

Sistema de Recomendação de Jogos da Steam

Jorge Sugano Suzart

Novembro 2025

1 Introdução

Este trabalho apresenta o desenvolvimento de um **sistema de recomendação de jogos da plataforma Steam**, que tem como objetivo auxiliar usuários a descobrirem novos títulos semelhantes aos seus jogos favoritos. O sistema aplica técnicas de *Processamento de Linguagem Natural* (PLN) e *Álgebra Linear* para analisar as descrições dos jogos e calcular a similaridade textual entre eles.

A base de dados utilizada (`steam.csv`) contém informações como nome, gênero, tags e descrições dos jogos disponíveis na Steam. Por meio do modelo **TF-IDF (Term Frequency–Inverse Document Frequency)** e da **similaridade do cosseno**, o sistema identifica jogos com maior proximidade temática.

O projeto foi desenvolvido na linguagem **Python**, utilizando as bibliotecas **Pandas**, **NumPy**, **scikit-learn**, **NLTK** e **ipywidgets**. Ele pode ser executado tanto em ambiente interativo (Google Colab) quanto diretamente pelo terminal.

2 Funcionamento do Sistema

O funcionamento do sistema de recomendação segue as seguintes etapas:

1. **Carregamento e preparação dos dados:** o arquivo `steam.csv` é lido com a biblioteca **Pandas**. As colunas `genres` e `steamspy_tags` são combinadas para formar uma descrição textual completa de cada jogo.
2. **Limpeza e pré-processamento:** são aplicadas expressões regulares para remover caracteres especiais, e o texto é convertido para letras minúsculas. Em seguida, são removidas *stopwords* em inglês e aplicada a lematização de cada palavra com o **WordNetLemmatizer** da biblioteca **NLTK**.
3. **Transformação vetorial:** o algoritmo **TfidfVectorizer** do **scikit-learn** é utilizado para converter o texto em vetores numéricos, onde cada palavra recebe um peso proporcional à sua importância dentro do corpus.
4. **Cálculo da similaridade:** a matriz de similaridade é obtida por meio da `cosine_similarity`, permitindo medir a proximidade entre cada par de jogos.
5. **Interface de interação:** o sistema oferece dois modos de uso:

- **Interface gráfica no Google Colab:** utilizando `ipywidgets`, o usuário digita o nome de um jogo e recebe uma tabela com os 10 jogos mais similares, colorida conforme o grau de similaridade.
- **Interface de terminal:** o usuário digita o nome de um jogo e visualiza no console uma tabela textual com os resultados ordenados.

3 Análise dos Ângulos entre Vetores

Cada jogo é representado por um vetor numérico que descreve sua composição textual. A similaridade entre dois jogos é calculada pelo **coseno do ângulo** entre seus vetores. Matematicamente, é dada por:

$$\cos(\theta) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \|\vec{B}\|}$$

onde \vec{A} e \vec{B} são os vetores TF-IDF de dois jogos. Assim:

- Se $\cos(\theta) \approx 1$, os jogos são muito semelhantes;
- Se $\cos(\theta) \approx 0$, eles compartilham poucos elementos em comum.

Essa relação vetorial é essencial para gerar recomendações mais precisas e baseadas em conteúdo textual.

4 Visualização das Recomendações

As recomendações geradas pelo sistema exibem:

- **Ranking** de similaridade (#1, #2, #3, ...);
- **Nome do jogo recomendado;**
- **Valor da similaridade** em formato percentual;
- No ambiente Colab, uma **tabela visual com gradiente de cor** (`style.background_gradient`) indica o nível de proximidade entre os jogos.

Essa forma de apresentação torna os resultados mais intuitivos, permitindo ao usuário identificar rapidamente quais jogos possuem características mais próximas do jogo pesquisado.

5 Estrutura do Projeto

Os principais componentes do projeto são:

- `main.py`: script principal que contém a lógica do sistema;
- `prepare_data()`: função responsável por limpar e lematizar o texto;

- `calculate_similarity_matrix()`: gera a matriz de similaridade TF-IDF;
- `get_similar_games()`: retorna os jogos mais parecidos com o título escolhido;
- `create_interface()`: cria a interface interativa para uso no Colab;
- `steam.csv`: dataset utilizado, contendo informações de cada jogo.

6 Como Executar o Projeto

Para executar o sistema, siga os passos abaixo:

1. Verifique se a versão do Python que está utilizando é a 3.10 ou superior.
2. Instale as dependências necessárias:

```
pip install pandas numpy scikit-learn nltk ipywidgets
```

3. Garanta que o arquivo `steam.csv` esteja no mesmo diretório do `main.py`.
4. Execute o programa:

```
python main.py
```

Em ambiente Colab, basta carregar o código e a base de dados para iniciar a interface gráfica.

7 Conclusão

O sistema de recomendação de jogos da Steam demonstra a aplicação prática de técnicas de PLN e Álgebra Linear para resolver um problema real: sugerir jogos similares de forma automatizada. Ao transformar descrições e tags em representações numéricas, o sistema é capaz de identificar padrões semânticos e fornecer recomendações relevantes.

Além de oferecer aprendizado prático sobre TF-IDF e similaridade do cosseno, o projeto também exemplifica como criar interfaces interativas com `ipywidgets` e integrar análises de texto em aplicações de recomendação.