

CYBER – a **C**ommunit**Y**-**B**ased
sEaRch engine

Yingguang Li

Lidan Shou

Kian-Lee Tan

Outline

- Introduction
- CYBER
- Related work
- Conclusion

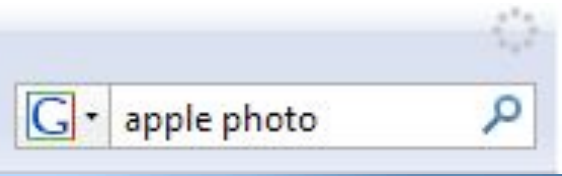
Introduction

- Relevance Feedback
 - Individual feedback
 - Community feedback

Introduction

- Observations
 - Commonality of communities
 - users with common interests tend to query on similar objects
 - Variety of communities
 - A user belongs to several communities
 - A community contains several sub-communities
 - Difference of communities
 - Recommendations are useless to other communities

Motivating Example



Apple fans	D ₁	Apple – iLife - iPhoto
	D ₂	Apple Journal- Photo Gallery (page one)
	D ₃	NY Apple Country Jonamac Apple Photo
photographers	D ₄	red apple photo ...
	D ₅	PC World - First Look: iPod ... Photos
	D ₆	Amazon.com: Apple 30 GB iPod Photo

User₃

User₁

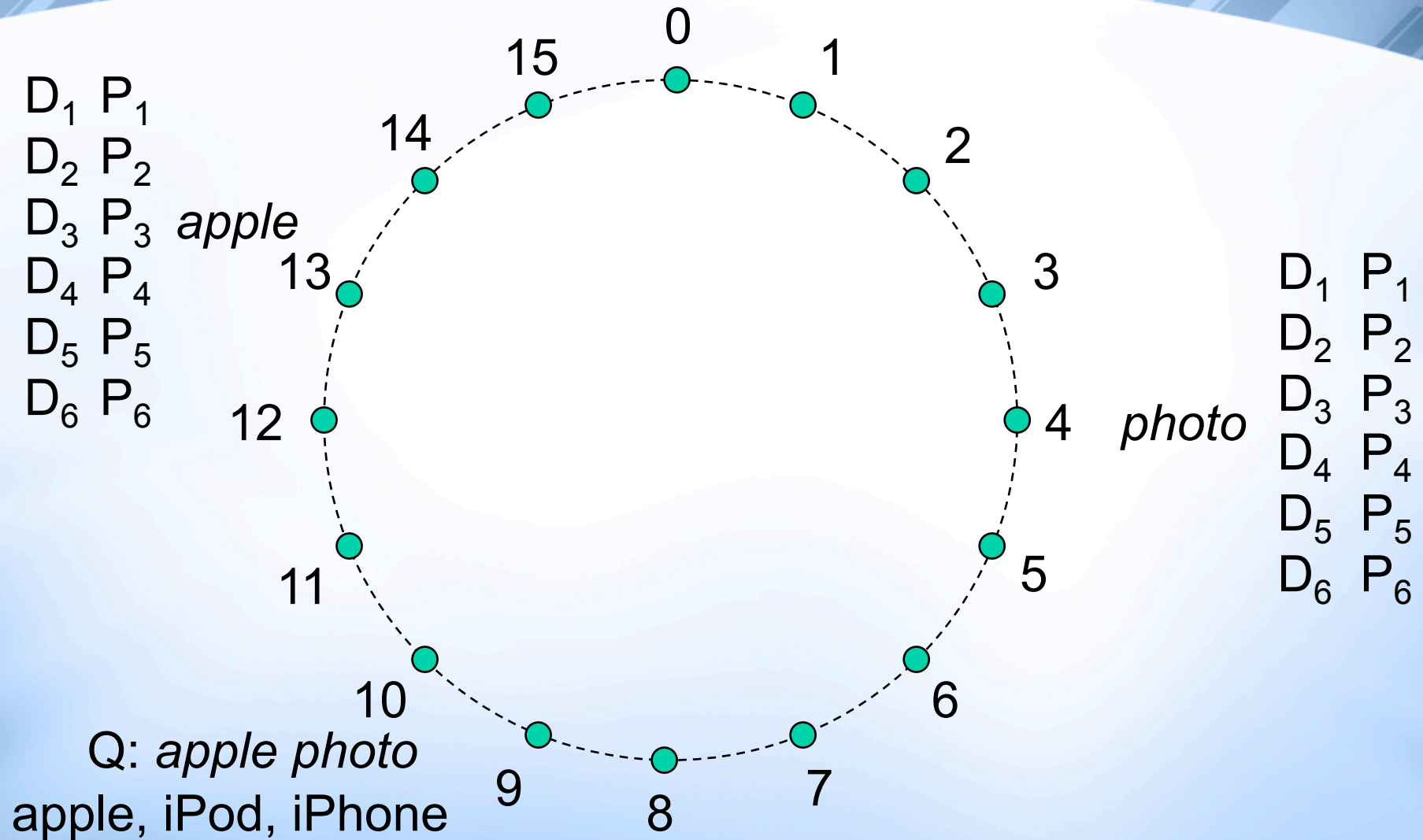
User₂

User₄

D ₁
D ₂
D ₃
D ₄
D ₅
D ₆

D ₁
D ₂
D ₃
D ₄
D ₅
D ₆

CYBER – Overview



CYBER – Profiles

- Document Profile

- *apple*

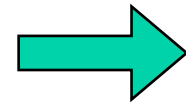
D ₁	apple iLife iPhoto	M ₁
D ₂	apple read color gallery	M ₂
D ₃	photo color red green	M ₃
D ₄	Red green yellow photo	M ₄
D ₅	iPod iPhone iMac	M ₅
D ₆	apple amazon iPod iPhone	M ₆

M_i: IP, Doc ID, TF, Doc length, Summary

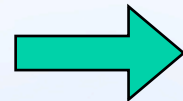
CYBER – Profiles

- User Profile

S₁ **S₂** **S₃**
S₄ **S₅** **S₆**
S₇ **S₈** **S₉**
S₁₀ **S₁₁** **S₁₂**



Apple iPod iPhone



Travel hostel ticket photo

CYBER – Query Processing

$$sim(Q, D_i) = \frac{\sum_{j=1}^n w_{Q,j} \times w_{i,j}}{\sqrt{\text{number_of_terms_in_} D_i}}$$

- Profile Similarity

$$vsim(V_u, V_D) = \frac{|V_u \cap V_D|}{|V_u|}$$

- New Term Weight

$$W_{i,j} = w_{i,j} \times (1 + vsim(V_u, V_D))$$

CYBER – Profile Updating

- Document Profile Updating

- *apple*

✓

D ₁	apple iLife iPhoto iPod
D ₂	apple read color gallery
D ₃	photo color red green
D ₄	Red green yellow photo
D ₅	iPod iPhone iMac apple
D ₆	apple amazon iPod iPhone

Q: apple photo (*apple iPod iPhone*)

CYBER – Profile Updating

- User Profile Updating
 - Cluster shared (S_i) and downloaded (D_j) documents

S_1	S_2	D_3
S_4	S_5	S_6
S_7	D_8	S_9
D_{10}	S_{11}	S_{12}

CYBER – Evaluation

- Dataset

- TREC9

- User click simulation

$$0.999 \times 0.9^{i-1}$$

Ranked list

Random()

Clicks

D₁	→	0.999	0.4	D₁
D₂	→	0.899	0.7	D₂
D₃				
D₄	→	0.728	0.8	
D₅				
D₆				
D₇	→	0.531	0.5	D₇
D₈				

CYBER – Evaluation

- Systems
 - A system with complete index but without feedback (CI)
 - A system with partial index and “single-user” feedback (PIF)
 - CYBER

$$\frac{\text{Feedback - CI}}{\text{CI}}$$

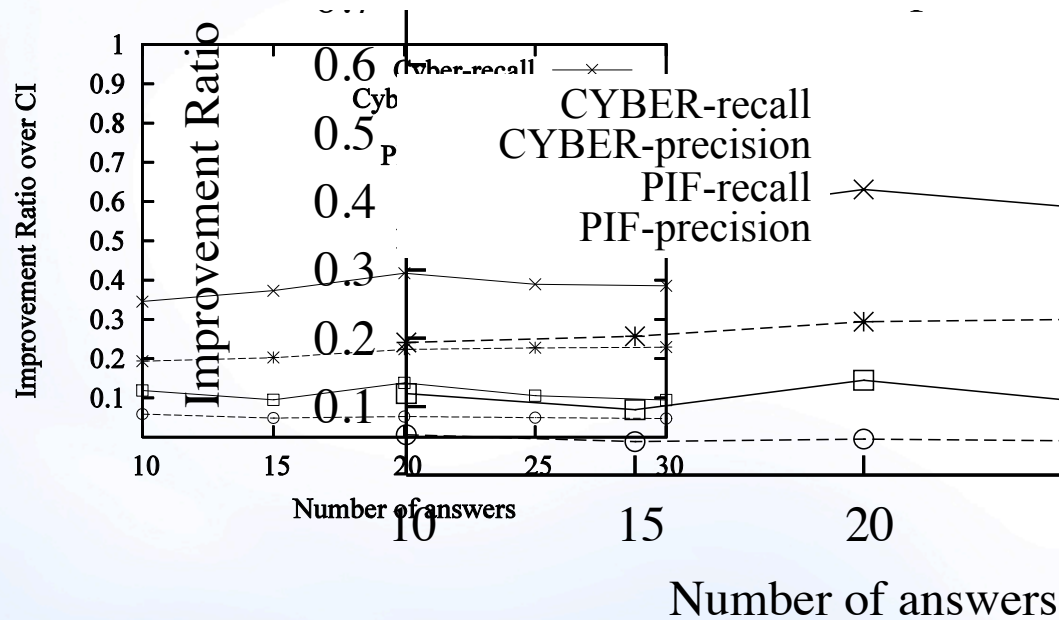
CYBER – Evaluation

- Experiment Settings

Parameters	Default values	Ranges
number of answers	20	[10, 30]
number of indexed terms	30	[10, 40]
document profile size	20	[5, 30]
user profile size	10	[5, 30]
number of user clicks	2	[1, 4]

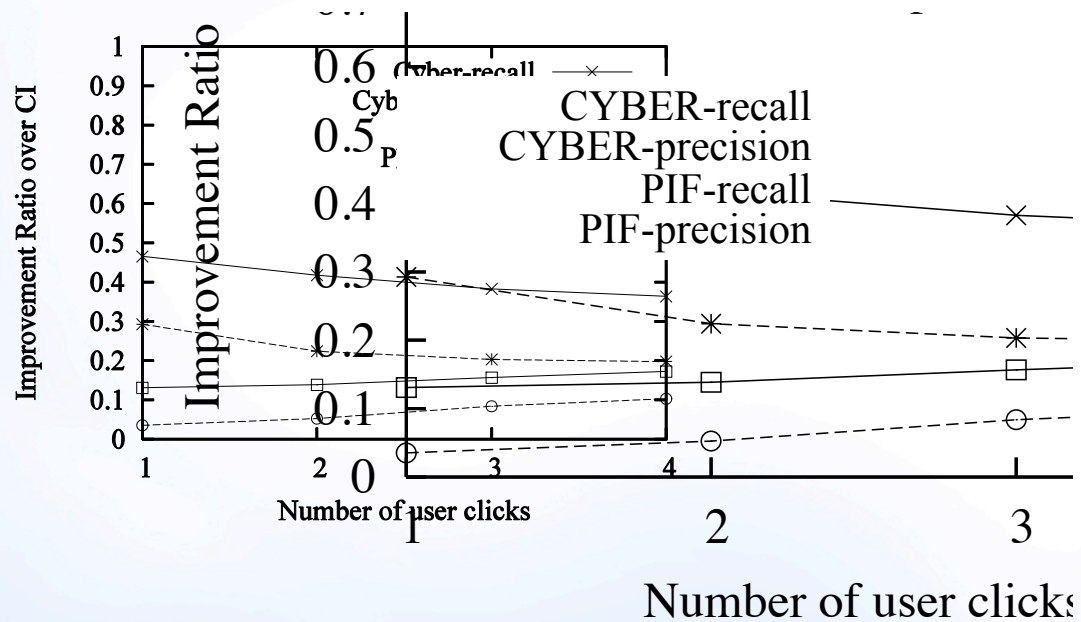
CYBER – Evaluation

- Varying number of answers



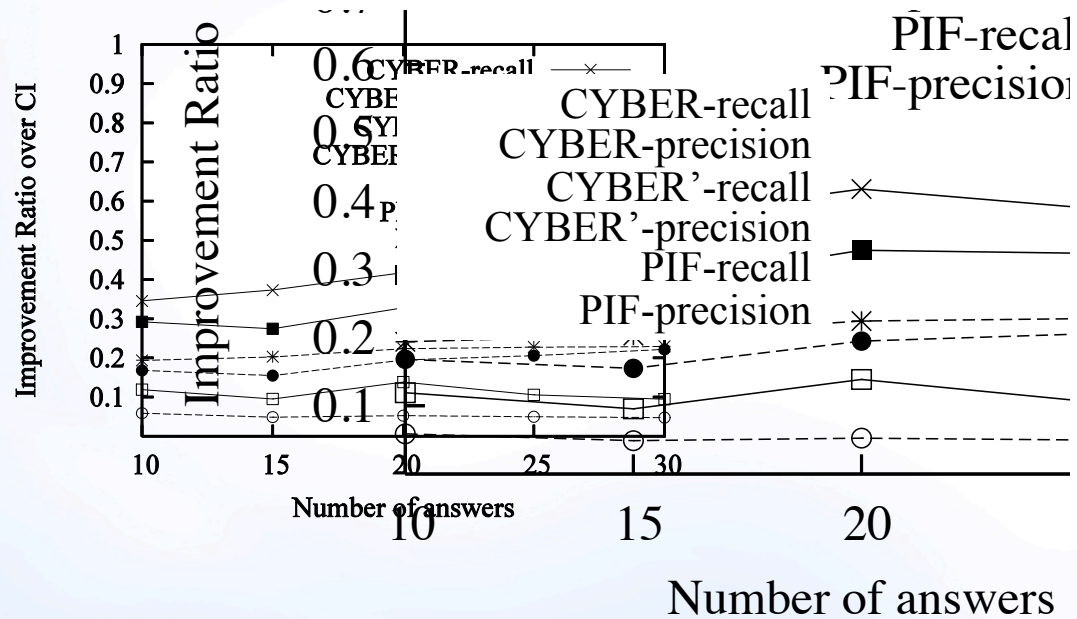
CYBER – Evaluation

- Varying number of clicked documents



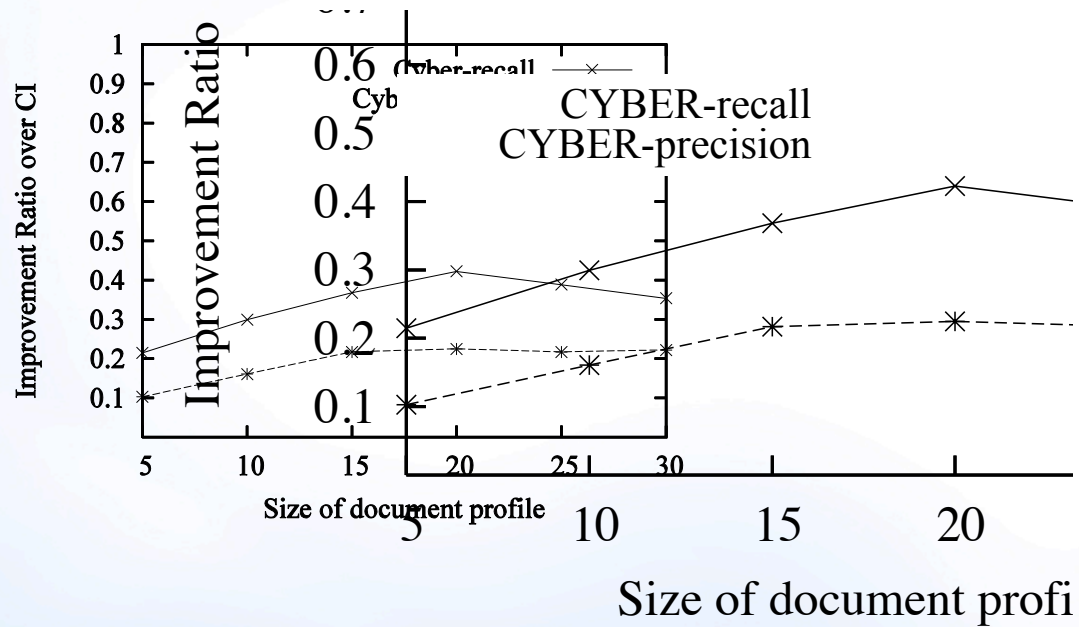
CYBER – Evaluation

- Changes in query pattern



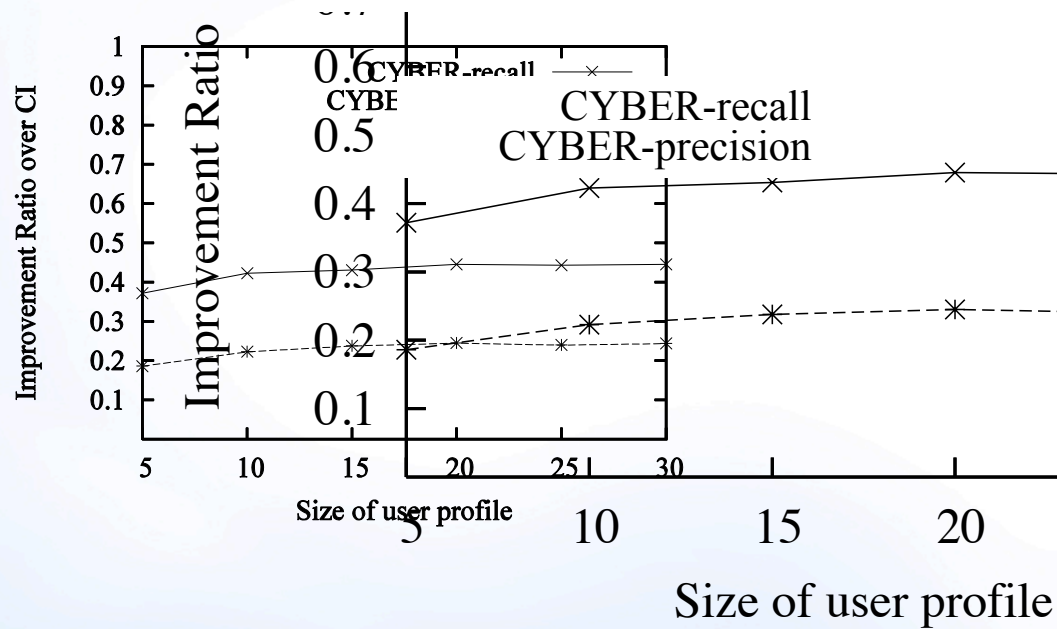
CYBER – Evaluation

- Varying size of document profiles



CYBER – Evaluation

- Varying size of user profiles



Related Work

- Social Networks
 - Ringo – “word of mouth” [CHI95]
 - Amazon, eBay
 - Caching results [HotNets06]
 - “Semantic annotation approach” [ESA07]
 - INGA – an unstructured P2P network [JSAC07]

Conclusion

- Summary of Contributions
 - Avoiding explicit user feedback
 - Managing profiles automatically
 - Better precision and recall