

Sistemas de recomendação utilizando o software R

Leonardo Filgueira

06 de julho de 2018

Motivação

- ▶ Lojas *on-line*.
- ▶ Variedade de informações ofertadas.
- ▶ Dados coletados.
- ▶ Aproveitamento dos dados.
- ▶ Aumento nas vendas das lojas.
- ▶ Aproveitamento do tempo do cliente.

Introdução

- ▶ Utilizado desde década de 90.
- ▶ Muito usado em ambientes *on-line*.
- ▶ Sugerir itens para o usuário.
- ▶ Calcular avaliação prevista.

Avaliações

- ▶ Avaliações em diferentes formatos:
 - ▶ Avaliações numéricas;
 - ▶ Avaliações qualitativas;
 - ▶ Avaliações binárias;
 - ▶ Avaliação unária.
- ▶ Representadas em matriz.

Matriz de avaliações

	Item 1	Item 2	...	Item m
Usuário 1	$r_{(1,1)}$...	
Usuário 2		$r_{(2,2)}$...	$r_{(2,m)}$
⋮	⋮	⋮	⋮	⋮
Usuário n			...	$r_{(n,m)}$

- ▶ Matriz esparsa.
- ▶ Vários dados faltantes.

Algoritmos

- ▶ Filtragem baseada em conteúdo.
- ▶ Filtragem colaborativa.
- ▶ Sistemas híbridos.

Filtragem baseada em conteúdo

- ▶ Uso do histórico do usuário.
- ▶ Associação entre itens.
- ▶ Buscar itens mais associados aos do histórico.
- ▶ Técnicas de *text mining* para associação.
- ▶ Não considera comportamento de outros usuários.

Filtragem colaborativa

- ▶ Uso das informações de outros usuários.
- ▶ Agrupamento de itens/usuários.
- ▶ Tipos:
 - ▶ Baseada no item,
 - ▶ Baseada no usuário.

Filtragem colaborativa baseada no item

- ▶ Similaridade entre itens i e j :
- ▶ Coeficiente de correlação de Pearson:

$$w_{i,j} = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_i)^2 \sum_{u \in U} (r_{u,j} - \bar{r}_j)^2}} \quad (1)$$

- ▶ U : conjunto de usuários que avaliaram os dois itens.
- ▶ $r_{u,i}$ o rating dado pelo usuário u ao item i .
- ▶ \bar{r}_i o rating médio recebido pelo item i dado por todos os usuários que o avaliaram.

Filtragem colaborativa baseada no item

- Cosseno entre vetores de avaliações:

$$w_{i,j} = \cos(\vec{r}_i, \vec{r}_j) = \frac{\vec{r}_i \cdot \vec{r}_j}{\|\vec{r}_i\| \times \|\vec{r}_j\|} = \frac{\sum_{u=1}^n r_{u,i} r_{u,j}}{\sqrt{\sum_{u=1}^n r_{u,i}^2 \sum_{u=1}^n r_{u,j}^2}} \quad (2)$$

Filtragem colaborativa baseada no item

- ▶ *rating* previsto:

$$p_{a,i} = \frac{\sum_{j \in k} r_{a,i} - w_{i,j}}{\sum_{j \in k} |w_{i,j}|} \quad (3)$$

- ▶ Onde k é a vizinhança do item i .

Filtragem colaborativa baseada no usuário

- ▶ Similaridade entre itens a e u .
- ▶ Coeficiente de correlação de Pearson:

$$w_{a,u} = \frac{\sum_{i \in I} (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i \in I} (r_{a,i} - \bar{r}_a)^2 \sum_{i \in I} (r_{u,i} - \bar{r}_u)^2}} \quad (4)$$

- ▶ Cosseno entre os vetores de avaliações:

$$w_{a,u} = \cos(\vec{r}_a, \vec{r}_u) = \frac{\vec{r}_a \cdot \vec{r}_u}{\|\vec{r}_a\| \times \|\vec{r}_u\|} = \frac{\sum_{i=1}^m r_{a,i} r_{u,i}}{\sqrt{\sum_{i=1}^m r_{a,i}^2 \sum_{i=1}^m r_{u,i}^2}} \quad (5)$$

Filtragem colaborativa baseada no usuário

- ▶ Cálculo do valor previsto da avaliação:

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u \in k} (r_{u,i} - \bar{r}_u) w_{a,u}}{\sum_{u \in k} |w_{a,u}|} \quad (6)$$

- ▶ Onde k é a vizinhança do usuário a .

Planos para Projeto Final II

- ▶ Incluir filtragem colaborativa *model based*, em especial ALS.
- ▶ Divisão em clusters antes de rodar algoritmos e comparar resultados.