Predicting User Interests from Contextual Information

Ryen W. White, Peter Bailey, Liwei Chen
Microsoft Corporation

Motivation

- Information behavior is embedded in external context
  - Context motivates the problem, influences interaction
- IR community theorized about context
  - Context sensitive search, user studies of search context

- User interest models can enhance post-query behavior & general browsing by leveraging contextual info.
  - e.g., personalization, information filtering, etc.

- Little is known about the value of different contextual sources for user interest modeling

Overview

- A systematic, log-based study of five contextual sources for user interest modeling during Web interaction
- Assume user has browsed to *URL*
- Evaluate the predictive value of five contexts of *URL*:
  - **Interaction:** recent interactions preceding *URL*
  - **Collection:** pages that link to *URL*
  - **Task:** pages sharing search engine queries with *URL*
  - **Historic:** long term interests of current user
  - **Social:** combined long-term interests those who visit *URL*
- Domain is website recommendation not search results

Data Sources

- Anonymized URLs visited by users of a widely-distributed browser toolbar
- 4 months of logs (Aug 08 – Nov 08 inclusive):
  - **Past:** Aug-Sep used to create user histories
  - **Present:** Oct-Nov used for current behavior and future interests
- 250K users randomly selected from a larger user pool once most active users (top 1%) were removed
  - Chosen users with at least 100 page visits in *Past*

Trails and Terminal URLs

- From logs we extracted millions of browse trails
  - Temporally-ordered sequence of URLs comprising all pages visited by a user per Web browser instance
  - Terminate with 30-minute inactivity timeout
- A set of 5M terminal URLs (u_t) obtained by randomly-sampling all URLs in the trails
  - Terminal URLs demarcate past and future events
- Task = Learn user interest models from contexts for u_t, use those models to predict future user interests

Building User Interest Models

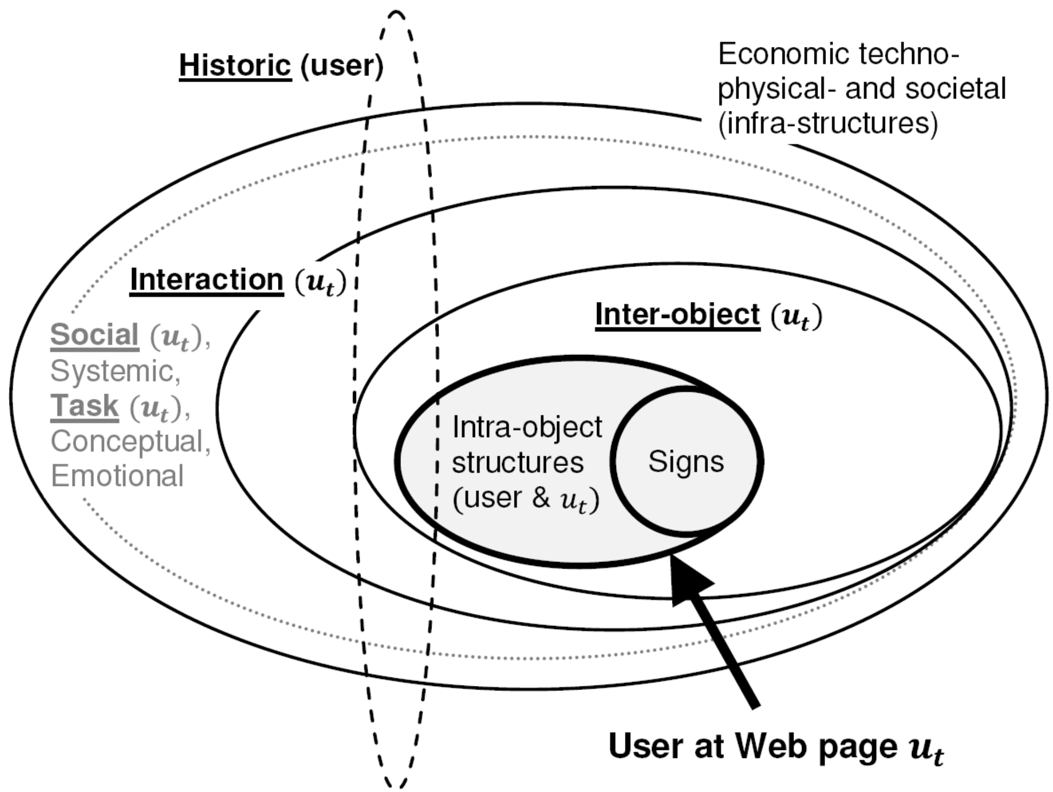
- Classified context URLs in the Open Directory Project human-edited Web directory (ODP, dmoz.org)
- Automatically assigned category labels via URL match
  - URL back-off used if no exact match obtained
- Represent interests as list of ODP category labels
  - Labels ranked in descending order by frequency
  - For example, for a British golf enthusiast, the top of their user interest profile might resemble:

ODP Category Labels	**Frequency**
Sports/Golf/Courses/Europe/United Kingdom	102
Sports/Golf/Driving Ranges	86
Sports/Golf/Instruction/Golf Schools	63

Selecting Contexts

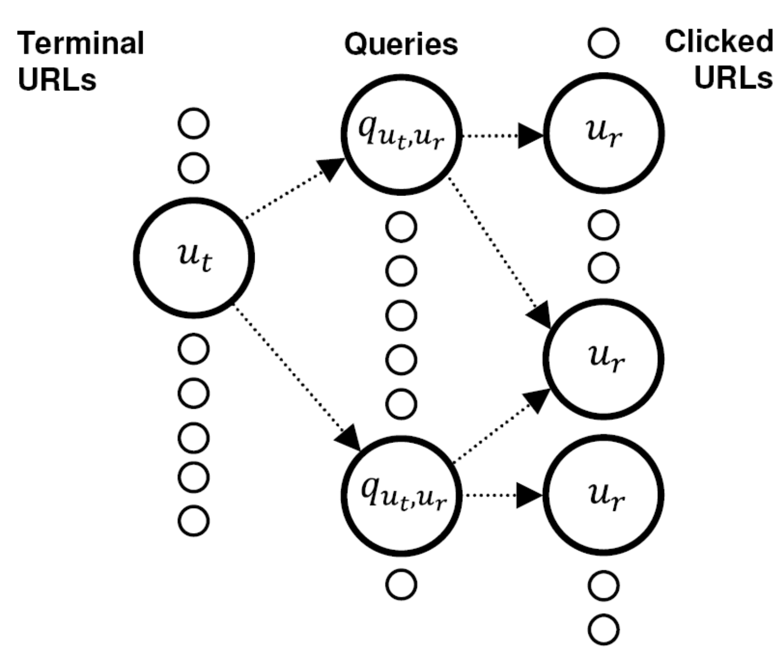
- Ingwersen and Järvelin (2005) developed nested model of context stratification representing main contextual influences on people engaged in information behavior

- **Dimensions used**
- Others challenging to model via logs
  - e.g., cognitive and affective state, infra-structures, etc.



Defining Contexts

- **None (*u_t* only):** Interest model for terminal URL
- **Interaction (*u_{t-5}* … *u_{t-1}*):** Interest model for five Web pages immediately preceding u_t
- **Task:** Interest model for pages encountered during the same or similar tasks
  - Walk on search engine click graph from u_t to queries and then back out to pages



Defining Contexts

- **Collection:** Interest model pages linking to u_t
    - We obtained a set of in-links for each u_t from a search engine index, built model from pages linking to u_t
- **Historic:** Interest model for each user based on their long-term Web page visit history
- **Social:** Interest model from combination of the historic contexts of users that also visit u_t

- What is the effectiveness of different context sources for user interest modeling?

Methodology

- Found instances of u_t in *Present* set (Oct-Nov 08 logs)
- Used all actions **after u_t** as source of future behavior
  - Futures specific to each user and each u_t
  - Used to gauge predictive value of each context
- Created three interest models representing future interests (ranked list of ODP labels & frequencies):
  - **Short:** within one hour of u_t
  - **Medium:** within one day of u_t
  - **Long:** within one week of u_t
- Filtered $\{u_t\}$ to help ensure experimental integrity
  - e.g., no more than 10 u_t per user

Methodology

- Divided filtered $\{u_t\}$ into 10 equally-sized runs
  - Each run contained at most one u_t from each user
- Experimental procedure:
  - For each u_t in each run:
    - Build ground truth for *short*-, *medium*-, and *long*-term future interest models
    - Build interest models for different contexts (and combinations)
    - Determine predictive accuracy of each model
- Used five measures to determine prediction accuracy
  - P@1, P@3, Mean Reciprocal Rank, nDCG, and F1
  - F1 tracked well with others - focus on that here

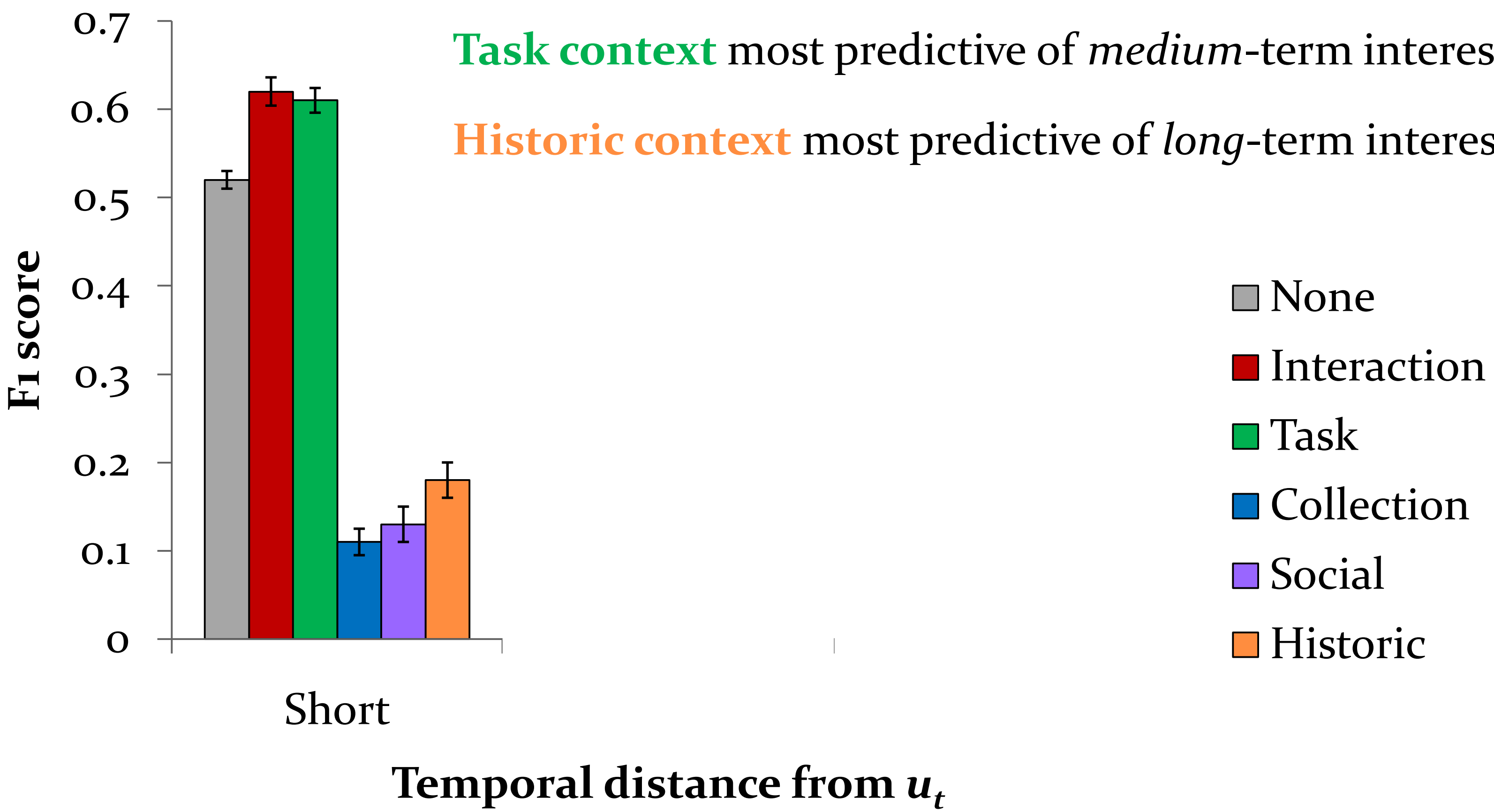
Findings – Context comparison

- Predictive performance of contextual sources for different futures

Interaction context & **Task context** most predictive of *short*-term interests

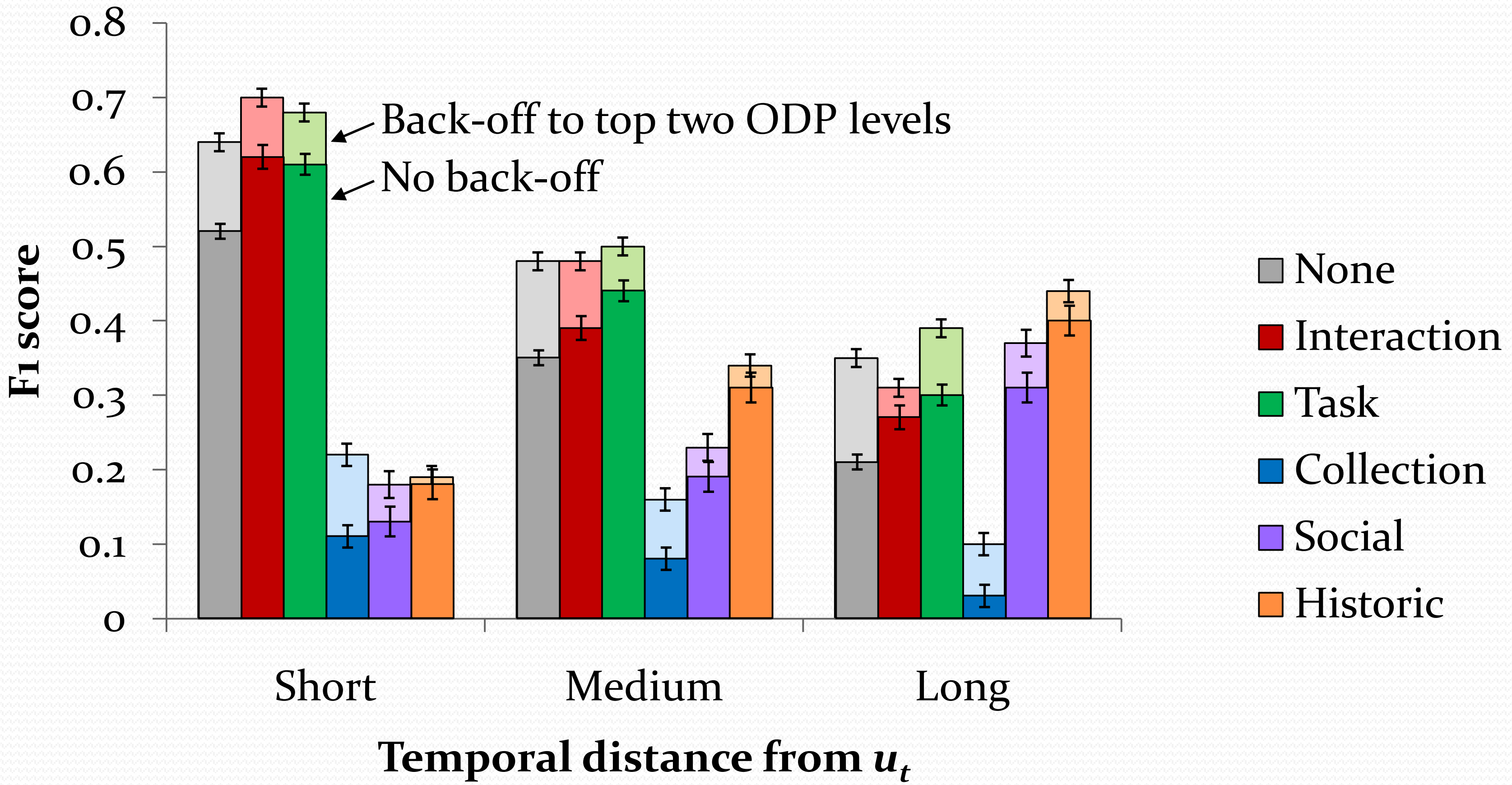
Task context most predictive of *medium*-term interests

Historic context most predictive of *long*-term interests



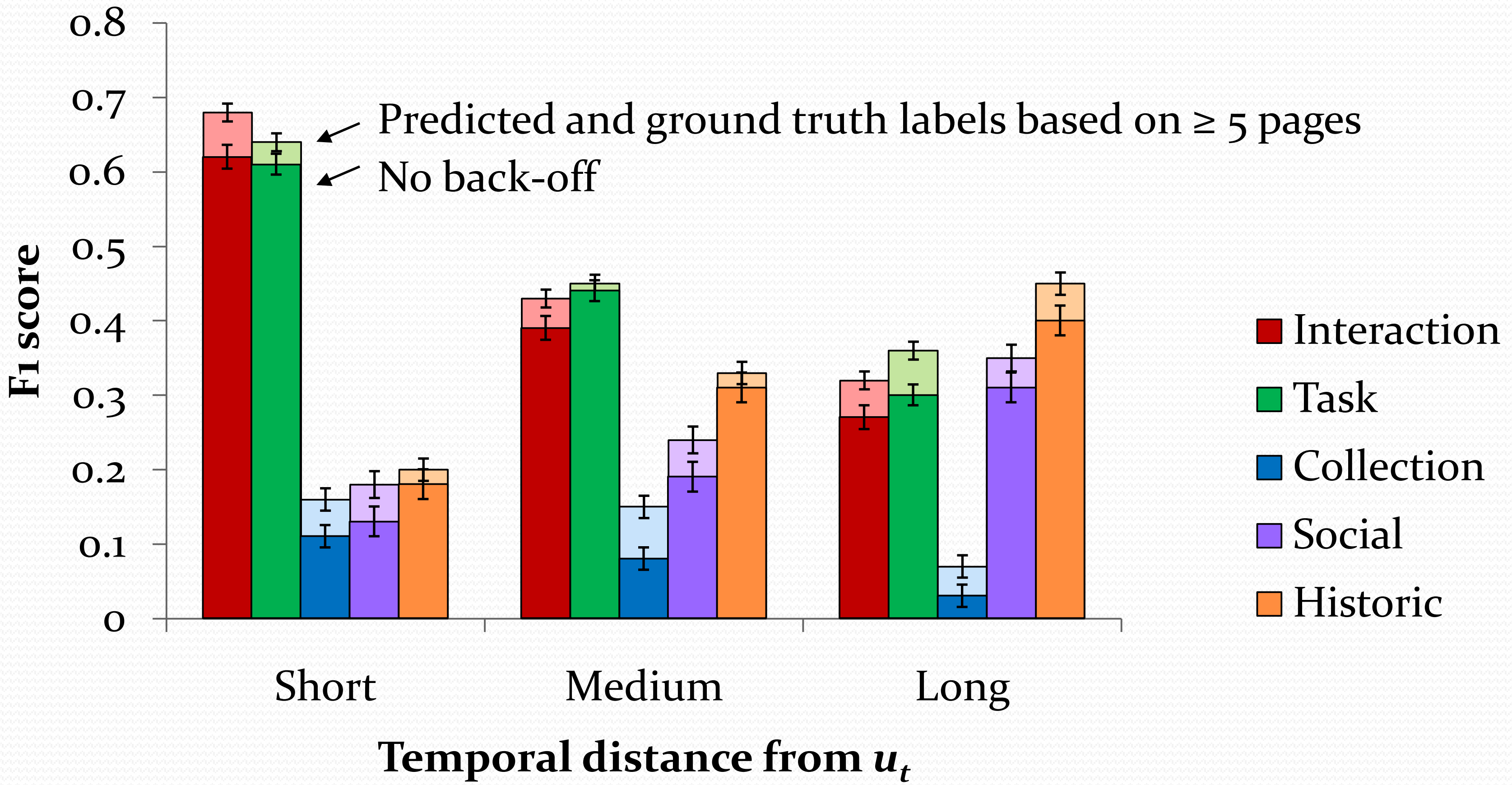
Findings – Handling near misses

- Near miss between prediction and ground truth regarded as total miss
  - Use one/two/three-level back-off on both ground truth and prediction



Findings – Improved confidence

- Basing predictions & ground truth on small # page visits may skew results
  - Repeat experiment & ignore labels based on < 5 page visits



Findings – Combining contexts

Rank	Short		Medium		Long	
	Sources	F1 score	Sources	F1 score	Sources	F1 score
1	n, **i**, **t**, h, s, c	0.72**	n, i, **t**, h, s, c	0.53**	n, i, t, s, **h**, c	0.45**
2	n, **i**, s, h, c	0.71**	n, i, **t**, h, c	0.52**	n, i, s, **h**, c	0.43**
3	n, **i**, **t**, h, c	0.71**	n, i, **t**	0.49**	n, i, t, **h**, c	0.43*
4	n, **i**, h, c	0.71**	n, i, s, h, c	0.48*	s, **h**	0.43*
5	n, **i**, s, **t**, c	0.69**	n, i, h, **t**	0.48*	n, i, s, **h**, t	0.42*

- Overlap beats single contextual sources
- Key contexts still important
  - Short = Interaction (**i**) and Task (**t**)
  - Medium = Task (**t**)
  - Long = Historic (**h**)
- Supports polyrepresentation theory (Ingwersen, 1994)
  - Overlap between sources boosts predictive accuracy

Summary of Findings

- Performance of context dependent on distance between u_t and end of prediction window
  - **Short**-term interests predicted by task/interaction contexts
    - Topical interest may not be highly dynamic, even if queries and information needs are
  - **Medium**-term interests best predicted by task context
    - More likely to include task variants appearing in next day
  - **Long**-term interests predicted by historic/social contexts
    - Interest may be invariant over time, users visiting same pages may have similar interests
- Overlap effective - many contexts reinforce key interests

Conclusions and Take-away

- Systematic study of context for user interest modeling
- Studied predictive value of five context sources
  - Value varied with duration of prediction
  - Short: interaction/task, Medium: task, Long: historic/social
- Overlap was more effective than any individual source
- Source must be tailored to modeling task
- Search/recommendation systems should not treat all contextual sources equally
  - Weights should be assigned to each source based on the nature of the prediction task