

# Reading China:

## Predicting Policy Change with Machine Learning

Julian TszKin Chan  
(Bates White)

Weifeng Zhong  
(AEI)

January 25, 2019  
Texas AI Summit

*The views expressed here are those of the authors and do not represent the views of AEI, Bates White, or their other employees.*

# Policy Change Index (PCI) for China

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

- Covers 1951 Q1 – 2018 Q4.
- Updated quarterly.

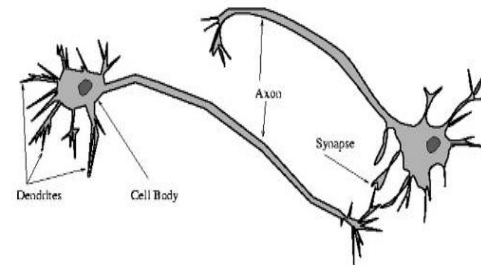
# How to predict policy changes?

Build a machine learning algorithm to

- “read” the *People's Daily*;
- detect changes in how it prioritizes policy issues.



Official newspaper, 1946-present



(Artificial) neural networks

# Source of predictive power

The Leninist tradition:

- “[T]he whole task of the Communists is to be able to **convince** the backward elements.” (1920)
- 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.” (in early USSR years)

# Source of predictive power

The Leninist tradition, in President Xi Jinping's words:

- The media should “uphold the correct **orientation** of public opinion in all aspects and at all stages.” (2016)

# 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?



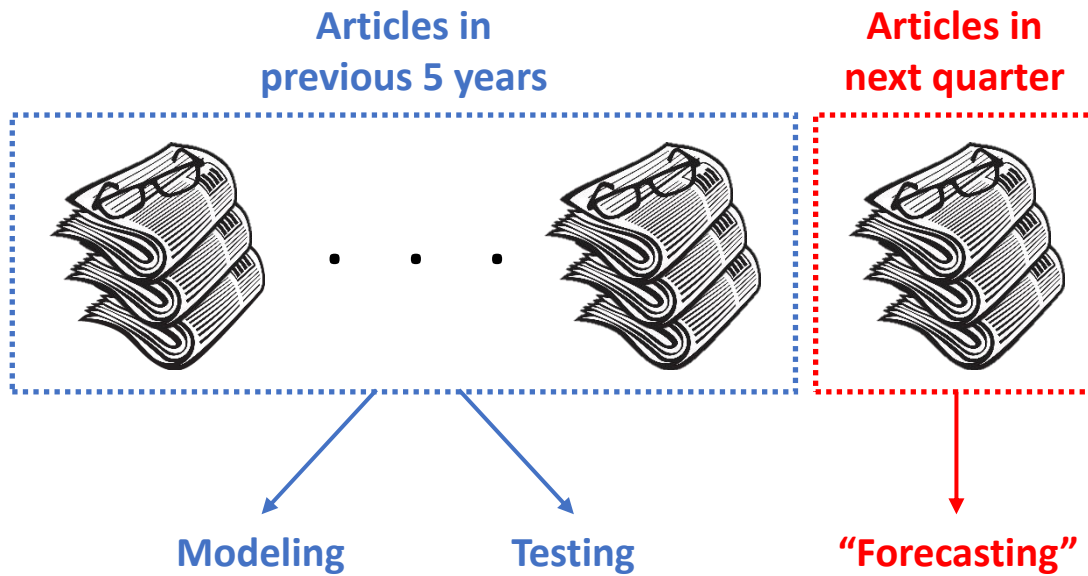
# Methodology

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.

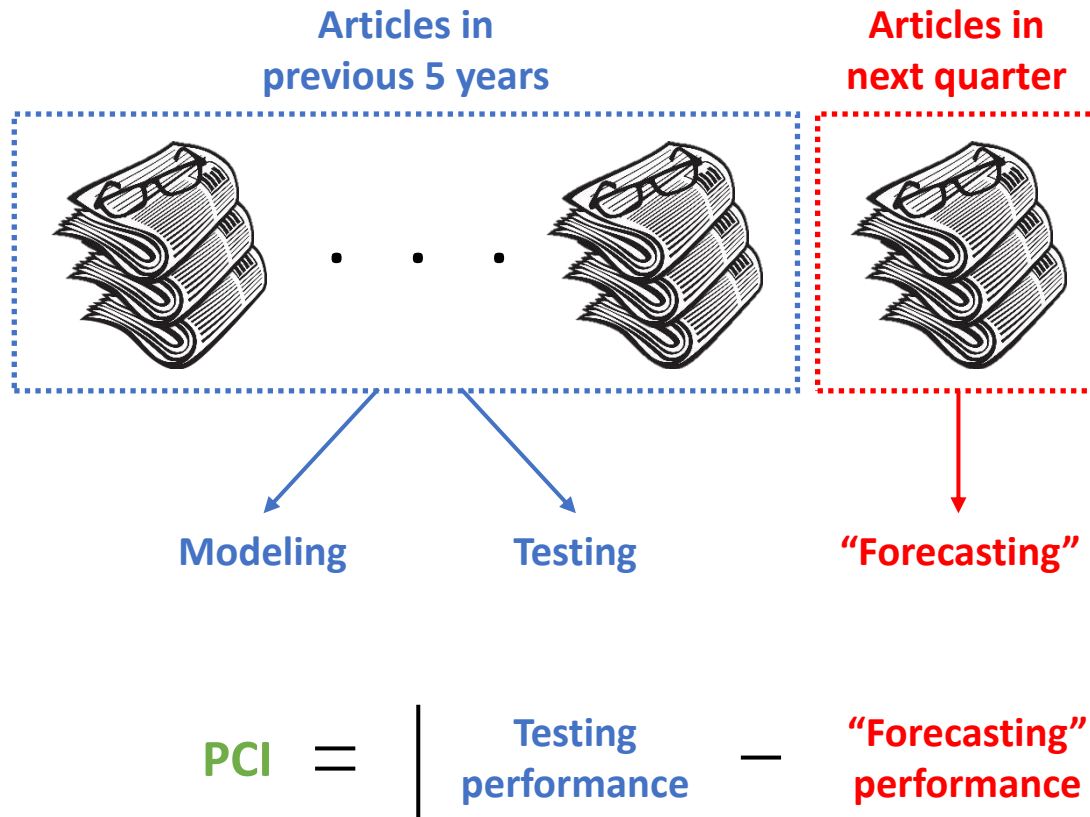


# Methodology



$$\text{PCI} = \left| \text{Testing performance} - \text{“Forecasting” performance} \right|$$

# Methodology



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

# Methodology: data

sample\_data

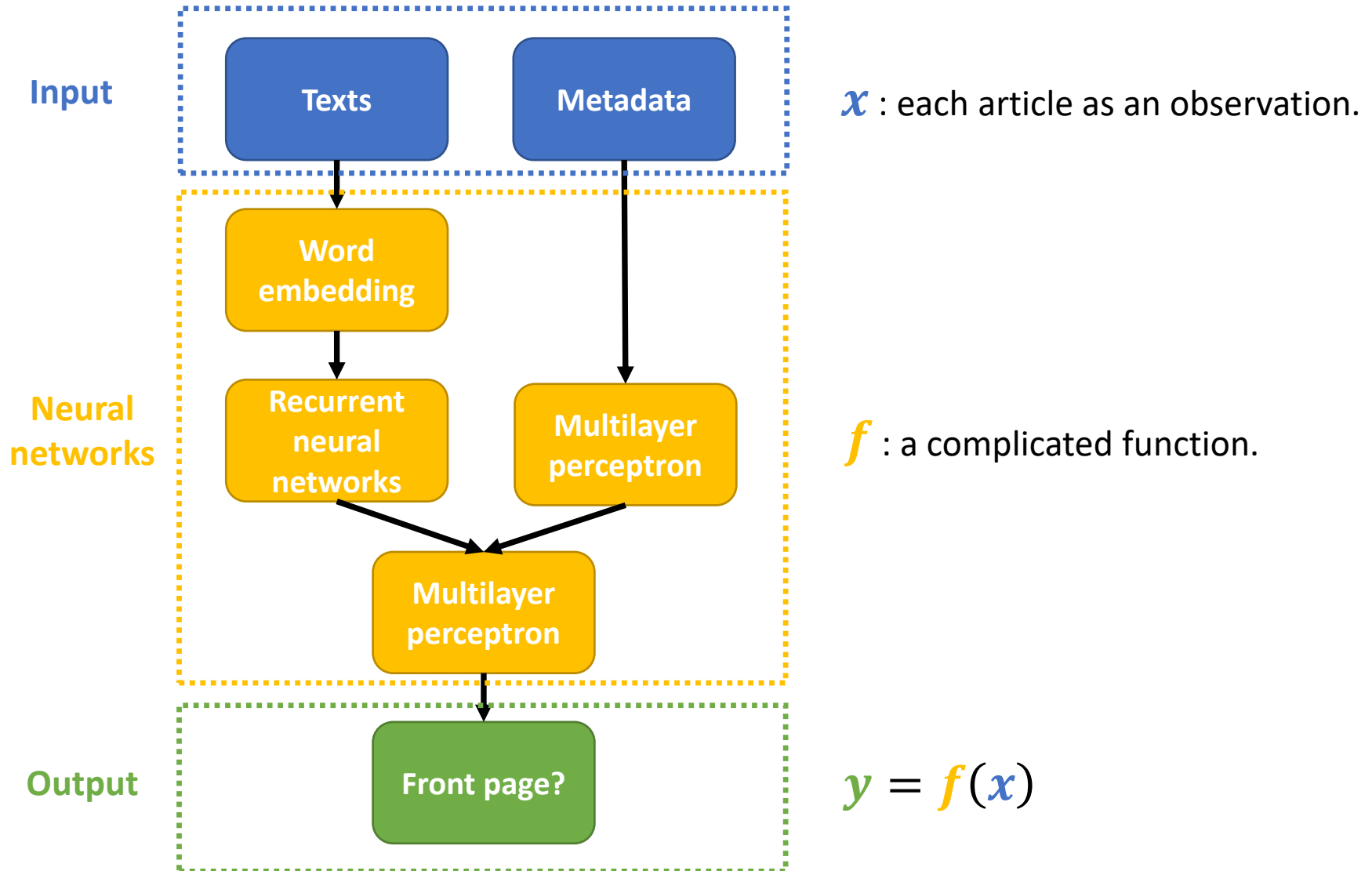
	date	year	month	day	page	title	body	id
0	2018-10-01	2018	10	1	1	习近平在会见四川航空“中国民航英雄机组”全体成员时强调 学习英雄事迹 弘扬英雄精神 将非凡英...	中共中央总书记、国家主席、中央军委主席习近平专门邀请四川航空“中国民航英雄机组”全体成...	2018100000
1	2018-10-01	2018	10	1	1	烈士纪念日向人民英雄敬献花篮仪式在京隆重举行 习近平李克强栗战书汪洋王沪宁赵乐际韩正王岐山出席	9月30日上午，党和国家领导人习近平、李克强、栗战书、汪洋、王沪宁、赵乐际、韩正、王岐山...	2018100001
2	2018-10-01	2018	10	1	1	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王岐山等...	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六十九周年...	2018100002
3	2018-10-01	2018	10	1	2	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	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强烈地震及...	2018100003
4	2018-10-01	2018	10	1	2	在庆祝中华人民共和国成立六十九周年招待会上的致辞 中华人民共和国国务院总理 李克强 （二〇...	各位来宾、各位朋友、同志们： 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...	2018100004
5	2018-10-01	2018	10	1	2	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	时间的年轮，刻印下奋斗者的足迹。当10月的阳光照耀大地，我们迎来了人民共和国69岁华诞。...	2018100005
6	2018-10-01	2018	10	1	2	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	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...	2018100006
7	2018-10-01	2018	10	1	2	谱写新时代乡村全面振兴新篇章 ——论学习习近平总书记关于实施乡村振兴战略重要讲话精神	本报评论员 乡村振兴既是一场攻坚战，更是一场持久战。必须坚定信心，咬定目标，苦干实干...	2018100007

# Methodology: data

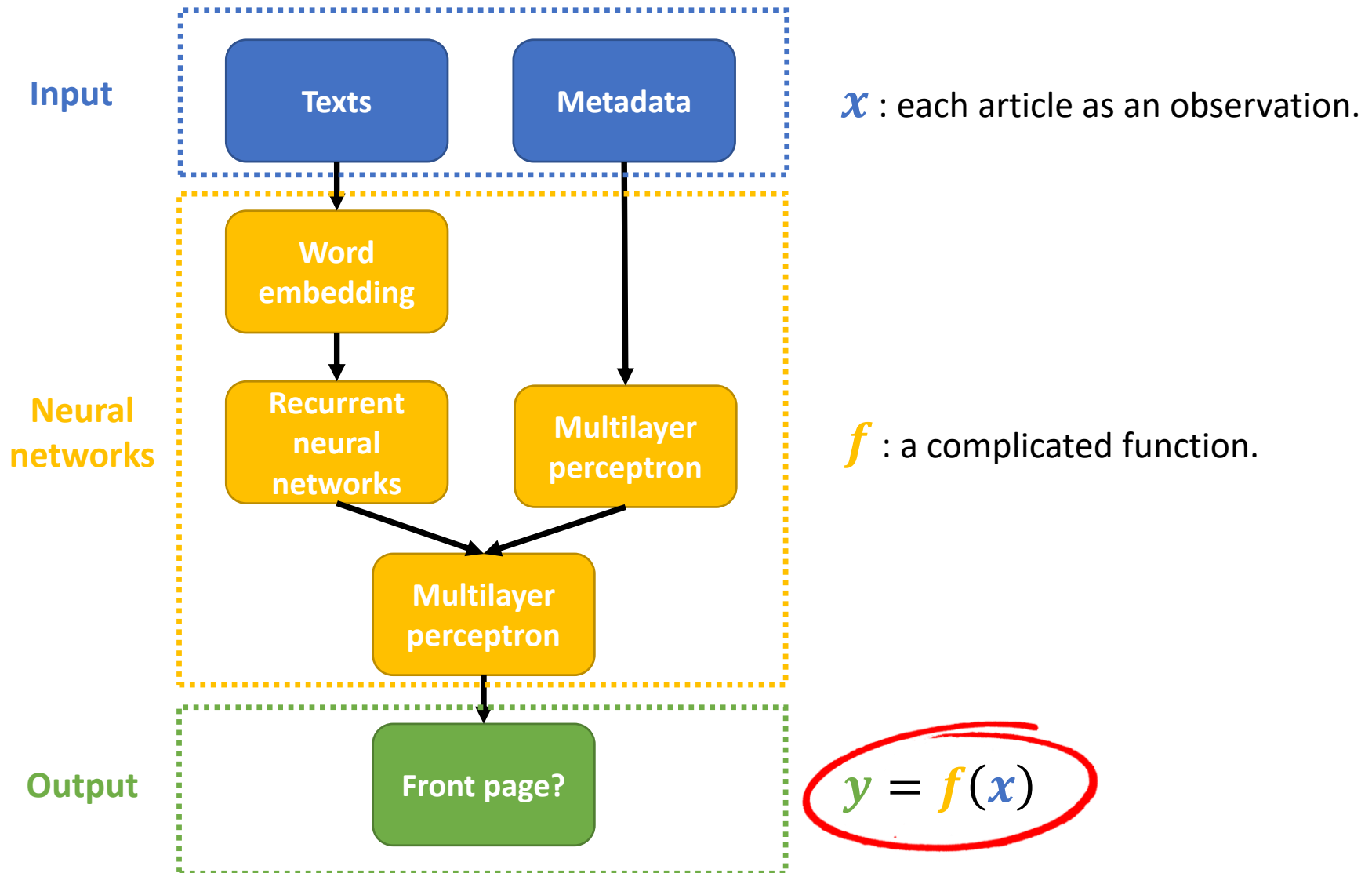
sample\_data

	date	year	month	day	page	title	body	id
0	2018-10-01	2018	10	1	1	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...	中共中央总书记、国家主席、中央军委主席习近平专门邀请四川航空“中国民航英雄机组”全体成员...	2018100000
1	2018-10-01	2018	10	1	1	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	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...	2018100001
2	2018-10-01	2018	10	1	1	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王岐山等...	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六十九周年...	2018100002
3	2018-10-01	2018	10	1	2	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	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强烈地震及...	2018100003
4	2018-10-01	2018	10	1	2	在庆祝中华人民共和国成立六十九周年招待会上的致辞 中华人民共和国国务院总理 李克强 （二〇...	各位来宾、各位朋友、同志们： 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...	2018100004
5	2018-10-01	2018	10	1	2	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	时间的年轮，刻印下奋斗者的足迹。当10月的阳光照耀大地，我们迎来了人民共和国69岁华诞。...	2018100005
6	2018-10-01	2018	10	1	2	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	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...	2018100006
7	2018-10-01	2018	10	1	2	谱写新时代乡村全面振兴新篇章 ——论学习习近平总书记关于实施乡村振兴战略重要讲话精神	本报评论员 乡村振兴既是一场攻坚战，更是一场持久战。必须坚定信心，咬定目标，苦干实干...	2018100007

# Methodology: modelling

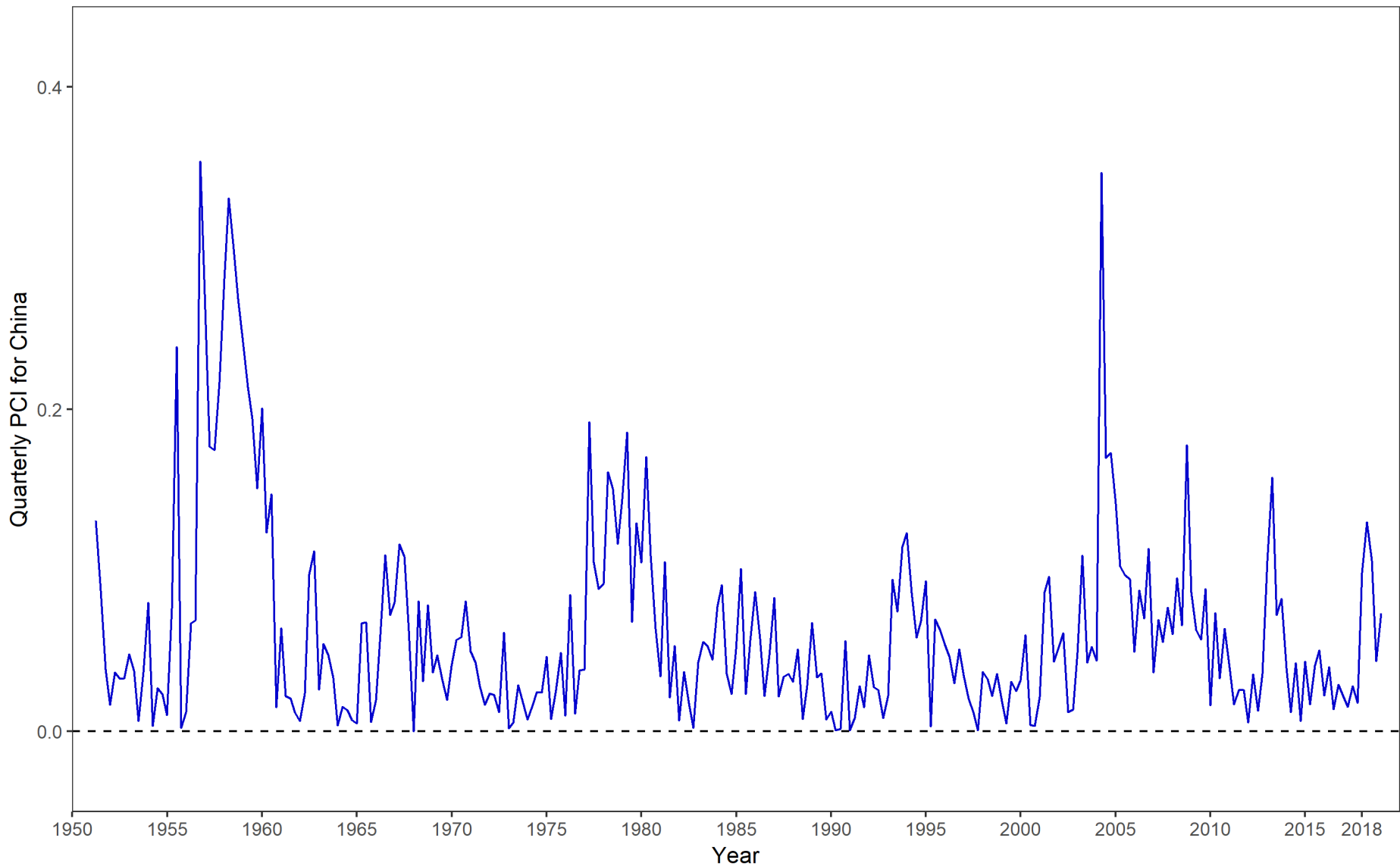


# Methodology: modelling



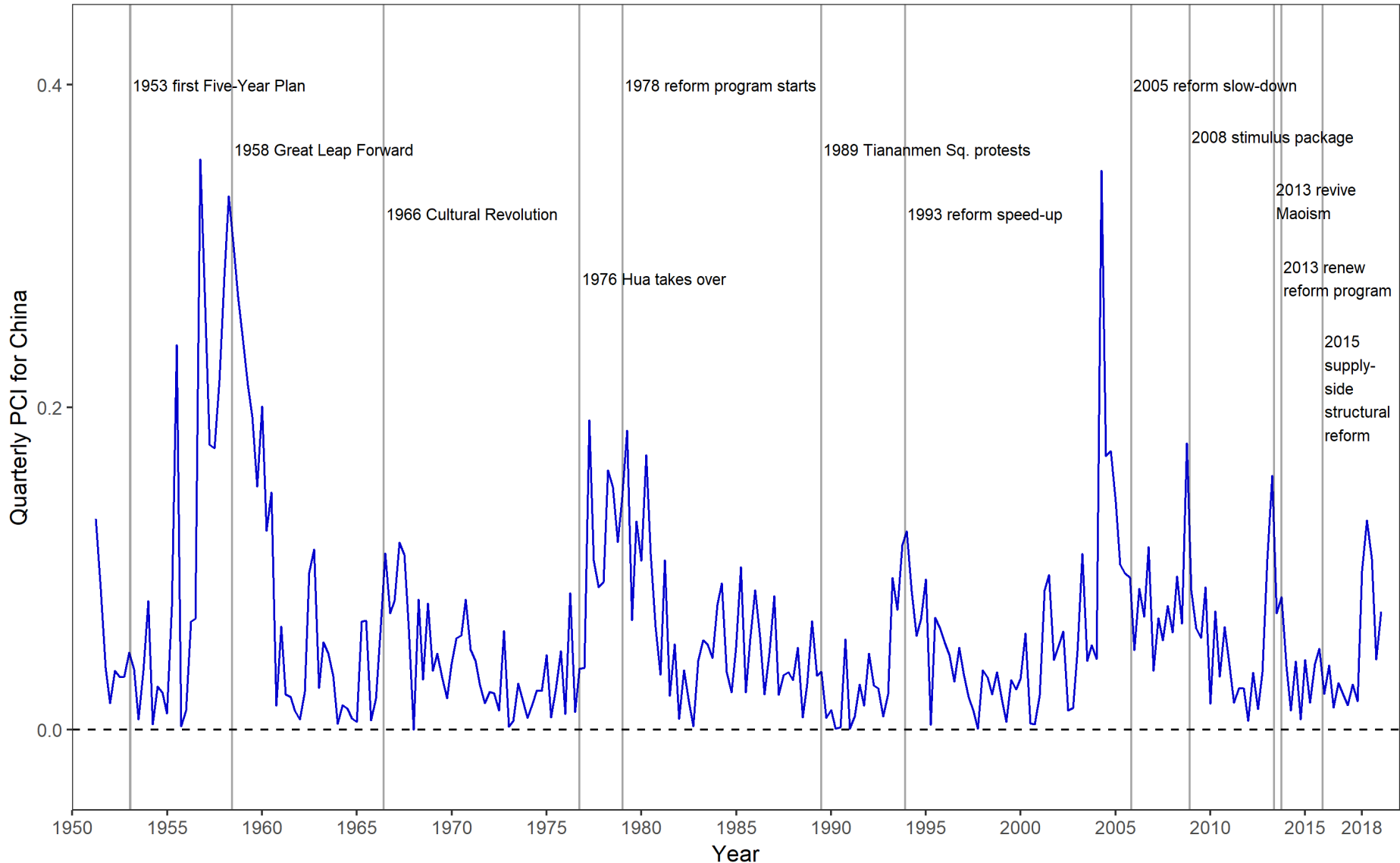
# Results

# Result: PCI

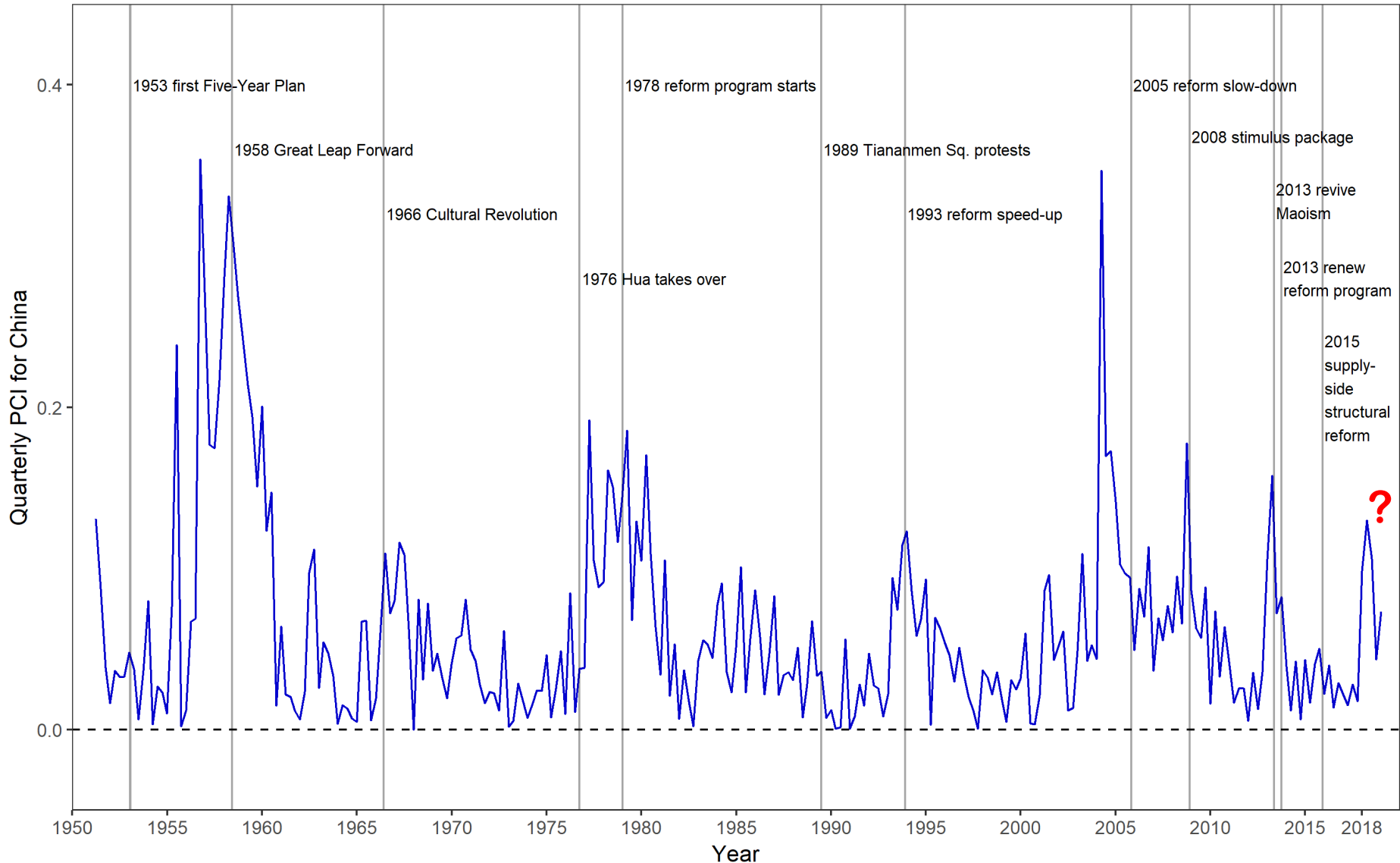




# Result: PCI — with ground truth



# Result: PCI — with ground truth



# Understanding substance of change

		Classified on front page?	
		No	Yes
Front page?	No	√	false positives
	Yes	false negatives	√

- Content of *mis*-classified articles has policy substance.

# Understanding substance of change

**Example: US-China trade conflict**

What does the 2018 Q1 uptick represent?

# Understanding substance of change

## Example: US-China trade conflict

What does the 2018 Q1 uptick represent?

- Strengthening party authority;
- Playing up nationalism and global leadership;
- Populist policies to maintain support.

# Understanding substance of change

## Example: US-China trade conflict

What does the 2018 Q1 uptick represent?

- Strengthening party authority;
- Playing up nationalism and global leadership;
- Populist policies to maintain support.

⇒ Curb your enthusiasm for trade talks.

**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  $g$ :
  - $g(\text{"pvt sector is important", front page}) = (\text{reform, high priority})$ ;
  - $g(\text{"central planning is great", front page}) = (\text{reform, low priority})$ ; ...

# The newspaper problem:

an *infeasible* approach

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

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

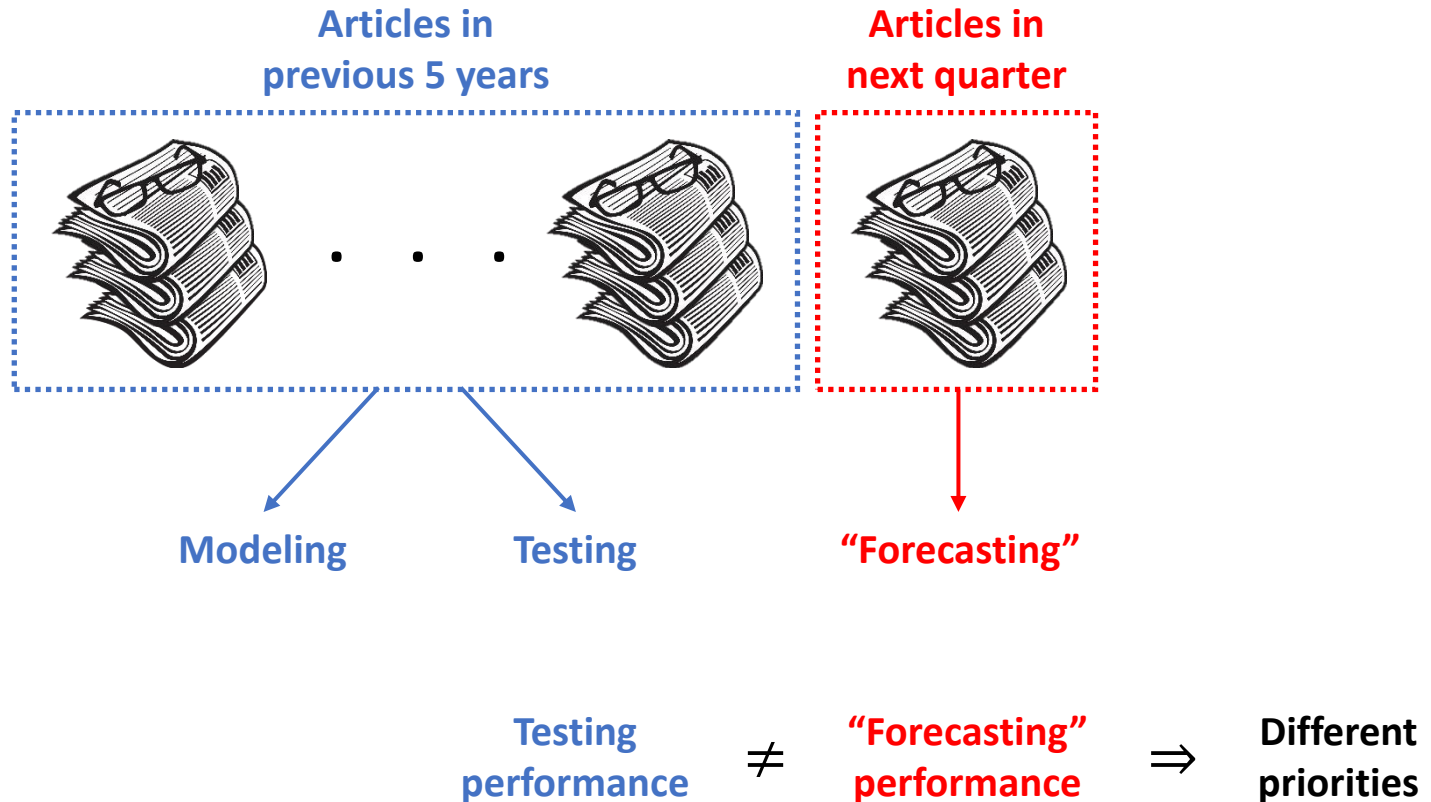
## a feasible approach

- 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



**Other applications**

# App 1: PCIs for other countries

Predicting other (ex-)Communist regimes' policies:

- Soviet Union's *Pravda*
- East Germany's *Neues Deutschland*
- Cuba's *Granma*
- North Korea's *Rodong Sinmun*
- Vietnam's *Nhân Dân*

# App 1: PCIs for other countries

Predicting other (ex-)Communist regimes' policies:

- Soviet Union's *Pravda*
  - East Germany's *Neues Deutschland*
  - Cuba's *Granma*
  - North Korea's *Rodong Sinmun*
  - Vietnam's *Nhân Dân*
- Work in progress



# App 2: predicting vote change

Predicting vote change in legislation:

- Newspaper texts → **Legislators' public statements**
- Page numbers → **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](http://policychangeindex.com) (newsletter sign-up)
- Paper: [policychangeindex.com/pdf/Reading\\_China.pdf](http://policychangeindex.com/pdf/Reading_China.pdf)
- Source code: [github.com/PSLmodels/PCI](https://github.com/PSLmodels/PCI)
  
- A [simulated example](#) to show how the PCI works.

**Questions?**