



Annotation Prediction in Social Media Mining

Hsin-Hsi Chen (陳信希)

Department of Computer Science and
Information Engineering
National Taiwan University



Social Media vs. Mass Media



- Mass Media
 - industrial media, traditional media, broadcast media
 - newspapers, television, film
- Social Media
 - Internet forums, weblogs, wikis, podcasts
- Characteristics
 - Reach – both media provide scale and enable anyone to reach a global audience
 - Accessibility
 - mass media: controlled by for-profit enterprises
 - social media: available to anyone at little or no cost
 - Usability
 - mass media: requires specialized skills and training
 - social media: anyone can operate the means of production
 - Recency
 - mass media: the time lag is relatively long
 - social media: virtually instantaneous responses

Applications of Social Media

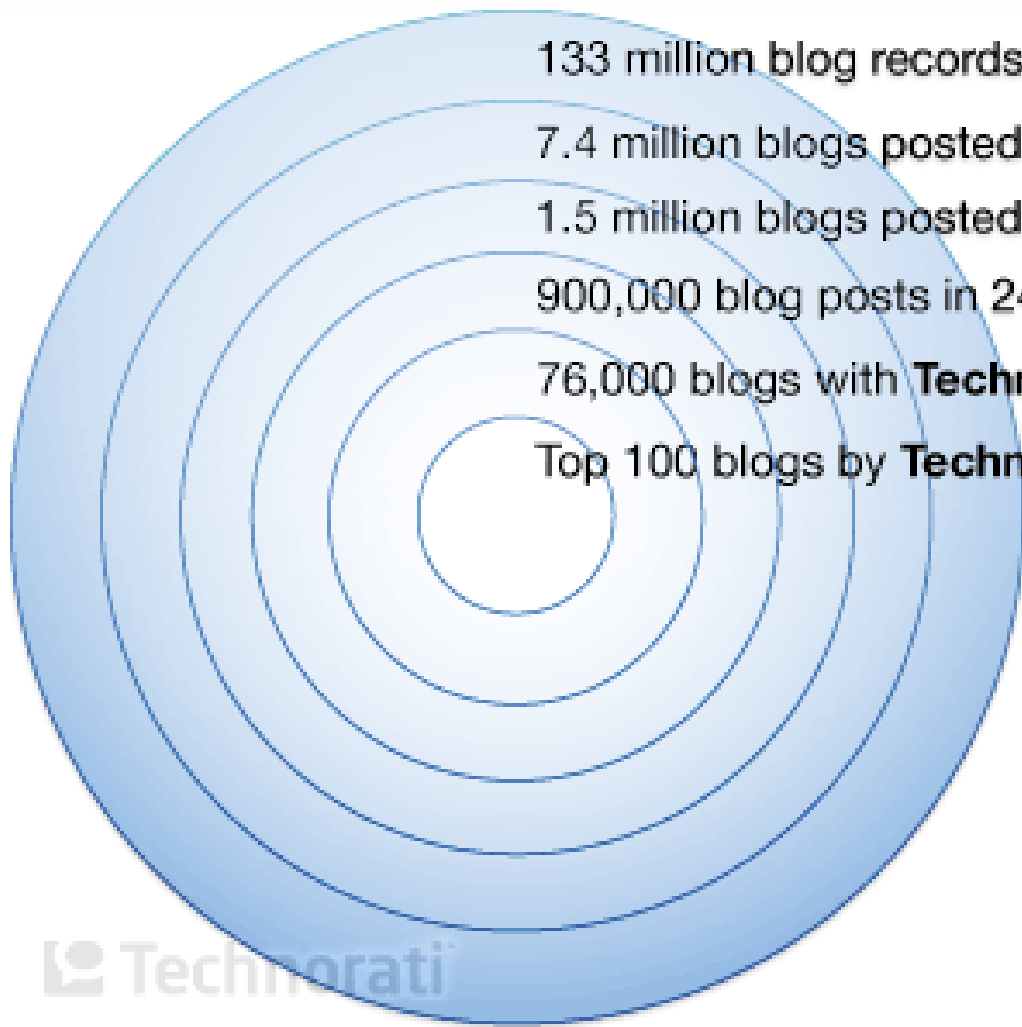


- Communication
 - Blogs (blogger, livejournal), ...
- Collaboration
 - Wikis (Wikipedia), Social bookmarking (Delicious), ...
- Multimedia
 - Photo sharing (Flickr), Video sharing (YouTube), ...
- Entertainment
 - Online gaming (World of Warcraft, 魔獸世界), ...

State of the Blogosphere / 2008



- Blogs are Pervasive and Part of Our Daily Lives
 - comScore MediaMetrix (August 2008)
 - Blogs: 77.7 million unique visitors in the US
 - Facebook: 41.0 million
 - MySpace: 75.1 million
 - Total internet audience 188.9 million
- eMarketer (May 2008)
 - 94.1 million US blog readers in 2007 (50% of Internet users)
 - 22.6 million US bloggers in 2007 (12%)
- Universal McCann (March 2008)
 - 184 million WW have started a blog | 26.4 US
 - 346 million WW read blogs | 60.3 US
 - 77% of active Internet users read blogs



133 million blog records indexed by Technorati since 2002

7.4 million blogs posted in last 120 days

1.5 million blogs posted in last 7 days

900,000 blog posts in 24 hours

76,000 blogs with **Technorati Authority** of 50+

Top 100 blogs by **Technorati Authority**

 Technorati

Technorati Authority is the number of blogs linking to a website in the last six months. The higher the number, the more Authority the blog has.

(<http://technorati.com/blogging/state-of-the-blogsphere/>)



From Taxonomy To Folksonomy

Collaborative Tagging:

Metadata of a resource is generated by both the creator and its readers collaboratively.

del.icio.us



del.icio.us

url	<input type="text" value="http://www.foodsubs.com/"/>
description	<input type="text" value="The Cook's Thesaurus"/>
notes	<input type="text" value="pictures of ingredients and substitutions"/>
tags	<input type="text" value="food reference"/>

recommended tags

export **food** guide search

your network

for:joshua for:jwhiting

popular tags

cooking **food** **reference** recipes thesaurus Dictionary cook



puppets ha noi

- [vietnam](#)
- [hanoi](#)
- [Ha](#)
- [Noi](#)
- [puppet](#)
- [play](#)
- [show](#)
- [water](#)
- [puppetry](#)
- [doll](#)



You See Berkeley



[Try out the NEW \(beta\) version of this page!](#)

Berkeley
University of California

Campus Life

Added: **September 06, 2007**

From: [ucberkeleycampuslife](#)

Subscribe

to ucberkeleycampuslife



Provided By:

[ucberkeleycampuslife](#)

DIRECTOR

You See Berkeley

Category [Sports](#)

Tags: [berkeley](#) [cal](#) [football](#) [you](#) [\(more\)](#)

URL

Embed

[Related](#)

[More from this user](#)

[Playlists](#)

回應：不負責任影評：《色·戒》

最精彩在於不色,不戒

更愛說到骨髓那句

男人因色生情而愛惜女人，

女人因戒惜情而愛上男人。

但不同意摸肚子假設懷孕的那一段 她想起腹中的孩子.....

真的評的太棒了

2007-10-06 22:56 | shihyunyin

Golden Lion at the Venice Film Festival



真的很棒

金吉利，你的影評真是很棒，我很多朋友都有情婦，我觀察他們之間的互動，其實跟電影裡面講的都完全一樣，所以這個世界沒什麼太大改變，我是一個生意人，平常很忙，沒時間也沒興趣去看電影，但是昨天我多年來第一次進電影院，電影真的很悶，張愛玲的小說也很悶，我很喜歡，因為很共鳴，男人在追求金錢與權力的道路上，非常辛苦，但是那是一般價值觀的社會肯定他的一種方式，你的影評對李安的用心很深刻，很精采，我有時間時，會再上來看您其他的作品，完全給你肯定啦！

2007-10-06 14:48 | 小祿

回應：不負責任影評：《色·戒》

我不喜歡這部電影,劇裏的角色沒有一個讓我感動或是讓我受到衝擊。我感受不到易先生該殺。(排除漢奸就該死的主觀前題)我感受不到女主角內心的衝擊,全片的主題也模模糊糊的。我很失望。

2007-09-30 13:56 | joe

回應：不負責任影評：《色·戒》

看電影多年,第一次,戲演完了,有人拍手

看電影多年,第一次,細看完了,過了那麼多天還在想劇情

至於其它的話就不需說了

只想給參予戲所有人說讚啦

2007-10-01 18:12 | JJ

回應：不負責任影評：《色·戒》

>>>>這種電影題材只會鼓勵更多壞男人出現！

這是什麼呀???

不如建議香港那些黑社會古惑仔電影導演全都殺頭好了@@
拍外過的強姦的更是都要宮刑...

講一部電影竟是這種說法，實在很污辱人

2007-09-29 20:50 | MAY



Tom so crazy but I'm so happy

分類：未分類資料夾

今天跟你約吃飯 不知為什麼特別緊張 😬

也許因為一陣子沒見吧

謝謝你請我吃飯 還送我禮物 📺

雖然一直叫你不要送我東西

但收到的時候還是很開心 😊

當打開禮物的時候 整個傻眼

居然送我 iPod 太誇張了

Tom so crazy but I'm so happy 😄

Yahoo! Taiwan News



Yahoo!奇摩首頁 | 信箱 | 拍賣

服務說明



新聞

會員登入
新使用者? 立即註冊

新聞 網頁

新聞首頁 | 政治 | 社會 | 地方 | 國際 | 財經 | 科技 | 運動 | 健康 | 教育 | 藝文 | 影劇 | 旅遊 | 生活

資訊3C | 科學發展 | 自然環境 | 照片故事 | 專輯 | 民調中心 | 軟體與環保

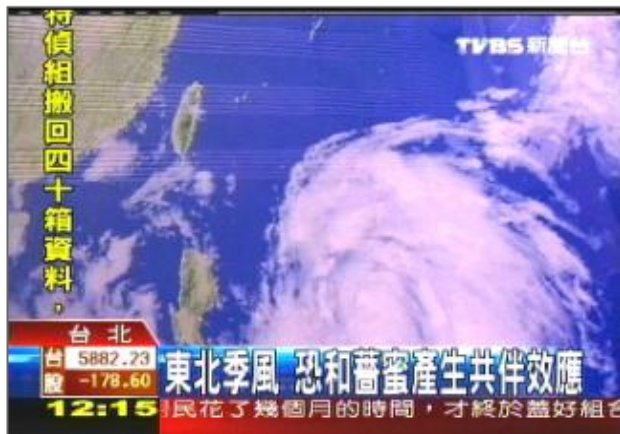
2008/10/23(星期四) 農曆(戊子)九月廿五日

新聞首頁 > 科技 > 科學發展 > TVBS

✉ 寄給朋友 | 友善列印 | 字級設定:

蕃蜜秋颱／秋颱蕃蜜 最快今海警、明晨陸警

TVBS 更新日期: 2008/09/26 12:20 雷凱雯



9月想趁著週末出門走走，卻幾乎每個禮拜都遇上颱風，就連最後這周末，也突然碰上個愛插花的中颱蕃蜜；氣象局表

廣告

精采多樣的生活
盡在HP光潮

YAHOO! 生活+

美食大推薦

- 香酥多汁脆皮雞排
- 300元以下吃到飽

超夯





休假日注定泡湯，但眼看災區重建工作還沒完成，薈蜜來勢洶洶，雨量恐怕不小的週末，真的得趕緊加快防颱準備工作。

這篇新聞讓你覺得？



新奇



溫馨



誇張



難過



實用



高興



無聊



生氣



送出

135人

推薦此新聞

推薦

- 若您願意推薦這則新聞，請點選左側的「推薦」按鈕，但推薦新聞需要收集更多網友意見，目前系統尚無法提供相關新聞。

儲存至「分享書籤」 | 傳訊給朋友

更多：科學發展 < 科技 < 新聞首頁

科學發展最多瀏覽

- 頭髮培養幹細胞 重症新希望 (自由)



- 戀愛新配方 <MiniTour> 兩人成行
- 哈日自由行3800 > 獨家送玩樂情懷誌
- 奧捷音樂會買1送1 > 再抽Gucci精品



- 歡樂金門趣 > 體驗台海戰地全新風情
- 可樂全球線上訂房獨享現金2%回饋
- 金廈小三通 <機票船票接駁> 3710起



休假日注定泡湯，但眼看災區重建工作還沒完成，薈蜜來勢洶洶，雨量恐怕不小的週末，真的得趕緊加快防颱準備工作。

這篇新聞讓你覺得？



新奇



1



溫馨



2



誇張



3



難過



4



實用



5



高興



6



無聊



7



生氣



8

送出

135人

推薦此新聞

推薦

• 若您願意推薦這則新聞，請點選左側的「推薦」按鈕，但推薦新聞需要收集更多網友意見，目前系統尚無法提供相關新聞。



- 戀愛新配方 <MiniTour> 兩人成行
- 哈日自由行3800 > 獨家送玩樂情懷誌
- 奧捷音樂會買1送1 > 再抽Gucci精品



- 歡樂金門趣 > 體驗台海戰地全新風情
- 可樂全球線上訂房獨享現金2%回饋
- 金廈小三通 <機票船票接駁> 3710起

1. Awesome
2. Heartwarming
3. Surprising
4. Sad
5. Useful
6. Happy
7. Boring
8. Angry

↓ 儲存至「分享書籤」 | 傳訊給朋友

更多：科學發展 < 科技 < 新聞首頁

科學發展最多瀏覽

- 頭髮培養幹細胞 重症新希望 (自由)

Yahoo! Taiwan News



休假日注定泡湯，但眼看災區重建工作還沒完成，薈蜜來勢洶洶，雨量恐怕不小的週末，真的得趕緊加快防颱準備工作。



共有 4214 人投票

[» 瀏覽更多態度投票結果](#)

135人
推薦此新聞

推薦

- 若您願意推薦這則新聞，請點選左側的「推薦」按鈕，但推薦新聞需要收集更多網友意見，目前系統尚無法提供相關新聞。



- 戀愛新配方 <MiniTour> 兩人成行
- 哈日自由行 3600 > 獨家送玩樂情報誌
- 奧捷音樂會買1送1 > 再抽Gucci精品



- 歡樂金門趣 > 體驗台海戰地全新風情
- 可樂全球線上訂房獨享現金2%回饋
- 金廈小三通 <機票船票接駁> 3710起

Social Annotation



- Multimedia
 - Text (news, forum, blog, etc)
 - Image
 - Video
- Explicit Tagging vs. Implicit Tagging
- Search, Recommendation, and so on



Social Annotation Enable Searching

From Search Point of View



- Content of resources (by indexing)
- Structures of resources (by link analysis)
- Linguistic resources (by concept expansion)
- User log (by concept recommendation)
- Social Annotation (by ???)

Motivation



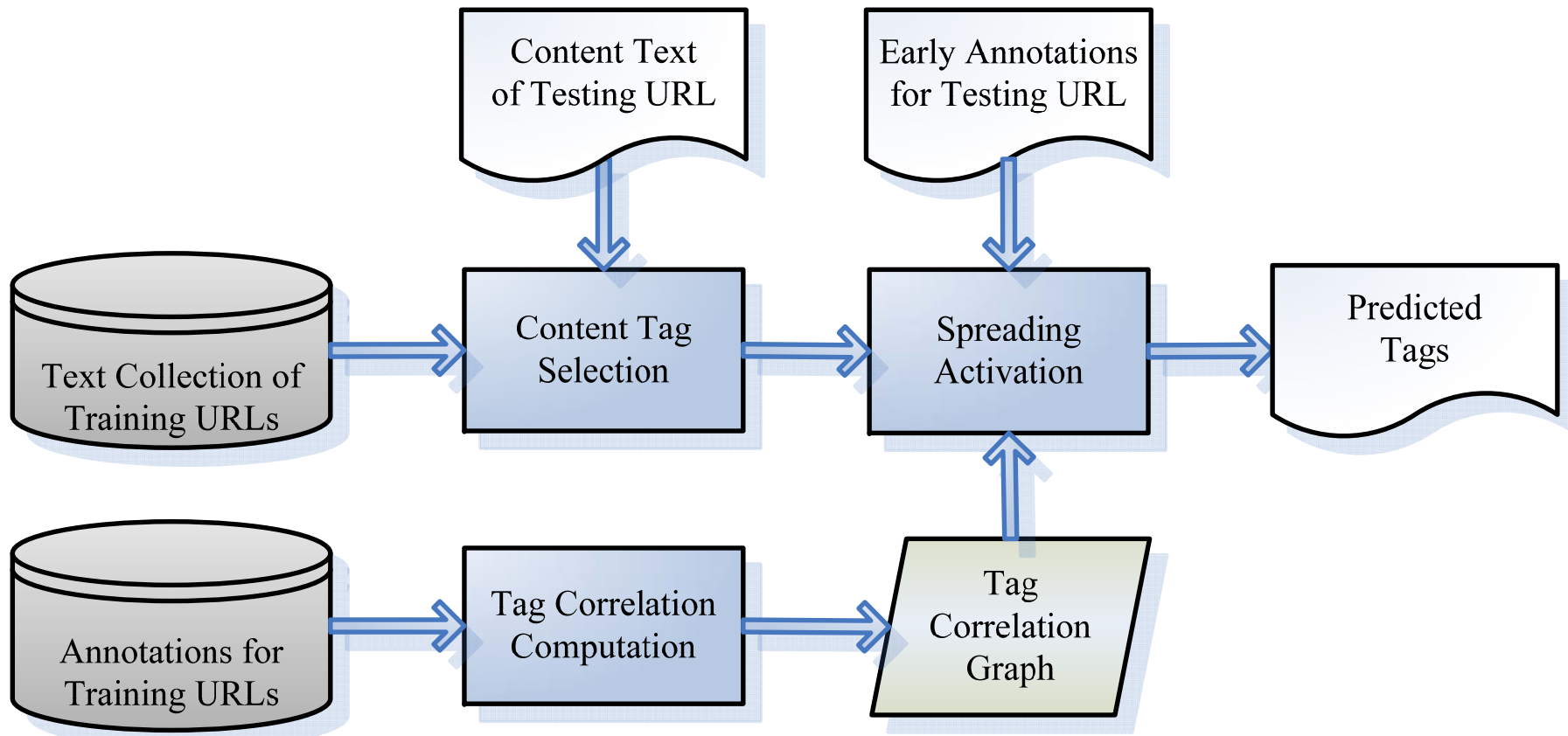
- Annotation may facilitate recommendation and effective retrieval.
- Not all resources can gain the benefits from that.
 - Ill-tagged period of URLs prevents them from being retrieved.
 - The retrieval performance for new-coming URLs degrades inevitably.
- How to predict a quality tagging set for a resource is an important issue.



Predicting Social Annotation

Indexing + Annotation Histories =
Prediction

Framework of Tag Prediction Algorithm



Tag Prediction Model (1/5)



- **Content Tag Selection**

- Follow the statistical translation model in information retrieval
- Estimate the probability that a query would be generated as a translation of a document
- Rank terms in a URL as candidate tags

$$\text{CTScr}(t_i | d) = \sum_{term_j \in d} \log\left(N \times \max(\text{P}_C(t_i | term_j), \frac{1}{N})\right)$$

- N : the number of documents in the training data
- $\text{P}_C(t_i | term_j)$: the probability of t_i as a *stabilized* tag when t_i and $term_j$ co-occur in the same document.
- D_j : the set of documents where $term_j$ occurs
- $D_{j,i}$: the set of documents in D_j with t_i as a stabilized tag.
- $\text{P}_C(t_i | term_j) = |D_{j,i}| / |D_j|$.
- All terms in the document are ranked according to their scores.

Tag Prediction Model (2/5)



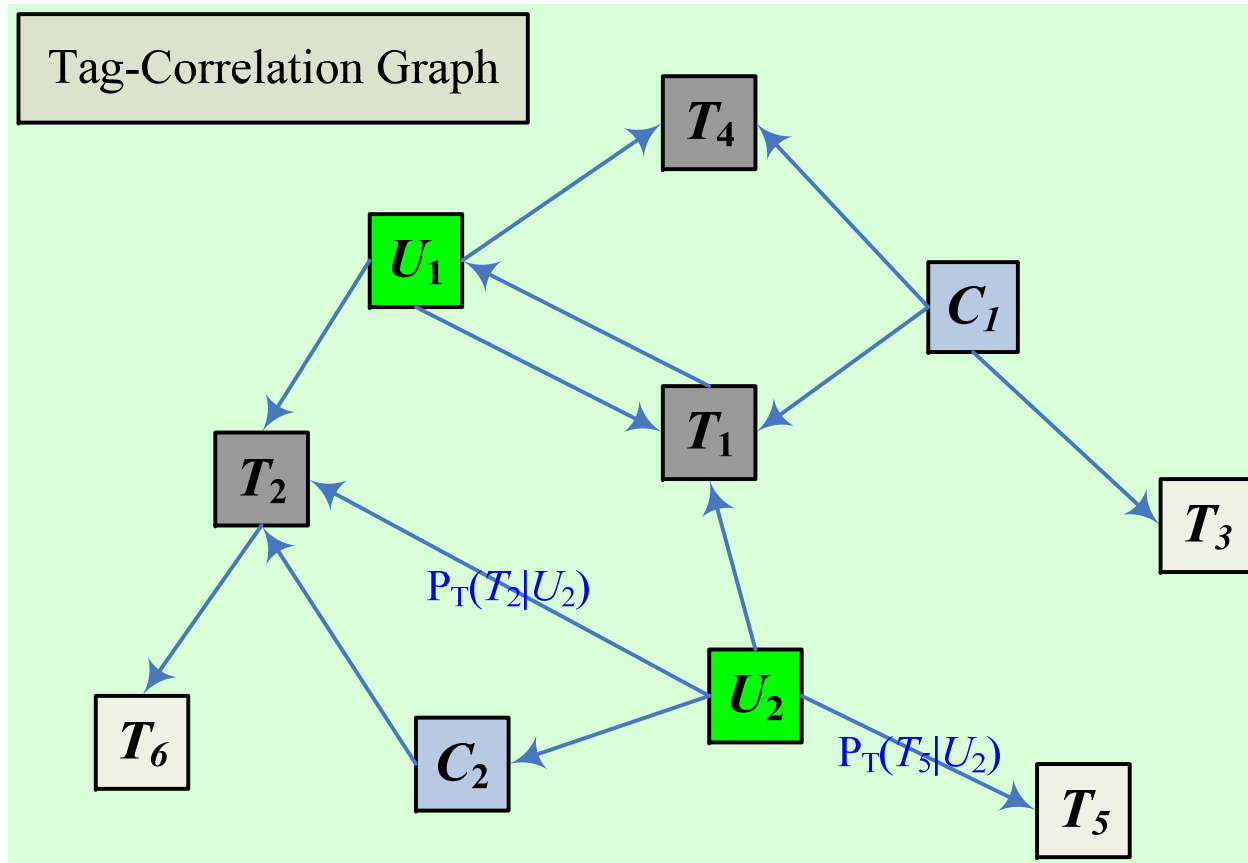
- Tag-correlation graph
 - node: tag
 - edge between two nodes: their correlation
- Correlation: the strength of the edge from t_j to t_i
 - asymmetric

$$P_T(t_i / t_j) = \frac{\text{\# of stabilized tag sets containing } t_i \text{ and } t_j}{\text{\# of stabilized tag sets containing } t_j}$$

Tag Prediction Model (3/5)



- Spreading Activation



Tag Prediction Model (4/5)



- Spreading Activation

- At each iteration, node t_j propagates a portion $(1 - \lambda)$ of its energy to its neighbors, and gains some energy from its neighbors.

$$E(t_i) = \lambda \times E(t_i) + (1 - \lambda) \times \sum_{t_j \in \text{Neighbor}(t_i)} (E(t_j) \times W(t_j, t_i))$$

$$W(t_j, t_i) = \frac{P_T(t_i / t_j)}{\sum_{t_k \in \text{Neighbor}(t_j)} P_T(t_k / t_j)}$$

- Spreading activation is performed for a fixed number of iterations and eventually tags of the highest energies are proposed as the predicted stabilized tags.

Tag Prediction Model (5/5)



- Spreading Activation

- A tag is called an *activation origin* if it is a selected candidate content tag or if it is an early user-annotated tag.
- We initialize the energy for each tag to 0 except those activation origins.
- For each candidate content tag t_i selected from the document d , its initial energy $E(t_i)$ is its content selection score $\text{CTScr}(t_i|d)$ normalized by the highest content selection score of terms co-occurring with t_i in d .
- For each of the early user-annotated tags, the initial energy is its total tagging by early users.
- For a content tag also annotated by early users, its initial energy is the sum of these two values.

Experimental Data



Dataset	# URLs	Comments
Training	59,090	<ul style="list-style-type: none">• 13,934 sufficiently-annotated• 45,156 annotated by 20+ users
Testing	2,000	<ul style="list-style-type: none">• sufficiently-annotated• top 25 stabilized tags

Experimental Results



Method	Metric	R-Precision	Recall	MAP
CTScr		0.2687	0.3569	0.1768
User-5		0.2359	0.2343	0.2077
CTScr + User-5		0.3972	0.4733	0.3439
CTScr+User-5+SA ($\lambda = 0.5$, iteration=2)		0.4493	0.6072	0.3879
User-10		0.3445	0.3436	0.3025
CTScr + User-10		0.4591	0.5360	0.4154

Tag Categorization



- Categorize the tags into three types according to the relations between tags and target URL.
 - Topic-description Tags
 - stabilized tags describe the implicit topic of the target URL
 - highly correlated to the terms in the URL content, and correlated to some other tags of the same type
 - tags such as ‘*programming*’ and ‘*java*’ for a URL introducing java programming
 - Function-related Tags
 - describe or conceptually related to the function of the target URL
 - some of them are highly correlated to other tags but most of them usually have no evident correlations with terms in the text content
 - “web 2.0” for a URL provides service to *rank* professors in universities
 - Personal-use Tags
 - occur in the stabilized tag set *arbitrarily*
 - E.g., *read* and *reference*

Tag Normalization



Canonical Form	Group of Tag Strings
common lisp	Commonlisp, clisp, common_lisp
drag and drop	drag-and-drop, draganddrop, dragndrop
online game	Onlinegames, online_game, olg
web 2.0	web2.0, web_2.0, web2, web20, web_20

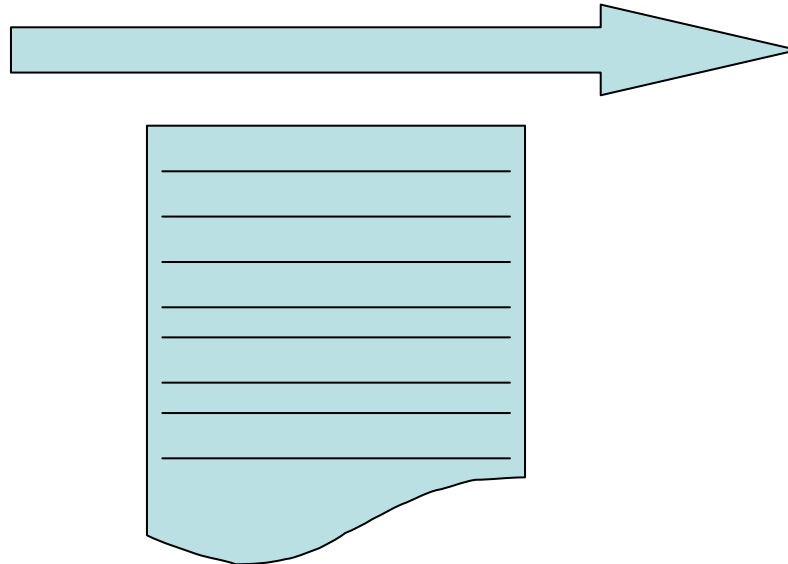


Emotion Tagging

Emotion Analysis



- Writer Emotion vs. Reader Emotion





Writer Emotion Analysis

Writer Emotion Analysis



- Bloggers express their strong feeling with a special set of emoticons
- Weblog as Corpora
 - Text provide the content
 - Emoticon is the gold standard

Positive Emotion			Negative Emotion		
Tagged Icon	Description	Sub Category	Tagged Icon	Description	Sub Category
	laughing	HAPPY		crying	SAD
	big grin			sad	
	happy				
	rose	JOY		angry	ANGRY
	blushing			phbbbbt	
	love struck				

Tom so crazy but I'm so happy

分類：[未分類資料夾](#)

今天跟你約吃飯 不知為什麼特別緊張 😬

也許因為一陣子沒見吧

謝謝你請我吃飯 還送我禮物 😄

雖然一直叫你不要送我東西

但收到的時候還是很開心 😊

當打開禮物的時候 整個傻眼

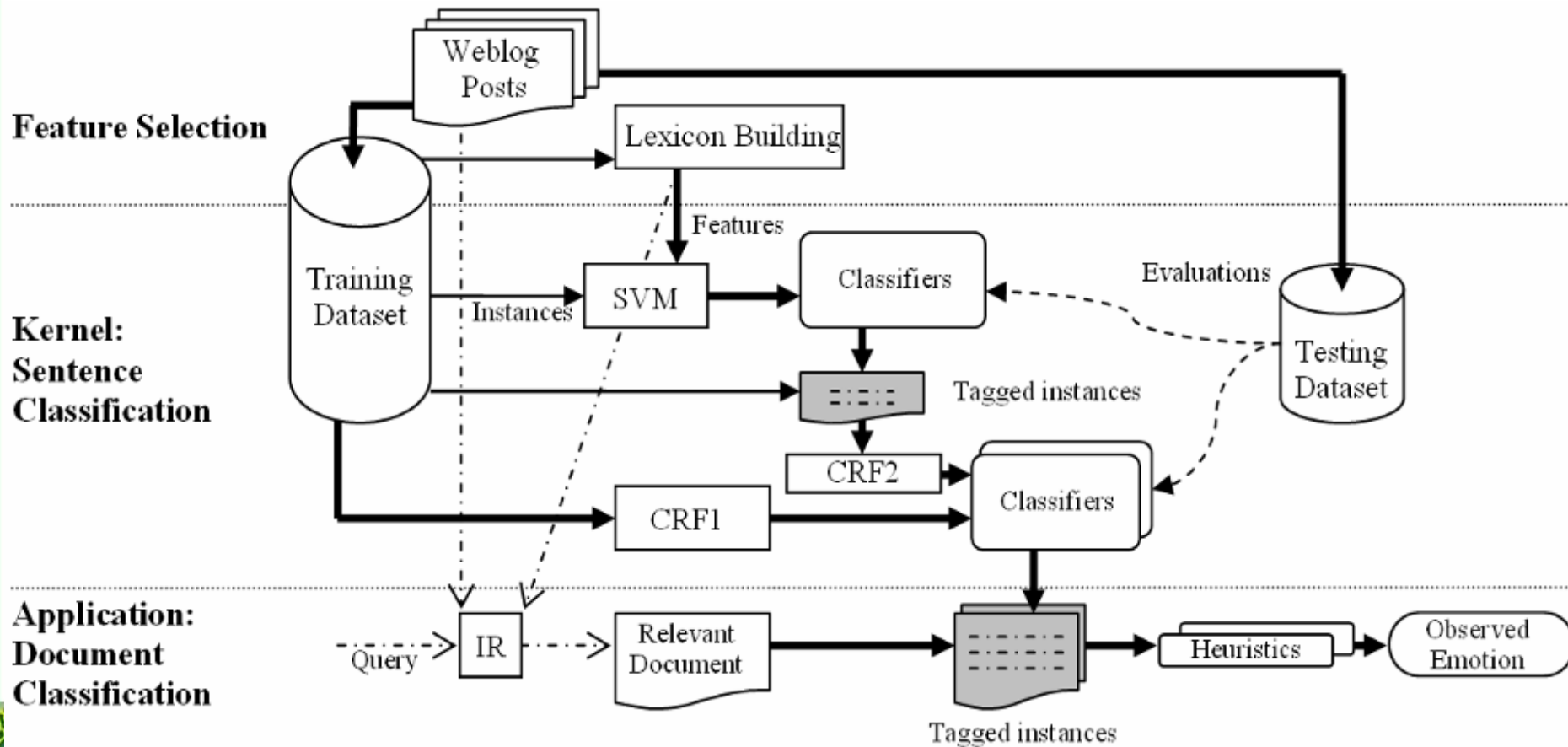
居然送我 iPod 太誇張了

Tom so crazy but I'm so happy 😄

System Framework



- Word Level, Sentence Level, and Document Level



Emotion Lexicon



Emotion	Description	Keywords
joy (喜)	polarity : positive energy : high	愛、幸福、可愛、喜歡、 謝謝、害羞、感動、寶貝
angry (怒)	polarity : negative energy : high	生氣、討厭、氣死、可惡、 幹、煩、罵、哼
sad (哀)	polarity : positive energy : low	哭、痛、鳴、難過、 淚、傷、慘、可憐
happy (樂)	polarity : negative energy : low	哈哈、開心、好笑、高興、 不錯、加油、很好、好玩

Sentence Level



- Blogger's Emotion Detection

今天的感動

分類：未分類資料夾

哇~~~我的生日快到了

今天丫...老闆居然偷偷塞了個生日禮物給我

哇...好感動...感動到不知如何說謝謝ㄟ

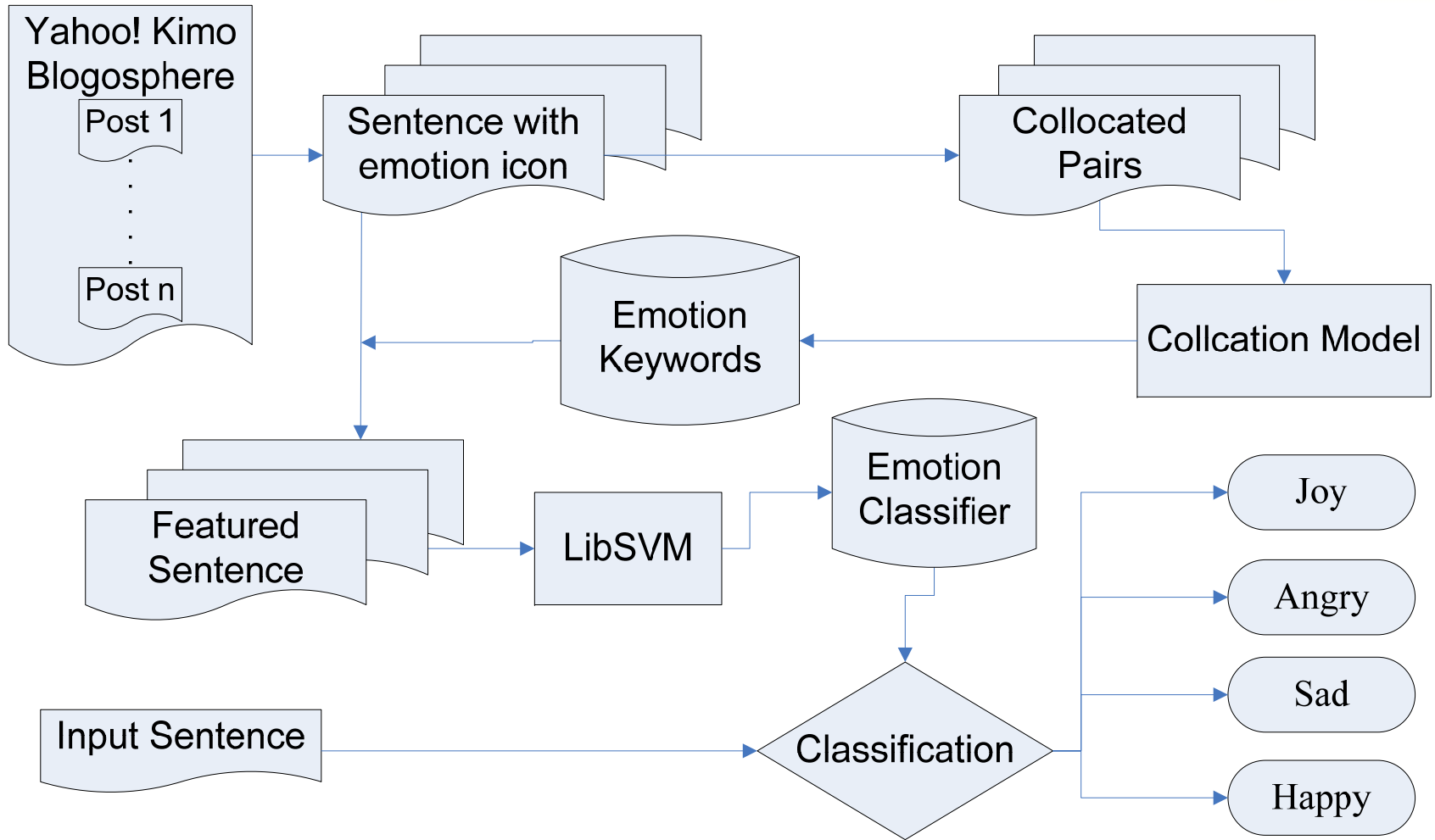
好笑的是..他交代:不能跟老娘(老闆娘)說

說了地就知道我偷藏私房錢..... 撲.....哈哈

謝謝你喔...my boss

- | | |
|-----|----------------------------|
| 1 | 哇~~~我的生日快到了 |
| 2 😊 | 今天丫...老闆居然偷偷塞了個生日禮物給我 |
| 3 🍷 | 哇...好感動...感動到不知如何說謝謝ㄟ |
| 4 😊 | 好笑的是..他交代:不能跟老娘(老闆娘)說 |
| 5 😊 | 說了地就知道我偷藏私房錢..... 撲.....哈哈 |
| 6 🍷 | 謝謝你喔...my boss |

Emotion Classification at Sentence Level



Emotion Classification with Blog Corpus

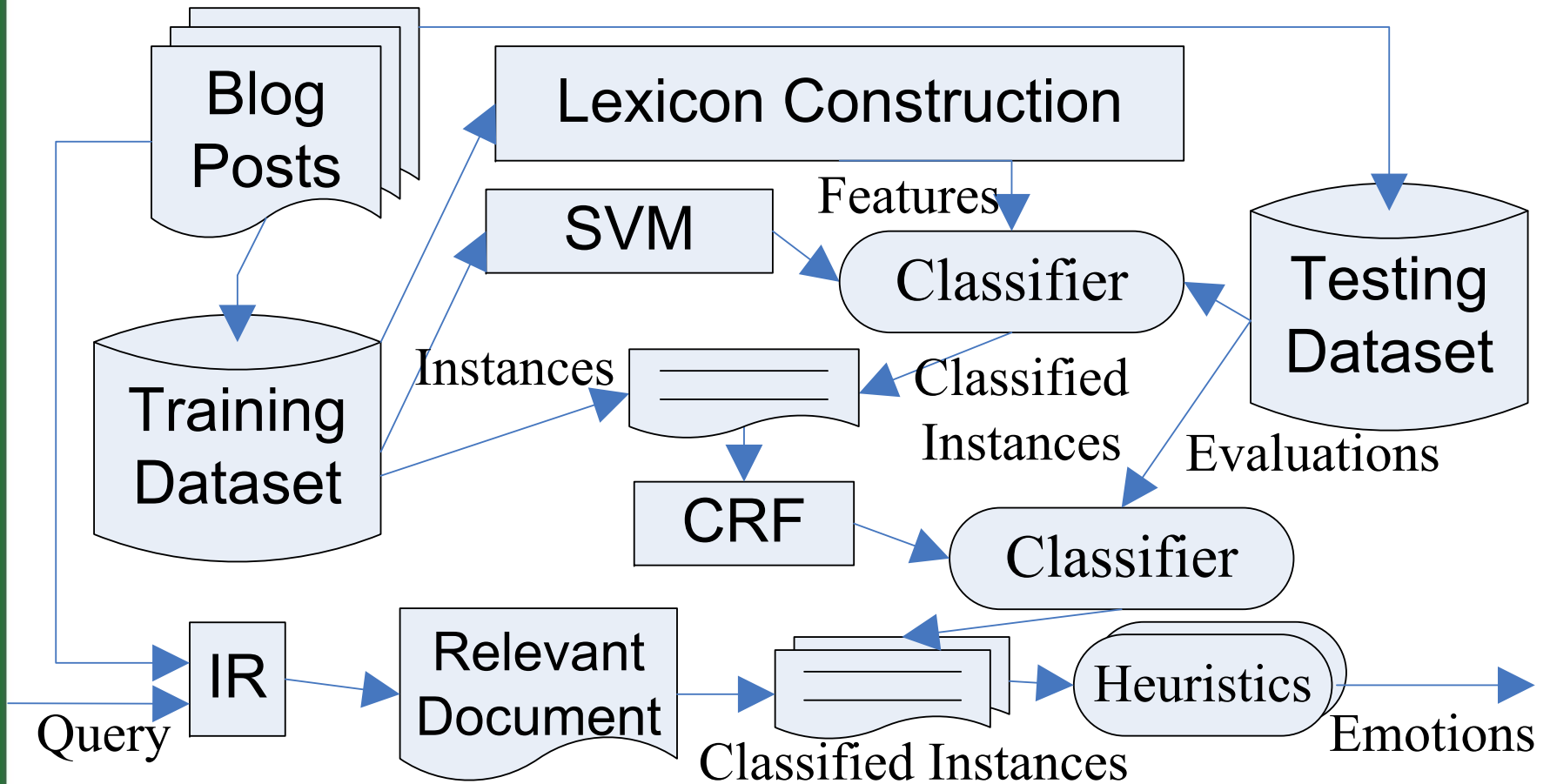
- Sentence X : blah blah blah ... 😄
- Sentence X+1: blah ... blah blah ... 😏
- A function from lexicons to emotion classes

$$FS = \{t_k \mid t_k, 1 \leq k \leq n, t_k \in Lex\}$$

$$S \xrightarrow{\text{extraction}} FS \xrightarrow{\text{classification}} \hat{e} \in \{e_1, \dots, e_n\}$$

- Machine Learning Methods (SVM, CRF)
 - Features: Keywords in blog articles
 - Ground-truth: Emotions of the blogger

System Framework

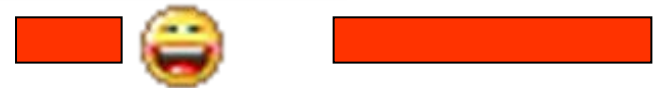


Dataset



Dataset (2006)	# of posts	# of tagged posts	Percentage	Avg. length of tagged	Avg. length of untagged
Training	4,176,250	575,009	13.77%	269.77 chrs.	468.14 chrs.
Testing	1,234,683	182,999	14.82%	281.42 chrs.	455.82 chrs.
Total	5,410,933	758,008	14.01%	272.58 chrs.	465.37 chrs.

Yahoo! Kimo Blog



- people use emoticons to replace certain portions of their text contents to make their articles more succinct

Experimental Results



Table 3. (E-1) on Coarser Emotion Categories.

Experiment Setup	Precision	Recall	F-Score
Bayesian 50 features	↑ 78.30%	31.38%	44.80%
Bayesian 100 features	74.66%	42.11%	53.85%
Bayesian 150 features	73.78%	46.72%	57.21%
SVM 50 features	↑ 78.67%	31.21%	↓ 44.69%
SVM 100 features	75.02%	41.67%	53.58%
SVM 150 features	↑ 74.02%	46.78%	↓ 57.33%

#test instances:
31,255

Table 4. (E-2) on Coarser Emotion Categories.

Experiment Setup	Precision	Recall	F-Score
SVM 50 features	↑ 78.22%	15.56%	25.96%
SVM 100 features	74.18%	26.33%	38.86%
SVM 150 features	72.41%	32.24%	44.61%
CRF 50 features	↑ 82.27%	16.37%	↓ 27.31%
CRF 100 features	80.34%	28.51%	↓ 42.09%
CRF 150 features	↑ 79.87%	35.56%	↓ 49.21%

#test instances:
17,887

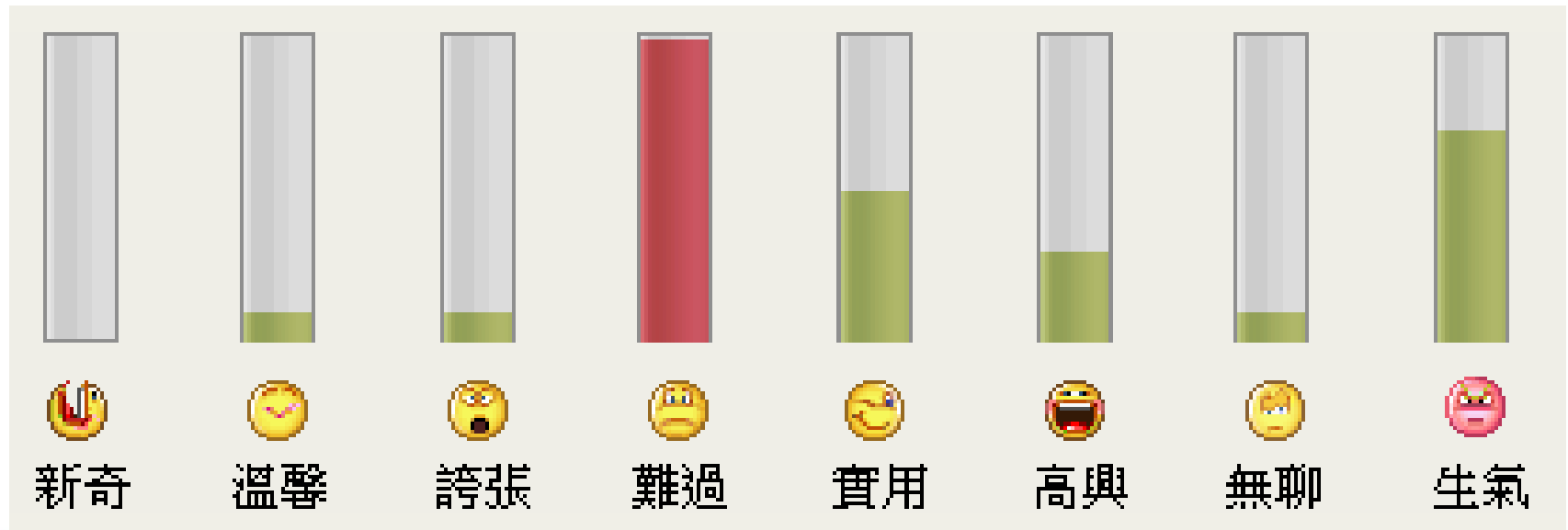


Reader Emotion Analysis

Research Objective



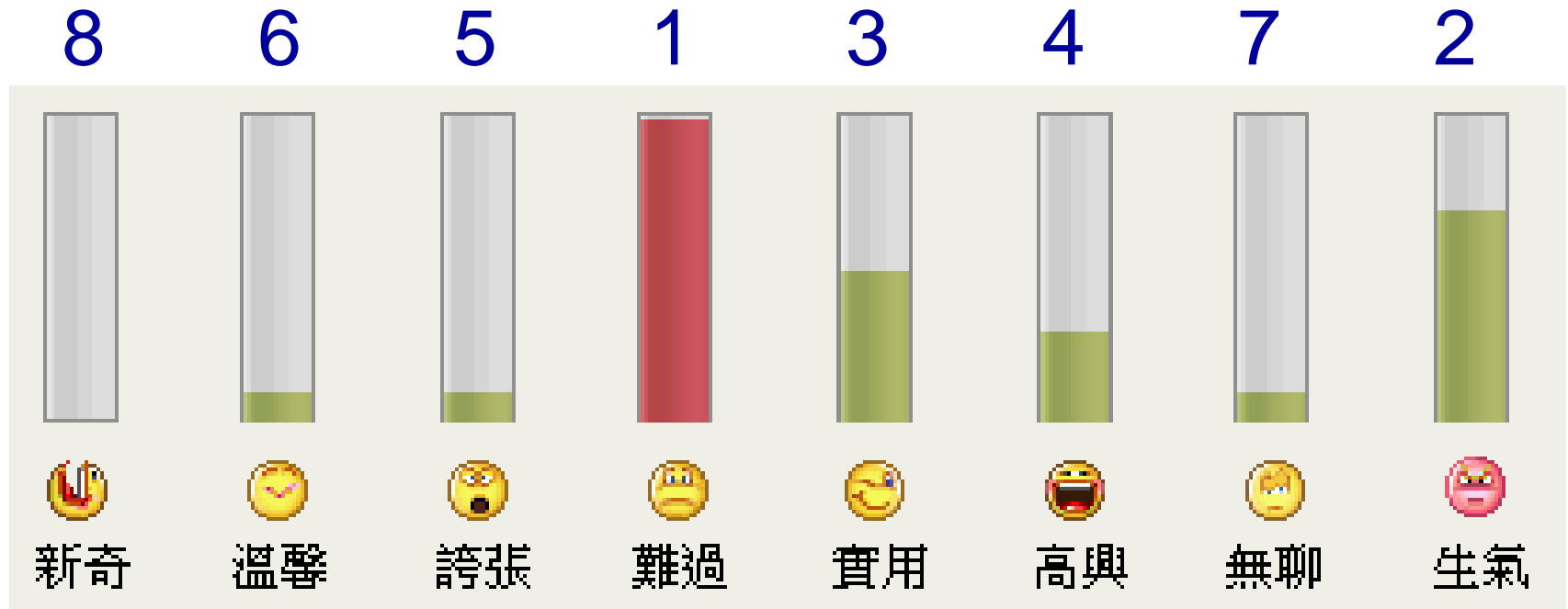
- Predict order of reader emotions



Research Objective



- Predict order of reader emotions



Research Objective



- Predict order of reader emotions

8

6

5

1

3

4

7

2



新奇



溫馨



誇張



難過



實用



高興



無聊



生氣

Corpus



- Yahoo! Taiwan News articles
- January 24 – August 7, 2007
- Training: 25,975 articles
- Test: 11,441 articles

Two Approaches



1. Pairwise Ranking
2. Regression

Pairwise Ranking



- Predict pairwise order of emotions
- Combine into a ranked list

Pairwise Ranking Example

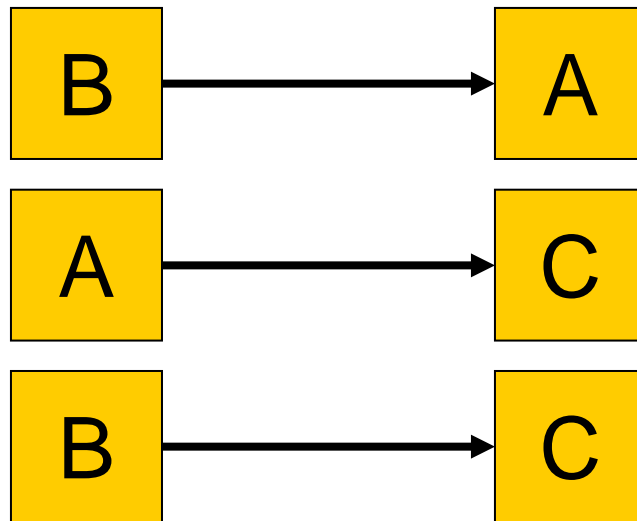


- Ranking three emotions A, B, C

Pairwise Ranking Example



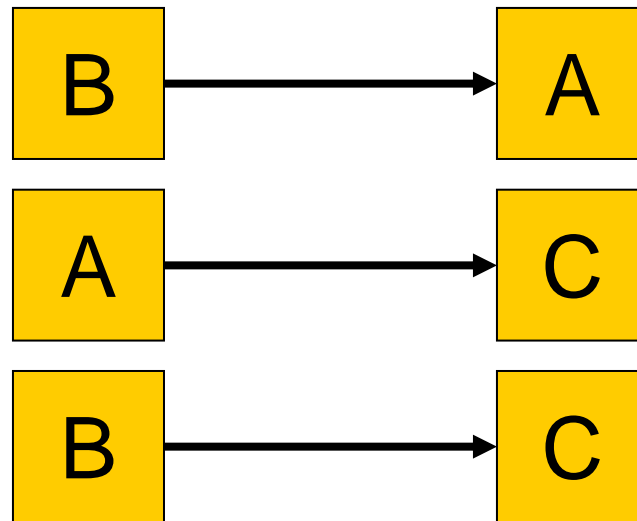
- Ranking three emotions A, B, C
- Prediction:



Pairwise Ranking Example



- Ranking three emotions A, B, C
- Prediction:



- Ranked List: B, A, C

Pairwise Ranking



- Use SVM to predict pairwise order

Pairwise Ranking



- Use SVM to predict pairwise order
- ➔ RankingSVM in information retrieval

Regression



- Use regression to predict voting percentage
- Rank emotions by voting percentage

Regression



- Use regression to predict voting percentage
- Rank emotions by voting percentage
- Prediction Result:

0% 2% 4% 37% 19% 10% 1% 27%



新奇



溫馨



誇張



難過



實用



高興



無聊



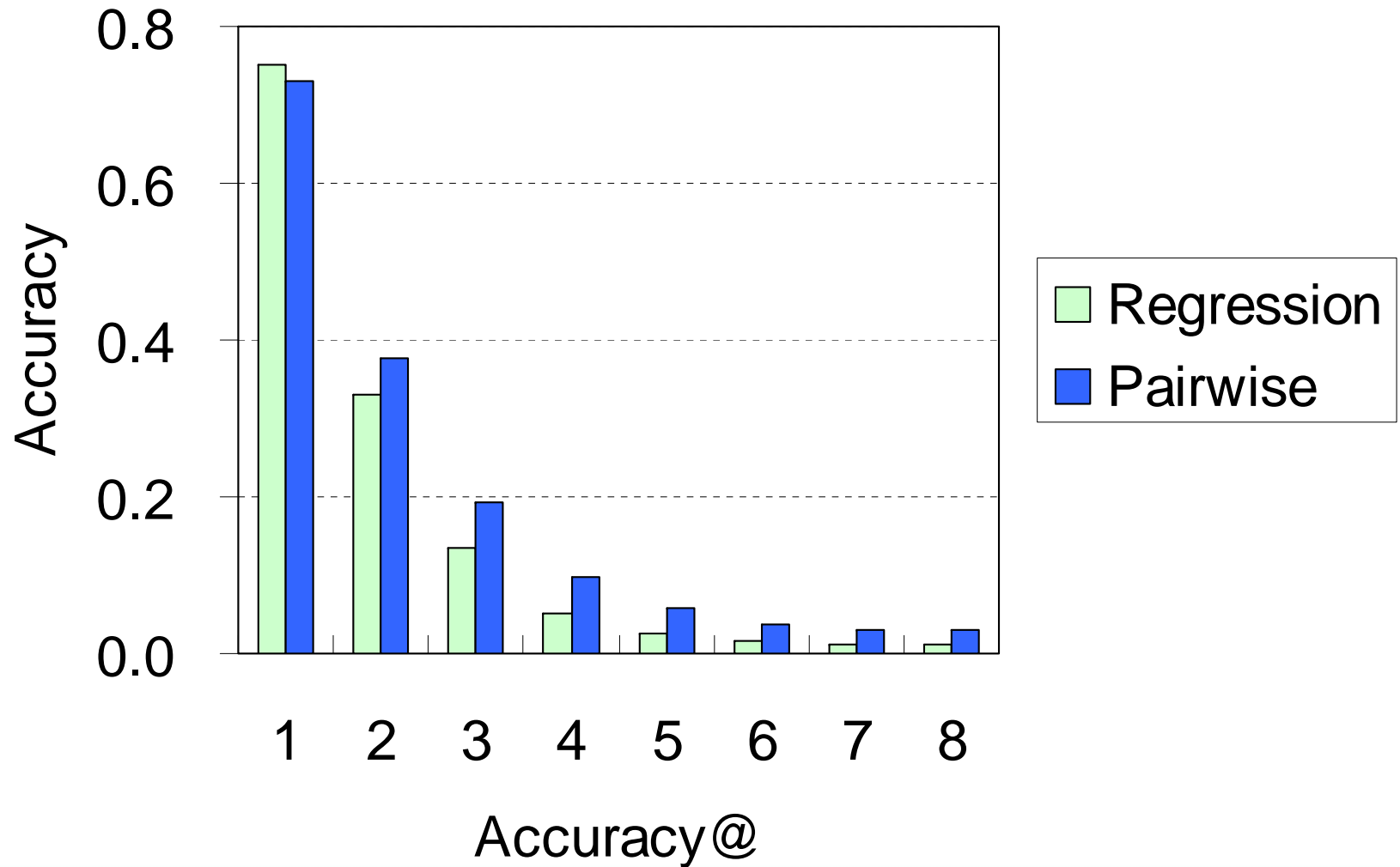
生氣

Regression

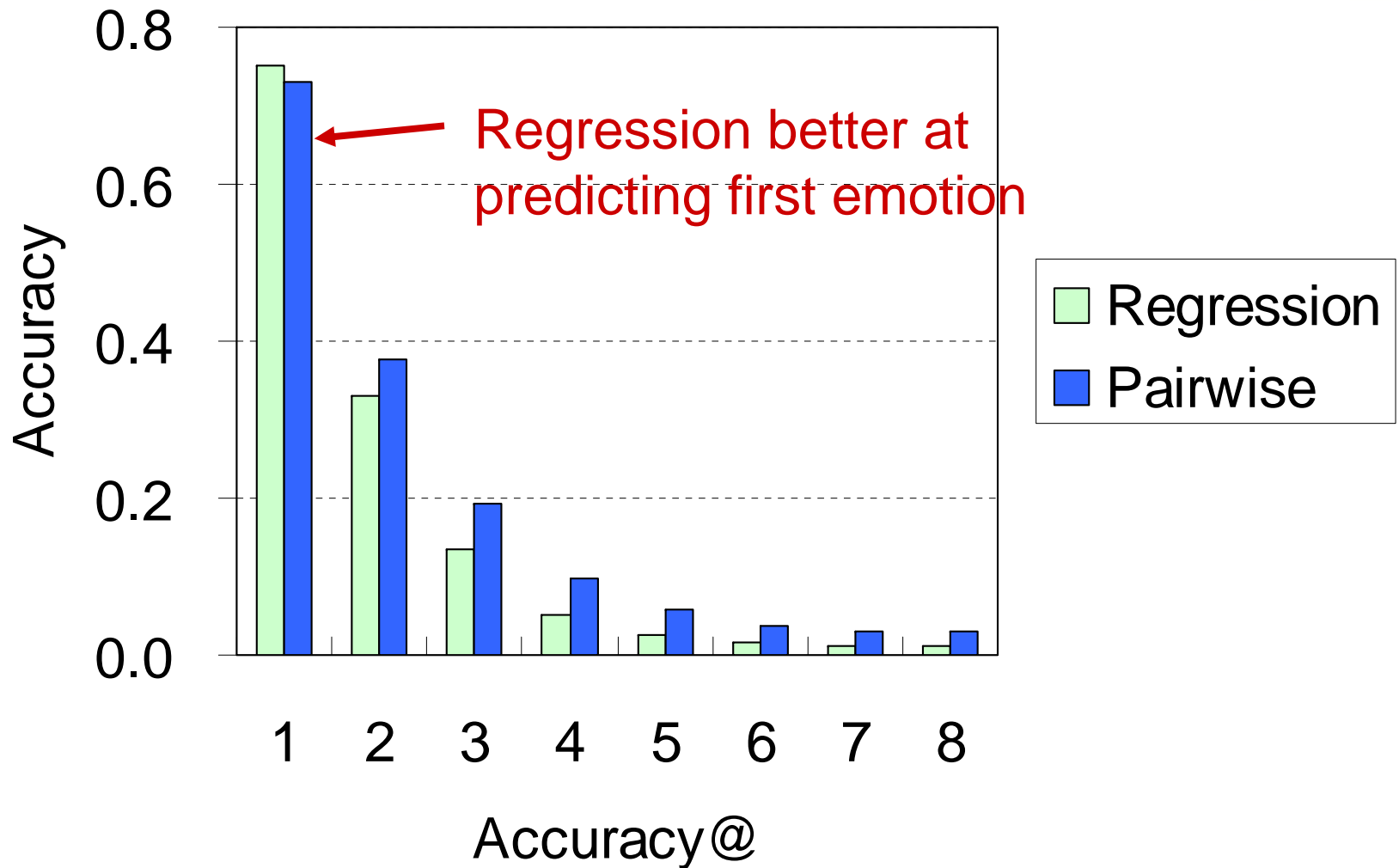


- Use Support Vector Regression (SVR)

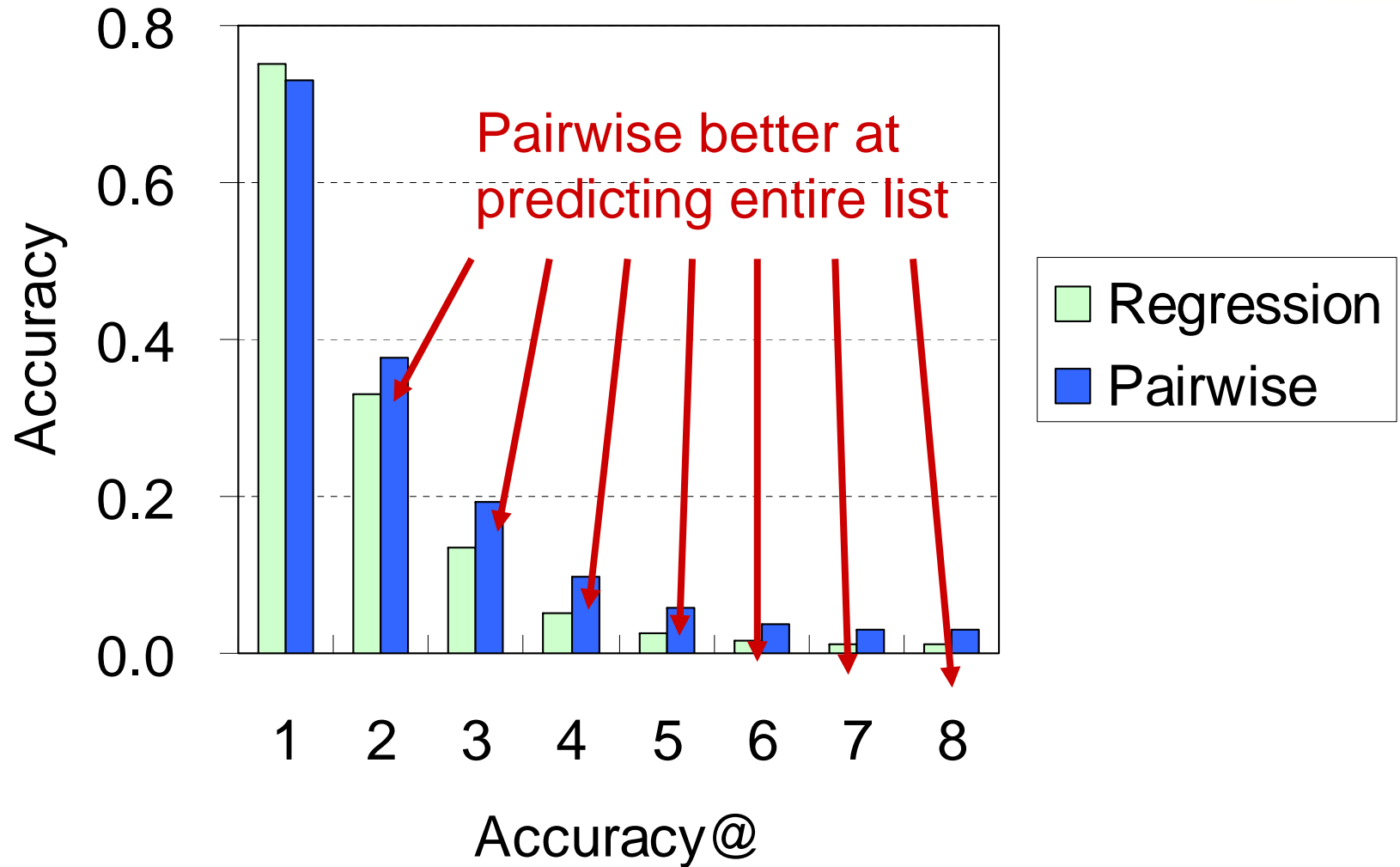
Results



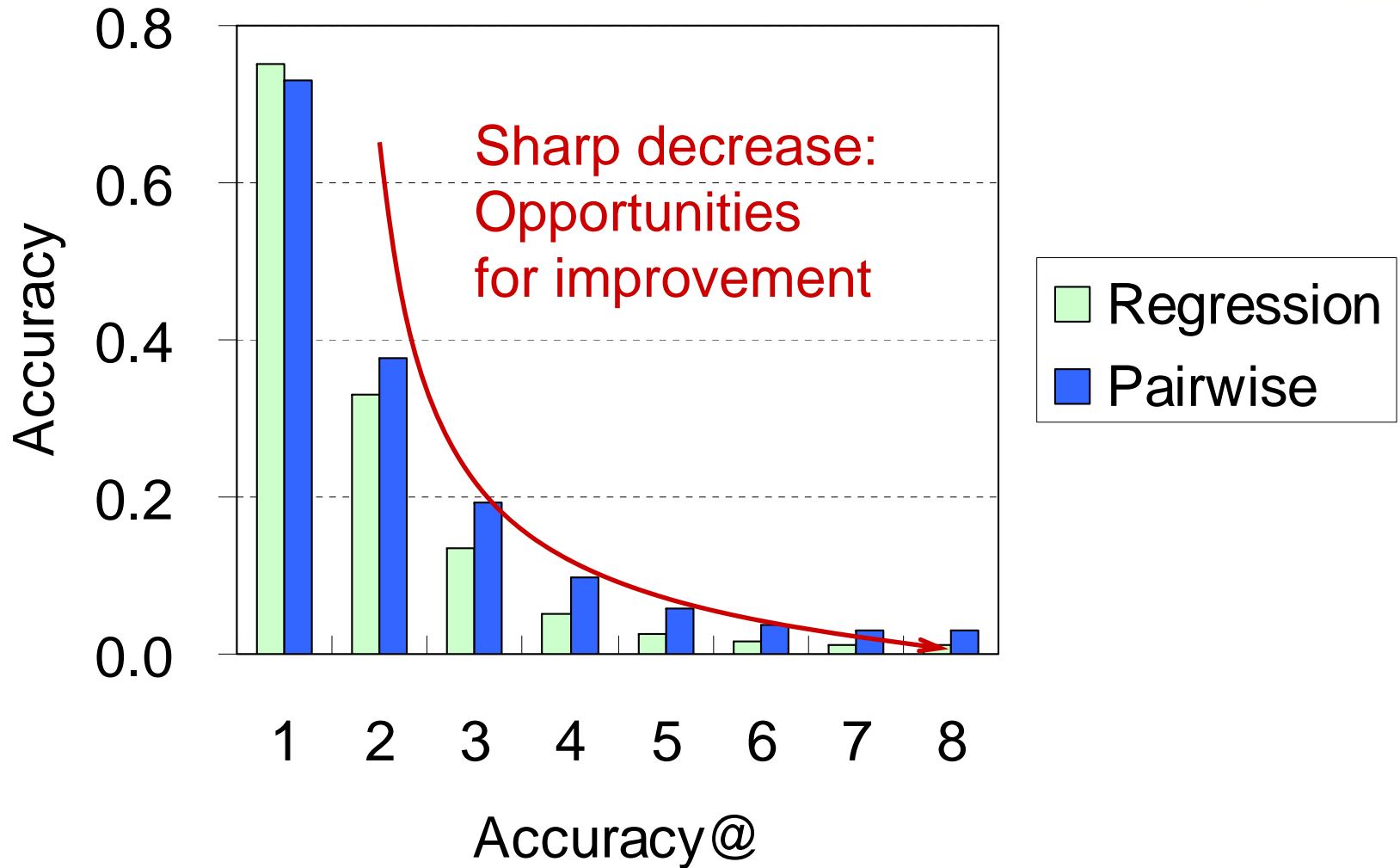
Results



Results



Results





Integration of Blog Browsing Experience into an Interactive Content Presenter

Motivation

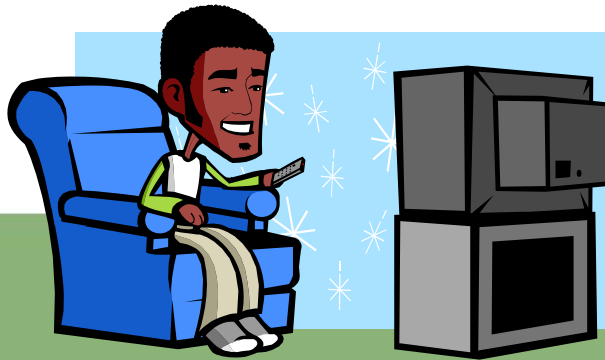
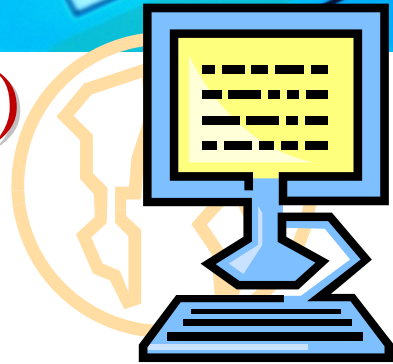
- Another **Emotion-Embedded Content (1)**

- BLOG - Text

- Integration of videos and audios (2)

- Replace BLOG browsing experience with the Video part

- **Physiological Preference Feedback (3)** applies as well



Emotion-Embedded Content



- Machine learning techniques applied

- Music and Photos

(Chen, Weng, Jeng & Chuang, 2008)

(Wu & Jeng, 2008)

- Blog Text

(Yang, Lin & Chen, 2007a, 2007b)



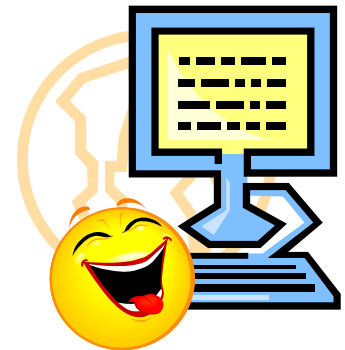
easy

happy



sublime

touching



- Focus on 8 emotion categories



From Blog Browsing to Video

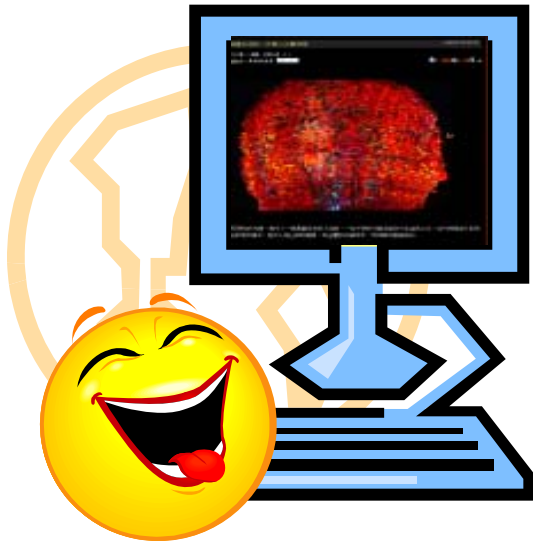
- Recorder
 - To record a browsing behavior while using the scroll
- Generator
 - 640×480 SHOT on blog page
 - Sampling Rate:
 - (3 pixels offset per SHOT)
 - (2 pixels offset per SHOT)



Integration with Music Slideshow Platform



- 8 (emotion classes) * 10 (blog pages)
=80 videos (in 640*480 wmv format)



		<table border="1"><tr><td>憤怒 Angry</td><td>害怕 Afraid</td><td>驚訝 Amazing</td><td>快樂 Happy</td></tr><tr><td>狂野 Wild</td><td>緊張 Tense</td><td>興奮 Exciting</td><td>幸福 Fortune</td></tr><tr><td>激勵 Inspiring</td><td>壯麗 Majestic</td><td>精彩 Wonderful</td><td>偉大 Grand</td></tr></table>	憤怒 Angry	害怕 Afraid	驚訝 Amazing	快樂 Happy	狂野 Wild	緊張 Tense	興奮 Exciting	幸福 Fortune	激勵 Inspiring	壯麗 Majestic	精彩 Wonderful	偉大 Grand				
憤怒 Angry	害怕 Afraid	驚訝 Amazing	快樂 Happy															
狂野 Wild	緊張 Tense	興奮 Exciting	幸福 Fortune															
激勵 Inspiring	壯麗 Majestic	精彩 Wonderful	偉大 Grand															
easy	happy	<table border="1"><tr><td>生氣 Angry</td><td>驚恐 Afraid</td><td>驚奇 Amazing</td><td>快樂 Happy</td></tr><tr><td>興奮 Exciting</td><td>緊張 Tense</td><td>幽默 Humorous</td><td>奇怪 Strange</td></tr><tr><td>令人 Inspiring</td><td>壯麗 Majestic</td><td>輕鬆 Relaxing</td><td>寧靜 Peaceful</td></tr><tr><td>偉大 Grand</td><td>精彩 Wonderful</td><td>容易 Easy</td><td>感動 Touching</td></tr></table>	生氣 Angry	驚恐 Afraid	驚奇 Amazing	快樂 Happy	興奮 Exciting	緊張 Tense	幽默 Humorous	奇怪 Strange	令人 Inspiring	壯麗 Majestic	輕鬆 Relaxing	寧靜 Peaceful	偉大 Grand	精彩 Wonderful	容易 Easy	感動 Touching
生氣 Angry	驚恐 Afraid	驚奇 Amazing	快樂 Happy															
興奮 Exciting	緊張 Tense	幽默 Humorous	奇怪 Strange															
令人 Inspiring	壯麗 Majestic	輕鬆 Relaxing	寧靜 Peaceful															
偉大 Grand	精彩 Wonderful	容易 Easy	感動 Touching															
		<table border="1"><tr><td>孤獨 Solitary</td><td>莊嚴 Majestic</td><td>溫柔 Tender</td><td>希望 Hopeful</td></tr><tr><td>高尚 Noble</td><td>悲傷 Sad</td><td>夢幻 Dreamy</td><td>感人 Touching</td></tr></table>	孤獨 Solitary	莊嚴 Majestic	溫柔 Tender	希望 Hopeful	高尚 Noble	悲傷 Sad	夢幻 Dreamy	感人 Touching								
孤獨 Solitary	莊嚴 Majestic	溫柔 Tender	希望 Hopeful															
高尚 Noble	悲傷 Sad	夢幻 Dreamy	感人 Touching															
sublime	touching																	

*slide mode

*blog mode

Related Publications



- Ming-Hung Hsu and Hsin-Hsi Chen (2008). “Tag Normalization and Prediction for Effective Social Media Retrieval.” *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence*, December 9-12, 2008, Sydney, Australia.
- Ming-Hung Hsu and Hsin-Hsi Chen (2008). “A Method to Predict Social Annotations.” *Proceedings of ACM 17th Conference on Information and Knowledge Management*, poster, October 26-30, 2008, Napa Valley, California.

Related Publications



- Changhua Yang, Kevin Hsin-Yih Lin and Hsin-Hsi Chen (2007). “Building Emotion Lexicon from Weblog Corpora.” *Proceedings of 45th Annual Meeting of Association for Computational Linguistics*, poster, June 23rd-30th, 2007, Prague, Czech Republic, 133-136.
- Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen (2007). “Emotion Classification Using Web Blog Corpora.” *Proceedings of 2007 IEEE/WIC/ACM International Conference on Web Intelligence*, November 2-5, 2007, Silicon Valley, 275-278.

Related Publications



- Kevin Hsin-Yih Lin, Changhua Yang and Hsin-Hsi Chen (2007). “What Emotions Do News Articles Trigger in Their Readers?” *Proceedings of 30th Annual International ACM SIGIR Conference*, poster, 23-27 July, 2007, Amsterdam, Netherland, 733-734.
- Kevin Hsin-Yih Lin and Hsin-Hsi Chen (2008). “Ranking Reader Emotions Using Pairwise Loss Minimization and Emotional Distribution Regression.” *Proceedings of EMNLP 2008: Conference on Empirical Methods in Natural Language Processing*, October 25-27, 2008, Honolulu, Hawaii.

Related Publications



- Kevin Hsin-Yih Lin, Changhua Yang, and Hsin-Hsi Chen (2008). “Emotion Classification of Online News Articles from the Reader’s Perspective.” *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence*, December 9-12, 2008, Sydney, Australia.
- Tien-Lin Wu, Hsuan-Kai Wang, Chien-Chang Ho, Yuan-Pin Lin, Tin-Tin Hu, Ming-Fang Weng, Li-Wei Chan, Chang-Hua Yang, Yi-Hsuan Yang, Yi-Ping Hung, Yung-Yu Chuang, Hsin-Hsi Chen, Homer Chen, Jyh-Horng Chen, Shyh-Kang Jeng (2008). “Interactive Content Presenter Based on Expressed Emotion and Physiological Feedback.” *ACM Multimedia Technical Demonstrations (ACM MM'08)*, October 27 – November 1, 2008, Vancouver, Canada.

Conclusion



- Prediction of Annotation in Social Media
- The experiment results show that our tag prediction model is able to predict a considerably large portion (~51%) of the stabilized tag set with only 5 user annotations.
- That facilitates recommendation and effective retrieval of new-coming resources.
- Social media provides training and testing set for emotion analysis.



Thank You!