Promote Epidemic Prevention and Control through Big Data Technology

Yan Shu[1]; Liu Siyuan[2]

At the beginning of 2020, the pneumonia brought by novel coronavirus broke in in full fury. Different from the SARS outbreak in 2003, human society has now entered the era of big data and can resort to this technology to better control the epidemic outbreak thanks to the rapid development of the new-generation information technology in more than ten years. City managers have also gained a lot of experience in epidemic prevention and city management in the process of building new smart cities.

Big data technology can assist in epidemic prevention, traceability, treatment and tracking, and city management, logistics, information release, problem solving, and even virus gene sequencing, etc.. This paper will briefly analyze the common methods, problems and corresponding suggestions in epidemic prevention and control supported by big data technology, with the focus on the future application direction of the city big data platform in promoting epidemic prevention and control.

Ⅰ. Common methods of big data technology in epidemic prevention and control

In the face of the surging public health emergency and massive data from many sources, how should the governments and enterprises be united to provide the public with more complete, continuous, accurate and timely epidemic prevention information, experts with methods to trace the source of the disease, and decision makers with the development trend of the epidemic through the scientific use of big data technology are the three important tasks.

(1) Big data can analyze the moving trajectory of the people involved in the epidemic outbreak.

Big data can analyze the trajectory of the people involved in the epidemic by integrating the information from the telecom operators, internet companies, transportation departments and other units. Specifically, by means of the data analysis, data mining and other technologies, we can, for one thing, analyze and draw the patient's trajectory on the basis of the data like the mobile phone signaling that includes the geographic location and timestamp. And for another thing, the close contacts of the patient can be inferred according to the patient's trajectory and his or her companions before the diagnosis date. A comprehensive analysis of the trajectory of confirmed patients, suspected patients and close contacts can help us accurately depict the movement of different categories of people entering and leaving certain regions, which can not only provide guidance for the accurate treatment but also help predict high-risk regions and potential high-risk regions.

[1] Yan Shu, Doctor of Science, senior engineer of Big Data Department, Cloud Computing & Big Data Research Institute, China Academy of Information and Communications Technology; research focus: big data and data flow; e-mail: yanshu@caict.ac.cn

[2] Liu Siyuan works on big data technology research in Big Data Department, Cloud Computing & Big Data Research Institute, China Academy of Information and Communications Technology; e-mail: liusiyuan@caict.ac.cn

(2) Big data can trace the source of the infectious disease.

A scientific modeling can be made through the emerging technologies such as artificial intelligence and in-depth learning based on a large number of data such as travel trajectory, social information, consumption data, and exposure and contact history. Then the space-time collision points can be located according to the patient's diagnosis sequence and other information like his or her close contacts, and thus it is expected to calculate the disease transmission path and provide theoretical basis for the traceability analysis.

(3) Big data can predict the development trend of the disease.

Based on the movement of the high-risk groups, namely the confirmed patients and their close contacts, and the number of the newly confirmed, suspected, and cured cases and that of the death cases, we can not only analyze and display the thermal distribution of the disease and risk thermal distribution of close contacts but also predict important information like the peak inflection point of the epidemic situation, by introducing the big data models and technologies, such as the transmission dynamics model, dynamic infection model, and regression model. Then the health department can strengthen sanitary measures in key areas and get epidemic prevention and control resources placed in advance in regions where the disease may spread, so as to avoid secondary, local and multi-point outbreaks. Besides, the forecast of epidemic development trend is vital for governments to determine the work resumption time and introduce public management and economic development measures. Only when the situation is known fairly well (through the big data) can the decision-makers or ordinary people take preventive measures ahead of time.

Ⅱ. Problems in applying big data to epidemic prevention and control

Although big data plays an important role in accurate epidemic prevention and decision support, there is still much room for exploration and improvement given China's response to the COVID-19 outbreak. In order to push epidemic prevention and control forward by means of big data, the following three problems need to be solved urgently.

(1) Data are collected in a primitive way and of low quality. Some government departments and grassroots organizations still adopt the human-sea tactics, such as filling the form and making calls, in collecting epidemic-related data, which not only increases the burden on grass-roots staff but also fails to guarantee the authenticity of the data. For example, the patient may fill his cellphone number or that of his family members in the form, for which the operator found inconsistencies between identification numbers and mobile phone numbers in the verification process. In the end, it becomes rather difficult to guarantee the accuracy and authenticity of the subsequent data analysis.

(2) Privacy is at risk in data flow. The original data reported in an unencrypted way by the relevant personnel have been leaked to varying degrees, which makes the life of the persons concerned rather inconvenient. They may suffer groundless harassment and even regional discrimination and physical threats from strangers. In order to meet COVID-19 prevention needs, the relevant departments lost no time in developing the inquiry tool for the same travel with the confirmed patients with personal private information hidden. However, the data were still leaked to varying degrees in the reporting process. Take Miss. Wu, a freshman at China

University of Geosciences (Wuhan Campus) who had returned to her hometown from Wuhan by the high-speed rail, as an example. On the evening of January 24, she saw a table in her family WeChat group entitled *Data Sheet of Persons Returning to Ningdu from Wuhan*. In the table, the information of 400-500 people, including their identity numbers, telephone numbers, specific home addresses, train information, etc., was exposed.

(3) It is difficult to achieve data flow among regions, levels and departments. The core of big data is connectivity. Cell phone signaling can only provide relatively rough moving trajectory, which has little influence on regional analysis results. However, the high-accuracy epidemic prevention scenes are in want of multidimensional, cross-level and cross-domain information, such as the information from trains, planes, shared bicycles, online car-hailing, search engines, social media, etc.. At present, the data are like "isolated islands" of different enterprises and government departments. Therefore the data should be integrated efficiently to be better applied to epidemic control.

Ⅲ. Strategies and suggestions for enhancing epidemic control by means of big data

Based on the above-mentioned problems and the spreading epidemic situation, it is suggested that relevant departments should improve their work in the following three aspects.

First of all, the data collection method and data verification should be improved. Various technical means should be introduced to avoid collection inefficiency. For example, when disease control departments in different regions launch the epidemiological investigation of patients, they should not only directly investigate and interview patients but also collect from relevant departments the location-based service data (LBS data), such as the mobile phone signaling data, travel data and even data of WeChat, Alipay. These data are often more accurate than those manually collected. In addition, multiparty data verification is needed to improve data quality from the source and lay a good foundation for subsequent data analysis.

Secondly, patients' privacy should be protected and safety practices standardized. The more urgent the situation is, the more it is necessary to strengthen the protection of data privacy, otherwise unnecessary speculation and panic would be caused among the public. Two principles should be followed when governments release the information. The first is the desensitization principle, which requires that the content to be released "cannot point at a specific individual or be recovered after being processed"; the second is the principle of non-disclosure if unnecessary, which requires that the content should be limited to those related to disease transmission and public epidemic prevention. For example, the refusal of Japan's Ministry of Health, Labor and Welfare to release the nationality of patients reflects this principle. Government departments should minimize data collection, and design data protection techniques, such as the anti-attack, anti-leakage, and anti-theft techniques, or data flow techniques, such as the secure multiparty computation and homomorphic encryption, in all aspects of data flow and use, in an effort to balance the need for personal information protection and social public interests within the legal and regulatory framework.

Thirdly, the "isolated islands" of data should be broken to make data flow efficiently. At present, it is of great urgency for government departments to smash the data isolation in all links and make a more profound comprehensive analysis and calculation by integrating

multidimensional data in health, communications, transportation, public security, population, meteorology, etc.. Besides, when analyzing and predicting the epidemic situation, we should constantly adjust and optimize the models and algorithms according to the updated data to increase the accuracy of the prediction models and algorithms.

Ⅳ. The edge tool for city epidemic prevention and control---city big data platform

Looking back at the information-based measures for epidemic prevention and control, we have a lot of experience to sum up and many lessons to be learned. Although many of our cities have been continuously strengthening the construction of smart cities, this sudden public health emergency is a big test for a city's true "intelligence level".

At present, the city grass-roots management relies on assault and campaign-style law enforcement for a long run. Temporary measures rather than permanent ones are usually taken. In this regard, we should further push forward smart city construction, set up the city big data platform, and even promote the concept of digital twin cities.

In the initial stage of smart city construction, due to the immature technical capability, standard framework and top-level design, most cities focus on construction rather than application, and as a result the data are usually "collected but not connected and if connected but not used".

However, in the face of the grim epidemic prevention and control, the accurate and detailed data collection and real-time accurate information dissemination are of great importance. Therefore, building the city big data platform is evidently significant.

The city big data platform is a comprehensive city data processing center where information related to city operation is collected, stored in a centralized way and processed. It also provides city data application services[3].

The platform can strengthen cross-sector and cross-industry organization and coordination efforts, improve the level of information resources integration, accelerate the orderly convergence of city information resources in an all-round way, promote the in-depth sharing, correlated analysis and efficient utilization, provide cross-level, cross-regional, cross-sector and cross-business collaborative services for governments, enterprises and citizens, and display the "wisdom" of the city to the greatest extent.

Digital twin cities will realize highly efficient collaborative governance based on big data integration and analysis. Real-time dynamic data, such as basic city data, government service data, medical treatment and transportation, are added to geographic entities which are then uploaded into the city information model (CIM), realizing the integration of entity, model and data. With such a city command center that gathers the global data, city commanders can gain insight into the city's operation and laws from an overall perspective, and will not be "out at the elbows" because of the data in epidemic prevention and control. By extracting city characteristics, city managers can timely grasp a variety of public management data including epidemic prevention and control, and comprehensively analyze the epidemic development

[3] White Paper on City Big Data Platform, China Institute of Information and Communications, June 2019.

trends from various angles, so as to facilitate a targeted and accurate policy implementation[4]. This is the key direction for all local governments to improve their social governance capabilities including the capability in epidemic prevention and control in the future.

This epidemic storm is a big test of the governance capability for local governments at all levels. Big data technology plays a key role in it, but there is still room for improvement. Just as the old saying goes, fortune and misfortune are two buckets in the same well. If we can sum up experience and lessons and make full use of big data technology in realizing scientific government decision-making, accurate social governance and efficient public services, we will surely better guarantee the people's strong sense of fulfillment, happiness and security.

[4] *The Importance of Smart City Construction Considering the COVID-19 Outbreak*, Xiao Peng, February 2020.