

Big Data Perspective on the Two Sessions: The Postprint of Percent's Exploration in Data Journalism

Authors: Chen Xuguan

Date: 2023-10-08T00:00:00+00:00

Abstract

During this year's Two Sessions, topics such as people's livelihood, internet, culture, entrepreneurship, environmental protection, and artificial intelligence garnered widespread attention. As news about the Two Sessions flooded media platforms, which topics were mentioned most frequently? How can we decipher China's future trajectory from this information deluge? Big data offers powerful solutions.

Full Text

Preamble

During this year's Two Sessions, topics such as people's livelihood, internet, culture, entrepreneurship, environmental protection, and artificial intelligence garnered widespread attention. As news about the Two Sessions flooded media platforms, which topics were mentioned most frequently? How can we decipher China's future trajectory from this information deluge? Big data offers powerful solutions.

Data journalism involves four key stages: collecting raw data, analyzing and filtering the data, visualizing the data, and producing concrete news reports.

In terms of data collection, Percentage Point Company employs a comprehensive public opinion monitoring system that covers online media extensively, captures information in real time, and analyzes semantics accurately. Coupled with public opinion analysis modules, this system enables deep understanding of public attitudes toward hot topics. The network public opinion data capture methodology includes: crawling relevant historical data from news outlets, forums, bulletin boards, Sina Weibo, and WeChat within specified timeframes; performing text segmentation and data preprocessing; utilizing TF-IDF models

to calculate term weights and generate term popularity metrics; clustering or classifying relevant information; and determining its positive, neutral, or negative sentiment orientation.

Monitoring period: the past six months. Statistical indicators: volume of discussion (attention level), topic heat, topic sentiment, etc. The process culminates in data visualization reports that analyze trending Two Sessions terms and the trajectory of public discussion volume.

Since 2016, Percentage Point Company has supported Xinhua News Agency's "Big Data View of Two Sessions" series, providing stable and reliable big data technical support through its public opinion monitoring and text analysis systems. The series and its visualization products have been published across Xinhua's "two micro-platforms and one app" (Weibo, WeChat, and mobile application).

In the 2017 "Big Data View of Two Sessions" collaboration between Xinhua News Agency and Percentage Point Company, the project reviewed hot topics from the 2014-2017 Two Sessions and conducted analysis across the following dimensions: topic attention rankings during each year's Two Sessions; concerns of different demographic groups by gender, age, and occupation, as well as geographic distribution; and which demographic groups heatedly discussed the most-liked (most-viewed) reports from each year's Two Sessions.

In 2017, Percentage Point Company also collaborated with People's Daily Central Kitchen, China Daily, and other media outlets to launch special features. Between March 1-9, 2017, Percentage Point's public opinion monitoring system analyzed data from Xinhua Net, People.com.cn, NetEase, Sina, Toutiao, Sohu, Sina Weibo, and other platforms across the entire network, revealing the following trends: Development is fundamental and the most important topic at the Two Sessions. Participation among Sina Weibo users in Beijing reached a new high of 24.1%, an increase of approximately 12% compared to 2016. Guangdong, Shanghai, Shandong, and Jiangsu remain active regions for Two Sessions discussions. The analysis examined regional differences in public attention toward Two Sessions topics.

The participation of Party and state leaders in delegation deliberations and discussions represents a topic of significant external focus. In the Two Sessions special feature produced with China Daily, Percentage Point Company first collected data based on relevant keyword queries, then applied model algorithms and text analysis systems to identify unstructured text, and conducted sentiment analysis, topic monitoring, and reputation detection regarding public opinion on leaders' participation in delegation meetings.

Percentage Point Company's collaborations with central media outlets such as People's Daily, Xinhua News Agency, and China Daily encompassed data collection before the Two Sessions to predict hot topics, through to supporting special coverage during the Two Sessions via public opinion systems, model algorithms, and text analysis systems. These partnerships demonstrate that

data journalism has become an urgent need in the media industry and a trend in media transformation and development. For media organizations, big data technology has permeated every stage of news reporting, including: mining news hot spots; tracking events to improve reporting quality; innovating news presentation forms through data visualization; implementing precise user profiling; quantifying and establishing dissemination impact metrics for media platforms and evaluating communication effectiveness; and analyzing user behavior patterns and characteristics through distributed computing, machine learning, and other methods.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.