

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

1° Júlia Cabral Diniz Braz

Departamento de Engenharia Elétrica
Universidade de Brasília
Brasília, Brazil
julia.braz@aluno.unb.br

2° Daniel Guerreiro e Silva

Departamento de Engenharia Elétrica
Universidade de Brasília
Brasília, Brazil
danielgs@ene.unb.br

Resumo—A utilização dos sistemas de recomendação em aplicações acadêmicas pode servir para auxiliar o aluno a selecionar disciplinas de acordo com suas habilidades, interesses e desempenho em disciplinas cursadas anteriormente. No presente trabalho propomos a implementação de um sistema de recomendação baseado em sistema imunológico artificial, tomando como base um banco de dados com informações acadêmicas de alunos de Engenharia de Redes de Comunicação. A utilização do sistema imunológico artificial confere uma diversidade na vizinhança que é tida como base para fazer a recomendação, devido a maior variedade entre os vizinhos que tenham alguma semelhança ao aluno alvo, é esperada uma abrangência maior nas opções de disciplinas que podem ser recomendadas. Experimentos foram conduzidos para avaliar o desempenho geral da abordagem proposta. Estudos iniciais mostraram que comparada a técnica básica de filtragem colaborativa, há um expressivo aumento do número de recomendações aliado a uma maior precisão, podemos inferir que a rede formada pelo AIS leva a obtenção de uma vizinhança que tem abrangência maior do conjunto de matérias relevantes.

I. INTRODUÇÃO

As meta-heurísticas bio-inspiradas 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 [1] e, sendo os métodos evolutivos são os mais utilizados dentre elas. Alguns representantes destas meta-heurísticas são: as estratégias evolutivas, a programação evolutiva e os algoritmos genéticos [2].

Outros algoritmos bio-inspirados têm recebido atenção nas últimas décadas como, por exemplo, as técnicas inspiradas no sistema imunológico. Ele possui algumas propriedades interessantes para apoiar a construção de sistemas computacionais inteligentes, são elas: robustez, tolerância, distribuição, memória, adaptabilidade, auto-organização e autonomia [3]. Dasgupta [4] 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”. Vale também mencionar que, dentro da grande diversidade de aplicações, encontramos os sistemas de recomendação como um pertinente campo para técnicas bio-inspiradas [2], [5], [6].

Em paralelo, a quantidade de opções de disciplinas oferecidas em cursos superiores de universidades aumentou [7].

Algumas instituições tentam manter um currículo aberto para que haja interdisciplinaridade, em outras palavras, o aluno pode cursar disciplinas que ele acha relevante a seus interesses. Disciplinas são ofertadas na condição de “optativas”, oferecendo uma especialização em uma certa área do curso.

Com base neste cenário, o presente trabalho propõe a 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. A utilização do sistema pode servir para auxiliar o aluno a selecionar disciplinas de acordo com suas habilidades, interesses e desempenho em disciplinas cursadas anteriormente. Dessa forma, o sistema utiliza a análise do histórico acadêmico do aluno, i.e. suas notas em matérias previamente cursadas, servirá para se recomendar matérias optativas em um grupo de outros alunos, com histórico parecido ao aluno alvo mas diversos entre si, tiveram rendimento satisfatório. Destaca-se também que o sistema utiliza uma estratégia baseada em modelos de concentração de anticorpos para produzir recomendações diversificadas ao aluno. Ademais, serão utilizadas métricas para avaliação da recomendação condizentes com as utilizadas na literatura, para fins de comparação.

O resto do trabalho se organiza conforme a seguir: a Seção II apresenta trabalhos relacionados a sistemas de recomendações e a utilização de tais sistemas no domínio acadêmico, a Seção III descreve as principais abordagens utilizadas na recomendação utilizadas na literatura, a Seção IV apresenta como o sistema imunológico artificial já foi utilizado para fins de recomendação e, finalmente, as Seções VI e VII descrevem os resultados experimentais e a conclusão, respectivamente.

II. TRABALHOS RELACIONADOS

Os sistemas de recomendação são algoritmos que auxiliam nas escolhas de atividades a serem exploradas por um indivíduo. Estes sistemas podem se basear em diversos fatores, como por exemplo preferências e histórico dos indivíduos [7]. São aplicados em diversos serviços, como em *streaming* de músicas e filmes, e.g. *Spotify* e a *Netflix*, que utilizam as classificações de itens favoritos pelos próprios usuários para lhe recomendarem novos itens.

Esse tipo de sistema é classificado como uma aplicação da inteligência artificial [8] 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 [9]. O interesse por essa área de pesquisa é motivado devido à sua vasta possibilidade de aplicações práticas e por auxiliar a analisar grandes massas de dados e, assim, poder fornecer recomendações personalizadas.

São várias as abordagens de recomendação já utilizadas para a seleção de disciplinas, sendo elas filtragem colaborativa, baseada em conhecimento, baseada em conteúdo e métodos híbridos [7]. Alguns trabalhos focam na predição de notas de disciplinas para recomendação, de modo a selecionar as disciplinas Top- n [10], [11]. Em [12], a necessidade de ter uma diversidade dos itens recomendados é abordada, são propostos alguns algoritmos de diversificação da lista de recomendação. Em [13] o algoritmo genético é utilizado para descobrir os critérios mais relevantes que podem afetar a recomendação de disciplinas eletivas para estudantes universitários, foram utilizados diferentes pesos para cada critério a fim de descobrir qual é a combinação de múltiplos critérios que fornece melhores resultados.

A acurácia de tais sistemas está atrelada à sua capacidade de indicar matérias relevantes. A avaliação das recomendações é baseada em medidas métricas amplamente utilizadas, como o erro médio absoluto (MAE, do inglês *mean absolut error*) ou a raiz do erro quadrado médio (RMSE, do inglês *root mean squared error*). Estas medidas expressam o erro médio da predição, em relação aos dados originais, além disso, medidas de recuperação de informação (*precision e recall*) também são usadas [8]. Trabalhos na área de avaliação desses sistemas levando em conta a dificuldade de previsão e a diversidade nas recomendações foram desenvolvidos em [12].

III. ABORDAGENS DE RECOMENDAÇÃO

Um 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 [14], há quatro classificações de técnicas, diferenciadas com base em sua fonte de conhecimento. Entre elas, destacamos as duas estratégias mais utilizadas:

- 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 dos itens de usuários semelhantes, ou seja, o sistema procura por usuários com as mesmas preferências e encontra itens através dessa relação de vizinhança.

A. Filtragem colaborativa

Esse método tenta prever quanto o item é relevante a um usuário individual com base nos itens avaliados por outros usuários. Deve-se identificar um subconjunto de usuários similares ao alvo, para assim avaliar a preferência de um indivíduo deste grupo a certo item. No problema de recomendação de disciplinas considerado, o sistema tenta buscar outros alunos com histórico de notas similar e, assim, recomendar disciplinas em que estes tenham obtido resultados satisfatórios.

Para a implementação deste método é gerada uma matriz, denominada “matriz usuário-item”, onde cada célula é a avaliação do item (linha) por um usuário (coluna). Este tipo de abordagem não precisa de informação nenhuma quanto às características dos itens, a recomendação é apenas baseada nas avaliações de usuários similares, ou itens semelhantes aos avaliados previamente pelo usuário. Diferentemente do método baseado em conteúdo, esses algoritmos têm a vantagem de conseguir recomendar itens com características diversas e não sofre com o problema de especificidade.

Outras vantagens desse tipo de abordagem é sua simples implementação e a recomendação pode ser mais facilmente explicada e mais intuitiva. No entanto, seu desempenho é bastante afetado quando a matriz usuário-item é muito esparsa, pois a falta de dados pode ser insuficiente para criar a vizinhança [8].

Já as abordagens baseadas em modelo utilizam as avaliações dos usuários para treinar um sistema preditor. Por meio deste busca-se inferir os valores de avaliações ausentes na matriz usuário-item, o que pode ser bastante eficiente na geração de recomendação, mas requer 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. Assim, a escolha da abordagem de recomendação depende da aplicação prática e não apenas do desempenho do sistema [15].

IV. USO DE SISTEMAS IMUNOLÓGICOS ARTIFICIAIS EM SISTEMAS DE RECOMENDAÇÃO

A inspiração no sistema imunológico 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 [3]. O sistema imune busca proteger o organismo de agentes danosos que se infiltram nele. No contexto biológico, quando isto ocorre, 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. Se a estratégia não for suficiente, outro tipo de resposta é ativada, responsável pela robustez do sistema ao combater especialmente substâncias estranhas. Esta é chamada de imunidade específica, adaptativa ou adquirida, ou seja, o sistema desenvolve seus próprios agentes protetores, os anticorpos, quando em contato com a substância estranha específica, o antígeno. Os anticorpos atuam reagindo quimicamente com o antígeno de forma a impedir malefícios ao organismo. Eles são antígeno-específicos, ou seja, para um

determinado antígeno, somente anticorpos que são capazes de reagir ao mesmo terão sua produção estimulada [3].

Em 1974, Jerne [16] apresenta 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 a já existentes na rede. 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.

No contexto de sistemas de recomendação, são três os recursos da rede imunológica que podem ser relevantes: (i) correspondência, que se refere à ligação entre anticorpos e antígenos; (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; e (iii) controle distribuído, pois o sistema imunológico é governado por interações locais entre células e não existe um controlador geral.

Há trabalhos que aplicam sistemas imunológicos artificiais (AIS, do inglês *Artificial Immune System*) no problema de recomendação, os quais geralmente adaptam a técnica para funcionar como uma filtragem colaborativa. Em [17], o AIS 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 com vizinhança para a predição na nota e a concentração do anticorpo também é utilizada como peso. Os resultados experimentais são validados na recomendação de filmes.

Também trabalhando com recomendação de filmes, Acilar e Arslan [18] utilizam-se da teoria da rede imunológica de modo híbrido com um algoritmo de filtragem colaborativa. O mecanismo de supressão do 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 [19] o AIS é utilizado na recomendação de disciplinas. O sistema é utilizado como uma técnica de *clustering* e a

afinidade na teoria da rede imunológica foi combinada com os conceitos do diagrama de expansão. A distância entre cada membro do conjunto de dados é medida pelo inverso de sua afinidade, uma afinidade mais alta entre dois membros implica que eles estarão localizados mais próximos. Cada aluno é tratado como um antígeno ou anticorpo. Anticorpos associados ao mesmo *cluster* criam as redes de anticorpos.

V. PROPOSTA DE SISTEMA

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 [12]. Assim, para explorar todo o conjunto de disciplinas relevantes, o conceito de rede imunológica regulada pelo mecanismo de concentração de anticorpos, semelhante ao apresentado em [17], será aplicado ao problema de recomendação de disciplinas, de modo a analisar o impacto na recomendação e a abrangência das disciplinas recomendadas.

O sistema aqui proposto baseia-se em [17] 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 AIS, 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 [17], a variação da concentração de cada anticorpo da rede a cada interação do AIS é em função da similaridade com o antígeno (aluno alvo da recomendação) e entre os demais anticorpos da rede. Ela é regida por

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

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.

1	Período	Código	Nome	Turma	Horário	Id_aluno	Nota	Situação
2	2	ENE009	ARQUITETURA E PROTOCOLOS DE REDES	A	2M34 4M34	aluno220	3,00	APROVADO
3	2	ECL042	MÉDIO AMBIENTE FÍSICO E ECOSISTEMAS	B	3M34 5M12	aluno225	3,00	APROVADO
4	2	ENE013	ALGORITMOS E ESTRUTURA DE DADOS	A	3M34 5M34	aluno224	4,00	APROVADO
5	2	ENE0177	ELETROMAGNETISMO 1	B	2745 4745	aluno224	4,00	APROVADO
6	2	ENE022	PROJETO TRANSVERSAL EM REDES DE COMUNICAÇÃO	A	3M12 5M12	aluno224	5,00	APROVADO
7	2	ENE043	LABORATÓRIO DE CIRCUITOS ELÉTRICOS	I	485 471	aluno184	3,00	APROVADO
8	2	IFD0177	FÍSICA 2 EXPERIMENTAL	D	672345	aluno184	4,00	APROVADO
9	2	MAT059	MÉTODOS MATEMÁTICOS DA FÍSICA 1	A	2M34 4M34 6M1	aluno184	4,00	APROVADO
10	2	FGA0160	PROJETO INTEGRADOR DE ENGENHARIA 1	E	2745 4745	aluno052	3,00	APROVADO
11	2	FGA0160	DESENHO INDUSTRIAL ASSISTIDO POR COMPUTADOR	D	1745 5745 6745	aluno052	3,00	APROVADO
12	2	FGA0184	GESTÃO DA PRODUÇÃO E QUALIDADE	A	3723 5723	aluno052	5,00	APROVADO
13	2	MAT075	CÁLCULO DE PROBABILIDADE 1	A	2M34 3M34 5M1	aluno326	3,00	APROVADO

Figura 1: Parte do conjunto de dados utilizado

Por esta equação 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.

A. Conjunto de dados

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 1.

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.

B. Algoritmo proposto

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 AIS é calculada a nova concentração dos anticorpos conforme a equação (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, ou seja, se nenhum anticorpo é eliminado da rede em 10 iterações.

Após a obtenção de uma rede estabilizada, como observado em [17], é necessário um procedimento de diferenciação das concentrações dos anticorpos. Esse passo é necessário pois anticorpos adicionados à rede há mais tempo atingem a saturação (concentração máxima $c_i = 100$) mais rápido e os mais recentes não tem o tempo necessário para maturação, estando assim em desvantagem. Uma nova iteração do AIS é feita, desta vez até que um dos anticorpos atinja a concentração

máxima, e ainda é possível que algum anticorpo seja eliminado da rede.

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. Neste sentido, a seguir são apresentadas decisões importantes para o funcionamento do sistema de recomendação.

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 completo & Não estável 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
31 fim
32 Selecionar todas as matérias com nota 4 ou maior a uma lista;
33 Selecionar destas quais foram cursadas por 2 ou mais anticorpos;
34 Retirar desta lista matérias obrigatórias e as optativas já cursadas
    pelo Aluno;

```

Algoritmo 1: Sistema de recomendação baseado no sistema imunológico artificial adaptado de [17].

1) *Representação do antígeno e anticorpos:* O algoritmo utilizado é uma técnica de filtragem colaborativa baseada em memória, 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: aluno - matéria - nota, outras informações disponíveis não afetam a

Tabela I: 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

recomendação. A Tabela I 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. O aluno0001 é representado como: [3,3, , , 4,], os espaços em branco representam matérias não cursadas pelo aluno.

2) *Cálculo da similaridade*: 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 relacionados [5], [17]–[19]:

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

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. 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) *Predição da menção*: Para prever as notas dos alunos em disciplinas que não fizeram, usando seus dados anteriores e a rede final de anticorpos, usamos

$$p_i = \bar{r} + \frac{\sum_{Ab=1}^k w_{Ag,Ab}(r_{Ab,i} - \bar{r}_{Ab})}{\sum_{Ab=1}^k w_{Ag,Ab}}, \quad (3)$$

onde p_i é a nota prevista do aluno alvo (Ag) na matéria i , \bar{r} é a média de notas do Ag, 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_{Ab} é o peso do anticorpo Ab , que é definido como

$$w_{Ag} = r_{Ag,Ab} \times c_{Ab}, \quad (4)$$

onde $r_{Ag,Ab}$ é a correlação entre o anticorpo e o antígeno e c_{Ab} é a concentração do anticorpo na rede final.

4) *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. 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 acontece 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.

C. 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 AIS como sistema de recomendação [17].

- Número de alunos visitados: contabiliza quantos alunos foram necessários avaliar antes da estabilização da rede de anticorpos.
- Número de anticorpos na rede final: é o número final de vizinhos na rede estabilizada, obtido após rodar novamente o AIS para a diferenciação das concentrações, passo que ainda pode eliminar anticorpos da rede.
- 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 foram cursadas pelo aluno alvo

$$\text{MAE} = \frac{1}{m} \sum_{i=0}^m |p_i - r_i| \quad (5)$$

onde m é o número de matérias que tiveram a nota predita.

- Número de recomendações: Número total de matérias aptas para recomendação, ou seja, se foi cursada por mais de um aluno presente na rede e a nota predita foi maior que 4.
- Taxa de sobreposição na recomendação: essa medida busca avaliar a precisão 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 cursadas. Em termos matemáticos:

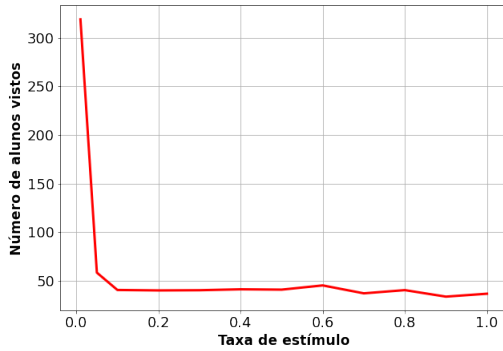
$$\frac{\#(\text{optativas cursadas} \cap \text{lista de recomendação})}{\# \text{optativas cursadas}} \quad (6)$$

VI. RESULTADOS EXPERIMENTAIS

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 de um conjunto de testes, para cada um destes o algoritmo é executado 5 vezes e são feitas as médias estatísticas dos resultados. O conjunto de teste contém um total de 64 alunos, o que representa 20% dos dados totais, enquanto a matriz usuário-item é composta pelos outros 80% do conjunto de dados e serve para o treinamento do sistema de recomendação.

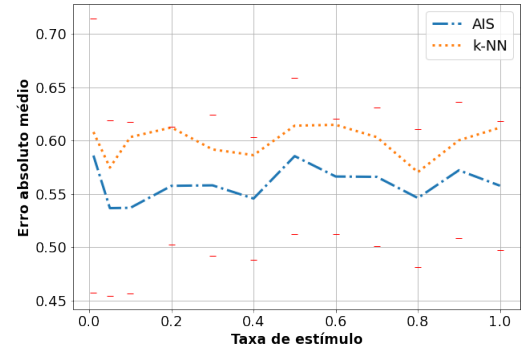
Para que a recomendação seja realizada é feita a exigência de que tenham cursado pelo menos 6 matérias, sendo 3 obrigatórias para ser possível achar correlações com outros alunos e 3 optativas, a fim de haver possibilidade de avaliar a precisão da recomendação, calculada pela taxa de sobreposição. Para cada um desses alunos a recomendação foi feita primeiramente com o AIS proposto, com o número máximo de anticorpos na rede igual a $n = 20$, e através de uma heurística simplificada

Efeito da taxa de estímulo no número de alunos visitados



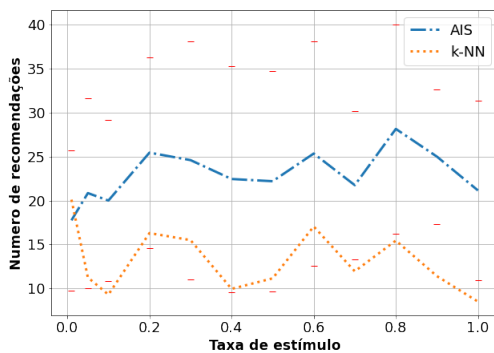
(a) Efeito da taxa de estímulo no número de alunos visitados.

Efeito da taxa de estímulo no erro absoluto médio



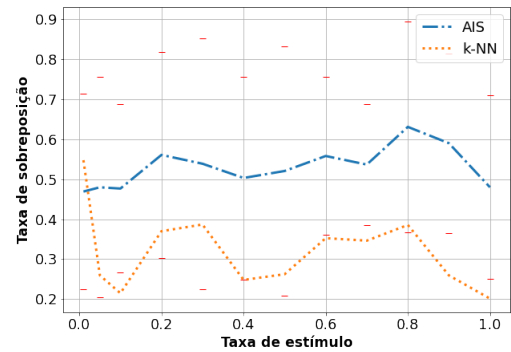
(b) Efeito da taxa de estímulo no erro absoluto médio.

Efeito da taxa de estímulo no número de recomendações



(c) Efeito da taxa de estímulo no número de recomendações.

Efeito da taxa de estímulo na taxa de sobreposição



(d) Efeito da taxa de estímulo na sobreposição da recomendação.

Figura 2: Análise do efeito da taxa de estímulo.

que seleciona os k alunos com maior similaridade. Para uma comparação justa, o número k de alunos mais similares é igual ao número de anticorpos presentes na rede final do AIS. As medidas observadas nos gráficos apresentados a seguir são estatísticas médias destas 20 recomendações.

A. Taxa de estímulo

Primeiramente é feita a análise da taxa de estímulo. Quando não há presença da taxa de supressão, ela é a responsável por incentivar padrões de similaridade do anticorpo com o antígeno. Pelo mecanismo de concentração de anticorpos do nosso sistema, a taxa de estímulo é a responsável por fornecer um limiar fixo de correlação entre o antígeno e o anticorpo. Mesmo 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 2a pode ser verificado que, em baixas taxas de estimulação, é 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 preencher o AIS e a rede 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 de correlação é importante pois opera na manutenção da qualidade de uma vizinhança, eliminando alunos pouco correlacionados.

Diferentemente do que foi apresentado em [17], 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, ainda foi possível completar a rede com o número máximo de anticorpos definido.

Este sistema gera um maior número de recomendações para qualquer taxa de estímulo. Este fenômeno ocorre mesmo sem

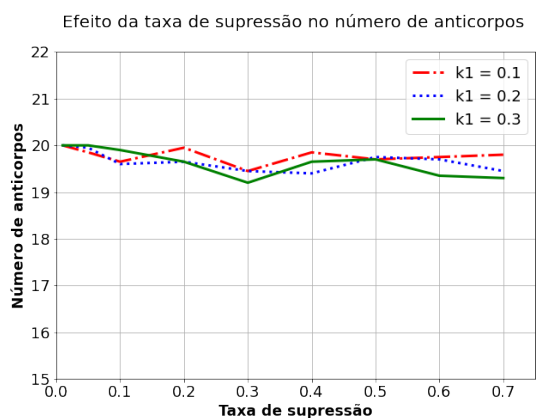


Figura 3: Efeito da taxa de supressão no número de anticorpos na rede.

a presença da taxa de supressão, pois alguma diversidade nos anticorpos já está garantida, uma vez que estes são adicionados aleatoriamente na rede, comportamento verificado no gráfico da Figura 2c. Se a taxa de estímulo for bem ajustada a rede é capaz de gerar boas predições.

No gráfico da Figura 2b verifica-se que, para qualquer taxa de estímulo, o MAE do AIS 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 2d mostra que, para taxas acima de 0,05, a taxa de sobreposição é maior. É verificado que as recomendações do AIS são significativamente melhores tanto em relação ao erro na predição da nota quanto na precisão da recomendação.

Quando a taxa de estímulo é muito baixa, 0,01, todos anticorpos para fazer 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 2c e 2d, tanto o número de recomendações quanto a taxa de sobreposição tem valores bem próximos. Apesar disso, o erro de predição pelo AIS 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.

B. Taxa de supressão

Quando adicionado o efeito da taxa de supressão, mais uma exigência é feita para que o anticorpo permaneça na rede. Ele deverá ser suficientemente diferente dos outros anticorpos na rede. Desse modo, a taxa de supressão é a responsável por adicionar diversidade na rede.

A taxa de supressão é então vista como um artifício que adiciona um refinamento na rede numa vizinhança já consi-

derada “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: 0,01, 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.

Primeiramente foi analisado se a taxa de supressão afeta o tamanho da rede final. No gráfico da Figura 3, 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 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. O efeito da presença desta taxa pode ser observado nos gráficos da Figura 4. Para a combinação de taxas ideal é esperado um MAE baixo e altas taxas de sobreposição na recomendação.

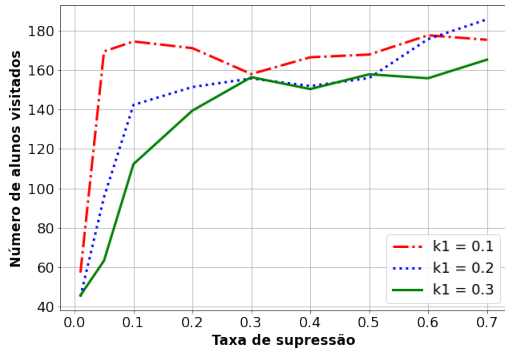
Como era esperado, pelo gráfico da Figura 4a é notado que o número de alunos visitados aumenta com o aumento da taxa de supressão, mas, após superar a taxa de estímulo, este número se estabiliza, o que pode significar que após esse patamar passa a ser mais importante que a rede seja diversa do que os anticorpos tenham alguma semelhança ao antígeno. Esse comportamento também pode ser atestado pelo gráfico da Figura 4c. 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 da recomendação é bastante afetado. Para valores menores que a taxa de supressão não é visto muita alteração no erro da predição, todos obtendo 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 4b. 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.

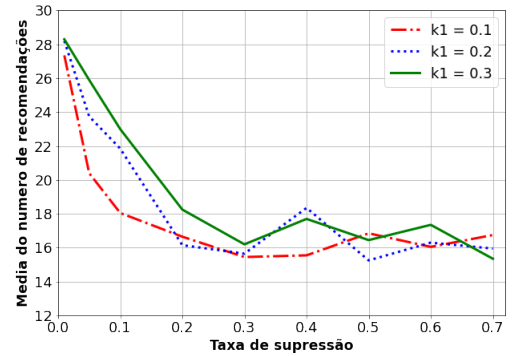
Segundo o gráfico da Figura 4d quando $k_1 = 0,3$ e $k_2 = 0,05$ o pico da sobreposição na recomendação é atingido, sendo esta a melhor combinação de taxas. Apesar do número de recomendações ser semelhante ao obtido sem a presença da supressão e o MAE não ser afetado com a presença da taxa, é visto que há uma melhora da taxa de sobreposição da recomendação. Deste modo, esta combinação é definida como ideal e deverá ser utilizada nos experimentos futuros.

Efeito da taxa de supressão no número de alunos visitados



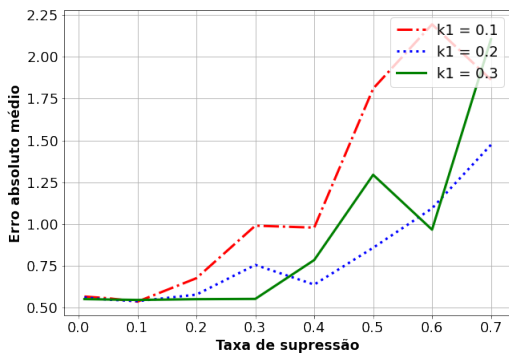
(a) Efeito da taxa de supressão no número de alunos visitados.

Efeito da taxa de supressão no número de recomendações



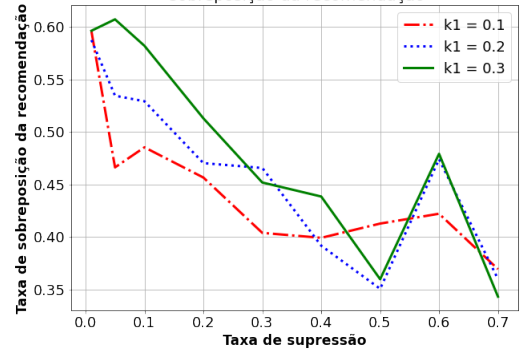
(b) Efeito da taxa de supressão no número de recomendações.

Efeito da taxa de supressão no erro absoluto médio



(c) Efeito da taxa de supressão no erro da predição.

Efeito da taxa de supressão na taxa de sobreposição da recomendação



(d) Efeito da taxa de supressão na taxa de sobreposição da recomendação.

Figura 4: Análise do efeito da taxa de supressão.

VII. CONCLUSÃO

Essa pesquisa utilizou dados de alunos do curso de Engenharia de Redes de Comunicação da Universidade de Brasília para avaliar 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 não foi estudado. 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 como 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 do de itens relevantes.

Neste trabalho são apresentados os resultados iniciais da utilização do AIS na recomendação de disciplinas. 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. Aliado também a 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, e a lista de matérias recomendadas pelo sistema proposto. Com isso, podemos inferir que a rede formada pelo AIS 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 utilizando o AIS, o que nos leva a concluir que a utilização da concentração como peso tem ligeira vantagem sobre a técnica básica também na predição.

Trabalhos futuros na análise da composição da vizinhança formada pelo AIS deverão 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, este efeito deve ser estudado. Além disso, é necessária a avaliação das limitações do algoritmo, por exemplo, em experimentos iniciais observou-se que a quantidade de matérias cursadas anteriormente pelo aluno alvo pode afetar a qualidade da recomendação.

REFERÊNCIAS

- [1] S. Voß, “Meta-heuristics: The state of the art,” in *Workshop on Local Search for Planning and Scheduling*. Springer, 2000, pp. 1–23.
- [2] D. Corne and M. A. Lones, *Evolutionary Algorithms*. Cham: Springer International Publishing, 2018, pp. 1–22.
- [3] G. P. Figueredo, H. S. Bernardino, and H. J. C. Barbosa, “Introdução aos sistemas imunológicos artificiais,” in *Meta-Heurísticas em Pesquisa Operacional*, 1st ed., H. S. Lopes, L. C. de Abreu Rodrigues, and M. T. A. Steiner, Eds. Curitiba, PR: Omnipax, 2013, ch. 8, pp. 113–128.
- [4] D. Dasgupta, S. Yu, and F. Nino, “Recent advances in artificial immune systems: Models and applications,” *Applied Soft Computing*, vol. 11, no. 2, pp. 1574 – 1587, 2011, the Impact of Soft Computing for the Progress of Artificial Intelligence.
- [5] A. Esteban, A. Zafra, and C. Romero, “Helping university students to choose elective courses by using a hybrid multi-criteria recommendation system with genetic optimization,” *Knowledge-Based Systems*, vol. 194, p. 105385, 2020. [Online]. Available: <https://doi.org/10.1016/j.knosys.2019.105385>
- [6] N. E. I. Karabadji, S. Beldjoudi, H. Seridi, S. Aridhi, and W. Dhifli, “Improving memory-based user collaborative filtering with evolutionary multi-objective optimization,” *Expert Systems with Applications*, vol. 98, pp. 153–165, 2018. [Online]. Available: <https://doi.org/10.1016/j.eswa.2018.01.015>
- [7] N. Lynn and A. Emanuel, “A review on recommender systems for course selection in higher education,” 06 2020.
- [8] G. Pavlidis, “Recommender systems, cultural heritage applications, and the way forward,” *Journal of Cultural Heritage*, vol. 35, pp. 183–196, 2019.
- [9] G. Adomavicius and A. Tuzhilin, “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*, vol. 17, no. 6, pp. 734–749, 2005.
- [10] M. Sheehan and Y. Park, “pgpa: A personalized grade prediction tool to aid student success,” *RecSys’12 - Proceedings of the 6th ACM Conference on Recommender Systems*, 09 2012.
- [11] A. Elbadrawy and G. Karypis, “Domain-aware grade prediction and top-n course recommendation,” in *Proceedings of the 10th ACM Conference on Recommender Systems*, 2016, pp. 183–190.
- [12] M. Zhang, “Enhancing diversity in top-n recommendation,” in *Proceedings of the third ACM conference on Recommender systems*, 2009, pp. 397–400.
- [13] A. Esteban, A. Zafra, and C. Romero, “A hybrid multi-criteria approach using a genetic algorithm for recommending courses to university students,” in *International Educational Data Mining Society*. International Educational Data Mining Society, jul 2018.
- [14] R. Burke and Robin, “Hybrid web recommender systems,” vol. 4321, 01 2007.
- [15] Y. Koren, “Factor in the neighbors: Scalable and accurate collaborative filtering,” *ACM Trans. Knowl. Discov. Data*, vol. 4, no. 1, Jan. 2010. [Online]. Available: <https://doi.org/10.1145/1644873.1644874>
- [16] N. Jerne, “Towards a network theory of the immune system,” *Annales d’immunologie*, vol. 125C, no. 1-2, p. 373–389, January 1974. [Online]. Available: <http://europepmc.org/abstract/MED/4142565>
- [17] S. Cayzer and U. Aickelin, “A recommender system based on idiotypic artificial immune networks,” *Journal of Mathematical Modelling and Algorithms*, vol. 4, no. 2, pp. 181–198, 2005.
- [18] A. Merve Acilar and A. Arslan, “A collaborative filtering method based on artificial immune network,” *Expert Systems with Applications*, vol. 36, no. 4, pp. 8324–8332, 2009. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417408007100>
- [19] P.-C. Chang, C.-H. Lin, and M.-H. Chen, “A hybrid course recommendation system by integrating collaborative filtering and artificial immune systems,” *Algorithms*, vol. 9, no. 3, p. 47, 2016.