Sistemas Recomendadores Híbridos

IIC 3633 - Sistemas Recomendadores

Denis Parra
Profesor Asistente, DCC, PUC CHile

Memo del Semestre

- **Tarea 1**: Deadline el Jueves 17 de Septiembre.
- **Lecturas en el semestre**: Ya fueron actualizadas en el sitio web del curso.
- **Proyecto Final**:
  - Entrega de abstract con a lo más 3 ideas el martes 22 de Septiembre, el 29 de septiebre se debe entregar propuesta final.

Lecturas Semana 2 (Evaluación)

- Aspectos débiles y críticos:
  - No contiene referencias a métodos de factorización matricial. Ref: Koren, Y., & Bell, R. (2011). Advances in collaborative filtering. In Recommender systems handbook (pp. 145-186). Springer US.
  - Métricas no son autoexplicativas. Especial crítica a cómo se describe métrica de diversity y novelty de Lathia (revisar el post de http://jpsf1971.tumblr.com/ por detalles)
  - Cobertura del capítulo, muy general. Poca diferencia entre machine learning, data mining y RecSys.
- Aspectos positivos:
  - Descripción de las fuentes de información.
  - Algunos términos que suenan novedosos para algunos usuarios (lurkers, trust)
- Aspectos para discusión:
  - Privacidad, mención de "venta de bases de datos de usuarios".
  - Robustez del recomendador a ataques / Reproducibility / Finalidad del RecSys.
  - Falta investigación que permita identificar técnicas que fallan para no repetir errores.
  - Combinación de la métricas: Cuál importa más? cómo encontrar la mejor combinación?

Resumen de Comentarios en Blogs II

Lecturas Semana 3 (Métricas de Evaluación)

- Aún faltan bastantes! Las revisaremos la próxima clase.

TOC

En esta clase

1. Motivación
2. Clasificación General
3. Modelos de Hibridización
4. Ejemplos

Motivación

Diferentes métodos tienen distintas debilidades y fortalezas

- Filtrado Colaborativo es preciso, pero sufre de sparsity, cold start y new item problem
- Filtrado Basado en contenido no sufre tanto por sparsity y permite con facilidad para extraer features del contenido. Sin embargo, también sufre de "new user problem", es menos preciso de el F.C. y presenta sobre-especialización.
- Knowledge-based: No los hemos visto hasta ahora. Casos típicos son Constrait-Based y Case-Based. Basados en un paradigma más interactivo, también los llaman “Conversacionales” (Burke, 2002). Su principal debilidad es el costo de mantener las reglas actualizadas.

Categorización de RecSys de Burke (2002)

Table I: Recommendation Techniques

Technique	Background	Input	Process
Collaborative	Ratings from **U** of items in **I**.	Ratings from **u** of items in **I**.	Identify users in **U** similar to **u**, and extrapolate from their ratings of **i**.
Content-based	Features of items in **I**	**u**'s ratings of items in **I**	Generate a classifier that fits **u**'s rating behavior and use it on **i**.
Demographic	Demographic information about **U** and their ratings of items in **I**.	Demographic information about **u**.	Identify users that are demographically similar to **u**, and extrapolate from their ratings of **i**.
Utility-based	Features of items in **I**.	A utility function over items in **I** that describes **u**'s preferences.	Apply the function to the items and determine **i**'s rank.
Knowledge-based	Features of items in **I**. Knowledge of how these items meet a user's needs.	A description of **u**'s needs or interests.	Infer a match between **i** and **u**'s need.

Ref: Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User modeling and user-adapted interaction, 12(4), 331-370.

7/33

Modelo Caja Negra de RecSys (Jannach et al. 2010)

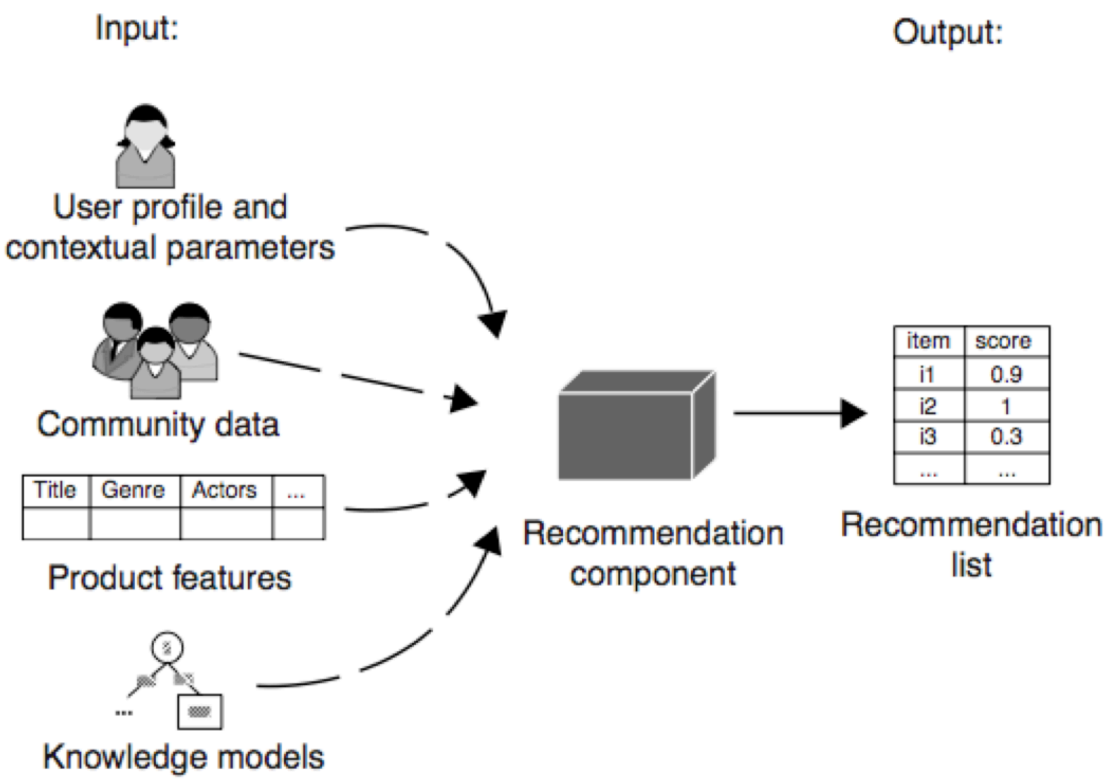


Figure 5.1. Recommender system as a black box.

Ref: Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). Recommender systems: an introduction. Cambridge University Press.

Combinar Métodos Content-based y Filtrado Colaborativo

Según (Adomavicius et al., 2005)

- Implementar métodos CF y CB separadamente y combinar las predicciones
- Incorporar características de métodos CB dentro de un método CF
- Incorporar características colaborativas dentro de modelo CB
- Construir un modelo que de manera unificada incorpore características basadas en contenido y colaborativas

Ref: Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on, 17(6), 734-749.

Cómo combinar Métodos de Recomendación?

Burke (2002) distingue 7 estrategias de hibridización

Jannach (2012) resume las 7 estrategias en 3 diseños generales

- Monolítico
- Paralelizado
- Pipeline

7 Estrategias de Hibridización (Burke 2002)

Table III: Hybridization Methods

Hybridization method	Description
Weighted	The scores (or votes) of several recommendation techniques are combined together to produce a single recommendation.
Switching	The system switches between recommendation techniques depending on the current situation.
Mixed	Recommendations from several different recommenders are presented at the same time
Feature combination	Features from different recommendation data sources are thrown together into a single recommendation algorithm.
Cascade	One recommender refines the recommendations given by another.
Feature augmentation	Output from one technique is used as an input feature to another.
Meta-level	The model learned by one recommender is used as input to another.

Hibridización Monolítica

Table III: Hybridization Methods

Hybridization method	Description
Weighted	The scores (or votes) of several recommendation techniques are combined together to produce a single recommendation.
Switching	The system switches between recommendation techniques depending on the current situation.
Mixed	Recommendations from several different recommenders are presented at the same time.
Feature combination	Featu [redacted] Monolitic [redacted] are thrown toget [redacted]
Cascade	One r [redacted] by another.
Feature augmentation	[redacted] Monolitic [redacted] o another.
Meta-level	The [redacted] ut to another.

Hibridización Paralela

Table III: Hybridization Methods

Hybridization method	Description
Weighted	The scores (or votes) of several recommendation techniques are combined together to produce a single recommendation.
Switching	T Parallel hniques d
Mixed	Recommendations from several different recommenders are presented at the same time
Feature combination	Features from different recommendation data sources are thrown together into a single recommendation algorithm.
Cascade	One recommender refines the recommendations given by another.
Feature augmentation	Output from one technique is used as an input feature to another.
Meta-level	The model learned by one recommender is used as input to another.

13/33

Hibridización Pipeline

Table III: Hybridization Methods

Hybridization method	Description
Weighted	The scores (or votes) of several recommendation techniques are combined together to produce a single recommendation.
Switching	The system switches between recommendation techniques depending on the current situation.
Mixed	Recommendations from several different recommenders are presented at the same time
Feature combination	Features from different recommendation data sources are thrown [redacted] n.
Cascade	Pipeline given by another.
Feature augmentation	Output from one technique is used as an input feature to another.
Meta-level	as input to Pipeline

14/33

Hibridización Monolítica

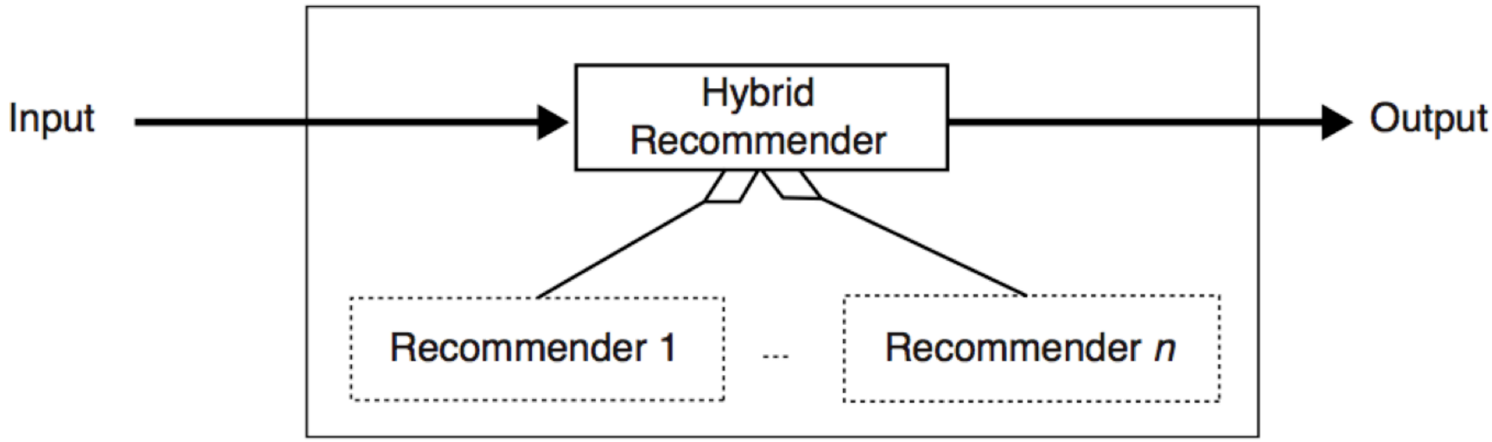


Figure 5.2. Monolithic hybridization design.

Estrategias de Combinación:

- Feature Combination
- Feature Augmentation

H. Monolítica: Feature Combination I

Table 5.3. *Hybrid input features*.

Feature	Alice	User1	User2	User3	User4
User likes many *mystery* books	true	true			
User likes some *mystery* books			true	true	
User likes many *romance* books					
User likes some *romance* books			true	true	
User likes many *fiction* books					
User likes some *fiction* books		true	true		true

H. Monolítica: Feature Combination II

Table 5.4. *Different types of user feedback.*

User	R_{nav}	R_{view}	R_{ctx}	R_{buy}
Alice	n_3, n_4	i_5	k_5	$\emptyset$
User1	n_1, n_5	i_3, i_5	k_5	i_1
User2	n_3, n_4	i_3, i_5, i_7	$\emptyset$	i_3
User3	n_2, n_3, n_4	i_2, i_4, i_5	k_2, k_4	i_4

H. Monolítica: Feature Augmentation

18/33

H. Monolítica: Feature Augmentation

Usar estas "features" en un nuevo recomendador

19/33

Hibridización Paralela

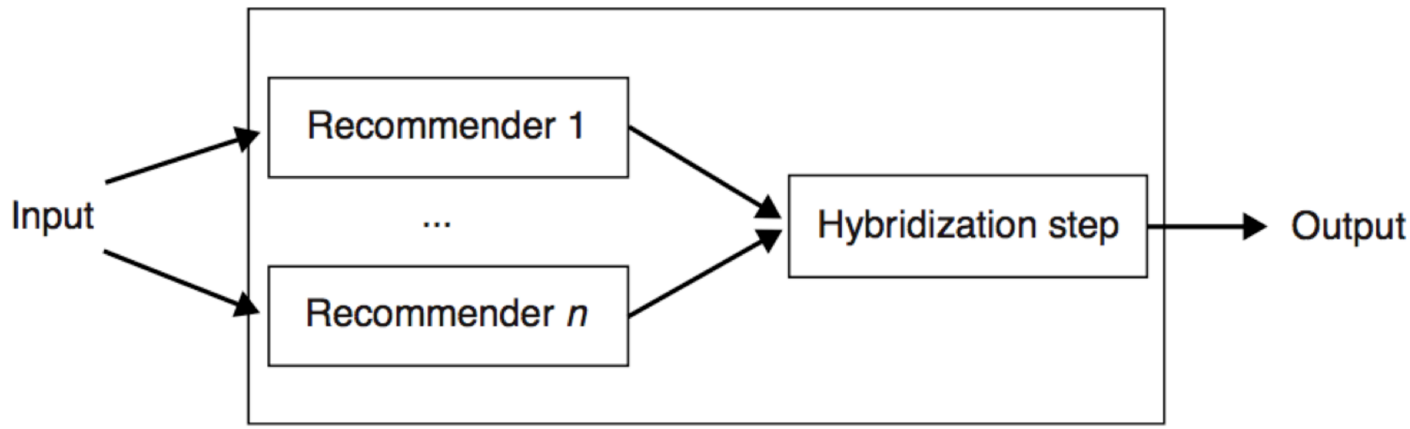


Figure 5.3. Parallelized hybridization design.

Tres mecanismos principales:

- Weighted
- Mixed
- Switching

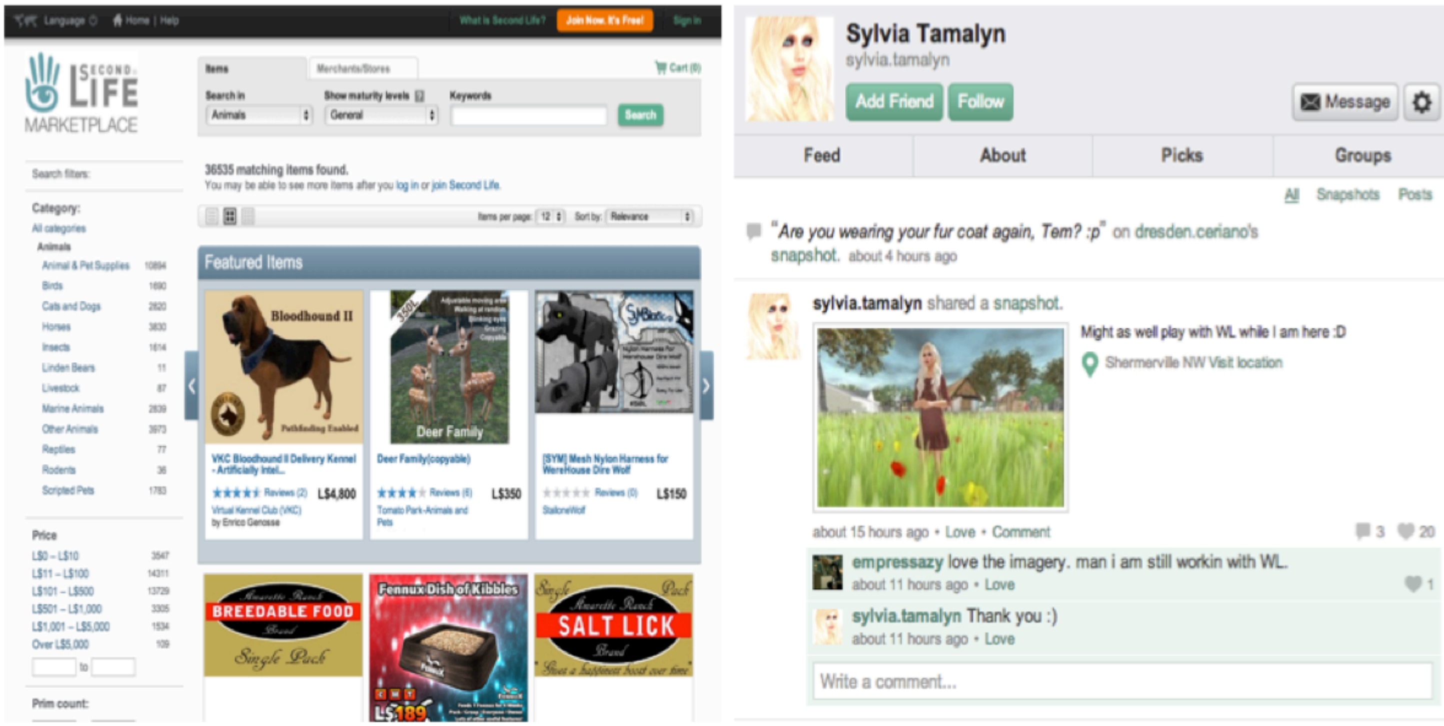
20/33

H. Paralela: Weighted I

$$W_{rec_i} = \sum_{s_j \in S} (W_{rec_i,s_j} \cdot W_{s_j})$$

rec_i	Item recomendado i
W_{rec_i}	Score combinado del item i
W_{rec_i,s_j}	Score del item I por el recomendador S_j
W_{s_j}	Peso del recomendador S_j

H. Paralela: Weighted II



(a) SecondLife store (b) SecondLife social stream

Fig. 1. Examples for a store in the marketplace and a social stream of an user in the online social network of the virtual world SecondLife.

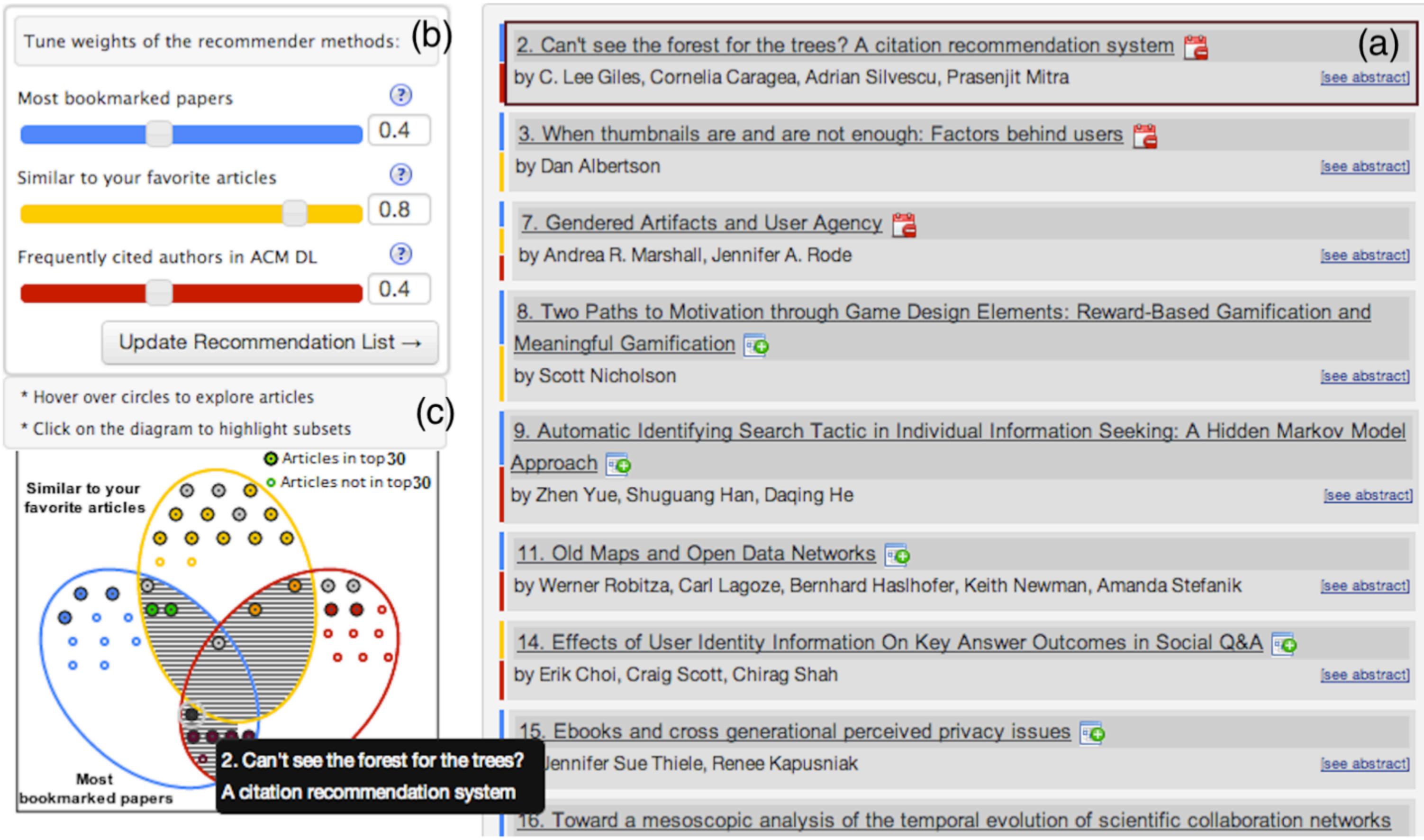
H. Paralela: Mixed I

$$src(rec_i) = \left[\sum_{m_m \in M} \frac{1}{rank_{rec_i, m_j}} \times \boxed{W_{m_j}} \right] \times |M_{rec_i}|$$

Slider weight

M: The set of all methods available to fuse
$rank_{reci,mj}$: rank–position in the list of a recommended item
rec_i : recommended item *i*
m_j, : recommendation method j
W_{mj} : weight given by the user to the method m_j using the controllable interface
$|M_{reci}|$ represents the number of methods by which item rec_i was recommended

H. Paralela: Mixed II



24/33

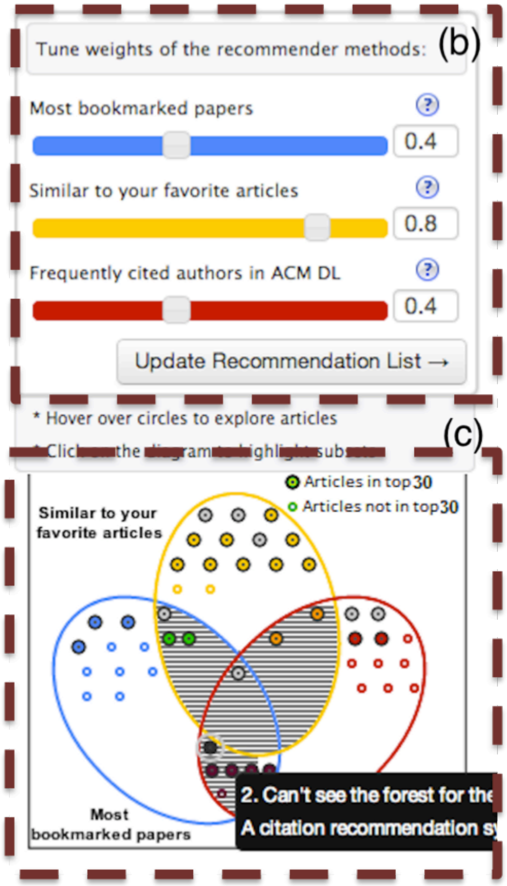
H. Paralela: Mixed III

Traditional Ranked List

Papers sorted by Relevance.
It combines 3 recommendation approaches.

25/33

H. Paralela: Mixed IV



Sliders

Allow the user to control the importance of each data source or recommendation method

Interactive Venn Diagram

Allows the user to inspect and to filter papers recommended. Actions available:

- Filter item list by clicking on an area
- Highlight a paper by mouse-over on a circle
- Scroll to paper by clicking on a circle
- Indicate bookmarked papers

26/33

H. Paralela: Switching I

- De un grupo de recomendadores, **activar un recomendador a la vez**.
- Podría ser especialmente útil **considerando los learning rate de algunos métodos**.
- **Ejemplo**: Elegir entre un clasificador Bayesiano y un recomendador Item-based como en: Ghazanfar, M., & Prugel-Bennett, A. (2010). **An Improved Switching Hybrid Recommender System Using Naive Bayes Classifier and Collaborative Filtering**.

$$P(C_j|d) = \frac{P(C_j)\prod_{i=1}^{h} P(F_i|C_j)}{P(F_1,\cdots,F_h)}.$$

27/33

H. Paralela: Switching II

Table 1: A comparison of proposed algorithm with existing in terms of cost (based on [31]), accuracy metrics, and coverage

Algorithm	On-line Cost	Best MAE		ROC-Sensitivity		Coverage	
		(ML)	(FT)	(ML)	(FT)	(ML)	(FT)
$UBCF_{DV}$	$O(M^2N) + O(NM)$	0.766	1.441	0.706	0.563	99.424	93.611
IBCF	$O(N^2)$	0.763	1.421	0.733	0.605	99.221	92.312
IDemo4	$O(N^2)$	0.749	1.407	0.739	0.621	99.541	94.435
Rec_{NBCF}	$O(N^2) + O(Mf)$	**0.696**	**1.341**	**0.778**	**0.657**	**100**	**99.992**
NB	$O(Mf)$	0.808	1.462	0.703	0.571	100	99.992
NH	$O(N^2) + O(Mf)$	0.785	1.438	0.712	0.586	100	99.992
CB	$O(M^2N) + O(NM) + O(Mf)$	0.721	1.378	0.741	0.611	100	**99.995**

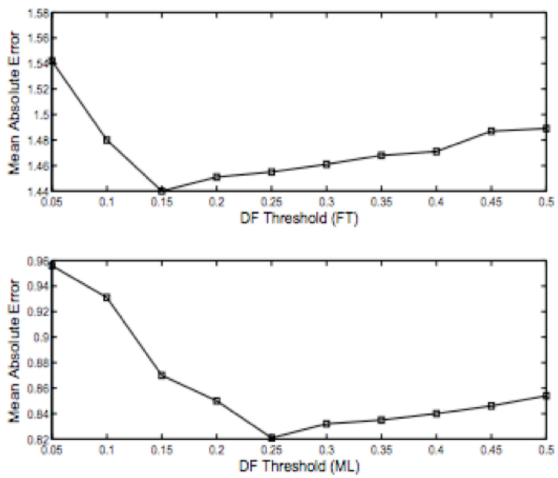


Figure 1: Determining the optimal value of DF.

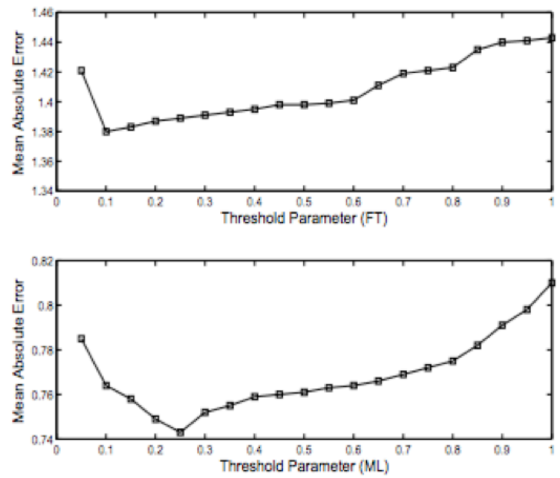


Figure 2: Determining the optimal value of α.

Hibridización Pipeline

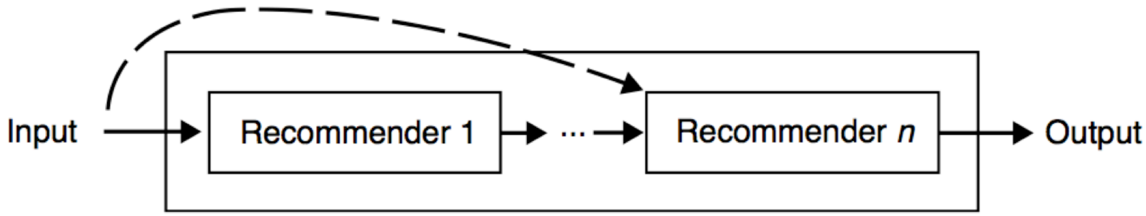


Figure 5.4. Pipelined hybridization design.

Dos mecanismos principales:

- Cascade
- Meta-Level

H. Pipeline: Cascade

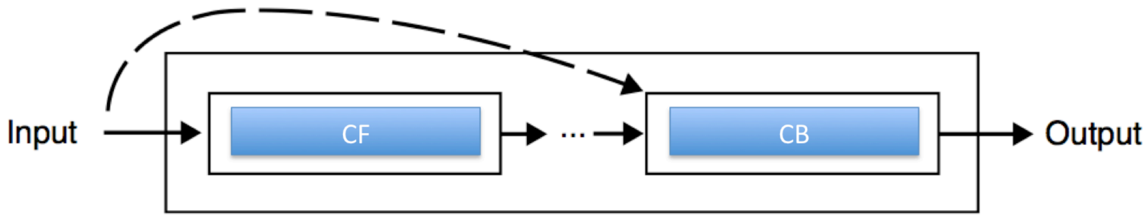
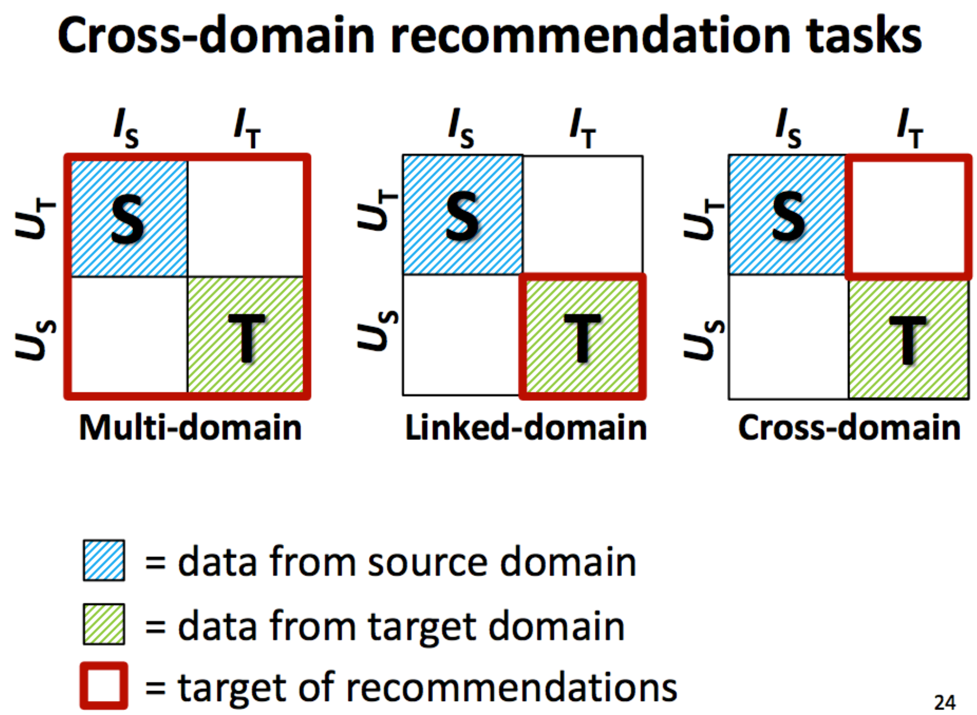


Figure 5.4. Pipelined hybridization design.

H. Pipeline: Meta-Level

El modelo aprendido por un recomendador es usado para un segundo recomendador. Transfer Learning:



Tutorial on Cross-domain recommender systemshttp://recsys.acm.org/wp-content/uploads/2014/10/recsys2014-tutorial-cross_domain.pdf

31/33

Referencias

- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User modeling and user-adapted interaction, 12(4), 331-370.
- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on, 17(6), 734-749.
- Burke, R. (2007). Hybrid web recommender systems. In The adaptive web (pp. 377-408). Springer Berlin Heidelberg.
- Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). Recommender systems: an introduction. Cambridge University Press. Chicago

33/33