



Recomendação de produtos baseado em palavras-chave

Matheus Cesário Mello dos Santos

Orientadora Profa. Dra. Simone do Prado



Informação

- ▶ 150 mil url's são criadas por dia (fonte: Verisign)
 - ▶ 1992: menos de 15 mil websites
 - ▶ 2008: 350 milhões de domínios
 - ▶ 1 trilhão de urls únicas indexadas pelo buscador Google
- ▶ Busca por informações
 - ▶ Muitos dados
 - ▶ Dificuldade em processá-los
- ▶ Web mais interativa
 - ▶ Web 1.0, 2.0 e 3.0



Livros

- ▶ Google Books: digitalizar 130 milhões de obras singulares.
- ▶ Suporte a busca, incentivo a leitura e descobertas
 - ▶ Caracterizar um item
 - ▶ Determinar itens parecidos
 - ▶ Melhores obras: definição de “melhor”?



Teoria da cauda longa



- ▶ A vantagem de uma página web sobre uma loja física
- ▶ O futuro dos negócios é vender pouco de muito (ANDERSON, 2006).



Motivação

- ▶ Disseminar conhecimento
- ▶ Atrair novos leitores
- ▶ Aproximar experiência e curiosidade

“Uma ferramenta que ajude a encontrar seu próximo livro favorito”



Sistema de recomendações

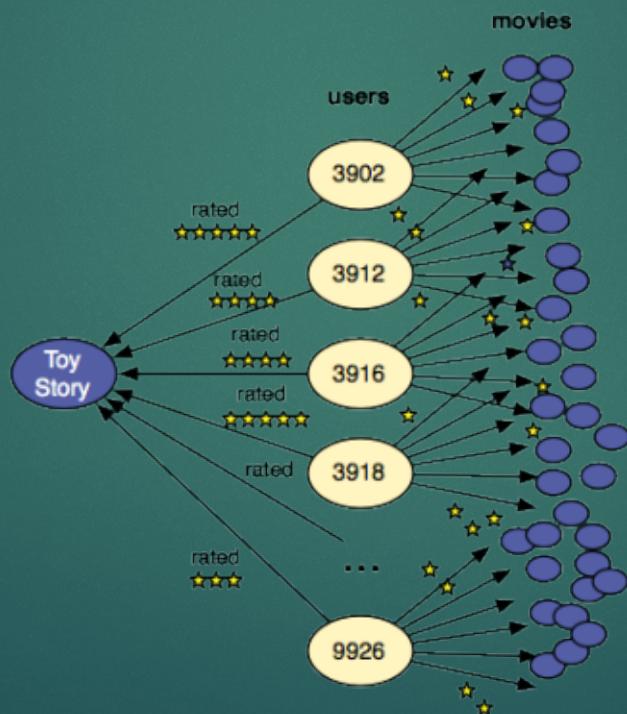
- ▶ O que é?
- ▶ Similaridade
 - ▶ Aspecto ou característica que se assemelha a outro(s).
- ▶ Aplicações
 - ▶ Comércio
 - ▶ Serviços
 - ▶ Entre outros



Abordagens

Filtragem colaborativa

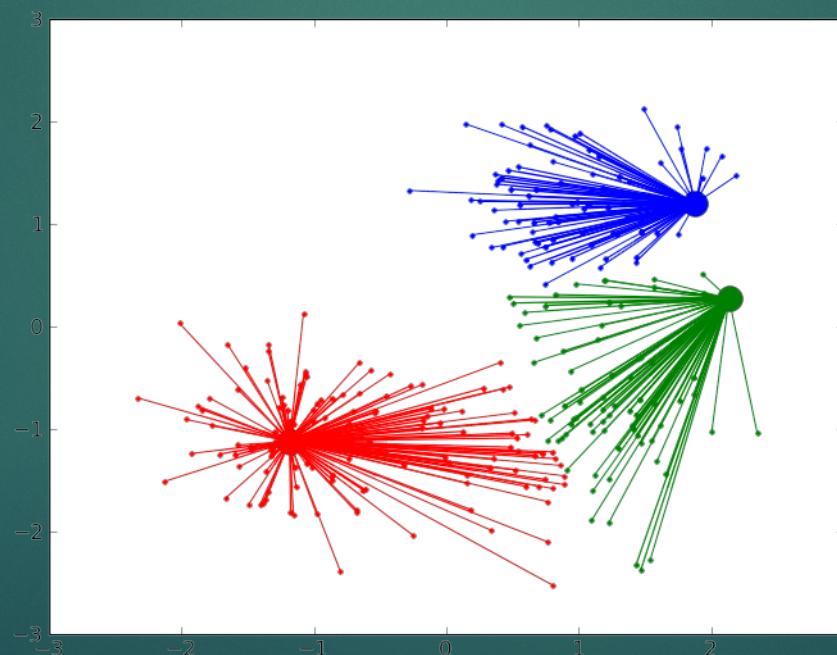
- ▶ Criar previsões baseado nas avaliações dos usuários





Abordagens Clustering

- ▶ Dividir base de dados em grupos que compartilham das mesmas características

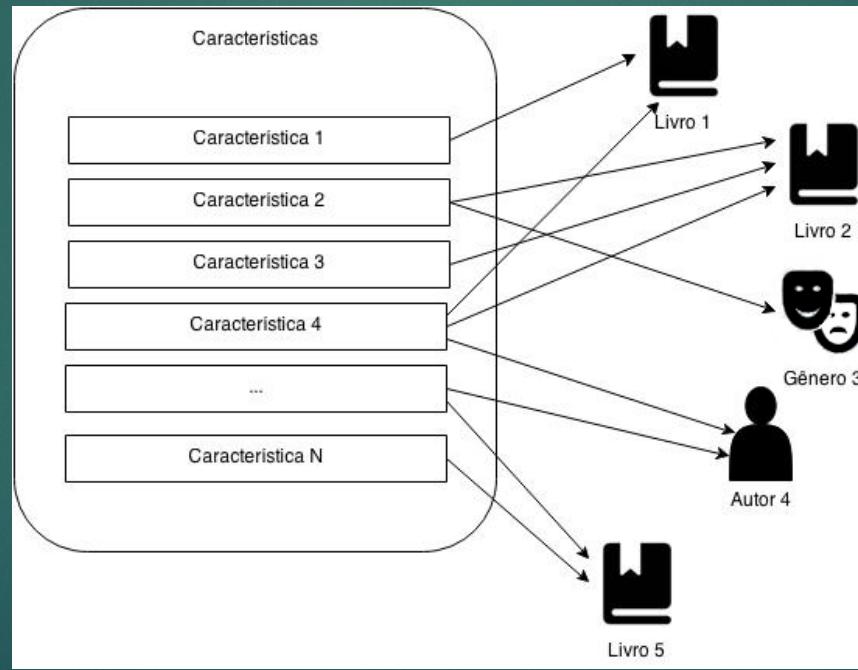




Abordagens

Baseado em conteúdo

- Recomendar um item baseado em suas características





Abordagens Hibridização

- ▶ Utilizar sistemas de suporte incrementa o poder de recomendação



Cálculo de similaridade

Cosseno

- Cosseno de vetores

$$\cos(\theta) = \frac{A \cdot B}{|A| |B|} = \frac{\sum_1^n A_i \times B_i}{\sqrt{\sum_1^n (A_i)^2} \times \sqrt{\sum_1^n (B_i)^2}}$$

$\therefore \cos(90) = 0$, $\cos(0) = 1$ e $\cos(180) = -1$



Outros

Tanimoto, Distância euclidiana, Pearson, Bayes

- ▶ Tanimoto

- ▶
$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

- ▶ Distância euclidiana

- ▶
$$d(p, q) = d(q, p) = \sqrt{(q_1 - p_1)^2 + \dots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

- ▶ Pearson

- ▶
$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{cov(x, y)}{\sqrt{var(x) \times var(y)}}$$

- ▶ Bayes (menção)

- ▶
$$p(x | \alpha) = \frac{\tau(a_0)}{\prod_{i=1}^n \tau(a_i)} \times \prod_{i=1}^n x_i^{a_i-1}$$



Ferramentas de desenvolvimento

- ▶ HTML
- ▶ Javascript
- ▶ CSS
- ▶ PHP
- ▶ MySQL





GNO

“WHAT READ NEXT?”



Gno

- ▶ Modelo MVC (modelo, visão e controle)
- ▶ Análise dos algoritmos
- ▶ Demonstração
- ▶ *Layouts*



16

Layouts

The screenshot shows a search results page for the term 'Filosofia'. At the top, there's a navigation bar with 'GNO' and a search icon. Below it, a large banner features the word 'Filosofia' in white on a teal background. The main content area displays several book cards:

- A book by Friedrich Nietzsche titled 'Além do bem e do mal' with a rating of 4.8 and 546 reviews.
- A book by Espinosa titled 'Além do bem e do mal'.
- A book by Deleuze titled 'Além do bem e do mal'.
- A book by Foucault titled 'Além do bem e do mal'.
- A book by Nietzsche titled 'Além do bem e do mal'.

Below the books, there's a sidebar with a profile picture of 'Walter White' and sections for 'Principais autores' and 'Ver todos'.

This screenshot shows a mobile version of the same search results page. The interface is designed for a smaller screen, with the main content area displaying the same book cards as the desktop version. The navigation bar at the top includes 'Pesquisar', 'My Shelf', and a user profile icon for 'Walter'.

The screenshot shows a feed of posts from various users. The top post is by 'Walter White' about a book by 'Chemistry'. Below it, there are posts from 'Hank Schrader' and 'Jesse Pinkman' with placeholder text. Further down, there are posts from 'Walter White' about 'Zygmunt Bauman' and 'Hank Schrader' about 'Chemistry'. Each post includes a timestamp, a small profile picture, and a comment section with a placeholder 'Deixe seu comentário...'.



Conclusões

- ▶ Problemas encontrados
 - ▶ Soluções propostas
- ▶ Futuros trabalhos
- ▶ Importância deste trabalho



18

Dúvidas?





Q&A

(discussão sobre o trabalho)



Método 1

Selecionando palavras-chave

Para cada par *<item, palavra-chave>*, faça

Se *item* = *ITEM* então

Listा.adicionar(<palavra-chave>)

FimSe;

FimPara

Retorna Lista



Método 2

Selecionando itens similares

Para cada par $\langle \text{item}, \text{palavra-chave} \rangle$, faça

Se $\text{Item.hashTable}(\langle \text{item} \rangle)$ não existe, então

Para cada palavra-chave de Lista, faça

Se $\langle \text{palavra-chave} \rangle = \text{palavra-chave}$, então

$\text{Item.adiciona}(\langle \text{item} \rangle)$

FimPara;

FimSe;

FimPara;



Método 3

Cálculo por cosseno

Função *COSSENO(item, ITEM)*:

V1 = palavrasChave(item); // <palavras-chave>

V2 = Lista; // (Palavras-chave de método 1)

*Retorna (V1 . V2 / |V1| * |V2|)*

FimFunção

Para cada item de Item, faça

Sim[item] = COSSENO(item, ITEM);

FimPara;

[Voltar](#)



Método 4

Classificando um usuário

Para cada interação de Interações, faça

 Para cada par <item, palavra-chave> dentro de interação, faça

 Se Lista.hashTable(item) não existe, então

 Se item = ITEM então

 Lista.adicionar(<palavra-chave>)

 FimSe;

 FimSe;

 FimPara;

FimPara;

[Voltar](#)



Prevendo um item Cosseno

Rating	Item 1	Item 2	Item 3
Usuario	1	0	?

Similarity	Item 1	Item 2
Item 3	0.9	0.2

Prediction (Usuario, Item3)

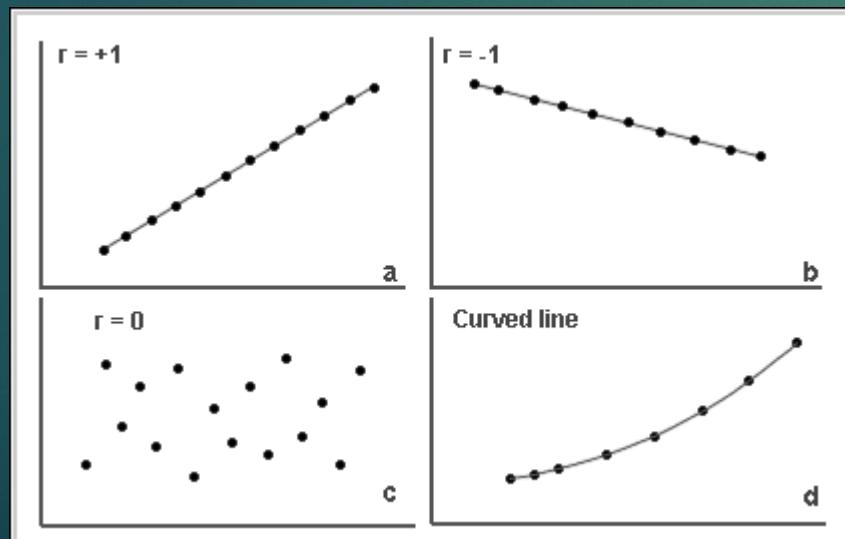
$$= \frac{\sum_{i \in ItensAvaliados} sim(i, p) * nota(i)}{\sum_{i \in ItensAvaliados} sim(i, p)} = \frac{0.9 * 1 + 0.2 * 0}{0.9 + 0.2} = 0.99$$



Prevendo um item

Correlação de Pearson

Rating	Item1	Item2	Item3	Item4
Ulisses	1	0	1	?
Perseu	1	0	-	1
Jasão	0	1	0	0



- Selecionar N usuários mais próximos.
No caso, Perseu.
- Previsão será intersecção formada entre a reta correspondente a nota dada por Perseu e a encontrada nos pontos I_{xy}



Métodos de recomendação

Filtragem colaborativa

Rating	Item1	Item2	Item3	Item4
Valjean	1	0	1	1
Cosette		1	0	0
Javert	1	1	0	
Fantine	0	1		0
Gavroche	1	1	?	0

- Selecionar usuários que tenham pelo menos 1 avaliação em comum
 - No caso, Valjean (1), Cosette (2) e Javert (2)
- Para cada usuário encontrado, calcular uma previsão
 - No exemplo, a previsão seria de que Gavroche não iria gostar do item.