

Recomenda  o Personalizada de Filmes: Desenvolvimento de Software e Prova de Conceito de Feature Weighting (Utilizando Streamlit)

Glenda Souza Fernandes Dos Santos; Isaac Oliveira Dias; Jo  o Luccas Lordelo Marques; Marcus Vinicius Lameu Lima; Paulo Henrique Pereira Araujo Piedade; Adailton de Jesus Cerqueira Junior.

Universidade Salvador – Salvador – BA – Brasil

INTRODU  O

O desafio da sobrecarga de informa  o em plataformas de m  dia    mitigado pelos Sistemas de Recomenda  o (SR). Este trabalho foca na arquitetura Baseada em Conte  do (Content-Based), que frequentemente falha ao atribuir pesos uniformes   s features dos itens. O problema de pesquisa    como otimizar a precis  o do SR Content-Based atrav  s da pondera  o din  mica e estrat  gica de caracter  sticas. O Objetivo Geral    desenvolver um sistema Content-Based aprimorado com um m  dulo de Pondera  o de Caracter  sticas (Feature Weighting) e uma interface web funcional. A metodologia    de Desenvolvimento Tecnol  gico e Experimental, utilizando Similaridade de Cosseno sobre vetores criados via One-Hot Encoding (dados IMDB). O projeto    dividido em Treinamento (constru  o da matriz de similaridade) e Infer  ncia (aplica  o dos pesos din  micos via Streamlit).

M  TODOS

Treinamento: O script inicial (treinamento_modelo_filmes.py) realiza o pr  -processamento de dados do IMDB, incluindo a Codifica  o One-Hot para transformar caracter  sticas categ  ricas (G  nero, Diretor, Ator, etc.) em vetores num  ricos. Em seguida, calcula a Matriz de Similaridade de Cosseno entre os filmes, essencial para o algoritmo. Infer  ncia e Usabilidade: O aplicativo (app_filmes.py), constru  do com Streamlit, implementa a fase de recomenda  o em tempo real (Usu  rio-para-Item). O sistema aplica a melhoria metodol  gica de Pondera  o de Caracter  sticas (Feature Weighting), atribuindo pesos vari  veis (e.g., G  nero pesa mais que Produtora) de forma din  mica, conforme a sele  o do usu  rio, para otimizar a precis  o das sugest  es. Embasamento: A arquitetura    fundamentada na Modelagem de Itens e na Vetoriza  o de Caracter  sticas, onde a similaridade    medida pela proximidade dos vetores no espa  o multidimensional.

RESULTADOS

Esta se  o apresenta e analisa o desempenho do Sistema de Recomenda  o Content-Based aprimorado. A coer  ncia l  gica das recomenda  es foi avaliada utilizando a M  dia de Jaccard, que compara as caracter  sticas definidas nos Perfis Sint  ticos com as caracter  sticas presentes nos filmes sugeridos. O   ndice m  dio encontrado foi de 0,2513 (25,13%), o que indica que aproximadamente um quarto das features fornecidas pelo usu  rio aparecem, em m  dia, nos filmes recomendados. Observou-se uma diferen  a clara entre os tipos de perfis avaliados: os Perfis Comuns, baseados em categorias frequentes no conjunto de dados, apresentaram os melhores   ndices de coer  ncia, pois o modelo encontra diversas inst  ncias semelhantes no treinamento, resultando em maior similaridade. Em contraste, os Perfis Desconexos, compostos por combina  es raras ou pouco representadas, geraram baixa Coer  ncia Jaccard. Isso ocorre devido    forte esparsidade dos vetores One-Hot e    depend  ncia direta do modelo em rela  o    distribui  o dos dados, o que reduz a qualidade das recomenda  es para categorias pouco frequentes. A capacidade de discrimina  o do sistema foi analisada por meio da Curva ROC, considerando filmes com rating igual ou superior a 7,0 como relevantes. O valor de AUC obtido foi de 0,5854, indicando que o modelo possui uma probabilidade de aproximadamente 59% de classificar corretamente um filme relevante acima de um irrelevante. Embora esse valor seja superior ao desempenho aleat  rio, ele revela que o sistema ainda necessita de ajustes na pondera  o das caracter  sticas e em sua representa  o vetorial para melhorar a precis  o e a relev  ncia das recomenda  es apresentadas ao usu  rio. Esses resultados demonstram tanto a efic  cia parcial do modelo quanto as oportunidades claras de aprimoramento para vers  es futuras.

REFER  NCIAS BIBLIOGR  FICAS

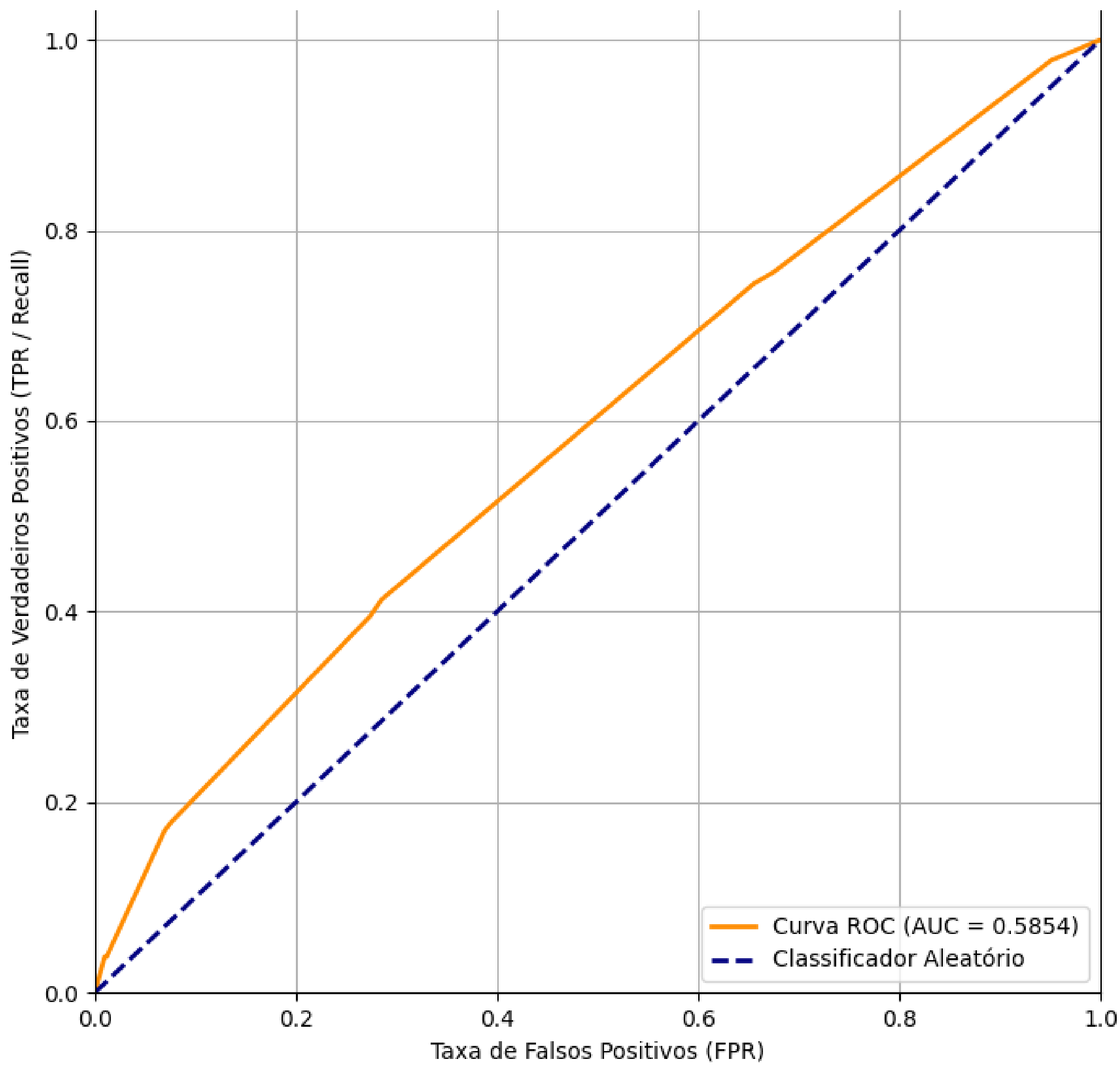
1. RICCI, Francesco et al. Recommender systems handbook. New York: Springer, 2015.

2. HAN, Jiawei; PEI, Jian; KAMBER, Micheline. Data Mining: Concepts and Techniques. 3. ed. Amsterdam: Elsevier, 2011.

Tabela: Coer  ncia Jaccard por Perfil de Entrada

ID Perfil	Tipo de Perfil	Jaccard	% Coer��ncia (Match)		
1	Perfis Comuns	0,3735	37,35%	Media Jaccard	0,2513
2	Perfis Comuns	0,3071	30,71%	% Coerencia	25,13%
3	Perfis Comuns	0,4141	41,41%		
4	Perfis Comuns	0,4604	46,04%		
5	Perfis Comuns	0,3764	37,64%		
6	Perfis Descone	0,2267	22,67%		
7	Perfis Comuns	0,3092	30,92%		
8	Perfis Comuns	0,2765	27,65%		
9	Perfis Descone	0,2090	20,90%		
10	Perfis Descone	0,2078	20,78%		
11	Perfis Descone	0,1709	17,09%		
12	Perfis Descone	0,1997	19,97%		
13	Perfis Descone	0,1661	16,61%		
14	Perfis Comuns	0,2671	26,71%		
15	Perfis Descone	0,1270	12,70%		
16	Perfis Comuns	0,2660	26,60%		
17	Perfis Descone	0,2366	23,66%		
18	Perfis Comuns	0,2724	27,24%		
19	Perfis Descone	0,2096	20,96%		
20	Perfis Descone	0,2316	23,16%		

Figura – Curva ROC para Avalia  o de Relev  ncia de Filmes (Rating ≥ 7.0)



CONCLUS  O

O trabalho desenvolveu e avaliou um Sistema de Recomenda  o Content-Based com pondera  o de caracter  sticas e Similaridade de Cosseno aplicado aos dados do IMDB. Os testes com perfis sint  ticos mostraram coer  ncia nas recomenda  es quando as caracter  sticas eram frequentes, por  m baixo desempenho em perfis com features raras devido    esparsidade dos dados. A Curva ROC (AUC = 0,5854) indicou capacidade moderada de discrimina  o entre filmes relevantes e irrelevantes, evidenciando a necessidade de ajustes nos pesos. O sistema e a interface em Streamlit configuram uma prova de conceito funcional e base para melhorias. Como trabalhos futuros, prop  em-se otimiza  o din  mica da pondera  o, integra  o com m  todos h  bridos e avalia  o de usabilidade com usu  rios reais.