

SCC0284 / SCC5966

Sistemas de Recomendação

Aula 03: Filtragem Baseada em Conteúdo

(mmanzato@icmc.usp.br)

Filtragem Baseada em Conteúdo (FBC)

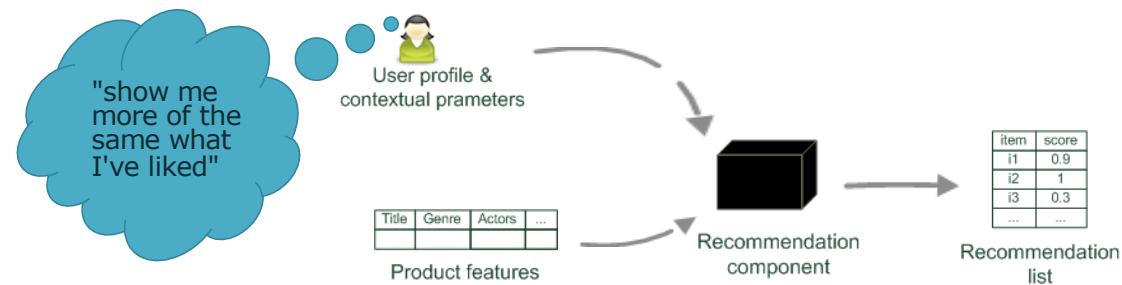
- FC não utiliza nenhuma informação sobre os itens
 - Apenas interações de usuários
 - Problemas de cold-start e esparsidade
- FBC calcula recomendações utilizando:
 - Descrições sobre os itens (metadados)
 - Perfil de usuário contendo o que ele gosta / não gosta

FBC



FBC

- Tarefa:
 - Coletar / aprender sobre as preferências do usuário
 - Buscar / recomendar itens que sejam similares às preferências do usuário



FBC

- Representação de itens
 - Utilização de metadados estruturados, semi-estruturados e/ou não-estruturados



| Title | Genre | Author | Type | Price | Keywords |
|----------------------|-------------------|-------------------|-----------|-------|--|
| The Night of the Gun | Memoir | David Carr | Paperback | 29.90 | Press and journalism, drug addiction, personal memoirs, New York |
| The Lace Reader | Fiction, Mystery | Brunonia Barry | Hardcover | 49.90 | American contemporary fiction, detective, historical |
| Into the Fire | Romance, Suspense | Suzanne Brockmann | Hardcover | 45.90 | American fiction, murder, neo-Nazism |

FBC baseada em vizinhos mais próximos (FBC-kNN)

- Dado um conjunto de itens I_u já avaliados pelo usuário
- Encontrar os n vizinhos mais próximos de um item i ainda não avaliado
 - Usar métricas de similaridade aplicadas em vetores de características
- Usar esses vizinhos para prever uma avaliação para i

FBC-kNN

- Exemplo



| | | | | | | |
|---------|---|---|---|---|---|---|
| Jessica | ? | 2 | 4 | 3 | 2 | 3 |
| Marta | 4 | 3 | ? | 4 | 3 | 2 |
| Jose | 1 | 5 | 3 | 4 | ? | 5 |
| Dave | 1 | ? | 2 | 3 | 4 | ? |

| | | | | | | |
|---------|---|---|---|---|---|---|
| Drama | 1 | 0 | 1 | 1 | 0 | 0 |
| Ação | 0 | 1 | 0 | 1 | 1 | 0 |
| Comédia | 1 | 1 | 0 | 0 | 0 | 1 |
| Sci-Fi | 0 | 0 | 1 | 0 | 1 | 0 |

FBC-kNN

- Exemplo
 - Calcular as similaridades entre itens usando os vetores de atributos
 - Métricas possíveis: Cosseno, Pearson, Jaccard, etc.
 - Selecionar os **k** itens mais similares a **i** que tenham sido avaliados por **u**
 - Calcular a predição usando a mesma estratégia da FC:

$$pred(u,i) = \frac{\sum_{j \in I_{ui}} sim(i,j) * r_{uj}}{\sum_{j \in I_{ui}} sim(i,j)}$$

I_{ui} : conj. dos **k** itens mais similares a **i** que foram avaliados por **u**

FBC-kNN

- Variações
 - Alterar o tamanho de vizinhos (k)
 - Utilizar limiares superiores e inferiores para similaridade
- Vantagens
 - Bom para modelar interesses de curtíssimo-prazo ou itens que formam uma história em sequência
 - Pode ser usado em combinação com outros métodos para modelar preferências de longo-prazo

Métodos Probabilísticos

- Calcular a probabilidade condicional de cada item ser relevante ou não-relevante de acordo com o conjunto de características que ele possui:

$$P(i = \textit{relevante} \mid X)$$

$$P(i = \textit{n\~ao.relevante} \mid X)$$

- onde X é o vetor de características do item i
- Teorema de Bayes:

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)} = \frac{\prod_{i=1}^d P(X_i \mid Y)P(Y)}{P(X)}$$

Métodos Probabilísticos

- Exemplo

Interações de um único usuário:

| ItemID | recommender | intelligent | learning | school | Label |
|--------|-------------|-------------|----------|--------|-------|
| 1 | 1 | 1 | 1 | 0 | R |
| 2 | 0 | 0 | 1 | 1 | N |
| 3 | 1 | 1 | 0 | 0 | R |
| 4 | 1 | 0 | 1 | 1 | R |
| 5 | 0 | 0 | 0 | 1 | N |
| 6 | 1 | 1 | 0 | 0 | ? |

Obs. Para evitar zerar a probabilidade, é possível aplicar uma suavização de Laplace, como:

$$P(t = 1 \mid \text{Label} = R) = (qtde + 1) / (\text{total} + |V|)$$

onde:

t : palavra-chave

|V| é o no. de palavras-chave

$$\begin{aligned} P(X \mid R) &= P(\text{recommender} = 1 \mid \text{Label} = R) \\ &\times P(\text{intelligent} = 1 \mid \text{Label} = R) \\ &\times P(\text{learning} = 0 \mid \text{Label} = R) \\ &\times P(\text{school} = 0 \mid \text{Label} = R) \\ &= (3+1)/(3+4) \times (2+1)/(3+4) \times (1+1)/(3+4) \\ &\times (2+1)/(3+4) = 0.0299 \end{aligned}$$

$$P(R \mid X) = 0.0299 \times 0.6 = 0.01794$$

$$\begin{aligned} P(X \mid R) &= P(\text{recommender} = 1 \mid \text{Label} = N) \\ &\times P(\text{intelligent} = 1 \mid \text{Label} = N) \\ &\times P(\text{learning} = 0 \mid \text{Label} = N) \\ &\times P(\text{school} = 0 \mid \text{Label} = N) \\ &= (0+1)/(2+4) \times (0+1)/(2+4) \times (1+1)/(2+4) \\ &\times (0+1)/(2+4) = 0.0015 \end{aligned}$$

$$P(N \mid X) = 0.0015 \times 0.4 = 0.0006$$

$P(X) \rightarrow \text{constante}$

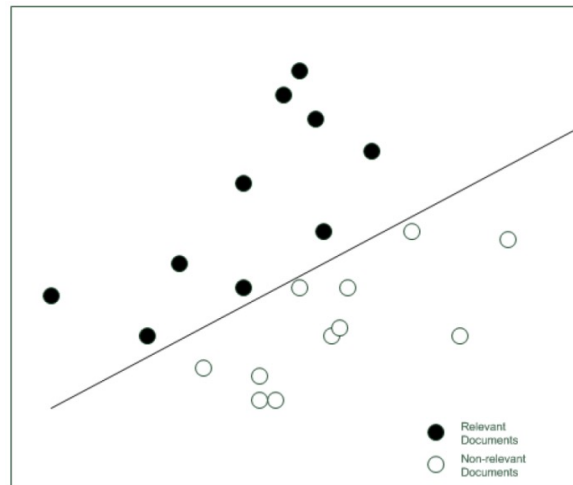
$P(Y)$:

$$P(\text{Label} = R) = 3/5 = 0.6$$

$$P(\text{Label} = N) = 2/5 = 0.4$$


Aprendizado de Máquina

- Recomendação baseada em conteúdo pode ser vista como um problema de classificação/regressão
- Um modelo linear pode discriminar itens entre as classes relevante e não-relevante



Aprendizado de Máquina

- Exemplo

| | | | | | | |
|---------|---|---|---|---|---|--|
| |  |  |  |  |  |  |
| Jessica | ? | 2 | 4 | 3 | 2 | 3 |
| Marta | 4 | 3 | ? | 4 | 3 | 2 |
| Jose | 1 | 5 | 3 | 4 | ? | 5 |
| Dave | 1 | ? | 2 | 3 | 4 | ? |

| | | | | | | |
|---------|---|---|---|---|---|---|
| Drama | 1 | 0 | 1 | 1 | 0 | 0 |
| Ação | 0 | 1 | 0 | 1 | 1 | 0 |
| Comédia | 1 | 1 | 0 | 0 | 0 | 1 |
| Sci-Fi | 0 | 0 | 1 | 0 | 1 | 0 |

Aprendizado de Máquina

- Exemplo



| | | | | | | | |
|---------|---|---|---|---|---|---|----------------|
| Jessica | ? | 2 | 4 | 3 | 2 | 3 | $\Theta^{(1)}$ |
| Marta | 4 | 3 | ? | 4 | 3 | 2 | $\Theta^{(2)}$ |
| Jose | 1 | 5 | 3 | 4 | ? | 5 | $\Theta^{(3)}$ |
| Dave | 1 | ? | 2 | 3 | 4 | ? | $\Theta^{(4)}$ |

| | | | | | | | |
|---------|-----------|-----------|-----------|-----------|-----------|-----------|-------|
| Drama | 1 | 0 | 1 | 1 | 0 | 0 | x_1 |
| Ação | 0 | 1 | 0 | 1 | 1 | 0 | x_2 |
| Comédia | 1 | 1 | 0 | 0 | 0 | 1 | x_3 |
| Sci-Fi | 0 | 0 | 1 | 0 | 1 | 0 | x_4 |
| | $x^{(1)}$ | $x^{(2)}$ | $x^{(3)}$ | $x^{(4)}$ | $x^{(5)}$ | $x^{(6)}$ | |

$$x^{(\text{The Princess Diary})} = \{1, 0, 1, 0, 1\}$$

Para cada usuário u , queremos aprender um perfil $\Theta^{(u)} \in \mathbb{R}^{m+1}$ (m = qtde. atributos)

Aprendizado de Máquina

- Exemplo

| | | | | | | | |
|---------|---|---|---|---|---|--|----------------|
| |  |  |  |  |  |  | |
| Jessica | ? | 2 | 4 | 3 | 2 | 3 | $\Theta^{(1)}$ |
| Marta | 4 | 3 | ? | 4 | 3 | 2 | $\Theta^{(2)}$ |
| Jose | 1 | 5 | 3 | 4 | ? | 5 | $\Theta^{(3)}$ |
| Dave | 1 | ? | 2 | 3 | 4 | ? | $\Theta^{(4)}$ |

| | | | | | | | |
|---------|---|---|---|---|---|---|-------|
| Drama | 1 | 0 | 1 | 1 | 0 | 0 | x_1 |
| Ação | 0 | 1 | 0 | 1 | 1 | 0 | x_2 |
| Comédia | 1 | 1 | 0 | 0 | 0 | 1 | x_3 |
| Sci-Fi | 0 | 0 | 1 | 0 | 1 | 0 | x_4 |

$x^{(1)} \quad x^{(2)} \quad x^{(3)} \quad x^{(4)} \quad x^{(5)} \quad x^{(6)}$

$x^{(\text{The Princess Diary})} = \{1, 0, 1, 0, 1\}$

Para cada usuário u , queremos aprender um perfil $\Theta^{(u)} \in \mathbb{R}^{m+1}$ (m = qtde. atributos)

A predição é realizada como:

$$\hat{r}_{ui} = (\theta^{(u)})^T x^{(i)}$$

Aprendizado de Máquina

- Exemplo

| | | | | | | | |
|---------|---|---|---|---|---|--|----------------|
| |  |  |  |  |  |  | |
| Jessica | ? | 2 | 4 | 3 | 2 | 3 | $\Theta^{(1)}$ |
| Marta | 4 | 3 | ? | 4 | 3 | 2 | $\Theta^{(2)}$ |
| Jose | 1 | 5 | 3 | 4 | ? | 5 | $\Theta^{(3)}$ |
| Dave | 1 | ? | 2 | 3 | 4 | ? | $\Theta^{(4)}$ |
| Drama | 1 | 0 | 1 | 1 | 0 | 0 | x_1 |
| Ação | 0 | 1 | 0 | 1 | 1 | 0 | x_2 |
| Comédia | 1 | 1 | 0 | 0 | 0 | 1 | x_3 |
| Sci-Fi | 0 | 0 | 1 | 0 | 1 | 0 | x_4 |
| | $x^{(1)}$ | $x^{(2)}$ | $x^{(3)}$ | $x^{(4)}$ | $x^{(5)}$ | $x^{(6)}$ | |

$x^{(\text{The Princess Diary})} = \{1, 0, 1, 0, 1\}$

Para cada usuário u , queremos aprender um perfil $\Theta^{(u)} \in \mathbb{R}^{m+1}$ (m = qtde. atributos)

A predição é realizada como:

$$\hat{r}_{ui} = (\theta^{(u)})^T x^{(i)}$$

Por exemplo, se o perfil de **Jéssica** foi treinado como: $\Theta^{(\text{Jessica})} = \{1.5, 0, -0.5, 2, 2\}$, então a predição para “**The Princess Diary**” será:

$$\text{pred} = 1.5*1+0*0-0.5*1+2*0+2*1 = 3$$

Aprendizado de Máquina

- Dados:
 - K : conjunto de pares (u,i) cujas notas são conhecidas
 - r_{ui} : nota dada pelo usuário u para o item i
 - $x^{(i)}$: vetor de características do item i
 - Predição: $\hat{r}_{ui} = (\theta^{(u)})^T x^{(i)}$
 - m, c, p : no. de características / usuários / itens
 - $\Theta^{(u)}$: vetor de parâmetros (perfil) do usuário u
 - $\Theta^{(u)} \in \mathbb{R}^{m+1}$

Aprendizado de Máquina

- Aprendizado de $\Theta^{(u)}$:

$$\min_{\theta^{(u)}} \frac{1}{2} \sum_{i:(u,i) \in K} \left((\theta^{(u)})^T (x^{(i)}) - r_{ui} \right)^2 + \frac{\lambda}{2} \sum_{k=1}^m \left(\theta_k^{(u)} \right)^2$$

Aprendizado de Máquina

- Aprendizado de $\Theta^{(u)}$:

$$\min_{\theta^{(u)}} \frac{1}{2} \sum_{i:(u,i) \in K} \left((\theta^{(u)})^T (x^{(i)}) - r_{ui} \right)^2 + \frac{\lambda}{2} \sum_{k=1}^m \left(\theta_k^{(u)} \right)^2$$

- Aprendizado de $\Theta^{(1)}, \Theta^{(2)}, \dots, \Theta^{(c)}$:

$$\min_{\theta^{(1)}, \dots, \theta^{(c)}} \frac{1}{2} \sum_{u=1}^c \sum_{i:(u,i) \in K} \left((\theta^{(u)})^T (x^{(i)}) - r_{ui} \right)^2 + \frac{\lambda}{2} \sum_{u=1}^c \sum_{k=1}^m \left(\theta_k^{(u)} \right)^2$$

Aprendizado de Máquina

- Função custo:

$$J(\theta^{(1)}, \dots, \theta^{(c)}) = \frac{1}{2} \sum_{u=1}^c \sum_{i: (u,i) \in K} \left((\theta^{(u)})^T (x^{(i)}) - r_{ui} \right)^2 + \frac{\lambda}{2} \sum_{u=1}^c \sum_{k=1}^m \left(\theta_k^{(u)} \right)^2$$

Aprendizado de Máquina

- Função custo:

$$J(\theta^{(1)}, \dots, \theta^{(c)}) = \frac{1}{2} \sum_{u=1}^c \sum_{i:(u,i) \in K} \left((\theta^{(u)})^T (x^{(i)}) - r_{ui} \right)^2 + \frac{\lambda}{2} \sum_{u=1}^c \sum_{k=1}^m \left(\theta_k^{(u)} \right)^2$$

- Derivadas:

$$\frac{\partial J}{\partial \theta_k^{(u)}} = \sum_{i:(u,i) \in K} \left((\theta^{(u)})^T (x^{(i)}) - r_{ui} \right) x_k^{(i)} + \lambda \theta_k^{(u)} \quad \text{para } k = 1, \dots, m$$

$$\frac{\partial J}{\partial \theta_k^{(u)}} = \sum_{i:(u,i) \in K} \left((\theta^{(u)})^T (x^{(i)}) - r_{ui} \right) x_k^{(i)} \quad \begin{array}{l} \text{para } k = m+1 \\ \text{(viés)} \end{array}$$

Aprendizado de Máquina

- Atualização via Gradiente Descendente Estocástico:

$$\varepsilon_{ui} = (\theta^{(u)})^T x^{(i)} - r_{ui}$$

$$\theta_k^{(u)} = \theta_k^{(u)} - \gamma \left(\varepsilon_{ui} x_k^{(i)} + \lambda \theta_k^{(u)} \right) \quad \text{para } k = 1, \dots, m$$

$$\theta_k^{(u)} = \theta_k^{(u)} - \gamma \varepsilon_{ui} x_k^{(i)} \quad \text{para } k = m+1 \text{ (viés)}$$

Aprendizado de Máquina

- Algoritmo

inicializar Θ

repita

para cada (u,i) com nota conhecida

calcular predição

calcular erro

para cada característica k:

atualizar Θ_{uk}

até convergir

Limitações da FBC

- Palavras-chave (características) podem não ser suficientes para julgar a qualidade ou relevância de um item
 - Falta de descrições, semântica, características não textuais, etc.
- Problema do novo usuário
- Sobre-especialização ou filtros bolha

Referências

- Dietmar Jannach, Markus Zanker, Alexander Felfernig, Gerhard Friedrich. *Recommender Systems: An Introduction*. Cambridge University Press, 2010.
- <https://www.coursera.org/learn/machine-learning/lecture/uG59z/content-based-recommendations>