

SISTEMAS DE RECOMENDAÇÃO

Prof. Dr. Sílvio César Cazella

Sumário

- Filtragem Colaborativa (FC)
 - Cálculo da Similaridade
 - Cálculo da Predição
 - Limitações da FC

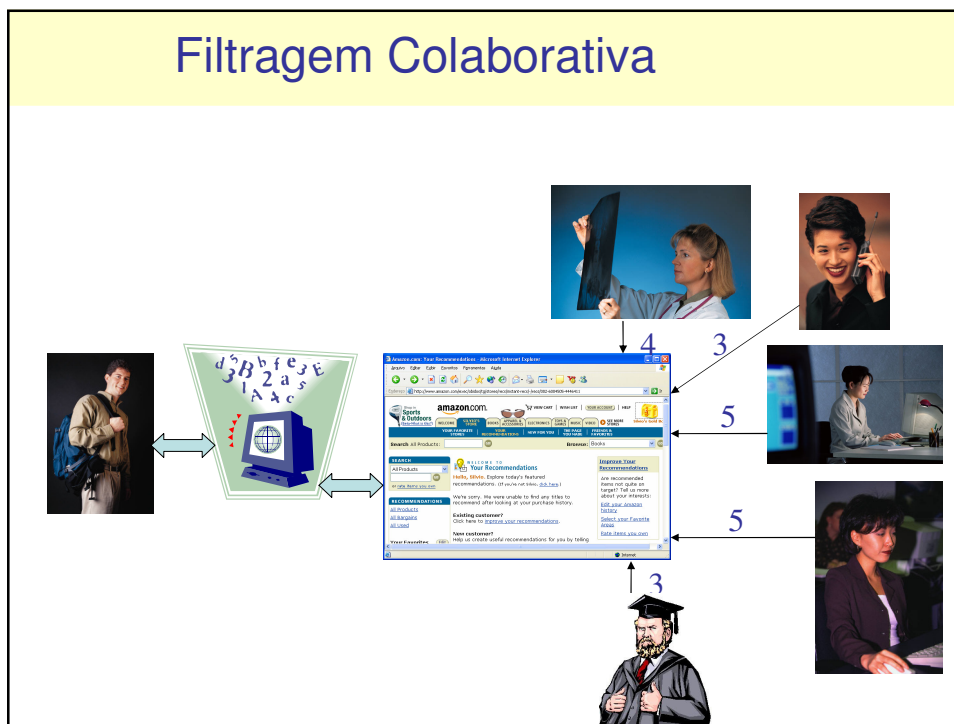
Filtragem Colaborativa

- A Filtragem Colaborativa não se constitui em algo novo e apresenta uma proposta diferenciada dos sistemas de filtragem uma vez que não requerem compreensão ou conhecimento de conteúdo dos itens.
- Nos sistemas colaborativos a essência está na troca de experiências entre as pessoas que possuem interesses comuns. Nestes sistemas os itens são filtrados baseado na avaliação feita pelos usuários daqueles itens, ao invés do conteúdo dos itens.
- O primeiro sistema criado com esta abordagem foi o Tapestry que permitia ao usuário especificar uma consulta como: "mostre-me todos os memorandos que uma determinada pessoa considera como importante". A idéia neste tipo de abordagem é a de que membros de uma dada comunidade podem ser beneficiados pela experiência de outros antes de decidir por consumir uma dada informação.

Filtragem Colaborativa

- Sistemas de Recomendação Colaborativa procuram prever a pontuação de um item para um consumidor em particular baseado em como outros usuários com gostos semelhantes ao usuário alvo previamente pontuaram o mesmo item.
- Mais formalmente, a pontuação $R(u,i)$ do item i para o usuário u é estimada baseada na pontuação $R(u',i)$ atribuída ao mesmo item i pelos usuários u' que são similares ao usuário em questão.
- O valor de uma pontuação desconhecida $r_{u,i}$ para o usuário u e item i é normalmente computado como uma agregação de pontuações de outros usuários (por exemplo, os N mais similares) para o mesmo item i .

Filtragem Colaborativa



Filtragem Colaborativa

- É um tipo de solução que libera o usuário do esforço maior de ter que procurar por itens de seu interesse.
- Baseada na busca de opiniões similares.
- A FC segue basicamente três etapas:
 1. Calcula-se o peso em relação a similaridade do usuário-alvo;
 2. Seleciona-se um subconjunto de usuários com maiores similaridades;
 3. Calcula-se as predições (Recomendações).

Filtragem Colaborativa

- Supondo a existência de um sistema de recomendação onde os usuários tem acesso a itens e podem avaliá-los quanto a satisfação. Para este *feedback* uma escala *Likert* de 5 pontos é disponibilizada (5 significa adorei e 1 significa detestei o item).
- Uma matriz de relação usuários x itens pode ser disponibilizada.

Filtragem Colaborativa

Usuário/ Item	A	B	C	D	E
U1	3		4	3	2
U2	4	5	4	4	2
U3	5	1	1	5	4
U4	4	5	3	3	3
U5	1	3		2	

Esta matriz nos apresenta a relação usuário x item, onde 5 usuários pontuaram 5 itens (A, B, C, D e E). A esparsidade desta matriz seria de 8%, ou seja, apenas 8% das avaliações possíveis não foram feitas. Estes itens não avaliados são potenciais recomendações.

Filtragem Colaborativa

A primeira etapa do FC:

- Calcula-se o peso em relação a similaridade do usuário-alvo: para tanto utiliza-se uma técnica denominada de “*k-nearest-neighbor*” ou “*user-based*”. A definição da similaridade pode ser realizada através de diversos coeficientes, sendo mais comumente aplicado o coeficiente de correlação de Pearson. Observa-se a necessidade de mais de uma avaliação em comum para que a correlação seja viável, e os resultados variam entre 1 para similaridade total, e -1 para total dissimilaridade.

Filtragem Colaborativa

•Cálculo da similaridade da FC utilizando *Pearson*:
assumindo na correlação $U1 = a$ e $U2 = b$, teríamos:

$$corr_{ab} = \frac{\sum_i (r_{ai} - \bar{r}_a)(r_{bi} - \bar{r}_b)}{\sqrt{\sum_i (r_{ai} - \bar{r}_a)^2 \sum_i (r_{bi} - \bar{r}_b)^2}}$$

$\bar{r}_a =$ Média das avaliações de *a* em comum com *b*

$\bar{r}_b =$ Média das avaliações de *b* em comum com *a*

$r_{ai} =$ Avaliação que *a* deu para o item *i*

$r_{bi} =$ Avaliação que *b* deu para o item *i*

Usuário/ Item	A	B	C	D	E
U1	3	?	4	3	2
U2	4	5	4	4	2
U3	5	1	1	5	4
U4	4	5	3	3	3
U5	1	3		2	

$\bar{r}_a = 3,0$

$\bar{r}_b = 3,5$

Usuário	Pearson (U1)
U2	1
U3	0,53
U4	0
U5	0

Alta similaridade

Média similaridade

Nada pode ser dito

Nada pode ser dito

$$corr_{ab} = \frac{(3-3)(4-3,5)+(4-3)(4-3,5)+(3-3)(4-3,5)+(2-3)(2-3,5)}{\sqrt{((3-3)^2+(4-3)^2+(3-3)^2+(2-3)^2)*((4-3,5)^2+(4-3,5)^2+(4-3,5)^2+(2-3,5)^2)}} = 1$$

Filtragem Colaborativa

A segunda etapa do FC:

- Seleciona-se um subconjunto de usuários com maiores similaridades :com base no resultado do coeficiente de Pearson defino este subconjunto (com foco no usuário alvo da recomendação – U1).

Usuário	Pearson (U1)
U2	1
U3	0,53
U4	0
U5	0

Filtragem Colaborativa

A segunda etapa do FC:

- Calcula-se as predições (Recomendações). : tendo um item a ser recomendado para um usuário alvo (item que este usuário não teve acesso). No nosso exemplo desejo calcular o quanto o usuário U1 daria de avaliação para o item B se tivesse acesso a ele.

Usuário/ Item	A	B	C	D	E
U1	3	?	4	3	2
U2	4	5	4	4	2
U3	5	1	1	5	4
U4	4	5	3	3	3
U5	1	3		2	

Filtragem Colaborativa

- Cálculo da predição da FC: assumindo $U1 = a$ como usuário alvo e B o item a ser predito, teríamos:

$$p_{ai} = \bar{r}_a + \frac{\sum_{b=1}^n (r_{bi} - \bar{r}_b) * corr_{ab}}{\sum_{b=1}^n |corr_{ab}|}$$

p_{ai} = Nota da predição de um item i para um usuário a
(considera a partir de 3 pontos)

\bar{r}_a = Média das avaliações do usuário-alvo a considerando os artigos em comum com todos os seus similares.

\bar{r}_b = Média das avaliações do usuário similar b considerando artigos em comum com o usuário-alvo a .

$corr_{ab}$ = Coeficiente de similaridade entre a e b .

r_{bi} = Avaliação dada por b ao item i .

$$\bar{r}_a = (3 + 4 + 3 + 2) / 4 = 3$$

$$p_{U1,B} = 3 + \left(\frac{((5 - 3,5) * 1) + ((1 - 3,75) * 0,53)}{1 + 0,53} \right) = 1,03 \text{ Recomendaria este item??}$$

Usuário/ Item	A	B	C	D	E
U1	3	?	4	3	2
U2	4	5	4	4	2
U3	5	1	1	5	4
U4	4	5	3	3	3
U5	1	3		2	

Usuário	Pearson (U1)
U2	1
U3	0,53
U4	0
U5	0

Filtragem Colaborativa

Limitações:

- **Problema do primeiro avaliador:**

Não existe maneira de um determinado item ser recomendado para o usuário por filtragem colaborativa até que mais informações sobre o item sejam obtidas através de outro usuário. Sendo assim é necessário realizar a avaliação inicial do item. A questão está em como obter esta primeira avaliação. Quem deveria avaliar e qual estratégia deveria ser aplicada?

Filtragem Colaborativa

Limitações:

-Similaridade

O funcionamento de um algoritmo de FC (clássico) está baseado em encontrar pessoas com gostos similares e a partir daí proceder com a busca de uma recomendação, baseando-se nas avaliações de pessoas com gostos similares.

Sendo assim, caso um usuário tenha gostos que variam do normal, o Sistema de Recomendação terá dificuldades para encontrar outros usuários com gostos similares, sendo assim as recomendações podem tornar-se pobres ou inviáveis.

Filtragem Colaborativa

Limitações:

-Novo Item

Na filtragem colaborativa, as recomendações são feitas baseada no perfil do usuário e de seus vizinhos mais próximos. Itens bem avaliados pelos vizinhos são recomendados ao usuário.

Entretanto, quando um novo item aparece no sistema e não faz parte do perfil de nenhum usuário, não é possível realizar recomendações ou predições sobre ele.

Devido ao cálculo da similaridade de Pearson, um item deve ser avaliado por pelo menos dois usuários, caso contrário, ele não entrará no cálculo da similaridade e será desconsiderado do perfil do usuário. A este problema é dado o nome de "Problema do Novo Item".

Filtragem Colaborativa

Limitações:

-Novo Usuário

Quando um usuário é cadastrado no sistema, ainda não existem avaliações realizadas por ele, portanto, seu perfil de avaliações está vazio. Como já visto, para as recomendações serem realizadas é necessário encontrar os vizinhos mais próximos do usuário, e como ele ainda não realizou avaliações, não existirão vizinhos. A este problema é dado o nome de “Problema do Novo Usuário”. Também é referenciado na literatura como “*Cold Start User*”.

Filtragem Colaborativa

Limitações:

-Escalabilidade

Um dos problemas mais críticos da FC é a escalabilidade. Quando o volume de usuários, itens e avaliações são muito grandes, os sistemas que fazem o cálculo da vizinhança (“*k-nearest-neighbor*”) de forma *on-line* (a cada cálculo de predição) podem chegar a um tempo de resposta inaceitável.

Filtragem Colaborativa

Limitações:

-Esparsialidade

Outra importante limitação de sistemas baseados em FC ocorre na medida que o número de itens na base de dados vai aumentando. Isso reduz as chances dos usuários possuírem itens em comum (informação necessária para o funcionamento dos Sistemas de Recomendação de FC), resultando na redução do tamanho médio da vizinhança dos usuários. Caso o número de usuários seja pequeno em relação ao volume de informações no sistema existe um grande risco das pontuações tornarem-se muito esparsas. Por conseguinte, o sistema terá menos vizinhos para realizar as predições, causando um impacto negativo na confiabilidade das recomendações.

Filtragem Colaborativa

Limitações:

-Superespecialização

Este problema refere-se ao fato do sistema só conseguir recomendar itens muito semelhantes àqueles que o usuário já avaliou. Por exemplo, um usuário do MovieLens que tem o perfil formado basicamente de filmes de guerra, receberá recomendações, em grande parte, de outros filmes de guerra. Isto pode tornar-se um problema, uma vez que os interesses dos usuários tendem a apresentar mudanças com o passar do tempo.

A superespecialização apresenta-se presente mais comumente em Sistemas de Recomendação baseados em conteúdo, porém a sua presença indesejada também é percebida em Sistemas de Recomendação de Filtragem Colaborativa. Para que um Sistema de Recomendação apresente o sucesso pretendido este necessita “cativar” o usuário, e a busca de conteúdo inesperado que venha a surpreender o usuário pode ser uma estratégia que venha a fidelizá-lo.

***That's all
folks!***