媒体计算技术

李泽超

计算机科学与工程学院


Now, Let us study a new media
Social Media


Writeboard
SHOUTWIRE
iKarma
CalendarHub
pando
zigtag
Findory
backfence
clipmarks
wayfaring
Rallypoint
Zoozio
blogbeat
Ziggs
zoto
vSocial
Opinity
Because your reputation matters!
reddit
measuremap
gumshoo
bluepulse
Ta-da Lists
FeedSky
jellyBarn.INC
PODZINGER
RSS MAD
FeedTier
phanfare
WIKIPEDIA
The Free Encyclopedia
Ning
Ookles
Strongspace
zoominfo
CASTPOST
FeedBurner
Bloglines
purevolume.com
FOTOLOG
Yub.com
Spot Runner
myspace
a place for friends
NewsAlloy
Gcast
blinkx
search
openomy
riffs
ajchat
alpha
jambo
NETWORKS
ROLLYO
PANDORA
looklater
beta
PROTOPAGE
PLAZES
Noodly
30
BETA
wondir
diigo
box
digg
del.icio.us
omni drive
AlmondRocks
Tagyu
Your tags, smarter
30GIGS.com
writely
SQUIDOO

Social Networks

Social networking accounts for 22% of all time spent online in the U.S.

Twitter averages almost 40 million ‘tweets’ per day

1 in 4 people over the age of 65 use social networking sites

Examples:

- Facebook
- Twitter
- Myspace
- LinkedIn
- Skype

“There are over 200 active sites using a wide variety of social networking models today.”

Recommendation

- General
- Personal

The Social Web

The Web not only as a marketplace for companies that want to sell, but as a social space for people.

In this social space, by sharing news, ideas, information, stories, etc., communities of like-minded individuals are created.

The Web as a publishing medium: everyone is an author.

The Web as a medium to harness “the wisdom of crowds”.

The participatory Web

Timeline of Social Web

December 1997: the word ‘Weblog’ was coined by Jorn Barger. In 1998, the word was pronounced “We blog” and shortened to “Blog”.

End of 1998 – 1999: Creation of several hosting services for blogs: OpenDiary, LiveJournal, Blogger.

January 2001: The launch of Wikipedia

August 2003: The launch of MySpace

February 2004: The launch of Flickr

December 2004: The launch of Digg

February 2005: The launch of YouTube

Categories of Social Web

Blogging

Social Networking (MySpace, Facebook)

Collaborative Knowledge Creation (Wikipedia)

Content Sharing (Flickr, YouTube)

Social Bookmarking (Digg, Del.icio.us)

Recommendation Engines (Amazon, Netflix, Last.fm)

Social Gaming

Open-source software (Linux, Apache, Python, etc.)

> *None of these would be interesting without massive user-generated content.*

Blogging

Transition from personal websites to blogs.

RSS (also known as Web feeds)

- Really Simple Syndication (Dave Winner)
- Rich Site Summary (Netscape)
- Subscribe to one page and get notified when the page changes.

Permalinks

Trackbacks

Free blog hosting services (Blogger, WordPress):

- Templating (no need to write HTML)
- rich editing (WYSIWYG)
- adding photos, music, video (the widget technology)

Collaborative Knowledge Creation

Examples: Wikipedia, WikiTravel, Wiktionary

Technology: wiki – a page that can be modified by anyone, using just a web browser.

wiki has been around since 1995

There are public and private Wikis.

Problems: Bias, accuracy, vandalism

Wikipedia ranks 8th in terms of global web traffic.

Content Sharing

Examples: Flickr, YouTube, Picassa

Content (videos, photos) is uploaded from users and made available to all.

Everyone can tag and comment on the content.

Tags can be used for retrieval of content.

Content can be embedded in other social sites through feeds or widgets.

YouTube is ranked 3rd (after Yahoo and Google) in terms of traffic ranking.

Social Bookmarking

Examples: Digg, Del.icio.us, StumbleUpon

Tags

Folksonomy versus Taxonomy

Thumbs-up, thumbs-down (a voting system)

- Features like “most e-mailed”, “most read” (in all other sites)

Risks:

- Users who promote sites that promote products and services
- The cascade process of reinforcement

Social Gaming

One of the first uses of Internet, well before Web 2.0.

Origin: Text-driven MUD (Multi-User Dungeon) (played by telnet)

Viral games: “Zombies”, “Vampires”, “Werewolfes”

Virtual Life (a social virtual world)

Google Lively

Facebook games:

- Who has the biggest brain?
- Scrabulous

Games in other platforms (Xbox, Playstation, Nintendo) with Internet connection.

Recommendation Engines

Examples: Amazon, Netflix, Last.fm

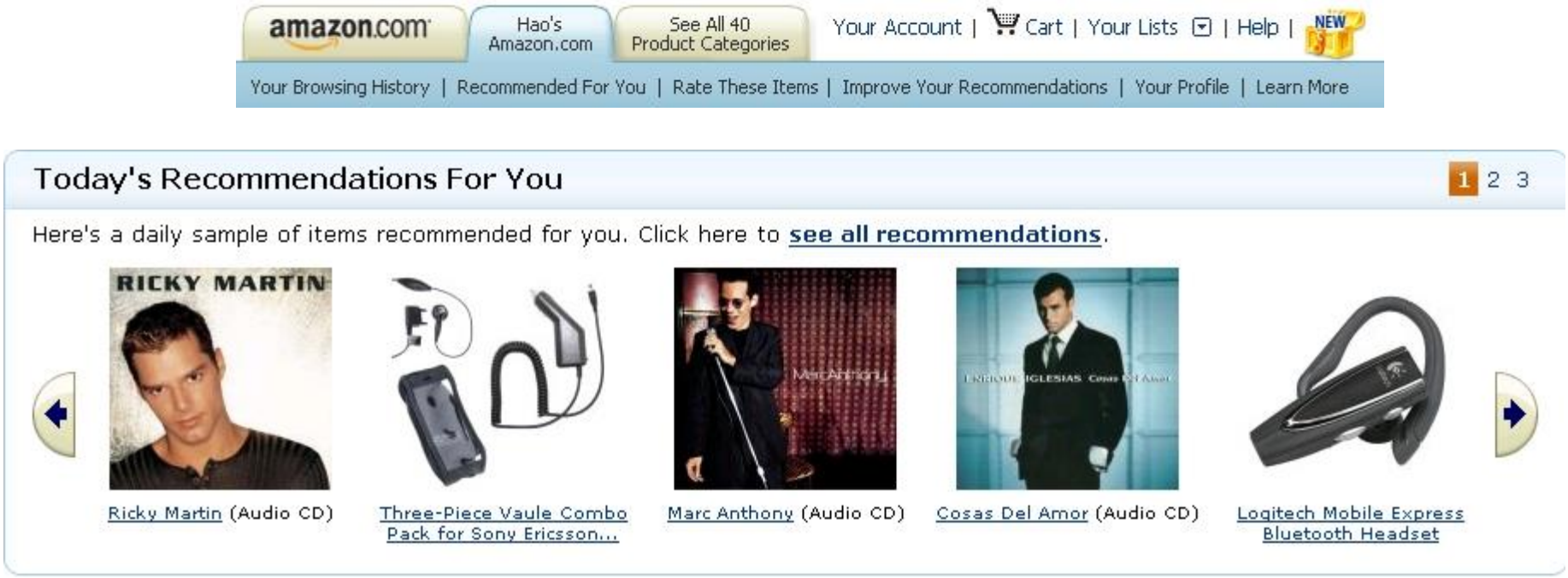
Personalized recommendation: recommend things based on the individual's paste behavior

Social recommendation: recommend things based on the past behavior of similar users

Item recommendation: recommend things based on the item itself

Background

Do you have this experience?



YAHOO! MOVIES

My Movies: gabe_ma Edit Profile

Recommendations For You

Receive Recommendations by Email

Movies in Theaters: 94089

Burn After Reading (R)
Showtimes & Tickets | Add to My Lists

Yahoo! Users: B- 4794 ratings
The Critics: B 14 reviews

Don't Recommend Again Seen It? Rate It!

Pride and Glory (R)
Showtimes & Tickets | Add to My Lists

Yahoo! Users: A- 59 ratings
The Critics: C+ 6 reviews

Don't Recommend Again Seen It? Rate It!

Fight Club (R)
Showtimes & Tickets | Add to My Lists

Yahoo! Users: B+ 52392 ratings
The Critics: B 12 reviews

Don't Recommend Again Seen It? Rate It!

Lakeview Terrace (PG-13)
Showtimes & Tickets | Add to My Lists

Yahoo! Users: B 3229 ratings
The Critics: C 12 reviews

Don't Recommend Again Seen It? Rate It!

Vicky Cristina Barcelona (PG-13)
Showtimes & Tickets | Add to My Lists

Yahoo! Users: B 1923 ratings
The Critics: B+ 13 reviews

Don't Recommend Again Seen It? Rate It!

The Duchess (PG-13)
Showtimes & Tickets | Add to My Lists

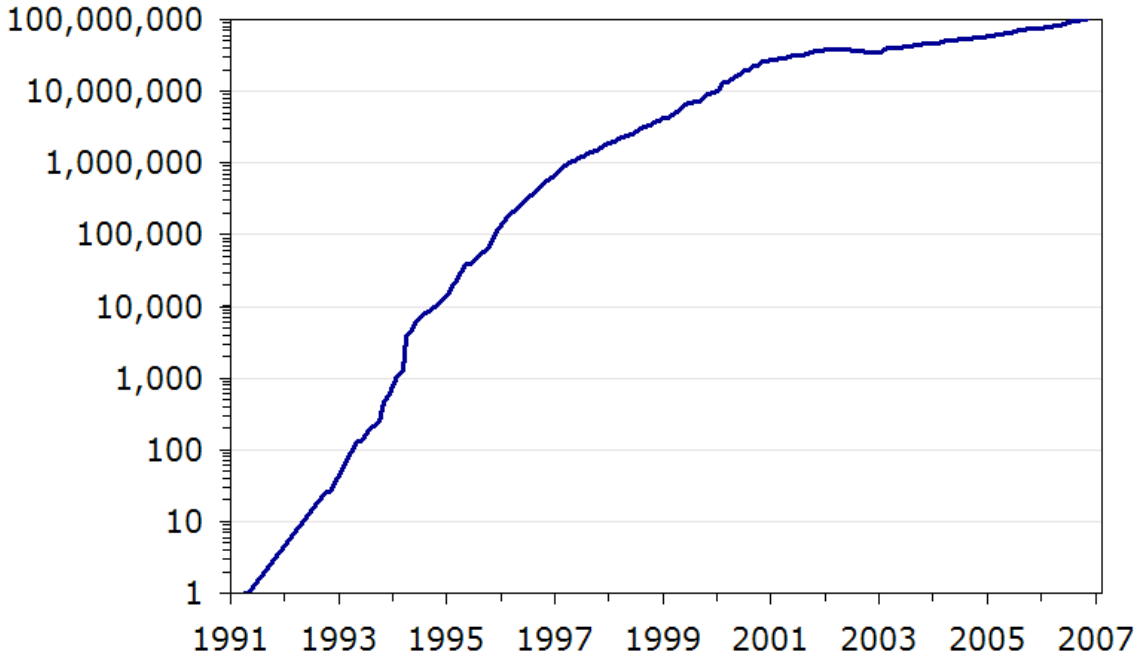
Yahoo! Users: B+ 953 ratings
The Critics: B- 10 reviews

Don't Recommend Again Seen It? Rate It!

See All Recommendations

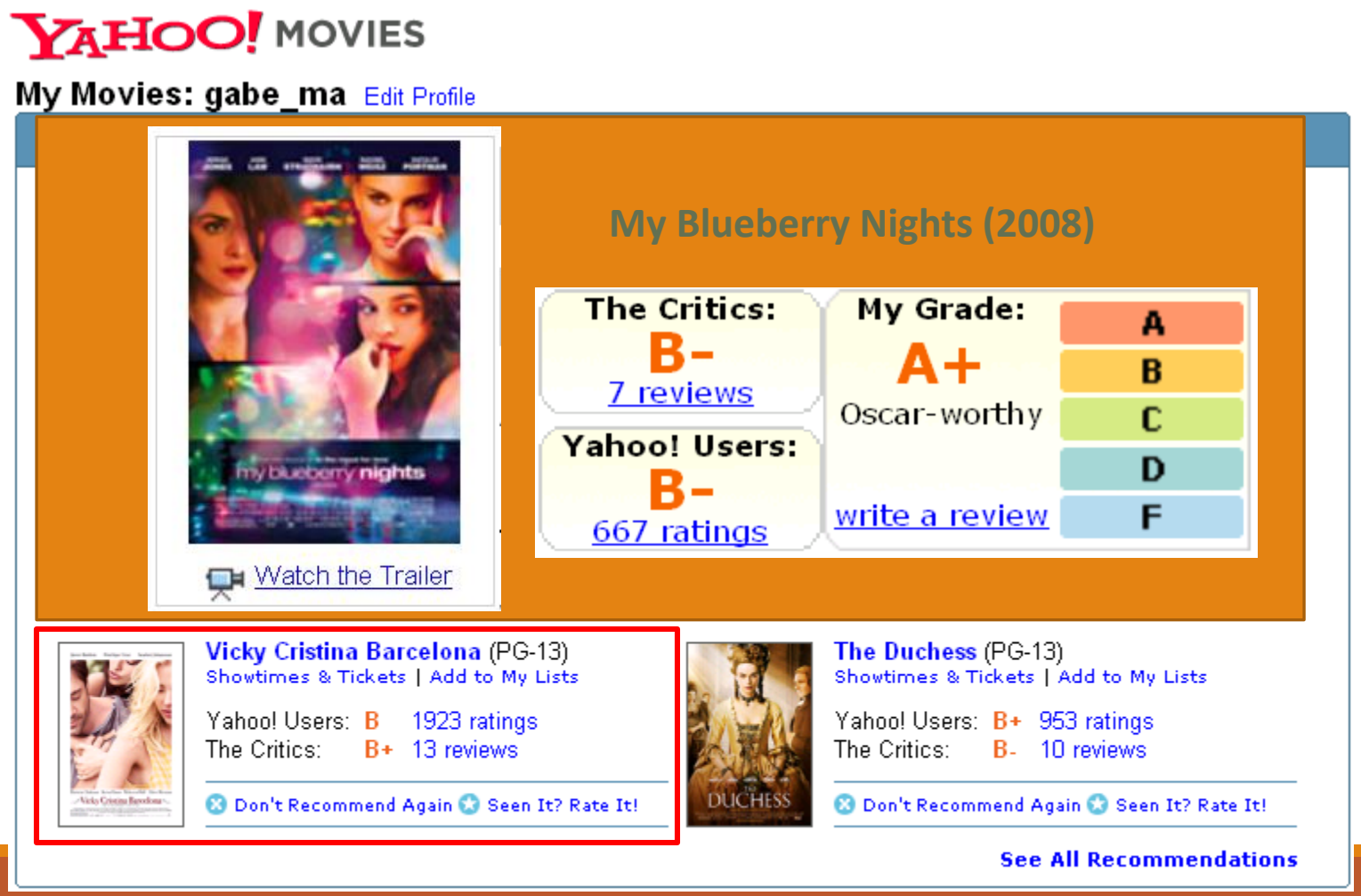
Background

Recommender Systems become more and more important

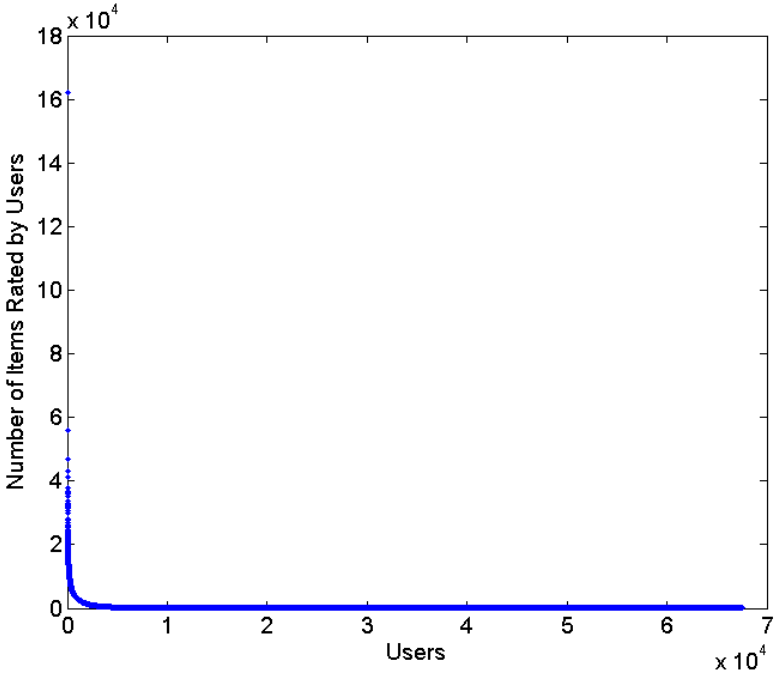


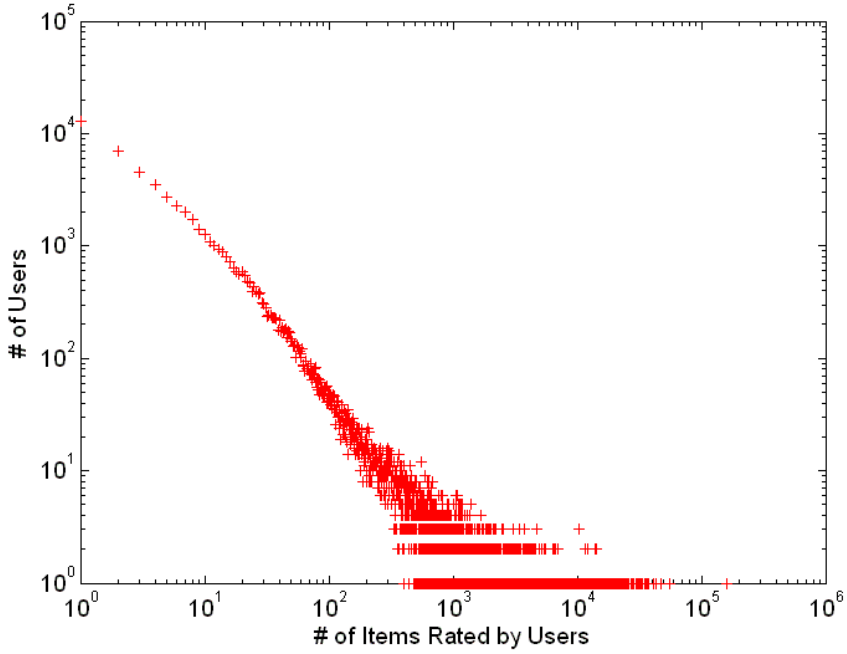
*The **number of Internet websites** each year since the Web's founding.*
From http://www.useit.com/alertbox/web-growth.html

Challenges
Data sparsity problem



Number of Ratings per User





Extracted From Epinions.com

114,222 users, 754,987 items and 13,385,713 ratings

Challenges

Traditional recommender systems ignore the social connections between users

Challenges

"Yes, there is a correlation - from social networks to personal behavior on the web"

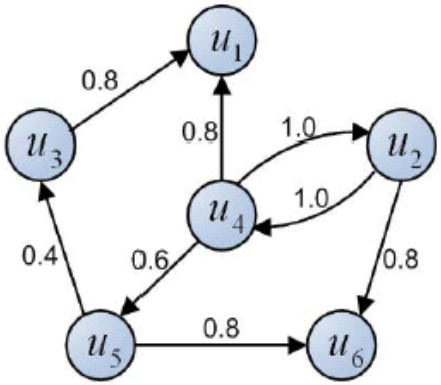
Parag Singla and Matthew Richardson (WWW'08)

- Analyze the who talks to whom social network over 10 million people with their related search results
- People who chat with each other are more likely to share the same or similar interests

Motivation

To improve the recommendation accuracy and solve the data sparsity problem, users' social network should be taken into consideration

Problem Definition



(a) Social Network Graph

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	5	2		3		4		
u_2	4	3			5			
u_3	4		2				2	4
u_4								
u_5	5	1	2		4	3		
u_6	4	3		2	4		3	5

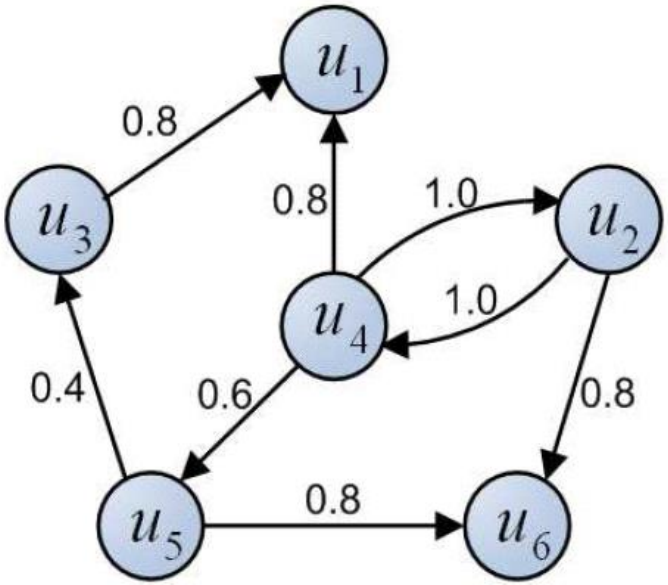
(b) User-Item Matrix

$$U = \begin{bmatrix} 1.55 & 1.22 & 0.37 & 0.81 & 0.62 & -0.01 \\ 0.36 & 0.91 & 1.21 & 0.39 & 1.10 & 0.25 \\ 0.59 & 0.20 & 0.14 & 0.83 & 0.27 & 1.51 \\ 0.39 & 1.33 & -0.43 & 0.70 & -0.90 & 0.68 \\ 1.05 & 0.11 & 0.17 & 1.18 & 1.81 & 0.40 \end{bmatrix},$$

$$V = \begin{bmatrix} 1.00 & -0.05 & -0.24 & 0.26 & 1.28 & 0.54 & -0.31 & 0.52 \\ 0.19 & -0.86 & -0.72 & 0.05 & 0.68 & 0.02 & -0.61 & 0.70 \\ 0.49 & 0.09 & -0.05 & -0.62 & 0.12 & 0.08 & 0.02 & 1.60 \\ -0.40 & 0.70 & 0.27 & -0.27 & 0.99 & 0.44 & 0.39 & 0.74 \\ 1.49 & -1.00 & 0.06 & 0.05 & 0.23 & 0.01 & -0.36 & 0.80 \end{bmatrix},$$

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	5	2	**2.5**	3	**4.8**	4	**2.2**	**4.8**
u_2	4	3	**2.4**	**2.9**	5	**4.1**	**2.6**	**4.7**
u_3	4	**1.7**	2	**3.2**	**3.9**	**3.0**	2	4
u_4	**4.8**	**2.1**	**2.7**	**2.6**	**4.7**	**3.8**	**2.4**	**4.9**
u_5	5	1	2	**3.4**	4	3	**1.5**	**4.6**
u_6	4	3	**2.9**	2	4	**3.4**	3	5

Social Network Graph Matrix Factorization



(a) Social Network Graph

$$p(C|U, Z, \sigma_C^2) = \prod_{i=1}^{m} \prod_{k=1}^{m} \mathcal{N}\left[\left(c_{ik}|g(U_i^T Z_k), \sigma_C^2\right)\right]^{I_{ik}^C}$$

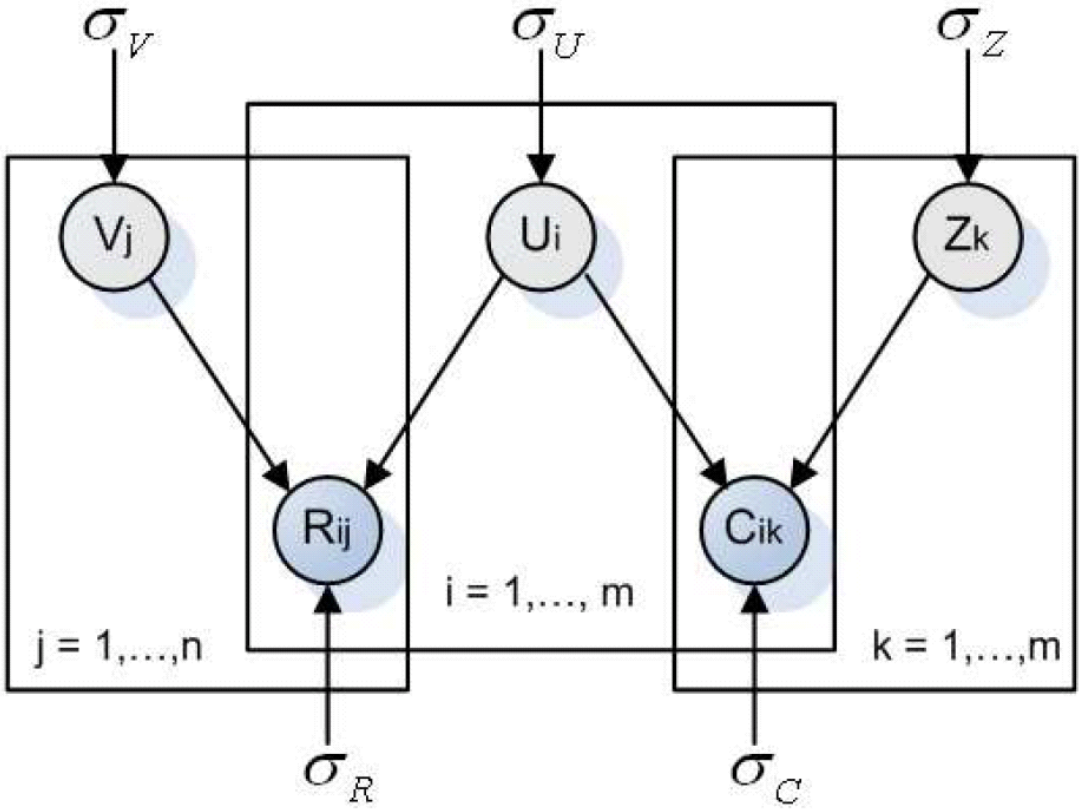
User-Item Rating Matrix Factorization

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	5	2		3		4		
u_2	4	3			5			
u_3	4		2				2	4
u_4								
u_5	5	1	2		4	3		
u_6	4	3		2	4		3	5

(b) User-Item Matrix

$$p(C|U,V,\sigma_R^2) = \prod_{i=1}^{m}\prod_{j=1}^{n}\mathcal{N}\left[\left(r_{ij}|g(U_i^T V_j),\sigma_R^2\right)\right]^{I_{ij}^R}$$

Social Recommendation



$$\mathcal{L}(R, C, U, V, Z) =$$

$$\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^{R}(r_{ij} - g(U_i^T V_j))^2 + \frac{\lambda_C}{2}\sum_{i=1}^{m}\sum_{k=1}^{m} I_{ik}^{C}(c_{ik}^{*} - g(U_i^T Z_k))^2$$

$$+\frac{\lambda_U}{2}\|U\|_F^2 + \frac{\lambda_V}{2}\|V\|_F^2 + \frac{\lambda_Z}{2}\|Z\|_F^2,$$

Gradient Descent

$$\frac{\partial \mathcal{L}}{\partial U_i} = \sum_{j=1}^{n} I_{ij}^R g'(U_i^T V_j)(g(U_i^T V_j) - r_{ij}) V_j$$

$$+ \lambda_C \sum_{j=1}^{m} I_{ik}^C g'(U_i^T Z_k)(g(U_i^T Z_k) - c_{ik}^*) Z_k + \lambda_U U_i,$$

$$\frac{\partial \mathcal{L}}{\partial V_j} = \sum_{i=1}^{m} I_{ij}^R g'(U_i^T V_j)(g(U_i^T V_j) - r_{ij}) U_i + \lambda_V V_j,$$

$$\frac{\partial \mathcal{L}}{\partial Z_k} = \lambda_C \sum_{i=1}^{m} I_{ik}^C g'(U_i^T Z_k)(g(U_i^T Z_k) - c_{ik}^*) U_i + \lambda_Z Z_k,$$

Complexity Analysis

For the Objective Function $O(\rho_R l + \rho_C l)$

For $\frac{\partial \mathcal{L}}{\partial U}$, the complexity is $O(\rho_R l + \rho_C l)$

For $\frac{\partial \mathcal{L}}{\partial V}$, the complexity is $O(\rho_R l)$

For $\frac{\partial \mathcal{L}}{\partial Z}$, the complexity is $O(\rho_C l)$

In general, the complexity of our method is linear with the observations in these two matrices

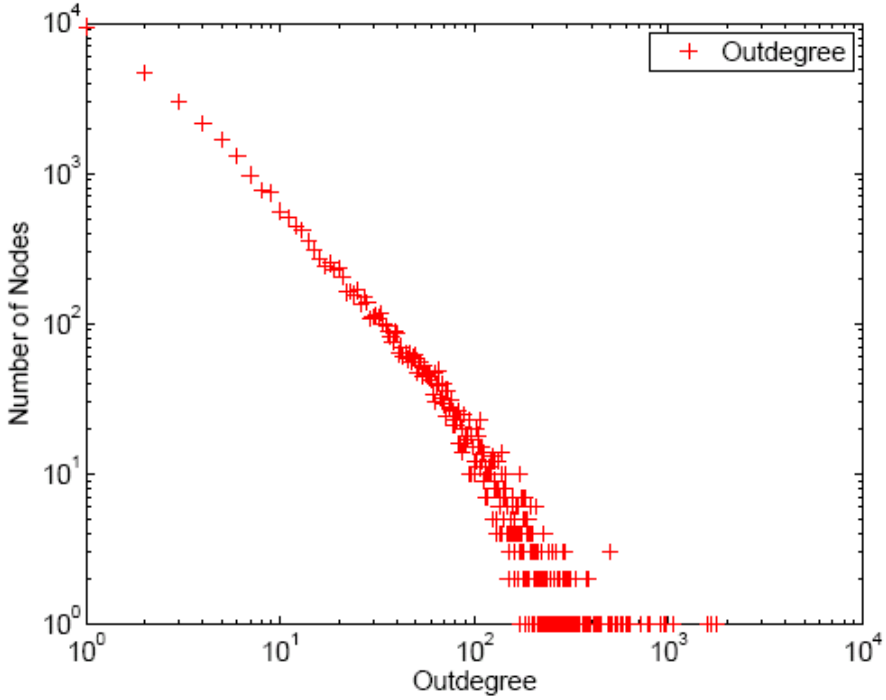
Epinions Dataset

40,163 users who rated 139,529 items with totally 664,824 ratings

Rating Density 0.01186%

18,826 users, representing 46.87% of the population, submitted fewer than or equal to 5 reviews

The total number of issued trust statements is 487,183



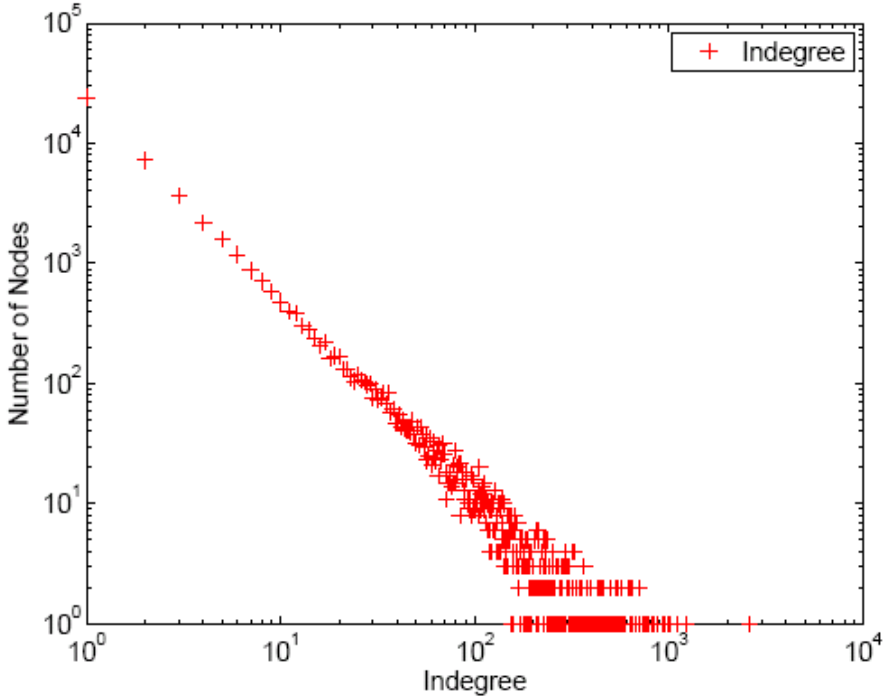


Figure 3: Degree Distribution of User Social Network

Metrics

Mean Absolute Error

$$MAE = \frac{\sum_{i,j} |r_{i,j} - \widehat{r}_{i,j}|}{N}$$

Personal Recommendation

- General
- Personal

Personalized

Each user has her personal preference

标签:
lawn
building
architecture
landscape

用户 1

标签:

white house
washington D. C.

用户 2

Introduction (2/3)

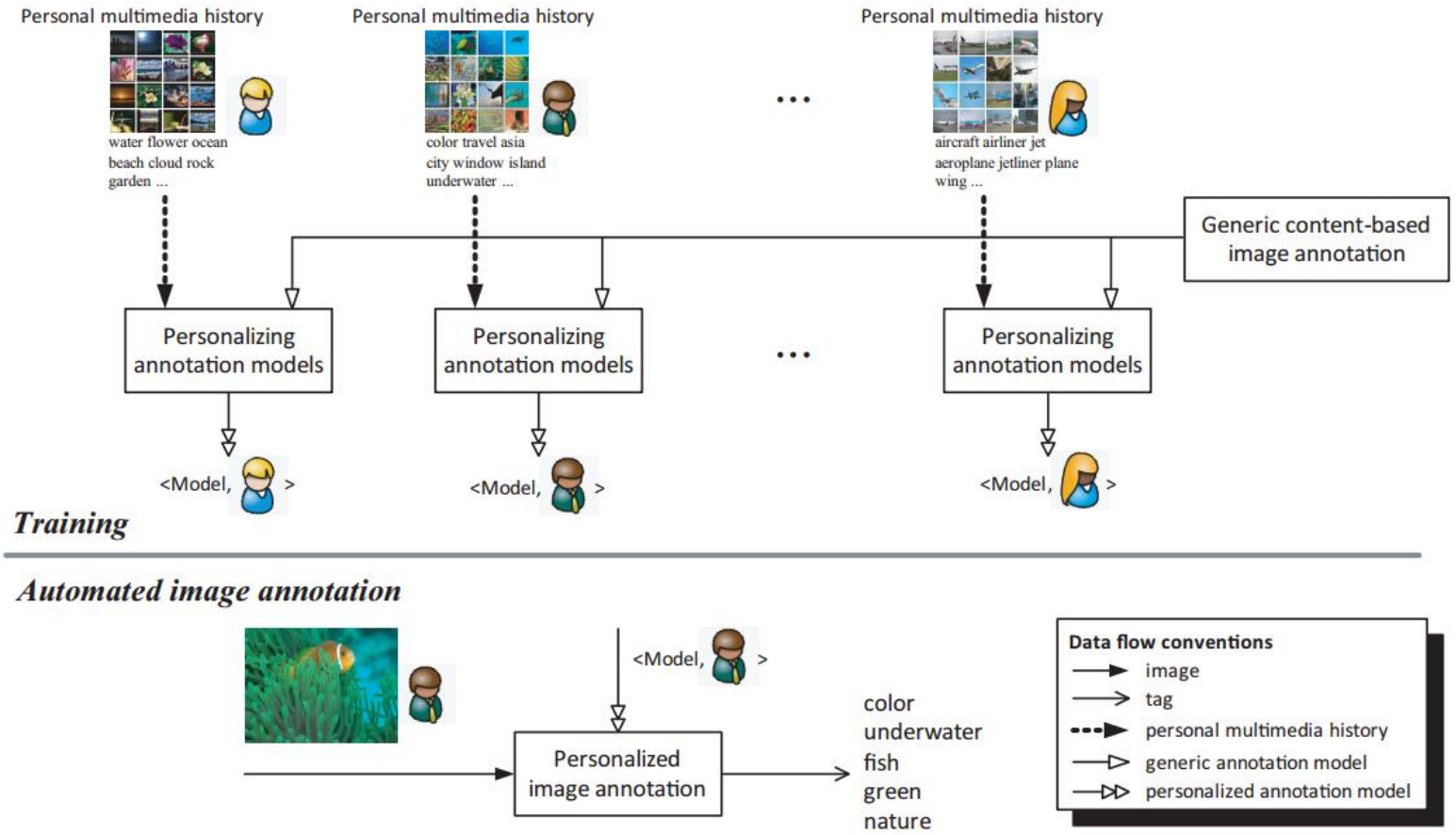
Why automatic tagging?

- Webpage are growth very fast
- Recommendation

Why personalization?

- Automatically generated tags have the drawback of presenting only a generic view

Personalizing Automated Image Annotation using Cross-Entropy



研究动机

- 装载GPS的摄像设备普及，个人照片迅速增长，如何管理个人照片成为一项迫切的任务
- 移动终端（如手机）操作不方便（屏小）
- 便捷快速地管理个人照片
- 自动标签推荐

研究动机

- 自动标签推荐
  - ✓个性化
  - ✓地理信息

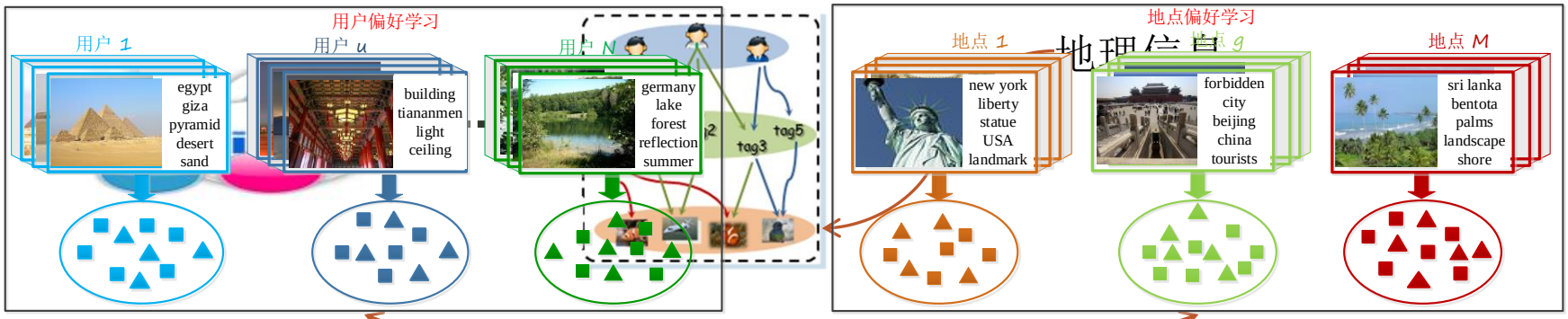
标签:
menara
kuala
lumper
tower

地点: Kuala Lumper

标签:
CCTV
tower
beijing C.

地点: Beijing

框架

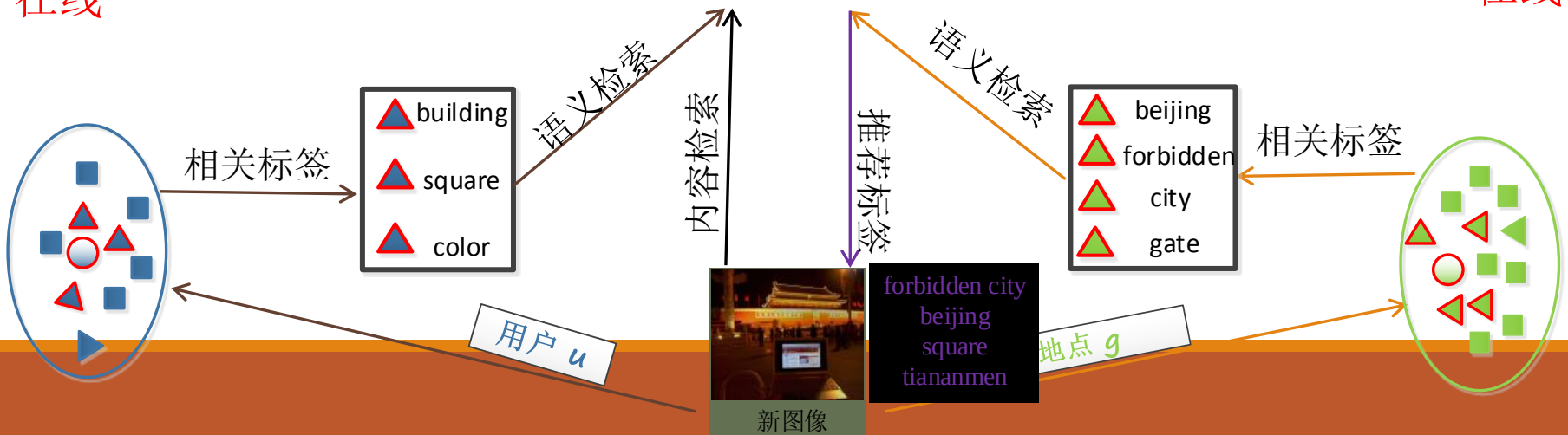

用户偏好学习
用户 1
用户 u
用户 N
egypt giza pyramid desert sand
building tiananmen light ceiling
germany lake forest reflection summer
tag5
tag3
地点偏好学习
地点 1
地点 g
地点 M
new york liberty statue USA landmark
forbidden city beijing china tourists
sri lanka bentota palms landscape shore

离线

离线

在线

在线


相关标签
building
square
color
语义检索
内容检索
推荐标签
语义检索
beijing
forbidden
city
gate
相关标签
用户 u
forbidden city
beijing
square
tiananmen
地点 g
新图像

数据

- 数据采集：Flickr

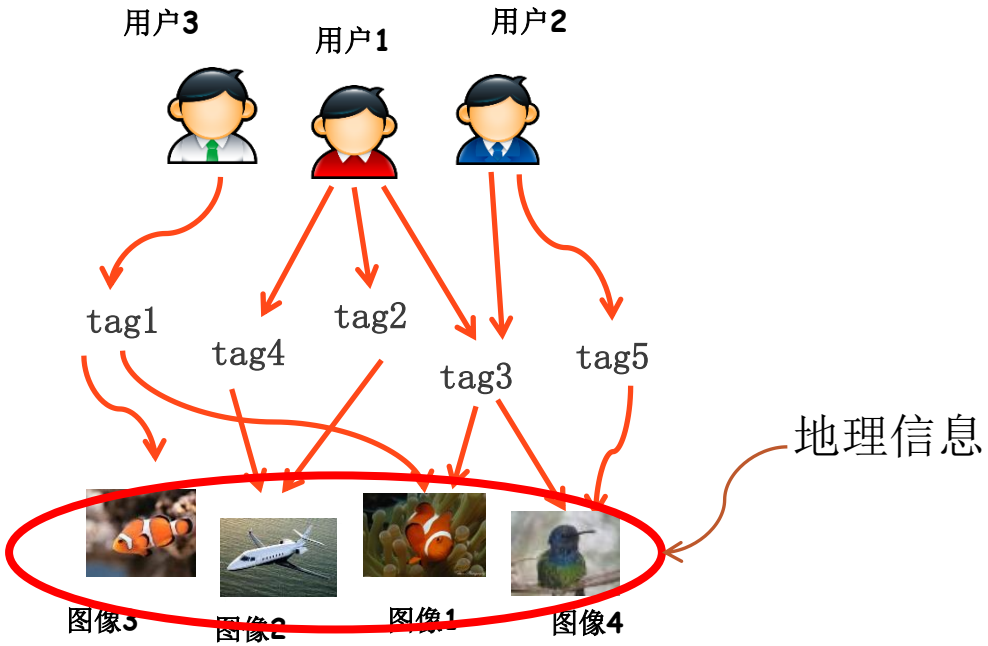


表 5-2 数据集的统计信息

图像个数	用户个数	地点个数	标签个数
2,727,312	559	1,351	15,554

Tag Suggestion on YouTube by Personalizing Content-based Auto-Annotation