Journal of Global Change Data & Discovery. 2020, 4(1): 16-23
DOI:10.3974/geodp.2020.01.03
www.geodoi.ac.cn

©2020 GCdataPR
Global Change Research Data Publishing & Repository

Knowledge Service System on Disaster Risk Reduction and Its Application in Social Media Analysis

Wang, J. L.[1,2*] Han, X. H.[1,2] Bu, K.[3] Zhang, M.[1,2] Wang, X. J.[4,1]
Yuan, Y. L.[1]

1. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;
2. College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China;
3. Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China;
4. School of Civil and Architectural Engineering, Shandong University of Technology, Zibo 255049, China

Abstract: Driven by the disaster risk reduction mission of the United Nations Educational, Scientific and Cultural Organization (UNESCO), Disaster Risk Reduction Knowledge Service System (DRRKS) was founded under the International Knowledge Center for Engineering Sciences and Technology (IKCEST) under the auspices of UNESCO. The system constructs disaster metadata standards and gathers and produces disaster risk reduction data products. Supported by geographic information technology, online knowledge applications for disaster risk reduction have been realized and opened to global users. DRRKS has established 16 online knowledge applications, to mine, analyze, and visualize disaster information based on big data, such as remote sensing and social media resources. There was an outbreak of pneumonia associated with the 2019 novel coronavirus (COVID-19) in Wuhan, Hubei province in China at the end of 2019, for which DRRKS constructed a knowledge application for public opinion analysis of disaster events quickly. The application acquired spatial information and topic semantic information from Sina Weibo texts. Based on the Latent Dirichlet Allocation model and machine learning algorithm, the spatial-temporal distribution of Weibo texts related to the novel coronavirus outbreak and public opinion were analyzed to provide information and application support for the prevention and control of the novel coronavirus pandemic.

Keywords: disaster risk reduction; knowledge service; social media; public opinion analysis; COVID-19

1 Introduction

Disaster risk reduction is an urgent global issue. The United Nations Educational, Scientific, and Cultural Organization (UNESCO) has long valued global cooperation in this field and has set up the Earth Sciences and Geohazards Risk Reduction Section in its Department of Natural Sciences. The International Knowledge Center for Engineering Science and Technology (IKCEST) was established by UNESCO in China in 2013, and operated by the Chi-

Received: 15-02-2020; **Accepted:** 28-02-2020; **Published:** 25-03-2020
Fundations: China Engineering Science and Technology Knowledge Center Construction Project (CKCEST-2020-2-4); China-Pakistan Joint Research Center on Earth Sciences (Y99R0900F2); Chinese Academy of Sciences (XXH13505-07).
***Corresponding Author:** Wang, J. L. R-8881-2016, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, wangjl@igsnrr.ac.cn
Article Citation: Wang, J. L., Han, X. H., Bu, K., *et al*. Knowledge service system on disaster risk reduction and its application in social media analysis [J]. *Journal of Global Change Data & Discovery*, 2020, 4(1): 16–23. DOI: 10.3974/geodp.2020.01.03.

nese Academy of Engineering. In combination with its positioning and mission, UNESCO sought close cooperation with IKCEST in disaster prevention and mitigation in 2015. Driven by UNESCO"s mission for disaster risk reduction, IKCEST launched the construction of a Disaster Risk Reduction Knowledge Service System (DRRKS) in 2016. It was sponsored by the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences.

2 Disaster Risk Reduction Knowledge Service System

2.1 Objectives and Vision

To meet the global demand for disaster prevention and mitigation, DRRKS focused on international/national metadata standards or best practices and established a global meta-database for disasters by consolidating data from different kinds of disasters under the unified standard system, as follows. DRRKS integrates disaster data and information at the national/regional scale and establishes a disaster database, based on the main types of disasters in China and its surrounding areas and typical regions of the world. It establishes an online platform with the support of big data mining and analysis techniques. Additionally, it mines database development methods and knowledge information service modes in disaster prevention, rescue, reconstruction, evaluation, and other aspects, and extensively implements special services, education and communication, international training, and cooperation in disaster prevention and mitigation. It also plays a fundamental role in UNESCO"s disaster prevention and mitigation program[1].

DRRKS vision is to provide a platform, technology, data, education, knowledge, and other services for the current global disaster risk reduction, including accumulating scientific, technological, and academic resources (such as disaster database, product database, and knowledge base); connecting domestic and international resources for disaster risk reduction; gathering international typical cases and application demonstrations for disaster prevention and mitigation; supporting the application of the Belt and Road Initiative in regional disaster risk reduction; becoming an important foundation and fulcrum for international cooperation in disaster risk reduction for UNESCO, and enhancing significantly the international influence of IKCEST.

2.2 System Architecture

With priority given to the use of open international technical standards and open-source web technologies, the DRRKS adopts the technical architecture of an information service platform with on-demand scalability and a modular mechanism. The iterative development model is adopted, so the system can be put into use in its developmental life cycle. Through this system, users can gain quick access to various disaster-related knowledge resources and subject-specific knowledge services, including data, maps, literature, experts, institutions, and videos. The overall architecture of DRRKS is shown in Figure 1.

The data storage scheme in the bottom layer adopts the Alibaba Cloud computing model to construct the file server, meta-database server, data server, map server, and web server, which are used to analyze user access. Supported by a series of open web technologies, a variety of editing, operation, and maintenance functions are realized, including data entry, information publishing, permission management. Functions like cartographic visualization, full-text literature retrieval, analysis of user behavior, and tag filtering for multiple disaster subjects can also be performed. This supports knowledge application functions related to the distribution of disaster organizations and institutions, disaster map browsing, and subject

application for disaster prevention and mitigation.

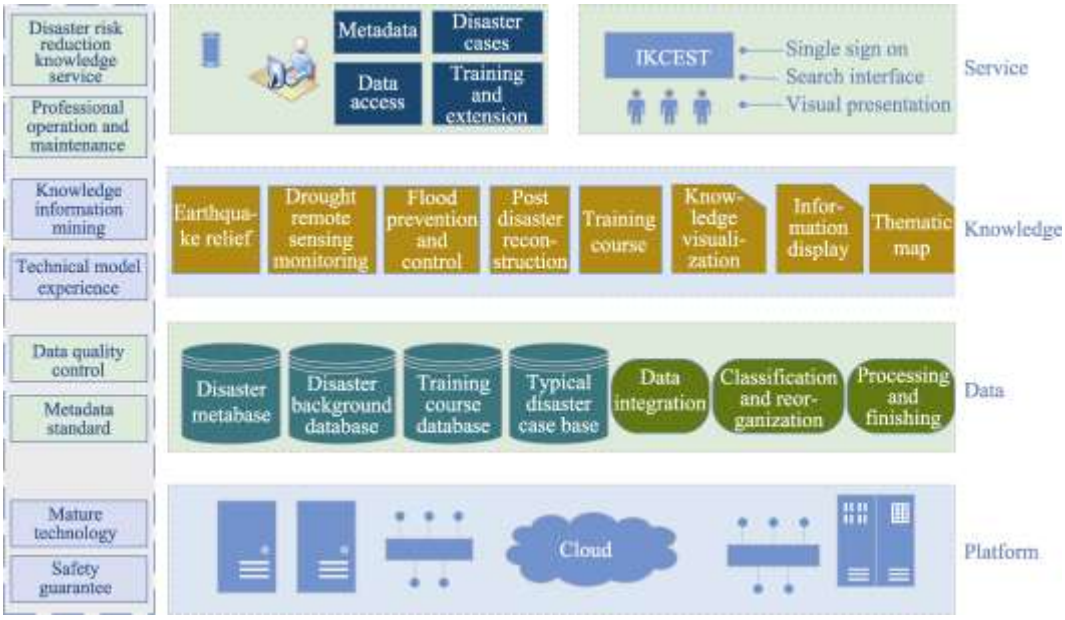

Figure 1 Platform architecture of the Disaster Risk Reduction Knowledge Service System

The DRRKS is developed based on the Browser/Server structure and the application framework of Python + Tornado + TorCMS. For front-end development, HTML5 and CSS3 technologies are utilized, combined with the jQuery and Bootstrap 3 framework. The backstage programming language is Python 3.4 or above. The PostgreSQL database is used to persist the data. The core attributes are mapped into the database field, and the extended attributes are stored in the JSONB fields of PostgreSQL. The backstage of WebGIS, which is applied to the visualization of disaster maps and spatial data, uses MapServer. On the front-end, the Leaflet and OpenLayers 3 JavaScript databases are adopted. The software, pycsw, is applied to metadata management. It is an OGC CSW server implementation written in Python. This CSW standard defines a set of unified interfaces for the retrieval, query, and browsing of spatial information and related data[2].

2.3 Product System

(1) Metadata standards and technical specifications. The disaster expert database, institution database, video courseware, and other resources are extended based on disaster core metadata standards research. Meanwhile, the technical specifications for disaster data management and open services are formulated.

(2) Global disaster metabase. Based on the global platform and professional online database related to disasters, the web crawler technology is used to obtain metadata information on disasters such as earthquakes, droughts, floods, typhoons, forest fires, high-temperature heat waves, etc. Natural language processing, information extraction, and other technologies are used to complete the word segmentation, filtering, keyword extraction, and other processing of disaster metadata information. Semantic tag extraction and disaster metadata classification are completed by combining the controlled vocabulary.

(3) China disaster map database. The disaster map database is composed of collected dis-

aster maps in China, which were reorganized, scanned, and processed by geographic information technology. Each map was re-numbered to ensure its uniqueness.

(4) Thematic disaster database. This database was established by remote sensing earth observation and historical statistical data mining, among others. These include obtaining land degradation data along the railway between China and Mongolia by object-oriented remote sensing interpretation, obtaining monthly historical meteorological regional disaster data from meteorological station data, obtaining and integrating the disaster datasets of typical megalopolis"s including Beijing, Shanghai, and Chongqing, etc, since 1949.

(5) The Belt and Road disaster database. This database integrates the background data of 65 countries and regions along with the Belt and Road project, including basic national conditions, natural resources, and social economy. Based on remote sensing data and network resources, the loss degree data from high-temperature heat waves, floods, and earthquakes along the China-Pakistan Economic Corridor are collected, mined, and compiled. The risk distribution dataset for rainstorms and flooding and the extreme precipitation event dataset in the adjacent areas of China and Russia are obtained using remote sensing and ground monitoring.

(6) The thematic database supporting SDGs. Addressing the problems of forest management, land degradation, and other major environmental and ecological resources in SDG15, the dataset products for SDG15.1 and SDG15.3 were obtained by using remote sensing and big data technology. For example, the database included regional forest classification data products in China, desertification data products in Mongolia, and land salinization and degradation data in the Yellow River Delta in China, etc.

(7) Construction of general resources. Regarding relevant data norms and standards, norms and standards, and disaster databases (expert, institution, literature, and video courseware), global engineering cases and other content construction will be carried out continuously and will be open to users through information sharing and knowledge application tools.

2.4 User Service

The DRRKS team carries out user services continuously. Main users are divided into five categories: 1) UNESCO and other international organizations or institutions; 2) relevant government agencies and management technicians for disaster prevention and mitigation; 3) scientific and technological researchers engaged in disaster prevention and mitigation; 4) teachers and students of higher education institutions; 5) the public. The number of user visits reached 13,000 per month, nearly 50% of which came from countries outside China, mainly the United States, Japan, India, the Philippines, the United Kingdom, and other countries.

3 Online Knowledge Application

Online knowledge application is a typical application mode provided by the DRRKS. Attracted by specific application needs, it provides user interaction and displays through data integration processing and visualization technology support. Currently, 16 online knowledge applications have been developed and deployed. These can be found and accessed on the homepage of DRRKS (http://drr.ikcest.org), shown in Figure 2 and Table 1.

4 Application of COVID-19 Public Opinion Analysis

4.1 Data Pre-processing

(1) Data collection. Sina-Weibo (http://us.Weibo.com), often referred to as Weibo, is one of the most popular social media platforms in China with over 516 million active users each month in 2019. Using Weibo Application Programming Interfaces (APIs), Weibo messages related to COVID-19 were collected with "pneumonia" and "coronavirus" as the keywords in this study. The following information was extracted finally: user ID, timestamp (i.e., message posted time), text, and location information.

(2) Data pre-processing. The original Weibo texts contain interfering information such as hyperlinks, spaces, punctuation marks, hashtags, and @users. Text filtering was thus necessary to eliminate noise and improve the efficiency of word segmentation. These types of interfering information were removed by regular expression operations ("re" module) in Python. Very short Weibo texts (less than four words) and duplicates were deleted. Word segmentation was necessary because there are no obvious separators between Chinese words. A Python package for Chinese text segmentation called "Jieba" was utilized. By building a user dictionary including keywords related to COVID-19, the package segmented words efficiently.

Figure 2 Picture of the homepage of DRRKS system

Table 1 DRRKS system online knowledge application list

No.	Name of knowledge application	Online address	Service function
1	Knowledge map service for major disaster risk reduction organizations	http://drr.ikcest.org/app/s8349	Obtain global disaster risk reduction organization list and provide online visualization and one-stop navigation services
2	Global earthquake daily distribution map service	http://drr.ikcest.org/app/s9834	Through the real-time USGS interface, global earthquake distribution data can be obtained and displayed visually online
3	Map visualization services of China's historical disasters	http://drr.ikcest.org/app/s7834	Obtain historical maps, publish them visually after scanning and correction, and provide editing function
4	Chinese and international experience in natural disaster relief	http://drr.ikcest.org/case/index.html	Collect global typical cases and show them in terms of pre-disaster prevention, disaster relief, and post-disaster reconstruction
5	Forest freezing, rain and snow disaster prevention and reduction in southern China	http://drr.ikcest.org/knowledge_service/forest.html	Use Anusplin software to perform spatial discretization and provide visualization services
6	Flood control in Songliao basin	http://drr.ikcest.org/knowledge_service/control_flood.html	Provides the spatial distribution display and analysis services of flood disaster data and information based on WebGIS
7	Spatio-temporal distribution of arable land drought along "the Belt and Road initiatives" project area	http://drr.ikcest.org/knowledge_service/drought.html	Establish a drought model of precipitation anomaly percentage to provide the display of cultivated land distribution and analysis of spatiotemporal sequences
8	Suspended solids concentration inversion from 2000 to 2013 in Poyang Lake, China	http://drr.ikcest.org/knowledge_service/poyang_lake.html	Data modeling and inversion of the four seasons of Poyang Lake, forming a visual analysis of spatio-temporal sequences over many years
9	Annual spatial distribution data for drought monitoring in the Mongolian Plateau (1981–2012)	http://drr.ikcest.org/knowledge_service/mongolian.html	A stable drought monitoring model was constructed based on the universal feature space of Ts-NDVI to realize the analysis of spatio-temporal sequences for many years
10	Spatial distribution of the seasonal chlorophyll-a concentration in Poyang Lake, China (2009–2012)	http://drr.ikcest.org/knowledge_service/poyang_yls.html	Semi-empirical and empirical methods were used to obtain the estimation model of chlorophyll-a concentration in Poyang Lake and render a visual analysis
11	Total factor data of land cover in disaster environment in Mongolia	http://drr.ikcest.org/knowledge_service/mongolian_lc.html	Distribution of various types of land cover elements is obtained using object-oriented interpretation methods and visual analysis
12	Spatio-temporal distribution of major historical disasters in the China-Mongolia-Russia Economic Corridor	http://drr.ikcest.org/knowledge_service/zmezl.html	Collection of multi-source disaster data and information and provide visual display and analysis
13	Temporal and spatial distribution of public sentiment on Shouguang flood	http://drr.ikcest.org/knowledge_service/shouguang.html	Topic extraction and classification using Weibo text big data, LDA topic model, and random forest algorithm
14	Hazard-formative environments knowledge service of "the Belt and Road initiatives" project	http://drr.ikcest.org/knowledge_service/the_belt_and_road.html	Obtain basic national condition information of countries along the Belt and Road project through multi-source means such as internet, text, statistics, etc., and display and service online
15	Grassland yield in "the Belt and Road initiatives" China-Mongolia-Russia Economic Corridor	http://drr.ikcest.org/knowledge_service/grassland_yield.html	Establish a grass yield estimation model along the China-Mongolia railway (Mongolia section) to obtain long-term serial products and visualize them
16	Public opinion analysis for COVID-19	http://drr.ikcest.org/knowledge_service/ncp.html	Obtain and visualize public opinion in China during the outbreak of COVID-19 based on Sina Weibo big data

4.2 Topic Extraction and Classification

A topic extraction and classification model combining the LDA model and the random forest (RF) algorithm was used to hierarchically process COVID-19-related Weibo texts[3]. As shown in Figure 3, the first step was to mine and generalize the topics from the COVID-19-related Weibo entries[4] based on the Gensim package in Python, from which the topic-terminology and document-topic lists were obtained. The first level topics were generalized into seven topics: events notification, popularization of prevention and treatment, government response, personal response, opinion, and sentiments, seeking help and making donations. Then, topic extraction results were utilized as training samples for the RF algorithm[5]

to classify the Weibo data. The RF algorithm was implemented by a machine learning package named "scikit-learn" in Python. Finally, a secondary classification was implemented to divide the broad topic into more detailed sub-topics.

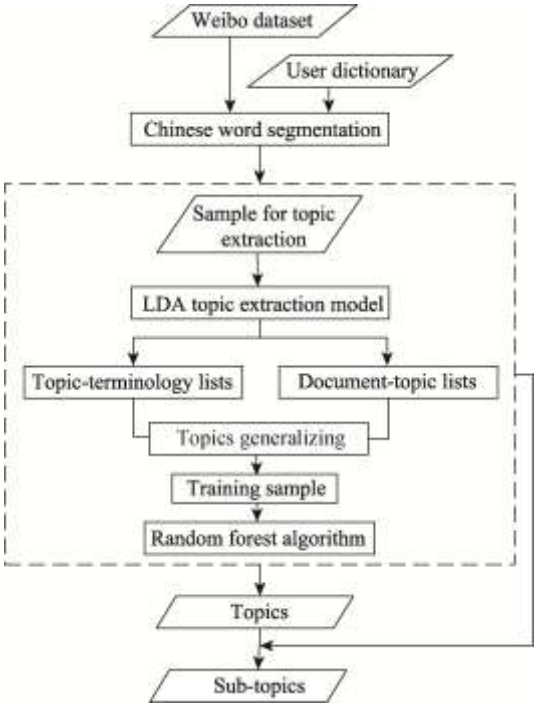

Figure 3 The processes of topic extraction and classification

4.3 Temporal-spatial Analysis

Taking the Weibo text information of the new coronavirus pneumonia from January 9, 0:00 to January 31, 24:00, 2020 as an example, 648,013 pieces of relevant information were initially obtained, including 55,260 pieces with geographic coordinates and located in China. Figure 4 is a time series analysis of the number of Weibo texts related to the pandemic. On January 9th, the pneumonia pathogen was initially identified as a novel coronavirus. The Weibo attention volume remained stable and slightly increased. On January 20th, the central government of China issued the highest instructions for novel pneumonia, and academician Zhong Nanshan confirmed that novel pneumonia had the characteristic of person to person transfer. The number of Weibo texts concerning the pandemic started to rise sharply, reaching a peak on the January 21st. Due to the impact of the Chinese Spring Festival holiday, the volume of attention dropped until January 29th. In this period, the response measures of the lockdown of Wuhan and some provinces such as Guangdong, Zhejiang, and Hunan also brought significant fluctuations in Weibo information. On January 31st, the WHO announced that the epidemic constituted a public health emergency of international concern, which also affected the change in the time trajectory of Weibo information.

Figure 5 is a spatial statistical map by province, showing that the main public opinion hotspots are concentrated in seven provinces including Hubei, Shandong, Henan, Jiangsu, Zhejiang, Sichuan, and Guangdong. Taking 200 km as the search radius and using the kernel density analysis method, a visual distribution map of public opinion related location information on the epidemic was formed (Figure 6). It shows that hotspots are prominently concentrated in the triangular high-value areas with the core hotspots in the Hubei-Henan border area, the Hebei-Shandong border area, the Jiangsu-Zhejiang-Anhui border area, as well as two independent hotspot areas in Sichuan and Guangdong.

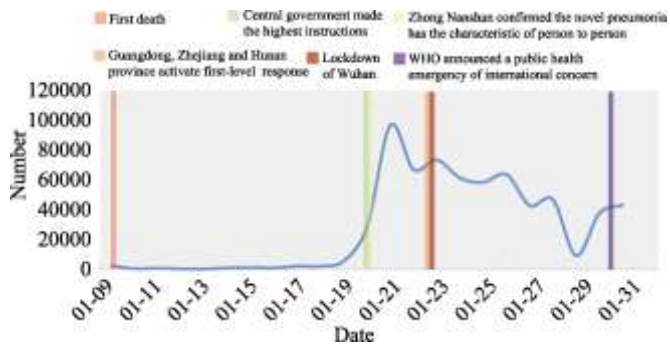

Figure 4 Time series of Weibo texts daily

5 Conclusion

The DRRKS has been activated online. Users can query, browse, and download resources for disaster risk reduction in DRRKS. As of the end of 2019, the DRRKS has provided the public with services of 167 datasets, 1,050 thematic maps, 90,000 metadata, 15 knowledge applications, and 220,000 documents. In response to the current urgent novel coronavirus pandemic, the DRRKS quickly built an online knowledge application of disaster public opinion analysis, publishing sharing, and visualizing services in the knowledge application module. This platform will continually provide support for experts and scholars in the fields of disaster mitigation and public health.

Acknowledgments

A special acknowledgment should be expressed to all the DRRKS team members, the Secretariat of IKCEST, and the experts of UNESCO's Earth Sciences and Geo-hazards Risk Reduction section.

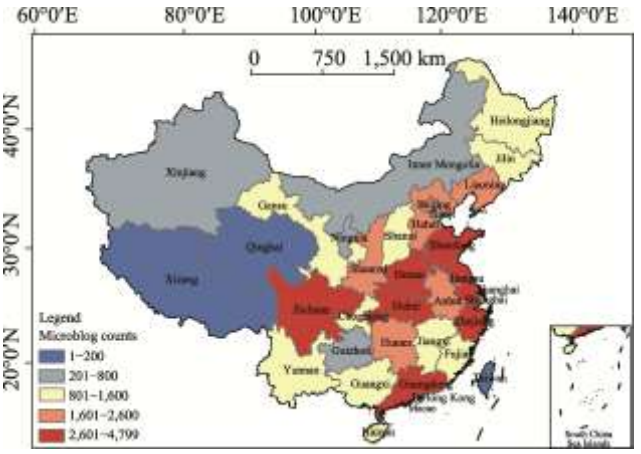

Figure 5 Spatial statistical map of epidemic-related microblogs by province

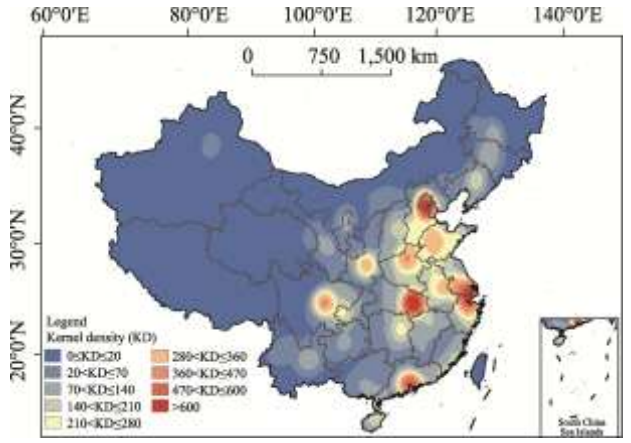

Figure 6 Kernel density analysis of epidemic-related Microblogs

References

[1] Wang, J., Bu, K., Yang, F., *et al*. Disaster risk reduction knowledge service: a paradigm shift from disaster data towards knowledge services [J]. *Pure and Applied Geophysics*, 2020, 177(1): 135–148.

[2] Wang, Y. J., Bu, K., Wang, J. L. Design and prototype implementation of disaster metadata management system based on open-source pycsw [J]. *E-science Technology* & *Application*, 2018, 9(2): 60–70.

[3] Han, X., Wang, J. Using social media to mine and analyze public sentiment during a disaster: a case study of the 2018 Shouguang city flood in China [J]. *ISPRS International Journal of Geo-Information*, 2019, 8(4): 185.

[4] Blei, D. M., Ng, A. Y., Jordan, M. I. Latent dirichlet allocation [J]. *Journal of Machine Learning Research*, 2003, 3(1): 993–1022.

[5] Breiman, L. Random forests [J]. *Machine Learnin*g, 2001, 45(1): 5–32.