

QCon
RIO DE JANEIRO

The background image shows a waterfall with two distinct streams of water falling from a dark, mossy rock cliff into a dark pool below. The upper part of the image is dominated by the green moss and the rocky texture of the cliff.

Algoritmos de Recomendação

Felipe Cruz

@felipejcruz

github.com/felipecruz

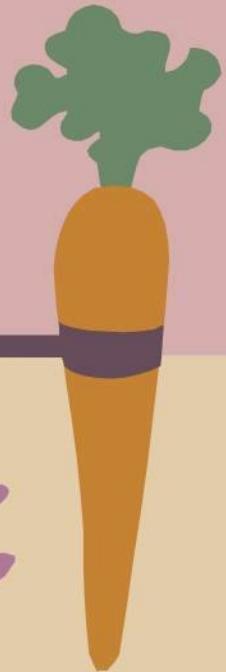


Problema

Algoritmos

CF & RBMs

Motivação?



**relevância
eficiência
personalização**



Problema

items

	w	x	y	z
a	4	3	?	?
b	?	4	?	1
c	?	?	3	4
d	2	4	?	?



Pacote Punta Cana

Punta Cana, DOM

Aéreo de 5 Cidades + Resort c/ Comidas e Bebidas à Vontade

1 Pessoa(s)

5 diárias

De R\$ 3039

Por apenas

R\$ 1979

**VER
PACOTE**





Quantos
usuários?

A photograph showing multiple rows of flip-flops displayed on shelves in a store. The flip-flops are arranged in a grid pattern, with various colors including pink, yellow, blue, green, and purple. Price tags are visible on the shelves, indicating prices like 14,99, 19,99, 22,99, and 22,90.

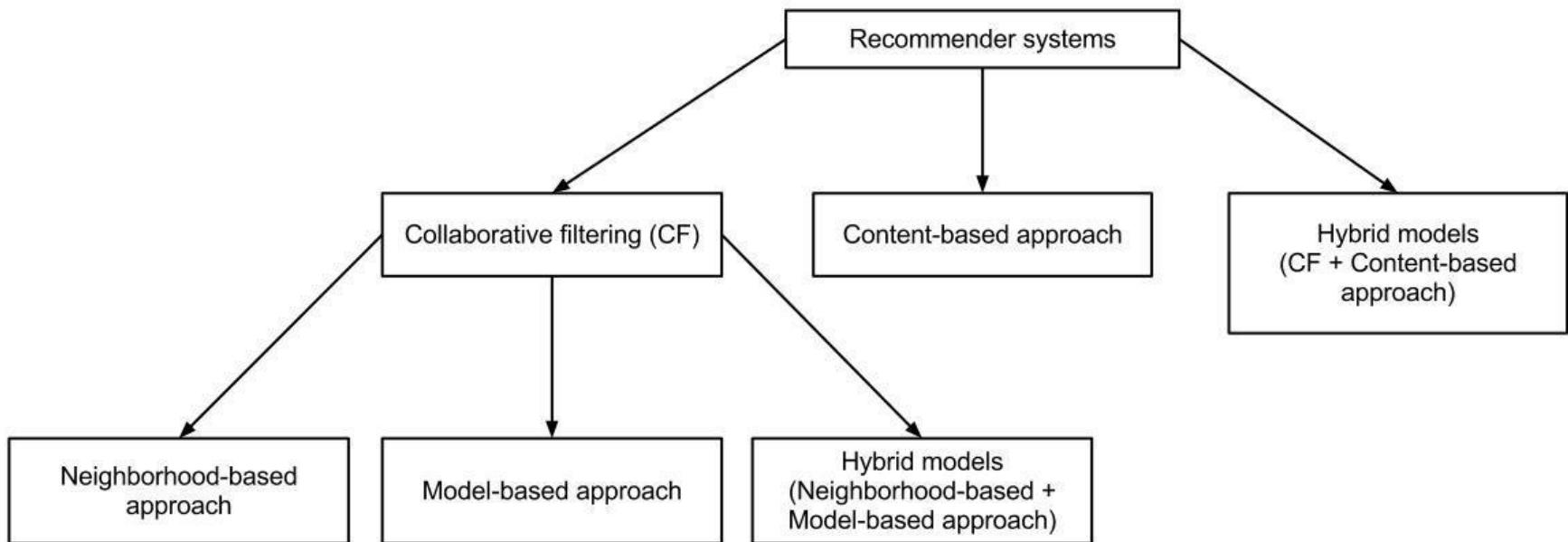
Quantos
produtos?

Recomendação não
serve apenas para
"Big Data", só que..

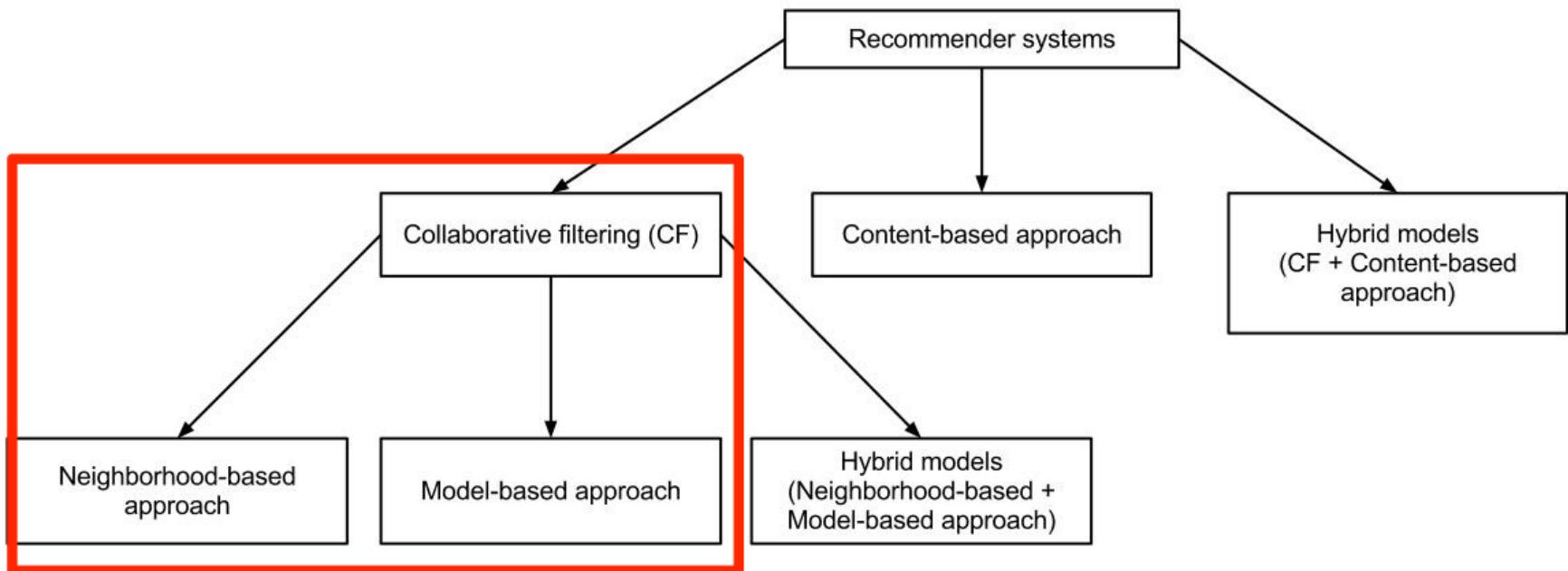
**quanto maior for a
matriz, mais
recursos são
necessários**

Problema **Algoritmos** CF & RBMs

Sistemas de Recomendação



Sistemas de Recomendação



Filtragem Colaborativa

A large, intricate geometric pattern known as the "Flower of Life" is centered behind the text. It consists of multiple overlapping circles of the same size, creating a complex arrangement of intersecting lines and points. This pattern is composed of 13 overlapping circles, which together form 7 larger equilateral triangles.

Similaridade

(neighborhood based)

Procedimento de treino

Para cada usuário



Descobre usuários que compraram produtos em comum



Procedimento de treino



Procedimento de treino



agora podemos calcular a similaridade

f é a função de similaridade

The diagram illustrates a function f mapping from two people to a gift box. The mapping is shown as follows:

- A person on the left is mapped to a wrapped gift box via a path labeled 4.5.
- A person on the right is mapped to the same wrapped gift box via a path labeled 4.
- The wrapped gift box is then mapped to a gift box with a percentage tag via a path labeled 3.
- The gift box with a percentage tag is mapped back to the person on the left via a path labeled 3.5.

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

Pearson correlation

$$\sum_{i \in I} (r_{a,i} - \bar{r}_a) (r_{u,i} - \bar{r}_u)$$

$$I = \{ \text{gift icon}, \text{gift icon with percentage tag} \}$$

nota dada pelo usuário a ao item i
menos média das notas do usuário a

$$\sum_{i \in I} (r_{a,i} - \overline{r_a}) (r_{u,i} - \overline{r_u})$$

$$[(4.5 - 4)(4 - 3.5)] + [(3.5 - 4)(3 - 3.5)]$$



$$[0.5 \times 0.5] + [-0.5 \times -0.5] = 0.5$$

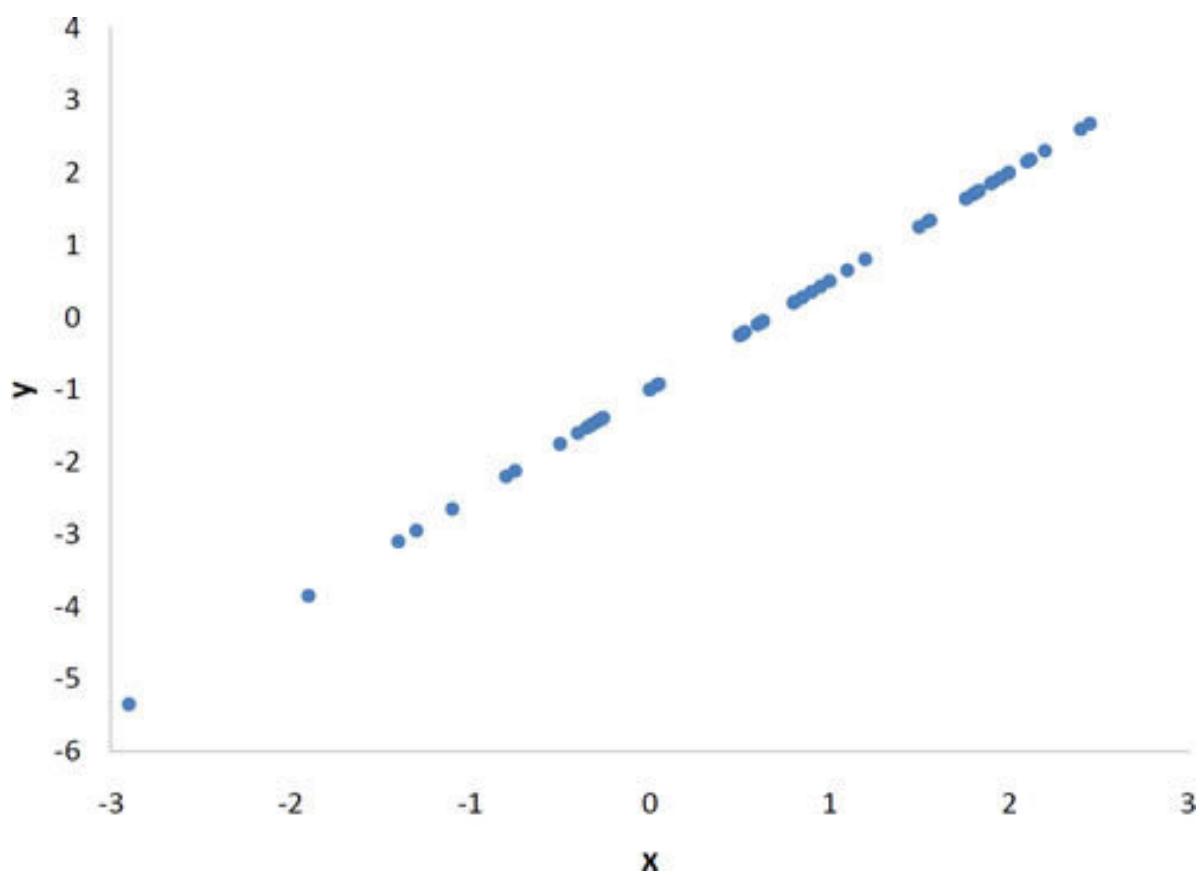
$$\sqrt{\sum_{i \in I} (r_{a,i} - \overline{r_a})^2} \sum_{i \in I} (r_{u,i} - \overline{r_u})^2$$

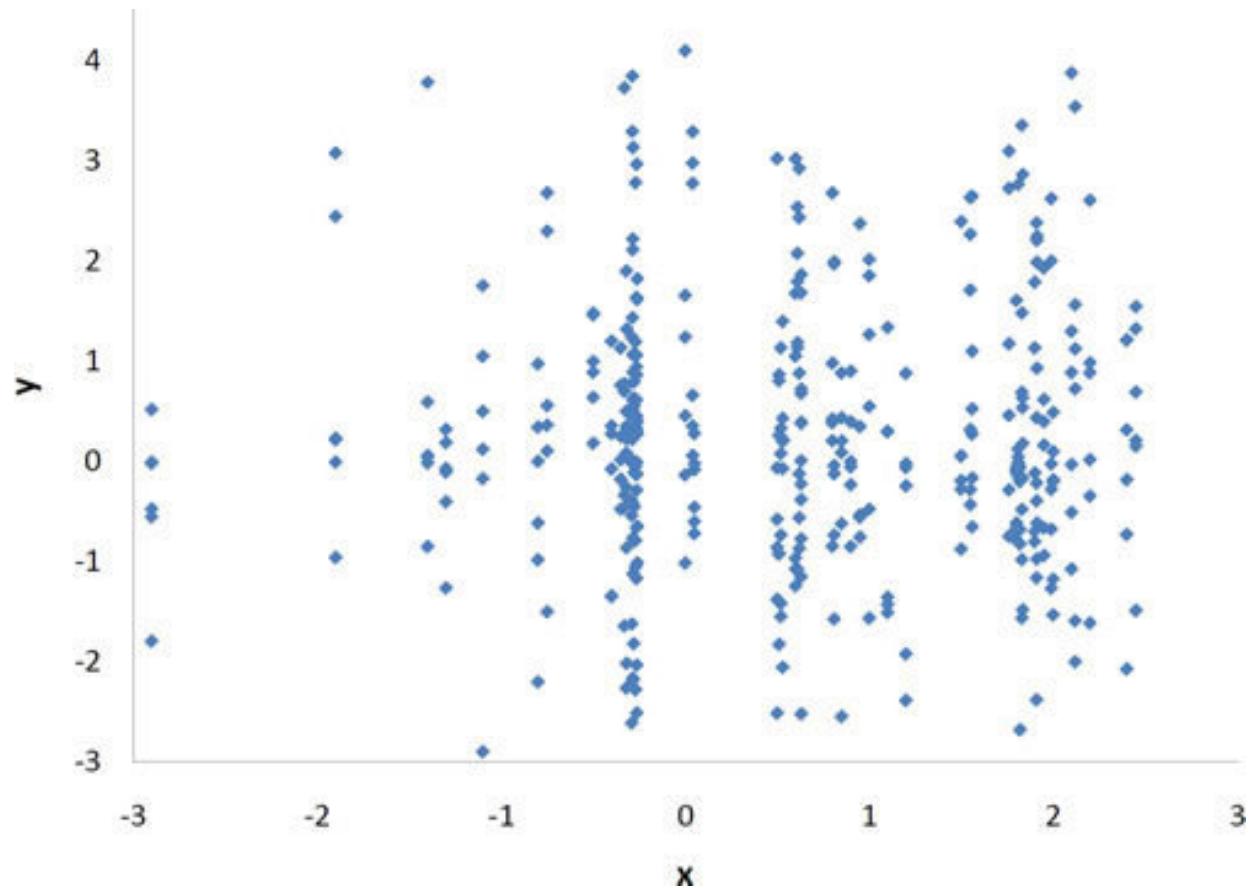
$$[(4.5 - 4)^2 + (3.5 - 4)^2] \times [(4 - 3.5)^2 + (3 - 3.5)^2]$$



$$[(0.5)^2 + (-0.5)^2] \times [(0.5)^2 + (-0.5)^2] = \sqrt{0.5^2} = 0.5$$

$$\frac{0.5}{0.5} = 1$$





Predição?



Predição

$$p_{a,i} = \bar{r_a} + \frac{\sum_{u \in K} (r_{u,i} - \bar{r_u}) \times w_{a,u}}{\sum_{u \in K} w_{a,u}}$$

Predição

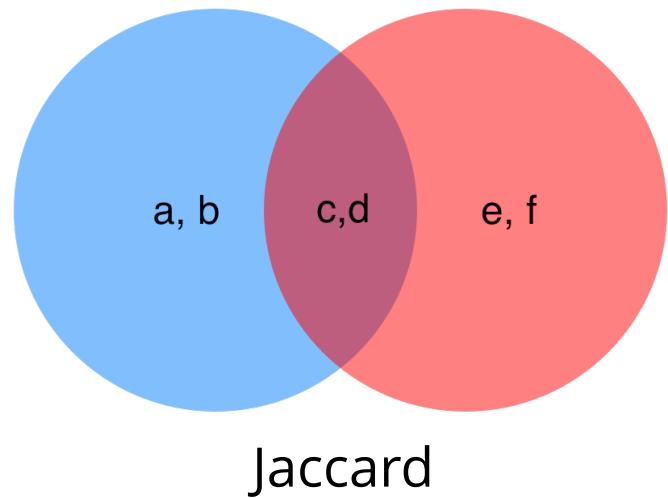
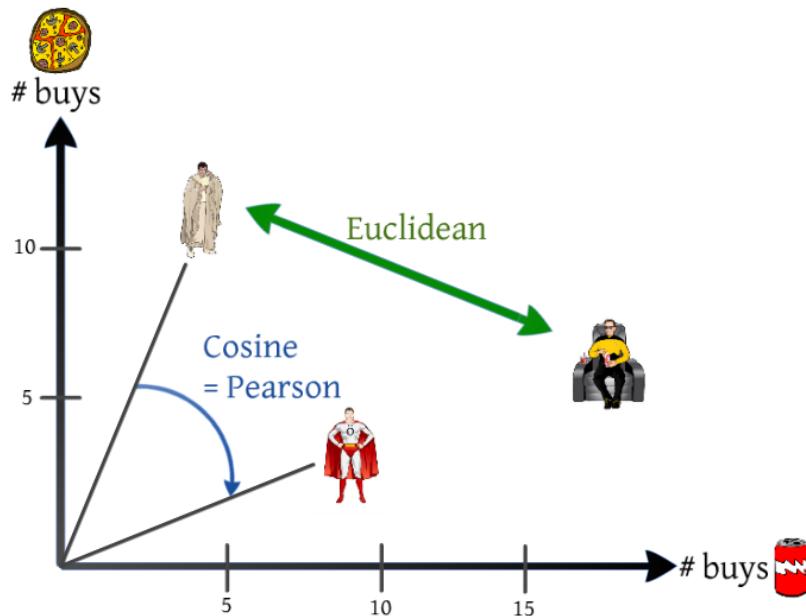
$$p_{a,i} = 4 + \frac{5 - 3.5 \times 1}{1} = 5.5$$



Template method - treino

- for u in usuarios:
 - $\text{similares} = \text{mesmos_produtos}(R, u)$
 - for s in similares:
 - $S[(u,s)] = \text{similaridade}(u, s)$

Outras medidas de similaridade



pausa...

Fator latente



(model based)

**Capturar informação
implícita**

Capturar informação
implícita

#comofaz

items

	w	x	y	z
a	4	3	?	?
b	?	4	?	1
c	?	?	3	4
d	2	4	?	?

Vamos dar um nome aquela matriz:

$$R_{ij} \in \mathbb{N}^{(u \times p)}$$

Model
Based?

MATTER $\int d(uv) = uv = \int u dv + \int v du.$ **WHY** $y'' + \frac{1}{2} \left[\frac{1}{x} + \frac{1}{x - \alpha_2} + \frac{1}{x - \alpha_3} \right] y' + \frac{1}{4} \left[\frac{(\alpha_2^2 + \alpha_3^2)g - p(p+1)x + \kappa x^2}{x(x - \alpha_2)(x - \alpha_3)} \right] y = 0$ **CHAOS** $\frac{d\delta x}{dt} = \begin{bmatrix} f_x(x_0, y_0) & f_y(x_0, y_0) \\ g_x(x_0, y_0) & g_y(x_0, y_0) \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \end{bmatrix}$ **QUANTUM** $K = -\frac{1}{c^2} \operatorname{sech}^4\left(\frac{v}{c}\right).$

LIVE $t = \int dt = \int \frac{dx}{\sqrt{2\lambda + x^2 + \frac{1}{2}x^4}}$ **FIELDS** $D^{1/2} c = c \lim_{\lambda \rightarrow 0} \frac{t^{1/2} \Gamma(\lambda + 1)}{\Gamma(\lambda + \frac{1}{2})} = \frac{c}{\sqrt{\pi t}}.$ **PRESENT** $\int_a^b p_j(x) W(x) dx = \sum_{i=1}^n w_i p_j(x_i).$ **HEAVEN** $(1 - \epsilon) F(n) < f(n) < (1 + \epsilon) F(n)$

EVERYTHING WE DO $T = -2 \sqrt{\frac{\alpha}{g}} \int_1^0 \frac{du}{\sqrt{1-u^2}} = 2 \sqrt{\frac{\alpha}{g}} [\sin^{-1} u]_0^1 = \pi \sqrt{\frac{\alpha}{g}},$ **IN MATH** $\Delta(0) =$ **IMPOSSIBLE TO ESCAPE** $y = e^{\pm iz} F_1\left(\frac{1}{2} \mp \frac{1}{2}iA; 1; \mp 2iz\right),$ **LOVE** $A = \int_a^b f(x) dx,$ **WHERE** $g_{kl} \frac{\partial g^{jk}}{\partial x^m} = -g^{jk} \frac{\partial g_{lm}}{\partial x^m}.$

THE MIND $\phi(x) = \sum_{j=0}^{\infty} \alpha_j p_j(x),$ **THINK ABOUT IT** $S = \int_0^b \int_0^y \sqrt{c^2 + u^2} du dv = \frac{1}{2} \theta \left[r \sqrt{c^2 + r^2} + c^2 \ln \left(\frac{r^2 + c^2}{c^2 + r^2} \right) \right]_0^b$ **REALIZED** $\frac{\partial f}{\partial y} y_x = \frac{df}{dt},$ **REALITY** $y_{xx} - \frac{\partial f}{\partial x},$ **LOVE** $\langle v, w \rangle = v^T H w,$ **ALL LIFE** $x(t) = x^*(t) + \sum_{i=1}^n c_i x_i, \quad \sum_{n=1}^{\infty} \frac{1}{n^s}$

CARTESIAN $f(x, y) = y^4 + x^2 y$ **BEING** $t_{12} = \int_{P_1}^{P_2} \frac{\sqrt{1+y'^2}}{\sqrt{2gy}} dx = \int_{P_1}^{P_2} \sqrt{\frac{1+y'^2}{2gy}} dx$ **ALL IS LOST** $\sum_{n=0}^{\infty} \alpha_n [b_n(r_1)]^n = \sum_{n=0}^{\infty} [r_1] \phi_n(r_1),$ **IMAGINATION** $F_1(x),$ **IMAGINATION** $F_2(x),$ **IMAGINATION** $F_3(x)$

ALL IS FOUND $y_{n+1} - y_n = h(q_n + \frac{1}{2} \nabla q_{n-1} + \frac{5}{12} \nabla^2 q_{n-2} + \dots + \frac{251}{720} \nabla^4 q_{n-4} + \frac{95}{208} \nabla^5 q_{n-5} + \dots),$ **IS IT POSSIBLE?** $\int f(x) dx = F(x) + C,$ **PATTERNS** $ds^2 = E dx^2 + 2F dx dv + G dv^2$ **THE SOLUTION** $D^\mu f(t) = D^{\mu_1} [D^{-(n-\mu_1)} f(t)],$ **FEAR** $\lim_{x \rightarrow \infty} f(x) = 0,$

WHO $SO(2) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{bmatrix}$ **DIE** $F(x, y, u, v, w) = x^3/3 - xy^2 + w(x^2 + y^2) - ux - vy$ **DIE** $d^2s^2 = g_{11}(dx^1)^2 + g_{22}(dx^2)^2 + g_{33}(dx^3)^2,$ **OUR EXISTENCE** $\int \frac{dx}{ax + b + cx^2}, \quad \nabla_{e_i} e_j = \sum_k \Gamma_{ij}^k e_k,$ **TO DESCRIBE** $z \mapsto \frac{i-z}{i+z}$

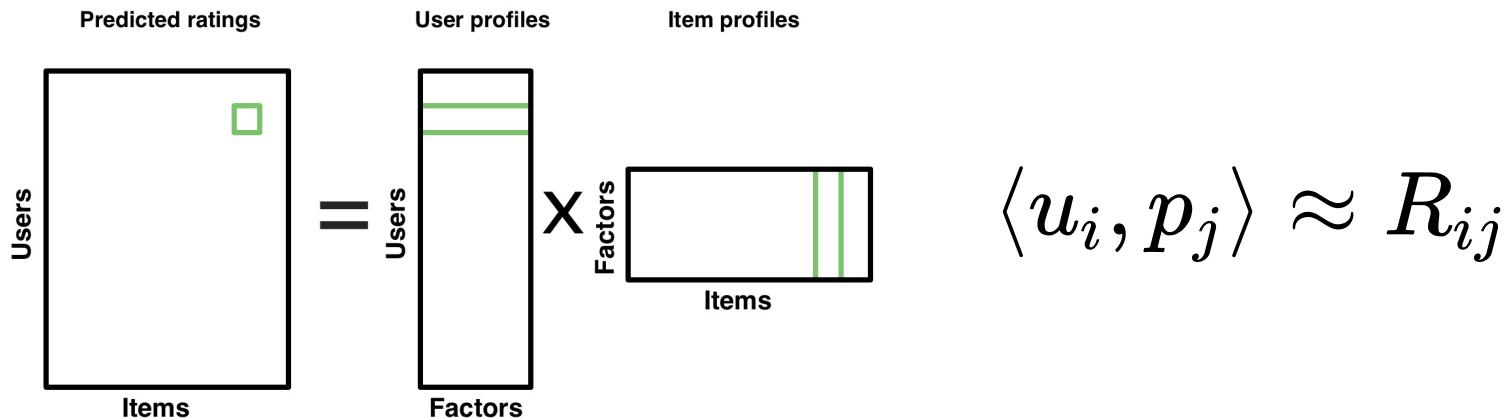
ALL DEATH $\phi(z) = cz + c_0 + c_1 z^{-1} + c_2 z^{-2} + \dots$ **USING ONLY** $\oint_Y f(z) \frac{g'(z)}{g(z)} dz = \sum_n f(\mu_n) - \sum_n f(\nu_n),$ **PUT** $T = \sqrt{\frac{\alpha}{g}} \int_{\theta_0}^{\pi} \frac{\sin(\frac{1}{2}\theta) d\theta}{\sqrt{\cos^2(\frac{1}{2}\theta_0) - \cos^2(\frac{1}{2}\theta)}}.$ **EXPERIMENTAL** $\frac{\partial f}{\partial y} - \frac{d}{dx} \left(\frac{\partial f}{\partial y_x} \right) = 0,$ **STRANGE** $y'' + \frac{k}{x} y' + \epsilon y^r y = 0,$ **ESCAPE**

BEYOND $\frac{df}{dx} - \frac{\partial f}{\partial y_x} y_{xx} - \frac{\partial f}{\partial x} - y_x \frac{d}{dx} \left(\frac{\partial f}{\partial y_x} \right) = 0$ **MISUNDERSTOOD** $_a D_\tau^{-\gamma} f(t) = \frac{1}{\Gamma(\gamma)} \int_a^t (t-\xi)^{\gamma-1} f(\xi) d\xi$ **PUT** $m x_i''(t) = k_c [x_{i-1}(t) - 2x_i(t) + x_{i+1}(t)] - k_p [x_i(t) - i\alpha - t\nu] - f F(f x_i'(t)),$ **PAST** $A = \frac{1}{2c} \sqrt{-q}, \quad \mathcal{U}_n(f) = \int_a^b f(x) K_n(x) dx,$ **PAST**

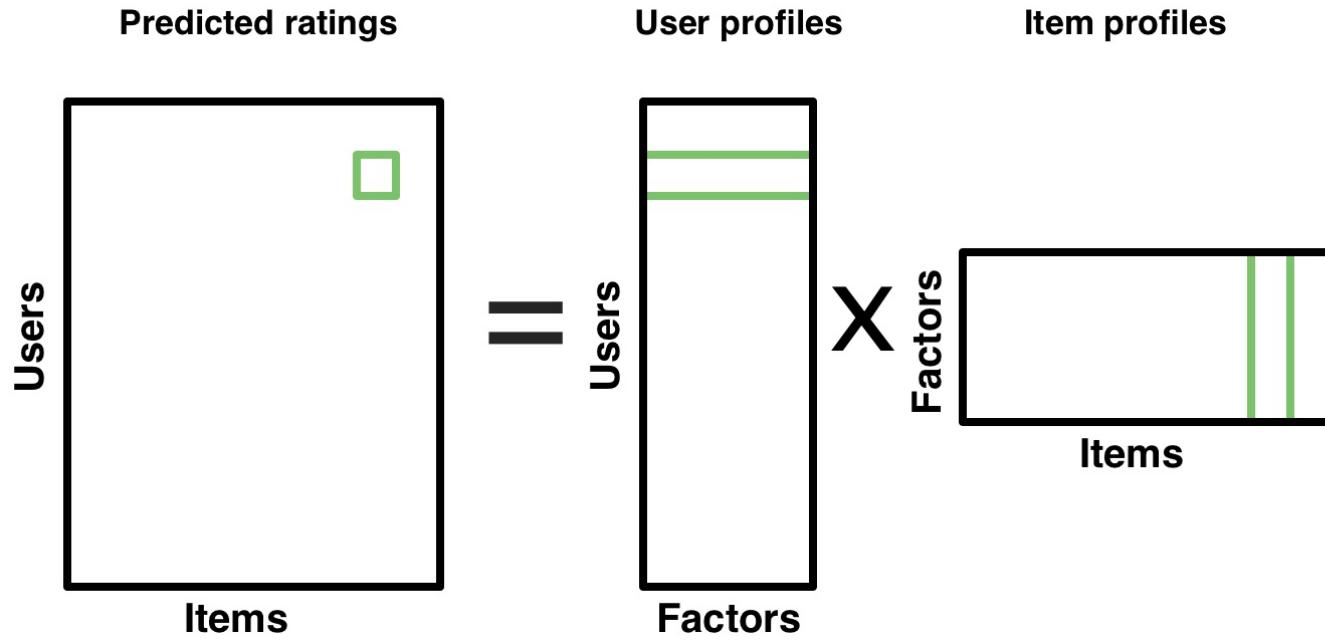
BEYOND $\int_0^t (\xi - \xi)^{\mu-1} d\xi \int_0^{\xi} (\xi - x)^{\nu-1} g(\xi, x) dx = \int_0^t dx \int_x^t (\xi - \xi)^{\mu-1} (\xi - x)^{\nu-1} g(\xi, x) d\xi,$ **NUMBERS AND PATTERNS** $\int_0^1 x^\alpha dx = \frac{1}{\alpha+1}$ **PAIN** $I_2 + [v^2 \lambda]_2^1 = \int_1^2 [v^2 (f_{yy} + \lambda) + 2v\nu (f_{yy} + \lambda) + v^2 f_{yy}] dt,$ **FORGET IT**

h(z_0) = \frac{1}{2\pi} \int_0^{2\pi} h(z_0 + e^{i\theta}) d\theta, **G(r_1, r_2) = \sum_{n=1}^{\infty} \frac{\phi_n(r_1) \phi_n(r_2)}{b^2 - n^2}, **POLAR** $(x, y) + (x', y') = (x + x', y + y')$**

Modelo



Fatorar matrizes



$$\bar{R}_{(u \times p)} = U_{(u \times k)} P_{(k \times p)}$$

Formalização do Problema

$$\min_{U, P} \sum_{i,j \in L} (r_{i,j} - u_i^T p_j)^2 + \lambda (\sum_i n_{u_i} \|u_i\|^2 + \sum_j n_{v_j} \|v_j\|^2)$$



$$\langle u_i, p_j \rangle$$

Formalização do Problema

$$\min_{U, P} \sum_{i,j \in L} (r_{i,j} - u_i^T p_j)^2 + \lambda \left(\sum_i n_{u_i} \|u_i\|^2 + \sum_j n_{v_j} \|v_j\|^2 \right)$$

Regularização contra *overfit*

Minimizar o erro

$$\min_{U, P} \sum_{i,j \in L} (r_{i,j} - u_i^T p_j)^2$$

$$E = \begin{pmatrix} 2 & 4 \\ 3 & 6 \end{pmatrix} - \begin{pmatrix} \langle u_1, p_1 \rangle & \langle u_1, p_2 \rangle \\ \langle u_2, p_1 \rangle & \langle u_2, p_2 \rangle \end{pmatrix}$$

$$R_{ij} \qquad \qquad \overline{R}_{ij}$$

Algoritmo ALS

1. Inicializa U com valores randômicos
2. Fixa U , e para cada linha de P
resolvemos um problema de regressão
linear para achar valores ótimos para P
3. Fixa P , e para cada linha de P
resolvemos um problema de regressão
linear para achar valores ótimos para U



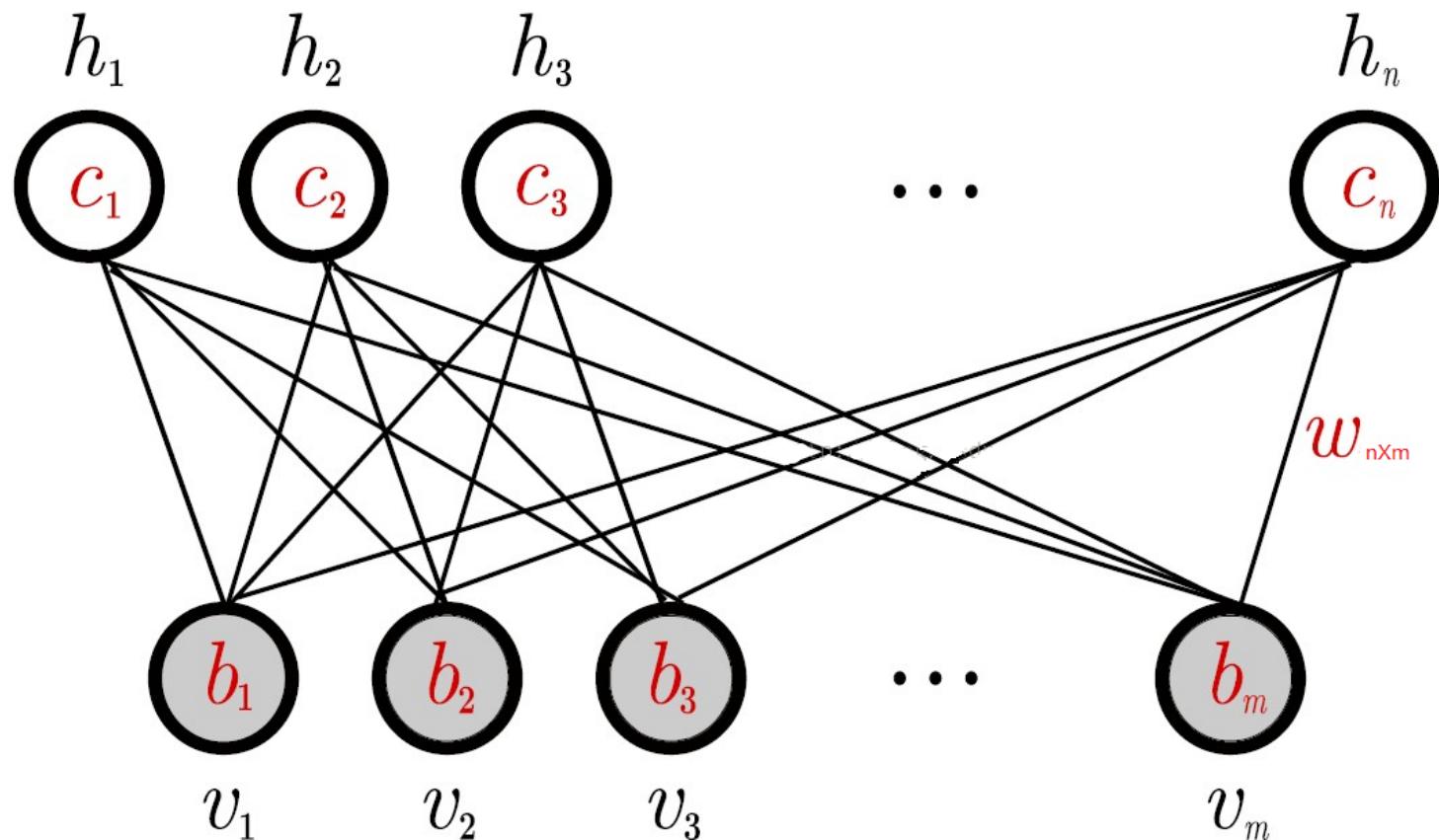
ALS

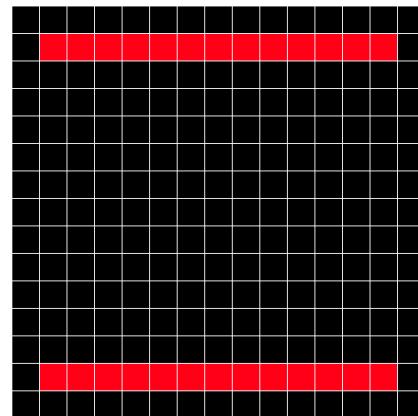
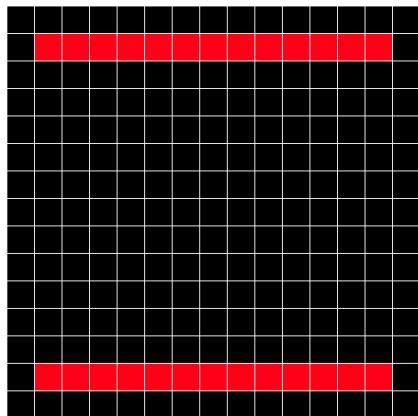
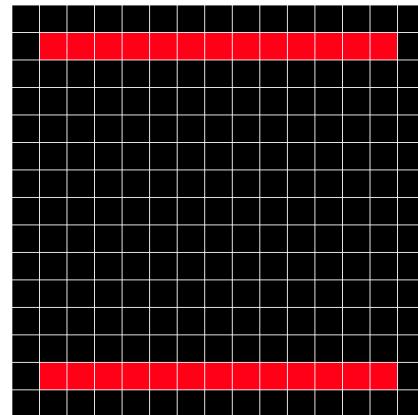
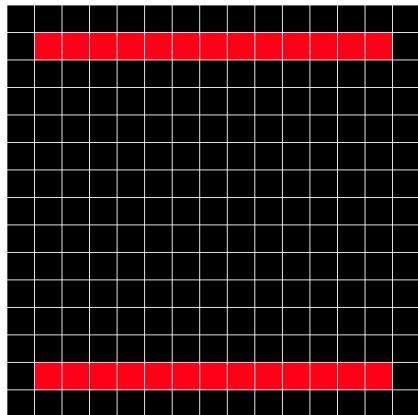
- | Ajusta traseira
- | Ajusta frente
- | Ajusta traseira
- | Ajusta frente

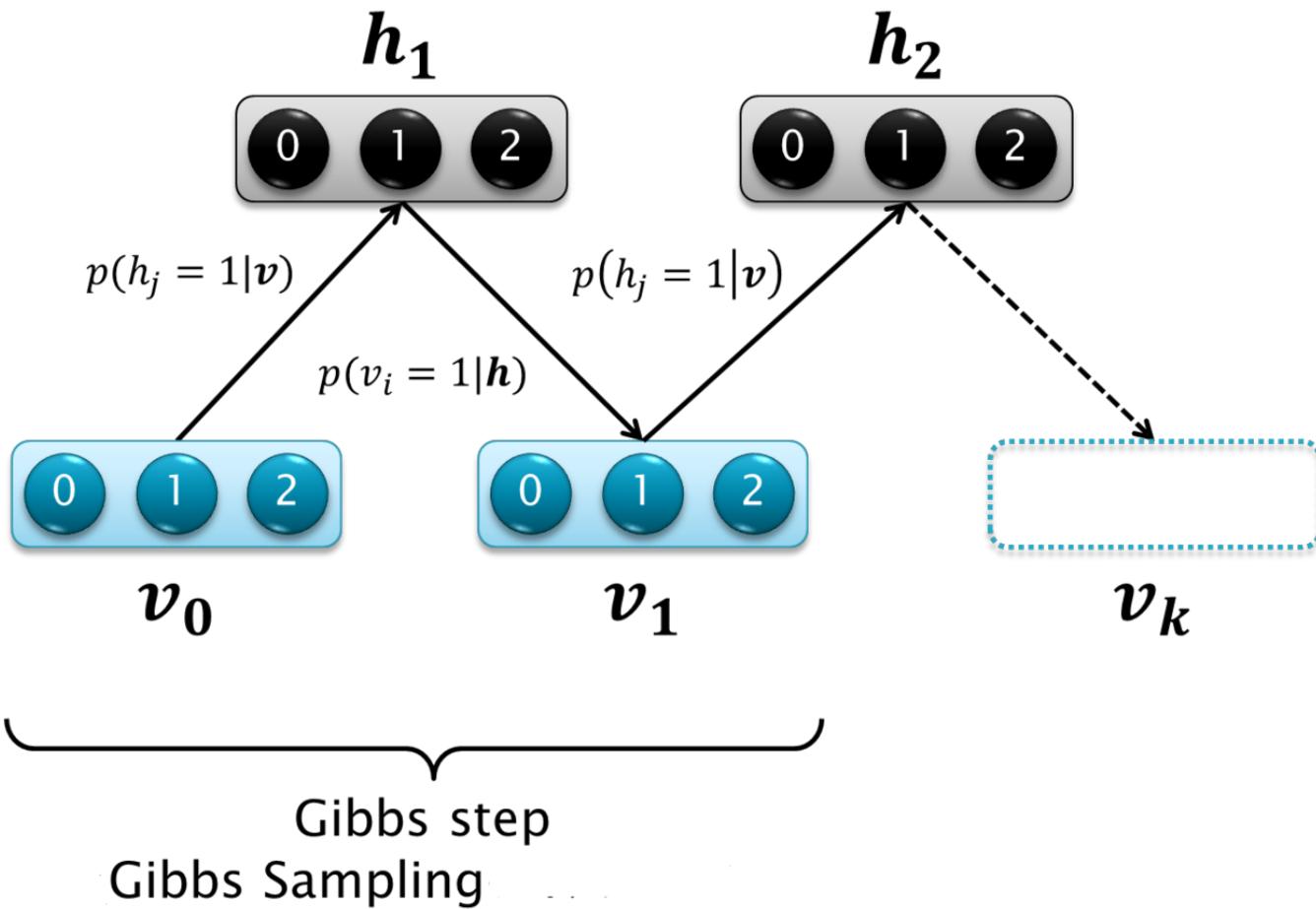
Problema Algoritmos **CF & RBMs**

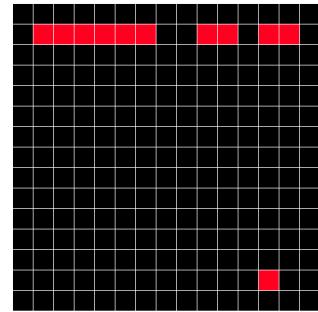
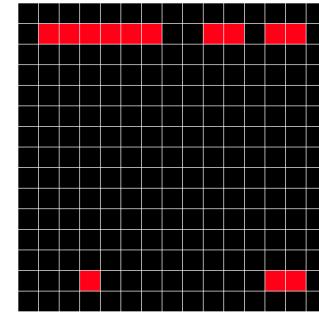
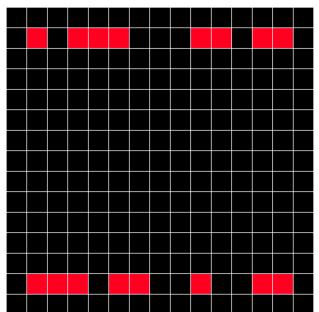
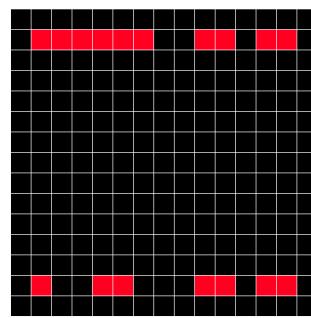
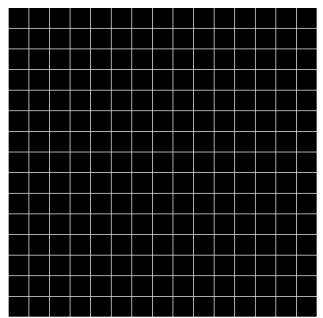
RBM

- Modelo generativo
- Reconhece e completa padrões
- Algoritmo de aprendizado eficiente









RBMs & CF

Modelo

- Cada destino é uma unidade visível
 - 1 se comprou
 - 0 se não comprou

Também é possível usar atributos do usuário no modelo. Ex: tem filhos? Só compra pacotes de luxo?

Predição - input

profile = [0, 0, 0, 1, 0, 0, 0, 1]

(buzios e ilha grande)

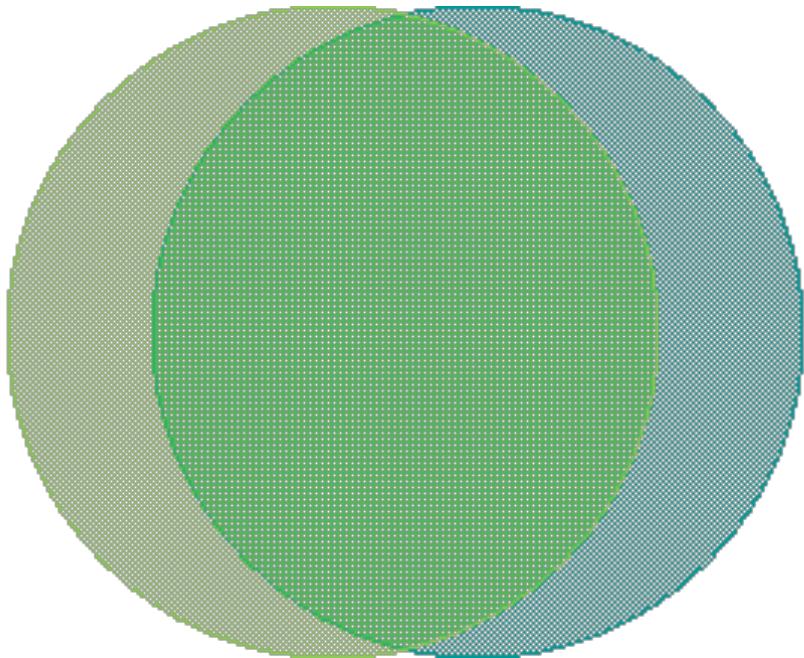
Predição - output

`pred = [0, 1, 0, 1, 0, 0, 1, 1]`

(buzios, ilha grande, **angra, balneário
camboriú**)

RBM

ALS



Na prática

- Funciona* bem com amostras do dataset
- Perfis vazios voltam com destinos **mais populares**
- Predição **não determinística**
- Pode ser usada antes de um modelo de Rating pra diminuir a quantidade de predições necessárias
- Ensemble!!!

<http://www.slideshare.net/akyrola/largescale-recommendation-systems-on-just-a-pc>

Be the first to clip this slide

Large-scale Recommender Systems on *Just a PC*

LSRS 2013 keynote
(RecSys '13 Hong Kong)



Apo Kyrola
Ph.D. candidate @ CMU
<http://www.cs.cmu.edu/~akyrola>
Twitter: @kyrpov

Big Data – small machine

Carnegie Mellon

Clip slide ◀ 1 of 41 ▶

Perguntas?

@felipejcruz

github.com/felipecruz