

Modelo Booleano

Wendel Melo

Faculdade de Computação
Universidade Federal de Uberlândia

Recuperação da Informação

Adaptado do Material da Profª Vanessa Braganholo - IC/UFF

Modelo Booleano

- Modelo simples;
- Baseado em teoria dos conjuntos e álgebra booleana;
- Documentos e consultas são representados como vetores binários;
- Consultas são especificadas como expressões booleanas

Modelo Booleano

- Cada documento é representado logicamente por um vetor de pesos binários ;
- Seja w_{ij} o peso do termo k_i no documento d_j . Assim:

$$w_{ij} = \begin{cases} 1, & \text{se } k_i \text{ aparece em } d_j, \\ 0, & \text{caso contrário} \end{cases}$$

- Desse modo, o número de componentes do vetor de pesos é dado pelo número de termos do vocabulário da base;
- Consultas também são representadas por vetores de pesos binários.

Exemplo 1

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

O índice invertido terá os seguintes termos (com eliminação de *stopwords*):

- ▶ amarela
- ▶ azul
- ▶ campo
- ▶ carro
- ▶ casa
- ▶ é
- ▶ linda
- ▶ marcelo

Exemplo 1

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

O índice invertido terá os seguintes termos (com eliminação de *stopwords*):

- ▶ amarela
- ▶ azul
- ▶ campo
- ▶ carro
- ▶ casa
- ▶ é
- ▶ linda
- ▶ marcelo

Cada documento será modelado como um vetor de pesos binários:

$$\begin{array}{c} \text{amarela} \\ \text{azul} \\ \text{campo} \\ \text{carro} \\ \text{casa} \\ \text{é} \\ \text{linda} \\ \text{marcelo} \end{array}$$
$$\overline{D1} = (1, 1, 1, 0, 1, 1, 1, 0)$$
$$\overline{D2} = (0, 1, 0, 1, 0, 1, 0, 1)$$

Exemplo 1

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

O índice invertido terá os seguintes termos (com eliminação de *stopwords*):

- ▶ amarela
- ▶ azul
- ▶ campo
- ▶ carro
- ▶ casa
- ▶ é
- ▶ linda
- ▶ marcelo

Cada documento será modelado como um vetor de pesos binários:

$$\begin{array}{c} \text{amarela} \\ \text{azul} \\ \text{campo} \\ \text{carro} \\ \text{casa} \\ \text{é} \\ \text{linda} \\ \text{marcelo} \end{array}$$
$$\overline{D1} = (1, 1, 1, 0, 1, 1, 1, 0)$$
$$\overline{D2} = (0, 1, \underbrace{0}_{\substack{\text{Peso do termo} \\ \text{"campo" em D2}}}, 1, 0, 1, 0, 1)$$

Modelo Booleano

- Em situações reais, com uma base razoável, cada vetor de pesos pode englobar muito mais zeros do que um's;
- Para otimizar o armazenamento desses vetores, pode ser utilizada alguma estrutura de dados que considere esparsidade:
 - É possível, por exemplo, armazenar apenas os índices dos vetores onde o peso 1 aparece.
 - Pode-se usar uma estrutura eficiente como tabela *hash* ou árvore para armazenar esses índices.

Modelo Booleano

- O modelo booleano suporta o uso de operadores lógicos na escrita de consultas;

Modelo Booleano

- O modelo booleano suporta o uso de operadores lógicos na escrita de consultas;
- Por exemplo, considerando uma consulta q envolvendo os termos ka , kb e kc :

$$q = ka \wedge (kb \vee \neg kc)$$

- Como um documento pode satisfazer à consulta acima?

Modelo Booleano

- O modelo suporta o uso de operadores lógicos na escrita de consultas;
- Ex: Seja q uma consulta envolvendo os termos ka , kb e kc :

$$q = ka \wedge (kb \vee \neg kc)$$

- Como um documento pode satisfazer à consulta acima?
 - Se ka , kb e kc estiverem presentes, OU
 - Se ka e kb estiverem presentes, OU
 - Se ka estiver presente e kc não estiver presente.

Modelo Booleano

- O modelo suporta o uso de operadores lógicos na escrita de consultas;
- Ex: Seja q uma consulta envolvendo os termos ka , kb e kc :

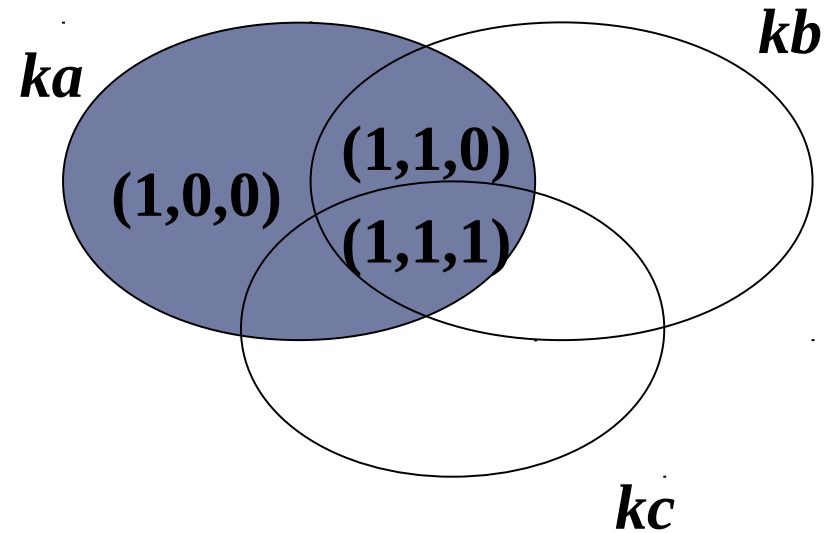
$$q = ka \wedge (kb \vee \neg kc)$$

- Como um documento pode satisfazer à consulta acima?
 - Se ka , kb e kc estiverem presentes $(1,1,1)$, OU
 - Se ka e kb estiverem presentes $(1,1,0)$, OU
 - Se ka estiver presente e kc não estiver presente $(1,1,0)$.

Modelo Booleano

$$q = ka \wedge (kb \vee \neg kc)$$

Esta consulta pode ser representada na forma normal disjuntiva da seguinte forma:



Modelo Booleano

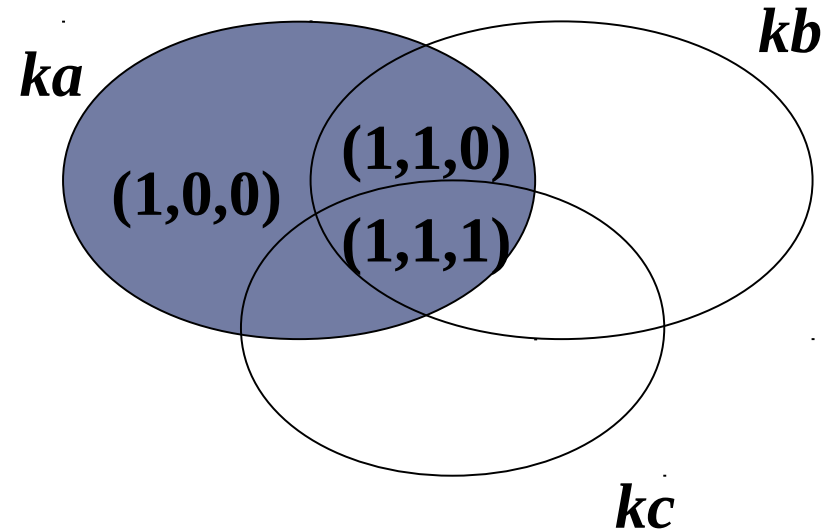
$$q = ka \wedge (kb \vee \neg kc)$$

Esta consulta pode ser representada na forma normal disjuntiva da seguinte forma:

$$q = (ka \wedge kb) \vee (ka \wedge \neg kc)$$

Assim:

$$q = (1,1,1) \vee (1,1,0) \vee (1,0,0)$$



Modelo Booleano

- Para que um documento seja classificado como relevante, basta que seu vetor de pesos case com o de alguma das disjunções da FND da consulta.
- Não há ranqueamento entre os documentos. Cada documento é apenas classificado como relevante ou não relevante.

Exemplo 2

D1

A casa de campo é linda, azul
e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

azul

Exemplo 2

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

azul

$$\begin{array}{cccccccc} & \text{amarela} & \text{azul} & \text{campo} & \text{carro} & \text{casa} & \text{é} & \text{linda} & \text{marcelo} \\ \overline{D1} & = & (1, & 1, & 1, & 0, & 1, & 1, & 1, & 0) \\ \overline{D2} & = & (0, & 1, & 0, & 1, & 0, & 1, & 0, & 1) \end{array}$$

$$\overline{q} = (_, 1, _, _, _, _, _, _, _)$$

Exemplo 2

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

azul

	amarela	azul	campo	carro	casa	é	linda	marcelo
$\overline{D1} =$	(1,	1,	1,	0,	1,	1,	1,	0)
$\overline{D2} =$	(0,	1,	0,	1,	0,	1,	0,	1)
$\overline{q} =$	(_,	1,	_,	_,	_,	_,	_,	_)

Resposta: D1 e D2

Exemplo 3

D1

A casa de campo é linda, azul
e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

campo \wedge é

Exemplo 3

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

campo \wedge é

$\overline{D1} = (1, 1, 1, 0, 1, 1, 1, 0)$
 $\overline{D2} = (0, 1, 0, 1, 0, 1, 0, 1)$

$\overline{q} = (_, _, 1, _, _, 1, _, _)$

Exemplo 3

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

campo \wedge é

	amarela	azul	campo	carro	casa	é	linda	marcelo
$\overline{D1}$	= (1, 1,	1, 0,	1,	1,	1,	1,	0)	
$\overline{D2}$	= (0, 1,	0,	1,	0,	1,	0,	1)	
\overline{q}	= (—, —,	1,	—,	—,	1,	—,	—)	

Resposta: D1

Exemplo 4

D1

A casa de campo é linda, azul
e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

azul \wedge \neg linda

Exemplo 4

D1

A casa de campo é linda, azul
e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

azul \wedge \neg linda

$$\begin{array}{l} \overline{D1} = (1, 1, 1, 0, 1, 1, 1, 0) \\ \overline{D2} = (0, 1, 0, 1, 0, 1, 0, 1) \\ \overline{q} = (_, 1, _, _, _, _, 0, _) \end{array}$$

*amarela
azul
campo
carro
casa é linda
marcelo*

Exemplo 4

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

azul \wedge \neg linda

$\overline{D1} = (1, 1, 1, 0, 1, 1, 1, 0)$
 $\overline{D2} = (0, 1, 0, 1, 0, 1, 0, 1)$
 $\overline{q} = (1, 1, 1, 1, 1, 1, 0, 1)$

amarela azul campo carro casa é linda marcelo

Resposta: D2

Exemplo 5

D1

A casa de campo é linda, azul
e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

$\text{azul} \wedge (\neg \text{casa} \vee \text{linda})$

Exemplo 5

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

$\text{azul} \wedge (\neg \text{casa} \vee \text{linda})$

Passamos a consulta à
forma normal disjuntiva

Exemplo 5

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

$\text{azul} \wedge (\neg \text{casa} \vee \text{linda})$

q'

$\text{azul} \wedge \neg \text{casa}$

q''

$\text{azul} \wedge \text{linda}$

Exemplo 5

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

$\text{azul} \wedge (\neg \text{casa} \vee \text{linda})$

q'

$\text{azul} \wedge \neg \text{casa}$

q''

$\text{azul} \wedge \text{linda}$

$\overline{\text{D1}} = (1, 1, 1, 0, 1, 1, 1, 0)$
 $\overline{\text{D2}} = (0, 1, 0, 1, 0, 1, 0, 1)$

$\overline{\text{q'}} = (_, 1, _, _, 0, _, _, _)$

Exemplo 5

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

$\text{azul} \wedge (\neg \text{casa} \vee \text{linda})$

q'

$\text{azul} \wedge \neg \text{casa}$

q''

$\text{azul} \wedge \text{linda}$

$\overline{\text{D1}} = (1, \text{amarela}, 1, \text{azul}, 0, 1, \text{campo}, 1, \text{carro}, 1, \text{casa}, 1, \text{é}, 1, \text{linda}, 0, \text{marcelo})$
 $\overline{\text{D2}} = (0, 1, 0, 1, 0, 1, 0, 1)$
 $\overline{\text{q'}} = (_, 1, _, _, 0, _, _, _)$

Resposta q': D2

Exemplo 5

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

$\text{azul} \wedge (\neg \text{casa} \vee \text{linda})$

q'

$\text{azul} \wedge \neg \text{casa}$

q''

$\text{azul} \wedge \text{linda}$

$\overline{D1} = (1, 1, 1, 0, 1, 1, 1, 0)$
 $\overline{D2} = (0, 1, 0, 1, 0, 1, 0, 1)$
 $\overline{q'} = (_, 1, _, _, 0, _, _, _)$

Resposta q': D2

$\overline{D1} = (1, 1, 1, 0, 1, 1, 1, 0)$
 $\overline{D2} = (0, 1, 0, 1, 0, 1, 0, 1)$
 $\overline{q''} = (_, 1, _, _, _, _, 1, _)$

Exemplo 5

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

$\text{azul} \wedge (\neg \text{casa} \vee \text{linda})$

q'

$\text{azul} \wedge \neg \text{casa}$

q''

$\text{azul} \wedge \text{linda}$

$\overline{\text{D1}} = (1, \text{amarela}, 1, \text{azul}, 0, \text{campo}, 1, \text{carro}, 1, \text{casa}, 1, \text{é}, 1, \text{linda}, 0, \text{marcelo})$
 $\overline{\text{D2}} = (0, 1, 0, 1, 0, 1, 0, 1)$
 $\overline{\text{q'}} = (_, 1, _, _, 0, _, _, _)$

Resposta q': D2

$\overline{\text{D1}} = (1, \text{amarela}, 1, \text{azul}, 0, \text{campo}, 1, \text{carro}, 1, \text{casa}, 1, \text{é}, 1, \text{linda}, 0, \text{marcelo})$
 $\overline{\text{D2}} = (0, 1, 0, 1, 0, 1, 0, 1)$
 $\overline{\text{q''}} = (_, 1, _, _, _, _, 1, _)$

Resposta q'': D1

Exemplo 5

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

Consulta q

$\text{azul} \wedge (\neg \text{casa} \vee \text{linda})$

q'

$\text{azul} \wedge \neg \text{casa}$

q''

$\text{azul} \wedge \text{linda}$

$\overline{\text{D1}} = (1, \text{amarela}, 1, \text{azul}, 0, \text{campo}, 1, \text{carro}, 1, \text{casa}, 1, \text{é}, 1, \text{linda}, 0, \text{marcelo})$
 $\overline{\text{D2}} = (0, 1, 0, 1, 0, 1, 0, 1)$
 $\overline{\text{q'}} = (_, 1, _, _, 0, _, _, _)$

Resposta q': D2

Resposta q'': D1

$\overline{\text{D1}} = (1, \text{amarela}, 1, \text{azul}, 0, \text{campo}, 1, \text{carro}, 1, \text{casa}, 1, \text{é}, 1, \text{linda}, 0, \text{marcelo})$
 $\overline{\text{D2}} = (0, 1, 0, 1, 0, 1, 0, 1)$
 $\overline{\text{q''}} = (_, 1, _, _, _, _, _, 1, _)$

Resposta final: D1, D2

Vantagens do Modelo Booleano

- Fácil compreensão e implementação;
- Semântica precisa: resultados previsíveis
- Suporte nativo ao operador *NOT*:
 - Muitas vezes, sistemas de RI baseados em outros modelos implementam a operação de MENOS em vez de *NOT* devido a dificuldade de fornecer suporte a esse operador.

Desvantagens do Modelo Booleano

- A semântica precisa faz com que o casamento entre documento e consulta precise ser exato para um doc ser considerado relevante;
 - Isto torna este modelo mais próximo à recuperação de dados que à recuperação de informação.
- O critério binário de decisão pode ser ruim para a recuperação: Se uma consulta incluir 10 termos e um documento d contiver apenas 9, a simples falta de um termo fará com que ele não entre no resultado;
 - Talvez d ainda seja relevante ao usuário, sobretudo se não houver nenhum doc que case exatamente com a consulta.

Desvantagens do Modelo Booleano

- Usuários podem se confundir com uso de operadores lógicos;
- Consultas booleanas formuladas pelos usuários frequentemente são simplistas;
- Em consequência, a quantidade de resultados retornado pode ser muito pequena ou muito grande;
- A falta de ranqueamento torna o modelo não prático;