



**UM SISTEMA IMUNO-INSPIRADO
DE RECOMENDAÇÃO DE DISCIPLINAS
PARA CURSOS SUPERIORES**

JÚLIA CABRAL DINIZ BRAZ

**DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**UM SISTEMA IMUNO-INSPIRADO
DE RECOMENDAÇÃO DE DISCIPLINAS
PARA CURSOS SUPERIORES**

JÚLIA CABRAL DINIZ BRAZ

Orientador: PROF. DR. DANIEL GUERREIRO E SILVA, ENE/UNB

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA

**PUBLICAÇÃO PPGENE.DM - PPGENE.DM 789/22
BRASÍLIA-DF, 27 DE JUNHO DE 2022.**

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**UM SISTEMA IMUNO-INSPIRADO
DE RECOMENDAÇÃO DE DISCIPLINAS
PARA CURSOS SUPERIORES**

JÚLIA CABRAL DINIZ BRAZ

DISSERTAÇÃO DE MESTRADO ACADÊMICO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA.

BANCA EXAMINADORA:

Prof. Dr. Daniel Guerreiro e Silva, ENE/UnB
Orientador

Prof. Dr. Francisco Assis, ENE/UnB
Examinador interno

Prof. Dr. Hugo Valadares Siqueira, DAELE/UTFPR
Examinador externo

BRASÍLIA, 27 DE JUNHO DE 2022.

FICHA CATALOGRÁFICA

JÚLIA CABRAL DINIZ BRAZ

Um sistema imuno-inspirado de recomendação de disciplinas para cursos superiores

2022xv, 57p., 201x297 mm

(ENE/FT/UnB, Mestre, Engenharia Elétrica, 2022)

Dissertação de Mestrado - Universidade de Brasília

Faculdade de Tecnologia - Departamento de Engenharia Elétrica

REFERÊNCIA BIBLIOGRÁFICA

JÚLIA CABRAL DINIZ BRAZ (2022) Um sistema imuno-inspirado de recomendação de disciplinas para cursos superiores. Dissertação de Mestrado em Engenharia Elétrica, Publicação PPGENE.DM 789/22, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 57p.

CESSÃO DE DIREITOS

AUTOR: JÚLIA CABRAL DINIZ BRAZ

TÍTULO: Um sistema imuno-inspirado de recomendação de disciplinas para cursos superiores.

GRAU: Mestre ANO: 2022

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de Mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor se reserva a outros direitos de publicação e nenhuma parte desta dissertação de Mestrado pode ser reproduzida sem a autorização por escrito do autor.

JÚLIA CABRAL DINIZ BRAZ

SQN 404 Bl. I Apto 201

Asa Norte

CEP: 70845-090 - Brasília/DF - Brasil

Agradecimentos

Agradeço aos meus pais, Aline e Rubens, que sempre me apoiaram em tudo, mesmo que as vezes de maneiras diferentes das quais eu esperava. Agradeço a minha irmã, Clara, que está sempre comigo e que, além de irmã, sempre será minha melhor amiga. Agradeço ao meu irmão que, mesmo tão diferente de mim, sempre me apoia. Agradeço ao meu grande amigo, Miguel Pachá pelo suporte e por sempre me ouvir. Por fim, agradeço ao meu orientador, Prof. Dr. Daniel Guerreiro e Silva, por ter me apoiado durante essa jornada.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)

Resumo

Sistemas de recomendação podem ser aplicados no meio acadêmico para auxiliar na escolha de disciplinas que se adaptem ao perfil de cada aluno. Isso porque algumas instituições tentam manter um currículo aberto para que haja flexibilidade de formações e interdisciplinaridade, em outras palavras, o aluno pode escolher quais disciplinas vai cursar, levando em consideração se ele acha relevante a seus interesses. Nesta dissertação apresentamos os fundamentos de duas classes importantes de métodos de recomendação: recomendação baseada em conteúdo e filtragem colaborativa. O sistema imunológico adquirido dos vertebrados é usado como inspiração para meta-heurísticas por sua robustez, auto-organização, e sua capacidade de aprendizado. Foi proposta a aplicação de um algoritmo de filtragem colaborativa baseado em um sistema imunológico artificial (SIA) ao problema de recomendar disciplinas a alunos de graduação. Sistemas imunológicos artificiais tem algumas características que os tornam adequados para implementação de sistema de recomendação. São elas: correspondência, que se refere à ligação entre anticorpos e antígenos, ao qual se pode traçar um paralelo entre o cálculo de semelhança de entre o alvo e outros usuários; diversidade, que se refere ao fato de que, para atingir a cobertura ideal do espaço de variações do antígeno, a diversidade de anticorpos deve ser incentivada, isso pode ser traduzido no sistema de recomendação como a capacidade de evitar a especificidade da recomendação; e memória, capacidade de "aprendizado", o sistema imunológico é capaz de lembrar de anticorpos capazes de combater antígenos específicos, esta característica pode que facilitar a geração de novas recomendações. Por fim, foram conduzidos experimentos para otimizar parâmetros do algoritmo e avaliar o desempenho geral da abordagem proposta. Para a análise foram utilizados dados históricos de alunos de Engenharia de Redes de Comunicação da Universidade de Brasília. Os resultados do algoritmo proposto mostram um expressivo aumento da quantidade de recomendações em comparação com uma técnica básica de filtragem, aliado a uma maior precisão e diversidade das recomendações.

Abstract

Recommender systems can be applied in the academic environment to assist in choosing subjects that adapt to the profile of each student. This is because some institutions try to keep an open curriculum, in other words, the student can choose which subjects to take, taking into account if he/she finds it relevant to his/her interests. In this dissertation we present the fundamentals of two important classes of recommendation methods: content-based recommendation and collaborative filtering. The immune system of vertebrates is used as inspiration for metaheuristics for its robustness, self-organization, and its ability to learn. It was proposed to apply a collaborative filtering algorithm based on an artificial immune system (AIS) to the problem of recommending courses to undergraduate students. Artificial immune systems have some characteristics that make them suitable for recommender system implementation. They are: correspondence, which refers to the link between antibodies and antigens, to which a parallel can be drawn between the calculation of similarity between the target of recommendation and other users; diversity, which refers to the fact that, in order to achieve the ideal coverage of the space of antigen variations, the diversity of antibodies must be encouraged, in a recommendation system this characteristic can avoid the specificity of the recommendation; and memory, "learning" capacity, the immune system is capable of remembering antibodies capable of fighting specific antigens, thinking about the recommendation system, this characteristic can facilitate the generation of new recommendations. Experiments were conducted to optimize algorithm parameters and evaluate the overall performance of the proposed approach. For the analysis, historical data from students of Communication Network Engineering at the University of Brasília were used. The results of the proposed algorithm show a significant increase in the number of recommendations compared to the basic filtering technique, combined with a great accuracy and diversity of recommendations.

SUMÁRIO

AGRADECIMENTOS	I
RESUMO	II
ABSTRACT	III
1 INTRODUÇÃO	1
1.1 OBJETIVOS.....	3
1.2 JUSTIFICATIVA	3
1.3 ORGANIZAÇÃO DA DISSERTAÇÃO	4
2 FUNDAMENTAÇÃO TEÓRICA	5
2.1 SISTEMAS DE RECOMENDAÇÃO	5
2.1.1 RECOMENDAÇÃO DE DISCIPLINAS.....	8
2.1.2 RECOMENDAÇÃO EXPLICÁVEL	9
2.1.3 ABORDAGENS DE RECOMENDAÇÃO	10
2.2 SISTEMA IMUNOLÓGICO ARTIFICIAL	17
2.2.1 PRINCÍPIOS DO SISTEMA IMUNOLÓGICO	18
2.2.2 APLICAÇÕES DO SISTEMA IMUNOLÓGICO ARTIFICIAL	22
2.3 TRABALHOS RELACIONADOS	23
2.4 CONSIDERAÇÕES FINAIS DO CAPÍTULO	25
3 METODOLOGIA	27
3.1 CONJUNTO DE DADOS	28
3.2 ALGORITMO PROPOSTO	29
3.2.1 CÁLCULO DA SIMILARIDADE.....	32
3.2.2 PREDIÇÃO DA NOTA	32
3.2.3 SELEÇÃO DE DISCIPLINAS	33
3.3 MÉTRICAS DE AVALIAÇÃO DO DESEMPENHO	33
3.4 CONSIDERAÇÕES FINAIS DO CAPÍTULO	35
4 RESULTADOS EXPERIMENTAIS	36
4.1 TAXA DE ESTÍMULO	37
4.2 TAXA DE SUPRESSÃO	40

4.3	DESEMPENHO DAS RECOMENDAÇÕES	43
4.4	ANALISE DA RECOMENDAÇÃO	47
5	CONCLUSÃO	49
5.1	TRABALHOS FUTUROS	50
5.2	TRABALHOS PUBLICADOS	50
	REFERÊNCIAS BIBLIOGRÁFICAS.....	51

LISTA DE FIGURAS

2.1	Esquema básico do processo de um sistema de recomendação	6
2.2	Esquema da recomendação baseada em conteúdo.....	11
2.3	Esquema da filtragem colaborativa.....	12
2.4	Filtragem colaborativa baseada em usuário	14
2.5	Filtragem colaborativa baseada em item.....	15
2.6	Processo de ativação do linfócito B	19
2.7	Processo de marcação do antígeno pelo anticorpo	20
2.8	Estabilização da rede imunológica.	21
3.1	Parte do conjunto de dados utilizado	29
4.1	Efeito a taxa de estímulo no número de alunos visitados.	37
4.2	Efeito da taxa de estímulo no número de recomendações.	38
4.3	Efeito da taxa de estímulo no erro absoluto médio.	39
4.4	Efeito da taxa de estímulo na taxa de sobreposição da recomendação.	39
4.5	Efeito da taxa de supressão no número de anticorpos na rede.	40
4.6	Efeito da taxa de supressão no número de alunos visitados.	41
4.7	Efeito da taxa de supressão no erro absoluto médio.....	41
4.8	Efeito da taxa de supressão no número de recomendações.	42
4.9	Efeito da taxa de supressão na taxa de sobreposição da recomendação.	42
4.10	Relação entre numero de disciplinas cursadas e MAE	44
4.11	Relação entre numero de disciplinas cursadas e o número de disciplinas re- comendadas	44
4.12	Relação entre numero de disciplinas cursadas e a precisão da recomendação... ..	45
4.13	Relação entre numero de disciplinas cursadas e a taxa de sobreposição da recomendação	45
4.14	Relação entre numero de disciplinas cursadas e a precisão da lista top-5	46
4.15	Relação entre numero de disciplinas cursadas e a taxa de sobreposição da lista top-5	47

LISTA DE TABELAS

2.1	Aplicações dos princípios do SI em problemas reais	22
3.1	Exemplo da matriz usuário-item.	30
4.1	Resultados da recomendação usando SIA.	48
4.2	Resultados da recomendação usando k -NN	48

Capítulo 1

Introdução

Recomendar é uma ação corriqueira. É comum sugerir todo tipo de coisa a amigos ou colegas, desde filmes e livros até atividades de lazer ou disciplinas de faculdade [East et al. 2005]. Sugestões são especialmente úteis diante de um grande quantidade de alternativas, a qual impossibilita uma avaliação cuidadosa de todas as possibilidades. Recomendações, portanto, restringem o universo de opções a um conjunto menor, funcionando como um filtro.

Como as pessoas fazem recomendações? Primeiramente, quando alguém se propõe a fazer uma recomendação, é normal levar em conta as preferências já conhecidas da pessoa a quem se recomenda. Assim, por exemplo, se sabemos que tal amigo gosta de filmes românticos, é natural que indiquemos filmes semelhantes. Um segundo critério que serve como base para recomendações é a semelhança entre as pessoas. Continuando com o exemplo dos filmes, este critério consiste em usar avaliações de pessoas com gostos parecidos para encontrar sugestões. Perceba que o primeiro critério se apóia na semelhança dos itens a serem recomendados, enquanto o segundo parte da semelhança dos indivíduos a quem se faz a sugestão [Das et al. 2017]. Em ambos os casos, a sugestão é em alguma medida explicável, ou seja: é possível motivar a recomendação de forma a facilitar a avaliação por quem a recebe [Zhang and Chen 2020].

Assim, para “imitar” este comportamento um sistema de recomendação deve ser capaz de entender as preferências do usuário e procurar por itens relevantes a ele. O uso desses sistemas tem crescido constantemente nos últimos anos, principalmente por empresas de varejo e comércio eletrônico, pois necessitam lidar com a sobrecarga de informações, ou seja, há um número muito grande de itens para se verificar a fim de gerar uma recomendação [Tsagkias et al. 2021].

Atualmente, a quantidade de opções de disciplinas oferecidas em cursos superiores de universidades aumentou [Lynn and Emanuel 2020]. Isso porque algumas instituições tentam manter um currículo aberto para que haja flexibilidade de formações e interdisciplinaridade, em outras palavras, o aluno pode escolher quais disciplinas vai cursar, levando em conside-

ração se ele acha relevante a seus interesses. Algumas dessas disciplinas são ofertadas na condição de eletivas, oferecendo uma especialização em uma certa área do curso. Então, alunos com listas extensas de opções podem gastar muito tempo procurando informações sobre as diferentes disciplinas disponíveis para escolher aquelas que melhor se encaixam em seus interesses pessoais e na jornada acadêmica. A escolha das disciplinas, portanto, pode não ser trivial, deste modo, a utilização de um sistema de recomendação pode ser indicada. Existem vários estudos da aplicação deste sistema no problema da seleção de disciplinas. Os métodos utilizados podem ser categorizados em três grupos diferentes, que são a filtragem colaborativa [Madani et al. 2019], baseados em conteúdo [Wang et al. 2018] e baseados em conhecimento [Saraswathi et al. 2014].

Mas um problema pouco explorado nesse contexto é a especificidade da recomendação, que é quando os itens da lista resultante acabam sendo muito semelhantes entre si. Então, a diversidade na recomendação deve ser considerada pois indivíduos tendem a ter interesses e habilidades diversas e uma lista redundante pode não ser de muito auxílio ao aluno.

Uma tendência consolidada é a utilização da inteligência computacional para implementar estes sistemas, pois tem grande capacidade de melhorar a qualidade da recomendação. Alguns trabalhos usam estas estratégias para enfrentar os problemas clássicos dos métodos de recomendação [Zhang et al. 2020].

Dentro de Inteligência Computacional, há a utilização de meta-heurísticas bio-inspiradas, que são adequadas para aplicações como planejamento, design, controle e classificação. Estas técnicas têm sua origem em diversos contextos, como a psicologia, biologia, física e neurologia [Voß 2000] sendo os métodos evolutivos e inteligência de enxame uns dos mais utilizados. Alguns representantes da classe métodos evolutivos são: as estratégias evolutivas, a programação evolutiva e os algoritmos genéticos [Corne and Lones 2018]. E de inteligência de exames os principais representantes são: enxame de partículas, colônias de formigas e colônias de abelhas artificiais [Nguyen et al. 2020]

Não obstante, há trabalhos que se utilizam destas meta-heurísticas para geração de recomendações. [Karabadjji et al. 2018] se utiliza de um algoritmo evolutivo para melhorar um algoritmo de filtragem colaborativa, com o objetivo de fazer recomendações que ao mesmo tempo sejam diversas (sem muitos itens semelhantes entre si) e precisas. Já [Esteban et al. 2020] propõe um sistema de recomendação híbrido que combina metodologias de Filtragem Colaborativa e Baseada em Conteúdo usando multi-critérios relacionados às informações do aluno e do curso para recomendar os cursos mais adequados aos alunos. Um Algoritmo Genético (GA) foi desenvolvido para descobrir automaticamente a configuração ótima do sistema de recomendação, que inclui tanto os critérios mais relevantes quanto a configuração dos demais parâmetros. Além disso, [Peška et al. 2019] fazem uma resumo dos principais trabalhos que aplicam inteligência de exames em sistemas de recomendação.

Outros algoritmos bio-inspirados têm recebido atenção nas últimas décadas. As técnicas inspiradas no sistema imunológico, por exemplo, possuem algumas propriedades in-

interessantes para apoiar a construção de sistemas computacionais inteligentes. São elas: robustez, tolerância, distribuição, memória, adaptabilidade, auto-organização e autonomia [Figueredo et al. 2013]. Dasgupta [Dasgupta et al. 2011] define os sistemas imunológicos artificiais como “sistemas adaptativos, inspirados na imunologia teórica e nas funções, princípios e modelos imunológicos observados, que são aplicados à resolução de problemas”.

[Chang et al. 2016] aplicam um algoritmo baseado no sistema imunológico para gerar as recomendações. Ele é usado para criar *clusters* de alunos que serão usadas na filtragem colaborativa para combinar um aluno alvo com grupos de alunos semelhantes.

Mas um dos processos do sistema imunológico que mais interessa para sistemas de recomendação é a auto-regulação da rede imunológica, que se refere ao fato de que, para atingir a cobertura ideal do espaço de variações do antígeno, a diversidade de anticorpos deve ser incentivada. Em um sistema de recomendação essa característica pode se traduzir na capacidade de gerar recomendações diversas. Será este o conceito explorado neste trabalho.

1.1 Objetivos

O presente trabalho propõe uma implementação de um sistema imuno-inspirado de recomendação, tomando como base um banco de dados com informações acadêmicas de alunos do curso de Engenharia de Redes de Comunicação, da Universidade de Brasília.

O sistema utiliza a análise do histórico acadêmico de um aluno alvo, i.e. suas notas em matérias previamente cursadas, para encontrar um grupo de alunos semelhantes a ele (com histórico parecido) mas diversos entre si. Deste grupo serão extraídas matérias optativas que serão recomendadas ao aluno. Ademais, serão utilizadas métricas para avaliação da recomendação condizentes com as utilizadas na literatura, para fins de comparação.

1.2 Justificativa

A utilização do sistema de recomendação pode servir para auxiliar o aluno a selecionar disciplinas de acordo com suas habilidades, interesses e desempenho em disciplinas cursadas anteriormente. Mas o impacto sugestões diversas ainda é pouco explorado no contexto da recomendação de disciplinas. Esta característica pode afetar os resultados da recomendação de maneira positiva, pois alunos tendem a ter interesses e habilidades diversas, não se assemelhando, a apenas um grupo de pessoas específico.

Assim a pesquisa apresentada neste trabalho tem a intenção de explorar melhor as possibilidades de recomendação aplicando o conceito de rede imunológica regulada pelo mecanismo de concentração de anticorpos, de modo a analisar o impacto na recomendação e a abrangência das disciplinas recomendadas. A utilização de sistemas imunológicos artificiais contribui com a pesquisa em sistemas de recomendação pois a diversidade é intrínseca

a esta classe de algoritmos e sua explicabilidade é mais simples, tendo a vantagem de dar melhor fundamento à sua saída produzida, o que por sua vez, resulta em um aumento da confiabilidade no sistema.

1.3 Organização da Dissertação

O restante da dissertação está organizado da seguinte maneira: Capítulo 2 apresenta os fundamentos teóricos empregados para a aplicação do sistema imunológico artificial no problema de recomendação de disciplinas. O Capítulo 3 apresenta o algoritmo proposto, os aprimoramentos aplicados e as respectivas métricas de desempenho consideradas para avaliar a qualidade da recomendação. O Capítulo 4 traz os resultados obtidos, junto a comparações com um sistema básico de filtragem colaborativa para validar as melhorias alcançadas. Por fim, o Capítulo 5 conclui este trabalho resumindo os resultados obtidos e propondo trabalhos futuros.

Capítulo 2

Fundamentação Teórica

A fim de explicar a relação entre sistemas de recomendação e o sistema imunológico artificial este capítulo primeiramente apresenta o que é um sistema de recomendação, mostrando suas principais abordagens e desafios enfrentados. Em seguida são apresentados os princípios do sistema imunológico dos vertebrados a fim de esclarecer como a sua incorporação artificial pode ser usado na recomendação de disciplinas. Para isso também são apresentados os principais algoritmos utilizado na literatura. Por fim, são apresentados trabalhos que já utilizaram o sistema imunológico artificial para gerar recomendações.

2.1 Sistemas de Recomendação

Sistemas de recomendação são programas computacionais automatizados que auxiliam nas escolhas de atividades a serem exploradas por um indivíduo. Estes sistemas podem se basear em diversos fatores (histórico de preferência de filmes, livros lidos anteriormente, curso de graduação, itens populares e até mesmo local de moradia do alvo) e são capazes de analisar e compreender o comportamento dos usuários de uma plataforma para fazer recomendações relevantes e personalizadas de conteúdos. Eles já foram aplicados com sucesso em diversos serviços, os mais famosos sendo os de *streaming* de músicas e filmes como o *Spotify* e a *Netflix*, que utilizam as classificações de itens favoritos pelos próprios usuários para lhes recomendar novos itens [Amatriain and Basilico 2015].

A Figura 2.1 ilustra o processo de um sistema de recomendação. O alvo é um indivíduo do qual o sistema deve analisar as suas preferências, que devem ser informadas, para que seja gerada uma lista de itens que o sistema classifica como relevantes. O sistema pode usar como base os itens disponíveis pra a recomendação e/ou o histórico de consumo de um grupo de usuários, que deve estar em um banco de dados. Este poderá ser usado como memória ou para a construção de modelos estatísticos de consumo. Após o alvo consumir a lista resultante, suas preferências também entram para o banco de dados e poderão ser usadas para gerar recomendações para próximos usuários.

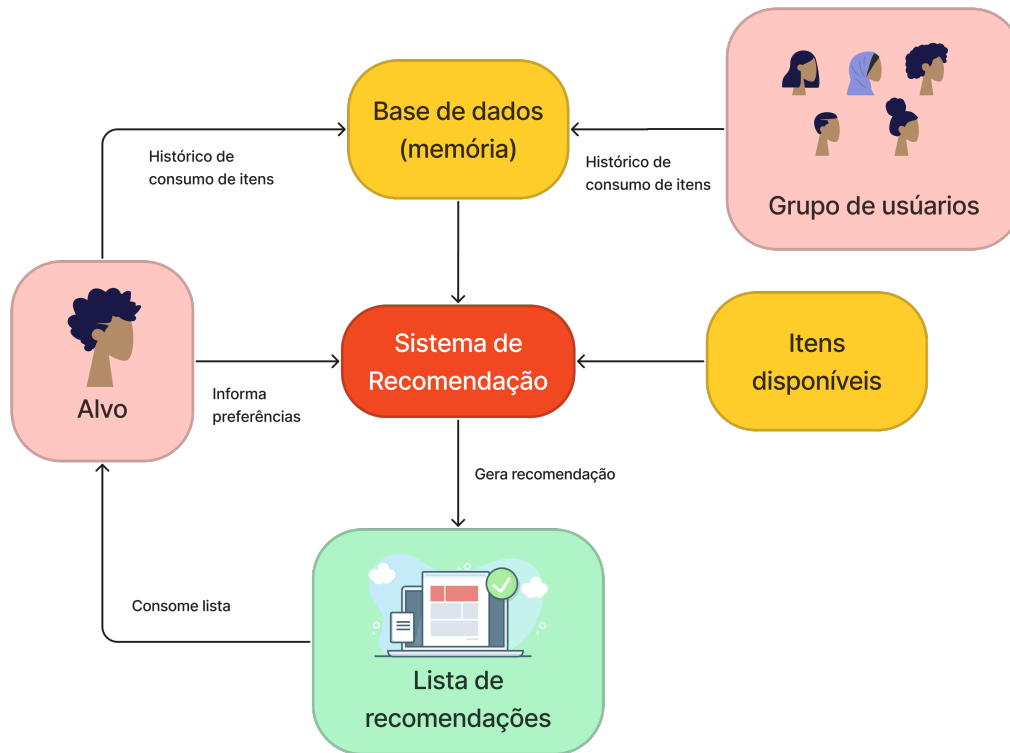


Figura 2.1: Esquema básico do processo de um sistema de recomendação

Esse tipo de sistema é classificado como uma aplicação da inteligência artificial [Pavlidis 2019] e sua origem pode ser atribuída a trabalhos em ciência cognitiva, teoria da aproximação e recuperação de informação, mas se tornou uma área de pesquisa independente na década de 90, quando o foco passou a ser em estruturas que dependem da classificação dos itens [Adomavicius and Tuzhilin 2005]. O interesse por essa área de pesquisa é motivado devido à sua vasta possibilidade de aplicações práticas pois auxiliam a analisar problemas que contém um universo grande de opções e assim podem fornecer recomendações personalizadas e certas [Isinkaye et al. 2015].

O trabalho em [Fayyaz et al. 2020] fornece uma visão geral do estado da arte em sistemas de recomendação, seus tipos, desafios, limitações e aplicações. Sua principal aplicação é no comércio eletrônico, com casos de sucesso em recomendações de filmes, músicas, imagens, livros, etc. mas sua aplicabilidade é extensa. Há várias outras áreas que podem se beneficiar de recomendações em seu processo de tomada de decisão, como agricultura [Kuanr et al. 2018], saúde [Prajapati and Brahmabhatt 2022] ou turismo [Mu et al. 2009]. Por exemplo, em [Rodriguez-Carrion et al. 2012] é estudado a possibilidade de utilizar padrões de mobilidade em transporte público para ajudar outros usuários a encontrar seu caminho ou prever sua próxima localização recomendando assim linhas de metrô e ônibus a fim dos usuários economizarem tempo e dinheiro.

Mensurar a qualidade de um sistema de recomendação não é simples devido à falta de um método explícito para determinar se uma recomendação é precisa ou não [Fayyaz et al. 2020].

Um sistema de recomendação pode prever como os usuários classificam um item ou qual item (ou lista de itens) é de interesse do usuário. Quando se trata da primeira opção a avaliação das recomendações é baseada em medidas amplamente utilizadas, como o erro médio absoluto (MAE) ou a raiz do erro quadrado médio (RMSE). No RMSE a diferença entre as avaliações reais do usuário e as previstas deve ser determinada. O MAE mede o desvio absoluto médio das avaliações previstas em relação as avaliações reais dos usuários. Valores mais baixos de RMSE e MAE indicam maior acurácia do sistema de recomendação. O RMSE é geralmente maior ou igual ao MAE, mas se ambas as métricas forem iguais, todos os erros terão a mesma magnitude.

Já na segunda opção, o qual deve se medir o quanto o sistema recomenda itens interessantes ao usuário, medidas de recuperação de informação (precisão e revocação) são usadas [Pavlidis 2019]. A precisão indica a fração de itens relevantes entre todos os itens recomendados a um usuário. Já a revocação representa o número de itens relevantes recomendados em relação o número total de itens que foram recomendados.

Em alguns casos um sistema de recomendação também deve levar em conta a diversidade, pois ter itens muito semelhantes em uma lista pode não ser de muito ajuda na perspectiva dos usuários. Isso porque as recomendações são baseadas em similaridades o que pode acarretar em recomendações que parecem redundantes e o usuário pode levar mais tempo para explorar a variedade de itens disponíveis. Isso ocorre quando o usuário é exposto a apenas a uma seleção mais restrita de itens, enquanto itens de outros nichos que também são relevantes podem ser completamente ignorados.

A diversidade de recomendações também permite que os usuários descubram objetos que eles não encontrariam facilmente por si mesmos. Entretanto se um algoritmo se concentra estritamente em aumentar a diversidade, a acurácia pode ser reduzida. O método mais utilizado para medir a diversidade usa a similaridade item-item, normalmente baseada no conteúdo do item. Pode-se medir a diversidade de uma lista com base na soma, média, mínima ou distância máxima entre os pares de itens.

Trabalhos na área de avaliação desses sistemas que levam em conta a dificuldade de precisão e a diversidade nas recomendações foram desenvolvidos ao longo do tempo. [Shani and Gunawardana 2011] avaliam como comparar os recomendadores com base em um conjunto de propriedades relevantes para suas aplicações, enquanto [Avazpour et al. 2014] faz de revisão uma série de métricas e medidas de avaliação, bem como algumas abordagens usadas para avaliar sistemas de recomendação, como, por exemplo, correção, novidade, cobertura etc. Em [Zhang 2009], o objetivo geral do trabalho é examinar maneiras de melhorar a diversidade de recomendações mantendo alta precisão. Neste sentido, o autor aborda a questão de como melhor avaliar as estratégias de diversificação e

propõe novos algoritmos de aprimoramento da diversidade.

Um dos esforços no estudo dos sistemas de recomendação é, naturalmente, a melhoria contínua de tais e, desta forma, [Adomavicius and Tuzhilin 2005] listaram linhas de progresso que moldaram pesquisas futuras. Algumas delas são:

- Criação de melhores métodos para representar o comportamento de usuários e informações de itens a serem recomendados;
- Incorporação de informações contextuais no processo de recomendação;
- Utilização de classificação multiobjetivo;
- Utilização medidas mais efetivas para determinar o desempenho do sistema.

2.1.1 Recomendação de disciplinas

Uma das áreas em que os sistemas de recomendação são utilizados é o meio acadêmico. O estudo em [Lynn and Emanuel 2020] descreve algumas áreas de atuação de sistemas de recomendação em domínio educacional: aplicação em *e-learning*, gerar materiais relevantes para os alunos devido grande quantidade de opções disponíveis para aprendizagem online, por exemplo, há sites como *Cousera*¹ e *Udemy*² que tem catálogos extensos de cursos; serviços de biblioteca, para ajudar estudantes, leitores e pesquisadores a encontrar facilmente livros apropriados de acordo com suas áreas de estudos; campo de pesquisas e publicações, os sistemas podem recomendar revistas e conferências adequadas aos autores; em institutos de ensino superior, pesquisadores podem usar sistemas de recomendação para prever probabilidade de abandono escolar, ajudar alunos a selecionar cursos de graduação, a selecionar disciplinas e identificar padrões de aprendizagem e desempenho [Sheehan and Park 2012].

Tratando-se especificamente da recomendação de disciplinas, são várias as abordagens de recomendação já utilizadas, sendo elas filtragem colaborativa, baseada em conhecimento, baseada em conteúdo, baseada na comunidade, demográfica, e métodos híbridos, como veremos mais adiante.

O trabalho feito por [Bendakir and Aïmeur 2006] desenvolveu um sistema baseado em regras de associação, que incorpora um processo de mineração de dados junto com as classificações dos usuários. A partir de um histórico de dados reais, descobre regras significativas que associam disciplinas acadêmicos seguidos por ex-alunos. Essas regras são usadas posteriormente para inferir recomendações. Já [Farzan and Brusilovsky 2006] apresentam um sistema baseado na comunidade, que fornece recomendações com base na avaliação da relevância de disciplinas por parte dos alunos levando em conta seus objetivos de carreira. Outro trabalho que envolve objetivos de carreira foi desenvolvido em [Jiang et al. 2019] que

¹<https://pt.coursera.org/>

²<https://www.udemy.com/pt/>

apresenta um sistema baseado em rede neural recorrente para sugerir disciplinas para ajudar os alunos a se prepararem para disciplinas-alvo, personalizados para seu histórico de conhecimento prévio estimado.

Alguns trabalhos focam na predição de notas de disciplinas para recomendação, de modo a selecionar as disciplinas top- n . [Sheehan and Park 2012] apresentam um protótipo com o objetivo de demonstrar o potencial da tecnologia de recomendação, fornecendo previsões de notas para os próximos cursos na carreira acadêmica de um aluno para apoiar a tomada de decisões de administradores, alunos, educadores e orientadores acadêmicos. Em [Elbadrawy and Karypis 2016] é investigado como as características acadêmicas do aluno e do curso influenciam os padrões de matrícula. Essas características são usadas para definir grupos de alunos e cursos em vários níveis de granularidade e, a partir destes, projetar modelos de previsão de notas e classificação de cursos top- n .

Há também o trabalho desenvolvido em [Park 2018] que propõe um método para prever o desempenho dos alunos em especializações do curso de computação. No trabalho o cálculo de similaridade entre alunos é feito ponderando as disciplinas com base na especialização em que disciplina pertence, como resultado produz previsões personalizadas das notas dos alunos em disciplinas necessárias para cada especialização em computação.

Outro problema abordado na recomendação de disciplinas é que, na maioria das abordagens, não é levada em conta a presença de pré-requisitos. O trabalho em [Parameswaran and Garcia-Molina 2009] tenta desenvolver algoritmos de recomendação que consideram esses pré-requisitos e suas recomendações contém cadeias de disciplinas.

2.1.2 Recomendação explicável

Problemas de recomendação se encaixam nos chamados problemas 5W (*when, where, who, what, why*, em inglês "quando, onde, quem, o quê e por quê") [Zhang and Chen 2020]. A recomendação explicável se refere a algoritmos que tentam responder o "por quê", ou seja, eles não apenas recomendam itens como também oferecem explicações de modo a esclarecer a razão da recomendação de tais itens. Na seleção de disciplinas essas explicações são um modo de aumentar a confiança nas recomendações do sistema. O trabalho em [Santos and Boticario 2010] adiciona a justificativa ao seu modelo de recomendação aplicado ao domínio educacional.

Um dos objetivos das explicações é transparência, ou seja, expor o raciocínio e os dados por trás de uma recomendação. Isso então se desdobra na ajuda que ela pode dar ao sistema ao inspirar confiança do usuário, aumentar a satisfação e persuadi-lo a experimentar ou comprar um item recomendado. Um trabalho extenso sobre a explicabilidade de um recomendação é feito em [Tintarev and Masthoff 2015], onde são apresentados vários estilos de explicação e como eles estão relacionados aos algoritmos.

A pesquisa em recomendação explicável considera duas filosofias. Uma delas considera

a explicabilidade dos métodos de recomendação e visa implementar modelos interpretáveis para aumentar a transparência e geralmente tem como saída resultados explicáveis, ou seja, essa filosofia tenta desenvolver modelos explicáveis intrínsecos e entender como funciona o processo de recomendação. A outra filosofia se concentra na explicabilidade apenas dos resultados, o modelo de recomendação é tratado como uma caixa-preta e são desenvolvidos modelos separados para explicar os resultados da recomendação [Zhang and Chen 2018].

2.1.3 Abordagens de recomendação

O sistema de recomendação tem como objetivo estimar uma nota para itens não avaliados pelos usuários e através delas oferecer uma boa recomendação ou lista de recomendações. De acordo com [Burke and Robin 2007] há quatro classificações de técnicas diferenciadas com base em sua fonte de conhecimento. São elas:

1. *Baseado em conteúdo*: A recomendação é obtida através da avaliação de itens similares aos presentes no conjunto de preferência do usuário, ou seja, o sistema "aprende" quais itens o usuário pode gostar com base nas características dos itens que ele já avaliou.
2. *Filtragem colaborativa*: A recomendação é obtida através das avaliações de itens por parte dos usuários, por exemplo, o sistema procura por usuários histórico similares de avaliações e encontra itens através dessa "vizinhança".
3. *Demográfico*: São utilizados dados demográficos dos usuários para gerar a recomendação, os itens recomendados para diferentes nichos demográficos podem ser gerados combinando as avaliações dos usuários desses nichos.
4. *Baseado no conhecimento*: A recomendação de itens tem como base inferências sobre as necessidades e preferências de um usuário, muitas vezes se terá informações explícitas de por quê o item é relevante a tal usuário.

Além dessas classificações alguns outros trabalhos utilizam de uma outra técnica, como definida por [Sankar et al. 2014]:

- 5) *Baseada na comunidade*: A recomendação é feita utilizando a preferência de usuários relacionados. A detecção de comunidade pode dividir os usuários em vários subgrupos com interesses e perfis semelhantes.

Essas técnicas ainda podem ser combinadas de forma a melhorar as recomendações, são as chamadas técnicas híbridas. Cada uma dessas técnicas foram muito exploradas nas últimas décadas e suas capacidades e limitações são conhecidas. As técnicas híbridas tem como objetivo contornar essas limitações. Agora detalhamos, a seguir, as estratégias mais utilizadas em sistemas de recomendação de disciplinas, pois este é o foco deste trabalho:

2.1.3.1 Baseadas em conteúdo

Neste método o objetivo é analisar os itens avaliados pelo usuário no passado de modo a encontrar características em comum entre eles e assim determinar um perfil de interesse, onde se armazena os interesses do usuário em relação às características dos itens, de modo que o sistema será capaz de filtrar itens que sejam de interesse e assim recomendá-los. Essa técnica tem origem nas pesquisas de recuperação e filtragem de informação [Adomavicius and Tuzhilin 2005]. A Figura 2.2 ilustra seu funcionamento.

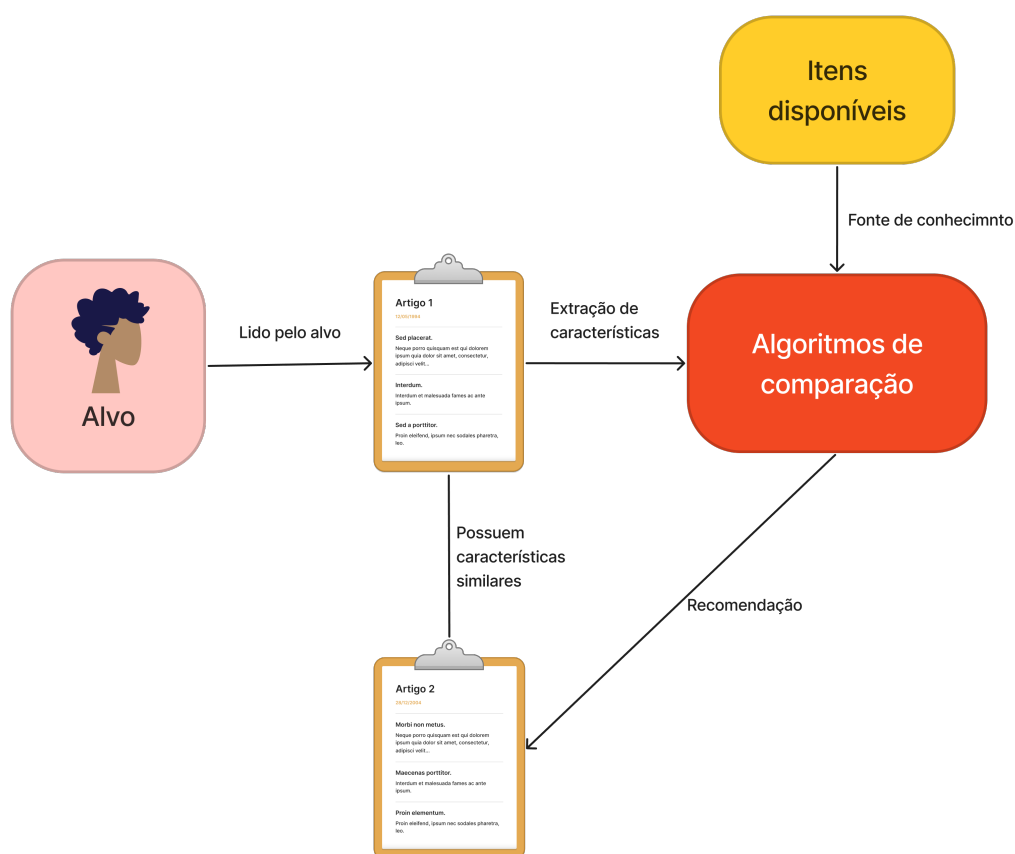


Figura 2.2: Esquema da recomendação baseada em conteúdo

Nesse tipo de técnica a recomendação se apoia na descrição do conteúdo dos itens e não se leva em conta as preferências de outros usuários, os algoritmos apenas exploram as informações sobre os itens avaliados pelo usuário. É dito que esse tipo de abordagem é independente e é devido a essas características que é possível especificar os motivos da recomendação e, conseqüentemente, aumentar a confiança do usuário no sistema. Outra vantagem dessa abordagem é que é possível indicar itens que nunca foram avaliados por outros usuários, ou seja, são novos no sistema. A descrição do item já seria suficiente para que ele seja recomendado a algum usuário que tenha interesse nas características desse item.

A problema enfrentado por esses algoritmos é o número de características escolhidas para serem analisadas ou que descrevam os itens. Se é sabido apenas parte das informações sobre os usuários, por exemplo se sabe o curso de um aluno mas não em qual semestre do currículo ele se encontra, é possível que a recomendação seja afetada. De mesmo modo, a quantidade de informação que temos sobre dado item pode afetar a recomendação, se temos poucas informações não é possível fazer boas recomendações mas, por outro lado, a grande quantidade poderia afetar o desempenho do sistema por causa do grande volume de dados a ser analisado.

Além dessa limitação há o problema da especificidade das recomendações. O usuário nunca receberá uma recomendação que se diferencie do perfil e a lista de recomendação poderá ter apenas itens parecidos aos avaliados anteriormente. Quando trabalhando com disciplinas, o sistema poderia recomendar apenas disciplinas parecidas às que o aluno já cursou e as opções de disciplinas de outros departamentos poderiam ficar de fora da lista.

2.1.3.2 Filtragem colaborativa

Esse método tenta prever quanto o item é relevante a um usuário particular baseando-se nos itens avaliados pelos outros. A abordagem consiste em identificar um subconjunto de usuários similares para assim avaliar as preferências deste grupo a certos itens. Para o problema de recomendação de disciplinas isso quer dizer que, para recomendar uma disciplina a certo aluno, o sistema tenta buscar outros alunos com histórico acadêmico similar e, assim, recomendar opções em que esses outros alunos tenham obtido resultados satisfatórios.

Para a implementação deste método é gerada uma matriz, denominada “matriz usuário-item”, onde os valores de cada célula da matriz são a avaliação do item por um usuário, a coluna corresponde a um usuário e a linha a um item, ou vice-versa. Este tipo de abordagem não precisa de informação nenhuma quanto às características dos itens, o item é tratado como uma “caixa-preta” e a recomendação é apenas baseada nas avaliações de usuários similares. Diferentemente do método baseado em conteúdo, esses algoritmos têm a vantagem

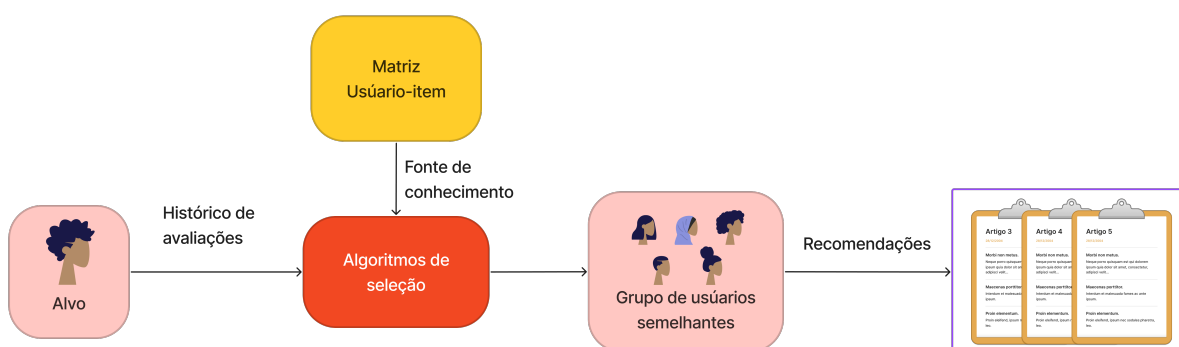


Figura 2.3: Esquema da filtragem colaborativa

de conseguir recomendar itens com características diversas e não sofrem com o problema de especificidade. A Figura 2.3 resume seu funcionamento.

Entretanto, há alguns desafios na implementação desse método [Su and Khoshgoftaar 2009]:

- *Esparsidade dos dados*: Nas maioria das vezes que um sistema de recomendação é aplicado apenas uma pequena fração dos itens são avaliados por um usuário e apenas uma parcela de todos os usuários avaliam um determinado item, o que leva a uma matriz usuário-item bastante esparsa, ou seja a matriz usuário-item utilizada tem muitas avaliações vazias ou desconhecidas, o que pode afetar o desempenho do sistema.
- *Cold-start*: este problema é decorrente da esparsidade, quando um item novo é adicionado ao sistema não há nenhuma informação quanto a ele e não haveria modos de recomendá-lo, pois não havia sido avaliado por nenhum usuário. O mesmo pode acontecer a um novo usuário que não pode ser direcionado a nenhum grupo de usuários e assim suas primeiras recomendações podem acabar não sendo muito relevantes.
- *Escalabilidade*: Sistemas de recomendação costumam ter bancos de dados constantemente modificados, novas avaliações de itens por usuários podem ser frequentes, o que leva a um novo perfil de interesse do usuário. O sistema precisa estar constantemente revendo as recomendações que são feitas, o que pode levar a uma sobrecarga, pois é necessário frequentemente analisar a qual grupo o usuário pertence e avaliar cada item candidato a recomendação. Este problema só ocorre em aplicações que envolvam grandes bases de dados e necessidade de requisições constantes a ela.
- *Sinônimos*: Há situações em que o mesmo item, ou itens muito semelhantes, são representados por identificadores diferentes.
- *Gray Sheep*: Ocorre quando as avaliações de um usuário não tem uma similaridade com nenhum grupo de usuário e o sistema não consegue fazer recomendações.

As técnicas de filtragem colaborativa podem ser dividida em duas classes que lidam de modos diferentes com esses problemas: baseada em memória e baseada em modelo.

Baseado em memória: Nessa abordagem as recomendações são computadas através da matriz usuário-item e precisam ter essa estrutura em sua memória. Os algoritmos mais básicos se baseiam no conceito do vizinho mais próximo, em inglês *K-nearest-neighbour*, KNN. Existem duas maneiras de definir esses vizinhos: *baseada em usuário*, em que ele se utiliza das ferramentas de cálculo de correlação para selecionar usuários com alta medida de similaridade a um usuário u , as avaliações destes usuários ao item i são ponderadas de acordo e utilizadas para calcular quanto um item i seria de interesse e assim fazer a recomendação. A outra maneira é dita *baseada em item*, que tenta predizer a relevância do item i para certo usuário u baseado nas preferências desse usuário u por itens similares ao item i , ou seja, o

método baseado em item considera o conjunto de itens que o usuário-alvo avaliou previamente e computam quão similares eles são com relação a um item-alvo i na matriz usuário-item, mas diferente da recomendação baseada em conteúdo ela não leva em consideração as características do item para calcular similaridade mas sim o padrão de avaliações na matriz usuário-item.

Para exemplificar melhor como funciona cada estratégia, nas Figuras 2.4 e 2.5 temos a mesma matriz usuário-item. Na Figura 2.4 é aplicado o método baseado em usuário, o alvo da recomendação é o Usuário 1 e seu vizinho mais próximo (usuário com padrão semelhante de avaliação) é o Usuário 3, logo o item a ser recomendado será o Item 3, que foi avaliado pelo Usuário 3 mas que ainda não foi consumido pelo Usuário 1. Já na Figura 2.5 o método é o baseado em item, o alvo continua o mesmo mas agora procuramos por itens com padrão de avaliações semelhantes aos já avaliados pelo Usuário 1, neste caso o Item 1 foi avaliado com nota mais alta e o item 4 tem padrão de avaliações bem semelhante a ele e por isso será selecionado para ser recomendado.

	Item 1	Item 2	Item 3	Item 4
Usuário 1	5	3	Adicionar	
similares Usuário 2		2	1	
Usuário 3	5	4	3	
Usuário 4	1			1
Usuário 5	5	Adicionar	1	5

Figura 2.4: Filtragem colaborativa baseada em usuário

Uma das etapas mais importantes da filtragem colaborativa baseada em memória é o cálculo da similaridade entre usuários ou itens. Para a técnica baseada em itens essa similaridade é calculada entre itens i e j , baseando-se nos usuários que avaliaram ambos os itens. Já para a técnica baseada em usuários a similaridade é entre os usuários u e v que avaliaram os mesmos itens. São vários os métodos utilizados para calcular a similaridade. O mais usado



Figura 2.5: Filtragem colaborativa baseada em item

deles é a correlação de Pearson [Al Hassanieh et al. 2018], que considera o quanto duas variáveis são relacionadas linearmente. A desvantagem em sua utilização é que o resultado é impreciso quando os dois usuários têm apenas uma avaliação em comum, ou quando um determinado usuário avaliou apenas um item. Também são utilizadas outras técnicas, por exemplo, o cosseno os vetores de avaliação de dois usuários ou itens com os quais se quer avaliar a similaridade, mas neste caso a desvantagem está no fato de que esta medida não considera as diferenças na média e variância das avaliações feitas pelos usuários. Assim a escolha do método de similaridade impacta na precisão e no desempenho do algoritmo de recomendação. O trabalho em [Al Hassanieh et al. 2018] ainda relata sobre várias outras medidas de similaridade que já foram utilizadas em filtragem colaborativa.

As principais vantagens da filtragem colaborativa são sua simples implementação e a maneira como a recomendação pode ser mais facilmente explicada e intuitiva. No entanto seu desempenho é bastante afetado quando o banco de dados é uma matriz muito esparsa, pois a falta de dados pode ser insuficiente para criar a vizinhança [Pavlidis 2019].

Baseada em modelo: abordagens baseadas em modelo utilizam as avaliações dos usuários para treinar um “modelo” capaz de fazer previsões. Os treinamentos são feitos com os itens que o usuário já avaliou. Técnicas de aprendizado de máquina ou algoritmos de mineração de dados são utilizados para a criação de um modelo que consiga identificar os padrões complexos existentes, prever classificações e gerar recomendações automáticas

[Pavlidis 2019].

Os métodos de aprendizado de máquina empregados nesta categoria incluem métodos bayesianos, modelos baseados em regras, árvores de decisão e modelos de fatores latentes [Pavlidis 2019]. Modelos utilizando fatores latentes são algoritmos que tentam explicar a preferência de um item para um usuário caracterizando os mesmos em fatores. Eles podem ser bastante eficientes na geração de recomendação mas devem ter uma grande quantidade de dados para treinamento, além disso, a construção do modelo pode ser custosa computacionalmente e, com novas avaliações pelos usuários o modelo deve ser reavaliado.

Essa abordagem pode gerar recomendações mais precisas mas por usar aspectos latentes e por isso ser uma técnica mais abstrata a explicação dos motivos da recomendação é mais difícil. Assim, a escolha da abordagem de recomendação depende da aplicação prática e não apenas do desempenho do sistema [Koren 2010].

2.1.3.3 Métodos híbridos

Os métodos híbridos se utilizam da junção de diversos algoritmos, podendo ou não ser com abordagens distintas, na tentativa de realçar o potencial de cada um e ao mesmo tempo compensar suas limitações.

Há varias maneiras de se fazer essa junção, que vão de simplesmente mesclar o resultado dos métodos mencionados anteriormente, ponderando a contribuição colaborativa de acordo com a interação usuário-item, até maneiras mais complexas como é o caso de aplicar métodos baseado em conteúdo para converter as matrizes esparsas em matrizes completas e então aplicar os métodos de filtragem colaborativa.

Em [Burke and Robin 2007] há uma extensa pesquisa de como a junção pode ser feita, a qual subdivide esses algoritmos em 7 tipos:

1. *Ponderado*: a pontuação de diferentes componentes de recomendação é combinada numericamente.
2. *Alternância*: o sistema escolhe entre os componentes de recomendação e aplica o selecionado.
3. *Misto*: as recomendações de diferentes recomendadores são apresentadas juntas.
4. *Combinação de dados*: dados derivados de diferentes fontes de conhecimento são combinados e fornecidos a um único algoritmo de recomendação. A ideia da combinação de dados é injetar dados de uma fonte (como recomendação colaborativa) em um algoritmo projetado para processar dados com uma fonte diferente (como recomendação baseada em conteúdo), ou seja, é capaz de tratar um dado "Usuário1 e Usuário2 gostaram Movie X" da mesma forma que o algoritmo trata dados como "Ator1 e Ator2 estrearam Movie X".

5. *Acréscimo de dados*: Uma técnica de recomendação é usada para calcular um dado ou conjunto de dados, que é então serão parte da entrada para a próxima técnica. Neste caso pode-se treinar um modelo baseado em conteúdo que gerará um conjunto de perfis que seja mais denso e mais útil para o estágio de filtragem colaborativa.
6. *Cascata*: As recomendações recebem prioridade estrita, com aquelas de prioridade mais baixa quebrando o empate na pontuação dos mais altos. Ou seja, um recomendador fraco não pode anular as decisões tomadas por um mais forte, pode apenas refiná-las.
7. *Meta-nível*: Uma técnica de recomendação é aplicada e produz algum tipo de modelo, que é então a entrada usada pela próxima técnica. É parecido como o acréscimo de dados mas neste caso a primeira estratégia de recomendação substitui completamente a fonte de conhecimento original por um modelo aprendido que o segunda estratégia deverá usar para gerar as recomendações.

[Trabelsi et al. 2021] apresentam uma pesquisa sobre o estado da arte na utilização de sistemas híbridos, segundo estas subdivisões.

2.2 Sistema Imunológico Artificial

Foram apresentados os conceitos relativos a sistemas de recomendação e a importância de uma estratégia que seja ao mesmo tempo robusta, escalável e explicável. A fim de esclarecer o vínculo entre os dois iremos apresentar nesta seção, primeiramente, o funcionamento do sistema imunológico de animais vertebrados, em seguida, estudaremos suas aplicações em problemas de otimização e recuperação de informações e, por fim, serão apresentados os principais algoritmos baseados nesta área da Inteligência Computacional.

O sistema imunológico humano possui algumas características que podem ser ideais para algoritmos de recomendação baseadas em filtragem colaborativa.

Para a aplicação da rede imunológica em um sistema de recomendação alguns recursos podem ser relevantes, são eles: (i) correspondência, que se refere à ligação entre anticorpos e antígenos, ao qual se pode traçar um paralelo entre o cálculo de semelhança de entre o alvo e outros usuários; (ii) diversidade, que se refere ao fato de que, para atingir a cobertura ideal do espaço de variações do antígeno, a diversidade de anticorpos deve ser incentivada, num sistema de recomendação essa característica pode evitar a especificidade da recomendação; (iii) memória, capacidade de "aprendizado", o sistema imunológico é capaz de lembrar de anticorpos capazes de combater antígenos específicos, pensando no sistema de recomendação esta característica pode que facilitar a geração de novas recomendações; e (iv) controle distribuído, pois o sistema imunológico é governado por interações locais entre células e não existe um controlador geral.

2.2.1 Princípios do Sistema Imunológico

A inspiração no sistema imunológico (SI) de vertebrados é interessante para a construção de sistemas computacionais inteligentes devido a algumas de suas propriedades, como: robustez, tolerância, distribuição, memória, adaptabilidade, auto-organização e autonomia [Figueredo et al. 2013].

Este sistema é estudado desde o século 18, mas foi apenas no século 19 que se descobriram os agentes infecciosos, estes chamados de microrganismos patogênicos (vírus, bactérias, fungos e parasitas) [Mettenleiter 2017]. Além disso, neste mesmo século, descobriu-se a existência das células brancas responsáveis pela destruição dos organismos invasores por serem capazes de produzir anticorpos [Addison 1843]. A função do sistema imune é proteger o animal dos agentes danosos que se infiltram nele. Ele é importantíssimo para a sobrevivência do animal na natureza pois é capaz de levantar informações sobre os agentes infecciosos de maneira a "*aprender*" a combatê-los e a "*memorizar*" estas informações para evitar novas infecções [De Castro and Von Zuben 1999].

No contexto biológico, diante de uma invasão de microrganismos, o SI apresenta dois níveis de barreiras [De Castro and Timmis 2002]. No primeiro ele tenta primeiramente combater a invasão com sua imunidade natural, as barreiras físico-químicas da pele, mucosas etc., este é um tipo de defesa amplo e não discriminado. Este é o sistema imune *inato*, fazem parte dele células de defesa prontas e sua resposta é idêntica e consistente em todos os vertebrados.

No segundo, se a estratégia inata não for suficiente, outro tipo de resposta é ativada, responsável pela robustez do sistema ao combater especialmente substâncias estranhas. Este é chamado de sistema imune *adaptativo*, específico ou adquirido, ou seja, o sistema desenvolve seus próprios agentes protetores, um deste são os anticorpos, que, quando em contato com a substância estranha específica, o antígeno (qualquer molécula do patógeno que o sistema imune consegue reconhecer), o marca para serem combatidos por células especializadas. Esse sistema é composto de células chamadas *linfócitos*, que podem ser do tipo B e T, e as células fagocitárias .

No combate ao antígeno há dois tipos de células fagocitárias: os macrófagos e as células dendríticas. O macrófago faz a fagocitose de células danificadas e velhas, assim como agentes estranhos e partículas inertes. As células dendríticas agem um pouco diferente, elas são capazes de reconhecer uma grande quantidade de diferentes agentes infectantes e tem como função capturar o antígeno e o apresentar ao linfonodo mais próximo. Estes são considerados o "centro de inteligência" do SI, segundo [Figueredo et al. 2013], pois é lá que é procurado o linfócito mais apto a combater o antígeno apresentado e onde se dá o início da resposta do sistema imune adaptativo.

Já os linfócitos são células que circulam constantemente pelo corpo e há uma grande diversidade de características entre estes, isso é decorrente do processo de produção das

células na medula (linfócito B) e no timo (linfócito T).

Durante seu trajeto entre corrente sanguínea e linfa os linfócitos podem encontrar um antígeno, quando isso acontece o linfócito aumenta de tamanho, replica seu DNA e dá origem a vários clones, células identicamente programadas para combater o agente infecciosos. Os linfócitos T são a maioria presente na sangue e seus clones podem se maturar em células com diversas funções diferentes, entre elas se destacam as *Células T de memória*. Elas são capazes de reconhecer o mesmo antígeno em um reinfecção e, assim, podem outra vez produzir anticorpos para evitar que a mesma doença se reestabeleça. Portanto, o tempo de ação do sistema imune *adaptativo* depende de se já houve uma exposição prévia a este agente.

Já os linfócitos B são os que iniciam a criação dos anticorpos, e estes atuam reagindo quimicamente com o antígeno. O processo se dá da seguinte maneira: quando os linfócitos B encontram um antígeno a célula é ativada e passa a sofrer um processo de clonagem, estes clones pode se se diferenciar em (I) células secretoras de anticorpos, os plasmócitos, e (II) células de memória de mesma especificidade antigênica que o linfócito original. Além disso cada clone de linfócitos B é responsável por um antígeno específico.

São as hipermutações somáticas que ocorrem nos linfócitos B ativados que confere a diversidade entre os anticorpos e, conseqüentemente, são responsáveis pelo refinamento da afinidade do anticorpo com relação a um antígeno específico, pois as modificações aleatórias são mais intensas quanto menor for a afinidade do anticorpo produzido com o corpo estranho. Assim, os linfócitos capazes de produzir anticorpos com poder de neutralização superior tendem a ser privilegiados no processo de escolha daquelas que irão se diferenciar em plasmócitos. Este fenômeno de seleção dos melhores linfócitos é chamado maturação de afinidade. A figura 2.6 ilustra o processo.

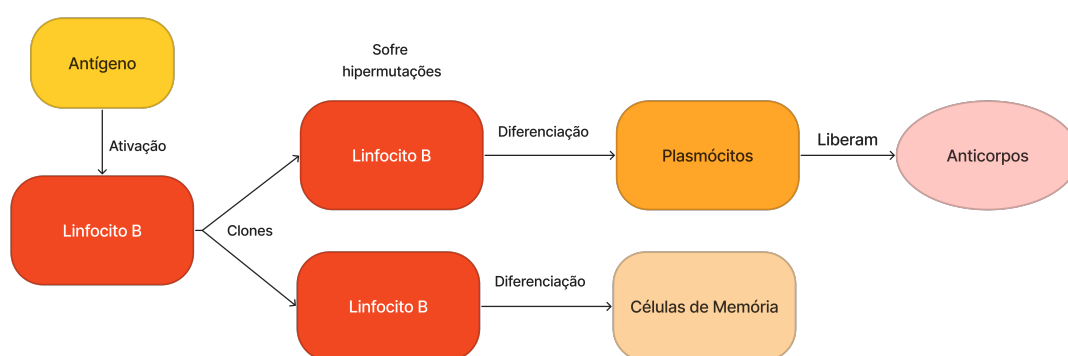


Figura 2.6: Processo de ativação do linfócito B

Os anticorpos são moléculas de forma específica que se ligam a um antígeno para marcar que este deve ser ingerido e destruído por células especializadas, os macrófagos. Como ilustra a Figura 2.7

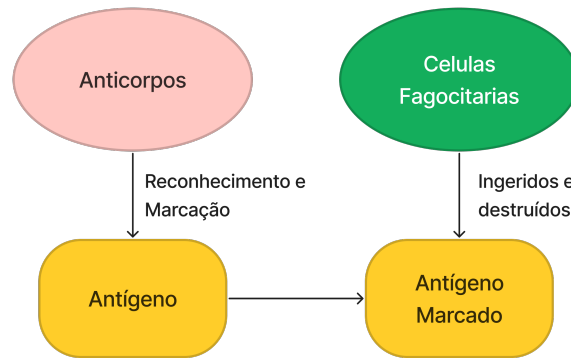


Figura 2.7: Processo de marcação do antígeno pelo anticorpo

Com todos estes mecanismos é visto que o sistema imunológico é um sistema construído de forma redundante, que tem uma série de componentes e mecanismos distintos, que são usados para proteger o indivíduo. Para nosso caso são as características do sistema imunológico adaptativo que interessam para aplicação em problemas de otimização, pois é nesta resposta que vemos o "aprendizado" do sistema imunológico. Uma teoria fundamental que descreve este processo é o de rede imunológica, que descrevemos a seguir.

Os processos presentes no sistema imunológico permitem muitas inspirações para a solução de problemas de otimização [Figueredo et al. 2013]. Descrevemos estes em mais detalhes a seguir.

2.2.1.1 Seleção Negativa

Linfócitos devem ser diversos o suficientes para reconhecer uma grande quantidade de antígenos, o problema é que algumas vezes eles podem acabar por reconhecer e atacar células do próprio organismo (i.e. doenças auto-imunes). Para que isso não aconteça há um processo chamado "seleção negativa", neste caso o sistema imunológico aprende através da otimização da distinção entre antígenos e as próprias células ou moléculas do corpo ("seleção de padrões") [Ji and Dasgupta 2007]. Este processo acontece quando o linfócito está passando pela maturação no timo.

O timo é um órgão em que não há a presença de agentes externos e assim as células presentes lá são amostras "próprias". Sendo então o ambiente ideal para que haja uma avaliação que verifica se os receptores resultantes de um processo de rearranjo genético pseudo-aleatório reagem a células do organismo [Hart et al. 2011]. Os Linfócitos T que reagem contra autoproteínas são destruídas; assim, apenas aqueles que não se ligam às autoproteínas podem deixar o timo. Ou seja, são testes de especificidade e afinidade do receptores do linfócito T que determinam se ele está apto a sobreviver ou se deve ser eliminado.

2.2.1.2 Seleção Clonal

O processo de seleção clonal ocorre quando o linfócito original é ativado ao reconhecer um antígeno [De Castro and Von Zuben 2002]. Tanto os linfócitos B quanto os T passam pela expansão clonal, mas as células B sofrem a hipermutação somática, ou seja, as novas células (clones) são submetidas a um processo com altas taxas de mutação. A seleção clonal ocorre quando linfócitos que apresentam maior reatividade ao elemento invasor se proliferam com maior velocidade em detrimento daqueles com menor capacidade de neutralização. Assim, as células B de memória são desenvolvidas para fazer uma resposta imune mais eficaz aos antígenos que foram encontrados.

2.2.1.3 Rede imunológica

Em 1974, Jerne [Jerne 1974] apresentou a teoria da rede imunológica. Nela, há uma rede idiotípica de células B, responsáveis pela produção de anticorpos e que são interconectadas para o reconhecimento do antígeno. Essas células também são capazes de se reconhecerem e, assim, podem gerar dois comportamentos distintos: estímulo, em que após a ativação da célula reconhecedora do antígeno, há a clonagem, hipermutação e secreção de anticorpos; e supressão, caracterizada pela eliminação de células muito semelhantes às já existentes na rede [de Castro and Von Zuben 2002]. Ou seja, as células estimulam e suprimem umas às outras, o que leva a uma estabilização da rede. Assim, o sistema imunológico pode ser visto como uma rede complexa na qual cada componente reconhece, é reconhecido por, interfere e sofre interferência de outros elementos.

O resultado do processo é visto na Figura 2.8. No lado esquerdo temos uma rede em que alguns anticorpos são bastante semelhantes entre si, onde a distância espacial entre os anticorpos representam seu grau de semelhança. Devido ao processo de reconhecimento entre si há a supressão de alguns anticorpos de modo que a rede estabilizada ainda seja capaz de reconhecer o antígeno mas que seja suficientemente diferentes uns dos outros. Como resultado temos a rede representada no lado direito da figura, onde todos os anticorpos tem

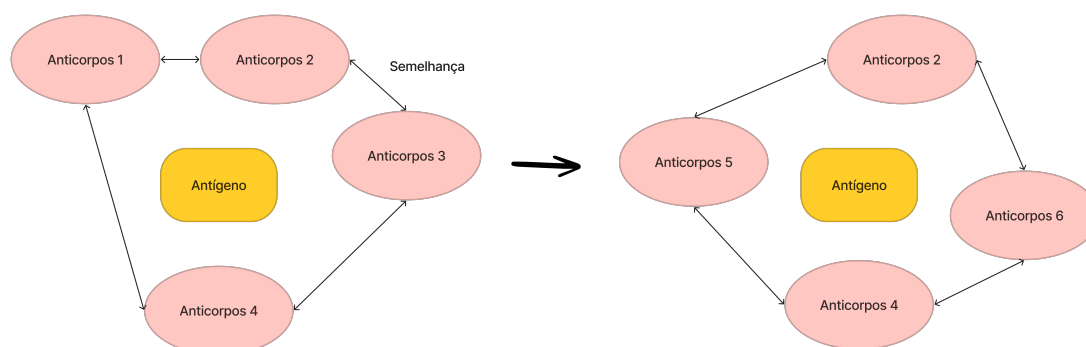


Figura 2.8: Estabilização da rede imunológica.

igualmente a capacidade de marcar o antígeno mas há uma diversidade entre eles.

2.2.2 Aplicações do Sistema Imunológico Artificial

Para ilustrar melhor como estes princípios são aplicados a problemas reais a Tabela 2.1 apresenta os principais algoritmos inspirados pelo sistema imunológico dos vertebrados e faz um paralelo entre o processo que inspirou o algoritmo e as principais áreas de aplicações.

Tabela 2.1: Aplicações dos princípios do SI em problemas reais

Processo	Algoritmo	Problema aplicado
Seleção negativa	Negative selection algorithm (NSA)	Deteção de anomalias, falhas e alterações
Rede imunológica	aiNet, cob-aiNet	Aprendizagem de máquina e otimização
Seleção clonal	CLONALG	Busca e otimização

Neste trabalho a inspiração advém da rede imunológica e, por isso, a seguir discutimos a fundo um exemplo de algoritmo. Apresentamos a aiNET [de Castro and Von Zuben 2002] e trabalhos posteriores que usam o conceito de concentração para regular a rede, como a cob-aiNET [Coelho and Von Zuben 2010].

2.2.2.1 Algoritmo aiNet:

O algoritmo aiNet é uma extensão das capacidades do algoritmo CLONALG [De Castro and Von Zuben 2002] e introduziu aspectos como a interação da rede entre soluções candidatas e adaptação do tamanho do conjunto de soluções candidatas. Ele foi primeiramente desenvolvido para problemas de *clustering* no qual pode identificar automaticamente o número adequado de grupos [de França et al. 2012].

O algoritmo se inicia com um conjunto de padrões (vetores) a serem reconhecidos (P) e o processo segue os seguintes passos [de Castro and Timmis 2002]:

1. A população inicial da rede (M) é gerada randomicamente.
2. Para cada padrão antigênico em P o algoritmo CLONALG é aplicado e retorna um conjunto de células de memória (M^*). Os seus passos são:
 - (a) Cada padrão de P_x , é apresentado à população M e sua afinidade (correspondência) com cada elemento da população M é calculada;
 - (b) São selecionados os elementos com afinidade mais alta de M e são geradas cópias desses indivíduos proporcionalmente à sua afinidade com o antígeno. Quanto maior a afinidade, maior o número de cópias.
 - (c) Todas estas cópias sofrem mutação com uma taxa inversamente proporcional à sua afinidade com o padrão de entrada: quanto maior a afinidade, menor a taxa de mutação.

- (d) Esses elementos mutantes são adicionados à população M e são escolhidos nesses elementos para serem mantidos como memórias do sistema.
3. Determina-se a afinidade (grau de semelhança) entre todos os elementos de M^*
 4. Há a eliminação de todos os elementos de M^* cuja afinidade é maior que um limiar. Este processo tem o objetivo de eliminar redundâncias presentes na rede suprimindo elementos que se reconhecem.
 5. Junta-se os elementos resultantes do passo anterior com todos elementos encontrados para cada padrão antigênico apresentado. O resultado será uma grande população de células de memória (M).
 6. A afinidade entre toda a população M é medida e todos os elementos que tem capacidade de se reconhecer são suprimido da rede, o que resultará em uma população final reduzida de células de memória que reconhecem os antígenos.
 7. Os passos 2 a 6 são repetidos até que um critério de parada pré-definido seja atendido, que pode ser um reconhecimento de padrão mínimo ou erro de classificação.

O algoritmo Cob-aiNet [Coelho and Von Zuben 2010] adiciona, em relação à aiNet o conceito de concentração. Ele é responsável pelo número de clones que é gerado para cada anticorpo. Este modelo permite que anticorpos com alta afinidade com os antígenos reduzam sua concentração se também apresentarem alta afinidade com outros anticorpos da população. A estrutura geral desse algoritmo é semelhante a outros algoritmos inspirados em redes imunes da literatura, exceto pela inclusão das etapas de atualização de concentração, o novo mecanismo de supressão baseado em concentração e a inclusão de um operador de busca local.

2.3 Trabalhos Relacionados

A inteligência artificial (IA), particularmente a inteligência computacional e métodos e algoritmos de aprendizado de máquina, tem sido aplicada no desenvolvimento de sistemas de recomendação para melhorar a acurácia das previsões e resolver os problemas apresentados na Seção 2.1.3 como a esparsidade de dados e o *cold-start*.

Com a utilização de técnicas de IA sistemas de recomendação alcançaram uma qualidade que antes não era possível, estas técnicas permitiram análises avançadas sobre os relacionamentos entre usuários e itens, representações de dados mais complexas e descoberta de conhecimento através de dados demográficos e contextuais. [Zhang et al. 2020] apresentam uma extensa revisão sobre as contribuições teóricas e práticas da aplicação de IA em sistemas de recomendação, além disso identificam áreas atuais de pesquisa e indicam novas direções. Aqui faremos um resumo breve sobre as principais aplicações apresentadas.

AI baseadas em Redes Neurais Profundas são mais utilizadas em problemas de classificação mas já foi aplicada com sucesso em sistemas de recomendação por [Salakhutdinov et al. 2007] que tinham como objetivo de enfrentar o problema que a filtragem colaborativa tem de não conseguir lidar com conjuntos de dados muito grande. Assim, um modelo de rede neural foi utilizado para explorar a propriedade ordinal das avaliações. A aplicação ganhou bastante visibilidade na competição promovida pela *Netflix* para achar o melhor algoritmo de filtragem colaborativa, pois conseguiu ser aplicada ao banco de dados disponibilizado que possui mais de 100 milhões de avaliações, como resultado conseguiram taxas de erro 6% menores que do sistema utilizado pela própria Netflix.

Mas Redes Neurais Profundas são melhores aplicadas nas áreas de processamento de linguagem natural, reconhecimento de fala e visão computacional. Com a disponibilidade de mais dados (por exemplo, comentários gerados pelo usuário ou fotos de itens), a necessidade de integrar todas as informações e fornecer recomendações para itens multimídia (imagens ou vídeos) levou ao desenvolvimento de mais sistemas de recomendação baseados em *deep learning*. [Cheng et al. 2016, Sedhain et al. 2015, Wu et al. 2014].

A utilização da Aprendizagem Ativa tenta enfrentar o problema de esparsidade de dados pois quanto maior o número de avaliações obtidas dos usuários, melhor o desempenho do sistema ao fornecer uma recomendação. Assim, a aprendizagem ativa é utilizada para ajudar a selecionar os itens mais representativos da base de itens disponíveis e entregá-los aos usuários para avaliar [Elahi et al. 2016]. Existem muitas estratégias de aprendizado ativo diferentes, como análise de impacto de avaliação [Mello et al. 2010] e *bootstrapping* [Golbandi et al. 2010], essas estratégias de aprendizado ativo foram integradas a modelos de recomendação comumente usados, como, por exemplo, árvores de decisão [Golbandi et al. 2011] e fatoração de matrizes [Karimi et al. 2011].

Além de tudo sistemas de recomendação tem a natureza de ser um processo iterativo entre usuário e o sistema o que é característica também da Aprendizagem por Reforço. Desse modo, os sistemas de recomendação baseados em aprendizado por reforço não tem como objetivo prever os interesses dos usuários em um momento específico, mas sim visam maximizar o engajamento e a satisfação dos usuários em longo prazo. Aqui o desafio está no fato de que o grande número de itens disponíveis aos usuários aumenta a complexidade do sistema, [Zhao et al. 2018, Zheng et al. 2018, Warlop et al. 2018]

Já Algoritmos Evolutivos podem ser principalmente usados para combinar as saídas de vários algoritmos de recomendação quando a recomendação é tratada como um problema de otimização multiobjetivo. Diversos trabalhos foram feitos considerando múltiplos indicadores de desempenho, por exemplo, acurácia, precisão e diversidade, etc [Rodriguez et al. 2012, Ribeiro et al. 2012]. Com o objetivo de encontrar tanto recomendações precisas e diversas [Karabadjji et al. 2018] melhorou um método filtragem colaborativa baseado em memória usando otimização multiobjetivo para encontrar vizinhos. Já [Esteban et al. 2020] apresentam um sistema de recomendação híbrido que combina Filtragem Colaborativa e Filtragem Baseada em Conteúdo usando vários critérios relacionados às

informações do aluno e das disciplinas para recomendar as disciplinas mais adequadas aos alunos, um algoritmo genético foi desenvolvido para descobrir automaticamente a configuração ótima do sistema.

Há alguns trabalhos que aplicam sistemas imunológicos artificiais (SIA) no problema de recomendação, os quais geralmente adaptam a técnica para funcionar como uma filtragem colaborativa. Em [Cayzer and Aickelin 2005], um SIA é utilizado como uma técnica de filtragem colaborativa completa, o antígeno é definido como o alvo da recomendação e os outros usuários são os possíveis anticorpos a serem testado contra o antígeno. Os autores trabalham com o conceito de concentração de anticorpos na rede: aqueles que tem maior correspondência com o antígeno tem maior concentração, enquanto os menos semelhantes possuem menor concentração e podem eventualmente ser removidos do sistema. Os anticorpos da rede são utilizados como conjunto de vizinhança para a predição na nota e a concentração do anticorpo também é utilizada como fator de ponderação. Os resultados experimentais são validados na recomendação de filmes.

Trabalhando com recomendação de filmes, Acilar e Arslan [Acilar and Arslan 2009] utilizam-se da teoria da rede imunológica de modo híbrido com um algoritmo de filtragem colaborativa. O mecanismo de supressão da rede imunológica é utilizado para redução de dimensão e, em seguida, os dados são agrupados através de um algoritmo *k-means* como filtragem colaborativa baseada em modelo. Esses conjuntos de dados agrupados são utilizados para produzir as recomendações ao usuário alvo. O problema da dispersão da avaliação de filmes também é abordado e é resolvido com avaliações implícitas, que tentam aumentar o número de avaliações por meio da observação do comportamento dos usuários. Essas avaliações são produzidas usando o mecanismo de hipermutação da rede imunológica.

Em [Chang et al. 2016] um SIA é utilizado na recomendação de disciplinas. O sistema é utilizado como uma técnica de *clustering* e a afinidade entre anticorpos antígenos foi combinada com os conceitos do diagrama de expansão, onde cada distância entre cada membro do conjunto de dados é medida pelo inverso de sua afinidade, ou seja, uma afinidade mais alta entre dois membros implica que eles estarão localizados mais próximos. Cada aluno é tratado como um antígeno ou anticorpo e anticorpos associados ao mesmo *cluster* criam as redes de anticorpos. Essas redes serão usadas na filtragem colaborativa para combinar antígenos (dados de teste) com grupos de usuários semelhantes.

2.4 Considerações finais do capítulo

Neste capítulo primeiramente foram apresentados os conceitos do sistema de recomendação e os elementos básicos que são necessários para a geração de uma recomendação personalizada, que podem ter como fonte de conhecimento os itens disponíveis para recomendação e/ou o histórico de consumo de usuários. Foi apresentado um resumo de sua aplicabilidade, citando trabalhos que utilizam estes sistemas em áreas como a agricultura,

saúde e turismo. Além disso, foi abordado também as principais formas de avaliação da qualidade das recomendações, que leva em consideração a acurácia da previsão de avaliações, o quanto o sistema retorna itens relevantes e a diversidade da recomendação.

Foi feito também uma apresentação do problema de recomendação de disciplinas. Discutimos trabalhos que aplicaram o sistema de recomendação de formas diferentes, alguns focam na predição de notas de disciplinas a fim de gerar uma lista top- n e outro tenta prever o desempenho de alunos em especializações de um curso. Também foi discutido a importância da explicabilidade do sistema a fim de que usuários tenham confiança nas recomendações.

Sobre as técnicas de recomendação focamos nas duas mais amplamente empregadas [Lynn and Emanuel 2020]: a recomendação baseada em conteúdo e filtragem colaborativa. Foi apresentado também um breve resumo de como estas duas abordagens podem ser combinadas para superar limitações inerentes à utilização de uma delas isoladamente.

Em seguida, a fim de deixar claro os paralelos aplicados no algoritmo proposto neste trabalho, foi feita uma apresentação do funcionamento do sistema imunológico dos vertebrados, dando foco nos principais processos que inspiraram os algoritmos mais utilizados na literatura, são eles: seleção negativa, rede imunológica e seleção clonal. Focamos a atenção principalmente na rede imunológica, apresentando também a aiNet, algoritmo que gerou diversas extensões sendo aplicados em problemas que vão de clusterização à otimização dinâmica [de França et al. 2012].

Por fim, foi feita uma revisão bibliográfica dos trabalhos que aplicam inteligência artificial a sistemas de recomendação. Foram apresentados trabalhos que tentam lidar de maneiras diferentes com as limitações das técnicas de recomendação. Estes trabalhos lidam com problemas como esparsidade dos dados, consideração de avaliações multimídias, natureza interativa do processo de recomendação e geração de recomendações multi-objetivo.

Vimos que a [Chang et al. 2016] faz uma aplicação bem sucedida do AIS a um sistema de recomendação de disciplinas mas sua inspiração utiliza apenas o conceito de afinidade entre anticorpo e antígeno para gerar a recomendação. Como foi discutido anteriormente isso pode causar uma especificidade na recomendação, ou seja, pode ser que a lista de disciplinas recomendadas sejam muito parecidas as que o aluno já cursou. Uma vizinhança diversa no contexto da recomendação de disciplinas pode afetar os resultados da recomendação de maneira positiva, visto que alunos tendem a ter interesses e habilidades diversas, não se assemelhando, portanto, a apenas um grupo de pessoas específico. Assim voltamos a olhar para o trabalho feito por [Cayzer and Aickelin 2005] para aplicar conceitos de regulação da rede para gerar a diversidade na recomendação de disciplinas.

No próximo capítulo apresentamos a metodologia utilizada para aplicar o sistema imunológico artificial ao problema de da recomendação de disciplinas.

Capítulo 3

Metodologia

Um problema comumente observado em algoritmos de recomendação é a concentração de suas recomendações em um subconjunto do conjunto completo de itens relevantes. Existem itens que podem ser relevantes ao usuário, mas que nunca (ou raramente) são recomendados pelo sistema [Zhang 2009]. Esta foi uma das áreas de estudo de sistemas de recomendação nos últimos anos. [Aytekin and Özge 2014] declaram que o algoritmo ser capaz de recomendar um conjunto diversificado de itens é importante para a satisfação do usuário, pois oferece um conjunto mais rico de itens para escolher e aumenta a chance de descobrir novos itens. Para solucionar o problema no trabalho é proposto um método que permite ajustes no nível de diversidade por parte do usuário. Na mesma linha de pesquisa [Karabadi et al. 2018] usa um sistema de recomendação baseado em algoritmo de otimização multiobjetivo evolucionário para obter um grupo de perfis que maximize a similaridade com o usuário ativo e a diversidade entre seus membros.

Assim, para estender a pesquisa na intenção de explorar todo o conjunto de disciplinas relevantes, o conceito de rede imunológica regulada pelo mecanismo de concentração de anticorpos será aplicado ao problema de recomendação de disciplinas, de modo a analisar o impacto na recomendação e a abrangência das disciplinas recomendadas. A utilização do sistema imunológico acrescenta a pesquisa pois a diversidade é intrínseca ao algoritmo e sua explicabilidade é mais simples, tendo a vantagem de conseguir embasar sua saída resultando em um aumento da confiabilidade no sistema.

O sistema aqui proposto baseia-se em [Cayzer and Aickelin 2005] e o estende para o contexto da recomendação de disciplinas, com o objetivo adicional de se avaliar o impacto de uma vizinhança diversa na qualidade da recomendação. A vantagem esperada na utilização do SIA, neste caso, é justamente a diversidade na recomendação final, devido a uma abrangência maior de vizinhos que tenham alguma semelhança com o aluno alvo, é esperada uma abrangência maior nas opções de disciplinas que podem ser recomendadas.

O impacto de uma vizinhança diversa ainda é pouco explorado no contexto da recomendação de disciplinas, mas esta característica pode afetar os resultados da recomendação de

maneira positiva, visto que indivíduos tendem a ter interesses e habilidades diversas, não se assemelhando, portanto, a apenas um grupo de pessoas específico, como a maioria dos sistemas de recomendação tendem a adotar como estratégia. Apesar da aplicação original bem-sucedida em recomendação de filmes, o método ainda não foi aplicado em conjuntos de dados educacionais, mas este parece ter aderência ao problema da recomendação de cursos uma vez que, devido à não-esparsidade dos dados, é possível encontrar alunos com vários níveis de semelhança entre si.

Neste trabalho, assim como em [Cayzer and Aickelin 2005], a variação da concentração de cada anticorpo da rede a cada interação do SIA é em função da similaridade com o antígeno (aluno alvo da recomendação) e entre os demais anticorpos da rede. Ela é regida pela equação 3.1.

$$\Delta c = k_1 m_i c_i c_a - \frac{k_2}{n} \sum_{j=1}^n m_{ij} c_i c_j - k_3 c_i, \quad (3.1)$$

Onde k_1 é a taxa de estímulo, m_i é a função de correspondência entre o anticorpo i e o antígeno, c_i é a concentração do anticorpo i , c_a é a concentração do antígeno, k_2 é a taxa de supressão, m_{ij} é a função de correspondência do anticorpo i ao anticorpo j , k_3 é a taxa de mortalidade e n é o número de anticorpos.

Pela equação 3.1 é possível verificar de onde vem a diversidade da recomendação final. Quando a taxa de supressão é zero, se o anticorpo for suficientemente semelhante ao antígeno e superar a taxa de mortalidade, sua concentração irá aumentar, caso contrário, eventualmente será suprimido da rede. Ao considerar uma taxa de supressão não-nula, se esse anticorpo for muito semelhante aos anticorpos já presentes na rede, ele terá um decréscimo em sua concentração. Assim, escolhendo valores de taxas adequados garantimos uma semelhança suficiente com o antígeno, uma diversidade nos anticorpos e, conseqüentemente, na recomendação.

3.1 Conjunto de dados

Neste trabalho, são utilizados os dados de 320 alunos do curso de Engenharia de Redes de Comunicação da Universidade de Brasília, entre os anos de 2015 a 2020, disponibilizados a estes autores já previamente anonimizados, conforme mostra a Figura 3.1.

No currículo da Universidade de Brasília (UnB) há particularidades que são ideais para que seja feita a recomendação de disciplinas. O currículo segue um sistema semi-seriado, que é uma estrutura acadêmica que combina características do sistema seriado, que consiste na distribuição das disciplinas do curso em blocos que deverão ser sucessivamente integrados pelo aluno e do sistema de créditos, onde um crédito corresponde a 15 horas-aula e o aluno deve obter um número mínimo de créditos para se formar. Ou seja, na UnB os estudantes em primeiro momento devem seguir lista pré-determinadas de disciplinas por semestre

	A	B	C	D	E	F	G	H
1	Período	Código	Nome	Turma	Horário	id_aluno	Nota	Situação
2	2	ENE0009	ARQUITETURA E PROTOCOLOS DE REDES	A	2M34 4M34	aluno220	3.00	APROVADO
3	2	ECL0042	MEIO AMBIENTE FÍSICO E ECOSSISTEMAS	B	3N34 6N12	aluno220	3.00	APROVADO
4	2	ENE0013	ALGORITMOS E ESTRUTURA DE DADOS	A	3M34 5M34	aluno224	4.00	APROVADO
5	2	ENE0177	ELETROMAGNETISMO 1	B	2T45 4T45	aluno224	4.00	APROVADO
6	2	ENE0022	PROJETO TRANSVERSAL EM REDES DE COMUNICAÇÃO	A	3M12 5M12	aluno224	5.00	APROVADO
7	2	ENE0043	LABORATÓRIO DE CIRCUITOS ELÉTRICOS	I	4M5 4T1	aluno184	3.00	APROVADO
8	2	IFD0177	FISICA 2 EXPERIMENTAL	D	6T2345	aluno184	4.00	APROVADO
9	2	MAT0059	MÉTODOS MATEMÁTICOS DA FÍSICA 1	A	2M34 4M34 6M34	aluno184	4.00	APROVADO
10	2	FGA0150	PROJETO INTEGRADOR DE ENGENHARIA 1	E	2T45 4T45	aluno052	3.00	APROVADO
11	2	FGA0168	DESENHO INDUSTRIAL ASSISTIDO POR COMPUTADOR	D	3T45 5T45 6T45	aluno052	3.00	APROVADO
12	2	FGA0184	GESTÃO DA PRODUÇÃO E QUALIDADE	A	3T23 5T23	aluno052	5.00	APROVADO
13	2	MAT0075	CÁLCULO DE PROBABILIDADE 1	A	2M34 3M34 5M34	aluno326	3.00	APROVADO

Figura 3.1: Parte do conjunto de dados utilizado

(ou ano letivo) e depois podem cursar disciplinas a escolha, observando os pré requisitos necessários. Esse tipo de sistema permite flexibilidade na construção do plano de estudo. A seguir é apresentado um resumo dos tipos de disciplinas presentes na currículo de um curso da UnB:

- **Disciplinas obrigatórias:** As disciplinas obrigatórias são aquelas que estão presentes na matriz curricular do curso e que devem ser cursadas por todos os alunos, pois este deverá se matricular e ser aprovado a fim de obter o diploma.
- **Disciplinas optativas:** As disciplinas optativas também fazem parte do currículo do seu curso, mas não são obrigatórias. Em caso de aprovação, as que foram escolhidas para serem cursadas serão contabilizadas para integralização curricular. Cada graduação tem uma quantidade mínima dessas disciplinas que precisam ser feitas.
- **Disciplinas de módulo livre:** As disciplinas de módulo livre são todas as disciplinas da UnB que não constam no currículo do curso e que não são obrigatórias ou optativas. É possível cursá-las em qualquer curso dentro da universidade, tendo um máximo de 24 créditos que podem ser obtidos.

No conjunto de dados, temos informações quanto à situação do aluno na matéria ao final de um semestre, a qual pode ser: aprovado, reprovado e trancamento. São utilizados apenas dados de matérias em que os alunos foram aprovados, para a geração da recomendação. Alunos aprovados podem ter notas 3, 4 ou 5, correspondentes às menções (conceitos) necessárias para aprovação na referida universidade, onde 3 é a nota mínima para aprovação. Considerando só as disciplinas com ao menos uma aprovação, temos uma total de 357 matérias distintas cursadas, sendo destas apenas 48 obrigatórias do curso de Redes de Comunicação.

3.2 Algoritmo proposto

Para apresentar o algoritmo proposto é necessário primeiramente expor como é feita a representação do antígeno e dos anticorpos.

Tabela 3.1: Exemplo da matriz usuário-item.

	aluno001	aluno002	aluno003	aluno004	aluno005
ADM0023	3	3	3		3
DEG0204	3			4	
BOT0004		2			
ENE0045	4		5		4
ENE0010			3	4	3

O algoritmo utilizado é uma técnica de filtragem colaborativa baseada em memória, definido na seção 2.1, ou seja, as recomendações são computadas através da matriz usuário-item e precisam ter essa estrutura em sua memória. Para a geração da recomendação no sistema proposto, são utilizadas apenas as seguintes informações: identificador do aluno, identificador da matéria e valor da nota obtida, outras informações disponíveis não foram consideradas para construir o sistema de recomendação. A Tabela 3.1 ilustra a matriz usuário-item para o problema, cada aluno é visto como um vetor de tamanho 1×357 , em que o comprimento é o número de matérias que podem ser cursadas e a posição identifica a matéria. Cada número do vetor representa a nota do aluno em uma matéria. Por exemplo, o aluno0001 é representado pelo vetor: $[3, 3, , , 4,]$, os espaços em branco representam matérias não cursadas pelo aluno. Os valores possíveis para as notas são 3, 4 e 5 e representam, respectivamente, as menções MM, MS e SS do método de avaliação da UnB. Sendo MM a nota mínima para aprovação e SS a nota máxima que pode ser obtida.

O aluno alvo da recomendação, tomado como antígeno, deverá ser representado da mesma maneira para que seja possível calcular a similaridade com os anticorpos, que são os demais alunos presentes na matriz usuário-item. Por exemplo, tomando o aluno001 como antígeno, os demais alunos (aluno002, aluno003, aluno004 e aluno005) serão os anticorpos, os quais, dependendo da similaridade com o antígeno e entre si, poderão fazer parte da rede de anticorpos final, i.e. a região de vizinhança.

O Algoritmo 1 descreve o pseudo-código do sistema de recomendação proposto. Para a criação da rede inicial, alunos foram escolhidos aleatoriamente entre os disponíveis na memória do sistema. A concentração inicial de cada anticorpo é definida como: $c_{i0} = 10$. A cada iteração do SIA é calculada a nova concentração dos anticorpos conforme a equação (3.1): se o valor for menor ou igual a zero, o anticorpo é eliminado da rede e outro é adicionado em seu lugar. A procura por novos anticorpos é interrompida quando não há mais alunos a serem analisados ou quando a rede for estabilizada, o que consideramos ocorrer quando nenhum anticorpo é eliminado da rede em 10 iterações. A busca por novos anticorpos também pode ser interrompida quando não houver mais alunos para serem analisados como um potencial anticorpo.

Após a obtenção de uma rede estabilizada, como observado em [Cayzer and Aickelin 2005], é necessário um procedimento de diferenciação das concentrações dos anticorpos. Esse passo é necessário pois anticorpos adicionados à rede há

Entrada: Aluno alvo, k_1 , k_2 , k_3 , n

Resultado: Lista de recomendação

```
1 Selecionar  $n$  alunos aleatoriamente para compor a rede inicial;
2 enquanto AIS não estabilizado & alunos disponíveis faça
3   | enquanto AIS com número máximo de anticorpos faça
4   |   | para cada Anticorpo faça
5   |   |   |  $c_i = c_i + \Delta c$ 
6   |   |   | se  $c_i \leq 0$  então
7   |   |   |   | remover anticorpo  $i$  da rede
8   |   |   | fim
9   |   |   | se  $c_i \geq 100$  então
10  |   |   |   |  $c_i = 100$ 
11  |   |   | fim
12  |   | fim
13  | fim
14  | Adicionar mais anticorpos na rede se ao final da interação algum anticorpo for removido
15 fim
16 Redefinir para a concentração inicial todos os anticorpos da rede final;
17 enquanto Nenhum anticorpo atingir concentração máxima faça
18  | para cada Anticorpo faça
19  |   |  $c_i = c_i + \Delta c$ 
20  |   | se  $c_i \leq 0$  então
21  |   |   | remover anticorpo  $i$  da rede
22  |   | fim
23  |   | se  $c_i \geq 100$  então
24  |   |   |  $c_i = 100$ 
25  |   | fim
26  | fim
27 fim
28 Fazer lista com todos as matérias presentes na rede;
29 para cada Matéria da rede faça
30  | Predizer menção conforme Equação 3.3
31 fim
32 Selecionar todas as matérias com nota 4 ou maior para uma lista;
33 Selecionar destas quais foram cursadas por 2 ou mais alunos;
34 Retirar desta lista matérias obrigatórias e optativas já cursadas pelo Aluno-alvo;

Algoritmo 1: Sistema de recomendação baseado no inumo-inspirado adaptado de [Cayzer and Aickelin 2005].


```

mais tempo atingem a saturação (concentração máxima $c_i = 100$) primeiramente aqueles não teriam o tempo necessário para maturação, estando assim em desvantagem. Uma nova iteração do SIA é feita, desta vez até que um dos anticorpos atinja a concentração máxima, e ainda é possível que algum anticorpo seja eliminado da rede.

A diferença no algoritmo usado no [Cayzer and Aickelin 2005] está no fato de que não estamos olhando para qual curso deve ter a nota prevista, e assim, ao contrário do que ele fez, não é necessário que o aluno tenha feito um curso específico para entrar na rede. A lista de recomendações será extraída da rede formada e os cursos selecionados para a lista deverão ter sido cursados por mais de um desses alunos.

Com a definição das concentrações finais, elas são utilizadas para a predição da nota de todas as matérias presentes na rede. Finalmente é aplicado um filtro que define quais matérias formam a lista de recomendação final. Essas estratégias, importantes para o funcionamento

do sistema de recomendação, são apresentadas a seguir.

3.2.1 Cálculo da similaridade

A similaridade aqui é a função *fitness* do algoritmo imunológico. Essa função é utilizada com o intuito de favorecer a reprodução dos anticorpos que tiveram contato com o antígeno e tiveram uma afinidade.

Para o cálculo da similaridade entre alunos foi utilizado o coeficiente de Pearson. Esta medida foi escolhida por ser aplicada com sucesso para calcular correlações em diversos trabalhos anteriores [Cayzer and Aickelin 2005, Chang et al. 2016, Esteban et al. 2020, Acilar and Arslan 2009]:

$$s_{x,y} = \frac{\sum_{i=1}^n (r_{x,i} - \bar{r}_x)(r_{y,i} - \bar{r}_y)}{\sqrt{\sum_{i=1}^n (r_{x,i} - \bar{r}_x)^2 \sum_{i=1}^n (r_{y,i} - \bar{r}_y)^2}}, \quad (3.2)$$

onde n é o número de disciplinas cursadas tanto pelo aluno x quanto pelo aluno y , \bar{r}_x denota a nota média de todas as matérias cursadas por cada aluno x , $r_{x,i}$ é a nota obtida pelo aluno x na matéria i . A magnitude de $s_{x,y}$ determina em que grau os alunos estão correlacionados, enquanto o sinal indica a direção da relação. Valores próximos a 1 indicam um alto grau de correlação entre os alunos, i.e. o padrão de notas dos alunos é semelhante, enquanto valores próximos de -1 indicam um padrão de notas oposto. Quando os alunos não cursaram nenhum curso em comum e $n = 0$, ou quando a variância for zero e $\sum_{i=1}^n (r_{x,i} - \bar{r}_x)^2 \sum_{i=1}^n (r_{y,i} - \bar{r}_y)^2 = 0$, não é possível calcular o coeficiente de Pearson, nestes casos a correlação é definida como $s_{x,y} = 0$.

3.2.2 Predição da nota

Para prever as notas dos alunos em disciplinas que eles não cursaram é adaptada uma fórmula que já foi utilizada também por [Chang et al. 2016] e [Adomavicius and Kwon 2012]. Usando os dados de histórico do aluno, media de notas em disciplinas, a similaridade do aluno alvo como os outros alunos da rede final de anticorpos e a concentração destes, calculamos a nota como uma media ponderada, Equação 3.3.

$$p_i = \bar{r} + \frac{\sum_{Ab=1}^k w_{Ag,Ab}(r_{Ab,i} - \bar{r}_{Ab})}{\sum_{Ab=1}^k w_{Ag,Ab}}, \quad (3.3)$$

Onde p_i é a nota prevista do aluno alvo (Ag) na matéria i , \bar{r} é a média de notas do aluno Ab, k é o número total de anticorpos da rede que cursaram a matéria i , $r_{Ab,i}$ é a nota do anticorpo Ab na matéria i e $w_{Ag,Ab}$ é o peso do anticorpo Ab , definido na Equação 3.4.

$$w_{Ag} = r_{Ag,Ab} \times c_{Ab}, \quad (3.4)$$

Onde $r_{Ag,Ab}$ é a similaridade entre o anticorpo e o antígeno e c_{Ab} é a concentração do anticorpo na rede final. Assim é a concentração dos anticorpo Ag na rede influencia no peso, de modo que mais peso é dado aos alunos (anticorpos) com maior concentração para a previsão de nota em um curso.

3.2.3 Seleção de disciplinas

Após a predição das notas de todas as matérias presentes na rede são selecionadas para recomendação final apenas matérias optativas do curso com nota predita acima de 4, isso porque a nota de mínima para aprovação é 3 e para ter certeza que a recomendação considera apenas disciplinas que o aluno terá bom desempenho esse limiar é ajustado. Também é exigido que no mínimo 2 alunos da rede tenham cursado a matéria, de modo a evitar disciplinas recomendadas que fujam muito do escopo do curso. Isto foi levando em conta pois alguns alunos tem em seu histórico dados de matérias que foram cursadas em um curso anterior de graduação. Outra vantagem de se utilizar este filtro é o fato de ter mais segurança na predição da nota, tendo mais de uma nota como base para fazer a predição.

Por fim, a lista de recomendação final não contém matérias já cursadas pelo aluno alvo mas presentes na rede, por razões óbvias.

3.3 Métricas de avaliação do desempenho

Selecionou-se um conjunto de métricas utilizadas para avaliar o sistema. Elas foram definidas com base nos trabalhos previamente citados sobre sistemas de recomendação de disciplinas e no trabalho que apresenta o SIA como sistema de recomendação [Cayzer and Aickelin 2005].

- Número de alunos visitados: contabiliza quantos alunos foram necessários avaliar antes da estabilização da rede de anticorpos. Essa métrica é importante para verificar se a taxa de estímulo está bem ajustada. Se o sistema precisou verificar muitos alunos quer dizer que nenhum anticorpo é suficientemente semelhante ao antígeno para fazer parte da rede final e se o número for pequeno a mínima semelhança é suficiente, sendo assim não haveria a procurar por alunos mais apropriados.
- Número de anticorpos na rede final: é o número final de vizinhos na rede estabilizada, obtido após rodar novamente o SIA para a diferenciação das concentrações, passo que ainda pode eliminar anticorpos da rede. Verifica se a rede está sendo preenchida.
- Erro médio absoluto (MAE): mede a magnitude média dos erros em um conjunto de previsões, é calculado sobre todas as matérias presentes na rede que já foram cursadas

pelo aluno alvo, Equação 3.5

$$\text{MAE} = \frac{1}{m} \sum_{i=0}^m |p_i - r_i| \quad (3.5)$$

onde m é o número de matérias que tiveram a nota predita. É a média das diferenças absolutas entre a previsão e a observação real, onde todas as diferenças individuais têm peso igual. É usada para verificar a acurácia das recomendações, valores baixos indicam que a nota predita e a nota real são semelhantes e que o sistema tem bom desempenho.

- Número de recomendações: Número total de matérias aptas para recomendação, ou seja, se foi cursada por pelo menos dois alunos presentes na rede e a nota predita foi maior que 4. Um número alto de recomendações indica que o sistema consegue alcançar uma cobertura maior de disciplinas e conseqüentemente consegue explorar melhor o conjunto de disciplinas relevantes.
- Taxa de sobreposição na recomendação (*Revocação*): essa medida busca avaliar a qualidade da recomendação. Para a geração da recomendação a um aluno alvo, a lista de matérias optativas cursada por ele é escondida do algoritmo. Assim, o sistema deve gerar recomendações apenas com base nos dados do histórico de disciplinas obrigatórias, ao final a lista resultante do processo é comparada com a lista de disciplinas realmente cursadas. Em termos matemático:

$$\frac{\#(\text{optativas cursadas} \cap \text{lista de recomendação})}{\#\text{optativas cursadas}} \quad (3.6)$$

onde $\#(\text{optativas cursadas} \cap \text{lista de recomendação})$ é o quantidade de disciplinas que foram cursadas pelo aluno e aparecem na lista de recomendação e $\#\text{optativas cursadas}$ são quantas disciplinas optativas o aluno cursou ao total. Altas taxas querem dizer que o sistema consegue prever as matérias relevantes aos alunos alvos e indica quão completos são os resultados. Um valor alto significa que um algoritmo retorna a maioria dos resultados relevantes.

- Precisão: É a medida de quantos cursos presentes na lista de recomendações são realmente relevantes para o aluno alvo:

$$\frac{\#(\text{optativas cursadas} \cap \text{lista de recomendações})}{\#\text{lista de recomendações}} \quad (3.7)$$

É uma medida que indica quão válidos são os resultados, ou seja, uma maior precisão significa que um algoritmo retorna resultados mais relevantes do que irrelevantes.

3.4 Considerações finais do capítulo

Neste capítulo foram descritas as principais decisões tomadas para implementação de um sistema de recomendação de disciplinas baseado no sistema imunológico artificial. O problema que este trabalho se propõe a resolver é a geração de recomendações diversas, o que ainda é pouco explorado no contexto da recomendação de disciplinas, mas pode oferecer resultados positivos uma vez que alunos possuem interesses diversos. O banco de dados utilizado para a geração da recomendação é referente às notas de alunos de graduação do curso de Engenharia de Redes de Comunicação da Universidade de Brasília, em que informações como identificador do aluno (anonimizado), identificador da matéria e valor da nota obtida foram usadas para formar a matriz usuário-item, e assim, construir o sistema de recomendação baseado em memória.

O algoritmo proposto foi apresentado e se utiliza da concentração de anticorpos e a similaridade destes com o antígeno para a predição de notas das disciplinas optativas a serem cursadas pelo aluno alvo. Ele se difere do apresentado por [Cayzer and Aickelin 2005] pois não olhamos para qual disciplina deve haver a nota prevista, e assim, não é necessário que o aluno tenha feito um matéria específica para entrar na rede. A rede será composta de vários anticorpos que cursaram matérias em comum com o antígeno, mas não necessariamente entre si e a lista de recomendações será extraída desta.

Para o cálculo da similaridade entre os componentes da rede foi utilizado o coeficiente de Pearson, medida que já foi aplicada com sucesso para calcular correlações em trabalhos anteriores. A previsão da nota é calculada usando os dados de histórico do aluno, media de notas em disciplinas, a similaridade do aluno alvo como os outros alunos da rede final de anticorpos e a concentração destes. Além disso foram discutidas as regras que devem ser levadas em conta para que uma disciplina esteja apta a ser recomendada, ou seja, ter sido cursada por pelo menos dois alunos que compõem a rede final e nota prevista acima de 4.

Por fim foram discutidos as medidas que são utilizadas para medir a qualidade da recomendação, são elas: Número de alunos visitados, Número de anticorpos na rede final, MAE, Número de disciplinas aptas a recomendação, Taxa de sobreposição na recomendação (Revocação) e Precisão.

No capítulo a seguir discutiremos os resultados da aplicação do algoritmo proposto.

Capítulo 4

Resultados Experimentais

Neste capítulo iniciaremos apresentando os experimentos feitos para encontrar as taxas ideais de estímulo e supressão. Primeiramente foram geradas recomendações sem a presença da taxa de supressão, a fim de se analisar os efeitos de taxa de estímulo isoladamente, os resultados são comparados com a recomendação gerada por uma técnica básica de filtragem colaborativa a fim de garantir que haja uma vantagem na utilização da taxa de estímulo.

Ao final do primeiro experimento, foram escolhidas os três valores que levaram a melhores desempenhos. Dai se passou para a definição da taxa de supressão, o que era necessário porque a interação entre as taxas poderia fazer com que o melhor valor de taxa de estímulo, encontrado no primeiro experimento, não continuasse como a melhor escolha pois é preciso levar em consideração os efeitos que as duas taxas exercem uma na outra. Desta maneira então se definem os valores que geram melhores resultados.

Por fim verificamos o desempenho do sistema de recomendação para um grupo de alunos de teste. Recomendações são geradas para cada um deles e os resultados para a lista de completa de recomendação, i.e. com todas as disciplinas de nota prevista acima de 4, são analisados levando em consideração as métricas dispostas no Capítulo 3. Também é analisada a qualidade da recomendação se considerando apenas a lista top-5.

Para que fosse possível gerar recomendações aos alunos disponíveis, garantimos que cada aluno cursou pelo menos 6 disciplinas, sendo 3 obrigatórias, para ser possível achar correlações com outros alunos e 3 optativas, a fim de haver possibilidade de avaliar a taxa de sobreposição da recomendação (Equação 3.6).

Em todos os experimentos a recomendação foi feita primeiramente com o SIA proposto, com o número máximo de anticorpos na rede igual a $n = 20$ (valor definido por experimentos empíricos feitos anteriormente) e através de uma heurística simplificada, o *k-nearest neighbors algorithm*, também conhecido como KNN ou *k*-NN, para a obtenção de uma vizinhança (grupos de usuários semelhantes).

O *k*-NN é aqui é uma técnica que se utiliza apenas da proximidade para definir uma vizinhança. Para determinar quais pontos de dados estão mais próximos de um determinado

ponto de consulta, se calcula a distância entre o ponto de consulta e os outros pontos de dados, aqui será usada a mesma medida que a utilizada no SIA, ou seja, o coeficiente de Pearson (Equação 3.2). Para uma comparação justa, o número k de alunos mais similares é igual ao número de anticorpos presentes na rede final do SIA.

4.1 Taxa de estímulo

Nesta primeira parte dos experimentos, foi utilizado um conjunto de dados de treinamento, do qual 80% dele compõe a matriz usuário-item. É com base nesta que serão feitas recomendações para o resto dos alunos deste conjunto. A fim de encontrar os valores ideais da taxa de estímulo e supressão, em cada taxa analisada foram feitas predições para 20 alunos selecionados aleatoriamente, para cada um destes o algoritmo é executado 5 vezes ao final são feitas as médias estatísticas dos resultados.

Primeiramente é feita a análise da taxa de estímulo quando não há presença da taxa de supressão, ou seja, ela é dita como igual a zero. Na equação da concentração de anticorpos do nosso sistema (Equação 3.1) a taxa de estímulo é a responsável por incentivar padrões de similaridade do anticorpo com o antígeno, assim, se o anticorpo for suficientemente similar ao antígeno sua proliferação será estimulada, logo terá concentração maior.

Assim a taxa de estímulo é a responsável por fornecer um limiar de correlação entre o antígeno e o anticorpo, pois sem a presença da taxa de supressão, a correlação de um anticorpo ponderada pela taxa de estímulo deverá superar a taxa de mortalidade para continuar na rede, caso contrário, será eventualmente eliminado.

Pelo gráfico da Figura 4.1 pode ser verificado que, em baixas taxas de estimulação, é

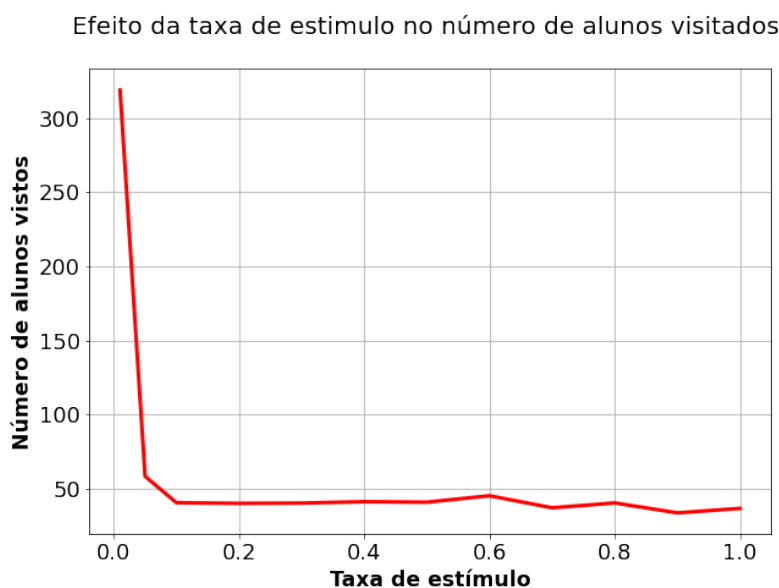


Figura 4.1: Efeito a taxa de estímulo no número de alunos visitados.

necessário visitar quase todos os alunos presentes na matriz em memória. Por outro lado, em taxas de estimulação muito altas, pode não ser necessário examinar outros alunos fornecidos para que o SIA construa e a rede a qual tende a se estabilizar rapidamente. Se o limiar de semelhança, ajustado pela taxa de estímulo, for muito baixo, a rede final tende a ser composta pelos primeiros anticorpos visitados que mostraram alguma semelhança ao antígeno. Isso é prejudicial ao sistema pois não dá espaço para a eventual adição de anticorpos mais semelhantes, pois eles sequer chegaram a ser visitados pelo algoritmo. A presença de um limiar mínimo de correlação é importante pois opera na manutenção da qualidade de uma vizinhança, eliminando alunos pouco correlacionados.

Diferentemente do que foi apresentado em [Cayzer and Aickelin 2005], o número de anticorpos na rede sempre foi preenchido ao máximo, uma vez que temos alunos suficientes semelhantes ao aluno alvo. Como estamos trabalhando com a recomendação de disciplinas para alunos do mesmo curso e que seguem um currículo de disciplinas obrigatórias definido pela universidade, não há o problema de preenchimento da matriz usuário-item comumente encontrado em sistemas de recomendação que trabalham com muitos itens e usuários, e conseqüentemente não há a dificuldade de achar usuários com alguma correlação. Mesmo com as taxas de estímulo relativamente mais baixas, ou seja, para que o anticorpo entre na rede é necessário que ele seja muito similar ao antígeno, ainda foi possível completar a rede com o número máximo de anticorpos definido.

Este sistema gera um maior número de recomendações, comprando aos resultados do k -NN, para qualquer taxa de estímulo. Este fenômeno ocorre mesmo sem a presença da taxa de supressão, alguma diversidade nos anticorpos já está garantida, uma vez que estes são adicionados aleatoriamente na rede, comportamento verificado no gráfico da Figura 4.2. Com uma taxa de estímulo ótima a rede é capaz de gerar retornar mais disciplinas na lista de

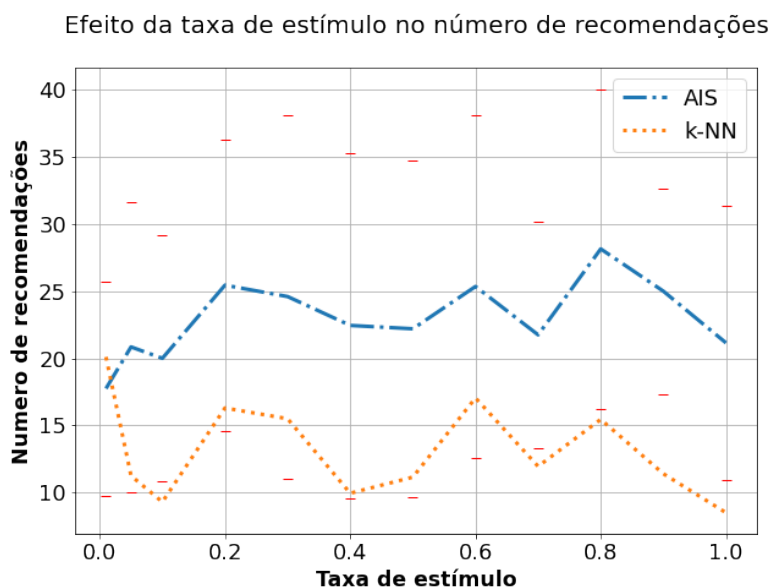


Figura 4.2: Efeito da taxa de estímulo no número de recomendações.

recomendação.



Figura 4.3: Efeito da taxa de estímulo no erro absoluto médio.

No gráfico da Figura 4.3 verifica-se que, para qualquer taxa de estímulo, o MAE do SIA é menor, atingindo valores entre 0,54 e 0,58, enquanto a predição feita pela heurística simplificada atingiu valores entre 0,57 e 0,61. O gráfico também mostra que as curvas tem tendências parecidas atingindo vales e picos nos mesmos pontos. O sistema também é avaliado em sua capacidade de precisão na recomendação, o gráfico da Figura 4.4 mostra que, para taxas acima de 0,05, a taxa de sobreposição é maior. É verificado que as recomendações do SIA são significativamente melhores que o k -NN tanto em relação ao erro na predição da nota quanto na precisão da recomendação.

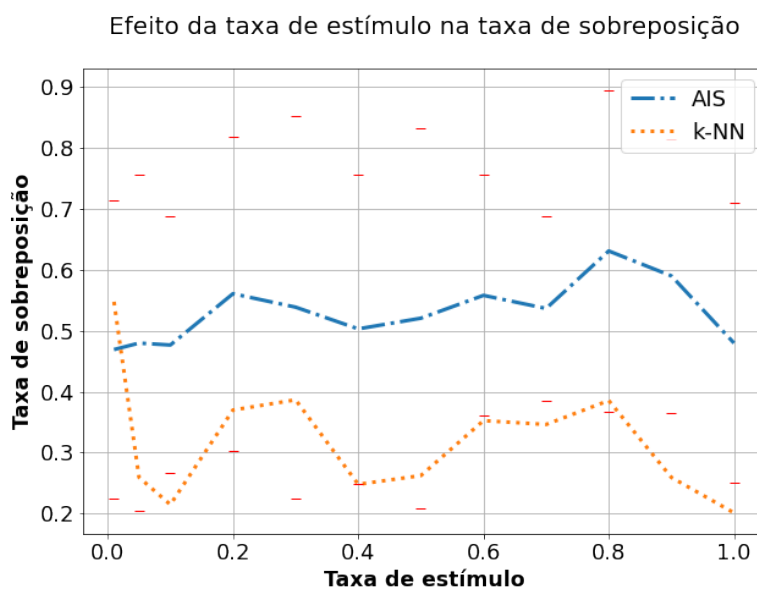


Figura 4.4: Efeito da taxa de estímulo na taxa de sobreposição da recomendação.

Quando a taxa de estímulo é muito baixa, 0,01, todos os anticorpos, para fazerem parte da rede, precisam ser bastante semelhantes ao antígeno e, sem a presença da taxa de supressão para garantir diversidade entre eles, é suposto que a vizinhança seja bastante semelhante à vizinhança composta pelos alunos mais semelhantes. Isso é comprovado pelos gráficos das Figuras 4.2 e 4.4, tanto o número de recomendações quanto a taxa de sobreposição tem valores bem próximos. Apesar disso, o erro de predição pelo SIA continua sendo melhor pois mesmo tendo os mesmos vizinhos, o peso no cálculo da nota é diferente. Isto nos leva a crer que a utilização da concentração como peso tem ligeira vantagem sobre a técnica básica na predição de notas.

4.2 Taxa de supressão

Quando adicionado o efeito da taxa de supressão cria-se, “mais uma exigência” para que o anticorpo permaneça na rede. Ele deverá ser suficientemente diferente dos seus pares. Desse modo, a taxa de supressão é a responsável por controlar diversidade na rede imunológica.

A taxa de supressão é então vista como um artifício que adiciona um refinamento na rede numa vizinhança já considerada “boa”, vizinhança definida pela taxa de estímulo. Para fazer a análise da taxa de supressão, foram escolhidos valores adequados para a taxa de estímulo, ou seja, 3 dos melhores valores para k_1 na primeira parte do experimento: 0,1, 0,2 e 0,3, pois com estes valores foram obtidos bons resultados em termos de MAE reduzido, taxas altas de sobreposição na recomendação e números razoáveis de alunos avaliados antes da estabilização da rede. Fixada a taxa de estímulo, a taxa de supressão foi variada entre 0,01 e 0,7.

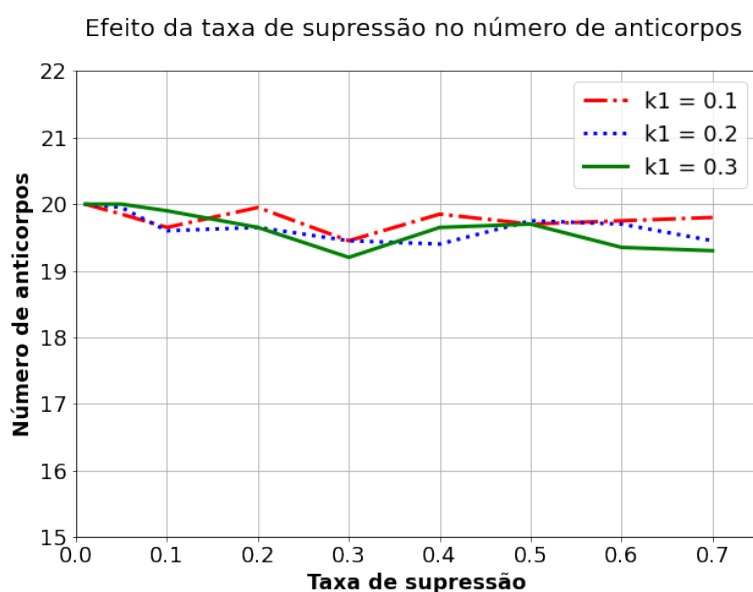


Figura 4.5: Efeito da taxa de supressão no número de anticorpos na rede.

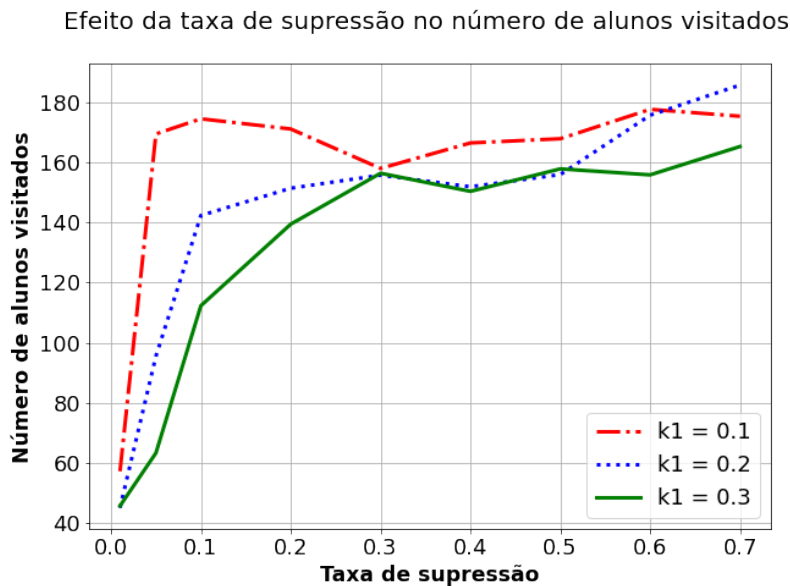


Figura 4.6: Efeito da taxa de supressão no número de alunos visitados.

Primeiramente foi analisado se a taxa de supressão afeta o tamanho da rede final. No gráfico da Figura 4.5, observa-se que há apenas uma pequena variação, ou seja é possível “encher” a rede mesmo com taxas mais altas de supressão. Isto ocorre pois é possível ter correlação negativa entre os anticorpos e o sistema consegue adicionar grande diversidade à rede. Apesar disso, uma exageradamente grande diversidade pode ser prejudicial ao sistema. Para ajustar essa taxa e avaliar o impacto da diversidade, novamente são avaliados o número de alunos visitados, o MAE, o número de recomendações e a taxa de sobreposição. Para a combinação de taxas ideal é esperado um MAE baixo e altas taxas de sobreposição na recomendação.



Figura 4.7: Efeito da taxa de supressão no erro absoluto médio.

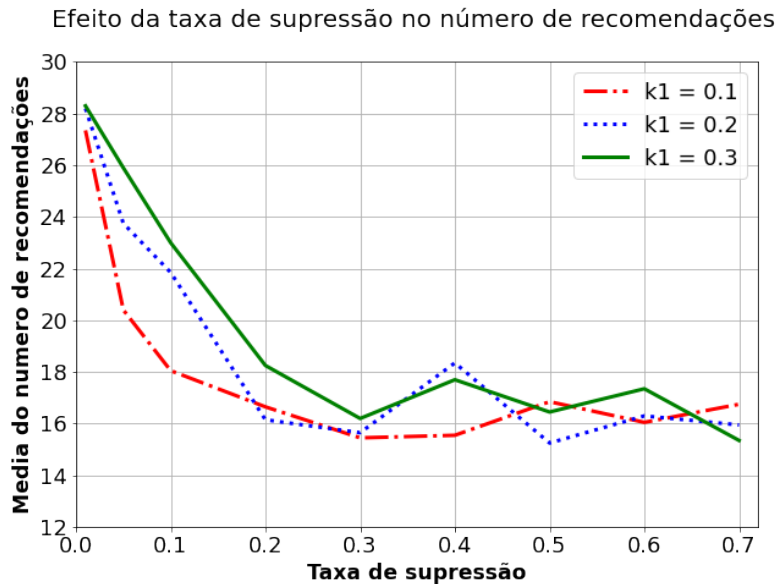


Figura 4.8: Efeito da taxa de supressão no número de recomendações.

Como era esperado, pelo gráfico da Figura 4.6 é notado que o número de alunos visitados é afetado com a presença da taxa de supressão, mas, após superar a taxa de estímulo, este número se estabiliza, o que ocorre é que quando a taxa de supressão é muito alta é mais importante para rede que ela seja diversa entre si do que que o antígeno o que é atingido após verificar aproximadamente 160 alunos.

Assim, os alunos da rede final acabam não sendo muito similares ao alvo da recomendação por isso o erro da predição da nota é alto. Esse comportamento também pode ser atestado pelo gráfico da Figura 4.7. Verifica-se que, nas três curvas, quando a taxa de supressão supera o valor da taxa de estímulo o MAE aumenta muito, ou seja, o desempenho

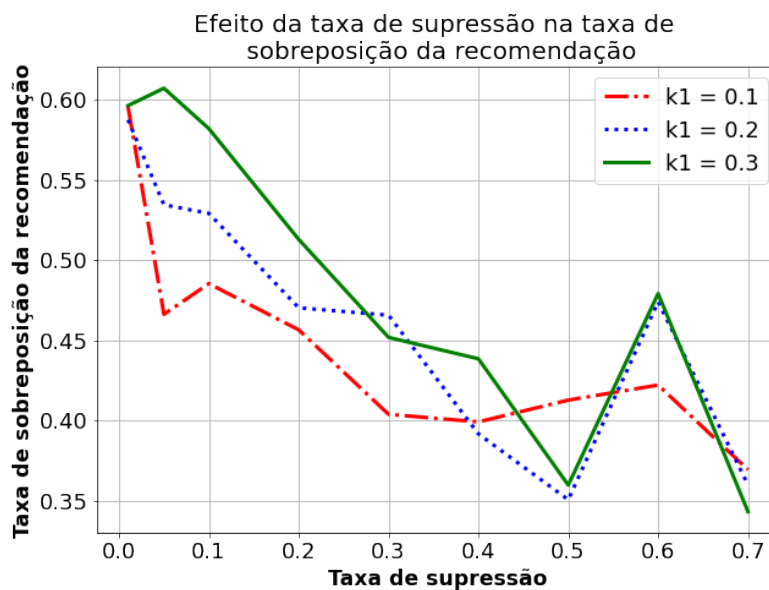


Figura 4.9: Efeito da taxa de supressão na taxa de sobreposição da recomendação.

da recomendação é bastante afetado. Para valores de taxa de estímulo menores que a taxa de supressão não se vê muita alteração no erro da predição, com valores de MAE em torno de 0,55, número semelhante ao obtido sem o efeito da taxa de supressão.

Altas taxas de supressão também acabam afetando de maneira acentuada o número de recomendações, comportamento verificado na Figura 4.8. Isso ocorre pois é mais difícil obter matérias que foram cursadas por mais de um aluno presente na rede, uma vez que há grande diversidade entre os alunos. É suposto que os padrões de comportamento dos alunos sejam bem diferentes.

Quando a taxa de supressão é mais baixa, um efeito positivo é observado, para $k_2 = 0.01$ o número médio de recomendações fica em torno de 28 matérias para as três taxas de estímulo, levemente superior ao encontrado sem a presença da taxa de supressão, que varia entre 23 e 25.

Segundo o gráfico da Figura 4.9 quando $k_1 = 0,3$ e $k_2 = 0,05$ o pico da sobreposição na recomendação é atingido, sendo esta a melhor combinação de taxas. Apesar do número de recomendações ser semelhante ao obtido sem a presença da supressão e o MAE não ser afetado com a presença da taxa, é visto que há uma melhora da taxa de sobreposição da recomendação. Deste modo, esta combinação é definida como ideal e será utilizada nos experimentos à seguir.

4.3 Desempenho das recomendações

Nesta seção apresentamos resultados gerais sobre a avaliação do desempenho do algoritmo proposto, além disso, foram realizados testes para verificar como a diversidade de anticorpos afeta a recomendação. O esquema Leave-One-Out [Wang et al. 2010] foi utilizado como estratégia de validação cruzada, ou seja, quando uma recomendação é feita a um aluno alvo, todos os demais alunos presentes no conjunto de dados fazem parte da matriz item-usuário e podem pertencer à rede final de anticorpos. Foram geradas recomendações para cada um dos usuários presentes no conjunto de dados.

Para cada aluno são calculados a taxa de sobreposição, número de recomendações, MAE, precisão e diversidade de recomendações. Devido à natureza estocástica do SIA, em que a execução mais de uma vez para um mesmo aluno pode gerar uma lista de recomendações diferentes, o algoritmo foi executado 10 vezes e, ao final, foram feitas medições estatísticas para os resultados.

Na Figura 4.10, que analisou o MAE versus o número de disciplinas já realizadas pelo aluno, é possível observar que o desempenho tanto do SIA quanto do k-NN são muito próximos, sendo o SIA um pouco melhor. Vê-se também que quando o aluno possui poucas disciplinas em seu histórico acadêmico, o cálculo da previsão pode ser difícil, erros relativamente altos são encontrados para os dois algoritmos avaliados, isso pode ser devido a quais

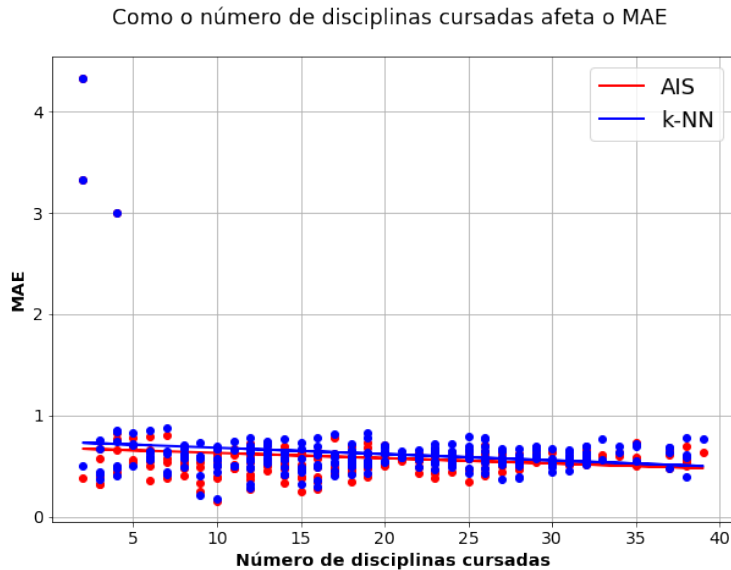


Figura 4.10: Relação entre numero de disciplinas cursadas e MAE

disciplinas obrigatórias foram cursadas pelo aluno alvo, como temos dados disponíveis de apenas um período de 5 anos, algumas disciplinas não foram cursados por muitos alunos na base de dados, no banco de dados disponíveis 50% das disciplinas foram cursadas por menos de 2 alunos, então, se os alunos-alvo cursaram apenas essas disciplinas é muito difícil encontrar equivalentes semelhantes.

Houveram 3 alunos que sofreram com este problema (*outliers* do gráfico da Figura 4.10), dois deles fizeram apenas duas disciplinas obrigatórias e o outro aluno apenas quatro. Para estes casos, verificou-se que o SIA visitou todos os alunos presentes na base antes da esta-

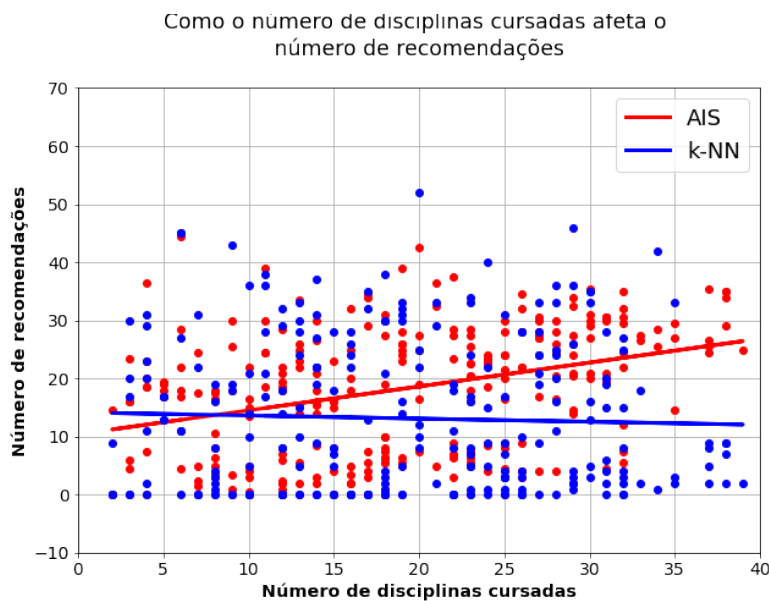


Figura 4.11: Relação entre numero de disciplinas cursadas e o número de disciplinas recomendadas

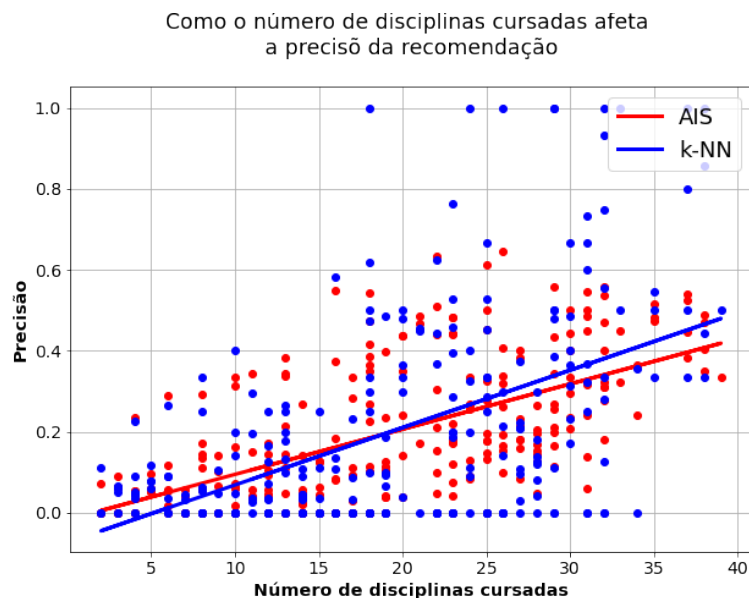


Figura 4.12: Relação entre numero de disciplinas cursadas e a precisão da recomendação

bilização da rede. Ainda analisando este gráfico, verifica-se que quanto maior o número de disciplinas no registro, menor o erro de previsão de ambos os algoritmos.

A maior diferença entre as duas estratégias de recomendação pode ser observada analisando o número de recomendações em relação à quantidade de disciplinas realizadas, na Figura 4.11. Como as estimativas de notas têm valores próximos e nos dois sistemas são recomendados apenas disciplinas com nota prevista acima de 4, parece que, para o SIA, quanto maior o número de disciplinas cursadas, mais o algoritmo tem acesso àquelas que podem ser relevantes para o aluno-alvo. Ou seja, há mais disciplinas presentes na vizinhança do SIA.

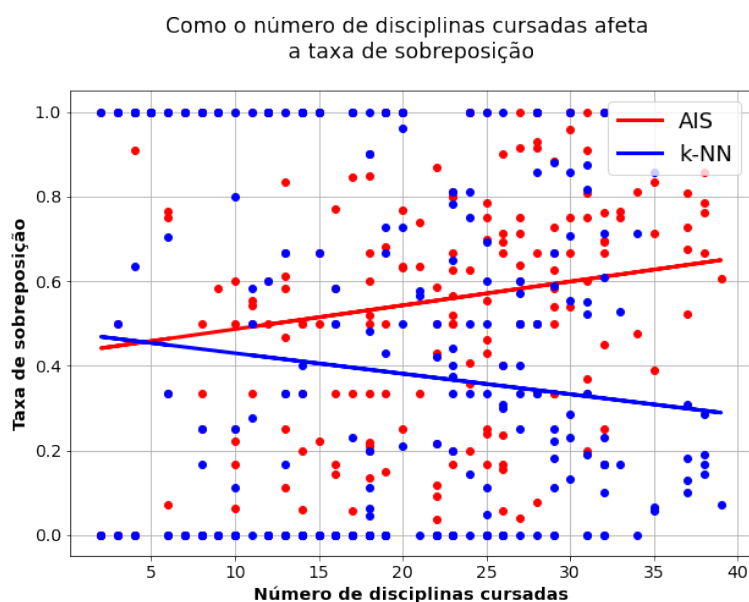


Figura 4.13: Relação entre numero de disciplinas cursadas e a taxa de sobreposição da recomendação

Verifica-se também que com o SIA é mais difícil obter recomendações vazias, sendo possível extrair recomendação de disciplinas para quase todos os alunos, com exceção dos 3 alunos mencionados no parágrafo anterior, onde nenhuma boa previsão pôde ser feita por nenhum dos algoritmos. Usando a heurística k-NN, listas de recomendações vazias foram retornadas para 19% dos alunos.

Mas um número maior de recomendações pode não levar a sugestões de alta qualidade, por isso é importante observar a taxa de sobreposição e a precisão. Na Figura 4.12 vê-se que a precisão dos dois algoritmos é maior quando o aluno possui mais disciplinas no histórico acadêmico, ou seja, a lista de recomendações contém resultados mais relevantes em ambos os casos.

Mas ao analisar a Figura 4.13 verifica-se que, para o SIA, o aumento do número de recomendações se alinha com o aumento da taxa de sobreposição, ou seja, o percentual de disciplinas relevantes devolvidas. O mesmo não ocorre com o k-NN, onde a taxa de sobreposição das recomendações tende a diminuir. Assim, observamos que o SIA, além de retornar mais recomendações, também fornece resultados mais precisos. Outra observação feita é que para 89 alunos, de um total de 250 alunos, o k-NN não conseguiu gerar nenhuma sobreposição, enquanto no SIA isso ocorreu para apenas 34 alunos.

Uma segunda análise é feita, mas agora o algoritmo retorna uma lista com apenas 5 disciplinas para recomendar. Estas são as 5 matérias com as maiores notas previstas. Neste caso, mesmo que levemente, observa-se um melhor desempenho da técnica k-NN ao analisar a revocação, veja a Figura 4.15. Para precisão, Figura 4.14, vê-se que as técnicas seguem a mesma tendência do cenário com a lista de recomendações completas, i.e. com todas as disciplinas de nota prevista acima de 4. Para ambas as técnicas, a taxa de sobreposição tende a ser inferior a 20%, mesmo quando o aluno frequentou um grande número de disciplinas

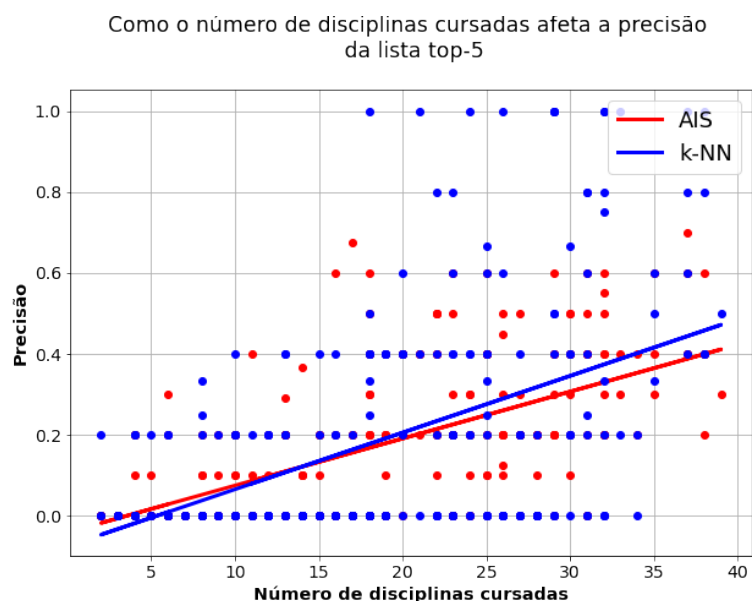


Figura 4.14: Relação entre numero de disciplinas cursadas e a precisão da lista top-5

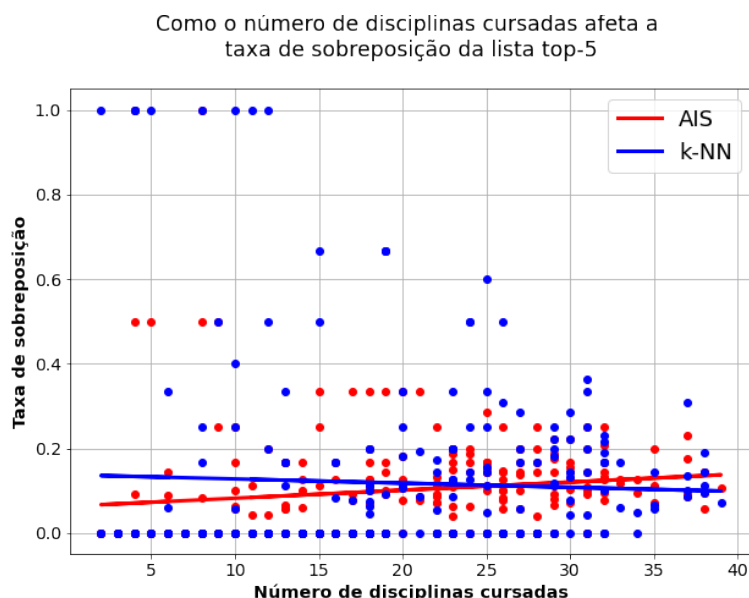


Figura 4.15: Relação entre número de disciplinas cursadas e a taxa de sobreposição da lista top-5

obrigatórias. Isso era esperado, uma vez que o fator de seleção de disciplina para um aluno não se baseia exclusivamente na nota a ser obtida, e esse foi o critério escolhido para ordenar a lista de recomendações. Assim, observa-se que a rede de anticorpos SIA realmente consegue acessar um conjunto maior de disciplinas relevantes, mas para ordenar essa lista outros fatores além da nota prevista precisariam ser levados em consideração.

4.4 Análise da recomendação

Nas tabelas 4.1 e 4.2 temos a ilustração dos resultados da recomendação para alunos que cursaram 3, 6, 20 e 30 disciplinas. Comparando as duas tabelas, vemos que o tanto o SIA quanto o k -NN, quando o aluno cursou poucas disciplinas, tem dificuldade em encontrar grupos de alunos semelhantes. É visto que para o Aluno271 o k -NN não conseguiu gerar nenhuma recomendação, já o SIA extrai 6 disciplinas da rede mas nenhuma delas relevantes para o aluno, veja que a taxa de sobreposição é 0. A diferença entre os alunos 066 e 271 está em quais disciplinas foram cursadas por eles, o 066 cursou matérias que vários outros também cursaram e, por isso, foi mais fácil encontrar vizinhos semelhantes tanto pelo k -NN, quanto pelo SIA.

Na Tabela 4.1 vemos que para alunos com poucas disciplinas cursadas o SIA, precisa visitar um número maior de alunos (de 74 a 197) antes da estabilização, este número pode variar devido à escolha aleatória para a entrada do anticorpo na rede. O desempenho do SIA é melhor quando o aluno cursou mais disciplinas. Mas, para quase todos os alunos o SIA conseguiu retornar mais disciplinas, aliado a maiores taxas de sobreposição e precisão. Ou seja, o SIA tem melhor capacidade de abrangência, conseguindo extrair maior número de

disciplinas relevantes.

Já o MAE das duas técnicas são bem semelhantes. O que mostra que a utilização da concentração para a predição da nota é eficiente para se levar em consideração o grau de semelhança do aluno em questão ao alvo da recomendação e ponderar sua participação na nota predita (Equação 3.3).

Tabela 4.1: Resultados da recomendação usando SIA.

Aluno	Obrigatórias cursadas	Número de anticorpos	Alunos visitados	MAE	Recomendações	Sobreposição	Precisão
aluno066	3	20.0	197.0	0.39	23.5	0.5	0.09
aluno271	3	20.0	74.5	0.67	6.0	0.0	0.0
aluno052	6	20.0	102.5	0.507	22.0	0.077	0.019
aluno160	6	20.0	115.0	0.66	44.5	0.76	0.29
aluno218	20	20.0	67.5	0.71	27.5	0.63	0.44
aluno099	20	20.0	66.0	0.52	19.0	0.64	0.35
aluno262	30	20.0	45.0	0.52	21.0	0.54	0.33
aluno096	30	20.0	45.5	0.53	20.0	0.67	0.44

Tabela 4.2: Resultados da recomendação usando k -NN

Aluno	Obrigatórias cursadas	MAE	Recomendações	Sobreposição	Precisão
aluno066	3	0.44	30	0.5	0.067
aluno271	3	0.75	0	0.0	0.0
aluno052	6	0.63	11	0.0	0.0
aluno160	6	0.64	45	0.71	0.27
aluno218	20	0.71	8	0.21	0.5
aluno099	20	0.54	22	0.73	0.36
aluno262	30	0.53	13	0.23	0.23
aluno096	30	0.57	3	0.13	0.67

Capítulo 5

Conclusão

Essa dissertação utiliza dados de alunos do curso de Engenharia de Redes de Comunicação da Universidade de Brasília para propor e desenvolver um sistema imuno-inspirado de recomendação de disciplinas. Enquanto sistemas de recomendação já foram aplicados no contexto de recomendação de disciplinas, o impacto na qualidade da recomendação ao se ter uma vizinhança diversa entre si havia sido estudado até então.

A utilização do sistema imunológico artificial, principalmente o conceito de rede imunológica regulada pelo mecanismo de concentração de anticorpos para a formação da vizinhança, tem como característica a criação de uma vizinhança que ao mesmo tempo se assemelha ao aluno alvo da recomendação, mas que é diversa entre si. A diversidade da rede é interessante na tentativa de evitar que o sistema de recomendação tenda a concentrar suas recomendações em um subconjunto de itens relevantes.

Foi apresentado uma metodologia de aplicação de um sistema imunológico artificial como técnica de filtragem colaborativa no problema de recomendação de disciplinas. O algoritmo proposto se utiliza da concentração de anticorpos e a similaridade destes com o antígeno para a predição de notas das disciplinas optativas a serem cursadas pelo aluno alvo. Para o cálculo da similaridade entre os componentes da rede foi utilizado o coeficiente de Pearson, medida amplamente utilizada na literatura.

Por fim, uma série de experimentos computacionais foram feitos para avaliar a utilização do SIA. Foi visto que, comparado com uma técnica básica de filtragem colaborativa onde são escolhidos os k alunos melhores correlacionados ao aluno alvo para fazerem parte da vizinhança, há um aumento expressivo no número de recomendações. Além de uma precisão maior na recomendação, calculada a partir da comparação entre as matérias optativas realmente cursadas pelo aluno alvo, mas escondidas do sistema de recomendação, com a lista de matérias recomendadas pelo sistema proposto. Verificou-se que a rede formada pelo SIA leva a obtenção de uma vizinhança que tem abrangência maior do conjunto de matérias relevantes.

O erro da nota predita também foi analisado e melhores resultados foram obtidos uti-

lizando o SIA, o que nos leva a concluir que a utilização da concentração como peso no cálculo da nota tem ligeira vantagem sobre a técnica básica também na predição da nota do aluno.

5.1 Trabalhos Futuros

Trabalhos futuros na análise da composição da vizinhança formada pelo SIA podem ser feitos. Devido à escolha aleatória de anticorpos para a composição da rede inicial é suposto que, ao fazer recomendações mais de uma vez ao mesmo aluno, pode-se esperar listas de disciplinas diferentes, logo, a sensibilidade a este processo precisaria ser estudada. Além disso, é necessária a avaliação mais aprofundada das limitações do algoritmo, pois se observou-se que, por exemplo, a quantidade de disciplinas cursadas anteriormente pelo aluno alvo pode afetar a qualidade da recomendação.

Além disso, a técnica aqui apresentada utiliza somente a nota do aluno para determinar se a disciplina é relevante. Outros objetivos poderiam ser empregados para a avaliação das disciplinas retornadas pela rede imunológica, a fim de serem formadas listas top-n que sejam ainda mais relevantes para o alunos considerando, por exemplo, sua dificuldade e ementa.

5.2 Trabalhos publicados

1. Braz, J. C. D. e Silva, D. G. (2021). *Um sistema imuno-inspirado de recomendação de disciplinas para cursos superiores*. In Anais do 15. Congresso Brasileiro de Inteligência Computacional, pages 1–9. SBIC.
2. Braz, J. C. D. e Silva, D. G. (**Em Revisão**) *An immune-inspired course recommendation system for higher education*. Journal of the Brazilian Computer Society.

Referências Bibliográficas

- [Acilar and Arslan 2009] Acilar, A. M. and Arslan, A. (2009). A collaborative filtering method based on artificial immune network. *Expert Systems with Applications*, 36(4):8324–8332.
- [Addison 1843] Addison, W. (1843). *Experimental and practical researches on inflammation, and on the origin and nature of tubercles of the lungs*. Churchill.
- [Adomavicius and Kwon 2012] Adomavicius, G. and Kwon, Y. (2012). Diversity Using Ranking-Based Techniques. 24(5):896–911.
- [Adomavicius and Tuzhilin 2005] Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749.
- [Al Hassanieh et al. 2018] Al Hassanieh, L., Abou Jaoudeh, C., Abdo, J. B., and Demerjian, J. (2018). Similarity measures for collaborative filtering recommender systems. In *2018 IEEE Middle East and North Africa Communications Conference (MENACOMM)*, pages 1–5. IEEE.
- [Amatriain and Basilico 2015] Amatriain, X. and Basilico, J. (2015). Recommender systems in industry: A netflix case study. In *Recommender systems handbook*, pages 385–419. Springer.
- [Avazpour et al. 2014] Avazpour, I., Pitakrat, T., Grunske, L., and Grundy, J. (2014). Dimensions and metrics for evaluating recommendation systems. In *Recommendation systems in software engineering*, pages 245–273. Springer.
- [Aytekin and Özge 2014] Aytekin, T. and Özge, M. (2014). Clustering-based diversity improvement in top- N recommendation. pages 1–18.
- [Bendakir and Aïmeur 2006] Bendakir, N. and Aïmeur, E. (2006). Using association rules for course recommendation. In *Proceedings of the AAAI workshop on educational data mining*, volume 3, pages 1–10.
- [Burke and Robin 2007] Burke, R. and Robin (2007). Hybrid web recommender systems. volume 4321.

- [Cayzer and Aickelin 2005] Cayzer, S. and Aickelin, U. (2005). A recommender system based on idiotypic artificial immune networks. *Journal of Mathematical Modelling and Algorithms*, 4(2):181–198.
- [Chang et al. 2016] Chang, P.-C., Lin, C.-H., and Chen, M.-H. (2016). A hybrid course recommendation system by integrating collaborative filtering and artificial immune systems. *Algorithms*, 9(3):47.
- [Cheng et al. 2016] Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhya, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., et al. (2016). Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*, pages 7–10.
- [Coelho and Von Zuben 2010] Coelho, G. P. and Von Zuben, F. J. (2010). A Concentration-Based Artificial Immune Network for Continuous Optimization. In *2010 IEEE Congress of Evolutionary Computation (CEC)*, pages 1–8.
- [Corne and Lones 2018] Corne, D. and Lones, M. A. (2018). *Evolutionary Algorithms*, pages 1–22. Springer International Publishing, Cham.
- [Das et al. 2017] Das, D., Sahoo, L., and Datta, S. (2017). A survey on recommendation system. *International Journal of Computer Applications*, 160(7).
- [Dasgupta et al. 2011] Dasgupta, D., Yu, S., and Nino, F. (2011). Recent advances in artificial immune systems: Models and applications. *Applied Soft Computing*, 11(2):1574 – 1587. The Impact of Soft Computing for the Progress of Artificial Intelligence.
- [De Castro and Timmis 2002] De Castro, L. N. and Timmis, J. (2002). Artificial Immune Systems : A Novel Approach to Pattern Recognition. *Artificial Neural networks in pattern Recognition*, pages 67–84.
- [de Castro and Timmis 2002] de Castro, L. N. and Timmis, J. (2002). Artificial immune systems: a novel approach to pattern recognition.
- [De Castro and Von Zuben 1999] De Castro, L. N. and Von Zuben, F. J. (1999). Artificial immune systems: Part i–basic theory and applications.
- [de Castro and Von Zuben 2002] de Castro, L. N. and Von Zuben, F. J. (2002). ainet: an artificial immune network for data analysis. In *Data mining: a heuristic approach*, pages 231–260. IGI Global.
- [De Castro and Von Zuben 2002] De Castro, L. N. and Von Zuben, F. J. (2002). Learning and optimization using the clonal selection principle. *IEEE transactions on evolutionary computation*, 6(3):239–251.

- [de França et al. 2012] de França, F. O., Coelho, G. P., Castro, P. A., and Von Zuben, F. J. (2012). Conceptual and practical aspects of the ainet family of algorithms. In *Nature-Inspired Computing Design, Development, and Applications*, pages 1–38. IGI Global.
- [East et al. 2005] East, R., Hammond, K., Lomax, W., and Robinson, H. (2005). What is the effect of a recommendation? *The Marketing Review*, 5(2):145–157.
- [Elahi et al. 2016] Elahi, M., Ricci, F., and Rubens, N. (2016). A survey of active learning in collaborative filtering recommender systems. *Computer Science Review*, 20:29–50.
- [Elbadrawy and Karypis 2016] Elbadrawy, A. and Karypis, G. (2016). Domain-aware grade prediction and top-n course recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 183–190.
- [Esteban et al. 2020] Esteban, A., Zafra, A., and Romero, C. (2020). Helping university students to choose elective courses by using a hybrid multi-criteria recommendation system with genetic optimization. *Knowledge-Based Systems*, 194:105385.
- [Farzan and Brusilovsky 2006] Farzan, R. and Brusilovsky, P. (2006). Social navigation support in a course recommendation system. In *International conference on adaptive hypermedia and adaptive web-based systems*, pages 91–100. Springer.
- [Fayyaz et al. 2020] Fayyaz, Z., Ebrahimian, M., Nawara, D., Ibrahim, A., and Kashef, R. (2020). Recommendation systems: Algorithms, challenges, metrics, and business opportunities. *applied sciences*, 10(21):7748.
- [Figueredo et al. 2013] Figueredo, G. P., Bernardino, H. S., and Barbosa, H. J. C. (2013). Introdução aos sistemas imunológicos artificiais. In Lopes, H. S., de Abreu Rodrigues, L. C., and Steiner, M. T. A., editors, *Meta-Heurísticas em Pesquisa Operacional*, chapter 8, pages 113–128. Omnipax, Curitiba, PR, 1 edition.
- [Golbandi et al. 2010] Golbandi, N., Koren, Y., and Lempel, R. (2010). On bootstrapping recommender systems. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1805–1808.
- [Golbandi et al. 2011] Golbandi, N., Koren, Y., and Lempel, R. (2011). Adaptive bootstrapping of recommender systems using decision trees. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 595–604.
- [Hart et al. 2011] Hart, E., McEwan, C., Timmis, J., and Hone, A. (2011). Advances in artificial immune systems. *Evolutionary Intelligence*, 4(2):67–68.
- [Isinkaye et al. 2015] Isinkaye, F. O., Folajimi, Y. O., and Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian informatics journal*, 16(3):261–273.

- [Jerne 1974] Jerne, N. (1974). Towards a network theory of the immune system. *Annales d'immunologie*, 125C(1-2):373—389.
- [Ji and Dasgupta 2007] Ji, Z. and Dasgupta, D. (2007). Revisiting negative selection algorithms. *Evolutionary computation*, 15(2):223–251.
- [Jiang et al. 2019] Jiang, W., Pardos, Z. A., and Wei, Q. (2019). Goal-based course recommendation. In *Proceedings of the 9th international conference on learning analytics & knowledge*, pages 36–45.
- [Karabadji et al. 2018] Karabadji, N. E. I., Beldjoudi, S., Seridi, H., Aridhi, S., and Dhifli, W. (2018). Improving memory-based user collaborative filtering with evolutionary multi-objective optimization. *Expert Systems with Applications*, 98:153–165.
- [Karimi et al. 2011] Karimi, R., Freudenthaler, C., Nanopoulos, A., and Schmidt-Thieme, L. (2011). Non-myopic active learning for recommender systems based on matrix factorization. In *2011 IEEE International Conference on Information Reuse & Integration*, pages 299–303. IEEE.
- [Koren 2010] Koren, Y. (2010). Factor in the neighbors: Scalable and accurate collaborative filtering. *ACM Trans. Knowl. Discov. Data*, 4(1).
- [Kuanr et al. 2018] Kuanr, M., Rath, B. K., and Mohanty, S. N. (2018). Crop recommender system for the farmers using mamdani fuzzy inference model. *International Journal of Engineering & Technology*, 7(4.15):277–280.
- [Lynn and Emanuel 2020] Lynn, N. and Emanuel, A. (2020). A review on recommender systems for course selection in higher education.
- [Madani et al. 2019] Madani, Y., Erritali, M., Bengourram, J., and Sailhan, F. (2019). Social collaborative filtering approach for recommending courses in an e-learning platform. *Procedia Computer Science*, 151:1164–1169. The 10th International Conference on Ambient Systems, Networks and Technologies (ANT 2019) / The 2nd International Conference on Emerging Data and Industry 4.0 (EDI40 2019) / Affiliated Workshops.
- [Mello et al. 2010] Mello, C. E., Aufaure, M.-A., and Zimbrao, G. (2010). Active learning driven by rating impact analysis. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 341–344.
- [Mettenleiter 2017] Mettenleiter, T. C. (2017). The first “virus hunters”. In *Advances in virus research*, volume 99, pages 1–16. Elsevier.
- [Mu et al. 2009] Mu, Z., Yi, C., Xiaohong, Z., and Junyong, L. (2009). Study on the recommendation technology for tourism information service. In *2009 Second International Symposium on Computational Intelligence and Design*, volume 1, pages 410–415. IEEE.

- [Nguyen et al. 2020] Nguyen, B. H., Xue, B., and Zhang, M. (2020). A survey on swarm intelligence approaches to feature selection in data mining. *Swarm and Evolutionary Computation*, 54:100663.
- [Parameswaran and Garcia-Molina 2009] Parameswaran, A. G. and Garcia-Molina, H. (2009). Recommendations with prerequisites. In *Proceedings of the third ACM conference on Recommender systems*, pages 353–356.
- [Park 2018] Park, Y. (2018). Predicting personalized student performance in computing-related majors via collaborative filtering. In *Proceedings of the 19th Annual SIG Conference on Information Technology Education*, pages 151–151.
- [Pavlidis 2019] Pavlidis, G. (2019). Recommender systems, cultural heritage applications, and the way forward. *Journal of Cultural Heritage*, 35:183–196.
- [Peška et al. 2019] Peška, L., Tashu, T. M., and Horváth, T. (2019). Swarm intelligence techniques in recommender systems - a review of recent research. *Swarm and Evolutionary Computation*, 48:201–219.
- [Prajapati and Brahmabhatt 2022] Prajapati, J. and Brahmabhatt, K. N. (2022). A systematic literature review on health recommender systems. In *Proceedings of the International e-Conference on Intelligent Systems and Signal Processing*, pages 557–572. Springer.
- [Ribeiro et al. 2012] Ribeiro, M. T., Lacerda, A., Veloso, A., and Ziviani, N. (2012). Pareto-efficient hybridization for multi-objective recommender systems. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 19–26.
- [Rodriguez et al. 2012] Rodriguez, M., Posse, C., and Zhang, E. (2012). Multiple objective optimization in recommender systems. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 11–18.
- [Rodriguez-Carrion et al. 2012] Rodriguez-Carrion, A., Garcia-Rubio, C., Campo, C., Cortés-Martín, A., Garcia-Lozano, E., and Noriega-Vivas, P. (2012). Study of lz-based location prediction and its application to transportation recommender systems. *Sensors*, 12(6):7496–7517.
- [Salakhutdinov et al. 2007] Salakhutdinov, R., Mnih, A., and Hinton, G. (2007). Restricted boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning*, pages 791–798.
- [Sankar et al. 2014] Sankar, A., Kiruthikaa, K., et al. (2014). Community based recommendation in e-learning systems. *Journal of e-Learning and Knowledge Society*, 10(1).
- [Santos and Boticario 2010] Santos, O. C. and Boticario, J. G. (2010). Modeling recommendations for the educational domain. *Procedia Computer Science*, 1(2):2793–2800.

- [Saraswathi et al. 2014] Saraswathi, S., Hemanth, M., Reddy, K. A., Kumar, S. U., and Suraj, M. (2014). Design of an online expert system for career guidance. *International Journal of Research in Engineering and Technology*, 03:314–319.
- [Sedhain et al. 2015] Sedhain, S., Menon, A. K., Sanner, S., and Xie, L. (2015). Autorec: Autoencoders meet collaborative filtering. In *Proceedings of the 24th international conference on World Wide Web*, pages 111–112.
- [Shani and Gunawardana 2011] Shani, G. and Gunawardana, A. (2011). Evaluating recommendation systems. In *Recommender systems handbook*, pages 257–297. Springer.
- [Sheehan and Park 2012] Sheehan, M. and Park, Y. (2012). pgpa: A personalized grade prediction tool to aid student success. *RecSys'12 - Proceedings of the 6th ACM Conference on Recommender Systems*.
- [Su and Khoshgoftaar 2009] Su, X. and Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009.
- [Tintarev and Masthoff 2015] Tintarev, N. and Masthoff, J. (2015). *Explaining Recommendations: Design and Evaluation*, pages 353–382. Springer US, Boston, MA.
- [Trabelsi et al. 2021] Trabelsi, F. Z., Khtira, A., and El Asri, B. (2021). Hybrid recommendation systems: A state of art. *ENASE*, pages 281–288.
- [Tsagkias et al. 2021] Tsagkias, M., King, T. H., Kallumadi, S., Murdock, V., and de Rijke, M. (2021). Challenges and research opportunities in ecommerce search and recommendations. In *ACM SIGIR Forum*, volume 54, pages 1–23. ACM New York, NY, USA.
- [Voß 2000] Voß, S. (2000). Meta-heuristics: The state of the art. In *Workshop on Local Search for Planning and Scheduling*, pages 1–23. Springer.
- [Wang et al. 2018] Wang, D., Liang, Y., Xu, D., Feng, X., and Guan, R. (2018). A content-based recommender system for computer science publications. *Knowledge-Based Systems*, 157:1–9.
- [Wang et al. 2010] Wang, Y., Wang, S., Stash, N., Aroyo, L., and Schreiber, G. (2010). Enhancing content-based recommendation with the task model of classification. volume 6317, pages 431–440.
- [Warlop et al. 2018] Warlop, R., Lazaric, A., and Mary, J. (2018). Fighting boredom in recommender systems with linear reinforcement learning. *Advances in Neural Information Processing Systems*, 31.
- [Wu et al. 2014] Wu, C.-Y., Diao, Q., Qiu, M., Jiang, J., and Wang, C. (2014). Jointly modeling aspects, ratings and sentiments for movie recommendation. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD*, volume 14, pages 193–202.

- [Zhang 2009] Zhang, M. (2009). Enhancing diversity in top-n recommendation. In *Proceedings of the third ACM conference on Recommender systems*, pages 397–400.
- [Zhang et al. 2020] Zhang, Q., Lu, J., and Jin, Y. (2020). Artificial intelligence in recommender systems. *Complex Intelligent Systems 2020 7:1*, 7(1):439–457.
- [Zhang and Chen 2018] Zhang, Y. and Chen, X. (2018). Explainable recommendation: A survey and new perspectives. *CoRR*, abs/1804.11192.
- [Zhang and Chen 2020] Zhang, Y. and Chen, X. (2020). Explainable recommendation: A survey and new perspectives. *Foundations and Trends® in Information Retrieval*, 14(1):1–101.
- [Zhao et al. 2018] Zhao, X., Xia, L., Zhang, L., Ding, Z., Yin, D., and Tang, J. (2018). Deep reinforcement learning for page-wise recommendations. In *Proceedings of the 12th ACM Conference on Recommender Systems*, pages 95–103.
- [Zheng et al. 2018] Zheng, G., Zhang, F., Zheng, Z., Xiang, Y., Yuan, N. J., Xie, X., and Li, Z. (2018). Drn: A deep reinforcement learning framework for news recommendation. In *Proceedings of the 2018 World Wide Web Conference*, pages 167–176.