

SCC0284 / SCC5966

Sistemas de Recomendação

Aula 05: Filtragem Híbrida

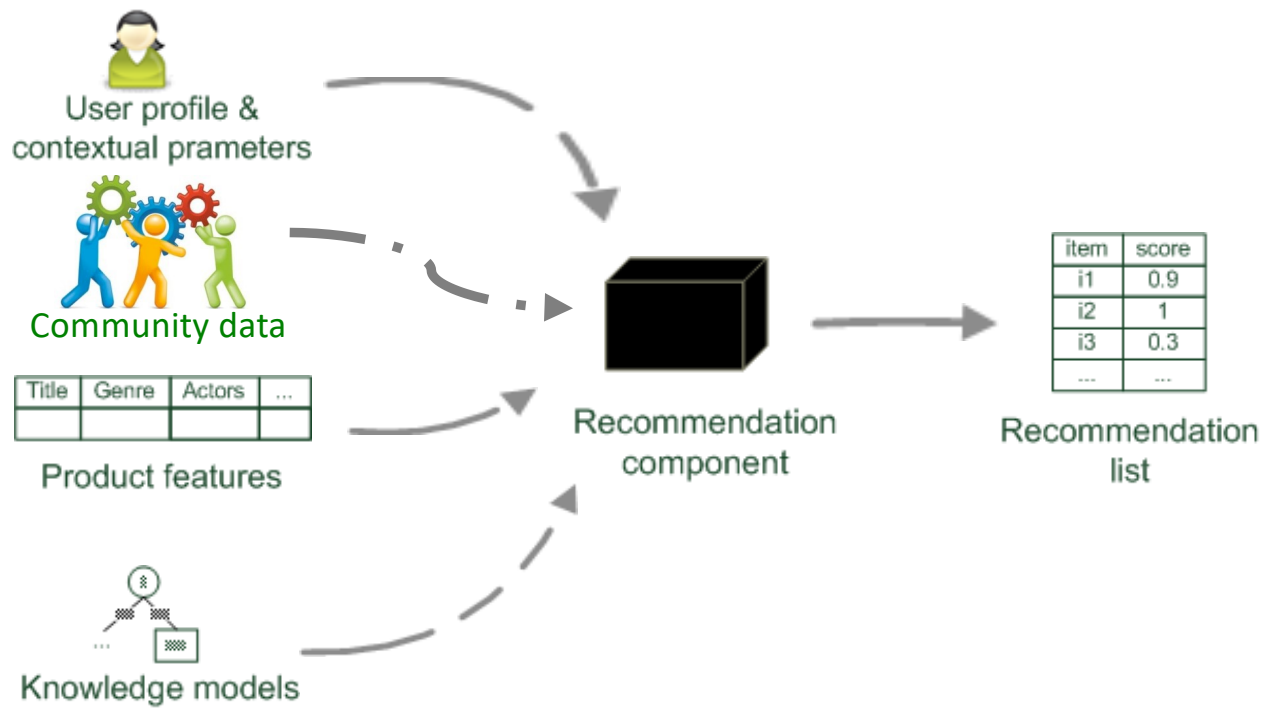
(mmanzato@icmc.usp.br)

Introdução

- Paradigmas de recomendação:
 - Baseado em conteúdo
 - Colaborativo
 - Baseado em conhecimento
- Desvantagens de cada uma?

Introdução

- Recomendação como uma caixa-preta



Introdução

- Requisitos de entrada

Paradigma	Perfil do usuário e parâmetros contextuais	Dados colaborativos	Características de produtos	Modelos de conhecimento
Colaborativo	Sim	Sim	Não	Não
Baseado em conteúdo	Sim	Não	Sim	Não
Baseado em conhecimento	Sim	Não	Sim	Sim

- Não há uma abordagem única que possa ser usada em qualquer domínio
 - Necessidade de hibridização!

Hibridização

- Muitas abordagens hoje em dia podem ser consideradas híbridas
- Sabe-se que combinar diferentes estratégias pode trazer resultados mais precisos
 - Prêmio Netflix: ganhador desenvolveu uma técnica de combinação de vários modelos de predição

Hibridização

- Projeto de hibridização envolve:
 - Escolher quais paradigmas de recomendação utilizar na combinação
 - Determinar a maneira com que eles serão combinados

Recomendação

- Um algoritmo de recomendação é representado como uma função de utilidade *rec*, tal que:

$$rec : U \times I \rightarrow R$$

- Um recomendador RS tem a tarefa de identificar *n* itens de um conjunto *I* com maiores escores:

$$RS(u, n) = \{i_1, \dots, i_k, \dots, i_n\} \quad \text{onde}$$

$$i_1, \dots, i_k, \dots, i_n \in I \quad \text{e}$$

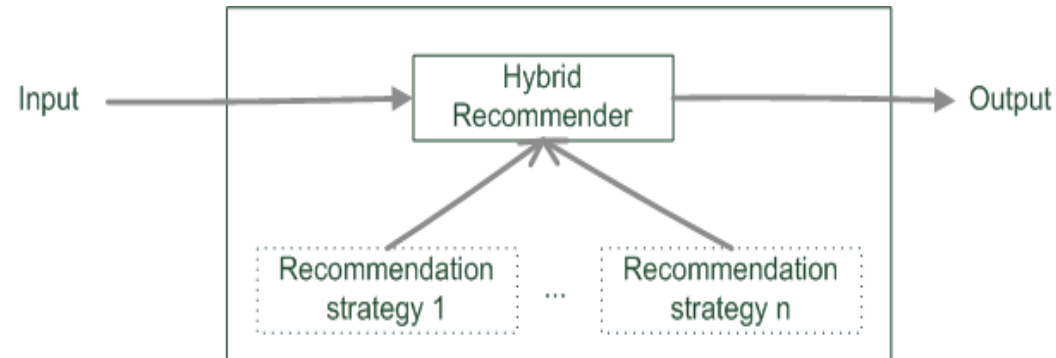
$$\forall k \quad rec(u, i_k) > 0 \wedge rec(u, i_k) > rec(u, i_{k+1})$$

Combinação

- Três maneiras para combinar algoritmos:
 - Híbridização **monolítica**
 - Híbridização **paralela**
 - Híbridização **canalizada**

Hibridização monolítica

- Consiste em um único componente recomendador que integra várias abordagens por meio de um pré-processamento e combinação de diferentes fontes de informação



Hibridização monolítica

- Duas formas:
 - **Combinação** de características
 - Diferentes tipos de dados são usados pelo algoritmo de recomendação
 - E.g. representação de itens com base em interações + metadados
 - **Enriquecimento** de características
 - Aplica transformações mais complexas nos dados de entrada
 - E.g. Exploração de revisões de usuários por meio de técnicas de *opinion mining* para melhorar a representação dos itens

Hibridização monolítica

- Exemplo (combinação)

						
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

Ideia:







- usando abordagem baseada em vizinhança de itens, combinar FC e FBC com base nos dados disponíveis

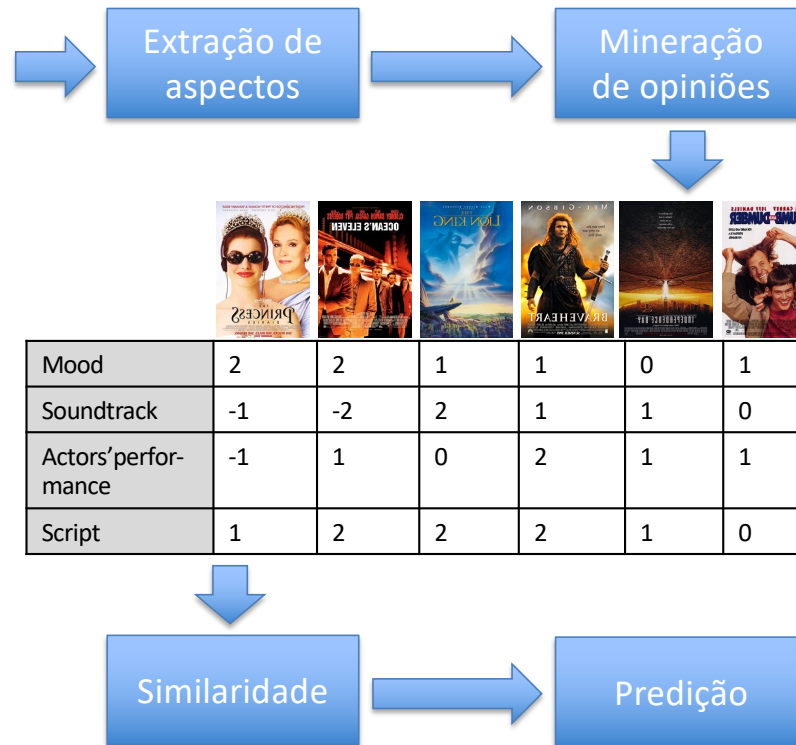
- Combinar similaridades baseadas em notas e em atributos

- Uma única função de predição deve ser usada!

Hibridização monolítica

- Exemplo (enriquecimento) (D'Addio & Manzato, 2015)

Filme	Revisões de usuários
	I really enjoyed this movie. It was fun to see a fairly clean movie that was fun to watch. The acting of Julie Andrews and Anne Hathaway is superb and the story shines out hope. The director takes a very unlikely event and makes it believable, teaching us all something in the process. It was refreshing to see a teen film without all the sexual innuendos and crude language.
	Personally, I have a theory that any movie that absolutely blows you away the first time you see it, will lose its entertainment value very quickly. Not necessarily because the movie was not good, but because it just seems to get old too fast. This is NOT one of those movies.
	This is a film that can entertain anyone young or old, I usually don't care for animated movies but this film is the real deal, this is one of disney's best animated movies. The animation is top notch and flawless. This film also features superb work from the vocal cast James Earl Jones, Jeremy Irons, Whoopi Goldberg. This is a standout.
	his has to be one of the best movies I have ever seen. I recently purchased it and have watched it at least five times since then, and each time i pick up on things I did not see the other times. The fight scenes are great, the plot is both interesting and thought provoking, there is romance and comedy. This is a movie that any person can appreciate at some level.
	Personally I think this is the best movie of the 90's. But then I think South Park is the best TV show of the 20th century. Followed by Twilight Zone and Star Trek tying for second and third place. While there's no accounting for taste in art, I can make reasoned arguments for my choices.
	This movie goes from joke to joke constantly and never misses a beat. I could not stop laughing the first time I saw it, and the tenth time I saw it, and the hundredth time I saw it. This movie never gets old. There is not one single dull moment. The movie takes off and never lets up, not even enough to let you breathe. There is something absolutely mesmerizing and hilarious about Jim Carrey and Jeff Daniels' performances as the lovable idiots Llovd Christmas(Jim Carrey) and Harry Dunne (Jeff Daniels).



Mood	2	2	1	1	0	1
Soundtrack	-1	-2	2	1	1	0
Actors' performance	-1	1	0	2	1	1
Script	1	2	2	2	1	0

A Sentiment-Based Item Description Approach for kNN Collaborative Filtering

Rafael M. D'Addio, Marcelo G. Manzato
 Institute of Mathematics and Computer Science
 University of São Paulo
 São Carlos, SP, Brazil
 {rdaddio, mmanzato}@icmc.usp.br

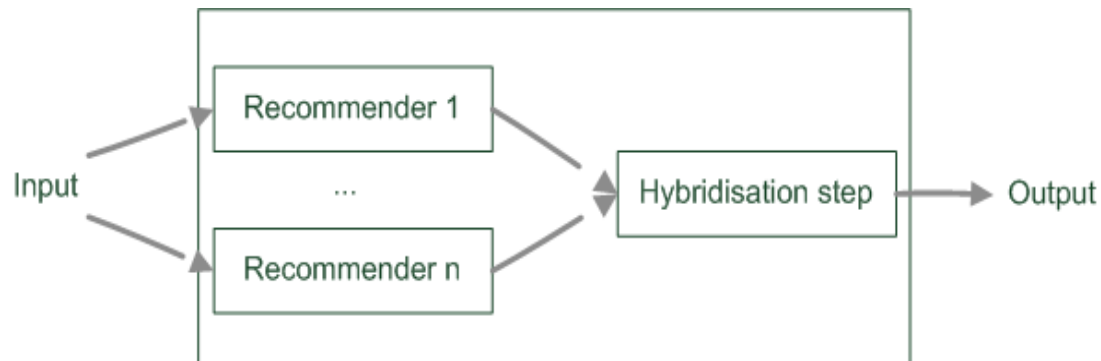
ABSTRACT

In this paper, we propose an approach based on sentiment analysis to describe items in a neighborhood-based collaborative filtering model. We use unstructured users' reviews to produce a vector-based representation that considers the overall sentiment of those reviews towards specific features. We propose and compare two different techniques to obtain and score such features from textual content, namely term-based and aspect-based feature extraction. Finally, our proposal is compared against structured metadata under the same recommendation algorithm, whose results show a significant improvement over the baselines.

Regarding the use of metadata to describe the content, there is a growing effort nowadays to consider unstructured data produced by the same or other users to improve the recommendation accuracy [9]. For instance, users' reviews are a rich information source that can support consumers to decide whether it is worth buying or consuming a particular item. Usually, those users manually check such source of information prior the consumption, but automatic techniques could smooth this task by incorporating such analysis into the filtering process, resulting in better recommendations. However, there is set of challenges that has to be dealt with when using unstructured textual content, in particular users' reviews, for representing items [1]. First, the reviews

Hibridização paralela

- Utiliza vários recomendadores lado-a-lado e emprega um mecanismo para combinar suas saídas



Hibridização paralela

- Três formas:
 - Modelagem **mesclada**
 - Modelagem **ponderada**
 - Modelagem **comutada**

Modelagem mesclada

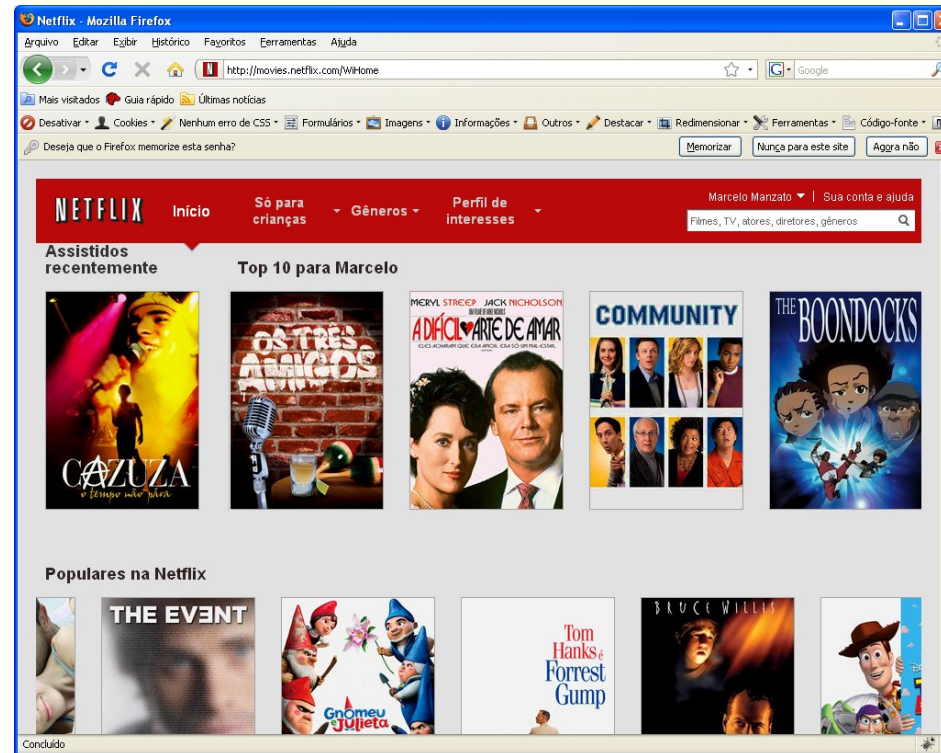
- Combina as saídas dos recomendadores em nível de interface do usuário
- Resultados são apresentados em conjunto

$$rec_m(u,i) = \bigcup_{k=1}^n \langle rec_k(u,i), k \rangle$$

- Itens com maiores escores de cada recomendador são apresentados para o usuário
 - Apresentação única baseada em heurísticas ou vários rankings exibidos ao mesmo tempo

Modelagem mesclada

- Exemplo: Netflix



Modelagem ponderada

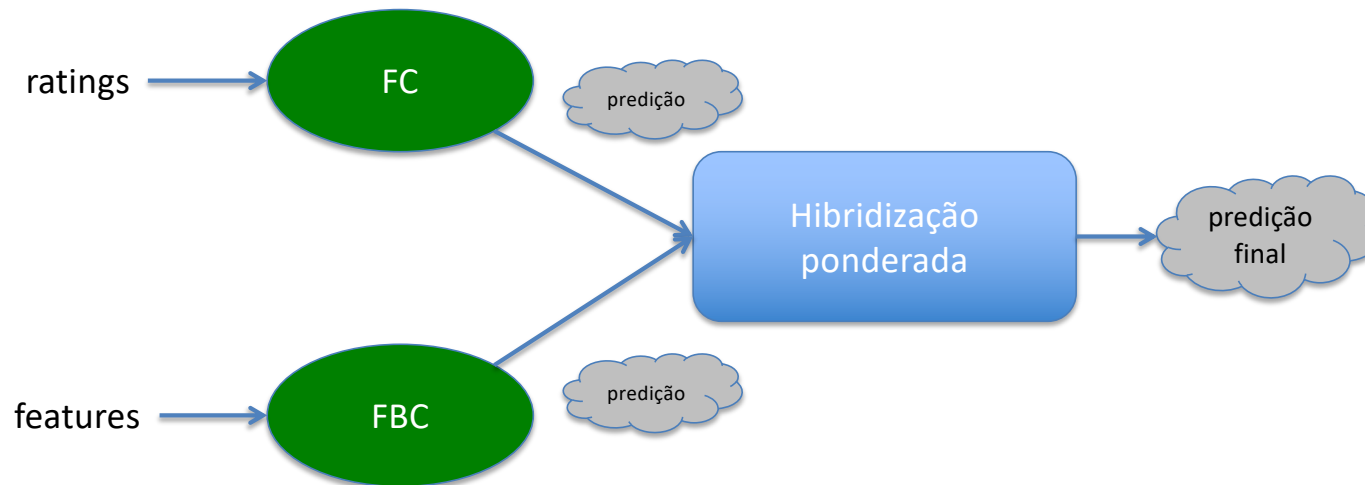
- Combina as recomendações usando uma soma ponderada de seus escores

$$rec_p(u,i) = \sum_{k=1}^n \beta_k \times rec_k(u,i)$$

- Sendo que:
 - Escores de cada recomendador precisam estar num intervalo fixo
 - Somatório de β 's tem que resultar em 1

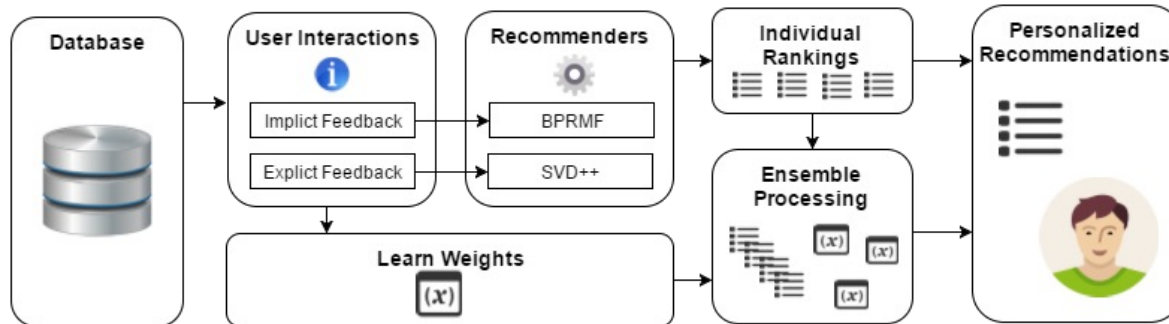
Modelagem ponderada

- Exemplo: combinação entre FC e FBC



Modelagem ponderada

- Outro exemplo: múltiplas interações (da Costa & Manzato, 2015):



Exploiting Multimodal Interactions in Recommender Systems With Ensemble Algorithms

Arthur F. da Costa and Marcelo G. Manzato

{fortes,mmanzato}@icmc.usp.br

Institute of Mathematics and Computer Science

University of São Paulo

São Carlos, SP, Brazil

Abstract

The increasing of products, information and services based on users' profiles has made recommender systems to be increasingly present, easing the selection and retention of users in services on the Web. However, optimizations must be performed in such systems mainly regarding the modeling of users' profiles. Preferences are generally modeled superficially, due to the scarcity of data collected, as notes or comments, or the inductive information susceptible to errors. This manuscript proposes a recommender tool with three ensemble approaches based on multimodal interactions that combines different types of users' feedback processed individually by traditional recommendation algorithms. The approaches have been developed to improve the quality of predictions in recommender systems, considering a large number of user information.

Modelagem comutada

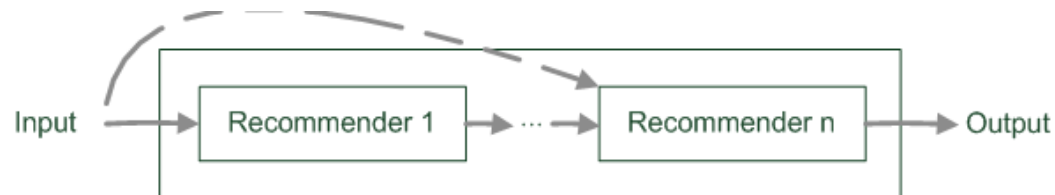
- Consiste em decidir qual recomendador deve ser usado em uma situação específica
 - A depender do perfil do usuário e/ou da qualidade dos resultados da recomendação

$$\exists k1...n : rec_c(u,i) = rec_k(u,i)$$

- Exemplo
 - Se usuário for novo, usar filtragem baseada em conteúdo, senão usar colaborativa

Hibridização canalizada

- Consiste em um processo em etapas, onde várias técnicas de filtragem são colocadas sequencialmente para gerar uma lista final de recomendações



Hibridização canalizada

- Duas formas:
 - Modelagem em cascata
 - Modelagem meta-nível

Modelagem em cascata

- Neste tipo, um recomendador subsequente apenas refina os resultados produzidos no passo anterior
- Itens que não foram selecionados num primeiro estágio não poderão ser recomendados pelo próximo algoritmo

$$rec_c(u,i) = rec_n(u,i)$$

- onde $\forall k \geq 2$ temos:

$$rec_k(u,i) = \begin{cases} rec_k(u,i) & : rec_{k-1}(u,i) \neq 0 \\ 0 & : otherwise \end{cases}$$

Modelagem em cascata

- Exemplo



Modelagem meta-nível

- Um recomendador constrói um modelo que é explorado por outro recomendador principal responsável por calcular as recomendações finais

$$rec_{meta}(u,i) = rec_n(u,i,\Delta_{rec_{n-1}})$$

- Onde o recomendador **n** explora um modelo **Δ** que foi construído pelo seu predecessor
- Geralmente, são utilizados apenas dois recomendadores, ou seja, $n = 2$.

Modelagem meta-nível

- Exemplos
 - Usar SVD para fatorar matriz de usuário/item e depois usar kNN na matriz Q
 - Construir um perfil do usuário com base em metadados, e depois usar kNN entre perfis
 - Usar matriz de vizinhança no modelo de fatoração de matrizes
 - Etc.

Conclusões

- Hibridização é usada para aumentar o poder de uso de várias fontes de dados
- Melhora a acurácia de sistemas existentes, por meio da complementação das vantagens de cada abordagem
- Métodos baseados em ensemble
 - Incorpora diversidade e reduz viés de cada modelo
 - Podem usar a mesma matriz de entrada (e.g. interações)
 - Ganhador do Netflix Prize: +100 modelos combinados

Referências

- Dietmar Jannach, Markus Zanker, Alexander Felfernig, Gerhard Friedrich. *Recommender Systems: An Introduction*. Cambridge University Press, 2010.
- D'Addio, R. M.; Manzato, M. G. A sentiment-based item description approach for kNN collaborative filtering. SAC'15, 2015.
- da Costa; Manzato, M. G. Exploiting multimodal interactions in recommender systems with ensemble algorithms. Information Systems, 2015.