Microsoft Research

Psychological Advertising:
Exploring User Psychology for Click Prediction in Sponsored Search

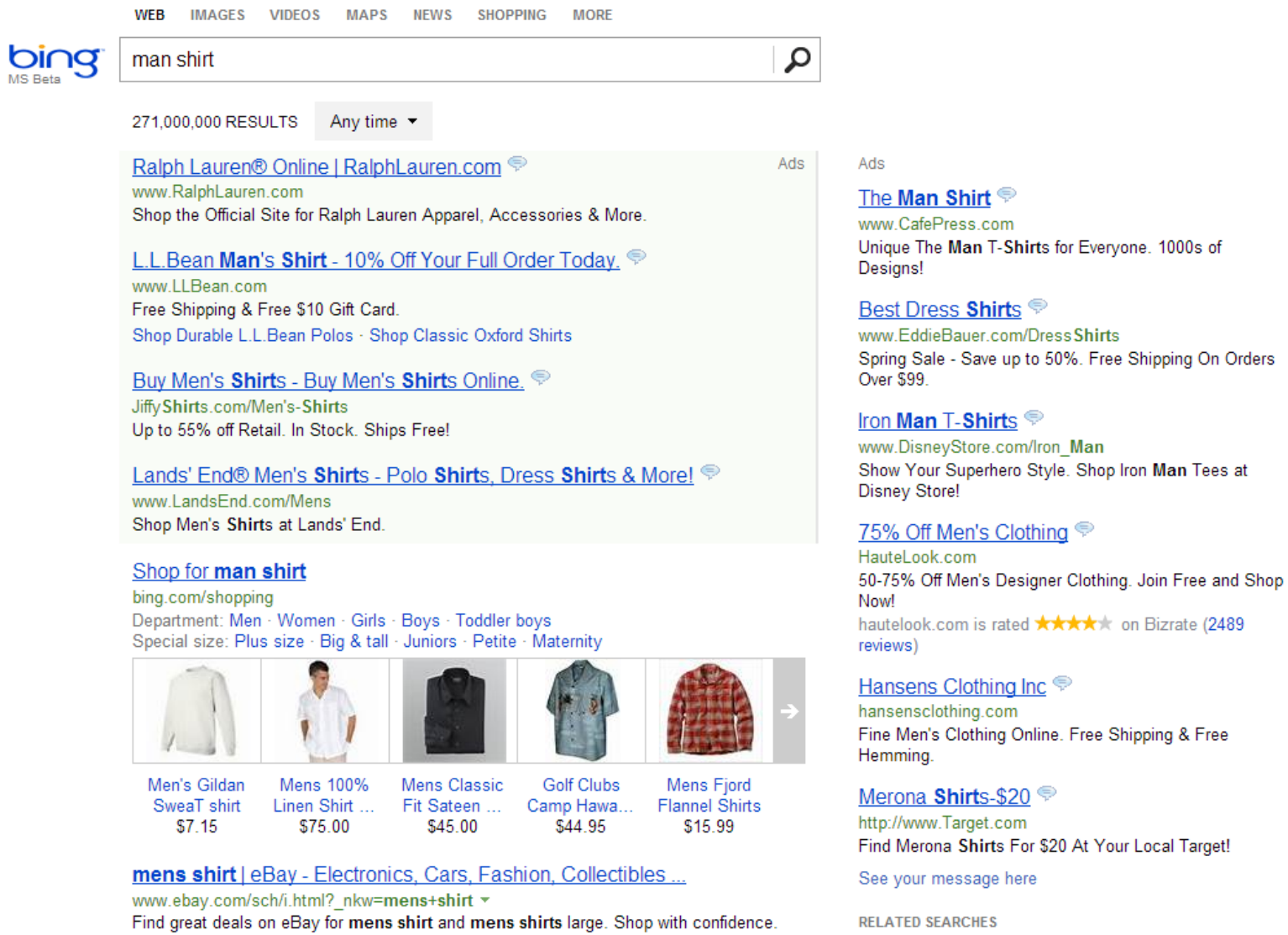
Taifeng Wang[1], Jiang Bian[1], Shusen Liu[2], Yuyu Zhang[3], Tie-Yan Liu[1]

[1]Microsoft Research Asia

[2]South China University of Technology

[3]Chinese Academy of Science

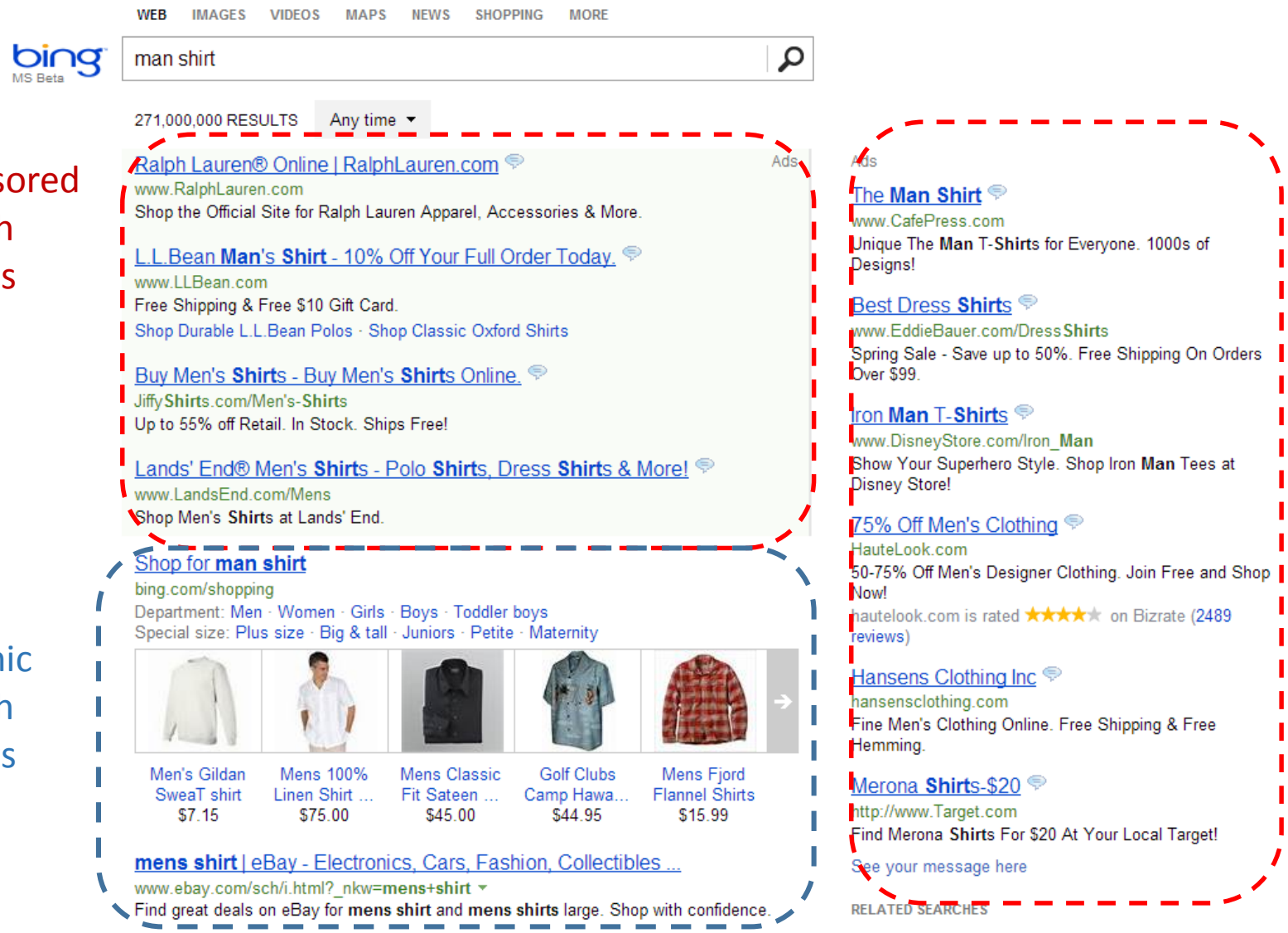
Sponsored Search



Sponsored Search

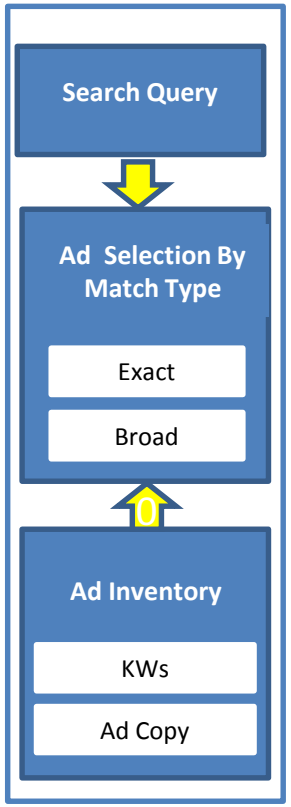
Sponsored search results

Organic search results

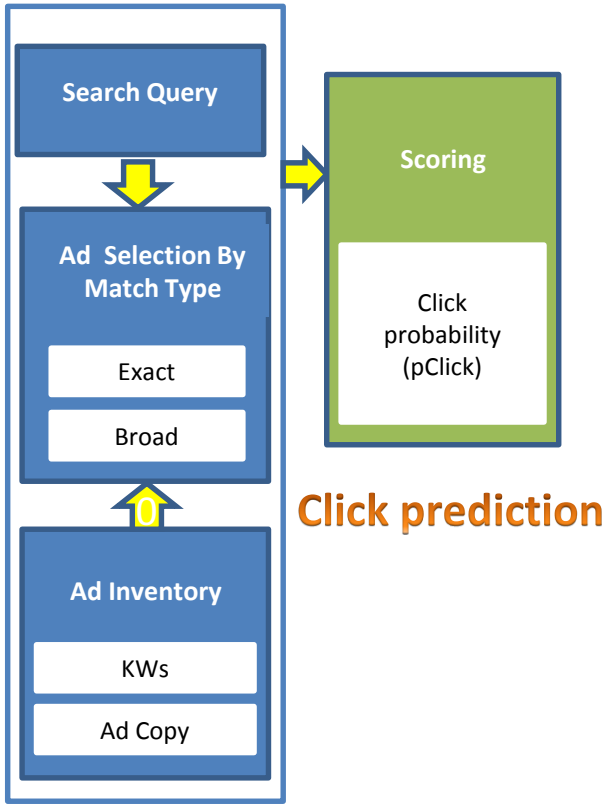


Click Prediction in Sponsored Search

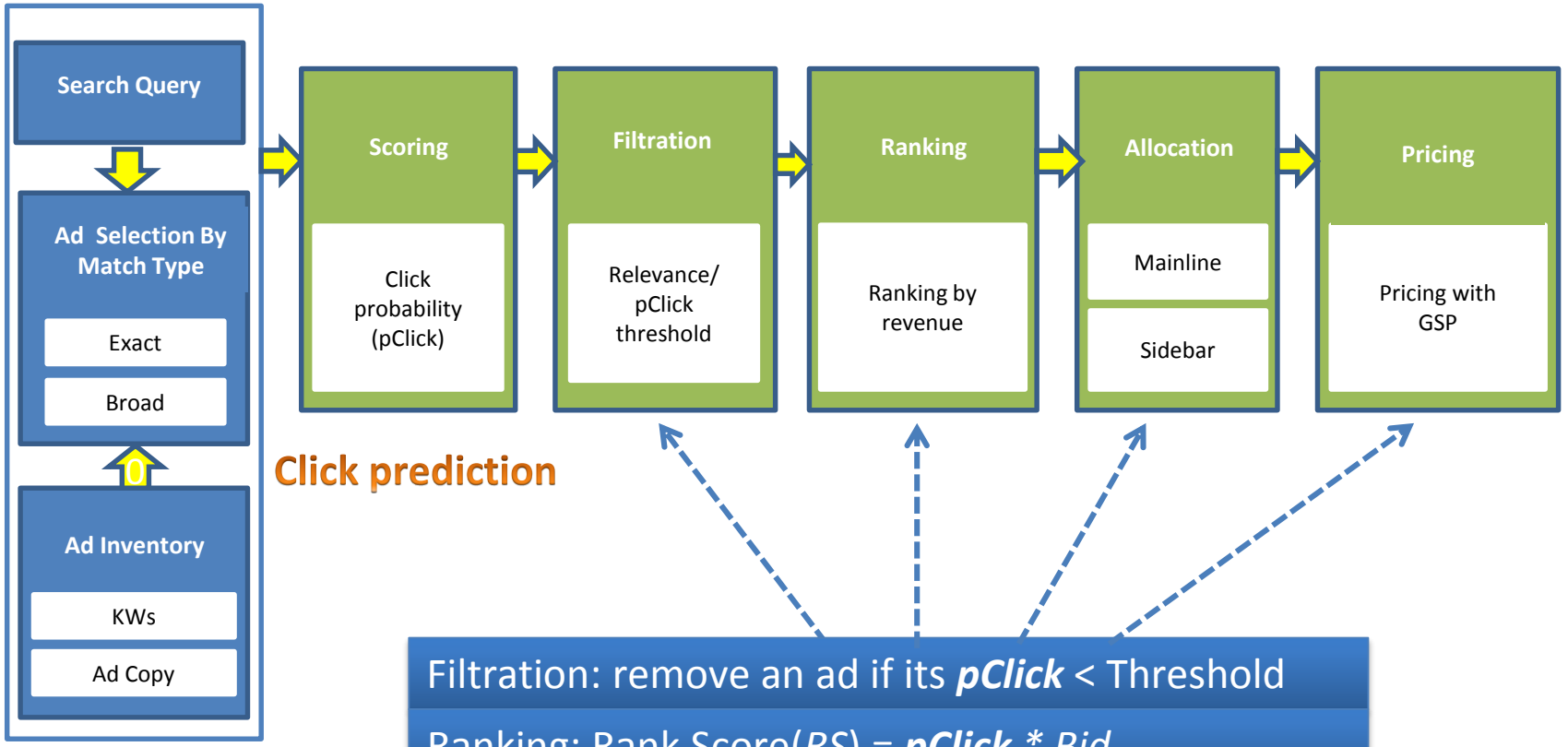
Click Prediction in Sponsored Search



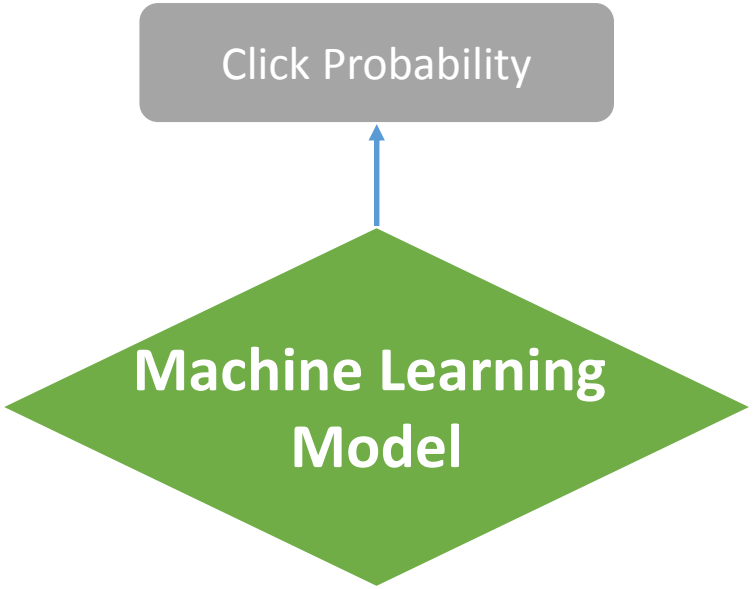
Click Prediction in Sponsored Search



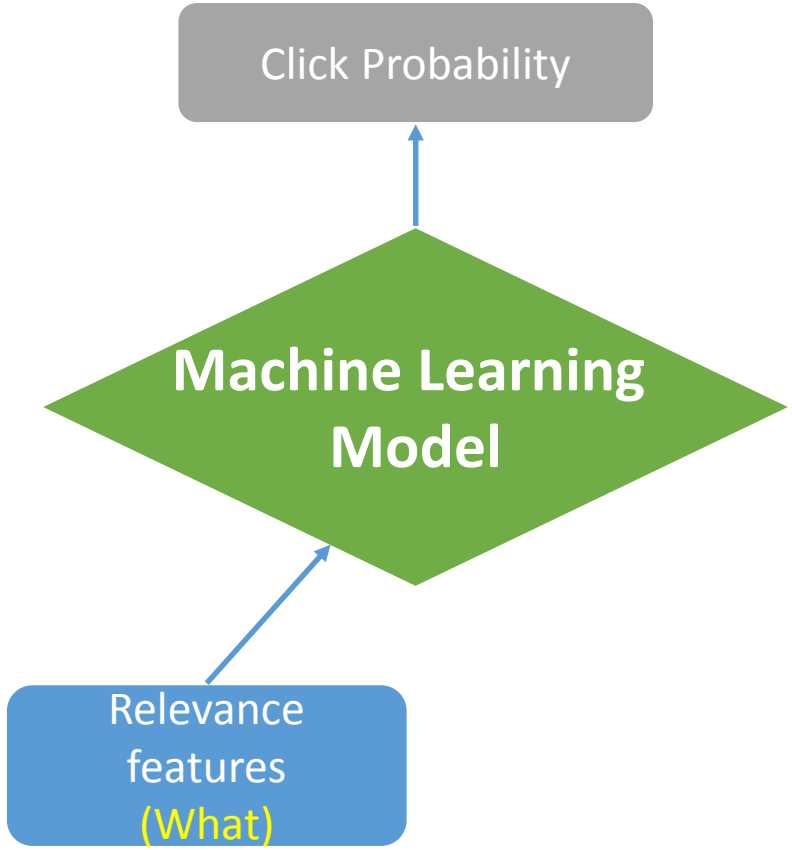
Click Prediction in Sponsored Search



State-of-The-Art Click Prediction Modeling

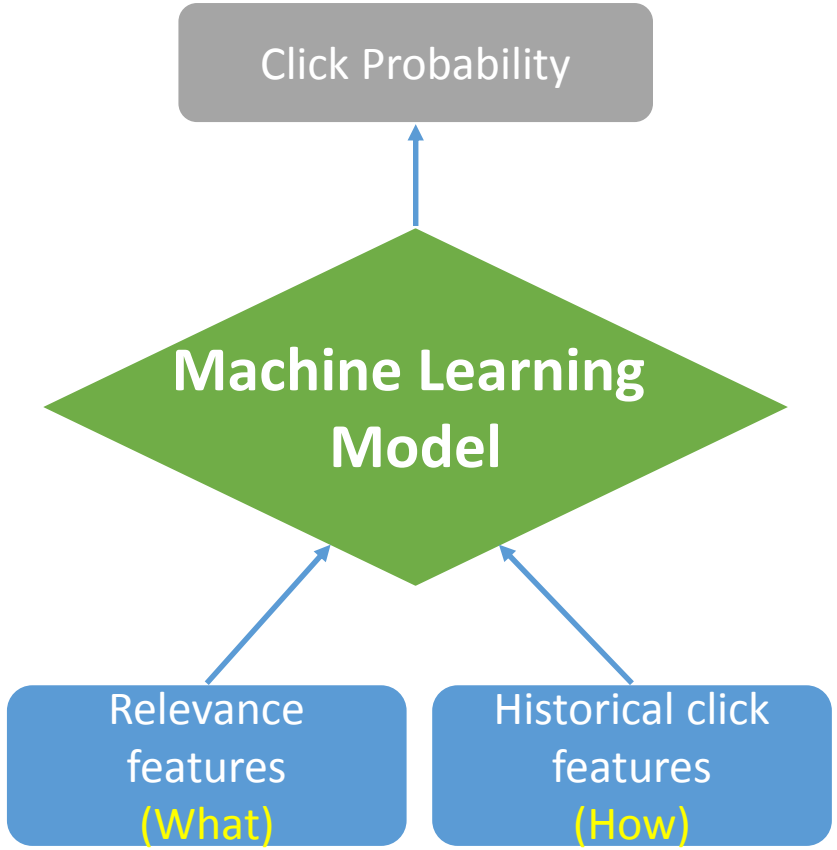


State-of-The-Art Click Prediction Modeling



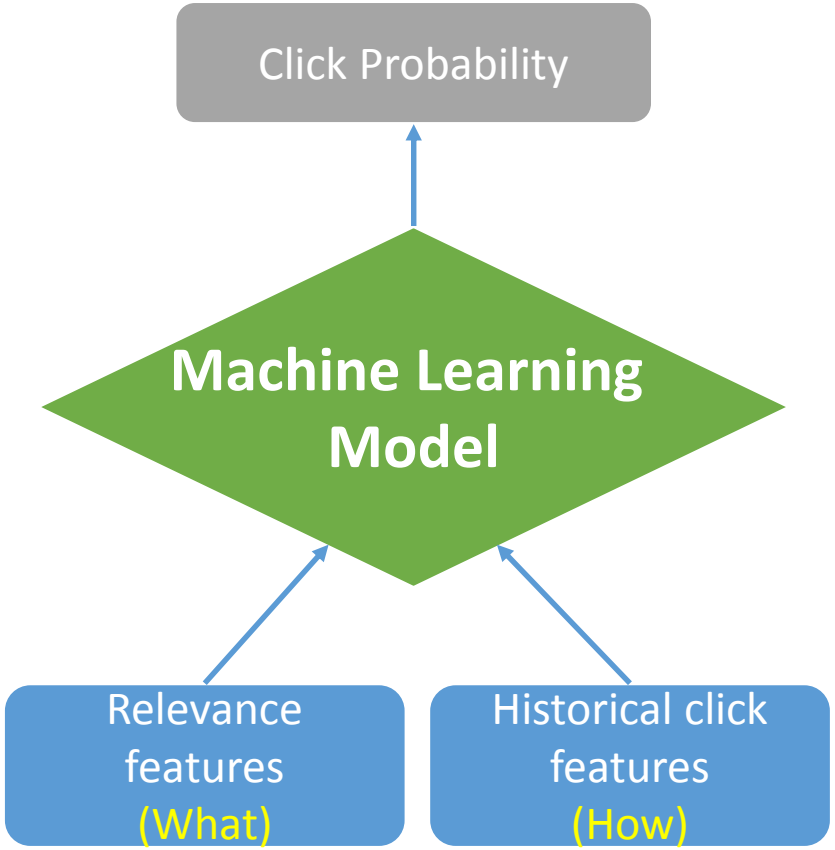
- Relevance features
  - ***What*** relevant content users seek to click

State-of-The-Art Click Prediction Modeling



- Relevance features
  - ***What*** relevant content users seek to click
- Historical click features
  - ***How*** users click

State-of-The-Art Click Prediction Modeling

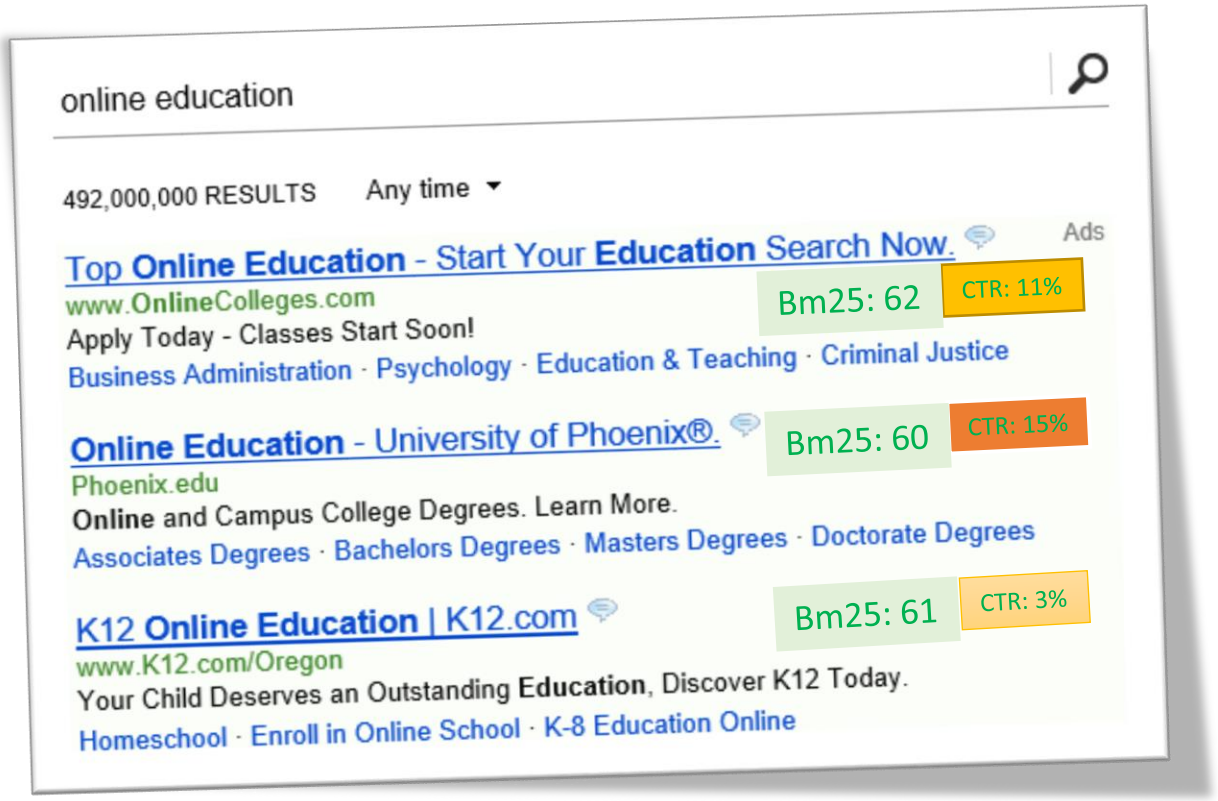


- Relevance features
  - ***What*** relevant content users seek to click
- Historical click features
  - ***How*** users click

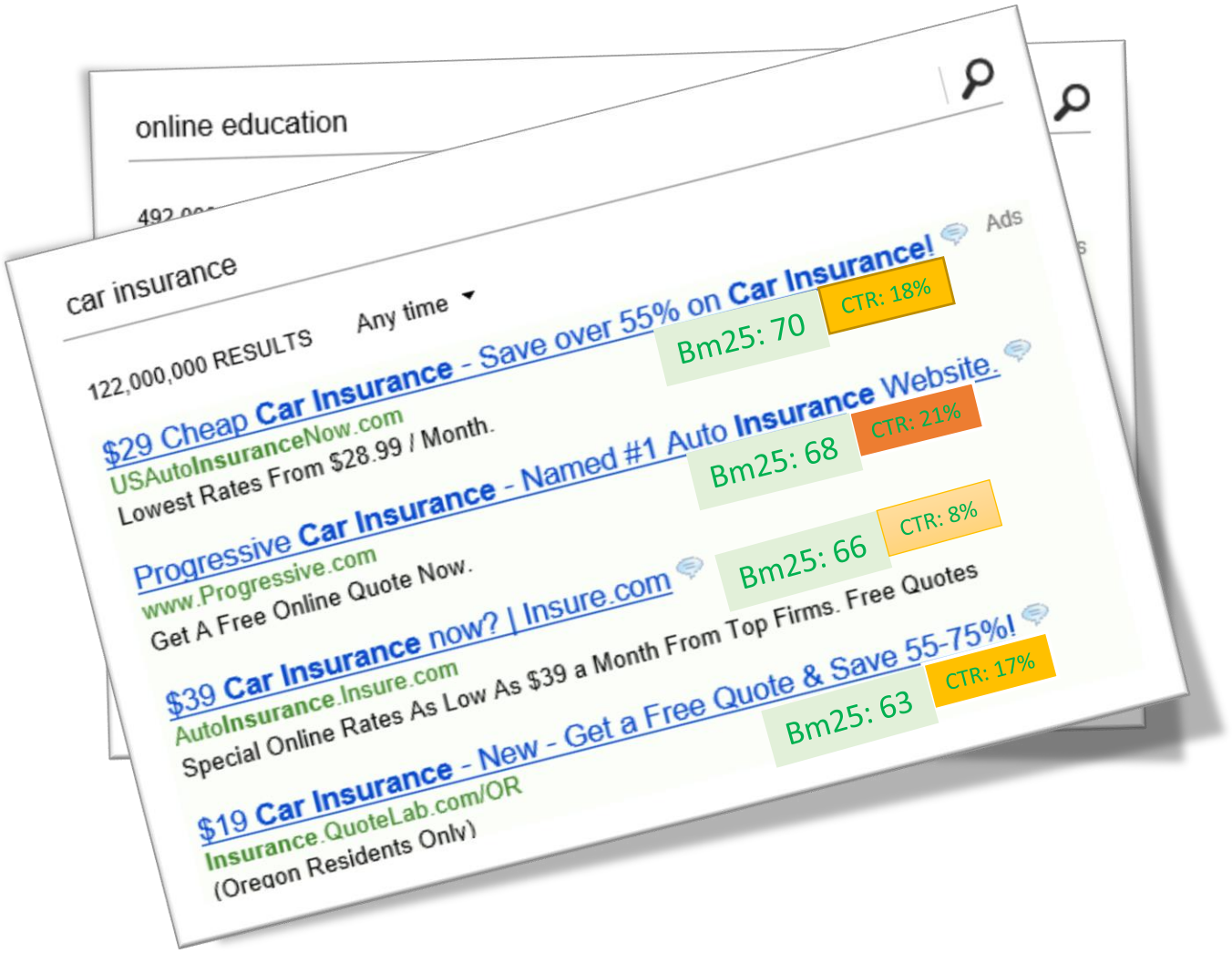
What can help us answer ***why*** users click?

Relevance Is Not Enough

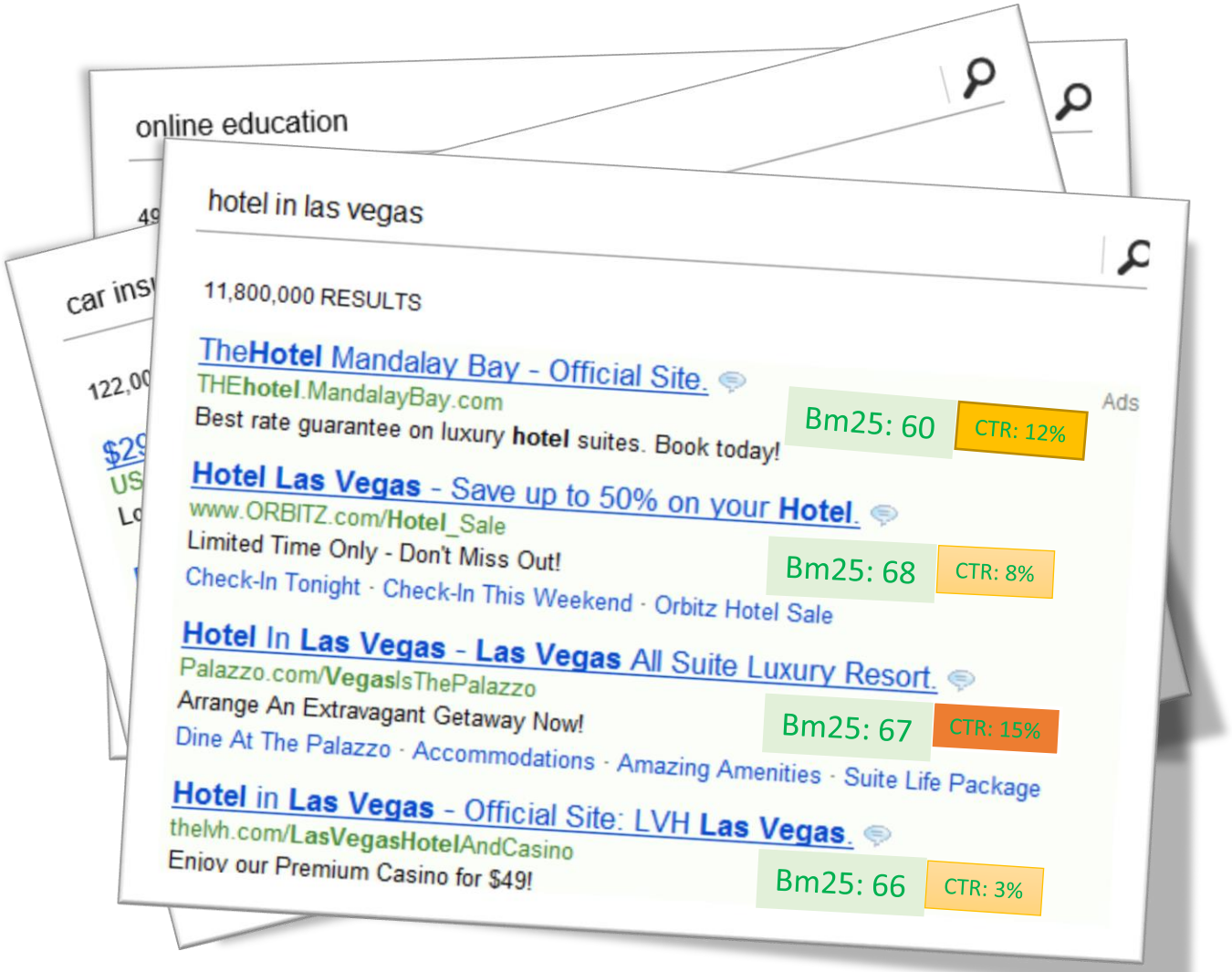
Relevance Is Not Enough



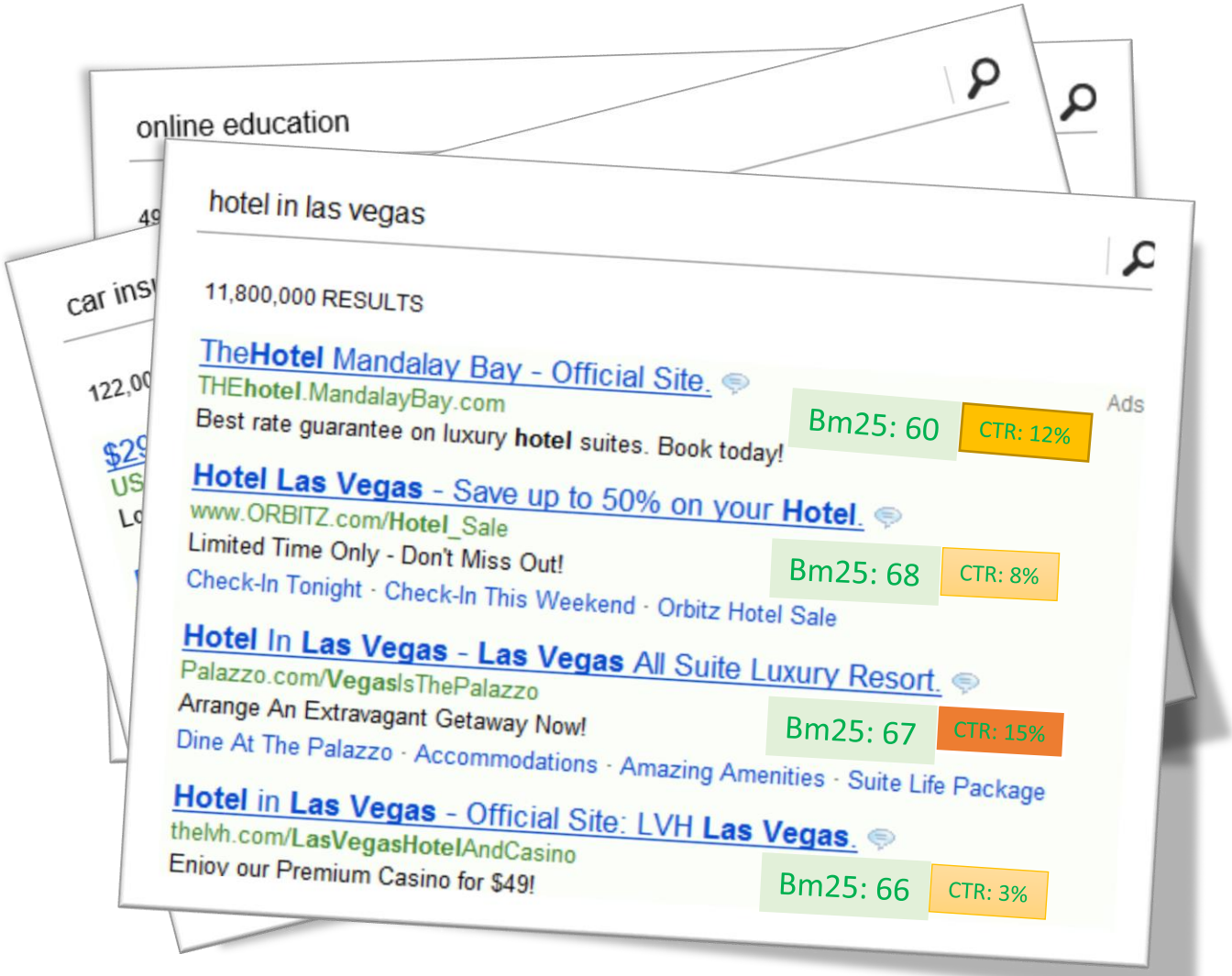
Relevance Is Not Enough



Relevance Is Not Enough



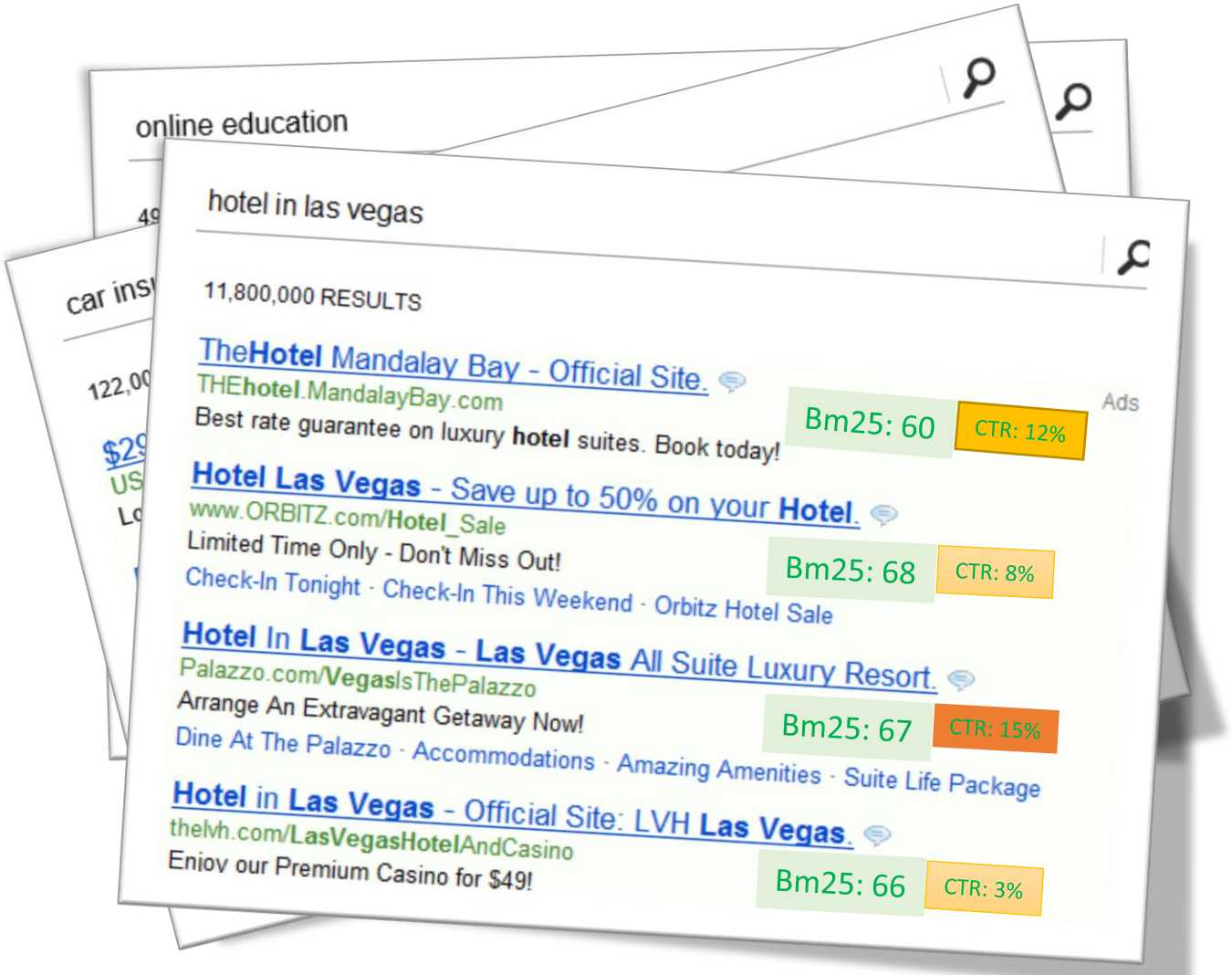
Relevance Is Not Enough



If directly using relevance score to predict clicks, the accuracy will be very low: AUC of BM25 is just 0.55 (very close to random guess!)

(Dataset: production pClick training data set with 15M ad impression, ~2M query event)

Relevance Is Not Enough



If directly using relevance score to predict clicks, the accuracy will be very low: AUC of BM25 is just 0.55 (very close to random guess!)

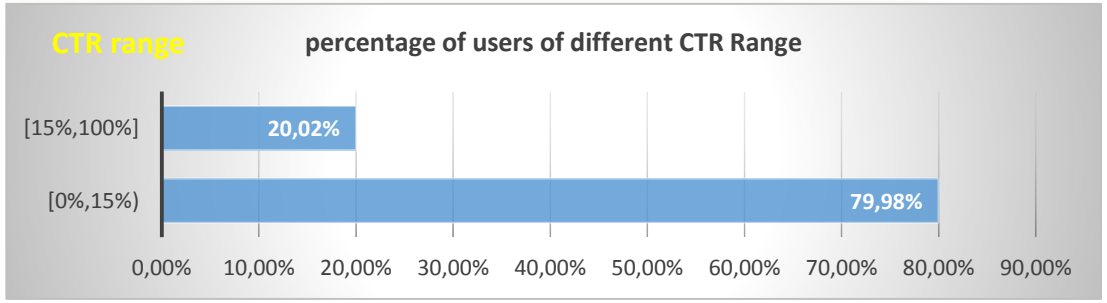
(Dataset: production pClick training data set with 15M ad impression, ~2M query event)

Explanations:

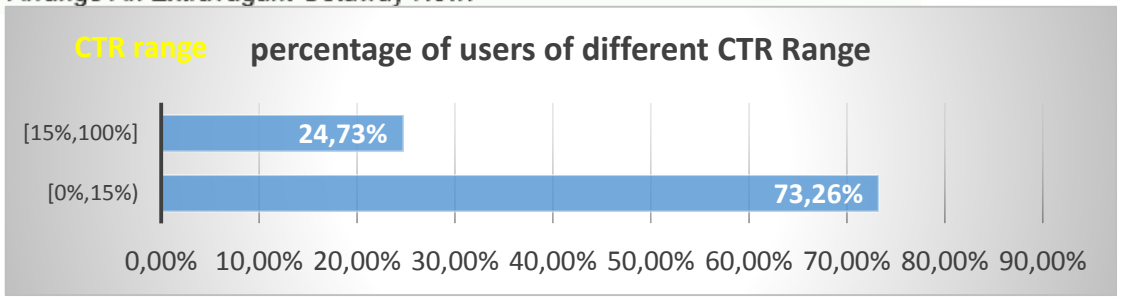
- Many ads, especially for popular queries, yield similar relevance and cannot be well distinguished by their relevance scores.
- Users come to search engine for info but not for ads (ads are pushed to them). Therefore users do not necessarily click on relevant ads because ads are not what they actively look for at all.

Other Click ↛ I Click

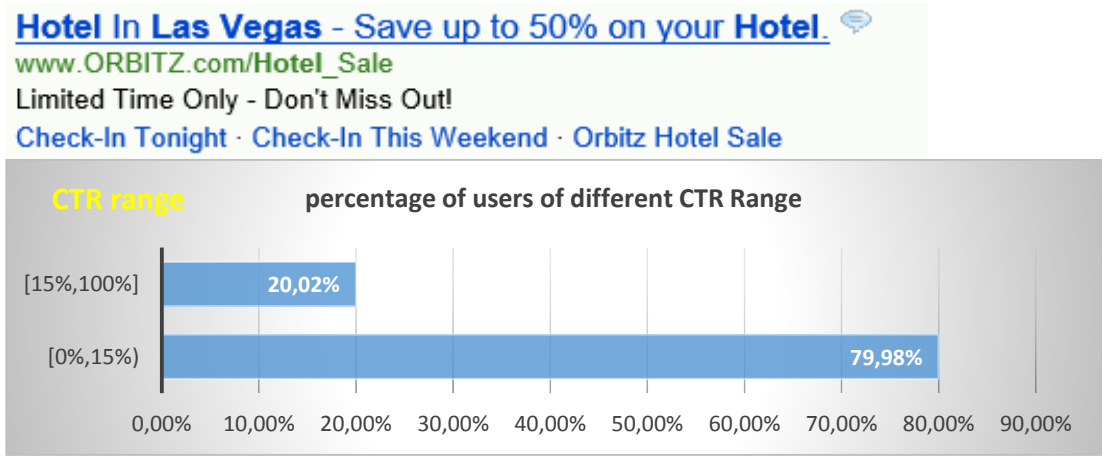
Hotel In **Las Vegas** - Save up to 50% on your **Hotel**.
www.ORBITZ.com/**Hotel**_Sale
Limited Time Only - Don't Miss Out!
Check-In Tonight · Check-In This Weekend · Orbitz Hotel Sale

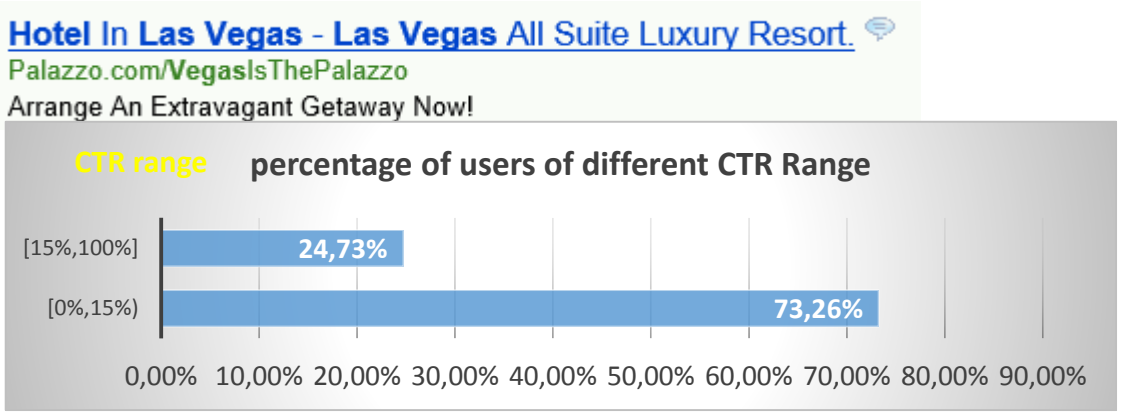


Hotel In **Las Vegas** - **Las Vegas** All Suite Luxury Resort.
Palazzo.com/**Vegas**IsThePalazzo
Arrange An Extravagant Getaway Now!



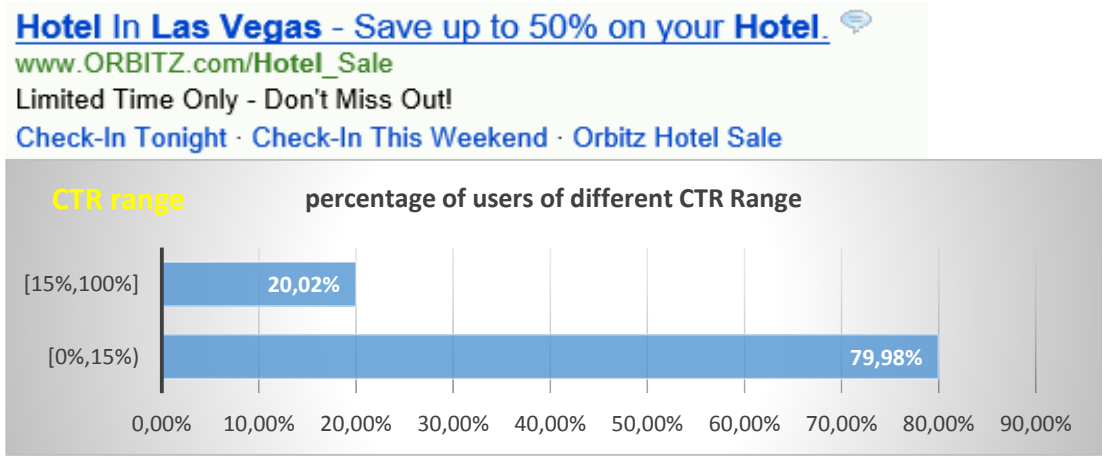
Other Click ↛ I Click

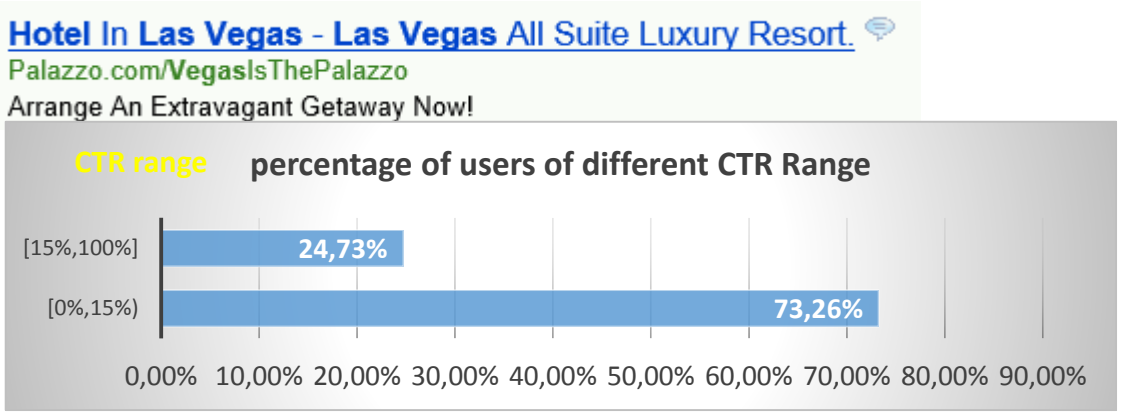




Given the same query-ad pair (with the same historical click counts), the variance of different users' CTRs can be very large.
Historical click counts, which reflect the mean of these CTRs, will lead to a significant prediction error, which corresponds to the variance of these CTRs.

Other Click ↛ I Click





Given the same query-ad pair (with the same historical click counts), the variance of different users' CTRs can be very large.
Historical click counts, which reflect the mean of these CTRs, will lead to a significant prediction error, which corresponds to the variance of these CTRs.

Explanations

- Users are not identical: their click behaviors (preferences on the same ad) are highly diverse, personal, and time-varying.
- One person may not necessarily click on an ad even if many other people click on it.

Then, What Else Should We Use?

- To answer ***why*** users click
  - Characterize the motivation of clicks beyond relevance
  - Distinguish diverse click behaviors of different users
- Our proposal
  - Leverage the study on user behaviors in the literature of behavior economics and advertising, and attribute the motivation of user clicks to the satisfaction of their

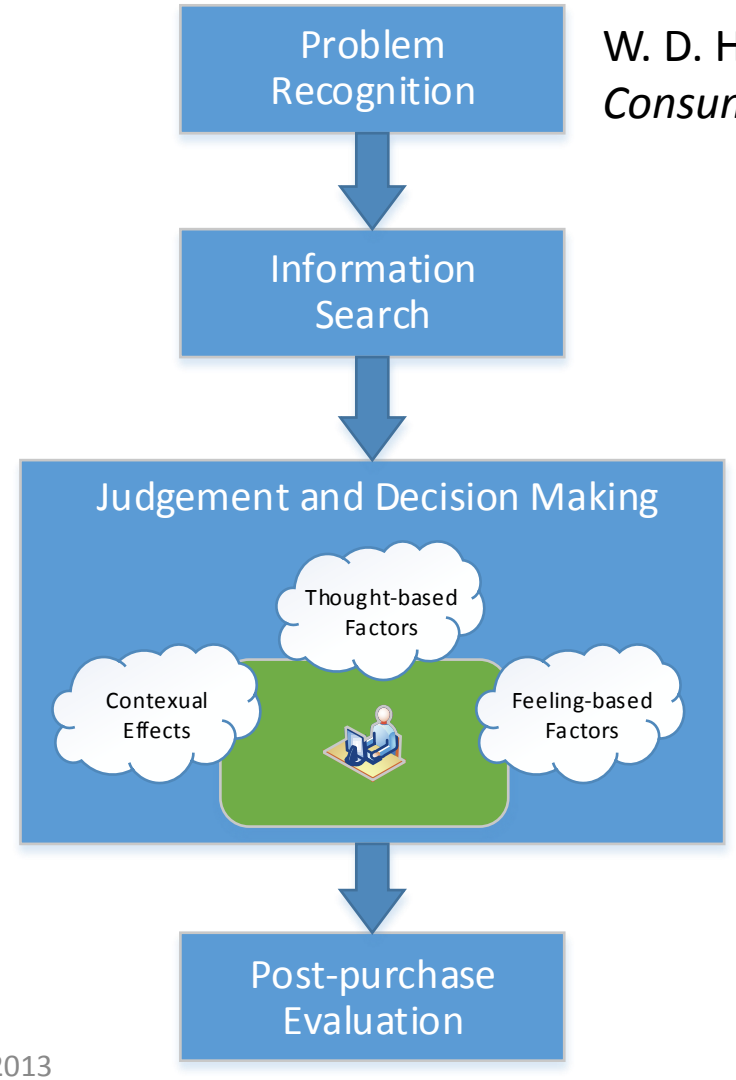
"Psychological Desire"

Psychological factors is one of the central design principles in the advertising industry

Roadmap

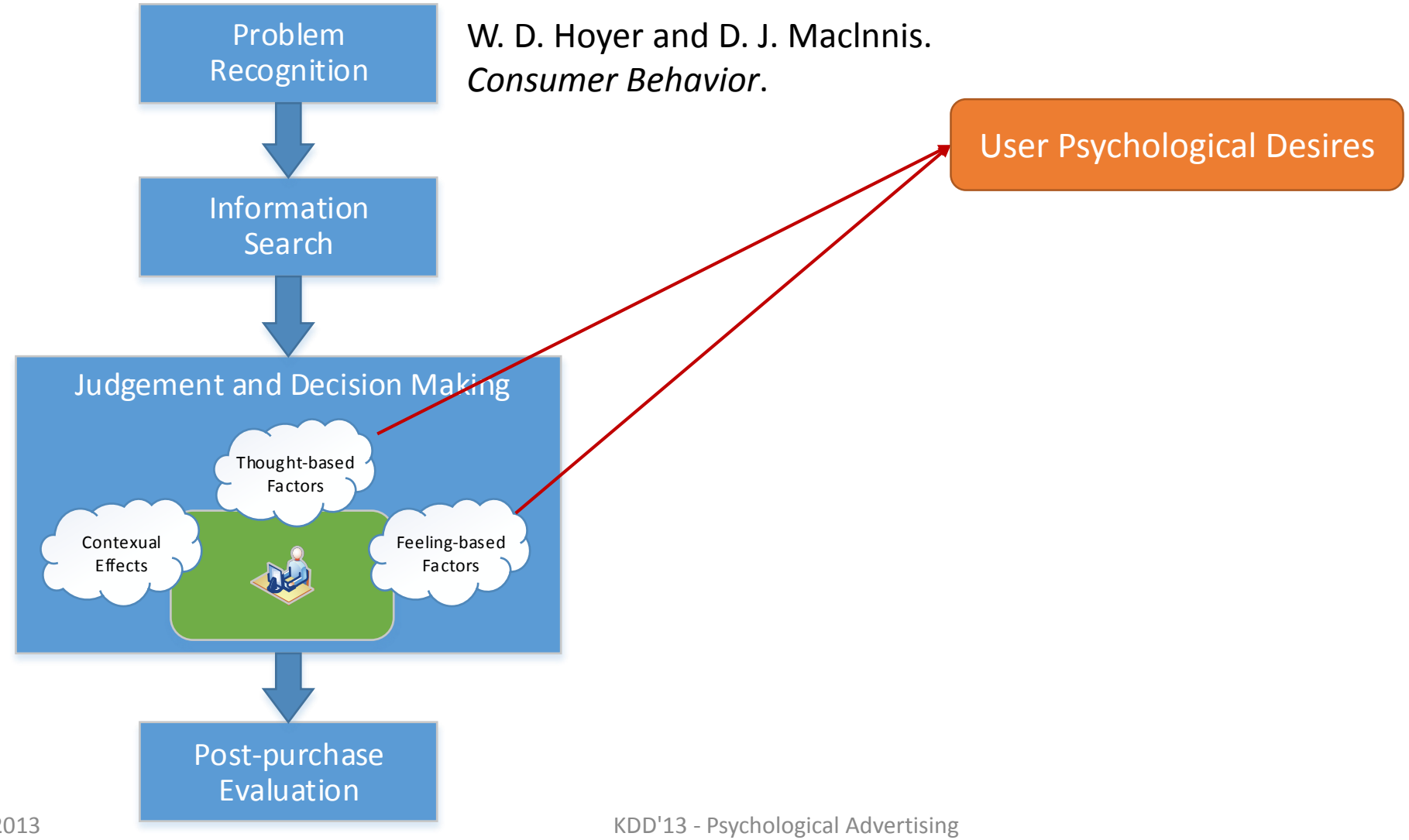
- Motivations
- Data Analysis on User Psychological Desires
- Discovering User Psychological Desire from Ads
- Click Prediction Modeling
- Experimental Evaluations

Psychological Desire in Consumer Decision Making Process



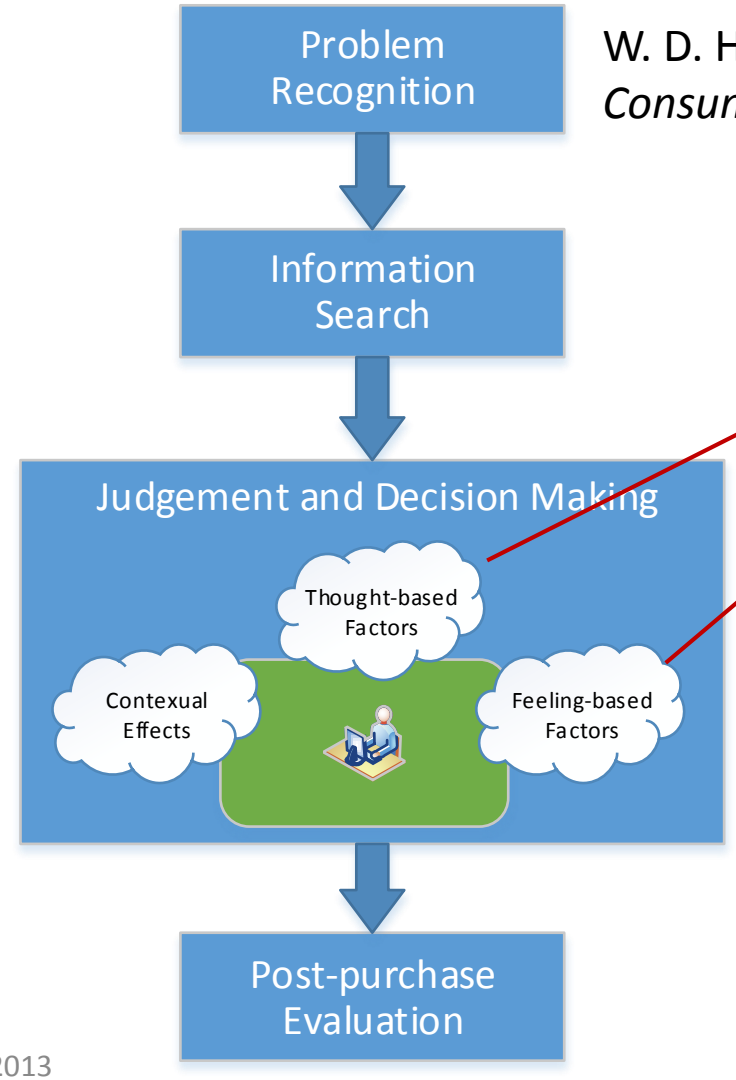
W. D. Hoyer and D. J. MacInnis.
Consumer Behavior.

Psychological Desire in Consumer Decision Making Process



W. D. Hoyer and D. J. MacInnis. *Consumer Behavior*.

Psychological Desire in Consumer Decision Making Process

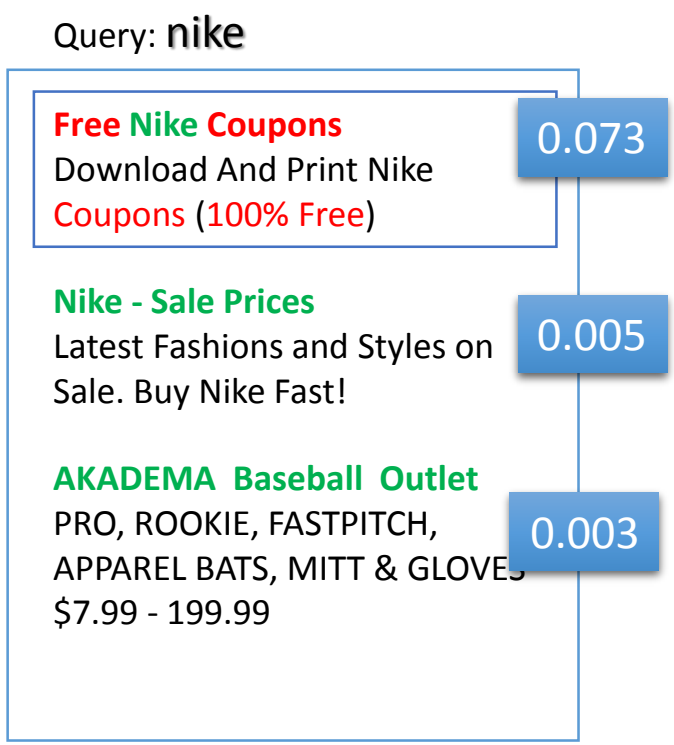


W. D. Hoyer and D. J. MacInnis. *Consumer Behavior*.

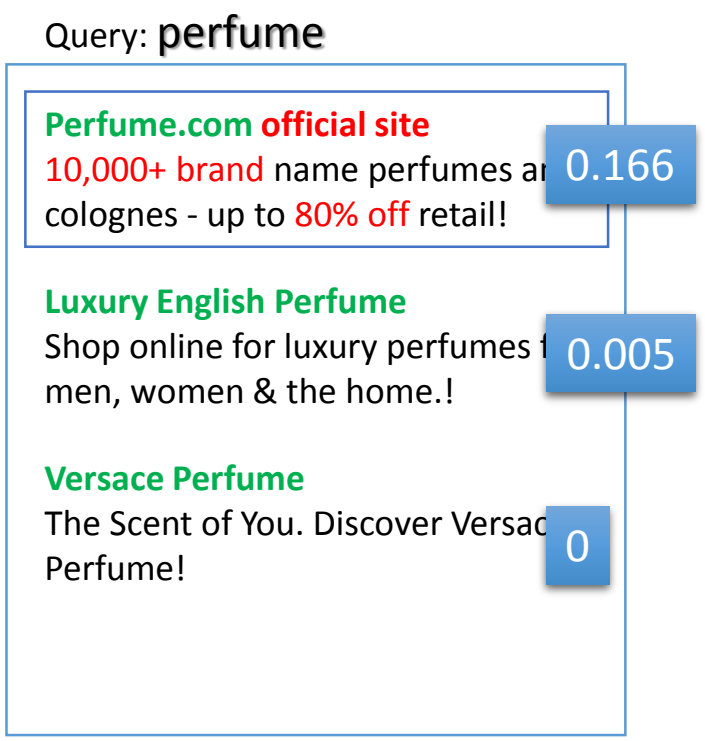
User Psychological Desires

Effects	User Desires	Examples
Thought-based (quantifiable)	Petty advantage	*big discount, good deal, coupon*
	Quantity/quality advantage	*Popular brand, large selection space*
	Extra convenience	*Quick delivering, flexible payment method*
Feeling-based (subjective)	Trustworthy	*Official seller, service with guarantee*
	Brand loyalty	*Ebay, Amazon*
	Luxury seeking	*First class brank*

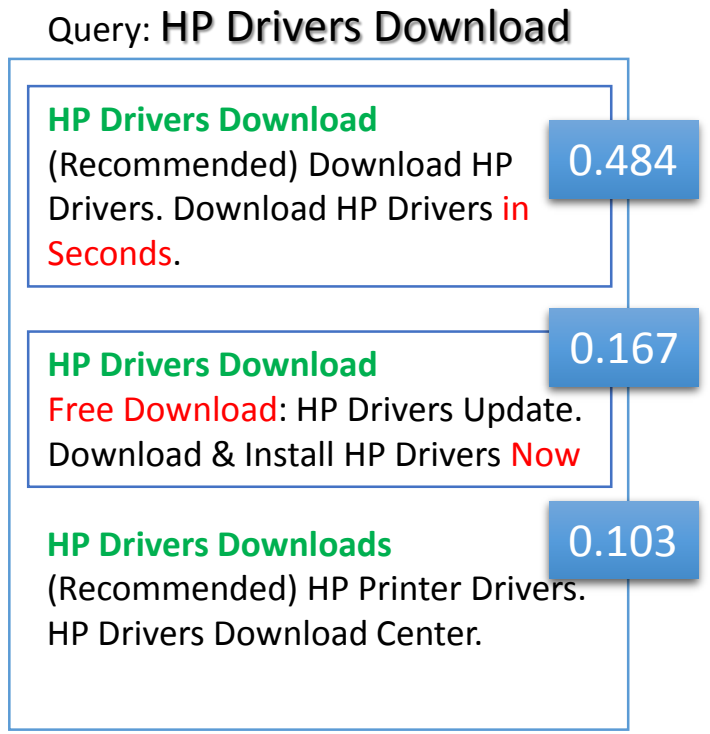
Psychological Desire in Ads Copy



Cheap/free, try to get something with less cost.

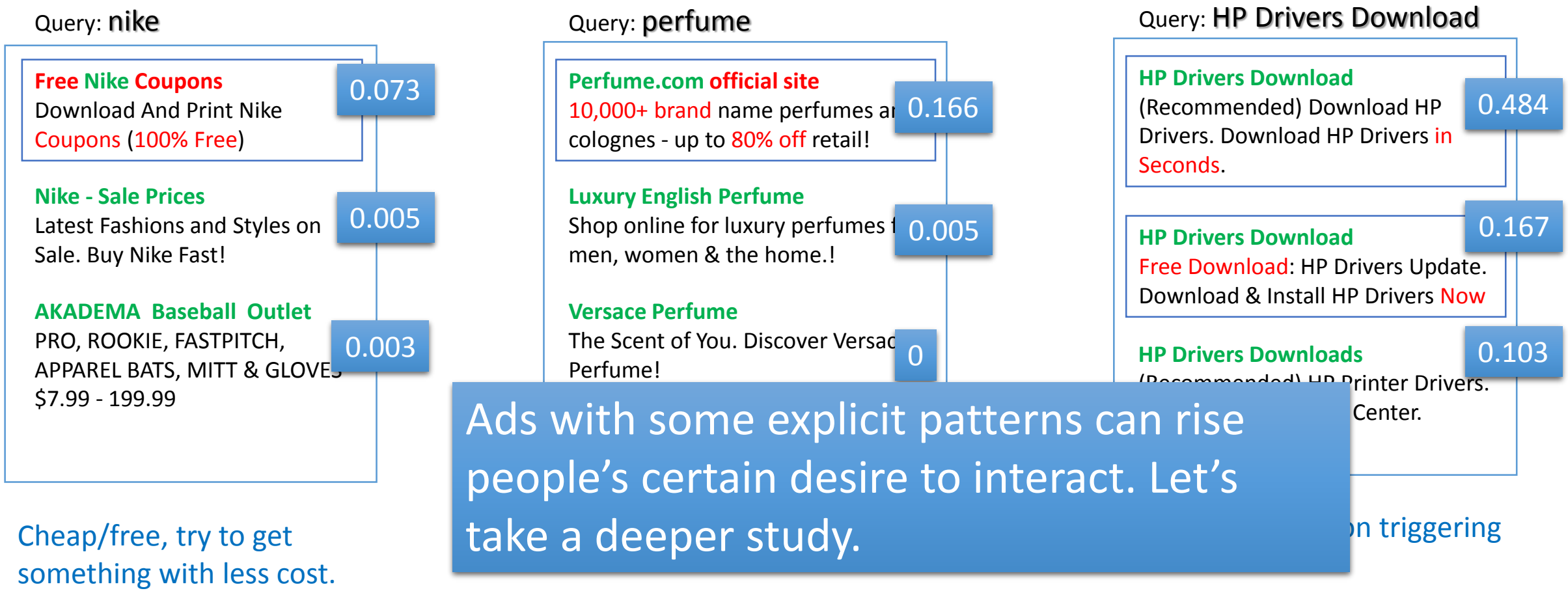


Authoritative and Trustable



Quick effect, action triggering

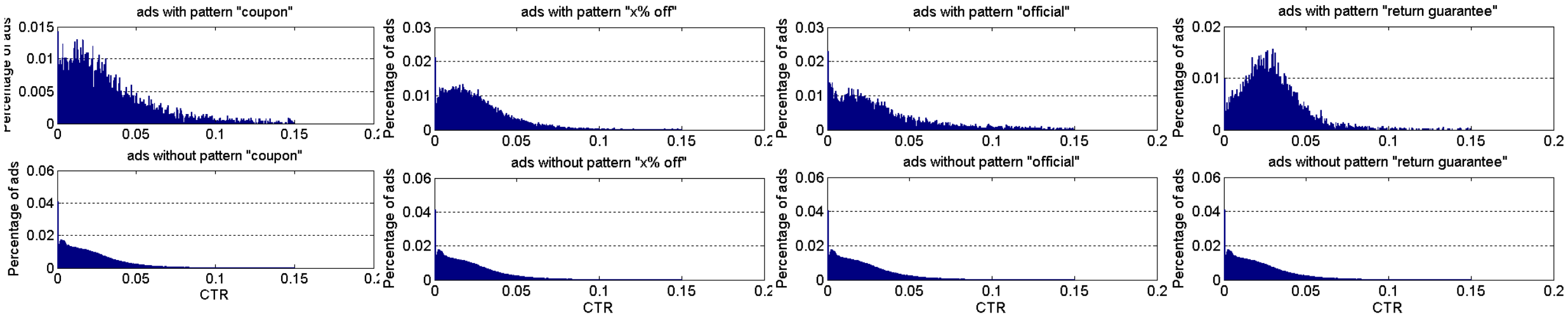
Psychological Desire in Ads Copy



Effects of User Psychological Desire

- CTR difference between the ads matched with user desire pattern and overall ads

	Desire pattern	Coverage of ads	CTR change
Thought-based	*"coupon"*	2.2%	+47.5%
	"x% off"	4.1%	+19.7%
Feeling-based	*"official"*	2.6%	+25.0%
	"return guarantee"	1.9%	+31.4%



Roadmap

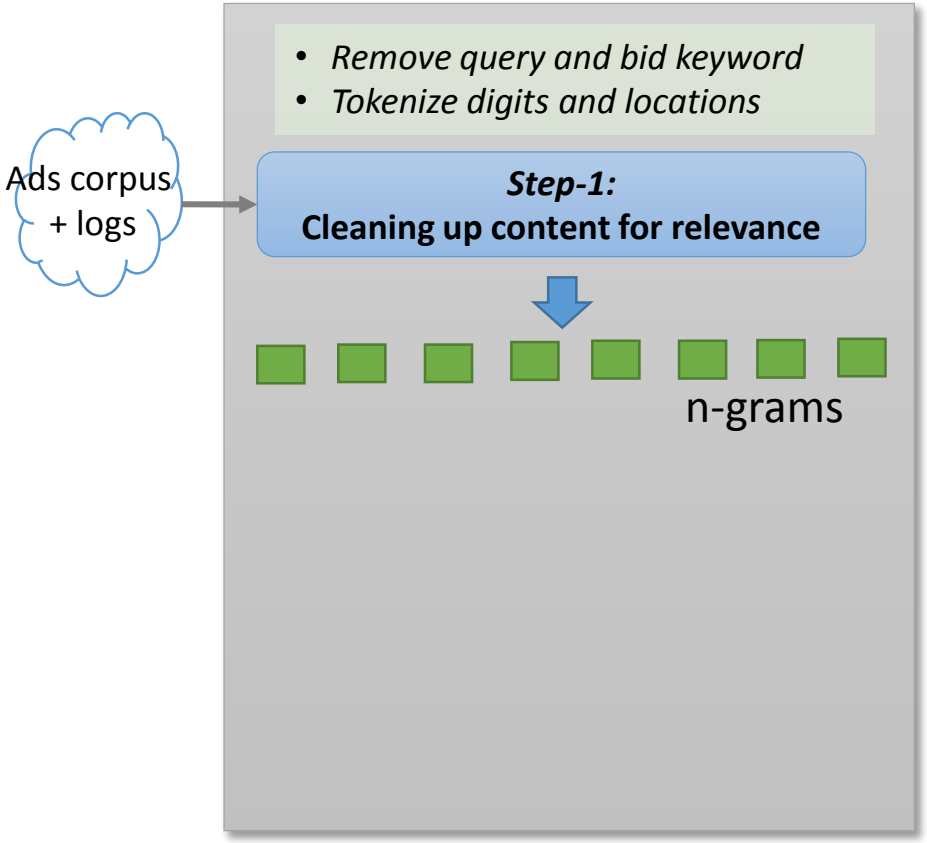
- Motivations
- Data Analysis on User Psychological Desires
- Discovering User Psychological Desire from Ads
- Click Prediction Modeling
- Experimental Evaluations

Discovering User Psychological Desire

- Principles to extract content reflecting user desires from ad texts:
  - Ad content = relevance + desire
  - Reflecting user desires in terms of CTR difference
  - Enough volume coverage in ad traffic
  - Generalized into ***user desire pattern –*** similar content reflecting the specific same desire
  - Content from experienced advertisers are more important

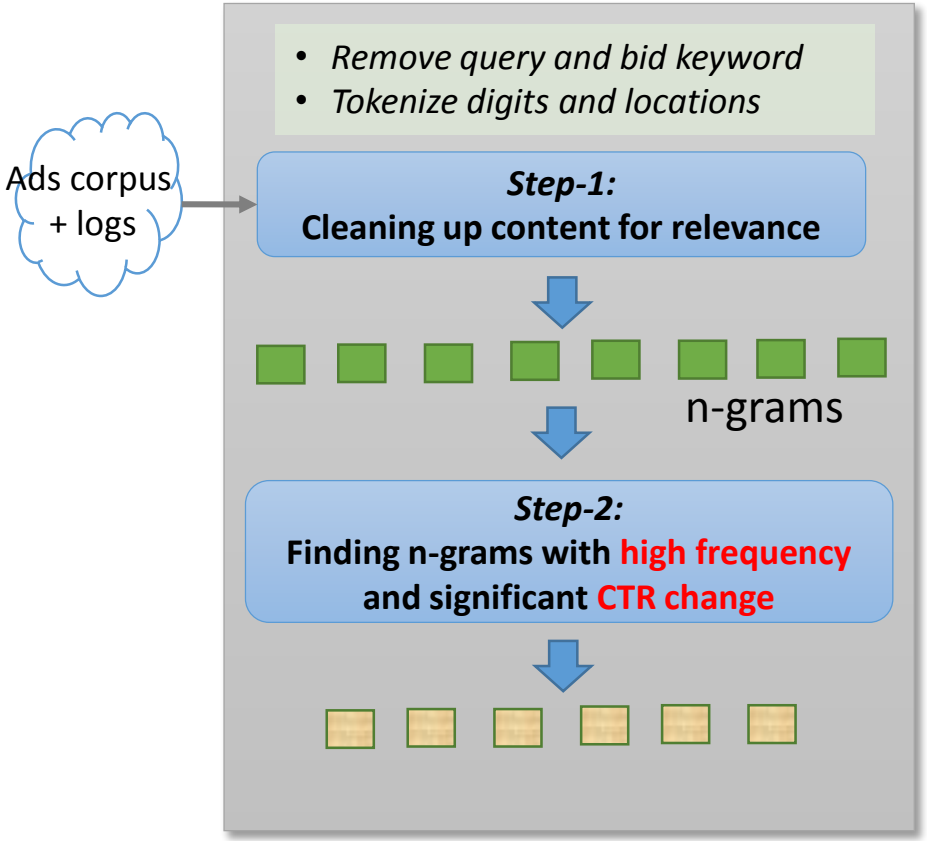
Discovering User Psychological Desire

- Principles to extract content reflecting user desires from ad texts:
  - Ad content = relevance + desire
  - Reflecting user desires in terms of CTR difference
  - Enough volume coverage in ad traffic
  - Generalized into ***user desire pattern –*** similar content reflecting the specific same desire
  - Content from experienced advertisers are more important



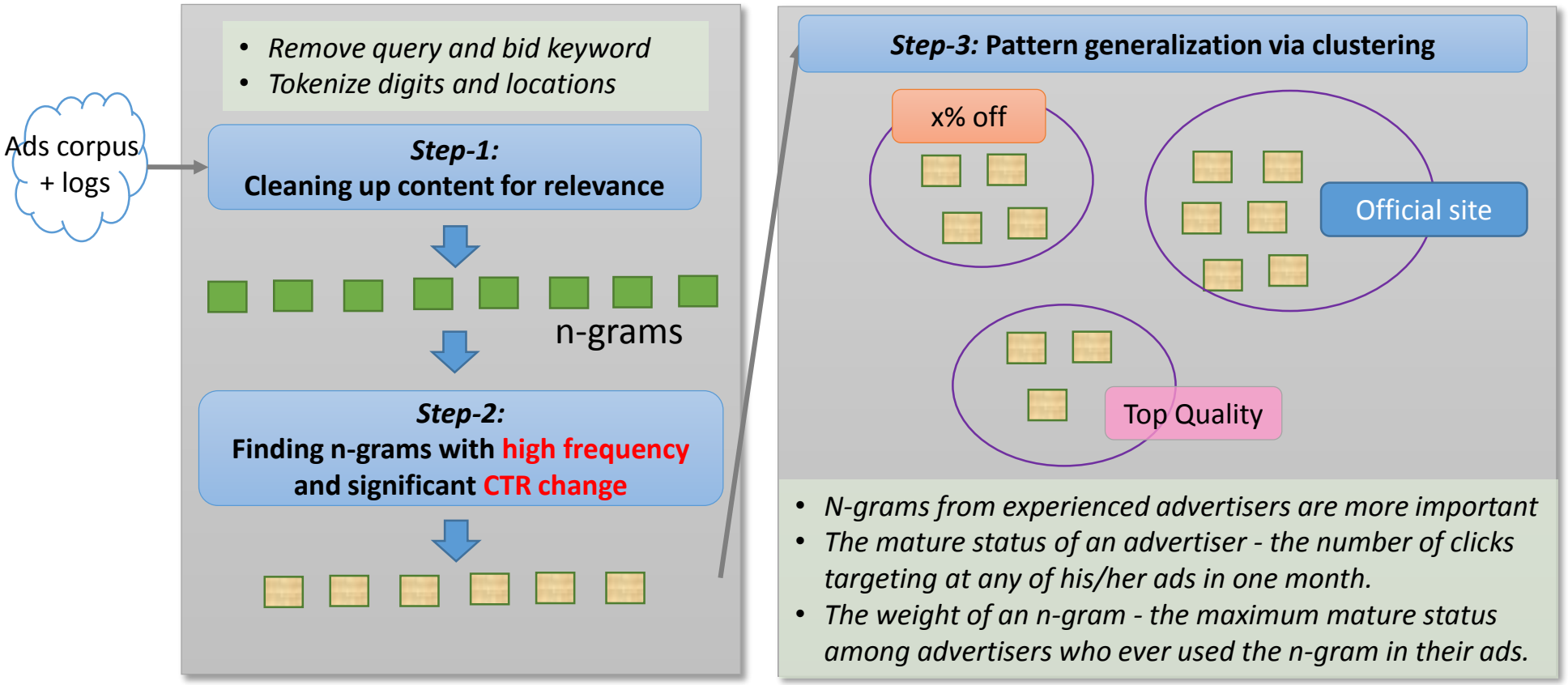
Discovering User Psychological Desire

- Principles to extract content reflecting user desires from ad texts:
  - Ad content = relevance + desire
  - Reflecting user desires in terms of CTR difference
  - Enough volume coverage in ad traffic
  - Generalized into ***user desire pattern –*** similar content reflecting the specific same desire
  - Content from experienced advertisers are more important



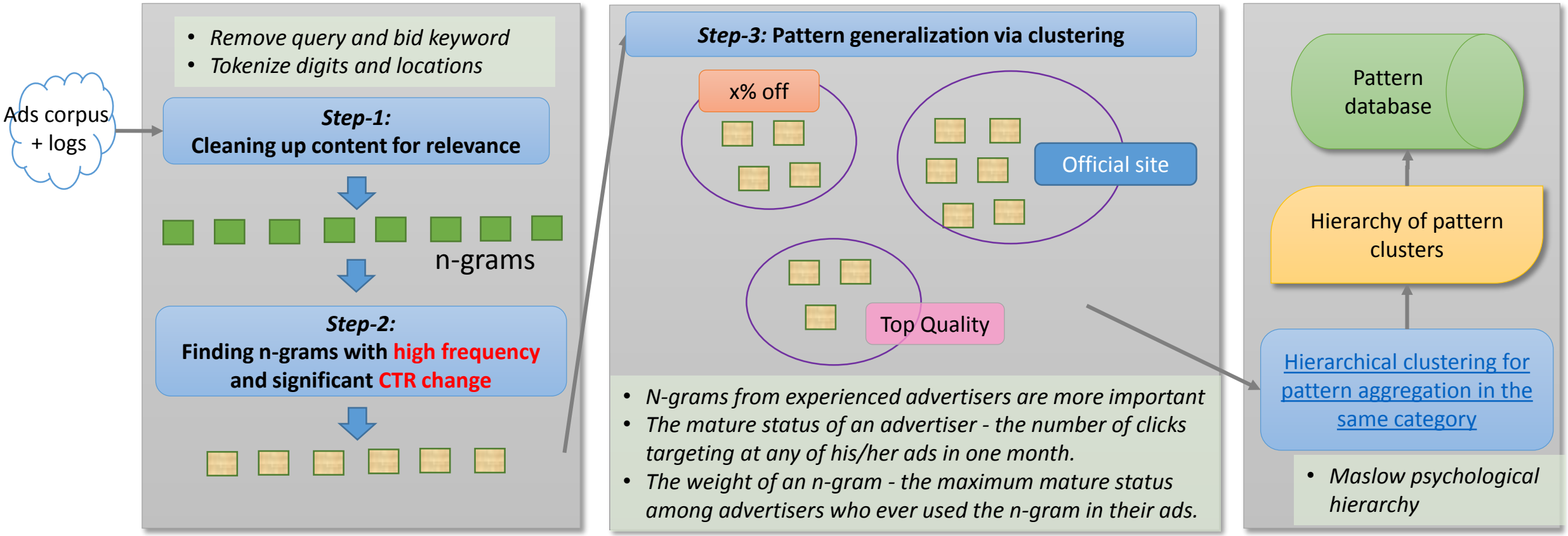
Discovering User Psychological Desire

- Principles to extract content reflecting user desires from ad texts:
  - Ad content = relevance + desire
  - Reflecting user desires in terms of CTR difference
  - Enough volume coverage in ad traffic
  - Generalized into ***user desire pattern –*** similar content reflecting the specific same desire
  - Content from experienced advertisers are more important

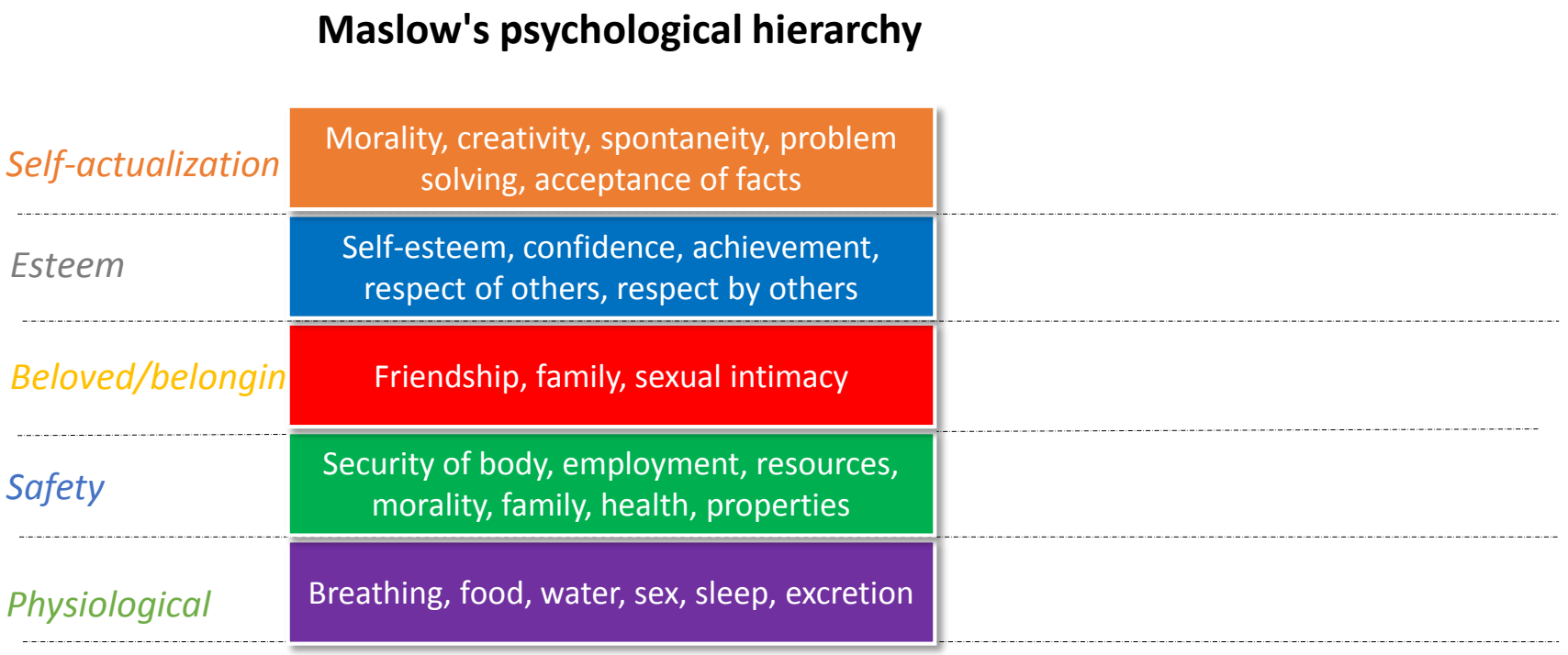


Discovering User Psychological Desire

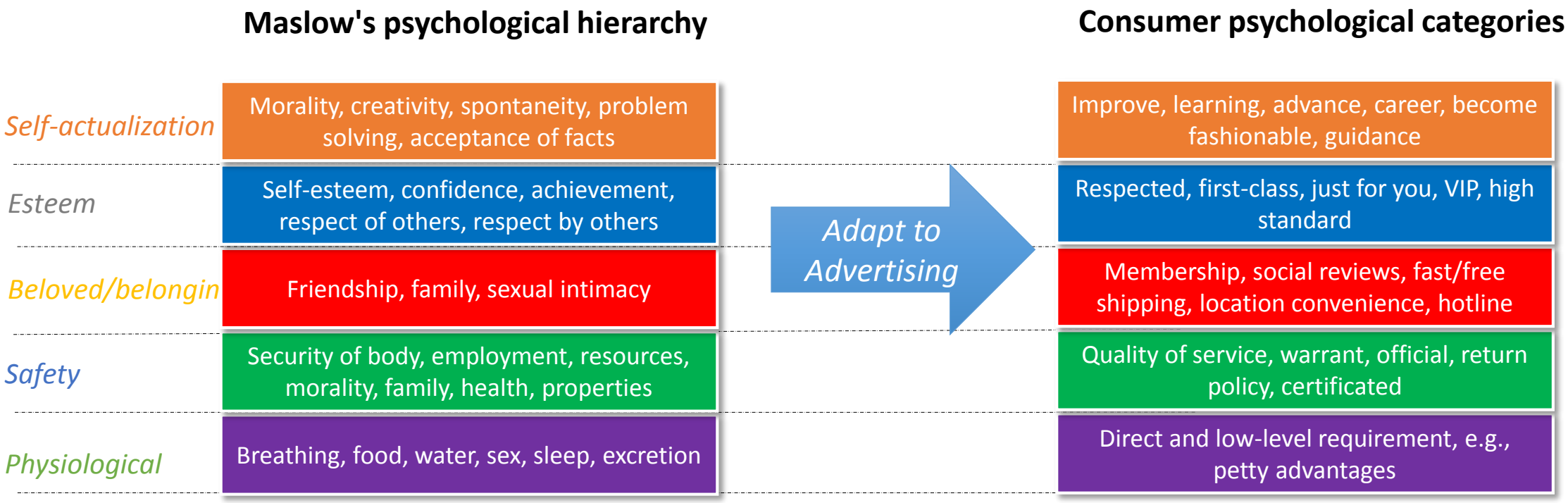
- Principles to extract content reflecting user desires from ad texts:
  - Ad content = relevance + desire
  - Reflecting user desires in terms of CTR difference
  - Enough volume coverage in ad traffic
  - Generalized into ***user desire pattern –*** similar content reflecting the specific same desire
  - Content from experienced advertisers are more important



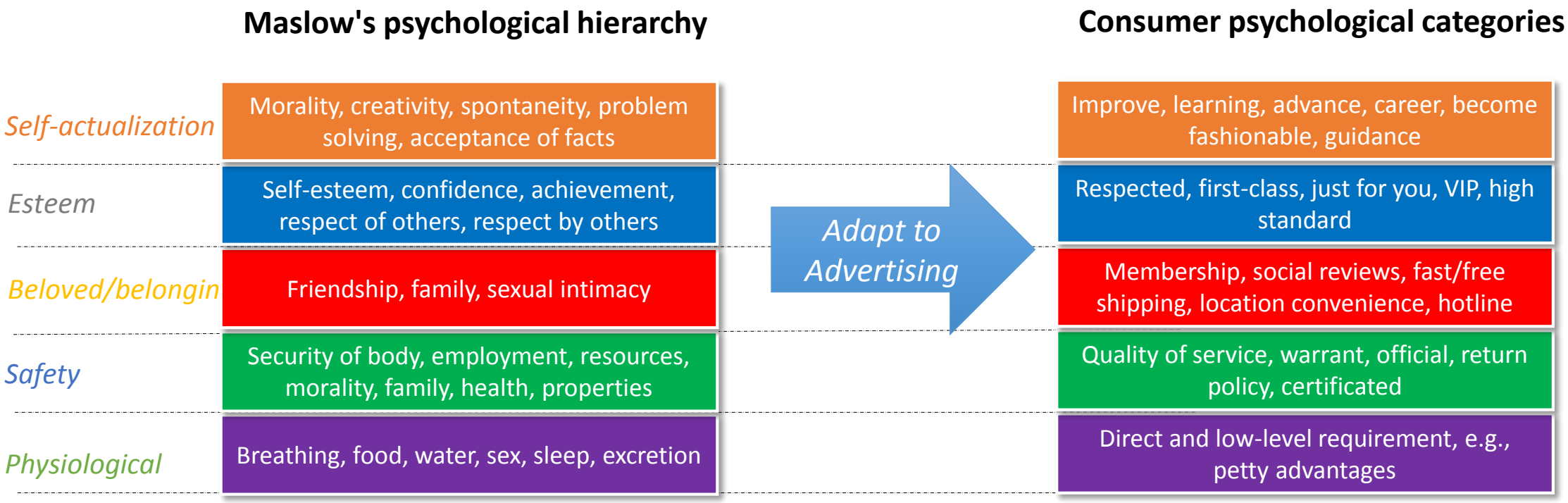
Hierarchy of User Psychological Desire



Hierarchy of User Psychological Desire

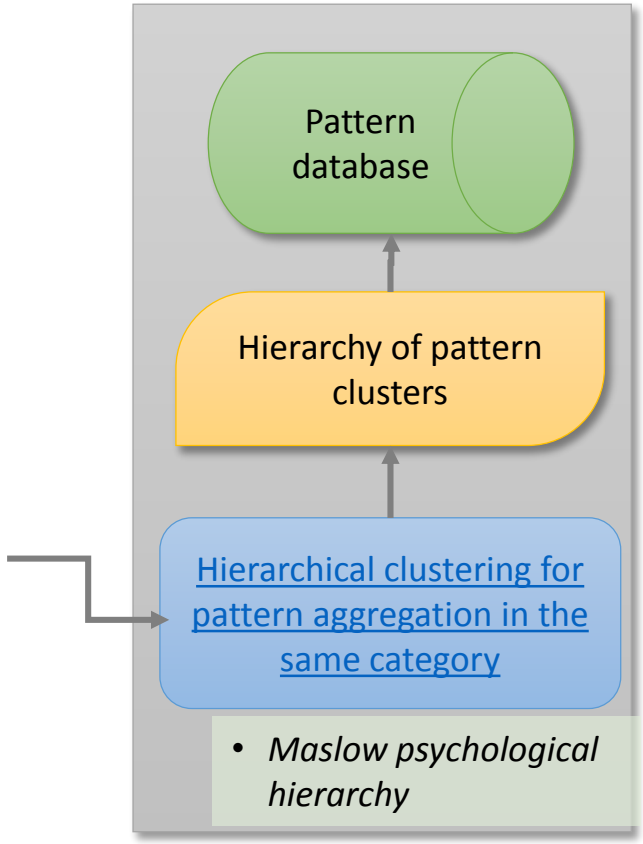


Hierarchy of User Psychological Desire

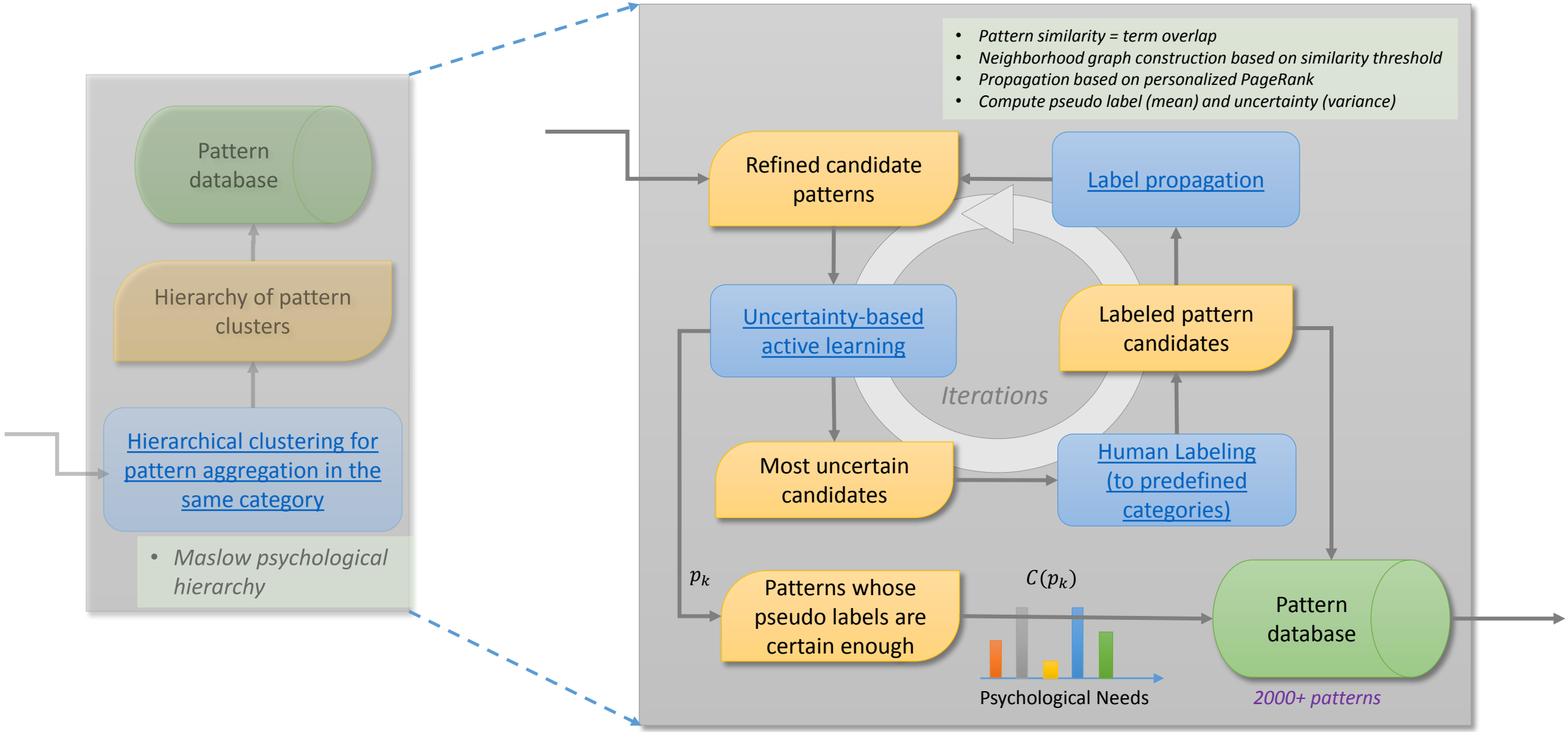


People in different situations (belonging to different financial classes, with different ages, in different regions, in different social communities, facing different products) may belong to different psychological categories.

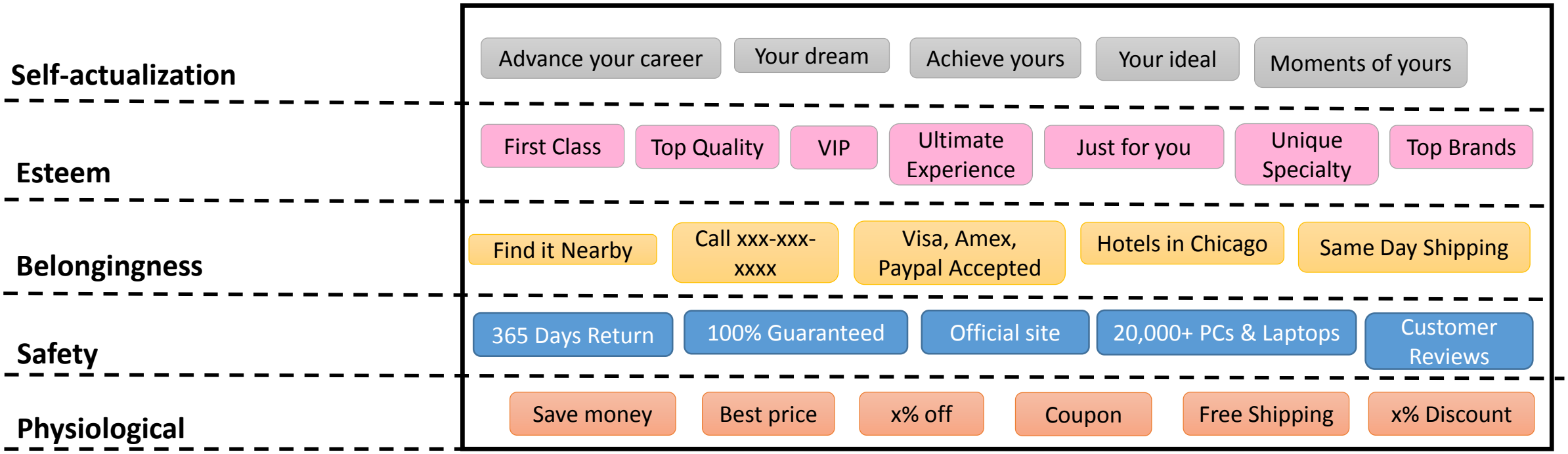
Active Psychological Needs Mining



Active Psychological Needs Mining



Hierarchy of User Psychological Desire



-- Succeed in reducing the sparsity of user desires for each individual ad

Roadmap

- Motivations
- Data Analysis on User Psychological Desires
- Discovering User Psychological Desire from Ads
- Click Prediction Modeling
- Experimental Evaluations

Click Prediction Modeling

Problem:

Calculate the probability of click $p(c|q$

$$\boldsymbol{w} = \text{argmax}_{\boldsymbol{w}}(\sum_{j=1}^{n} \log\big(p(c_i|q_i, a_i, u_i)\big) + \log(p(\boldsymbol{w})))$$

Click Prediction Modeling

Problem:

Calculate the probability of click $p(c|q$

- The maximum entropy model is well suited since its strength in combing diverse forms of features

$$p(c|q,a,u) = \frac{1}{1+\exp(\sum_{i=1}^{d} \omega_i f_i(q,a,u))}$$

i-th feature

weight

$$\boldsymbol{w} = \text{argmax}_{\boldsymbol{w}}(\sum_{j=1}^{n} \log\big(p(c_i|q_i,a_i,u_i)\big) + \log(p(\boldsymbol{w})))$$

Feature set:

- Relevance features
- Historical click features

References:

- Cheng et al. WSDM 2010'
- Hillard et al. IRJ 2011'
- Richardson et al. WWW 2007'

Click Prediction Modeling

Problem:

Calculate the probability of click $p(c|q$

- The maximum entropy model is well suited since its strength in combing diverse forms of features

$$p(c|q,a,u) = \frac{1}{1+\exp(\sum_{i=1}^{d} \omega_i f_i(q,a,u))}$$

i-th feature

weight

- The maximum entropy model learns the weight vector $\boldsymbol{w}$ by maximizing the likelihood of exponential models as:

$$\boldsymbol{w} = \text{argmax}_{\boldsymbol{w}}(\sum_{j=1}^{n} \log\big(p(c_i|q_i,a_i,u_i)\big) + \log(p(\boldsymbol{w})))$$

Feature set:

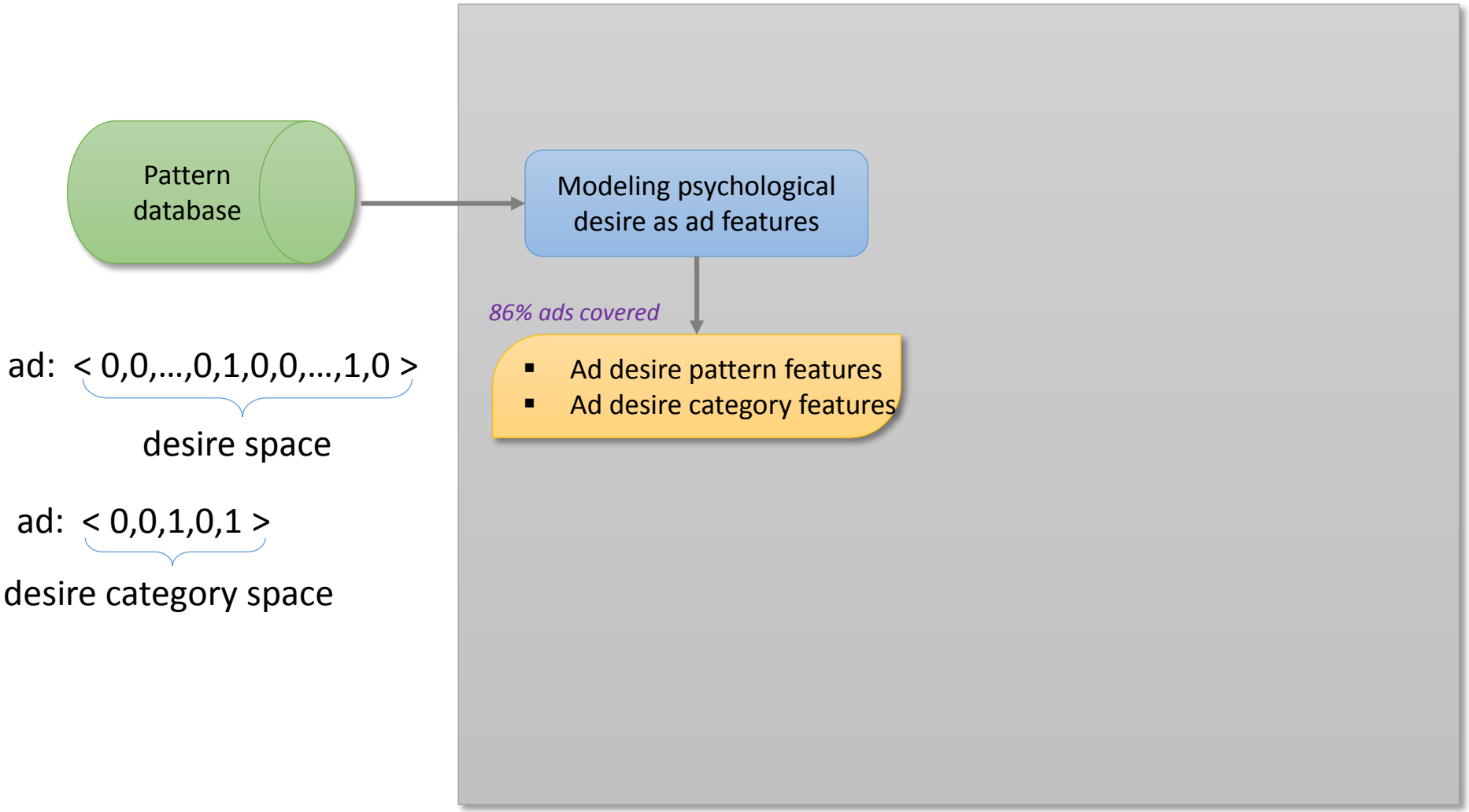
- Relevance features
- Historical click features

References:

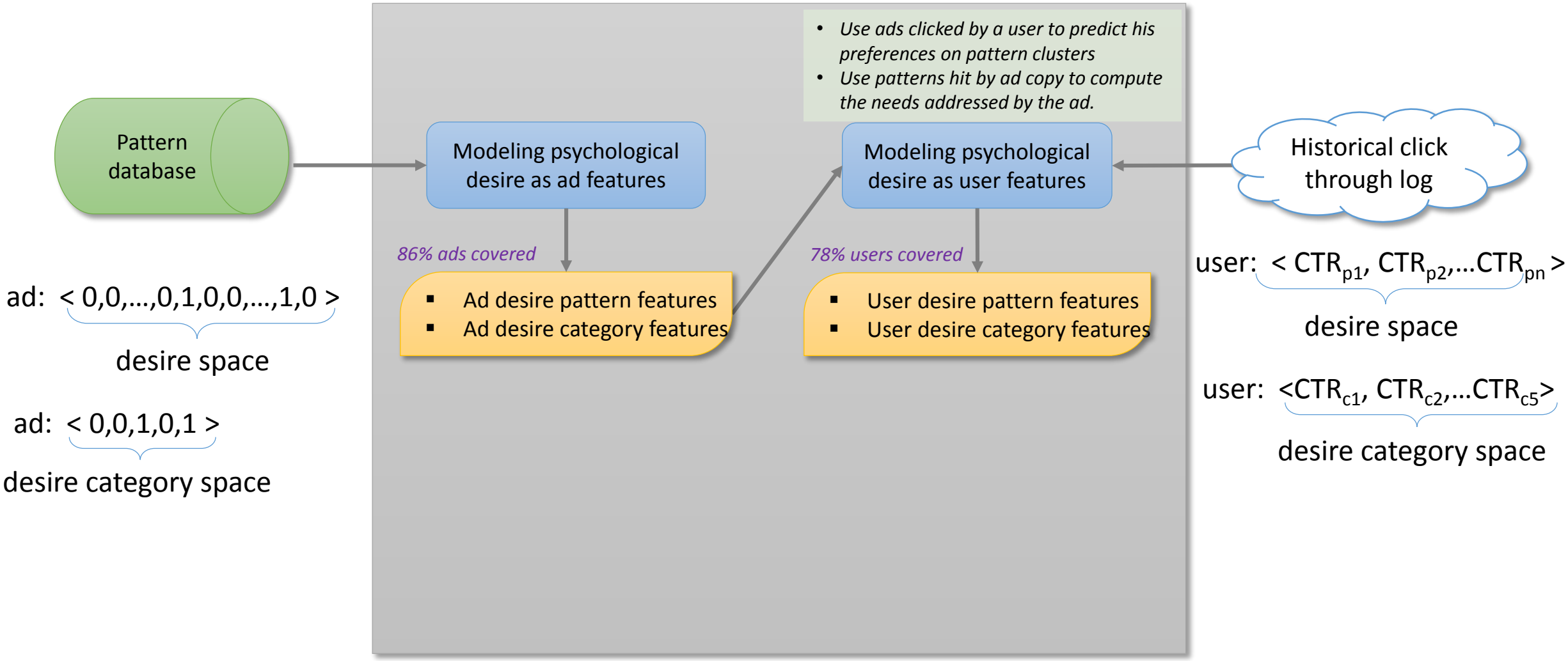
- Cheng et al. WSDM 2010'
- Hillard et al. IRJ 2011'
- Richardson et al. WWW 2007'

Integrating User Psychological Desires into Click Prediction

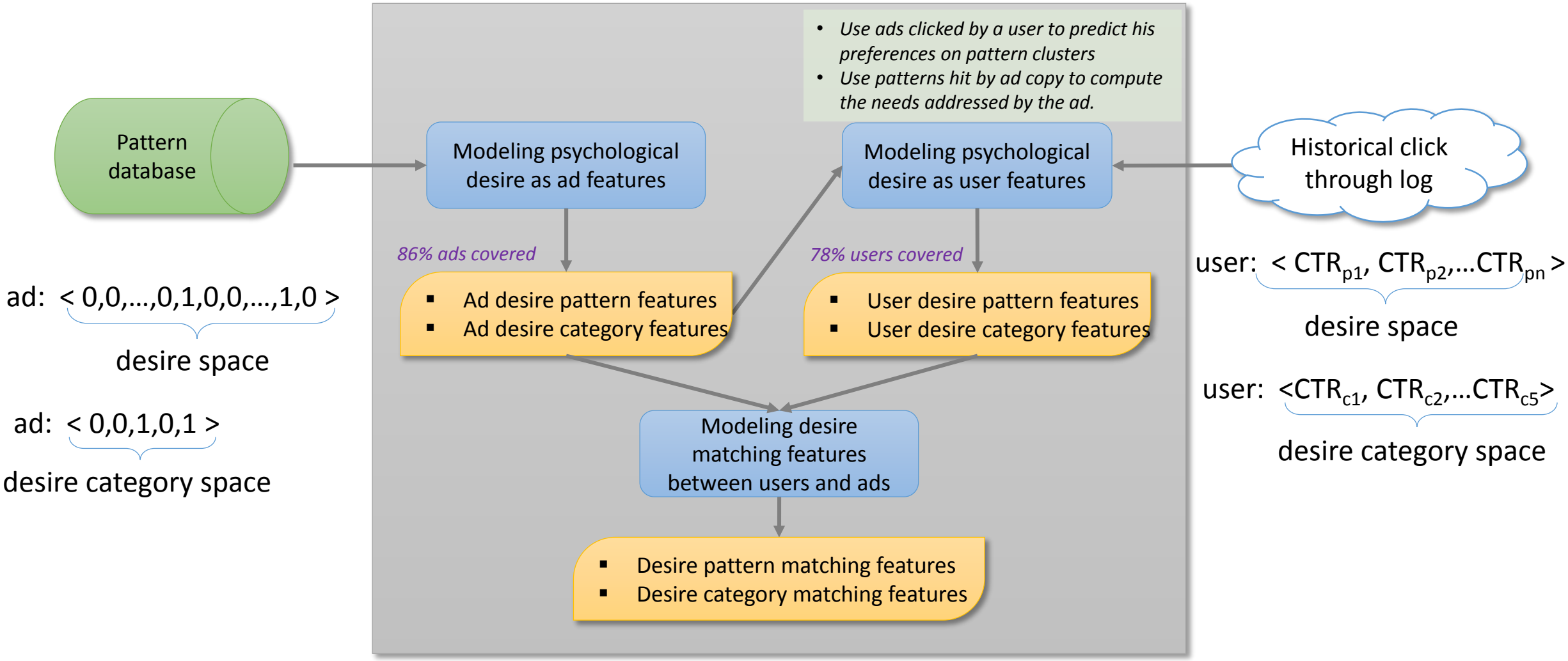
Integrating User Psychological Desires into Click Prediction



Integrating User Psychological Desires into Click Prediction



Integrating User Psychological Desires into Click Prediction



Roadmap

- Motivations
- Data Analysis on User Psychological Desires
- Discovering User Psychological Desire from Ads
- Click Prediction Modeling
- Experimental Evaluations

Experimental Settings

- Data set:
  - A sample of click-through logs in a two-week period from a commercial search engine

	Ad impressions	# of unique ads	# of unique queries
Training (1st week)	20.8M	4.3M	2.6M
Testing (2nd week)	19.8M	5.3M	2.5M

- Compared Methods

	Relevance features	Historical click features	Desire pattern features	Desire category features
HF (base)		X		
HF-RF	X	X		
HF-DPF		X	X	
HF-DPLF		X	X	X
HF-RF-DPF	X	X	X	
HF-RF-DPLF	X	X	X	X

Experimental Settings

- Evaluation Metrics
  - Relative Information Gain (RIG) (T. Graepel, et al. ICML 2010)

$$\mathrm{RIG} = \frac{\mathrm{LogScore+Entropy(CTR)}}{\mathrm{Entropy(CTR)}} = 1 + \mathrm{NormalizedLogScore}$$

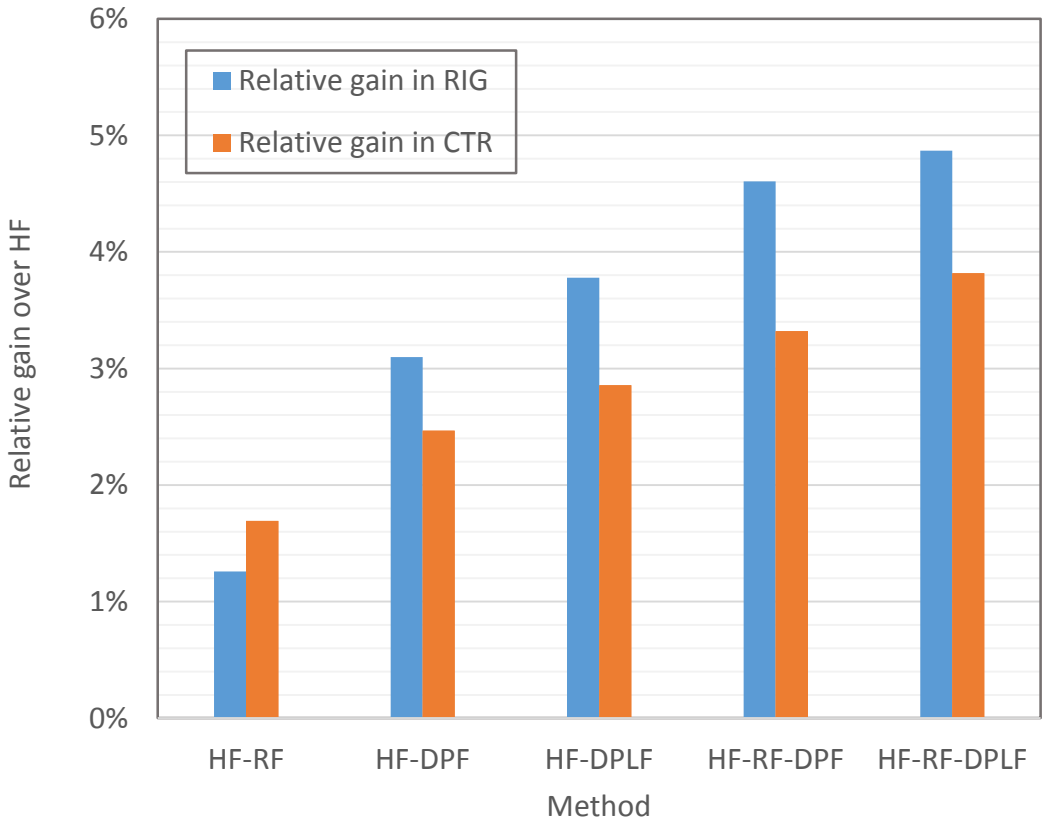
$$\mathrm{LogScore} = \frac{1}{N}\sum_i y_i \ln p_i + (1-y_i)\ln(1-p_i)$$

$$\mathrm{Entropy(CTR)} = -\mathrm{CTR}\ln\mathrm{CTR} - (1-\mathrm{CTR})\ln(1-\mathrm{CTR}),\ \text{with}\ \mathrm{CTR} = \frac{1}{N}\sum_i y_i$$

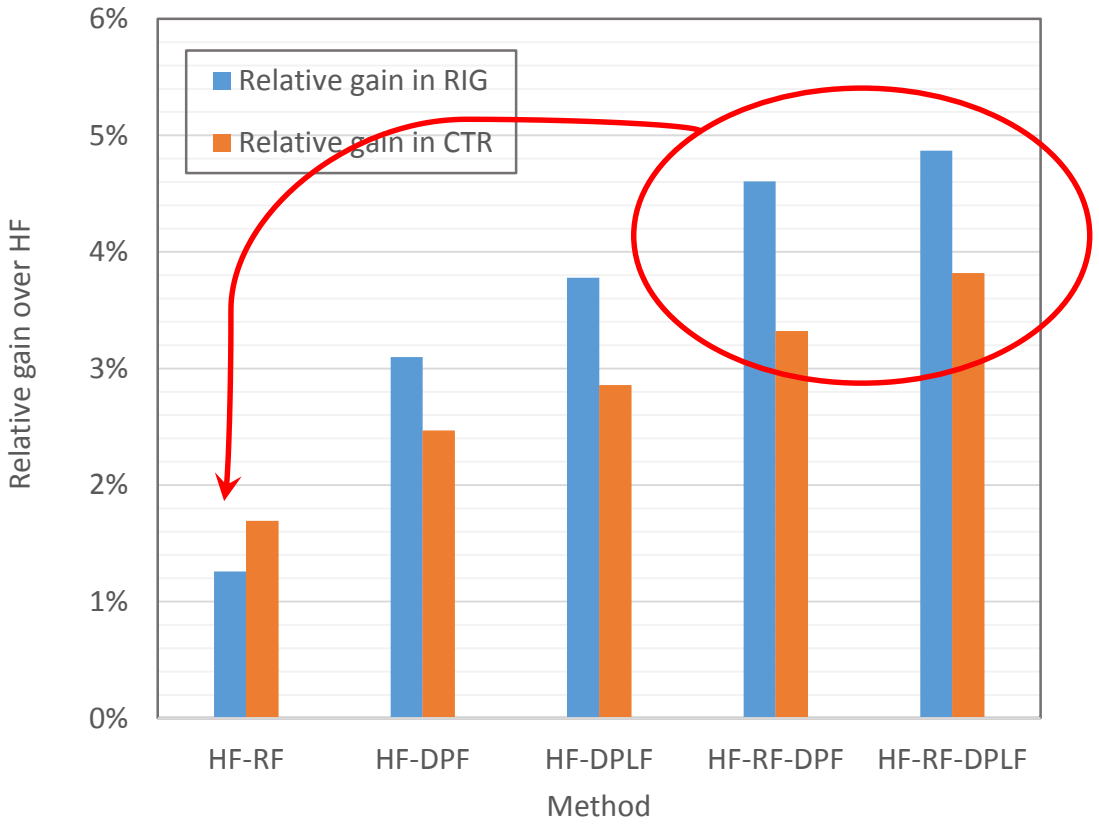
- Where
  - p_i are the pClick scores
  - y_i values are the Click=1/NoClick=0 labels.
    > **RIG@ML-1** and **AllPositionRIG**
  - Since $p_i = 0$ and $p_i = 1$ values can produce infinity values, p_i is clamped to lie between $p_i \in [\varepsilon, 1-\varepsilon]$, where $\varepsilon = 10^{-5}$. Both p_i and c_i values are from the "test" set.

  - Simulated Click-through rate (CTR)
    - Replay-based simulation: re-rank ads and use real clicks as ground-truth

Overall Performance

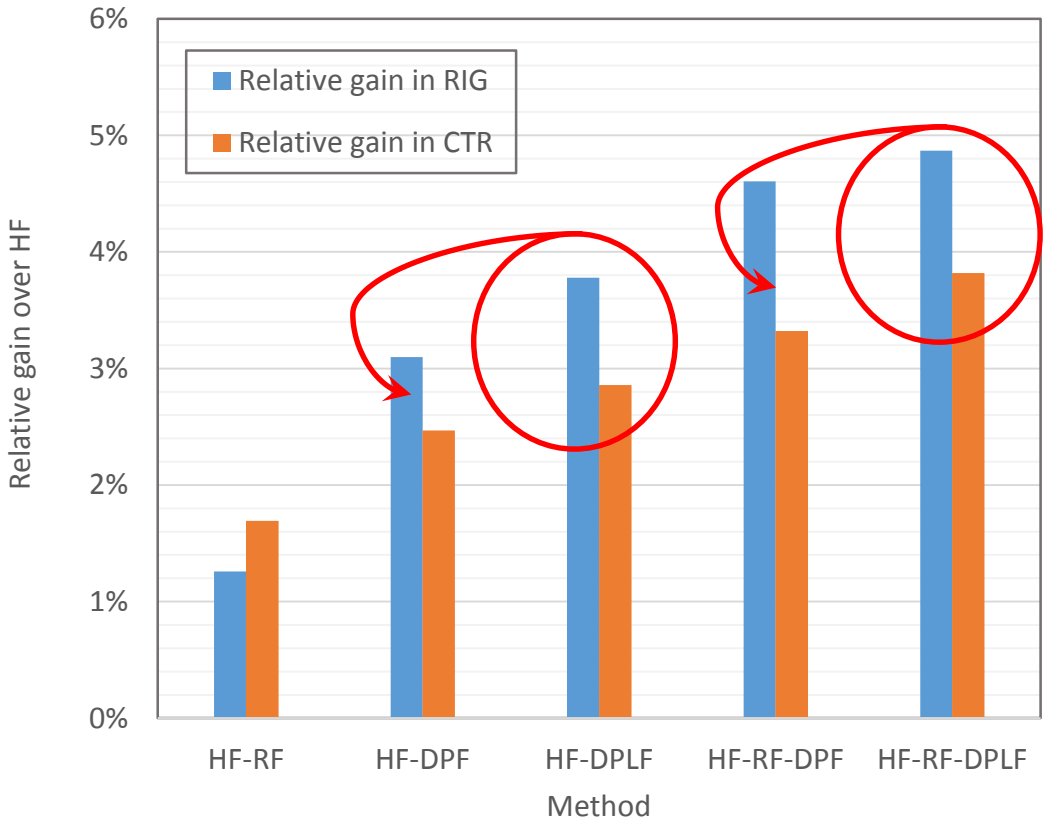


Overall Performance



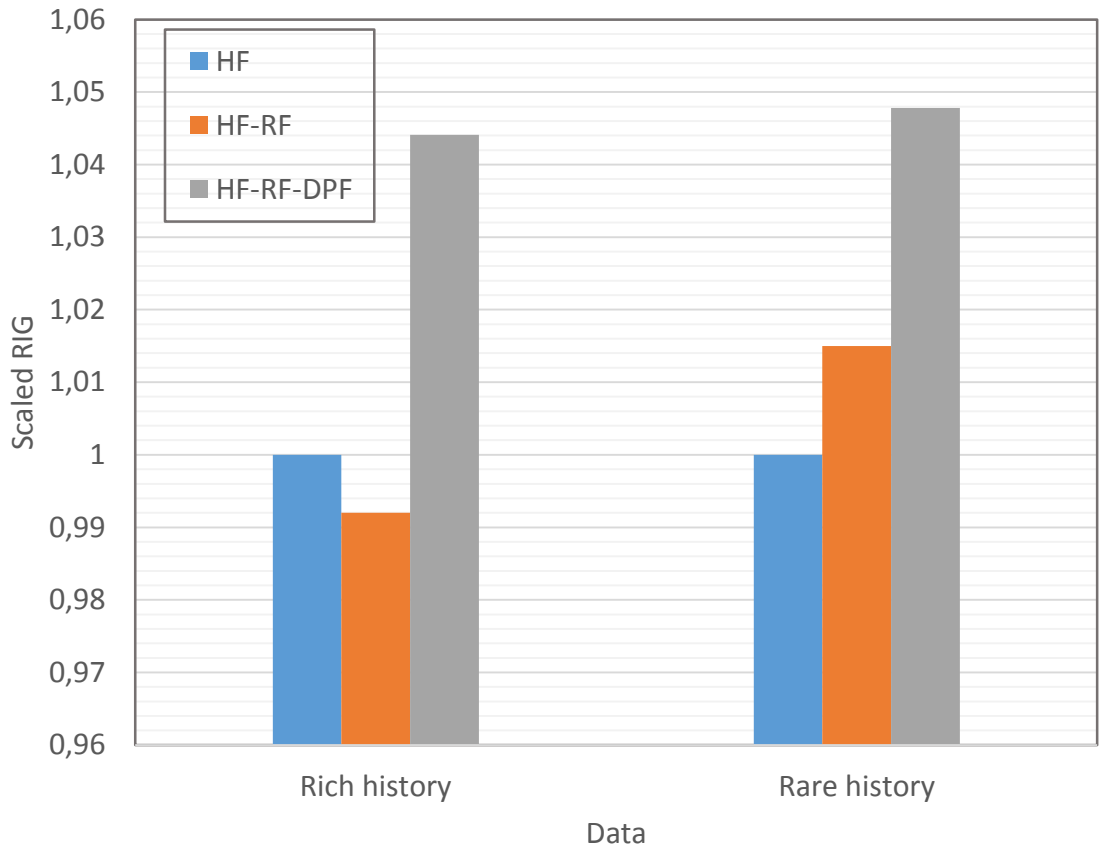
- User desire features can lead to significant improvement on click prediction

Overall Performance



- User desire features can lead to significant improvement on click prediction
- Desire category features are good complements to desire pattern features
  - It can reduce the sparsity of desire features for individual ads

Impacts on Ads with Rich v.s. Rare History



- Rich history set
  - Relevance features may not very beneficial
  - Since advertisers may adjust ad text, it is beneficial to always leverage updated desire patterns for click prediction

- Rare history set:
  - Click prediction is mainly based on understanding users' click intents ***(why)***
  - Relevance features may not indicate users' potential for consuming
  - User psychological desires can more effectively reflect users' potential for consuming

Conclusion and Future Work

- Summary
  - *Asking "why" of user clicks on ads and designing strong features according to the answer*
  - Connect click prediction with user behavior analysis
  - Embrace user psychological desire in ad click prediction
  - Promising experimental results over a large scale data set from real world
- Future Work
  - Context-aware user psychological desire
    - If user desire is dependent with queries or other kinds of search context
  - Structured user desire
    - If the hierarchical relationship of user psychological desire can be leveraged in click prediction
  - Temporal psychological desire
    - Modeling users' temporal psychological desire and detecting their emerging interests in terms of desire at real-time

Microsoft Research

Thanks!

Taifeng Wang[1], Jiang Bian[1], Shusen Liu[2], Yuyu Zhang[3], Tie-Yan Liu[1]

[1]Microsoft Research Asia
[2]South China University of Technology
[3]Chinese Academy of Science