Research of Big Data Analysis Platform Based on Positioning Information

Xiong Li and Shiyun Wang
Information Technology College, Shanghai Jianqiao University, 1111 Hucheng Ring Road, Nanhui new city, Pudong New Area, Shanghai, China
Email: 17032@gench.edu.cn; 15201715019@163.com

Abstract. With the wide spread of intelligent terminals, a variety of location-based business applications continue to emerge. We began to think about the analysis and application of location-based information big data, using big data technology to transform the existing industry, so as to better deal with epidemic disasters and other issues. For the novel Coronavirus, if artificial influenza virus infection or infection is detected by artificial means, it is easy to cause a series of problems, such as low efficiency, potential infection, and deliberately concealing travel history and contact history, which will have certain impact on epidemic prevention and control. Therefore, with the huge amount of data and the rapid growth of service demand, it is very urgent and necessary to isolate and confirm suspected cases based on big data analysis of location information. The big data analysis platform of location information can analyse the user’s location information, form the user’s trace, and provide more comprehensive information for epidemic control.

Keywords. Location information; LBS; CDRs data.

1. Introduction
A novel Coronavirus infected pneumonia in Shandong deliberately concealed travel history and contact history, resulting in 68 medical workers and 49 other personnel being quarantined. Coincidentally, a confirmed case in Changchun deliberately concealed work and life experience in the key prevention and control area, and repeatedly took the initiative to contact and eat with others, resulting in direct infection of 5 people, and many people were isolated for observation. They deliberately conceal their work and life experience in key prevention and control areas, and are afraid of being tested for virus infection. This behaviour leads to more people being infected. Their concealment make countless people’s efforts go to waste. In the case of insufficient manpower, intentional concealment will seriously affect the overall situation of epidemic prevention.

Therefore, in order to solve the above problems and realize the query of ID card number or mobile phone number to understand the historical trend of personnel and the 14-day historical track of personnel, we designed a big data analysis platform based on the positioning information. Fourteen days is the incubation period of the virus. By understanding the flow route of personnel in these fourteen days, we can confirm whether to isolate and observe them, and also ensure that there are more people in the severely affected areas.

The principle of querying ID card number or mobile phone number based on positioning information to understand the historical trend of personnel is to provide positioning data software through mobile phone, and the positioning data can measure more accurate position information.
through satellite or WIFI fixed access point. We can analyse the data of the flow of potential infected people, better arrange the isolation work and increase the number of staff.

2. Principle of Positioning

2.1. Location Based Services

Location-based service can obtain the location information (geographic coordinates, or geodetic coordinates) of mobile terminal users through the radio communication network (such as GSM network, CDMA network) or external positioning mode (such as GPS) of Telecom mobile operators, and offers corresponding services for users with the support of GIS platform.

It includes two meanings: first, to determine the location of mobile devices or users; second, to provide various location related information services. It refers to all kinds of positioning related service systems, referred to as "positioning services", another method is MPs mobile position services, also known as "mobile positioning services" system [1]. For example, find the current location of mobile phone users, and find the names and addresses of hotels, cinemas, libraries, gas stations, etc.

Generally speaking, LBS is composed of mobile communication and computer network, and the two networks interact with each other through gateway. The mobile terminal sends out the request through the mobile communication network and passes it to the LBS service platform through the gateway; the service platform processes according to the user’s request and the user’s current location. The result is returned to the user through the gateway, as shown in figure 1.

![Figure 1. LBS positioning principle.](image)

The mobile terminal can be mobile phone, personal digital assistant, PDA, Pocket PC, or desktop PC. The service platform mainly includes web server, location server and lightweight directory access protocol server.

At present, in all kinds of application software installed on our mobile devices, most of them need to obtain the geographic location information of users. As we know, LBS application system in IoT has been involved in various fields, such as transportation, medical treatment, travel, social networking, entertainment, etc. [2]. In the shopping application, LBS is used to get the user’s location, which not only saves the tedious process of the user’s manual input of location information, but also provides the basis for the selection of the distribution warehouse; in the navigation application, LBS obtains the user’s location information in real time and returns it to the user, making the acquisition and query of road information more intuitive and simple [3].
At present, geohash is a better algorithm for distance calculation of mobile geographic location. It is much more efficient than using longitude and latitude directly. Its principle is to convert a longitude and latitude information into a string code that can be sorted and compared.

But there are still disadvantages to this algorithm: when geohash algorithm is used to search for the address nearby, two points on both sides of the grid boundary are very close, but the encoding method is completely different in different grids, so an error occurs. In practical application, we have to search 8 grids around the current grid at the same time to solve this problem.

2.2. Global Positioning System

GPS is a global positioning system. The system determines the position of objects by satellite positioning. It has high accuracy in the process of practical application [4].

Positioning principle: at an altitude of 12000 kilometres from the ground, 24 GPS satellites orbit the earth in a 12-hour period, so that at any time, at any point on the ground, more than four satellites can be observed at the same time.

Because we know the precise data of the distance between the satellite and the Earth. In GPS observation, we can get the distance from the satellite to the receiver. Using the formula of 3D coordinate distance, the distance of three satellites can form three equations, and finally obtain the observation point (x, y, z). However, there is an error between the clock of the satellite and the clock of the receiver. In fact, there are four unknowns, x, y, Z and clock difference. Therefore, it is necessary to introduce a fourth satellite to solve the four equations, so as to obtain the specific position of the observation point, as shown in figure 2.

![Figure 2. GPS positioning principle.](image)

$$
\begin{align*}
[(x_1-x)^2+(y_1-y)^2+(z_1-z)^2]^{1/2}+c(x_1-x_0)-c_1 = 0 \\
[(x_2-x)^2+(y_2-y)^2+(z_2-z)^2]^{1/2}+c(x_2-x_0)-c_2 = 0 \\
[(x_3-x)^2+(y_3-y)^2+(z_3-z)^2]^{1/2}+c(x_3-x_0)-c_3 = 0 \\
[(x_4-x)^2+(y_4-y)^2+(z_4-z)^2]^{1/2}+c(x_4-x_0)-c_4 = 0
\end{align*}
$$

3. Source of Location Information Data

3.1. CDRs Data

Call detail records (CDRs) are data records generated by telephone exchange or other telecommunication equipment, which record the detailed information of telephone calls or other telecommunication transactions through the equipment (e.g. sending SMS, connecting to the Internet, etc.). CDRs mainly record various attributes of call or other telecommunication transactions, such as initiation time, duration, completion status, source number and destination number. Generally speaking, mobile network operators will keep CDRs for at least 3 months. Therefore, through CDRs, it is very useful for both pre tracking and continuous tracking. However, the disadvantage of this method
is that the representation of people, children and the elderly with low socio-economic status may be insufficient, and the geographical location resolution of remote rural areas is low.

3.2. Bat’s Application Software
Bat is the top three Internet companies in China, including Baidu, Alibaba and Tencent. The representative software that can provide location service has Baidu map under Baidu, Gao de map under Ali, WeChat and QQ under Tencent.

Baidu map, the location point is the AGPs location data generated when the “Baidu map” software is running. Baidu Map will store these location data in the software database, forming the mobile phone footprint recorded by the software itself. However, when Baidu map is not running, AGPs data will not be generated, which cannot fully reflect the trace of the whole working time of the mobile phone.

Gao de Map navigation uses a variety of methods for positioning, including IP address, GPS and other sensors that can provide relevant information (for example, it may provide information about adjacent devices and base stations for Gaud). When using the navigation service, Gaud map will collect the user’s vehicle information, address information and track.

WeChat can obtain the user’s location using the positioning system embedded in the mobile device [5]. When users agree to use some functions of WeChat, the software will collect some sensitive information of users. For example, when using the functions of people near WeChat, shaking and face-to-face group building, the software will record the geographic location information. When WeChat payment is used, the information on the payment bill includes time and place.

3.3. Other Applications Based on Mobile Location Services
Meituan obtains the location information of users through IP address, GPS and other sensors (such as nearby equipment, Wi-Fi access point and base station information) that can provide relevant information.

Metro Metropolis is the “Metro plus” Internet service launched by Shanghai Metro. There will be a ride record during the trip, which will show the station and time of entry and exit. When you use a service with location function, the app will use all kinds of technologies to collect and process information about the actual location of the user, for example, sensor data from user equipment can provide data from nearby base stations.

4. Main Obstacles of Big Data Analysis Platform

4.1. Sharing the above Data with a Third Party
Mobile phone number and ID card are bound to each other, and the historical trace of ID card searcher is realized through the positioning information in the software. First of all, these application software companies should be willing to share the above data with third parties (government agencies, research institutions, etc.). In today’s era of big data, data becomes an invisible resource. For an ordinary enterprise, the enterprise not only has valuable customer data, but also has supplier data, internal financial, design, manufacturing, management and other data, which involves their vital interests.

Secondly, although most of the application software in mobile phones is bound to the mobile phone number, the application software is a lightweight database that is encrypted and protected one by one. The storage methods and locations of location data in their respective database structures are different. It is extremely important to break through the encryption protection of database, analyse the data structure, and find the storage location of location data in the massive data It’s difficult. We need to crack a large number of databases with different structures in mobile phones, and collect and organize the location data of various application software together to form a complete historical trace.
4.2. Risks of Personal Information Protection
The most important thing is that there has risks of personal information protection in the shared data. Because we need to pass the 14-day historical track of the ID card number or mobile phone number, there is no anonymous data record, that is, the data will contain personal information such as ID card number.

5. Solution
In order to solve the above two problems, the proposal is as follows: government agencies take the lead in building a data sharing platform, different software companies update their positioning information on the platform in real time, do not need to break different databases and reorganize the positioning information, but actively send encrypted information to the platform, the platform integrates the historical positioning of the same user through algorithms. After that, he formed a complete historical trace. In addition, through big data analysis of the historical traces of the infected people, the platform can predict the next severe area of disease outbreak, which will better allocate limited medical staff.

As for the security maintenance of sensitive data, the measure is that only users at a specific level on a specific computer can query through ID card number or mobile phone number, so that the access to the barrier is generally set between cities or between districts. If the encrypted file is transferred outside the authorized scope of the unit through QQ, e-mail, mobile storage device and other means, it will display garbled code when it is opened and cannot be read or used normally. The data security is guaranteed through the encryption software. In addition, it is necessary to supervise, track and record all the operations of all the office staff in the whole process, and trace back the whole process of leakage in real time. Through real-time supervision and regulation of staff behaviour, to ensure information security.

6. Conclusion
In this paper, through positioning information to obtain the data of 14-day historical track of personnel, to help epidemic prevention and control. Data can work hard on “accuracy” and try to be right. Using technology to reduce the burden of prevention and control of the epidemic, using ID card or mobile phone number to query the historical trace, more secure isolation of suspected infected persons, and through the location of data analysis to determine the severely affected areas, so as to better arrange the manpower in case of shortage of manpower. For the sharing of historical positioning data, the government should strengthen the promotion and improve the data security. The design of big data analysis platform still needs to be optimized, and the concept is still not perfect. The following issues need to be further studied:

(1) On the technical level, different enterprises use different databases, such as Oracle, Sybase, mysql, etc. How to integrate different databases to form a unified standard is the next step to continue to improve.

(2) On the data level, the information provided by different companies needs to be sorted out according to the same user, and the duplicate data needs to be deleted, otherwise a large amount of storage space will be wasted. Therefore, we need an efficient algorithm to sort out a large number of messy data. As for the security of information, the information of historical trace in different software should be transmitted to the platform of historical trace query system through encryption algorithm, so in order to keep private information from leaking, the encryption algorithm should be complex enough and hard to be cracked.

References
[1] Xiaoguang Y, Qisong Z and Guojun S 2018 Research on classification of LBS service facilities based on rough sets neural network 2018 Chinese Control and Decision Conference.
[2] Sun G, Liao D, Li H, Yu H and Chang V 2016 L2p2: A location-label based approach for privacy preserving in lbs Future Generation Computer Systems 74 375-384.
[3] Li Z, Lv H and Liu Z 2019 Noise-added selection method for location-based service using differential privacy in internet of things Advances in Mechanical Engineering 11 (1).

[4] Xie G and Zhong C 2018 Research on mobile positioning method and accuracy Modern Information Technology 002 (007) 36-37.

[5] Wenqi S, Xiangyang L, Fan Z, Ziru P, Qingfeng C and Yong G 2018 Geolocating a wechat user based on the relation between reported and actual distance International Journal of Distributed Sensor Networks 14 (4) 155014771877446.