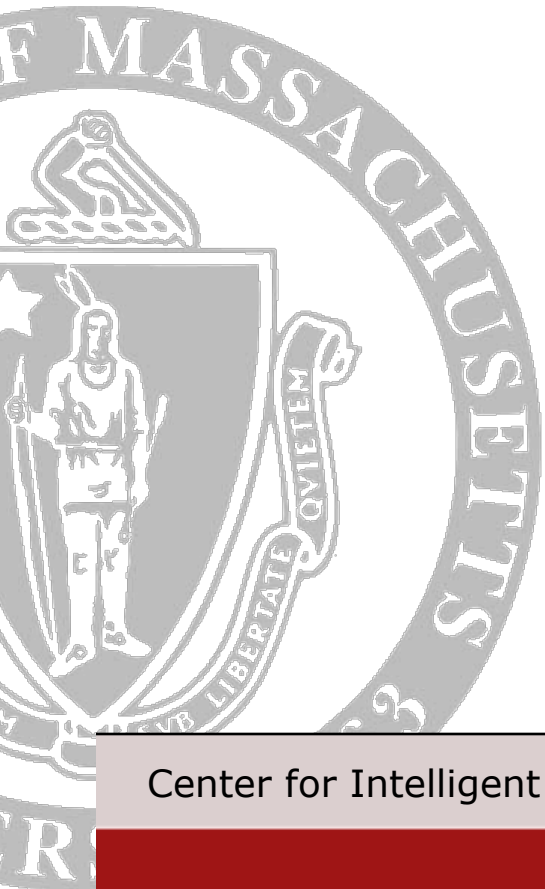


Predicting Searcher Frustration

Henry Feild (UMass Amherst)
James Allan (UMass Amherst)
Rosie Jones (Akamai*)

July 20, 2010

* work done while at Yahoo!



Satisfaction vs. Frustration vs. Success



- Dissatisfactory:
 - Getting a red light

- Frustrating:
 - Getting every single red light between your house and the airport



- Success
 - Reaching the airport in time to catch your flight



- Take away:
 - You can be dissatisfied and not frustrated
 - You can be successful but still frustrated along the way*

* Ceaparu et al. (Journal of HCI, 2004)

Real search example

- What was the best selling TV model in 2008?
 - Actual search sequence from UMass study:
 - television set sales 2008
 - "television set" sales 2008
 - "television" sales 2008
 - google trends
 - "television" sales statistics 2008
- ← user got frustrated starting here

Questions:

1. Can we detect when users get frustrated?
2. Can we do something to help users once we know they are frustrated?

Real search example

- What was the best selling TV model in 2008?
 - Actual search sequence from UMass study:
 - television set sales 2008
 - “television set” sales 2008
 - “television” sales 2008
 - google trends
 - “television” sales statistics 2008
- ← user got frustrated starting here

Questions:

1. Can we detect when users get frustrated?
2. Can we do something to help users once we know they are frustrated?

Outline

- Ways of detecting frustration
- User study overview
- Models
- Conclusion

Ways of detecting frustration

- Physical sensors
 - camera
 - predicts 6 mental states
 - pressure sensitive mouse
 - pressure sensors around mouse
 - pressure sensitive chair
 - pressure sensors on back and seat of chair

- Intelligent tutoring systems
 - user cognitive state prediction [Cooper et al. (UMAP 2009)]
 - frustration prediction [Kapoor et al. (J. of Human-Computer Studies, 2007)]
 - when will the user click an “I’m frustrated” button



Ways of detecting frustration

- Query logs

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

Ways of detecting frustration

- Query logs
 - search level
 - *query + navigation*

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

Ways of detecting frustration

- Query logs
 - search level
 - *query + navigation*
 - task level
 - *all searches related to an information need*

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

Ways of detecting frustration

■ Query logs

- search level
 - *query + navigation*
- task level
 - *all searches related to an information need*
- user level
 - *'personalization'*
 - *aggregate stats over previous tasks*

Where's the nearest cafe?

television set sales 2008

<click>

<scroll>

"television set" sales 2008

<click>

"television" sales 2008

<click>

<back>

...

What are the best grad school programs in CS?

When's the next time Dave Matthews is playing in Boston?

Ways of detecting frustration

- Query logs

- search level
 - *query + navigation*
- task level
 - *all searches related to an information need*
- user level
 - *'personalization'*
 - *aggregate stats over previous tasks*

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

- Search engine switching (White & Dumais, CIKM 2009)
- Next action prediction (Downey, ICAI, 2007)
- Task satisfaction (Huffman & Hochster, SIGIR 2007; Fox et al. TIS, 2005)

Outline

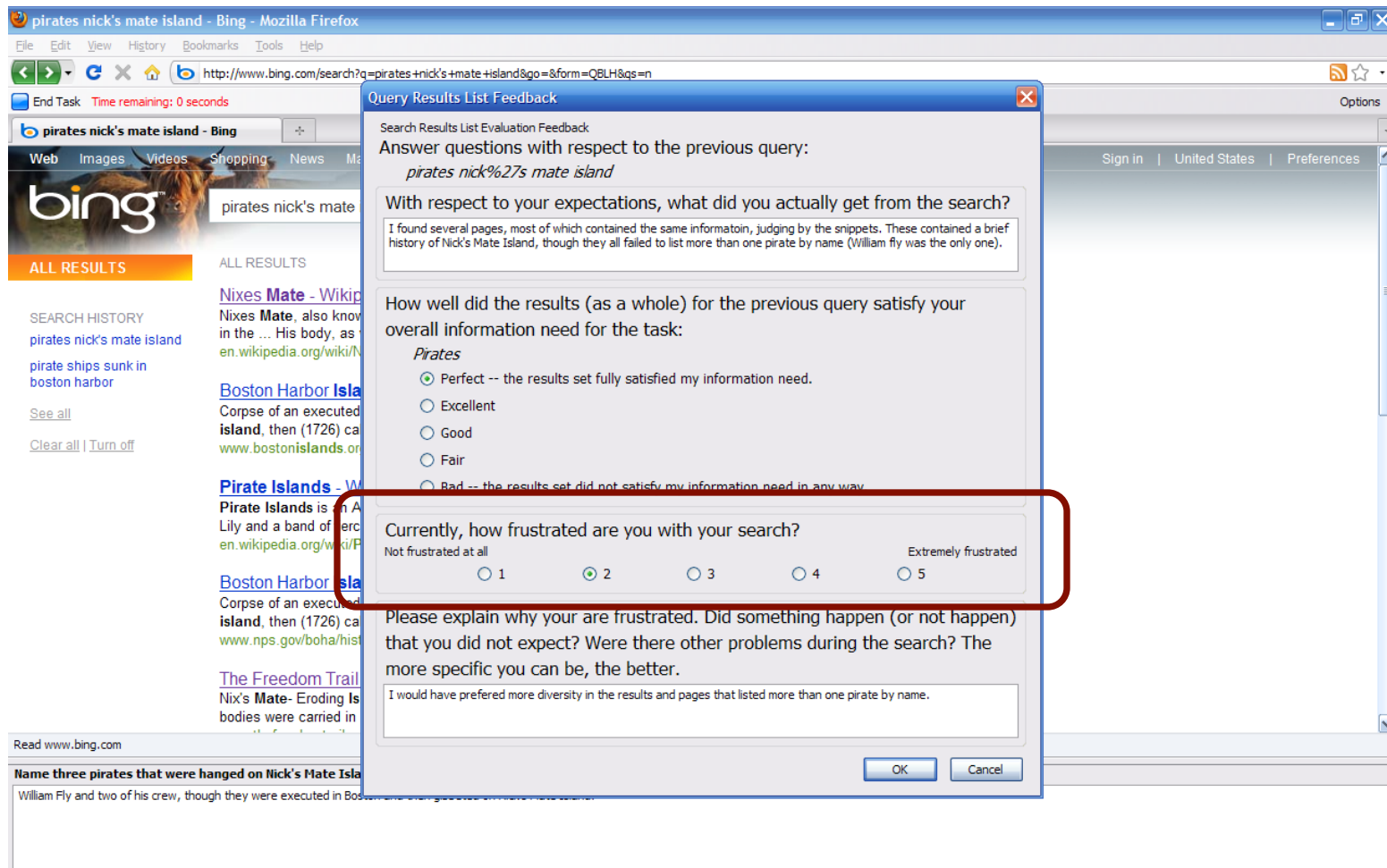
- Ways of detecting frustration
- **User study overview**
- Models
- Conclusion

User study

- 30 users
- assigned 7—8 pre-defined tasks
- searched the web
 - Google, Yahoo!, Bing, Ask.com
- prompted for feedback
- logged sensor readings + web browsing



Frustration reporting dialog



Frustration reporting dialog

Currently, how frustrated are you with your search?

Not frustrated at all Extremely frustrated

1 2 3 4 5

Frustration labels

Frustration Level

Search

1

television set sales 2008

1

"television set" sales 2008

1

"television" sales 2008

2

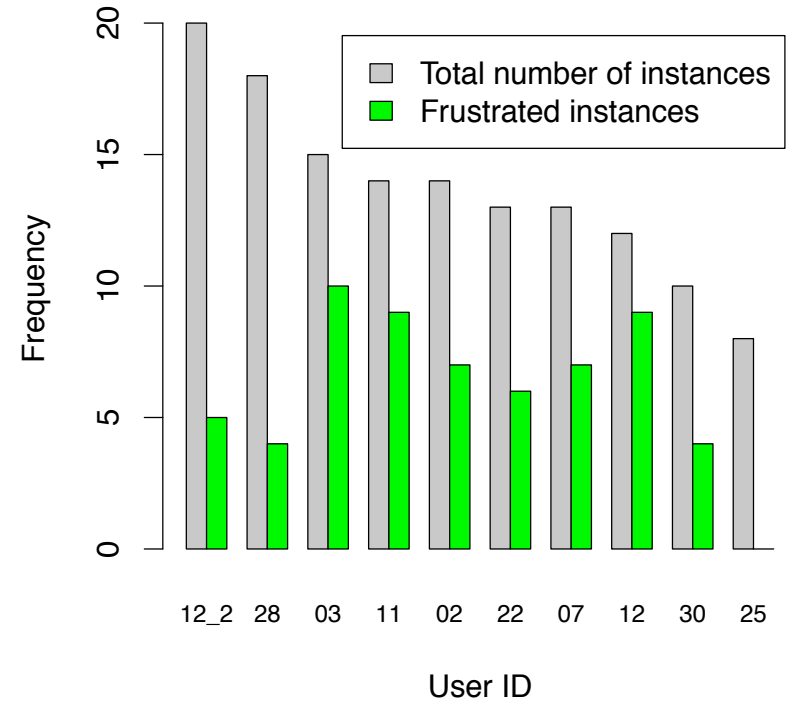
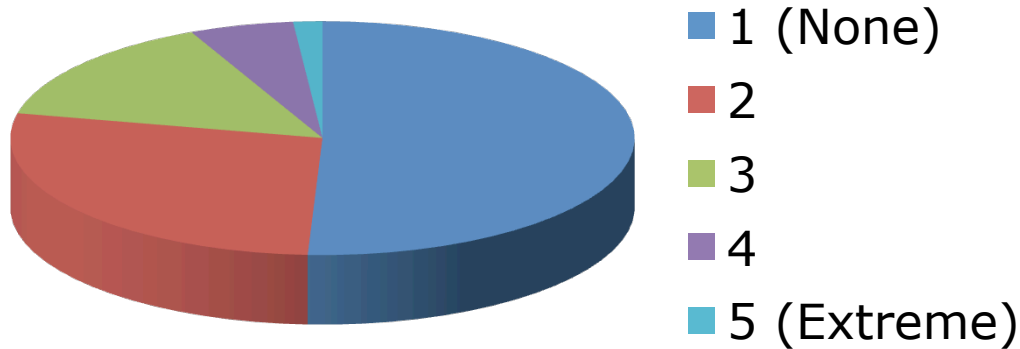
google trends

3

"television" sales statistics 2008

Statistics

Frustration Level Frequency



	Frustration	No Frustration
Success	46	85
Failure	72	8

Outline

- Ways of detecting frustration
- User study overview
- **Models**
- Conclusion

Sensor features

- 240 total

Sensor features

- 240 total
 - 10 sensor readings (from camera, mouse, & chair)

Sensor features

- 240 total
 - 10 sensor readings (from camera, mouse, & chair)
 - min, max, mean, std-dev

Sensor features

- 240 total
 - 10 sensor readings (from camera, mouse, & chair)
 - min, max, mean, std-dev
 - over time windows preceding frustration judgment:

Sensor features

- 240 total
 - 10 sensor readings (from camera, mouse, & chair)
 - min, max, mean, std-dev
 - over time windows preceding frustration judgment:
 - 30 seconds

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

Sensor features

- 240 total
 - 10 sensor readings (from camera, mouse, & chair)
 - min, max, mean, std-dev
 - over time windows preceding frustration judgment:
 - 30 seconds
 - search

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

Sensor features

- 240 total
 - 10 sensor readings (from camera, mouse, & chair)
 - min, max, mean, std-dev
 - over time windows preceding frustration judgment:
 - 30 seconds
 - search
 - entire task

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

Sensor features

- 240 total
 - 10 sensor readings (from camera, mouse, & chair)
 - min, max, mean, std-dev
 - over time windows preceding frustration judgment:
 - 30 seconds
 - search
 - entire task
 - two versions of each:
 - including time spent responding to prompts
 - excluding time spent responding to prompts

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

Query log features

- 43 total
 - search-level
 - task-level
 - user-level

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

Query log features

- 43 total
 - **search-level**
 - task-level
 - user-level

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

Query log features

- 43 total
 - **search-level**
 - search duration
 - query length
 - average word length in query
 - pages clicked...
 - task-level

 - user-level

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

Query log features

- 43 total
 - search-level
 - search duration
 - query length
 - average word length in query
 - pages clicked...
 - **task-level**
 - user-level

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

Query log features

- 43 total
 - search-level
 - search duration
 - query length
 - average word length in query
 - pages clicked...
 - **task-level**
 - task duration
 - # of searches
 - average query length...
 - user-level

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

Query log features

- 43 total
 - search-level
 - search duration
 - query length
 - average word length in query
 - pages clicked...
 - task-level
 - task duration
 - # of searches
 - average query length...
 - **user-level**

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

Where’s the nearest cafe?

What are the best grad school programs in CS?

When’s the next time Dave Matthews is playing in Boston?

Query log features

- 43 total
 - search-level
 - search duration
 - query length
 - average word length in query
 - pages clicked...
 - task-level
 - task duration
 - # of searches
 - average query length...
 - **user-level**
 - average # of URLs visited per task
 - average # of actions per task...

television set sales 2008

<click>

<scroll>

“television set” sales 2008

<click>

“television” sales 2008

<click>

<back>

...

Where’s the
nearest cafe?

What are the best
grad school
programs in CS?

When’s the next time
Dave Matthews is
playing in Boston?

Modeling

- logistic regression
 - binarize instances:
 - 1 = “not frustrated”
 - 2–5 = “frustrated”

Models

- all features
 - query log + sensors
- Sequential Forward Selection (SFS) over:
 - all features
 - 7 features automatically chosen
 - query log features
 - 5 features automatically chosen
 - sensor features
 - 3 features automatically chosen
- search engine switching [White & Dumais, CIKM 2009]
 - 5 query log features
- Markov Model Likelihood (event patterns) [Hassan et al. WSDM 2009]

Features from two of the models

SFS-QL+Sensors: SFS over query log and sensor features

1. task duration
2. proportion of unique queries in task
3. mean of **'unsure'**, **30-sec**, **no prompts**
4. minimum of **'unsure'**, **search**, **prompts**
5. stddev of **'concentrating'**, **30-sec**, **no prompts**
6. minimum of **'net-back-change'**, **search**, **no prompts**
7. minimum of **'concentrating'**, **search**, **no prompts**

W&D: Model used by White & Dumais (CIKM 2009) to detect switching between search engines

[task] task duration

[user] average number of URL's visited per task

[search] character length of most recent query

[search] average token length of most recent query

[task] number of actions performed in task

Results

Model	Accuracy	$F_{\beta=0.5}$	Mean Average Precision
W&D	0.75	0.80	0.87
SFS-QL+Sensors	0.69	0.72	0.85
SFS-QL	0.69	0.73	0.80
W&D+MML-time	0.66	0.69	0.76
MML-time	0.56	0.62	0.65
SFS-Sensors	0.55	0.61	0.65
QL+Sensors	0.54	0.49	0.59
Always frustrated	0.44	0.55	---

Outline

- Ways of detecting frustration
- User study overview
- Models
- **Conclusion**

Conclusions

- Searcher frustration is detectable
- Sensors are not helpful using our processing methods
- Best prediction criteria:
 - long task duration
 - user tends to visit few URLs per task
 - few clicks and other actions are performed
 - the most recent query is long, but has very short words

Future work

- What models work best in real search environments?
- How can we help frustrated searchers?

Results

