
FINAL

Data Science

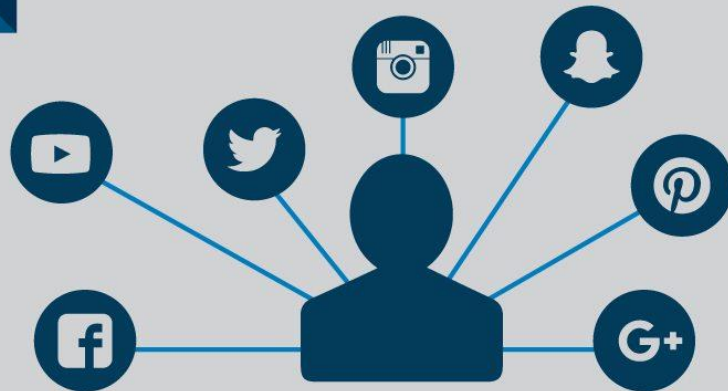
Sistemas de recomendação

Turma TN05

10/02/2020



USER-GENERATED CONTENT



69% OF PEOPLE HAVE SOME FORM OF SOCIAL MEDIA ACCOUNT



CONSUMERS ARE **83%** MORE LIKELY TO TRUST PEERS OVER THE BRAND ITSELF





You know you're suffering from information overload when

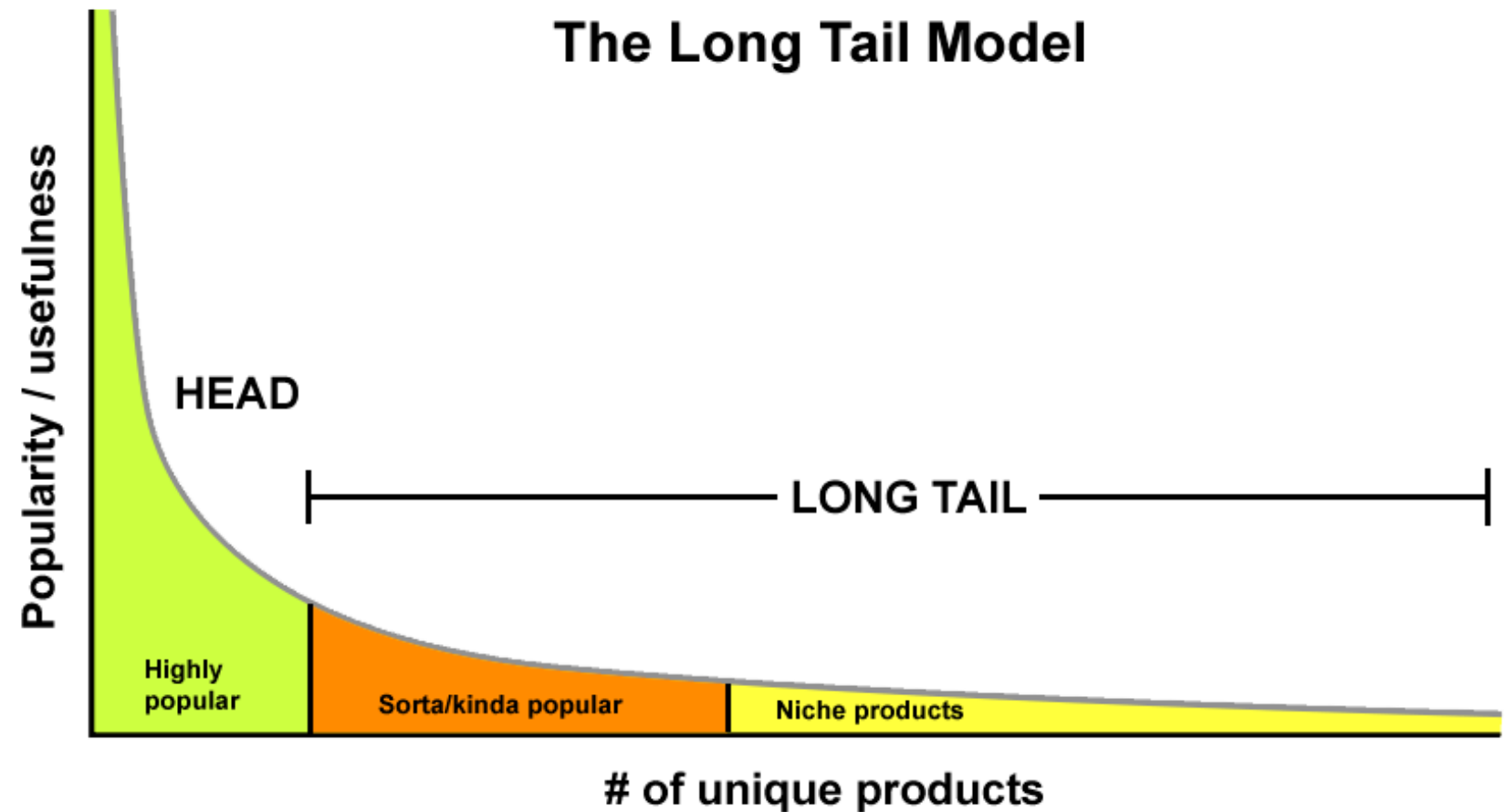
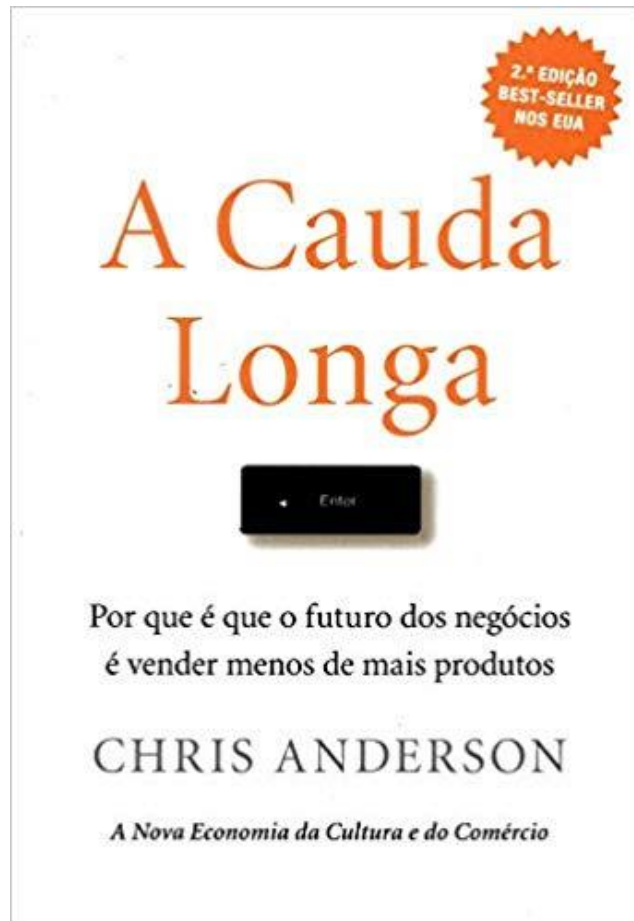


N:ISE TO SIGNAL
Rob Cottingham - socialsignal.com/n2s

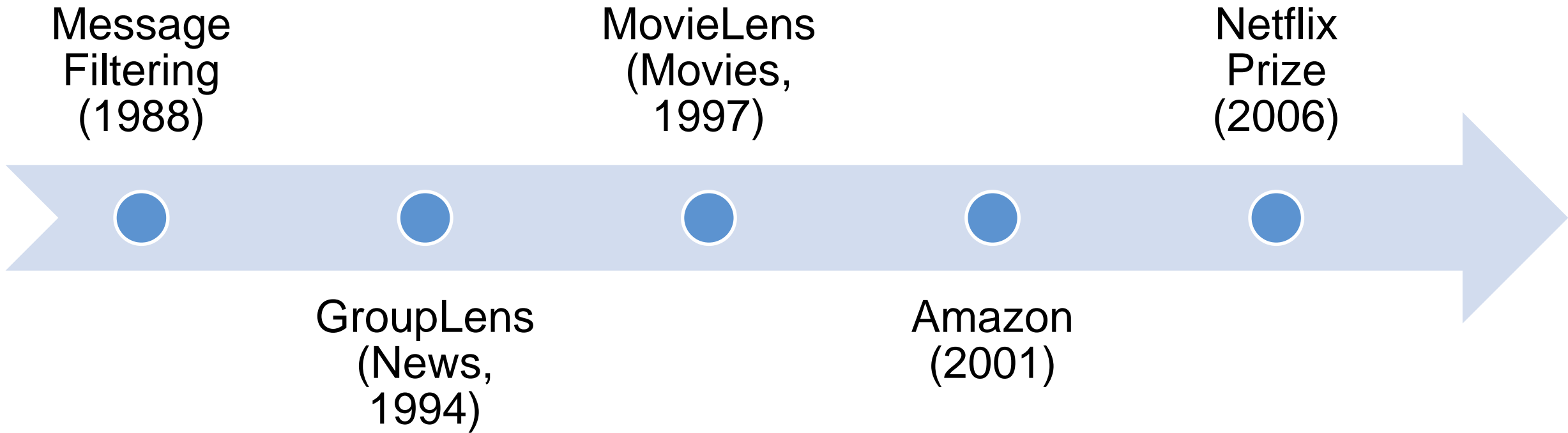


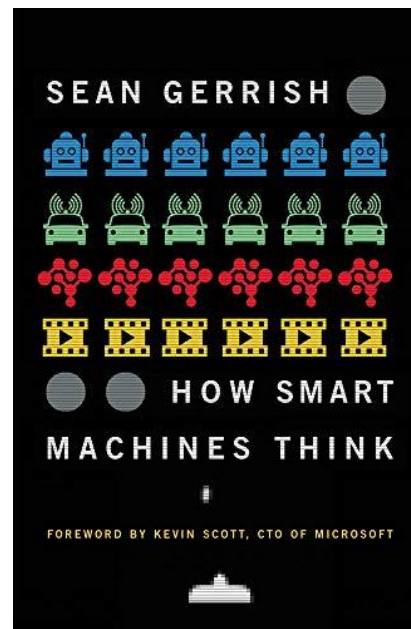
O aumento de opções para o usuário / consumidor não necessariamente implica em uma maior satisfação das pessoas





Sistemas de Recomendação aumentam as vendas, principalmente com a expansão de produtos – quase que infinita - que podem ser ofertados online

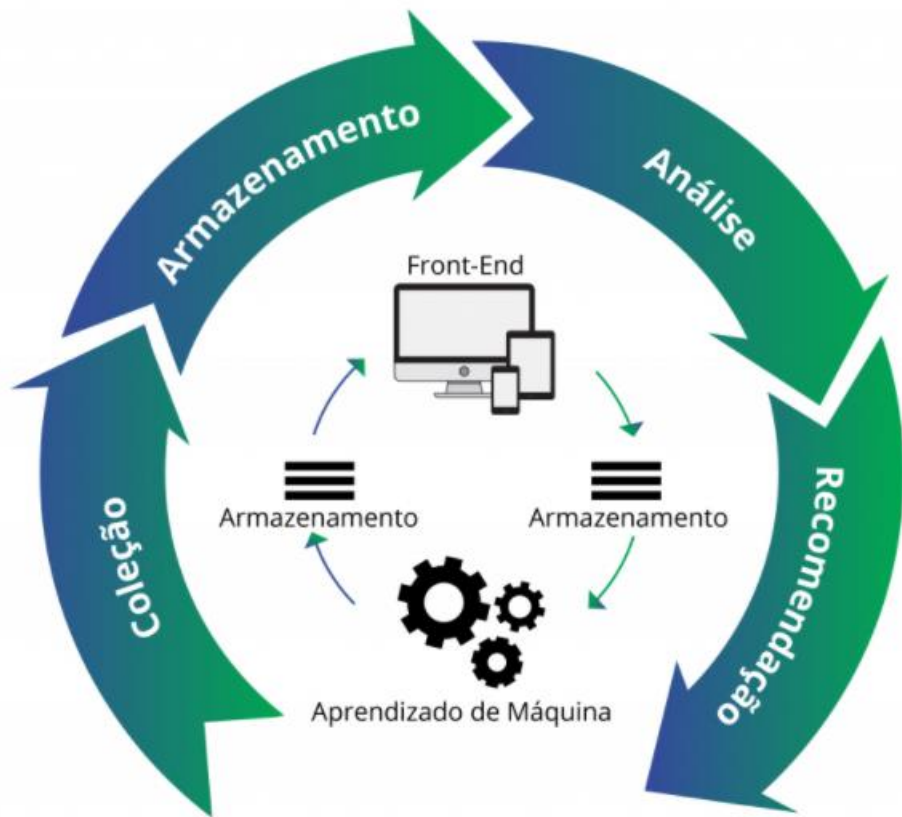




Advances in Ensemble Learning from the Netflix Prize Competition

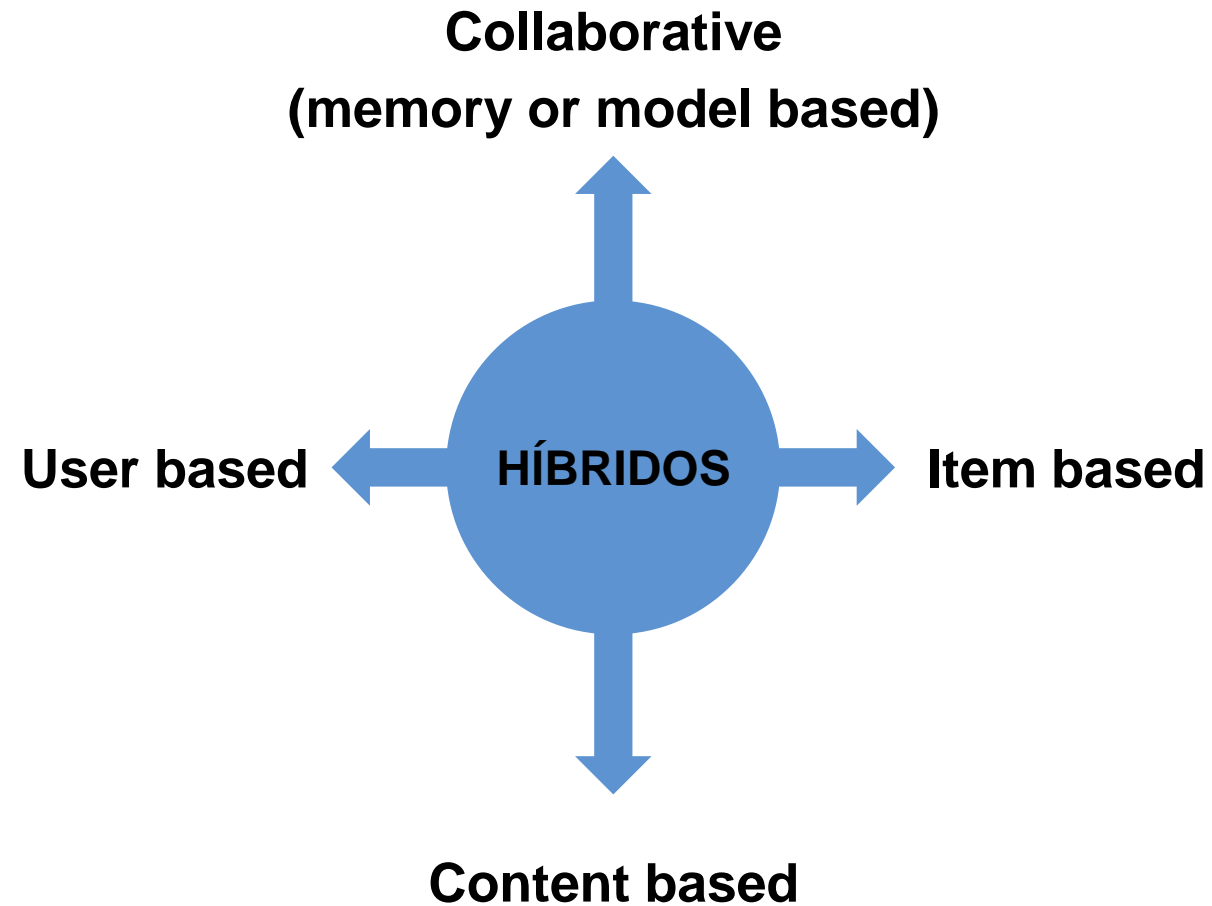
Joe Sill
June 28, 2010

Em 2006 a Netflix – ainda uma companhia que somente enviava DVDs a clientes pelo correio – lançou um desafio com prêmio de 1 milhão de dólares para quem implementasse um algoritmo com performance 10% acima do seu motor de recomendação da época - Cinematch.



<https://www.smarthint.co/como-ia-e-sistemas-de-recomendacao-de-produto-trabalham-juntos/>

A **coleta** de dados pode ser **explícita** (perguntas ao usuário) ou **implícita** (através de seu comportamento)



- Baseam-se nas interações passadas entre usuários e itens, gerando a **Matriz de Interações – Usuários x Itens**.
- Vai sendo enriquecido com as interações de usuários e itens ao longo do tempo.
- Não necessitam de informações específicas dos usuários ou dos itens, portanto podem ser utilizados em diferentes contextos.

	i_1	i_2	i_3	i_4	i_5
u_1	5		4	1	
u_2		3		3	
u_3		2	4	4	1
u_4	4	4	5		
u_5	2	4		5	2

have interacted with a lot of common items in the same way (similar ratings)



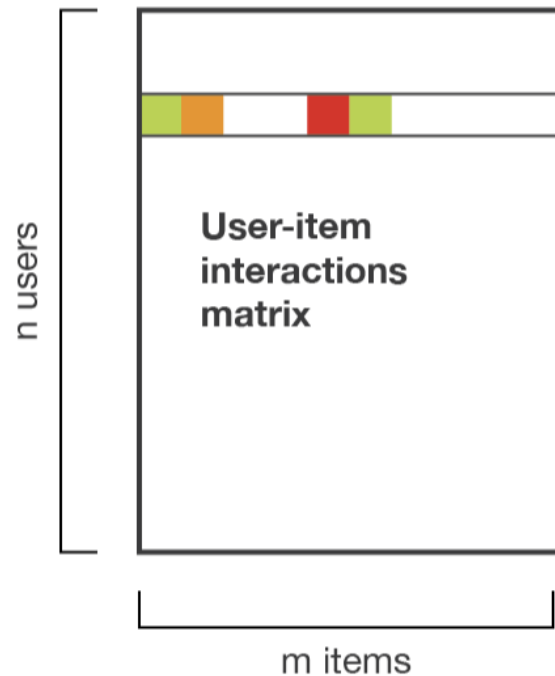
positive interactions



neutral interactions



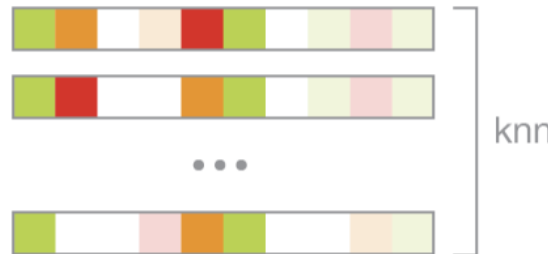
negative interactions



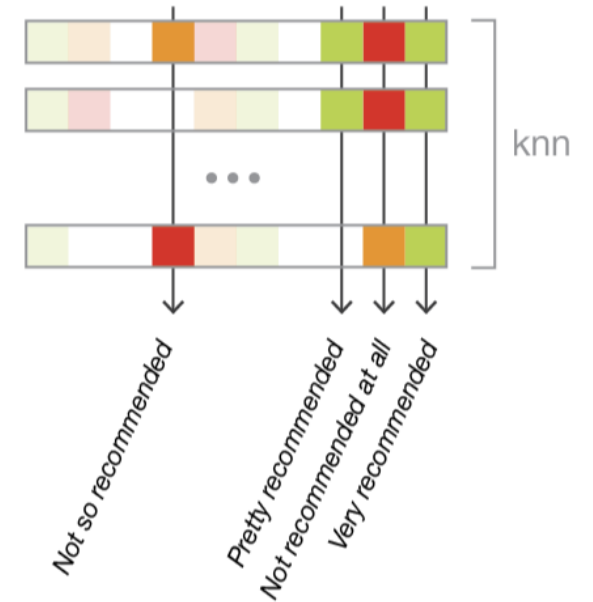
User we want to make a recommendation for is represented by its row in the matrix...



... and we search the K nearest neighbours of this user in the matrix



We can then recommend the most popular items among the K nearest neighbours



have interacted with a lot of common items in the same way (similar ratings)



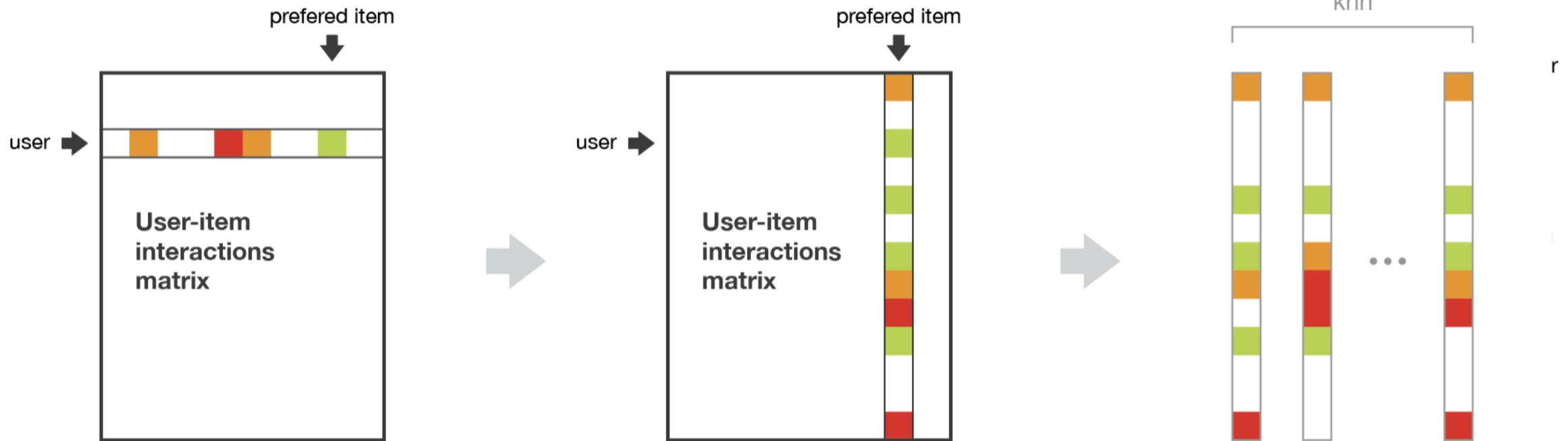
positive interactions



neutral interactions



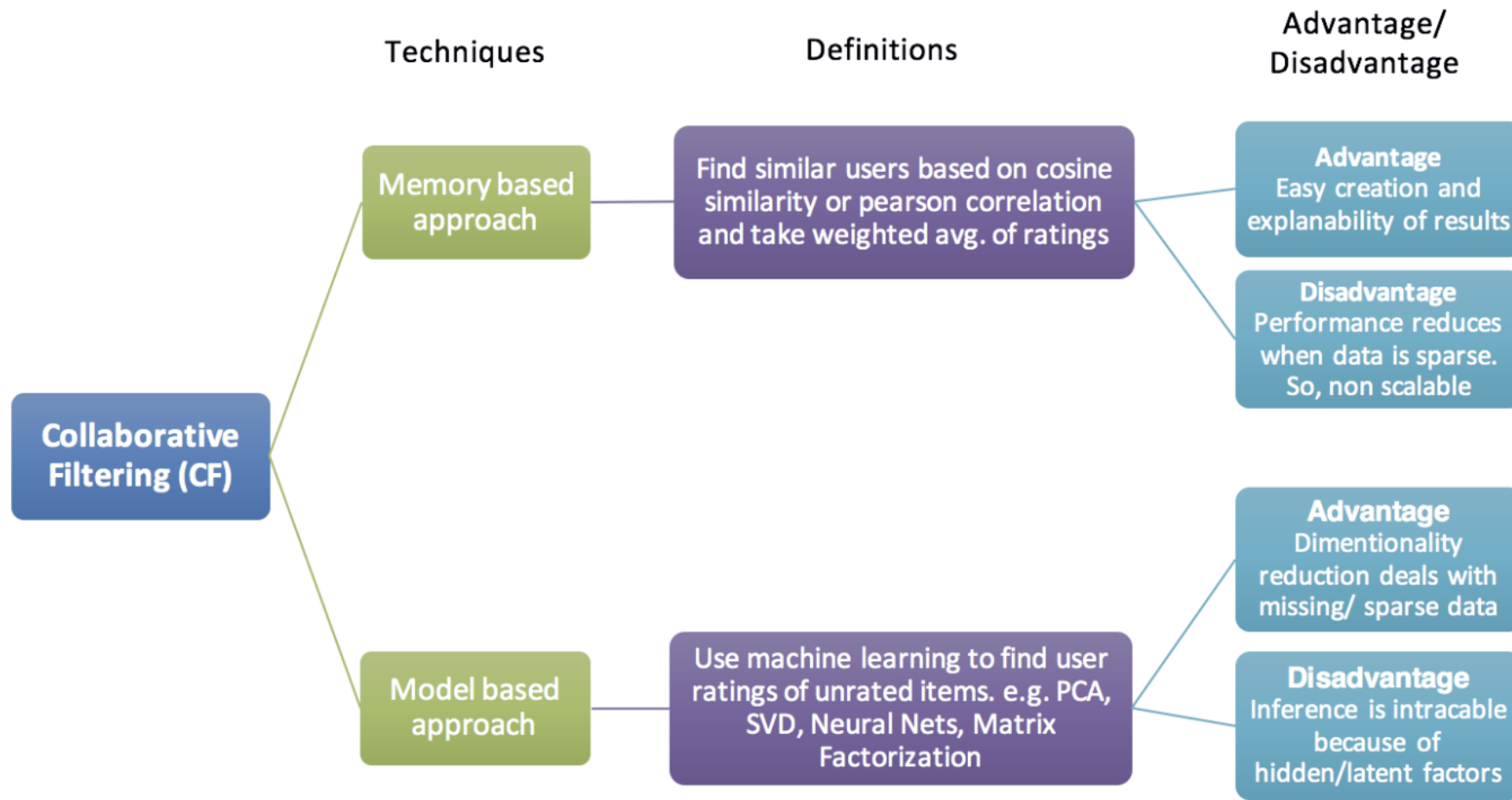
negative interactions



We identify the preferred item of user we want to make recommendation for.

The preferred item is represented by its column in the matrix.

We can search and recommend the K nearest items to this "preferred item"

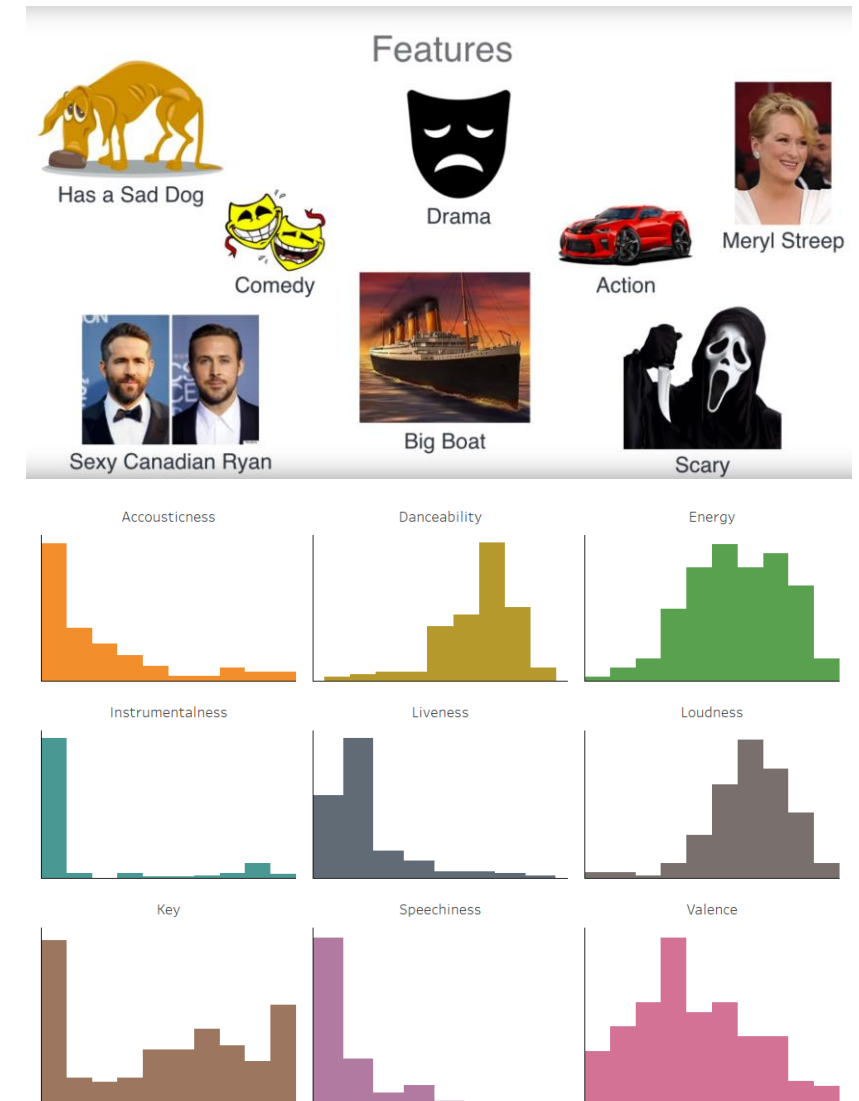


*RecSys colaborativos sofrem **Cold Start** – quando inicialmente não temos informações de itens ou usuários.*

Memory base, podem ter problemas de escalabilidade por tratarem com matrizes esparsas

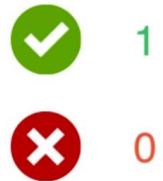


- Usa **características adicionais**, além da Matriz de Interações – Usuários x Itens para alimentar o modelo que faz as recomendações.
- As recomendações são normalmente realizadas por modelos de **Classificação** ou de **Regressão** através dos algoritmos tradicionais (Árvores Decisão, Class. Bayes, Regressão Logística, Deep Learning, etc).



Matrix Factorization

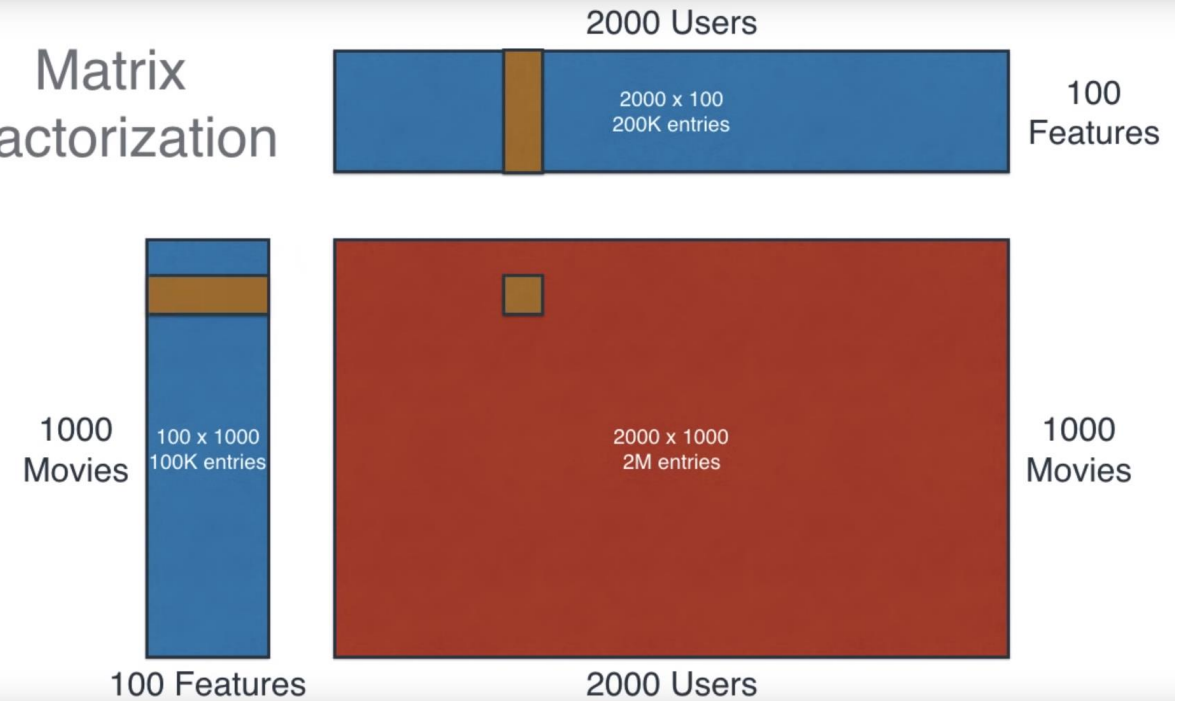
	M1	M2	M3	M4	M5
Comedy	3	1	1	3	1
Action	1	2	4	1	3



	Comedy	Action
A	1	0
B	0	1
C	1	0
D	1	1

	M1	M2	M3	M4	M5
A	3	1	1	3	1
B	1	2	4	1	3
C	3	1	1	3	1
D	4	3	5	4	4

Matrix Factorization



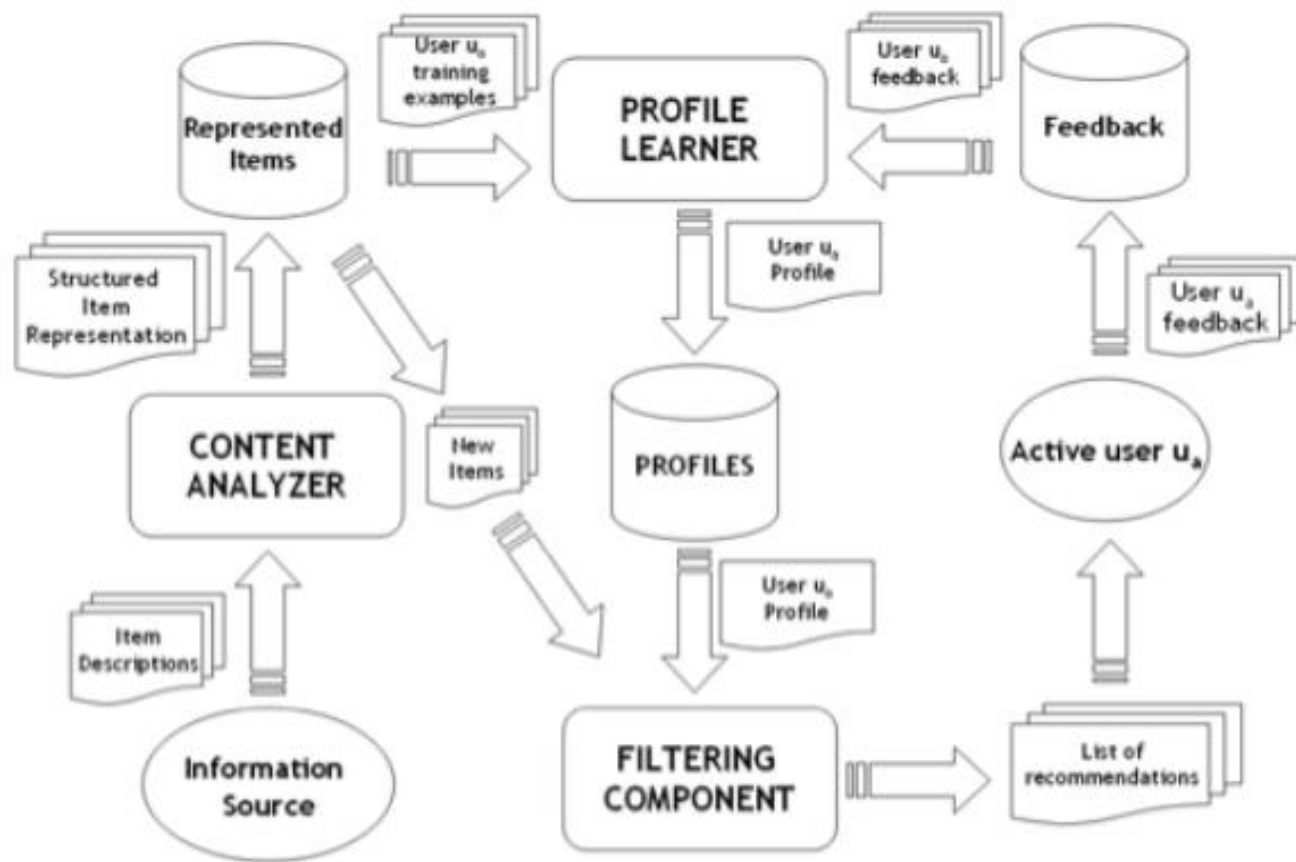


Fig. 3.1: High level architecture of a Content-based Recommender

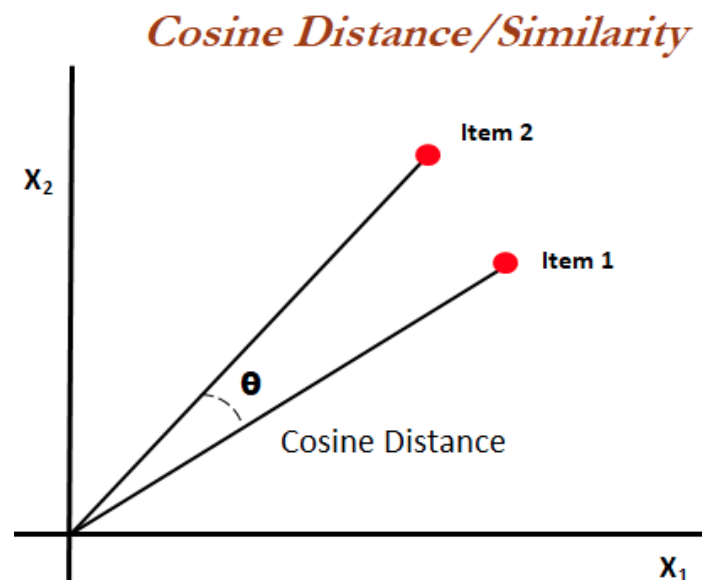
- Permite a utilização de diversas técnicas (NLP, Computer Vision, etc) – para enriquecer as informações de usuários e itens. **Principalmente com as descrições em texto.**
- Não sofre com os problemas de *cold start*, porém podem ser tão especializados/específicos que acabam recomendando itens sempre muito parecidos (*filter bubble*).

CORRELAÇÃO DE PEARSON

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

SIMILARIDADE COSSENOS

JACCARD INDEX/SIMILARITY



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

Hybrid Recommenders

