

# Detecting Online Commercial Intention (OCI)

---

Honghua Dai et al.

Microsoft

AdCenter Lab

Presented by Yi ZHU

2010-09-28

# Outline

- Motivation
- Problem Definition
- Online Commercial Intention (OCI) Detecting
  - Web Page OCI
  - Query OCI
- Evaluation and Experiment
- Conclusion and Future Work

# Motivation

- Serving ads more effective and less annoying, when user has intent to purchase
- Different services strategies for different users
- Detect user's commercial value

# Search Intention

## □ Navigational

- Reach a particular web site

## □ Informational

- Acquire information on web pages

## □ Transactional / Resource

- Perform some “web-mediated” activity

|                          | Commercial        | Non-Commercial |
|--------------------------|-------------------|----------------|
| Navigational             | Wal-Mart          | Hotmail        |
| Informational            | Digital camera    | San Francisco  |
| Transactional / Resource | U2 music download | Lyrics         |

# Problem Definition

- Definition of Online Commercial Intention (OCI)
  - Function of query or web page, to a binary value: Commercial or Non-Commercial (Binary Classification)
  - Commercial: general purpose of submitting query or visiting a web page is to commit a commercial activity (purchase, auction, selling, paid service)

# Problem Definition (cont.)

- T: the set of all possible terms.
- Q: the set of search queries.
- P: the set of all web pages on the web.
- Online Commercial Intention

$$OCI : Q \rightarrow \{\text{Commercial, Non-Commercial}\}$$

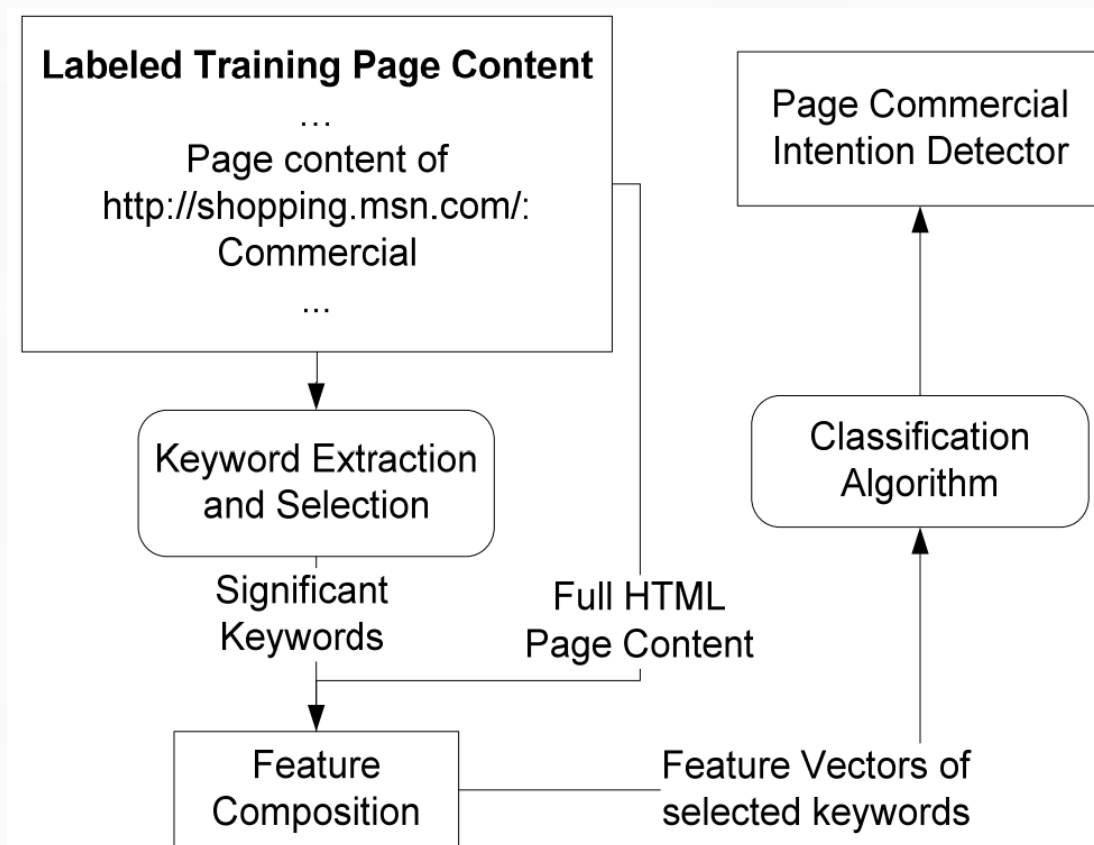
$$OCI : P \rightarrow \{\text{Commercial, Non-Commercial}\}$$

# OCI Detecting

- Web Page OCI
- Query OCI

# Web Page OCI

## □ Web Page OCI Detector





# Web Page OCl (cont.)

## □ Keywords Selection

□ Significance 
$$Sig(k) = \frac{Max\{\Pr(k | C_+), \Pr(k | C_-)\}}{\Pr(k | C_-) + \Pr(k | C_+)} \times 2 - 1$$

□ Frequency 
$$Freq(k) = \Pr(k | C_+ \cup C_-)$$

□  $\Pr(k | C)$  is the probability the  $k$  occurring in a web page belonging to class  $C$

□  $C_+$  and  $C_-$  are positive and negative class respectively

# Web Page OCI (cont.)

- keyword set  $K = \{k_1, \dots, k_n\}$
- Keyword  $k$  appearance rate in inner text in  $p$   $nit(k, p)$
- Keyword  $k$  appearance rate in tag attributes in  $p$   $nta(k, p)$
- Combine all  $nit(k, p)$  and  $nta(k, p)$  to obtain a vector of  $2*n$  dimensions (Input features for classification)
- Smoothed by power of 0.125
- SVM

# Web Page OCI (cont.)

## □ Input:

- A set of labeled pages for training

## □ Output:

- OCI value of web page
- Label of web page

# Query OCI

## □ Data Sources

- Constituent terms of search query
- Content of top landing pages recommended by search engine
- Content of search result page
- Click-through data

# Query OCI (cont.)

- Query
- Search result page
- Search result landing page
- Query snippet

# Query OCI (cont.)



Online Commercial Intention

query

Search

[Advanced Search](#)

Search:  the web  pages from Hong Kong

Web

[+ Show options...](#)

Results 1 - 10 of about 17,100,000

[Detecting Online Commercial Intention: Audience Intelligence ...](#)

Web page searches display two levels of **commercial intent**: informational and transactional. This tool can detect customer **intent** to acquire information or ...

[adlab.msn.com/Online-Commercial-Intention/](#) - [Cached](#) - [Similar](#)

[Online Commercial Intent | Online Commercial Intent | BloggerVenue](#)

12 May 2008 ... Well, for this, microsoft has a free tool called **Online Commercial Intent**, OCI for short. This tool analyses determined keywords that have ...

[bloggervenue.com/online-commercial-intent/](#) - [Cached](#) - [Similar](#)

[Online Commercial Intent \(OCI\) Problem](#)

If you've got the time, could you take a stab at this question. What is the correlation between **Online Commercial Intent** and Number of Searches (exact).

[www.warriorforum.com/.../61010-online-commercial-intent-oci-problem.html](#) - [Cached](#) - [Similar](#)

[Online Commercial Intention](#)

Microsoft adCente labs file, created by Honghua (Kathy) Dai, Zaiqing Nie, Lee Wang, Lingzhi Zhao, Ji-Rong Wen, and Ying.

[www.slideshare.net/.../online-commercial-intention-presentation](#) - United States - [Cached](#) - [Similar](#)

[Detecting online commercial intention \(OCI\)](#)

by HK Dai - 2006 - [Cited by 41](#) - [Related articles](#)

We call the commercial intentions behind a user's online activities as OCI (**Online Commercial Intention**). We also propose the notion of "Commercial Activity ...

[portal.acm.org/citation.cfm?id=1135902](#) - [Similar](#)

Landing Page Link

Query Snippet

# Query OCI (cont.)

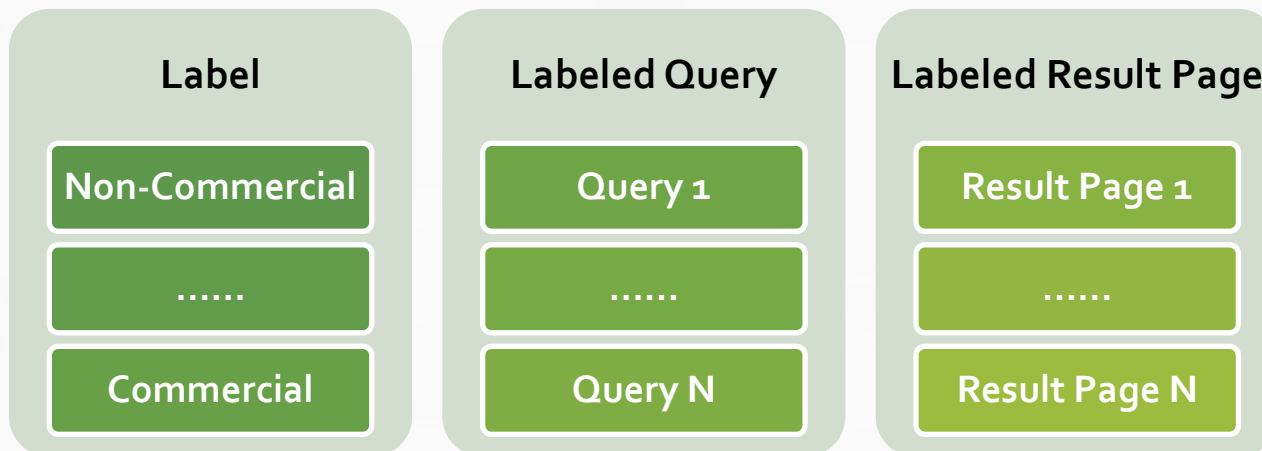
- Top landing pages recommended by search engine
  - Compute OCIs for all top landing pages

$$OCI(TLP_q) = \sum_{i=1}^N \alpha_i \times OCI(p_q^i)$$

- $p_q^i$  is the Web page that has rank  $i$  in the search result of query  $q$
- $\alpha_i$  is learning parameter learnt by SVM

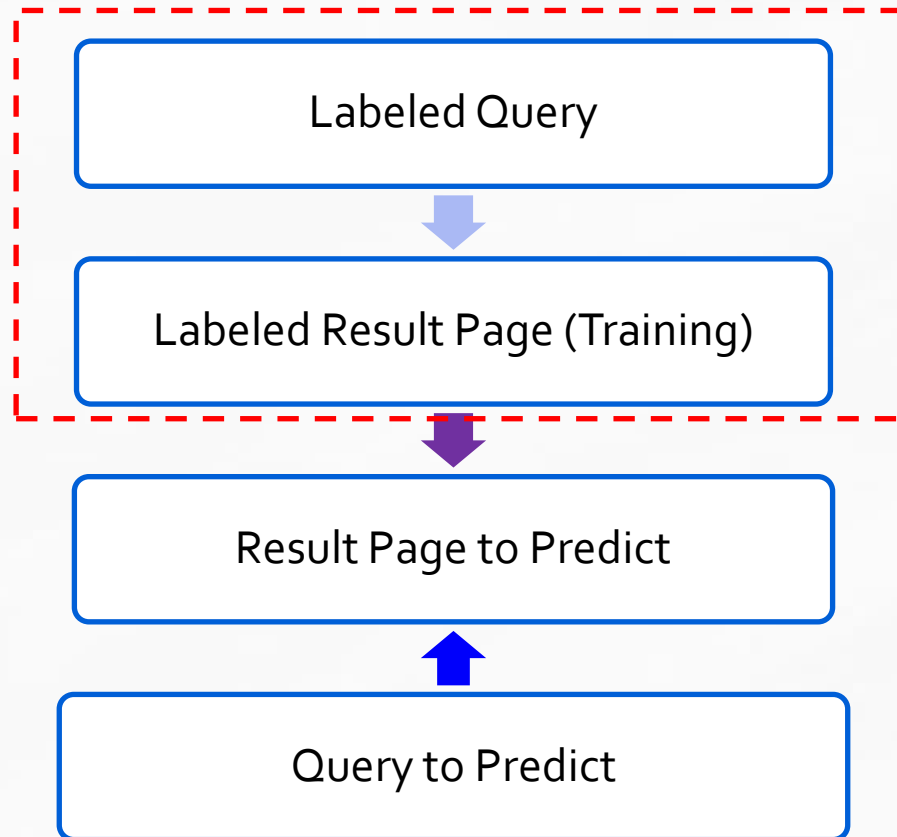
# Query OCI (cont.)

- First search result page
  - Use labeled query to get search result
  - The search result page will be assigned the same label as the query (labeled search result page)
  - Use the labeled search result pages as training set to build a Web page OCI detector
  - Predict query OCI





# Query OCI (cont.)



# Evaluation

## □ Manually label

- 3 human labelers, majority vote (commercial, non-commercial, confused)
- Randomly pick (page, query)
- Labeled commercial page ?

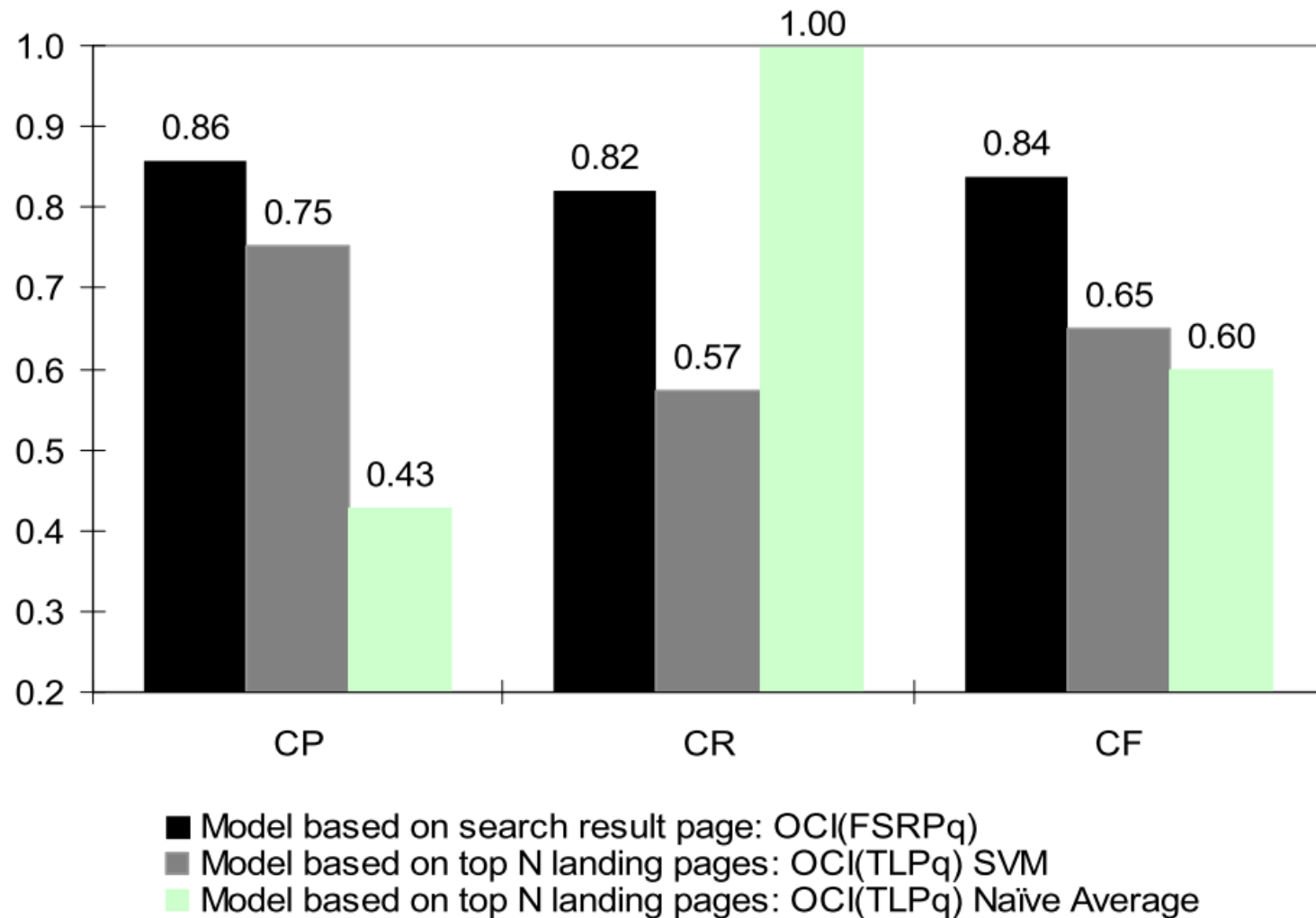
|                | Pages | Queries |
|----------------|-------|---------|
| Commercial     | 4074  | 602     |
| Non-Commercial | 21823 | 790     |
| Confused       | 289   | 16      |
| Total          | 26186 | 1408    |

|                | Training | Test | Total |
|----------------|----------|------|-------|
| Commercial     | 2820     | 2936 | 5756  |
| Non-Commercial | 2555     | 2653 | 5208  |
| Total          | 5375     | 5589 | 10964 |

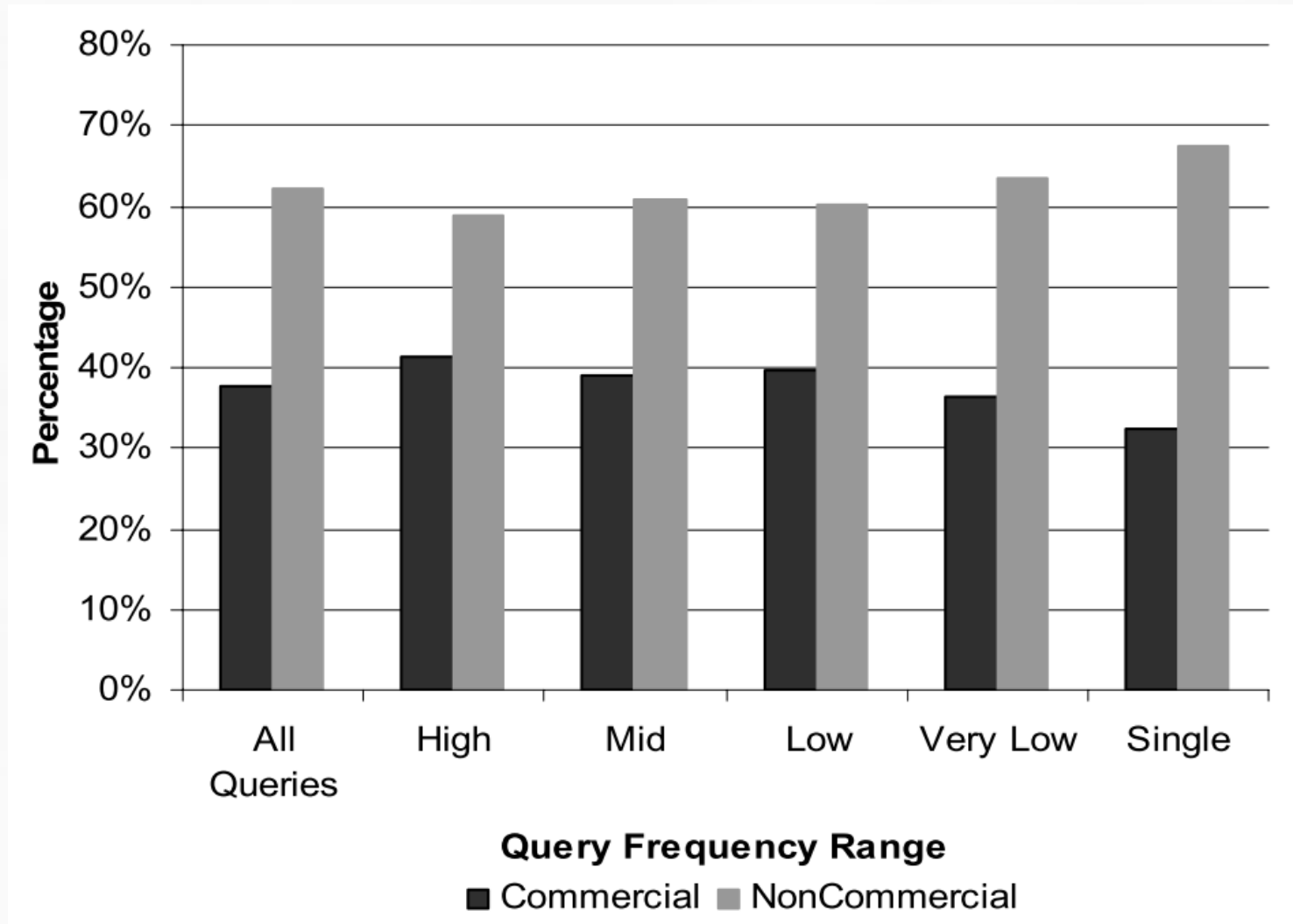
# Experiment (Web Page OCl)

| Threshold  | Keyword No. | Precision    | Recall       | F-measure    |
|------------|-------------|--------------|--------------|--------------|
| 0.01       | 4523        | 0.814        | 0.907        | 0.858        |
| 0.03       | 1712        | 0.956        | 0.884        | 0.919        |
| 0.05       | 989         | 0.948        | 0.899        | 0.923        |
| 0.075      | 600         | 0.934        | 0.918        | 0.926        |
| <b>0.1</b> | <b>391</b>  | <b>0.930</b> | <b>0.925</b> | <b>0.928</b> |
| 0.15       | 179         | 0.921        | 0.923        | 0.922        |
| 0.2        | 100         | 0.916        | 0.905        | 0.910        |
| 0.3        | 25          | 0.893        | 0.840        | 0.865        |
| 0.4        | 6           | 0.848        | 0.791        | 0.819        |

# Experiment (Query OCI)



# Experiment (cont.)



# Conclusion and Future Work

## □ Conclusion

- Formal definition of OCI
- Three methods to detect OCI of web page and query

## □ Future Work

- Less labeling effort
- Utilize click through data
- Utilize the context of individual user

Thanks!  
Q&A