sim(i, m)$ é a similaridade entre os itens i e m e $a(u, m)$ é a avaliação que o usuário u atribuiu ao item m.

$$p(u,i) = \frac{\sum_{m \in M} sim(i,m) * a(u,m)}{\sum_{m \in M} sim(i,m)} \tag{2}$$

3.1.3 Filtragem Baseada em Conteúdo (FBC)

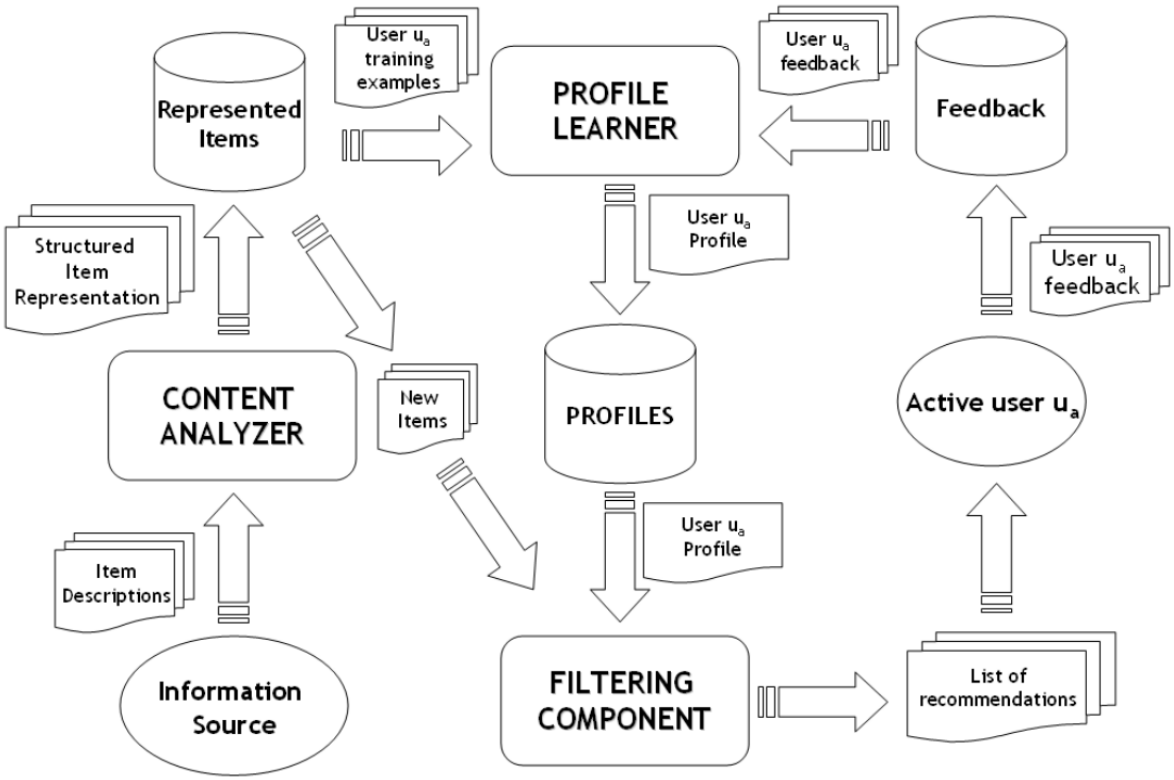
Semelhantemente à Filtragem Colaborativa Baseada em Itens, a Filtragem Baseada em Conteúdo, do inglês *Content-Based Filtering*, parte da premissa que usuários tendem a se interessar por itens semelhantes aos que lhes interessaram no passado (LOPS; DE GEMMIS; SEMERARO, 2011). Entretanto, modelos de Filtragem Baseada em Conteúdo são focados nas características dos itens consumidos e/ou avaliados pelos usuário, e não simplesmente na transação efetuada. Sendo assim, um item é composto por um conjunto de características que o define e o perfil de um usuário é construído com base nas características dos itens que o mesmo consome.

Segundo (LOPS; DE GEMMIS; SEMERARO, 2011), Sistemas de Recomendação por Filtragens Baseadas em Conteúdo produzem recomendação através de um processo dividido em três etapas, que são:

a) **Análise de Conteúdo** – Nesta etapa, os itens a serem recomendados devem ser modelados por suas características relevantes, para que possam ser consumidos pelas etapas seguintes. Estes itens podem ser estruturados ou não-estruturados. Itens estruturados possuem características definidas, como por exemplo, em um contexto de recomendação de filmes onde os itens podem ser modelados por características como: gênero, duração, atores principais, atores coadjuvantes, diretores, estúdio, produtora, idioma, ano de lançamento, prêmios, etc. Por outro lado, itens não-estruturados precisam passar por um pré-processamento para que informações relevantes sejam extraídas de forma estruturada. Notícias é um exemplo de itens não-estruturados, onde um método de pré-processamento valido poderia ser a aplicação de um modelo de processamento de linguagem natural no texto da notícia para extrair suas palavras chaves.
b) **Construção de Perfil** – A construção do perfil de um usuário é baseada nos itens já consumidos por ele, visando generalizar o conteúdo previamente consumido para identificar as características de itens que agradam o usuário, assim como as que o desagradam. Esta generalização normalmente é feita através de modelos de aprendizado de máquina.
c) **Filtragem** – Por fim, a etapa de Filtragem é responsável por encontrar itens relevantes à um usuário com base na correspondência de características dos itens com

o seu perfil, calculada através de métricas de similaridade. Sendo que os itens que possuem maior similaridade com o perfil do usuário são recomendados.

Figura 3 – Arquitetura em alto nível apresentando as etapas que compõem um Sistema de Recomendação por Filtragem Baseada em Conteúdo



Fonte: (LOPS; DE GEMMIS; SEMERARO, 2011).

A Figura 3, originada em (LOPS; DE GEMMIS; SEMERARO, 2011), apresenta as três etapas descritas acima em uma arquitetura em alto nível de Sistema de Recomendação por Filtragem Baseada em Conteúdo, mostrando o processo de recomendação e como as etapas são interligadas.

Também segundo (LOPS; DE GEMMIS; SEMERARO, 2011), pode-se listar as seguintes vantagens de modelos de Filtragem Baseada em Conteúdo em relação com à Filtragem Colaborativa:

a) Filtragem Baseada em Conteúdo baseia-se exclusivamente no comportamento do próprio usuário para construir o seu perfil e gerar recomendações, enquanto modelos de recomendação por Filtragem Colaborativa precisam identificar os itens populares entre usuários similares.
b) As recomendações realizadas por modelos de Filtragem Baseada em Conteúdo são mais facilmente justificáveis devido ao fato de serem geradas com base na correspondência de características. Por outro lado, em Filtragem Colaborativa, esta

justificativa se torna mais complexa por depender do comportamento de usuários similares, sendo que essa similaridade nem sempre é simples de se identificar.

c) Filtragem Baseada em Conteúdo não sofre de problemas de *cold-start* para novos itens que são inseridos no sistema, por não dependerem da avaliação de outros usuários para serem recomendados.

Além disso, algumas desvantagens também podem ser listadas:

a) Pelo fato de ser baseado nas características dos itens, este tipo de modelo precisa que estas características sejam suficientes para discriminar itens que são ou não do interesse do usuário.
b) Filtragem Baseada em Conteúdo possuí uma tendência a produzir recomendação com baixo nível de inovação, restringindo o conjunto de itens a serem recomendados.
c) Apesar de não sofrer com problemas de *cold-start* para novos itens, novos usuário sofrem desse problema, já que a construção do seu perfil depende do mesmo consumir ou avaliar uma determinada quantidade de itens para que as recomendações sejam geradas de modo mais assertivo.

3.1.4 Sistemas de Recomendação Baseados em Conhecimento (SRBC)

Ambos os modelos de Filtragem Colaborativa e Filtragem Baseada em Conteúdo necessitam de uma grande quantidade de dados históricos, seja para identificar a semelhança entre usuários ou construir o perfil de interesses de um usuário (BURKE, 2000). Estes modelos são eficazes em domínios como comércio digital, filmes, músicas e entretenimento em geral, onde usuários frequentemente interagem com o sistema e consomem itens diversos. Porém, estes modelos tendem a não ser adequados em cenários de recomendação onde os itens são consumidos de modo infrequente pelos usuários, como em compra de imóveis e automóveis (FELFERNIG et al., 2011).

Em alternativa, modelos de Sistemas de Recomendação Baseados em Conhecimento tem como objetivo gerar recomendações através da exploração de conhecimentos estruturados sobre o domínio do problema e das especificações das necessidades atuais do usuário, sem a necessidade de analisar dados históricos (BURKE, 2000). Estas características fazem com que este tipo de modelo seja capaz de gerar recomendações relevantes, além de complementar os modelos citados anteriormente por ser capaz de performar mesmo quando os demais não podem devido à ausência de dados. Em contrapartida, os Sistemas de Recomendação Baseados em Conhecimento sofrem do fenômeno conhecido como "gargalo de aquisição de dados"(do inglês *knowledge acquisition bottleneck*), que representa o alto esforço necessário para que um profissional da computação (normalmente um engenheiro do conhecimento) possa extrair os conhecimentos de especialistas do domínio do problema e persisti-los de forma estruturada para serem consumidos por modelos computacionais (FELFERNIG et al., 2011).

Em (FELFERNIG et al., 2011), apesar de terem um funcionamento similar, pelo fato gerarem recomendações através da exploração de conhecimentos e partindo das necessidades dos usuários, os SRBCs são divididos em duas categorias, que diferem essencialmente pela forma como as recomendações são calculadas. Essas categorias são:

a) **Sistemas de Recomendação Baseados em Casos** — Provenientes de modelos de Raciocínio Baseado em Casos (*Case-Based Reasoning*), que tem como objetivo utilizar problemas previamente solucionados como base para resolução de novos problemas semelhantes através da adaptação das soluções antigas (BRIDGE et al., 2005; LORENZI; RICCI, 2003).

b) **Sistemas de Recomendação Baseados em Restrições** — utilizam Bases de Conhecimento, que são compostas por regras explicitas, para identificar os itens mais adequados aos usuários através de técnicas oriundas de soluções de Problemas de Satisfação de Restrições (do inglês *Constraint Satisfaction Problems*) (FELFERNIG et al., 2011)

3.1.5 Sistemas de Recomendação Sensíveis ao Contexto (SRSC)

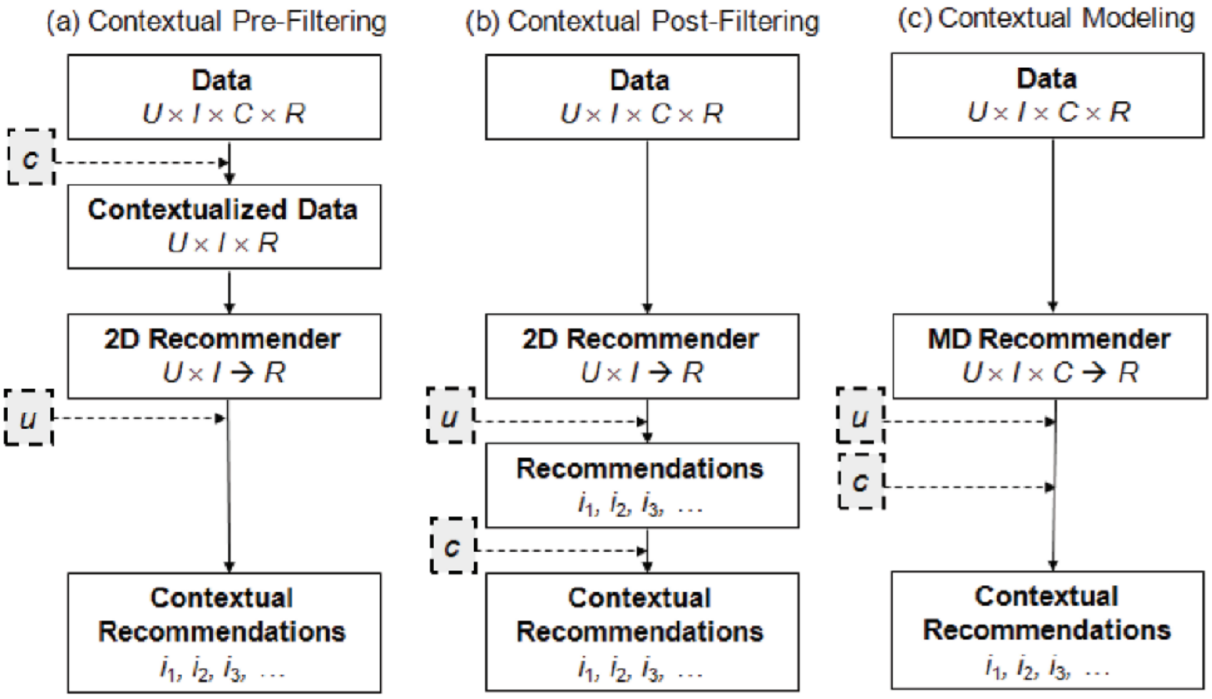
Esta abordagem parte da premissa de que as recomendações devem ser personalizadas em relação ao contexto no qual a recomendação é realizada (ADOMAVICIUS; TUZHILIN, 2011). Considerando um cenário de recomendação em uma aplicação de *delivery* de comida, informações como o horário e dia da semana podem caracterizar informações de contexto. Neste cenário, essas informações contextuais podem ser utilizadas, por exemplo, para identificar se o usuário está realizando um pedido para o café da manhã, almoço ou jantar, ou se o pedido está sendo realizado em um dia útil ou no final de semana. Com isso, o sistema é capaz de identificar as preferências do usuário em situações especificas. Por exemplo, se um usuário come pizzas frequentemente aos sábados a noite, ou tem o costume de pedir sobremesas no horário de almoço.

Segundo (ADOMAVICIUS; TUZHILIN, 2011), é difícil estabelecer uma definição universal para o termo “contexto”, pois é um conceito amplo que é discutido por diversas áreas de conhecimento, sendo que cada área tende a enviesar este conceito dependendo do objetivo do estudo. Sendo assim, o trabalho em (ADOMAVICIUS; TUZHILIN, 2011) apresenta algumas definições de contexto, oriundas de diversas áreas, que podem ser relevantes à Sistemas de Recomendação, sendo que a principal taxonomia adotada foi proposta por Dourish em (DOURISH, 2004). Esta taxonomia de Dourish (DOURISH, 2004) divide informações contextuais em duas categorias:

a) **Representacional** — Onde as informações contextuais são definidas por um conjunto finito de características observáveis, definidas a priori e que não mudam significantemente através do tempo.

b) **Interacional** — Onde as informações contextuais não são necessariamente observáveis e tem uma relação direta com o comportamento dos usuários. Ou seja, além

Figura 4 – Fluxograma que ilustram os paradigmas de incorporação de informações contextuais em Sistemas de Recomendação.



Fonte: (ADOMAVICIUS; TUZHILIN, 2011).

do contexto influenciar as ações dos usuários, novos contextos podem ser criados levando em conta determinados padrões de comportamento..

Em (ADOMAVICIUS; TUZHILIN, 2011), também são apresentadas algumas abordagens para obtenção de informações contextuais, que são as seguintes:

a) **Explicitamente** — Interagindo diretamente com usuários e outras fontes para coletar informações relevantes ao contexto, através de questionários ou entrevistas.

b) **Implicitamente** — Utilizando dados concretos dos usuários, tais como hora, temperatura, coordenadas de GPS e dados sensoriais, que podem ser obtidos sem interação direta.

c) **Inferidamente** — Deduzindo informações contextuais, através de técnicas de mineração de dados, baseando-se no comportamento dos usuários. Por exemplo, inferir a faixa etária do usuário com base nos filmes que ela assiste.

Por fim, semelhantemente aos Sistemas de Recomendação Baseados em Conhecimento, os SRSCs também são considerados complementares aos demais métodos, pois não apresentam uma nova abordagem para gerar recomendações, e sim exploram os benefícios da integração de informações contextuais no processo de recomendação. Com isso em vista, pode-se destacar três metodologias para agregar contexto em Sistemas de Recomendação (ADOMAVICIUS; TUZHILIN, 2011), ilustradas na Figura 4, que são:

a) **Pré-Filtragem Contextual** — Também chamado de Contextualização de Entrada de Recomendação, este método considera informações contextuais dos dados de

entrada do Sistema de Recomendação para filtra-los. Em seguida, recomendações são geradas através de métodos de recomendação tradicionais.

b) **Pós-Filtragem Contextual** — Também referenciado como Contextualização de Saída de Recomendação, este método inicialmente desconsidera as informações contextuais e gera as recomendações através de métodos tradicionais. Posteriormente, as recomendações são contextualizadas.

c) **Modelagem Contextual** — Também conhecido como Contextualização da Função de Recomendação, este método utiliza informações contextuais diretamente no processo de recomendação.

3.1.6 Sistemas de Recomendação Híbridos (*Ensemble*)

Um técnica muito popular para aumentar a precisão de modelos de Aprendizado de Máquina é chamada de *Ensemble* (DIETTERICH, 2000). O aumento de precisão é atingido através da combinação de dois ou mais modelos Aprendizado de Máquina. Similarmente, em SRs, modelos de recomendação também são frequentemente combinados para a obtenção de sistemas mais precisos. Entretanto, apesar de Sistemas de Recomendação serem considerados modelos de Aprendizado de Máquina, a nomenclatura adotada pela literatura para a combinação de modelos em SRs é Sistemas de Recomendação Híbridos (BURKE, 2002).

Outro benefício dos Sistemas de Recomendação Híbridos está relacionado com as diferenças entre as categorias de SR, apresentadas nas Seções 3.1.2, 3.1.3, 3.1.4 e 3.1.5. Cada tipo de SR oferece vantagens e desvantagens específicas. Desta forma, Sistemas de Recomendação Híbridos buscam combinar estes sistemas para usufruir de suas vantagens minimizando suas desvantagens individuais (BURKE, 2002).

Segundo Burke (BURKE, 2007), Sistemas de Recomendação podem ser combinados através de sete metodologias, que são:

a) **Ponderada** — Onde pesos são atribuídos à cada um dos modelos, definindo a sua importância.

b) **Comutada** — Com base em um critério pré-estabelecido, o sistema analisa informações sobre a recomendação que será realizada, tais como o usuário ou o contexto, e seleciona o modelo mais adequado para este cenário.

c) **Mista** — As listas de itens a serem recomendados geradas por cada um dos modelos individualmente são combinadas em uma única lista.

d) **Características Combinadas** — Onde características extraídas de uma categoria de SR é utilizada como dado de entrada para um modelo de outra categoria de SRs que será responsável por calcular as recomendações.

e) **Características Aumentadas** — Este método é semelhante à combinação por Características Combinadas. Porém, neste caso, os dados de treinamento são aumentados por um modelo de SR e depois são utilizados para treinar outro modelo de SR.

f) **Cascata** — Combinação em cascata mantém uma relação de hierarquia entre os modelos de recomendação, onde um modelo de hierarquia mais baixa é responsável por aprimorar as recomendações realizadas por modelos de hierarquia mais alta, nunca desfazendo decisões definidas pelos modelos acima na hierarquia.

g) **Meta-level** — Neste método, modelos gerados a partir de uma categoria de SR são utilizados como entrada para outra categoria de SR gerar as recomendações. A combinação Meta-level difere da combinação por Características Aumentadas pelo fato de o modelo que gera as recomendações não utilizar os dados originais como entrada, utilizando apenas os modelos gerados pelo SR que, por sua vez, pré-processou os dados originais.

3.1.7 Métricas de Similaridade

Muitos dos modelos de SR apresentados até aqui exploram a similaridade entre elementos de um cenário de recomendação para calcular os itens mais adequados para cada usuário. Usualmente, os usuário e itens de um SR são representados por vetores N-dimensionais, favorecendo a utilização de métricas matemáticas para o cálculo de similaridade entre elementos. Em (AMATRIAIN et al., 2011), as métricas mais populares utilizadas para o cálculo de similaridade em Sistemas de Recomendação são apresentadas. Estas métricas serão descritas na sequência.

Primeiramente, a similaridade $sim(x, y)$ entre dois elementos em um SR, aqui representados pelos vetores N-dimensionais x e y, é frequentemente calculada com base na distância $d(x, y)$ entre os mesmos.

$$d(x, y) = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \tag{3}$$

A Distância Euclidiana é uma das métricas mais populares, representada pela Equação (3), principalmente para cenários onde a proximidade geométrica é relevante, sendo i a i-ésima dimensão dos vetores x e y. Outra característica relevante para a utilização desta métrica é a escala das dimensões dos vetores nos quais a distância será calculada, sendo adequada em cenários onde as dimensões possuem a mesma escala. Quando as dimensões possuem escalas diferentes, as dimensões que apresentam valores maiores têm mais impacto no cálculo da Distância Euclidiana, o que muitas vezes não é desejado. Uma alternativa para isso seria aplicar uma normalização dos vetores, igualizando as escalas de sus dimensões. Alguns métodos populares de transformação de dados serão apresentadas na Seção 3.3.

$$d(x, y) = \sum_{i=1}^{N} |x_i - y_i| \tag{4}$$

A Distância de *Manhattan*, dada pela Equação (4), é uma alternativa à Distância Euclidiana, pois também preserva uma relação geométrica entre os elementos. Está métrica apresenta uma desvantagem computacional pela necessidade de se calcular o módulo da diferença entre

valores, que também é uma operação custosa.

$$d(x, y) = 1 - \cos(x, y) = 1 - \frac{(x \cdot y)}{\|x\|\|y\|} \tag{5}$$

Outra métrica popular em SRs é a Distância dos Cossenos, definida pela Equação (5). Neste caso, a distância entre os elementos é determinada pela direção e sentido de seus vetores.

$$d(x, y) = \sqrt{(x - y)\sigma^{-1}(x - y)^T} \tag{6}$$

A Distância de Mahalanobis, definida pela Equação (6), é outra alternativa popularmente adotada em SRs, sendo σ a matriz de covariância dos dados. Esta métrica difere das demais principalmente pelo fato de considerar as correlações entre as dimensões. Outra característica desta métrica é ser invariante à escala, ou seja, não sofre impacto quando aplicada em vetores que possuem dimensões em escalas distintas.

$$sim(x, y) = \frac{1}{1 + d(x, y)} \tag{7}$$

Para métricas de distância não-normalizadas, usualmente utiliza-se a Equação (7) para transformar a distância entre os elementos em similaridade.

$$sim(x, y) = 1 - d(x, y) \tag{8}$$

Algumas métricas de distância, como a Distância dos Cossenos, calculam a distância entre elementos em uma escala entre 0 e 1, e podem ser convertidas para similaridade através da Equação (8).

$$Pearson(x, y) = \frac{cov(x, y)}{\sigma_x \times \sigma_y} \tag{9}$$

Outro método para calcular a similaridade entre elementos é através da correlação, sendo que a métrica mais popular em SRs para este tipo de cálculo é a Correlação de Pearson, definida pela Equação (9).

Ademais, métricas de similaridade especificas foram propostas para lidar com variáveis binárias. Estas métricas são baseadas na quantidade de correspondências dos valores de x e y em cada dimensão. Desta forma, existem quatro possíveis combinações de valores de x e y, representadas por M_{00}, M_{01}, M_{10} e M_{11}; sendo M_{00} a quantidade de dimensões onde ambos x e y apresentam valor 0, M_{01} a quantidade de dimensões onde x apresenta valor 0 e y apresenta valor 1, M_{10} a quantidade de dimensões onde x apresenta valor 1 e y apresenta valor 0 e M_{11} a quantidade de dimensões onde ambos x e y apresentam valor 1.

$$SMC(x, y) = \frac{M_{00} + M_{11}}{M_{00} + M_{01} + M_{10} + M_{11}} \tag{10}$$

Uma das métricas para variáveis binárias é a *Simple Matching Coefficient* (SMC), dado pela Equação (10).

$$JC(x, y) = \frac{M_{11}}{M_{01} + M_{10} + M_{11}} \tag{11}$$

Outra métrica popular para variáveis binárias é o Coeficiente de Jaccard (JC), definido pela Equação (11).

$$TC(x, y) = \frac{x \cdot y}{\|x\|^2 \|y\|^2 - x \cdot y} \tag{12}$$

Por fim, o coeficiente de Jaccard possuí uma versão estendida, conhecida também como Coeficiente de Tanimoto (TC), definida pela Equação (12), proposta para ser aplicada também para valores continuos.

3.1.8 Avaliação

Em (GUNAWARDANA; SHANI, 2015), são apresentados diversos métodos para avaliar e comparar Sistemas de Recomendação. Pode-se destacar três destes métodos, que são:

a) **Experimentos *Offline*** — Onde experimentos são realizados a partir da simulação do comportamento dos usuários no sistema com base em dados históricos, partindo da premissa de que estes usuários tendem a manter o seu padrão de comportamento futuramente.

b) **Estudos com Usuários** — Neste método, uma seleção de voluntários são recrutados para interagir com o sistema. Em seguida, estudos são realizados com base em entrevistas, questionários e análise dos dados de interação com o sistema utilizando técnicas oriundas de Interface Humano-Computador.

c) **Avaliação *Online*** — Onde pode-se medir tanto a influência do Sistema de Recomendação no comportamento dos usuários quanto a qualidade das recomendações realizadas, através da análise de comportamento e *feedback* dos usuários.

Além disso, ao avaliar um SR, aspectos específicos do sistema podem ou não ser selecionados para serem medidos, dependendo de sua relevância para o cenário no qual será aplicado (GUNAWARDANA; SHANI, 2015). Alguns dos principais atributos são:

a) **Acurácia Preditiva** — É o atributo mais amplamente explorado pela comunidade acadêmica, considerando que a maioria dos modelos de SR são baseados na predição da transação. Por isso, será abordado com mais profundidade na Seção 3.1.8.1. De um modo geral, este atributo avalia as predições realizadas para produzir as recomendações e mede o quão próximos foram dos valores reais.

b) **Cobertura** — Em alguns casos, o SR pode não conseguir gerar recomendações para um usuário ou não recomendar um determinado item para nenhum usuário, em geral por falta de dados históricos relacionados aos mesmos. Sendo assim, este atributo mede a porcentagem de itens do sistema que são recomendados em relação com o total de itens disponíveis, ou a porcentagem de usuários para os quais o sistema é capaz de produzir recomendações. Um problema bem conhecido de SRs relacionado à Cobertura é o *cold-start*, que representa a incapacidade dos SRs de recomendar itens recém adicionados no sistema, ou gerar recomendações para novos usuários.

c) **Confiança** — Representa o nível de confiança do SR em relação a uma determinada recomendação produzida. De um modo geral, quanto maior a quantidade de dados relacionados aos itens recomendados e aos usuários que recebem as recomendações, maior é a Confiança do sistema.

d) **Confiabilidade** — De modo análogo à Confiança, a Confiabilidade está relacionada com a confiança que os usuários têm nas recomendações que recebem do sistema. Alguns dos principais fatores que impactam a Confiabilidade são o quanto os usuários gostam das recomendações geradas e o quanto compreendem as razões pelas quais receberam-as. Por exemplo, um usuário que recebe muitas recomendações que não agradam o seu gosto impacta negativamente a Confiabilidade do sistema. Por outro lado, um SR que é capaz de explicar aos usuários como foram produzidas as recomendações tendem a ter a Confiabilidade elevada.

e) **Originalidade** — Mede a capacidade do SR de recomendar itens novos, ainda não consumidos pelo usuário. Um estratégia simples para elevar a Originalidade é não recomendar itens conhecidamente já consumidos pelos usuários. Em contrapartida, este atributo pode ser complexo de ser avaliado, pois usualmente os usuários não identificam explicitamente ao sistema todos os itens já consumidos no passado.

f) **Serendipidade** — Este atributo também está relacionado com itens novos que são recomendados aos usuários. Porém, o foco neste caso é avaliar o quão surpreendente foram as recomendações, no sentido de inesperadamente positivas. Por exemplo, supondo o caso de um usuário que sempre lê as obras de um determinado autor, recomendar um novo livro escrito por este autor para este usuário pode ser considerada uma recomendação nova e relevante, mas não inesperada. Por outro lado, se o SR consegue identificar outros autores que o usuário gosta e recomenda suas obras, estas recomendações podem ser consideradas surpreendentes.

g) **Utilidade** — Representa os benefícios obtidos pela adoção do SR. Por exemplo, o aumento de vendas de produtos ou a redução da taxa de evasão de usuários de uma determinada ferramenta.

h) **Risco** — Avalia o possibilidade de produzir recomendações ruins, levando em conta o impacto dessas recomendações dependendo do cenário em que o SR é aplicado. Por exemplo, uma análise de risco associado com a recomendação de investimentos em ações na bolsa de valores ou com a recomendação de medicamentos para pacientes.

i) **Robustez** — Mede a capacidade do SR em lidar com ruído, inserção de dados errados e ataques que tentem enviesar o sistema.

j) **Privacidade** — Este atributo está relacionado com questões de segurança da informação, medindo a capacidade do SR em manter os dados dos usuários anônimos.

k) **Adaptabilidade** — Mede a capacidade do sistema em se adaptar a mudanças rápidas no domínio no qual são aplicados. Por exemplo, um SR de notícias deve se

adaptar rapidamente a novas notícias importantes que podem acontecer a qualquer momento.

l) **Escalabilidade** — Avalia a capacidade do sistema em lidar com o crescimento da quantidade de dados no qual ele deve manipular.

3.1.8.1 Acurácia Preditiva

Nesta seção, serão apresentadas as principais métricas matemáticas utilizadas para avaliar a capacidade preditiva dos SRs. Segundo (GUNAWARDANA; SHANI, 2015), a capacidade preditiva de um SR pode ser avaliada através de três vertentes, que são: (1) precisão ao prever a avaliação dos usuários; (2) precisão ao prever a adoção dos usuários; e (3) precisão no ranqueamento de itens. Estas vertentes serão discutidas na sequência.

3.1.8.1.1 Precisão ao Prever a Avaliação dos Usuários

Frequentemente, usuários tem a possibilidade de avaliar os itens nos quais eles consumiram respeitando alguma unidade de medida, podendo ser, por exemplo, uma nota entre 1 e 5, ou apontando os itens que o usuário gostou ou não gostou. Um método amplamente adotado em SRs para gerar recomendações foca em prever a avaliação que usuários fariam para os itens e, assim, recomendar os itens que possuem avaliações preditas mais elevadas (GUNAWARDANA; SHANI, 2015). Portanto, o SR é tido como uma técnica de Aprendizado de Máquina para regressão neste cenário. Consequentemente, pode-se avaliar um SR através da comparação da avaliação predita com a avaliação real feita pelos usuários. Tradicionalmente, este tipo de avaliação é feita com base em métricas de erro.

$$MSE = \frac{1}{|\mathcal{T}|} \sum_{(u,i)\in\mathcal{T}} (\hat{r}_{ui} - r_{ui})^2 \tag{13}$$

$$RMSE = \sqrt{\frac{1}{|\mathcal{T}|} \sum_{(u,i)\in\mathcal{T}} (\hat{r}_{ui} - r_{ui})^2} \tag{14}$$

$$MAE = \frac{1}{|\mathcal{T}|} \sum_{(u,i)\in\mathcal{T}} |\hat{r}_{ui} - r_{ui}| \tag{15}$$

As Equações (13), (14) e (15) definem, respectivamente, as métricas de Erro Quadrático Médio (do inglês *Mean Squared Error* - MSE), Raiz do Erro Quadrático Médio (do inglês *Root Mean Squared Error* - RMSE) e Erro Médio Absoluto (do inglês *Mean Absolute Error* - MAE), que são as principais métricas de erro utilizadas para avaliar Sistemas de Recomendação; onde $\mathcal{T}$ representa o conjunto de avaliações preditas, $\hat{r}_{ui}$ representa a previsão da avaliação do usuário u ao item i e r_{ui} representa a avaliação real observada do usuário u ao item i.

Tabela 1 – Matriz de Confusão que relaciona os itens recomendados pelo Sistema de Recomendação com os itens "úteis" aos usuários, ou seja, os itens que os usuários de fato consumiram, se interessaram e/ou avaliaram positivamente.

	Recomendados	**Não Recomendados**
Úteis	Verdadeiros Positivos (VP)	Falsos Negativos (FN)
Não Úteis	Falsos Positivos (FP)	Verdadeiros Negativos (VN)

3.1.8.1.2 Precisão ao Prever a Adoção dos Usuários

Outra forma de avaliar Sistemas de Recomendação é baseada na adoção das recomendações pelos usuários, ou seja, a taxa de recomendações realizadas pelo SR que foram úteis ou não aos usuários. Uma recomendação pode ser considerada útil quando o usuário consome, gosta ou demonstra algum tipo de interesse pelo item recomendado. Por exemplo, para um SR aplicado em *e-commerce*, várias atividades realizadas pelo usuário no sistema podem indicar se um produto foi útil ou não ao mesmo, tais como se os produtos foram visualizados, adicionados ao carrinho, comprados ou avaliados positivamente ou negativamente após a compra. Logo, neste tipo de cenário, o SR é tido como uma técnica de Aprendizado de Máquina para classificação, onde os itens devem ser classificados como úteis ou não-úteis aos usuários.

Este tipo de avaliação é realizada com base na Matriz de Confusão, ilustrada na Tabela 1, composta por quatro valores, que são: verdadeiros positivos (VP), falsos positivos (FP), verdadeiros negativos (VN) e falsos negativos (FN); onde VP representa os itens recomendados pelo SR que foram úteis ao usuário, FP os itens recomendados pelo SR que não foram úteis ao usuário, FN os itens não recomendados pelo SR que foram úteis ao usuário e VN os itens não recomendados que não foram úteis ao usuário.

Muitas métricas utilizam os valores da Matriz de Confusão para avaliar características específicas de modelos de Aprendizado de Máquina. Segundo (GUNAWARDANA; SHANI, 2015), este tipo de avaliação é aplicada em SRs baseando-se em três métricas principais, que são:

a) **Precisão** — definida pela Equação (16), representa a porcentagem de itens recomendados que foram úteis aos usuários, desconsiderando os itens que não foram recomendados.

$$Precisão = \frac{VP}{VP + FP} \tag{16}$$

b) **Sensibilidade** — definida pela Equação (17), representa a porcentagem de itens úteis recomendados em relação ao total de itens úteis aos usuários, desconsiderando os itens que não são úteis aos mesmos. A Sensibilidade também é referenciada como **Revocação**, **Taxa de Acerto** e **Taxa de Verdadeiros Positivos**.

$$Sensibilidade = \frac{VP}{VP + FN} \tag{17}$$

c) **Taxa de Falsos Positivos (TFP)** — definida pela Equação (18), representa a porcentagem de itens não-úteis recomendados em relação ao total de itens não-úteis

aos usuários, desconsiderando os itens que são úteis aos mesmos.

$$TFP = \frac{FP}{FP + VN} \tag{18}$$

A Precisão é comumente adotada em cenários onde a lista de recomendações possuí um tamanho fixo pré-determinado (GUNAWARDANA; SHANI, 2015). Desta forma, a avaliação da Precisão do SR para uma lista de recomendação de tamanho N é conhecida como "Precisão em N"($P@N$) (GUNAWARDANA; SHANI, 2015).

Em contrapartida, em cenários onde a lista de recomendações não possuí restrição quanto ao seu tamanho, a avaliação do SR é feita através da composição de curvas, variando-se o tamanho N da lista de recomendações (GUNAWARDANA; SHANI, 2015). Além da métrica $P@N$, pode-se também compor curvas que relacionam as três métricas citadas anteriormente, sendo que as principais curvas são:

a) **Curvas de Precisão-Revocação** — Como o nome já diz, este tipo de curva relaciona a Precisão com a Revocação (Sensibilidade) do SR. Este método é adequado para avaliar dados desbalanceados, ou seja, quando há muitos dados de uma determinada classe e poucos de outra, pelo fato de ser sensível a grande quantidade de falsos positivos (DAVIS; GOADRICH, 2006). Portanto, Curvas de Precisão-Revocação são basante populares em SRs, uma vez que normalmente uma parcela muito pequena dos itens disponíveis em um cenário de recomendação são úteis a um determinado usuário.

b) **Curva ROC** — Proveniente do termo em inglês "*Receiver Operating Characteristic*" (Característica de Operação do Receptor), este tipo de curva relaciona a Taxa de Verdadeiros Positivos (Sensibilidade) com a Taxa de Falsos Positivos (TFP). Diferentemente das Curvas de Precisão-Revocação, este método é mais adequado a problemas de classificação onde as classes são balanceadas, ou seja, quando as classes possuem uma quantidade semelhante de observações entre si (GUNAWARDANA; SHANI, 2015). Ademais, a Curva ROC é útil em cenários onde a recomendação de falsos positivos está agregada com algum tipo de risco ou impacto negativo. Por exemplo, em um cenário de recomendação de medicação, onde um medicamento inapropriadamente sugerido pode trazer efeitos colaterais e até mesmo danos à saúde do paciente; ou em um cenário de recomendação de investimentos, onde um falso positivo pode acarretar em perda de capital.

$$F_1 = 2 \cdot \frac{Precisão \cdot Sensibilidade}{Precisão + Sensibilidade} \tag{19}$$

Estas curvas podem ser sumarizadas, através de algumas métricas, para comparar o desempenho de diferentes algoritmos. Uma das métricas mais populares utilizadas para este fim é a chamada F-*Measure*, também referenciada como F-*Score* ou F_1, definida pela Equação (19). Como a F_1 considera a Precisão e a Revocação, ela só pode ser relacionada com a Curva Precisão-Revocação, e não com a Curva ROC. Outra métrica bastante popular é a Área Sob a

Curva, do inglês *Area Under Curve* (AUC) (GUNAWARDANA; SHANI, 2015), que pode ser adotada tanto para a Curva Precisão-Revocação quanto para a Curva ROC.

3.1.8.1.3 Precisão no Ranqueamento de Itens

Por fim, Sistemas de Recomendação podem ser avaliados com base na ordem em que os itens são apresentados na lista de recomendações (GUNAWARDANA; SHANI, 2015). O processo de definição da ordem dos itens a serem recomendados é conhecido como Ranqueamento. Neste tipo de avaliação, compara-se a lista de recomendações geradas pelo SR com uma lista de recomendações ideal, composta pelos itens que usuário mais gostou, sendo que os itens melhor avaliados sempre ficam à frente da lista.

Uma métrica bastante popular para avaliação de ranqueamento em SRs é baseada no chamado *Discounted Cumulative Gain* (DCG), definido pela Equação (20), onde os itens sofrem um desconto logaritmo do ganho em que proporcionam ao usuário com base na sua posição na lista de recomendação. Esta avaliação parte da premissa que a probabilidade de um usuário verificar um item da lista de recomendações diminuí conforme o item é apresentado numa posição posterior da lista. Portanto, os itens que oferecem maior ganho ao usuário devem ser recomendados primeiro.

$$DCG = \frac{1}{N} \sum_{u \in N} \sum_{j=1}^{J} \frac{g_{ui_j}}{\log_b(1+j)} \tag{20}$$

$$nDCG = \frac{DCG}{DCG^*} \tag{21}$$

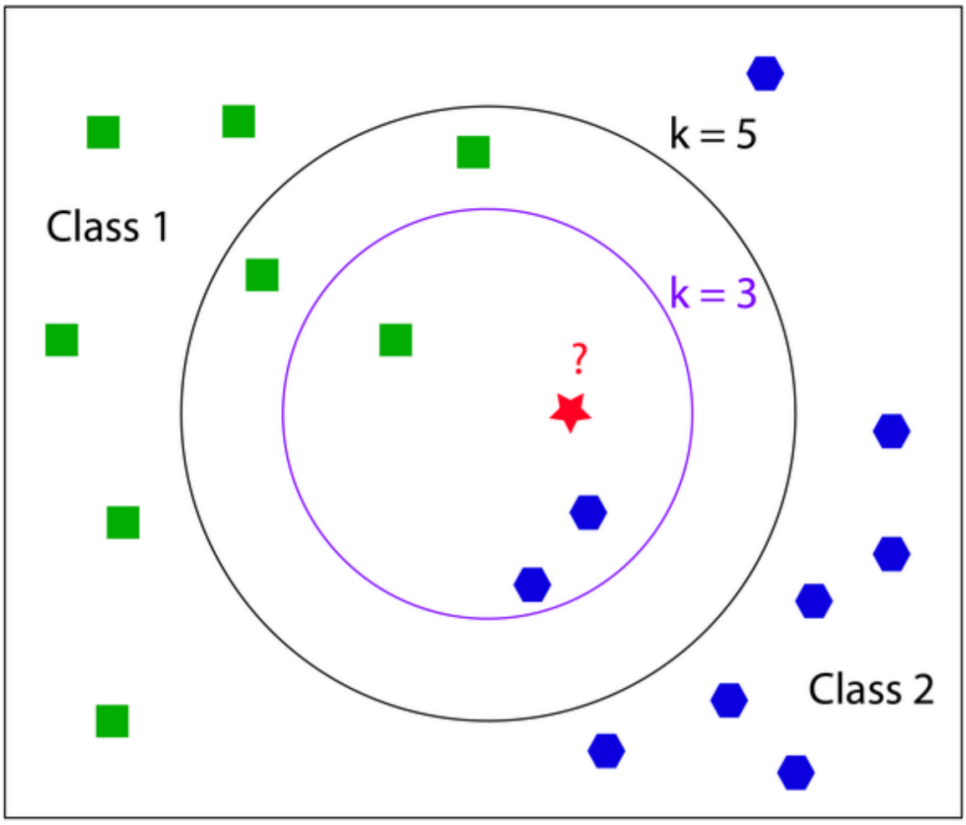
Na Equação (20), sendo N um conjunto de usuários que recebem uma lista com J recomendações, g_{ui_j} representa o ganho proporcionado ao usuário u pela recomendação do item i na j-ésima posição da lista de recomendações e b é a base do logaritmo, que deve ser pré-determinada. Após o cálculo do DCG, calcula-se o chamado *Normalized Discounted Cumulative Gain* (nDCG), definido pela Equação (21), onde DCG^* representa a lista ideal de recomendações, que nada mais é do que o cálculo do DCG aplicado aos mesmos J itens, porém ordenados de forma decrescente pelo g_{ui_j}. Desta forma, o DCG^* representa o maior DCG possível a ser obtido para para os J itens a serem recomendados. Portanto, o $nDCG$ assume um valor entre 0 e 1, que representa a porcentagem de ganho obtido pelas recomendações em relação ao ganho ótimo que poderia ser obtido.

3.2 *K*-NEAREST NEIGHBORS (*K*NN)

Proposto em (COVER; HART et al., 1967), o kNN, ou k-Vizinhos Mais Próximos, é um modelo tradicional de classificação e regressão. Este modelo parte do princípio que uma determinada observação deve ser classificada com base nas classes das observações mais próximas. Portanto, este algoritmo calcula os k vizinhos mais próximos de um elemento, com base em

alguma métrica de distância. A Figura 5 apresenta um exemplo do funcionamento do kNN para classificar um elemento de entrada entre duas classes.

Figura 5 – Exemplo do funcionamento do modelo kNN para diferentes valores de k. Sendo a observação objetivo, representada por $\star$, classificada como $\hexagon$ para $k = 3$ e como $\square$ para $k = 5$.



Fonte: (COVER; HART et al., 1967)

O kNN recebe os seguintes parâmetros de entrada:

a) Uma lista de observações anotadas para treinamento, representadas por vetores n-dimensionais, usadas como referência para aplicação da classificação/regressão;
b) Uma observação não-anotadas, representada por um vetor n-dimensional, para a qual o modelo deverá calcular a classificação/regressão;
c) Um valor k que define a quantidade de vizinhos mais próximos a serem considerados para cálculo da classificação/regressão para cada observação objetivo; e
d) A métrica de similaridade que deverá ser utilizada para calcular a distância entre a observação objetivo e as observações para treinamento.

Uma característica do modelo kNN é que ele não demanda nenhum processamento computacional para treinamento, sendo que toda a computação esta concentrada na etapa de predição, onde são calculados os vizinhos mais próximos (COVER; HART et al., 1967). A vantagem desta característica é que o modelo é capaz de considerar novos dados de treinamento sem nenhum impacto. Em contrapartida, é um modelo que demanda mais espaço de armazenamento,

pois todos os dados de treinamento precisão estar disponíveis sempre que uma predição for calculada.

3.3 PRÉ-PROCESSAMENTO DE DADOS

Principalmente em cenários onde deve-se manipular dados do mundo real, normalmente estes dados devem passar por um tratamento para prepará-los para a utilização de modelos de Aprendizado de Máquina (AMATRIAIN et al., 2011). Este tratamento pode ser aplicado para mitigar o impacto da má qualidade dos dados, como dados aberrantes (*outliers*) e dados faltantes (*missing values*), assim como para adequar os dados a determinados modelos, evidenciar a correlação entre dados e viabilizar o processamento de grandes massas de dados.

Nas seções subsequentes, serão apresentadas as principais técnicas de pré-processamento de dados que serão adotadas neste trabalho.

3.3.1 Validação Cruzada *k-Fold*

A avaliação cruzada *k-fold* é uma técnica tradicional em Mineração de Dados e Aprendizado de Máquina, amplamente adotada também em Sistemas de Recomendação, utilizada para averiguar a capacidade preditiva e de generalização dos modelos (AMATRIAIN et al., 2011).

No *k-fold*, os dados a serem processados são divididos em k amostras mutualmente exclusivas de mesmo tamanho. Em seguida, k modelos são construídos, cada um utilizando uma das k amostras para validação e as demais para treinamento do modelo. Por fim, cada modelo m treinado é avaliado, utilizando sua respectiva amostra de validação, podendo-se utilizar qualquer métrica de avaliação ϵ, tais como as métricas apresentadas na Seção 3.1.8. Portanto, a avaliação final $\mathcal{E}$ é definida pela média da avaliação individual dos k modelos, conforme definido na Equação (22).

$$\mathcal{E} = \frac{1}{k} \sum_{i=1}^{k} \epsilon(m_i) \tag{22}$$

3.3.2 Transformação de Dados

3.3.2.1 Transformação One-Hot Encoding

O *one-hot encoding* é uma técnica de transformação de variáveis popular em aprendizado de máquina, também referenciada como *dummy variable* em estatística, onde cada valor de uma variável categórica é transformado em uma nova variável binária (HARRIS; HARRIS, 2015). Desta forma, esta técnica permite que cada valor categórico seja representado por uma dimensão única e independente. Por outro lado, o *one-hot encoding* tem a desvantagem de aumentar o volume de dados a ser processado pelo modelo, justamente pelo fato de produzir novas variáveis.

Tabela 2 – Exemplo da transformação *one-hot encoding* aplicado a uma variável categórica que representa o tipo sanguíneo, sendo (a) a variável original e (b) o conjunto de variáveis resultantes.

Tipo Sanguíneo
O
A
B
B
A
AB

(a)

$\xrightarrow{\textit{one-hot encoding}}$

Tipo Sanguíneo			
O	**A**	**B**	**AB**
1	0	0	0
0	1	0	0
0	0	1	0
0	0	1	0
0	1	0	0
0	0	0	1

(b)

A Tabela 2 exemplifica esta transformação aplicada à uma variável categórica que armazena o tipo sanguíneo de um grupo de indivíduos, sendo que a Tabela 2.a representa a variável categórica original e a Tabela 2.b representa o grupo de variáveis resultantes da aplicação da transformação *one-hot encoding*.

3.3.2.2 Normalização Min-Max

Alguns modelos de Aprendizado de Máquina exigem que as variáveis dos dados a serem analisados estejam dentro da mesma escala numérica. Nestes casos, aplica-se a normalização min-max, definida pela Equação (23), para transformar uma distribuição D com uma escala variada para uma nova distribuição D' respeitando um intervalo entre um valor mínimo a e um valor máximo b pré definidos (PATRO; SAHU, 2015). Frequentemente, os valores 0 e 1 são utilizados como mínimo e máximo.

$$D' = \left\{ (b-a)\frac{x - \min(D)}{\max(D) - \min(D)} + a;\ \forall\, x \in D \right\} \tag{23}$$

3.3.2.3 Normalização Logarítmica

A normalização logarítmica é utilizada para diminuir o impacto de valores aberrantes sobre a análise de dados. Além disso, esta transformação pode ser aplicada em distribuições de dados enviesadas, ou seja, distribuições que possuem um histograma que apresenta grande concentração de observações em uma das suas extremidades e pouca concentração de dados na outra. A Equação (24) descreve a transformação logarítmica aplicada a uma distribuição D, resultando em uma nova distribuição D'. Sendo que a base do logaritmo b pode ser ajustada conforme as características dos dados a serem analisados para melhor ajustá-los ao problema em questão.

$$D' = \{\log_b(x);\ \forall\, x \in D\} \tag{24}$$

3.4 BASE DE DADOS

Um fator impactante que tem incentivado pesquisas em SRSs são os Históricos Médicos Eletrônicos (HMEs) (ZHANG et al., 2017; YANG et al., 2017; HORS-FRAILE et al., 2017; WIESNER; PFEIFER, 2014; LÓPEZ-NORES et al., 2012; WIESNER; PFEIFER, 2010; NACHAWATI et al., 2012; KONDYLAKIS et al., 2014, 2015; BISSOYI; MISHRA; PATRA, 2016; BATEJA; DUBEY; BHATT, 2018). Através desta tecnologia, prontuários médicos são padronizados e armazenados digitalmente, criando grandes massas de dados de pacientes. É impraticável que um médico analise manualmente as informações históricas de milhares (ou até mesmo milhões) de pacientes, contidas nos HMEs, sempre que tiver que tomar uma decisão. Por outro lado, introduzindo o contexto popular de *big data*, sistemas computacionais são capazes de consumir e processar milhares de registros médicos em segundos, contribuindo para o desenvolvimento de modelos para suporte à tomada de decisão clínica mais robustos, capazes de sugerir alternativas mais assertivas aos médicos.

Sendo assim, para o desenvolvimento deste trabalho, adotou-se uma base de dados médica construída com base em HMEs, fornecida por um hospital público de São Paulo. Esta base é composta por registros médicos de pacientes que apresentam cardiopatias congênitas e que foram submetidos à cirurgia, contendo informações desde a admissão do paciente até a sua saída do hospital.

Os dados desta base podem ser separados em três etapas do tratamento do paciente, sendo: pré-operatório, intra-operatório e pós-operatório; onde as variáveis de cada uma das etapas e os seus respectivos tipos serão listadas a seguir:

a) Pré-operatório
   - Binário:
     - Diagnóstico Pré-Natal (Se o paciente foi diagnosticado antes do seu nascimento);
     - Prematuridade;
     - Mãe Diabética;
     - Edema Pré-Operatório;
   - Categórico:
     - Sexo;
     - Anomalias (Cada paciente pode conter até cinco. Ex.: *Down Syndrome*.);
     - Grupo de Diagnóstico (Categoria de diagnóstico que engloba outros diagnósticos mais específicos. Ex.: *Single Ventricle*);
     - Diagnóstico (Especificação do diagnóstico. Cada paciente pode conter até dez. Ex.: *Aortic Arch Hypoplasia*);
     - Procedimentos Prévios (Procedimentos cirúrgicos aos quais o paciente foi submetido previamente. Cada paciente pode conter até dez.);

    - Fatores de Risco Pré-Operatórios (Cada paciente pode conter até vinte. Ex.: *Cardio-Pulmonary Resuscitation*.);
  - Data/Hora:
    - Data de Nascimento do Paciente;
    - Data de Admissão do Paciente no Hospital;
  - Numérico Contínuo:
    - Peso (em quilogramas);
    - Altura (em centímetros);
    - Superfície Corpórea;
    - Saturação de Oxigênio ($SatO_2$) Pré-Operatória;
  - Numérico Inteiro:
    - Número de Internações Prévias;
    - Número de Cirurgias Prévias;
    - Número de Admissões Prévias na UTI;
    - Hematócrito Pré-Operatório;

b) Intra-Operatório
  - Binário:
    - Mortalidade Operatória;
    - Ecocardiograma Intra-Operatório;
    - Complicação Intra-Operatória - Sangramento;
    - Complicação Intra-Operatória - Arritmia;
    - Complicação Intra-Operatória - Circulação Extracorpórea (CEC);
    - Complicação Intra-Operatória - Membrana Extracorpórea (ECMO);
  - Categórico:
    - Procedimentos (Procedimentos cirúrgicos aos quais o paciente foi submetido. Cada paciente pode conter até dez. Ex.: "*AVC (AVSD) repair, Complete (CAVSD)*".);
  - Data/Hora:
    - Data de Início da Cirurgia;
    - Data de Término da Cirurgia;
    - Data da Morte;
  - Numérico Contínuo:
    - *Aristotle* Procedimentos (Estratificação de risco correspondente a cada um dos procedimentos executados);
    - Maior *Aristotle* (Maior Estratificação de risco dentre os procedimentos executados);
  - Numérico Inteiro:

    - Rachs Procedimentos (Estratificação de risco correspondente a cada um dos procedimentos executados);
    - Maior Rachs (Maior Estratificação de risco dentre os procedimentos executados);

c) Pós-Operatório
  - Binário:
    - Re-Intubação;
    - Traqueostomia;
    - Drogas na Saída do Centro Cirúrgico;
    - Drogas nas Primeiras 12 Horas Após;
    - Drogas nas Primeiras 12 Horas Após Cirurgia;
    - Sepse;
    - Sepse - Identificação de Micro-Organismo;
    - Infecção Local;
    - Pneumonia Devido à Ventilação Mecânica;
  - Categórico:
    - Tipo de Infecção Local;
    - Complicações (Cada paciente pode conter até dez. Ex.: *Wound infection*.);
    - Eventos (Cada paciente pode conter até dez. Ex.: *Stridor / Sub-glottic Stenosis*.);
    - Desfecho da UTI (Ex.: Alta para Enfermaria, Óbito);
    - Desfecho Alta (Ex.: Alta para Casa);
  - Data/Hora:
    - Data de Entrada na UTI;
    - Data de Extubação;
    - Data de Desfecho da UTI;
    - Data de Alta;
  - Numérico Contínuo:
    - Lactato Pós-Operatório;
    - Droga Dopamina Saída;
    - Droga Dobutamina Saída;
    - Droga Adrenalina Saída;
    - Droga Noradrenalina Saída;
    - Droga Milrinona Saída;
    - Droga Levosimendana Saída;
    - Score de Droga Saída;
    - Droga Dopamina 12 Horas Pós;

    - Droga Dobutamina 12 Horas Pós;
    - Droga Adrenalina 12 Horas Pós;
    - Droga Noradrenalina 12 Horas Pós;
    - Droga Milrinona 12 Horas Pós;
    - Droga Levosimendana 12 Horas Pós;
    - *Score* de Droga 12 Horas Pós;
  - Numérico Inteiro:
    - Hematócrito Pós-Operatório;
    - Quantidade de Re-Intubações;

Esta base de dados possuí um total de 155 atributos (sendo 61 pré-operatórios, 41 intra-operatórios e 53 pós-operatórios) e 2009 observações.

4 METODOLOGIA

Conforme apresentado no Capítulo 2, trabalhos recentes têm explorado modelos de SR aplicados na área médica, os chamados Sistemas de Recomendação para Saúde (SRSs). Em (WIESNER; PFEIFER, 2014), os SRSs são divididos em duas categorias: SRSs para profissionais da saúde; e SRSs para pacientes. Com base nesta classificação, o trabalho aqui proposto enquadra-se em SRSs para profissionais da saúde. Assim, pode-se listar ao menos três justificativas principais para este direcionamento do trabalho, que serão apresentadas na sequência.

Primeiramente, o desenvolvimento de pesquisas nesta área, em especial em países subdesenvolvidos e com defasagens na qualidade dos serviços de saúde, como o Brasil, tem o potencial de proporcionar benefícios aos sistemas saúde, como a diminuição de erros médicos, a melhoria na qualidade de vida dos pacientes, a otimização de alocação de recursos e a redução de gastos públicos, tendo assim, impactos diretos na sociedade.

Segundo, com base nas constatações exibidas na Seção 2.1, esta área ainda oferece muitas lacunas abertas à investigação, tais como a baixa quantidade de resultados experimentais formalmente publicados, a dificuldade de comparação entre trabalhos devido à alta divergência em relação à adoção de métricas de avaliação de sistemas e à dificuldade de acesso às bases de dados, que normalmente são privadas, a adoção dos chamados HMEs em SRSs e os desafios em relação à modelagem de perfil dos pacientes.

Terceiro, o desenvolvimento de SRSs para pacientes em um cenário de recomendação de tratamento pode não ser adequadamente abordado neste trabalho, em virtude das competências técnicas dos autores e da gravidade das consequências de recomendações indevidas em um cenário desta natureza.

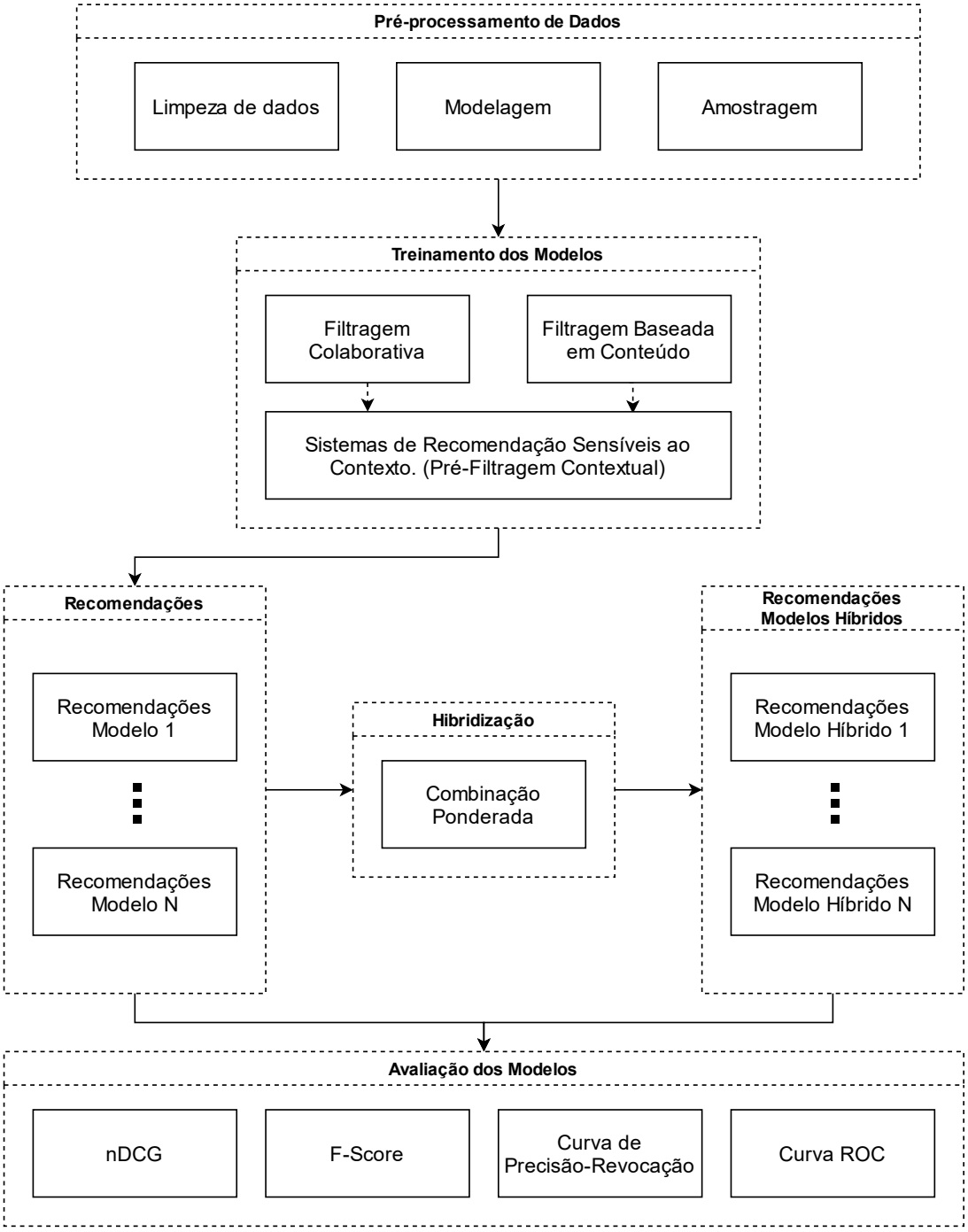
Sendo assim, o enfoque deste trabalho foi direcionado para a aplicação e avaliação de técnicas de SRs em um cenário de recomendação de tratamento. Este estudo foi realizado utilizando um cenário de recomendação de procedimentos de tratamento cirúrgicos para pacientes que apresentam cardiopatias congênitas, utilizando a base de dados descrita na Seção 3.4.

Este trabalho adotou como base uma metodologia tradicional para modelos de aprendizado de máquina, dividida em 4 etapas: Pré-processamento de dados; Treinamento dos modelos; Criação de recomendações; e Avaliação dos modelos. Ademais, esta metodologia pode ser estendida com uma nova etapa, entre as etapas de criação de recomendações e avaliação dos modelos, conhecida como *ensemble*, na área de aprendizado de máquina, e como hibridização, na área de SRs. A Figura 6 mostra um diagrama que representa esta metodologia em quatro etapas, que serão descritas na sequência.

4.1 PRÉ-PROCESSAMENTO DE DADOS

A falta de qualidade dos dados, conforme discutido na Seção 3.3, é uma causa frequente de problemas em modelos de aprendizado de máquina, devido a fenômenos clássicos como

Figura 6 – Ilustra as etapas da metodologia desenvolvida.



Fonte: Autor.

ruído, valores aberrantes (*outliers*), e não-resposta (*missing values*). Isto posto, a etapa de pré-processamento de dados visa tratar estes problemas, melhorar a qualidade dos dados e adequá-los para serem consumidos nas etapas seguintes.

O primeiro passo do pré-processamento de dados é a análise exploratória das variáveis, com o intuito de entender a distribuição dos dados e identificar os problemas citados anteriormente, assim como as técnica mais adequadas para solucioná-los. Em seguida, o pré-processamento realizado neste trabalho pode ser dividido em três tarefas: a limpeza de dados, a modelagem dos dados e a amostragem, que serão descritas na sequência.

4.1.1 Limpeza de Dados

Através da analise inicial da base de dados adotada, descrita na Seção 3.4, 133 observações foram descartadas, passando de um total de 2009 observações para 1876 observações, seguindo os seguintes critérios:

a) 80 registros sem data de admissão ou com a data de admissão inválidas. Estes registros foram descartados por não ser possível identificar a idade do paciente;
b) 4 registros de pacientes que não continham nenhum diagnóstico ou procedimento cirúrgico efetuado;
c) 49 registros de pacientes que apresentaram um grupo de diagnóstico, diagnóstico ou procedimento cirúrgico que nenhum outro paciente apresentou. Estes registros foram descartados para garantir que todos os grupos de diagnóstico, diagnósticos e procedimentos sejam incluídos no mínimo em duas amostras diferentes da validação-cruzada k-Fold (Seção 3.3.1), fazendo com que sempre haja ao menos uma ocorrência destes dados no treinamento dos modelos. Este cuidado foi tomado para evitar problemas de *cold-start*, pois o tratamento deste tipo de problema está fora do escopo deste trabalho;

Analisando os valores não-resposta, apenas variáveis binárias apresentaram valores nulos. Sendo assim, todos os valores nulos foram considerados como 0.

Além disso, algumas variáveis foram criadas a partir das variáveis do tipo Data/Hora contidas na base, com o intuito de utilizá-las de maneira mais significativa no perfil dos pacientes. A lista de variáveis criadas foi a seguinte:

a) Idade do Paciente na Cirurgia (em anos) - Composta pelo intervalo entre a Data de Início da Cirurgia e a Data de Nascimento do Paciente;
b) Período Pré-Operatório (em dias) - Composto pelo intervalo entre a Data de Início da Cirurgia e a Data de Admissão do Paciente no Hospital;

4.1.2 Modelagem dos Dados

Apesar de muitas soluções e caminhos terem sido propostos na área de SRSs para profissionais da saúde, conforme apresentado na Seção 2.1, muitos desafios ainda estão abertos à

investigação. Um dos principais desafios está relacionado com a modelagem dos dados médicos e a construção do perfil dos pacientes. É comum que a aplicação de algoritmos de Aprendizado de Máquina dependa de uma estruturação específica dos dados de entrada. No caso de SRs, esta adequação é justamente a construção destes perfis, permitindo o cálculo de similaridade entre pacientes e/ou entre procedimentos de tratamento, utilizados para discriminar os itens a serem recomendados.

Conforme abordado na Seção 3.4, a base de dados adotada contém dados sobre o processo de tratamento dos pacientes dividido em 3 etapas, que são: pré-operatório, intra-operatório e pós-operatório. Como o objetivo deste trabalho é avaliar as técnicas de SRs na recomendação de procedimentos cirúrgicos (contidos dentre os dados intra-operatórios), o perfil dos pacientes foi construído com base nos dados pré-operatórios dos mesmos.

Além disso, para que a comparação entre os perfis seja adequada, a modelagem dos dados foi realizada através da aplicação de algumas transformações de variáveis.

Para as variáveis qualitativas (ou categóricas), foi aplicada a técnica de transformação *one-hot encoding*, descrita na Seção 3.3.2.1. Esta técnica foi utilizada para gerar uma variável binária para cada valor de variável categórica existente na base de dados, de modo a facilitar o processamento para calcular a similaridade entre perfis e discriminar os procedimentos a serem recomendados. Como resultante, a base de dados totalizou 318 variáveis pré-operatórias, que compõem o perfil dos pacientes, e 113 procedimentos cirúrgicos. A lista de variáveis transformadas e a quantidade de variáveis resultantes foi a seguinte:

a) Sexo - 2 variáveis;
b) Grupo Diagnóstico - 29 variáveis;
c) Diagnóstico - 102 variáveis;
d) Anomalias - 31 variáveis;
e) Fatores de Risco Pré-Operatórios - 29 variáveis;
f) Procedimentos Prévios - 110 variáveis;
g) Procedimentos Cirúrgicos - 113 variáveis;

As variáveis quantitativas precisam de um tratamento específico conforme a sua utilização no modelo. Quando utilizadas para o cálculo de similaridade, uma característica que deve ser observada em variáveis numéricas é a dispersão e a relação de diferença entre os valores. Por exemplo, supondo que dois pacientes possuem uma diferença de uma hora no tempo de permanência em UTI. Se um paciente P_1 permaneceu na UTI por cinco minutos e outro paciente P_2 por uma hora e cinco minutos (65 minutos), essa diferença de uma hora pode ser considerada grande. Em contrapartida, se um paciente P_3 permanece na UTI por vinte dias (28800 minutos) e outro paciente P_4 permanece na UTI por vinte dias e uma hora (28860 minutos), essa mesma diferença de uma hora já não se mostra tão relevante. Para equalizar este tipo de comparação, tradicionalmente adota-se uma normalização logarítmica, apresentada na Seção 3.3.2.3. No exemplo citado acima, se aplicarmos a função logarítmica com base 2 no período de permanência na UTI em minutos dos pacientes citados, a diferença entre os valores dos pacientes P_1 e P_2

aproximadamente 3,7 e entre os pacientes P_3 e P_4 seria aproximadamente 0,003, conforme resolução apresentada nas Equações (25) e (26). Desta forma, imprimindo numericamente a maior relevância na diferença de uma hora do primeiro caso em comparação com a mesma diferença de uma hora do segundo caso.

$$|\log_2(65) - \log_2(5)| \cong |6{,}022 - 2{,}322| = 3{,}7 \tag{25}$$

$$|\log_2(28860) - \log_2(28800)| \cong |14{,}817 - 14{,}814| = 0{,}003 \tag{26}$$

A normalização logarítmica também ajuda a tratar valores aberrantes, pois ela diminuí a dispersão da distribuição, aproximando os valores. A lista de variáveis submetidas à normalização logarítmica foi a seguinte:

a) $SatO_2$ Pré-Operatória;
b) Hematócrito Pré-Operatório;
c) Superfície Corpórea;
d) Altura (em centímetros);
e) Peso (em quilogramas);
f) Número de Cirurgias Prévias;
g) Número de Internações Prévias;
h) Idade do Paciente na Cirurgia (em anos);
i) Período Pré-Operatório (em dias);

Outra característica impactante de variáveis numéricas está relacionada com a escala de valores. Quando utilizadas para o cálculo de similaridade entre perfis, dependendo da métrica de similaridade utilizada, estas variáveis devem ser transformadas para uma mesma escala numérica, para equalizar os pesos das variáveis nesta situação. Esta normalização é conhecida como min-max, descrita na Seção 3.3.2.2. Por exemplo, supondo que a similaridade entre pacientes seja calculada com base na altura em metros e no peso em quilos dos mesmos, o fato destas variáveis possuírem escalas muito distintas tem por consequência a variável de maior escala ter mais impacto no cálculo da similaridade, quando adotadas métricas geométricas, como a distância euclidiana. Por outro lado, se considerarmos a similaridade dos cossenos como métrica, a magnitude dos vetores não impacta no resultado do cálculo de similaridade. Portanto, a normalização min-max foi aplicada para equalizar todas as variáveis para a escala entre 0 e 1.

4.1.3 Amostragem de Dados

A etapa de amostragem consiste na separação dos dados em amostras de treinamento e teste, que serão consumidas pelas etapas subsequentes da metodologia. Neste trabalho, a base de dados foi separada em 5 amostras mutuamente exclusivas de tamanhos aproximados para que, posteriormente, seja aplicada a validação cruzada *k-fold*, apresentada na Seção 3.3.1, onde os dados são divididos em cinco variações de conjuntos de treinamento e de teste. As observações são distribuídas de forma aleatória entre as amostras. O único cuidado que foi tomado

na construção das amostras foi em garantir que cada um dos Grupos de Diagnóstico, Diagnóstico e Procedimentos Cirúrgicos existentes na base fossem incluídos em ao menos 2 amostras diferentes, com o intuíto de evitar problemas de *Cold-Start*.

O objetivo da adoção deste método é avaliar a capacidade de generalização do modelo em relação aos dados utilizados.

4.2 TREINAMENTO DOS MODELOS

A etapa de treinamento consiste no processo de aprendizado do modelo, onde padrões do conjunto de dados conhecidos são extraídos para que, posteriormente, as recomendações sejam efetuadas de forma preditiva para novos dados. Nesta etapa, os modelos de SRs são treinados utilizando as amostras de treinamento, construídas na etapa de amostragem, descrita na Seção 4.1.3.

Conforme apresentado na revisão bibliográfica, apresentada no Capítulo 2, os modelos de Filtragem Colaborativa (HASSAN; SYED, 2010; DAVIS et al., 2008; HUSSEIN et al., 2014; THONG et al., 2015b,a; KOMKHAO; LU; ZHANG, 2012; KOMKHAO; HALANG, 2013; DUAN; STREET; XU, 2011; GRÄSSER et al., 2017; HORS-FRAILE et al., 2016) e Filtragem Baseada em Conteúdo (HORS-FRAILE et al., 2016; LI; JONES, 2015; ZHANG et al., 2017; NACHAWATI et al., 2012), especificados respectivamente nas Seções 3.1.2 e 3.1.3, figuram dentre as principais técnicas de SRs tradicionais utilizadas em trabalhos de SRSs. Sendo que, Filtragem Colaborativa e Filtragem Baseada em Conteúdo são adotadas, respectivamente, por 53%, 12% dos trabalhos na área de SRSs para profissionais da saúde considerados neste trabalho. Como também pode ser observado na Seção 2.1, raramente há intersecção destes modelos em um mesmo artigo.

Além disso, um dos modelos mais modernos que compõe o estado da arte em SRs tradicionais, os chamados Sistemas de Recomendação Sensíveis ao Contexto, abordado na Seção 3.1.5, ainda não foram amplamente explorados em SRSs, visto que, dentre os trabalhos aqui citados, este modelo é adotado apenas em (MONTEIRO et al., 2015).

Portanto, considerando que os mesmos compõem a lista das principais técnicas de sistemas de recomendação tradicionais atualmente (RICCI; ROKACH; SHAPIRA, 2015), serão adotados neste trabalho os modelos de Filtragem Colaborativa, Filtragem Baseada em Conteúdo e Sistemas de Recomendação Sensíveis ao Contexto, para serem avaliados e comparados entre si em um cenário de SRS.

Os modelos de recomendação de procedimentos cirúrgicos foram construídos com base na combinação de 3 componentes: Perfil do Paciente; Perfil do Procedimento Cirúrgico; e Perfil do Contexto. Cada um desses componentes é um vetor N-dimensional composto por um conjunto específico das variáveis pré-operatórias, pré-processadas na etapa anterior, Seção 4.1. As variações de construção destes componentes, assim como a forma como os mesmos são utilizados para gerar as recomendações serão descritas nas seções seguintes.

4.2.1 Perfil de Contexto

O Perfil de Contexto é o conjunto de variáveis pré-operatórias que são considerados para determinar o contexto de recomendação no qual o paciente está inserido, utilizado nos Sistemas de Recomendação Sensíveis ao Contexto. Foram propostos 6 variações de Perfis de Contexto do Paciente, que são:

a) Contexto de Grupo de Diagnóstico - Composto pelas 29 variáveis de Grupo Diagnóstico;
b) Contexto de Anomalias - Composto pelas 31 variáveis de Anomalias;
c) Contexto de Fatores de Risco - Composto pelas 29 variáveis de Fatores de Risco Pré-Operatórios;
d) Contexto de Procedimentos Prévios - Composto pelas 110 variáveis de Procedimentos Prévios;
e) Contexto Físico - Composto pelas demais variáveis pré-operatórias que não foram utilizadas nos contextos previamente descritos, com exceção das variáveis de Diagnóstico. As 16 variáveis incluídas neste contexto são:
   - Diagnóstico Pré-Natal;
   - Prematuridade;
   - Mãe Diabética;
   - Edema Pré-Operatório;
   - Sexo (2 variáveis);
   - Peso;
   - Altura;
   - Superfície Corpórea;
   - $SatO_2$ Pré-Operatória;
   - Número de Internações Prévias;
   - Número de Cirurgias Prévias;
   - Número de Admissões Prévias na UTI;
   - Hematócrito Pré-Operatório;
   - Idade do Paciente na Cirurgia;
   - Período Pré-Operatório;
f) Contexto Completo - Composto por todas as 215 variáveis pré-operatórias consideradas nos contextos descritos previamente.

Conforme abordado na Seção 3.1.5, os Sistemas de Recomendação Sensíveis ao Contexto são utilizados para complementar outros modelos de SRs. Tendo isso em vista, as variáveis de Diagnóstico não foram utilizadas em nenhum contexto pelo fato de serem consideradas fundamentais para a recomendação de procedimentos de tratamento, pois devem ser adotadas pelo SR mesmo quando o sistema não seja complementado com contexto.

Dentre os perfis de contexto citados acima, apenas o Perfil de Contexto Completo é utilizado para comparação com os modelos que não utilizam contexto. Os modelos que utilizam os demais Perfis de Contexto foram utilizados para avaliar e comparar o impacto individual que cada contexto exerce na recomendação.

4.2.2 Perfil de Paciente

Similarmente ao Perfil de Contexto, o Perfil de Paciente é composto por um conjunto de variáveis pré-operatórias utilizadas para o cálculo de similaridade entre os pacientes. Foram utilizadas 2 variações de Perfil do Paciente, que são:

a) Perfil Diagnóstico - Composto pelas 102 variáveis de Diagnóstico.
b) Perfil Completo - Composto por todas as variáveis pré-operatórias, excluindo as variáveis consideradas no Perfil de Contexto, quando aplicável. Por exemplo, num SR que não considera contexto, o Perfil de Paciente Completo é composto por todas as 318 variáveis pré-operatórias. Em contrapartida, considerando um Sistema de Recomendação Sensível ao Contexto que utiliza o Perfil de Contexto de Grupo de Diagnóstico, as 29 variáveis do Perfil de Contexto são desconsideradas no Perfil de Paciente Completo, totalizando 289 variáveis neste cenário.

4.2.3 Pré-Filtragem Contextual

O paradigma de integração de contexto em SRs adotado é a Pré-Filtragem Contextual, descrito na Seção 3.1.5. Neste tipo de abordagem, primeiramente utiliza-se as informações contextuais para filtrar a base de dados. Na sequência, os dados filtrados são consumidos por um modelo de SR tradicional que produz as recomendações.

Neste trabalho, a filtragem é realizada através do cálculo de similaridade entre o Perfil de Contexto do paciente para o qual a recomendação será produzida, referenciado nesse trabalho como **paciente alvo**, e os Perfis de Contexto dos pacientes que compõem o conjunto de treinamento do Sistema de Recomendação, esta similaridade será referenciada como **similaridade contextual**. Desta forma, pacientes com baixa similaridade contextual sofrem uma inibição do seu impacto no processo de geração de recomendações.

4.2.4 Perfil de Procedimento Cirúrgico

Conforme discutido na Seção 3.1.3, os modelos de Filtragem Baseada em Conteúdo necessitam tanto do Perfil do Usuário quanto do Perfil do Item para produzir as recomendação. Portanto, o Perfil de Procedimento Cirúrgico deve ser construído para que este tipo de modelo seja utilizado.

O Perfil de Procedimento Cirúrgico é composto pelas mesmas variáveis presentes no Perfil do Paciente, a diferença está na forma como o perfil é construído. Neste trabalho, o Perfil de um determinado Procedimento Cirúrgico é composto pela média do Perfil dos Pacientes (pertencentes ao conjunto de treinamento) que foram submetidos ao procedimento em questão, definido pela Equação (27), onde PPC_{proc} representa o Perfil do Procedimento Cirúrgico *proc* e P representa o conjunto de Perfis de Pacientes que foram submetidos ao procedimento *proc*, ressaltando que cada Perfil de Paciente é um vetor N-dimensional.

$$PPC_{proc} = \frac{\sum_{p \in P} \vec{p}}{|P|} \tag{27}$$

No caso da integração da Pré Filtragem Contextual, apresentada na Seção 4.2.3, há uma pequena diferença na construção dos Perfis de Procedimentos Cirúrgicos. Neste caso, primeiro calcula-se a similaridade entre o Perfil de Contexto do paciente alvo e o Perfil de Contexto dos Pacientes presentes no conjunto de treinamento do modelo. Em seguida, o Perfil de um determinado Procedimento Cirúrgico é composto pela média ponderada do Perfil dos Pacientes (pertencentes ao conjunto de treinamento) que foram submetidos ao procedimento pela similaridade contextual. Este cálculo é definido pela Equação (28), onde PPC_proc representa o Perfil do Procedimento Cirúrgico *proc*, P representa o conjunto de Perfis de Pacientes que foram submetidos ao procedimento *proc* e $sim(\vec{u_c}, \vec{p_c})$ representa a similaridade entre o Perfil do Contexto do paciente u, para o qual a recomendação será gerada, e o paciente p pertencente ao conjunto P.

$$PPC_{proc} = \frac{\sum_{p \in P} sim(\vec{u_c}, \vec{p_c}) * \vec{p}}{\sum_{p \in P} sim(\vec{u_c}, \vec{p_c})} \tag{28}$$

Assim como para os Perfis de Pacientes, os Perfis de Procedimentos Cirúrgicos também utilizam as duas variações de composição de perfil, o Perfil Diagnóstico e o Perfil Completo, que são construídos exatamente da mesma maneira, conforme descrito na Seção 4.2.2.

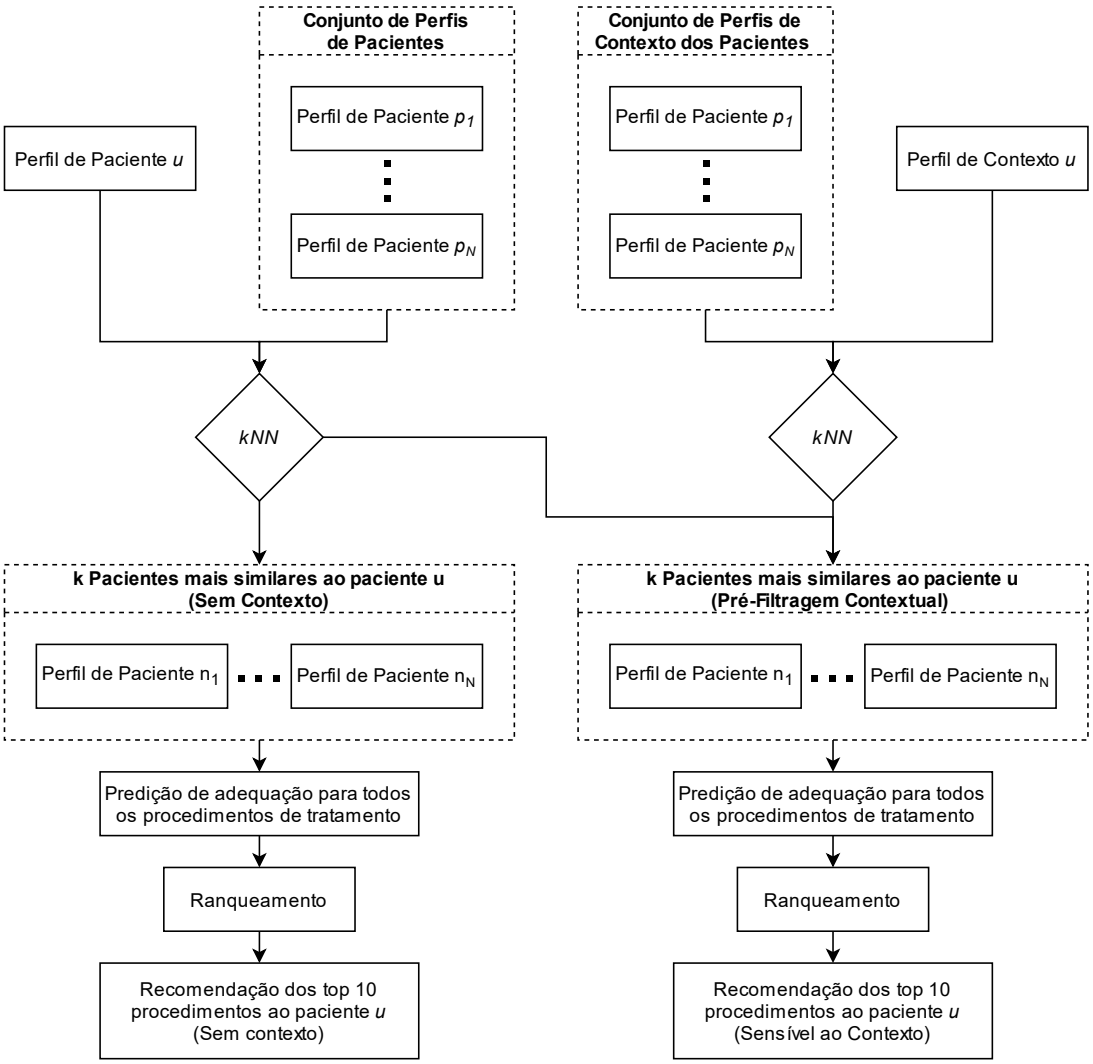
4.2.5 Filtros Colaborativos

Um modelo de Filtragem Colaborativa, apresentado na Seção 3.1.2, aplicado no cenário de recomendação adotado, utiliza a similaridade entre os Perfis de Pacientes para produzir as recomendações, onde o conjunto de pacientes mais similares ao paciente alvo são considerados, sendo que este conjunto dos k pacientes mais similares é calculado através do algoritmo kNN, apresentado na Seção 3.2, utilizando como base uma métrica de similaridade pré-definida.

Seguindo a Equação (1), apresentada na Seção 3.1.2, para cada Procedimento Cirúrgico i na base de dados, é calculada a predição da adequação do procedimento i para o paciente u ($p(u, i)$), sendo que N representa o conjunto de pacientes (pertencentes ao conjunto de treinamento) mais similares ao paciente u e $a(n, i)$ representa a variável binária que indica se o paciente $n \in N$ foi ou não submetido ao procedimento i.

Integrando Pré-Filtragem Contextual, definida na Seção 4.2.3, aos Filtros Colaborativos, deve-se também calcular a similaridade entre o paciente alvo e os pacientes do conjunto

Figura 7 – Diagrama que ilustra o cálculo das recomendações através da Filtragem Colaborativa, com e sem a adoção de dados contextuais, dentro da metodologia proposta.



Fonte: Autor.

de treinamento em relação ao Perfil de Contexto. Desta forma, a similaridade entre pacientes ($sim(u, n)$) é definida pelo produto da similaridade entre os Perfis de Paciente ($sim_p(u, n)$) pela similaridade entre os Perfis de Contexto ($sim_c(u, n)$), definida pela Equação (29). Com essa alteração da similaridade, a predição de adequação de procedimento segue sendo realizada através da Equação (1)

$$sim(u, n) = sim_p(u, n) * sim_c(u, n) \tag{29}$$

Por fim, os procedimentos são recomendados na ordem decrescente em relação à predição de adequação, ou seja, os procedimentos com maior valor de $p(u, i)$ são recomendados primeiro. Ressaltando-se que apenas os procedimentos que apresentam valores de $p(u, i) > 0$ são recomendados e os procedimentos com valor de $p(u, i) = 0$ são descartados.

Neste cenário, foi definido um limite de dez recomendações de procedimentos para cada paciente, por duas razões principais:

a) A estrutura da base de dados limita a quantidade de procedimentos de tratamento aplicados a cada paciente em dez.
b) Na grande maioria dos casos, os pacientes apresentam até três procedimentos adotados em seus tratamentos.

A Figura 7 apresenta um diagrama de como os modelos de Filtragem Colaborativa foram construídos.

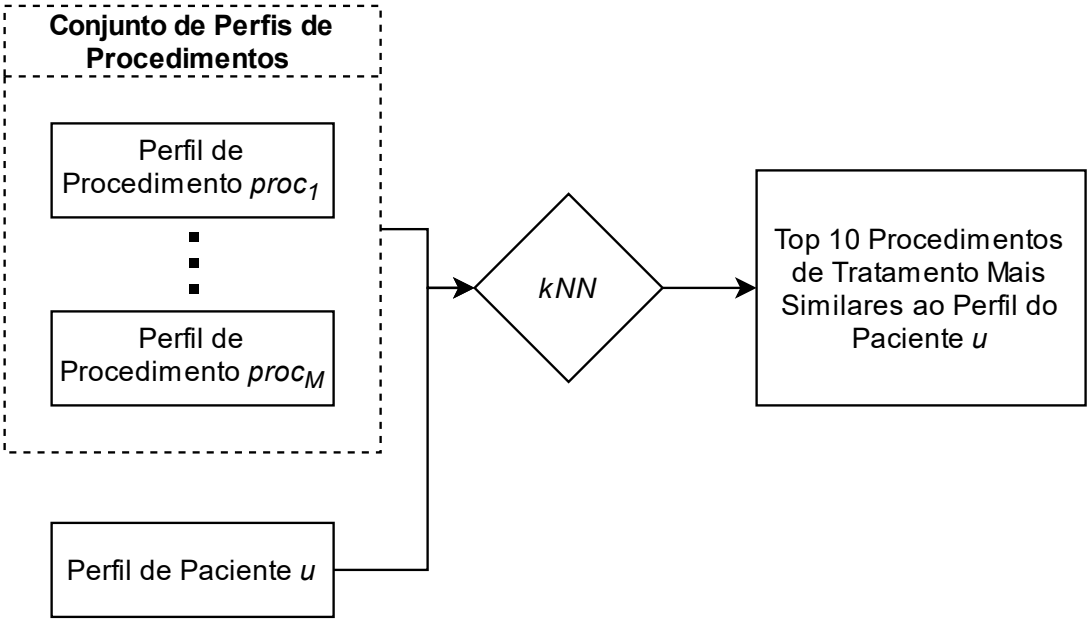
No total, combinando as variações de Perfil de Paciente e Perfil de Contexto, foram construídos 13 modelos de Filtros Colaborativos, que são:

a) Filtro Colaborativo com Perfil de Paciente Diagnóstico sem Contexto (FC-PD);
b) Filtro Colaborativo com Perfil de Paciente Diagnóstico e Contexto de Grupo Diagnóstico (FC-PD-CGD);
c) Filtro Colaborativo com Perfil de Paciente Diagnóstico e Contexto de Anomalias (FC-PD-CA);
d) Filtro Colaborativo com Perfil de Paciente Diagnóstico e Contexto de Fatores de Risco (FC-PD-CFR);
e) Filtro Colaborativo com Perfil de Paciente Diagnóstico e Contexto de Procedimentos Prévios (FC-PD-CPP);
f) Filtro Colaborativo com Perfil de Paciente Diagnóstico e Contexto Físico (FC-PD-CF);
g) Filtro Colaborativo com Perfil de Paciente Completo e sem Contexto (FC-PC);
h) Filtro Colaborativo com Perfil de Paciente Completo e Contexto de Grupo Diagnóstico (FC-PC-CGD);
i) Filtro Colaborativo com Perfil de Paciente Completo e Contexto de Anomalias (FC-PC-CA);
j) Filtro Colaborativo com Perfil de Paciente Completo e Contexto de Fatores de Risco (FC-PC-CFR);
k) Filtro Colaborativo com Perfil de Paciente Completo e Contexto de Procedimentos Prévios (FC-PC-CPP);
l) Filtro Colaborativo com Perfil de Paciente Completo e Contextos Físico (FC-PC-CF);
m) Filtro Colaborativo com Perfil de Contexto Completo (FC-CC);

4.2.6 Filtros Baseados em Conteúdo

Nos modelos de Filtragem Baseada em Conteúdo, descritos na Seção 3.1.3, aplicados no cenário de recomendação de procedimentos cirúrgicos adotado neste trabalho, a lista de recomendações é construída diretamente através do cálculo de similaridade entre o Perfil do Paciente

Figura 8 – Diagrama que ilustra o cálculo das recomendações através da Filtragem Baseada em Conteúdo dentro da metodologia proposta.



Fonte: Autor.

Alvo e o Perfil dos Procedimentos Cirúrgicos disponíveis na base de dados. Ou seja, para os modelos de FBC, adota-se que $sim(u, i) = p(u, i)$.

Isto posto, o modelo recomenda o conjunto dos k procedimentos mais similares ao paciente alvo, determinados também pela aplicação do algoritmo kNN, através de uma métrica de similaridade pré determinada. Assim como os modelos de FC, os procedimentos com valor de $p(u, i) = 0$ são desconsiderados da lista de recomendação.

Seguindo o mesmo princípio dos Filtros Colaborativos, foram construídos 13 variações de modelos de Filtros Baseados em Conteúdo, que são:

a) Filtro Baseado em Conteúdo com Perfil de Paciente Diagnóstico sem Contexto (FBC-PD);
b) Filtro Baseado em Conteúdo com Perfil de Paciente Diagnóstico e Contexto de Grupo Diagnóstico (FBC-PD-CGD);
c) Filtro Baseado em Conteúdo com Perfil de Paciente Diagnóstico e Contexto de Anomalias (FBC-PD-CA);
d) Filtro Baseado em Conteúdo com Perfil de Paciente Diagnóstico e Contexto de Fatores de Risco (FBC-PD-CFR);
e) Filtro Baseado em Conteúdo com Perfil de Paciente Diagnóstico e Contexto de Procedimentos Prévios (FBC-PD-CPP);
f) Filtro Baseado em Conteúdo com Perfil de Paciente Diagnóstico e Contexto Físico (FBC-PD-CF);
g) Filtro Baseado em Conteúdo com Perfil de Paciente Completo e sem Contexto (FBC-PC);

h) Filtro Baseado em Conteúdo com Perfil de Paciente Completo e Contexto de Grupo Diagnóstico (FBC-PC-CGD);
i) Filtro Baseado em Conteúdo com Perfil de Paciente Completo e Contexto de Anomalias (FBC-PC-CA);
j) Filtro Baseado em Conteúdo com Perfil de Paciente Completo e Contexto de Fatores de Risco (FBC-PC-CFR);
k) Filtro Baseado em Conteúdo com Perfil de Paciente Completo e Contexto de Procedimentos Prévios (FBC-PC-CPP);
l) Filtro Baseado em Conteúdo com Perfil de Paciente Completo e Contexto Físico (FBC-PC-CF);
m) Filtro Baseado em Conteúdo com Perfil de Contexto Completo (FBC-CC);

A Figura 8 apresenta um diagrama de como os modelos de Filtragem Baseada em Conteúdo foram construídos.

4.3 HIBRIDIZAÇÃO

Por fim, os modelos de SRs descritos previamente foram combinados em um Sistema de Recomendação Híbrido, através do método de combinação ponderada, apresentado na Seção 3.1.6. Esta técnica visa criar um modelo mais assertivo através da combinação de dois ou mais modelos. Particularmente, o método de *ensemble* por combinação linear permite analisar a relevância de cada um dos modelos individualmente na composição do sistema híbrido final, com base nos pesos atribuídos à cada modelo.

Considerando a Equação (1), a predição de adequação para o SR Híbrido por combinação ponderada ($p_{cp}(u, i)$) é definido pela Equação (30); onde M representa o conjunto de modelos que compõem o sistema híbrido, α_m representa o peso calculado para o modelo $m \in M$ e $p_m(u, i)$ representa a predição de adequação do item i ao usuário u resultante do modelo m.

$$p_{cp}(u,i) = \frac{\sum_{m \in M} \alpha_m * p_m(u, i)}{\sum_{m \in M} \alpha_m} \tag{30}$$

O Sistema de Recomendação Híbrido proposto é composto por três variações de modelos de Filtragens Colaborativas e três variações de modelos de Filtragens Baseadas em Conteúdo, apresentados nas Seções 4.2.5 e 4.2.6, que são:

a) FC-PD
b) FC-PC
c) FC-CC
d) FBC-PD
e) FBC-PC
f) FBC-CC

Por fim, o Algoritmo 1 descreve o método utilizado para otimizar os pesos de cada um dos modelos que compõem os Sistemas de Recomendação Híbridos. Neste método, os pesos de

Algoritmo 1 – Descrição do cálculo de pesos para cada um dos modelos que compõem o Sistema de Recomendação Híbrido. A função *RMSE* representa a raíz do erro quadrado médio, descrito pela Equação (14), da predição de adequações $p_{cp}(u, i)$ em relação à adequação observada $a(u, i)$, para todos os pacientes u e procedimentos i contidos na base de dados. A constante λ é utilizada para limitar a porção do peso do modelo a ser transferida em cada iteração; neste trabalho, foi adotado o valor de $\lambda = 0.3$.

1 **Entrada:** Lista de Modelos M, constante λ
2 **Saída:** Lista de Pesos De Cada Modelo α
3 variável α = Lista de tamanho $|M|$;
4 **para cada** *variável* $m \in M$ **faça**
5 | $\alpha_m = 1/|M|$;
6 **fim**
7 variável $menorError = RMSE(M, \alpha)$;
8 variável $i = 0$;
9 **enquanto** $i < 10000$;
10 **faça**
11 | a = aleatório($a \in M$);
12 | b = aleatório($b \in M, b \neq a$);
13 | r = aleatório($0 \leq r \leq 1$);
14 | $\Delta = \alpha_a * \lambda * r$;
15 | $\alpha_a = \alpha_a - \Delta$;
16 | $\alpha_b = \alpha_b + \Delta$;
17 | variável $erro = RMSE(M, \alpha)$;
18 | **se** $erro < menorErro$ **então**
19 | | $menorErro = erro$;
20 | **fim**
21 | **senão**
22 | | $\alpha_a = \alpha_a + \Delta$;
23 | | $\alpha_b = \alpha_b - \Delta$;
24 | **fim**
25 | $i = i + 1$;
26 **fim**
27 **retorna** $alpha$

todos os modelos são inicializados com pesos iguais. Em cada iteração, uma parcela do peso de um modelo é transferido para outro modelo, onde os modelos que vão ter os pesos modificados são selecionados aleatoriamente, assim como a parcela transferida que também é definida aleatoriamente, limitada pela constante λ, definida como 0.3 neste trabalho. Ao final de cada iteração, calcula-se o erro através da Raiz do Erro Quadrado Médio, definido pelo Equação (14), para definir se a transferência de peso entre os modelos resultou numa melhoria ao sistema híbrido; se essa melhoria for constatada, mantêm-se a nova composição de pesos entre os modelos, caso contrário, a transferência de pesos é desfeita.

4.4 AVALIAÇÃO DOS MODELOS

Por fim, a etapa de avaliação utiliza os modelos treinados na etapa anterior para realizar a recomendação de procedimentos para novos pacientes, desconhecidos pelos modelos. Estes novos dados são compostos pelas amostras de teste obtidas na etapa de amostragem, conforme definido na Seção 4.1.

Os modelos de recomendação foram avaliados em duas vertentes, apresentadas na Seção 3.1.8, que são: acurácia preditiva e cobertura. Sendo que duas categorias de acurácia preditiva serão consideradas, sendo: predição de adoção do usuário e ranqueamento de itens, discutido na Seção 3.1.8.1.

Pelo fato dos procedimentos aplicados aos pacientes que compõem as amostras de teste já serem conhecidos, porém não terem sido considerados na etapa de treinamento, é possível avaliar a acurácia preditiva dos modelos. Esta avaliação é realizada através da comparação entre as recomendações geradas e os procedimentos realmente realizados.

Apesar da divergência na adoção de métricas de avaliação de SRSs pela academia, ressaltada na Seção 2.1, avaliações oriundas da matriz de confusão, aparecem com frequência neste cenário. As métricas de avaliação de predição de adoção dos usuários em SRs tradicionais serão utilizadas neste trabalho para averiguar se os procedimentos recomendados pelos modelos são condizentes com os procedimentos de fato executados pelos profissionais de saúde. Portanto, as métricas provenientes da matriz de confusão, descritas na Seção 3.1.8.1.2, que foram adotadas são:

a) *F-Score*;
b) Curva ROC;
c) Curva Precisão-Revocação;

As métricas de ranqueamento de itens de SRs tradicionais são utilizados para avaliar se a priorização das recomendações realizadas pelo modelo são condizentes com a priorização real. Por exemplo, supondo que um determinado paciente foi submetido à cinco procedimentos, que são ordenados pela sua relevância, este tipo de avaliação analisa se a ordem da recomendação realizada pelo modelo é condizente com a relevância determinada pelo profissional da saúde. No caso deste trabalho, esta tipo de avaliação foi utilizado para apurar se os procedimentos que foram adotados no tratamento dos pacientes encontraram-se nas primeiras posições da lista de recomendações ou não. Uma das métricas mais populares em SRs para este fim é o chamado nDCG, que foi adotada.

Finalmente, a métrica de cobertura avalia a porcentagem de procedimentos de tratamento disponíveis que o modelo efetivamente recomenda.

4.4.1 Descrição Experimental

Nesta seção, será descrita a parametrização dos modelos apresentados nas seções anteriores para o desenvolvimento dos experimentos deste trabalho.

Na etapa de treinamento dos modelos, discutida na Seção 4.2, os modelos de FC e FBC foram treinados para produzir recomendações de procedimentos cirúrgicos conforme o cenário descrito. Cada modelo foi submetido uma variação dos seguintes parâmetros: tamanho da lista de similaridade e métrica de similaridade, que servem como entrada do algoritmo kNN, descrito na Seção 3.2.

O tamanho da lista de similaridade representa o valor k de vizinhos mais próximos a serem calculados pelo algoritmo. Neste trabalho, foram adotados os valores de $k \in \{1, 3, 5, 10, 25, 50\}$ para os modelos de Filtragens Colaborativas. Já para os Filtragens Baseadas em Conteúdo, foi adotado apenas o valor de $k = 10$, uma vez que o resultado direto do kNN é a lista de recomendações de procedimentos cirúrgicos.

Ademais, foram adotadas as seguintes métricas de similaridade, formalizadas na Seção 3.1.7:

a) Distância Euclidiana;
b) Distância Mannhatan;
c) Similaridade dos Cossenos;
d) Coeficiente de Jaccard;
e) Coeficiente de Tanimoto;

Essas variações paramétricas, em conjunto com as variações de utilização de contexto, foram efetuadas para avaliar os seguintes aspectos de SRSs:

a) Os impactos do aumento na quantidade de vizinhos considerados, para identificar até onde o aumento na quantidade de dados que influenciam as recomendações melhoram o desempenho dos modelos ou agregam ruído;
b) Quais métricas de similaridade são mais adequadas para modelar os padrões dos dados em questão;
c) Qual o impacto de incorporar informações contextuais no processo de recomendação.

5 RESULTADOS

Neste capitulo, são apresentadas as avaliações empíricas dos modelos apontados na metodologia proposta, descrita no Capítulo 4.

Considerando o cenário de recomendação de Procedimentos de Tratamento, os modelos devem recomendar uma lista de dez procedimentos em ordem decrescente de relevância, ou ganho que o procedimento proporciona ao paciente; ou seja, o ganho do procedimento recomendado na posição i da lista de recomendação (R_i) deve ser igual ou superior ao ganho do item recomendado na posição j (R_j), para qualquer valor de $j > i$, seguindo a equação: $r_i \geq r_j \forall j > i$.

Como parâmetro de avaliação, são adotados como padrão-ouro os Procedimentos de Tratamento aos quais os pacientes da base de dados, apresentada na Seção 3.4, foram submetidos. Pelo fato da base apenas constar se um determinado procedimento foi ou não realizado no tratamento do paciente, sem incluir nenhuma informação quanto à sua relevância no tratamento, os procedimentos aplicados ao paciente serão atribuídos com valor de ganho um, enquanto os não aplicados serão atribuídos com valor de ganho zero. Portanto, o objetivo dos modelos é não apenas incluir na lista de recomendação os procedimentos que foram de fato adotados pelos médicos, mas também recomendá-los antes dos procedimentos que não foram adotados no processo de tratamento real.

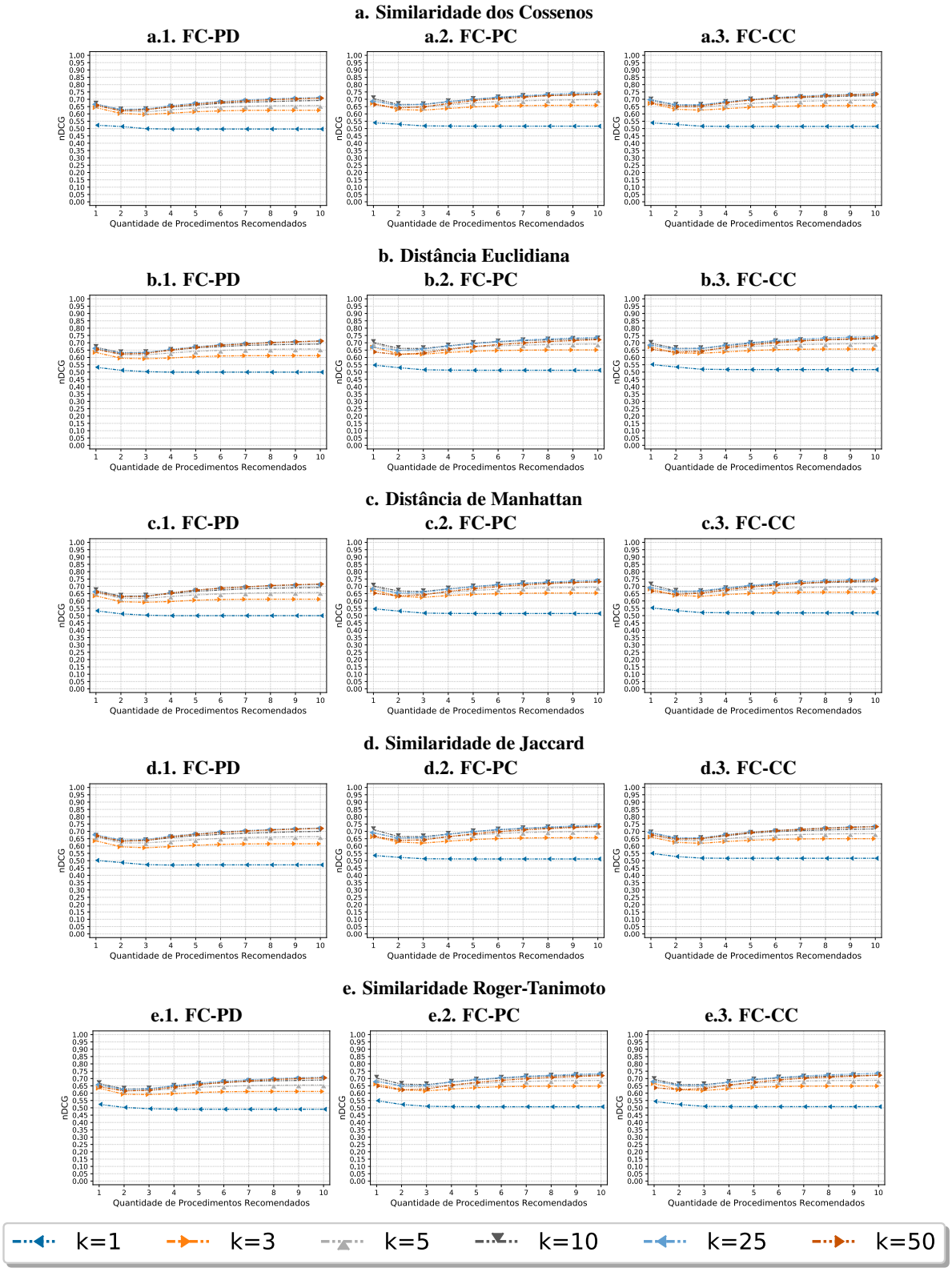
Para medir a performance dos modelos e compará-los entre si, conforme levantado na Seção 4.4, foram adotadas as métricas $nDCG$, F-$Score$, Curva de Precisão-Revocação, *F-Score* e Curva ROC, descritas na Seção 3.1.8.

Apesar de algumas métricas adotadas serem originadas da Matriz de Confusão, vale ressaltar que este cenário não se enquadra a um problema tradicional de classificação. Isto se dá pelo fato de cada paciente poder figurar simultaneamente em mais de uma classe, pois pode ser submetido a mais de um procedimento no seu processo de tratamento, sendo que a quantidade ideal de procedimentos varia para cada paciente. Ademais, há um grande desbalanceamento de classes neste cenário, pois uma quantidade muito pequena de classes são verdadeiras para cada paciente.

5.1 QUANTIDADE DE PACIENTES MAIS SIMILARES

Conforme definido na Seção 4.2.5, os modelos de Filtragem Colaborativa utilizam o algoritmo kNN para determinar os k pacientes mais similares a cada indivíduo, que são utilizados para calcular suas recomendações de procedimentos de tratamento. Sendo assim, a primeira etapa de avaliação dos modelos de FC baseia-se em analisar o impacto da quantidade de pacientes similares (k) utilizados. A Figura 9 apresenta o desempenho dos modelos de Filtragem Colaborativa em relação ao $nDCG$ médio das recomendações geradas, variando-se o valor de k para todos os modelos e métricas de similaridade adotadas.

Figura 9 – Avaliação, através do *n*DCG, do impacto da quantidade de pacientes similares (k) adotada pelos modelos de Filtragem Colaborativa (FC) em relação a cada uma das métricas de similaridade consideradas.



Fonte: Autor.

Os resultados apresentados na Figura 9 mostram que, independente do modelo de FC e da métrica de similaridade que foram utilizados, algumas características das recomendações em relação ao valor de k foram semelhantes, que são:

a) Os modelos com valor de $k = 1$ apresentam os piores desempenhos de forma evidente;
b) Nas primeiras posições da lista de recomendação, os modelos com $k > 1$ apresentam resultados próximos.
c) Nas últimas posições da lista de recomendação, a diferença entre os resultados torna-se mais discriminada, onde há uma melhora evidente nos resultados conforme aumenta-se o valor de k até $k = 10$ e não apresentam grandes diferenças para valores de $k > 10$.
d) Apesar da pequena diferença, os melhores resultados nas últimas posições da lista são obtidos pelos modelos que adotam $k = 25$.

O crescimento do desempenho dos modelos, obtido através do aumento da quantidade de pacientes similares utilizados, caracteriza a melhoria de performance inerente do aumento da quantidade de informações consumidas pelo modelo. Em contrapartida, a queda na qualidade das recomendações, sobretudo nas primeiras posições da lista, para alguns modelos que utilizaram $k = 50$ caracteriza o efeito da adoção de ruído nos modelos.

Portanto, para o restante deste trabalho, foi adotado o valor de $k = 25$ para os modelos de Filtragem Colaborativa, devido ao seu desempenho estável ao longo de todas as posições da lista de recomendação e também pelo seu destaque, mesmo que pouco expressivo, nas últimas posições das listas de recomendações para todas as variações de modelos de FC e métricas de similaridade adotados.

Seguindo a estruturação dos modelos de Filtragem Baseada em Conteúdo, apresentado na Seção 4.2.6, este modelo não sofre uma variação do valor de k, uma vez que não é realizado o cálculo de similaridade diretamente entre pacientes.

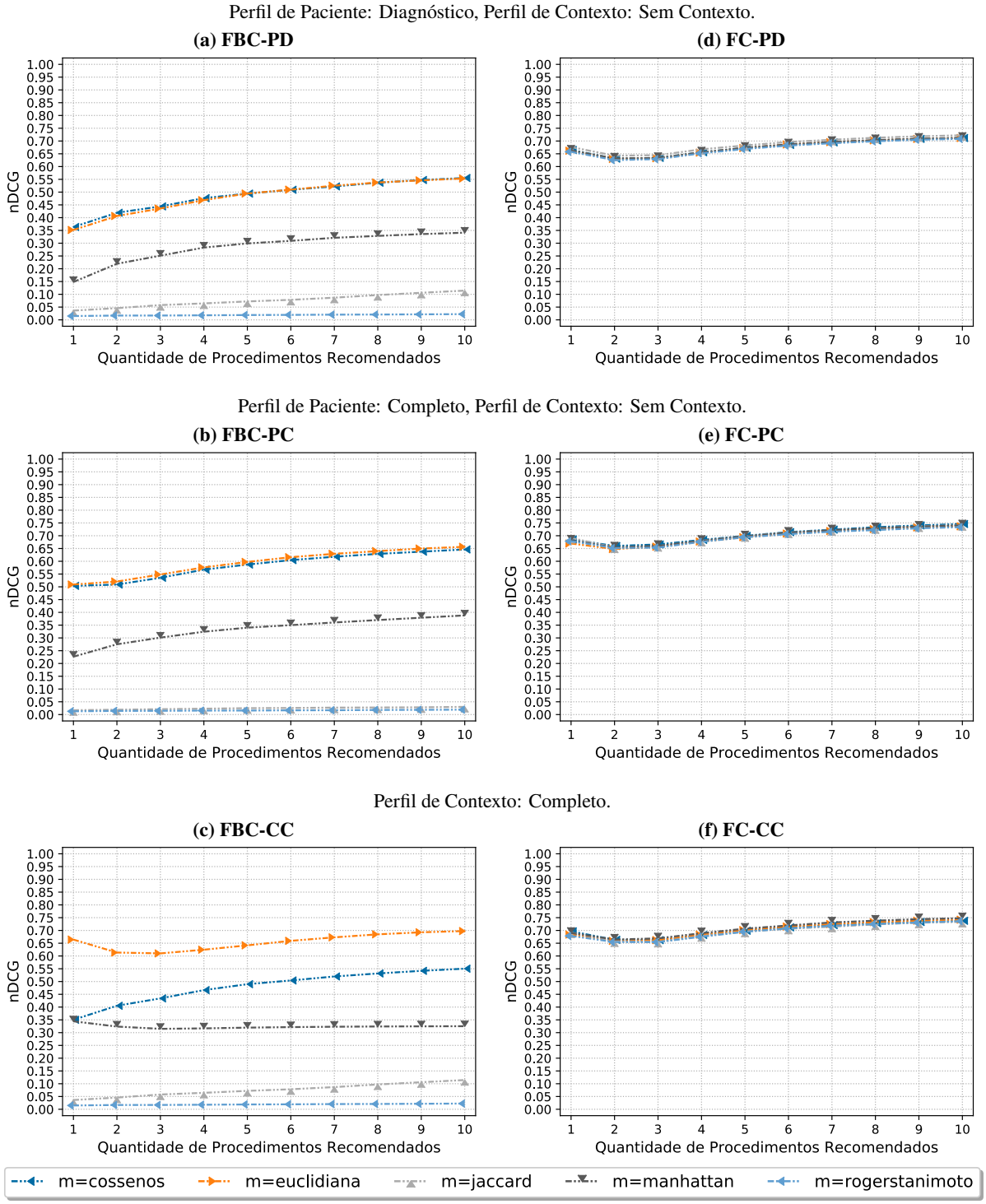
5.2 DESEMPENHO ATRAVÉS DO NDCG

A Figura 10 apresenta os resultados dos modelos de Filtragem Baseada em Conteúdo e Filtragem Colaborativa em relação ao $nDCG$ médio das recomendações geradas para todos os pacientes.

Comparando as Figuras 10.a e 10.b, pode-se observar que o modelo com Perfil de Paciente Completo promove uma melhoria significativa na qualidade das recomendações em comparação com o Perfil de Paciente Diagnóstico, na ausência de utilização de contexto. Ademais, os modelos de Filtragem Baseada em Conteúdo que adotam a Distância Euclidiana e Distância dos Cossenos apresentam resultados equiparados entre si e melhores do que os modelos de FBC que adotam as demais métricas de similaridade.

Entretanto, a utilização de contexto exerce um efeito diferente em cada um dos modelos de FBC que se destacaram anteriormente, conforme evidenciado na Figura 10.c. Enquanto

Figura 10 – Avaliação dos modelos de recomendação de procedimentos cirúrgicos através da métrica *n*DCG.



Fonte: Autor.

a adoção do Perfil de Contexto Completo em conjunto com a Distância Euclidiana melhorou a qualidade das recomendações, atingindo os melhores resultado dentre todos os modelos de Filtragem Baseada em Conteúdo, a adoção do Perfil de Contexto Completo juntamente com a Distância dos Cossenos resultou num desempenho pior comparado ao Perfil de Paciente Completo.

Diferentemente dos resultados dos modelos de Filtragem Baseada em Conteúdo, os modelos de Filtragem Colaborativa se mostram pouco sensíveis às métricas de similaridade adotadas no cenário de recomendação explorado, conforme ilustrado na Figura 10.

Esta estabilidade em relação às métricas de similaridade pode ser explicada pelas características dos dados que compõem o Perfil dos Pacientes, descritos na Seção 4.2.2. O Perfil dos Pacientes é composto quase integralmente por variáveis binárias com alto grau de esparsidade, ou seja, a grande maioria dessas variáveis binárias assume valor zero e poucas assumem valor um. Com isso, ao calcular-se o conjunto de pacientes mais próximos à cada indivíduo, através do algoritmo *kNN* (Seção 4.2.5), a tendencia é que os conjuntos resultantes não sofram grandes alterações em relação a quais pacientes os compõem, mesmo adotando-se diferentes métricas de similaridade. Apesar desta característica de os k pacientes mais similares a um determinado paciente serem basicamente os mesmos, a mudança da métrica adotada resulta em uma variação na diferença entre o fator de similaridade destes pacientes. Esta variação explica a pequena diferença entre os resultados dos modelos de FC que utilizam cada uma das métricas adotadas.

A diferença entre o desempenho dos modelos FC-PD, FC-PC e FC-CC, ilustrados respectivamente pelas Figuras 10.e, 10.e e 10.f, também não apresenta grandes variações. O modelo com Perfil de Paciente Completo apresentou resultados ligeiramente melhores que o modelo que adota o Perfil de Paciente Diagnóstico. Além disso, a melhoria obtida pela adoção do Perfil de Contexto Completo em relação ao modelo que utiliza o Perfil de Paciente Completo é ainda menos expressiva.

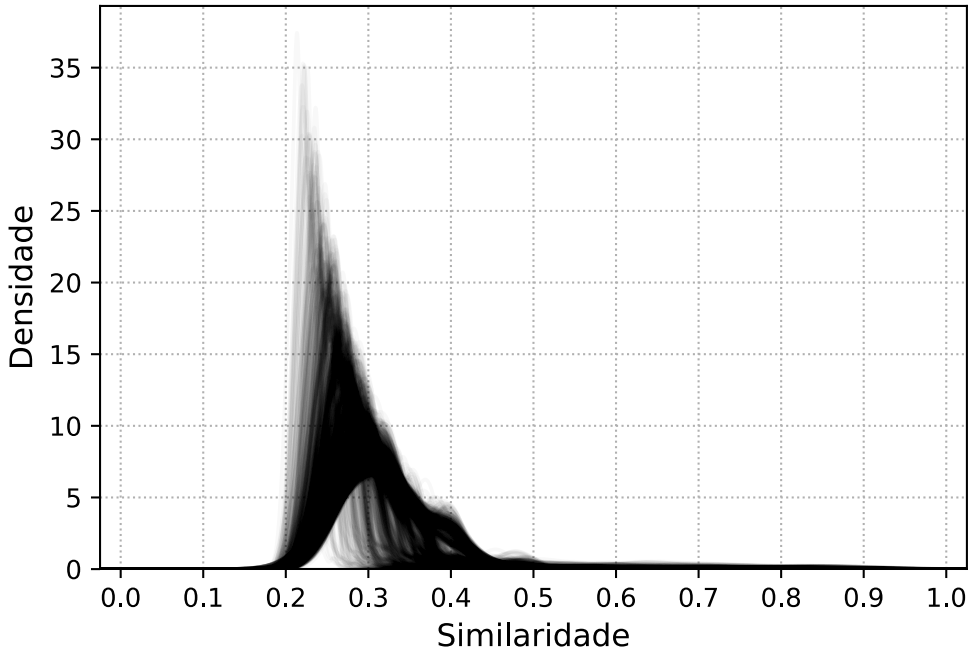
Comparando os resultados dos modelos de Filtragem Colaborativa em relação aos modelos de Filtragem Baseada em Conteúdo, apresentados na Figuras 10, os modelos de FC superam os resultados dos modelos de FBC em todas as variações.

5.2.1 Impacto das Métricas de Similaridade nos Modelos de Filtragem Baseada em Conteúdo Sensíveis ao Contexto

Considerando as métricas de similaridade que se destacaram para os modelos de FBC, a diferença entre a Distância Euclidiana e a Distância dos Cossenos para os dados utilizados nesta análise podem ser evidenciados pelas Figuras 11 e 12, que apresentam uma suavização das curvas de distribuição da similaridade do Perfil de Contexto de cada paciente pertencente às amostras de treinamento em relação aos pacientes nas suas respectivas amostras de treinamento.

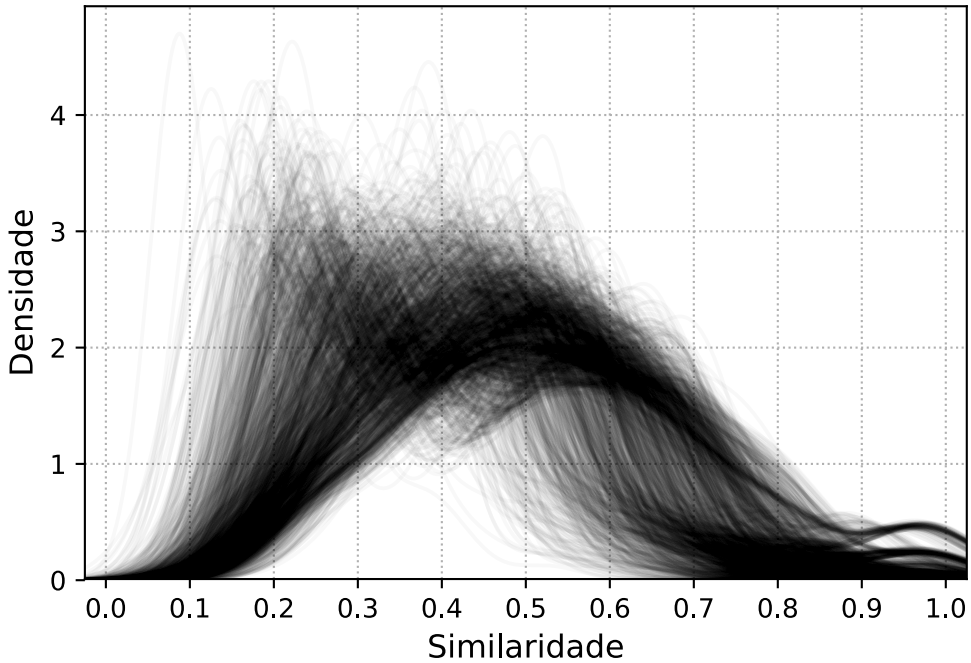
Nas Figuras 11 e 12, com o objetivo de destacar os pontos de concentração de similaridade entre pacientes para cada uma das métricas, as curvas de todos os pacientes foram sobre-

Figura 11 – Sobreposição das curvas suavizadas de distribuição suavizadas de similaridade entre o Perfil de Contexto Completo de cada paciente em relação aos demais através da Distância Euclidiana.



Fonte: Autor.

Figura 12 – Sobreposição das curvas suavizadas de distribuição de similaridade entre o Perfil de Contexto Completo de cada paciente em relação aos demais através da Distância dos Cossenos.



Fonte: Autor.

postas, onde cada curva representa a distribuição de similaridade de um paciente. Desta forma, as regiões mais escuras representam pontos de maior incidência de pacientes que possuem uma determinada quantidade de pacientes similares em um grau de similaridade específico.

Com base na Figura 11, pode-se observar que as grandes concentrações de pacientes apresentam baixo grau similaridade, entre 0.2 e 0.5, enquanto há poucos pacientes com alto grau de similaridade, acima de 0.5. Ou seja, a Pré-Filtragem Contextual com a Distância Euclidiana faz com que a influência exercida pela maioria dos pacientes no processo de criação dos Perfis de Procedimento de Tratamento seja mitigada.

Em contrapartida, a Figura 12 indica que a Pré-Filtragem Contextual através da Distância dos Cossenos frequentemente resulta em curvas de distribuição de similaridade mais dispersas, com maior desvio-padrão. Consequentemente, o impacto na construção dos Perfis de Procedimento de Tratamento é mais bem distribuído através dos pacientes da amosta de treinamento, causando um aumento na quantidade de pacientes com alto impacto na construção destes perfis em comparação com a pré-filtragem contextual através da Distância Euclidiana.

5.2.2 Impacto dos Perfis de Contexto

Ao observar o impacto de cada Perfil de Contexto individualmente aos modelos de Filtragem Baseada em Conteúdo, ilustrado pela Figura 13, três fenômenos podem ser destacados:

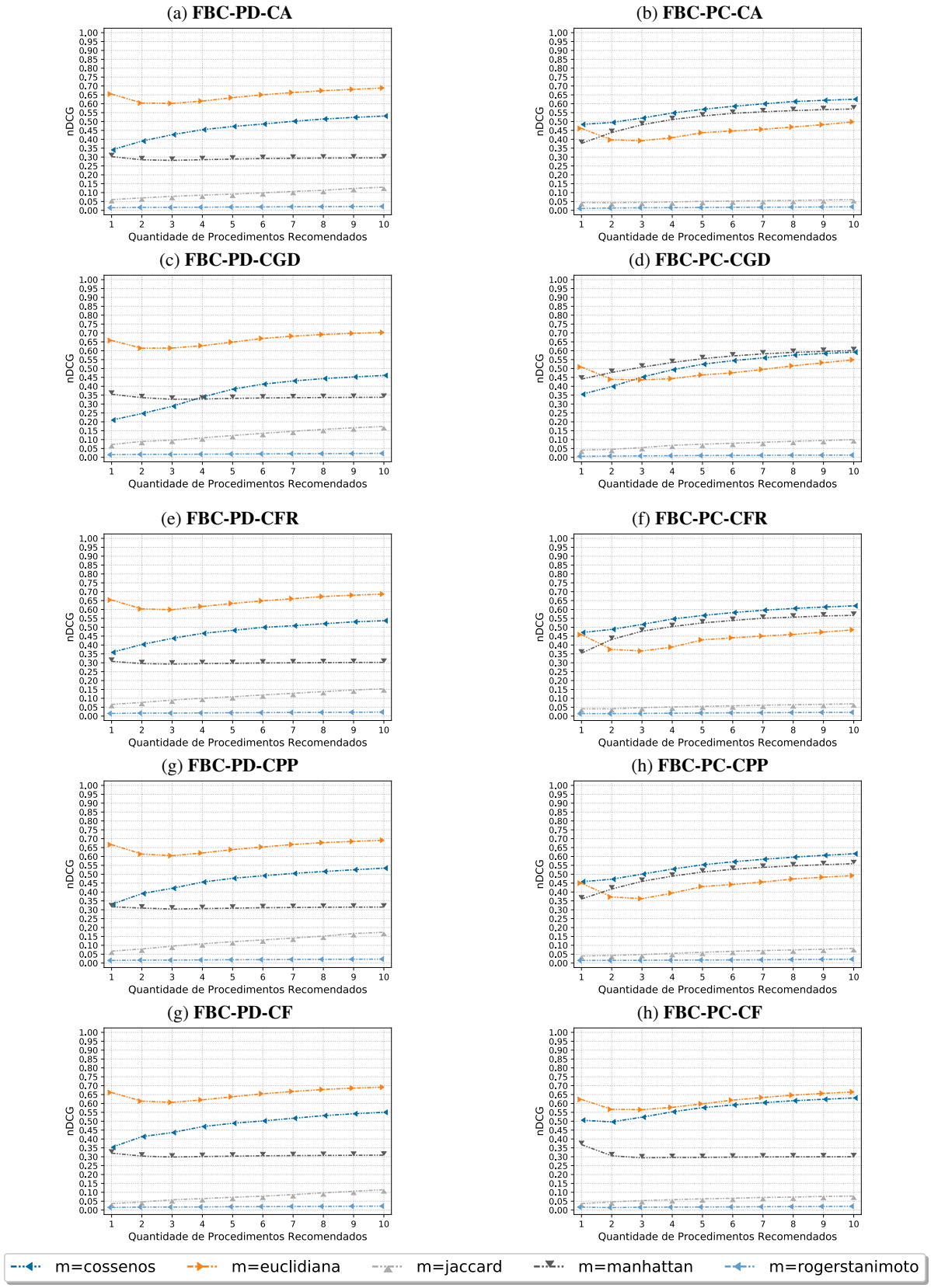
a) Nos modelos que utilizam o Perfil de Paciente Diagnóstico, os modelos que adotam a Distância Euclidiana não apenas destacam-se dos demais, como também atingem resultados próximos aos do modelo que utiliza o Perfil de Contexto Completo (Figura 10.c) em todos os Perfis de Contexto.

b) Nos modelos que utilizam o Perfil de Paciente Completo, com exceção do Perfil de Contexto Físico, os modelos que adotam a Distância Euclidiana obtêm resultados inferiores ao modelo que não utiliza contexto (Figura 10.b). Sugerindo assim que a utilização do Perfil de Paciente Completo juntamente com a Distância Euclidiana, para o cálculo de similaridade dos pacientes em relação aos procedimentos de tratamento, agrega ruído ao modelo, uma vez que todos os resultados com o Perfil de Paciente Diagnóstico superam distintamente os resultados com o Perfil de Paciente Completo.

c) Apesar da queda de desempenho dos modelos com Perfil de Paciente Completo combinados com a Distância Euclidiana e a utilização de contexto, destacada anteriormente, o modelo FBC-PC-CF com a Distância Euclidiana (Figura 10.h) foi o que obteve os melhores resultados dentre os modelos de Filtragem Baseada em Conteúdo que utilizam Perfil de Paciente Completo. Destacando assim o impacto positivo do uso do Perfil de Contexto Físico, sobretudo com a Distância Euclidiana.

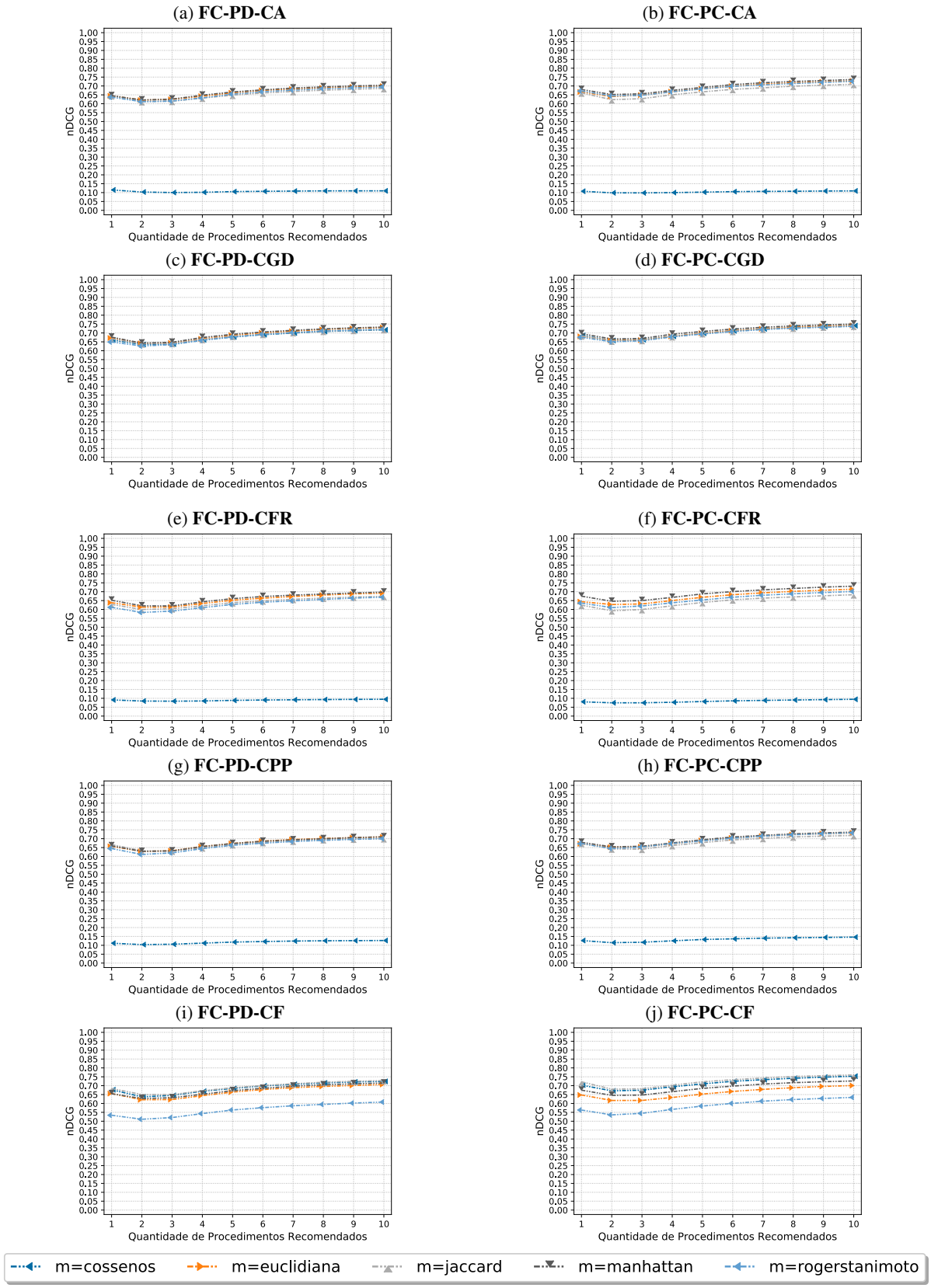
Os fenômenos destacados acima sugerem que a Distância Euclidiana é a métrica que melhor pondera a importância dos pacientes para a construção dos Perfis dos Procedimentos de

Figura 13 – Avaliação, através da métrica *n*DCG, do impacto dos perfis de contexto específicos na recomendação de procedimentos cirúrgicos através dos modelos de Filtragem Baseada em Conteúdo.



Fonte: Autor

Figura 14 – Avaliação, através da métrica *n*DCG, do impacto dos perfis de contexto específicos na recomendação de procedimentos cirúrgicos através dos modelos de Filtragem Baseada em Conteúdo.



Fonte: Autor

Tratamento com base nos dados contextuais para os modelos de Filtragem Baseada em Conteúdo.

Em relação ao impacto de cada Perfil de Contexto aos modelos de Filtragem Colaborativa, a Figura 14 mostra que não há nenhum perfil que se destaca em melhorar a qualidade das recomendações. Entretanto, alguns Perfis de Contexto específicos são mais sensíveis às métricas de similaridade. Por exemplo, para os modelos que adotam os Perfis de Contexto de Fatores de Risco e Físicos, apresentados nas Figuras 14.e, 14.f, 14.i e 14.j, pode-se observar uma maior variação de desempenho dos modelos que utilizam diferentes métricas de similaridade, sugerindo que algumas métricas são mais adequadas para representar a similaridade entre pacientes em alguns contextos específicos.

Outra característica que pode ser destacada pela Figura 14 é o desempenho muito inferior dos modelos que utilizam a similaridade dos cossenos nos Perfis de Contexto de Anomalias, Fatores de Risco e Procedimentos Prévios. Este fenômeno ocorre devido ao fato destes Perfis de Contextos representarem características mais raramente observadas nos pacientes da base de dados adotada, fazendo com que os perfis de contexto destes pacientes sejam representados por vetores nulos, resultando numa similaridade de contexto igual a zero em relação a qualquer outro vetor. Seguindo a metodologia de utilização de contexto nos modelos de Filtragem Colaborativa, apresentados na Seção 4.2.5, a Equação (29) mostra que uma similaridade entre perfis de contexto igual a zero faz com que a similaridade entre os pacientes também seja igual a zero, resultando em listas de recomendações vazias para estes pacientes.

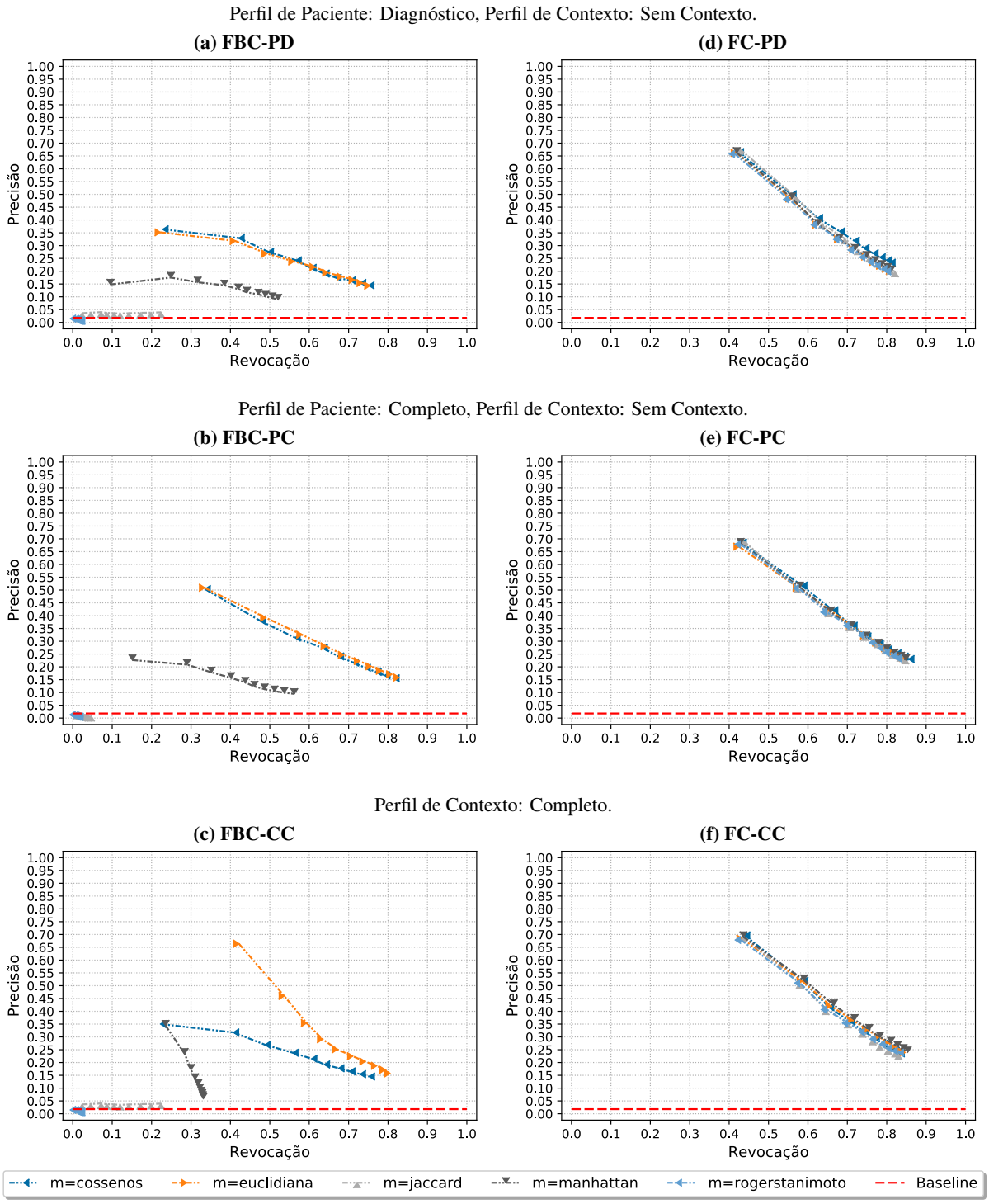
5.3 DESEMPENHO ATRAVÉS DA CURVA DE PRECISÃO-REVOCAÇÃO

Nesta seção, são apresentados os resultados obtidos pelos modelos de recomendação de procedimentos de tratamento através da Curva de Precisão-Revocação, exibidos na Figura 15. Um ponto importante a ser destacado é a relação entre o desempenho dos diferentes modelos apresentados, que permanece a mesma quando comparada às avaliações através do $nDCG$, apresentados na Seção 5.2; ou seja, os modelos com os melhores e piores desempenhos permanecem os mesmos. A diferença está nas características dos resultados que são exploradas por cada métrica.

Comparando o comportamento do desempenho dos modelos em relação à Curva de Precisão-Revocação, ilustrada na Figura 15, com os resultados do $nDCG$ (Figura 10) pode-se observar que, conforme aumenta-se o tamanho da lista de recomendações, apesar dos resultados se manterem estáveis em relação ao $nDCG$, a Precisão diminuí e a Revocação aumenta gradualmente. Este fenômeno ocorre devido à diferença de sensibilidade a falsos positivos e verdadeiros positivos entre as métricas.

Neste cenário, onde o ganho das recomendações assume apenas os valores um (positivo) ou zero (negativo), o $nDCG$ só é afetado negativamente pelos falsos positivos que aparecem antes de todos os valores verdadeiros serem recomendados. Por outro lado, mesmo que todos

Figura 15 – Avaliação dos modelos de recomendação de procedimentos cirúrgicos através da Curva de Precisão-Revocação.



Fonte: Autor.

os valores positivos já tenham sido corretamente apresentados nas primeiras posições da lista de recomendações, a Precisão decresce conforme a lista de recomendações aumenta. Como cada paciente é submetido a uma quantidade pequena de procedimentos, o aumento da lista de recomendações implica no aumento da proporção de falsos positivos em relação ao total de procedimentos recomendados, consequentemente diminuindo o valor da Precisão.

Em contrapartida, os verdadeiros positivos têm maior impacto sobre o $nDCG$ nas primeiras posições da lista de recomendação, já que o ganho dos procedimentos é mitigado gradualmente, em ordem logarítmica, em relação a sua posição na lista de recomendações, conforme definido na Equação (21). Por sua vez, o impacto dos verdadeiros positivos na Revocação é o mesmo em todas as posições da lista de recomendação. Portanto, o aumento do tamanho da lista de recomendação permite que mais procedimentos sejam recomendados para um mesmo paciente, consequentemente aumentando a chance de verdadeiros positivos serem recomendados.

Esta discussão mostra como aspectos diferentes da performance dos Sistemas de Recomendação são medidos através de cada uma das métricas, que podem ser priorizadas dependendo do cenário onde os sistemas são aplicados.

Por exemplo, supondo que o SR seja utilizado para produzir recomendações de medicamentos diretamente aos pacientes, uma performance de Precisão alta pode ser priorizada, já que uma recomendação indevida pode resultar em riscos à saúde do indivíduo.

Por outro lado, supondo que o SR seja utilizado para ajudar o médico no processo de diagnóstico diferencial, valores de Revocação e $nDCG$ mais altos podem ser mais apropriados. Neste caso, pode ser importante ter uma quantidade maior de verdadeiros positivos recomendados, mesmo que a Precisão seja baixa, pois o médico terá o senso crítico de avaliar as recomendações geradas pelo sistema e selecionar as que sejam adequadas. Ademais, quanto antes o sistema for capaz de recomendar os diagnósticos corretos, mais benéfico o sistema será para o médico.

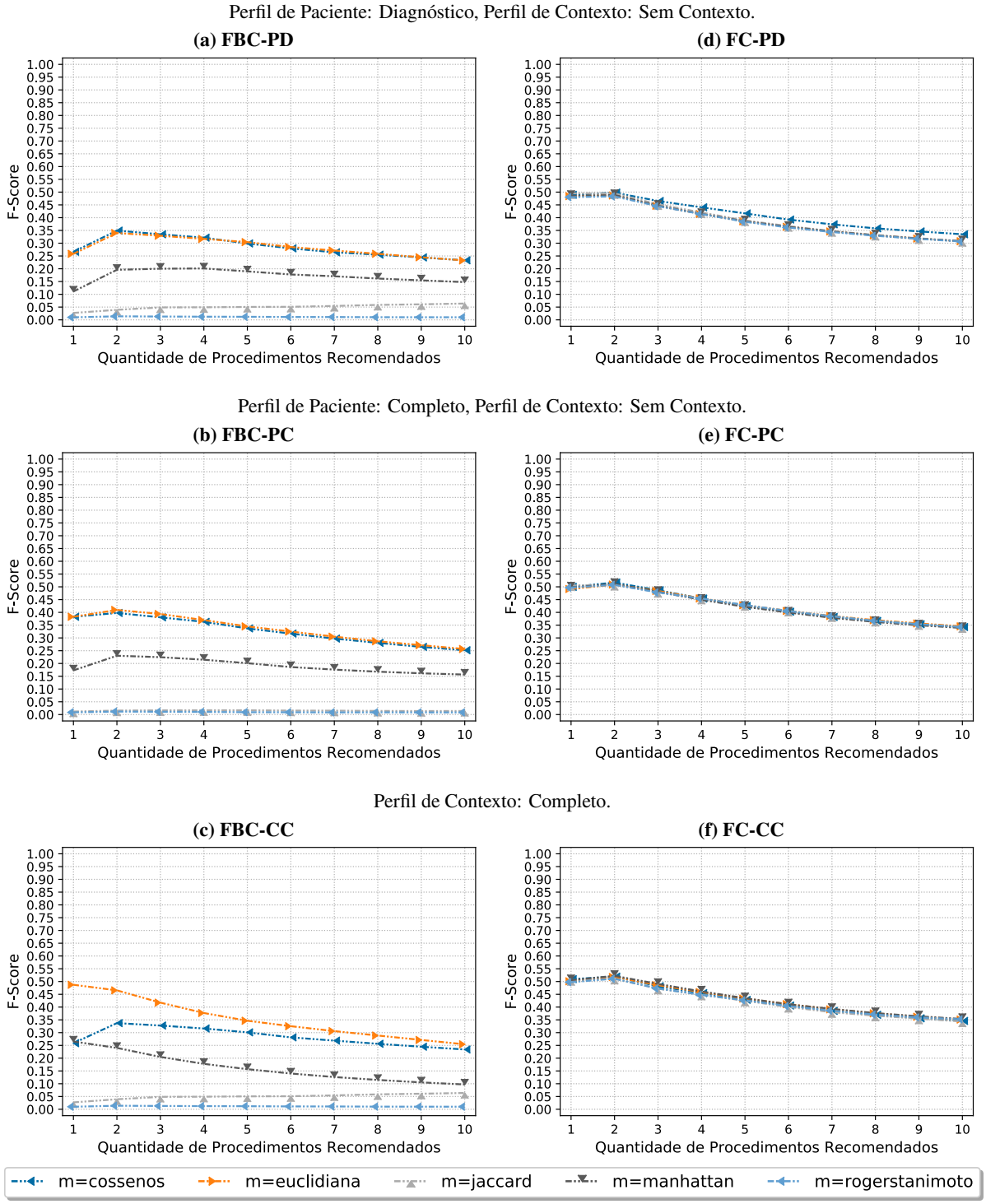
5.4 DESEMPENHO ATRAVÉS DO *F-SCORE*

A Figura 16 apresenta a avaliação dos Sistemas de Recomendação através do *F-Score* (F_1). Conforme apresentado na Seção 3.1.8.1.2, o *F-Score* combina a Precisão e a Revocação em um valor único.

Considerando que a imagem da função *F-Score* é $F_1 \in R, 0 \leq F_1 \leq 1$, onde melhor é o desempenho do modelo quanto maior é o valor de F_1, pode-se observar que até os modelos com os melhores resultados não são capazes de atingir $F_1 \geq 0.55$.

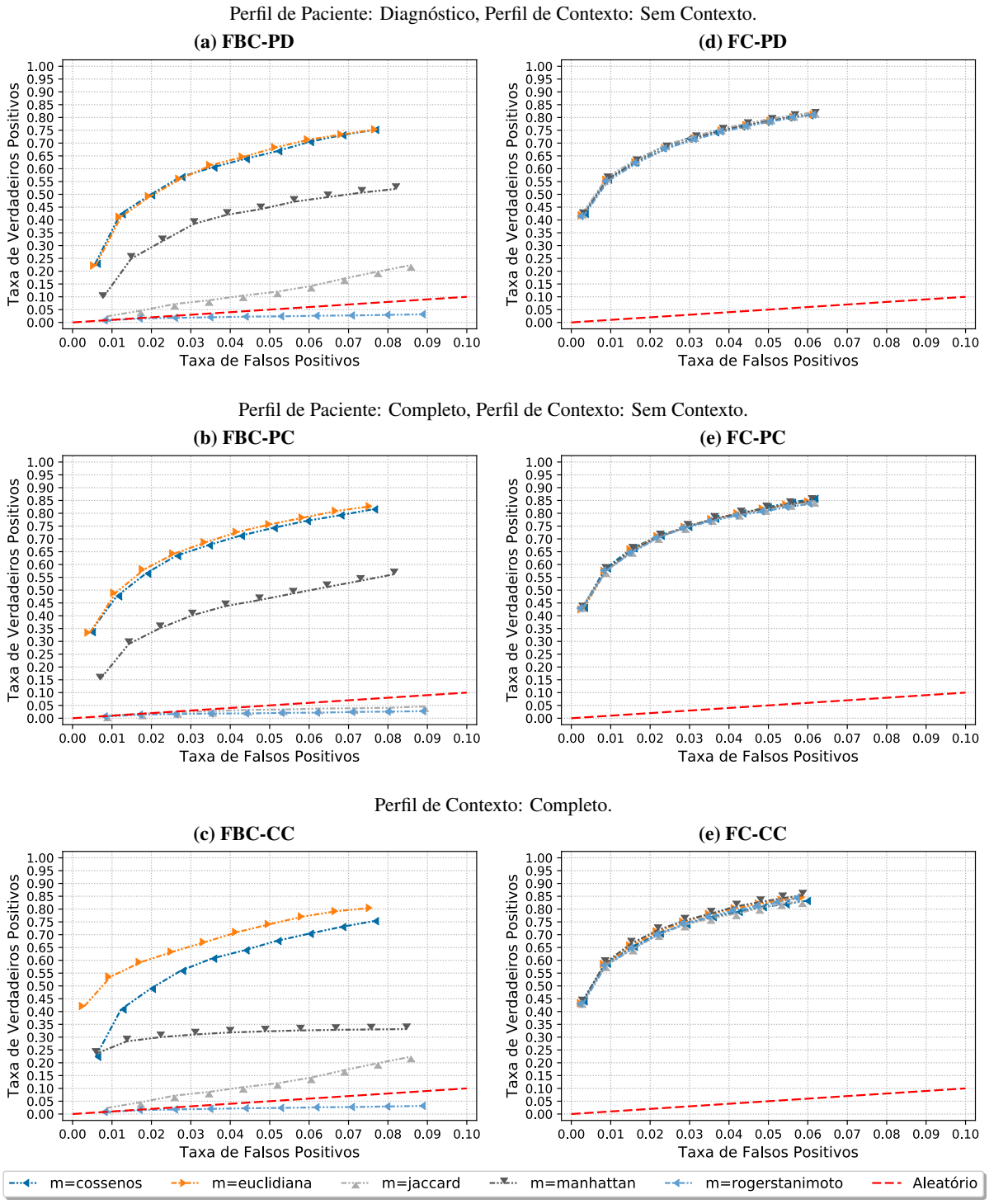
Equiparando o *F-Score* com a avaliação através da Curva Precisão-Revocação, apresentada na Seção 5.3, que também relaciona a Precisão e a Revocação, o padrão de valores de Precisão inversamente proporcional à Revocação indicados pela Curva de Precisão-Revocação, apresentada na Figura 15, justifica o desempenho ruim dos modelos em relação ao *F-Score*.

Figura 16 – Avaliação dos modelos de recomendação de procedimentos cirúrgicos através do *F-Score*.



Fonte: Autor.

Figura 17 – Avaliação dos modelos de recomendação de procedimentos cirúrgicos através da Curva ROC.



Fonte: Autor.

5.5 DESEMPENHO ATRAVÉS DA CURVA ROC

Por fim, a Figura 17 exibe a avaliação dos Sistema de Recomendação através da Curva ROC, que relaciona a Revocação (ou Taxa de Verdadeiros Positivos) com a Taxa de Falsos Negativos, sendo outro método tradicional de comparação entre modelos de classificação.

Um ponto importante a ser destacado nesta avaliação são os valores baixos de Taxa de Falsos Positivos apresentados, entre 0.0 e 0.1, apesar dos modelos atingirem precisão abaixo de 0.2 nas listas de recomendações maiores. Isto ocorre mais uma vez devido ao fato de cada paciente ser submetido a uma quantidade limitada de procedimentos, fazendo com que a quantidade de verdadeiros negativos para cada paciente seja muito grande.

Por exemplo, considerando que a base de dados contém um total de 113 procedimentos cirúrgicos, supondo que um determinado paciente tenha sido submetido a três procedimentos e nenhum deles seja recomendado até a décima posição da lista de recomendação, seguindo a Equação (18), apresentada na Seção 3.1.8.1.2, o valor da Taxa de Falsos Positivos (TFP) na décima posição da lista de recomendação será:

$$TFP = \frac{10}{10 + 100} = 0.09$$

Portanto, apesar de ainda poder-se utilizar a Curva ROC para comparar os modelos neste cenário, este método é mais adequado para avaliar modelos em cenários que apresentam um balanceamento maior de classes, ou seja, quando a quantidade de valores positivos e negativos são próximos.

5.6 DESEMPENHO DOS SISTEMAS HÍBRIDOS

Conforme abordado na Seção 3.1.6, a combinação de modelos de recomendação em um sistema híbrido é uma técnica frequentemente utilizada para a obtenção de um modelo mais assertivo.

Seguindo a metodologia apresentada na Seção 4.3, foram criados cinco modelos híbridos para recomendação de procedimentos de tratamento, um para cada métrica de similaridade adotada. A Tabela 3 apresenta os pesos calculados para cada variação de modelo de recomendação, separados por métrica de similaridade, de modo a otimizar o desempenho final do sistema híbrido.

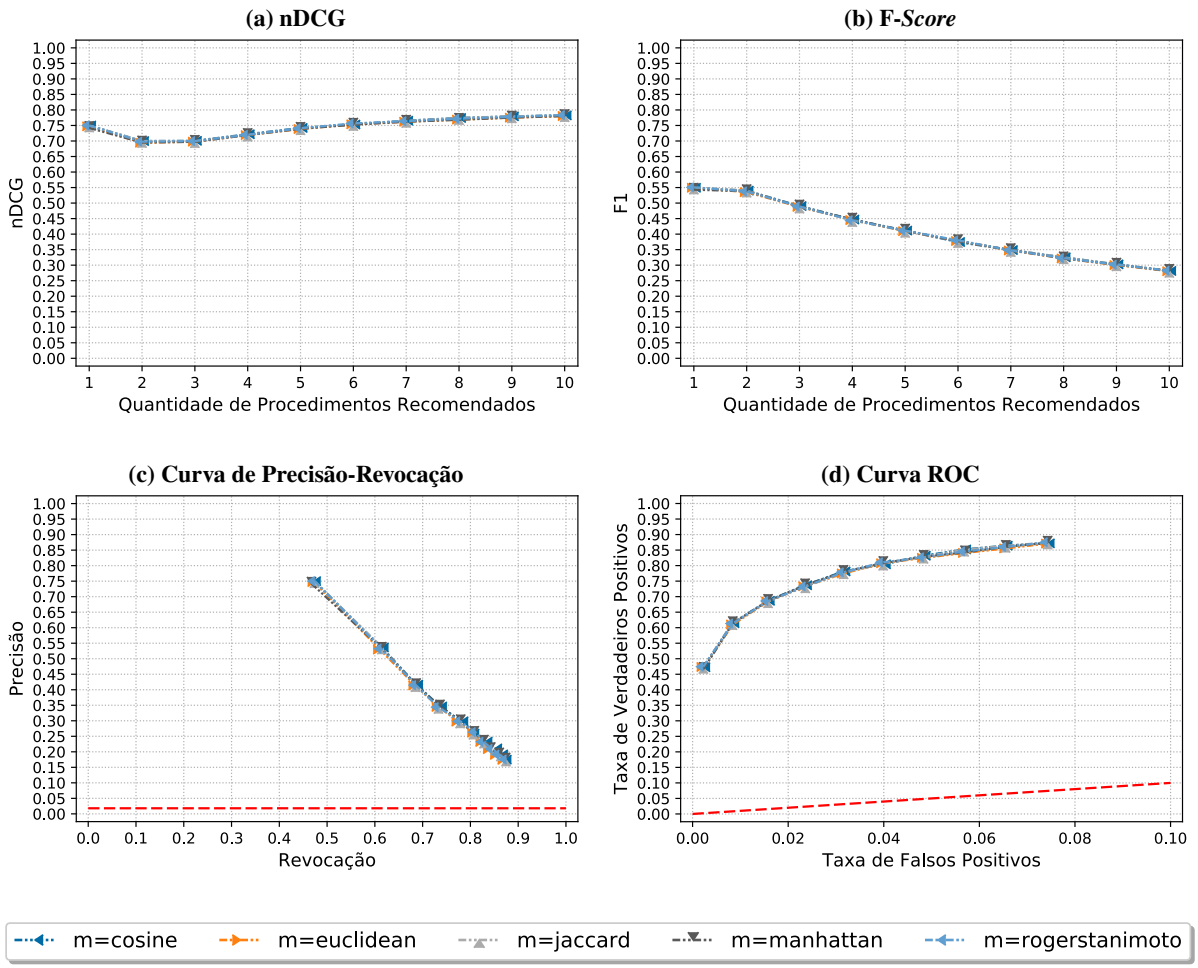
Dados os pesos de cada modelo individual na composição do modelo híbrido, apresentados na Tabela 3, fica evidente a prevalência dos modelos de Filtragem Colaborativa sobre os modelos de Filtragem Baseada em Conteúdo.

Os modelos de FBC obtiveram pesos próximos de zero nos sistemas híbridos para todas as métricas de similaridade adotadas, obtendo uma relevância máxima de 1,432%, obtida nos sistema híbrido com a similaridade dos cossenos. Até mesmo o modelo FBC-CC combinado com a Distância Euclidiana não mostrou-se relevante nos sistemas híbridos, apesar de ter se destacado dentre os modelos de FBC, conforme apontado nas seções anteriores.

Tabela 3 – Aproximação com 5 casas decimais dos pesos de cada modelos de recomendação que compõem os modelos híbridos, separados pela métrica adotada para o cálculo de similaridade entre os perfis.

	Modelo de Recomendação					
Métrica de Similaridade	**FBC-PD**	**FBC-PC**	**FBC-CC**	**FC-PD**	**FC-PC**	**FC-CC**
Cossenos	.01432	.00000	.00000	.13193	.55911	.29464
Jaccard	.00000	.00148	.00000	.21921	.55466	.22465
Rogers Tanimoto	.00000	.00158	.00000	.28649	.49087	.22106
Euclidiana	.00000	.00657	.00000	.21768	.32936	.44639
Manhattan	.00453	.00000	.00000	.20971	.43165	.35410

Figura 18 – Avaliação comparativa entre os sistemas híbridos que utilizam diferentes métricas para cálculo de similaridade. As avaliações são realizadas através do *nDCG*, F-*Score*, Curva de Precisão-Revocação e Curva ROC.



Fonte: Autor.

Este comportamento sugere que os modelos de FC não apenas são mais assertivos em comparação aos modelos de FBC no cenário de recomendação apresentado, mas também que os acertos dos modelos de FBC são majoritariamente um subconjunto dos acertos dos modelos de FC.

Outro ponto que pode ser destacado da análise dos pesos dos modelos apresentados na Tabela 3 é a variação da proporção dos pesos dos diferentes modelos de Filtragem Colaborativa (FC) para cada métrica de similaridade adotada. Com exceção do sistema híbrido que utiliza a Distância Euclidiana, o modelo FC-PC foi o que recebeu os maiores pesos, em alguns casos assumindo uma relevância maior que 50% na composição dos sistemas híbridos. Apesar do destaque dos modelos de FC-PC, os demais modelos de FC sempre assumem valores de peso expressivos, variando de 13% a 35% aproximadamente da formação do sistema híbrido.

Mais um destaque que pode ser apontado é o fato do modelo FC-CC ter atingido o maior fator de importância para o sistema híbrido que utiliza a distância Euclidiana, representando 44,6% do mesmo. Esta característica reforça a hipótese de que a Distância Euclidiana é a métrica que melhor representa a similaridade contextual para os dados utilizados no cenário de recomendação deste trabalho, conforme apontado na Seção 5.2.2.

A Figura 18 apresenta a comparação entre a performance dos sistemas híbridos que utilizam cada uma das métricas de similaridade adotadas neste trabalho. Nesta visualização, fica evidente que os modelos são equivalentes entre si. Considerando que os sistemas híbridos são basicamente compostos pela combinação de três modelos de Filtragem Colaborativa, esta equivalência ressalta os apontamentos levantados na Seção 5.2 sobre a baixa sensibilidade dos modelos de FC em relação às métricas de similaridade para este cenário de recomendação.

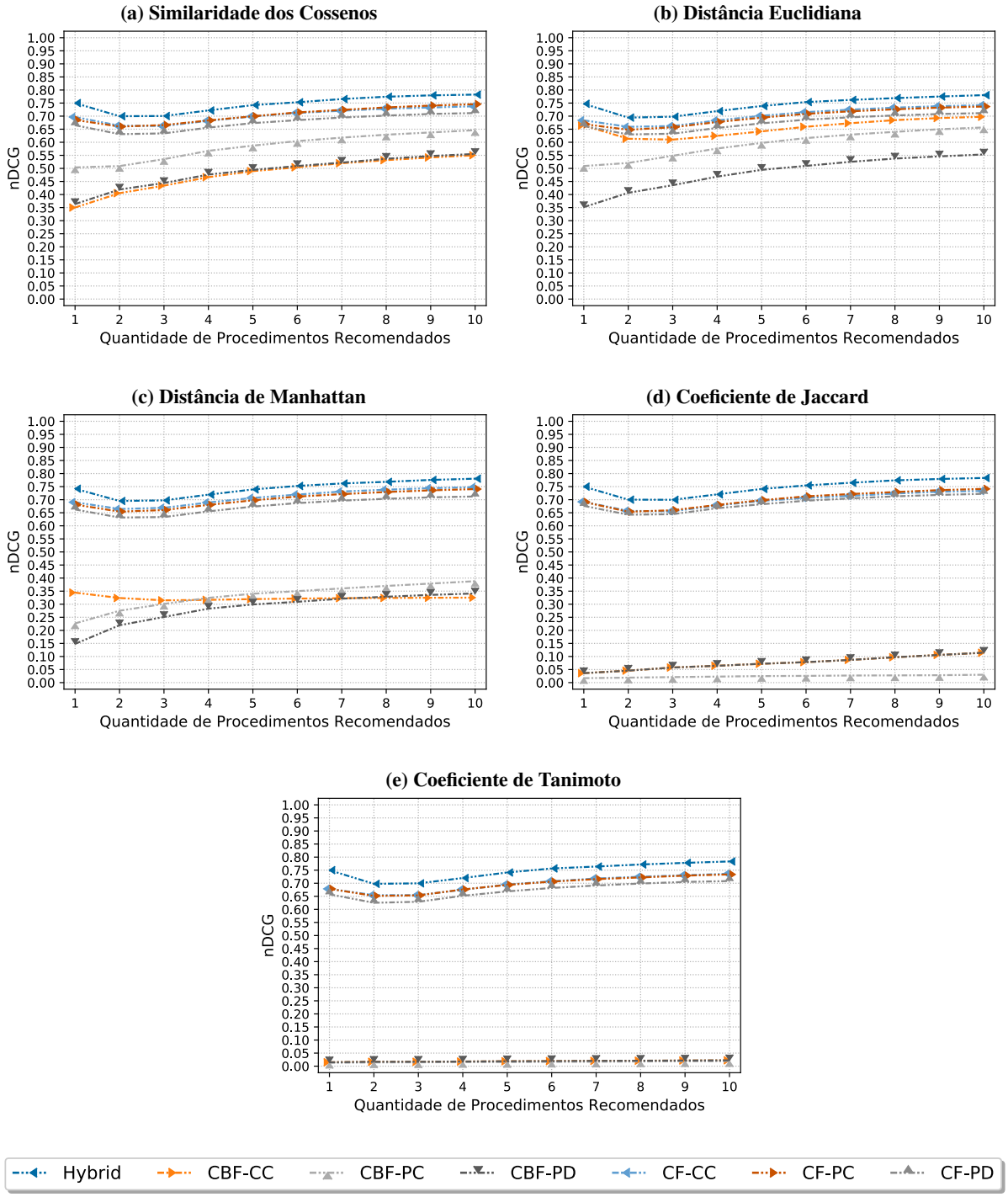
Por fim, as Figuras 19, 20, 21 e 22 apresentam uma comparação entre o desempenho dos sistemas híbridos com o desempenho dos demais modelos de recomendação, apresentados nas seções anteriores, que utilizam a mesma métrica de similaridade.

A comparação de desempenho através do $nDCG$, apresentada na Figura 19, mostra que os sistemas híbridos superam todos os demais modelos de recomendação em todas as situações consideradas neste trabalho. Esta melhoria, obtida principalmente pela combinação dos modelos de FC, mostra que cada uma das diferentes variações de modelos de FC desenvolvidas foram capazes de representar algumas características diferentes entre si ao gerar as recomendações de procedimento de tratamento para os pacientes.

A investigação do desempenho dos sistemas híbridos através do F-*Score*, ilustrada na Figura 20 mostra que há uma melhoria em relação aos demais nas primeiras posições da lista de recomendação. Entretanto, embora os resultados dos sistemas híbridos tenham claramente superado os demais modelos na análise do $nDCG$, o desempenho do sistema híbrido em relação ao F-*Score* apresenta uma queda conforme aumenta-se o tamanho da lista de recomendação, deixando de superar alguns dos outros modelos de recomendação nas ultimas posições. A explicação para este fenômeno está relacionada com o comportamento dos modelos em relação à Curva de Precisão-Revocação e à Curva ROC.

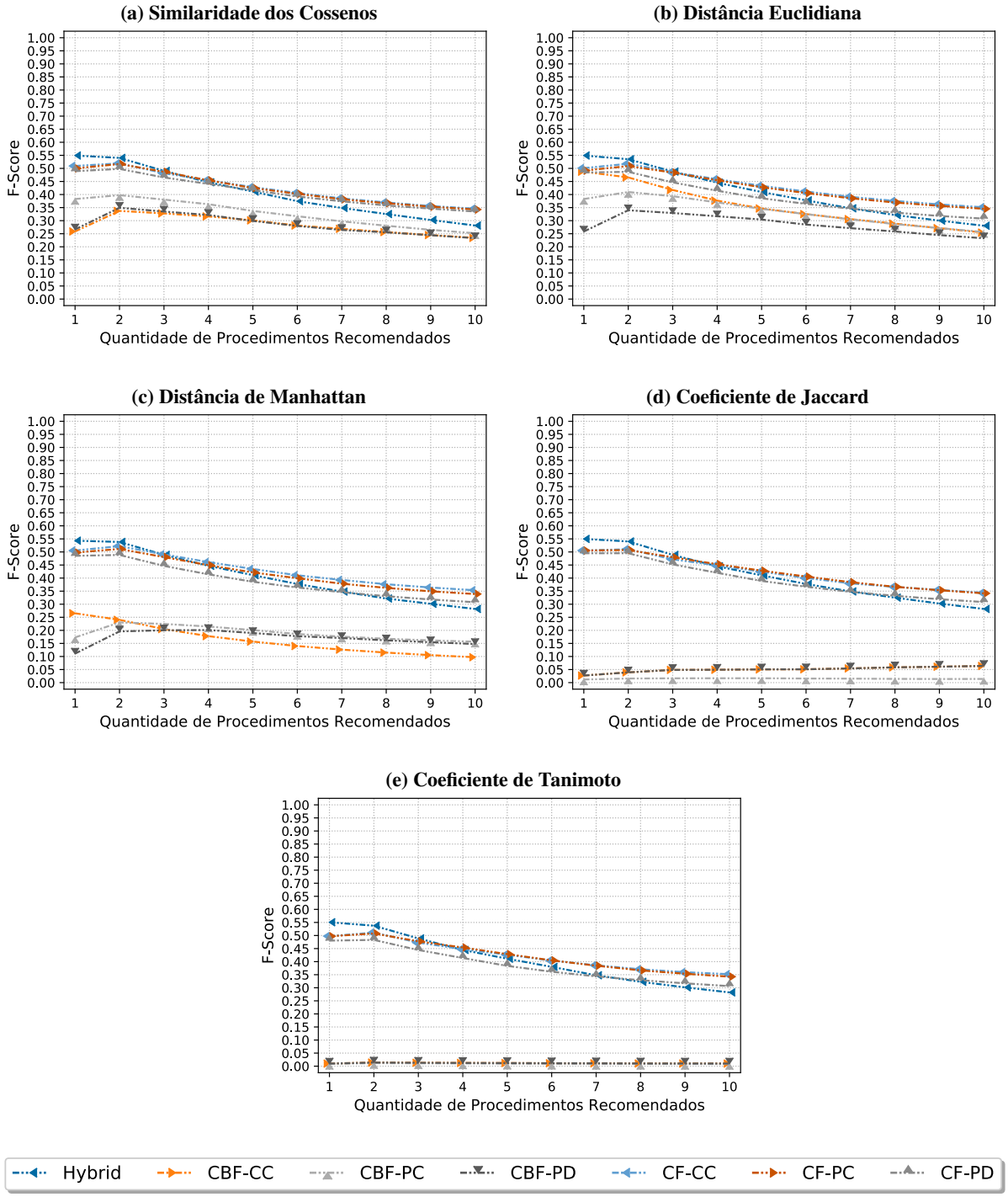
Pela observação dos resultados da Figura 21, pode-se notar que o aumento da quantidade de procedimentos recomendados faz com que o sistema híbrido atinja valores mais altos de Revocação. Contudo, esse aumento também causa uma queda notável no desempenho do sistema híbrido em relação a Precisão.

Figura 19 – Avaliação dos sistemas híbridos de recomendação de procedimentos cirúrgicos através do $nDCG$.



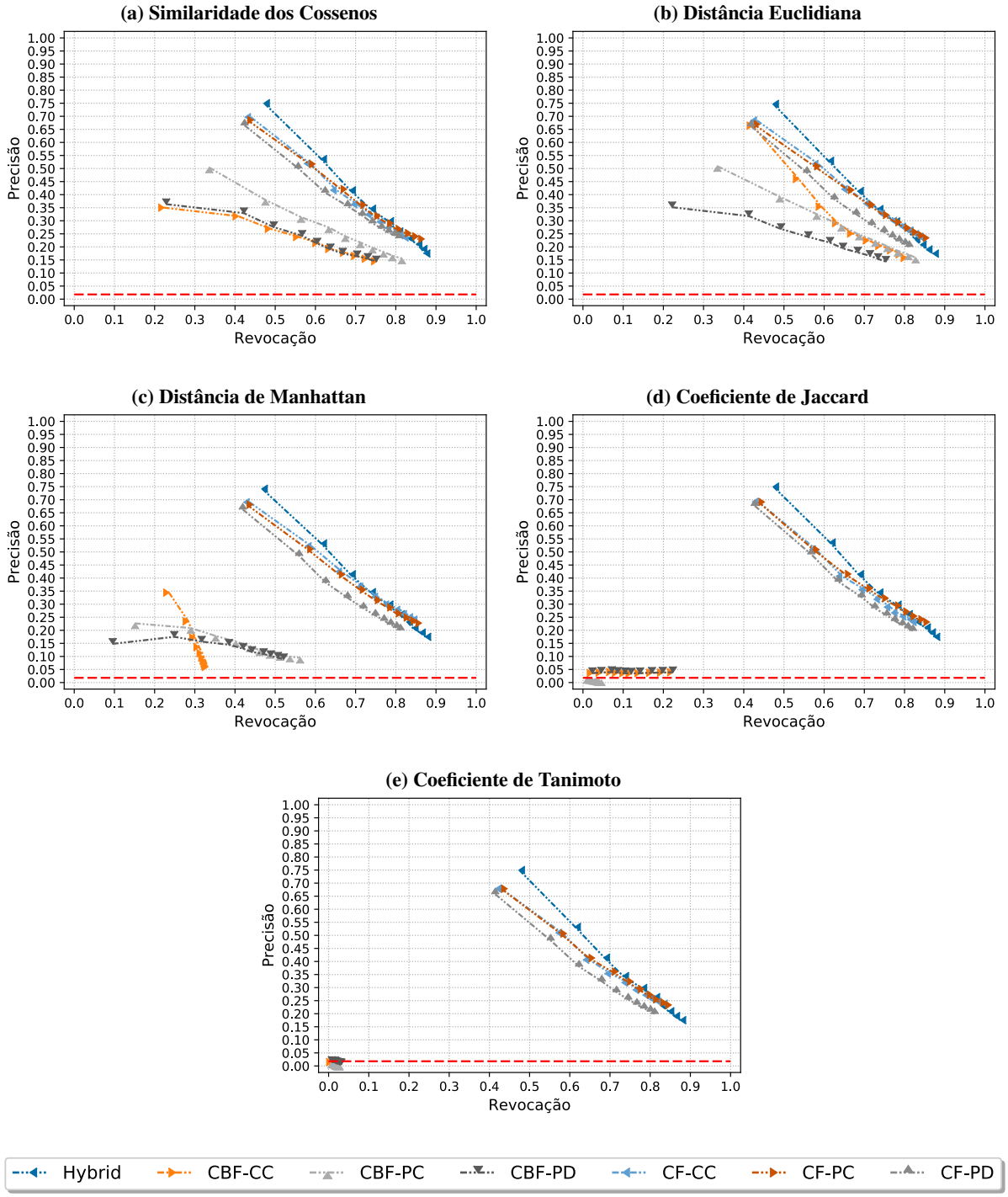
Fonte: Autor.

Figura 20 – Avaliação dos sistemas híbridos de recomendação de procedimentos cirúrgicos através do F-*Score*.



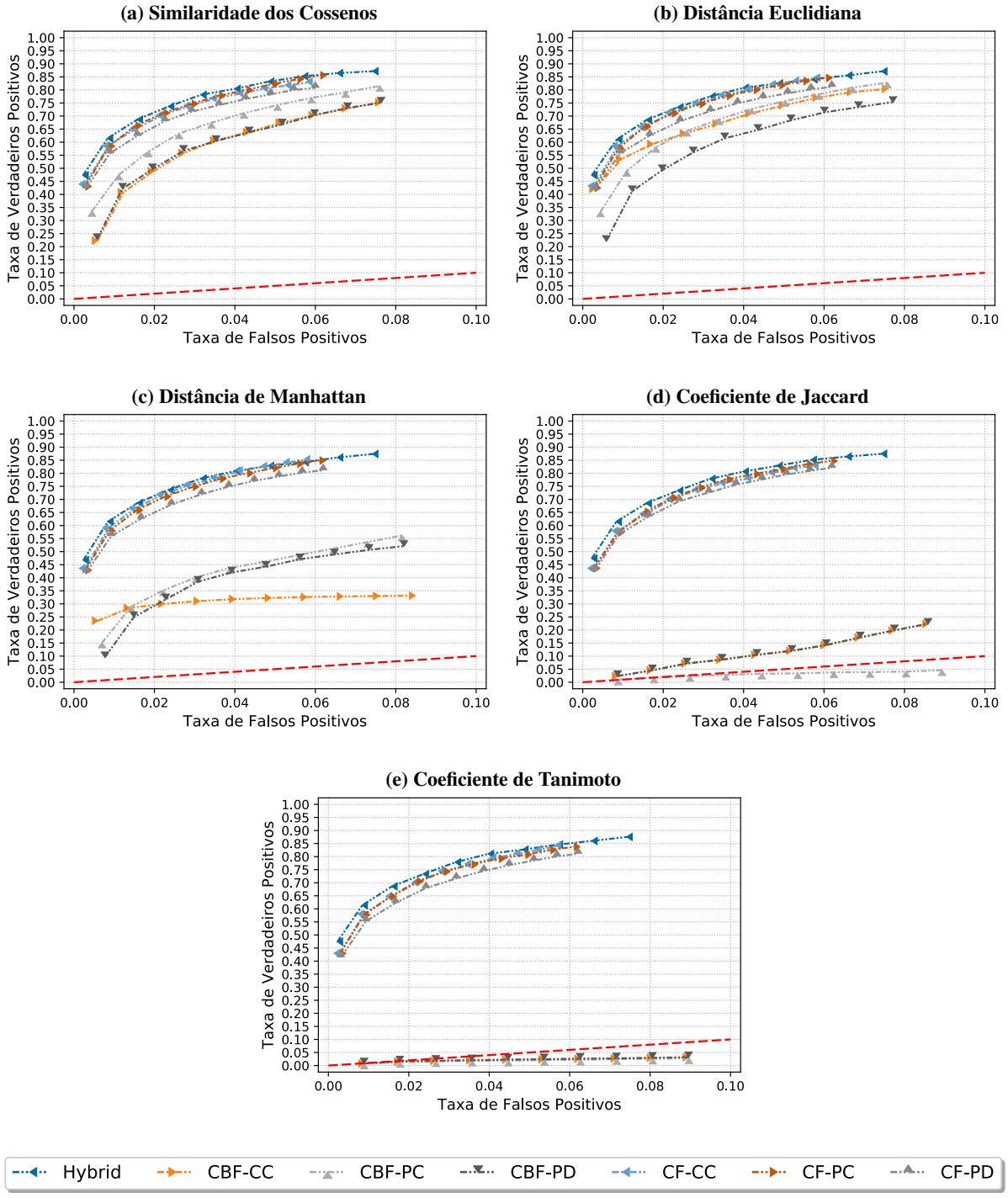
Fonte: Autor.

Figura 21 – Avaliação dos sistemas híbridos de recomendação de procedimentos cirúrgicos através da Curva de Precisão-Revocação.



Fonte: Autor.

Figura 22 – Avaliação dos sistemas híbridos de recomendação de procedimentos cirúrgicos através da Curva ROC.



Fonte: Autor.

Analisando o desempenho dos sistemas híbridos em relação a Curva ROC exibidos na Figura 22, o aumento da Revocação com o aumento da lista de revocação também pode ser constatado. Da mesma forma, também é constatado um aumento da taxa de falsos positivos.

Portanto, o aumento da Revocação e da Taxa de Falsos Positivos em conjunto da diminuição da Precisão são uma consequência do sistema híbrido ser capaz de recomendar mais procedimentos do que os demais modelos individualmente. Este comportamento se da pelo fato de que os modelos apenas recomendam procedimentos de tratamento que possuem uma previsão de avaliação maior que zero ($p(u,i) > 0$), conforme definido nas Seções 4.2.5 e 4.2.6. Deste modo, a combinação dos modelos em um sistema híbrido é capaz de tirar proveito de alguns procedimentos que apenas assumiriam um valor de $p(u, i) > 0$ em um modelo específico e $p(u, i) = 0$ nos demais. Em contrapartida, este mesmo fenômeno faz com que mais procedimentos que não integram o conjunto de validação sejam recomendados, causando as quedas de desempenho apontadas anteriormente.

Finalmente, apesar das quedas de performance em relação à Revocação, à Taxa de Falsos Positivos e, consequentemente, ao F-*Score* dos sistemas híbridos, a sua superioridade aos demais modelos em relação ao $nDCG$ e à Revocação indicam não só que os sistemas híbridos são capazes de recomendar mais procedimentos de tratamento, como eles também são capazes de aumentar a quantidade de procedimentos corretos (que integram o tratamento dos pacientes dos conjuntos de validação) que são recomendados, além de predominantemente recomendá-los antes dos procedimentos incorretos.

6 CONCLUSÃO

Sistemas de Recomendação tem sido evidenciados nas últimas décadas, seja como um campo de pesquisa, ramificado da Inteligência Artificial e Aprendizado de Máquina, ou como um conjunto de ferramentas aplicadas em diversos nichos da indústria (RICCI; ROKACH; SHAPIRA, 2015). Atualmente, SRs podem ser considerados como uma área já bem estabelecida, onde a eficácia de suas técnicas já foi amplamente investigada, sobretudo em estudos e aplicações relacionadas a plataformas de varejo e entretenimento (RICCI; ROKACH; SHAPIRA, 2015).

Estudos mais recentes surgiram buscando analisar o potencial da aplicação de modelos de SR em cenários voltados à saúde (WIESNER; PFEIFER, 2010). O crescimento das discussões nesse sentido resultaram na formalização de em um novo campo de pesquisa, nomeado Sistemas de Recomendação para Saúde (SRSs), do inglês *Health Recommender Systems* (WIESNER; PFEIFER, 2014).

Uma das vertentes de pesquisa em SRS é voltada para produção de recomendações para apoiar profissionais da saúde na tomada de decisão, sobretudo no processo de diagnóstico e tratamento de pacientes. Dentre os principais desafios dos chamados SRS para Profissionais da Saúde, conforme apontado na Seção 2.1, pode-se destacar:

a) A adoção dos chamados Históricos Médicos Eletrônicos (HMEs) no desenvolvimento de SRSs;
b) A construção do perfil do paciente para ser utilizado em um cenário de recomendação na área médica.
c) A escassez de bases de dados médicas públicas disponíveis para pesquisa;
d) A carência de publicações que apresentam análises experimentais de SRSs.

Isto posto, este trabalho apresentou um estudo experimental, avaliando empiricamente o desempenho de modelos tradicionais de SRs em um cenário de recomendação de procedimentos cirúrgicos.

Os experimentos foram realizados utilizando uma base de dados privada, extraída de HMEs, fornecida por um hospital de São Paulo. Esta base de dados apresenta informações pré, intra e pós operatórias de pacientes que sofrem de cardiopatias congênitas.

Os resultados apresentados no Capítulo 5 mostraram que os modelos de FC se sobressaíram em relação aos modelos de Filtragem Baseada em Conteúdo (FBC) no cenário explorado, obtendo melhores resultados em todas as comparações realizadas. Além disso, com base nos dados da Tabela 3, o fato dos modelos de FBC praticamente não apresentarem relevância na composição dos Sistemas de Recomendação Híbridos é outro indicativo de maior adequação dos modelos de FC neste cenário.

Outra comparação apresentada está relacionada com o impacto das métricas de similaridade adotados sobre os modelos no cenário de recomendação proposto. Conforme os resultados apresentados nas Seções 5.2, 5.3, 5.4 e 5.5, as diferenças entre as curvas de desempenho dos modelos de FBC para cada métrica de similaridade são evidentes. Por outro lado, o desempenho

dos modelos de FC que utilizam diferentes métricas de similaridade são muito próximos. Este comportamento indica que os modelos de FBC são sensíveis à métrica de similaridade utilizada, diferentemente dos modelos de FC, que mantém um desempenho estável para todas as métricas de similaridade.

Os Sistemas de Recomendação Sensíveis ao Contexto (SRSCs) aplicados neste estudo mostraram exercer um impacto positivo no desempenho das recomendações pelos seguintes fatores:

a) Ter obtido os melhores resultados dentre todos os modelos de FBCs, atingindo os valores de $nDCG$ médio de 0,6642 e 0,6976 respectivamente nas posições um e dez da lista de recomendações, enquanto o melhor modelo de FBC sem a utilização de contexto atingiu os valores 0,5091 e 0,6567.
b) Ter também obtido os melhores resultados dentre todos os modelos de FC, apesar da diferença ser menos expressiva em comparação com a diferença obtida nos modelos de FBC. Na primeira posição da lista de recomendações, o melhor modelo de FC que utiliza contexto obteve $nDCG = 0{,}6967$, enquanto o melhor modelo de FC que não utiliza contexto obteve $nDCG = 0{,}6914$. Por sua vez, na última posição da lista de recomendações, o modelo de FC atingiu $nDCG = 0.7480$ com contexto e $nDCG = 0.7459$ e sem contexto, no melhor cenário.
c) Ter sido relevante na composição de todas as variações de SR Híbridos, assumindo pesos que variam entre 22,106% e 44,639%, dependendo da métrica de similaridade adotada.

Outro aspecto discutido no Capítulo 5 foram as diferentes características expressadas por cada um dos métodos utilizados na avaliação de SRs. Deste modo, métricas como a Precisão e a Taxa de Falsos Positivos podem ser preferíveis para avaliar modelos em cenários onde recomendações indevidas possuem um risco alto atrelado, como por exemplo em um cenário de recomendação de medicamentos onde o paciente seria o usuário direto do SR. Em contrapartida, em situações onde é mais importante aumentar a quantidade de recomendações relevantes e recomendá-las antecipadamente em relação aos outros conteúdos de menor relevância, métricas como $nDCG$ e Revocação podem ser mais adequadas.

Na perspectiva do $nDCG$ e da Revocação neste cenário, a criação de Sistemas de Recomendação Híbridos resultou em modelos que apresentam melhores resultados, por aumentar a quantidade de recomendações adequadas, assim como por trazer uma maior quantidade delas para as primeiras posições da lista de recomendação, conforme ilustrado nas Figuras 19 e 21. A Tabela 4 apresenta os resultados obtidos pelos modelos híbridos no melhor cenário.

Por outro lado, apesar dos resultados para as primeiras posições da lista de recomendação também indicarem melhorias com os SRs Híbridos, a queda de desempenho em relação à Precisão, Taxa de Falsos Positivos e F-*Score* é mais acentuada com o aumento da quantidade de procedimentos recomendados, conforme apresentado nas Figuras 20, 21 e 22. Isso ocorre pelo fato dos SRs Híbridos serem capazes de recomendar mais itens que os demais modelos, causando um aumento de Falsos Positivos como consequência.

	Posição na Lista de Recomendação	
Métrica de Avaliação	*1*	*10*
$nDCG$	0.7500	0.7834
Revocação	0.4743	0.8754

Tabela 4 – Tabela que apresenta os melhores resultados dos Sistemas de Recomendação Híbridos em relação ao $nDCG$ e à Revocação na primeira e na última posição da lista de recomendações.

A queda de desempenho dos SRSs propostos em relação ao F-Score, principalmente devido ao alto número de Falsos Positivos, mostram um caminho para novos trabalhos explorarem métodos para otimizar a metodologia proposta, ou mesmo criar novas metodologias mais focadas no problema da queda de Precisão. Explorar a definição de um limiar de corte de procedimentos que não atingirem uma adequação predita pré-definida pode ser um caminho para esta otimização. Conforme definido na Seção 4, os procedimentos com valor de $p(u, i) = 0$ foram descartados. Com isso, estudos futuros pode focar na variação deste limiar para analisar o impacto desta parametrização nos resultados dos modelos.

Neste trabalho, foi adotado como padrão ouro os procedimentos de tratamento que foram de fato aplicados aos pacientes. Apesar de ter sido o critério que guiou as avaliações dos modelos neste trabalho, não se pode afirmar que os únicos tratamentos adequados são os que integram o nosso padrão ouro. Possivelmente, outros procedimentos poderiam também ser considerados adequados em determinadas situações, mas os médicos decidiram por outro procedimento, seguindo algum critério técnico que está fora do escopo e das competências dos autores deste trabalho. Portanto, outro aspecto que poderia enriquecer esta análise seria realizar um estudo com profissionais médicos especialistas em cardiopatias congênitas, para avaliar a adequação dos procedimentos recomendados pelos modelos.

Por fim, outra visão de trabalhos futuros é a construção de uma base de conhecimentos focada em cardiopatias congênitas, o que permitiria o desenvolvimento e integração de modelos de Sistemas de Recomendação Baseados em Conhecimento (SRBCs) na metodologia proposta.

REFERÊNCIAS

ACHANANUPARP, Palakorn; WEBER, Ingmar. Extracting food substitutes from food diary via distributional similarity. **arXiv preprint arXiv:1607.08807**, 2016.

ADOMAVICIUS, Gediminas; TUZHILIN, Alexander. Context-aware recommender systems. In: RECOMMENDER systems handbook. [S.l.]: Springer, 2011. p. 217–253.

AGU, Emmanuel; CLAYPOOL, Mark. Cypress: A cyber-physical recommender system to discover smartphone exergame enjoyment. In: PROCEEDINGS of the ACM workshop on engendering health with recommender systems. [S.l.: s.n.], 2016.

ALHARBI, Reham Faisal; BERRI, Jawad; EL-MASRI, Samir. Ontology based clinical decision support system for diabetes diagnostic. In: IEEE. SCIENCE and Information Conference (SAI), 2015. [S.l.: s.n.], 2015. p. 597–602.

ALI, Mumtaz; THANH, Nguyen Dang; VAN MINH, Nguyen et al. A neutrosophic recommender system for medical diagnosis based on algebraic neutrosophic measures. **Applied Soft Computing**, Elsevier, 2017.

ÁLVAREZ, Santiago; SALAZAR, Oscar; OVALLE, Demetrio. User-Centered Ubiquitous Multi-Agent Model for e-Health Web-Based Recommender Applications Development. In: SPRINGER. INTERNATIONAL Conference on Practical Applications of Agents and Multi-Agent Systems. [S.l.: s.n.], 2013. p. 429–440.

AMATRIAIN, Xavier et al. Data mining methods for recommender systems. In: RECOMMENDER systems handbook. [S.l.]: Springer, 2011. p. 39–71.

BATEJA, Ritika; DUBEY, Sanjay Kumar; BHATT, Ashutosh. Health Recommender System and Its Applicability with MapReduce Framework. In: SOFT Computing: Theories and Applications. [S.l.]: Springer, 2018. p. 255–266.

BAYYAPU, Karunakar Reddy; DOLOG, Peter. Tag and neighbour based recommender system for medical events. **Proceedings of MEDEX**, 2010.

BENNETT, James; LANNING, Stan et al. The netflix prize. In: NEW YORK, NY, USA. PROCEEDINGS of KDD cup and workshop. [S.l.: s.n.], 2007. v. 2007, p. 35.

BISSOYI, Swarupananda; MISHRA, Brojo Kishore; PATRA, Manas Ranjan. Recommender systems in a patient centric social network—A survey. In: IEEE. SIGNAL Processing, Communication, Power and Embedded System (SCOPES), 2016 International Conference on. [S.l.: s.n.], 2016. p. 386–389.

BOCANEGRA, Carlos Luis Sanchez et al. HealthRecSys: A semantic content-based recommender system to complement health videos. **BMC medical informatics and decision making**, BioMed Central, v. 17, n. 1, p. 63, 2017.

BÖTTIGER, Ylva et al. SFINX—a drug-drug interaction database designed for clinical decision support systems. **European journal of clinical pharmacology**, Springer, v. 65, n. 6, p. 627–633, 2009.

BRIDGE, Derek et al. Case-based recommender systems. **The Knowledge Engineering Review**, Cambridge University Press, v. 20, n. 3, p. 315–320, 2005.

BURKE, Robin. Hybrid recommender systems: Survey and experiments. **User modeling and user-adapted interaction**, Springer, v. 12, n. 4, p. 331–370, 2002.

_____. Hybrid web recommender systems. In: THE adaptive web. [S.l.]: Springer, 2007. p. 377–408.

_____. Knowledge-based recommender systems. **Encyclopedia of library and information systems**, v. 69, Supplement 32, p. 175–186, 2000.

CHEN, YY; GOH, KN; CHONG, K. Rule based clinical decision support system for hematological disorder. In: IEEE. SOFTWARE Engineering and Service Science (ICSESS), 2013 4th IEEE International Conference on. [S.l.: s.n.], 2013. p. 43–48.

CHO, Jason HD et al. Resolving healthcare forum posts via similar thread retrieval. In: ACM. PROCEEDINGS of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics. [S.l.: s.n.], 2014. p. 33–42.

COVER, Thomas M; HART, Peter E et al. Nearest neighbor pattern classification. **IEEE transactions on information theory**, Menlo Park, v. 13, n. 1, p. 21–27, 1967.

DAVIS, Darcy A et al. Predicting individual disease risk based on medical history. In: ACM. PROCEEDINGS of the 17th ACM conference on Information and knowledge management. [S.l.: s.n.], 2008. p. 769–778.

DAVIS, Jesse; GOADRICH, Mark. The relationship between Precision-Recall and ROC curves. In: ACM. PROCEEDINGS of the 23rd international conference on Machine learning. [S.l.: s.n.], 2006. p. 233–240.

DEAN, Jeffrey; GHEMAWAT, Sanjay. MapReduce: simplified data processing on large clusters. **Communications of the ACM**, ACM, v. 51, n. 1, p. 107–113, 2008.

DHEERU, Dua; KARRA TANISKIDOU, Efi. **UCI Machine Learning Repository**. [S.l.: s.n.], 2017. Accessed: 2018-08-16. Disponível em: <http://archive.ics.uci.edu/ml>.

DIETTERICH, Thomas G. Ensemble methods in machine learning. In: SPRINGER. INTERNATIONAL workshop on multiple classifier systems. [S.l.: s.n.], 2000. p. 1–15.

DOUALI, Nassim et al. Clinical decision support system based on fuzzy cognitive maps. **Journal of Computer Science & Systems Biology**, OMICS Publishing Group, v. 8, n. 2, p. 112, 2015.

DOURISH, Paul. What we talk about when we talk about context. **Personal and ubiquitous computing**, Springer-Verlag, v. 8, n. 1, p. 19–30, 2004.

DUAN, Lian; STREET, W Nick; XU, E. Healthcare information systems: data mining methods in the creation of a clinical recommender system. **Enterprise Information Systems**, Taylor & Francis, v. 5, n. 2, p. 169–181, 2011.

EKSTRAND, Jennifer D; EKSTRAND, Michael D. First do no harm: Considering and minimizing harm in recommender systems designed for engendering health. In: ACM. ENGENDERING Health Workshop at the RecSys 2016 Conference. [S.l.: s.n.], 2016. p. 1–2.

FELFERNIG, Alexander et al. Developing constraint-based recommenders. In: RECOMMENDER systems handbook. [S.l.]: Springer, 2011. p. 187–215.

FERNANDES, Ana S et al. A clinical decision support system for breast cancer patients. In: SPRINGER. DOCTORAL Conference on Computing, Electrical and Industrial Systems. [S.l.: s.n.], 2010. p. 122–129.

GARG, Amit X et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. **Jama**, American Medical Association, v. 293, n. 10, p. 1223–1238, 2005.

GE, Mouzhi; RICCI, Francesco; MASSIMO, David. Health-aware food recommender system. In: ACM. PROCEEDINGS of the 9th ACM Conference on Recommender Systems. [S.l.: s.n.], 2015. p. 333–334.

GE, Mouzhi et al. Using tags and latent factors in a food recommender system. In: ACM. PROCEEDINGS of the 5th International Conference on Digital Health 2015. [S.l.: s.n.], 2015. p. 105–112.

GOLDBERG, David et al. Using collaborative filtering to weave an information tapestry. **Communications of the ACM**, Association for Computing Machinery, Inc., v. 35, n. 12, p. 61–71, 1992.

GRÄSSER, Felix et al. Neighborhood-based Collaborative Filtering for Therapy Decision Support. In: SECOND International Workshop on Health Recommender Systems. [S.l.: s.n.], 2017. p. 22.

GUDADHE, Mrudula; WANKHADE, Kapil; DONGRE, Snehlata. Decision support system for heart disease based on support vector machine and artificial neural network. In: IEEE. COMPUTER and Communication Technology (ICCCT), 2010 International Conference on. [S.l.: s.n.], 2010. p. 741–745.

GUNAWARDANA, Asela; SHANI, Guy. Evaluating recommender systems. In: RECOMMENDER systems handbook. [S.l.]: Springer, 2015. p. 265–308.

HAMED, Ahmed Abdeen et al. T-Recs: Time-aware Twitter-based Drug Recommender System. In: IEEE. ADVANCES in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on. [S.l.: s.n.], 2012. p. 1027–1031.

HARRIS, Sarah; HARRIS, David. **Digital design and computer architecture: arm edition**. [S.l.]: Morgan Kaufmann, 2015.

HASSAN, Shahzaib; SYED, Zeeshan. From netflix to heart attacks: collaborative filtering in medical datasets. In: ACM. PROCEEDINGS of the 1st ACM international health informatics symposium. [S.l.: s.n.], 2010. p. 128–134.

HERBERT, Joseph P; YAO, Jingtao. Game-theoretic rough sets. **Fundamenta Informaticae**, IOS Press, v. 108, n. 3-4, p. 267–286, 2011.

HORS-FRAILE, Santiago et al. Analyzing recommender systems for health promotion using a multidisciplinary taxonomy: A scoping review. **International journal of medical informatics**, Elsevier, 2017.

HORS-FRAILE, Santiago et al. Design of two combined health recommender systems for tailoring messages in a smoking cessation app. **arXiv preprint arXiv:1608.07192**, 2016.

HU, Chenhui et al. Clinical decision support for Alzheimer's disease based on deep learning and brain network. In: IEEE. COMMUNICATIONS (ICC), 2016 IEEE International Conference on. [S.l.: s.n.], 2016. p. 1–6.

HUNT, Dereck L et al. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. **Jama**, American Medical Association, v. 280, n. 15, p. 1339–1346, 1998.

HUSSAIN, Sajjad; ABIDI, Samina Raza; ABIDI, Syed Sibte Raza. Semantic web framework for knowledge-centric clinical decision support systems. In: SPRINGER. CONFERENCE on artificial intelligence in medicine in europe. [S.l.: s.n.], 2007. p. 451–455.

HUSSEIN, Asmaa S et al. Smart collaboration framework for managing chronic disease using recommender system. **Health Systems**, Springer, v. 3, n. 1, p. 12–17, 2014.

JASPERS, Monique WM et al. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. **Journal of the American Medical Informatics Association**, BMJ Group, v. 18, n. 3, p. 327–334, 2011.

JIANG, Hongxun; XU, Wei. How to find your appropriate doctor: An integrated recommendation framework in big data context. In: IEEE. COMPUTATIONAL Intelligence in Healthcare and e-health (CICARE), 2014 IEEE Symposium on. [S.l.: s.n.], 2014. p. 154–158.

KARIM, Shazia; BAJWA, Imran Sarwar. Clinical decision support system based virtual telemedicine. In: IEEE. INTELLIGENT Human-Machine Systems and Cybernetics (IHMSC), 2011 International Conference on. [S.l.: s.n.], 2011. v. 1, p. 16–21.

KAWAMOTO, Kensaku et al. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. **Bmj**, British Medical Journal Publishing Group, v. 330, n. 7494, p. 765, 2005.

KOMKHAO, Maytiyanin; HALANG, Wolfgang A. Recommender systems in telemedicine. **IFAC Proceedings Volumes**, Elsevier, v. 46, n. 28, p. 28–33, 2013.

KOMKHAO, Maytiyanin; LU, Jie; ZHANG, Lichen. Determining pattern similarity in a medical recommender system. In: DATA and Knowledge Engineering. [S.l.]: Springer, 2012. p. 103–114.

KONDYLAKIS, Haridimos et al. PMIR: A Personal Medical Information Recommender. In: MIE. [S.l.: s.n.], 2014. p. 1193.

KONDYLAKIS, Haridimos et al. Semantically-enabled Personal Medical Information Recommender. In: INTERNATIONAL Semantic Web Conference (Posters & Demos). [S.l.: s.n.], 2015.

KUPERMAN, Gilad J et al. Medication-related clinical decision support in computerized provider order entry systems: a review. **Journal of the American Medical Informatics Association**, BMJ Group BMA House, Tavistock Square, London, WC1H 9JR, v. 14, n. 1, p. 29–40, 2007.

LAGE, Ricardo Gomes et al. Applicability of recommender systems to medical surveillance systems. In: ACM. PROCEEDINGS of the second international workshop on Web science and information exchange in the medical web. [S.l.: s.n.], 2011. p. 1–6.

LI, Wei; JONES, Gareth JF. Enhancing Medical Information Retrieval by Exploiting a Content-Based Recommender Method. In: SPRINGER. INTERNATIONAL Conference of the Cross-Language Evaluation Forum for European Languages. [S.l.: s.n.], 2015. p. 142–153.

LIKERT, Rensis. A technique for the measurement of attitudes. **Archives of psychology**, 1932.

LIU, Ximeng et al. Privacy-preserving patient-centric clinical decision support system on naive Bayesian classification. **IEEE journal of biomedical and health informatics**, IEEE, v. 20, n. 2, p. 655–668, 2016.

LÓPEZ-NORES, Martín et al. Property-based collaborative filtering for health-aware recommender systems. **Expert Systems with Applications**, Elsevier, v. 39, n. 8, p. 7451–7457, 2012.

LOPS, Pasquale; DE GEMMIS, Marco; SEMERARO, Giovanni. Content-based recommender systems: State of the art and trends. In: RECOMMENDER systems handbook. [S.l.]: Springer, 2011. p. 73–105.

LORENZI, Fabiana; RICCI, Francesco. Case-based recommender systems: A unifying view. In: SPRINGER. IJCAI Workshop on Intelligent Techniques for Web Personalization. [S.l.: s.n.], 2003. p. 89–113.

MACEDO, Alessandra Alaniz et al. A Health Surveillance Software Framework to deliver information on preventive healthcare strategies. **Journal of biomedical informatics**, Elsevier, v. 62, p. 159–170, 2016.

MAGO, Vijay Kumar et al. Clinical decision support system for dental treatment. **Journal of Computational Science**, Elsevier, v. 3, n. 5, p. 254–261, 2012.

MARCO-RUIZ, Luis et al. Publication, discovery and interoperability of clinical decision support systems: a linked data approach. **Journal of biomedical informatics**, Elsevier, v. 62, p. 243–264, 2016.

MARTÍNEZ-ROMERO, Marcos et al. NCBO Ontology Recommender 2.0: an enhanced approach for biomedical ontology recommendation. **Journal of biomedical semantics**, BioMed Central, v. 8, n. 1, p. 21, 2017.

MBANDU, Angolo Shem et al. Recommender Privacy Preserving Reputation Based Medical Services Scheme Using a Variant of ElGamal. **International Journal of Computer and Communication Engineering**, IACSIT Press, v. 4, n. 4, p. 234, 2015.

MONTEIRO, Eriksson J Melicio et al. A recommender system for medical imaging diagnostic. In: MIE. [S.l.: s.n.], 2015. p. 461–463.

NACHAWATI, M Omar et al. Social Sifter: An Agent-Based Recommender System to Mine the Social Web. In: CITESEER. STIDS. [S.l.: s.n.], 2012. p. 125–128.

NING, Xia; DESROSIERS, Christian; KARYPIS, George. A comprehensive survey of neighborhood-based recommendation methods. In: RECOMMENDER systems handbook. [S.l.]: Springer, 2015. p. 37–76.

O'CONNOR, Patrick J et al. Impact of electronic health record clinical decision support on diabetes care: a randomized trial. **The Annals of Family Medicine**, Annals Family Med, v. 9, n. 1, p. 12–21, 2011.

OLUWAGBEMI, Olugbenga; OLUWAGBEMI, Folakemi; ABIMBOLA, Oluyemi. Ebinformatics: Ebola fuzzy informatics systems on the diagnosis, prediction and recommendation of appropriate treatments for Ebola virus disease (EVD). **Informatics in Medicine Unlocked**, Elsevier, v. 2, p. 12–37, 2016.

PATRO, S; SAHU, Kishore Kumar. Normalization: A preprocessing stage. **arXiv preprint arXiv:1503.06462**, 2015.

RESNICK, Paul; VARIAN, Hal R. Recommender systems. **Communications of the ACM**, Association for Computing Machinery, Inc., v. 40, n. 3, p. 56–59, 1997.

RICCI, Francesco; ROKACH, Lior; SHAPIRA, Bracha. Recommender systems: introduction and challenges. In: RECOMMENDER systems handbook. [S.l.]: Springer, 2015. p. 1–34.

SCHÄFER, Hanna et al. Towards health (aware) recommender systems. In: ACM. PROCEEDINGS of the 2017 international conference on digital health. [S.l.: s.n.], 2017. p. 157–161.

SCHAFER, J Ben; KONSTAN, Joseph; RIEDL, John. Recommender systems in e-commerce. In: ACM. PROCEEDINGS of the 1st ACM conference on Electronic commerce. [S.l.: s.n.], 1999. p. 158–166.

SCHAFER, J Ben et al. Collaborative filtering recommender systems. In: THE adaptive web. [S.l.]: Springer, 2007. p. 291–324.

SEZGIN, Emre; OZKAN, Sevgi. A systematic literature review on Health Recommender Systems. In: IEEE. E-HEALTH and Bioengineering Conference (EHB), 2013. [S.l.: s.n.], 2013. p. 1–4.

SITTIG, Dean F et al. Grand challenges in clinical decision support. **Journal of biomedical informatics**, Elsevier, v. 41, n. 2, p. 387–392, 2008.

SODSEE, Sunantha; KOMKHAO, Maytiyanin. Evidence-based medical recommender systems: A review. **Int J Inf Process Manag**, v. 4, n. 6, p. 114–20, 2013.

SUBBALAKSHMI, G; RAMESH, K; RAO, M Chinna. Decision support in heart disease prediction system using naive bayes. **Indian Journal of Computer Science and Engineering (IJCSE)**, v. 2, n. 2, p. 170–176, 2011.

SUGANTHI, Muthusamy; MADHESWARAN, Muthusamy. An improved medical decision support system to identify the breast cancer using mammogram. **Journal of medical systems**, Springer, v. 36, n. 1, p. 79–91, 2012.

SUN, Leilei et al. Data-driven automatic treatment regimen development and recommendation. In: ACM. PROCEEDINGS of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S.l.: s.n.], 2016. p. 1865–1874.

THANH, Nguyen Dang; ALI, Mumtaz et al. A novel clustering algorithm in a neutrosophic recommender system for medical diagnosis. **Cognitive Computation**, Springer, v. 9, n. 4, p. 526–544, 2017.

THONG, Nguyen Tho et al. HIFCF: An effective hybrid model between picture fuzzy clustering and intuitionistic fuzzy recommender systems for medical diagnosis. **Expert Systems with Applications**, Elsevier, v. 42, n. 7, p. 3682–3701, 2015.

_____. Intuitionistic fuzzy recommender systems: an effective tool for medical diagnosis. **Knowledge-Based Systems**, Elsevier, v. 74, p. 133–150, 2015.

VAGHELA, Chaitali; BHATT, Nikita; MISTRY, Darshana. A survey on various classification techniques for clinical decision support system. **International Journal of Computer Applications**, Foundation of Computer Science, v. 116, n. 23, 2015.

VALDEZ, André Calero et al. Recommender systems for health informatics: State-of-the-art and future perspectives. In: MACHINE Learning for Health Informatics. [S.l.]: Springer, 2016. p. 391–414.

WIESNER, Martin; PFEIFER, Daniel. Adapting recommender systems to the requirements of personal health record systems. In: ACM. PROCEEDINGS of the 1st ACM International Health Informatics Symposium. [S.l.: s.n.], 2010. p. 410–414.

_____. Health recommender systems: concepts, requirements, technical basics and challenges. **International journal of environmental research and public health**, Multidisciplinary Digital Publishing Institute, v. 11, n. 3, p. 2580–2607, 2014.

YANG, Sen et al. A Data-driven Process Recommender Framework. In: ACM. PROCEEDINGS of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S.l.: s.n.], 2017. p. 2111–2120.

YAO, JingTao; AZAM, Nouman. Web-based medical decision support systems for three-way medical decision making with game-theoretic rough sets. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 23, n. 1, p. 3–15, 2015.

YOON, Jinsung; DAVTYAN, Camelia; SCHAAR, Mihaela van der. Discovery and clinical decision support for personalized healthcare. **IEEE journal of biomedical and health informatics**, IEEE, v. 21, n. 4, p. 1133–1145, 2017.

ZHANG, Yutao et al. Leap: Learning to prescribe effective and safe treatment combinations for multimorbidity. In: ACM. PROCEEDINGS of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S.l.: s.n.], 2017. p. 1315–1324.

ZUCCOTTI, G et al. Reducing risk with clinical decision support. **Applied clinical informatics**, Schattauer GmbH, v. 5, n. 03, p. 746–756, 2014.