

Active Learning en Sistemas Recomendadores

Denis Parra

IIC 3633 – Sistemas Recomendadores

PUC Chile

27 de Octubre de 2015

Agenda Semestral

6 - 8 Oct	User centric evaluation + User interfaces	Prof. Denis Parra
13 - 15 Oct	Context-aware recommenders	Prof. Denis Parra
20 - 22 Oct	Context-aware recommenders II	Prof. Denis Parra
27 - 29 Oct	Active/Reinforcement Learning Recommender Systems	Gabriel della Maggiora y Javier Machin
3 - 5 Nov	Graph-based recommendation	Juan Pablo Salazar y Christopher Arenas
10 - 12 Nov	Applications: music	Miguel Fadic
17 - 19 Nov	Modelos graficos probabilisticos para sistemas recomendadores	Laura Cruz (invitada)

Proyecto #2

- A Recommender System for the ALMA Science Portal
- Problema: Sistema actual sólo permite búsquedas, no Navegar/Explorar proyectos
- Solución: Construir un item-based recommender basado en contenido (abstracts) e información espacial
- Dataset: proyectos de alma (Cycle-0 to 3)
- Métricas: No se indican, sólo un potencial ground-truth basado en logs.

Observaciones Proyecto #2

- Qué técnicas se usarán para encontrar similitud en:
 - Información textual (reducción de dimensionalidad)
 - Información espacial (clarificar mejor)
- Métrica: clarificar. Dado que no hay ratings, usar Mean Percentile Rank.
- Evaluar con usuarios (hacer un estudio pequeño con 5-10 personas)

Proyecto #7

- Recomendación de Inscripción de Cursos a Estudiantes UACH
- Problema: Dificultad de los alumnos para elegir cantidad y qué cursos tomar al comienzo del semestre. Objetivo es reducir riesgo de perder becas, terminando carrera a tiempo.
- Dataset: Pregrado UACH, cubre 11 años
- Métricas: Realizar una relación entre ratings/notas, intentar predecir nota o probabilidad de aprobación.
- Antecedentes: Papers de recomendaciones con requisitos.

Observaciones Proyecto #7

- No queda claro con el abstract cómo conectar el problema de optimización con restricciones (donde se intenta maximizar el flujo) a un modelo clásico de recomendación híbrida como el que se propone en el trabajo => ¿cómo maximizar el flujo usando UB y IB ?
- Se podría simplificar el problema en una primera parte: Si mi modelo predice muy bien las notas (bajo RMSE o MAE), pero predice que la nota va a ser baja, el sistema debería recomendar no tomar el ramo. Además de una buena predicción debe estar acompañado de decisiones alternativas si la predicción es mala (nota bajo 4) y viceversa.
- Otra estrategia podría ser la de J. Machin: usando Active Learning, ¿cuál deber ser el próximo item a ratear para disminuir el RMSE, penalizando en el modelo cada vez que se tomar un curso nuevo?
- Dar tratamiento especial a usuarios con beca en el modelo.
- Métrica: Porcentaje de usuarios a quienes se les recomendaron items que aprobaron –o- items que reprobabron pero no los hicieron atrasar su carrera.

Proyecto #8

- Explotando Clustering para mejorar(?) de recomendaciones
- Problema: Limitación de métricas como Pearson y Cosine, no consideran la “estructura del cluster” y de métodos de clustering que restringen el espacio de búsqueda.
- Solución: Usar índices de similaridad que consideren la distancia basada en la estructura misma del clúster
- Dataset(s): MovieLens y Jester
- Métricas: MAE/RMSE, Precision/Recall y nDCG.

Observaciones Proyecto #8

- No se menciona claramente qué motivó a Sarwar (1999) y a O'Connor y Herlocker a utilizar clustering. ¿mejorar performance? ¿mejorar escalabilidad?
- No es claro por qué se puede esperar que “usando índices de similaridad que consideren la distancia basada en la estructura misma del clúster” deberían dar mejores resultados. ¿hay alguna investigación previa que generar esta hipótesis?
- ¿Cuál será el baseline? ¿contra qué métodos se compara?
- ¿Cuál es el criterio para la elección de los dataset? Quizás un análisis previo de los mismos datasets pueda ayudar a identificar un par más donde es posible encontrar las diferencias esperadas.
- Si se desea considerar la estructura local de los datos, ¿hay alguna expectativa de usar Spectral Clustering? Revisar poster de Bellogin, 2012.

Active Learning en Sistemas Recomendadores

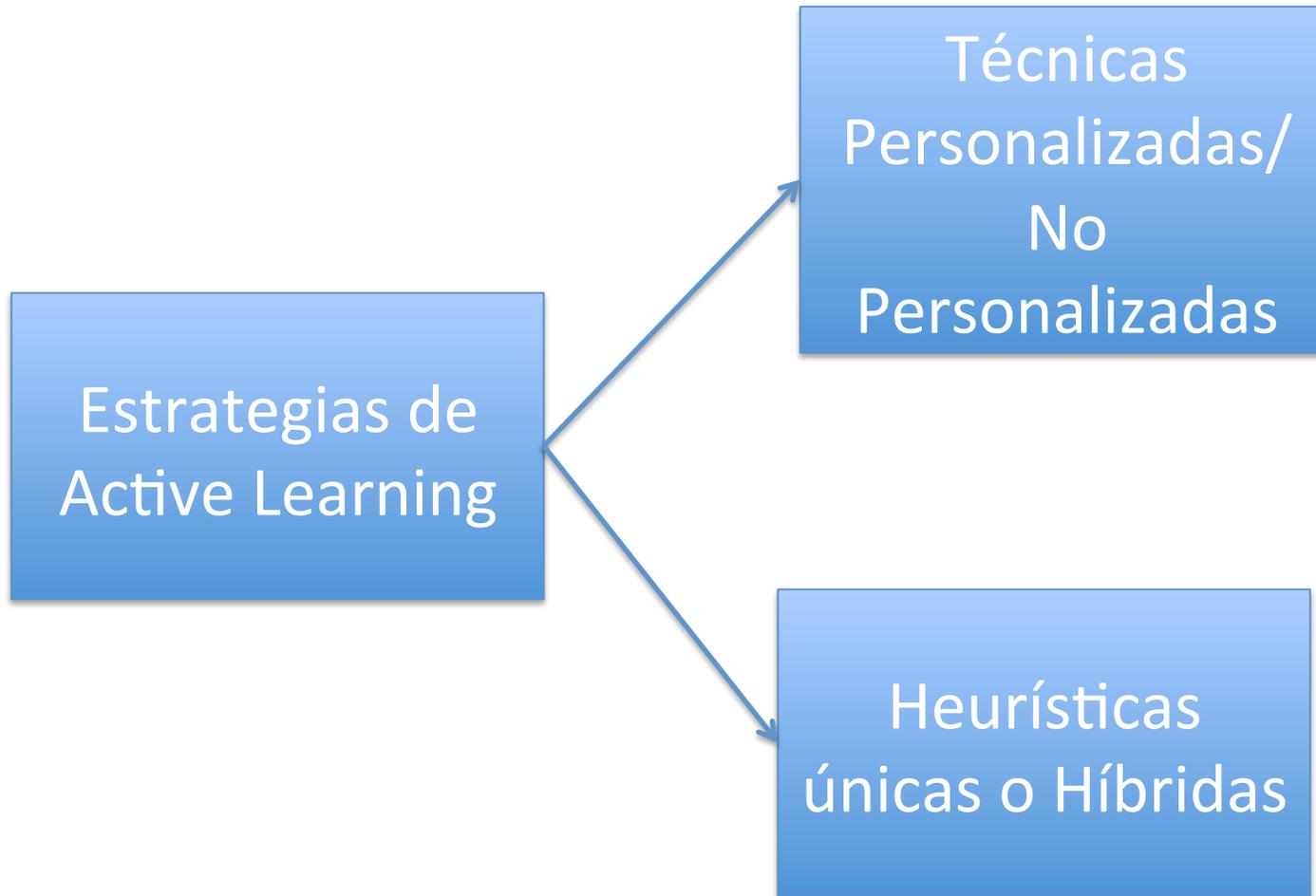
El Problema de Recomendación

- ... hasta ahora, es predecir ratings or rankear items dado un dataset existente
- Sin embargo, los sistemas evolucionan de forma dinámica
- Una estrategia de Aprendizaje Activo en Filtrado Colaborativo implica una estrategia precisa para seleccionar items de forma que el usuario provea ratings y eso nos permita mejorar nuestro sistema.

A.L. en el problema de Recomendación

- New User
- New Item
- Costo de Obtener Feedback
- Adaptación de métodos de Aprendizaje Activo

Clasificación de Técnicas Existentes



Técnicas Personalizadas/No Personalizadas

- No Personalizadas: Solicitar a los usuario dar feedback (ratings) a los mismos items. Esto podria ayudar a mejorar el sistema en general pero quizás en detrimento de preferencias de usuarios.
- Personalizadas: Seleccionar los items considerando los gustos o preferencias del usuario, o que permitan de forma más efectiva obtener esta información

Herísticas únicas o Hibridizadas

- Single-Heuristic: Se considera sólo una regla de selección, la cual se utiliza para seleccionar items.
- Combined-Heuristic: Permitan construir una estrategia Híbrida al agregar o combinar varias opciones con potencial de mejorar el rendimiento del sistema y, por ende, ser usadas para ratings de usuarios.

Ejemplo Ilustrativo

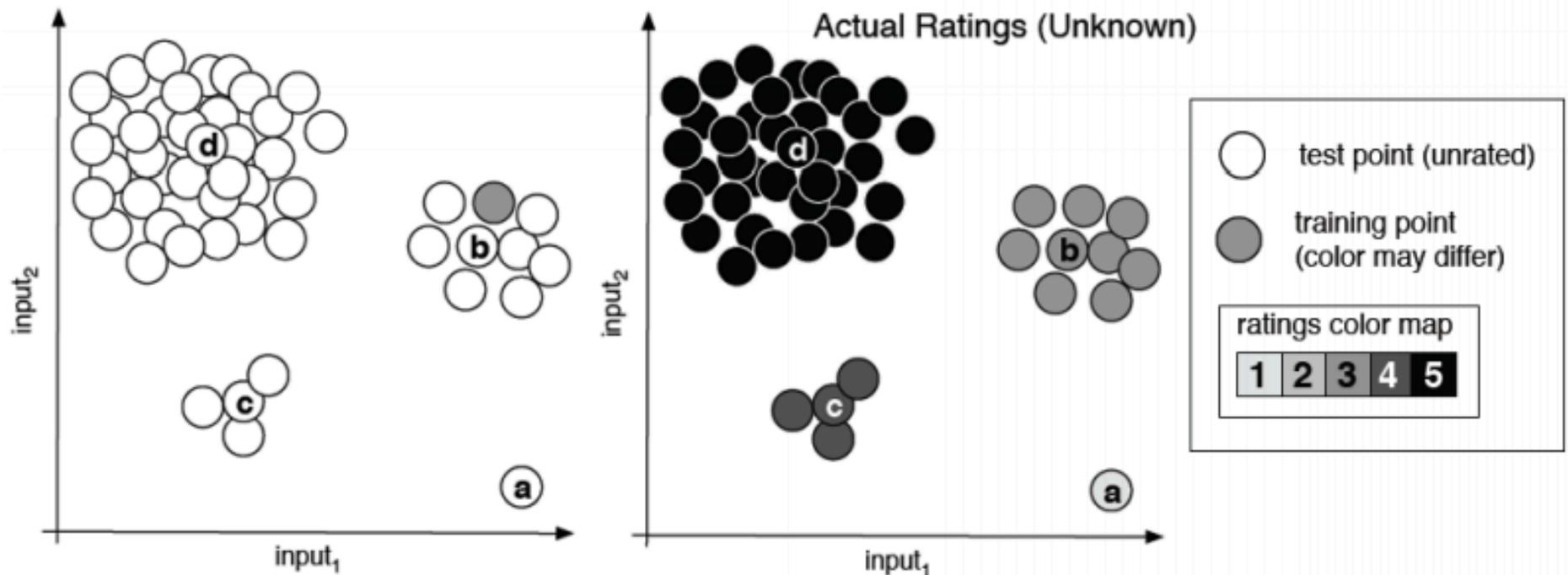
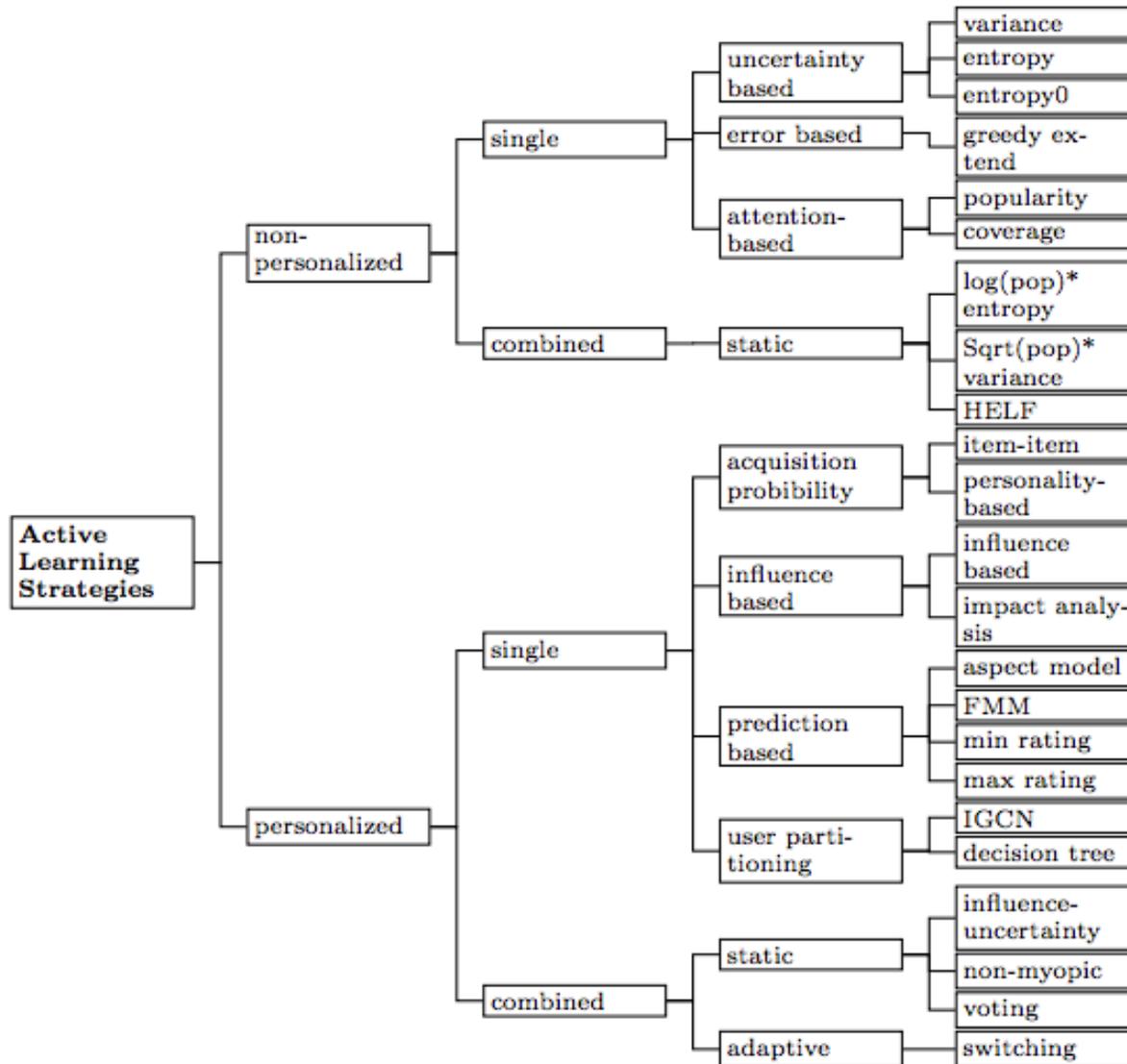


Fig. 1. Active learning, an illustrative example [19]

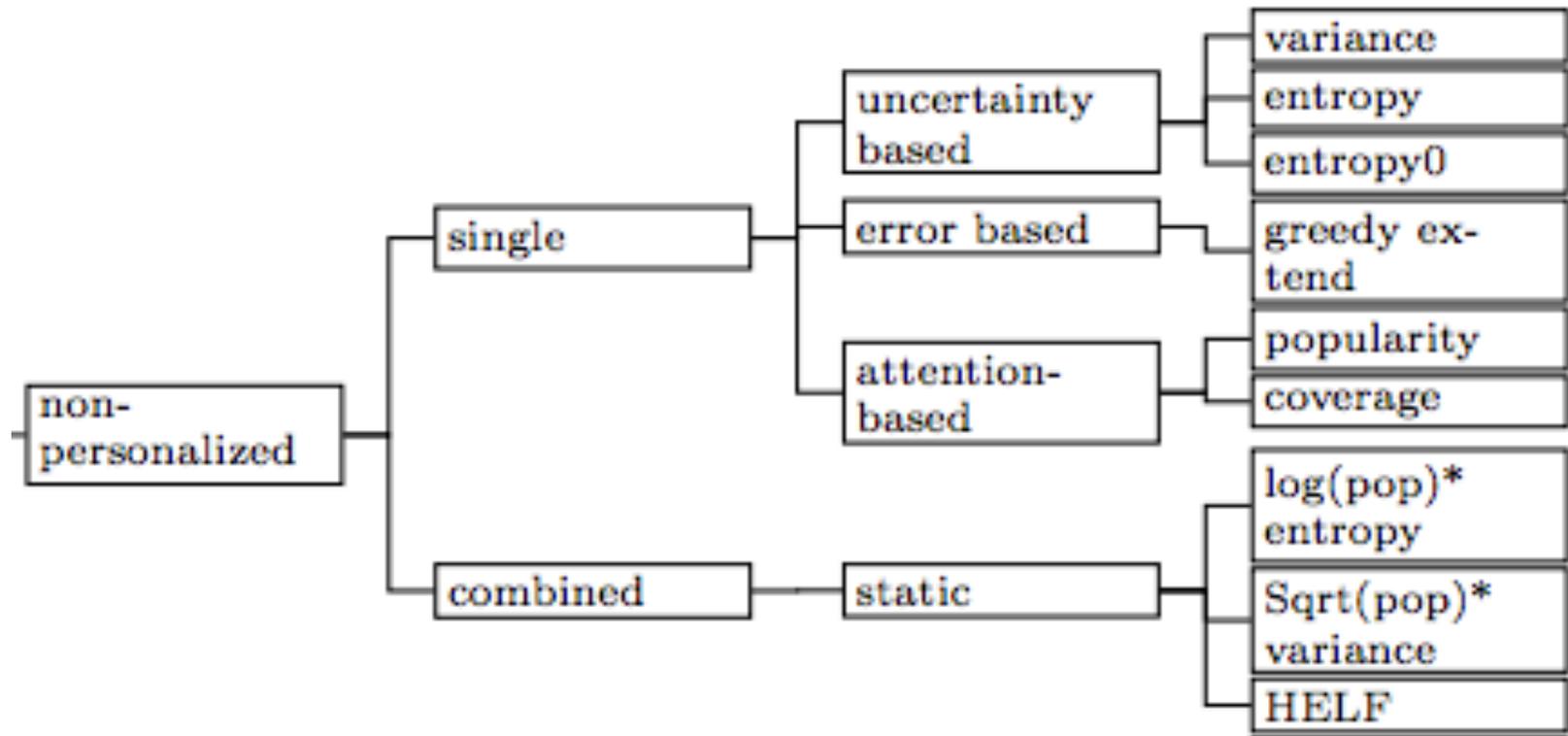
Propiedades de los Datos

- (R1) Representado (e.g. Trilogía Star Wars, disminuir redundancia)
- (R2) Representativo (e.g. elegir items que representen tipos generales: Drama mejor que Zombie)
- (R3) Resultados: Elegir este punto mejorará mi predicción o algún otro objetivo?

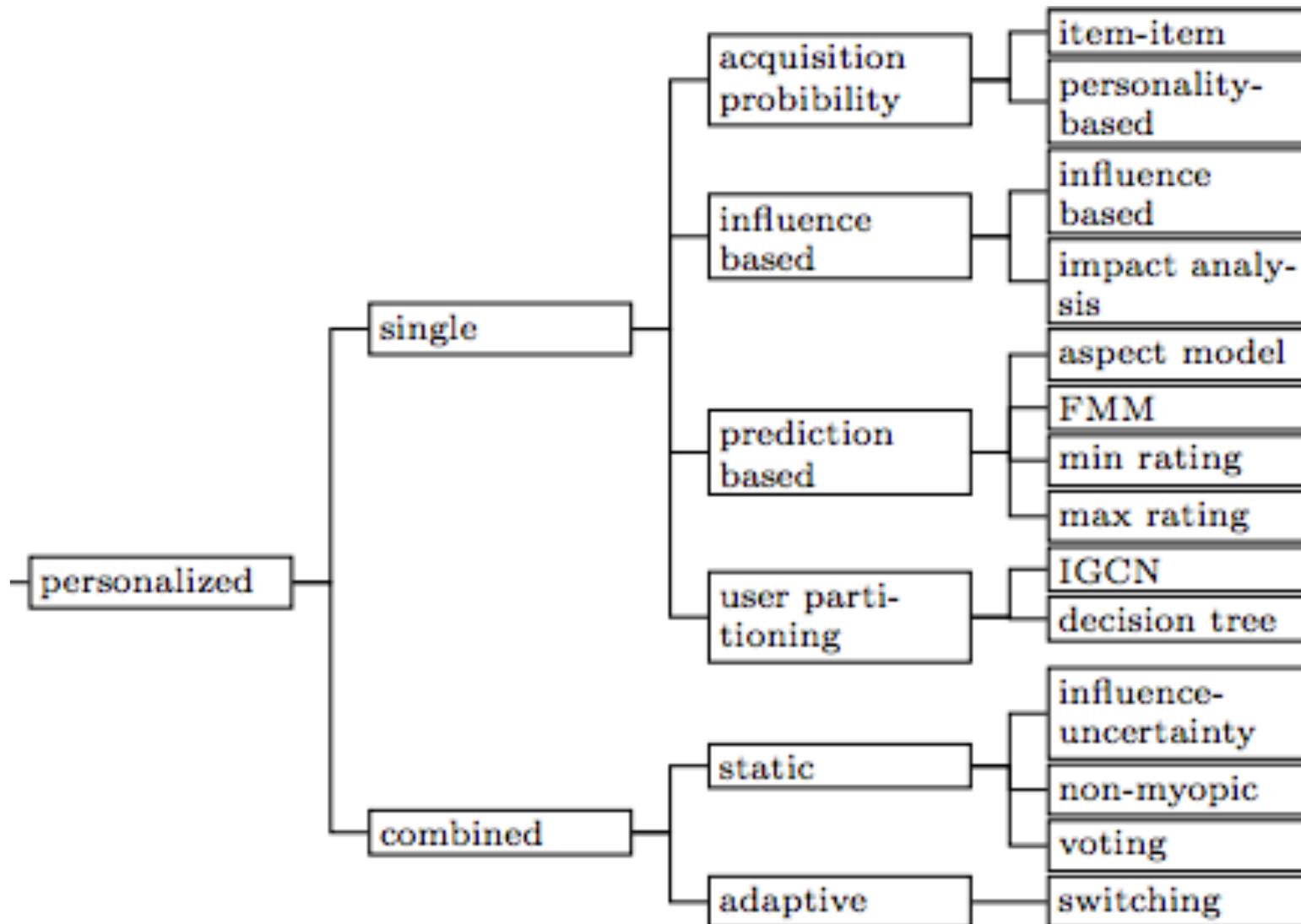
Jerarquía de Estrategias



Non-Personalized (Several Heuristics)



Personalized (Several Heuristics)



Aprendizaje Activo No-Personalizado

- Estrategias de Heurística única
 - Basadas en Incerteza
 - Basadas en Error
 - Basadas en Atención

Estrategias Basadas en Incerteza

- Varianza

$$\text{var}(y|x) = \frac{1}{|Y_x|} \sum_{y \in Y_x} (y - \bar{y}_x)^2,$$

- Entropía ~ Entropía0

$$H = - \sum p(x) \log p(x)$$

Estrategias Basadas en Error

- Mejorar directamente la exactitud de predicción del sistema
- Greedy Extend (Golbandi et al. 2010)

$$S = \operatorname{argmin}_{S \subset \mathcal{I}, |S|=k} \mathcal{F}(\mathcal{A}(S))$$

Estrategias Basadas en Atención

- Fáciles de implementar, propuestas iniciales para resolver el cold-start
 - Popularidad: Elegir items que han recibido la mayor cantidad de items
 - Coverage: Elegir items que han sido co-rated por muchos usuarios

$$Coverage(i) = \sum_j n_{ij}$$

Aprendizaje Activo No-Personalizado

- Estrategias de Heurística Combinada (Estáticas)
 - $\text{Log}(\text{pop}) * \text{entropy}$
 - $\text{SQRT}(\text{pop}) * \text{entropy}$
 - HELF

$F(\text{pop}) * \text{entropy}$

- $\text{Log}(\text{pop}) * \text{entropy}$: Intenta combinar el efecto de popularidad con la entropía de los ratings
- $\text{Sqrt}(\text{pop}) * \text{entropy}$: Variación usa raíz cuadrada en lugar de $\text{Log}()$ y en algunos casos, varianza en lugar de entropía

HELF

- Harmonic Mean de Entropía y Logaritmo de la Frecuencia
- Combina popularidad con “informativeness”. Funciona bien porque la entropía tiende a seleccionar items que rara vez han sido rated.

$$HELF(i) = \frac{2 \times LF_i \times H(i)}{LF_i + H(i)}$$

Comparación

- Basada en paper “On Bootstrapping Recommender Systems” de Golbandi et al. (2010)

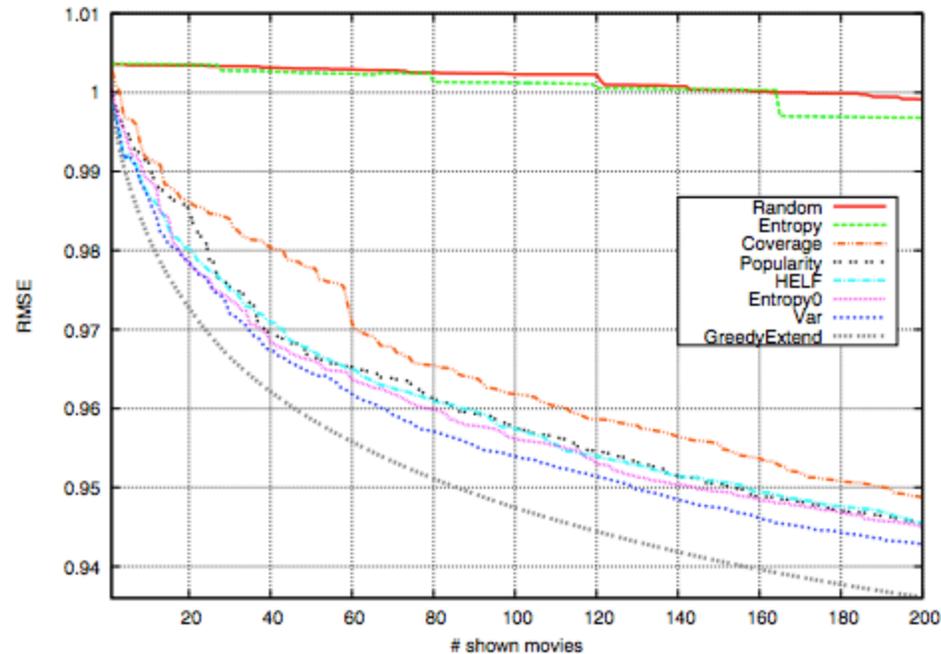


Figure 2: The test error rate vs. number of displayed items (=size of seed set), for various methods of selecting seed set items. Methods that disregard item popularity (Random and Entropy) significantly lag in performance. GreedyExtend delivers the best performing seed sets by guiding the set creation process with a suitable cost function. Note that the legend orders methods by their performance.

Aprendizaje Activo Personalizado

- Estrategias de Heurística única
 - Probabilidad de Adquisición
 - Basadas en Influencia
 - Basadas en Predicción
 - Particionamiento de los Usuarios

Probabilidad de Adquisición

- Estrategia Item-Item
- Predicción Binaria Basada en Personalidad

Basada en Influencia

- Basada Directamente en Influencia
- Análisis de Impacto

Basada en Predicción

- Modelo de Aspecto
- MinRating
- MinNorm

Particionamiento de los Usuarios

- IGCN: Information Gain a través de clustering neighbors
- Árbol de Decisión

Aprendizaje Activo Personalizado

- Estrategias de Heurística Combinada
 - Estáticas
 - Incerteza
 - No miópica
 - Votación
 - Adaptativas
 - Switching

A.A. Personalizado: Incerteza

- Incerteza propiamente
- No miópica
- Votación
- Switching

