

Overview of Fusion Technology in Food Safety Big Data

Yajie Wang^{1,2}, Bing Yang^{1,2,*}, Hong Yang¹, Miao Hao^{1,2}, Chengmei Zhang^{1,2} and Heng Tao¹

¹ Guizhou Academy of Sciences Big Data Co. LTD, Guiyang, Guizhou, 550000, China

² Guizhou Academy of Testing and Analysis, Guiyang, Guizhou, 550000, China

Abstract. Large amounts of organised and unstructured data from various sources and origins are increasingly being handled by technology. Big data is a term used to describe these technologies, which open up new fields of research and applications that will have an expanding impact on all aspects of our society. Because of its potential to turn massive amounts of data into insights for informed business and operational choices, big data has found uses in a variety of industries. In the food industry, advanced techniques have been used for providing Food Safety. We present an overview of how and to what extent big data is being used in the food safety domain in this study. Mobile phones as food safety detecting devices and the use of social media as an early warning system for food safety issues are just two examples of the new advancements made feasible by big data.

1 Introduction

The growth of data in practically every sector of our society has ushered in the big data era, which is a hot topic influencing many parts of our life [10]. Nearly every area of society, including business, government, health care, and research fields such as natural sciences, biological sciences, engineering, humanities, and social sciences, produces vast amounts of data. However, the aggregation and speed with which big data is generated necessitates overcoming obstacles connected to data collection, storage, and processing. There are many big data applications which are available in market.

Food safety data and information are dispersed throughout the food, health, and agriculture sectors, hence the term "big data" is rarely used in this context. With the help of data mining [6], patterns representing knowledge implicitly hidden in massive data can be automatically extracted for assisting human decision making and can be used for production optimization and quality and safety assurance of food. Traditional food safety statistics, such as national monitoring data, are limited but highly structured, but they are not always harmonised between regions. In this study we analyse if and to which extent big data play a role in food safety [14]. For example, JECFA [13] Evaluations Database is used to identify and monitor the hazards of food fraud and food safety. Databases from government agencies, international authorities, and scientific fields are usually high in data credibility.

2 Data Availability

Several databases are available have been used to increase our understanding of food-related topics such as food safety and fraud, as well as the link between food and disease in complex food systems. Various sorts of sources, such as (online) databases, the internet, omics profiling, mobile phones, and social media, can include or provide information important for food safety. The task at hand is to find important data inside a data source and connect it to other data sources. There are numerous data sources available in the market like, RASFF [4] for notification from Rapid Alert System for Food and Feed, USDA Production supply and Production online [1] official USDA data for food commodity and agricultural items supply, Global G.A.P [7] for the producers, which can be used for analyzing and ensuring food safety.

3 Process

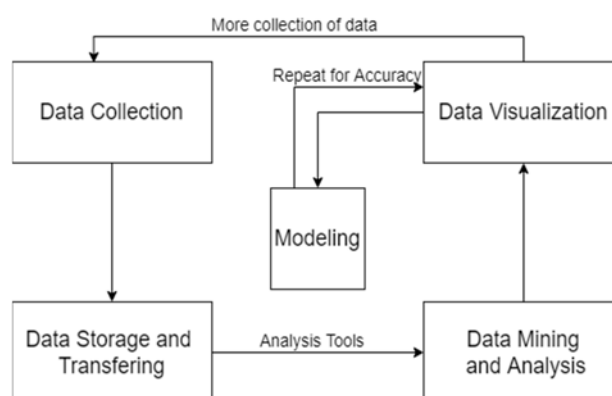


Figure 1. Big Data Architecture

* Corresponding author: yangbing@gzbd.com

We can see the whole big data structure in Figure 1, which consists of collecting more and more publically available data, and then using software to store massive amounts of data for subsequent processing and analysis. Then, following the analysis, data visualisation is utilised to gain a better understanding of the data and to model it.

3.1 Data Collection

The main channels for gathering data linked to food safety include (online) databases, the internet, omics profiling, sensors, mobile phones, and social media. In addition, sophisticated surveillance systems using new technologies to collect data on food safety have been introduced. Video surveillance, Internet of Things (IoT) sensors and portable devices, Geographic Information System (GIS) [3], satellite images, and blockchain technology [17] are examples of such technologies.

3.2 Data Storage and Transferring

Data management systems such as MySQL, Oracle, and PostgreSQL are commonly used to store data as shown in Table 1. However, such systems are insufficient for storing and processing large amounts of data. In these instances, much more speed, flexibility, and dependability are required than standard systems can provide.

Table1. Data Storage and Transfer Technologies

Technology	Tool	Type
SQL	MySQL, Oracle PostgreSQL	Data Storage
NoSQL	HBase Cassandra MongoDB Big Table	Data Storage
Computational Technologies	Hadoop MapReduce Spark	Data Preprocessing
Data Transfer	Hive, Aspera Apache Flume Talend	Data Transfer

As a result, next-generation databases, known as NoSQL, have been developed that are nonrelational, open source, and horizontally scalable. MongoDB, Cassandra, and HBase are examples of such systems. Following storage, moving large data from many sources into a NoSQL cluster for processing is the next hurdle. This necessitates the usage of data transfer software, such as Aspera and Talend, which are examples of such software used to handle huge data.

3.3 Data Mining and Analysis

Data mining and analysis is a collection of techniques in the process of knowledge discovery [5]. Data cleaning (to prepare specific data to be mined), data integration (to consolidate data from different sources into a single dataset), data selection (to obtain relevant data), and data

transformation (to combine unstructured data with structured data to analyze) are the four primary steps of data mining and analysis, as shown in Figure 2.

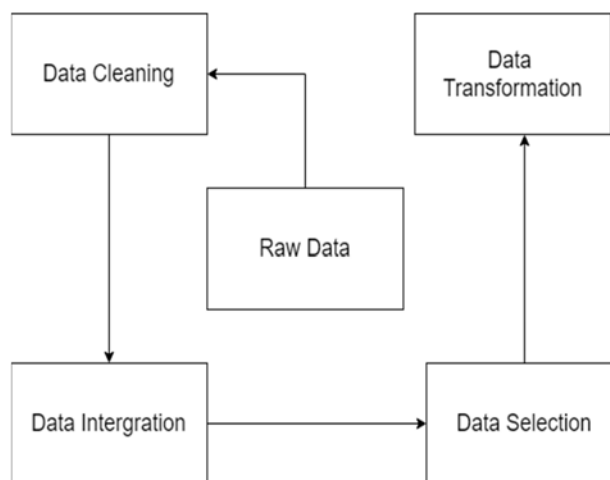


Figure 2. Mining and Analysis Process

3.4 Data Visualization

There are several visualisation tools available to evaluate and show summaries of large amounts of data, each with its own set of benefits and drawbacks. R and Cicos are the most often utilised. R is an open source programming language for visualising and analysing data in data science. It has plot and network plot capabilities. Cicos allows you to investigate relationships between objects or positions by visualising data in a circular arrangement. IBM Many Eyes and Tableau are strong choices for commercial visualisation software that does not require programming knowledge. Various visualization tools are: - R, Cytoscape, Tableau, PowerBI, GraphViz and Fusion charts.

4 Food Safety Applications of Big Data

4.1 Recommendation System

In e-commerce, recommendation systems are commonly used to target recommendations by filtering users' preferences, interests, or behaviour, however this method has rarely been employed in food safety [11]. Using the beef supply chain as an example, a recommendation system based on social media and Big Data analytics to notify supply-chain (SC) decision makers about food safety and quality issues.

4.2 Surveillance of food safety and fraud

Food safety is essential for guaranteeing food security and long-term food systems. It also plays an important role in public health. According to the Centers for Disease Control and Prevention, an estimated 800 foodborne outbreaks occur in the United States each year, resulting in 15,000 illnesses, 800 hospitalizations, and 20 deaths (CDC, 2018) [15]. Foodborne disease outbreaks continue

to represent a major concern to public health and the food industry, despite the development of various improved surveillance and monitoring systems. Another issue that generates government concerns about food consumption in the context of food safety is food fraud [9].

An algorithm for text classification To uncover hidden trends in disease outbreaks and find correlations between food types and outbreak locations, DT was proposed. They identified vehicles and locations related with specific etiologies using text data from the CDC Outbreak Surveillance Data. The information gathered can assist policymakers in developing effective food handling, preparation, and consumption behaviours interventions.

4.3 Econometric analysis

Food safety issues in international food trade based on food refusal reports or border inspection instances, such as hidden trade protectionism or the influence of import rejection risk of border inspection on food safety, are commonly studied using econometric analysis [2,8]. Scientists employed econometric analysis to investigate the influence of import refusal (i.e., a food safety issue) on developing countries' reputation and economy. Another controversial topic is seeing food safety as a hidden trade barrier. In recent years, several scholars have focused on analysing border sampling cases to give evidence of the non-tariff trade barrier.

4.4 Hygiene Connect

Although hand contact with food has long been recognised as a typical mechanism for infections to enter and spread throughout the food chain, there has been no efficient means to monitor – and so enforce – hand washing until recently. Hygiene Connect, a system of networked devices that delivers real-time insight into hand hygiene, anonymously monitors hand washing habits and gives feedback to both users and management, influencing behaviour and compliance. Hand washing compliance in businesses has improved from 46% to 70% on average, with some companies reporting compliance rates of over 85%.

4.5 Analysing food-buying patterns

In the event of a foodborne disease epidemic, timely tracing and identification of contaminated items is critical to minimising health and economic consequences, as well as identifying goods from which to collect samples for analysis.

Food sales data, which is collected automatically by grocery stores, can be used to detect food goods that are likely to be contaminated. Using a technique known as the probability-based method, a model of the chance of specific foods being ingested at different areas may be built and mapped against the sites of food borne disease outbreaks. This can aid businesses in swiftly identifying tainted merchandise in their stores.

4.6 FOSCOLLAB

The WHO has lately adopted a big data strategy to enhance food safety decision-making, resulting in the creation of the food safety platform "FOSCOLLAB," which integrates data from several disciplines. Structured and unstructured data from a variety of sectors, including animal, agriculture, food, public health, and economic indicators, are merged and made available to users via numerous specific dashboards on this platform [12]. The Joint FAO/WHO Expert Committee on Food Additives (JECFA) and Joint FAO/WHO Meeting on Pesticide Residues (JMPR) databases on chemical risk assessment, the WHO database on Collaborating Centres, and the GEMS food databases on food consumption and chemical occurrence in food are among the data sources connected in FOSCOLLAB [16]. As the FOSCOLLAB platform evolves, information from more data sources will be added, assisting actors involved in food and feed risk analysis. The Figure 1 depicts the several stages that can be identified when managing big data and has been adapted for food safety from the health sciences.

5 Challenges

Text data obtained from web sites, blogs, and social media, unlike structured data obtained through instrumental detectors, is unstructured and difficult to interpret. Despite the fact that social media is the most promising data source with the greatest applications, the ambiguities in social media data make it challenging to extract and interpret meaning, even with cutting-edge approaches. The nature of indigenous unstructured information is typically in unclear format due to missing, mislabeled, inaccurate, or potentially false information representation. Another issue that could explain the low adoption of IoT technology in food safety is that the data generated by IoT devices today is difficult to comprehend, convey, and share due to a lack of defined communication protocols. Using FAIR guiding principles in IoT devices will enable the Internet of Data and services to identify, communicate, and remain available for data exchange and reuse. Furthermore, there are various difficulties related with IoT security in food safety, such as insecure hardware and software. Any insecure IoT nodes in the food supply chain can be a security risk for the entire IoT system as well as the rest of the internet. The use of blockchain technology in food safety is promising, and it is projected to deliver safer and more transparent food chains in the near future, but it is still in its early stages and difficult to implement due to its complexity. Its use in food safety is currently restricted to traceability, but concerns like data integrity and governance still require further attention.

6 Conclusion

This research looked at the most current breakthroughs in Big Data applications in the field of food safety. The key sources of data for food safety are online databases, the Internet, sensors, mobile phones, and social media, according to this study. In the last seven years, new

technologies such as video monitoring, sensors and portable devices employing IoT technology, GIS, satellite imagery, and blockchain technology have been integrated in smart monitoring systems to collect data linked to food safety.

The above overview also demonstrates how the usage of social media data has enabled us to “see” hidden patterns through the use of a “big data scope.” However, poor data quality and privacy concerns limit the widespread use of social media applications. Other minor sources of text data have been leveraged to construct helpful information systems for assisting consumers in purchasing, cooking, and eating decisions. Overall, the findings revealed that this technology has significant potential, and successful applications for predicting, monitoring, and controlling food safety in the food supply chain have been documented. Many more are likely to follow soon, but it is evident that various obstacles, including sociological, governmental, and technical challenges, must be overcome in order to fully realise its promise.

Acknowledgement

This work was Supported by Guizhou Provincial Science and Technology Projects([2015]4005 & [2018]5404) and Science and Technology Projects of Guiyang National High-tech Industrial Development Zone (GXYP-2017-005) .

References

1. Inbal Becker-Reshef, Chris Justice, Mark Sullivan, Eric Vermote, Compton Tucker, Assaf Anyamba, Jen Small, Ed Pak, Ed Masuoka, Jeff Schmaltz, et al. Monitoring global croplands with coarse resolution earth observations: The global agriculture monitoring (glam) project. *Remote Sensing*, **2(6)**:1589–1609, (2010).
2. Matthias Beestermöller, Anne-Célia Disdier, and Lionel Fontagné. Impact of european food safety border inspections on agri-food exports: Evidence from chinese firms. *China Economic Review*, **48**:66–82, (2018).
3. Kang-Tsung Chang. Geographic information system. *International Encyclopedia of Geography: People, the Earth, Environment and Technology*, pages 1–10, (2016).
4. Ilija Djekic, Danijela Jankovic, and Andreja Rajkovic. Analysis of foreign bodies present in european food using data from rapid alert system for food and feed (rasff). *Food Control*, **79**:143–149, (2017).
5. Jiawei Han. Data mining techniques. In *Proceedings of the 1996 ACM SIGMOD international conference on Management of data*, page 545, (1996).
6. Jiawei Han, Micheline Kamber, and Jian Pei. Data mining concepts and techniques third edition. The Morgan Kaufmann Series in Data Management Systems, **5(4)**:83–124, (2011).
7. Spencer Henson, Oliver Masakure, and John Cranfield. Do fresh produce exporters in sub-saharan africa benefit from globalgap certification? *World Development*, **39(3)**:375–386, (2011).
8. Marie-Agnès Jouanjean, Jean-Christophe Maur, and Ben Shepherd. Reputation matters: Spillover effects for developing countries in the enforcement of us food safety measures. *Food Policy*, **55**:81–91, (2015).
9. Yasunobu Maeda, Norio Kurita, and Saburo Ikeda. An early warning support system for food safety risks. In *Annual Conference of the Japanese Society for Artificial Intelligence*, pages 446–457. Springer, (2005).
10. Hans JP Marvin, Esmée M Janssen, Yamine Bouzembrak, Peter JM Hendriksen, and Martijn Staats. Big data in food safety: An overview. *Critical reviews in food science and nutrition*, **57(11)**:2286–2295, (2017).
11. Rajhans Mishra, Pradeep Kumar, and Bharat Bhasker. A web recommendation system considering sequential information. *Decision Support Systems*, **75**:1–10, (2015).
12. World Health Organization et al. Guidance document for the integration of data in foscollab. Technical report, World Health Organization, (2015).
13. AG Renwick. Toxicology databases and the concept of thresholds of toxicological concern as used by the jecfa for the safety evaluation of flavouring agents. *Toxicology letters*, **149(1-3)**:223–234, (2004).
14. Aybuke Ceyhun Sezgin and Nevin Sanlier. Street food consumption in terms of the food safety and health. *Journal of Human Sciences*, **13(3)**:4072–4083, (2016).
15. John Spink and Douglas C Moyer. Defining the public health threat of food fraud. *Journal of food science*, **76(9)**: R157–R163, (2011).
16. D Stuckler. Understanding the relationship between alternative pension systems and healthy ageing: cross-national and quasi-natural experimental analyses of 27 eu countries. *European Journal of Public Health*, **30(Supplement_5)**: ckaal65– 766, (2020).
17. Dylan Yaga, Peter Mell, Nik Roby, and Karen Scarfone. Blockchain technology overview. arXiv preprint arXiv:1906.11078, (2019).