Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
1. Introduction

In recent years, air pollution has attracted increasing attention globally as it significantly impacted global health (Gheorghe & Ion, 2011; Yuan, Zhang, Wang, Wang, & Zuo, 2017). Worldwide air pollution-related mortality was estimated at around 6.5 million in 2012 (Influenza-like illness concept, 2014). In Taiwan, urbanization and the rapid development of various industries have made air pollution an critical concern. Obtaining real-time air quality information is of great importance for air pollution prevention. Thus, it is necessary to perform air pollution analysis and prediction to determine changing trends in air pollution and to propose prevention strategies. However, since the collected air pollution data is not only dynamically evolving and vast, a lot of computation resources are needed. In addition, it is difficult to analyze the data effectively and reliably in stand-alone computation environments. Therefore, an integrated platform of air pollution monitoring was constructed by using Spark (Spark yarn, 2016; Geng & Zhang, 2017; Yang, Chan, Liu, & Lou, 2017; Yang, Chen, Den, Wang, & Kristiani, 2018), R (Liu et al., 2018) and ELK Stack (LangiWidyawan et al., 2015, pp. 181–186; Prakash, Kakkar, & Patel, 2016) based on Hadoop framework to provide a real-time and easy-to-use big data analysis and visualization environment. In this work, the real-time data of air pollution and influenza-like illness were collected from the open data websites provided by the environmental protection administration and the Taiwan CDC open data portal, respectively. In addition, HDFS and Alluxio software were employed to build the cloud storage framework. The Ceph object storage environment was also developed as the backup storage framework. Through the dashboard created by ELK Stack software, users could quickly realize the association analyses between air pollution and influenza-like illness in Taiwan in a visualized way.

The main contributions of this study are summarized as follows:

- We successfully confirmed the associations between air pollution and influenza-like illness in Taiwan by building an integrated platform based on Hadoop Distributed File System (HDFS), Alluxio,
Spark, Ceph and ELK Stack software to provide a cluster, visualization, and backup storage environment. The development of such systems enables clinician, policy maker, and administration staff to better predict and prepare for a possible influenza outbreak.

- We demonstrated the lag-time associations between air pollution and influenza-like illness, especially three weeks lag-time. This work makes the current understanding of air pollution-related influenza outbreak in a real-time visualization way more easily, which is of great importance for multidisciplinary cooperation and global engage this severe problem.

The remainder of this paper was organized as follows. In Section 2, we reviewed the background literature and related works. In Section 3, we presented the system architecture and methods. The experimental results were shown in Section 4. In Section 5, we discuss concerning the associations between air pollution and influenza-like illness was presented. Finally, concluding remarks were given in Section 6.

2. Background review and related works

2.1. Hadoop distributed file system (HDFS)

HDFS is a decentralized storage system which provides reliable, scalable and faults tolerant data storage (Jena, Gourisaria, Rautaray, & Pandey, 2017). It works with MapReduce computing framework based on the requests and queries (Matsuzaki, 2017; Park, Min, & Shim, 2017; Pulgar-Rubio, Rivera-Rivas, Prez-Godoy, Gonzalez, Carmona, & del Jesus, 2017; Sha, Liang, Jiang, Chen, & Zhuge, 2017, pp. 585–592). HDFS can accept data in several formats, like text, images, videos ... etc. The important benefit of HDFS is fault tolerance which means that when nodes on a network fail, services can be provided continuously. The general architecture is presented in Fig. 1.

In the HDFS architecture, NameNode serves as the master while DataNode serves as the slaves, which are introduced as follows:

- NameNode: NameNode is mainly responsible for storing and managing the metadata of files. This metadata is cached in main memory to provide faster access to read/write requests. Files are broken into blocks which are then transferred and stored in DataNodes. In addition, the NameNode also responsible for management of the DataNodes' input and output tasks.
- DataNode: DataNode is a daemon which runs on the slave node in Hadoop Cluster. It is responsible for storing and managing the data blocks on the slave nodes. In addition, the DataNode also responsible for performing the data read/write operations to disks. Data blocks stored in DataNodes are replicated and distributed across clusters of storage devices to provide reliability and high availability of accessing data.
- Secondary NameNode: Secondary NameNode is mainly responsible for periodically reading the file system, logging the changes and updating the file management system.

2.2. Ceph object storage

Ceph is object storage based on free software storage platform that stores data on a single distributed computer cluster and provides interfaces for an object, block, and file-level storage (Yang et al., 2017b; Zhan & Piao, 2016, pp. 719–725). Ceph storage clusters are designed to run on commodity hardware, using an algorithm called CRUSH (Controlled Replication Under Scalable Hashing) to ensure data is evenly distributed across the cluster and that all cluster nodes can retrieve data quickly without any centralized bottlenecks.

The Ceph storage system is designed to be self-healing, self-managing and strives to reduce the operational cost. The storage system uniquely delivers object, block, and file storage in one unified system. Ceph object storage is accessible like Amazon S3 (Persico, Montieri, & Pescape, 2016, pp. 113–118; Shetty & Naik, 2017, pp. 507–512). Open Stack Swift provides a native API for integration with software applications. Ceph block storage saves the data on Ceph Block Device, which is a virtual disk that can be installed on Linux-based servers or virtual machines. A Ceph Storage Cluster is shown in Fig. 2.

The Ceph object storage consists of three service components, which are the Reliable, Autonomic Distributed Object Store (RADOS) Gateway, the Block Device and the CephFS. First, the RADOS Gateway is a bucket-based on Representational State Transfer (REST) gateway, which is compatible with S3 and Swift. Second, the RBD is a reliable and fully-distributed block device, with a Linux kernel client and QEMU/KVM driver. Finally, the CephFS is a Portable Operating System Interface (POSIX)-compliant distributed file system.

2.3. Alluxio system

Alluxio (formerly known as Tachyon) is an open source, in-memory virtual distributed storage system. It unifies data access and bridges...
In Big Data ecosystem, Alluxio mainly lies between in computation frameworks or jobs, such as Apache Spark, Apache Map Reduce, or Apache Flink, and various kinds of storage systems, for examples Amazon S3, Google Cloud Storage, Open Stack Swift, GlusterFS, HDFS, and Ceph. With Alluxio, the performance of big-data ecosystem has been significantly improved. In addition, Alluxio enables new workloads across different storage systems. Various frameworks can share data efficiently with each other. Alluxio’s unified namespace allows applications to interact with data on any storage system.

2.4. Influenza-like illness

Based on the definitions of the World Health Organization (WHO) global influenza surveillance standards case (Influenza-like illness concept, 2014) for ILI and severe acute respiratory infections (SARI) as below:

- Influenza infection causes a clinical syndrome not easily distinguished from other respiratory infections.
- The case definitions for ILI and SARI are not necessarily intended to capture all cases but to describe trends over time.
- Using one common case definition globally will allow national health authorities to interpret their data in an international context.

**Influenza-like illness case definition**

Acute respiratory infection with:

1. Measured fever of 38 C and Cough
2. The occurrence within the last ten days

One of the countries that greatly effects Influenza-Like Illness is Taiwan. Its data trend for ILI is shown in Fig. 3.

Based on the data, almost 10–15% population is divided into six big areas infected by Influenza-LikeIllness Symptoms.

2.5. Related work

Hadoop and Spark are the most popular open source large-scale data analysis frameworks iteratively, which mean that reading the same data with multiple times and sharing intermediate results across worker nodes. In the Spark environment, one of the weaknesses is that data transformation and distribution is implicitly managed by Hadoop Distributed File System (HDFS). Data locality does not guarantee for iterative machine learning algorithms which read the same data multiple times. The resulting data shuffling becomes a bottleneck when iteratively reading such RDDs. It needs to write to HDFS first and continue to the next iteration. YinMem et al. in (Huang, Yesha, Halem, Yesha, & Yinmem, 2016) proposed a parallel distributed indexed in-memory computation system, bridged the gap between Hadoop ecosystem and HPC by replacing MapReduce with MPI while obtaining the advantages of the distributed data storage.

Ceph distributed storage system is a demanding concentrated multi-purpose with saving by object storage daemon. In (Iop science, 2016), authors’ motivations are to extend the STAR online compute farm from a simple job processing tool for data taking into a multi-purpose resource equipped with a large storage system would lead any dedicated resources to become an extremely efficient and an attractive multi-purpose facility. The Ceph storage system is used to achieve this goal on a 10 GB backbone network for eliminating the network as a limitation. With further acquisition of large fast drive (1 TB SSDs), they showed how they can customize the data placement options. Ceph has to offer such as primary affinity, mounting OSD journals on SSDs and Cache Tiering along with no-Ceph related local disk caching techniques.

Apache Spark has advantages over Hadoop MapReduce for analysis of real-time data using time-series analysis. While Hadoop MapReduce is a widely used and famous execution engine for working with the storage and analysis of large datasets. In MapReduce, the data is read from the disk and the computing results are written to the HDFS after a particular iteration. Then, the data is read from the HDFS for the next iteration. The whole process consumes much time and a large amount of space, which incur the issues of high latency and fault tolerance to the entire system. Maheshwar & Haritha (2016), pp. 721–725 survey on the high performance of Apache Spark to overcome the issues and disadvantages of MapReduce.

Air pollution in Beijing, especially PM2.5, has received increasing attention in the past years. Although exposure to PM2.5 has been linked to many health issues, few studies have quantified the impact of PM2.5 on the risk of ILI. This study aims to investigate the association between daily PM2.5 and ILI risk in Beijing, by utilizing a generalized additive model. Feng, Li, Sun, Zhang, & Wang (2016) use daily PM2.5, meteorological factors, and ILI counts from January 1st, 2008 to December 31st, 2014 were retrieved. An inverse Gaussian generalized additive model with log link function was used to flexibly model the nonlinear relationship between the PM2.5 (single- and multiday lagged exposure) and ILI risk, adjusted for the weather conditions, seasonal and yearly trends.

During influenza season, acute respiratory illness due to influenza is difficult to distinguish from other influenza-like illnesses, but testing should be reserved for situations when timely results will influence management or infection control measures. Humiston, MPH, & Pham (2016) state immunization status and timing of disease onset notwithstanding, a neuraminidase inhibitor should be offered immediately for particular high-risk children. Neuraminidase inhibitor treatment should be considered if shorter illness duration is warranted or an at-risk sibling may be protected. Antipyretics and cough control may be useful. Immunization with an age-appropriate dose of an inactivated influenza vaccine is the cornerstone of prevention for healthcare personnel and our patients.

Vietnam studies provide opportunities to understand ILI transmission, but data from (sub)tropical developing countries are scarce. Nguyen et al. (2016) determine the viral etiology and epidemiology of ILI in households. ILI was detected by active case finding amongst a cohort of 263 northern Vietnam households between 2008 and 2013. Health workers collected nose and throat swabs for virus detection by multiplex real-time RT-PCR. ILI was detected at least once in 219 (23.7%) of 945 household members. 271 (62.3%) of 435 nose/throat swabs were positive for at least one of the 15 viruses tested. Six viruses predominated amongst positive swabs: Rhinovirus (28%), Influenza virus (17%), Coronavirus (8%), Enterovirus (5%), Respiratory syncytial virus (3%), Metapneumovirus virus (2.5%) and Parainfluenza virus 3 (1.8%). There was no clear seasonality, but 78% of episodes occurred in Winter/Spring for Influenza compared to 32% for Rhinovirus.
Participants, on average, suffered 0.49 ILI, and 0.29 virus-positive ILI episodes, with no significant effects of gender, age, or household size. In contrast to the US and Australian community studies, the frequency of ILI decreased as the number of household members aged below five years increased (p = 0.006). The findings indicate the need for tailored ILI control strategies, and for a better understanding of how local childcare practices and seasonality may influence transmission and the role of children.

3. System architecture

Air Quality Index (AQI) is a standard of air condition value in Taiwan. The system architecture of this work is presented in Fig. 4. The architecture divided into two parts, which are the data storage and the data processing, respectively. For data storage, we used MySQL as the main gate of RDBMS, Hadoop HDFS, and Alluxio as a temporary cloud data storage for processing and Ceph is utilized for backup storage.

A data processing architecture is shown in Fig. 5. The detailed statements are described as follows:

1. The air quality data are downloaded from the open data websites in Taiwan automatically.
2. The air quality data are combined with the healthcare data from one of Taiwan’s hospitals.
3. The data resources are stored in MySQL RDBMS.
4. Sqoop plugin over JDBC will transfer data from RDBMS to HDFS. Sqoop allows to import the full table, selected columns and even allows the flexibility of specifying free form queries to extract data and to write into HDFS.
5. The data will be sent to Alluxio as In-memory temporary saving.
6. Through Spark processing, the data will proceed, and the output will be sent back to HDFS and Ceph storage as backup devices.
7. Through the Sqoop plugin, the data will then be dumped into MySQL to virtualize on ELK Stack and Web.

For data processing, the data is automatically sent from HDFS to Alluxio (In-Memory level), and the Spark software will still read data from Alluxio for further steps. The general integrated framework of Alluxio and Spark is presented in Fig. 6.

To prevent data corruption or data loss in the system, we use Ceph software to construct the backup storage system. Its architecture is shown in Fig. 7. In addition, for providing a real-time air quality index information and population statistics of every city more clearly, the concentration of pollutants and their corresponding air quality index values are presented in Table 1. Fig. 8 shows the flowchart of data processing. Moreover, Fig. 9 shows the flowchart of data backup processing.

3.1. Provided services

This cloud system provides a real-time air quality index information and population statistics each city with various kinds of the graph on the dashboard. This data can be exported to the excel file for further analysis. In addition, more detailed data of AQI are itemized as follows:
4. Experimental results

4.1. Setup of the experimental environment

In this section, the experimental results are presented to illustrate the association analyses between air pollution and influenza-like illness. First, concerning the setup of the hardware environment, 11 servers are deployed to serve as the master server, the ELK stack server, the web server, building for Spark and Ceph environments. Each server has the same specifications concerning CPU, Memory, Hard disk and Operating System. In addition, for performing the analysis more efficiently, we construct a cluster environment by using the Cloudera Manager which was served as a cluster server. The Cloudera Manager consists of packages of related services, such as Hadoop, HDFS, MapReduce, Spark, HBase, Hive and so on.

Second, in the step of data collection, there are mainly two types of data, which are the air pollution data and the influenza-like illness data, respectively. Air pollution provides hourly, daily and weekly data, but influenza-like illness offers weekly data only. Therefore, we need to summarize data into weekly data to let users analyze efficiently. Air pollution data is also not completed. Not all monitoring stations have 24 h of complete data every day. These happen because of sometimes some of the monitoring stations are being not run well.

4.2. The data visualization platform

The collected data on air quality and influenza-like illness from 2016 to 2017 are shown in Fig. 10. The weekly mean value is calculated and used to compare with these two data.

Fig. 11 shows the visualization of the weekly data on air quality and influenza-like illness collected in 2016. The advantage of visualization offers the researchers to find the associations easily. The visualization of the weekly data on air quality and influenza-like illness collected in 2017 are shown in Fig. 12.

Fig. 13 shows the visualization of higher and lower values of air quality and influenza-like illness from 2016 to 2017, to avoid the effect of extremities.

We also delay ILI time for one until four weeks than the air quality index to see the associations as shown in Fig. 14. Vertical charts are influenza-like illness, and line charts are air quality index. After repeated trials of 1–4 weeks, we found that the apparent associations in 3 weeks delay ($p < 0.005$). There is a lag-time association between air quality index and influenza-like illness.
5. Discussion

There are many influenza cases in a defined region are associated with environmental factors, such as air quality index trend, which is compatible with the previous study in Beijing (Feng et al., 2016). Moreover, PM$_{2.5}$ has a good practical association with the incidence of ILL and incidence of ILL is higher in the winter season and southern Taiwan, which is associated with PM$_{2.5}$. Our work provides a real-time gathering of data set and a social event. The results demonstrate an association of multi-factor air pollution and influenza cases, which is the combined effect of several mechanisms. First, high air pollution has been proved to harm both innate and cell-mediated immunity. It may take a period to develop symptoms and signs of influenza after the insults. Second, air pollution is associated with the abundance of micro-organisms. The size of influenza virus about 0.08–0.12 μm is smaller than the PM$_{2.5}$ so that airborne pollution particles provide condensation nuclei to which virus droplets attach. The effect of incubation period could play a role in the lag-time effect. In the incubation period is 5–7 days in the cases of influenza, so the lag-effect cannot be explained by organisms inoculation alone. Population susceptibility and

| Air Poll. | $O_3$, 8 h | $O_3$ | PM$_{2.5}$ | PM$_{10}$ | CO | SO$_2$ | NO$_2$
|-----------|------------|--------|------------|----------|----|-------|-----|
| Unit      | Ppm        | Ppm    | g/m$^3$    | g/m$^3$  | Ppm| Ppb   | Ppb |
| 050       | 0.00000.054| –      | 0.0 15.4   | 054      | 04.4| 035   | 053 |
| 51100     | 0.0550.070 | –      | 15.535.4   | 55125    | 4.5 | 9.512 | 76  |
| 101150    | 0.0710.085 | 0.1250 | 16.164.4   | 255      | 9.5 | 76-185| 101 |
| 151200    | 0.086-0.105| 0.165  | 0.204     | 54.150.4 |15.5| 12.515.4| 66 |
| 201300    | 0.0100.320 | 0.205  | 0.404     | 35524    |15.5| 15.530.4| 65 |
| 301400(2) | 0.0450.504 | 0.425  | 0.250     |1254     |15.5| 12.515.4| 65 |
| 401500(2) | 0.5050.604 | 0.350  | 0.5500.4  | 1850     |15.5| 12.515.4| 65 |

Fig. 7. The Ceph object storage architecture.

Fig. 8. The flowchart of Data processing.

Table 1

The concentration of pollutants and air quality index values.

C.-T. Yang et al. Computers in Human Behavior 100 (2019) 266–274
immunogenically modulated by genetic factor and environment stimuli would be possible underlying mechanisms in the phenomena. Our study provides a frontier work for the further study of population genetic, host-environment-microorganisms interactions, and epigenetics.

6. Conclusion

In this work, we propose a complete big data ecosystem environment that consists of database relational to collect air quality and Influenza-Like Illness data. For data analysis, based on our trend charts air pollution has an impact on Influenza-Like Illness disease, actually for PM$_{2.5}$ and O$_3$ particulate matters. Some of the other parameters like humidity, temperature, PM$_{10}$, NO$_2$ etc. also have an impact to this disease. We only research a significant component that influences ILI. On the infrastructure side, we choose the best performance ways for the data operation. First, we use Indexing for the relational database. Second, data transfer between MySQL to HDFS, we optimize the Sqoop tool with generating multi-file and with direction parameters. Third, for
data processing, Hadoop Distributed File System (HDFS) as a temporary data storage and through Alluxio in-Memory environment feeds data to Spark is faster than using HDFS only. For the performance comparison, we can conclude the speed performance averagely grows to 40% than the previous condition.

6.1. Future works

In this work, the cycle of data processing is smoothly running well based on end-user requirement. All data include resource and output data which are stored in Ceph Environment for the backup level. As time passed by the data growth up, the backup level needs more space.
and better speed up to store data securable, we hope Ceph environment could be further developed till using SSD level to mount Object Storage Daemon (OSD), and the experiment can be extended to cluster SSD level.

For the application, besides the web page can give us the information about Taiwan’s air pollution, it also can provide notification via email or SMS to end-user on a real day. These messages are notified for an emergency level of air pollution.

Conflicts of interest

Chao-Tung Yang declares that he has no conict of interest. Cai-Jin Chen declares that he has no conict of interest. Yu-Tse Tsan declares that he has no conict of interest. Po-Yu Liu declares that he has no conict of interest. Wei-Chen Chan declares that he has no conict of interest. Yu-Wei Chan declares that he has no conict of interest. Po-Yu Liu declares that he has no conict of interest.

Acknowledgments

This work was funded by the Ministry of Science and Technology (MOST), Taiwan, under grant number MOST 106-3114-M-029-001, and MOST 107-2221-E-029-008. In addition, this work was also funded in part by the Taichung Veterans General Hospital (TCVGH), Taiwan, under grant number TCVGH-T1077803.

References

Feng, C., Li, J., Sun, W., Zhang, Y., & Wang, Q. (2016). Impact of ambient fine particulate matter (pm2.5) exposure on the risk of influenza-like illness: A time-series analysis in Beijing, China. Environmental Health: A Global Access Science Source, 15(1).

Geng, Y., & Zhang, J. (2017). Optimized parallelization of binary classification algorithms based on spark. Vol. 2, 83–86.

Gheorghe, I. P., & Ion, B. (2011). The effects of air pollutants on vegetation and the role of vegetation in reducing atmospheric pollution. In M. K. Khallaf (Vol. Ed.), The impact of air pollution on health, economy, environment and agricultural sources: Vol. 0(8) Rijeka: InTech chapter 12.

Huang, Y., Yesha, Y., Halem, M., Yesha, Y., & Yinnem, S. Zhou. (2016). A distributed parallel indexed in-memory computation system for large scale data analytics. 2016 IEEE international conference on big data (big data) (pp. 214–222).

Humiston, F. Sharon G., MPH, M. D., & Pham, T. N. (2016). Influenza-like illness diagnosis and management in the acute care setting. Pediatric Emer Care, 32, 883–884 2016.

Influenza-like illness concept: http://www.who.int/influenza/surveillance_monitoring/. Top science http://oapsciences.org/article/10.1086/1742-6596/762/1/012025.

Jena, B., Gourisaria, M., Rautary, S., & Pandey, M. (2017). A survey work on optimization techniques utilizing map reduce framework in hadoop cluster. International Journal of Intelligent Systems and Applications, 9(4), 61–68.

Langi, P., Widyawaran, Najib, W., & Aji, T. (2015). An evaluation of twitter river and logstash performances as elasticsearch inputs for social media analysis of twitter. Liu, P.-Y., Tsan, Y.-T., Chan, Y.-W., Chan, W.-C., Shi, Z.-Y., Yang, C.-T., et al. (2018). Associations of pm2.5 and aspergillosis: Ambient fine particulate air pollution and population-based big data linkage analyses. Journal of Ambient Intelligence and Humanized Computing, 1–11.

Maheshwari, R., & Haritha, D. (2016). Survey on high performance analytics of bigdata with Apache spark.

Matsuzaki, K. (2017). Functional models of hadoop mapreduce with application to scan. International Journal of Parallel Programming, 45(2), 362–381.

Nguyen, D. N. T., Mai, L. Q., Bryant, J. E., Hang, N. L. K., Hoa, L. N. M., Nadjm, B., et al. (2016). Epidemiology and etiology of influenza-like illness in households in vietnam; its not all about the kids.. Journal of Clinical Virology, 82(0), 126–132.

Park, Y., Min, J.-K., & Shim, K. (2017). Efficient processing of skyline queries using mapreduce. IEEE Transactions on Knowledge and Data Engineering, 29(5), 1031–1044.

Persico, V., Montieri, A., & Pescape, A. (2016). On the network performance of amazon s3 cloud storage service.

Prakash, T., Kakkar, M., & Patel, K. (2016). Geo-identification of web users through logs using elk stack. 606–610.

Pulgar-Rubio, F., Rivera-Rivas, A., Prez-Godoy, M., Gonzalez, P., Carmona, C., & del Jesus, M. (2017). Mefasd-bd: Multi-objective evolutionary fuzzy algorithm for subgroup discovery in big data environments - a mapreduce solution. Knowledge-based Systems, 117, 70–78.

Shia, E.-M., Liang, Y., Jiang, W., Chen, X., & Zhuge, Q. (2017). Optimizing data placement of mapreduce on ceph-based framework under load-balancing constraint.

Shetty, M., & Naik, S. (2017). Ontology representation of amazon s3 objects for metadata enrichment.

Spark yarn: http://blog.csdn.net/BlockheadLS/article/details/53329323.

Yang, C.-T., Chen, C.-J., & Chen, T.-Y. (2017b). Implementation of ceph storage with big data for performance comparison. Lecture Notes in Electrical Engineering, 424, 625–635.

Yang, C.-T., Chen, S.-T., Den, W., Wang, Y.-T., & Kristiani, E. (2018). Implementation of an intelligent indoor environmental monitoring and management system in cloud. Future Generation Computer Systems. https://doi.org/10.1016/j.future.2018.02.041.

Yuan, X., Zhang, M., Wang, Q., Wang, Y., & Zuo, J. (2017). Evolution analysis of en terainment data for performance comparison.