

Aprendizado de Máquina

Aula 9: Raciocínio Baseado em Casos e Sistemas de Recomendação (Parte 3)

André C. P. L. F de Carvalho
ICMC/USP

andre@icmc.usp.br



Tópicos do módulo

- Introdução
- Raciocínio baseado em casos
- Sistemas de recomendação
- Principais abordagens
- Filtragem por conteúdo
- Filtragem colaborativa
- Métodos baseados em modelos
- Métodos baseados em memória

Tópicos do módulo

- Introdução
- Raciocínio baseado em casos
- Sistemas de recomendação
- Principais abordagens
- Filtragem por conteúdo
- Filtragem colaborativa
- Métodos baseados em modelos
- Métodos baseados em memória

Métodos baseados em memória

- Usam algoritmos baseados em vizinhança
 - Primeiras versões de filtragem colaborativa
- Baseiam-se nas seguintes premissas:
 - Usuários semelhantes avaliam itens de forma semelhante
 - Itens com propriedades similares são avaliados de forma similar
 - Preferências de usuários se mantêm estável e consistente ao longo do tempo
- Podem ser:
 - Baseados em usuário (usuário-usuário)
 - Baseados em item (item-item)

Métodos baseados em usuário

- Recomendam a um usuário ativo **u** uma avaliação para um item alvo **i**
 - Usam avaliações feitas por um grupo de usuários semelhantes a **u** para itens diferentes de **i**
 - Usuários são semelhantes quando possuem preferências (avaliações) semelhantes às do usuário **u** para itens diferentes de **i**
 - Usuários vizinhos
 - Avaliação recomendada é definida pela combinação das avaliações destes usuários
 - Ex.: Média ponderada

Métodos baseados em usuário

- Usam avaliações feitas por usuários similares para prever uma avaliação
 - Ex.: Dada a matriz de avaliações abaixo, prever minha avaliação para o item 5

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4	4	?
Usuário1	3	1	2	3	3
Usuário2	4	3	4	3	5
Usuário3	3	3	1	5	4
Usuário4	1	5	5	2	1

Métodos baseados em usuário

- Usam avaliações feitas por usuários similares para prever uma avaliação
 - Ex.: Dada a matriz de avaliações abaixo, prever minha avaliação para o item 5

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4	4	?
Usuário1	3	1	2	3	3
Usuário2	4	3	4	3	5
Usuário3	3	3	1	5	4
Usuário4	1	5	5	2	1

Métodos baseados em usuário

- Usam avaliações feitas por usuários similares para prever uma avaliação
 - Ex.: Dada a matriz de avaliações abaixo, prever minha avaliação para o item 5

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4	4	?
Usuário1	3	1	2	3	3
Usuário2	4	3	4	3	5
Usuário3	3	3	1	5	4
Usuário4	1	5	5	2	1

Métodos baseados em usuário

- Usam avaliações feitas por usuários similares para prever uma avaliação
 - Ex.: Dada a matriz de avaliações abaixo, prever minha avaliação para o item 5

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4	4	4,5
Usuário1	3	1	2	3	3
Usuário2	4	3	4	3	5
Usuário3	3	3	1	5	4
Usuário4	1	5	5	2	1

Métodos baseados em usuário

- Necessário transformar o espaço usuário-item para usuário-usuário
 - Transforma matriz $n \times m$ em uma matriz $n \times n$
 - Onde n é o número de usuários e m o número de itens

Matriz de similaridades

- Transforma a matriz de avaliações em uma matriz de similaridades

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4	4	
Usuário1	3	1	2	3	3
Usuário2	4	3	4	3	5
Usuário3	3	3	1	5	4
Usuário4	1	5	5	2	1



	Eu	Usuário1	Usuário2	Usuário3	Usuário4
Eu	1	p1	p2	p3	p4
Usuário1	p1	1	p5	p6	p7
Usuário2	p2	p5	1	p8	p9
Usuário3	p3	p6	p8	1	p10
Usuário4	p4	p7	p9	p10	1

Matriz de avaliações 5x5 R

Matriz de similaridades 5x5 S

Métodos baseados em usuário

- Necessário transformar o espaço usuário-item para usuário-usuário
 - Transforma matriz $n \times m$ em uma matriz $n \times n$
 - Onde n é o número de usuários e m o número de itens
 - Em geral é usada correlação de Pearson ou similaridade cosseno

$$similaridade_{Pearson}(p, j) = \frac{\sum_{i \in I} (R_{pi} - \bar{R}_p) \cdot (R_{ji} - \bar{R}_j)}{\sqrt{\sum_{i \in I} (R_{pi} - \bar{R}_p)^2 \cdot \sum_{i \in I} (R_{ji} - \bar{R}_j)^2}}$$

$$similaridade_{Cosseno}(p, j) = \frac{\sum_{i \in I} (R_{pi} - \bar{R}_i) \cdot (R_{ji} - \bar{R}_i)}{\sqrt{\sum_{i \in I} (R_{pi} - \bar{R}_i)^2 \cdot \sum_{i \in I} (R_{ji} - \bar{R}_i)^2}}$$

Medem similaridade entre dois usuários na matriz de avaliações R

Métodos baseados em usuário

- Predição de novas avaliações
 - Seleciona os k usuários mais parecidos ao usuário ativo u_A , que rotulem o item alvo i_A
 - Ponderar cada avaliação dada a i_A por cada usuário selecionado pela similaridade do usuário com u_A

$$\hat{R}_{u_A i_A} = \sum_{\text{Para-cada-}u\text{-parecido-com-}u_A} \text{Similaridade}(u, u_A) \times R_{ui_A}$$

- Predição é escalada (normalizada) pela soma das similaridades

$$\hat{R}_{u_A i_A} = \frac{\sum_{\text{Para-cada-}u\text{-parecido-com-}u_A} \text{Similaridade}(u, u_A) \times R_{ui_A}}{\sum_{\text{Para-cada-}u\text{-parecido-com-}u_A} |\text{Similaridade}(u, u_A)|}$$

Métodos baseados em usuário

- Matriz de avaliações não precisa estar totalmente preenchida
 - Em geral, medida de distância usa apenas itens com valores preenchidos

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4		?
Usuário1		1	2	3	3
Usuário2	4	3	4	3	5
Usuário3		3	1		4
Usuário4		5	5	2	1

Métodos baseados em usuário

- Matriz de avaliações não precisa estar totalmente preenchida
 - Em geral, medida de distância usa apenas itens com valores preenchidos

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4		?
Usuário1		1	2	3	3
Usuário2	4	3	4	3	5
Usuário3		3	1		4
Usuário4		5	5	2	1

Métodos baseados em usuário

- Matriz de avaliações não precisa estar totalmente preenchida
 - Em geral, medida de distância usa apenas itens com valores preenchidos

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4		3
Usuário1		1	2	3	3
Usuário2	4	3	4	3	5
Usuário3		3	1		4
Usuário4		5	5	2	1

Métodos baseados em item

- Usam avaliações feitas para **itens** similares para prever uma avaliação
 - Ex.: Dada a matriz de avaliações abaixo, prever minha avaliação para o produto 5

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4	4	?
Usuário1	3	1	2	3	3
Usuário2	4	3	4	3	5
Usuário3	3	3	1	5	4
Usuário4	1	5	5	2	1

Métodos baseados em item

- Usam similaridade entre avaliações recebidas por itens para prever uma avaliação
 - Ex.: Dada a matriz de avaliações abaixo, prever minha avaliação para o produto 5

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4	4	?
Usuário1	3	1	2	3	3
Usuário2	4	3	4	3	5
Usuário3	3	3	1	5	4
Usuário4	1	5	5	2	1

Métodos baseados em item

- Usam similaridade entre avaliações recebidas por itens para prever uma avaliação
 - Ex.: Dada a matriz de avaliações abaixo, prever minha avaliação para o produto 5

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4	4	4,5
Usuário1	3	1	2	3	3
Usuário2	4	3	4	3	5
Usuário3	3	3	1	5	4
Usuário4	1	5	5	2	1

Métodos baseados em item

- Necessário transformar o espaço usuário-item para item-item
 - Transforma matriz $n \times m$ em uma matriz $m \times m$
 - Onde n é o número de usuários e m o número de itens

Matriz de similaridades

- Transforma a matriz de avaliações em uma matriz de similaridades

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4	4	
Usuário1	3	1	2	3	3
Usuário2	4	3	4	3	5
Usuário3	3	3	1	5	4
Usuário4	1	5	5	2	1



	Item1	Item2	Item3	Item4	Item5
Item1	1	p1	p2	p3	p4
Item2	p1	1	p5	p6	p7
Item3	p2	p5	1	p8	p9
Item4	p3	p6	p8	1	p10
Item5	p4	p7	p9	p10	1

Matriz de avaliações 5x5 R

Matriz de similaridades 5x5 S

Métodos baseados em item

- Matriz não precisa estar totalmente preenchida
 - Medida de distância usa usuários com valores preenchidos

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4		?
Usuário1		1	2	3	3
Usuário2	4	3	4	3	5
Usuário3		3	1		4
Usuário4		5	5	2	1

Métodos baseados em item

- Necessário transformar o espaço usuário-item para item-item
 - Transforma matriz $n \times m$ em uma matriz $m \times m$
 - Onde n é o número de usuários e m o número de itens
 - Pode usar correlação de Pearson ou similaridade cosseno

$$similaridade_{Pearson}(p, j) = \frac{\sum_{u \in U}^m (r_{pu} - \bar{R}_p) \cdot (r_{ju} - \bar{R}_j)}{\sqrt{\sum_{u \in U}^m (r_{pu} - \bar{R}_p)^2 \cdot \sum_{u \in U}^m (r_{ju} - \bar{R}_j)^2}}$$

$$similaridade_{Cosseno}(p, j) = \frac{\sum_{u \in U}^m (R_{pu} - \bar{R}_u) \cdot (R_{ju} - \bar{R}_u)}{\sqrt{\sum_{u \in U}^m (R_{pu} - \bar{R}_u)^2 \cdot \sum_{u \in U}^m (R_{ju} - \bar{R}_u)^2}}$$

Medem similaridade entre dois itens na matriz de avaliações R

Métodos baseados em item

- Predição de novas avaliações
 - Seleciona os k itens mais parecidos ao item alvo i_A , que foram rotulados pelo usuário ativo u_A
 - Ponderar cada avaliação dada por u_A para cada item selecionado pela similaridade do item com o item alvo i_A

$$\hat{R}_{u_A i_A} = \sum_{\text{Para-cada-}i\text{-parecidos-com-}i_A} \text{Similaridade}(i, i_A) \times r_{u_A i}$$

- Predição pode ser escalada (normalizada) pela soma das similaridades

$$\hat{R}_{u_A i_A} = \frac{\sum_{\text{Para-cada-}i\text{-parecidos-com-}i_A} \text{Similaridade}(i, i_A) \times r_{u_A i}}{\sum_{\text{Para-cada-}i\text{-parecidos-com-}i_A} |\text{Similaridade}(i, i_A)|}$$

Problema de cold start

- Não existem avaliações suficientes para o processo de recomendação
- Número de avaliações inicialmente disponíveis é pequeno
- Vários métodos têm sido propostos para lidar com cold start

Sistemas híbridos

- Combinam aspectos positivos de diferentes abordagens
 - Pega o melhor que tem em cada uma delas
- Permite criar sistemas mais robustos em uma ampla variedade de situações
- Pode ter diferentes estruturas
 - Paralela
 - Pipeline

Conclusão

- Métodos baseados em proximidade
- Raciocínio baseado em casos
- Sistemas de recomendação
- Principais abordagens
- Filtragem por conteúdo
- Filtragem colaborativa
- Exemplos

Fim do
módulo