Reading China: Predicting Policy Change with Machine Learning

Julian TszKin Chan
(Bates White)

Weifeng Zhong
(AEI)

December 19, 2018
US Naval Research Lab

The views expressed here are solely those of our own and do not represent the views of the American Enterprise Institute, Bates White Economic Consulting, or their other employees.
Policy Change Index (PCI) for China

The first *leading* indicator of China’s policy changes.

- Covers 1951 Q1 – 2018 Q3.
- Can be updated in the future.
How to predict policy changes?

Build a machine learning algorithm to

• “read” the People's Daily;
• detect changes in how it prioritizes policy issues.
Source of predictive power

The Leninist tradition:

- “[T]he whole task of the Communists is to be able to convince the backward elements.”

- It is a fundamental necessity “to transform the press from an organ which primarily reports the political news of the day into a serious organ for the economic education of the mass of the population.”
Source of predictive power

*People's Daily:* nerve center of China’s propaganda system

Propaganda often precedes policies.

Detect changes in newspaper’s priorities

≈

Predict changes in gov’t policies
Source of predictive power

People's Daily: nerve center of China's propaganda system

Propaganda often precedes policies.

Detect changes in newspaper’s priorities ≈ Predict changes in gov’t policies

Front page?
Imagine an avid reader of the *People’s Daily* who

1. reads recent articles;
2. forms a paradigm about front-page content;
3. tests the paradigm on new articles.
Methodology

Articles in previous 5 years
- Modeling
- Testing

Articles in next quarter
- "Forecasting"

PCI = Testing performance - "Forecasting" performance
Methodology

“Language-free” — it does **not** require the reading of the Chinese text.
Methodology: data

<table>
<thead>
<tr>
<th>date</th>
<th>year</th>
<th>month</th>
<th>day</th>
<th>page</th>
<th>title</th>
<th>body</th>
<th>id</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018-10-01</td>
<td>2018</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>习近平在会见四川航空“中国民航英雄机组”全体成员时强调 学习英雄事迹 弘扬英雄精神 将非凡英雄精神 用到日常工作中……………</td>
<td>中共中央总书记、国家主席、中央军委主席习近平专程前往四川航空“中国民航英雄机组”全体成员慰问，强调……………</td>
<td>2018100000</td>
</tr>
<tr>
<td>2018-10-01</td>
<td>2018</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>烈士纪念日向人民英雄敬献花篮仪式在京隆重举行    习近平李克强栗战书汪洋王沪宁赵乐际韩正王岐山出席</td>
<td>9月30日上午，党和国家领导人习近平、李克强、栗战书、汪洋、王沪宁、赵乐际、韩正、王岐山等……………</td>
<td>2018100001</td>
</tr>
<tr>
<td>2018-10-01</td>
<td>2018</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>庆祝中华人民共和国成立69周年 国务院举行国庆招待会    习近平栗战书汪洋王沪宁赵乐际韩正王岐山出席</td>
<td>9月30日晚，国务院在北京人民大会堂举行国庆招待会，热烈庆祝中华人民共和国成立69周年……………</td>
<td>2018100002</td>
</tr>
<tr>
<td>2018-10-01</td>
<td>2018</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>习近平就印度尼西亚中苏拉威西省地震海啸向印尼总统佐科致慰问电</td>
<td>新华社北京9月30日电，9月30日，国家主席习近平就印度尼西亚中苏拉威西省发生地震海啸向印尼总统佐科致慰问电……………</td>
<td>2018100003</td>
</tr>
<tr>
<td>2018-10-01</td>
<td>2018</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>在庆祝中华人民共和国成立69周年国庆招待会上的致辞    中华人民共和国国务院总理 李克强 （三〇四）</td>
<td>各位来宾、各位朋友、同事们：今天，我们隆重庆祝中华人民共和国成立六十九周年。新中国波澜壮阔……………</td>
<td>2018100004</td>
</tr>
<tr>
<td>2018-10-01</td>
<td>2018</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>用奋斗成就复兴伟业（社论）——热烈庆祝中华人民共和国成立69周年</td>
<td>时间的年轮，刻印下奋斗者的足迹。当10月的阳光照耀大地，我们迎来了中华人民共和国69岁华诞……………</td>
<td>2018100005</td>
</tr>
<tr>
<td>2018-10-01</td>
<td>2018</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>国务院印发《决定》 进一步压减工业产品生产许可证管理目录和简化审批程序</td>
<td>新华社北京9月30日电，经李克强总理签批，国务院日前印发《关于进一步压减工业产品生产许可证管理目录……………</td>
<td>2018100006</td>
</tr>
<tr>
<td>2018-10-01</td>
<td>2018</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>谱写新时代乡村振兴新篇章——论学习习近平总书记在中央政治局第十八次集体学习时重要讲话……………</td>
<td>本报评论员：乡村振兴既是一场攻坚战，更是一场持久战，必须咬定目标，奋进前行……………</td>
<td>2018100007</td>
</tr>
</tbody>
</table>
Methodology: data
Methodology: modelling

\[ x \] : each article as an observation.

\[ y = f(x) \]

\[ f \] : a complicated function.
Methodology: modelling

\[ x : \text{each article as an observation.} \]

\[ f : \text{a complicated function.} \]

\[ y = f(x) \]
Results
Result: PCI
Result: PCI — with ground truth

- 1953 first Five-Year Plan
- 1958 Great Leap Forward
- 1966 Cultural Revolution
- 1976 Hua takes over
- 1978 reform program starts
- 1989 Tiananmen Sq. protests
- 1993 reform speed-up
- 2005 reform slow-down
- 2008 stimulus package
- 2013 revive Maoism
- 2013 renew reform program
- 2015 supply-side structural reform

Quarterly PCI for China

Year:
- 1950
- 1955
- 1960
- 1965
- 1970
- 1975
- 1980
- 1985
- 1990
- 1995
- 2000
- 2005
- 2010
- 2015
- 2018
Understanding substance of change

<table>
<thead>
<tr>
<th>Front page?</th>
<th>Classified on front page?</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>False positives</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Front page?</th>
<th>Classified on front page?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>False negatives</td>
</tr>
<tr>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

Content of *mis*-classified articles has policy substance.
Supervised learning: a digression
Supervised learning

mapping : $X \rightarrow Y$

• Trained on $\{x_i, y_i\}_{i \in \text{training}}$.

• Goal: from $\{x_j\}_{j \in \text{new}}$, to predict $\{y_j\}_{j \in \text{new}}$.

• Challenge: need lots of training data.
The newspaper problem:  
an infeasible approach

\[ g : \{(Article, FrontPage)\} \rightarrow \{(Policy, Priority)\} \]

- With the learned function \( g \):
  - \( g("pvt sector is important", \text{front page}) = \text{(reform, high priority)}; \)
  - \( g("central planning is great", \text{front page}) = \text{(reform, low priority)}; \ldots \)
The newspaper problem: an *infeasible* approach

\[ g : \{(Article, FrontPage)\} \rightarrow \{(Policy, Priority)\} \]

• With the learned function \(g\):
  
  • \(g("pvt sector is important", \text{front page}) = (\text{reform, high priority});\)
  
  • \(g("central planning is great", \text{front page}) = (\text{reform, low priority});\) ...

• But where are the training data?
The newspaper problem: a feasible approach

- Think of priorities as a latent variable:

\[ f_{\{(Policy,Priority)\}} : \{Article\} \rightarrow \{FrontPage\} \]
The newspaper problem: 
a feasible approach

• Think of priorities as a latent variable:

\[ f\{ (Policy, Priority) \} : \{ Article \} \rightarrow \{ FrontPage \} \]

• Lots of training data to learn each function \( f \).

• Difference in function \( \Rightarrow \) difference in priorities.

• “Language-free!”
The newspaper problem:
a feasible approach

Articles in previous 5 years

Articles in next quarter

Modeling

Testing

“Forecasting”

Testing performance ≠ “Forecasting” performance ⇒ Different priorities
Other applications
App 1: PCIs for other countries

Predicting other (ex-)Communist regimes’ policies:

• Soviet Union’s *Pravda*
• East Germany’s *Neues Deutschland*
• North Korea’s *Rodong Sinmun*
• Cuba’s *Granma*
• Vietnam’s *Nhân Dân*
App 1: PCls for other countries

Predicting other (ex-)Communist regimes’ policies:

• Soviet Union’s *Pravda*
• East Germany’s *Neues Deutschland*
• North Korea’s *Rodong Sinmun*
• Cuba’s *Granma*
• Vietnam’s *Nhân Dân*
Measuring media bias in the US:

- Replicate the same analysis on US newspapers.
- Compare $\text{PCI}_{WaPo}$, $\text{PCI}_{NYT}$, $\text{PCI}_{WSJ}$, ...
- Divergence among $\text{PCI}$'s $\rightarrow$ Polarization in media
App 3: predicting vote change

Predicting vote change in legislation:

- Newspaper texts $\rightarrow$ Legislators’ public statements
- Page numbers $\rightarrow$ Legislators’ names
- What if Sen. A’s statement is mistaken as Sen. B’s?

Source of predictive power:

- Political necessity to justify vote changes by making different statements *in advance*. 
Other apps

Omitted here. See our research paper.
Interested in DIY?

- Website: policychangeindex.com (newsletter sign-up)
- Source code: github.com/PSLmodels/PCI

- A simulated example to show how the PCI works.
Questions?