NEWS RECOMMENDATION
(REAL-TIME, CROSS-LINGUAL, SOCIAL MEDIA)

Marko Grobelnik, Blaž Fortuna
Artificial Intelligence Laboratory
Jožef Stefan Institute

Contributors: Pat Moore, Andrej Muhic, Jan Rupnik, Tadej Stajner

LSOLDM Workshop, Cumberland Lodge, Sep 24th 2013
Outline

- Real-time personalized news recommendation
- Cross-lingual news recommendation
- News recommendation for social media
Exclusive: Banks in Libor probe consider group settlement-sources

By Katharina Bart and Diane Bartz
ZURICH/WASHINGTON | Fri Jul 20, 2012 4:21am EDT

(Reuters) - A group of banks being investigated in an interest-rate rigging scandal are looking to pursue a group settlement with regulators rather than face a Barclays-style backlash by going it alone, people familiar with the banks’ thinking said.

Such discussions are preliminary, and it is unclear if regulators will enter these talks, aimed at resolving allegations that banks attempted to manipulate the London interbank offered rate, or Libor, a benchmark that underpins hundreds of trillions of dollars in contracts.

Still, there are powerful incentives for the banks to enter joint negotiations.

Barclays Plc was the first to settle with U.S. and British regulators, paying a $453 million penalty and admitting to its role in a deal announced June 27. Its chief executive, Bob Diamond, abruptly quit the next week, bowing to public pressure and erosion of the bank’s reputation.

The sources told Reuters that none of the banks involved now want to be second in line for fear that they will get similarly hostile treatment from politicians and the public. Bank discussions about a group settlement initially took place before the Barclays agreement, and picked back up in the aftermath.

It is unclear which banks are involved in the potential settlement talks. More than a dozen banks are being investigated in the scandal, including Citigroup, HSBC, Deutsche Bank and JPMorgan Chase. They all declined

Related News

Central bankers eyeing whether Libor needs scrapping
Thu, Jul 19 2012

Geithner defends U.S. response to 2008 Libor concerns
Wed, Jul 18 2012

BoE’s King: Barclays was in denial over Libor manipulation
Tue, Jul 17 2012

Bernanke says Fed’s hands were tied on Libor
Tue, Jul 17 2012

UPDATE 4-BoE’s King tells how Bob Diamond took a bullet
Tue, Jul 17 2012
Coney Island, Renewed but Recognizable

Laura Friedman came up with a list of things she wanted to do before leaving Brooklyn for law school in California, and near the top was an outing at Coney Island.

So the other day, she revisited the restored Wonder Wheel and the Steeplechase roller coaster in Coney Island’s spruced-up amusement park and chomped on a cob of corn from one of its legendary food stands. Afterward she enjoyed a beer with a friend at Ruby’s Bar and Grill, an 81-year-old Boardwalk dive that itself has been refurbished as part of Coney Island’s revitalization.

“Having never been here before, it still feels honky-tonk to me, but in a good way,” said Ms. Friedman, 23. “I like a little bit of grit.”

Shoppers who feared that replacing or freshening the woobagno Boardwalk shops, restaurants and rides would mean the Disneyfication of Coney Island — a swapping of its distinctive sizzle for a more standardized grit — may have to wait awhile to see their fears realized. While the overhaul engineered by the city is evident throughout the
Americans Hold Dimmest View On Economic Outlook Since January

The most Americans in six months said the economy in July was getting worse, indicating the slowdown in hiring is dimming moods as the third quarter begins.

The share of households viewing the U.S. as heading in the wrong direction rose to 36 percent, the highest since January, from 33 percent in June. The Bloomberg monthly expectations gauge was minus 11, matching June as the lowest level since January. The weekly Bloomberg Consumer Comfort Index fell to minus 37.9 in the period ended July 15, the lowest in a month.

Limited wage gains and unemployment stuck above 8 percent risk further slowing consumer spending and leaving the U.S. more vulnerable to a global slowdown. There is also growing pessimism that is being done in Washington to avoid the so-called fiscal cliff at the end of the year, when higher taxes and automatic spending cuts kick in, raising the risk of recession.

“A soft labor market and political tensions surrounding potential changes in tax policy are weighing on consumer sentiment,” said Joseph Brusuelas, a senior economist at Bloomberg LP in New York. “Consumers are concerned about their incomes and have become much more cautious about spending. The economy is limping into the third quarter.”

---

Read Jeffrey Goldberg Now
Why news recommendation?

- “Increase in engagement”
  - Keeping a user on a web site
- “User experience”
  - Users return to the site

Predominant success metric is the **attention span** of a user expressed in terms of time spent on site and number of page views.
Why news recommendation is hard?

- Cold start
  - Recent news articles have little usage history
  - Harder for articles not accessible from a homepage or section fronts
    - ...but are still relevant for a particular user segment

- Recommendation model must be able to generalize over to new articles
Example: Bloomberg.com

- Access logs analysis shows, that half of the articles read are less than 10 hours old
Example: Bloomberg.com

- Access logs analysis shows, that half of the articles read are less then 10 hours old
- Weekends are exception
Approaches

• Keep track of users’ history

• Generate recommendations using
  • Collaborative features
    • “Users that read similar articles also read ...”
  • Content features
    • Recommend semantically/topically similar articles
    • Typically relies on article categories
    • Meta-data
  • Demographic features
    • ...not always available
NEWS RECOMMENDATION
CASE STUDY
News Recommender Architecture

Logging

Recommendation Engine (RecEngine)

Archive

Web

Amazon
News Recommender Architecture

Log visit

Web

Amazon

Recommendation Engine (RecEngine)

Logging

Archive
News Recommender Architecture

Log visit → Logging

Logging → Store

Store → Archive

Web

Recommendation Engine (RecEngine)

Amazon
News Recommender Architecture

Web

Log visit

Logging

Update index

Recommendation Engine (RecEngine)

Store

Archive

Amazon
News Recommender Architecture

- Web
- Amazon

- Recommendation Engine (RecEngine)
- Logging
- Archive
News Recommender Architecture

Web

Amazon

Recommendation Engine (RecEngine)

Logging

Archive
News Recommender Architecture

- **Web**
- **Recommend**
- **Amazon**

- **Recommendation Engine (RecEngine)**
- **Logging**
- **Archive**
Architecture Considerations

- Speed is important
  - Speed of generating recommendations (<100ms)
  - Speed of updating user profiles (in minutes)

- Integration
  - Tight integration with CMS vs. loose JavaScript integration
Data Collection

- JavaScript code on each article and video page
  - New visitors are tagged using a cookie
  - JavaScript creates a link to a single-pixel image
  - Results in one log entry per visit
- Usage logs gathered on Amazon S3
  - Updates sent to RecEngine once per minute
- Article crawling
  - Article is crawled and preprocessed by RecEngine when first spotted in logs
Example


Request header fields:


User-Agent: Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.11 (KHTML, like Gecko) AppleWebKit/536.11 Chrome/20.0.1132.57 Safari/536.11
Example

User ID cookie: 1234567890
IP: 95.87.154.251 (Ljubljana, Slovenia)
Requested URL:
Referring URL: http://www.bloomberg.com/
Date and time: 2009-08-25 08:12:34
Device: Chrome, Windows, PC
Sample User profile

- History
  - Time
  - Article
Sample User profile

- History
  - Time
  - Article

History:
- Top 2% Not Job Creators or Millionaires in Tax Debate
- Looming Copper Surplus Contracting as Mining Fails: Commodities
- Mayer Becomes Highest-Profile Pregnant Woman Hired as CEO
- Zuckerberg’s Loan Gives New Meaning to the 1%
- Brees and Saints Agree on New Contract 3 Days Prior to Deadline
- JPMorgan's Drew Forfeits 2 Years' Pay as Managers Ousted
- JPMorgan's $4.4 Billion CIO Loss Drives Profit Down 9%
- JPMorgan’s London Whale Could Use New Nickname
Sample User profile

- History
  - Time
  - Article
- Current request:
  - Location
  - Requested page
  - Referring page
  - Local Time
  - Device/Agent

History:
- Top 2% Not Job Creators or Millionaires in Tax Debate
- Looming Copper Surplus Contracting as Mining Fails: Commodities
- Mayer Becomes Highest-Profile Pregnant Woman Hired as CEO
- Zuckerberg’s Loan Gives New Meaning to the 1%
- Brees and Saints Agree on New Contract 3 Days Prior to Deadline
- JPMorgan’s Drew Forfeits 2 Years’ Pay as Managers Ousted
- JPMorgan’s $4.4 Billion CIO Loss Drives Profit Down 9%
- JPMorgan’s London Whale Could Use New Nickname
Sample User profile

- History
  - Time
  - Article
- Current request:
  - Location
  - Requested page
  - Referring page
  - Local Time
  - Device/Agent

### History:
- Top 2% Not Job Creators or Millionaires in Tax Debate
- Looming Copper Surplus Contracting as Mining Fails: Commodities
- Mayer Becomes Highest-Profile Pregnant Woman Hired as CEO
- Zuckerberg’s Loan Gives New Meaning to the 1%
- Brees and Saints Agree on New Contract 3 Days Prior to Deadline
- JPMorgan's Drew Forfeits 2 Years' Pay as Managers Ousted
- JPMorgan’s $4.4 Billion CIO Loss Drives Profit Down 9%
- JPMorgan’s London Whale Could Use New Nickname

### User and current visit:
- **User:** 20208908727046187
- **Country:** US [NC]
- **City:** High Point
- **#Visits:** 8
Recommendation Algorithm
Recommendation Algorithm
Recommendation Algorithm
Recommendation Algorithm
Recommendation Algorithm

time
Recommendation Algorithm
Recommendation Algorithm

$T_0$

time
Recommendation Algorithm

\[ T_0 - \Delta T \]

\[ T_0 \]

\( T_0 \) - \( T_0 - \Delta T \)

Time
Recommendation Algorithm

$T_0 - \Delta T$

$T_0$

Time
Recommendation Algorithm

$T_0 - \Delta T$

$T_0$

Time
Recommendation Algorithm

\[ T_0 - \Delta T \]

\[ T_0 \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]

\[ <, >= \]
Recommendation Algorithm

\[ T_0 - \Delta T \]

\[ \text{time} \]

\[ \text{ranking} \]
Recommendation Algorithm

\[ T_0 - \Delta T \]

\[ T_0 \]

time
Recommendation Algorithm

\[ T_0 - \Delta T \]

\[ T_0 \]
Features

- Features:
  - Content – *what a user is reading?*
  - Collaborative – *what similar users are reading?*
  - Demographic – *what we know about a user?*
CONTENT FEATURES
What do we know about an article?

- **News-source:**
  - www.bloomberg.com

- **Article URL:**

- **Author:**
  - Elizabeth Lopatto

- **Produced at:**
  - New York

- **Editor:**
  - Reg Gale

- **Publish Date:**
  - Jan 17, 2011 6:00 AM

- **Topics:**
  - U.S., Health Care, Media, Technology, Science

---

About 9 percent of children play such long hours of video games that they are pathological gamers, increasing risks of anxiety, depression, bad grades and social phobia, a study in Singapore found.

The compulsive gamers played for a weekly average of 31 hours compared with 19 for kids not deemed pathological, according to research released today by the journal *Pediatrics*. Overall, 83 percent of 3,034 children in the study played video games at least occasionally.

Gamers are considered pathological when their playing interferes with everyday life, and their behavior is described as being similar to that of gambling addicts, according to background information in the paper. The gaming isn’t merely a symptom of disorders such as depression, anxiety and social phobia, today’s study found. Rather, gaming can cause and reinforce those maladies.

“Although children who are depressed may retreat into gaming, the gaming increases the depression,” wrote the study authors, led by Douglas A. Gentile, a psychologist at Iowa State University, in Ames.

The study, of children in grades 3, 4, 7 and 8, lasted two years. Kids who stopped being pathological gamers during the study period showed lower levels of depression, anxiety and social phobia compared with peers who didn’t stop, the researchers said.

To contact the reporter on this story: Elizabeth Lopatto in New York at elopatto@bloomberg.net.

To contact the editor responsible for this story: Reg Gale at rgale5@bloomberg.net.
What can be derived from content?

Topics (e.g. DMoz):
- Health/Mental Health/
- Health/Mental Health/Disorders/Mood
- Games/Game Studies

Keywords (e.g. DMoz):
- Health, Mental Health, Disorders, Mood, Games,
- Video Games, Depression, Recreation, Browser
- Based, Game Studies, Anxiety, Women, Society,
- Recreation and Sports

Locations:
- Singapore (sws.geonames.org/1880252/)
- Ames (sws.geonames.org/3037869/)

People:
- Duglas A. Gentile

Organizations:
- Iowa State University (dbpedia.org/resource/Iowa_State_University)
- Pediatrics (journal)

---

Video Gamers’ Prolonged Play Raises Risk of Depression, Anxiety

By Elizabeth Lopatto - Jan 17, 2011 6:00 AM GMT+0100

About 9 percent of children play such long hours of video games that they are pathological gamers, increasing risks of anxiety, depression, bad grades and social phobia, a study in Singapore found.

The compulsive gamers played for a weekly average of 31 hours compared with 19 for kids not deemed pathological, according to research released today by the journal Pediatrics. Overall, 83 percent of 3,034 children in the study played video games at least occasionally.

Gamers are considered pathological when their playing interferes with everyday life, and their behavior is described as being similar to that of gambling addicts, according to background information in the paper. The gaming isn’t merely a symptom of disorders such as depression, anxiety and social phobia, today’s study found. Rather, gaming can cause and reinforce those maladies.

“Although children who are depressed may retreat into gaming, the gaming increases the depression,” wrote the study authors, led by Douglas A. Gentile, a psychologist at Iowa State University in Ames.

The study, of children in grades 3, 4, 7 and 8, lasted two years. Kids who stopped being pathological gamers during the study period showed lower levels of depression, anxiety and social phobia compared with peers who didn’t stop, the researchers said.

To contact the reporter on this story: Elizabeth Lopatto in New York at elopatto@bloomberg.net.

To contact the editor responsible for this story: Reg Gale at rgale5@bloomberg.net.
Extracting triples from text and linking to LOD (DBpedia, OpenCyc, Yago) with Enrycher (http://enrycher.ijs.si/)

Diego Maradona Semantics:
owl:sameAs: http://dbpedia.org/resource/Diego_Maradona
owl:sameAs: http://sw.opencyc.org/concept/Mx4rvofERZwpEbGdrcN5Y29ycA
rdfs:type: http://dbpedia.org/class/yago/ArgentinaInternationalFootballers
rdfs:type: http://dbpedia.org/class/yago/ArgentineExpatriatesInItaly
rdfs:type: http://dbpedia.org/class/yago/ArgentineFootballManagers
rdfs:type: http://dbpedia.org/class/yago/ArgentineFootballers

Robbie Keane Semantics:
owl:sameAs: http://dbpedia.org/resource/Robbie_Keane
rdfs:type: http://dbpedia.org/class/yago/CoventryCityF.C.Players
rdfs:type: http://dbpedia.org/class/yago/ExpatriateFootballPlayersInItaly
rdfs:type: http://dbpedia.org/class/yago/F.C.InternazionaleMilanoPlayers

“Enrycher” is available as a web-service generating PoS tags, Deep parsing, SPO-Graph, LOD links, Entities, Keywords, DMOZ-Categories, Summarization, Sentiment
Content features

#1 – Bag-of-words
- Article title and body

#2 – Bag-of-”meta-data”
- Author
- Editorial categories
- Automatic categories
- Extracted entities
COLLABORATIVE FEATURES
Article Popularity

- Number of page views

Problems:
- Bias towards older articles
- Bias towards homepage and section front articles
Co-visited articles
Co-visited articles
Co-visited articles
Co-visited articles

time
Co-visited articles
Co-visited articles
Co-visited articles

\[ T - \Delta t \] \hspace{1cm} \text{browsing session} \hspace{1cm} \[ T + \Delta t \]
Co-visited articles

\[ \{ \} \]

\[ T - \Delta t \quad \text{browsing session} \quad T + \Delta t \]

\( \text{time} \)
Article readers
Collaborative features

#3 – Bag-of-“articles”
- Articles read within same browsing session

#4 – Bag-of-“readers”
- Users who read the article

#5 – Popularity
- Number of page views
DEMOGRAPHIC FEATURES
User Data (NYTimes)

- Provided only for registered users
  - ~20% unique users in our case
  - Can we generalize to other users?

- Each registered users described with:
  - Gender
  - Year of birth
  - Household income

- Noisy
User Data (NYTimes)

- Provided only for registered users
  - ~20% unique users in our case
  - Can we generalize to other users?

- Each registered user is described with:
  - Gender
  - Year of birth
  - Household income

- Noisy
User Modeling

- **Feature space**
  - Extracted from subset of fields
  - Using vector space model
  - Vector elements for each field are normalized

- **Training set**
  - One visit = one vector
  - One user = a centroid of all his/her visits
  - Users from the segment form positive class
  - Sample of other users form negative class

- **Classification algorithm**
  - Support Vector Machine
  - Good for dealing with high dimensional data
  - Linear kernel
  - Stochastic gradient descent
    - Good for sampling
Estimating user’s “Gender”

- Male
  - ≥2: 50.00%
  - ≥10: 55.00%
  - ≥50: 60.00%
- Female
  - ≥2: 65.00%
  - ≥10: 70.00%
  - ≥50: 75.00%
Estimating user’s “Age”

<table>
<thead>
<tr>
<th>Age Range</th>
<th>≥2</th>
<th>≥10</th>
<th>≥50</th>
<th>#visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-30</td>
<td></td>
<td></td>
<td>35%</td>
<td>45%</td>
</tr>
<tr>
<td>31-40</td>
<td>30%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41-50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51-60</td>
<td>20%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61-80</td>
<td>40%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Estimating user’s “Income” (≥10 visits)
EVALUATION
Performance comparison

Our solution

Random

Old solution

Control, JSI SVM, Random, JSI CF, DailyMe Personalized, Most Popular, Contextual
Which features matter?

- SVM weights for combining features:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2-10</th>
<th>11-50</th>
<th>51-</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Popularity</strong></td>
<td>1.28</td>
<td>0.58</td>
<td>-0.10</td>
<td>-0.30</td>
</tr>
<tr>
<td><strong>Content [h]</strong></td>
<td>0.17</td>
<td>0.85</td>
<td>1.09</td>
<td>0.66</td>
</tr>
<tr>
<td><strong>Meta-data [h]</strong></td>
<td>0.31</td>
<td>0.58</td>
<td>0.72</td>
<td>1.09</td>
</tr>
<tr>
<td><strong>Co-visits [h]</strong></td>
<td>0.67</td>
<td>0.39</td>
<td>-0.40</td>
<td>-0.29</td>
</tr>
<tr>
<td><strong>Users [h]</strong></td>
<td>4.45</td>
<td>10.33</td>
<td>11.59</td>
<td>8.56</td>
</tr>
<tr>
<td><strong>Content [r]</strong></td>
<td>0.17</td>
<td>-0.06</td>
<td>-0.47</td>
<td>-0.30</td>
</tr>
<tr>
<td><strong>Meta-data [r]</strong></td>
<td>0.31</td>
<td>-0.04</td>
<td>-0.20</td>
<td>-0.19</td>
</tr>
<tr>
<td><strong>Co-visits [r]</strong></td>
<td>0.67</td>
<td>1.03</td>
<td>1.35</td>
<td>1.23</td>
</tr>
<tr>
<td><strong>Users [r]</strong></td>
<td>4.45</td>
<td>1.95</td>
<td>0.24</td>
<td>1.05</td>
</tr>
</tbody>
</table>
### Which features matter?

- SVM weights for combining features:

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>1</th>
<th>2-10</th>
<th>11-50</th>
<th>51-100</th>
</tr>
</thead>
<tbody>
<tr>
<td># history</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Popularity</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content [h]</td>
<td>0.17</td>
<td>0.58</td>
<td>1.09</td>
<td>0.66</td>
</tr>
<tr>
<td>Meta-data [h]</td>
<td>0.31</td>
<td>0.72</td>
<td>1.09</td>
<td>-0.29</td>
</tr>
<tr>
<td>Co-visits [h]</td>
<td>0.07</td>
<td>0.39</td>
<td>0.40</td>
<td>-0.30</td>
</tr>
<tr>
<td>Users [h]</td>
<td>10.33</td>
<td>11.59</td>
<td>8.56</td>
<td></td>
</tr>
<tr>
<td>Content [r]</td>
<td>0.17</td>
<td>-0.06</td>
<td>-0.47</td>
<td>-0.30</td>
</tr>
<tr>
<td>Meta-data [r]</td>
<td>0.31</td>
<td>-0.04</td>
<td>-0.20</td>
<td>-0.19</td>
</tr>
<tr>
<td>Co-visits [r]</td>
<td>0.67</td>
<td>1.03</td>
<td>1.35</td>
<td>1.23</td>
</tr>
<tr>
<td>Users [r]</td>
<td>4.45</td>
<td>1.95</td>
<td>0.24</td>
<td>1.05</td>
</tr>
</tbody>
</table>

Two main observations:
- First time users tend to react on ‘what is popular’
- Loyal users react on particular topics of interest
Conclusions

- Diverse features help
  - Pick as many signals as possible

- Speed is important
  - Model pre-computed (RankSVM)
  - Recent article feature vectors updated each few minutes
  - User feature vectors computed on the fly
    - ...up to 50ms per user with full feature set

- Actual performance depends also on other factors
  - ...e.g. page design (placement of recommendations)
CROSS-LINGUAL NEWS RECOMMENDATION
What is cross-lingual recommendation?

- Providing alternative views for a given topic
  - ...cross-linguality is needed in countries with different language
- Publisher can define itself as a hub for news
- Big publishers’ fear is sending users away from their site – will the users come back?
  - ...evaluation shows, the fear is not necessarily justified
Taxes to Rise for Workers as Budget Deal Still Elusive

By Kathleen Hunter & Roxana Tiron - Dec 31, 2012 2:36 PM GMT+0100

With taxes set to increase for almost every U.S. worker at midnight, Congress hasn't reached a budget deal that Democrats and Republicans say is necessary to prevent a blow to the U.S. economy.

Private talks between Senate Majority Leader Harry Reid and Minority Leader Mitch McConnell that began Dec. 28 stalled yesterday because of disputes over income tax rates, the estate tax and other issues.
How CL-recommendations are being done?

- We collect news from the whole world, knowing for each publisher from which country is coming from
  - ...~40,000 publishers accessible through RSS feeds

- Using cross-lingual similarity we map crawled articles to a target article
  - ...using KCCA trained model for ~100 languages

- Each publisher has its own style of publishing
  - ...crawled articles with a similar content and similar style are recommended
Empirical evaluation of CL-Recommendation

Average Weekly Click Rate Over Time by Test Bucket

Week
1 2 3 4 5 6

Control

XLike CL-recommendation

2.7% 3.1% 2.9% 2.8% 2.5% 2.2%
3.1% 3.9% 3.0% 3.0% 3.2% 3.1%
Content acquisition from RSS feeds
http://newsfeed.ijs.si

Few 100k news stories per day from 40k main stream news sources
Capturing publisher’s style “Bloombergeness”

- Goal: determine if a (e.g. German) document is relevant to topics covered in Bloomberg stream
  - Works for English(Hub)/German/Spanish/French/Italian
  - Mapping to common cross-lingual space
- The result is a number between 0 and 1 telling how close a given non-Bloomberg article is to the Bloomberg style of publishing
  - Kernel density estimation of probability distribution
Comparing publishing styles: Logdensity of Bloomberg and German newsfeed

Plot shows the distributions

X-axis: logdensity

Y-axis: number of documents
NEWS RECOMMENDATION FOR SOCIAL MEDIA
What is news recommendation for social media?

- Publishers push their content through social media channels (like Twitter, Facebook, ...)  
  ...mainly to increase users’ engagement

- The problem is which articles to select at which time for a particular geographical region to maximize click-through  
  ...current relevance in different countries could be significantly different

- Using global real-time news feed and cross-lingual similarity we estimate relevance for a selected moment and for a selected country
Determining article relevance

Relevance of an article to be pushed into a social media channel for a particular region is a combination of several scores:

- Top local topics
- Age of an article in combination with how evergreen the content is
  - (e.g. earnings report have shorter lifecycle compared to opinion articles)
- Last time similar topic was published to the feed
- Popularity of articles according to Bloomberg web logs for monitored regions
## Prototype: Relevant Bloomberg articles for different countries

### Germany
- [862] Pope Francis Got His Name Right
- [599] Francis Toughened by Argentine Politics Ready for Papal Test
- [247] K+S Quarterly Sales Miss Estimate as Late Winter Cut Salt Demand
- [245] HeidelbergCement Debt Falls More Than Expected on Higher Profit
- [231] ASMI Shares Plunge After Plan to Cut Holding in ASM Pacific
- [186] Xi Assumes China Presidency to Cement Rapid Transition of Power
- [183] Irish Banks Face Penalties for Failing to Tackle Arrears
- [162] Generali Reports Fourth-Quarter Loss After Writedowns Increase
- [137] Kenya's Muthaura Didn't Bribe, Threaten Witnesses, Lawyer Says
- [131] Hugo Boss Forecasts Higher 2013 Profit, Led by Growth in Retail

### France
- [2616] Kerry Properties Pays HK$11.7 Billion for Hong Kong Land Site
- [1811] HSBC Boosts Mortgage Rates as Hong Kong Cools Property Market
- [1205] ILH Wins HK$5 Billion Bid to Build Private Hospital in Hong Kong
- [179] Pope Francis Got His Name Right
- [123] Francis Toughened by Argentine Politics Ready for Papal Test
- [106] Henry Tang to Sell $3.7 Million Wine From Personal Cellar
- [97] Wall Street Banks Eligible for Expanded RQFI, PBOC's Li Says
- [63] Hong Kong Stocks Fall Third Day as China's Builders Drop
- [60] Hong Kong Stocks Fall Third Day as China's Builders Drop
- [41] Kenya's Muthaura Didn't Bribe, Threaten Witnesses, Lawyer Says

### Spain
- [291] Pope Francis Got His Name Right
- [148] Netflix Adds Social Options, Facebook Ties for U.S. Users