Social Information Filtering

Hang Li

Noah's Ark Lab

Huawei Technologies

Talk Outline

- Huawei Noah's Ark Lab
- Our Vision and Approach
- Social Information Filtering
- Weibo Robot

Huawei Noah's Ark Lab

- Mission:
  - Building "Noah's Ark" to overcome the deluge of information
- Research Areas:
  - Data Mining
  - Machine Learning
  - Natural Language Processing
  - Information Retrieval
  - Social Media
  - Mobile Intelligence
  - Human Computer Interaction
- Location: Hong Kong, Shenzhen, Beijing
- Established: June 2012

Our Vision

- Vision
  - From Big Data to Deep Knowledge
- Goal
  - Building intelligent assistants on mobile devices using big data on the cloud
  - Passing Turing tests in specific domains

Our Approach

- Big Machine Learning
  - Acquiring knowledge by large scale machine learning
  - Example: Google, learning concept of cat from millions of images
- Human Computation
  - Humans and machines jointly accomplish tasks
  - Example: Univ. of Washington, solving large scale gene problem by leveraging human power
- Lifelong Learning
  - Continuously accumulating knowledge by machine learning
  - Example: CMU, NELL (Never Ending Language Learning)

Example: Intelligent Mobile Device

- Video
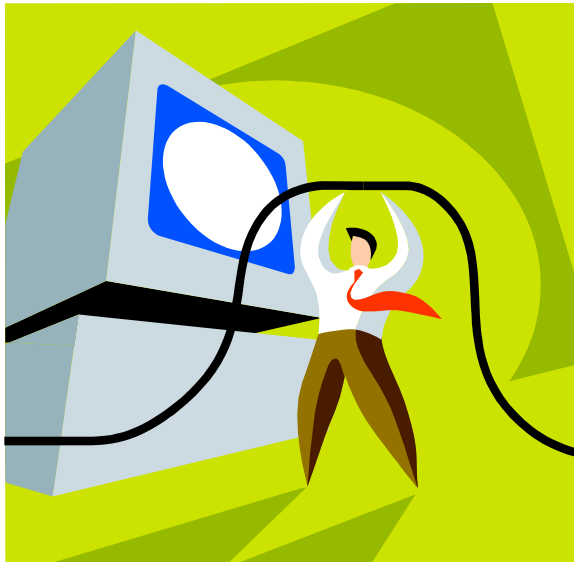
Two Challenges: Information Overload and Information Shortage

Social Information Filtering

- Valuable information is shared on social media
- Can get information by following key people
- Taking social media as information filter
- Building information assistant for each user or each group of users based on the 'social information filter'
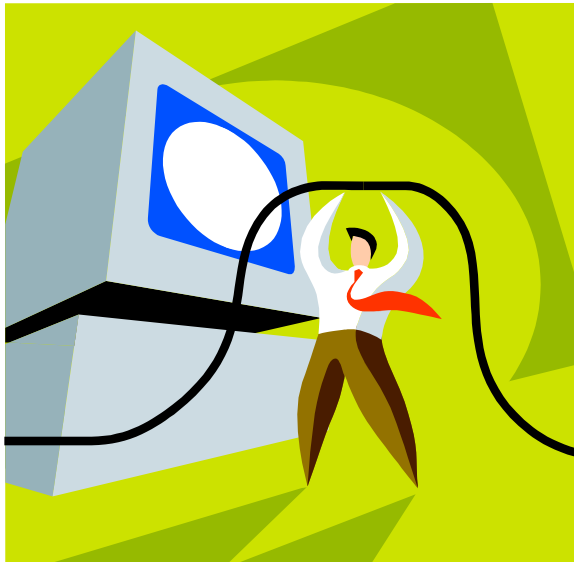
Weibo Robot

- Intelligent Information Assistant
- Built on social media
- Each user has her own Weibo Robot
- Each community has their own Weibo Robot
- Weibo Robot can understand user or community's information needs
- Weibo Robot can help user or community to collect, analyze, disseminate information

Weibo Robot

- Video

Weibo Robot: Technical Challenges

- Social media data is unstructured, dynamic, and noisy
- Scale is very large
- Need to provide the right information to the right people at the right time
- Language understanding and generation is challenging

Main Features of Weibo Robot V1

- Features Developed
  - Following People
  - Re-Tweeting (Forwarding Tweets)
  - Generating Simple Comments
- Features To Be Developed
  - Generating Tweets based on Analysis of Weibo Trends
  - Generating Tweets based on Hot Topics in Other Information Sources

Weibo Robot V1: Current Status

- Weibo Robot Version 1.0 released on Jan 1, 2013
- Number of users to follow = 400 as of Aug 17, 2013
- In the community of NLP and IR
- Number of followers = 368 as of Aug 17, 2013
- Capacity of System
  - Crawling 15M tweets / day
  - Monitoring 4k users online / day
  - Processing data of 120k users / day
- Received private messages from real people
  - "Who are you? I would like to connect with you…"

Microsoft Research

微软亚洲研究院终于有 **60000** 位粉丝啦。

你们的每一条转发和评论主页君都会认真浏览。
亲爱的@小诺_pinocchio @明天你是top @蓝色梦旅人
@谢涛TaoXie ，
你们有着最高的互动率，想必就是传说中的"真爱粉"了！

正如知名IT评论人炳叔所说：不公知不五毛不卖萌不传销不
淘宝，搞到6万粉丝，还真是难啊。

Accuracies of Current Models

	Precision	Recall
Following People	0.94	0.63
Re-Tweeting	0.85	0.64
Generating Marks (Thumbup, etc)	0.80	0.50

Accuracies of Information Recommendation
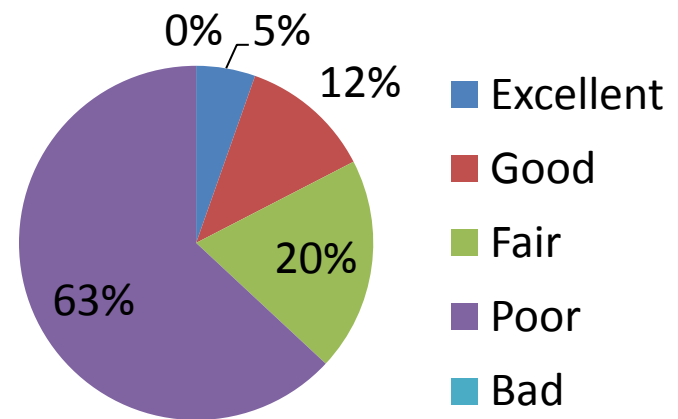
- Human judgment on forwarded tweets
- Relevance of forwarded tweets
- 5: Excellent, 4: Good, 3 Fair, 2 Poor, 1 Bad
- Average score = 3.5

Forwarded Tweets

5%   0%

20%   39%

36%

- Excellent
- Good
- Fair
- Poor
- Bad
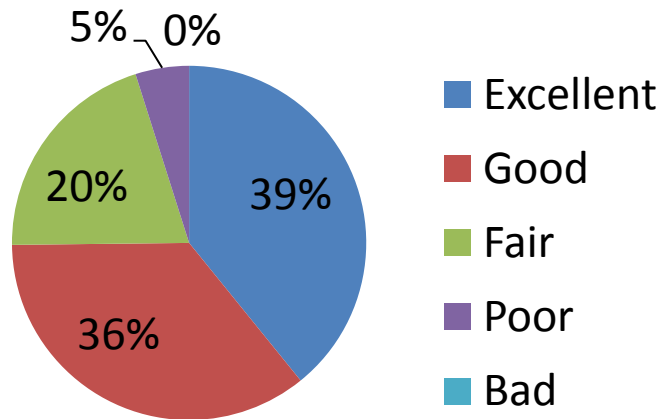
Average 3.49

Skipped Tweets

0% 5%

12%

63%   20%

- Excellent
- Good
- Fair
- Poor
- Bad

Average 2.39
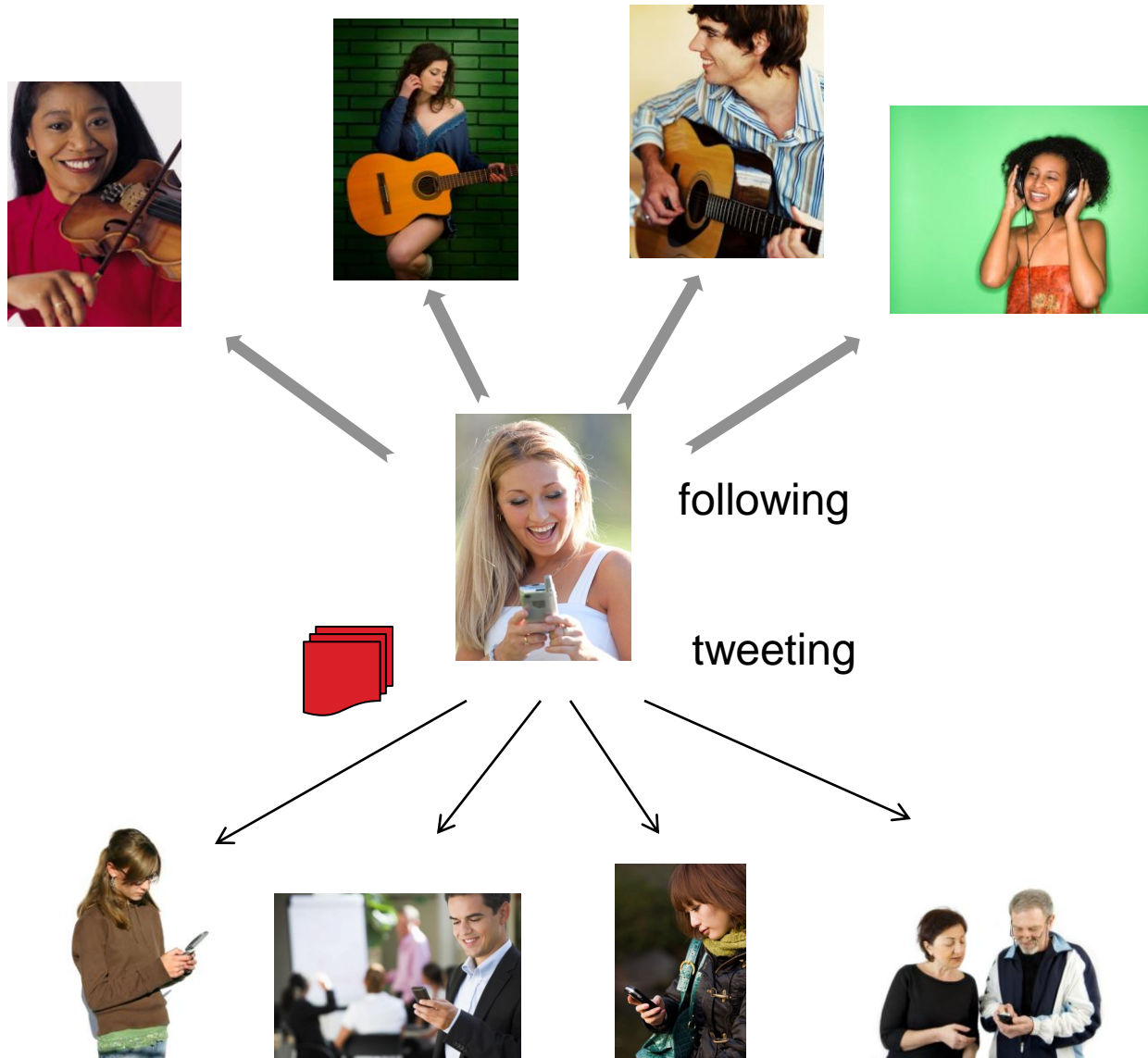
Technical Problems (Current Focus)

- 1. Following People and Re-Tweeting
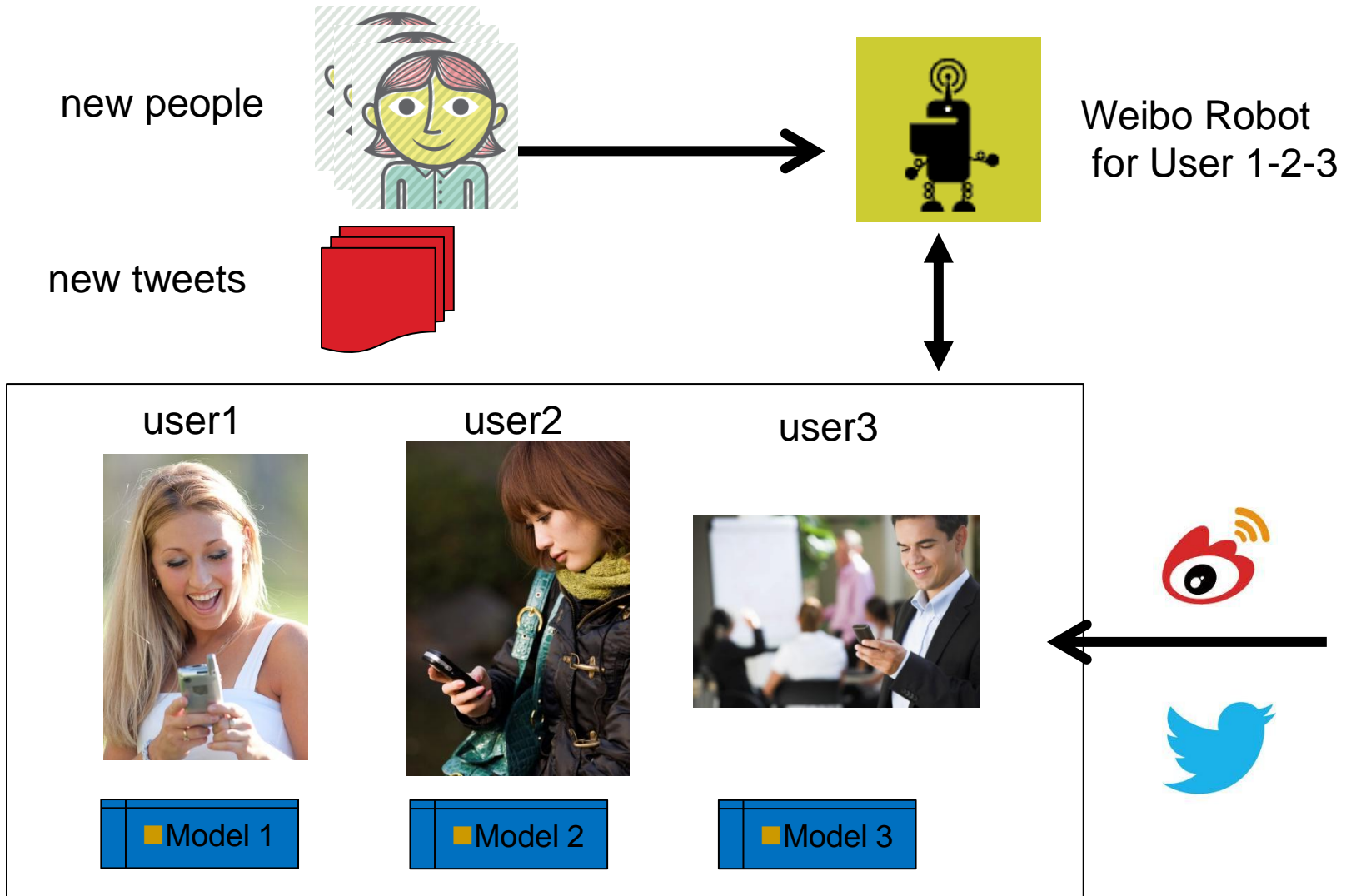- 2. Generating Simple Comments

Our Approach to Following People and Re-Tweeting

- Building User Models from Social Media
- Models Represent Information Needs of Users
- Assumption:
  - Can understand user's needs from her tweets
  - Can understand user's needs from people she follows
- Recommending People and Tweets based on User Models
  - People and tweets as items to recommend
  - Indexing of user models
  - Ranking of people and tweets

Understanding User from Social Media

following

tweeting

Following People and Re-Tweeting = Recommending People and Tweets based on User Models

new people

new tweets

Weibo Robot
for User 1-2-3

user1

user2

user3

Model 1

Model 2

Model 3

Our Approach to Generating Simple Comments

- Formalizing Problem as 'Semantic Matching'
- Matching between Tweets and Comments
- Assumption
  - Similar tweets have similar comments
- Learning Matching Model from Large Number of Tweet-Comment Pairs
- Finding Best Comment Given New Tweet

Similar Tweets Have Similar Comments

Our paper entitled learning to rank has been accepted by ACL .

Congratulations! It is a great achievement

We are lucky. Our paper has been accepted by SIGIR this year. We are going to present it.
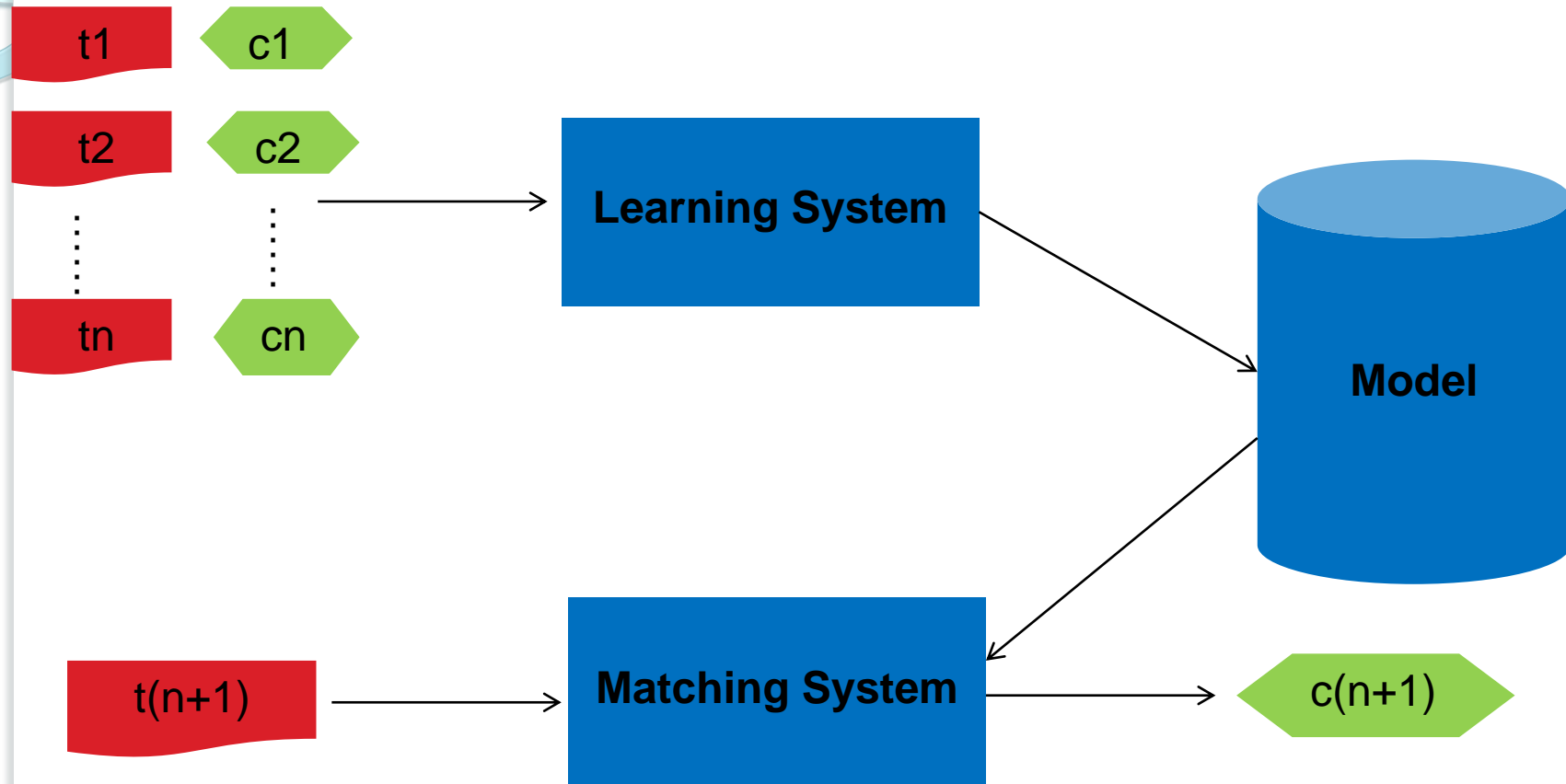
Great news! Please accept my congrats!

The PC of WSDM noticed us that our paper has been accepted.

Awesome! It is a great achievement

⋮

⋮

Semantic Matching

Summary

- Our Vision on Big Data
  - From Big Data to Deep Knowledge
  - Big Machine Learning, Human Computation, Lifelong Learning
- Social Information Filtering
  - Information Assistant
  - Weibo Robot

Acknowledgement

- Joint work with Yibo Zhang, Wenyang Zang, Huawei Li, Jie Zhang, Zhengdong Lu, Jun Xu, Qiang Yang
- Many constructive inputs are from Qiang Yang and Zhengdong Lu

THANK YOU!

Contact: hangli.hl@huawei.com