

# Exploiting Temporal Locality to Determine User Bias in Microblogging Platforms

---

Pedro H. Calais Guerra, Loïc Cerf, Thiago C. Porto  
Adriano Veloso, Wagner Meira Jr. and Virgilio Almeida

Universidade Federal de Minas Gerais, Brazil

InWeb – Brazilian National Institute of  
Science and Technology for the Web



October 4th, 2011



Blogs, microblogs and online social networks are flooded with real-time opinions about a multitude of topics:

- Politics - *"candidate X is the best"*
- Sports competitions - *"Team Y is favorite to win today!"*
- Other topics and current "buzz"

**Bias** is inherent to most humans [Watson 1991], since they:

- take particular position regarding a subject
- have a personal interest from the arguer in the outcome of the argument or discussion.
- lack proper balance and neutrality in argumentation
- lack proper critical doubt

**Bias** is inherent to most humans [Watson 1991], since they:

- take particular position regarding a subject
- have a personal interest from the arguer in the outcome of the argument or discussion.
- lack proper balance and neutrality in argumentation
- lack proper critical doubt

**Bias and opinions are not independent!**

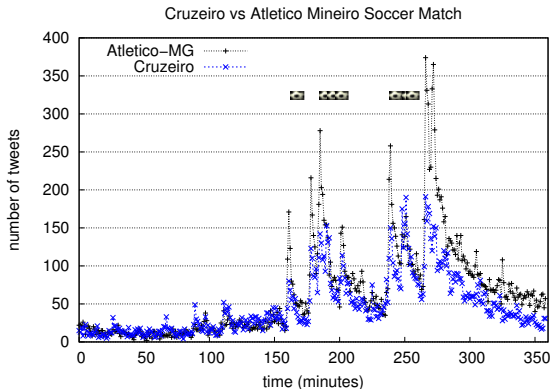
- Supporters of a candidate are likely to issue positive opinions on him/her
- Soccer team supporters act similarly



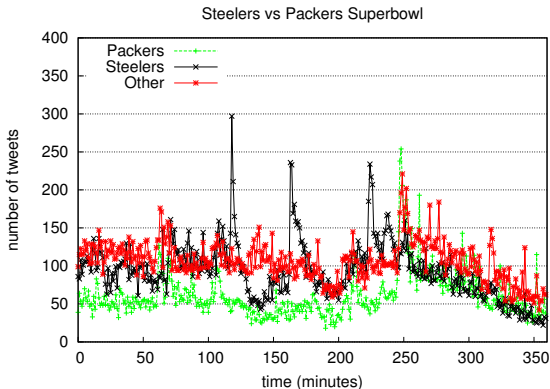
Knowing social media user preferences and inclinations is useful to

- sentiment analysis [Guerra et. al 2011]
- offer personalized content, recommendation

- Web users now may comment political speeches, sports competitions, or any buzz, while they occur
- They **react** to real-life events (e.g., in Twitter)



- Web users now may comment political speeches, sports competitions, or any buzz, while they occur
- They **react** to real-life events (e.g., in Twitter)



- Web users now may comment political speeches, sports competitions, or any buzz, while they occur
- They **react** to real-life events (e.g., in Twitter)

Can temporal behavior unveil what a Twitter user thinks about a topic?

assumption: microblog users having similar biases are prone to send messages at about the same time

- each user is represented as a bias vector  $B_u$  of size  $|C|$ :  
 $[b_1, b_2, b_3, \dots, b_k]$
- for some users, their bias vector is known (supervised approach!)
- we need to estimate the bias vector of unknown users

assumption: microblog users having similar biases are prone to send messages at about the same time

- we divide the time period in intervals of  $X$  minutes each
- for each interval, we compute a vector  $V$  with the count of number of users belonging to class/team in  $V_i$
- the bias vector of user  $U$  is the sum of the bias vector of all intervals in which  $U$  has posted a tweet
- normalization!

---

**Algorithm 1** Learning biases by temporal locality.

---

```

 $S \leftarrow (0, \dots, 0)$ 
while true do
   $U \leftarrow \text{GETUSERSINAGROUP}()$ 
   $k \leftarrow 0$ 
   $V \leftarrow (0, \dots, 0)$ 
  for all  $u \in U$  do
    if  $u$  is a key of  $B_{\text{known}}$  then
       $V \leftarrow V + B_{\text{known}}[u]$ 
       $k \leftarrow k + 1$ 
    end if
  end for
   $S \leftarrow S + k.V$ 
  for all  $u \in U$  do
     $B[u] \leftarrow B[u] + V$ 
  end for
end while

```

---



**Pedro H Calais**  
**@pedrocalais** Belo Horizonte, MG  
*Bacharel e mestre em Ciência da Computação. Aluno de doutorado em Ciência da Computação na UFMG. Mineirinho de sangue azul nas veias. :-)* *Cruzeiro, sempre!*  
<http://www.dcc.ufmg.br/~pcalais>

About @pedrocalais

<b>5,230</b> Tweets	<b>911</b> Following	<b>337</b> Followers
------------------------	-------------------------	-------------------------

Recent Images · [view all](#)

233 posts in Twitter about Brazilian Soccer in 1.5 year (0.42 post/day)



if we divide time periods in intervals of one minute, we will see that **@pedrocalais** usually post together with:

- 30.131 **Cruzeiro** supporters
- 12.191 **Atletico Mineiro** supporters
- 7.211 **Corinthians** supporters
- 6.674 **Flamengo** supporters
- ...

twitter



- Brazilian 2010 Soccer League (**12 classes**)
- 2010 NFL Season (**22 classes**)
- Brazilian 2010 Elections (**2 classes**)

twitter



- 2010 Brazilian Soccer League
  - 35 mi tweets
  - 6.5 mi users
  - 98 tweets / minute

twitter



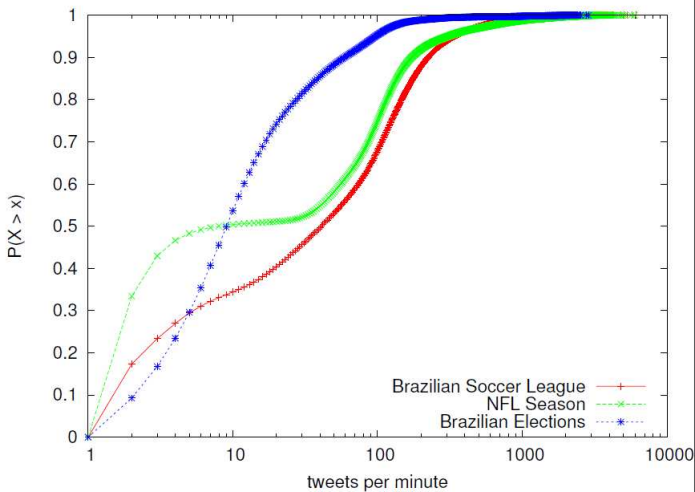
- 2010 NFL Season
  - 23 mi tweets
  - 4.2 mi users
  - 88 tweets / minute

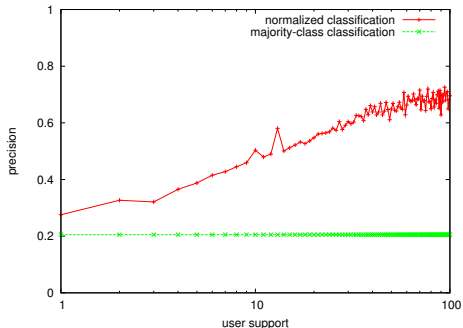
twitter



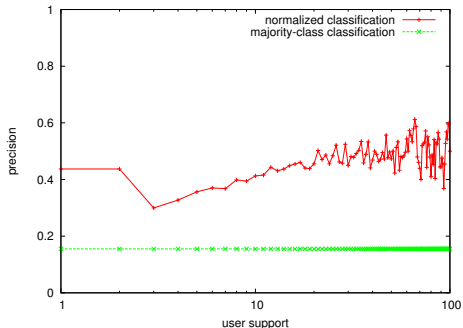
- 2010 Brazilian Presidential Elections
  - 10 mi tweets
  - 1.2 mi users
  - 26 tweets / minute

# Some periods are bursty!

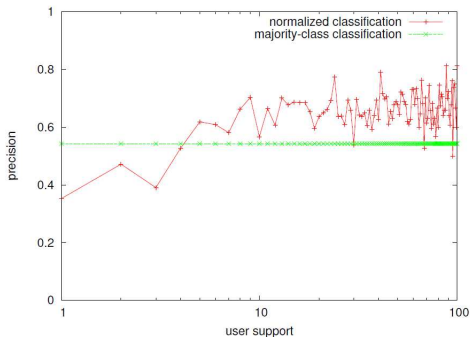




Accuracy of up to 75% for users posting at least 0.3 tweet a day;



Accuracy of up to 60% for users posting at least 0.3 tweet a day



Even on a more “static” scenario, some temporality patterns do exist (retweets!)

Running time to execute our algorithm:

- 10 minutes, 1.5 year data (Soccer)
- 7 minutes, 1 year data (NFL)
- 4 minutes, 1 year data (Elections)

Running time to execute our algorithm:

- 10 minutes, 1.5 year data (Soccer)
- 7 minutes, 1 year data (NFL)
- 4 minutes, 1 year data (Elections)

### Performance vs Accuracy

- Much more efficient than content-based and graph-based methods ( $O(|U|)$  and  $O(|C|)$ )
- Future Work: combining content, social-network data with temporal behavior to achieve best accuracy/performance trade-off

- we show that microbloggers do not express their opinions at random times but follow a temporal “pattern” that reveals their bias;
- we propose an efficient linear time algorithm that solely relies on temporal information to learn these biases;
- we apply and validate our approach using messages from three topics that drive live reactions in the Twitter microblogging platform: soccer, NFL and Elections.

- Combining temporal locality with other evidences of bias
- Building personalized applications: recommendation, advertisement
- Temporal-Context based Sentiment Analysis
- 2012 US Elections and 2012 Olympics

# Exploiting Temporal Locality to Determine User Bias in Microblogging Platforms

---

---

Pedro H. Calais Guerra, Loïc Cerf, Thiago C. Porto  
Adriano Veloso, Wagner Meira Jr. and Virgilio Almeida

Universidade Federal de Minas Gerais, Brazil

InWeb – Brazilian National Institute of  
Science and Technology for the Web