iPS Institute of Policy Studies

Symposium
on
Media and Internet Use
During General Election 2015

Wednesday, 27 January 2016

Conference Room, Level 1, Oei Tiong Ham Building

Lee Kuan Yew School of Public Policy

National University of Singapore

Using Text Analytics in Analysing GE2015 Online Content

Professor Lim Ee-Peng

Director
Living Analytics Research Centre
School of Information Systems
Singapore Management University

Engaging Minds, Exchanging Ideas

Using Text Analytics in Analysing GE2015 Online Content

LIM Ee-Peng
School of Information Systems
Singapore Management University

IPS Symposium on Media and Internet Use During General Election 2015

The Team

- Aek Palakorn, SMU LARC
- Dai Bingtian, SMU LARC
- Adrian Vu, SMU LARC

- Tan Tarn How, NUS IPS
- Carol Soon, NUS IPS
- Siti Nadzirah Binte Samsudin, NUS IPS
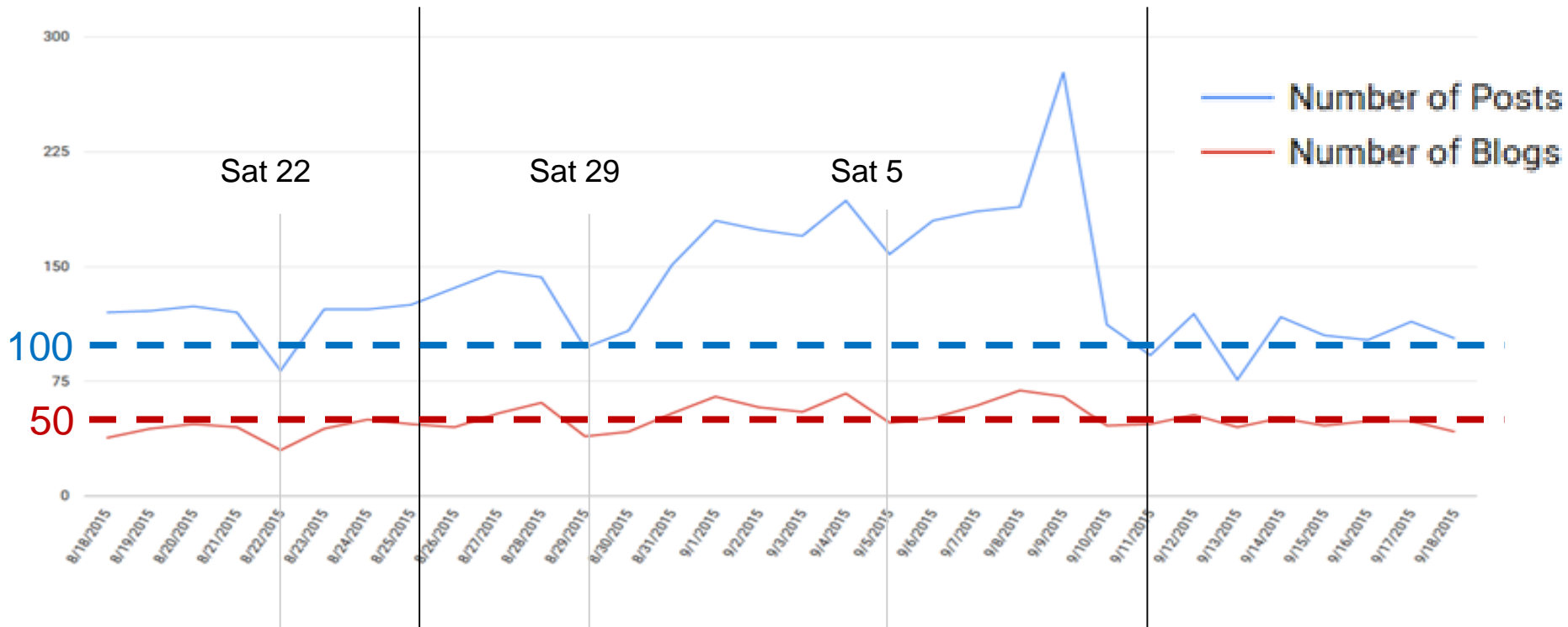- Shawn Goh Ze Song, NUS
- Valerie Yeo Jiewen, NUS

Motivation

- Manual judgement requires much time and effort.

- Online content is generated at a fast pace.
  - More than 200 blogs relevant to Singapore
  - More than 100 posts each day

- During GE2015, as many as 275 posts were generated in one single day.

Overall Trends
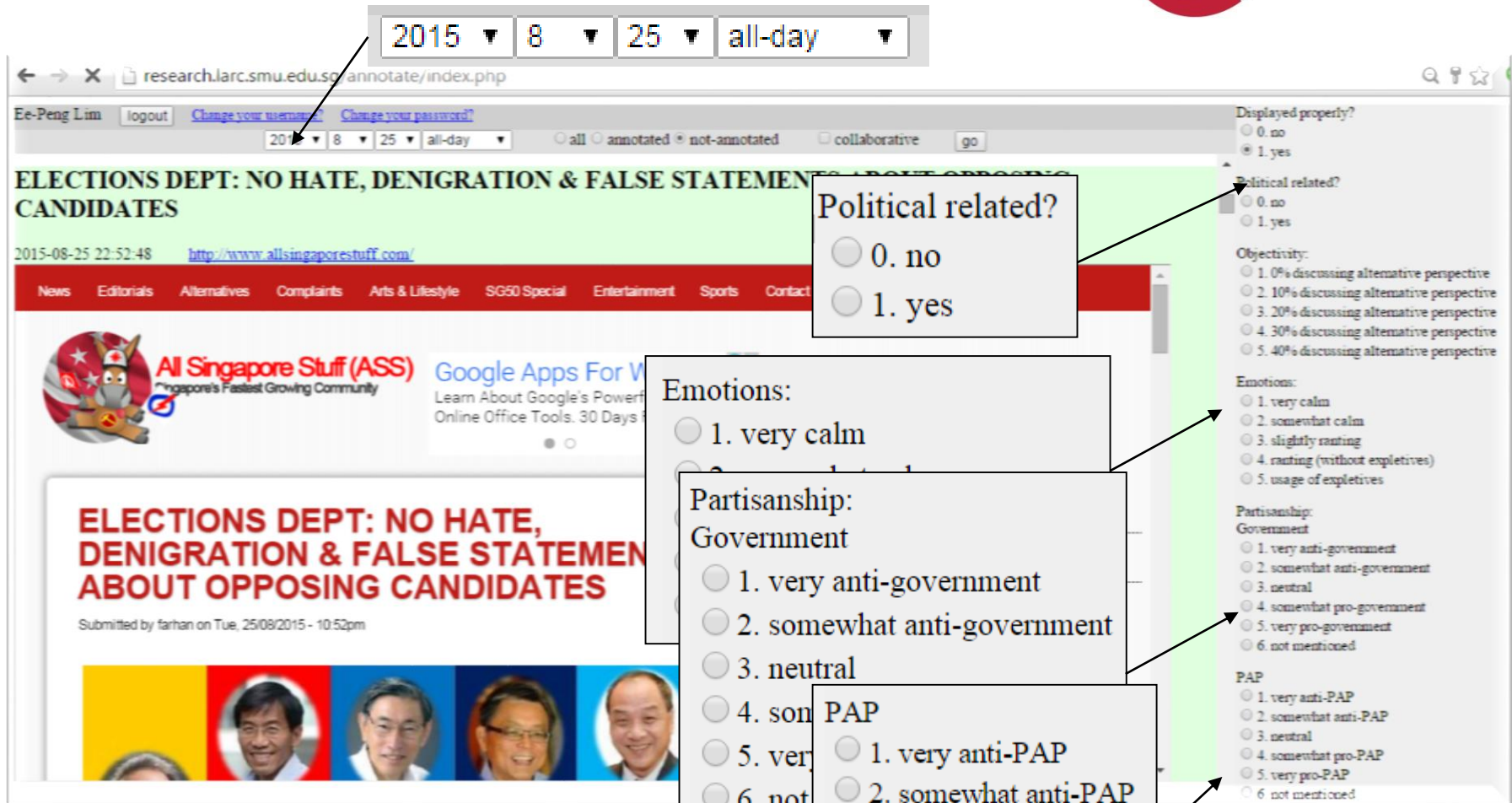
4365 blog posts were generated during the above period.

Research Questions

- Can text analytics help?
  - Need human experts to provide some training data

- Can we derive useful insights quickly from online content using text analytics?

- Can we create text analytics systems to help us find interesting content?

Blog Post Coding Tool

Coding Period: May 20 -

Coding Schemes

Scale	Political?	Label
0	No	Non-Political
1	Yes	Political

Scale	Partisanship	Label
1	Very anti-government/PAP	Partisan
2	Somewhat anti-government/PAP	Partisan
3	Neutral	Non-Partisan
4	Somewhat pro-government/PAP	Partisan
5	Very pro government/PAP	Partisan

Coding Schemes

Scale	Emotionality	Label
1	Very calm	Non-Emotional
2	Somewhat calm	Non-Emotional
3	Slightly ranting	Emotional
4	Very ranting	Emotional
5	Usage of expletives	Emotional

Training The Classifiers

- Using the human labeled posts before GE2015 was announced, we trained classifiers to assign **political, partisanship** and **emotionality** labels using blog post content.
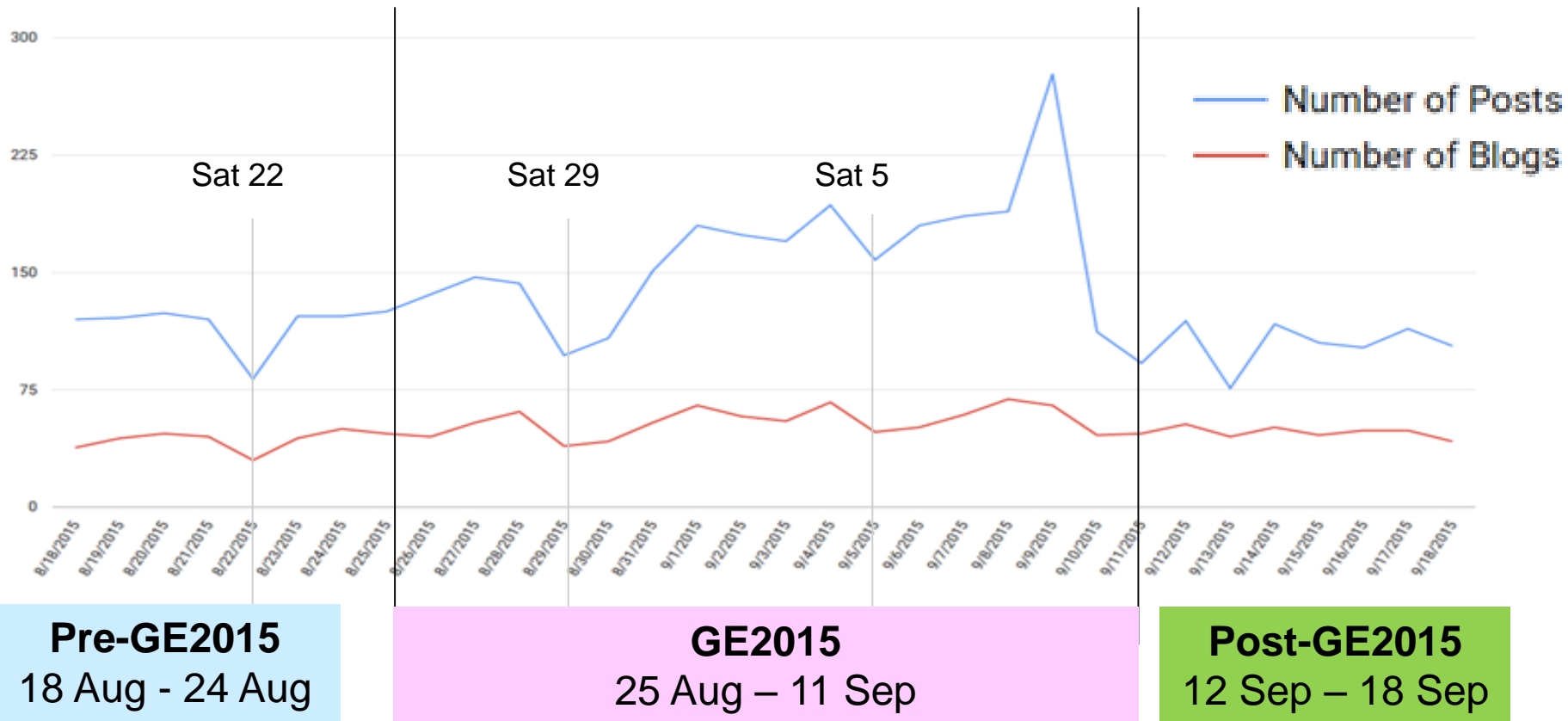
Labeled Data	Size (num of docs)	# Positive Examples	# Negative Examples	Accuracy (F1,10-CV)
Political	914	476	438	0.86
Partisan	461	278	183	0.78
Emotionality	462	174	289	0.71

- We then applied the classifiers to predict **political, partisanship** and **emotionality** labels to the posts during the GE2015 period.

Overall Trends

Classified Posts

Classified Posts	Pre-GE2015 18 Aug - 24 Aug (# Posts)	GE2015 25 Aug - 11 Sep (# Posts)	Post-GE2015 12 Sep - 18 Sep (# Posts)
Political Posts	393	1756	353
Political Posts (% of all posts)	55%	74%	56%

Total Number of Blog Posts: 4365

- About 55% of all posts are political before and after GE2015.

- Proportion of political posts increases significantly to 74% during GE2015.

Classified Posts

Label	Pre-GE2015 18 Aug - 24 Aug (# Posts)	GE2015 25 Aug - 11 Sep (# Posts)	Post-GE2015 12 Sep - 18 Sep (# Posts)
Political	393	1756	353
Partisan	315 (80%)	1426 (81%)	315 (89%)
Emotional	131 (33%)	664 (38%)	189 (54%)

(x% of classified political posts)

- Most political posts are partisan oriented.
  - Partisan posts increase slightly after GE2015.
- About 1/3 political posts are emotional.
  - Emotional posts increase significantly after GE2015.

Trends: Political, Partisan, and Emotional Posts
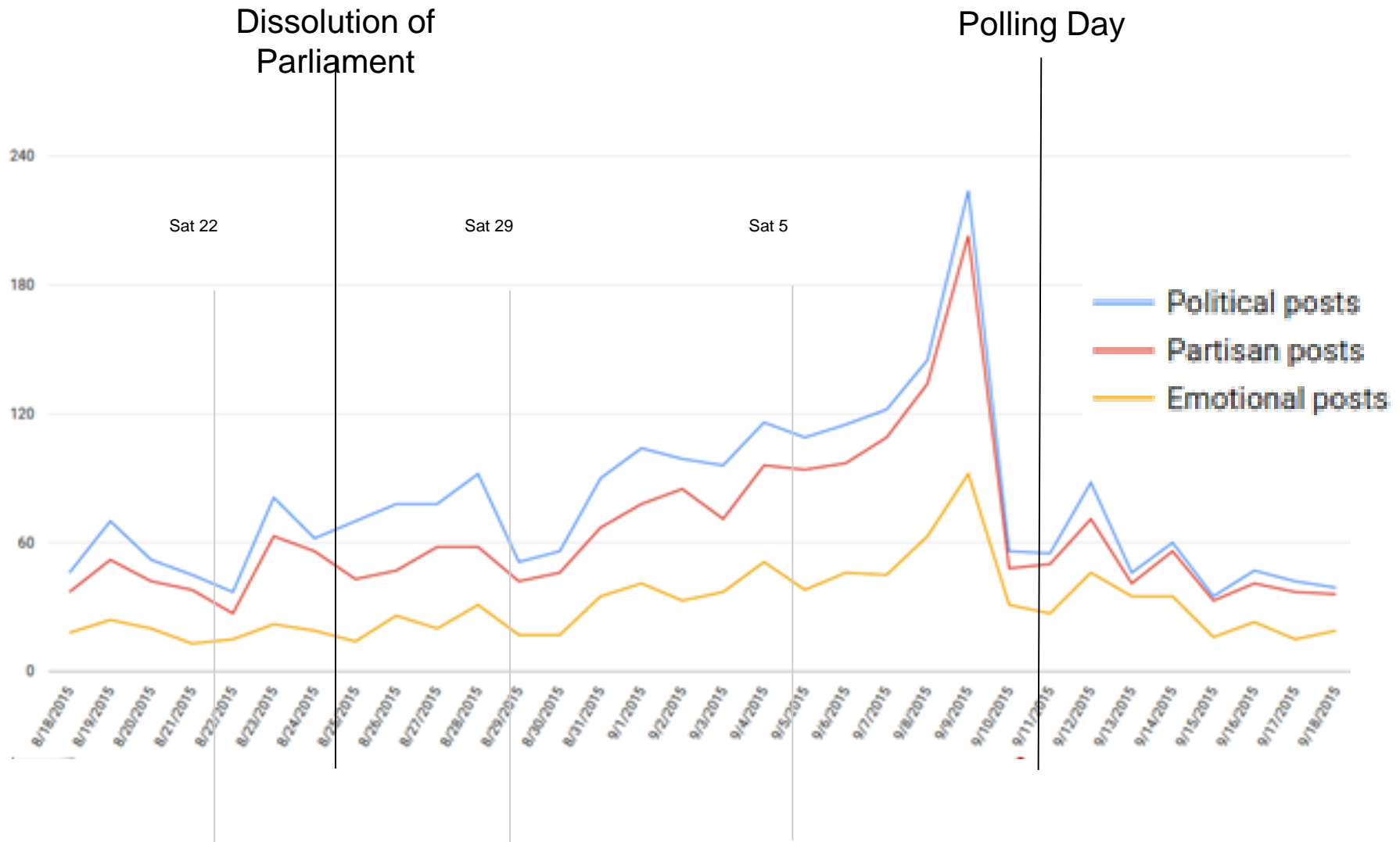
Political Post Classification: Top Features

- Minister
- PAP
- Singaporean
- Vote
- Prime
- Ministry
- Policy
- Wage
- Lee
- Parliament

- Amos
- Opposition
- Development
- Officer
- Society
- Yew
- Lim
- Authority
- Education
- Security

Partisan Post Classification: Top Features

- Population
- Income
- Award
- Conservative
- Medical
- Active
- Breach
- Unfair
- Singaporean
- Expense

- Salary
- PAP
- Vote
- Prison
- Battle
- War
- Passenger
- Tax
- Game
- Power

Emotional Post Classification: Top Features

Words:
- Bed
- Malay
- Warning
- Bank
- Political
- Heart
- Taxpayer
- Singaporean
- Content
- PAP
- Game
- Age
- Vote
- Solution
- Council
- Island
- NEA
- Malaysia

Stylometric:
Punctuations
- #
- *
- %
- ?
- &

Stylometric:
Repeating Chars
- _
- h
- x
- w

Stylometric:
Political Sentiment Words
- Afraid
- Careful
- Warn
- Hate
- Fair
- Critic
- Strang*
- Complain
- Well
- Shield
- Aristocra*

Exclusive Word Clouds: Political Posts

Election

Post Election

GE2015 Analytics System

#GE2015 Persons ▾ Parties ▾ Constituencies ▾ Issues ▾ | Blog Posts | Media Posts | ⓘ

the Bangla terrorists, I will take my chance to write about racial politics. I am a keen admirer of Mdm Speaker Halimah Yacob, and a strong advocate for her to ... Continue reading →

thesingaporebeacon.com

21 Jan, 01:03 am

🕊 👍 0 👎 0

Doubt on the necessity of Media Literacy Council's existence - The Online Citizen

Are you (nationality) first, and (ethnicity)

The SkillsFuture credits are in. Now what? - The Middle Ground

themiddleground.sg

20 Jan, 08:56 pm

🕊 👍 0 👎 0

WP's airborne trooper MAJ (NS) Pritam Singh shows up in court in No. 4 | Mothership.SG

My nephew the newly minted police officer | Gintai_昇泰

Last Saturday (16/1/16), my nephew graduated from the 3-month residential Police Officer Basic Course (NS) at Home Team Academy. He was amongst the 600 over National Servicemen enlisted to serve in the police undergoing the 3-month training. There ...

gintai.wordpress.com

20 Jan, 11:28 pm

🕊 👍 0 👎 0

20

GE2015 Analytics System

http://research.larc.smu.edu.sg/ge2015/

21

Conclusion

- Can text analytics help?
  - Accuracy depends on amount of training data.
  - Training data requires some efforts but software can help to reduce these efforts

- Insights from text analytics?
  - During GE2015, there are significantly more and larger proportion of political posts.
  - The last day of GE2015 campaigning saw a significant increase in political blog posts.
  - Cooling day reduces political blog posts.
  - Right after GE2015 result announcement, partisan and emotional posts increased by proportion.

Looking Forward

- Bloggers ≠ User population

- Social media users
  - Social media users may be silent
  - Users may read topics different from what they write about.

- Combine surveys with observed social media analytics

Wei Gong, Ee-Peng Lim, Feida Zhu. *Characterizing Silent Users in Social Media Communities*. In International AAAI Conference on Web and Social Media (ICWSM), Oxford, 2015.

Wei Gong, Ee-Peng Lim, Feida Zhu. Posting Topics Are Not Reading Topics: On Discovering Posting and Reading Topics in Social Media, In International School and Conference on Network Science (NetSciX), 2016.

Thank you

iPS Institute of Policy Studies

ENGAGING MINDS, EXCHANGING IDEAS

Symposium
on
Media and Internet Use
During General Election 2015

Wednesday, 27 January 2016

Conference Room, Level 1, Oei Tiong Ham Building

LKY Lee Kuan Yew School of Public Policy

National University of Singapore