Cross-Device Search

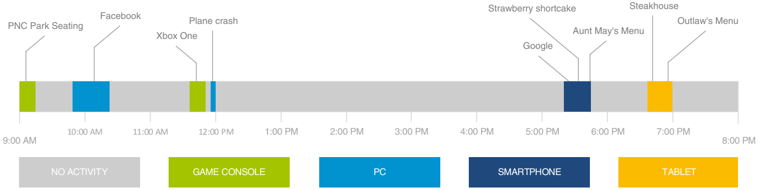
George D. Montañez, Ryen White, Xiao Huang

Carnegie Mellon University
Microsoft Research
Microsoft

CIKM 2014

Nov. 6th, 2014

Why Look At Cross-Device Search?



- The number of smart and web capable devices is rapidly increasing.
- Multi-device users on a commercial search engine: **5%** of all users.
- Multi-device user query volume: **16%** of total volume.
- Previous research has not investigated search on gaming consoles, or focused on device-device transitions.

Search Log Dataset

- Looked at several months of search and mobile log data, for English speaking US market.
- Obtained users who were logged in to search engine during this period, with unique identifier for each user.
- Filtered abnormal queries (e.g., from automated bots) and queries without necessary metadata.
- $\sim$**2 billion** total queries, $\sim$**33 million** users.

Query Counts

Table: Query Counts

Description	Total	%
Queries	2,271,142,893	100.00
Multi-Device User Queries	370,865,428	16.33
PC Queries	2,074,083,054	91.32
Smartphone Queries	53,939,886	2.38
Tablet Queries	137,979,833	6.08
Xbox Queries	1,854,422	0.08

Device-Type Counts

Table: User Device-type Counts

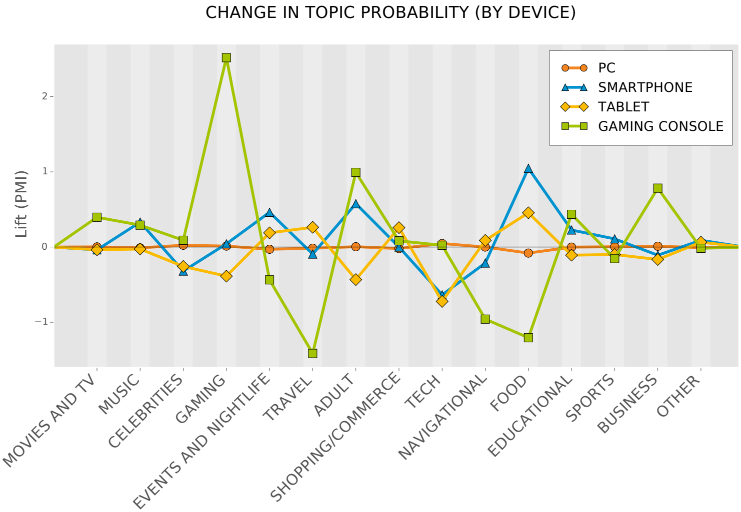
Device(s)	**Users**	**%**
Any Device(s)	33,221,253	100.00
More Than One Device	1,675,272	5.04
One Device	31,545,981	94.96
Two Devices	1,585,018	4.77
Three Devices	89,834	0.27
Four Devices	420	<0.01
PC	31,770,955	95.63
Smartphone	1,301,717	3.92
Tablet	1,863,783	5.61
Xbox	50,744	0.15

Cross-Device Analysis

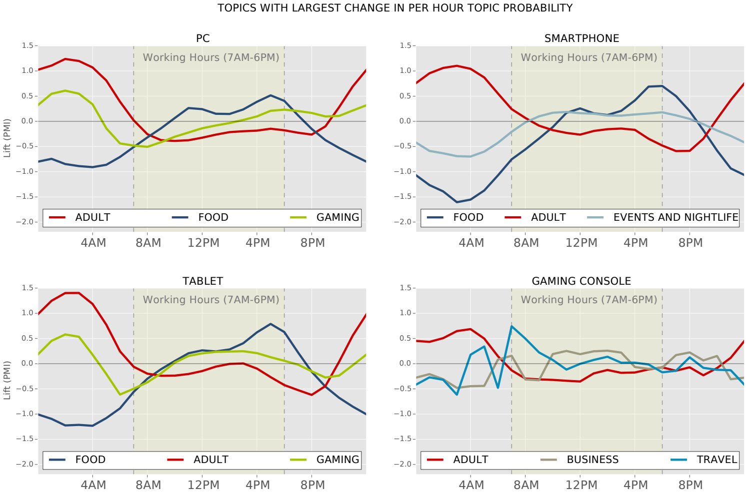
We characterized:

- Search behavior for different device types, i.e., topical interests on different devices
- How topical interests shifted over the course of the day
- Device-transition behavior, i.e., how did users transition among devices

Topical Interest by Device-Type



Changes in Topic Distributions Per Device, Per Hour



Markov Transition Probability Graphs

- Maximum likelihood estimates for device transitions conditioned on the previous device
- Most transitions are self-transitions

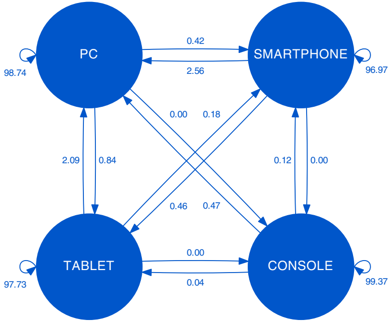


Figure: Device transition probabilities %

Markov Transition Probability Graphs - Cont.

- Only consider transitions between different devices
- Most transition mass leads to PC
- Significant interplay between PC-smartphone and PC-tablet

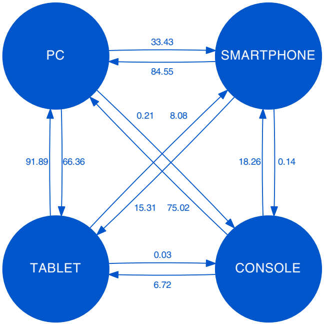
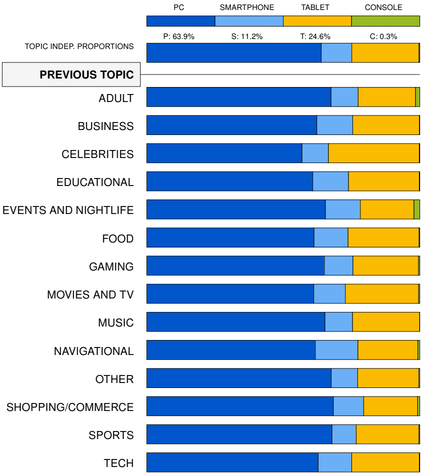


Figure: Cross-device transition probabilities %

Previous Query Topic To Next Device

- Relationship between previous query topic and the device of next query
- Events and Nightlife query increases likelihood of using a gaming console
- Celebrities query increases likelihood of using a tablet



Predicting the Next Device

- Given a previous query, time of day and associated topical, spatial and temporal context, can we predict the next device a user will use?

Predicting the Next Device

- Given a previous query, time of day and associated topical, spatial and temporal context, can we predict the next device a user will use?

- **Answer:** Yes, with high accuracy, precision and recall.

Predicting the Next Device

- Given a previous query, time of day and associated topical, spatial and temporal context, can we predict the next device a user will use?
- **Answer:** Yes, with high accuracy, precision and recall.
- Knowing which device is next can allow for device-appropriate content gathering and better query-sense disambiguation.

Prediction Datasets

- Created three datasets from primary dataset: Main, Balanced (50/50 mix of cross-device/same-device transitions), and Cross-Device Only.
- Features: Previous Device, Previous Query Topics, Previous Query Length, Previous Location, etc.
- Computed global and user-specific transition statistics as features from separate historical data.

Features

Type	Features and Counts
Previous Device	Previous Device Flag (x4)
Query Length	Query Length
Hour	Previous Query Local Hour Flag (x24)
Topic	Previous Query Topic Flags (x15)
	Current Query Topic Flags (x15)
Global Stats	Global Device Probabilities (x4)
	Global Device-Device Transition Probabilities (x16)
	Global Cross-Device Transition Probabilities (x12)
Global Temporal Stats	Global Device Avg. Transition Delay (x4)
	Global Device-Device Avg. Transition Delay (x16)

Features (Cont.)

Type	**Features and Counts**
User Stats	Number of Historical User Samples
	Number of Historical User Cross-Device Samples
	User Device Probabilities (x4)
	User Device-Device Transition Probabilities (x16)
	User Cross-Device Transition Probabilities (x12)
	User Cross-Device Destination Device Probabilities (x4)
	User IsDeviceDominant Flags (x4)
	User IsCrossDeviceDominant Flags (x4)
User Temporal Stats	User Device Avg. Transition Delay (x4)
	User Device-Device Avg. Transition Delay (x16)

Experimental Set-up

- 5-Fold cross-validation on subsample of 250,000 records from each set.
- **Baseline Methods:** Most Frequent Label, Uniform Guessing, Stratified Guessing.
- **Models:** L1 Logistic Regression and Gradient Boosting Trees Ensemble Classifiers.
- **Metrics:** Accuracy, Precision, Recall, F1-score, Multiclass Log-Loss.

Predict Next Device - Main

- Both models show significant improvement over baseline methods
- Previous device state is the strongest feature for next device prediction

Method	**Accuracy**	**Avg. Precision**	**Avg. Recall**	**Avg. F1-Score**	**Log-Loss**
Baseline Method - Most Frequent Label	0.648	0.420	0.648	0.510	12.157
Baseline Method - Stratified	0.489	0.489	0.489	0.489	0.883
Baseline Method - Uniform	0.250	0.488	0.250	0.309	1.386
Gradient Boosting Trees Classifier	0.982***	0.982***	0.982***	0.982***	0.097***
L1 Logistic Regression	0.983***	0.983***	0.983***	0.983***	0.089***

(Statistical significance assessed by two-tailed paired t-test, with: * $< \alpha = .05$, ** $< \alpha = .01$, *** $< \alpha = .001$)

Predict Device Switch (Balanced)

- Predict whether a user switched device between queries.
- Balanced transition dataset: transition instances equally split among the "same device" and "different device" class label
- A non-trivially important task in scenarios when a new query is issued without corresponding device-type info

Method	Accuracy	Avg. Precision	Avg. Recall	Avg. F1-Score	Log-Loss
Baseline Method - Most Frequent Label	0.499	0.249	0.499	0.332	17.320
Baseline Method - Stratified	0.498	0.498	0.498	0.498	0.693
Baseline Method - Uniform	0.501	0.501	0.501	0.501	1.386
Gradient Boosting Trees Classifier	0.787***	0.788***	0.787***	0.786***	0.462***
L1 Logistic Regression	0.779***	0.782***	0.779***	0.778***	0.484***

(Statistical significance assessed by two-tailed paired t-test, with: * $< \alpha = .05$, ** $< \alpha = .01$, *** $< \alpha = .001$)

Predict Device Switch - Feature Ablations

Feature Group	Accuracy	Avg. Precision	Avg. Recall	Avg. F1-Score	Log-Loss
Baseline	0.589	0.591	0.589	0.587	0.676
Baseline+Q	0.589	0.591	0.589	0.587	0.674***
Baseline+Q+PT	0.599***	0.602***	0.599***	0.596***	0.665***
Baseline+Q+PT+PH	0.608***	0.609***	0.608***	0.607***	0.661***
Baseline+Q+PT+PH+U1	0.660***	0.661***	0.660***	0.660***	0.615***
Baseline+Q+PT+PH+U1-2	0.665***	0.666***	0.665***	0.664***	0.610***
Baseline+Q+PT+PH+U1-4	0.668***	0.668***	0.668***	0.668***	0.607***
Baseline+Q+PT+PH+U1-6	0.673***	0.674***	0.673***	0.673***	0.601***
Baseline+Q+PT+PH+U1-6+G	0.674***	0.674***	0.674***	0.673***	0.602***
Baseline+Q+PT+PH+U1-6+G+CT	0.675***	0.675***	0.675***	0.675***	0.599***
Baseline+Q+PT+PH+U1-6+G+DT	0.778***	0.782***	0.778***	0.778***	0.485***
All Features	0.779***	0.782***	0.779***	0.778***	0.484***

Baseline = Previous Device
Q = Previous Query Length
G = Global Transition Stats Features
PT = Previous Topic features
CT = Current Topic Features
PH = Previous Query Hour
U1 = Number of historical samples for the user and the number of cross-device samples for the user
U1-2 = U1 and user-specific device probabilities
U1-3 = U1, U2 and device-device pair transition probabilities
U1-4 = U1, U2, U3 and destination device transition probabilities
U1-5 = U1 through U4 and device conditioned transition probabilities
U1-6 = U1 through U5 and user-specific average transition times for device-device pairs
DT = Delay time (in seconds) between previous and current queries
All Features = All of the above feature sets

(Statistical significance assessed by two-tailed paired *t*-test, with: * $< \alpha = .05$, ** $< \alpha = .01$, *** $< \alpha = .001$)

Predict Next Device Given Device Switch

- A more difficult task
- The feature-based methods continue to significantly outperform the baselines

Method	**Accuracy**	**Avg. Precision**	**Avg. Recall**	**Avg. F1-Score**	**Log-Loss**
Baseline Method - Most Frequent Label	0.455	0.207	0.455	0.284	18.829
Baseline Method - Stratified	0.370	0.371	0.370	0.371	1.046
Baseline Method - Uniform	0.250	0.370	0.250	0.292	1.386
Gradient Boosting Trees Classifier	0.931***	0.931***	0.931***	0.931***	0.197***
L1 Logistic Regression	0.934***	0.933***	0.934***	0.933***	0.193***

(Statistical significance assessed by two-tailed paired t-test, with: * $< \alpha = .05$, ** $< \alpha = .01$, *** $< \alpha = .001$)

Predict Next Device Given Device Switch - Feature Ablations

- Previous device and searchers' own transition histories were the pirmary factors in the prediction
- Compactness is important of large scale search engines.

Feature Group	Accuracy	Avg. Precision	Avg. Recall	Avg. F1-Score	Log-Loss
Baseline+U1-6	0.932***	0.931***	0.932***	0.931***	0.196***
Baseline	0.781	0.642	0.781	0.703	0.496
Baseline+Q	0.781	0.642	0.781	0.703	0.496
Baseline+Q+PT	0.781	0.642	0.781	0.703	0.495
Baseline+Q+PT+PH	0.781	0.679	0.781	0.703	0.493
Baseline+Q+PT+PH+U1	0.781	0.716	0.781	0.703	0.491
Baseline+Q+PT+PH+U1-2	0.903***	0.903***	0.903***	0.898***	0.281***
Baseline+Q+PT+PH+U1-3	0.928***	0.927***	0.928***	0.927***	0.203***
Baseline+Q+PT+PH+U1-4	0.932***	0.932***	0.932***	0.931***	0.195***
Baseline+Q+PT+PH+U1-5	0.932***	0.932***	0.932***	0.931***	0.195***
Baseline+Q+PT+PH+U1-6	0.933***	0.932***	0.933***	0.932***	0.194***
Baseline+Q+PT+PH+U1-6+G	0.933***	0.932***	0.933***	0.932***	0.194***
Baseline+Q+PT+PH+U1-6+G+CT	0.933***	0.932***	0.933***	0.932***	0.194***
Baseline+Q+PT+PH+U1-6+G+DT	0.933***	0.933***	0.933***	0.932***	0.194***
All Features	0.934***	0.933***	0.934***	0.933***	0.193***

For Feature Group legend, see Table 19

(Statistical significance assessed by two-tailed paired t-test, with: * $< \alpha = .05$, ** $< \alpha = .01$, *** $< \alpha = .001$)

Summary

- We characterized temporal and topical aspects of cross-device search.
- Found that query content differed among device types and hour of day, often in intuitive ways.
- These differences and patterns can be successfully exploited to reliably predict the next device a user will transition to.
- Previous device can signal the next device, even when the devices differ.

Opportunities For Future Work

- Delay time prediction (anticipated time between devices)
- Session-level rather than query level predictions
- Qualitative investigations to complement our quantitative analysis
- Supporting device transitions (e.g., pre-fetching content suitable for the next device used)

The End

Thank You!