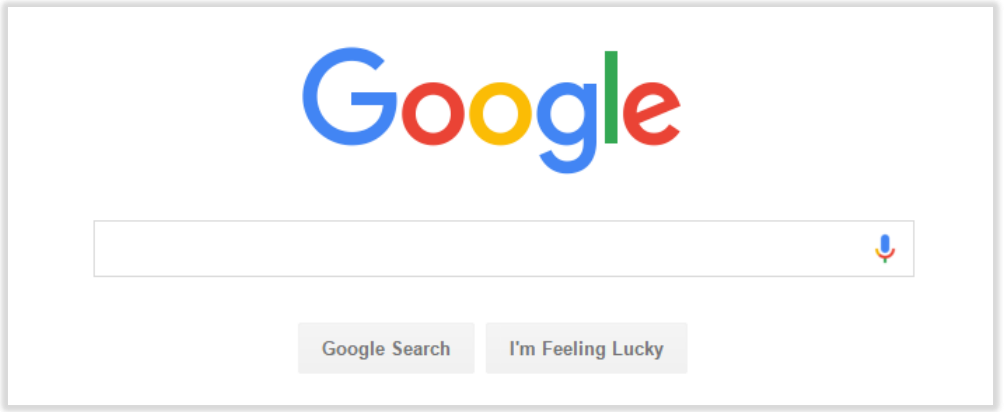
FOCUS: Shedding Light on the High Search Response Time in the Wild

Dapeng Liu, Youjian Zhao, Kaixin Sui, Lei Zou, Dan Pei Qingqian Tao, Xiyang Chen, Dai Tan

清華大學
Tsinghua University

Baidu 百度

Web Search Engines



Search Response Time (SRT)

t_1 A search query is submitted

t_2 The result page Is rendered

$$SRT = t_2 - t_1$$

Search Response Time Matters

Bing

+500ms revenue ↓ 1.2%

[Eric Schurman, Bing]

Google

+100ms~400ms queries ↓ 0.2%~0.6%

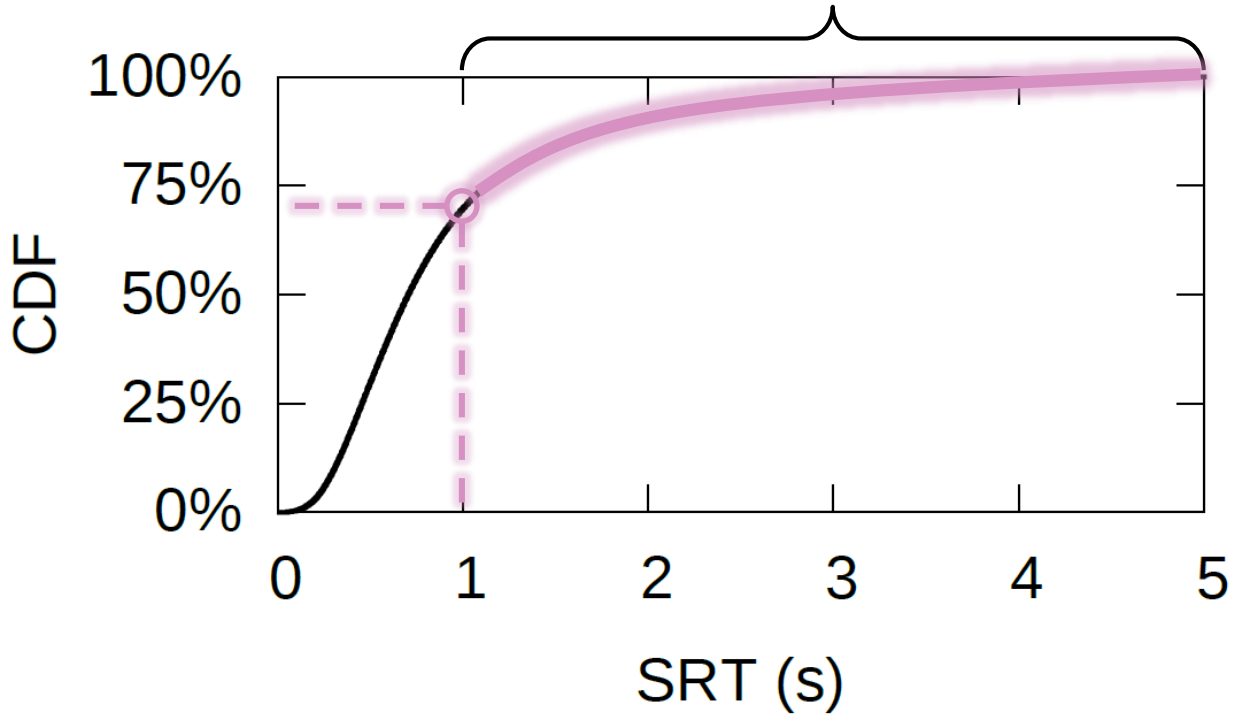
[Jake Brutlag, Google]

Given two content-wise identical search result pages, users are more likely to perform clicks on the fast page

[SIGIR 2014]

Search Response Time in the Wild

User's flow of thought is interrupted
if pages take **longer than 1s** to load



Why?

https://www.nngroup.com/articles/response-times-3-important-limits/

Monitoring SRT: Search Logs

Measurable attributes that can potentially impact SRT

SRT	User's ISP	Browser engine	# of Images	Ads	Server Load	…
800ms (Low SRT)	China Unicom	WebKit	10	Yes	1000 queries/s	…
1200ms (High SRT)	China Telecom	Trident 5.0	5	No	500 queries/s	…
……						

Goal of FOUCS

Measurable attributes that can potentially impact SRT

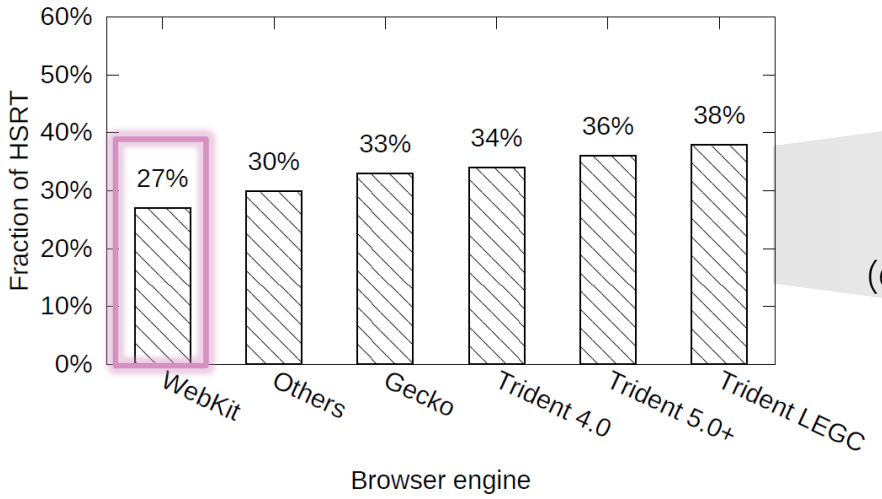
SRT	Client ISP	Browser engine	# of Images	Ads	Server Load	…
800ms (Low SRT)	China Unicom	WebKit	10	Yes	1000 queries/s	…
1200ms (High SRT)	China Telecom	Trident 5.0	5	No	500 queries/s	…
……						

We propose **FOCUS**, a search log analysis system to answer the following questions:

- Under what conditions **HSRT （High SRT）** is more likely to happen?
- Which HSRT conditions are similar (HSRT condition types)?
- How does each attribute affect SRT in HSRT condition types?

Challenges

Limited visibility of naïve single-dimension analysis



What we can see

WebKit is a good condition, where HSRT is **only 27%**

(e.g. used by Chrome and Safari)

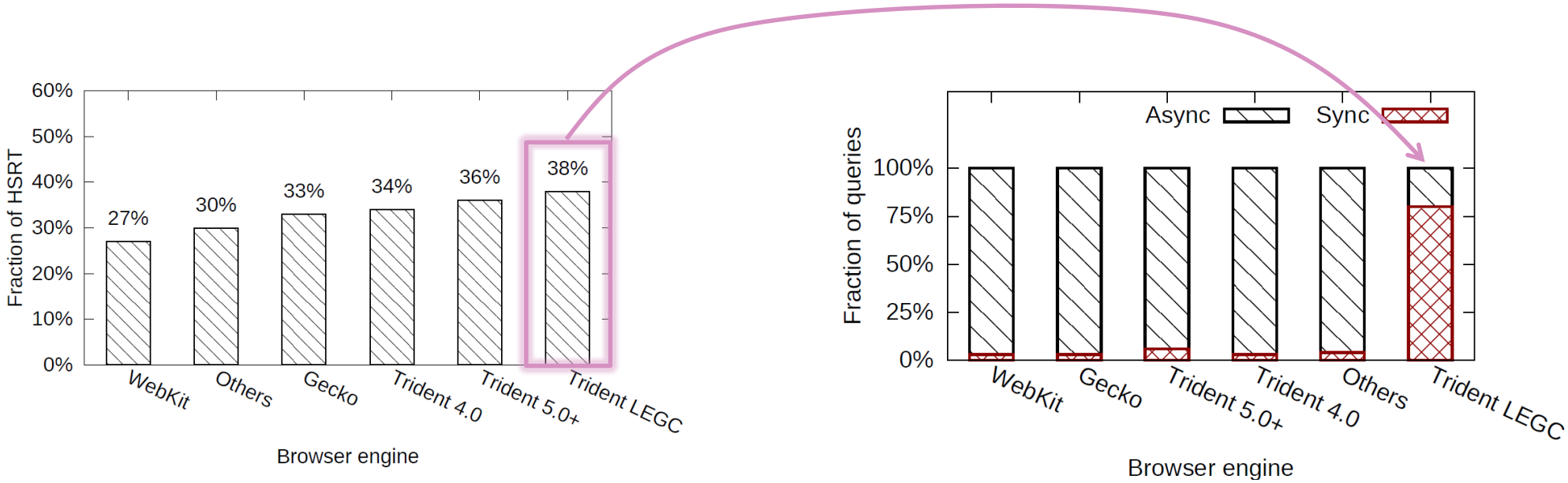
What we **cannot** see

HSRT is **more than 38%**
when **"WebKit + #Images >30"**

Challenges

Limited visibility of naïve single-dimension analysis

Interdependencies between attributes



Which one should be blamed? Legacy Trident or sync page loading?

Challenges

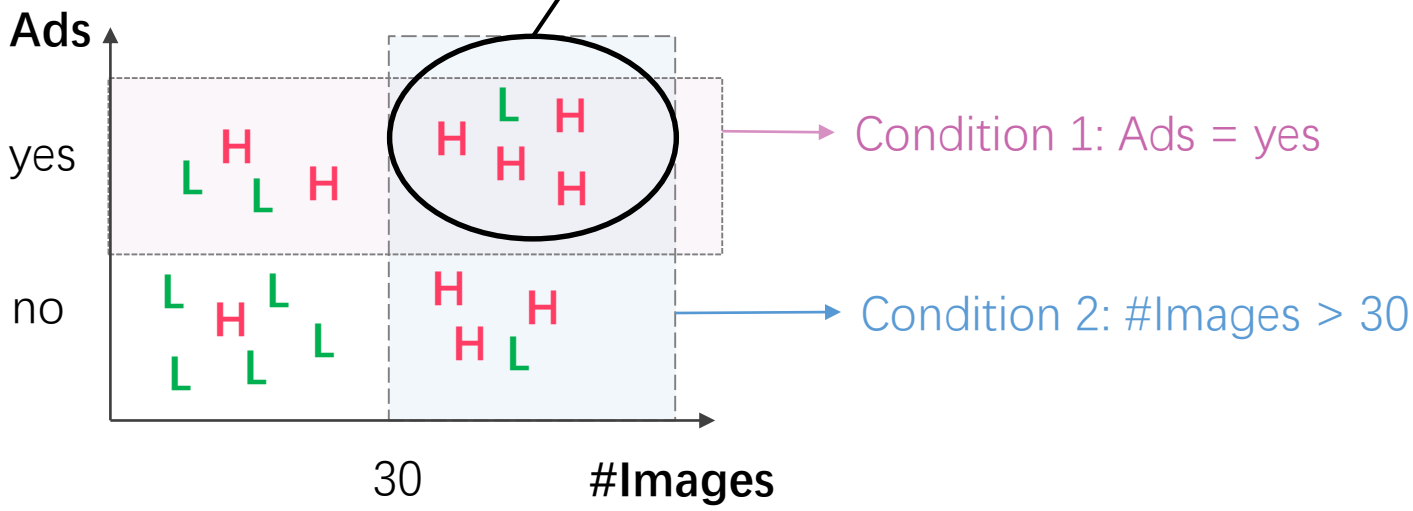
Limited visibility of naïve single-dimension analysis

Interdependencies between attributes

Overlapped HSRT conditions

For example: **H**(igh SRT)
L(ow SRT)

HSRT in the overlapped part will be explained by more than one condition, but which one is better?



Challenges

Limited visibility of naïve single-dimension analysis

Interdependencies between attributes

Overlapped HSRT conditions

Key Idea of FOCUS

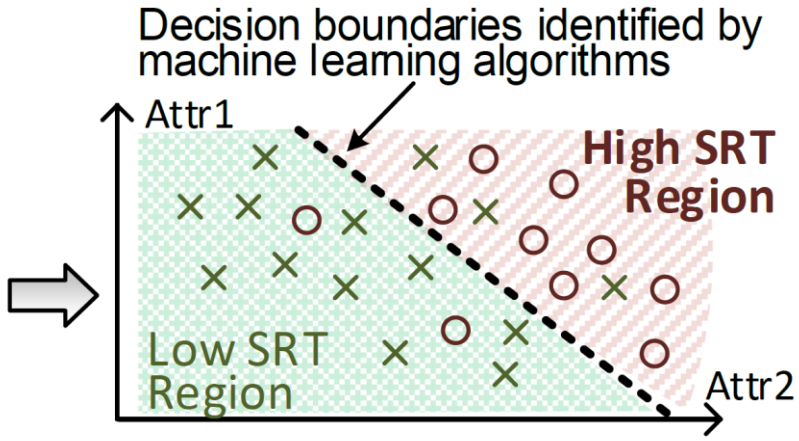
Limited visibility of naïve single-dimension analysis → Multi-dimension analysis

Interdependencies between attributes → Work with interdependencies

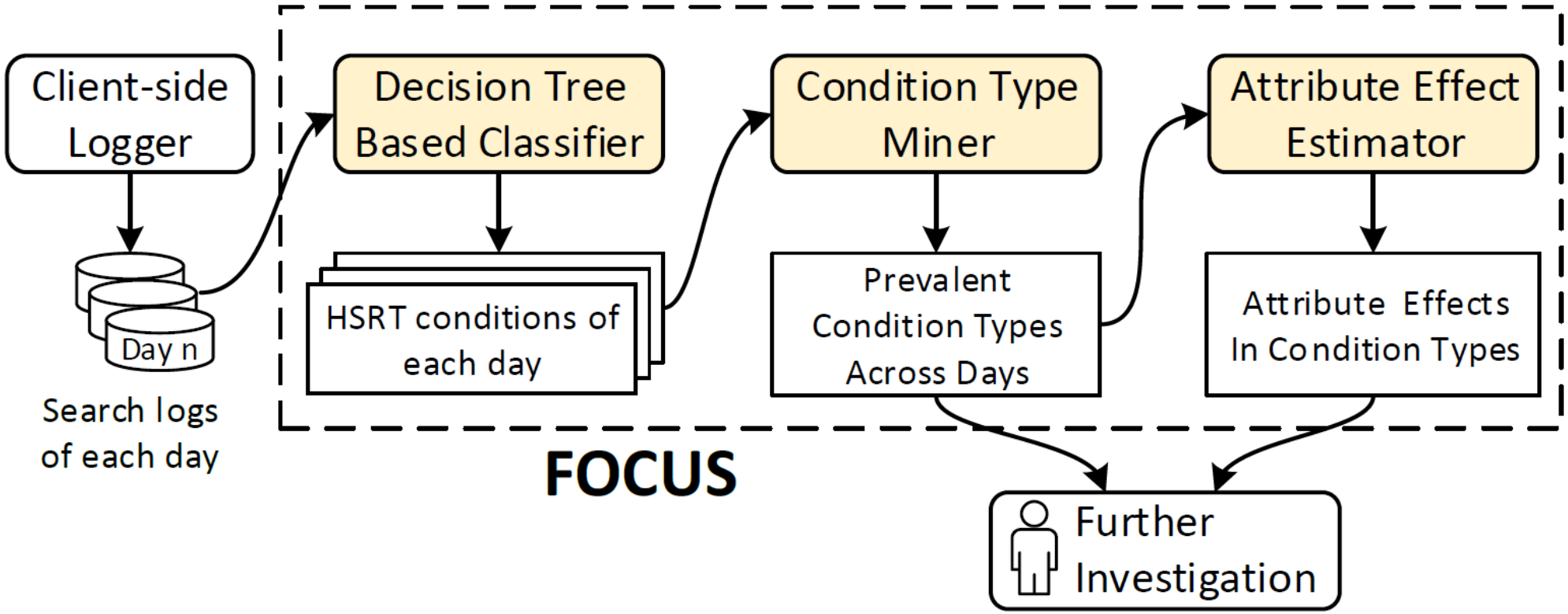
Overlapped HSRT conditions → Classification is non-overlap

- Model it as a classification problem
- Solve it using decision trees

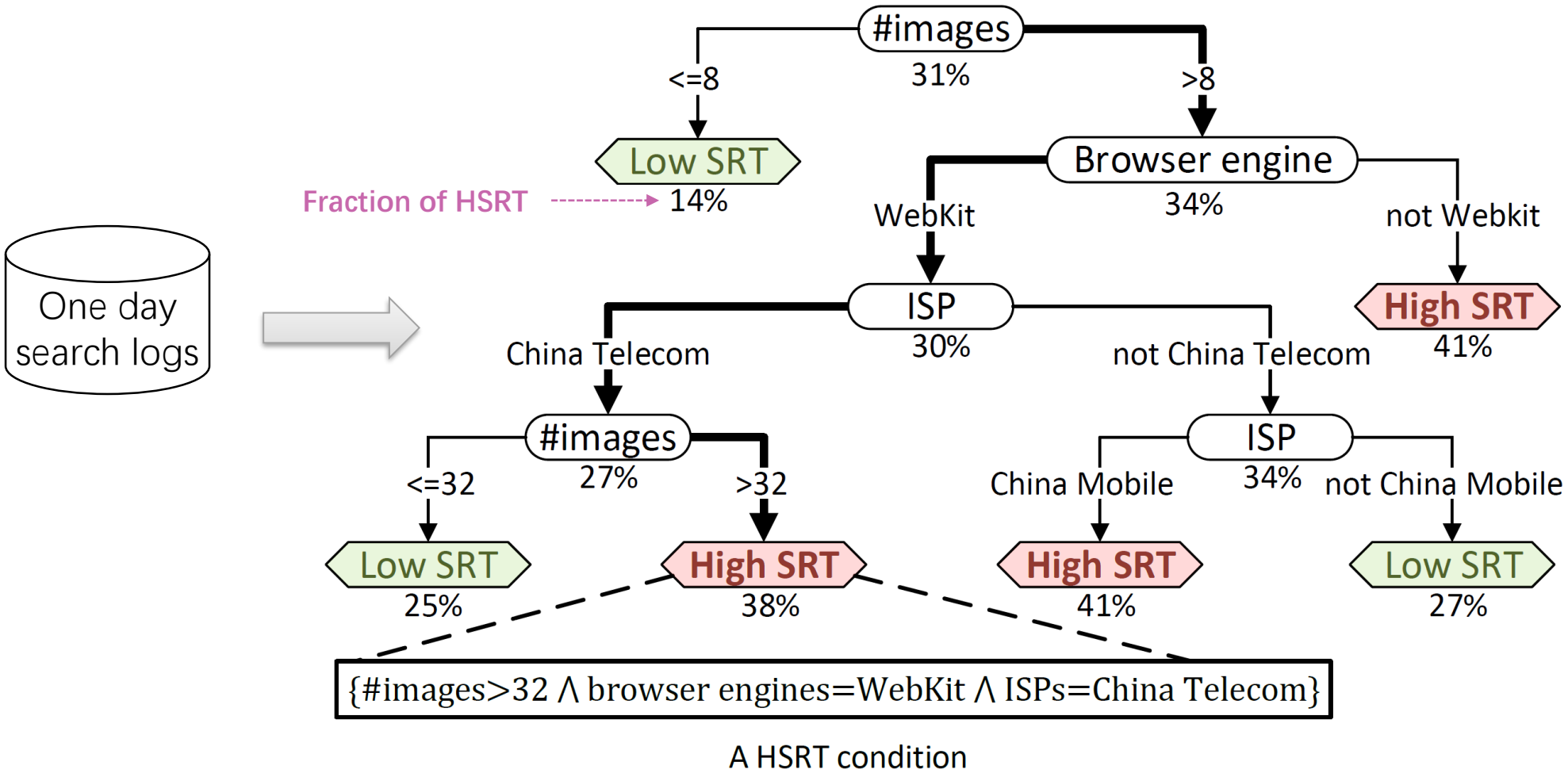
Attr1	Attr2	Label
…	…	High SRT ○
…	…	Low SRT ×
…	…	Low SRT ×
…	…	…



FOCUS Overview

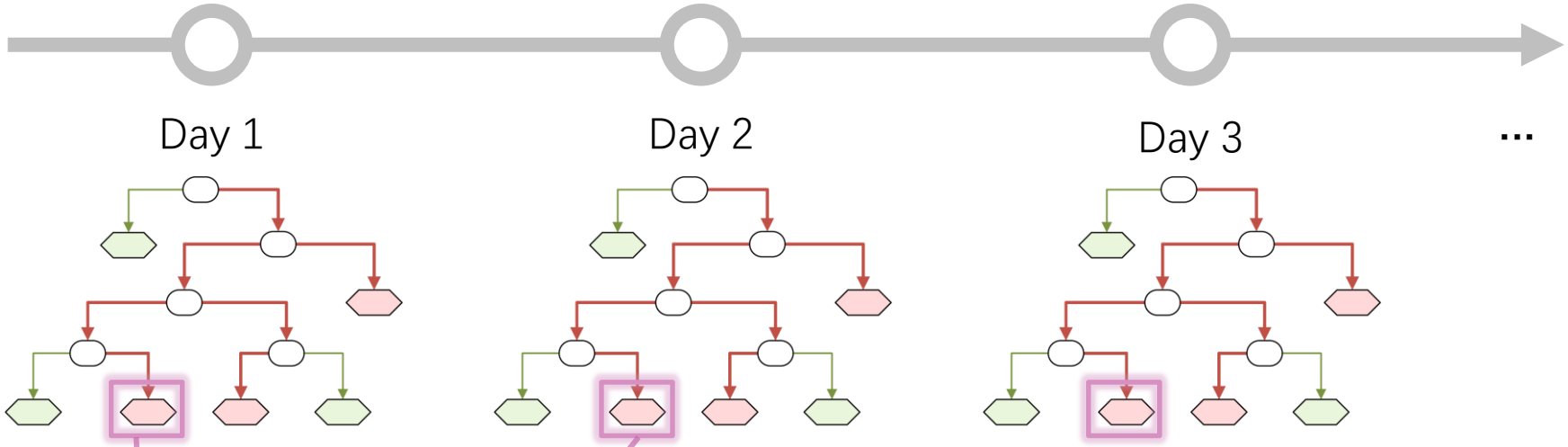


Identify HSRT Conditions Based on a Decision Tree



To build a reasonable tree, we **tailor** the mechanisms of decision trees (Details are in the paper)

Find Similar HSRT Conditions (HSRT Condition Types)



ID	HSRT Conditions		
	#Images	Browser engine	Ads
1	> 9	Not WebKit	no
2	> 10	Not WebKit	no

HSRT Condition Type		
#Images	Browser engine	Ads
$> i, i \in \{9,10\}$	Not WebKit	no

Hierarchical clustering

- Same combination of attributes
- Same value for each categorical attribute
- **Similar** value for each numeric attribute

Estimate the Impact of Each Attribute

Inspired by controlled experiment

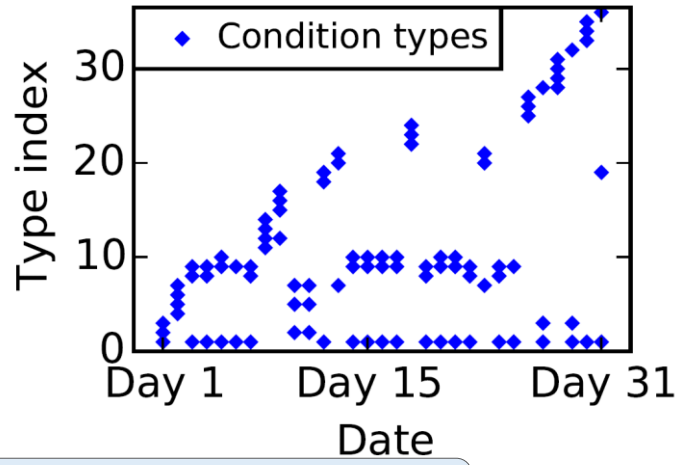
- **Control group:** the original HSRT contrition types
- **Experimental group:** changing one attribute at a time

Compare performance in historical logs

ID	HSRT Condition Type		
	#Images	Browser engine	Ads
C	$> i, i \in \{9,10\}$	Not WebKit	no
C_1	$\leq i, i \in \{9,10\}$	Not WebKit	no
C_2	$> i, i \in \{9,10\}$	WebKit	no
C_3	$> i, i \in \{9,10\}$	Not WebKit	yes

Results of FOCUS：Prevalent HSRT Condition Types

- Find 36 HSRT condition types in one month of search logs
- **Four** of them (11%) appear in more than five days



Images are the main bottleneck

(**Attributes** in bold have a bad effect on SRT)

Condition type ID	Prevalent condition type	Prevalence (days)
1	$\#\mathbf{images} > i, i \in \{5,6,7,8,9\} \wedge \mathbf{browser\ engine = not\ WebKit}$	21
2	$\#\mathbf{images} > i, i \in \{5,6,7,8,9\} \wedge \mathbf{ISP = not\ China\ Telecom} \wedge \text{browser engine} = \text{WebKit}$	15
3	$\#\mathbf{images} > i, i \in \{25,26,27\} \wedge \text{ISP} = \text{China Telecom} \wedge \text{browser engine} = \text{WebKit}$	7
4	$\#\mathbf{images} > i, i \in \{5,6,8\} \wedge \text{ISP} = \text{China Telecom} \wedge \text{browser engine} = \text{WebKit} \wedge \mathbf{ads = yes}$	6

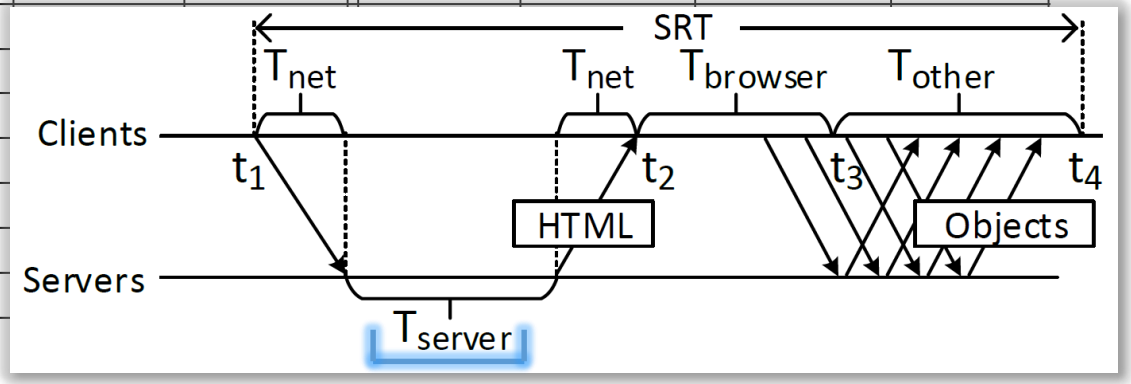
Results of FOCUS：Attribute Effects

Row#	Category	Condition type ID	Attribute condition to be flipped	Performance variations after flipping an attribute condition					
				HSRT%	SRT	T_{net}	T_{server}	$T_{browser}$	T_{other}
1	Images	1	**#images** $> i, i \in \{5,6,7,8,9\}$	-61%	-39%	-26%	+33%	-43%	-83%
2		4	**#images** $> i, i \in \{5,6,8\}$	-59%	-36%	-29%	+43%	-40%	-78%
3		2	**#images** $> i, i \in \{5,6,7,8,9\}$	-53%	-32%	-29%	+42%	-36%	-77%
4		3	**#images** $> i, i \in \{25,26,27\}$	-33%	-20%	-21%	+37%	-22%	-39%
5	Browsers	1	**browser engine = not WebKit**	-24%	-14%	-7%	-3%	-63%	-5%
6	ISPs	2	**ISP = not China Telecom**	-22%	-12%	-14%	-21%	-7%	-6%
7	Ads	4	**ads = yes**	-19%	-12%	-19%	-3%	-27%	-9%
8	ISPs	3	ISP = China Telecom	+22%	+12%	+10%	+28%	+7%	+8%
9		4	ISP = China Telecom	+27%	+12%	+14%	+26%	+5%	+4%
10	Browsers	3	browser engine = WebKit	+27%	+13%	+5%	+7%	+174%	-1%
11		2	browser engine = WebKit	+28%	+14%	+7%	+2%	+168%	+3%
12		4	browser engine = WebKit	+40%	+21%	+13%	+8%	+194%	-1%

Condition type ID	Prevalent condition type	Prevalence (days)
1	**#images** $> i, i \in \{5,6,7,8,9\} \wedge$ **browser engine = not WebKit**	21
2	**#images** $> i, i \in \{5,6,7,8,9\} \wedge$ **ISP = not China Telecom** $\wedge$ browser engine = WebKit	15
3	**#images** $> i, i \in \{25,26,27\} \wedge$ ISP = China Telecom $\wedge$ browser engine = WebKit	7
4	**#images** $> i, i \in \{5,6,8\} \wedge$ ISP = China Telecom $\wedge$ browser engine = WebKit $\wedge$ **ads = yes**	6

Observations by investigating the results of FOCUS

Row#	Category	Condition type ID	Attribute condition to be flipped	Performance variations after flipping an attribute condition					
				HSRT%	SRT	T_{net}	T_{server}	$T_{browser}$	T_{other}
1	Images	1	**#images** $> i, i \in \{5, 6, 7, 8, 9\}$	-61%	-39%	-26%	+33%	-43%	-83%
2		4	**#images** $> i, i \in \{5, 6, 8\}$	-59%	-36%	-29%	+43%	-40%	-78%
3		2	**#images** $> i, i \in \{5, 6, 7, 8, 9\}$	-53%	-32%	-29%	+42%	-36%	-77%
4		3	**#images** $> i, i \in \{25, 26, 27\}$	-33%	-20%	-21%	+37%	-22%	-39%
5	Browsers	1	**browser engine = not WebKit**	-24%	-14%	-7%	-3%	-63%	-5%
6	ISPs	2	**ISP = not China Telecom**						
7	Ads	4	**ads = yes**						
8	ISPs	3	ISP = China Telecom						
9		4	ISP = China Telecom						
10	Browsers	3	browser engine = WebKit						
11		2	browser engine = WebKit						
12		4	browser engine = WebKit						



Popular queries tend to have more images in their result pages,
but they have lower SRT because their HTML files are cached better by servers

Observations by investigating the results of FOCUS

Query Frequency	HTML Cache Ratio	Average #Images	SRT (ms)	T_{net} (ms)	T_{server} (ms)	$T_{browser}$ (ms)	T_{other} (ms)
$[1, f]$	32%	19	785	132	400	71	182
$(f, 10f]$	75%	22	663	121	250	86	206
$(10f, 100f]$	95%	28	659	127	205	93	234
$(100f, \infty]$	99%	32	643	114	191	93	244

Decreased

Decreased

Increased

Popular queries tend to have more images in their result pages,
but they have lower SRT because their HTML files are cached better by servers

Observations by investigating the results of FOCUS

Query Frequency	HTML Cache Ratio	Average #Images	SRT (ms)	T_{net} (ms)	T_{server} (ms)	$T_{browser}$ (ms)	T_{other} (ms)
$[1, f]$	32%	19	785	132	400	71	182
$(f, 10f]$	75%	22	663	121	250	86	206
$(10f, 100f]$	95%	28	659	127	205	93	234
$(100f, \infty]$	99%	32	643	114	191	93	244

Decreased

Increased

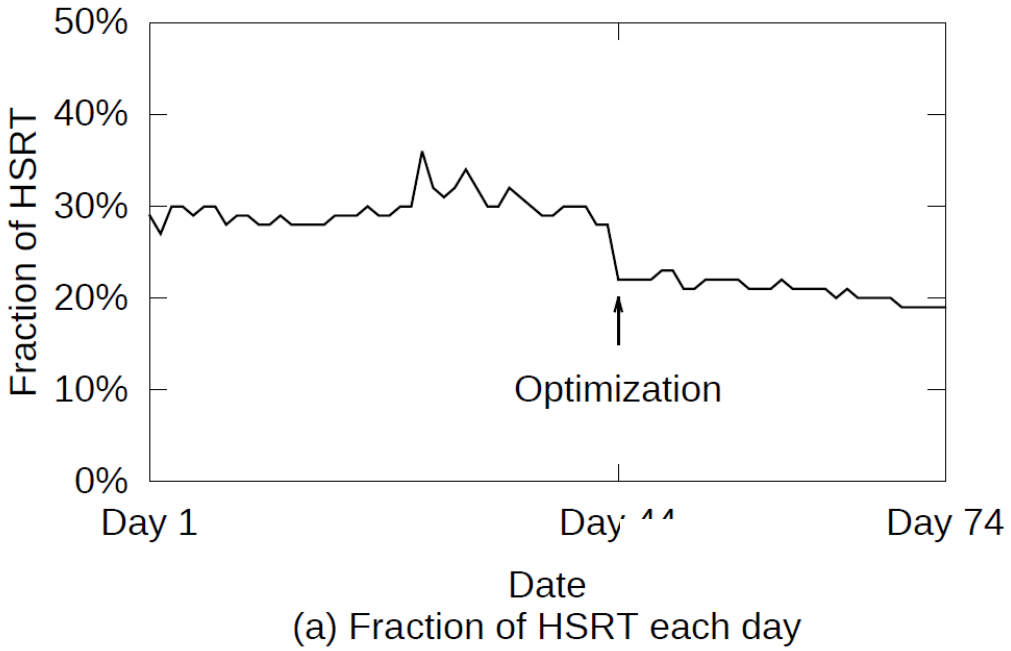
Confounding factors

Popular queries tend to have more images in their result pages,
but they have lower SRT because their HTML files are cached better by servers

More observations are in the paper

Real-world Optimization

- 1st month results of FOCUS → images are the main bottleneck of SRT
- Deploy “image base64 encoding" to improve the transmission time of images



(a) Fraction of HSRT each day

The fraction of HSRT is reduced by 30%

Conclusion

- FOCUS can
  - Narrow down the debugging space of High SRT in search logs
  - Analyze the effects of each attribute (potential improvements)

- With the output of FOCUS
  - We make several interesting observations
  - Deploy a solution in practice and greatly improve SRT

- FOCUS is a general method for analyzing multi-attribute logs
  - Web applications other than search engines
  - Performance of mobile apps
  - …

Thank you
Q&A

Dapeng Liu
liudp10@mails.tsinghua.edu.cn

清華大學
Tsinghua University

Baidu 百度