Temporal Dynamics and Information Retrieval

Susan Dumais
Microsoft Research

http://research.microsoft.com/~sdumais

In collaboration with:
Eric Horvitz, Jaime Teevan, Eytan Adar, Jon Elsas, Dan Liebling,
Richard Hughes, Krysta Svore, Kira Radinsky
Change is Everywhere in IR

- Change is everywhere in digital information systems
  - New documents appear all the time
  - Document content changes over time
  - Queries and query volume change over time
  - What’s relevant to a query changes over time
    - E.g., *U.S. Open 2012* (in May vs. Sept)
  - User interaction changes over time
    - E.g., anchor text, “likes”, query-click streams, social networks, etc.
  - Relations between entities change over time
    - E.g., President of the US is <> [in 2008 vs. 2004 vs. 2000]

- Change is pervasive in digital information systems
  ... yet, most retrieval systems ignore it!

Stanford InfoSeminar 3/9/12
Digital Dynamics Easy to Capture

- Easy to capture

- But ... few tools support dynamics
Content Changes

User Visitation/ReVisitation

Today's Browse and Search Experiences

But, ignores ...
Overview

- **Change on the Web**
  - Content changes over time
  - User interaction varies over time (queries, re-visitation, anchor text, query-click stream, “likes”)
  - Tools for understanding Web change (e.g., Diff-IE)

- **Improving Web retrieval using dynamics**
  - Query trends over time
  - Retrieval models that leverage dynamics
  - Task evolution over time
Overview

- Change on the Web
  - Content changes over time
  - User interaction varies over time (queries, re-visitation, anchor text, query-click stream, “likes”)
  - Tools for understanding Web change (e.g., Diff-IE)

- Improving Web retrieval using dynamics
  - Query trends over time
  - Retrieval models that leverage dynamics
  - Task evolution over time
Characterizing Web Change

Content Changes

- Large-scale Web crawls, over time
  - Revisited pages
    - 55,000 pages crawled hourly for 18+ months
    - Unique users, visits/user, time between visits
  - Pages returned by a search engine (for ~100k queries)
    - 6 million pages crawled every two days for 6 months

[Adar et al., WSDM 2009]
Measuring Web Page Change

- Summary metrics
  - Number of changes
  - Amount of change
  - Time between changes

- Change curves
  - Fixed starting point
  - Measure similarity over different time intervals

- Within-page changes
Measuring Web Page Change

Summary metrics
- Number of changes
- Amount of change
- Time between changes

33% of Web pages change
66% of visited Web pages change
- 63% of these change every hr.

Avg. Dice coeff. = 0.80
Avg. time bet. change = 123 hrs.

.edu and .gov pages change infrequently, and not by much
.com pages change at an intermediate rate, but by a lot
popular pages change more frequently, but not by much
Measuring Web Page Change

- **Summary metrics**
  - Number of changes
  - Amount of change
  - Time between changes

- **Change curves**
  - Fixed starting point
  - Measure similarity over different time intervals

---

![Change curve for http://losangeles.craigslist.org/hsh/
Knot point at 10 hours](image)

![Change curve for http://allrecipes.com/
Knot point at 261 hours](image)
Measuring Within-Page Change

- **Term-level changes**
  - Divergence from norm
    - cookbooks
    - salads
    - cheese
    - ingredient
    - bbq
    - ...
  - “Staying power” in page
Example Term Longevity Graphs

- allrecipes.com
- craigslist.org LA
- bestbuy.com

(time since 10)
Revisitation on the Web

- Revisitation patterns
  - Log analyses
  - Toolbar logs for *revisitation*
  - Query logs for *re-finding*
  - User survey to understand intent in revisitations

User Visitation/ReVisitation

*What was the last Web page you visited?*
*Why did you visit (re-visit) the page?*
Measuring Revisitation

- Summary metrics
  - Unique visitors
  - Visits/user
  - Time between visits

- Revisitation curves
  - Histogram of revisit intervals
  - Normalized

- Many motivations for revisits
- 60-80% of Web pages you visit, you’ve visited before

Stanford InfoSeminar 3/9/12
Four Revisitation Patterns

- **Fast**
  - Hub-and-spoke
  - Navigation within site
- **Hybrid**
  - High quality *fast* pages
- **Medium**
  - Popular homepages
  - Mail and Web applications
- **Slow**
  - Entry pages, bank pages
  - Accessed via search engine
Relationships Between Change and Revisitation

- Interested in change
- Monitor
- Effect change
- Transact
- Change unimportant
- Re-find old
- Change can interfere with re-finding
Revisitation and Search
(Re-finding)

- 60-80% of the Web page visits are re-revisits
- 33%-43% of queries are re-finding

- Repeat query (33%)
  - Q: *microsoft research*
  - Click same or different URLs

- Repeat click (39%)
  - [http://research.microsoft.com/](http://research.microsoft.com/)
  - Q: *microsoft research; msr*

- Big opportunity (43%)
  - 24% “navigational revisits”

---

Stanford InfoSeminar 3/9/12

[Teevan et al., SIGIR 2007]
[Tyler et al., WSDM 2010]
[Teevan et al., WSDM 2011]
Diff-IE toolbar

Changes to page since your last visit

2009. During his career at Microsoft, Gates held the positions of CEO and chief software architect, and remains the largest individual shareholder with more than 8 percent of the common stock. He has also authored or co-authored several books.

[Teevan et al., UIST 2009]
Interesting Features of Diff-IE

- Always on
- New to you
- Non-intrusive
- In-situ

Examples of Diff-IE in Action
Expected New Content
Monitor
Serendipitous Encounters

The tax man cometh
Get ready

All fun and games
Take 5 minutes to play

Popular now
Toyota recall · KONY 2012 · Children living in bus · Solar flare
Unexpected Important Content
Understand Page Dynamics
Studying Diff-IE

- Internal study of Diff-IE
- Logging
  - URLs visited
  - Amount of change when revisited
- Feedback buttons
- Survey
  - Prior to installation
  - After a month of use
- Experience interview

[In situ]
Representative Experience
Longitudinal
People Revisit More

- Perception of revisitation remains constant
  - How often do you revisit?
  - How often are revisits to view new content?
- Actual revisitation increases
  - First week: 39.4% of visits are revisits
  - Last week: 45.0% of visits are revisits
- Why are people revisiting more with DIFF-IE?
Revisited Pages Change More

- Perception of change increases
  - What proportion of pages change regularly?
  - How often do you notice unexpected change?

- Amount of change seen increases
  - First week: 21.5% revisits changed, by 6.2%
  - Last week: 32.4% revisits changed, by 9.5%

- Diff-IE is driving visits to changed pages
  - It supports people in understanding change

Stanford InfoSeminar 3/9/12
Other Examples of Dynamics and User Experience

**Content changes**
- **Diff-IE** (Teevan et al., 2008)
- **Zoetrope** (Adar et al., 2008)
- **Diffamation** (Chevalier et al., 2010)
- Temporal summaries and snippets ...

**Interaction changes**
- Explicit annotations, ratings, “likes”, etc.
- Implicit interest via interaction patterns
  - Edit wear and read wear (Hill et al., 1992)
Overview

- Change on the Web
  - Content changes over time
  - User interaction varies over time (queries, re-visititation, anchor text, query-click stream, “likes”)
  - Tools for understanding Web change (e.g., Diff-IE)

- Improving Web retrieval using dynamics
  - Query trends over time
  - Retrieval models that leverage dynamics
  - Task evolution over time

Questions?
Overview

- **Change on the Web**
  - Content changes over time
  - User interaction varies over time (queries, re-visitation, anchor text, query-click stream, “likes”)
  - Tools for understanding Web change (e.g., Diff-IE)

- **Improving Web retrieval using dynamics**
  - Query trends over time
  - Retrieval models that leverage dynamics
  - Task evolution over time
Temporal Retrieval Models 1
(content-based)

- Current retrieval algorithms look only at a single snapshot of a page.
- But, Web page content changes over time.
- Can we leverage this to improved retrieval?
  - Pages have different rates of change.
    - Different priors (using change rate vs. link structure).
  - Terms have different longevity (staying power).
    - Some are always on the page; some transient.
  - Language modeling approach to ranking.

\[
P(D | Q) = P(D) \cdot P(Q | D)
\]

Change prior

Term longevity

Elsas et al., WSDM 2010

Stanford InfoSeminar 3/9/12
Relevance and Page Change

- **Page change** is related to relevance
  - Human relevance judgments
    - 5-point scale - Perfect/Excellent/Good/Fair/Bad
  - Rate of Change -- 60% Perfect pages; 30% Bad pages

- Use change rate as a document prior (vs. priors based on link structure like Page Rank)
  - Shingle prints to measure change

\[
P(D|Q) = P(D) \cdot P(Q|D)
\]
Relevance and Term Change

- **Terms patterns** vary over time
- Represent a document as a mixture of terms with different “staying power”
  - Long, Medium, Short

\[
P(Q|D) = \lambda_L P(Q|D_L) + \lambda_M P(Q|D_M) + \lambda_S P(Q|D_S)
\]

\[
P(D|Q) = P(D) \cdot P(Q|D)
\]
Evaluation: Queries & Documents

- 18K Queries, 2.5M Judged Documents
  - 5-level relevance judgment (Perfect ... Bad)
- 2.5M Documents crawled weekly for 10 wks

- Navigational queries
  - 2k queries identified with a “Perfect” judgment
  - Assume these relevance judgments are consistent over time
- Measure changes in nDCG
Experimental Results

The graph shows the comparison of different models in terms of % Gain/Loss in nDCG. The x-axis represents the nDCG cutoff, ranging from 0 to 10. The y-axis represents the % Gain/Loss in nDCG, ranging from -10 to 20.

- **Baseline Static Model**: This model performs consistently across different nDCG cutoffs, indicating no significant gain or loss.
- **Dynamic Model**: This model shows a steady improvement as the nDCG cutoff increases, indicating better performance.
- **Dynamic Model + Change Prior**: This model performs significantly better than the Baseline model and Dynamic model, reaching the highest % Gain in nDCG.
- **Change Prior**: This model also shows improvement compared to the Baseline model but not as much as the Dynamic Model + Change Prior.

Stanford InfoSeminar 3/9/12
Temporal Retrieval Models 2
(behavior-based)

- **Initial evaluation**
  - Navigational queries; assume relevance is “static” over time

- **But, relevance often changes over time**
  - E.g., *Super Bowl* -- in 2012 vs. in 2011
  - E.g., *US Open 2012* -- in May (golf) vs. in Sept (tennis)
  - E.g., *March madness 2012* -- before/during/after event
    - Before event: Schedule and tickets, e.g., stubhub
    - During event: Real-time scores, e.g., espn, cbssports
    - After event: General sites, e.g., wikipedia, ncaa

- **Current evaluation**
  - Collect explicit and implicit relevance judgments, query frequency, interaction data, and page content over time
Relevance over Time

- **Query:** football [season Sep - Jan]
- **Need to model time of query, pages and events**

Stanford InfoSeminar 3/9/12
Relevance over Time

- **Query:** sigir
- **Why is old content ranked higher?**
  - User interaction data more prevalent for older documents
  - E.g., query-clicks, anchor text, etc.
- **Need to weight user behavior signals appropriately**
Experimental Setup

- **Data**
  - Queries and clicked URLs, over 4 months

- **Types of queries**
  - General
  - Periodic
  - Temporal (Dynamic, Alternating, Temporal Reformulation)

- **Ground truth**
  - Actual user search behavior over time (implicit measure)

- **Model**
  - Temporal dynamics of behavior

- **Use**
  - Model to improved ranking
Time Series Modeling

- Model search behavior as time series
  - Assume that the series of behavioral observations $Y_1 \ldots Y_n$ is generated sequentially based on some underlying structure (e.g., a sequence of state vectors)

- Linear State Space Model (SSM)
  - Let $X_t$ be a state vector at moment of time $t$, then a semi-linear state space model is defined by:
    
    $$ Y_t = w(\theta)X_t + \epsilon_t $$  
    (observation eqn.)
    $$ X_t = F(\theta)X_{t-1} + G(\theta)\epsilon_t $$  
    (state transition eqn.)

- Model state with Holt-Winters
  - Smoothing
  - Trend (+Level)
  - Periodic/Seasonal
Experimental Details

- **Train**: Learn time series models
- **Predict**: Future query and click behavior
- **Ranking models**
  - Predicted clicks as the *only* feature for ranking
  - Temporal features (+other features) as input to learned ranker
- **Three types of features**
  - No user behavior (i.e., just content)
  - Historical average of user behavior
    - Uniform, Linear, Power
  - Temporal models of user behavior
    - Smoothing, +Trend, +Trend+Periodicity
- **Measure**: Correlation (predicted vs. actual) rankings
Experimental Results

- Predicted clicks as the only feature

<table>
<thead>
<tr>
<th>Query Type</th>
<th>Baselines</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
</tr>
<tr>
<td>General</td>
<td>0.91</td>
</tr>
<tr>
<td>Tail</td>
<td>0.18</td>
</tr>
<tr>
<td>Periodic</td>
<td>0.91</td>
</tr>
<tr>
<td>Dynamic</td>
<td>0.28</td>
</tr>
<tr>
<td>Alternating</td>
<td>0.80</td>
</tr>
<tr>
<td>Temp Reform</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Table 2: Pearson correlation on ordering of our temporal models compared to baseline models. Statistically significant differences based on a paired t-test (p < .05) are shown in bold.

- Ranker trained with temporal features

- Best-performing queries

![Table 2: Pearson correlation on ordering of our temporal models compared to baseline models. Statistically significant differences based on a paired t-test (p < .05) are shown in bold.](image1)

![Table 4: Pearson Correlation on ranking using Base features without user behavior, with statistical significance using our temporal models. Statistically significant differences based on a paired t-test (p < .05) are shown in bold.](image2)

![Figure 6: Dominant query shapes for queries where temporal model yielded better rankings than baseline rankers.](image3)
Temporal IR Summary

- **Goal:** Improve Web retrieval by modeling temporal dynamics

- **Content-based models**
  - Rate of page change
  - Detailed term-level changes

- **Behavior-based models**
  - Query frequency over time
  - Click patterns over time

- **Ongoing work**
  - Combine content and behavior features
  - Surprise detection
  - Snippet generation
  - ...

Stanford InfoSeminar 3/9/12
Other Examples of Dynamics and Information Systems

- **Temporal retrieval models**
  - Radinski et al. (submitted); Elsas & Dumais (2010); Liu & Croft (2004); Efron (2010); Aji et al. (2010)

- **Document dynamics, for crawling and indexing**
  - Adar et al. (2009); Cho & Garcia-Molina (2000); Fetterly et al. (2003)

- **Query dynamics**
  - Kulkarni et al. (2011); Jones & Diaz (2004); Diaz (2009); Kotov et al. (2010)

- **Extraction of temporal entities within documents**

- **Protocol extension for retrieving versions over time**
  - E.g., Memento (Van de Sompel et al., 2010)
Web content changes: page-level, term-level

Relating revisitation and change allows us to
- Identify pages for which change is important
- Identify interesting components within a page

People revisit and re-find Web content

Temporal IR:
Leverages change for improved IR

Diff-IE:
Supports (and influences) interaction and understanding
Challenges and Opportunities

- Temporal dynamics are pervasive in information systems
- Influence many aspect of information systems
  - Systems: protocols, crawling, indexing, caching
  - Document representations: meta-data generation, information extraction, sufficient statistics at page and term-level
  - Retrieval models: term weights, document priors, etc.
  - User experience and evaluation
- Better supporting temporal dynamics of information
  - Requires digital preservation and temporal metadata extraction
  - Enables richer understanding of the evolution (and prediction) of key ideas, relations, and trends over time
- Time is one important example of context for IR
  - Others include: location, individuals, tasks ...
Think Out of Search (Research) Boxes

User Context

Query Words

Document Context

Task/Use Context

Ranked List
Thank You!

- Questions/Comments ...

Diff-IE ... try it!