Journal of Artificial Intelligence Practice (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2025.080306
ISSN 2371-8412 Vol. 8 Num. 3

Dynamic Public Opinion Hotspot Analysis System Architecture Based on Multi-Source Fusion

Jiaojiao Lu[1,2], Yanyan Qiao[1,2]

[1]*School of Information & Intelligence Engineering, University of Sanya, Sanya, Hainan, 572022, China*
[2]*Academician Chunming Rong Workstation, Sanya, Hainan, 572022, China*

Keywords: Multi-Source Fusion; Public Opinion Hotspot Analysis; Public Opinion Monitoring

Abstract: With the rapid proliferation of social media platforms, the impact of online public opinion on social governance has become increasingly pronounced. To overcome the limitations of traditional public opinion analysis methods-particularly their poor real-time performance and insufficient data coverage-this study proposes a dynamic public opinion hotspot analysis system architecture based on multi-source fusion, which integrates multi-platform data streams utilizing automated collection, intelligent sentiment analysis, and geospatial mapping technologies. The system enables real-time monitoring, adaptive early-warning with configurable thresholds, and AI-driven visual analytics. This research provides government agencies and enterprises with an efficient public opinion monitoring tool, demonstrating significant practical value for enhancing crisis response capabilities and improving social governance precision through data-driven insights.

1. Introduction

With the rapid development of social media, the influence of online public opinion has grown significantly. While positive public opinion can promote social progress and information dissemination, negative public opinion has the potential to mislead public judgment, trigger the escalation of public opinion incidents, and even damage corporate images and social stability. Therefore, effectively monitoring, accurately analyzing, and promptly warning about online public opinion has become a critical research topic in modern social governance.

In academic research, Li Xiaoxiong [1] innovatively applied machine learning (including algorithms such as SVM and LSTM) to construct an online public opinion classification model. By optimizing text feature extraction and classification algorithms, the model's superiority in accuracy and efficiency was experimentally verified. This work provides automated solutions for government and corporate public opinion monitoring, enhancing the timeliness and accuracy of its management. Reference [2] established a sentiment analysis model for social media texts based on the SIS model, proposing a multidimensional governance system that encompasses government evaluation and early warning, platform regulation, and user media literacy improvement. Reference [3] developed a big data public opinion analysis system based on the Hadoop architecture. Utilizing data mining technology, it achieves risk assessment, trend prediction, and real-time early warning for public

opinion incidents throughout their early, middle, and late stages, providing technical support and a decision-making basis for scientifically guiding online public opinion. Zhou C [4] advanced the research further by proposing a method combining Long Short-Term Memory (LSTM) networks with the Expectation-Maximization (EM) algorithm, which significantly improved sentiment analysis accuracy.

These studies demonstrate that domestic and international research in online public opinion analysis possesses distinct strengths and offers opportunities for mutual learning, jointly promoting the continuous advancement of the technology. Compared to traditional methods relying on manual data collection and analysis, constructing a multi-platform online public opinion hotspot analysis system enables automated processing and analysis of massive data. This provides real-time monitoring, early warning, and analysis of trending public opinion, along with response strategies to help relevant parties address the impact of negative public opinion.

2. Online Public Opinion Data Collection, Preprocessing, and Storage

2.1 Online Public Opinion Data Collection

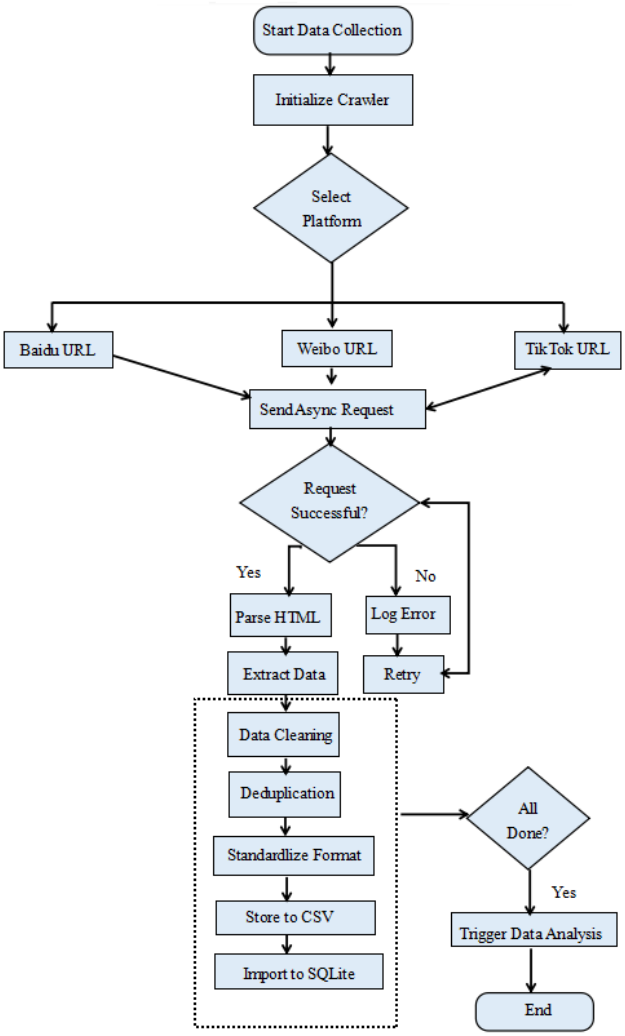

Figure 1: Data Collection Process

Data collection for online public opinion analysis is the primary step in building a hotspot analysis system model. The integrated multi-platform data collection solution designed in this study

can obtain real-time trending data from major social platforms including Baidu, Weibo, and TikTok, achieving efficient data storage through standardized processing. Implementation employs platform-specific strategies: Weibo and TikTok data are acquired by configuring third-party APIs using the aiohttp asynchronous HTTP library for high-concurrency requests, enabling rapid data retrieval, while Baidu Hot Search data is obtained through direct HTML page content parsing. During data reception, since formats differ across platforms, the system uniformly processes heterogeneous data sources by executing standardized mapping of key fields such as titles, popularity values, and URLs to ensure data format consistency. The collected data is initially stored in platform-categorized CSV files, then processed using the pandas library for deduplication and popularity ranking, with text segmentation performed via the jieba library to prepare for subsequent text analysis. Ultimately, the processed data is imported into time-series tables in an SQLite database, automatically triggering the analysis mechanism. The data collection workflow is illustrated in Figure 1.

2.2 Preprocessing of Online Public Opinion Data

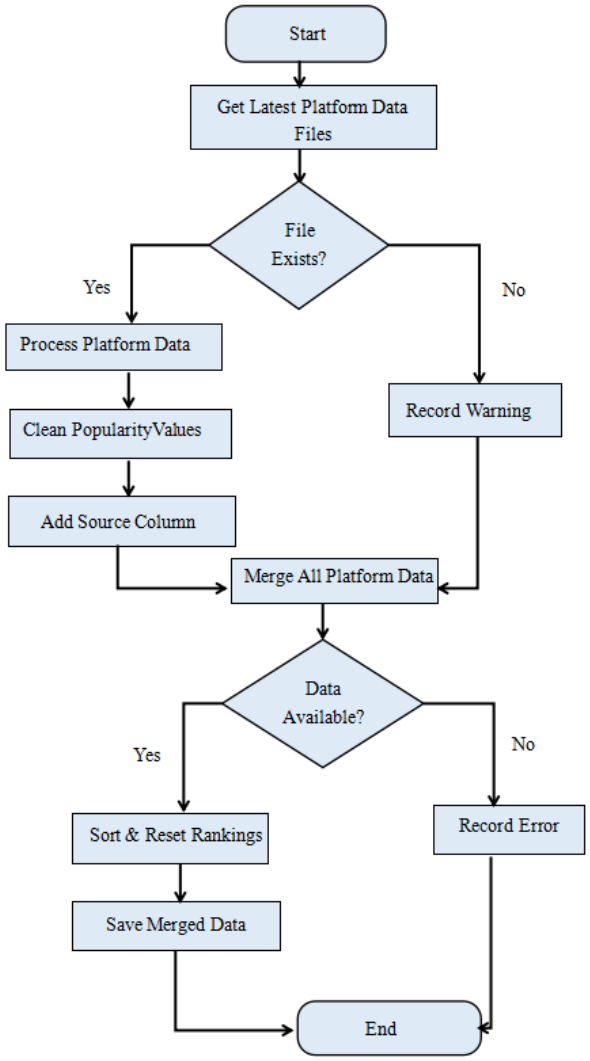

Figure 2: Data Processing Workflow

The collected data frequently exhibits duplicate records, inconsistent formats, and missing fields, requiring systematic preprocessing. The model implemented in this study enables unified format

conversion for multi-source heterogeneous data through predefined standardization protocols. Furthermore, the integrated system annotates each data entry with its source platform-a key capability that facilitates subsequent data provenance analysis while providing essential references for visualizing public opinion propagation paths. This functionality allows users to visually identify dissemination patterns across platforms. The preprocessing workflow for online public opinion data is illustrated in Figure 2.

2.3 Online Public Opinion Data Storage

Following collection and preprocessing, data enters the storage phase. The system model employs a layered storage strategy: initially preserving raw platform-specific data as CSV files, then performing cross-platform data fusion, before persisting the consolidated datasets into a SQLite relational database. This solution ensures data traceability while providing structured support for subsequent time-series analysis and trend prediction. The online public opinion data storage process is depicted in Figure 3.

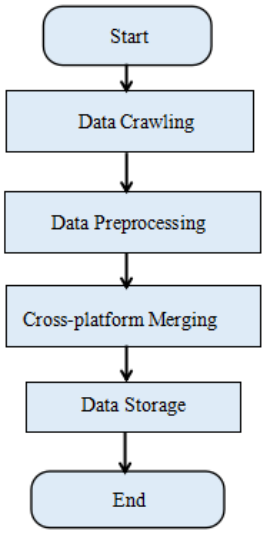

Figure 3: Data Storage Flowchart

3. Online Public Opinion Analysis

The online public opinion analysis system adopts a modular architecture that stores structured analysis results in the file system by generating independent CSV report files for each data platform. In the sentiment analysis module, the system first extracts trending data from the database, filters it by platform, and then uses an optimized sentiment analysis model to perform multidimensional sentiment evaluation of text data. For each trending topic title, the system calculates dual indicators: a sentiment score (ranging from -1 to 1) and a sentiment level (positive/neutral/negative), storing the results in DataFrame. To visually display the distribution of different sentiment tendencies, the system provides dual visualization options through word clouds and bar charts. Word clouds intuitively showcase the emotional attributes of public opinion keywords through word size and color coding (red for negative, green for positive), while bar charts quantitatively display the proportional distribution of different sentiment levels, allowing users to quickly grasp the overall sentiment distribution at a glance. The system supports free selection of visualization graphics, enabling users to rapidly understand the overall public opinion situation.

3.1 Public Opinion Early Warning Analysis

The system adopts a multi-dimensional early warning mechanism that integrates real-time data from Weibo, TikTok, Baidu and other platforms to establish an intelligent public opinion monitoring system. The interactive warning interface displays sentiment analysis results of trending topics across platforms, automatically sorting data by sentiment scores from low to high and highlighting negative public opinion below preset thresholds in red for rapid risk identification. The system supports customizable warning thresholds - when sentiment scores fall below user-defined values, the warning mechanism is automatically triggered. Warning results support multi-dimensional sorting, allowing users to sort by either sentiment scores or topic popularity to meet different analytical needs. To enhance user experience, the system embeds deep links between titles and original content - a single click on any trending title directly jumps to the corresponding platform's details page for quick information access. Additionally, flexible data source switching enables users to selectively view warning information from specific platforms for targeted monitoring. This multi-platform, multi-dimensional warning mechanism ensures both comprehensive coverage and precise analysis. The detailed workflow is shown in Figure 4.

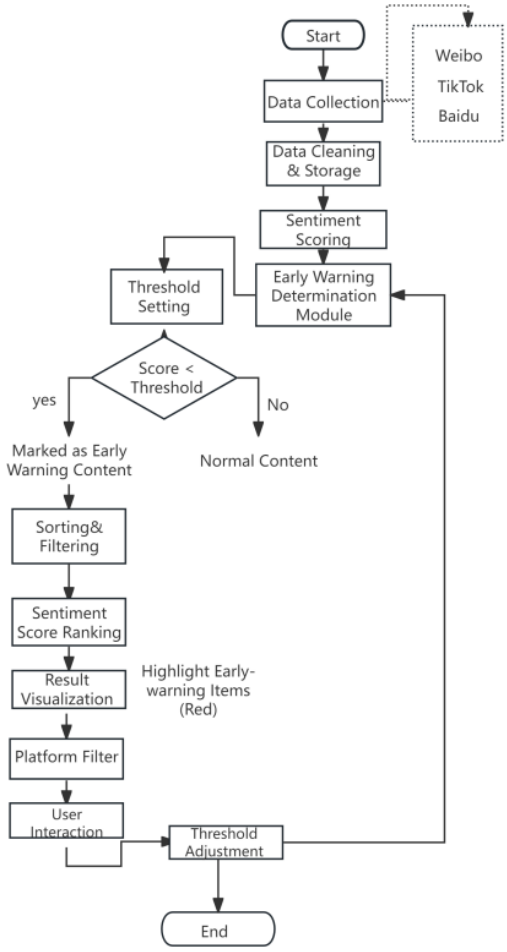

Figure 4: Public Opinion Early Warning Flowchart

3.2 Public Opinion Report Analysis

The public opinion report analysis module utilizes intelligent generation technology to achieve

efficient report output through the integration of multiple AI large models. The system supports users in flexibly invoking large models with different performance characteristics such as DeepSeek V3, DeepSeek R1 or Qwen2.5 VL72B for in-depth analysis based on real-time monitored trending data. The report generation process adopts a dual mechanism combining "AI analysis with template-driven" approach, ensuring both professional quality and efficient output. Users can select different AI models to generate reports with varying analytical emphases according to their actual needs: DeepSeek V3 is suitable for comprehensive analysis, DeepSeek R1 focuses on rapid response, while Qwen2.5 VL72B specializes in in-depth interpretation. The entire report generation process automates the workflow from data input to final report output. The reports are presented with clearly hierarchical headings and paragraph structures to facilitate quick browsing and focused reading. The specific workflow for public opinion report analysis is shown in Figure 5.

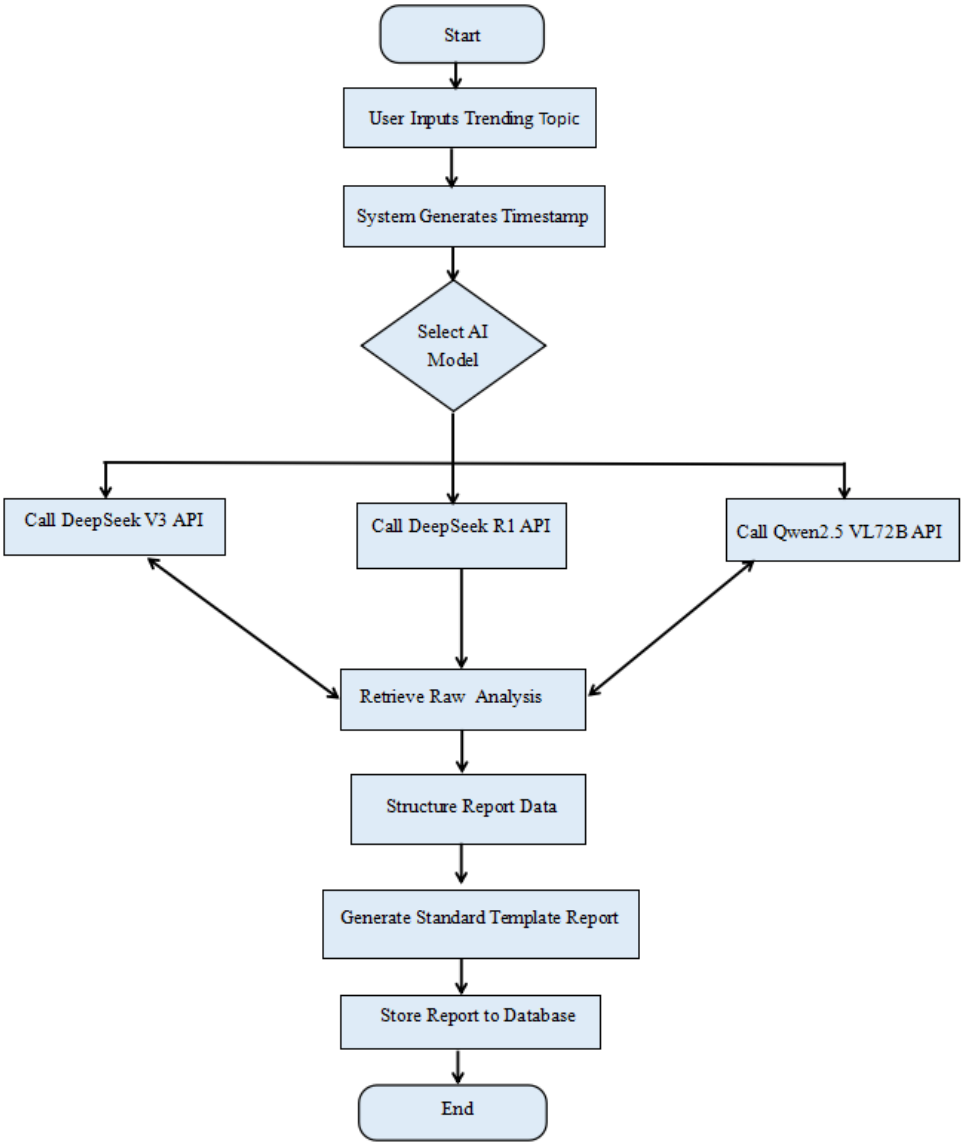

Figure 5: Public Opinion Analysis Process Diagram

3.3 IP Address Analysis

The system model includes a regional public opinion monitoring module. When users input keywords, the system automatically calls APIs from platforms such as Weibo, TikTok, and Baidu to

obtain comment data related to trending topics. Through a professional data cleaning process, the system first filters invalid IPs and abnormal data while retaining genuine user geographic location information. It then utilizes a high-precision IP address database for geocoding, accurately mapping IP addresses to provincial-level administrative divisions and calculating the distribution ratios of comment volumes across provinces. This module visually presents regional attention disparities in public opinion events through heatmaps and other visualization methods. For example, when an incident occurs in a specific region, the system can clearly demonstrate that discussion activity from netizens in that province significantly surpasses other areas. Such location-based data analysis provides governmental departments with insights into geographic distribution characteristics of public opinion and helps enterprises identify market focus areas, supporting the formulation of more targeted public opinion response strategies to enhance the precision and timeliness of public opinion management.

4. Conclusion

Online public opinion analysis serves as a crucial tool for social governance and public decision-making, with its research significance becoming increasingly prominent. This study innovatively proposes a multi-platform integrated online public opinion hotspot analysis system model to address prevalent issues in current opinion analysis, including low efficiency in manual processing, limited data coverage, and inadequate real-time response capabilities. The system enables fully automated end-to-end processing from data collection, preprocessing, and storage to in-depth analysis and visual presentation, aiming to enhance both the efficiency and accuracy of public opinion analysis. Future work will focus on continuously optimizing system performance, with particular emphasis on developing real-time online public opinion monitoring and early warning functions. By establishing a more intelligent monitoring framework, the system will achieve rapid identification and early warning of public opinion information, providing timely and accurate data support for scientific decision-making by government departments and enterprise risk management, thereby effectively mitigating potential social impacts of negative public sentiment.

Acknowledgement

This research was supported by the Hainan Provincial Higher Education Institutions Scientific Research Project of China (Hnky2023-40); The Hainan Provincial Natural Science Foundation of China (621QN0900); The Hainan Provincial Higher Education Institutions Scientific Research Project of China (Hnky2024-46); The Hainan Provincial Higher Education Institutions Scientific Research Project of China (Hnky2025-38)；The Hainan Provincial Higher Education Institutions Scientific Research Project of China (Hnky2025-29)；The Sanya University's Product-Thinking Oriented Characteristic Curriculum Reform Project. (SYJKCP2023150)

References

[1] LI Xiaoxiong. Research on machine learning-based online public opinion classification algorithms[D]. Central South University, 2022.

[2] You G, Gan S, Guo H, et al. Public opinion spread and guidance strategy under COVID-19: A SIS model analysis[J]. Axioms, 2022, 11(6): 296.

[3] Bi J, Chen X, Wang S. Development of Network Public Opinion Analysis System in Big Data Environment Based On Hadoop Architecture[J]. Procedia Computer Science, 2023, 228: 291-299.

[4] Zhou C.Online Public Opinion Analysis Model Based on Long Short-term Memory Network and Expectation Maximization[J]. Applied Mathematics and Nonlinear Sciences, 2024, 9(1).