

# Sistemas de Recomendação com Inteligência Artificial

Rafael Barbolo Lopes

Escola Politécnica da Universidade de São Paulo

[rafael.barbolo@poli.usp.br](mailto:rafael.barbolo@poli.usp.br)

<http://www.poli.usp.br>

**Resumo.** Um sistema de recomendação pode ser implementado através de técnicas de Inteligência Artificial, misturando busca, filtragem, mineração de dados, entre outras. Os dois modelos de sistema de recomendação mais populares são baseados em filtragem colaborativa e análise de conteúdo. Este artigo explica, com exemplos, como um sistema de recomendação baseado nessas técnicas pode ser construído.

**Palavras-chave:** inteligência artificial, recomendação, filtragem colaborativa, análise de conteúdo, similaridade

## 1 Introdução

Em um mundo onde somos bombardeados diariamente com milhares de informações disponibilizadas em vídeos, músicas, livros, notícias, imagens, literatura e artigos científicos, programas de TV e páginas de Internet, sistemas de recomendação surgem para influenciar nossas escolhas de maneira ágil e assertiva, permitindo que nosso tempo seja melhor distribuído em outras atividades.

Um sistema de recomendação tem como função sugerir itens de interesse a um usuário. Saber o que sugerir é de importância crucial para empresas que oferecem produtos e serviços. Ao realizar uma sugestão acertada, a possibilidade de conversão de venda é maior e o usuário se beneficia gastando menos tempo procurando o que precisa. Segundo dados levantados por (Lamere & Celma, 2007), 2/3 dos aluguéis na Netflix são de filmes recomendados, 38% de cliques a mais são gerados no Google News através de recomendação de notícias, e 35% das vendas da Amazon são provenientes de recomendação de compras.

Através de técnicas inteligentes para descobrir similaridades entre diferentes itens, um sistema de recomendação é construído. A similaridade entre dois itens é, frequentemente, definida com base no conteúdo de cada item analisado ou com base nas ações de usuários sobre cada um dos itens.

Este artigo descreve técnicas que podem ser utilizadas na construção de um sistema de recomendação. Como exemplo de implementação, é apresentado um sistema de recomendação de filmes.

## 2 Sistemas de recomendação

Para entender um algoritmo de recomendação, é preciso ter claro o conceito de similaridade. A similaridade é uma heurística que determina quão próximos itens são uns dos outros ao realizar recomendações. Por exemplo, uma heurística simples para determinar a similaridade entre dois livros poderia ser o tamanho do conjunto intersecção entre os autores dos dois livros. Desta forma, livros escritos pelo mesmo autor seriam considerados similares e poderiam ser sugeridos para um leitor que gostou de ler um livro do autor.

Uma outra forma de pensar em similaridade é através do conceito de distância. A similaridade está inversamente relacionada com a distância: dados dois itens, quanto mais similares forem, menos distantes estão; por outro lado, quanto menos similares forem, mais distantes estão. Portanto, uma boa maneira de encontrar funções que determinam similaridades entre itens é procurando funções que calculam suas distâncias. Um exemplo simples de como usar a distância para determinar a similaridade entre itens é visto analisando a recomendação de amigos em uma rede social: um fator que determina a similaridade entre usuários é a distância geográfica entre eles, de forma que quanto menos distantes dois usuários estão, maior é a chance da recomendação de amizade ser feita com sucesso.

Existem duas categorias principais de algoritmos de recomendação. A primeira, filtragem colaborativa, consiste em analisar as interações entre os usuários do sistema e os itens fornecidos pelo sistema. Como exemplos de interação, pode-se pensar em avaliações ou comentários que usuários fazem sobre itens. A segunda, análise de conteúdo, consiste em entender os perfis dos usuários e dos itens que compõem o sistema, e construir recomendações com base em perfis de itens que devem agradar perfis de usuários.

### 2.1 Filtragem colaborativa (*collaborative filtering*)

A filtragem colaborativa é uma técnica de recomendação baseada na seguinte premissa: se dois usuários possuem os mesmos interesses nos itens oferecidos pelo sistema, então os mesmos itens deveriam ser indicados para os dois. Esses interesses podem ser encontrados observando as interações dos usuários com o sistema, por exemplo, através de avaliações e comentários feitos em itens. Tomando como exemplo uma rádio de músicas online, se dois usuários avaliam e fazem marcações similares sobre as mesmas músicas, então quando um desses usuários avalia positivamente uma música nova, esta música deve ser recomendada para o outro usuário.

As recomendações devem ser atualizadas constantemente e por isso o algoritmo sofre com problemas de desempenho em grandes sistemas. Analisando o site last.fm, que é a maior rede online de músicas do mundo, seu grande trunfo é o sistema de recomendações de músicas chamado Audioscrobbler. Esse sistema é baseado em filtragem colaborativa e é capaz de recomendar novas músicas com base nas músicas que um usuário ouviu ou avaliou. Um dos grandes problemas

resolvidos pelo sistema foi a escalabilidade, pois existem mais de 40 milhões de usuários cadastrados no site que recebem frequentemente novas recomendações.

A seguir são apresentados a heurística de similaridade, e os modelos de recomendação com filtragem colaborativa baseados em usuários similares e em itens similares.

### Heurística de similaridade

Para sistemas de recomendação baseados em filtragem colaborativa, uma boa heurística para determinar a similaridade entre dois usuários é identificar quão próximas são as suas avaliações, ou seja, quanto mais itens os dois usuários avaliam da mesma maneira (com a mesma nota, por exemplo), mais similares os usuários são. Pensando analogamente para determinar a heurística de similaridade entre os itens do sistema, quanto mais avaliações similares dois itens recebem dos mesmos usuários, mais eles são similares.

Para exemplificar o cálculo de similaridade, deve-se supor que o sistema analisado é uma rádio de músicas online. Para calcular a similaridade entre dois usuários, A e B, que avaliaram as músicas x e y de acordo com a tabela 1, pode-se calcular o cosseno entre o ângulo formado pelos vetores de suas avaliações, de acordo com a fórmula 1, e obter os resultados observados na equação 2. A similaridade obtida é um valor entre 0 e 1, sendo que dois itens similares devem ter valor de similaridade próximo de 1.

$$sim(i, j) = \cos(\mathbf{i}, \mathbf{j}) = \frac{\mathbf{i} \cdot \mathbf{j}}{\|\mathbf{i}\|_2 * \|\mathbf{j}\|_2} \quad (1)$$

$$sim(A, B) = \cos((3, 2), (5, 1)) = (3*5+2*1)/(\sqrt{3^2 + 2^2} * \sqrt{5^2 + 1^2}) = 0.925 \quad (2)$$

De forma análoga, as similaridades entre duas músicas pode ser determinada calculando-se o cosseno entre os ângulos formados pelos seus vetores de avaliações.

### Recomendações baseadas em usuários similares

A recomendação baseada em usuários similares consiste em estimar a avaliação que um usuário faria sobre um item olhando para as avaliações que usuários similares (também chamados de vizinhos ou amigos) fizeram para esse item. Uma boa heurística para estimar a avaliação de um usuário sobre um item é realizar a soma ponderada da multiplicação das avaliações de cada vizinho pelo peso de sua similaridade com o usuário. Quanto maior for a avaliação estimada para um item, maior é a chance da recomendação ser feita com sucesso para o usuário analisado.

No exemplo da rádio de músicas online, suponha que existam três usuários, A, B e C, e duas músicas, x e y. Suponha que as avaliações de cada usuário sobre cada música sejam as apresentadas na tabela 1, e que as similaridades entre um usuário D e os usuários A, B e C seja fornecida pela tabela 2.

Usuário	Música	Avaliação
A	x	3
A	y	2
B	x	5
B	y	1
C	x	0
C	y	4

**Tabela 1.** Avaliações dos usuários A, B e C sobre as músicas x e y

Usuário	Similaridade com D
A	0.52
B	0.93
C	0.12

**Tabela 2.** Similaridade dos usuários A, B e C medidas para o usuário D

Pode-se calcular as estimativas de avaliações do usuário D sobre as músicas x e y como mostrado na tabela 3.

Música	Estimativa de avaliação
x	$(0.52 * 4 + 0.93 * 5 + 0.12 * 0) / (0.52 + 0.93 + 0.12) = 4.29$
y	$(0.52 * 2 + 0.93 * 1 + 0.12 * 4) / (0.52 + 0.93 + 0.12) = 1.56$

**Tabela 3.** Estimativas de avaliações do usuário D sobre as músicas x e y

A partir das estimativas de avaliações do usuário D sobre as músicas x e y, pode-se recomendar ao usuário D a música x com alta chance de sucesso.

### Recomendações baseadas em itens similares

A filtragem colaborativa baseada em itens similares funciona de maneira muito parecida com a baseada em usuários similares, exceto que a similaridade entre usuários é substituída pela similaridade entre itens. Para entender como é o funcionamento desse algoritmo, o exemplo da rádio de músicas online é usado novamente. Suponha que deseja-se realizar recomendações de músicas para o usuário A.

Assuma que o usuário A avaliou as músicas x e y de acordo com a tabela 4 e que as similaridades entre as músicas x, y, z e w sejam dadas pela tabela 5.

Música	Avaliação do usuário A
x	4
y	2

**Tabela 4.** Avaliações do usuário A sobre as músicas x e y

Música 1	Música 2	Similaridade entre músicas 1 e 2
x	y	0.59
x	z	0.32
x	w	0.98
y	z	0.22
y	w	0.09
z	w	0.02

**Tabela 5.** Similaridades entre músicas x, y, z e w (relação de similaridade é simétrica)

Com base nas informações apresentadas, pode-se calcular as estimativas de avaliação para as músicas z e w, como mostrado na tabela 6.

Música	Estimativa de avaliação
z	$(0.32 * 4 + 0.22 * 2) / (0.78 + 0.22) = 1.72$
w	$(0.98 * 4 + 0.09 * 2) / (0.98 + 0.09) = 3.83$

**Tabela 6.** Estimativas de avaliações do usuário A sobre as músicas z e w

A partir das estimativas de avaliações do usuário A sobre as músicas z e w, pode-se recomendar ao usuário A a música w com boa chance de sucesso.

## 2.2 Recomendações baseadas em análise de conteúdo

Recomendações baseadas em análise de conteúdo são provenientes de similaridades de conteúdos entre usuários, entre itens ou entre usuários e itens. Deve-se conhecer as estruturas que definem um usuário ou um item, e prever itens com perfis que se encaixam ao perfil de usuário que receberá a recomendação.

Diferentemente da filtragem colaborativa, uma recomendação estritamente baseada em análise de conteúdo não utiliza informações de interações entre usuários e itens do sistema. Uma boa heurística para definir a similaridade entre dois itens é a média das similaridades dos atributos que compõem cada item.

Como exemplo de como uma recomendação baseada em análise de conteúdo é utilizada, pode-se analisar um jogo de vídeo game que avalia os perfis dos personagens que podem ser escolhidos para jogar. Em um jogo fictício de futebol, um personagem pode ter 5 atributos: força, velocidade, inteligência, precisão no chute e resistência cardiovascular, sendo que cada atributo recebe um valor entre 0 e 100. Como indicado na figura 1, o perfil de um personagem pode ser mapeado em um gráfico de radar.

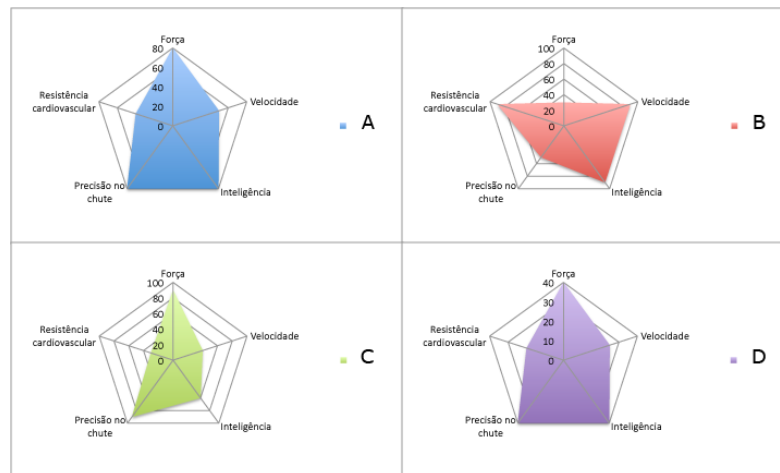


**Fig. 1.** Perfil de um jogador visto em um gráfico de radar

Para jogadores A, B, C e D com atributos apresentados na tabela 7, os gráficos indicados na figura 2 são gerados.

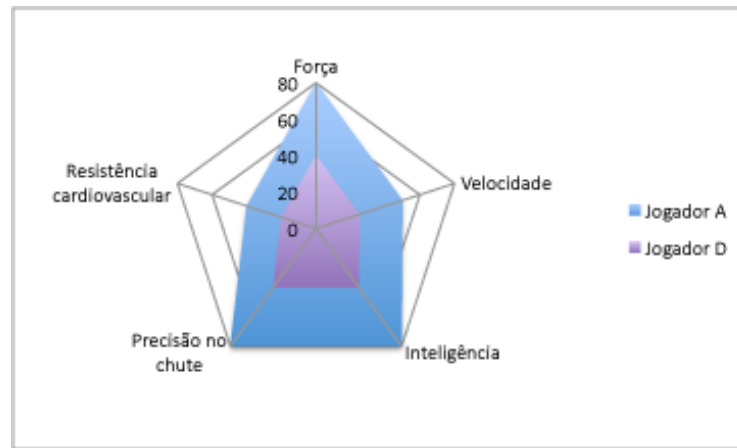
Jogador	Força	Velocidade	Inteligência	Precisão no chute	Resistência cardiovascular
A	80	50	80	80	40
B	30	90	90	50	90
C	90	40	60	90	30
D	40	25	40	40	20

**Tabela 7.** Atributos dos jogadores A, B, C e D



**Fig. 2.** Perfis dos jogadores A, B, C e D visualizados em gráficos de radar

Procurando por gráficos com curvas de formatos semelhantes, encontram-se jogadores que possuem perfis de habilidades similares. Por exemplo, os jogadores A e D possuem habilidades equivalentes, porém em intensidades diferentes, como pode ser visualizado na figura 3. Dependendo da heurística de similaridade escolhida, esses dois jogadores podem ser considerados similares.



**Fig. 3.** Comparação de atributos entre jogadores A e D

Conhecendo jogadores similares, o jogo de futebol pode realizar substituições de jogadores de maneira autônoma durante partidas. Esse tipo de recomendação pode ser usada diretamente pelo sistema em uma partida em que o computador joga, ou então pode ser enviada para um usuário que fará substituição de jogador durante o jogo.

### 3 Exemplo: sistema de recomendação de filmes

Um sistema de recomendação de filmes foi implementado para ilustrar o uso dos conceitos apresentados neste artigo. Esse sistema é híbrido, usando filtragem colaborativa e análise de conteúdo para encontrar similaridades entre filmes. Com as similaridades calculadas, o sistema recomenda filmes a partir de um filme escolhido pelo usuário.

O sistema foi desenvolvido utilizando as linguagens de programação Ruby e C, e o banco de dados MySQL. Foi utilizada uma base pública com cerca de 12 mil filmes, 70 mil usuários e 10 milhões de avaliações, disponibilizada pelo (GroupLens Research, 2010).

A filtragem colaborativa é usada para calcular a similaridade entre dois filmes com base nas avaliações feitas pelos usuários. É calculado o cosseno do ângulo formado pelos vetores de avaliações dos dois filmes comparados. Quanto maior

o cosseno, mais similares os dois filmes são. Colocando de forma mais simples, se dois filmes recebem as mesmas avaliações pelos mesmos usuários, então eles são considerados similares e por isso são boas sugestões de recomendação entre si.

A análise de conteúdo é aplicada para calcular a similaridade entre dois filmes com base nas categorias em que os filmes se encontram. Um filme pode estar localizado em mais de uma categoria, por exemplo, um desenho que também é uma comédia. A heurística adotada calcula o tamanho do conjunto complemento da intersecção das categorias dos filmes; em outras palavras, quanto mais categorias diferentes os dois filmes possuírem, mais distantes e menos similares eles serão considerados.

Após utilizar os métodos descritos para calcular a similaridade entre todos os filmes, a similaridade final entre dois filmes é calculada como a média ponderada entre a similaridade obtida pela filtragem colaborativa que tem peso 9 e a similaridade obtida pela análise de conteúdo que tem peso 1. Por fim, é realizada uma normalização para o intervalo  $[0,10]$  das similaridades calculadas, de forma que o pior resultado de similaridade recebe nota 0 e o melhor recebe nota 10.

Para encontrar os filmes similares a um outro, é realizada uma busca pelos filmes que possuem as maiores similaridades com o filme analisado. O código fonte completo desse exemplo pode ser encontrado em (Rafael Barbolo Lopes, 2010).

## 4 Conclusões

Sistemas de recomendação envolvem diferentes áreas de IA, como por exemplo busca, filtragem, mineração de dados, processamento de textos, interações entre usuários, entre outras. Além disso, o desenvolvimento de sistemas dessa natureza tem alto impacto em vendas de serviços e produtos.

Um grande problema enfrentado pelos sistemas atuais é de escalabilidade. Em sistemas com milhões de usuários e de itens, e com centenas de atributos sendo avaliados para cada item e usuário, é quase impossível realizar cálculos de similaridades entre os itens em tempo real. Ou seja, os usuários devem receber recomendações com certo atraso.

Este artigo apresentou as técnicas de recomendação baseadas em filtragem colaborativa e em análise de conteúdo, que são as mais usadas na área. Como exemplo prático da importância desse tipo de sistema, apresentou a implementação de um sistema de recomendação de filmes. Com a teoria disposta, é possível iniciar a criação de um sistema de recomendação para qualquer tipo de aplicação.

## Referências

1. S. Russel and P. Norvig. Artificial Intelligence: A Modern Approach. Prentice Hall, Upper Saddle River, USA. 2nd. Edition, (2003).
2. H. Marmanis and D. Babenko. Algorithms of the Intelligent Web. Manning Publications, 1st. Edition, (2009).



3. B. Sarwar. Item-Based Collaborative Filtering Recommendation Algorithms. Proc. In: INTERNATIONAL WORLD WIDE WEB CONFERENCE, 2001, Hong Kong. Proceedings of the 10th international conference on World Wide Web. New York: ACM Press, p. 285-295, (2001).
4. P. Lamere and O. Celma. MUSIC RECOMMENDATION TUTORIAL ISMIR 2007. TU Wien - Freihaus, Wiedner Hauptstrasse 8-10, 1040 Wien, Room HS6, (2007).
5. M. Deshpande and G. Karypis. Item-based top-n recommendation algorithms. ACM Trans. Inf. Syst., (2004).
6. <http://www.grouplens.org/node/73>. GroupLens Research - MovieLens Data Sets, (2010).
7. <https://github.com/barbolo/Movie-Recommender> - Movie Recommender - Código fonte do projeto de recomendação de filmes criado por Rafael Barbolo Lopes, (2010).