

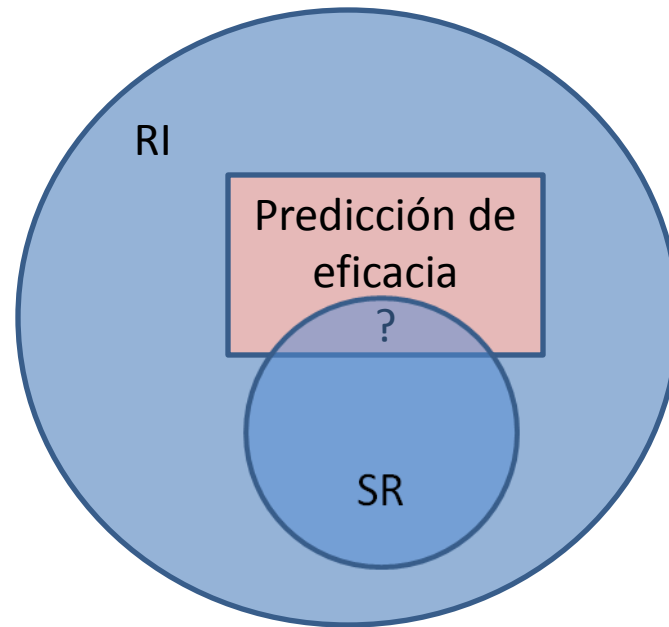
# Performance prediction in recommender systems: application to the dynamic optimisation of aggregative methods

Alejandro Bellogín Kouki

Tutor: Pablo Castells Azpilicueta

Departamento de Ingeniería Informática  
Escuela Politécnica Superior  
Universidad Autónoma de Madrid

# Introducción



RI : Recuperación de Información

SR : Sistemas de Recomendación

# Índice

- Motivación y objetivos
- Trabajo relacionado y contexto
- Marco formal
  - Fusión dinámica en sistemas de recomendación
  - Predictores de eficacia para componentes de sistemas de recomendación
- Ponderación de vecinos en filtrado colaborativo
  - Descripción
  - Experimentos
- Ponderación en recomendación híbrida
  - Descripción
  - Experimentos
- Conclusiones

# Motivación

- En RI entran en juego distintas fuentes de información
- Uso de predicción de eficacia en RI: ajustar las estrategias de recuperación en función de la predicción
  - Típicamente entendida como eficacia de la consulta
  - Aplicaciones: expansión de consultas, fusión de rankings, RI distribuida, etc.

# Objetivos

- Estudio y análisis del estado del arte en predicción de eficacia en Recuperación de Información
- Investigar la aplicación de estas técnicas en sistemas de recomendación y su utilidad
- Definición de predictores de eficacia en sistemas de recomendación
- Aplicación de predictores a combinaciones dinámicas en sistemas de recomendación

# Predicción de eficacia en RI: expansión de consultas

- Decidir si una consulta debe ser expandida o no, es decir, dados un predictor  $\gamma$  y una consulta  $q$ :
  - si  $\gamma(q)$  es mayor que un cierto número (umbral) entonces la consulta funcionará bien y no se debe expandir,
  - en otro caso, se expande la consulta.
- Problemas:
  - Definición del predictor (cuanto mayor es el valor, mejor funciona la consulta)
  - Definición del valor del umbral (se puede buscar el óptimo)
- Miran la ambigüedad, vaguedad, o especificidad de la consulta
  - $q_A$ : “carrera”
  - $q_B$ : “coche de carreras”, “carrera de caballos”  
$$\gamma(q_B) > \gamma(q_A)$$
- Trabajos: Cronen-Townsend et al. 2002, He & Ounis 2004, Diaz & Jones 2004, Mothe & Tanguy 2005, Jensen et al. 2005, Amati et al. 2004, Zhou & Croft 2006, Zhou & Croft 2007, Carmel et al. 2006

# Predictores de eficacia de consulta en RI

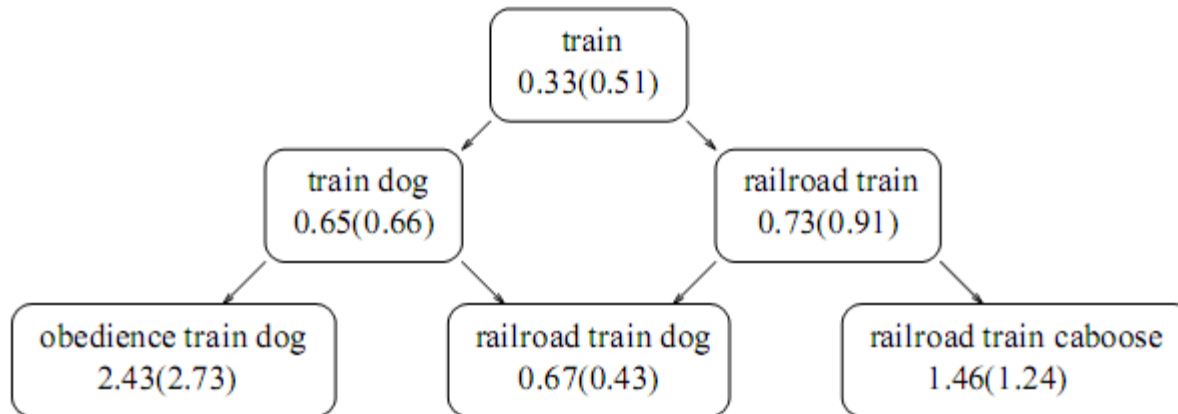
- Claridad: distancia (entropía relativa) entre los modelos de lenguaje de la consulta y de la colección

$$\text{clarity}(q) = \sum_{w \in V} P(w|q) \log_2 \frac{P(w|q)}{P_{coll}(w)}$$

$$P(w|q) = \sum_{d \in R} P(w|d)P(d|q), \quad P(q|d) = \prod_{w_q \in q} P(w_q|d)$$

$$P(w|d) = \lambda P_{ml}(w|d) + (1 - \lambda) P_{coll}(w)$$

- Ejemplo:



# Predicción de eficacia en RI: agregación de rankings

- Aplicación a la agregación de rankings
  - Obtener una única ordenación a partir de distintos rankings (p.e., meta-búsqueda, búsqueda distribuida, etc.) (Fox & Shaw 1993)
- En esta área, la predicción de eficacia se utiliza para ponderar cada uno de los rankings a combinar (según su eficacia estimada)
- Referencias: Yom-Tov et al. 2005, Aslam & Pavlu 2007, Castells et al. 2005, Diaz 2007
  
- Ejemplo:
  - “gripe” →  $0.8 * \text{PubMed} + 0.2 * \text{MathSciNet}$
  - “integral” →  $0.2 * \text{PubMed} + 0.8 * \text{MathSciNet}$



# Contexto: sistemas de recomendación

- En este trabajo: aplicación a sistemas de recomendación

objetos

usuarios

	$i_1$		$i_k$		$i_m$
$u_1$	$r_{11}$		$r_{1k}$		$r_{1m}$
$u_j$	$r_{j1}$		?		$r_{jm}$
$u_n$	$r_{n1}$		$r_{nk}$		$r_{nm}$

¿cómo se predice el valor de  $r_{jk}$ ?

# Contexto: sistemas de recomendación

- Filtrado colaborativo (FC) basado en usuarios

objetos

	$i_1$		$i_k$		$i_m$
usuarios	$u_1$	$r_{11}$		$r_{1k}$	$r_{1m}$
	$u_j$	$r_{j1}$		?	$r_{jm}$
	$u_n$	$r_{n1}$		$r_{nk}$	$r_{nm}$

# Contexto: sistemas de recomendación

- Filtrado colaborativo (FC) basado en objetos

objetos

usuarios

	$i_1$		$i_k$		$i_m$
$u_1$	$r_{11}$		$r_{1k}$		$r_{1m}$
$u_j$	$r_{j1}$		?		$r_{jm}$
$u_n$	$r_{n1}$		$r_{nk}$		$r_{nm}$

# Contexto: sistemas de recomendación

- Recomendación basada en contenido (BC)

objetos

▲      ■      ▲

usuarios

	$i_1$		$i_k$		$i_m$
$u_1$	$r_{11}$		$r_{1k}$		$r_{1m}$
$u_j$	$r_{j1}$		?		$r_{jm}$
$u_n$	$r_{n1}$		$r_{nk}$		$r_{nm}$

# Contexto: sistemas de recomendación

- Tanto FC como BC tienen desventajas:

Problema	Descripción	FC	BC
Oveja negra	Un usuario cuyos intereses son raros comparados con el resto de la comunidad	×	
Baja densidad (sparsity)	El número de ratings disponible es pequeño	×	
Nuevo objeto	Los objetos a ser recomendados debe ser puntuados por un número importante de usuarios	×	
Nuevo usuario	Un usuario tiene que votar un número suficiente de objetos para poder inferir sus preferencias	×	×
Análisis de contenido	Los objetos a recomendar deben tener datos disponibles con respecto a sus propiedades		×
Especialización	Los objetos recomendados son todos similares		×

- Una alternativa a estos problemas es utilizar estrategias híbridas:
  - En cascada
  - Usando ponderación (combinación lineal o esquemas de votación)

# Operaciones de combinación en SR

- Generación de recomendaciones en FC

$$g(u_m, i_n) = \frac{\sum_{u_j \in N[u_m]} \text{sim}(u_m, u_j) \times r_{j,n}}{\sum_{u_j \in N[u_m]} |\text{sim}(u_m, u_j)|}$$

- Recomendación híbrida

$$g(u_m, i_k) = \lambda \times g_{CB}(u_m, i_k) + (1 - \lambda) \times g_{CF}(u_m, i_k)$$

- Operaciones de combinación en RI personalizada

$$s(d) = \lambda \times s_q(d) + (1 - \lambda) \times s_p(d)$$

- Pregunta a investigar:

¿Se podría aplicar la predicción de eficacia a estos problemas?

¿Cómo?

# Marco formal

- Teoría de la utilidad: formulamos RI como encontrar  $g : \mathcal{D} \times \Omega \rightarrow \mathbb{R}$ , tal que  $d_1 \leq_q d_2 \Leftrightarrow g(d_1, q) \leq g(d_2, q)$
- Un sistema compuesto de RI está formado por subsistemas, donde cada uno se encarga de un criterio o estrategia:

$$\vec{g} = g(d, q) = \varphi (g_1(d, q), \dots, g_n(d, q))$$

- La combinación lineal es una forma muy general de composición en RI:

$$\varphi (s_1, \dots, s_n) = \alpha_1 s_1 + \dots + \alpha_n s_n = \vec{\alpha} \cdot \vec{s}$$

- Ejemplos:

- Cálculo de recomendaciones en FC
- Recomendación híbrida
- Búsqueda personalizada (combinación lineal entre la búsqueda estándar y la personalización) :  $\lambda \times \text{score}_q (d) + (1 - \lambda) \times \text{score}_p (d)$
- Fusión de rankings:  $\text{score}(d) = \sum_{\tau} \text{score}_{\tau} (d)$  (CombSUM)

# Marco formal

- Esta combinación se puede mejorar usando los mejores pesos posibles, es decir:

$$\vec{\alpha}^* = \arg \max_{\vec{\alpha} \in \mathbb{R}^n} (\rho(\vec{g}))$$

- $\rho$  representa una medida de la calidad de la decisión  $\vec{g}$
- El coeficiente  $\alpha_j$  determina cuánto debe dominar  $g_j$  en la decisión combinada  $\vec{g}$
- De esta manera, se puede aumentar la eficacia de  $g$  favoreciendo los  $g_j$  que darían mayor calidad a la salida en cada situación
- Es necesario “predecir” la eficacia de cada componente:
  - Función (o predictor)  $\gamma_j(d, q, \Omega)$
- Así:  $\alpha_j = \psi(\gamma_j(d, q, \Omega))$ 
  - Simplificación:  $\psi(x) = x$



# Hipótesis

1. La combinación lineal es una manera apropiada y general de construir sistemas compuestos en RI
2. Una ponderación dinámica adecuada de los componentes permite mejorar la eficacia de sistemas de RI compuestos
3. Los predictores de eficacia del campo de RI pueden adaptarse a SR y obtener predictores efectivos
4. La bondad de los predictores de eficacia se puede medir usando la correlación con respecto a unas medidas de eficacia adecuadas
5. La eficacia de un sistema de RI, o de un componente, es monótonamente decreciente con respecto a la cantidad de incertidumbre involucrada en la tarea de RI

# Fusión dinámica en sistemas de recomendación

- En este caso:  $\mathcal{D} = \mathcal{I}$  y  $\Omega = (\mathcal{U}, r)$ , es decir, el espacio de búsqueda son los objetos a recomendar, mientras que el de entrada es el conjunto de usuarios y  $r : \mathcal{S} \rightarrow \mathcal{R}$  con  $\mathcal{S} \subset \mathcal{U} \times \mathcal{I}$ , ratings de un usuario a un objeto (indicando cuánto le gusta a dicho usuario ese objeto)

- Normalmente, se centra en encontrar el mejor objeto (top 1):

$$g : \mathcal{U} \times \mathcal{I} \rightarrow \mathcal{R}$$

$$\forall u \in \mathcal{U}, i_u^* = \arg \max_{i \in \mathcal{I}} g(u, i)$$

- La utilidad de un objeto es equivalente al rating (real o predicho por el sistema)

# Predictores de eficacia para sistemas de recomendación

- Basados en claridad
  - De usuario vs objeto
  - Usando usuarios (UUC, UIC) u objetos (IUC, IIC)
- Basados en Teoría de la Información
  - Ganancia de información de un usuario (UIG) y de un objeto (IIG)
- Heurísticos
  - Contar frecuencias (inspirado en el IDF de RI):
    - Frecuencia inversa de usuario (IUF) y de objeto (IIF)
    - Frecuencia de ratings inversa de usuario (IURF) y de objeto (IIRF)
    - Características de un usuario (UF) y de un objeto (IF)

# Ponderación de vecinos en filtrado colaborativo

- Formalización:

$$g(u_m, i_k) = \varphi_{N[u_m]}(g_1(u_m, i_k), \dots, g_n(u_m, i_k))$$

$$g_j(u_m, i_k) = \text{sim}(u_m, v_j) \times r(u_j, i_k)$$

$$\alpha_j = \frac{\gamma_j}{\sum_{v \in N[u_m]} |\text{sim}(u_m, v)|}$$

$$g(u_m, i_n) = \frac{\sum_{u_j \in N[u_m]} \text{sim}(u_m, u_j) \times r_{j,n}}{\sum_{u_j \in N[u_m]} |\text{sim}(u_m, u_j)|}$$

# Ponderación de vecinos en filtrado colaborativo

- Predictores utilizados:

- *Item-based user clarity (IUC):*

$$\gamma_j = \gamma_j(v_j) = \text{IUC}(u) = \sum_{i \in I} p(i|u) \log_2 \frac{p(i|u)}{p_c(i)}$$
$$p(i|u) = \lambda \frac{\text{rat}(u,i)}{5} + (1-\lambda) p_c(i)$$
$$p_c(i) = \frac{1}{|I|}$$

- *User-based user clarity (UUC):*

$$\gamma_j = \gamma_j(v_j) = \text{UUC}(u) = \sum_{v \in U} p(v|u) \log_2 \frac{p(v|u)}{p_c(v)}$$
$$p(v|u) = \sum_{i: \text{rat}(u,i) \neq 0} p(v|i) p(i|u)$$
$$p(v|i) = \lambda \frac{\text{rat}(v,i)}{5} + (1-\lambda) p_c(v)$$
$$p_c(v) = \frac{1}{|U|}$$

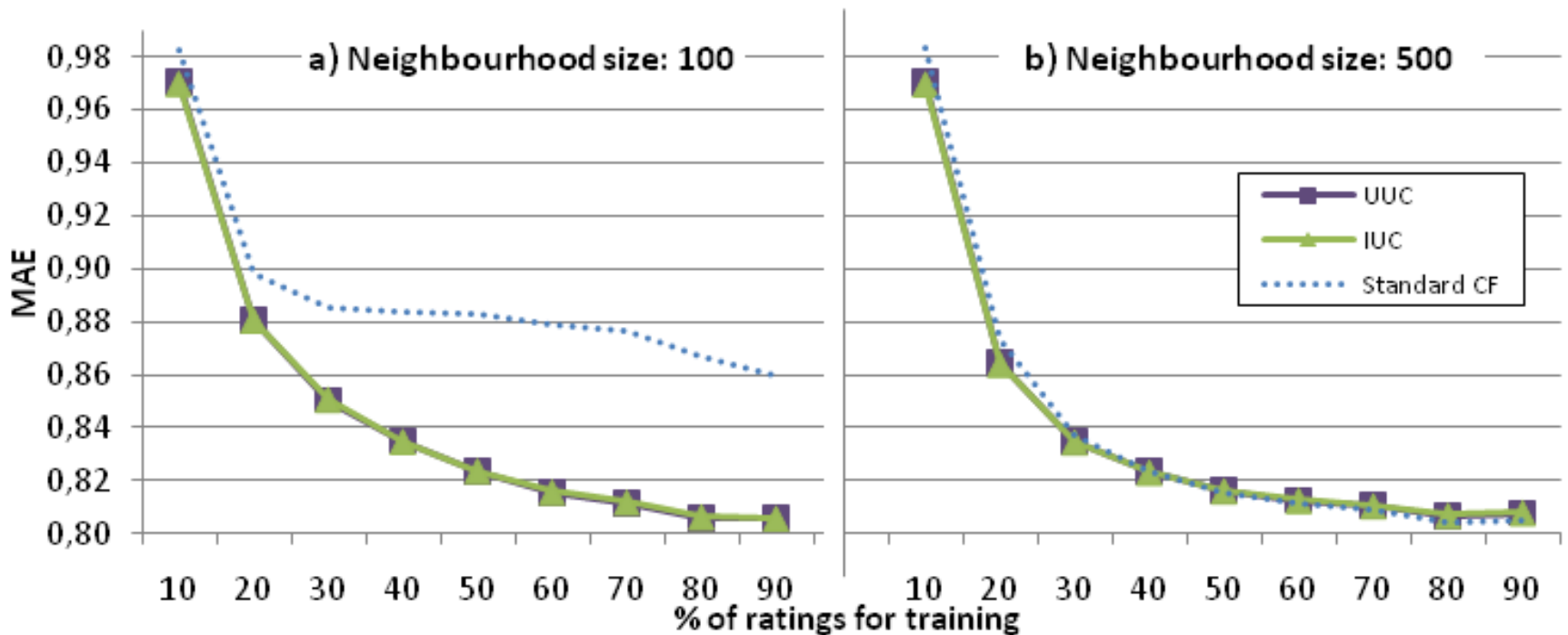
# Ponderación de vecinos en filtrado colaborativo

## ■ Resultados:

- Medida de evaluación: MAE (error medio promedio)

$$\text{MAE} = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N |r_{m,n} - p_{m,n}|$$

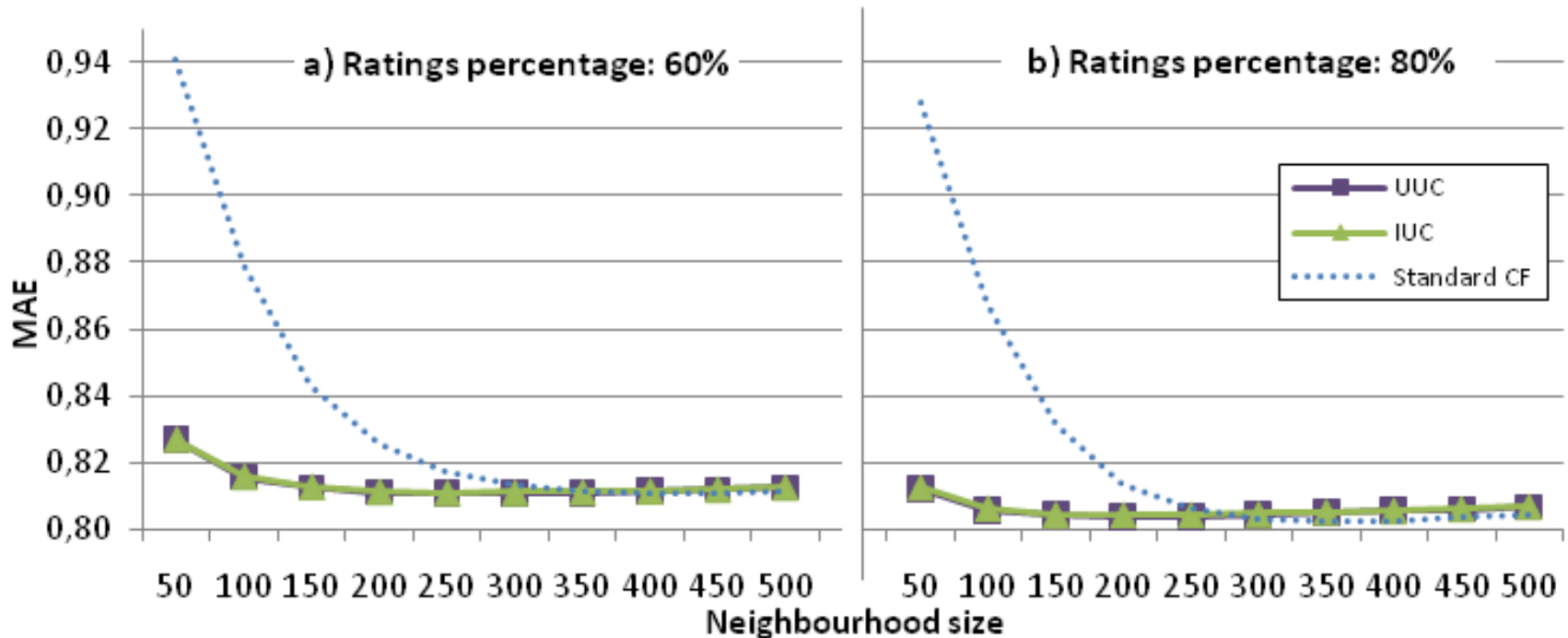
- Comportamiento según la densidad de ratings:



# Ponderación de vecinos en filtrado colaborativo

- Resultados (cont.):

- Comportamiento según el tamaño del vecindario (tamaño de  $N[u]$ )



# Ponderación de vecinos en filtrado colaborativo

- Resultados (cont.):

- Análisis de correlación con respecto a una nueva medida: *Neighbour Goodness* (bondad del vecino)

$$\text{neighbour\_goodness}(u) = \text{MAE}(\mathcal{U} - \{u\}, \mathcal{R} - \mathcal{R}(u)) - \text{MAE}(\mathcal{U} - \{u\}, \mathcal{R})$$

- Correlación de Pearson a nivel de 5%

Predictor de eficacia	% de ratings								
	10	20	30	40	50	60	70	80	90
UUC	-0.23	0.21	0.26	0.22	0.21	0.20	0.19	0.18	0.15
IUC	-0.24	0.17	0.19	0.15	0.14	0.13	0.13	0.13	0.09



# Ponderación en recomendación híbrida

- Formalización:

$$g(u_m, i_k) = \alpha_{CB} \times g_{CB}(u_m, i_k) + \alpha_{CF} \times g_{CF}(u_m, i_k)$$

$$\alpha_{CB} = \gamma_{CB}$$

$$\alpha_{CF} = \gamma_{CF}$$

$$g(u_m, i_k) = \lambda \times g_{CB}(u_m, i_k) + (1 - \lambda) \times g_{CF}(u_m, i_k)$$

# Ponderación en recomendación híbrida

## ▪ Predictores utilizados:

### • Componente basada en contenido ( $\gamma_{CB}$ )

– *Item features* (IF):

$$\gamma_{CB} = \gamma_{CB}(i_k) = \text{IF}(i_k) = \sum_{t \in \mathcal{I}} \text{TF-IDF}(i, t)$$

– *Item information gain* (IIG):

$$\gamma_{CB} = \gamma_{CB}(i_k) = \text{IIG}(i_k) = \frac{p_1^i - p_0^i}{p_1^i}$$

– *Item-based user clarity* (IUC):

$$\gamma_{CB} = \gamma_{CB}(v_j) = \text{IUC}(v_j)$$

### • Componente colaborativa ( $\gamma_{CF}$ )

– *Item-based item clarity* (IIC):

$$\gamma_{CF} = \gamma_{CF}(i_k) = \text{IIC}(i_k) = \sum_{u \in U} p(u|i) \log_2 \frac{p(u|i)}{p_c(u)}$$

– *User-based user clarity* (UUC):

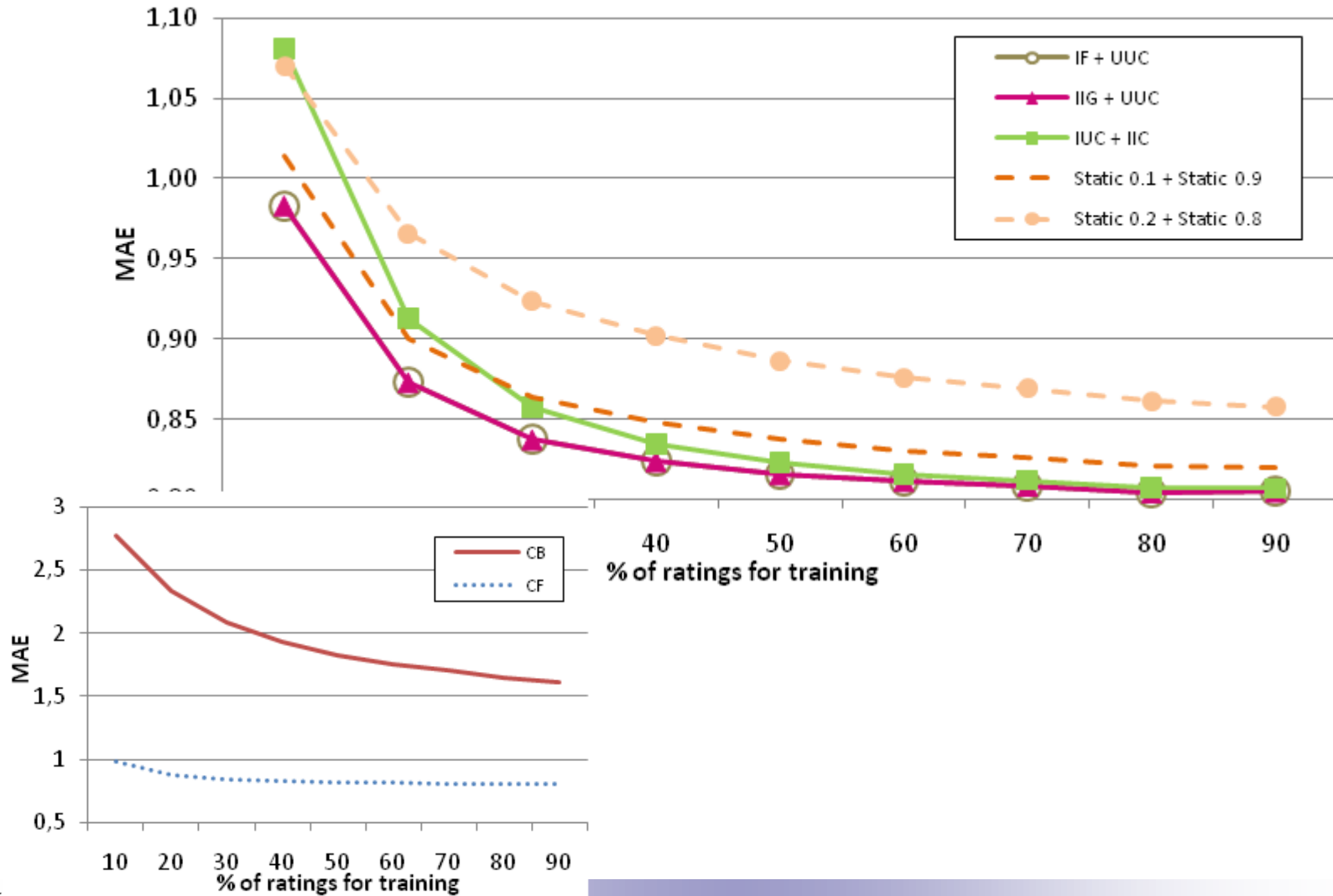
$$\gamma_{CF} = \gamma_{CF}(v_j) = \text{UUC}(v_j)$$

## ▪ También consideramos la referencia estática como predictor.

– Referencia estática, de parámetro  $\lambda$  :  $\gamma_{CB} = \gamma_{CF} = \gamma^\lambda = \lambda$

# Ponderación en recomendación híbrida

## ■ Resultados

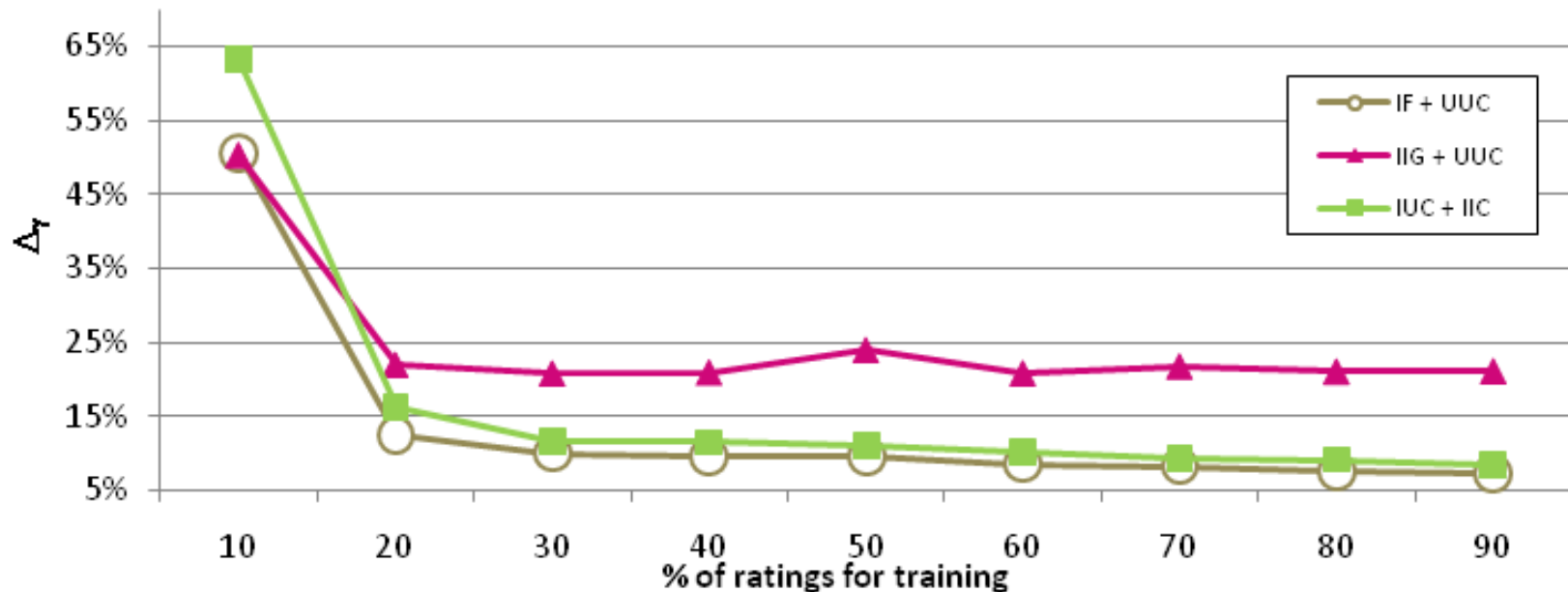


# Ponderación en recomendación híbrida

- MAE no es suficientemente discriminativo (en este contexto)
- Nueva medida: reflejar la mejora producida al escoger los pesos según la combinación dinámica frente a la estática

$$\Delta_\gamma = \frac{100}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{\text{MAE}(S(\alpha_u(u))) - \text{AE}_\gamma(u)}{\text{MAE}(S(\alpha_u(u)))}$$

$$\alpha_u = \frac{\alpha_{CB}}{\alpha_{CB} + \alpha_{CF}}$$
$$\gamma = (\gamma_{CB}, \gamma_{CF})$$



# Ponderación en recomendación híbrida

- **Análisis de correlación con respecto al error promedio** (por usuario / objeto)
  - Correlación con respecto a FC

Predictor de eficacia	% de ratings								
	10	20	30	40	50	60	70	80	90
<b>UUC</b>	-0.44	0.07	0.19	0.20	0.18	0.17	0.17	0.16	0.12
<b>IIC</b>	0.21	0.34	0.31	0.27	0.24	0.21	0.18	0.16	0.13

- Correlación con respecto a BC

Predictor de eficacia	% de ratings								
	10	20	30	40	50	60	70	80	90
<b>IIG</b>	-0.09	-0.07	-0.08	-0.08	-0.10	-0.10	-0.11	-0.10	-0.10
<b>IUC</b>	0.17	0.18	0.16	0.15	0.14	0.14	0.12	0.10	0.09
<b>IF</b>	0.06	0.00	N/A	N/A	-0.07	-0.07	-0.01	-0.03	0.08

- Los predictores basados en claridad: correlación positiva tanto en CB como en CF
- IIG e IF tienen valores negativos, aunque poco significativos

# Conclusiones

- Contribuciones del trabajo:
  - Marco formal para introducir predicción de eficacia en sistemas de recomendación
  - Adaptación de técnicas de claridad a sistemas de recomendación
  - Definición de nuevos predictores para sistemas de recomendación, basados en Teoría de la Información
  - Aplicación a dos problemas:
    - Ponderación de vecinos en filtrado colaborativo
    - Ponderación de recomendación híbrida
  - Validación experimental de la propuesta:
    - Analizando los resultados en el sistema final, donde los predictores han sido introducidos
    - Estudiando la correlación entre predictores y medidas de precisión
  - Propuesta de dos nuevas medidas de precisión:  $NG$  y  $\Delta_\gamma$

# Conclusiones

- Trabajo futuro:
  - Mejorar los predictores ya existentes y definir nuevos (basados en JSD o WIG)
  - Análisis de los predictores (cambiando la definición de  $\psi$ )
  - Creación de conjuntos de datos específicos
  - Experimentos usando conjuntos de datos mayores
  - Investigar medidas de precisión (propiedades, comportamiento)
  - Extensión del marco formal
  - Nuevas áreas de interés: búsqueda personalizada, RI basada en contexto, meta-búsqueda, búsqueda distribuida

Gracias



# Referencias

- (Amati et al. 2004) Amati, G., Carpineto, C. & Romano, G. (2004), ‘Query difficulty, robustness, and selective application of query expansion’, *Advances in Information Retrieval* pp. 127–137.
- (Aslam & Pavlu 2007) Aslam, J. A. & Pavlu, V. (2007), Query hardness estimation using Jensen-Shannon divergence among multiple scoring functions, in ‘ECIR’, pp. 198–209.
- (Carmel et al. 2006) Carmel, D., Yom-Tov, E., Darlow, A. & Pelleg, D. (2006), What makes a query difficult?, in ‘SIGIR ’06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval’, ACM, New York, NY, USA, pp. 390–397.
- (Castells et al. 2005) Castells, P., Fernández, M., Vallet, D., Mylonas, P. & Avrithis, Y. (2005), Self-tuning personalized information retrieval in an ontology-based framework, in R. Meersman, Z. Tari & P. Herrero, eds, ‘SWWS’05: In proceedings of the 1st International Workshop on Web Semantics’, Vol. 3762, Springer, pp. 977–986.
- (Cronen-Townsend et al. 2002) Cronen-Townsend, S., Zhou, Y. & Croft, B. W. (2002), Predicting query performance, in ‘SIGIR ’02: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval’, ACM Press, New York, NY, USA, pp. 299–306.
- (Diaz 2007) Diaz, F. (2007), Performance prediction using spatial autocorrelation, in ‘SIGIR ’07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval’, ACM, New York, NY, USA, pp. 583–590.

# Referencias

- (Diaz & Jones 2004) Diaz, F. & Jones, R. (2004), Using temporal profiles of queries for precision prediction, *in* 'SIGIR '04: Proceedings of the 27th annual international conference on Research and development in information retrieval', ACM Press, pp. 18–24.
- (He & Ounis 2004) He, B. & Ounis, I. (2004), Inferring query performance using pre-retrieval predictors, *in* 'String Processing and Information Retrieval, SPIRE 2004', pp. 43–54.
- (Jensen et al. 2005) Jensen, E. C., Beitzel, S. M., Grossman, D., Frieder, O. & Chowdhury, A. (2005), Predicting query difficulty on the web by learning visual clues, *in* 'SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval', ACM, New York, NY, USA, pp. 615–616.
- (Mothe & Tanguy 2005) Mothe, J. & Tanguy, L. (2005), Linguistic features to predict query difficulty, *in* 'Predicting Query Difficulty - Methods and Applications, SIGIR 2005'.
- (Yom-Tov et al. 2005) Yom-Tov, E., Fine, S., Carmel, D. & Darlow, A. (2005), Metasearch and federation using query difficulty prediction, *in* 'Predicting Query Difficulty - Methods and Applications, SIGIR 2005'.
- (Zhou & Croft 2006) Zhou, Y. & Croft, B. W. (2006), Ranking robustness: a novel framework to predict query performance, *in* 'CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management', ACM, New York, NY, USA, pp. 567–574.
- (Zhou & Croft 2007) Zhou, Y. & Croft, B. W. (2007), Query performance prediction in web search environments, *in* 'SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval', ACM, New York, NY, USA, pp. 543–550.