



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS CRATEÚS
CURSO DE GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

JONAS DE CASTRO GOMES

**PREDIÇÃO DE DESEMPENHO DE ALUNOS NO PRÓXIMO PERÍODO BASEADO NO
HISTÓRICO DE NOTAS**

CRATEÚS

2022

JONAS DE CASTRO GOMES

PREDIÇÃO DE DESEMPENHO DE ALUNOS NO PRÓXIMO PERÍODO BASEADO NO
HISTÓRICO DE NOTAS

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Sistemas de Informação
do Campus Crateús da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Sistemas de Informação.

Orientador: Prof. Me. Lívio Antônio
Melo Freire

Coorientador: Prof. Me. Marciel Barros
Pereira

CRATEÚS

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

G614p Gomes, Jonas de Castro.
Predição de desempenho de alunos no próximo período baseado no histórico de notas / Jonas de Castro
Gomes. – 2022.
41 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Crateús,
Curso de Sistemas de Informação, Crateús, 2022.
Orientação: Prof. Me. Lívio Antônio Melo Freire.
Coorientação: Prof. Me. Marciel Barros Pereira.

1. Filtragem Colaborativa. 2. Filtragem Colaborativa Neural. 3. Redes Neurais. 4. Fatoração de Matriz. 5.
Sistema de Recomendação. I. Título.

CDD 005

JONAS DE CASTRO GOMES

PREDIÇÃO DE DESEMPENHO DE ALUNOS NO PRÓXIMO PERÍODO BASEADO NO
HISTÓRICO DE NOTAS

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Sistemas de Informação
do Campus Crateús da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Sistemas de Informação.

Aprovada em:

BANCA EXAMINADORA

Prof. Me. Lívio Antônio Melo Freire (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Me. Marciel Barros Pereira (Coorientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. José Wellington Franco da Silva
Universidade Federal do Ceará - (UFC)

Prof. Me. Amanda Drielly Pires Venceslau
Universidade Federal do Ceará - (UFC)

À Deus, minha família e amigos.

AGRADECIMENTOS

Ao Prof. Me. Lívio Antônio Melo Freire por me orientar em meu trabalho de conclusão de curso e contribuir com seu conhecimento e experiência.

Ao Prof. Me. Marciel Barros Pereira, que fez parte da construção desse trabalho e sem dúvida foi de grande importância.

À Prof. Me. Lisieux Marie Marinho dos Santos Andrade por me orientar na disciplina de Projeto de Pesquisa Científica e Tecnológica.

Ao Doutorando em Engenharia Elétrica, Ednardo Moreira Rodrigues, e seu assistente, Alan Batista de Oliveira, aluno de graduação em Engenharia Elétrica, pela adequação do *template* utilizado neste trabalho para que o mesmo ficasse de acordo com as normas da biblioteca da Universidade Federal do Ceará (UFC).

A minha esposa e meus pais pelo incentivo e apoio, sem eles essa caminhada seria mais árdua.

Aos amigos que foram de grande ajuda em toda a graduação apoiando e incentivando, contribuindo direto ou indiretamente.

Agradeço a todos os professores por me proporcionar o conhecimento e por tanto que se dedicaram e apoiaram.

Por fim, principalmente agradeço a Deus pela graça e por me proporcionar forças para realizar esse sonho.

“Assim, fixamos os olhos, não naquilo que se vê, mas no que não se vê, pois o que se vê é transitório, mas o que não se vê é eterno.”

(2 CORÍNTIOS 4:18)

RESUMO

Na contramão do crescimento gradual do número de matrículas na graduação, está a problemática da evasão e retenção de alunos. De acordo com os dados do Censo da Educação Superior 2018 do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), apenas 37,9% dos estudantes ingressantes em 2010 concluíram os estudos e 5,3% seguiam na graduação após 6 anos (INEP, 2018). Nesse sentido, a capacidade de prever o desempenho de alunos nas disciplinas é fundamental para auxiliar nas decisões educacionais das Instituições de Ensino Superior (IES). O presente estudo tem como principal objetivo prever o desempenho de alunos utilizando Filtragem Colaborativa Neural (FCN) além de comparar esse modelo com modelos clássicos de Filtragem Colaborativa (FC). Os alunos serão classificados como aprovados ou reprovados. Os modelos geram recomendações para os alunos com base no histórico de notas. Os modelos foram submetidos a dados não sensíveis da Universidade Federal do Ceará (UFC) campus Crateús. Como métrica para analisar o desempenho dos modelos e compará-los foi utilizado a média das métricas: *accuracy*, *precision*, *recall* e *F1_score*. Os modelos testados tiveram bons resultados e dentre eles o modelo clássico Fatoração Matricial Generalizada (FMG) teve o melhor desempenho indicando *Accuracy* de 0,891, *Precision* de 0,931, *Recall* de 0,871 e *F1_score* de 0,867 baseado nos dados utilizados para o treinamento, teste e validação dos modelos. Por fim, o modelo Rede Perceptron Multicamada (RPM) teve os seguintes resultados: *Accuracy* de 0,841, *Precision* de 0,841, *Recall* de 0,913 e *F1_score* de 0,840. O modelo Fatoração de Matrizes Neurais (FMNeu), por sua vez teve os seguintes resultados: *Accuracy* de 0,863, *Precision* de 0,896, *Recall* de 0,870 e *F1_score* de 0,848.

Palavras-chave: Filtragem Colaborativa. Filtragem Colaborativa Neural. Redes Neurais. Fatoração de Matriz. Sistema de Recomendação

ABSTRACT

Contrary to the gradual growth in the number of undergraduate enrollments, there is the problem of student evasion and retention. According to data from the 2018 Higher Education Census by INEP, only 37.9% of students entering in 2010 completed their studies and 5.3% were continuing their studies after 6 years (INEP, 2018). In this sense, the ability to predict student performance in subjects is fundamental to help in the educational decisions of Higher Education Institutions (HEI). The main objective of the present study is to predict student performance using Neural Collaborative Filtering (NCF) in addition to comparing this model with classical models of Collaborative Filtering (CF). Students will be classified as passing or failing. Templates generate recommendations for students based on grade history. The models were submitted to non-sensitive data from the Federal University of Ceará (FUC) Crateús campus. As a metric to analyze the performance of the models and compare them, the average of the metrics was used: *accuracy*, *precision*, *recall* and *F1_score*. The tested models had good results and among them the classic model Generalized Matrix Factorization (GMF) had the best performance indicating *Accuracy* of 0.891, *Precision* of 0.931, *Recall* of 0.871 and *F1_score* of 0.867 based on the data used for training, testing and validating the models. Finally, the Multi-Layer Perceptron (MLP) model had the following results: *Accuracy* of 0.841, *Precision* of 0.841, *Recall* of 0.913 and *F1_score* of 0.840. The Neural Matrix Factorization (NMF) model, in turn, had the following results: *Accuracy* of 0.863, *Precision* of 0.896, *Recall* of 0.870 and *F1_score* of 0.848.

Keywords: Collaborative Filtering. Neural Collaborative Filtering. Neural Networks. Deep Learning. Matrix Factorization. Recommendation System.

LISTA DE FIGURAS

Figura 1 – Matriz de usuário-item	24
Figura 2 – Espaço latente do usuário	24
Figura 3 – Estrutura da filtragem colaborativa neural	25
Figura 4 – Modelo de fatoração de matriz neural	27
Figura 5 – Modelo de filtragem colaborativa neural customizado	35

LISTA DE TABELAS

Tabela 1 – Comparativo de trabalhos relacionados	32
Tabela 2 – Número de alunos e disciplinas por semestre	33
Tabela 3 – Divisão dos dados de treinamento e teste	34
Tabela 4 – Hiperparâmetros utilizados nos modelos RPM e FMNeu	36
Tabela 5 – Média das métricas para cada hidden states	37
Tabela 6 – Média das métricas para cada hidden states / hidden layer	38
Tabela 7 – Média das métricas do modelo FCN	38
Tabela 8 – Melhores métricas de cada modelo	38

LISTA DE ABREVIATURAS E SIGLAS

INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
IES	Instituições de Ensino Superior
FCN	Filtragem Colaborativa Neural
FC	Filtragem Colaborativa
UFC	Universidade Federal do Ceará
FMG	Fatoração Matricial Generalizada
RPM	Rede Perceptron Multicamada
FMNeu	Fatoração de Matrizes Neurais
HEI	Higher Education Institutions
NCF	Neural Collaborative Filtering
CF	Collaborative Filtering
FUC	Federal University of Ceará
GMF	Generalized Matrix Factorization
MLP	Multi-Layer Perceptron
NMF	Neural Matrix Factorization
SRE	Sistemas de Recomendações Educacionais
SR	Sistemas de Recomendações
FCBM	Filtragem Colaborativa Baseada em Modelo
FM	Fatoração de Matrizes
GAD	Grafo Acíclico Direcionado
SVD	Singular Value Decomposition
RNA	Redes Neurais Artificial
RNP	Redes Neurais Profundas
ReLU	Unidade linear retificada

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Contextualização	15
1.2	Justificativa	15
1.3	Objetivo Geral	16
1.4	Objetivos Específicos	16
1.5	Organização do Trabalho	16
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Filtragem Colaborativa	17
2.1.1	<i>Baseadas em Memória</i>	18
2.1.1.1	<i>Predição baseada em usuário</i>	18
2.1.1.2	<i>Predição baseada em item</i>	19
2.1.1.3	<i>Similaridade</i>	19
2.1.1.3.1	<i>Correlação de Pearson</i>	20
2.1.1.3.2	<i>Cosseno</i>	20
2.1.1.3.3	<i>Jaccard</i>	21
2.1.2	<i>Baseadas em Modelo</i>	21
2.2	Filtragem Colaborativa Neural	22
2.2.1	<i>Rede Neural Artificial (RNA)</i>	23
2.2.2	<i>Limitação da fatoração de matrizes</i>	23
2.2.3	<i>Estrutura da filtragem colaborativa neural</i>	25
2.2.4	<i>Fatoração Matricial Generalizada</i>	26
2.2.5	<i>Fatoração de Matrizes Neurais</i>	26
2.3	Resumo do Capítulo	28
3	TRABALHOS RELACIONADOS	29
3.1	Resumo do Capítulo	31
4	PROPOSTA	33
4.1	Aquisição e Tratamento dos Dados	33
4.2	Metodologia de Teste e Predição do Desempenho	34

4.3	Definição dos modelos a serem utilizados	35
4.4	Arquitetura	35
5	RESULTADOS	37
5.1	Resultados da Fatoração Matricial Generalizada	37
5.2	Resultados da Rede Perceptron Multicamada	37
5.3	Resultados da Filtragem Colaborativa Neural	38
5.4	Melhores resultados dos modelos	38
6	CONCLUSÕES E TRABALHOS FUTUROS	39
	REFERÊNCIAS	40

1 INTRODUÇÃO

1.1 Contextualização

Dentre as preocupações das IES está manter o ensino de qualidade, minimizando possíveis insucessos e retenções durante a formação dos alunos, melhorando, então, a taxa de conclusão do ensino superior. De acordo com os dados do Censo da educação superior 2018 do INEP, apenas 37,9% dos estudantes ingressantes em 2010 concluíram os estudos e 5,3% seguiam na graduação após 6 anos (INEP, 2018).

Um conjunto de técnicas que vem tomando grande proeminência no meio científico, com relação a descoberta de conhecimento e recomendação, são os Sistemas de Recomendações Educacionais (SRE). De acordo com (BURKE, 2002), Sistemas de Recomendações (SR) podem ser definidos como qualquer sistema que produz recomendações individualizadas como saída ou tem o efeito de orientar o usuário de maneira personalizada para objetos interessantes ou úteis em um amplo espaço de opções possíveis.

O presente estudo, questiona se é possível prever o desempenho educacional dos alunos a partir do histórico de notas utilizando-se de SR. Dessa forma, influenciar diretamente nas decisões educacionais das IES de modo a contribuir com ações focadas nos alunos com maior dificuldade.

1.2 Justificativa

De acordo com os dados do Censo da Educação Superior 2018 (INEP, 2018) o número de matrículas na educação superior continua crescendo, atingindo a marca de 8,45 milhões de alunos. Na contramão do crescimento gradual do número de matrículas na graduação, está a problemática da evasão e retenção. Essa problemática tem resultado negativo sobre o aluno e a instituição educacional. Apesar deste ser um problema antigo e já existirem pesquisas na área, a necessidade de trabalhar no desenvolvimento de pesquisas relacionadas, ainda é persistente.

Nesse sentido, a utilização de técnicas de SR tem como objetivo auxiliar a tomada de decisão, fornecendo sugestões de itens a serem recomendados. Através do conhecimento descoberto com a aplicação das técnicas de SR, é possível identificar e traçar perfis semelhantes de alunos,

que, por sua vez, permitem prever o desempenho dos mesmos com base nessa semelhança. Essa abordagem é importante porque pode auxiliar as IES no desenvolvimento de ações que visam aprimorar a proposta educacional criando condições de melhoria no desempenho dos alunos com maior dificuldade (BURKE, 2002).

Com isso, este trabalho tem como principal contribuição, apresentar um diagnóstico sobre em quais áreas os alunos terão um menor desempenho e necessitarão de maior atenção, possibilitando ao corpo docente da universidade, traçar medidas de prevenção contra o número de incidência de casos de evasão e retenção dos alunos. Além de permitir aos alunos conhecer seu possível desempenho no próximo semestre de forma a auxiliar suas escolhas.

1.3 Objetivo Geral

Utilizar métodos de FCN para prever o desempenho dos alunos com base no histórico de notas.

1.4 Objetivos Específicos

- Analisar as abordagens de resolução já propostas para o problema de previsão de desempenho de alunos;
- Estabelecer conjunto de testes para predição dos desempenhos;
- Aplicar métodos de FCN para predição desempenho dos alunos para o semestre seguinte;
- Avaliar o desempenho da FCN comparando com técnicas clássicas.

1.5 Organização do Trabalho

O presente trabalho está organizado em 6 capítulos. O Capítulo 2 refere-se a fundamentação teórica, apresentando conceitos importantes dos principais temas abordados na presente pesquisa. Entre eles, FC e FCN. O Capítulo 3 discorre sobre os trabalhos relacionados aos assuntos abordados nessa pesquisa. O Capítulo 4 apresenta a proposta do trabalho, onde são descritas as soluções propostas para alcançar os objetivos. O Capítulo 5 apresenta os resultados obtidos. Por fim, o Capítulo 6 discute os resultados, limitações e trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como objetivo apresentar conceitos importantes dos principais temas abordados na presente pesquisa, de modo a familiarizar o leitor com os mesmos. Com isso, o capítulo se organiza da seguinte forma: A Seção 2.1 apresenta os principais conceitos relacionados a FC. Por fim, a Seção 2.2 apresenta os principais conceitos relacionados a FCN.

2.1 Filtragem Colaborativa

Para a predição e sugestões, grande parte dos SR tem como principais abordagens métodos baseados em conteúdo e FC. Os métodos baseados em conteúdo tem como base o *feedback* do usuário alvo, na qual se deseja calcular uma predição, por outro lado a FC considera o *feedback* de um conjunto de usuários. Os *feedbacks* podem ser considerados explícitos, onde existe a participação ativa do usuário ou implícitos, é analisado o comportamento do usuário.

De acordo com (SCHAFER *et al.*, 2007), FC é o processo de filtrar ou avaliar itens por meio das opiniões de outras pessoas. Intencionando prever o desempenho de alunos a partir do histórico de notas, pode-se compreender o processo de filtragem colaborativa da seguinte forma: dado um conjunto de alunos e um conjunto de disciplinas que compreende uma matriz de grau esparsa cuja a dimensão é o tamanho do conjunto de alunos por o conjunto de disciplinas, o objetivo é prever um subconjunto de células vazias na matriz.

Um dos principais desafios enfrentados pela FC é o *Cold Start*, causado pela escassez de dados. De acordo com (SCHAFER *et al.*, 2007), o problema de *Cold Start* descreve situações em que um recomendador é incapaz de fazer recomendações significativas devido à falta inicial de classificações, degradando significativamente o desempenho da FC. Sabendo que a FC recomenda itens com base nas classificações dos usuários, para que se tenha recomendações precisas faz-se necessário que o número de notas ou classificações de itens seja suficiente para o sistema traçar um perfil para o usuário.

As principais técnicas de FC podem ser divididas em duas categorias: baseado em memória, detalhada na Seção 2.1.1 e Filtragem Colaborativa Baseada em Modelo (FCBM), detalhada na Seção 2.1.2. Os algoritmos baseados em memória se utilizam de uma medida de similaridade entre os usuários ou itens para que se possa efetuar a recomendação, já os algoritmos de FCBM,

utiliza-se de técnica de *machine learning* para construir um modelo capaz de fazer previsões.

Para uma melhor compreensão do uso das técnicas de FC, considere uma matriz de avaliações $M : a \times d$, em que a representa os alunos e d as disciplinas. Cada célula m_{ij} da matriz M representa uma nota do i -ésimo aluno em relação a j -ésima disciplina. Observe que um aluno a_i é representado por um vetor numérico de dimensão n da i -ésima linha da matriz M e que uma disciplina d_j é representada por um vetor da j -ésima coluna da matriz M . Com isso, tendo como entrada a matriz M , o objetivo é prever a nota de um aluno a_i em relação a uma disciplina d_j ainda não cursada por a .

2.1.1 Baseadas em Memória

Os algoritmos baseados em memória ou baseados em vizinhança se utilizam de toda a base de dados de usuários e itens para gerar a previsão. Para cada recomendação é calculado uma medida de similaridade que define a semelhança entre usuários ou entre itens. Com isso, tendo o conjunto de vizinhos, é realizada uma previsão ou recomendação baseada nas classificações ou notas dos n vizinhos mais similares ao usuário alvo.

Existem duas vertentes para a realização da previsão, dentro dessa classe de algoritmos, sendo elas: *baseado em usuário* e *baseado em item*.

2.1.1.1 Previsão baseada em usuário

Os métodos de recomendação de vizinhança baseados no usuário preveem a propensão de um usuário u para um item i , avaliando o *feedback* dado a i pelos usuários mais semelhantes a u (RICCI *et al.*, 2011).

Considere a matriz de avaliações $M : a \times d$ descrita na Seção 2.1. No contexto da previsão baseada em usuário, o primeiro passo a ser dado é o cálculo da similaridade entre o aluno alvo a_i e os demais alunos da matriz M . O cálculo da similaridade entre dois alunos a_i e a_n é descrito na Seção 2.1.1.3. Após calcular a similaridade entre o aluno alvo e os demais alunos, são definidos os k vizinhos mais próximos do aluno alvo a_i . Para prever a aptidão do aluno alvo a_i em relação a uma disciplina d_j , calcula-se a média ponderada das notas dos alunos mais similares ao aluno alvo.

A equação 2.1 representa a previsão do aluno a em relação a disciplina d . Considere A_a

como o conjunto de alunos mais similares ao aluno alvo a , \bar{r}_a a média das notas do aluno a e r_{bd} representa a nota de um aluno $b \in A_a$ em uma disciplina d .

$$pred(a, d) = \bar{r}_a + \frac{\sum_{b \in A_a} sim(a, b)(r_{bd} - \bar{r}_b)}{\sum_{b \in A_a} sim(a, b)} \quad (2.1)$$

Alguns dos problemas da técnica de vizinhança de usuários é a dificuldade na recomendação de itens, como consequência de uma base muito esparsa, além do problema de escalabilidade, ou seja, o crescimento computacional depende do número de usuários e itens.

2.1.1.2 Predição baseada em item

Enquanto os métodos baseados no usuário dependem da opinião de usuários com ideias semelhantes para prever uma classificação, as abordagens baseadas em itens analisam as classificações dadas aos itens semelhantes (RICCI *et al.*, 2011).

Sabendo que o objetivo é prever a nota de um aluno representado pelo vetor a_i em relação a uma disciplina representado pelo vetor d_j . Considere a matriz $M : a \times d$ descrita na Seção 2.1. Essa predição é calculada na equação 2.2, onde D_d representa o conjunto de k disciplinas mais similares a disciplina alvo e que foram cursadas pelo aluno a_i . A função $sim(d, v)$ representa a similaridade entre a disciplina d e v . Por fim, r_{aj} representa a nota do aluno a para o item j .

$$pred(a, d) = \frac{\sum_{j \in D_d} sim(d, j)(r_{aj})}{\sum_{j \in D_d} sim(d, j)} \quad (2.2)$$

A predição baseada em item tem o propósito de solucionar o problema de escalabilidade da predição baseado em usuário.

2.1.1.3 Similaridade

A similaridade define a proximidade entre um conjunto k de usuários e o usuário alvo, de modo que o conjunto k tenha avaliado o item a ser predito. Além disso, a similaridade pode definir a proximidade entre um conjunto de itens avaliados por um usuário alvo em relação ao item alvo, no caso da predição baseada em item.

Existem vários meios de calcular a similaridade, as principais formas são apresentadas a seguir.

2.1.1.3.1 Correlação de Pearson

A similaridade entre os itens pode ser dada pela sua correlação que mede a relação linear entre os objetos. Embora existam vários coeficientes de correlação que podem ser aplicados, a correlação de Pearson é a mais usada (RICCI *et al.*, 2011).

Considerando a matriz $M : a \times d$, a equação 2.3, apresenta como é calculado a similaridade entre dois usuários ou alunos. O aluno a e v , são representados por vetores da matriz M , enquanto D_{av} representa o conjunto de disciplinas avaliadas tanto por a como por v , r_{ad} representa a nota de a na disciplina d , por fim, \bar{r}_a e \bar{r}_v representam a média das notas de a e de v , respectivamente.

$$sim(a, v) = \frac{\sum_{d \in D_{av}} (r_{ad} - \bar{r}_a)(r_{vd} - \bar{r}_v)}{\sqrt{\sum_{d \in D_{av}} (r_{ad} - \bar{r}_a)^2} \sqrt{\sum_{d \in D_{av}} (r_{vd} - \bar{r}_v)^2}} \quad (2.3)$$

O cálculo da similaridade entre itens é semelhante ao cálculo da similaridade entre usuários. Considere a matriz $M : a \times d$ para calcular a similaridade entre duas disciplinas d e j , onde o conjunto A_{dj} representa os alunos que cursaram ambas as disciplinas. O cálculo da similaridade entre essas disciplinas é apresentado na equação 2.4. Observe que r_{ad} e r_{aj} representam o nota do usuário a para a disciplina d e j respectivamente.

$$sim(d, j) = \frac{\sum_{a \in A_{dj}} (r_{ad} - \bar{r}_d)(r_{aj} - \bar{r}_j)}{\sqrt{\sum_{a \in A_{dj}} (r_{ad} - \bar{r}_d)^2} \sqrt{\sum_{a \in A_{dj}} (r_{aj} - \bar{r}_j)^2}} \quad (2.4)$$

2.1.1.3.2 Cosseno

O cálculo da similaridade por cosseno usa o valor de cosseno de dois vetores para indicar o grau de similaridade entre eles (ZHANG *et al.*, 2014).

Considerando a matriz $M : a \times d$. A similaridade entre dois alunos a e v é obtida pelo cosseno dos vetores r_{ad} que representa a lista de disciplinas cursadas por a e r_{vd} que representa a lista de disciplinas cursadas por v . O cálculo da similaridade é representada na equação 2.5

$$\text{sim}(a, v) = \cos(r_{ad}, r_{vd}) = \frac{r_{ad} \cdot r_{vd}}{\|r_{ad}\| \cdot \|r_{vd}\|} \quad (2.5)$$

2.1.1.3.3 Jaccard

Outra métrica para calcular a similaridade é o coeficiente Jaccard ou coeficiente Tanimoto. Essa métrica calcula a similaridade entre dois conjuntos através da interseção dividido pelo tamanho da união dos conjuntos de amostras (ZHANG *et al.*, 2014).

A equação 2.6, apresenta o cálculo da similaridade entre a e b , onde D_{ab} representa o conjunto de disciplinas cursadas tanto pelo aluno a como pelo aluno b . U_{ab} representa a cardinalidade da união das disciplinas cursadas pelos dois alunos.

$$\text{sim}(a, v) = \frac{|D_{ad}|}{|U_{ab}|} \quad (2.6)$$

2.1.2 Baseadas em Modelo

Criados com o propósito de resolver os problemas dos algoritmos baseado em memória, os algoritmos FCBM usam o conjunto de classificações para construir um modelo capaz de fazer predições. Esses algoritmos se utilizam de uma abordagem probabilística e produzem recomendações rapidamente. Alguns dos métodos de aprendizado de máquina mais conhecidos para o processo de modelagem são: *Redes bayesianas*, *Árvore de decisão*, *Redes neurais* e *Fatoração de Matrizes (FM)*.

A técnica de *redes neurais* implementa padrões de detecção e algoritmos de aprendizado de máquina para construir modelos de predição para bases de dados histórico em larga escala (PINHEIRO, 2008).

A técnica de *árvore de decisão* consiste em um modelo de predição que pode ser visualizado como uma estrutura de árvores. Especificamente, cada ramo da árvore representa um tópico de classificação, e as folhas da árvore são partições de conjuntos de dados com suas respectivas classificações (PINHEIRO, 2008).

As *redes bayesianas* fornece uma representação gráfica dos relacionamentos probabilísticos entre um conjunto aleatório de variáveis. Os dois elementos chaves de uma *rede bayesiana* são:

Grafo Acíclico Direcionado (GAD) e uma tabela de probabilidades (TAN *et al.*, 2009).

A FM é um dos métodos mais clássicos da FCBM. Nesse modelo os usuários ou itens são associados a um vetor de fatores latentes de menor dimensão. A fatoração é feita em uma matriz de usuários e itens onde cada célula representa uma interação do usuário com o item. A recomendação é feita quando se encontra um item que tenha fatores similares a um usuário. Para sistemas de recomendação, existem diversas técnicas de FM.

Uma das principais técnicas é a *Singular Value Decomposition (SVD)*. A SVD se preocupa com a redução da dimensionalidade da matriz. A decomposição de uma matriz $M : a \times b$ é a fatoração de M em três matrizes: $M = U\Sigma Q^T$. Sendo que $\Sigma_{f \times f}$ é uma matriz diagonal de valores singulares decrescentes, $U^T U = I_a$ e $Q^T Q = I_b$ são matrizes ortogonais, onde f é o posto da matriz M .

Como exemplo, considere a matriz de entrada $M : a \times d$. O primeiro passo a ser feito é a decomposição de M no produto de outras três matrizes, sendo elas $U_{a \times f}$, $\Sigma_{f \times f}$ e $Q_{f \times d}^T$, onde f representa o número de fatores latentes. Observe que cada linha da matriz U representa o quanto cada aluno está associado a cada um dos fatores, a matriz Q^T , por sua vez, representa o quanto cada disciplina está associada a cada um dos fatores. Tendo como objetivo prever a nota de um aluno a em relação a uma disciplina d , considere o vetor u_a de tamanho f da matriz U , onde u_a representa a relação do aluno a com um conjunto de f fatores e o vetor q_d^T representa a relação da disciplina d com um conjunto de n fatores. Utilizando-se da equação 2.7, temos \hat{r}_{ad} que representa o valor predito da nota de a em relação a uma disciplina d .

$$\hat{r}_{ad} = u_a \Sigma_f q_d^T \quad (2.7)$$

Algumas das vantagens dessa abordagem é trabalhar melhor com uma base de dados muito esparsa em comparação com a técnica de vizinhança de usuários, sendo isso consequência da redução da dimensionalidade, além de ocupar menos espaço do que uma matriz de similaridades.

2.2 Filtragem Colaborativa Neural

A estrutura geral, FCN, apresentada por (HE *et al.*, 2017), substitui o produto interno dos recursos latentes de usuários e itens, utilizado na FM, por uma arquitetura neural. Com isso,

a FCN expressa e generaliza a FM em sua estrutura. Além disso, a FCN busca aprender a função de interação item-usuário por meio de uma rede neural perceptron de várias camadas, de modo a prover um alto nível de não linearidades a modelagem de FCN.

Na Seção 2.2.1, é apresentado uma introdução sobre os conceitos de Redes Neurais Artificial (RNA). A Seção 2.2.2, expõe a limitação da FM. A Seção, 2.2.3, mostra a estrutura da FCN que tem como objetivo eliminar a limitação da FM. A Seção, 2.2.4, denota como a FCN é uma generalização da FM. Por fim, a Seção 2.2.5, tem por finalidade apresentar a FMNeu.

2.2.1 Rede Neural Artificial (RNA)

RNA pode ser entendida como um conjunto de unidades de processamento interligadas entre si que formam uma rede. Essas unidades são chamadas de neurônios e as conexões entre essas unidades são chamadas de sinapses (PINHEIRO, 2008).

Os componentes básicos das RNA são os conjuntos de conexões, nos quais cada conexão está associada a um parâmetro que determina o peso da conexão. Na primeira etapa do processo, as entradas são multiplicadas pelos pesos das conexões em cada uma das unidades, ou neurônios. Na segunda etapa, as entradas ponderadas pelos pesos das conexões são somadas, resultando em um valor, chamado de valor de ativação. Na terceira e última etapa, uma função de ativação é aplicada sobre o valor de ativação obtido na saída do neurônio. O comportamento da rede varia de acordo com a função de ativação escolhida (PINHEIRO, 2008).

2.2.2 Limitação da fatoração de matrizes

Os métodos clássicos de fatoração de matrizes apresentados na Seção 2.1.2, sofrem limitações causadas pelo produto escalar dos vetores latentes de usuários e itens. Essas limitações são superadas pelas redes neurais. Para compreender a limitação da FM, considere a matriz apresentada na Figura 1. As linhas da matriz representam os usuários, as colunas representam os itens e os valores definem a relação dos usuários com os itens.

Calculando a similaridade do usuário u_4 em relação aos demais é possível realizar o mapeamento destes no espaço latente ilustrado na Figura 2 onde P_n é o vetor latente que representa o usuário n .

Figura 1 – Matriz de usuário-item

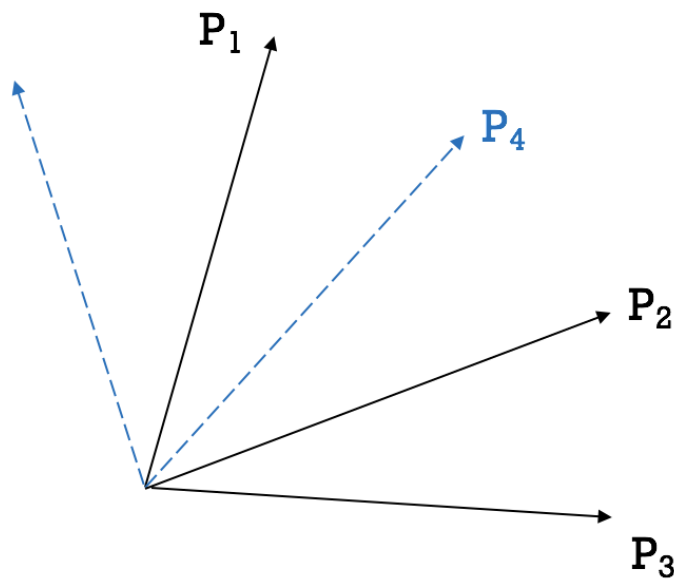
	i_1	i_2	i_3	i_4	i_5
u_1	1	1	1	0	1
u_2	0	1	1	0	0
u_3	0	1	1	1	0
u_4	1	0	1	1	1

← items →

↑ users

Fonte: (HE *et al.*, 2017)

Figura 2 – Espaço latente do usuário



Fonte: (HE *et al.*, 2017)

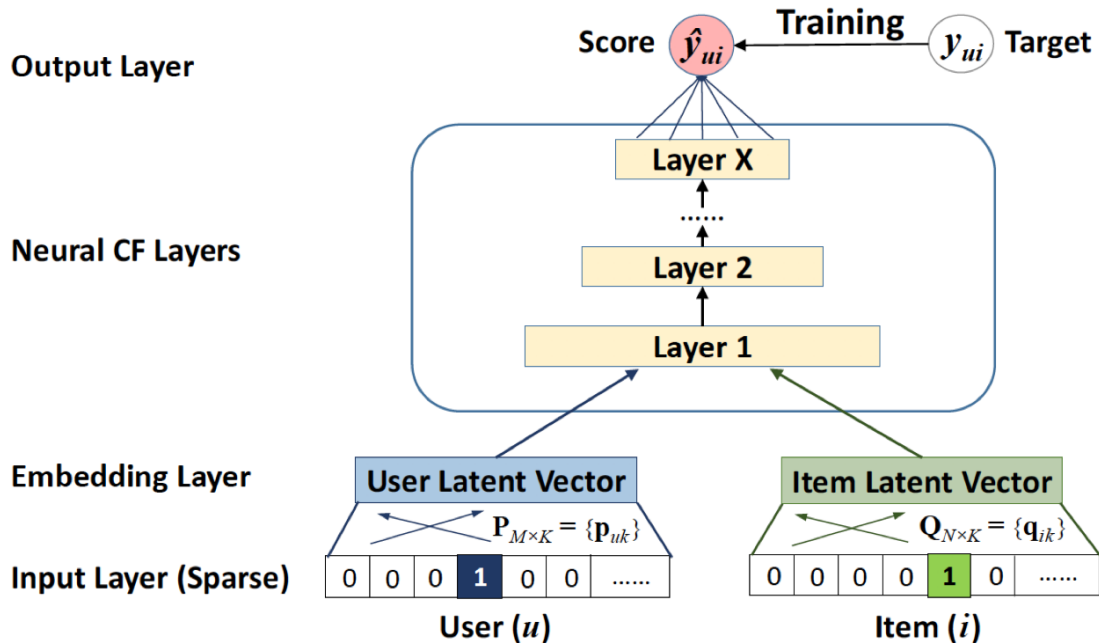
Sabendo que $sim(u, v)$ representa a similaridade entre o usuário u e v . Com o cálculo da similaridade do usuário u_4 em relação aos demais usuários, entende-se que $sim(u_1, u_4) > sim(u_3, u_4) > sim(u_2, u_4)$. Observe que no espaço latente ilustrado na Figura 2, a proximidade do

vetor latente P_4 ao vetor P_1 torna o vetor P_4 mais próximo de P_2 do que do vetor P_3 , causando uma perda de classificação, já que P_4 é mais similar a P_3 do que P_2 . Com isso, pode-se perceber a limitação do produto interno, devido a perda de classificação. O objetivo da FCN é mitigar esse problema de classificação utilizando redes neurais.

2.2.3 Estrutura da filtragem colaborativa neural

A Figura 3 é uma representação da estrutura da FCN, onde a saída de uma camada serve como entrada da próxima camada.

Figura 3 – Estrutura da filtragem colaborativa neural



Fonte: (HE *et al.*, 2017)

A primeira camada, denominada camada de entrada, consiste em um vetor binarizado esparsa com codificação *one-hot* que representa um aluno u e uma disciplina i , representados por $User(u)$ e $Item(i)$, respectivamente. Acima da camada de entrada está a camada de incorporação, que é uma camada totalmente conectada que projeta a representação esparsa em um vetor denso. A incorporação de usuário ou item obtido pode ser vista como o vetor latente do usuário ou item no contexto do modelo de fator latente. Essas incorporações são direcionadas para a camada *Neural FC*, uma arquitetura neural de várias camadas que mapeia os vetores latentes para as pontuações de

previsão. Cada uma das camadas *Neural FC* pode ser personalizada para descobrir certas estruturas latentes de interações entre usuários e itens. A dimensão da última camada oculta X determina a capacidade do modelo. Por fim, a camada final de saída é a pontuação prevista, e o treinamento é realizado minimizando a perda pontual entre \hat{y}_{ui} e o valor alvo y_{ui} (HE *et al.*, 2017).

A equação 2.8 apresenta como o modelo preditivo da FCN é formulado para calcular a interação usuário-item. Observe que a matriz P com dimensão $M \times K$ denota a matriz de fatores latentes para usuários, Q com dimensão $N \times K$ denota a matriz de fatores latentes para itens e Θ_f representa os parâmetros do modelo da função de interação f .

$$\hat{y}_{ui} = f(P^T v_u^T, Q^T v_i^I | P, Q, \Theta_f) \quad (2.8)$$

2.2.4 Fatoração Matricial Generalizada

A FMG é um componente da FCN. A generalização ocorre com a substituição da camada *Neural CF* por uma camada com operação de multiplicação dos elementos de entrada e com a definição dos pesos das bordas da camada de saída. A equação 2.9, expressa a saída prevista da FCN, onde a_{out} representa a função de ativação e h representa os pesos das bordas da camada de saída e \otimes simboliza a multiplicação por elementos - produto interno. Observe que p_u e q_i são vetores latentes de usuário e item, respectivamente, onde p_u substitui $P^T v_u^T$ e q_i substitui $Q^T v_i^I$ na equação 2.8. Perceba que a equação 2.9 é equivalente a equação 2.7 apresentada na Seção 2.1.2 que fala de FM. Com isso, é possível considerar que a FM é um caso especial da FMG (HE *et al.*, 2017).

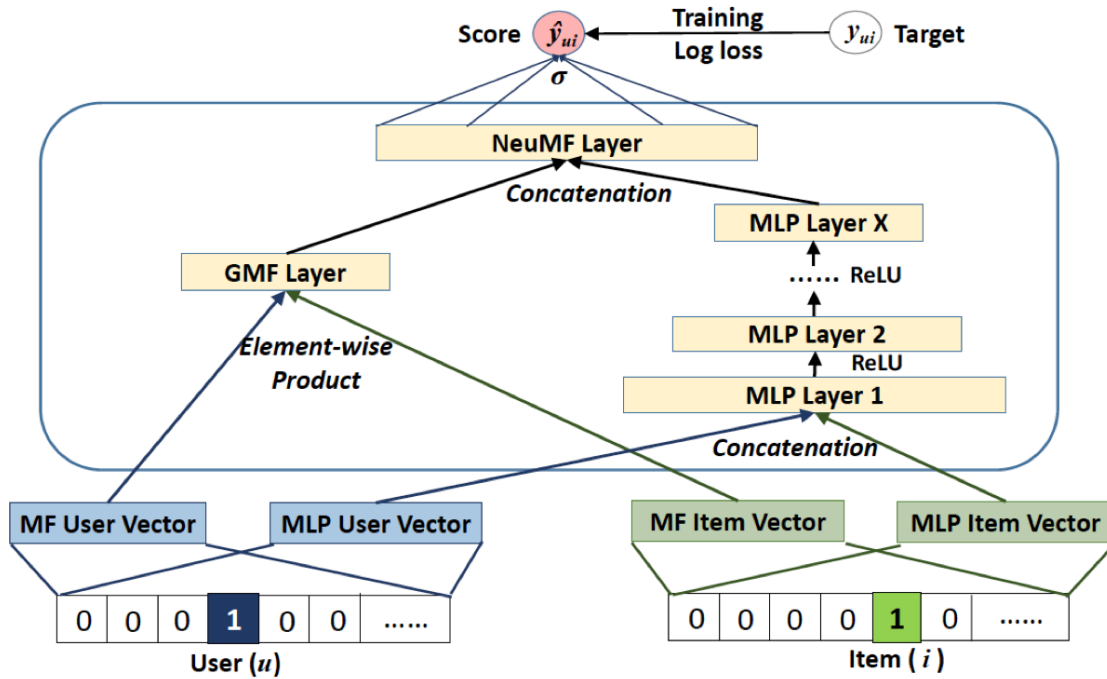
$$\hat{y}_{ui} = a_{out}(h^T(p_u \otimes q_i)) \equiv p_u^T \cdot q_i \quad (2.9)$$

2.2.5 Fatoração de Matrizes Neurais

Esta técnica tem como objetivo, melhorar a modelagem das complexas interações usuário-item, introduzindo não linearidade à estrutura FCN. Para isso, o componente FMG compartilha a camada de incorporação com o RPM, além de concatenar as saídas de suas funções de interação, como apresentado na Figura 4 e formulado na equação 2.10. Com isso, tem-se a combinação de um modelo flexível e não linear das Redes Neurais Profundas (RNP) e a linearidade

do FM, para aprender as interações entre os vetores latentes. Esse modelo é denominado FMNeu (HE *et al.*, 2017).

Figura 4 – Modelo de fatoração de matriz neural



Fonte: (HE *et al.*, 2017)

$$\begin{aligned} \phi^{FMG} &= p_u^{FMG} \otimes q_i^{FMG}, \\ \phi^{RPM} &= a_L \left(W_L^T \left(a_{L-1} \left(\dots a_2 \left(W_2^T \begin{bmatrix} p_u^{RPM} \\ q_i^{RPM} \end{bmatrix} + b_2 \right) \dots \right) \right) + b_L \right), \\ \hat{y}_{ui} &= \sigma \left(h^T \begin{bmatrix} \phi^{FMG} \\ \phi^{RPM} \end{bmatrix} \right) \end{aligned} \quad (2.10)$$

FCN usa FMG com sigmóide como a função de ativação e aprende h (os pesos das arestas) dos dados com perda de log.

FCN usa Unidade linear retificada (ReLU) como uma função de ativação para sua parte RPM.

Na equação 2.10, p_u^{FMG} e p_u^{RPM} denotam a *embeddings* do usuário para FMG e RPM, respectivamente. Enquanto, q_i^{FMG} e q_i^{RPM} reference aos *embeddings* de itens. Assim é possível

observar que FMG e RPM têm *embeddings* de usuário e item separados fornecendo mais flexibilidade ao modelo. A RPM tem como entrada a concatenação de vetores latentes de itens do usuário, utilizando a função *ReLU* como função de ativação e a FMG reproduz a FM, multiplicação por elemento do vetor de usuário-item, tendo como função de ativação a função *sigmoid*. Na camada final as saídas de FMG e RPM são concatenadas.

2.3 Resumo do Capítulo

O atual Capítulo 2, apresentou as principais abordagens da presente pesquisa, bem como uma visão geral da FC, seus objetivos e principais técnicas, como FC baseada em memória, FC baseada em modelo, discutindo definições, limitações e comparações entre elas. Além de discutir sobre FCN na Seção 2.2, apresentando sua estrutura e abordando conceitos como RNP, FMG, FMNeu, bem como limitações e desafios da FM.

3 TRABALHOS RELACIONADOS

Esse capítulo, tem como objetivo apresentar os trabalhos que estão relacionados a predição de desempenho de alunos, além de realizar uma comparação em relação ao presente trabalho.

O estudo de KOTSIANTIS *et al.* 2003, tem como propósito, analisar o uso de técnicas de aprendizado de máquina na previsão de evasão de alunos. O estudo avalia o uso das principais técnicas de aprendizado de máquina sendo elas: *Decision Trees*, *Neural Networks*, *Naive Bayes algorithm*, *Instance-Based Learning Algorithms*, *Logistic Regression* e *Support Vector Machines*. As análises mostram que o algoritmo *Naive Bayes* teve a melhor precisão entre os outros analisados. Por fim, os resultados mostram que os algoritmos de aprendizagem são capazes de prever o abandono de alunos com precisão de 63% nas previsões iniciais baseadas em dados demográficos e superior a 83% antes do meio do período acadêmico.

NGHE *et al.* 2007, investigaram a aplicação de técnicas de mineração de dados na previsão do desempenho acadêmico dos estudantes. Para isso, foram utilizados dados de dois institutos acadêmicos diferentes. Os conjuntos de dados de um dos institutos incluíam registros dos alunos e a média de notas no final do segundo ano para previsão no desempenho no terceiro ano, o outro caso utilizou informações de admissão para previsão do desempenho no final do primeiro ano. O estudo se utiliza de dois algoritmos de mineração de dados, sendo eles *decision Tree* e *bayesian network algorithms*. Como resultado, identificou-se que as *decision Tree* apresentaram melhores precisões.

MANHÃES *et al.* 2012, avalia o uso de técnicas de mineração de dados para identificar alunos em risco de evasão, a partir das primeiras notas semestrais. Com esse propósito, os autores realizaram três experimentos aplicando dez algoritmos de classificação. Os algoritmos utilizados são OneR, JRip, *Decision Table*, *Simple Cart*, J48, *Random Forest*, *Simple Logistic*, *Multilayer Perceptron*, *Naive Bayes* e *Bayes Net*. Como resultado, mostrou-se que é possível identificar a situação final dos alunos com uma precisão de 80%.

BRITO *et al.* 2014, propõe a utilização de métodos de aprendizagem de máquina para predição de desempenho de alunos do primeiro período utilizando-se das notas de ingresso. Para a realização do trabalho foi utilizado cinco algoritmos de aprendizado de máquina sendo eles, *naive*

bayes, IBk , SMO, *random forest e multilayer perceptron*. Como resultado, observou-se que é possível inferir o desempenho dos estudantes com uma acurácia superior a 70%.

Sweeney *et al.* 2015, propõe um sistema cuja tarefa é prever as notas dos cursos dos alunos para o próximo período. Duas classes de métodos para previsão são explorados: *simple baselines e matrix factorization based methods*. Para essa realização, o estudo afirma que o modelo *factorization machine*, um algoritmo de fatoração matricial de uso geral, é o mais adequado para a tarefa, de modo que possam prever notas para alunos novos e recorrentes e para cursos novos e existentes.

Meier *et al.* 2015, propõem um algoritmo que faz uma previsão personalizada e oportuna da nota de cada aluno em uma classe, utilizando-se de pontuações obtidas pelos alunos nas avaliações de desempenho iniciais. O estudo investiga dois tipos de previsões, sendo elas, configuração de regressão, onde a pontuação geral de cada aluno é prevista e definição de classificação onde os alunos são classificados em duas ou mais categorias. Como resultado, mostrou-se que, além de ser capaz de aprender com dados passados, o algoritmo supera os algoritmos de referência em relação à precisão e acurácia nas configurações de classificação e regressão.

ELBADRAWY *et al.* 2015, analisam a eficácia de uma classe de modelos de regressão colaborativa para predição do desempenho dos alunos. Os modelos de regressão são avaliados usando as informações históricas de todos os alunos. Como resultado, identificou-se que um modelo colaborativo de regressão múltipla tem um desempenho melhor que a regressão linear única. Isso se justifica pela capacidade de capturar diferenças pessoais dos alunos.

BYDŽOVSKÁ; HANA 2015, relatam o uso de diferentes métodos de FC na previsão do desempenho do aluno com base na história do aluno. Os resultados são comparados com os resultados obtidos por algoritmos de classificação, que são algoritmos também utilizados para prever o desempenho dos alunos. A pesquisa mostra que os resultados dos algoritmos são muito semelhantes. Os autores afirmam que a abordagem de FC tem como principal vantagem, necessitar apenas das notas dos alunos para previsão, dados esses que estão disponíveis em todas as universidades.

O trabalho de ASIF *et al.* 2017, tem como objetivo analisar o desempenho de estudantes que cursam um programa de bacharelado de 4 anos nas disciplinas de TI, combinando três abordagens da mineração de dados educacionais: *prediction, clustering e distillation of data for human judgment*. O estudo descreve em três pontos, a abordagem para atingir esse objetivo: O primeiro,

vários classificadores são gerados para prever o desempenho dos estudantes. O segundo, com os classificadores, é possível obter cursos que possam servir como indicadores efetivos do desempenho dos alunos. O terceiro, em que é investigado como o desempenho acadêmico dos alunos progride ao longo do programa de quatro anos como uma espécie de triangulação. Dois métodos são usados neste trabalho: classificação e *clustering*. Para a previsão do desempenho foram utilizados vários algoritmos de classificação, sendo eles, *decision trees*, *rule induction*, *artificial neural networks*, *k-nearest neighbours*, *Naive Bayes* e *random forest trees*.

MORSY; KARYPIS 2017, propõem uma modelagem para o problema de previsão de notas para o próximo ano a fim de ajudar os alunos a tomar decisões sobre quais cursos se inscrever. O modelo é descrito como um modelo de regressão baseado no conhecimento, que acumula o desempenho de um aluno em todos os cursos já realizados por ele, assumindo a existência de um espaço de componentes de conhecimento que descrevem o currículo geral. No trabalho é investigado três maneiras diferentes de definir a estrutura do espaço do componente de conhecimento, sendo dois deles baseados em um espaço latente, e o terceiro baseia-se nas descrições textuais dos cursos. Como resultado, foi identificado que os modelos que usam as descrições textuais dos cursos podem ser usados para identificar o conhecimento necessário para que os alunos tenham bom desempenho nos cursos.

3.1 Resumo do Capítulo

O presente Capítulo 3, tem como objetivo apresentar os trabalhos que estão relacionados a análise e predição de desempenho, utilizando métodos de *machine learning*, abordagens da mineração de dados educacionais, além de diferentes métodos de FC. A Tabela 1 apresenta um comparativo entre as principais características dos trabalhos relacionados e o presente trabalho.

Tabela 1 – Comparativo de trabalhos relacionados

Trabalho	Contexto	Técnica
KOTSIANTIS <i>et al.</i>	Previsão de evasão de alunos	<i>Decision Trees, Neural Networks, Naive Bayes algorithm, Instance-Based Learning Algorithms, Logistic Regression e Support Vector Machines.</i>
NGHE <i>et al.</i>	Previsão do desempenho acadêmico dos estudantes	<i>Decision Tree e bayesian network algorithms</i>
MANHÃES <i>et al.</i>	Identificar alunos em risco de evasão	<i>OneR, JRip, Decision Table, Simple Cart, J48, Random Forest, Simple Logistic, Multilayer Perceptron, Naive Bayes e Bayes Net</i>
BRITO <i>et al.</i>	Predição de desempenho de alunos do primeiro período	<i>Naive bayes, IBk , SMO, random forest e multilayer perceptron</i>
Sweeney <i>et al.</i>	Prever as notas dos cursos dos alunos para o próximo período	<i>Simple baselines e matrix factorization based methods</i>
Meier <i>et al.</i>	Previsão personalizada da nota de alunos	Configurações de classificação e regressão.
ELBADRAWY <i>et al.</i>	Analisar a eficácia de uma classe de modelos de regressão colaborativa para predição do desempenho dos alunos	Modelo colaborativo de regressão múltipla e regressão linear única
BYDŽOVSKÁ; HANA	Previsão do desempenho do aluno com base no histórico do aluno	Filtragem colaborativa
ASIF <i>et al.</i>	Analisar o desempenho de estudantes	<i>Decision trees, rule induction, artificial neural networks, k-nearest neighbours, Naive Bayes e random forest trees.</i>
MORSY; KARYPIS	Previsão de notas para o próximo ano	Modelo de regressão baseado no conhecimento.
Este Trabalho	Predição de desempenho de alunos no próximo período baseado no histórico de notas	Filtragem Colaborativa Neural

Fonte: Autoria Própria.

4 PROPOSTA

Este capítulo tem como objetivo descrever as soluções propostas para prever o desempenho dos alunos com base no histórico de notas, apresentando os experimentos e descrevendo o conjunto de dados. Assim, o capítulo se organiza da seguinte forma: A Seção 4.1 apresenta a base de dados utilizada, bem como foi obtida e pré-processada. A Seção 4.2 discorre sobre as metodologias de testes e predição do desempenho. Por fim, a Seção 4.3 define o modelo a ser utilizado na predição do desempenho de alunos e 4.4 define com mais detalhes a arquitetura do modelo utilizado.

4.1 Aquisição e Tratamento dos Dados

Nessa etapa, utilizou-se a base de dados da UFC Campus Crateús, composta pelos alunos e disciplinas do curso de Engenharia Civil, obtidas no formato *csv*. O curso de Engenharia Civil foi escolhido por ter o maior número de alunos, além de sua estrutura curricular ser bem definida e com poucas mudanças.

Após a obtenção da base de dados, foi realizado o pré-processamento onde os dados sensíveis foram removidos. As principais informações utilizadas são as notas, semestre e a disciplina cursada de modo que não identifique o aluno. Os dados utilizados para a previsão são referentes aos semestres 2016.1, 2016.2, 2017.1, 2017.2, 2018.1, 2018.2, 2019.1, 2019.2, 2020.1, 2020.2 e 2021.1. A Tabela 2 apresenta o número de alunos e disciplinas para cada semestre, sendo um total de 251 alunos e 54 disciplinas.

Tabela 2 – Número de alunos e disciplinas por semestre

Semestre	N.º de alunos	N.º de disciplinas
2016.1	74	17
2016.2	78	21
2017.1	106	27
2017.2	114	33
2018.1	142	38
2018.2	150	44
2019.1	178	53
2019.2	182	53
2020.1	207	54
2020.2	180	54

Fonte: Autoria Própria.

4.2 Metodologia de Teste e Predição do Desempenho

A predição do desempenho do aluno como aprovado ou reprovado se caracteriza como um problema de FC, onde 0 representa a reprovação e 1 aprovação. A base de dados pré-processada foi processada e utilizada para predizer o desempenho de um aluno com base em n outros alunos. Esses dados foram utilizados para a construção e teste do modelo. O principal objetivo é analisar o quão bem o modelo irá apresentar os resultados. Cada modelo é treinado n vezes para um conjunto de hiperparâmetros.

Nesse problema, o conjunto de dados é dividido em duas partes onde cada uma é utilizada em fases distintas, sendo elas: Treinamento e teste. A Tabela 3 apresenta a divisão dos dados de treinamento e teste para cada semestre.

Tabela 3 – Divisão dos dados de treinamento e teste

Semestre	Treinamento	Teste
2016.1	47.79%	52.21%
2016.2	59.05%	40.95%
2017.1	55.83%	44.17%
2017.2	67.44%	32.56%
2018.1	70.45%	29.55%
2018.2	76.68%	23.32%
2019.1	78.31%	21.69%
2019.2	81.10%	18.90%
2020.1	83.57%	16.43%
2020.2	89.01%	10.99%

Fonte: Autoria Própria.

Durante o treinamento e teste os dados foram submetidos ao modelo de duas formas: Por disciplina, o modelo recebe o aluno e a disciplina e deve definir se o aluno foi aprovado ou reprovado. Por grupo de disciplinas, o modelo recebe o aluno e todas as disciplinas que ele se matriculou e dentre as disciplinas, o modelo deve retornar em quais o aluno foi aprovado ou reprovado.

A Seção 2.2 apresentou a estrutura da FCN que é composta por dois outros modelos FMG e RPM. Neste trabalho, durante o processo de treinamento e teste, os modelos FCN, FMG e RPM foram submetidos a diferentes hiperparâmetros de forma individual, os hiperparâmetros são melhor detalhados do Seção 4.4, como exemplo, os modelos foram submetidos a diferentes *hidden states*, que define quantos fatores latentes devem ser utilizados para representar o aluno e as disciplinas.

Desse modo, será analisado o comportamento dos modelos durante a predição com diferentes hiperparâmetros e com as duas formas de submissão descritas. Para isso utilizou-se a ferramenta *colab* com configuração máxima de 12,68 GB de memória RAM e 107,72 GB de disco.

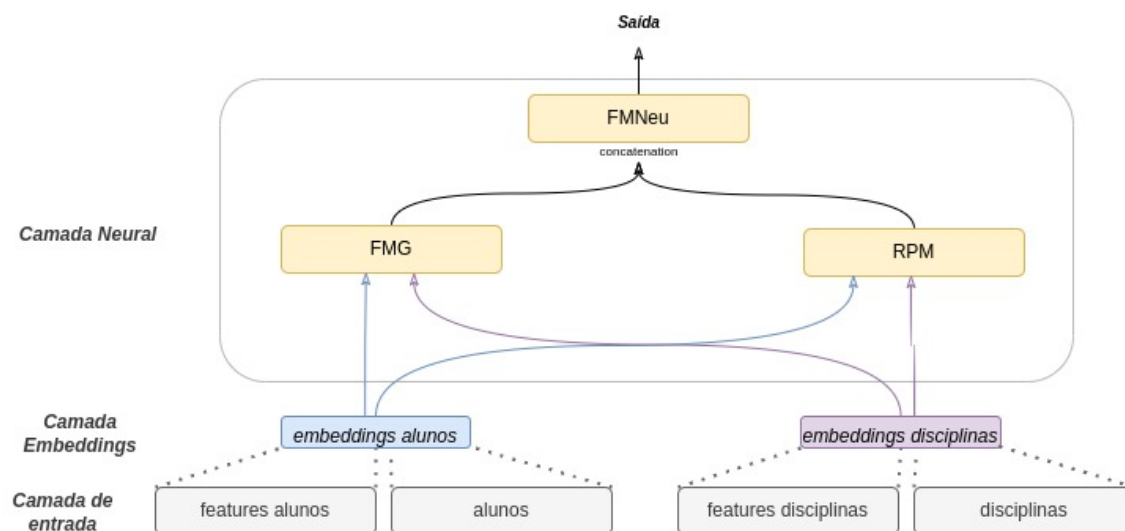
4.3 Definição dos modelos a serem utilizados

Os dados apresentados na Seção 4.1 foram submetidos a FCN de modo a modelar as interações entre os vetores latentes, ou seja, entre os alunos e disciplinas, podendo assim identificar similaridades entre os mesmos e poder prever o nível de afinidade dos alunos em relação às disciplinas. Os resultados da FCN são comparados aos resultados da aplicação de técnicas clássicas de FC, como FM que calcula o produto interno do recursos latentes de usuários e itens.

4.4 Arquitetura

Baseado no modelo FMNeu apresentado na Seção 2.2.5, o modelo de FCN utilizado foi customizado para o problema atual. A Figura 5 é uma representação do modelo utilizado que tem como entrada vetores que representam os alunos e disciplinas com identificadores numéricos, além de vetores de features, descrita com mais detalhes a seguir, que tornam a camada de *embeddings* mais completa para representação dos alunos e disciplinas e possibilitando extrair mais características latentes dos mesmos.

Figura 5 – Modelo de filtragem colaborativa neural customizado



Fonte: Aatoria Própria

As *features* do aluno são vetores que representam as disciplinas que o aluno cursou e vetores com valores binários que definem se o aluno foi aprovado ou reprovado. As *features* da disciplina são representadas por vetores que identificam os alunos que cursaram a disciplina e vetores com valores binários que definem se o aluno foi aprovado ou reprovado.

Ainda na Figura 5, a última camada recebe os *embeddings* que são manipulados pelos modelos FMG e RPM e suas saídas são concatenadas e utilizadas como entrada para o modelo FMNeu que, por sua vez, tem como saída o resultado previsto.

Todos os modelos utilizam como hiperparâmetro o *Optimizer Adam*. A Tabela 4 apresenta os hiperparâmetros fixos utilizados nos modelos RPM e FMNeu.

Tabela 4 – Hiperparâmetros utilizados nos modelos RPM e FMNeu

Hiper parâmetro fixado	valor
Optimizer	Adam
Learning Rate	1e-3
Activation function	ReLU

Fonte: Autoria Própria.

5 RESULTADOS

Nesta seção, serão apresentados os resultados que permitem avaliar o desempenho dos modelos FCN, FMG e RPM com diferentes hiperparâmetros e as duas formas de submissão dos dados, sendo elas por disciplina e por grupo de disciplinas, como mencionado na Seção 4.2. Com isso, será possível identificar os melhores resultados de cada modelo e entre os modelos qual teve o melhor desempenho para o atual problema de predição com o corrente conjunto de dados.

Para analisar o desempenho dos modelos e compará-los foi utilizado a média das métricas: *accuracy*, *precision*, *recall* e *f1_score*.

5.1 Resultados da Fatoração Matricial Generalizada

Abaixo, a Tabela 5 apresenta o desempenho do modelo FMG para cada valor do hiperparâmetro *hidden states* que identifica o número de fatores latentes para representar o aluno e disciplina. É possível observar que FMG, com a média 0,890, teve um melhor desempenho utilizando *hidden states 16* para cada aluno e disciplina por vez.

Tabela 5 – Média das métricas para cada *hidden states*

<i>hidden states</i>	Disciplinas Agrupadas	Disciplina por vez
8	0,878	0,887
16	0,889	0,890
32	0,884	0,889

Fonte: Aatoria Própria.

5.2 Resultados da Rede Perceptron Multicamada

O modelo RPM foi submetido a diferentes valores dos hiperparâmetros *hidden states* e *hidden layer* com o objetivo de identificar os melhores parâmetros. A seguir, a Tabela 6 apresenta os melhores resultados do modelo RPM com os principais valores dos hiperparâmetros *hidden states* e *hidden layer*.

É possível observar que o modelo RPM apresentou um melhor desempenho, com média 0,858 para os hiperparâmetros *hidden states 16* com 8 camadas ocultas e para cada aluno com o grupo de disciplinas em que ele está matriculado.

Tabela 6 – Média das métricas para cada hidden states / hidden layer

hidden states / hidden layer	Disciplinas Agrupadas	Disciplina por vez
16 / 8	0,858	0,809
8 / 8	0,785	0,852

Fonte: Aatoria Própria.

5.3 Resultados da Filtragem Colaborativa Neural

O modelo FCN foi treinado e avaliado tendo como base os hiperparâmetros com melhores decorrências dos modelos FMG e RPM. Os resultados obtidos são apresentados na Tabela 7. É possível observar que o modelo FCN teve um melhor desempenho quando utilizado o aluno e uma disciplina por vez.

Tabela 7 – Média das métricas do modelo FCN

Disciplinas Agrupadas	Disciplina por vez
0,854	0,869

Fonte: Aatoria Própria.

5.4 Melhores resultados dos modelos

A seguir, a Tabela 8 apresenta os dados para comparação dos melhores resultados entre os modelos FMG, RPM e FCN:

Tabela 8 – Melhores métricas de cada modelo

Métricas	FCN	RPM	FMG
Accuracy	0,863	0,841	0,891
Precision	0,896	0,841	0,931
Recall	0,870	0,913	0,871
F1_score	0,848	0,840	0,867
Média	0,869	0,858	0,890

Fonte: Aatoria Própria.

6 CONCLUSÕES E TRABALHOS FUTUROS

Os resultados apresentados na Seção 5 expressam que os modelos testados tiveram bons resultados na predição do desempenho de alunos com base no histórico de notas. Era esperado que o modelo FCN apresentasse os melhores resultados para esse objetivo. Contudo, é evidente que, baseando-se nos testes e resultados obtidos com o atual conjunto de dados, o modelo clássico FMG teve um melhor desempenho quando comparado aos modelos FMG e RPM. É válido observar que os modelos foram submetidos a um conjunto de dados relativamente pequeno o que pode ter influenciado nos resultados, definindo o modelo FMG como o melhor dentre os testados.

Como trabalho futuro, percebe-se a necessidade de refinamento do modelo para previsão de desempenho no primeiro semestre e análise do modelo com a utilização de um conjunto de dados mais robusto. Além disso, o presente trabalho utilizou *features* binárias para representar os alunos e disciplinas, *features* que definem se o aluno foi aprovado ou reprovado. Sabendo disso, um segundo trabalho seria analisar o uso das notas em um modelo de regressão comparando com o atual modelo de classificação.

REFERÊNCIAS

- ASIF, R.; MERCERON, A.; ALI, S. A.; HAIDER, N. G. Analyzing undergraduate students' performance using educational data mining. **Computers Education**, v. 113, p. 177 – 194, 2017. ISSN 0360-1315. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0360131517301124>.
- BRITO, D. M.; JÚNIOR, I. A. de A.; QUEIROGA, E. V.; RÊGO, T. G. do. Predição de desempenho de alunos do primeiro período baseado nas notas de ingresso utilizando métodos de aprendizagem de máquina. **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)**, v. 25, n. 1, p. 882, 2014. ISSN 2316-6533. Disponível em: <https://br-ie.org/pub/index.php/sbie/article/view/3024>.
- BURKE, R. Hybrid recommender systems: Survey and experiments. **User Modeling and User-Adapted Interaction**, v. 12, 11 2002.
- BYDŽOVSKÁ; HANA. Are collaborative filtering methods suitable for student performance prediction? In: PEREIRA, F.; MACHADO, P.; COSTA, E.; CARDOSO, A. (Ed.). **Progress in Artificial Intelligence**. Cham: Springer International Publishing, 2015. p. 425–430. ISBN 978-3-319-23485-4.
- ELBADRAWY, A.; STUDHAM, R. S.; KARYPIS, G. Collaborative multi-regression models for predicting students' performance in course activities. In: **Proceedings of the Fifth International Conference on Learning Analytics And Knowledge**. New York, NY, USA: Association for Computing Machinery, 2015. (LAK '15), p. 103–107. ISBN 9781450334174. Disponível em: <https://doi.org/10.1145/2723576.2723590>.
- HE, X.; LIAO, L.; ZHANG, H.; NIE, L.; HU, X.; CHUA, T.-S. Neural collaborative filtering. In: **Proceedings of the 26th International Conference on World Wide Web**. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2017. (WWW '17), p. 173–182. ISBN 9781450349130. Disponível em: <https://doi.org/10.1145/3038912.3052569>.
- INEP, I. N. D. E. E. P. E. A. T. Censo educacao superior. 2018. Disponível em: http://download.inep.gov.br/educacao_superior/censo_superior/documentos/2019/censo_da_educacao_superior_2018-notas_estatisticas.pdf.
- KOTSIANTIS, S. B.; PIERRAKEAS, C. J.; PINTELAS, P. E. Preventing student dropout in distance learning using machine learning techniques. In: PALADE, V.; HOWLETT, R. J.; JAIN, L. (Ed.). **Knowledge-Based Intelligent Information and Engineering Systems**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003. p. 267–274. ISBN 978-3-540-45226-3.
- MANHÃES, L.; CRUZ, S. da; COSTA, R. M.; ZAVALETA, J.; ZIMBRÃO, G. Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)**, v. 1, n. 1, 2012. ISSN 2316-6533. Disponível em: <https://www.br-ie.org/pub/index.php/sbie/article/view/1585>.
- Meier, Y.; Xu, J.; Atan, O.; Schaar, M. v. d. Personalized grade prediction: A data mining approach. In: **2015 IEEE International Conference on Data Mining**. [S. l.: s. n.], 2015. p. 907–912.

MORSY, S.; KARYPIS, G. Cumulative knowledge-based regression models for next-term grade prediction. In: _____. **Proceedings of the 2017 SIAM International Conference on Data Mining**. [S. n.], 2017. p. 552–560. Disponível em: <https://epubs.siam.org/doi/abs/10.1137/1.9781611974973.62>.

NGHE, N. T.; JANECEK, P.; HADDAWY, P. A comparative analysis of techniques for predicting academic performance. In: **2007 37th Annual Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports**. [S. l.: s. n.], 2007. p. T2G–7–T2G–12.

PINHEIRO, C. A. R. **Inteligência Analítica: Mineração de Dados e Descoberta de Conhecimento**. 1. ed. <https://www.lcm.com.br/site/livros/detalhesLivro/f/inteligencia-analitica—mineracao-de-dados-e-descoberta-de-conhecimento.html>: Editora Ciência Moderna, 2008. v. 1. ISBN 9788573937077.

RICCI, F.; ROKACH, L.; SHAPIRA, B. Introduction to recommender systems handbook. In: _____. **Recommender Systems Handbook**. Boston, MA: Springer US, 2011. p. 1–35. ISBN 978-0-387-85820-3. Disponível em: https://doi.org/10.1007/978-0-387-85820-3_1.

SCHAFFER, J. B.; FRANKOWSKI, D.; HERLOCKER, J.; SEN, S. Collaborative filtering recommender systems. In: _____. **The Adaptive Web: Methods and Strategies of Web Personalization**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 291–324. ISBN 978-3-540-72079-9. Disponível em: https://doi.org/10.1007/978-3-540-72079-9_9.

Sweeney, M.; Lester, J.; Rangwala, H. Next-term student grade prediction. In: **2015 IEEE International Conference on Big Data (Big Data)**. [S. l.: s. n.], 2015. p. 970–975.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introdução ao datamining: mineração de dados**. 1. ed. <https://www.lcm.com.br/site/livros/detalhesLivro/f/introducao-ao-data-mining—mineracao-de-dados.html>: Editora Ciência Moderna, 2009. ISBN 9788573937619.

ZHANG, Y.; YANG, C.; NIU, Z. A research of job recommendation system based on collaborative filtering. In: **2014 Seventh International Symposium on Computational Intelligence and Design**. [S. l.: s. n.], 2014. v. 1, p. 533–538.