

Use of community mobile phone big location data to recognize unusual patterns close to a pipeline which may indicate unauthorized activities and possible risk of damage

Shao-Hua Dong¹ · He-Wei Zhang¹ · Lai-Bin Zhang¹ · Li-Jian Zhou² · Lei Guo²

Received: 4 October 2016 / Published online: 25 April 2017
© The Author(s) 2017. This article is an open access publication

Abstract Damage caused by people and organizations unconnected with the pipeline management is a major risk faced by pipelines, and its consequences can have a huge impact. However, the present measures to monitor this have major problems such as time delays, overlooking threats, and false alarms. To overcome the disadvantages of these methods, analysis of big location data from mobile phone systems was applied to prevent third-party damage to pipelines, and a third-party damage prevention system was developed for pipelines including encryption mobile phone data, data preprocessing, and extraction of characteristic patterns. By applying this to natural gas pipelines, a large amount of location data was collected for data feature recognition and model analysis. Third-party illegal construction and occupation activities were discovered in a timely manner. This is important for preventing third-party damage to pipelines.

Keywords Pipeline · Big location data · Third-party damage · Model · Prevention

1 Introduction

The risk caused by third-party damage is an important issue during the entire life of pipelines. During 2001–2015, 30%–40% of pipeline accidents in China were caused by third-party damage. According to European accident statistics, 52% of pipeline accidents in Europe were due to third-party external damage during 1984–1992 (Dong 2015); 40.4% in the USA and Europe according to the PHMAS latest statistics. Accidents caused by third-party construction accounted for ~20% in 1993–2010. More than 702 leakage accidents occurred during 2010–2016, and 177 of those accidents were caused by third-party damage (external force or excavation by third party), accounting for 25.21%.

Typical third-party accidents in China had a great impact and caused huge economic losses. Several accidents have been reported: On October 6, 2004, because of mechanical failure, pipeline leakage occurred during third-party construction on the Shaanxi–Beijing pipelines in Shenmu Town, Yulin City, Shaanxi Province. On December 30, 2009, the Lan-Zheng-Chang oil products pipeline leaked because of third-party construction, leading to diesel fuel being spilt into the Weihe River. On May 2, 2010, third-party construction caused pipeline rupture on No. 223 pile of the East-Huang oil pipeline in Jiulong Town, Jiaozhou City, leading to leakage of 240 tonnes of crude oil. On July 28, 2010, the propylene pipeline in the Qixia District in Nanjing City exploded because of third-party construction failure. More than 13 people were killed, 28 people were seriously injured, and more than 100 people were slightly injured. On June 30, 2014, because of third-party unauthorized excavation, a leakage accident occurred on 14# + 700 m of Xingang–Songgang pipe of Xin-Da pipeline, and the oil spilt into

✉ Shao-Hua Dong
shdong@cup.edu.cn

¹ The Pipeline Technology Research Center, China University of Petroleum (Beijing), Beijing 102249, China

² PetroChina R&D Center, Langfang 065000, Hebei, China

Edited by Yan-Hua Sun

the municipal sewer network. On September 16, 2015, a medium-pressure gas PE pipeline leaked due to the construction in Xujiawan, Gansu Province, near the Lanyaqinn River.

At present, pipeline patrol is the main measure for monitoring third-party activities and preventing damage; however, because these activities are hidden and random, the patrol monitoring is not effective, especially for third-party mining on pipelines. Illegal activities such as oil and gas stealing are often carried out during the rest time of line patrol officers. For fiber optic early warning and third-party intrusion detection technologies with a high false alarm rate, a large number of databases should be built. This is because cable vibration caused by mining action on site is used to determine third-party activities. However, many similar activities take place, and it is difficult to accurately determine damage. At the same time, some places have different cable and pipeline trenches, thus limiting the applicability of the technology.

Big location data (BLD) have been widely utilized. BLD have become an important resource to observe human community activity and analyze geographical conditions. By analyzing the BLD of oil and gas transport vehicles, human social attribute and relationship with the environment can be extended from a simple positioning data, and a type of intelligent and social application is formed (Daggitt et al. 2016; Doornik and Hendry 2015; Duan et al. 2014; Ettinger-Dietzel et al. 2016; Hashem et al. 2016; Narayanan and Cherukuri 2016; Teli et al. 2016; Tsou 2015).

IBM used mobile phone signals and a signal tower to locate the specific personnel, thus timely accessing the information as to whether the specific personnel came to the region, and established models to perform complex analyses. Then, some information related to the specific personnel was obtained, including the mobile phone behavior of people together with their location, to determine future behavior and help to analyze their movement (Hashem et al. 2016).

Inspired by the above analysis, big location data were used to help prevent third-party damage in this study and to solve the problems in the current third-party damage identification such as real-time deficiencies and small monitoring scope. By establishing the location relationship between a specific cell phone signal and signal towers along the pipeline and obtaining the mobile phone GPS location information, the data of mobile phone signals were analyzed, and third-party damage behavior was evaluated. An area of about 10 km on a pipeline suffering from a higher third-party risk was selected for monitoring using the BLD to uninterruptedly determine the existing excavation and construction activities. A big data association

model of mobile phone signal position was developed to provide timely alarms.

2 Extraction of big location data

Big data are a combination of large complex datasets. The scale and complexity of these datasets exceeds the capabilities of current database management software and traditional data processing technology in acquisition, management, retrieval, analysis, mining, and visualization (Liu 2012).

2.1 Features of BLD

An important part of the big data is BLD. The location data are a combination of geographical data and human social information data containing the space position and time identification. Here, the space position can be accurate geographical coordinates and also can be a conventional place or position (Guo et al. 2013, 2014).

The features of BLD are as follows:

- (1) BLD are multiple, heterogeneous, and rapidly changing with typical characteristics such as a large volume, rapid update speed, diversity, and low density.
- (2) The common characteristic of BLD is space–time identification; this can be described by absolute location, coordinate, relative position, and language. In addition, the space–time identification of the location data should be accurate and reliable. Accuracy, reliability, and credibility are required in processing and analyzing the location data.
- (3) This has a feature called “complex but sparse”. Because of the constraint in data acquisition technology, BLD may not reflect the overall picture of the object.

Analysis of BLD means extraction of clues from the local research object and establishment of several characteristic patterns based on a single area r_i or moving object o_i . The extraction methods for a feature model can be divided into two categories as follows:

- (1) First-order characteristics: this refers to characteristics that can be easily calculated from the location records, map data, or historical track of moving objects in the region, such as the mean and variance.
- (2) Second-order characteristics: this refers to characteristics where the hybridity of original observation data can be eliminated to a certain extent. These features are processed by higher-order statistics.

2.2 Extraction features of mobility pattern in a bar area

Mobility pattern (MP) φ_{mp} : take one or two (peer) moving objects o as the observation target, and the aspects over a period of time include the mobility uniqueness feature, randomness and periodic features, metastatic nature, static and dynamic intermittence, and expectations of movement (Pan et al. 2013; Quinlan 1993a, b).

(1) Uniqueness feature, f_{uniq}

The mobility uniqueness feature can be used to distinguish moving objects and defined as the probability of a track tra_i that can be determined according to the number of given regions $\|F\|$, average size of a region $\overline{F_{size}}$, and interval of statistical time $\overline{F_{time}}$:

$$P_F\{|tra_i| \leq 2|F_{size}, F_{time}|, \|F\|\} \tag{1}$$

When $\overline{F_{size}}$ and $\overline{F_{time}}$ are relatively appropriate, the activities of the bar area are considered. For example, the probability to determine a unique path is very high in an area with a length of 200 m and width of 50 m on both sides of the pipeline ($\overline{F_{size}} = 0.02 \text{ km}^2$, $\overline{F_{time}} = 0.5 \text{ h}$), and it is only about 8 regions ($\|F\| = 8$) (De Montjoye et al. 2013) When $\|F\|$ is fixed, similar power-law relationships of probability with $\overline{F_{size}}$ and $\overline{F_{time}}$ are established.

$$f_{uniq} = \alpha - (\overline{F_{size}})^\beta \tag{2}$$

$$f_{uniq} = \alpha - (\overline{F_{time}})^\beta$$

β is a power exponent and linear with $\|F\|$:

$$\beta = \lambda_1 - \lambda_2\|F\| \tag{3}$$

By observing a small number of regions with abnormal activities surrounding the pipeline, third-party damage by the relevant personnel or tracks of third-party construction users can only be determined. This shows that individual mobility has a high degree of regularity and also shows that the mobility behavior significantly differs among different populations.

(2) Periodic features, f_{peri}

For a moving object, o_i , a discrete Fourier transform was conducted for the binarization of its access region's sequence F_j (1 means visiting, 0 means not visiting). By observing the frequency of the largest Fourier transform coefficient, the cycle of position TP_j^i can be obtained (Liu et al. 2010)

It is supposed that a group of regions $A = \{F_1, F_2, \dots, F_{\|F\|}\}$ with the same access period $TP = \{T_1, T_2, \dots, T_Q\}$ is divided into Q time slots. Thus, the detailed probability distribution matrix of each individual mobility $P = [P_1, P_2, \dots, P_j]$ can be obtained.

Among them, $P_j = [P_r(F_1|T = T_j), P_r(F_2), \dots, P_r(F_{\|F\|})]$ represents the column probability vector. The location record of the T time period in BLD is generated into $[T/TP] = m$ probability distribution matrix $\{P_1, P_2, \dots, P_m\}$ according to the cycle of TP . Then, the periodic behavior of moving objects can be analyzed by calculating their Kullback–Leibler (KL) divergence (Yuan et al. 2013).

The more precise standard location entropy can be obtained:

$$H(P) = - \sum_{t_j=1}^Q \sum_A P_r(F_i|T = T_j) \log_2 P_r(F_i|T = T_j) \tag{4}$$

Then, the entropy of relative distribution is:

$$KL(P_1||P_2) = \sum_{t_j=1}^Q \sum_A P_{rP_1}(F_j) \log_2 \frac{P_{rP_1}(F_j)}{P_{rP_2}(F_j)} \tag{5}$$

According to the order of relative entropy, hierarchical cluster, the probability distribution of n continuous or discontinuous location $\{P_1, P_2, \dots, P_n\}$, several clusters frequently matching with each other and having the same period (possibly maximum) could be obtained. This represented several typical periodic motion patterns of moving objects o_i (Song et al. 2010). During the clustering, the position probability distribution for associating two clusters C_i and C_j can be calculated as follows:

$$P^{New} = \frac{|C_i|}{|C_i| + |C_j|} P_i + \frac{|C_j|}{|C_i| + |C_j|} P_j \tag{6}$$

3 Privacy protection for location data

Location information is generally formed by the identification and location information. Identification information is used to describe the user-specific attributes and characteristics that can be uniquely identified by the user. Location information represents a current specific location or track within a certain time of the user.

The privacy protection measures are as follows: When users submitted a service request to the server, accurate location information was provided by the mobile client, and the user's real identity was hidden at the same time. This method can provide high-quality location service to the user according to the location information (Wang 2015). The relationship is shown in Fig. 1.

4 Techniques used in the BLD detection of third parties along the pipeline

- (1) Acquisition technology of third-party intrusion signal and GPS signal data

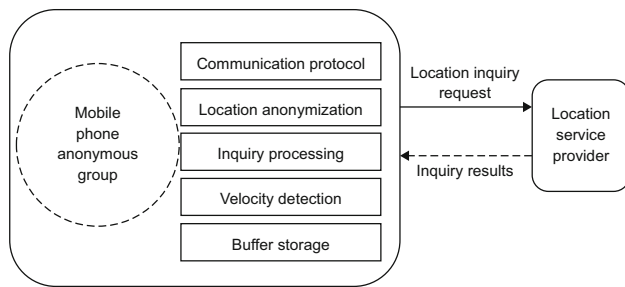


Fig. 1 Location privacy protection

The mobile data and GPS signals of third-party personnel activities along the pipeline were continually collected for 24 h. The signals were used to establish the location relationship between specific cell phone signals and signal towers along the pipeline and to obtain information related to mobile phone GPS location and cell phone towers. The data collected from the mobile equipment (including unique device ID, latitude, longitude, and time stamp) were stored in a database or loaded into the Hadoop platform.

(2) Storage technology for BLD

A computational framework model such as Hadoop, efficient space–time index and distributed analysis technology for flow media, map data, and track data were established. Because BLD are nonrelational, database storage technologies were used, such as Hbase, Big SQL, and Mango.

(3) Preprocessing technology of third-party mobile data

The filtering, integrity, reduction, and discretization methods for third-party communication mobile data were established as the pretreatment. Then, data mining, machine learning, and other methods were used for further processing and mining of location data to analyze the correlation.

By the pretreatment of map and location trace data, the plane map for continuous space was discretized and divided into several regions based on the BLD of map or road network data. The main methods include grid division, division according to road network, division according to position density, and division according to reference sites (Thiessen polygon) (Ester et al. 1996; Li et al. 2013; Pan et al. 2013; Liu et al. 2010; Yuan et al. 2012; Zheng et al. 2013; Zhu et al. 2013), as shown in Figs. 2, 3 and 4. In the analysis of BLD, especially the track data, the dataset should have a high sampling rate to make a simple linear interpolation in the track data. ST-matching, IVMM, Passby, and other algorithms and methods were used to relate the track data and map data (Lou et al. 2009; Liu et al. 2012; Tang et al. 2012; Yuan et al. 2010).

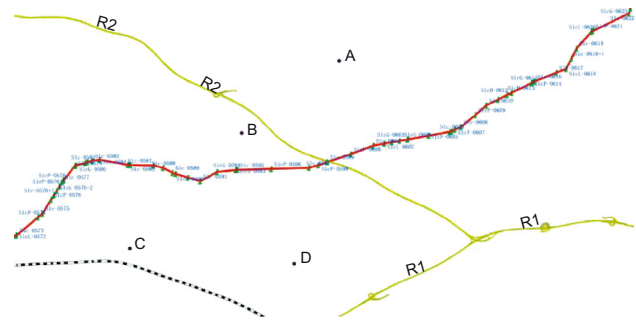


Fig. 2 Location distribution: road, traffic, and village network diagram near the long-distance transportation pipeline

(4) Technology for feature extraction of third-party damage risk.

The feature model between the mobile phone locations and risk of third-party damage was established according to the time feature, which was used to extract the valuable information and following three types of features: (a) Regional static characteristics. Taking a certain area as the observation object, the indexes related to the map were extracted, including the road network characteristics and change rate of concerned points. (b) Mechanical characteristics of regional position movement. The behavior of the moving group targets in the area such as the time evolution of the regional traffic mobility was extracted. (c) Movement patterns characteristics of individuals/groups in different periods. Taking the moving individual/group as the observation object, the mobile behavior characteristics of individual/group within a period of time were extracted. The second-order statistical characteristics and their application to the service calculation of the specific location were studied (Duan et al. 2014). By establishing the model, the signs of risk of third-party damage and destruction were identified.

With the acquisition of BLD, the data quantity gradually increased, and the pattern recognition methods were constantly updated. Logistic regression, support vector machine (SVM), random and uncertain analysis model, wavelet transform, and neural network model were used to analyze the BLD. Combining the behavior of third-party personnel with pipeline risk characteristics, the precision of the forecast warning model was improved.

(5) Visualization methods for third-party damage risk based on BLD.

A statistical chart was used to show the results or data trends in data processing. Based on the characteristics of large scale and diversity, visualization methods were developed to accurately simulate the development state and motional tendency of third-party intrusion along the pipeline.

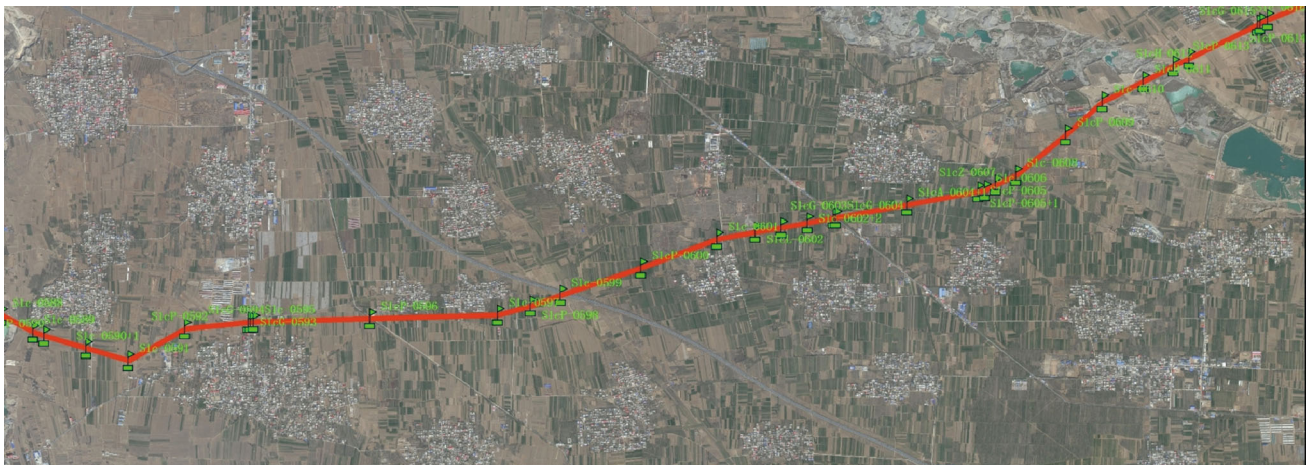


Fig. 3 Personnel activities

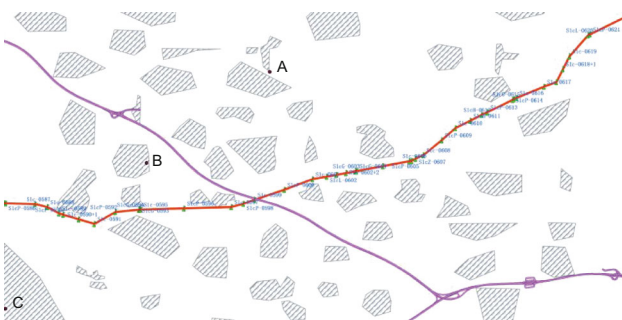


Fig. 4 Discrete reference point map along the pipeline

- (6) Forecast warning system for third-party damage to pipeline.

Through the abovementioned research, a third-party forecasting and early warning system for pipeline were established, including data acquisition, data storage, data analysis and modeling, data risk visualization, and trend analysis.

5 Case study

5.1 Application steps

The length of the pipeline in this case is 9.8 km. By accessing the mobile phone signals, important results were obtained in the modeling of third-party damage prevention. Specific steps for mobile phone BLD analysis are as follows:

- (1) Data acquisition

This is the first step. Wireless service providers are responsible for collecting location information. A mobile phone provides services using a group of mobile phone

signal towers. Its specific location can be obtained by triangulation to the distance from the nearby towers, and the position accuracy is less than 20 m. Most smart phones can even provide more accurate GPS location information (the accuracy is about 20 m). Location data including latitude and longitude require 26 bytes if all this information are stored. If you are dealing with 2 million users and store their position information per minute, the size is about 0.1 TB per day.

In this case, the particular personnel can be three types of people: pipeline managers who have periodic and frequent activities on pipeline base, station, and line; planned construction personnel along the pipeline section, who report to the management. Their activities are clear to managers, illegal excavation, construction and sabotage persons are the focus of the monitoring data analysis.

In practical engineering applications, mobile signals within a distance of ± 50 m from the mobile tower to pipeline have been accessed from mobile companies. Mobile companies encrypt the data, changing mobile signals into specific codes. The movement of these codes is under analysis, not involving personal privacy and security.

- (2) Big data storage and processing

Because of the nonrelational BLD, database storage technologies such as Hbase, Big SQL, Mango, and others were used to establish Hadoop analysis (Fig. 5).

- (3) Dimension reduction analysis

For the dimension reduction treatment of a BLD network in a space scale, the core is to reduce the nodes (region) or edge (regional association) of the network and obtain global features by analyzing the key components. The main methods are dimensionality according to super betweenness and dimension reduction according to principal components. For the time scale, the dimension is

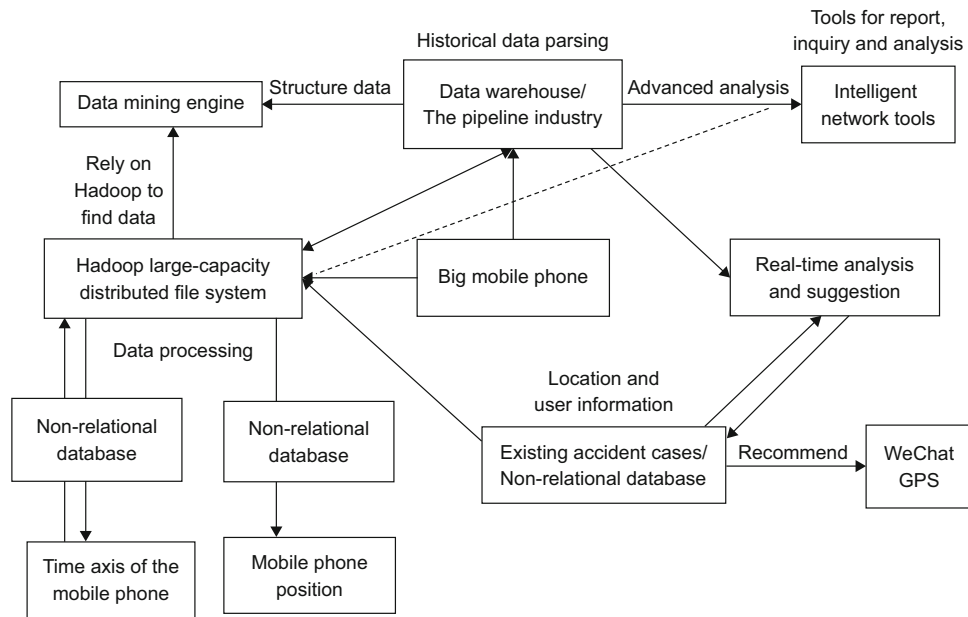


Fig. 5 Hadoop distributed storage hardware integration for big location data

mainly about time discretization, which reduces the similarity between different time periods.

According to the time dimension (determined by the maximum frequency of the occurrence of third-party damage to pipeline), the time periods were shortened to 20:00–22:00, 12:00–14:00, and 2:00–4:00 with a higher risk. For space dimension reduction, the location data in the range of 30 m around the pipeline showed the range of activity.

(4) Feature extraction and modeling of local location data

For the hybrid of BLD, extraction of the static data of mobile phone users should take a certain region as the object of observation and obtain some indicators related to landforms and maps of the area including the road network features, change rate of points, and other static characteristics. Based on the technology for extracting the mobility pattern features in a bar area, the trajectory of the relevant personnel of third-party damage risk or construction can only be determined through the feature probability extraction of individual location and two or more co-locations.

The model for feature probability extraction $H(P)$ is:

$$H_1(P) = - \sum_{t_j=1}^Q \sum_A P_r(F_i|T = T_j) \log_2 P_r(F_i|T = T_j) \quad (7)$$

$$H_2(P) = - \sum_{t_n=1}^Q \sum_A P_r(F_m|T = T_n) \log_2 P_r(F_m|T = T_n) \quad (8)$$

$$H(P) = H_1(P) \cap H_2(P) \quad (9)$$

where Q is equal to 3 (Time periods are 20:00–22:00, 12:00–14:00, and 2:00–4:00); A is the strip area for 9.8 km and 20 m within the scope of the pipeline; $H_1(P)$ is the location probability for individual 1; $H_2(P)$ is the location probability for individual 2; $H(P)$ is the intersection degree of the location probability for the two people in the same area. Generally, warning is needed when it is greater than 90%.

In this case, the model of a third-party damage critical region was developed according to the analysis of accident statistics. The accident statistics show that 85% of the third-party accidents have the same features: more than two people, more than two times, and each static time for 0.5 h. All these elements appeared in the same region.

(5) Data analysis

The mobile phone data were tested for 30 days, and 253,708 bar location data were collected. Then, all the data were screened as follows: in accordance with two or more people (not limited to the same person), at least arriving at the same place twice (with two) above, and each static time more than 0.5 h. After the screening and statistical analysis, the final statistical data were 232, as shown in Table 1.

The statistical analysis in Fig. 6 shows two high risk points of abnormal personnel situation during 22:00–24:00 and 2:00–4:00, and they are the highest risk. The level of personnel risk appearing at the wasteland, hills, and gullies is medium. The level of personnel risk appearing at the fields, railways, water conservancy project, and sites is low.

Table 1 Statistics of mobile phone location data

Time	Location										Statistics
	Gully	Field	Wasteland	Hill	Railway	Highway	Site	Water conservancy project	River	Forest	
6:00–8:00	1	13	0	0	0	1	0	0	0	1	16
8:00–10:00	2	25	0	0	1	2	3	2	3	2	40
10:00–12:00	0	37	1	0	1	2	4	3	2	0	50
12:00–14:00	1	11	2	2	1	1	0	1	1	1	21
14:00–16:00	0	13	1	0	0	3	3	2	2	1	25
16:00–18:00	0	25	1	0	2	5	4	3	2	2	44
18:00–20:00	0	21	0	0	1	2	2	3	3	1	33
22:00–24:00	0	0	0	1	1	0	0	0	0	0	2
24:00–2:00	0	0	0	0	0	0	0	0	0	0	0
2:00–4:00	0	0	1	0	0	0	0	0	0	0	1
4:00–6:00	0	0	0	0	0	0	0	0	0	0	0
Statistics	4	145	6	3	7	16	16	14	13	8	

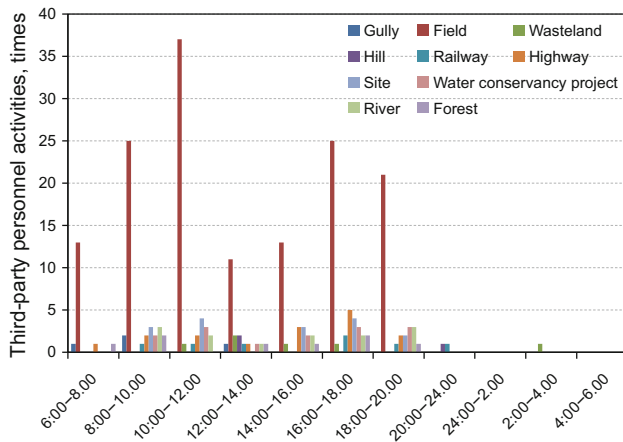


Fig. 6 Diagram of third-party personnel activities and time

After analysis, most people working in the fields around the pipeline, about 145, belong to normal production. The gully data were verified as returning farmland to forest plant operation; however, the abnormal data at 2:00–4:00 were verified as illegal construction for green houses near the pipeline and confirmed as not reporting to the pipeline protection department. An illegal earth borrowing occurred on the hill at 22:00–24:00, and the railway construction near the pipeline belongs to emergency inspection at night.

12:00–14:00 is lunch time, attributing 21 times to the model: 11 of them are involved in field farming; one of them on gully land is involved in forest operation. The wasteland, railway, highway, water conservancy, rivers, and woodland account for five times in total and belong to normal operation; however, the construction lacking normal monitoring on hills and wasteland work along the pipeline account for four times.

The data analysis shows one illegal construction on hills around the pipeline, one construction of a greenhouse at the edges of the wasteland, and other situations belong to normal work (fishing by the river). By analyzing the BLD, the cross-projects along the pipeline would be understood, and abnormal situations would be rapidly detected and monitored.

5.2 Brief summary

(1) Comparison for technologies

By studying these technologies, the following characteristics are given in Table 2.

Through comparison, several limitations were observed in the existing third-party prevention technologies. For example, the monitoring range of optical fiber early warning is small, and the prediction function is not present. The warning occurs after the occurrence of mining behavior. The big data have the features of forecast warning and protection. By collecting and analyzing the real-time data within 50 m of the pipeline, maintenance personnel can reach the scene to prevent third-party construction damage.

(2) Scientific problems to be solved

By analyzing big data, the early warning problem of the risk of third-party damage for bar area pipeline facilities was solved. With the established intersection degree model of location probability, the characteristics of the risk of third-party damage to pipelines can be accurately defined. Furthermore, the technology can also be extended to third-

Table 2 Comparison of prevention methods for preventing third-party damage to pipelines

Indicators	Optical fiber vibration	Remote sensing recognition	Actual patrol	BLD early warning
Monitoring distance	30 km	A wide range	5 km	100 km
Real-time performance	Real time	Not real time	Interval	Real time
False alarm rate	High	Middle	None	Low
Accuracy range	5 m	0.61 m	500 m	50 m
Features	(1) Vibration signal is complex (2) False alarm rate is higher (3) Investment costs a lot (4) Characteristics of third-party activities are not obvious	(1) Data acquisition time of remote sensing image recognition is long (2)The analysis is difficult (3) Not applicable for third-party monitoring	Patrol is intermittent	(1) A full-time monitoring (2) It can provide warning of the third-party construction work, damage, and destruction along the pipeline (3) The disadvantage is that a lot of data is needed, and the model should be constantly improved

party monitoring for railways, highways, and electricity networks.

6 Conclusions

- (1) For the first time, BLD technology was used to reduce the risk of third-party damage to pipelines. A set of BLD acquisition technologies was established, including encryption technology, data preprocessing technology, third-party damage pattern feature extraction technology, and third-party damage risk visualization methods. A prediction and warning system was developed for third-party damage to pipelines based on BLD.
- (2) The case study shows that illegal third-party construction around the pipeline can be rapidly found using this technique. Early detection of risks and automatic classification of the system can help to control the third-party risk to pipelines.
- (3) Through time and regional dimension reduction to reduce the nodes in the mobile data network, the periods with high third-party risk can be extracted, thus effectively solving the discretization problem of third-party location data.
- (4) The developed method in this study has overcome the deficiency of other methods, such as the uncertainty and false alarm rate of optical fiber vibration and remote sensing image analysis. By analyzing the

data, a three-dimensional network of enterprise defense can be gradually established.

- (5) The method can be used in pipeline safety management and increase the strength of research and application.

Acknowledgements This work was supported by Pipeline Management Data Analysis and Typical Model Research [Grant Number 2016B-3105-0501] and CNPC (China National Petroleum Corporation) project, Research on Oil and Gas Pipeline Safety and Reliability Operating [Grant Number 2015-B025-0628].

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Daggitt ML, Noulas A, Shaw B, et al. Tracking urban activity growth globally with big location data. *R Soc Open Sci.* 2016;3(4):150688. doi:10.1098/rsos.150688.
- De Montjoye YA, Hidalgo CA, Verleysen M, et al. Unique in the crowd: the privacy bounds of human mobility. *Sci Rep.* 2013;3:1376. doi:10.1038/srep01376.
- Dong SH. Pipeline integrity management technology and practice. Beijing: Sinopec Press; 2015. p. 2–15 (in Chinese).
- Doornik JA, Hendry DF. Statistical model selection with “big data”. *Cogent Econ Finance.* 2015;3(1):1045216. doi:10.1080/23322039.2015.1045216.

- Duan R, Hong O, Ma G. Semi-supervised learning in inferring mobile device locations. *Qual Reliab Eng Int.* 2014;30(6):857–66. doi:[10.1002/qre.1701](https://doi.org/10.1002/qre.1701).
- Ester M, Kriegel HP, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the 2nd international conference on knowledge discovery and data mining (KDD-96)*. 1996, August. 96(34): p. 226–31.
- Ettinger-Dietzel SA, Dodd HR, Westhoff JT, et al. Movement and habitat selection patterns of smallmouth bass *Micropterus dolomieu* in an Ozark river. *J Freshw Ecol.* 2016;31(1):61–75. doi:[10.1080/02705060.2015.1025867](https://doi.org/10.1080/02705060.2015.1025867).
- Guo C, Fang Y, Liu JN, et al. Study on social awareness computation methods for location-based service. *J Comput Res Dev.* 2013;50(12):2531–42 (in Chinese).
- Guo C, Fang Y, Liu JN, et al. Analysis and processing of large data processing research. *J Wuhan Univ (Information Science Edition)*. 2014;39(4):379–85. doi:[10.13203/j.whugis20140210](https://doi.org/10.13203/j.whugis20140210) (in Chinese).
- Hashem IAT, Chang V, Anuar NB, et al. The role of big data in smart city. *Int J Inf Manage.* 2016;36(5):748–58. doi:[10.1016/j.ijinfomgt.2016.05.002](https://doi.org/10.1016/j.ijinfomgt.2016.05.002).
- Li Z, Ding B, Han J, et al. Mining periodic behaviors for moving objects. In: *The 16th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM. 2013, April. p. 1099–108. doi:[10.1145/1835804.1835942](https://doi.org/10.1145/1835804.1835942).
- Liu JN. The recent progress on high precision applications of Beidou navigation satellite system. Report of the stanford's 2012 PNT challenges and opportunities symp. (SCPNT 2012), 2012. (in Chinese).
- Liu K, Li Y, He F, et al. Effective map-matching on the most simplified road network. In: *The 20th international conference on advances in geographic information systems*. ACM. 2012, November. p. 609–12. doi:[10.1145/2424321.2424429](https://doi.org/10.1145/2424321.2424429).
- Liu S, Liu Y, Ni LM, et al. Towards mobility-based clustering. In: *The 16th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM. 2010, July. p. 919–28. doi:[10.1145/1835804.1835920](https://doi.org/10.1145/1835804.1835920).
- Lou Y, Zhang C, Zheng Y, et al. Map-matching for low-sampling-rate GPS trajectories. In *The 17th ACM SIGSPATIAL international conference on advances in geographic information systems*. ACM. 2009, November. p. 352–61. doi:[10.1145/1653771.1653820](https://doi.org/10.1145/1653771.1653820).
- Narayanan M, Cherukuri AK. A study and analysis of recommendation systems for location-based social network (LBSN) with big data. *IIMB Manag Rev.* 2016;28(1):25–30. doi:[10.1016/j.iimb.2016.01.001](https://doi.org/10.1016/j.iimb.2016.01.001).
- Pan G, Qi G, Wu Z, et al. Land-use classification using taxi GPS traces. *IEEE Trans Intell Transp Syst.* 2013;14(1):113–23. doi:[10.1109/TITS.2012.2209201](https://doi.org/10.1109/TITS.2012.2209201).
- Quinlan JR. Combining instance-based and model-based learning. In: *The tenth international conference on machine learning*. 1993a. p. 236–43. doi:[10.1016/B978-1-55860-307-3.50037-X](https://doi.org/10.1016/B978-1-55860-307-3.50037-X).
- Quinlan JR. C4. 5: programs for machine learning. San Mateo: Morgan Kaufmann Publishers; 1993b.
- Song C, Qu Z, Blumm N, et al. Limits of predictability in human mobility. *Science.* 2010;327(5968):1018–21. doi:[10.1126/science.1177170](https://doi.org/10.1126/science.1177170).
- Tang Y, Zhu AD, Xiao X. An efficient algorithm for mapping vehicle trajectories onto road networks. In: *The 20th international conference on advances in geographic information systems*. ACM. 2012, November. p. 601–4. doi:[10.1145/2424321.2424427](https://doi.org/10.1145/2424321.2424427).
- Teli P, Thomas MV, Chandrasekaran K. Big data migration between data centers in online cloud environment. *Proced Technol.* 2016;24:1558–65. doi:[10.1016/j.protcy.2016.05.135](https://doi.org/10.1016/j.protcy.2016.05.135).
- Tsou MH. Research challenges and opportunities in mapping social media and big data. *Cartogr Geogr Inf Sci.* 2015;42(sup1):70–4. doi:[10.1080/15230406.2015.1059251](https://doi.org/10.1080/15230406.2015.1059251).
- Wang XY. Analysis and processing method of location service data and privacy protection. *J Jixi Univ (Integrated Edition)*. 2015;15(7):51–3 (in Chinese).
- Yuan J, Zheng Y, Xie X. Discovering regions of different functions in a city using human mobility and POIs. In: *The 18th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM. 2012, August. p. 186–94. doi:[10.1145/2339530.2339561](https://doi.org/10.1145/2339530.2339561).
- Yuan J, Zheng Y, Xie X, et al. T-drive: enhancing driving directions with taxi drivers' intelligence. *IEEE Trans Knowl Data Eng.* 2013;25(1):220–32. doi:[10.1145/1869790.1869807](https://doi.org/10.1145/1869790.1869807).
- Yuan J, Zheng Y, Zhang C, et al. An interactive-voting based map matching algorithm. In: *The 2010 eleventh international conference on mobile data management*. IEEE Computer Society. 2010, May. p. 43–52. doi:[10.1109/MDM.2010.14](https://doi.org/10.1109/MDM.2010.14).
- Zheng Y, Liu F, and Hsieh HP. U-Air: when urban air quality inference meets big data. In: *The 19th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM. 2013, August. 1436–44. doi:[10.1145/2487575.2488188](https://doi.org/10.1145/2487575.2488188).
- Zhu B, Huang Q, Guibas L, et al. Urban population migration pattern mining based on taxi trajectories. In: *3rd international workshop on mobile sensing: the future, brought to you by big sensor data*, Philadelphia, USA. 2013, April.