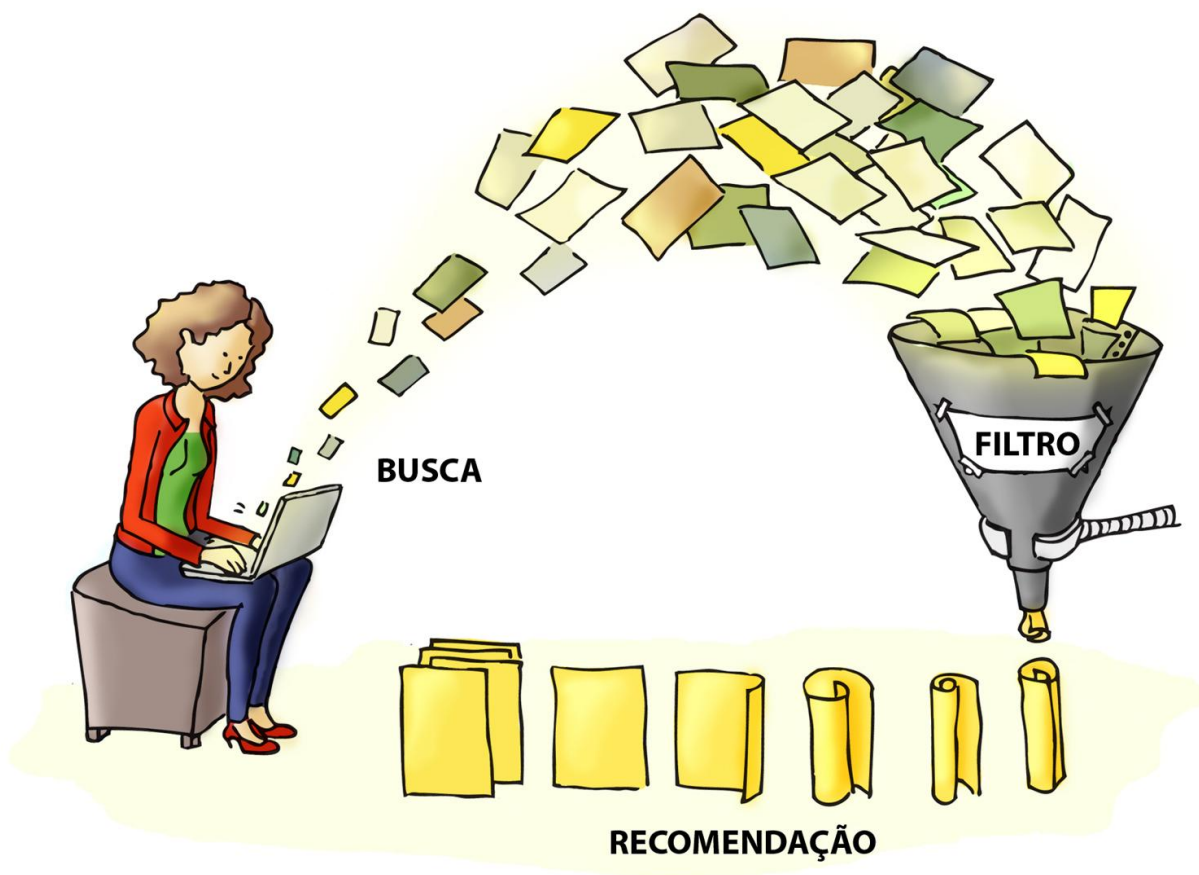


SISTEMA DE RECOMENDAÇÃO DE FILMES COM SVD

(Trabalho Final - Computação científica e análise de dados)



Riquelme Freitas Gomes

Universidade Federal do Rio de Janeiro (UFRJ)

INTRODUÇÃO

Nos últimos anos, o crescimento exponencial da quantidade de informações disponíveis na internet tem tornado cada vez mais desafiador para os usuários encontrarem conteúdos relevantes para seus interesses. Nesse contexto de sobrecarga e aparente infinidade de opções, os **sistemas de recomendação (SR)** mostram-se como ferramentas valiosas, capazes de filtrar a grande quantidade de informações e apresentar recomendações personalizadas, buscando tornar a experiência de cada usuário mais satisfatória e eficiente.

Assim, através da análise de dados, padrões de comportamento e preferências, esses sistemas conseguem identificar e prever quais itens são mais relevantes para cada usuário. Atualmente, os SR têm um papel fundamental em sites como Amazon, YouTube, Netflix, Walmart, Spotify e nas principais redes sociais, como Facebook, Twitter e Instagram. Mas como é definido esse tipo de sistema?

- “Sistemas de Recomendação são conjuntos de algoritmos capazes de analisar e identificar padrões de comportamento dos usuários de determinada plataforma, com objetivo de fornecer sugestões personalizadas, com base em seus interesses e histórico de navegação”. [5]

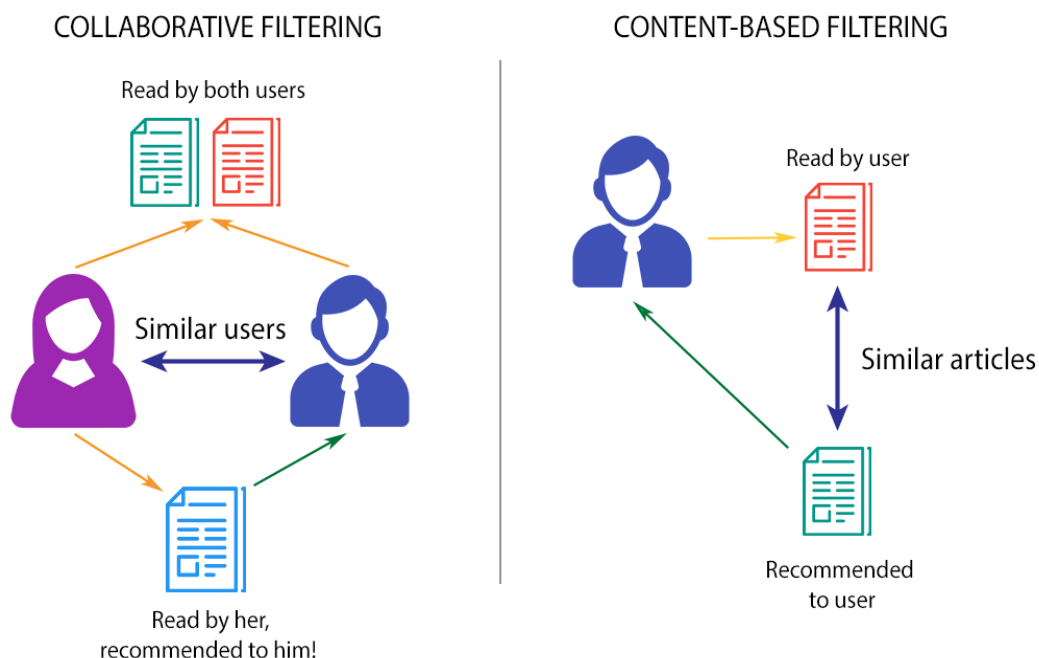
A partir disso, é simples inferir que os SR podem ser utilizados para recomendar diversos tipos de informações, conteúdos ou produtos (como páginas web, filmes, músicas, livros, medicamentos, lojas, artigos), além da

possibilidade de serem implementados com o uso de diferentes técnicas e ferramentas. No entanto, neste trabalho irei ter como foco principal o sistema de recomendação de **filmes**, utilizando o método da **Singular Value Decomposition (SVD)**, um método clássico da álgebra linear que está se tornando bem popular no campo da ciência de dados e do aprendizado de máquina, muito pela sua aplicação no desenvolvimento de sistemas de recomendação.

Portanto, em síntese, pretendo utilizar como entrada conjuntos de dados de filmes e avaliações desses filmes por usuários para, com a ajuda da SVD, obter como saída uma recomendação ordenada de filmes para o usuário.

METODOLOGIA

Inicialmente, é importante entendermos que existem duas abordagens principais para filtragem em um sistema de recomendação: **Filtragem colaborativa** e **Filtragem baseada em conteúdo**.



A **Filtragem Colaborativa** é um método de geração de recomendação que tenta prever o grau de interesse de um usuário em determinados produtos a partir de correlações entre as avaliações feitas por este e pelos outros usuários. Como avaliações refletem os gostos das pessoas, a hipótese subjacente é que pessoas que avaliaram um grande conjunto de produtos de maneira semelhante, pelo menos num futuro próximo, devem continuar avaliando de maneira semelhante novos produtos. Desta forma, se quisermos sugerir um produto a um usuário, devemos utilizar os dados de uma pessoa que venha avaliando ao longo do tempo de maneira semelhante a este cliente, ou até mesmo utilizar os dados de um produto que foi avaliado ao longo do tempo de modo similar a um produto já consumido pelo usuário. Note que neste método é considerada a experiência/avaliação de vários ou todos os usuários. [6]

Já a **Filtragem baseada em conteúdo** se ampara em atribuições de um conteúdo sem depender necessariamente de uma interação de outro usuário. Nesse caso, as indicações são baseadas em atributos dos itens recomendados, e o mecanismo cria uma espécie de perfil genérico de usuário que deve se interessar por temas semelhantes. Logo, essa filtragem é fundamentada na descrição ou em alguns dados fornecidos para aquele produto. O sistema encontra a semelhança entre os produtos com base em seu contexto ou descrição. O histórico anterior do usuário é levado em consideração para encontrar produtos semelhantes que o usuário possa gostar, logo considera-se muito a experiência do usuário alvo. [7]

Além dessas duas, existem mais algumas abordagens para construir

sistemas de recomendação, como sistemas de recomendação multicritério, sistemas de recomendação com reconhecimento de risco, sistemas de recomendação móvel e sistemas de recomendação híbridos (combinando filtragem colaborativa e filtragem baseada em conteúdo). No entanto, a que optei por utilizar neste trabalho foi a primeira citada, a **Filtragem Colaborativa**. Mais especificamente a Filtragem Colaborativa Item-Item, já que ao invés de considerarmos a similaridade entre usuários, estaremos considerando a similaridade entre os filmes (por meio das avaliações). Tal escolha se deve ao fato de ser possível implementá-la utilizando conjuntos de dados anteriores de filmes e avaliações disponíveis de forma pública, sendo que a qualidade dessa recomendação tende a melhorar com o aumento do número de avaliações e de usuários englobados no processo.

SINGULAR VALUE DECOMPOSITION (SVD)

A Decomposição em Valores Singulares é um método de fatoração de matrizes, que, de forma resumida, efetua a decomposição de uma matriz qualquer $A_{M \times N}$ em três outras matrizes:

$$A = U \Sigma V^T$$

The diagram illustrates the SVD equation $A = U \Sigma V^T$. Below the equation, three labels are connected to the matrices by lines: 'Orthogonal' is connected to U , 'Rectangular Diagonal' is connected to Σ , and 'Orthogonal' is connected to V^T .

Onde $U_{M \times M}$ é uma matriz singular esquerda ortogonal, $S_{M \times N}$ (ou Σ) é uma matriz retangular diagonal que possui os valores singulares de A em sua diagonal e $V_{N \times N}$ é uma matriz singular direita ortogonal que, nesse caso, será útil para nos permitir calcular a similaridade entre os filmes pelas avaliações. Com base nisso, a SVD reduz a dimensão da matriz de utilidade A , na medida que mapeia cada usuário e cada filme. Esse mapeamento facilita uma representação clara das relações entre usuários e filmes e é isso utilizado para implementar nosso sistema de recomendação. Logo, vendo pelo nosso contexto, a SVD realiza a fatoração da nossa estrutura de matriz (usuários \times filmes), onde os elementos dessa matriz são as avaliações dadas aos filmes pelos usuários. [10]

Devido ao escopo deste relatório e a grande variedade de interpretações e utilidades da Decomposição em Valores Singulares, não tenho como objetivo explicar, de modo mais aprofundado, o seu funcionamento e os cálculos que o código realiza para, a partir da decomposição mencionada, chegar nas recomendações. É válido citar apenas que a SVD decompõe nossa matriz de avaliações e, a partir disso, podemos calcular a “semelhança” entre os filmes usando o algoritmo de correlação da similaridade por cosseno e então, dado um filme, recomendamos os mais similares. Sendo assim, o que precisamos ter conhecimento é que, nessa situação, a SVD está sendo utilizado para implementar a técnica da filtragem colaborativa.

DADOS UTILIZADOS

Para fazermos as recomendações de filmes, utilizamos dados disponíveis no site GroupLens [1], um site de Pesquisa em Computação Social da Universidade de Minnesota. Mais especificamente dados do projeto MovieLens [2], que, além de ser um site que ajuda as pessoas a encontrar filmes para assistir, também conduz experimentos de campo online nas áreas de recomendação automatizada de conteúdo, interfaces de recomendação, recomendações e interfaces baseadas em marcação, bancos de dados mantidos por membros e design de interface de usuário inteligente. Utilizei o dataset: “MovieLens 1M Dataset” [3] (1 milhão de avaliações de aproximadamente 6.000 usuários em 4.000 filmes, lançado em 2/2003.), uma vez que não é pequeno mas também não é tão grande ao ponto de ser muito custoso manipulá-lo.

Tal conjunto de dados contém 2 arquivos que serão por mim utilizados: “movies.dat” e “ratings.dat”.

1. movies.dat

Contém dados no formato **MovieID::Title::Genres**. Onde MovieID é um código (número inteiro) identificador do filme. Title é o título do filme. Genres é o conjunto de gêneros do filme, selecionados entre 18 gêneros possíveis.

id_filme	titulo	genero
1	Toy Story (1995)	Animation Children's Comedy
2	Jumanji (1995)	Adventure Children's Fantasy
3	Grumpier Old Men (1995)	Comedy Romance
4	Waiting to Exhale (1995)	Comedy Drama
5	Father of the Bride Part II (1995)	Comedy
6	Heat (1995)	Action Crime Thriller
7	Sabrina (1995)	Comedy Romance
8	Tom and Huck (1995)	Adventure Children's
9	Sudden Death (1995)	Action
10	GoldenEye (1995)	Action Adventure Thriller

2. ratings.dat

Contém dados no formato **UserID::MovieID::Rating::Timestamp**. Onde UserID e MovieID são, respectivamente, o código do usuário e do filme. Rating é uma avaliação/nota dada ao filme pelo usuário feita em uma escala de 5 estrelas (de 1 a 5 e somente valores inteiros), sendo que cada usuário contribuiu com pelo menos 20 avaliações. Timestamp armazena a data da avaliação e é representado em segundos.

id_usuario	id_filme	nota	tempo
1	1193	5	978300760
1	661	3	978302109
1	914	3	978301968
1	3408	4	978300275
1	2355	5	978824291
1	1197	3	978302268
1	1287	5	978302039
1	2804	5	978300719
1	594	4	978302268
1	919	4	978301368

A partir desses dados e utilizando o código implementado no Google Colaboratory [4] em Python, conseguimos, uma vez escolhido o código de um filme, gerar recomendações de outros N filmes que, com base nas amostras, mais se assemelham ao filme escolhido.

RESULTADOS

O sistema e as recomendações foram feitos na linguagem de programação Python pelo Google Colab [4] com as bibliotecas “Numpy” e

“Pandas”. Portanto, para utilizar basta baixar e fazer a leitura dos arquivos mencionados e rodar todas as células do código.

Buscando checar o funcionamento do programa, vamos sortear, de forma aleatória, um filme disponível no dataset e pedir a recomendação de $N = 10$ filmes similares a ele. Saída:

```
Recomendações para: Alien: Resurrection (1997)
```

```
0º: Alien: Resurrection (1997)
1º: Star Trek VI: The Undiscovered Country (1991)
2º: Star Trek: Insurrection (1998)
3º: Sphere (1998)
4º: Star Trek: First Contact (1996)
5º: Event Horizon (1997)
6º: Blade (1998)
7º: Starship Troopers (1997)
8º: Deep Rising (1998)
9º: Lost in Space (1998)
```

O filme definido foi “Alien: Resurrection (1997)”, que possui ação, terror e ficção científica como gêneros e aborda uma temática espacial. Podemos observar que o primeiro filme a ser sugerido nas recomendações (com índice 0) foi o próprio, isso mostra que as recomendações estão fazendo certo sentido já que um filme é sempre o mais similar possível a si mesmo. Além disso, todos os outros filmes recomendados têm pelo menos um gênero em comum com o filme sorteado (geralmente ação ou ficção científica), além de boa parte apresentar também a temática espacial

citada. Isto nos sugere que a recomendação não foi feita de forma aleatória e que o cálculo das similaridades aparenta estar funcionando. Mas será que os gêneros dos filmes recomendados serão sempre parecidos? Vamos repetir o procedimento sorteando outro filme, porém agora explicitando abaixo os gêneros de cada filme para facilitar a análise.

Saída:

```
Recomendações para: What Happened Was... (1994)
{'Comedy|Drama|Romance'}
```

```
0º: What Happened Was... (1994)
{'Comedy|Drama|Romance'}
```

```
1º: Stuart Saves His Family (1995)
{'Comedy'}
```

```
2º: Abyss, The (1989)
{'Action|Adventure|Sci-Fi|Thriller'}
```

```
3º: American in Paris, An (1951)
{'Musical|Romance'}
```

```
4º: Excalibur (1981)
{'Action|Drama|Fantasy|Romance'}
```

```
5º: Quiet Man, The (1952)
{'Comedy|Romance'}
```

```
6º: Big Squeeze, The (1996)
{'Comedy|Drama'}
```

```
7º: Supercop (1992)
{'Action|Thriller'}
```

```
8º: For Whom the Bell Tolls (1943)
{'Adventure|War'}
```

```
9º: Sgt. Bilko (1996)
{'Comedy'}
```

O filme escolhido foi “What happened was...(1994)” dos gêneros “comédia”, “drama” e “romance”. Podemos notar que nem todos os filmes sugeridos na lista possuem algum desses gêneros. Isso faz sentido pois nada impede que, nas amostras, parte significativa dos usuários tenham avaliado de forma parecida filmes de gêneros completamente diferentes, além do fato de que um mesmo usuário não avalia todos os filmes mas apenas alguns. Por consequência, esse aspecto pode ser refletido nas recomendações e isso não é necessariamente um problema pois os filmes podem estar relacionados no gosto dos usuários de outras formas que não por seus gêneros.

VANTAGENS E DESVANTAGENS OBSERVADAS

Como já mencionado, sistemas de recomendação de filme podem ser feitos utilizando várias abordagens, técnicas e métodos diferentes. No entanto, como meu trabalho se restringiu à filtragem colaborativa item-item com o uso da SVD, pude observar mais diretamente os principais problemas e benefícios desse formato.

Dentre as principais desvantagens, vale ressaltar que a Filtragem Colaborativa é um método computacionalmente caro, pois a matriz de avaliações deve ser gerada e manipulada, e para que as recomendações sejam satisfatórias é necessária uma grande quantidade de dados. Isso aliado ao fato da matriz ser esparsa (cada usuário avalia apenas alguns filmes) fazem com que o método tenha alta complexidade computacional.

Outra desvantagem é que o método requer que os filmes sejam avaliados para entrar no cálculo da similaridade, o que gera o problema de “Cold Start”. Ou seja, no caso de novos filmes, o sistema não irá recomendá-los até que eles possuam uma quantidade considerável de avaliações.

Quanto às vantagens, cabe destacar que, como são utilizadas avaliações de outros usuários, não há problema em fazer recomendações para usuários novos. Diferente de outros procedimentos, não temos a exigência de analisar o comportamento e o feedback dos novos consumidores para podermos recomendá-los filmes. Além disso, as recomendações não se restringem a serem parecidas com o filme consumido e, por isso, refletem maior diversidade em suas características. Por fim, por fazermos uso da Singular Value Decomposition (SVD), conseguimos amenizar o problema da esparsidade reduzindo a dimensionalidade da matriz inicial com a decomposição.

CONCLUSÃO

É evidente, portanto, que os sistemas de recomendação desempenham um papel crucial na evolução e otimização da experiência do usuário em diversas áreas da sociedade atual. No geral, acredito que o trabalho apresentado me proporcionou uma consistente compreensão da construção e funcionamento de sistemas de recomendação, principalmente quando com o uso de filtragem colaborativa e SVD. Com ele, pudemos reduzir a dimensionalidade do conjunto de dados e extrair informações valiosas, tornando o processo de recomendação mais rápido e eficiente. Isso nos permitiu identificar padrões ocultos e associações relevantes entre usuários e filmes, proporcionando sugestões de filmes mais alinhadas com as avaliações dos usuários.

No entanto, é importante destacar que, como em qualquer sistema de recomendação, existem desafios e limitações a serem considerados. Em nosso contexto, por exemplo, a escassez de dados nos leva a recomendações menos precisas. Por isso, a melhor, e mais utilizada, solução, é a busca pelo emprego de métodos híbridos, que conseguem ser mais flexíveis e adaptáveis ao contexto, permitindo aproveitar mais as vantagens de cada tipo de sistema de recomendação.

REFERÊNCIAS

- [1] GroupLens. <https://grouplens.org>
- [2] MovieLens. <https://movielens.org>
- [3] MovieLens 1M Dataset. <https://grouplens.org/datasets/movielens/1m/>
- [4] Sistema de recomendação de filmes implementado no Google Colaboratory. <https://colab.research.google.com>
- [5] Sistemas de recomendação: como personalizar a experiência de seus usuários? <https://blog.xpeducacao.com.br/sistemas-de-recomendacao/>
- [6] Sistemas de Recomendação. <https://sistemascolaborativos.uniriotec.br>
- [7] Como funcionam os sistemas de recomendação?
<https://tecnoblog.net/como-funcionam-os-sistemas-de-recomendacao/>
- [8] Brief on Recommender System. <https://towardsdatascience.com>
- [9] Singular Value Decomposition (SVD) & Its Application In Recommender System. <https://analyticsindiamag.com/singular-value-decomposition-svd/>
- [10] SVD Visualized, Singular Value Decomposition explained.
<https://www.youtube.com/watch?v=vSczTbgc8Rc>
- [11] Simple Movie Recommender Using SVD.
<https://alyssaq.github.io/2015/simple-movie-recommender-using-svd/>