Leveraging Media Big Data to Build a Healthy China–Deconstruction of the Health Food Risk Early Warning Platform Postprint

Authors: Wang Chong

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

Abstract

China' s sustained rapid economic growth and the enhancement of public health awareness have further propelled the rapid development of the health food industry. However, frequent irregularities within the industry have triggered public concerns regarding health food products and engendered distrust toward regulatory authorities. China Health Media Group has launched a health food risk early warning platform to assist regulatory bodies in standardizing the industry and safeguarding public health.

Full Text

Using Media Big Data to Help Build a Healthy China–Deconstructing the Health Food Risk Early Warning Platform

Author: Wang Chong

Abstract

China' s sustained economic growth and rising public health consciousness have further propelled the rapid development of the health food industry. However, frequent irregularities in the industry have sparked public concerns about health foods and eroded trust in regulatory authorities. In response, China Health Media Group has launched a health food risk early warning platform to assist regulatory agencies in standardizing the health food industry and safeguarding public health.

Keywords: big data; health food; risk early warning

Introduction

With rapid economic development and continuously improving living standards, public health awareness has gradually strengthened, with increasing emphasis on health preservation and wellness. As foods with specific health functions that can regulate bodily functions, health foods have gained growing popularity among the public. The *Health Food Registration and Filing Management Measures* officially implemented on July 1, 2016, defines "health food" as foods that claim to have specific health functions or are intended to supplement vitamins or minerals—suitable for specific populations, capable of regulating bodily functions, not intended for disease treatment, and that do not cause any acute, subacute, or chronic harm to the human body [1].

In recent years, health foods have emerged rapidly, offering more choices to consumers but also creating numerous problems, including illegal drug additives during production, sales of counterfeit and substandard products, and exaggerated product efficacy claims. These issues have caused significant public distress and created substantial difficulties for market supervision. To help the public correctly understand health foods and assist regulatory authorities in strengthening health food management, China Health Media Group took the lead in launching a health food risk early warning platform. As the first risk early warning platform in China' s health food sector, this platform represents a knowledge achievement that closely integrates Chinese-language-centric natural language technology with health food business applications. Relying on deep learning techniques and combining computational linguistics with domain knowledge, social computing with knowledge, and humanities computing with knowledge, the platform undergoes systematic and in-depth research through sample training, machine learning, and professional judgment based on human experience to continuously improve. Through business cooperation and data API services, the platform provides data and application services for food and drug safety regulatory authorities and related institutions.

Platform Overview: The "Smoke Index" Concept

Internally, the platform is known as the "Health Food Smoke Index."The "Smoke Index" concept derives from the analogy that "smoke in a forest signals an impending fire," using this vivid metaphor to infer the relationship between the smoke index and risk levels—namely, judging risk levels based on the smoke index values of products and enterprises. The platform significantly reduces operational costs compared to traditional supervision methods while improving regulatory coverage and efficiency [2]. Through big data collection and analysis technologies, the "Health Food Smoke Index" early warning platform monitors health foods around the clock without gaps. Based on industry characteristics, it collects raw data at scale from e-commerce website reviews, online forums, social media, and other sources, integrates industrial and commercial data and court enforcement data, and classifies and stores information according to domestic/import status, functional efficacy, and target populations within the

platform. Through distributed computing on an in-memory computing platform, the data undergoes preprocessing including cleaning and integration to transform datasets into structured data suitable for analysis. Analytical algorithms then perform big data calculations based on established risk prediction models, generating risk rankings (top N) across dimensions such as brand, product, and price through an indexed approach, enabling true early warning before risks materialize.

In 2017, to fully implement the *Food Safety Law of the People' s Republic of China* and the "Four Strictest" requirements, the Office of the State Council Food Safety Committee, together with relevant departments, formulated the *Food and Health Food Fraud and False Advertising Rectification Plan* to further strengthen supervision over food and health food production, operation, and import units, crack down on illegal marketing and false efficacy claims that mislead and deceive consumers, and promote the implementation of territorial management responsibilities. On January 8, 2019, 13 departments including the State Administration for Market Regulation, Ministry of Industry and Information Technology, Ministry of Public Security, and National Health Commission held a teleconference, deciding to launch a nationwide 100-day joint campaign to rectify "health product" market chaos starting from January 8, 2019. This campaign intensified post-event supervision of key industries, fields, and behaviors in the "health product" market nationwide, and legally cracked down on various illegal activities including false advertising, fake advertisements, manufacturing and sales of counterfeit products, and other market order disruptions and consumer fraud.

Technical Architecture

Data Collection and Integration

The data collection platform supports batch import of Excel and other format data and connects with internal systems (data from the State Administration and local regulatory bureaus) through interfaces. By analyzing research on health food activity patterns, characteristics, and evolution, the platform broadens information acquisition channels. The monitoring platform integrates information resources from various regions and departments, promoting interoperability among industrial and commercial market entity disclosure information, internet data, and court case judgment and enforcement information, while capturing multiple characteristic information for comparison.

Database Design. For the China Health Food Risk Early Warning Platform, database tables are reasonably designed for collected enterprise data information. Data table items are graded according to the data dimensions required by models in the model management platform to maximally satisfy model computation needs.

Internal Data Access. The platform automatically captures and extracts relevant industry, enterprise, and policy information from designated websites

or data sources, displaying it in corresponding areas as links or text to enable intelligent public opinion analysis, including sentiment analysis, viewpoint extraction, geographic distribution, and article classification, providing a visual public opinion monitoring platform.

Third-party Data Access. The platform interfaces with data providers such as Meituan and Ele.me.

Industrial and Commercial Data Integration. The platform connects with enterprise industrial and commercial information within target scopes, structurally storing various enterprise information to ensure data quality and scale, with accuracy reaching 100% consistency with State Administration for Market Regulation data. Data update rules are established with clear update frequencies, and existing enterprise data is regularly updated.

E-commerce Data Collection. Collected fields include original links, product names, product IDs, prices, origins, specifications (different specifications correspond to different products), brand enterprises, main ingredients, gross weight, target populations, product formulations, symptoms, blue hat certification marks, domestic/import status, shelf life, JD.com self-operated status, stores, shipping locations, store names, promotional images, positive review rates, review tags, total review counts, image review counts, follow-up review counts, positive/neutral/negative review counts, etc. Promotional images are collected to enable image traceability.

Natural Language Processing

With the surge in artificial intelligence, major domestic enterprises have begun deploying AI fields, creating various intelligent terminals such as humanoid robots, autonomous vehicles, smart TVs, and smart refrigerators. These intelligent terminals share a common characteristic—they can understand human language, communicate with humans, and further execute human commands. This remarkable technology is realized through a core AI processing technology: NLP. NLP (Natural Language Processing) is a discipline that studies language issues in human-computer interaction and an important subfield of artificial intelligence [3]. Simply put, NLP enables machines to "understand" the structure and meaning of natural language used by people, translate it into machine language form, process it (summarization, syntactic analysis, etc.), and return it to users in natural language. It involves numerous contents and technologies, including text-to-speech/speech synthesis, speech recognition, Chinese word segmentation, part-of-speech tagging, syntactic analysis, natural language generation, text classification, information retrieval, information extraction, text proofreading, question-answering systems, machine translation, automatic summarization, textual entailment, etc.

Since the early days of AI development, NLP technology has demonstrated tremendous potential. In 1949, Edmund Berkeley wrote in his book *Giant Brains Or Machines That Think*: "Much has been said recently about strange

giant machines that process information with great speed and skill···This machine resembles a brain, composed of hardware and wires rather than flesh and nerves. The machine can process information, calculate, draw conclusions, make choices, and perform reasonable operations based on information. In short, this machine can think" [4]. As one of the core AI technologies, natural language processing has increasingly gained favor among technology companies. In the *New Generation Artificial Intelligence Development Plan* issued by the State Council, natural language processing technology is listed as a key common technology.

Risk Analysis Models

Risk Monitoring and Analysis Mining Model Design. The object-oriented risk monitoring analysis mining model—the OFV Risk Analysis Model (Monitoring Object O - Risk Feature F - Feature Value/Feature Degree V)—is supported by rich underlying resource libraries and mining logic expressions. The underlying resource libraries include a monitoring object library, business knowledge classification library, and monitoring element network. Risk features are a series of special grammar generation patterns, with a finite set containing system default elements and various elements from resource libraries. System default elements (IR_) include content, title, time, region, keywords, etc.; resource library elements include monitoring objects (O_), business knowledge (K_), and monitoring elements (E_).

Index Model Design. The index model comprehensively considers blue hat certification marks, sales volume, negative review counts, star ratings, public opinion dissemination data, and other factors, training on corpora to obtain final values.

Risk Grading Model Design. Scores calculated from risk assessments are divided into several early warning levels: normal monitoring for health foods with risk indices of 0-40; key monitoring for 41-60; high-risk warning for 61-80; and investigation intervention for 81-100.

By analyzing various data and past experiences, the platform establishes health food risk assessment models using data mining, statistical analysis, and other technical methods. The system performs repeated iterations and pattern matching on massive accessed data according to final formed models, identifying suspicious behaviors in data, triggering corresponding early warnings, and establishing a risk index system to analyze and characterize health food risk levels.

Advanced Features. The construction of China' s health food risk early warning platform shifts the risk control threshold forward, achieving a leap from "post-event public opinion monitoring" to "pre-event risk warning," providing important decision-making references for regulatory authorities and holding significant meaning for enhancing government credibility, promoting healthy industry development, and improving social co-governance.

Platform Significance

China Health Media Group is a central cultural enterprise funded by the State Council with the Ministry of Finance acting as the investor representative, with administrative management and party leadership relations 隶属于 the National Medical Products Administration. With the mission of serving food and drug regulation and public health, the group' s business scope covers books, newspapers, periodicals, internet (public opinion monitoring and big data), education and training, exhibitions, film and television, and health services. The group currently owns multiple news, publishing, and media entities including China Pharmaceutical News, China Medical Science and Technology Press, China Health Media Group Tianchuang Science and Technology Information Technology Co., Ltd., China Health Media Group Tianhe Exhibition Co., Ltd., Tianxiang (Beijing) Health Technology Development Co., Ltd., China Food and Drug Regulation Magazine, and China Food and Drug Network. Looking to the future, China Health Media Group is committed to building a leading enterprise in China' s health media field, deeply integrating news, publicity, and publishing resources under the National Medical Products Administration, developing more new media businesses adapted to market demands, and comprehensively interpreting new trends in food, drugs, health foods, cosmetics, medical devices, and the broader health industry from all perspectives and with new visions to serve regulatory upgrades, industry development, and public health.

By leveraging its profound resources to build China' s health food risk early warning platform and prevent health food "black swan" events, the group plays an important role in helping regulatory authorities identify potential hidden dangers in the health food industry early and standardize the health food industry. Health food supervision is a long-term, complex, and arduous systematic project requiring high attention from governments at all levels and relevant departments. From the perspective of maintaining stable economic development and social harmony and stability, it is essential to strengthen the sense of urgency, responsibility, and mission for preventing and handling health food market chaos, unify thoughts and actions with central government decisions and deployments, and treat prevention and handling of health food chaos as an important task affecting the overall situation. Efforts must be intensified to maximize risk prevention and resolution, effectively protect the legitimate rights and interests of the people, resolutely curb the high incidence and spread of health food market chaos, firmly maintain the bottom line of preventing regional and systemic risks, and provide strong guarantees for sustained and healthy economic and social development.

The platform's construction can provide fee-based services such as big data analysis and personalized data customization for party and government institutions, public institutions, and enterprises in need. Simultaneously, it establishes the group' s big data brand, provides demonstration effects for the industry, and enhances brand influence. It can provide powerful data support for leaders to grasp health food dynamics, assess situations, make scientific decisions, and

guide work.

References

[1] China Food and Drug Administration. *Health Food Registration and Filing Management Measures* [Z]. 2016-02-26.

[2] Li Chonggang, Xu Huiquan. Smoke Index: Big Data Monitoring of Internet Financial Risks [J]. *Big Data Research*, 2018, 4(4): 76-84.

[3] Pang Yu. Research on Key Issues in Computational Linguistics Based on Chinese NLP [J]. *Computer Products and Circulation*, 2019(1): 236.

[4] Berkeley E C, Frankel S P. Giant Brains or Machines That Think[J]. *Physics Today*, 1950, 3(5): 39.

Author' s Affiliation: China Health Media Group Co., Ltd.

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