Predicting Policy Change with Text Data

Julian TszKin Chan (Bates White Economic Consulting)
Weifeng Zhong (Mercatus Center at George Mason University)

O’Reilly AI Conference (10/17/2019)

The views expressed here are the authors’ and do not represent the views of Bates White, the Mercatus Center, or their other employees.
Our use case: China

The Policy Change Index (PCI): a machine learning model that predicts *policy changes* by “reading” this:

policychangeindex.org
How can it possibly work?! 

Because propaganda often precedes policies.
How can it possibly work?! 

Because propaganda often precedes policies.
The “value” of propaganda

People's Daily: central to China’s propaganda system

Propaganda often precedes policies

Changes in propaganda content \approx Future changes in actual policies

Machine learning

policychangeindex.org
Challenge: lack of labels

• Propaganda is abstract and subjective.
• We narrow it down to editorial priorities.
• Still, no labels—editorial priorities are in the editor’s mind!
What’s in the editor’s mind??

Transform estimation problems to anomaly detection problems.
What’s in the editor’s mind??

Transform estimation problems to anomaly detection problems.

Trick #2
Estimation: infeasible

• Newspaper text comes with page number.

• The infeasible estimation problem:

  \[
g : \{(text, FrontPage)\} \rightarrow \{"editorial priorities"\}
  \]
  
  • \(g("pvt sector is great", \text{on front page}) = \text{(reform is high-priority)}\);
  
  • \(g("central planning is great", \text{on front page}) = \text{(reform is low-priority)}\).

• Only if \(g\) can be learned...
Anomaly detection: feasible

- Treat “editorial priorities” as a latent variable:
  \[ f \{"editorial priorities"\} : \{text\} \rightarrow \{FrontPage\} \]

- Lots of labels for \( f \).
- Detecting changes in \( f \): a feasible anomaly detection problem.
- Change in \( f \) \( \Rightarrow \) change in propaganda priorities \( \Rightarrow \) change in policy!
Policy Change Index for China (PCI-China)

Articles in 5 years before $t$

Training & validation

Testing

Articles in quarter $t$

“Forecast”

PCI-China in quarter $t$ = Test performance − “Forecast” performance

Large PCI-China ⇒ Structural change in editorial priorities
PCI-China: result vs ground truth
PCI-China: result v ground truth

Trade-war implications
Understanding changes

Content of *misclassified* articles has policy substance:

• False negatives suggest new policies;

• False positives suggest phasing-out policies.
Understanding changes: 2018 Q1 uptick
Trade-war implications

Two upticks in 2018 Q1 and 2019 Q1:

• Internally: strengthening party authority;
• Externally: nationalism and global leadership.

A trade deal with structural reforms? Curb your enthusiasm...
How can these be useful to you?
Remember Trick #2?

When there’s no label for the target variable, transform the estimation problem to an anomaly detection problem.

Many other use cases share the same structure!
Future PCI projects

Stay tuned for more in the PCI pipeline:

• PCIs for other countries: North Korea, Russia, Cuba,…
• PCI sub-indices, by subject: trade policy, fiscal policy,…
• And more!
Propaganda often precedes policies.

More generally, words often foreshadow actions.
Consider the Hong Kong protests

• Will the protests be met with a military crackdown by China?
  ▮ How hard is the *People’s Daily* slamming Hong Kong?

• There’s a unique, well-formed precedent: Tiananmen crackdown.
Policy Change Index for Crackdown (PCI-Crackdown)

1989 articles leading up to Tiananmen crackdown

Recent articles on HK protests

Train a date classifier

Deploy

PCI-Crackdown: mapping each current date to an “as-if” date
PCI-Crackdown: Hong Kong protests
Takeaways

1. Words often foreshadow actions.

2. No labels? Try transforming estimation problems to anomaly detection problems.

Website: policychangeindex.org
(papers, code, newsletter,...)