

Como funciona parte 2

Transcrição

Qual nota acha que Guilherme deu para o livro Trem da Alegria? Achamos que seria a nota 3, sabe por quê? Observe a tabela novamente:

--	Metamorfose	Mochileiro	Trem	Xuxa
Guilherme	9	8	?	3
Joana	2		8	9
João				9
Daniela	9		3	

De forma simples, nosso pensamento foi o seguinte: Guilherme deu nota 9 para Metamorfose e Joana deu nota 2, o que demonstra que Guilherme e Joana são bem diferentes visto que suas notas para o mesmo livro atestam opiniões bem reversas. Por outro lado, Daniela deu a mesma nota que Guilherme para o Metamorfose. Baseado nisso e sabendo que a Daniela deu nota 3 para o Trem da Alegria podemos chutar pela *similaridade* que Guilherme dará nota perto de 3 para o Trem da Alegria.

--	Metamorfose	Mochileiro	Trem	Xuxa
Guilherme	9	8	3	3
Joana	2		8	9
João				9
Daniela	9		3	

Claro que estamos simplificando bastante o raciocínio, pois em matemática é preciso definir claramente o que é similaridade, igualdade, desigualdade e também escrever fórmulas para garantir que em todos os casos estaremos seguindo uma mesma lógica, que é a base do algoritmo.

Lembrando que o resultado desse raciocínio não é exato, mas sim uma estimativa. Em recomendações para restaurantes, por exemplo, precisa-se conhecer os gostos de cada uma das pessoas para que com base nisso possamos escolher um restaurante que seja bom para todos. O mesmo seria para balada e outros.

Seguindo com essa lógica, quais notas o João daria para os livros Metamorfose e Trem da Alegria? Vamos observar mais uma vez a tabela:

--	Metamorfose	Mochileiro	Trem	Xuxa
Guilherme	9	8	3	3
Joana	2		8	9
João	?		?	9
Daniela	9		3	

Se você aplicar o mesmo raciocínio chegará as notas 2 e 8, respectivamente. Por quê? Porque são as mesmas notas da Joana e eles são parecidos, uma vez que deram a mesma nota para o livro da Biografia da Xuxa.

--	Metamorfose	Mochileiro	Trem	Xuxa
Guilherme	9	8	3	3
Joana	2		8	9
João	2		8	9
Daniela	9		3	

Seguindo esses passos, chegaremos a um caso misterioso. Vejamos:

--	Metamorfose	Mochileiro	Trem	Xuxa
Guilherme	9	8	3	3
Joana	2	?	8	9
João	2	?	8	9
Daniela	9	8	3	3

Pelo raciocínio de similaridade que elaboramos, completamos quase toda a tabela, porém, temos o caso de Joana e o João, que são similares, mas não avaliaram o livro Mochileiro. O que fazer neste caso? Buscamos um segundo elemento similar!

O segundo similar, neste caso, seria alguém que também leu os outros títulos e avaliou - sem ser processado pelas nossas estimativas. Ao observar chegamos no Guilherme que encaixa-se neste caso, ele avaliou o livro da Xuxa com nota 3 e, por associação, assimilamos que Joana e João avaliarão o Mochileiro com a nota 8.

Lembrando que apenas a avaliação da Joana se baseou no Guilherme, a do João foi um segundo passo, baseado na Joana.

Perceba, a relação de similaridade mudou? Buscar segundos ou terceiros similares exige que a regra seja adaptada. Antes avaliamos apenas a nota do livro, agora estamos levando em consideração o fato de duas pessoas terem lido o mesmo livro, independente da nota, pois neste caso eram notas não muito parecidas.

Quais os problemas da abordagem que utilizamos? Podemos listar pelo menos uma vantagem e uma desvantagem: - Vantagem: De alguma forma estamos estimando uma recomendação. - Desvantagem: Podemos estar subestimando ou

superestimando essa recomendação.

Outra possível abordagem para resolver esse problema seria utilizar médias de notas entre alguns livros ou entre todos os livros. Porém, também encontraríamos as mesmas vantagens e desvantagens que listamos antes.

Estes casos são interessantes, mas não serão resolvidos por nós neste curso, pois o algoritmo já implementa regras que cuida desses casos de forma mais genérica.

É importante frisar que os erros são normais e que não precisam ser levados com tanta consideração. O mais importante é prestar atenção na taxa de erro! Caso os sistemas de recomendações errem muito, temos um problema, por exemplo, a alta taxa de erro pode levar a Netflix a perder diversos usuários.

No caso da situação levantada na questão dos livros usamos um algoritmo que se baseia na similaridade dos usuários, que em inglês chama-se: **User Based Similarity**.