Recommendation and Information Retrieval: Two sides of the same coin?

Prof.dr.ir. Arjen P. de Vries

arjen@acm.org

CWI, TU Delft, Spinque

CWI

Outline

- Recommendation Systems
  - Collaborative Filtering (CF)
- Probabilistic approaches
  - Language modelling for Information Retrieval
  - Language modelling for log-based CF
  - Brief: adaptations for rating-based CF
- Vector Space Model (“Back to the Future”)
  - User and item spaces, orthonormal bases and “the spectral theorem”

CWI

Recommendation

- Informally:
  - Search for information “without a query”
- Three types:
  - Content-based recommendation
  - Collaborative filtering (CF)
    - Memory-based
    - Model-based
  - Hybrid approaches

CWI

Recommendation

- Informally:
  - Search for information “without a query”
- Three types:
  - Content-based recommendation
  - Collaborative filtering
    - Memory-based
    - Model-based
  - Hybrid approaches

Today’s focus!

Online News

Content-based

deVolkskrant

Paradiso begint eigen platenlabel

Het Amsterdamse poppodium Paradiso begint een eigen platenlabel: de Paradiso Vinyl Club. Hiermee brengt het podium alleen werk uit van beginnende Nederlandse muziekacts.

Door: Jelmer Luimstra 11 februari 2015, 17:17 3

Het idee: bands leveren zelf de opnamen en het artwork, en Paradiso brengt hun single uit. De popzaal wil acht keer per jaar een 7-inch single op de

CWI

Arjen P. De Vries
17 mins ·

First public announcement: I will be moving jobs (and house) to beautiful Nijmegen!!

Proud to be taking up the chair of Information Retrieval, even though I feel also sad to leave behind so many friends in Amsterdam and Utrecht.

Oh!!! Nijmegen is REALLY CLOSE BY people!!

Arjen de Vries appointed as full professor of Information Retrieval

Arjen P. de Vries appointed as full professor of Information Retrieval at Radboud University.

WWW.RU.NL

Like Comment Share

Jeremy Pickens, Max Hinne, Martine Zwiers and 22 others like this.

Mounia Lalmas-Roelleke Fantastic news.
Unlike · Reply · 1 · 15 mins

Miriam Gravemaker Gefeliciteerd!
Unlike · Reply · 1 · 9 mins

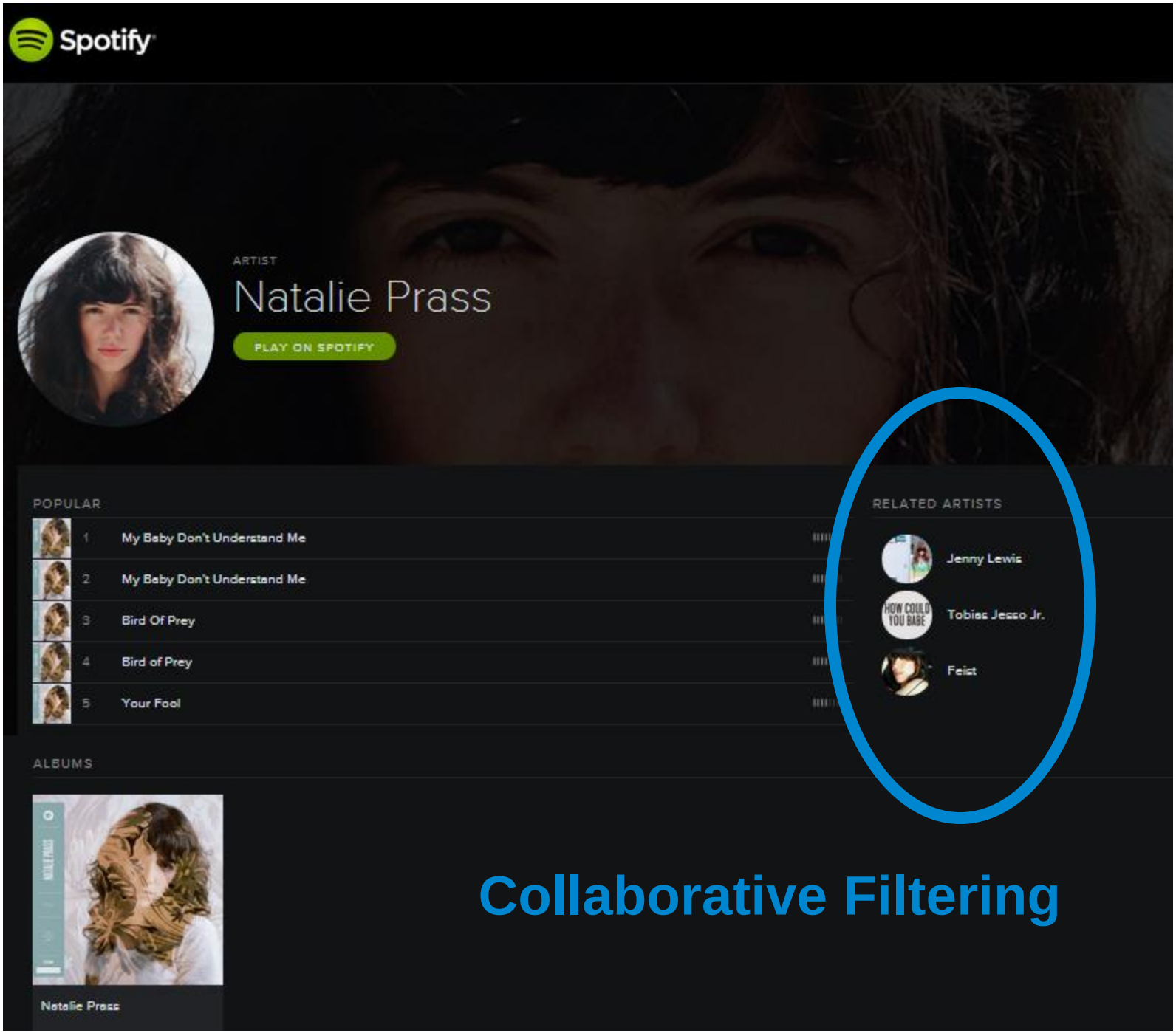
Mani Zandifar Congratulations!
Unlike · Reply · 1 · 8 mins

Katya Mourits Congratulations! And wow, big change!
Unlike · Reply · 1 · 7 mins

Raffaele Perego Great! Congratulations Arjen!
Unlike · Reply · 1 · 6 mins

CWI

Music



Collaborative Filtering

Collaborative Filtering

- Collaborative filtering (originally introduced by Patti Maes as "social information filtering")
  1. Compare user judgments
  2. Recommend differences between similar users

- Leading principle:
  People's tastes are not randomly distributed
  - *A.k.a.* "You are what you buy"

CWI

Rating Matrix

	 Die Hard	 Mission: Impossible	 GoldenEye	 Casino Royale	 Titanic	 Notting Hill	 Bridget Jones's Diary	 Love Actually
Boris		★★★★☆	★★★★★		★★★★☆	★☆☆☆☆	★★★☆☆	
Dave		★★★★★	★★★★★	★★★★★	★★☆☆☆			★☆☆☆☆
Will					★★★★★	★★★★★	★★★☆☆	★★★★☆
George	★★★★☆	★★★★★		★★★★☆		★★☆☆☆	★★★★☆	★★☆☆☆

CWI

Users

	 Die Hard	 Mission: Impossible	 GoldenEye	 Casino Royale	 Titanic	 Notting Hill	 Bridget Jones's Diary	 Love Actually
Boris		★★★★☆	★★★★★		★★★★☆	★☆☆☆☆	★★★☆☆	
Dave		★★★★★	★★★★★	★★★★★	★★☆☆☆			★☆☆☆☆
Will					★★★★★	★★★★★	★★★☆☆	★★★★☆
George	★★★★☆	★★★★★		★★★★☆		★★☆☆☆	★★★★☆	★★☆☆☆

CWI

Items

	 Die Hard	 Mission: Impossible	 GoldenEye	 Casino Royale	 Titanic	 Notting Hill	 Bridget Jones's Diary	 Love Actually
Boris		★★★★☆	★★★★★		★★★★☆	★☆☆☆☆	★★★☆☆	
Dave		★★★★★	★★★★★	★★★★★	★★☆☆☆			★☆☆☆☆
Will					★★★★★	★★★★★	★★★☆☆	★★★★☆
George	★★★★☆	★★★★★		★★★★☆		★★☆☆☆	★★★★☆	★★☆☆☆

CWI

Rating

	 Die Hard	 Mission: Impossible	 GoldenEye	 Casino Royale	 Titanic	 Notting Hill	 Bridget Jones's Diary	 Love Actually
Boris		★★★★☆	★★★★★		★★★★☆	★☆☆☆☆	★★★☆☆	
Dave		★★★★★	★★★★★	★★★★★	★★☆☆☆			★☆☆☆☆
Will					★★★★★	★★★★★	★★★☆☆	★★★★☆
George	★★★★☆	★★★★★		★★★★☆		★★☆☆☆	★★★★☆	★★☆☆☆

CWI

User Profile

	 Die Hard	 Mission: Impossible	 GoldenEye	 Casino Royale	 Titanic	 Notting Hill	 Bridget Jones's Diary	 Love Actually
Boris		★★★★☆	★★★★★		★★★★☆	★☆☆☆☆	★★★☆☆	
Dave		★★★★★	★★★★★	★★★★★	★★☆☆☆			★☆☆☆☆
Will					★★★★★	★★★★★	★★★☆☆	★★★★☆
George	★★★★☆	★★★★★		★★★★☆		★★☆☆☆	★★★★☆	★★☆☆☆

CWI

Item Profile

	 Die Hard	 Mission: Impossible	 GoldenEye	 Casino Royale	 Titanic	 Notting Hill	 Bridget Jones's Diary	 Love Actually
Boris		★★★★☆	★★★★★		★★★★☆	★☆☆☆☆	★★★☆☆	
Dave		★★★★★	★★★★★	★★★★★	★★☆☆☆			★☆☆☆☆
Will					★★★★★	★★★★★	★★★☆☆	★★★★☆
George	★★★★☆	★★★★★		★★★★☆		★★☆☆☆	★★★★☆	★★☆☆☆

CWI

Unknown Rating

	 Die Hard	 Mission: Impossible	 GoldenEye	 Casino Royale	 Titanic	 Notting Hill	 Bridget Jones's Diary	 Love Actually
Boris		★★★★☆	★★★★★		★★★★☆	★☆☆☆☆	★★★☆☆	?
Dave		★★★★★	★★★★★	★★★★★	★★☆☆☆			★☆☆☆☆
Will					★★★★★	★★★★★	★★★☆☆	★★★★☆
George	★★★★☆	★★★★★		★★★★☆		★★★☆☆	★★★★☆	★★☆☆☆

Collaborative Filtering

	 Die Hard	 Mission: Impossible	 GoldenEye	 Casino Royale	 Titanic	 Notting Hill	 Bridget Jones's Diary	 Love Actually
Boris		★★★★☆	★★★★★		★★★★☆	★☆☆☆☆	★★★☆☆	?
Dave		★★★★★	★★★★★	★★★★★	★★☆☆☆			★☆☆☆☆
Will					★★★★★	★★★★★	★★★☆☆	★★★★☆
George	★★★★☆	★★★★★		★★★★☆		★★★☆☆	★★★★☆	★★☆☆☆

If user Boris watched *Love Actually*, how would he rate it?

Collaborative Filtering

- Standard item-based formulation
  (Adomavicius & Tuzhilin 2005)

$$\mathrm{rat}(u,i) = \sum_{j \in I_u} \frac{\mathrm{sim}(i,j)}{\sum_{j \in I_u} \mathrm{sim}(i,j)} \mathrm{rat}(u,j)$$

		 Die Hard	 Mission: Impossible	 GoldenEye	 Casino Royale	 Titanic	 Notting Hill	 Bridget Jones's Diary	 Love Actually
▶	Boris		★★★★☆	★★★★★		★★★★☆	★☆☆☆☆	★★★☆☆	?
	Dave		★★★★★	★★★★★	★★★★★	★★☆☆☆			★☆☆☆☆
	Will					★★★★★	★★★★★	★★★☆☆	★★★★☆
	George	★★★★☆	★★★★★		★★★★☆		★★★☆☆	★★★★☆	★★☆☆☆

Collaborative Filtering

- Benefits over content-based approach
  - Overcomes problems with finding suitable features to represent e.g. art, music
  - Serendipity
  - Implicit mechanism for qualitative aspects like style
- Problems: large groups, broad domains

Prediction vs. Ranking

- Original formulations focused on modelling the users' item *ratings*: **rating prediction**
  - Evaluation of algorithms (e.g., Netflix prize) by Mean Absolute Error (MAE) or Root Mean Square Error (RMSE) between predicted and actual ratings

CWI

Recency-based

Netflix Never Used Its $1 Million Algorithm Due To Engineering Costs

BY CASEY JOHNSTON, ARS TECHNICA 04.16.12 | 8:20 AM | PERMALINK

Share 0 Tweet 0

Netflix awarded a $1 million prize to a developer team in 2009 for an algorithm that increased the accuracy of the company's recommendation engine by 10 percent. But it doesn't use the million-dollar code, and has no plans to implement it in the future, Netflix announced on its blog Friday. The post goes on to explain why: a combination of too much engineering effort for the results, and a shift from movie recommendations to the "next level" of personalization caused by the transition of the business from mailed DVDs to video streaming.

Netflix notes that it does still use two algorithms from the team that won the first Progress Prize for an 8.43 percent improvement to the recommendation engine's root mean squared error (the full $1 million was awarded for a 10 percent improvement). But the increase in accuracy on the winning improvements "did not seem to justify the engineering effort needed to bring them into a production

netflixprize.jpg

By that time, the company had moved on anyway

Prediction vs. Ranking

- Original formulations focused on modelling the users' item *ratings*: **rating prediction**
  - Evaluation of algorithms (e.g., Netflix prize) by Mean Absolute Error (MAE) or Root Mean Square Error (RMSE) between predicted and actual ratings
- For the end user, the **ranking of recommended items** is the essential problem: **relevance ranking**
  - Evaluation by precision at fixed rank (P@N)

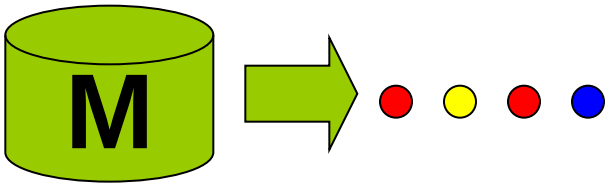
CWI

Relevance Ranking

- *Core problem of Information Retrieval!*

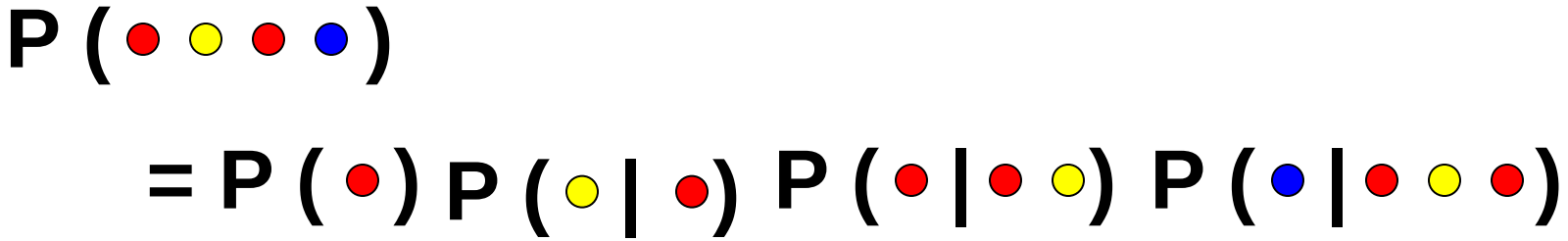
Generative Model

- A statistical model for generating data
  - Probability distribution over samples in a given ‘language’



$$P(\bullet\circ\bullet\bullet \mid M) = P(\bullet \mid M)\, P(\circ \mid M, \bullet)\, P(\bullet \mid M, \bullet\circ)\, P(\bullet \mid M, \bullet\circ\bullet)$$

Unigram models etc.

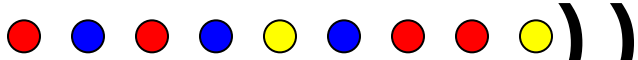


- Unigram Models

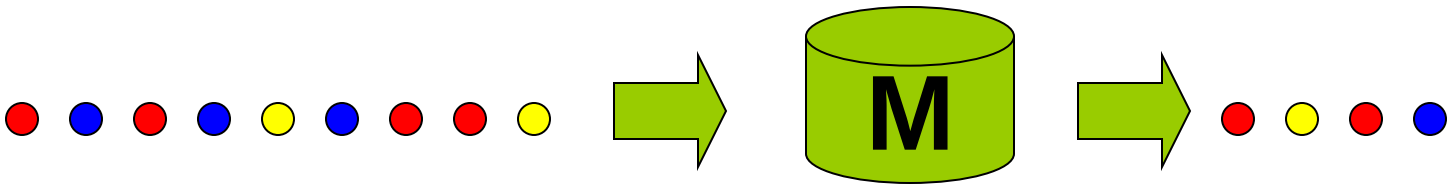
- N-gram Models (here, N=2)

Fundamental Problem

- Usually we don't know the model **M**
  - But have a sample representative of that model

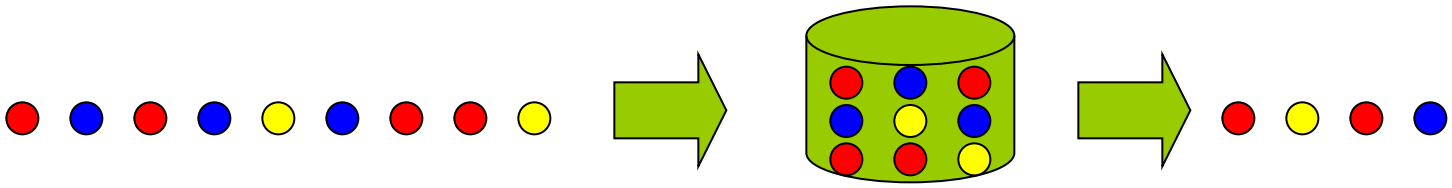
P ( **| M (**  **))**

- First estimate a model from a sample
- Then compute the observation probability



Language Models…

- Unigram Language Models (LM)
  - Urn metaphor



- P(● ○ ● ●) ~ P (●) P (○) P (●) P (●)

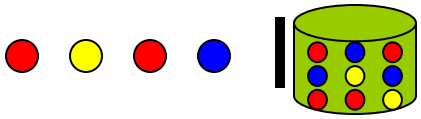
  = 4/9 * 2/9 * 4/9 * 3/9

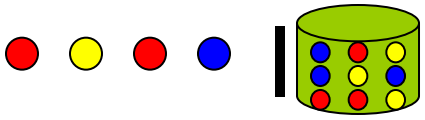
CWI

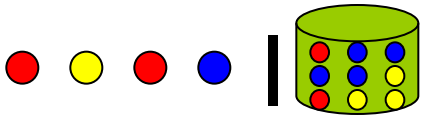
… for Information Retrieval

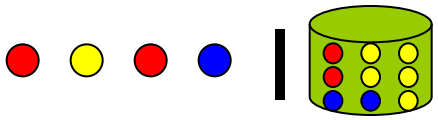
- Rank models (documents) by probability of generating the query:

  Q: 

  P( **)** = 4/9 * 2/9 * 4/9 * 3/9 = 96/9

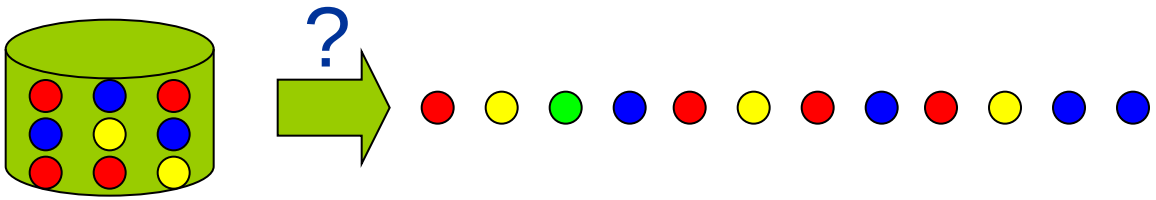
  P( **)** = 3/9 * 3/9 * 3/9 * 3/9 = 81/9

  P( **)** = 2/9 * 3/9 * 2/9 * 4/9 = 48/9

  P( **)** = 2/9 * 5/9 * 2/9 * 2/9 = 40/9

Zero-frequency Problem

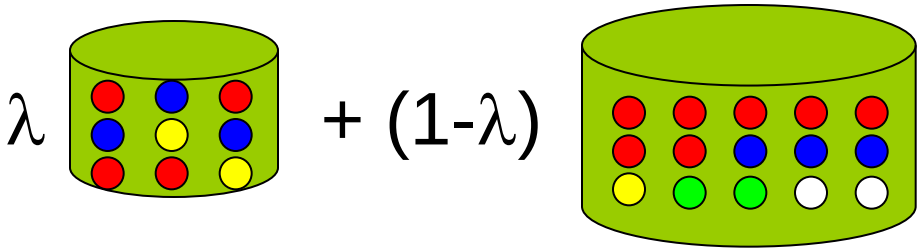
- Suppose some event not in our example
  - Model will assign zero probability to that event
  - And to any set of events involving the unseen event
- Happens frequently in natural language text, and it is incorrect to infer zero probabilities
  - Especially when dealing with incomplete samples



CWI

Smoothing

- Idea:
  Shift part of probability mass to unseen events
- Interpolate document-based model with a background model (of “general English”)
  - Reflects expected frequency of events
  - Plays role of IDF



CWI

Relevance Ranking

- *Core problem of Information Retrieval!*
  - Question arising naturally:
    Are CF and IR, from a modelling perspective, really two different problems then?

Jun Wang, Arjen P. de Vries, Marcel JT Reinders, A User-Item Relevance Model for Log-Based Collaborative Filtering, ECIR 2006

User-Item Relevance Models

- Idea: CF by a probabilistic retrieval model

User-Item Relevance Models

- Idea: CF by a probabilistic retrieval model
- Treat **user profile as query** and answer the following question:
  - **“ What is the probability that *this* item is relevant to *this* user, given his or her profile”**
- Hereto, apply the language modelling approach to IR as a formal model to compute the user-item *relevance*

Implicit or explicit relevance?

- Rating-based CF:
  - Users explicitly rate “items”

    We use “items” to represent contents (movie, music, etc.)

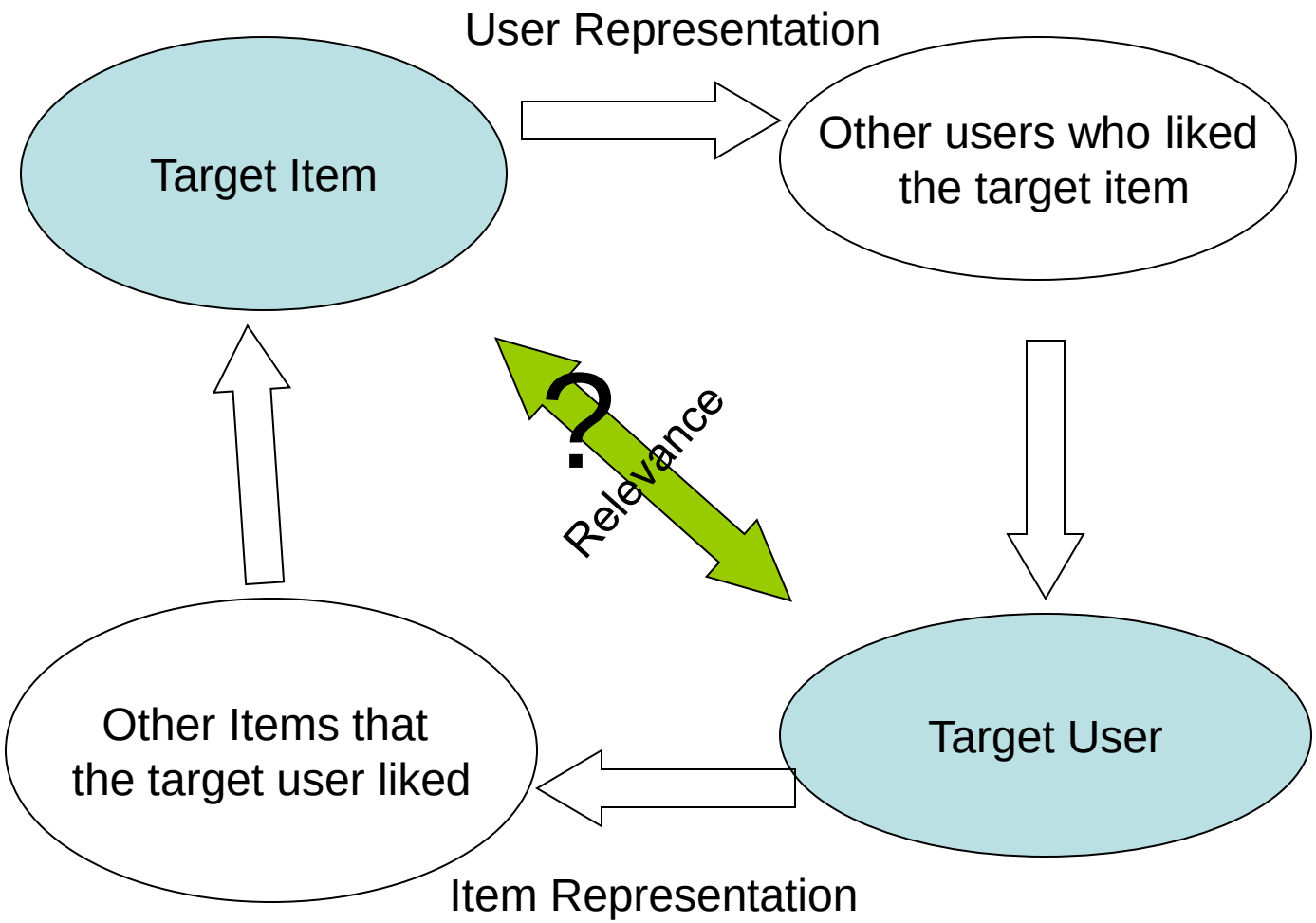
- Log-based CF:
  - User profiles are gathered by logging the interactions. Music play-list, web surf log, etc.

User-Item Relevance Models

- Existing User-based/Item-based approaches
  - Heuristic implementations of “word-of-mouth”
  - Unclear how to best deal with the sparse data!

- User-Item Relevance Models
  - Give probabilistic justification
  - Integrate smoothing to tackle the problem of sparsity

User-Item Relevance Models



User-Item Relevance Models

- Introduce the following random variables

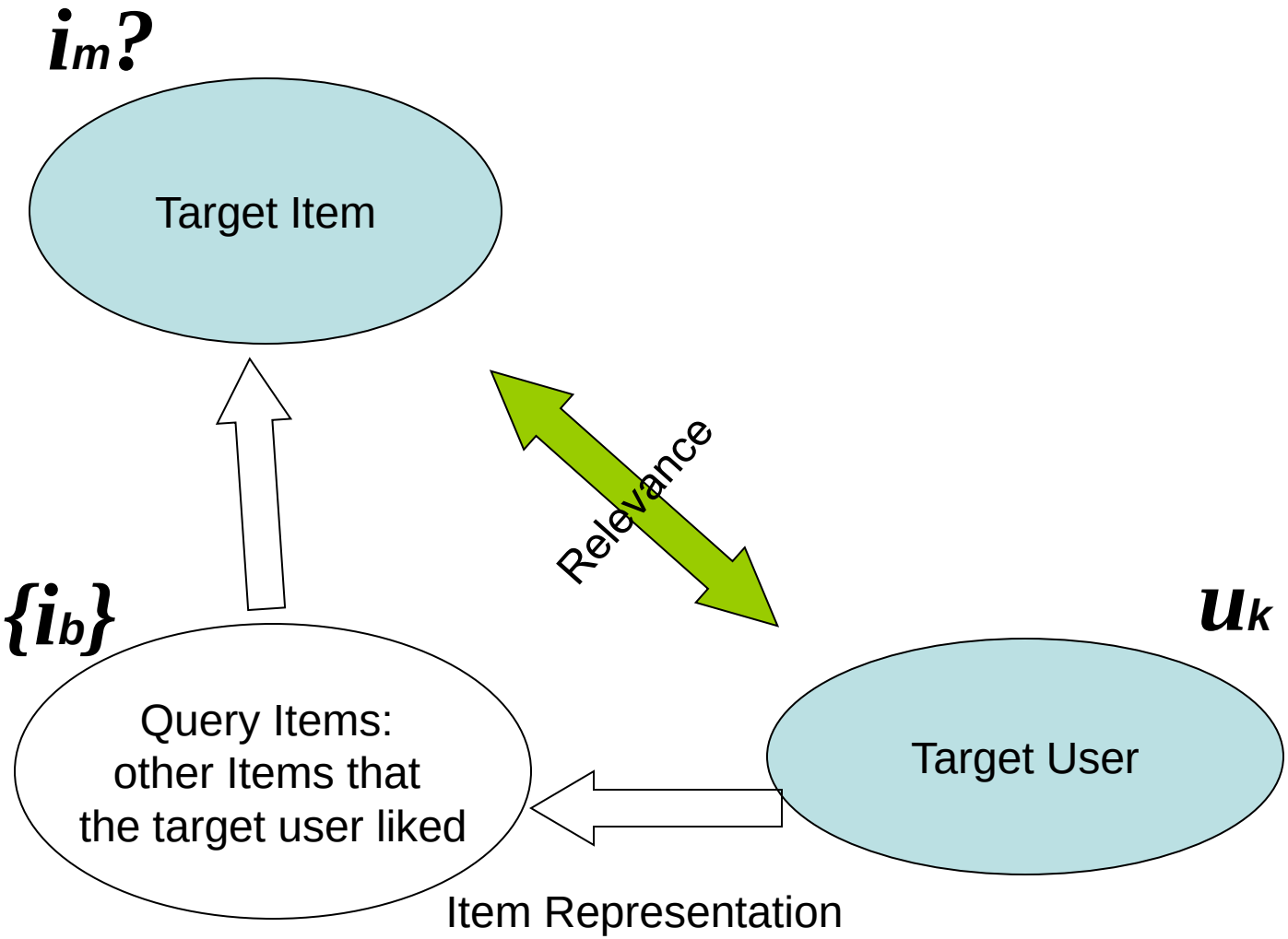
$$\text{Users: } U \in \{u_1, ..., u_K\} \text{ Items: } \mathrm{I} \in \{i_1, ..., i_M\}$$

$$\text{Relevance}: R \in \{r, \bar{r}\}\ ,\ r \text{ "relevant", } \bar{r} \text{ "not relevant"}$$

- Rank items by their log odds of relevance

$$\mathrm{RSV}_U(I) = \log \frac{P(R = r \mid U, I)}{P(R = \bar{r} \mid U, I)}$$

Item Representation



User-Item Relevance Models

- Item representation
  - Use *items that I liked* to represent target user
  - Assume the item "ratings" are independent
  - Linear interpolation smoothing to address sparsity

$$P(i_b \mid i_m, r) = (1-\lambda) P_{ml}(i_b \mid i_m, r) + \lambda P_{ml}(i_b \mid r)$$

$$\text{RSV}_{u_k}(i_m) = \log \frac{P(r \mid i_m, u_k)}{P(\bar{r} \mid i_m, u_k)} = \log \frac{P(u_k \mid r, i_m) P(r \mid i_m)}{P(u_k \mid \bar{r}, i_m) P(\bar{r} \mid i_m)}$$

...

$$= \sum_{\forall i_b : i_b \in L_{u_k} \cap c(i_b, i_m) > 0} \log\left(1 + \frac{(1-\lambda) P_{ml}(i_b \mid i_m, r)}{\lambda P(i_b \mid r)}\right) + \log P(i_m \mid r)$$

$\lambda \in [0,1]$ is a parameter to adjust the strength of smoothing

User-Item Relevance Models

- Probabilistic justification of Item-based CF
  - The RSV of a target item is the combination of its **popularity** and its **co-occurrence** with items (*query items*) that the target user liked.

Co-occurrence

popularity

$$\mathrm{RSV}_{u_k}(i_m) = \sum_{\forall i_b : i_b \in L_{u_k} \cap c(i_b, i_m) > 0} \log\left(1 + \frac{(1-\lambda) P_{ml}(i_b \mid i_m, r)}{\lambda P(i_b \mid r)}\right) + \log P(i_m \mid r)$$

User-Item Relevance Models

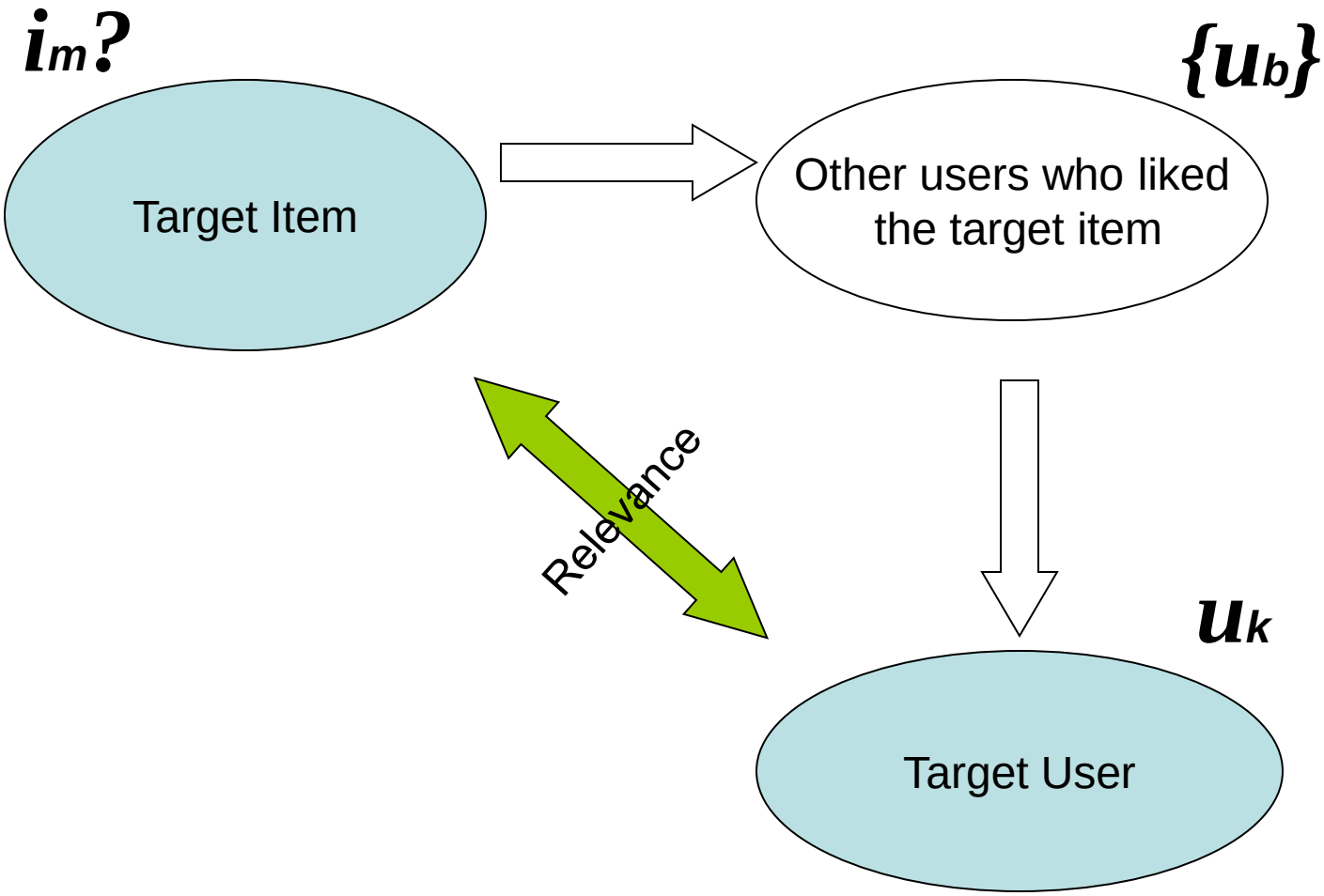
- Probabilistic justification of Item-based CF
  - The RSV of a target item is the combination of its **popularity** and its **co-occurrence** with items (*query items*) that the target user liked
    - Item co-occurrence should be **emphasized** if more users express interest in both target & query item
    - Item co-occurrence should be **suppressed** when the popularity of the query item is high

$$\mathrm{RSV}_{u_k}(i_m) = \sum_{\forall i_b : i_b \in L_{u_k} \cap c(i_b, i_m) > 0} \log\left(1 + \frac{(1-\lambda) P_{ml}(i_b \mid i_m, r)}{\lambda P(i_b \mid r)}\right) + \log P(i_m \mid r)$$

Co-occurrence between target item and query item

Popularity of query item

User Representation



User-Item Relevance Models

- User representation
  - Represent target item by users who like it
  - Assume the user profiles are independent
  - Linear interpolation smoothing to address sparsity

$$P_{ml}(u_b \mid u_k, r) = (1-\lambda)P_{ml}(u_b \mid u_k, r) + \lambda P_{ml}(u_b \mid r)$$

$$\mathrm{RSV}_{u_k}(i_m) = \log \frac{P(r \mid i_m, u_k)}{P(\overline{r} \mid i_m, u_k)} = \log \frac{P(i_m \mid r, u_k)P(r \mid u_k)}{P(i_m \mid \overline{r}, u_k)P(\overline{r} \mid u_k)}$$

...

$$= \sum_{\forall u_b : u_b \in L_{i_m}} \log\left(1 + \frac{(1-\lambda)P_{ml}(u_b \mid u_k, r)}{\lambda P(u_b \mid r)}\right)$$

$\lambda \in [0,1]$ is a parameter to adjust the strength of smoothing

User-Item Relevance Models

- Probabilistic justification of User-based CF
  - The RSV of a target item towards a target user is calculated by the target user's **co-occurrence** with other users who liked the target item
    - User co-occurrence is **emphasized** if more items liked by target user are also liked by the other user
    - User co-occurrence should be **suppressed** when this user liked many items

Co-occurrence between the target user and the other users

$$\mathrm{RSV}_{u_k}(i_m) = \sum_{\forall u_b : u_b \in L_{i_m}} \log\left(1 + \frac{(1-\lambda) P_{ml}(u_b \mid u_k, r)}{\lambda P(u_b \mid r)}\right)$$

Popularity of the other users


CWI

CWI

Empirical Results

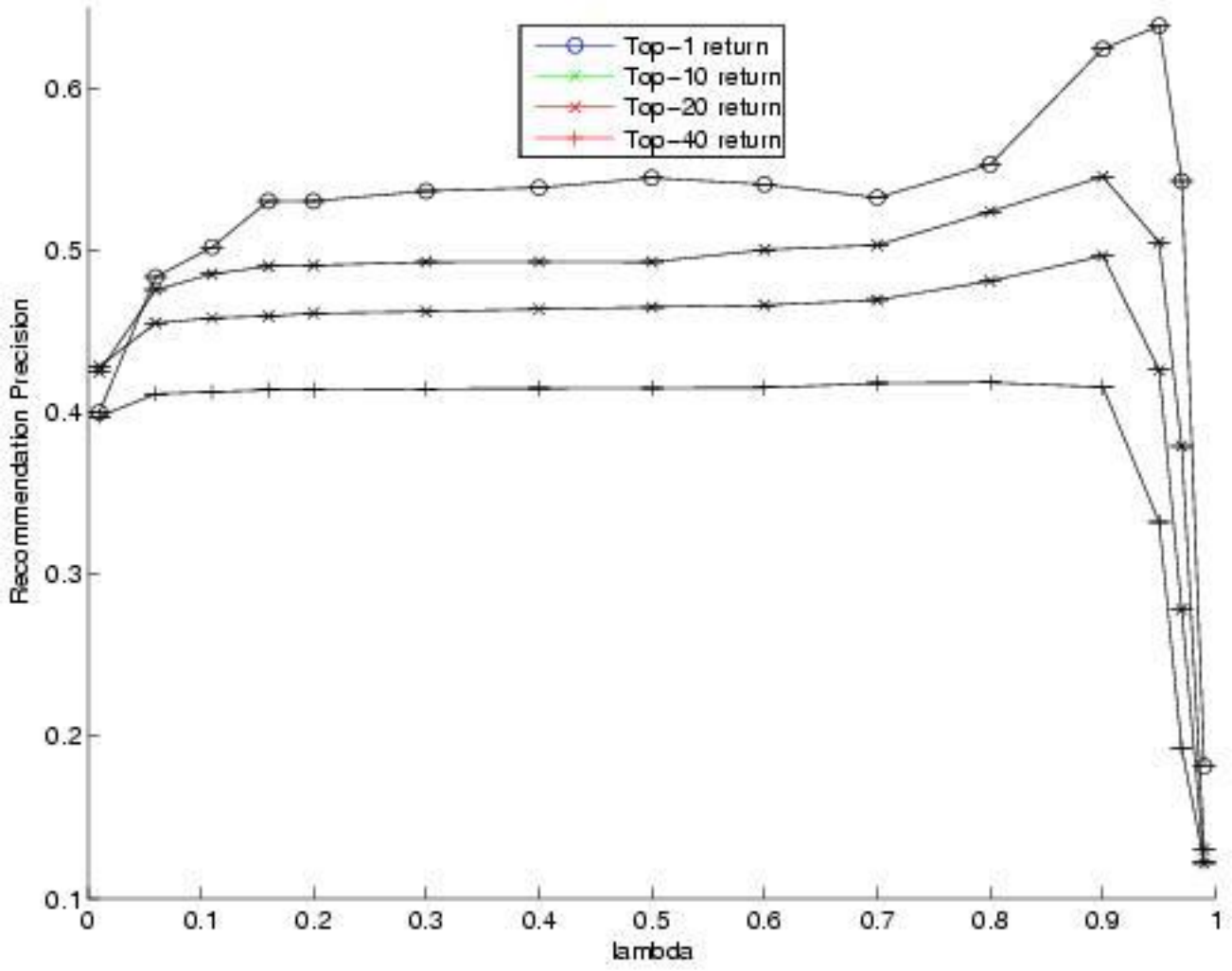
- Data Set:
  - Music play-lists from audioscrobbler.com
  - 428 users and 516 items
  - 80% users as training set and 20% users as test set.
  - Half of items in test set as ground truth, others as user profiles

- Measurement
  - Recommendation Precision:

    (num of corrected items)/(num. of recommended)
  - Averaged over 5 runs
  - Compared with the suggestion lib developed in GroupLens

CWI

P@N vs. lambda



CWI

Effectiveness (P@N)

	Top-1 Item	Top-10 Item	Top-20 Item	Top-40 Item
UIR-Item	0.62	0.52	0.44	0.35
Item-TFIDF	0.55	0.47	0.40	0.31
Item-CosSim	0.56	0.46	0.38	0.31
Item-CorSim	0.50	0.38	0.33	0.27
User-CosSim	0.55	0.42	0.34	0.27

(a) Precision

	Top-1 Item	Top-10 Item	Top-20 Item	Top-40 Item
UIR-Item	0.02	0.15	0.25	0.40
Item-TFIDF	0.02	0.15	0.26	0.41
Item-CosSim	0.02	0.13	0.22	0.35
Item-CorSim	0.01	0.11	0.19	0.31
User-CosSim	0.02	0.15	0.25	0.39

(b) Recall

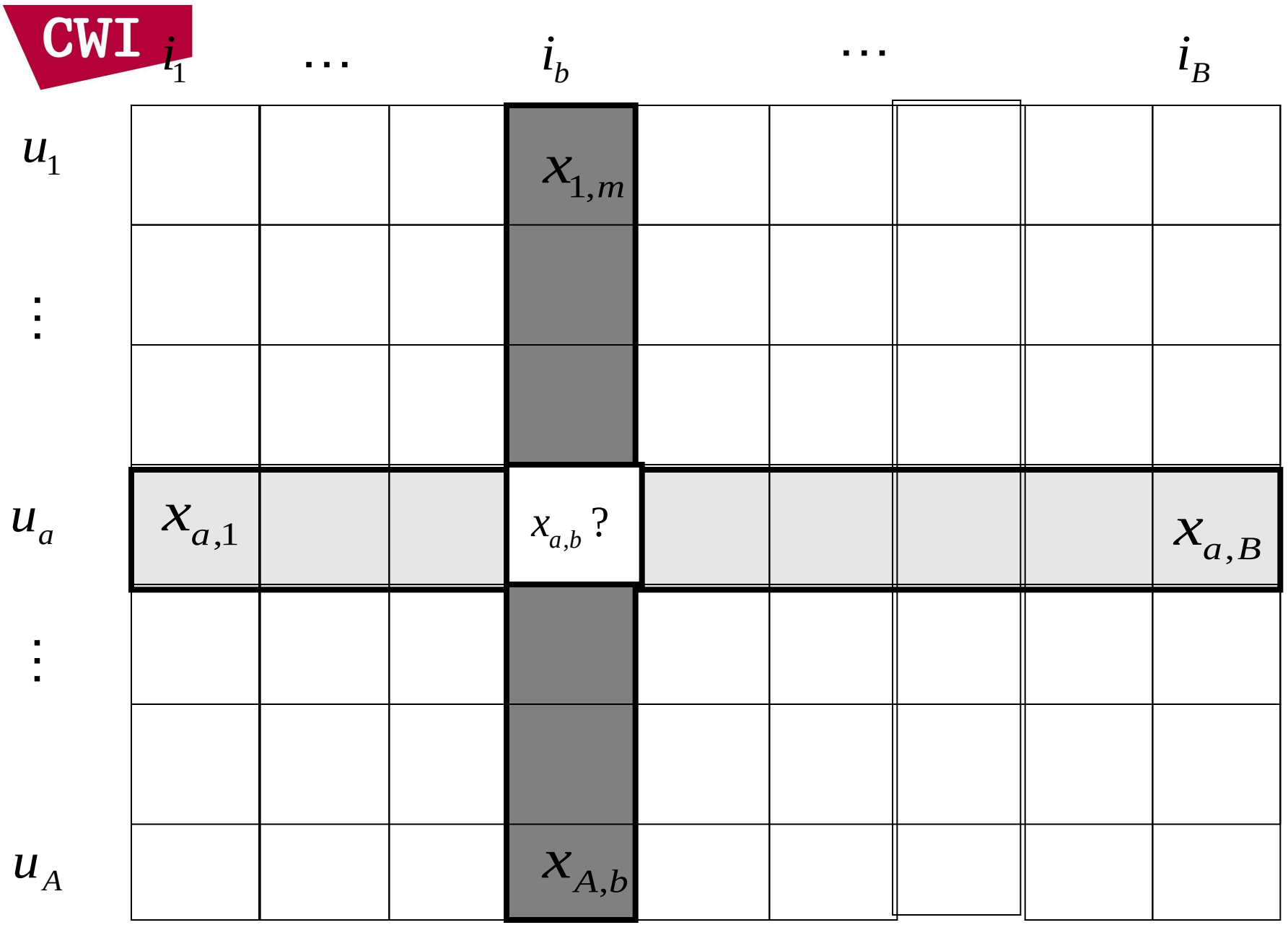
CWI

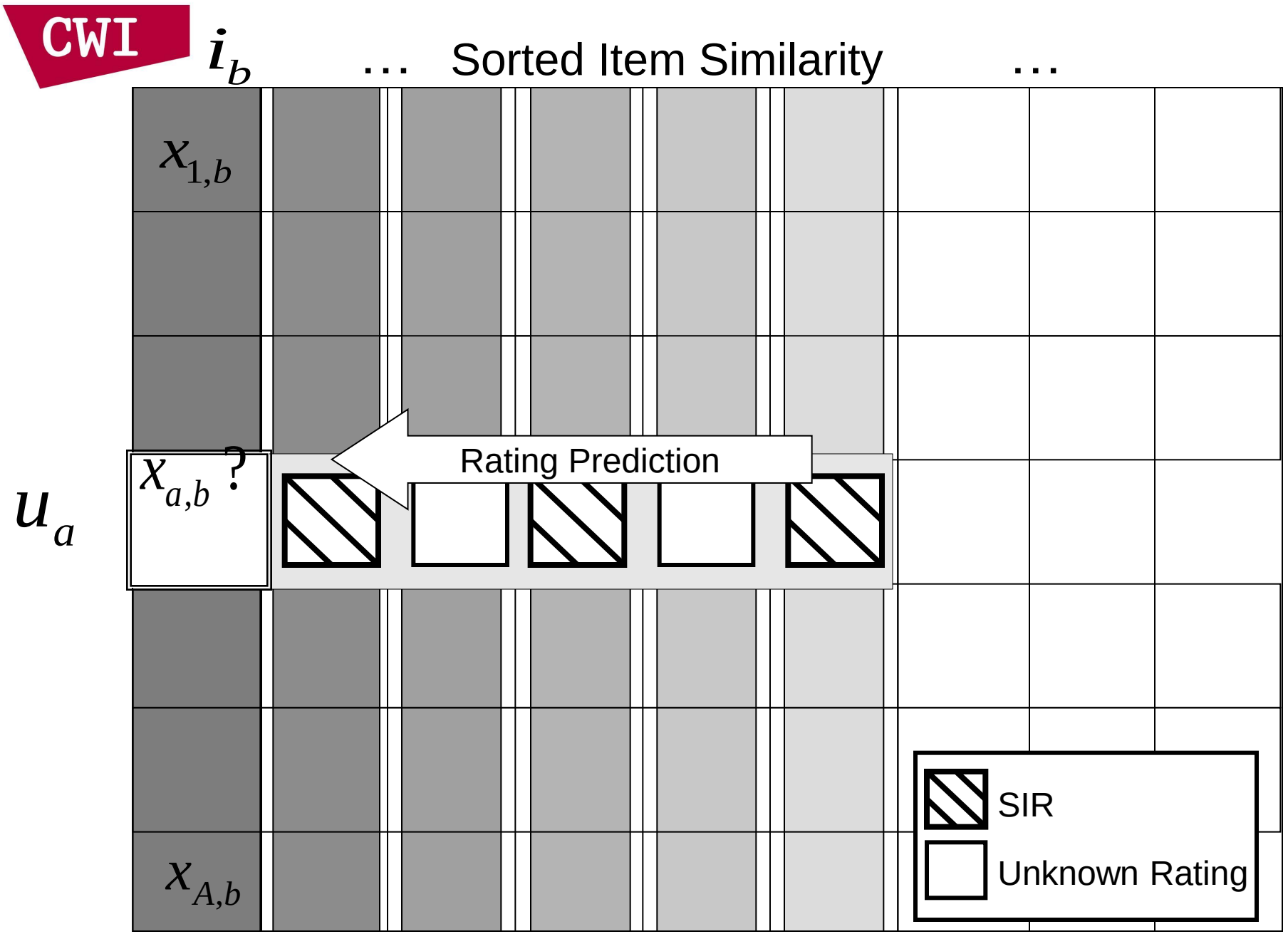
So far…

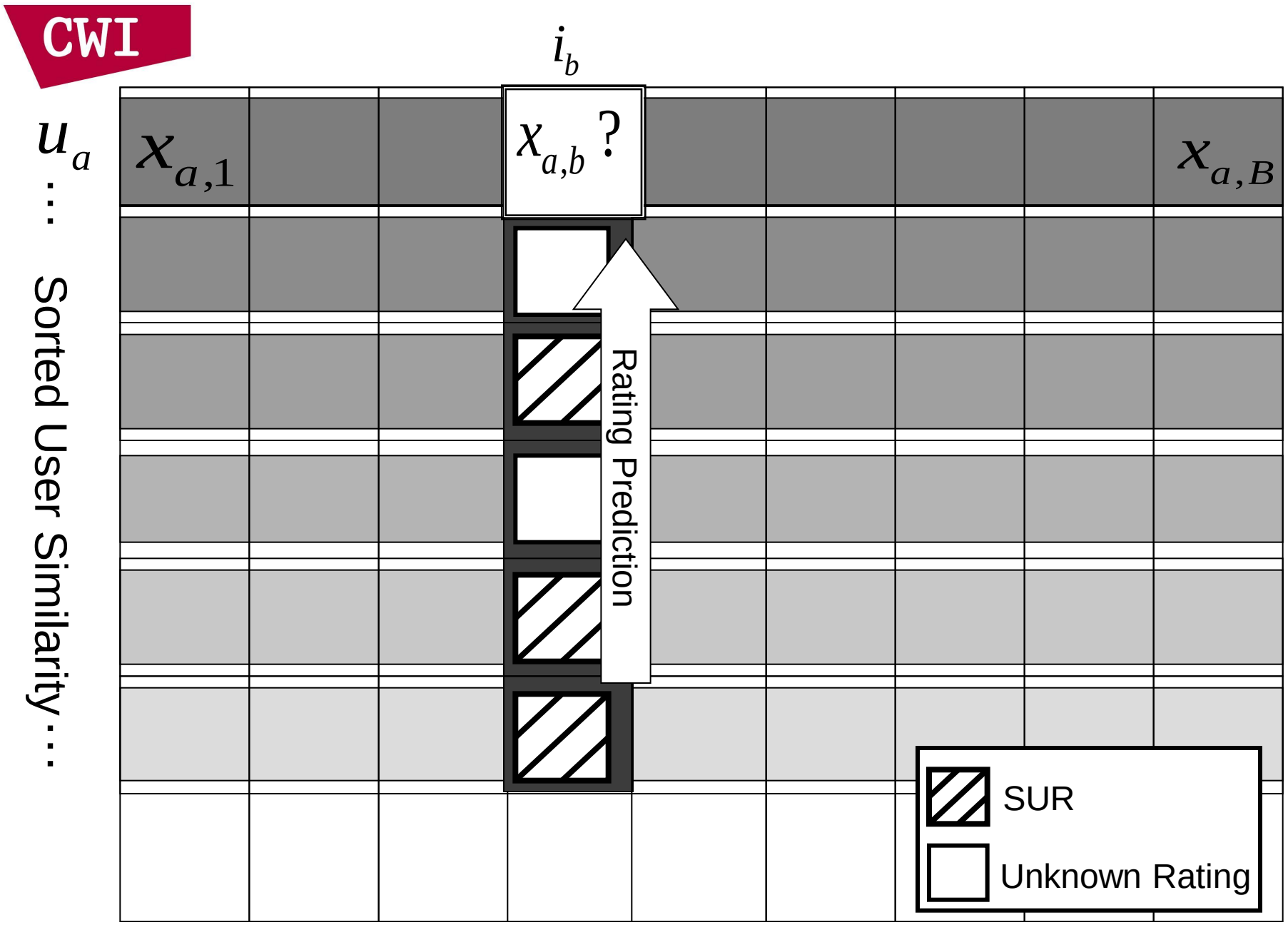
- User-Item relevance models
  - Give a probabilistic justification for CF
  - Deal with the problem of sparsity
  - Provide state-of-art performance

Rating Prediction?

- Previous log-based CF method predicts nor uses rating information
  - Ranks items solely by usage frequency
  - Appropriate for, e.g., music recommendation in a service like Spotify or personalised TV


CWI
i_1
…
i_b
…
i_B
u_1
$x_{1,m}$
u_a
$x_{a,1}$
$x_{a,b}$?
$x_{a,B}$
u_A
$x_{A,b}$


CWI
i_b
… Sorted Item Similarity …
$x_{1,b}$
u_a
$x_{a,b}$?
Rating Prediction
$x_{A,b}$
SIR
Unknown Rating


CWI
i_b
u_a
$x_{a,1}$
$x_{a,b}$?
$x_{a,B}$
Sorted User Similarity…
Rating Prediction
SUR
Unknown Rating

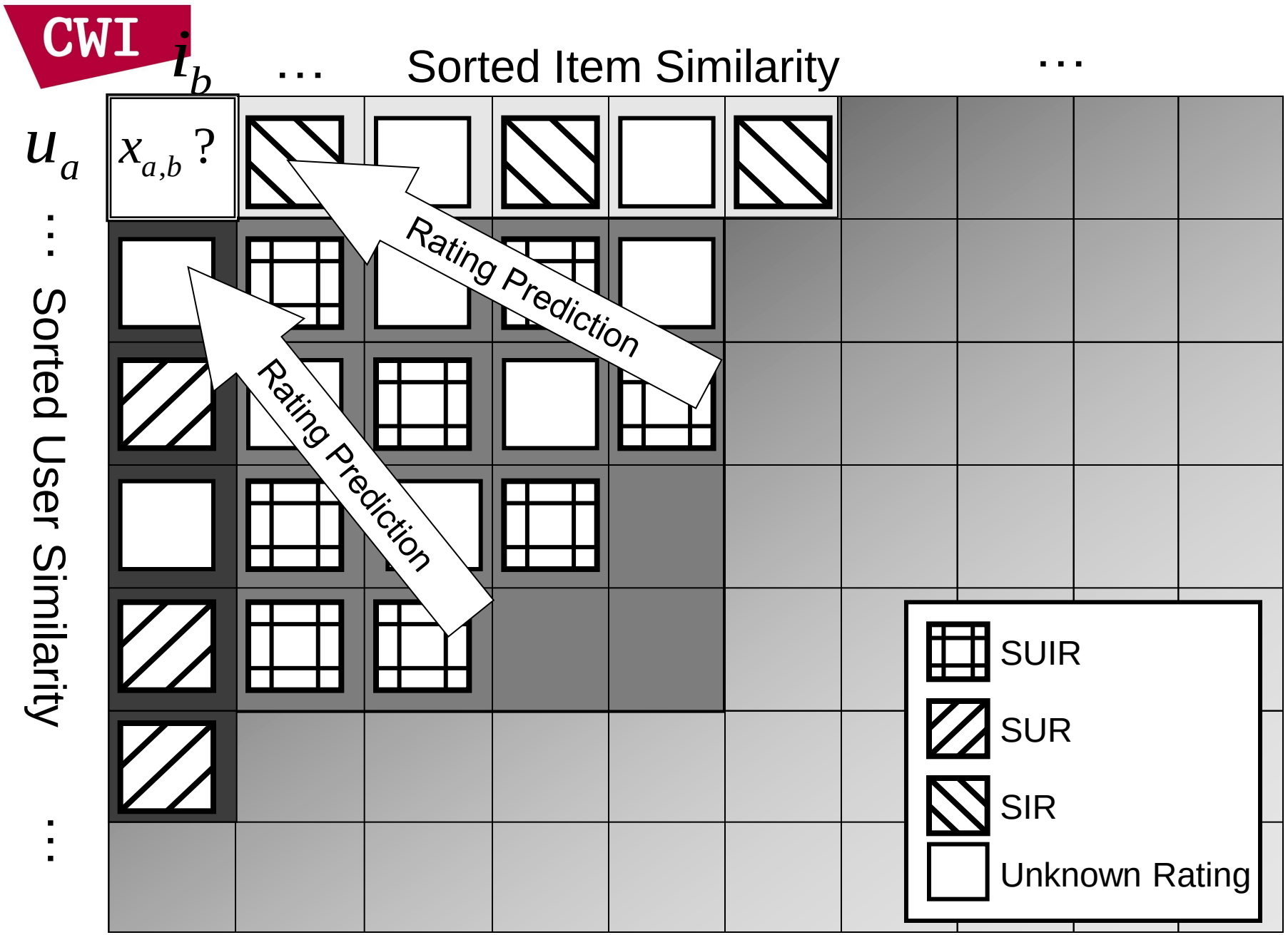
CWI

Sparseness

- Whether you choose SIR or SUR, in many cases, the neighborhood extends to include “not-so-similar” users and/or items

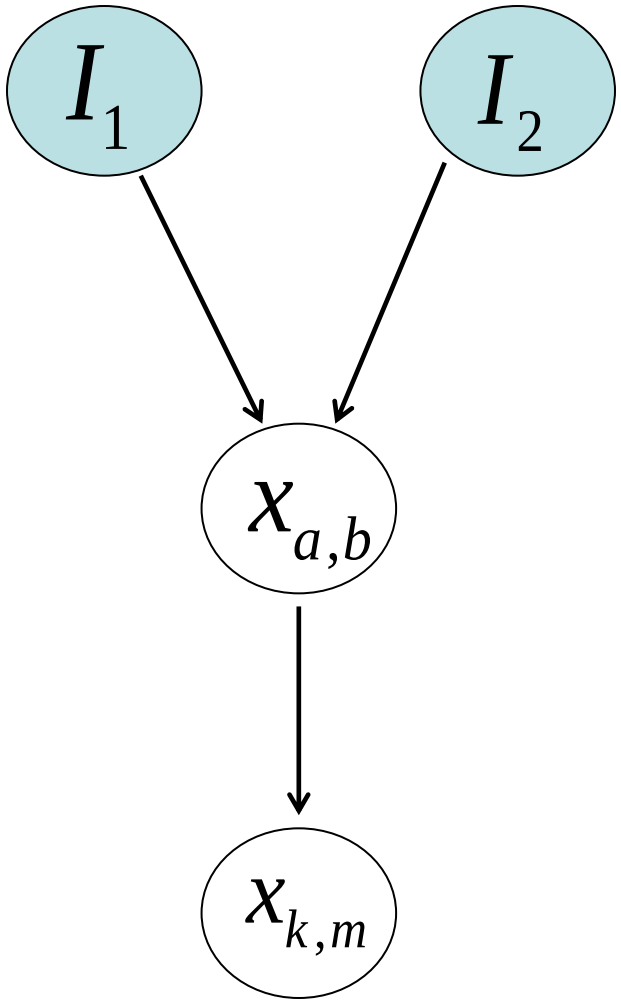
- Idea:
  Take into considerations the similar item ratings made by similar users as extra source for prediction

Jun Wang, Arjen P. de Vries, Marcel JT Reinders, Unifying user-based and item-based collaborative filtering approaches by similarity fusion, SIGIR 2006


CWI
i_b
…
Sorted Item Similarity
…
u_a
$x_{a,b}$?
Rating Prediction
Rating Prediction
… Sorted User Similarity …
SUIR
SUR
SIR
Unknown Rating

CWI

Similarity Fusion



	$I_1 = 0$	$I_1 = 1$
$I_2 = 0$	$x_{a,b} \in SIR$	$x_{a,b} \in SUR$
$I_2 = 1$	$x_{a,b} \in SUIR$	$x_{a,b} \in SUIR$

Sketch of Derivation

$$P(x_{k,m} \mid SUR, SIR, SUIR)$$

$$= \sum_{I_2} P(x_{k,m}, I_2 \mid SUR, SIR, SUIR) P(I_2)$$

$$= P(x_{k,m}, I_2 = 1 \mid SUR, SIR, SUIR) P(I_2 = 1) +$$

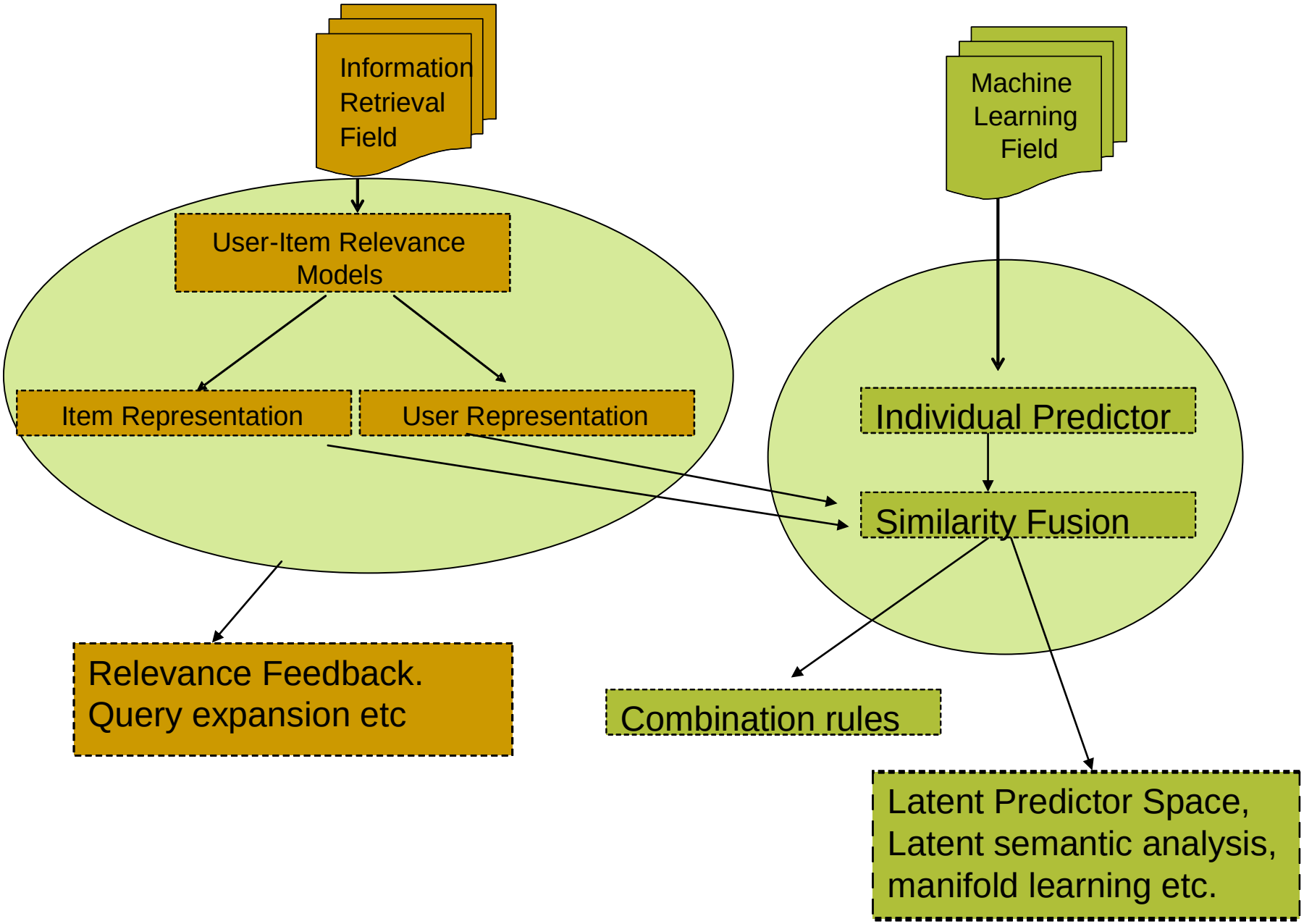
$$P(x_{k,m}, I_2 = 0 \mid SUR, SIR, SUIR)(1 - P(I_2 = 1))$$

$$= P(x_{k,m} \mid SUIR)\delta + P(x_{k,m} \mid SUR, SIR)(1 - \delta)$$

$$= P(x_{k,m} \mid SUIR)\delta + (P(x_{k,m} SUR)\lambda + P(x_{k,m} SIR)(1 - \lambda))(1 - \delta)$$

See SIGIR 2006 paper for more details

Theoretical Level



Remarks

- SIGIR 2006 paper estimates probabilities directly from the similarity distance given between users and items
- TOIS 2008 paper below applies Parzen window kernel density estimation to the rating data itself, to give a full probabilistic derivation
  - Shows how the “kernel trick” let’s us generalize the distance measure; such that a **cosine (projection) kernel** (length-normalized dot product) can be chosen, while keeping Gaussian kernel Parzen windows

Jun Wang, Arjen P. de Vries, and Marcel J. T. Reinders. Unified relevance models for rating prediction in collaborative filtering. *ACM TOIS* 26 (3), June 2008

Relevance Feedback

- Relevance Models for query expansion in IR
  - Language model estimated from known relevant or from top-k documents (Pseudo-RFB)
  - Expand query with terms generated by the LM
- Application to recommendation
  - User profile used to identify neighbourhood; a Relevance Model estimated from that neighbourhood used to expand the profile
  - Deploy probabilistic clustering method PPC to construct the neighbourhood
  - Very good empirical results on P@N

Javier Parapar, Alejandro Bellogín, Pablo Castells, Álvaro Barreiro. *Relevance-Based Language Modelling for Recommender Systems.* Information Processing & Management 49 (4), pp. 966-980

CWI

CF =~ IR?

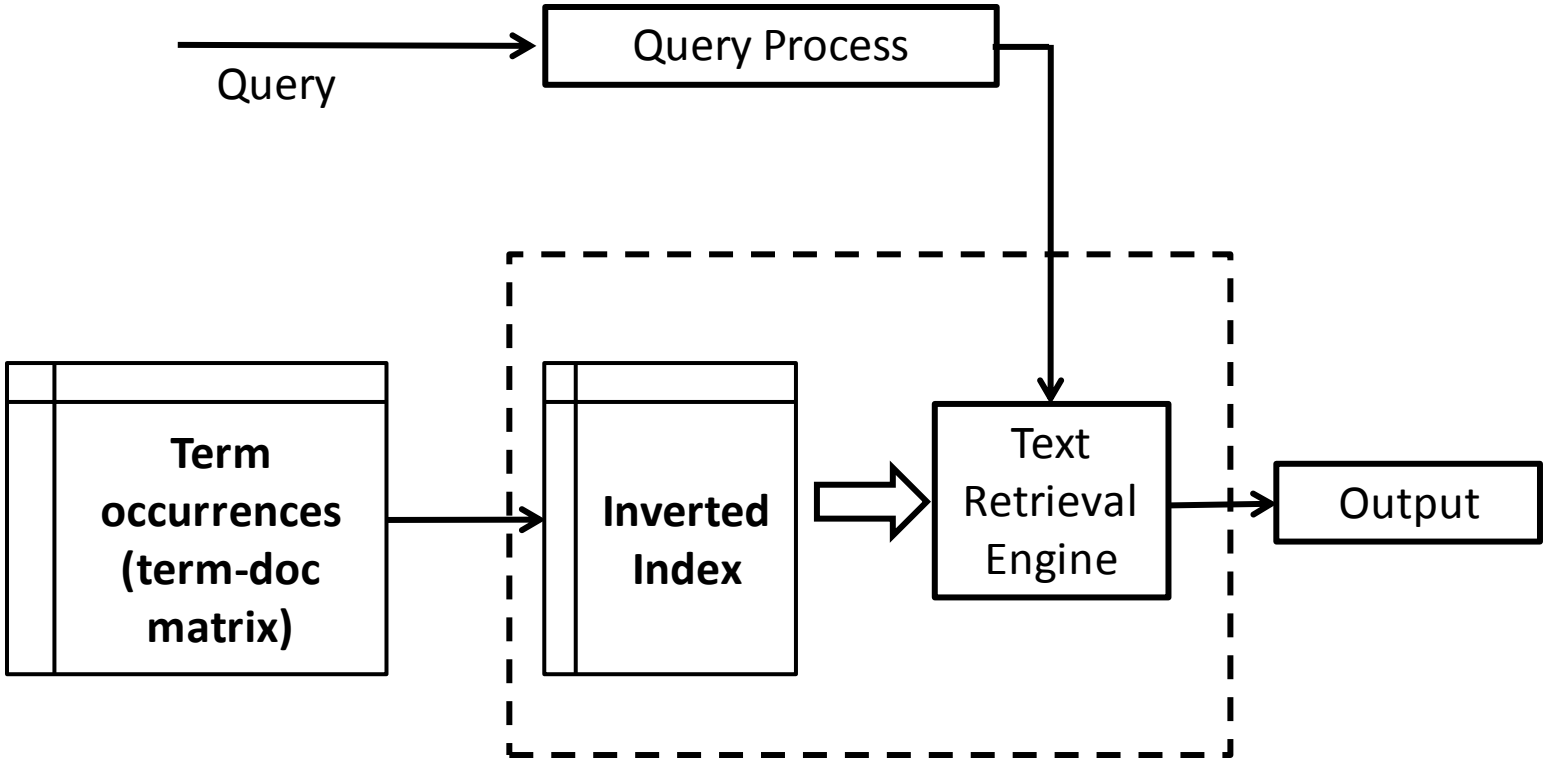
Follow-up question:

Can we go beyond “model level” equivalences observed so far, and actually cast the CF problem such that we can use the full IR machinery?

Alejandro Bellogín, Jun Wang, and Pablo Castells.*Text Retrieval Methods for Item Ranking in Collaborative Filtering.* ECIR 2011

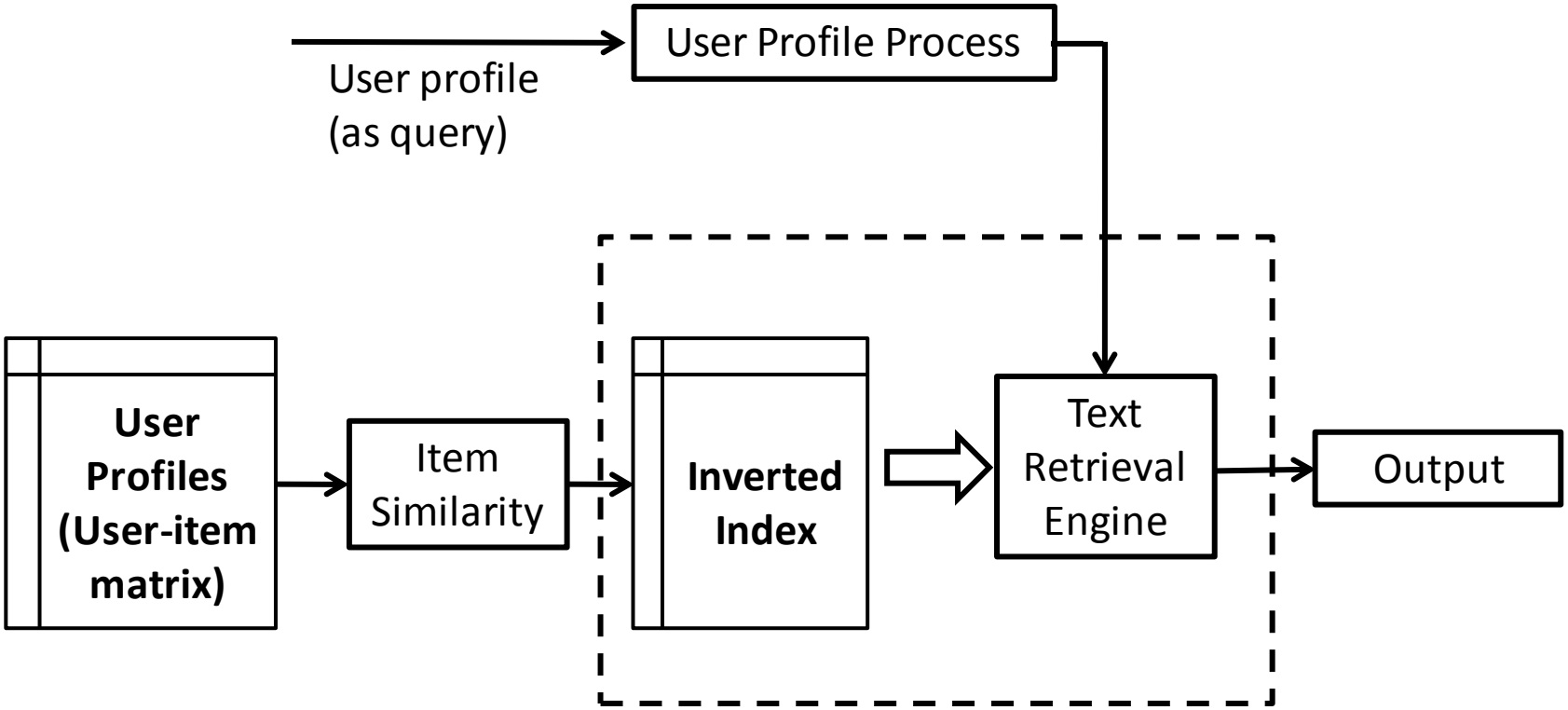
CWI

IR System



CWI

CF RecSys?!



Collaborative Filtering

- Standard item-based formulation

$$\text{rat}(u,i) = \sum_{j \in I_u} \frac{\text{sim}(i,j)}{\sum_{j \in I_u} \text{sim}(i,j)} \text{rat}(u,j)$$

- More general

$$\text{rat}(u,i) = \sum_{j \in g(u)} f(u,i,j) = \sum_{j \in g(u)} f_1(u,j)\, f_2(i,j)$$

Table 2. User and item components for function f in user- and item-based CF. E represents the space where e belongs, that is, $e \in E$.

Approach	f_e^u	f_e^i	E	w_e^u	w_e^i
User-based	$\frac{\text{sim}(u,e)}{\sum_{e \in N[u]} \lvert \text{sim}(u,e) \rvert}$	r_i^e	users	$\text{sim}(u,e)$	r_i^e
Item-based	r_e^u	$\frac{\text{sim}(i,e)}{\sum_{e \in I_u} \lvert \text{sim}(i,e) \rvert}$	items	r_e^u	$\text{sim}(i,e)$

CWI

Text Retrieval

- In (Metzler & Zaragoza, 2009)

$$s(q,d) = \sum_{t \in g(q)} s(q,d,t)$$

  - In particular: *factored form*

$$s(q,d,t) = w_1(q,t)\, w_2(d,t)$$

CWI

Text Retrieval

- Examples
  - TF:

$$w_1(q,t) = \mathrm{qf}(t)$$
$$w_2(d,t) = \mathrm{tf}(t,d)$$

  - TF-IDF:

$$w_1(q,t) = \mathrm{qf}(t)$$
$$w_2(d,t) = \mathrm{tf}(t,d)\log\left(\frac{N}{\mathrm{df}(t)}\right)$$

  - BM25:

$$w(q,t)_1 = \frac{(k_3+1)\mathrm{qf}(t)}{k_3+\mathrm{qf}(t)}$$
$$w(d,t)_2 = \log\left(\frac{N-\mathrm{df}(t)+0.5}{\mathrm{df}(t)+0.5}\right)\frac{(k_1+1)\mathrm{tf}(t,d)}{k_1\left((1-b)+b\cdot\mathrm{dl}(d)/\overline{\mathrm{dl}}\right)+\mathrm{tf}(t,d)}$$

IR =~ CF?

- In item-based Collaborative Filtering

$$\mathrm{tf}(t,d) = \mathrm{sim}(i,j)$$

$$\mathrm{qf}(t) = \mathrm{rat}(u,j)$$

$$t \approx j$$
$$d \approx i$$
$$q \approx u$$

- Apply different models
  - With different normalizations and norms: s_{qd}, L_1 and L_2

s_{qd}		Document: No norm	Document: Norm (/|D|)
Query	No norm	s_{00}	s_{01}
	Norm (/|Q|)	s_{10}	s_{11}

CWI

IR =~ CF!

- TF L1 s01 is **equivalent to item-based CF**

$$s(q,d) = \sum_{t \in g(q)} w_1(q,t)\, w_2(d,t) = \sum_{t \in g(q)} \mathrm{qf}(t) \frac{\mathrm{tf}(t,d)}{\sum_{t \in g(q)} \mathrm{tf}(t,d)}$$

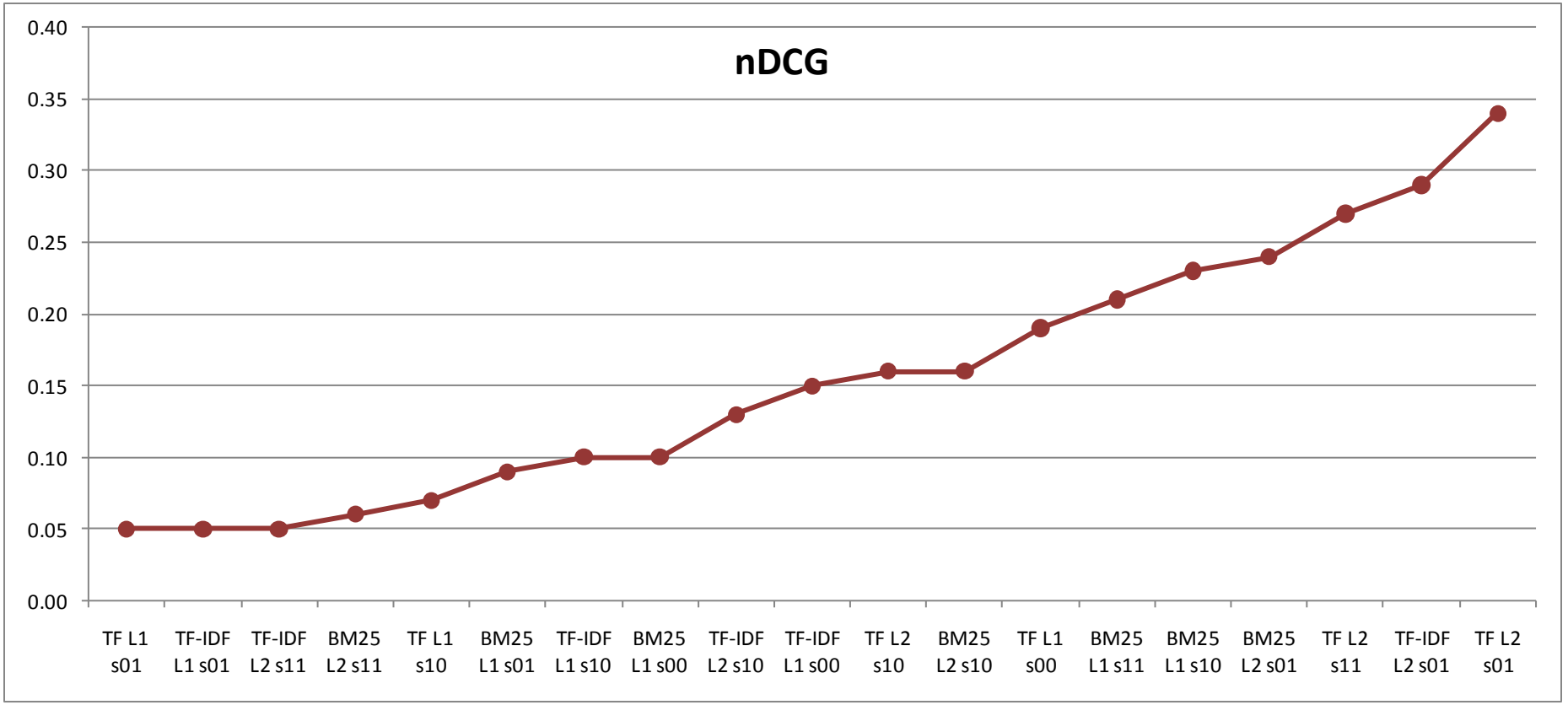
$$\mathrm{rat}(u,i) = \sum_{j \in I_u} \mathrm{rat}(u,j) \frac{\mathrm{sim}(i,j)}{\sum_{j \in I_u} \mathrm{sim}(i,j)}$$

$$\mathrm{tf}(t,d) = \mathrm{sim}(i,j)$$

$$\mathrm{qf}(t) = \mathrm{rat}(u,j)$$

Empirical Results

- Movielens 1M
  - Movielens100k: comparable results



CWI

Vector Space Model

- Challenge:
  - No shared “words” to relate documents to queries
- Solution:
  - First project users and items in a common space
- Two **extreme settings**:
  - Project users into a space with dimensionality of the number of items
  - Project items into a space with dimensionality of the number of users

A. Bellogín, J. Wang, P. Castells. *Bridging Memory-Based Collaborative Filtering and Text Retrieval.* Information Retrieval Journal

CWI

Item Space

- User $\mathbf{u}_I = (r_{u1}, \cdots, r_{uk}, \cdots, r_{un})$
- Item $\mathbf{i}_I = (s_{i1}, \cdots, s_{ik}, \cdots, s_{in})$
- Rank $\text{Score}(\mathbf{u}_I, \mathbf{i}_I) = \sum_k r_{uk} \cdot s_{ik}$
- Predict rating:

$$\hat{r}(u,i) = \frac{\sum_{k=1}^{n} r_{uk} \cdot s_{ik}}{\sum_{\forall k: r_{uk} \neq 0} s_{ik}} = \frac{\text{Score}(\mathbf{u}_I, \mathbf{i}_I)}{\sum_{\forall k: r_{uk} \neq 0} s_{ik}} = \frac{\text{Score}(\mathbf{u}_I, \mathbf{i}_I)}{\text{Score}\left(\delta\left(\mathbf{u}_I\right), \mathbf{i}_I\right)}$$

CWI

User space

- User $\mathbf{u}_U = (s_{u1}, \cdots, s_{uk}, \cdots, s_{um})$
- Item $\mathbf{i}_U = (r_{1i}, \cdots, r_{ki}, \cdots, r_{mi})$
- Rank $\text{Score}(\mathbf{u}_I, \mathbf{i}_I) = \sum_k r_{uk} \cdot s_{ik}$
- Predict rating:

$$\hat{r}(u,i) = \frac{\sum_{k=1}^{m} r_{ki} \cdot s_{uk}}{\sum_{\forall k: r_{ki} \neq 0} s_{uk}} = \frac{\text{Score}(\mathbf{u}_U, \mathbf{i}_U)}{\sum_{\forall k: r_{ki} \neq 0} s_{uk}} = \frac{\text{Score}(\mathbf{u}_U, \mathbf{i}_U)}{\text{Score}\left(\mathbf{u}_U, \delta\left(\mathbf{i}_U\right)\right)}$$

CWI

Linear Algebra

- Users and items in shared orthonormal space:

$$\mathbf{u^v} = \lambda_1^v e_1 + \cdots + \lambda_l^v e_l = (\lambda_1^v, \cdots, \lambda_l^v)$$
$$\mathbf{i^j} = \mu_1^j e_1 + \cdots + \mu_l^j e_l = (\mu_1^j, \cdots, \mu_l^j)$$

- Consider covariance matrix

$$C_I = cov(X) \qquad a_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$$
$$\mu_i = E(X_i)$$

- Spectral theorem now states that an orthonormal basis of eigenvectors exists

orthonormal basis of eigenvectors of dimension n of C_I.

CWI

Linear Algebra

- Use this basis to represent items and users:

$$\mathbf{i^j} = \mu_1^j e_1 + \cdots + \mu_n^j e_n = (\mu_1^j, \cdots, \mu_n^j)$$

$$\begin{aligned}\mathbf{u^v} &= r_{v1}\mathbf{i^1} + \cdots + r_{vn}\mathbf{i^n} = r_{v1}\sum_j \mu_j^1 e_j + \cdots + r_{vn}\sum_j \mu_j^n e_j \\ &= (r_{v1}\mu_1^1 + \cdots + r_{vn}\mu_1^n, \cdots, r_{v1}\mu_n^1 + \cdots + r_{vn}\mu_n^n) \\ &= (r_{v1} + \cdots + r_{vn}) \cdot C_I\end{aligned}$$

- The dot product then has a remarkable form (of the IR models discussed):

$$\mathbf{u^v} \cdot \mathbf{i^j} = \sum_{p=1}^{n} \mu_p^j (r_{v1}\mu_p^1 + \cdots + r_{vn}\mu_p^n) = \sum_{p=1}^{n} \mu_p^j \sum_{k=1}^{n} r_{vk}\mu_p^k = \sum_{k=1}^{n} r_{vk} \sum_{p=1}^{n} \mu_p^j \mu_p^k$$

CWI

Subspaces…

- Number of items (n) vs. number of users (m):
  - If n < m, a linear dependency must exist between users in terms of the item space components
  - In this case, it has been known empirically that item-based algorithms tend to perform better
    - Dimension of sub-space key for the performance of the algorithm?
    - ~ better estimation (more data per item) in the probabilistic versions

CWI

Subspaces…

- Matrix Factorization methods are captured by assuming a lower-dimensionality space to project items and users into (usually considered “model-based” rather than “memory-based”)

$$\mathbf{u}^{\mathbf{v}} \cdot \mathbf{i}^{\mathbf{j}} = \sum_{p=1}^{n} \mu_p^j (r_{v1}\mu_p^1 + \cdots + r_{vn}\mu_p^l) = \sum_{p=1}^{l} \mu_p^j \sum_{k=1}^{n} r_{vk}\mu_p^k$$

~ Latent Semantic Indexing (a VSM method replicated as pLSA and variants)

Ratings into Inverted File

	Tom	John
Item 1	5	1
Item 2	1	5
Item 3	5	1
Item 4	1	5

User-item matrix

	Item 1	Item 2	Item 3	Item 4
Item 1	1.0	-1.0	1.0	-1.0
Item 2	-1.0	1.0	-1.0	1.0
Item 3	1.0	-1.0	1.0	-1.0
Item 4	-1.0	1.0	-1.0	1.0

Similarity matrix

ID	Index	Candidate. Doc
1	Item 1	Item 3 (similarity 1)
2	Item 2	Item 4 (similarity 1)
3	Item 3	Item 1 (similarity 1)
4	Item 4	Item 2 (similarity 1)

Inverted index (similarity stored as term frequency)

- Note: distribution of item occurrences not Zipfian like text, so existing implementations (including choice of compression etc.) may be sub-optimal for CF runtime performance

CWI

Weighting schemes

Table 2 Weighting schemes under the unified framework for item-based CF. The rating from the (query) user u is denoted as r_{uk}, the similarity between the target item and item k is s_{ik}, N is the number of items, N_k is the number of items similar to item k, $\text{il}(i)$ is the number of similar items of the target item, and $\bar{\text{il}}$ is the average il.

Method	w_k^u	w_k^i
Binary	1 if rated	1 if similar
TF	r_{uk}	s_{ik}
TF-IDF	r_{uk}	$s_{ik} \log\left(\frac{N}{N_k}\right)$
BM25	$\frac{(k_3+1)r_{uk}}{k_3+r_{uk}}$	$\log\left(\frac{N-N_k}{N_k}\right) \frac{(k_1+1)s_{ik}}{k_1\left((1-b)+b\cdot\text{il}(i)/\bar{\text{il}}\right)+s_{ik}}$
Language Model (Jelinek-Mercer)	r_{uk}	$(1-\lambda)p(k\|i) + \lambda p(k\|\mathcal{C})$
Language Model (Dirichlet)	r_{uk}	$\frac{s_{ik}}{\text{il}(i)+\mu} + \mu\frac{p(k\|\mathcal{C})}{\text{il}(i)+\mu}$

CWI

Empirical results 1M

Table 8 Performance results in the user space for the item ranking task (*Movielens 1M*).

Methods	P@5	P@10	nDCG@3	nDCG@5	nDCG@10	MAP
User-based CF	0.0274	0.0252	0.0224	0.0232	0.0224	0.0139
TF $\|\|n_{10}\|\|_1$	0.0274	0.0252	0.0224	0.0232	0.0224	0.0139
MF	0.0623	0.0586	0.0592	0.0606	0.0602	0.0307
BM25 $\|\|n_{01}\|\|_2$	0.0983	0.0863	0.0964	0.0914	0.0883	0.0379
TF-IDF $\|\|n_{01}\|\|_2$	0.0984	0.0862	0.0963	0.0913	0.0882	0.0379
Dirichlet n_{00}	0.1013	0.0892	0.1020	0.0953	0.0921	0.0395
BM25 n_{00}	0.1013	0.0892	0.1020	0.0953	0.0921	0.0395
Jelinek-Mercer n_{00}	0.1013	0.0892	0.1020	0.0953	0.0921	0.0395
TF-IDF n_{00}	0.1013	0.0892	0.1020	0.0953	0.0921	0.0395
BM25 $\|\|n_{10}\|\|_2$	0.1038	0.0902	0.1049	0.0982	0.0942	0.0397
TF-IDF $\|\|n_{10}\|\|_2$	0.1041	0.0902	0.1051	0.0987	0.0944	0.0399

CWI

Empirical results 10M

Table 10 Performance results in the user space for the item ranking task (*Movielens 10M*).

Methods	P@5	R@5	nDCG@5	MAP	MRR	bpref
User-based CF	0.0124	0.0018	0.0102	0.0090	0.0425	0.4972
TF $\|\|n_{10}\|\|_1$	0.0124	0.0018	0.0102	0.0090	0.0425	0.4972
MF	0.0456	0.0103	0.0467	0.0162	0.1210	0.3303
BM25 $\|\|n_{01}\|\|_2$	0.0865	0.0272	0.0773	0.0381	0.2177	0.5983
TF-IDF $\|\|n_{01}\|\|_2$	0.0865	0.0272	0.0773	0.0381	0.2177	0.5983
Dirichlet n_{00}	0.0913	0.0279	0.0826	0.0388	0.2251	0.5800
BM25 n_{00}	0.0913	0.0279	0.0826	0.0388	0.2251	0.5800
Jelinek-Mercer n_{00}	0.0913	0.0279	0.0826	0.0388	0.2251	0.5800
TF-IDF n_{00}	0.0913	0.0279	0.0826	0.0388	0.2251	0.5800
TF-IDF $\|\|n_{10}\|\|_2$	0.0927	0.0275	0.0848	0.0382	0.2281	0.5705
BM25 $\|\|n_{10}\|\|_2$	0.0928	0.0277	0.0850	0.0382	0.2285	0.5716

Rating prediction

Table 12 Results for the rating prediction task (*Movielens 1M*).

Item-based			User-based		
Method	MAE	RMSE	Method	MAE	RMSE
Item-based CF	0.8210^{a}	1.0255^{a}	User-based CF	0.9443^{a}	1.2138^{a}
MF	0.6747^{b}	0.8687^{b}	MF	0.6747^{b}	0.8687^{b}
BM25	0.8236^{a}	1.0408^{c}	BM25	0.9443^{a}	1.2138^{a}
TF-IDF	0.8256^{c}	1.0301^{a}	TF-IDF	0.9443^{a}	1.2138^{a}
Dirichlet	0.8284^{d}	1.0359^{d}	Dirichlet	0.9443^{a}	1.2138^{a}
Jelinek-Mercer	0.8290^{d}	1.0358^{d}	Jelinek-Mercer	0.9443^{a}	1.2138^{a}

CWI

Concluding Remarks

- The probabilistic models are elegant (often deploying impressive maths), but what do they *really* add in understanding IR & CF – i.e., beyond the (often claimed to be “ad-hoc”) approaches of the VSM?

Concluding Remarks

- Clearly, the models in CF & IR are closely related
- Should these then really be studied in two different (albeit overlapping) communities, RecSys vs. SIGIR?

CWI

Meanwhile at TREC…

Contextual Suggestions

- Given a user profile and a context, make suggestions
  - *AKA* Context-aware Recommendation, zero-query Information Retrieval, …

CWI

“Entertain me”

- Recommend “things to do”, where
  - User profile consists of opinions about attractions
  - Context consists of a specific geo-location

TREC-CS (1/3)

- Given a user profile
  - 70 – 100 POIs represented by a title, description and URL (situated in Chicago / Santa Fe)
  - Rated on a scale 0 – 4

125, Adler Planetarium & Astronomy Museum, "Interactive exhibits & high-tech sky shows entertain stargazers -- lakefront views are a bonus.", http://www.adlerplanetarium.org/
131,Lincoln Park Zoo,"Lincoln Park Zoo is a free 35-acre zoo located in Lincoln Park in Chicago, Illinois. The zoo was founded in 1868, making it one of the oldest zoos in the U.S. It is also one of a few free admission zoos in the United States.", http://www.lpzoo.org/

700, **125**, 4, 4
700, **131**, 0, 1

CWI

TREC CS (2/3)

- … and a context
  - Corresponding to a metropolitan area in the USA, e.g., 109, Kalamazoo, MI

TREC CS (3/3)

- Suggest Web pages / snippets
  - From the Open Web, or from ClueWeb

700, 109 ,1,"About KIA History Kalamazoo Institute of Arts KIA History","The Kalamazoo Institute of Arts is a nonprofit art museum and school. Since , the institute has offered art classes and free admission programming, including exhibitions, lectures, events, activities and a permanent collection. The KIAs mission is to cultivate the creation and appreciation of the visual arts for the communities",clueweb12-1811wb-14-09165

CWI

Common approach

$$P_{rel}(u, s) = P(s) \cdot (\lambda \cdot SIM(u^{+}, s) - (1 - \lambda) \cdot SIM(u^{-}, s))$$

Candidate Selection Prior

Personalization

CWI

References

- Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE TKDE 17(6), 734-749 (2005)
- Alejandro Bellogín, Jun Wang, and Pablo Castells.*Text Retrieval Methods for Item Ranking in Collaborative Filtering.* ECIR 2011.
- Metzler, D., Zaragoza, H.: Semi-parametric and non-parametric term weighting for information retrieval. ECIR 2009.
- Javier Parapar, Alejandro Bellogín, Pablo Castells, Álvaro Barreiro. *Relevance-Based Language Modelling for Recommender Systems.*Information Processing & Management 49 (4), pp. 966-980
- A. Bellogín, J. Wang, P. Castells. *Bridging Memory-Based Collaborative Filtering and Text Retrieval.*Information Retrieval (to appear)
- Jun Wang, Arjen P. de Vries, Marcel JT Reinders, Unifying user-based and item-based collaborative filtering approaches by similarity fusion, SIGIR 2006
- Jun Wang, Arjen P. de Vries, Marcel JT Reinders, A User-Item Relevance Model for Log-Based Collaborative Filtering, ECIR 2006
- Jun Wang, Arjen P. de Vries, and Marcel J. T. Reinders. Unified relevance models for rating prediction in collaborative filtering. *ACM TOIS* 26 (3), June 2008.
- Jun Wang, Stephen Robertson, Arjen P. de Vries, and Marcel J.T. Reinders. Probabilistic relevance ranking for collaborative filtering. Information Retrieval 11 (6):477-497, 2008

CWI

Thanks

- Alejandro Bellogín
- Jun Wang
- Thijs Westerveld
- Victor Lavrenko