Classifying the Political Leaning of News Articles and Users From User Votes

Xiaodan Daniel Zhou
Paul Resnick
Qiaozhu Mei
US Political News & Opinion

By Sean Munson

Top US political news and opinion stories from the last 48 hours, based on links from over 500 blogs. Links are highlighted to show the average political leaning of the blogs that are talking about each item, but you can disable this in the settings.

< 100 users | ★★★★★ 0 ratings

Add it now

Share this gadget

http://tinyurl.com/dy53jq
Politics News

172

Why Julian Assange is 2010's Person Of The Year
prosebeforephotos.com — For striking fear into the heart of every average politician and banker and for showing us the gift—and the curse—of too much information, Julian A... 1 hr 32 min ago
via rudymoman 36 Comments Save Bury

208

Elsberg: “EVERY attack now made on WikiLeaks and Julian Assange was made against me and the rele...
elsberg.net — 71 min ago
via jamesg546 36 Comments Save Bury

145

VIDEO: Maddow And Olbermann Slam Obama's Tax Cut Compromise
tpmcc.talkingpointsmemo.com — If you thought you were mad about Obama’s tax cut compromise, you should watch Keith Olbermann and Rachel Maddow slam it. 7 hr 12 min ago
via wonkettemagan 147 Comments Save Bury

Hot Stories on Digg

HOT STORY
The Escapist: Video Galleries: Zero Punctuation: Assassin’s Creed: Brotherhood
escapistmagazines.com

Chosen by digg staff. Learn more.
Text Classification

- (Oh et al, ICWSM 2009)
- SVM
- 87% accuracy on blog posts
Politics News

Did President Obama Receive an Illegal Contribution?
blogs.browardpalmbeach.com — Thanks to a disgraced politician in South Florida, President Obama's campaign may have unknowingly benefited from a $15,000 check. 37 min ago

Poll: Voters like Clinton-backed candidates more than Obama
articles.latimes.com — The poll found that 53% of Democrats said Clinton's personal endorsement would carry weight compared to 48% who said Obama's would. 1 hr 18 min ago

The 'Must Have' Gift For The Holidays
nowspoint.com/holiday — At under 1lb, learn why the Samsung Galaxy Tab with 7" touchscreen, dual cameras, and flash capabilities, is a must for your wish list this season.
Bizarre Avatar-themed weddings take place in China

Avatar-themed weddings in Wulingyuan, a scenic area in Hunan Province, China become popular. 1 day 7 hr ago

via meetjoewhitel <i>1 Comment</i>  
<i>Save</i>  
<i>Bury</i>

2011 Looks Gloomy For Renewable Energy

Natural Resource Report.com — With the 2010 elections behind us, environmental experts worry that the political shift in the U.S. House of Representatives may reduce federal spe... 1 day 6 hr ago

via kimballo <i>7 Comments</i>  
<i>Save</i>  
<i>Bury</i>

A Drop of Water at 2,000 Frames Per Second

That Video Site.com — Not what you would expect to see 1 day 20 hr ago

via wappe <i>6 Comments</i>  
<i>Save</i>  
<i>Bury</i>

The Ultimate Elisha Cuthbert Gallery

Top Cultured.com — Elisha Cuthbert is hot, no doubt about it. But how hot? Hot
Massive Censorship Of Digg Uncovered

A group of influential conservative members of the behemoth social media site Digg.com have just been caught red-handed in a widespread campaign of censorship, having multiple accounts, upvote padding, and deliberately trying to ban progressives. An undercover investigation has exposed this effort, which has been in action for more than one year.
We need to continue the fight... not just on Digg but networking with other conservative sites and getting the word out.

Positive Attitude! We're conservatives and our philosophy is based on the concept that if a person works hard, keeps faith, a positive attitude and his or her name is the groundwork that anything is possible. Liberals are full of envy and the belief that someone owes them. They have no clue what responsibility or independence is. They've been brainwashed into thinking that they are independent thinkers because they don't subscribe to a faith (no offense to any agnostic or atheist among us) while subscribing to faith in government and they don't even realize it. Their religion is governmental statism and they don't realize their philosophy really sucks.
Expanding Coverage By Propagation
Expanding Coverage By Propagation
Expanding Coverage By Propagation
Expanding Coverage By Propagation
A Semi-Supervised Learning Problem

• Some items labeled
• Learn to label the rest
  – Not based on item content features
  – Based on distances/relationships between items
Outline

- Propagation algorithms
- Tuning parameters
- Tuning with additional data sources
- Testing against held out data
Zhou et al 2004

• "Learning with Global Consistency"
Mapping to Our Case

• Choose a continuous (red, blue) score for each node that minimizes a cost function
  – Cost for color distance on adjacent nodes
  – Cost for changing originally labeled nodes

• Equivalently: fixed point of color propagation, with mix of last iteration and original
Desired Behavior (Axiom)

• Color item blue if disproportionately liked by blue people
  – More blues than reds like it

• Color person blue if disproportionately likes blue items
  – Likes more blues than reds
Blue or Red?
Desired Behavior (Axiom)

• Color item blue if disproportionately liked by blue people
  – More blues than reds like it
  – Higher proportion of blues like it than the proportion of reds who like it

• Color person blue if disproportionately likes blue items
  – Likes more blues than reds
  – Likes a higher fraction of blue items than of red items
Random Walk With Restart
Random Walk With Restart

• Classify
  – blue if $\text{Score}_b > \theta \text{Score}_{all}$
  AND NOT $\text{Score}_r > \theta \text{Score}_{all}$
TUNING PARAMETERS
Tuning $\alpha$

- Ten-fold cross-validation
Tuning $\theta$

- Very few labeled nodes as both/neither
  - $\text{Score}_b > \text{Score}_{\text{all}} \iff \text{Score}_r < \text{Score}_{\text{all}}$
- Maximize accuracy at $\theta=1$
  - Everything classified as red XOR blue
- Higher $\theta$ yields more "purple" classifications
  - Might be useful if training data had some purples
TUNING WITH ADDITIONAL DATA SOURCES
Externally Classified Blogs

Blog Directory
The following blogs are currently indexed by Wonkosphere unless prohibited by the blog, as noted.
Counts by classification: Conservative: 813, Liberal: 433, Independent: 63
Know of a blog we should index? Want us to correct something in this list? Use this link.
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z 0-9

"A" Blogs
<table>
<thead>
<tr>
<th>Blog</th>
<th>Classification</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>a bird and a bottle</td>
<td>Liberal</td>
<td></td>
</tr>
<tr>
<td>A Blog Around The Clock</td>
<td>Liberal</td>
<td></td>
</tr>
<tr>
<td>A Blog For All</td>
<td>Conservative</td>
<td></td>
</tr>
<tr>
<td>A Buck for Huck</td>
<td>Conservative</td>
<td></td>
</tr>
<tr>
<td>A Chequer-Board of Nights and Days</td>
<td>Conservative</td>
<td></td>
</tr>
<tr>
<td>A Daily Briefing on Iran</td>
<td>Conservative</td>
<td></td>
</tr>
<tr>
<td>A Few Shiny Pebbles</td>
<td>Conservative</td>
<td></td>
</tr>
<tr>
<td>A Good Choice</td>
<td>Conservative</td>
<td></td>
</tr>
</tbody>
</table>
Items In Those Blogs Inherit Labels
Items In Those Blogs Inherit Labels
Blog Links as Votes
## Additional Datasets

- **Baseline**: only labels for known raters

<table>
<thead>
<tr>
<th>Add</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>“items from labeled blogs”</td>
<td>94.8%</td>
<td></td>
</tr>
<tr>
<td>“blog links as votes”</td>
<td></td>
<td>95.4%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>94.8%</td>
<td>92.9%</td>
</tr>
</tbody>
</table>
Additional Domain Data

• "Domain" nodes and "Friend/follower" links
### Results

<table>
<thead>
<tr>
<th>“friendship” links</th>
<th>“domain” links</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>94.9%</td>
</tr>
<tr>
<td>Yes</td>
<td>95.1%</td>
</tr>
</tbody>
</table>
Weighted Domain Links

- Domain link = $n$ digg votes
TESTING AGAINST HELD OUT DATA
Test Data

• Stories
  – Read and labeled by Turkers

• People
  – User's comments read by research team
'Cadillac Tax' in Health Plan Would Hit Middle Class Hard

7 months ago

Is the article above:

**Liberal?** The article needs to have some part of it that satisfies any of the following:
-- argues from a liberal perspective, and/or uses liberal ideological arguments (e.g., inequality is bad, social safety net is important, government should not interfere with personal life style choices, etc.)
-- presents facts selectively to promote liberal perspective
-- attacks conservative positions, perspectives, or people from a liberal perspective

**Conservative?** The article needs to have some part of it that satisfies any of the following:
-- argues from a conservative perspective, and/or uses conservative ideological arguments (e.g., big government is bad, taxes should be low, government should not interfere with markets, etc.)
-- presents facts selectively to promote conservative perspective
-- attacks liberal positions, perspectives, or people from a conservative perspective

**Both?** The article has a part that satisfies the criteria for "liberal" and a part that satisfies the criteria for "conservative"

**Neither (non-partisan)?** The article needs to satisfy any of the following:
-- does not satisfy criteria of "liberal" or "conservative"
-- presents simple facts with no hint of political opinions

**Broken link?** The article needs to satisfy any of the following:
-- links to nothing
-- links to something other than what its title suggested

Want to work on this HIT?  Want to see other HITs?

[Accept HIT]  [Skip HIT]
Stories Labeled by Mechanical Turkers

- 1000 stories
- Qualification test to hire turkers
- 3 self-identified liberal and 3 self-identified conservative for each story
- Only used 307 labeled stories with 6/6 agreement that are either red or blue (73 red, 234 blue)
Labeling Users

• 220 Users with
  – >15 comments;
  – >5 diggs
  – >½ of recent comments on political stories

• 2 undergrad coders
  – One code for each user
  – Code as red (conservative), blue (liberal), gray (unclear or signs of both)

• Kappa score: 0.69
  – .89 excluding users coded by either as gray
  – 6 pre-labeled users included for quality control
    • all coded correctly

• 131 users where coders agreed on pure red or blue
  – (69 red, 62 blue)
# Accuracy

<table>
<thead>
<tr>
<th>Cross-validation on training data</th>
<th>94.90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>On test data</td>
<td><strong>94.5% overall</strong>, 92.8% for stories, 98.5% for users</td>
</tr>
</tbody>
</table>
Conclusion

- Can propagate political leanings from a few labeled items and people
- Much room for tuning
  - Algorithm: definition of "disproportionately liked"
  - Data to include
    - ✓ Diggs as votes
    - ✓ Source blog ➔ item label
    - x Blog references as votes
    - x Friend links
    - ✓ Source links