

Enabling Personalized Recommendation on the Web based on User Interests and Behaviors

Y. H. Wu

Y. C. Chen

A. L. P. Chen

**Database Laboratory
Department of Computer Science
National Tsing Hua University**

12/22/2000

Outline

- Introduction**
- User Interests and Behaviors**
- Incremental Mining**
- Dynamic Clustering**
- Personalized Recommendation**
- Performance Evaluation**
- Conclusion**

Introduction

□ Motivation

- ▶ Searching techniques used on the WWW

- ☞ information retrieval

- ☞ information filtering

□ Goal

- ▶ Profile derivation

- ☞ mining

- ▶ Profile matching

- ☞ clustering

1

Introduction

□ Approaches

- ▶ Content-based filtering

- ☞ similarity between documents and profiles

- ▶ Collaborative filtering

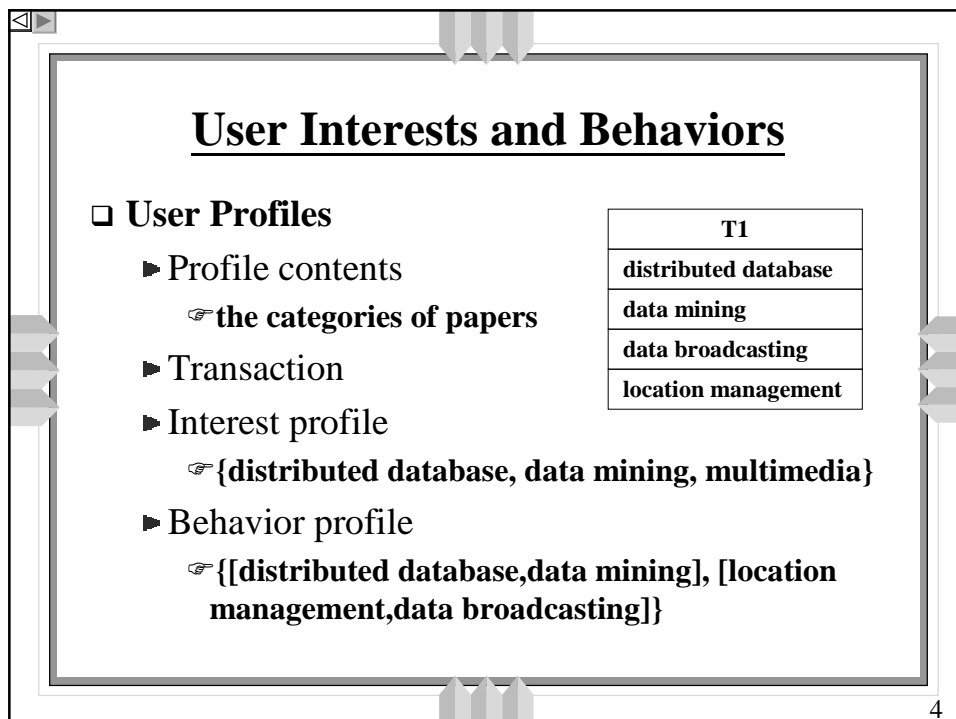
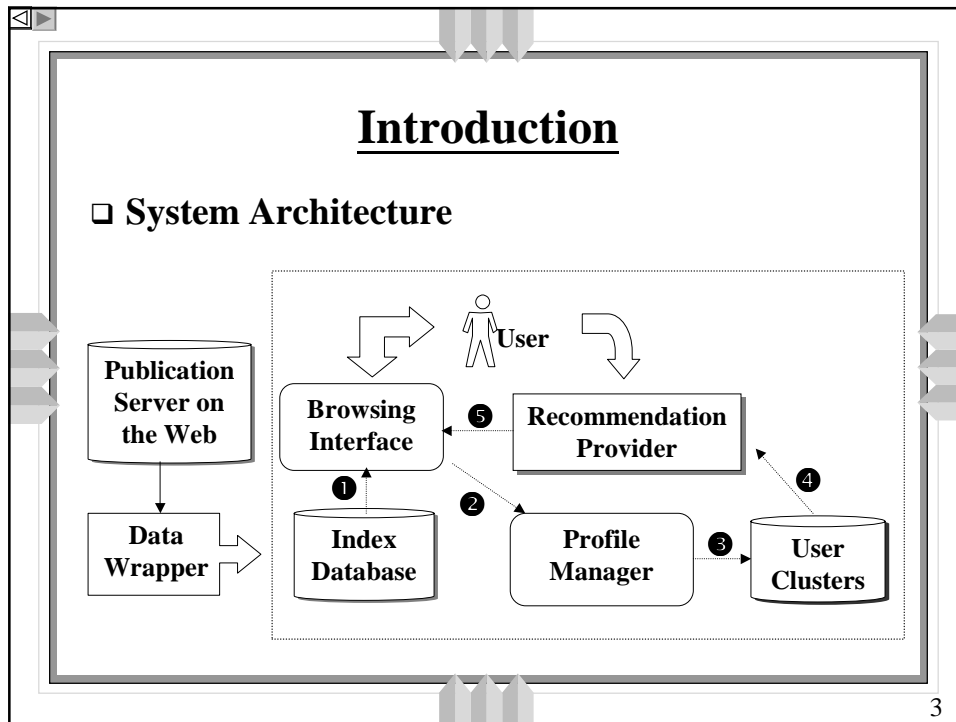
- ☞ similarity between profiles

□ Contributions

- ▶ Both kinds of recommendation services

- ▶ User profiles based on the browsing history

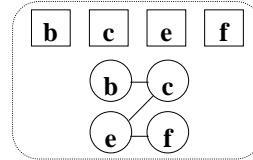
2



User Interests and Behaviors

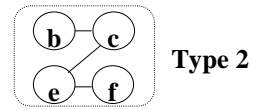
□ I-B Diagram

- ▶ $I = \{b, c, e, f\}$
- ▶ $B = \{[b,c], [c,e], [e,f]\}$
- ☞ $\mathbf{B} = \{b, c, e, f\}$

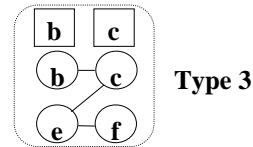


□ Profile Types

- ▶ Type 1: ($I = \emptyset$) and ($\mathbf{B} = \emptyset$)
- ▶ Type 2: ($I = \emptyset$) and ($\mathbf{B} \neq \emptyset$)
- ▶ Type 3: ($I \neq \emptyset$) and ($\mathbf{B} \neq \emptyset$) and ($I \subseteq \mathbf{B}$ or $\mathbf{B} \subseteq I$)



Type 2



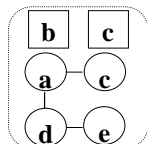
Type 3

User Interests and Behaviors

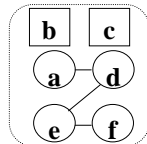
□ Profile Types (Continued)

- ▶ Type 4: ($I \neq \emptyset$) and ($\mathbf{B} \neq \emptyset$) and ($I \not\subseteq \mathbf{B}$) and ($\mathbf{B} \not\subseteq I$) and ($I \cap \mathbf{B} \neq \emptyset$)
- ▶ Type 5: ($I \neq \emptyset$) and ($\mathbf{B} \neq \emptyset$) and ($I \cap \mathbf{B} = \emptyset$)
- ▶ Type 6: ($I \neq \emptyset$) and ($\mathbf{B} = \emptyset$)

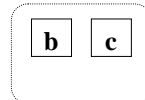
Type 4



Type 5



Type 6



Incremental Mining

Interest Table

- ▶ First and Count columns: support measure
- ▶ Last column: data update

$$\text{Support}(c) = \frac{\text{Count}(c)}{T_c - \text{First}(c) + 1}$$

Category	First	Last	Count	Support
a	T1	T1	1	N/A
b	T2	T4	2	67%
c	T1	T4	3	75%
d	T3	T4	2	100%
e	T1	T3	3	75%
f	T2	T3	2	67%

Transactions
T1 {a, c, e}
T2 {b, c, e, f}
T3 {d, e, f}
T4 {b, c, d}

7

Incremental Mining

Mining Thresholds

- ▶ Minimum support threshold $\alpha=75\%$
- ▶ Minimal count $\gamma=2$
- ▶ Expired time $\lambda=4$

Category	First	Last	Count	Support
a	T1	T1	1	N/A
b	T2	T5	3	75%
c	T1	T5	4	80%
d	T3	T4	2	67%
e	T1	T5	4	80%
f	T2	T5	3	75%
g	T5	T5	1	N/A

Transactions
T1 {a, c, e}
T2 {b, c, e, f}
T3 {d, e, f}
T4 {b, c, d}
T5 {b, c, e, f, g}

8

Incremental Mining

□ Behavior Table

- ▶ 2-Category set column
- ▶ Minimum support threshold $\beta=60\%$

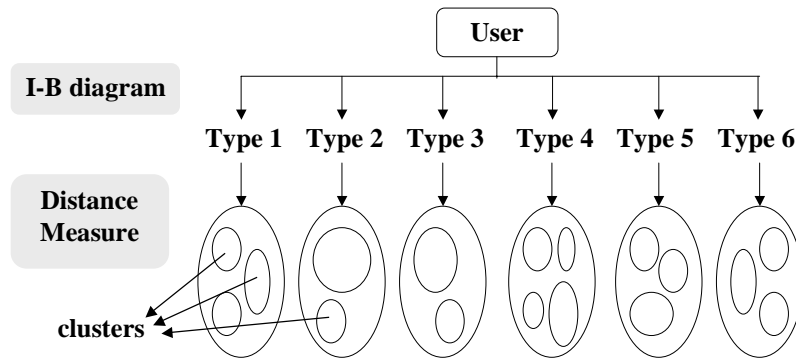
Transactions		2-Category Set	First	Last	Count	Support
T1 {a, c, e}	→	[a,c]	T1	T1	1	N/A
T2 {b, c, e, f}		[a,e]	T1	T1	1	N/A
T3 {d, e, f}		[b,c]	T2	T4	2	67%
T4 {b, c, d}		[b,d]	T4	T4	1	N/A
		[b,e]	T2	T2	1	N/A
		[b,f]	T2	T2	1	N/A
	

9

Dynamic Clustering

□ Framework

- ▶ Two-phase clustering



10

Dynamic Clustering

□ Distance Measure

▶ I-B matrix

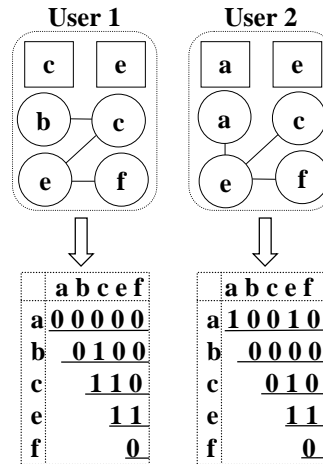
☞ number of dimension

▶ I-B vector

☞ User 1: 000000100110110

☞ User 2: 100100000010110

$$\text{Distance}(A,B) = \sqrt{\sum_{i=1}^n (A_i - B_i)^2}$$



11

Dynamic Clustering

□ Clustering Threshold

▶ Minimum distance threshold δ

□ Centroid-based Approach

▶ No cluster existed

☞ create a new cluster and the centroid

▶ Some clusters existed

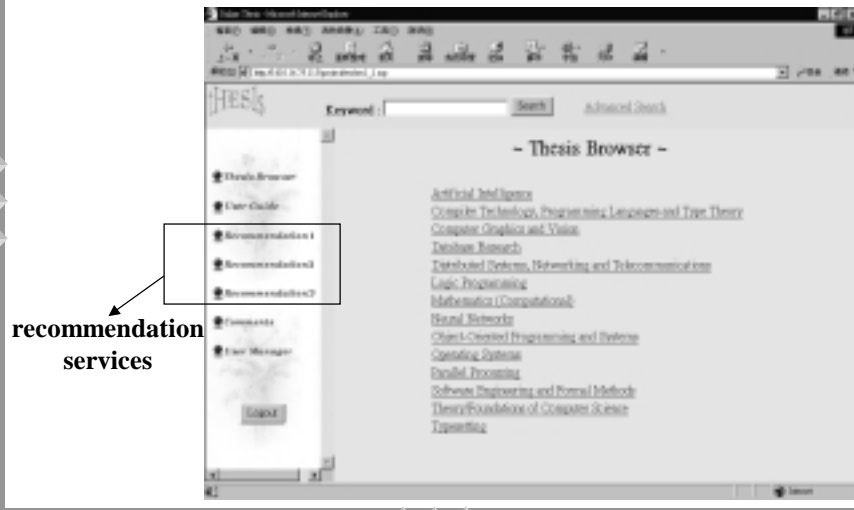
☞ select the centroid with a minimal distance d

✓ $d \leq \delta$: assign to the cluster and update the centroid

✓ $d > \delta$: create a new cluster and the centroid

12

Personalized Recommendation



13

Personalized Recommendation

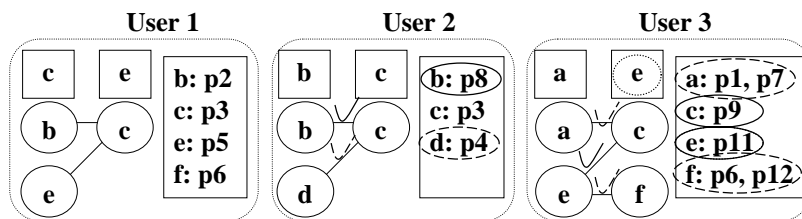
- **Recommendation 1**
 - ▶ Interesting papers (content-based filtering)
- **Recommendation 2**
 - ▶ From the users with similar interests
- **Recommendation 3**
 - ▶ From the users with similar behaviors
- **Recommendation 4**
 - ▶ Potentially interesting papers

14

Personalized Recommendation

□ Example for User 1

- ▶ Recommendation 2: p11 from User3
- ▶ Recommendation 3: p8 (User2), p9, p11 (User3)
- ▶ Recommendation 4: p4 (User2), p1, p7 (User3), p6, p12 (User3, category f)



15

Performance Evaluation

□ Experimental Setting

- ▶ 230 papers, 14 fields each with some categories
 - ☞ title, URL, author name, keyword, category, ...
- ▶ 13 students, 3 levels of feedbacks
 - ☞ good, moderate, bad
 - ☞ Ng, Nm, Nb ⇒ hit ratio, average ratio, miss ratio
- ▶ Parameters
 - ☞ $\alpha=40\%$, $\beta=25\%$, $\gamma=2$, $\lambda=10$
 - ☞ $\delta=1.6, 2.4, 3.2, 4.0, 4.8$

16

Performance Evaluation

□ Clustered Results

Cluster	$\delta=1.6$	$\delta=2.4$	$\delta=3.2$	$\delta=4.0$	$\delta=4.8$	R2	$\delta=2.4$	$\delta=3.2$	$\delta=4.0$	$\delta=4.8$
C1	U1	U1	U1,U2	U1-U3	U1-U11	U1		100%	100%	50%
C2	U2	U2				U2		33%	25%	40%
C3	U3	U3	U3			U3				100%
C4	U4	U4,U5	U4,U5	U4-U7		U4	50%	50%	52%	42%
C5	U5					U5	67%	67%	80%	31%
C6	U6	U6,U7	U6,U7			U6	50%	50%	20%	43%
C7	U7					U7	19%	19%	24%	50%
C8	U8	U8	U8	U8,U9		U8				38%
C9	U9	U9	U9			U9			11%	17%
C10	U10	U10	U10	U10		U10				42%
C11	U11	U11	U11	U11		U11				21%
C12	U12	U12	U12	U12	U12,U13	U12				50%
C13	U13	U13	U13	U13		U13				0%

17

Performance Evaluation

□ Comparisons ($\delta=3.2$)

R4	$\delta=2.4$	$\delta=3.2$	$\delta=4.0$	$\delta=4.8$
U1		50%	50%	50%
U2		25%	25%	36%
U3		63%		
U4	63%	62%	38%	41%
U5	67%	59%	40%	30%
U6	50%	42%	22%	38%
U7	42%		57%	43%
U8			33%	38%
U9			33%	44%
U10				29%
U11				20%
U12				0%
U13				0%

	Hit Ratio	Average Ratio	Miss Ratio	Accept Ratio
R1	48.5%	46.1%	5.4%	94.6%
R2	53.1%	35.4%	11.5%	88.5%
R3	48.7%	22.8%	28.5%	71.5%
R4	33.2%	37.3%	29.5%	70.5%

18

Conclusion

□ Summary

- ▶ Define and mine user interests and behaviors
- ▶ Type and cluster user profiles
- ▶ Design four kinds of recommendation services

□ Future Issues

- ▶ The extent of the user profiles
- ▶ The effects of N-category sets ($N > 2$)
- ▶ The settings of the thresholds
- ▶ Complexity analysis and comparisons