

Introduction of the “Chinese Industrial Policy Attention Dataset” (CIPAD)

Shiyang Xiao

January 29, 2024

"Chinese Industrial Policy Attention Dataset" (CIPAD) is an original dataset which contains 612 central-level and 1907 provincial-level Chinese industrial policies from 2001 to 2019, along with their allocation of attention across 155 finely segmented manufacturing industry categories.

A novel design of the CIPAD is that, by using computational text analysis techniques, full text of each industrial policy is transformed into a distribution-of-attention vector. A distribution-of-attention vector can be written as $(x_1, x_2, \dots, x_{155})$, with $x_i \in [0,1]$ and $\sum_1^{155} x_i = 1$. Each vector describes the attention allocation of an industrial policy to the 155 industry categories in manufacturing sector, and x_i equals the proportion of attention paid to the i th industry category. The 155 industry categories are based on the three-digit codes in Chinese Industrial Classification for National Economic Activities (GB/T 4754-2002). To the best of our knowledge, this dataset stands out as one of the first to identify industry categories as granular as three-digit levels in full texts of industrial policies.

By transforming each policy full text into a distribution-of-attention vector, the CIPAD allows researchers to accomplish the following tasks: 1) Quickly identify the most frequently mentioned industry categories in each policy; 2) compare the similarities of targeted industry categories between policies; 3) track changes in government’s attention to different industry categories over time; 4) compare similarities of policy attention between different governments. Below I will demonstrate how the CIPAD helps researchers to achieve the above tasks.

1. Quickly Identify the Most Frequently Mentioned Industry Categories in Each Policy

The CIPAD transforms full text of each industrial policy into a distribution-of-attention vector. This allows researchers to fast identify industry categories that receive the most attention from an industrial policy. Take the policy “Made in China 2025” issued by Chinese State Council in 2015 as an example. Based on the CIPAD, we can fast extract the top-15 industry categories that are most frequently mentioned by “Made in China 2025.” According to Figure 1, the industry category that received the most attention is “special equipment manufacturing” which receives eight percent of attention from the policy. The industry categories that rank the second and the third are “aerospace manufacturing” and “communication equipment manufacturing,” which obtain five and four percent of the policy's attention respectively. Industry categories outside the top-15 chart account for 45% of the attention, showing that “Made in China 2025” is a highly comprehensive industrial policy that targets diverse industry categories.

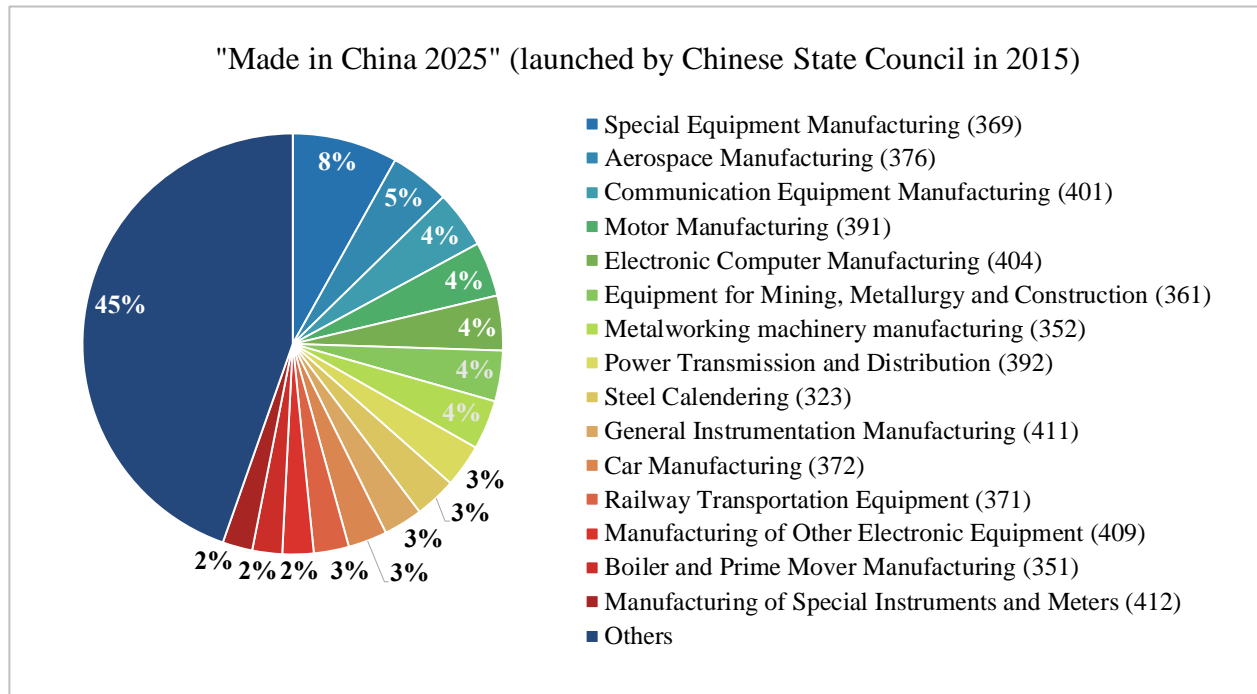


Figure 1: Top-15 Industry Categories Most Frequently Mentioned in “Made in China 2025”

2. Compare the Similarities of Targeted Industry Categories between Industrial Policies

CIPAD also enables researchers to fast compare similarities between industrial policies based on the industry categories they mentioned. By transforming policy full texts into distribution-of-attention vectors, researchers can measure policy similarities by calculating cosine similarity between policies' corresponding vectors. Cosine similarity ranges from 0 to 1, where 0 indicates no similarity at all, and 1 indicates complete similarity. In table 1, we demonstrate the top-3 provincial industrial policies that are the most and the least similar to “Made in China 2025” and the cosine similarities between the above provincial industrial policies and “Made in China 2025.”

Table 1: Provincial Industrial Policies Most and Least Similar to “Made in China 2025”

Panel A: Provincial industrial policies that are most similar to “Made in China 2025”			
Title	Time	Issue Agency	Cosine Simi.
"Made in China 2025" Five-Year Action Plan	2015.11	Hunan Provincial Government	0.891
Opinions on Further Promoting "Made in China 2025"	2015.11	Henan Provincial Government	0.881
Implementation Plan for Transformation and Upgrading of Equipment Manufacturing Industry	2018.12	Shandong Provincial Government General Office	0.881
Panel B: Provincial industrial policies that are least similar to “Made in China 2025”			
Title	Time	Issue Agency	Cosine Simi.
Implementation Opinions on Promoting Supply-side Structural Reform in the Liquor Industry	2016.9	Guizhou Provincial Government General Office	0.001
Measures to Support The High-quality Development of the Tobacco Industry	2019.10	Yunnan Provincial Government	0.006
Guiding Opinions on Promoting the Inheritance and Development of the Silk Industry	2015.11	Zhejiang Provincial Government General Office	0.020

3. Track Changes in Government’s Attention to Different Industry Categories over Time

Based on the CIPAD, we can not only identify the attention allocation of each industrial policy, but can also calculate the government's attention allocation during a period of time by calculating the average distribution-of-attention vector based on all the policies adopted by the government. Figure 2 shows the top 10 industry categories that gained or lost the most attention from the Chinese central government after 2013 (i.e., 2006-2012 vs. 2013-2019). The gray words in parentheses are high-frequency key phrases corresponding to each industry category. For example, the increasing attention paid to “Other Special Equipment” after 2013 is largely due to Chinese central government’s increasing attention to robots.

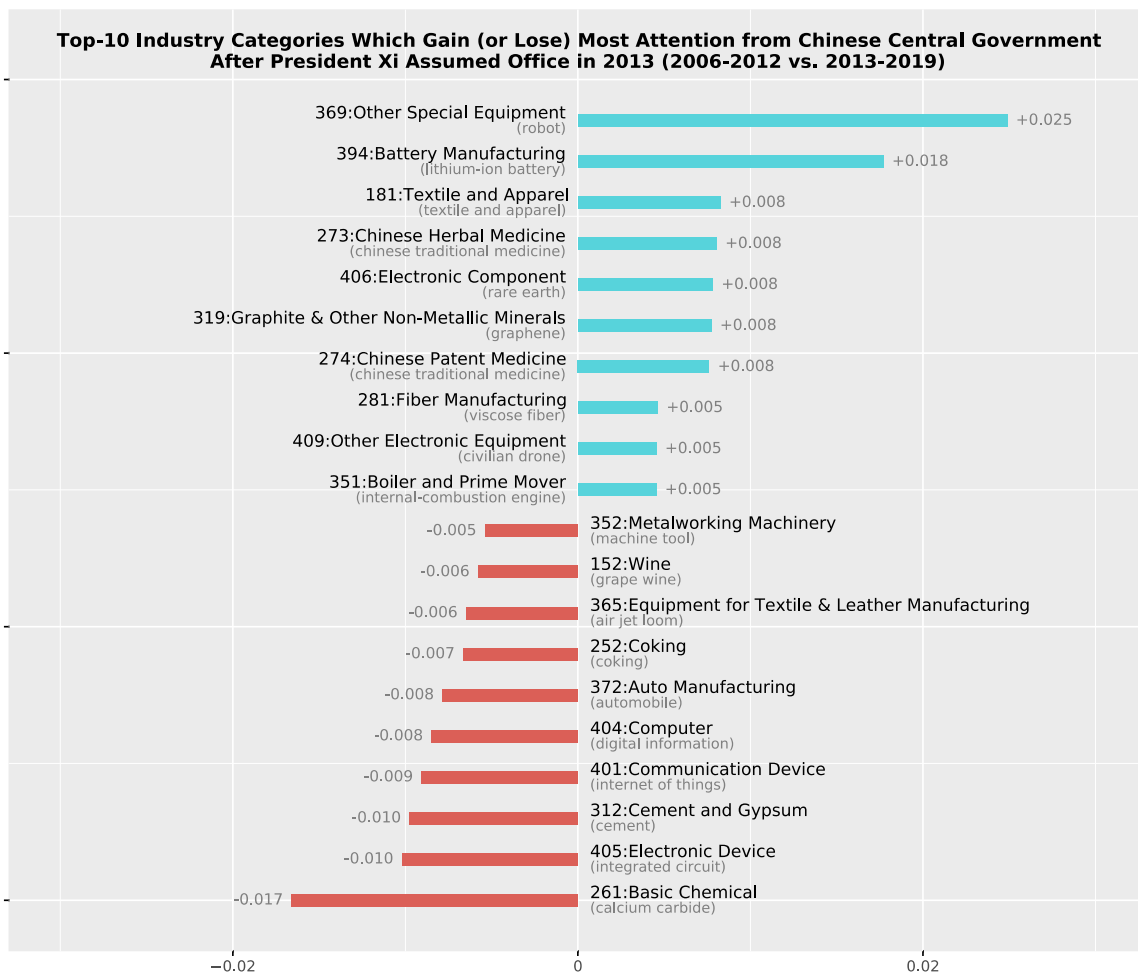


Figure 2: Industry Categories Gain or Lose Most Attention from Chinese Central Government

4. Compare Similarities of Policy Attention between Different Governments

By calculating the average distribution-of-attention vector for each government during a period of time, the CIPAD enables us to quickly compare similarities of policy attention between different governments. In figure 3, we compare similarities of policy attention between provincial governments in China during 2006-2012 and during 2013-2019. Each node represents a provincial-level government. Provincial governments in different regions are highlighted by different colors. Two nodes are connected by a gray line if the two corresponding provincial governments share highly similar distribution-of-attention vectors (cosine similarity>0.8) during a period of time. It can be shown from figure 3 that provincial governments' attention allocation became more similar to one another during 2013-2019 than they were during 2006-2012.

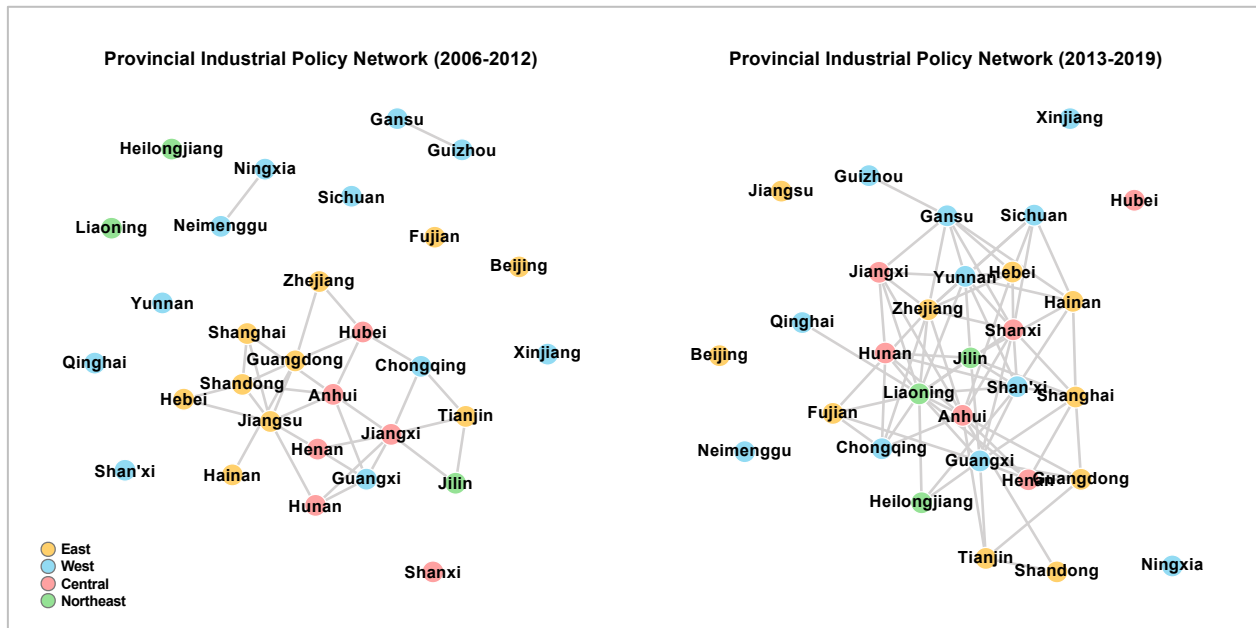


Figure 3: Similarities of Policy Attention between Different Provincial Governments