Research article

Connotation, characteristics and framework of coal mine safety big data

Wanguan Qiao\textsuperscript{a,b,}\*, Xue Chen\textsuperscript{b}

\textsuperscript{a} School of Economics and Management, Jiangsu Vocational Institute of Architectural Technology, Xuzhou, 221116, China
\textsuperscript{b} School of General Course, Jiangsu Vocational Institute of Architectural Technology, Xuzhou, 221116, China

\section*{ARTICLE INFO}

\textbf{Keywords:}
Data-driven
CMSBD
Management connotation
Safety features
Research paradigm

\section*{ABSTRACT}

With the continuous development of automation and information technology, large amounts of safety data are produced in the processes of coal production. Most enterprises simply focus on statistics and do not conduct systematic big data analyses. Therefore, it is necessary to study the theory of coal mine safety while using big data systematically. This paper expounds on the changes in coal mine safety that have been driven by big data from three aspects: the connotation, characteristics and research framework. First, the connotation of coal mine safety big data (CMSBD) is redefined by changing the safety entities and methods. Second, the advantages and disadvantages of the big data model are compared from the perspective of feature analysis. Finally, the research paradigm and technical framework of CMSBD are designed. The results show that the management connotation of CMSBD focuses on the role of big data in coal mine safety. Compared with coal mine safety small data (CMSSD), CMSBD has both advantages and disadvantages. Therefore, CMSBD must be combined with a small data method. The research paradigm emphasizes the intersection of the research, the relevance of safety thinking, the importance of safety data analysis, and the fusion of big data with traditional small data models.

\section*{1. Introduction}

The major role of coal energy in energy production and consumption in China will not change for the foreseeable future. The output of coal will continue to ensure the rapid development of the national economy (Liu et al., 2019b). However, coal mine production safety is fundamental for the stable, sustainable and high-speed development of the coal industry, which not only guarantees the safety of the miners but also prevents the country from suffering heavy losses.

To realize the safety of production in coal mines, it is necessary to introduce new methods and concepts so that coal mining enterprises can shift from post-analysis management to prevention management and reduce the incidence of accidents. In the current coal industry, the use of computer control technology, network technology, big data analysis and other information technologies has become a trend for ensuring safe production. The 13th Five-Year Plan of China for production safety states that it is necessary to comprehensively promote the application of information technologies such as big data in production safety and to improve the performance in monitoring major hazards, identifying hidden dangers, and managing risk and control. Especially for some rare events, it is difficult to predict the risk through expert experience and direct information. At this time, the big data technology method is applied to the security field to mine the associative relationship between non-explicit data to make a more accurate prediction (Yang et al., 2015). Therefore, the use of information technologies to improve coal mine safety management in China has become a trend of current research.

Many scholars have also attempted to apply the Internet of Things, big data, information and other technologies to current safety research (He et al., 2017; Huang et al., 2018; Rivas et al., 2011). Sanmiquel et al. (2015) used the data mining method to analyse coal mine accidents in Spain from 2003 to 2012 and formulated improvement policies according to the laws of mining to minimize the occupational hazard rate of the mining sector. Cheng and Yang (2012) used rough sets and the support vector machine (SVM) algorithm in data mining to build a comprehensive early warning model for a mine ventilation system. The model can not only deal with a large amount of data that is generated by the mine ventilation system in a timely manner but also help coal mine managers make safety decisions. Qiao et al. (2018) used a decision tree algorithm in data mining to classify unsafe behaviours of miners. The results show that training, attendance, experience and age are all factors that affect the frequency of unsafe behaviours of human beings, among which training factors have the greatest impact on unsafe behaviour.

The emergence of big data has highly impacted the traditional safety field. Scholars have applied big data to safety in different industries (such as...
as coal mining, transportation, aviation, marine, and construction) and have conducted big data research in the fields of public safety, food safety, traffic safety and others (Liu et al., 2022; Mannering et al., 2020; Talari et al., 2021). However, few scholars have studied the basic principles of big data application to the field of safety science, and scholars insist that traditional safety theory is not compatible with big data theory. Traditional safety theory emphasizes causality between factors, while big data focuses more on the correlation between factors (Wang and Wang., 2021). Scholars have pointed out that the current safety concepts need to be improved to adapt to the development of the current data era (Huang et al., 2018). In this case, how researchers should assess the relationship between big data and safety management is particularly important. Whether to abandon the traditional safety theory or to effectively combine it with traditional coal mine safety not only poses theoretical challenges but also poses application challenges in CMSBD.

In the current research on CMSBD, we identified three problems. First, most of the current research is limited to one aspect of big data in coal mine safety and is not conducted from the overall perspective. Second, most of the current research is on the application of big data technology in coal mine safety; the supporting theory of coal mine safety management that is driven by big data is lacking. Third, big data mining methods are used directly to conduct coal mine risk management, and there is a lack of connectivity with the theoretical research that is driven by the original coal mine safety management model.

Based on the above analysis, this paper attempts to explain the changes in coal mine safety management that are driven by big data from three aspects: the connotation, characteristics and research framework. Section 2 redefines the connotation of big data in coal mine safety management according to changes in safety management entities and methods. Section 3 analyses the advantages and disadvantages of big data models and small data models in coal mine management from the perspective of feature comparison. Section 4 designs the functional and technical framework of coal mine safety management research that is driven by big data. Finally, the conclusions of this study are presented in Section 5.

2. New connotation of CMSBD

2.1. Traditional connotation of coal mine safety

A coal production system is a complex social technology system, and its dynamic and nonlinear characteristics are affected by internal and external factors (He and Song, 2012; Zhang et al., 2020). The occurrence of a coal mine accident depends not only on human factors, machine factors, and environmental factors but also on more important management factors that restrict these three factors (Qiao, 2021; Liu et al., 2019a). We believe that the direct causes of an accident are unsafe human behaviours, unsafe states of machinery, and unsafe environmental states, and the essential reason for these unsafe states is management defects (Figure 1).

Based on the above accident causation model, the connotation of traditional coal mine safety management refers to the use of limited safety management resources and the correlation between various departments by functional departments to reduce the number and risk of hidden dangers that are caused by human, machinery, and environmental factors.

2.2. New connotation of CMSBD

In the context of big data, the connotation of coal mine safety management focuses more on the management of safety big data. Coal mine enterprises conduct multiple tasks and procedures, and they use the cooperative relationships between departments to manage coal mine safety tasks. In this process, a large amount of safety data will be generated during the execution of tasks and procedures in various departments, and these safety data can reflect the safety of the departments. The essential strategy for coal mine safety in the context of big data is to use coal mine big data to replace traditional safety information and to use a variety of data mining and analysis methods to increase safety efficiency. The connotation of coal mine safety under the background of big data should focus on the collection, cleaning, processing and analysis of safety big data to obtain valuable safety management rules to provide a realistic basis for enterprise managers to make safety decisions. Therefore, the further refinement of the big data content for coal mine safety management can be reflected in two aspects:

2.2.1. Change in the direct entity of coal mine safety

Under the background of big data, coal mine safety entities are transformed from traditional physical entities (e.g., departments, teams, miners, equipment, and the environment) to safety management big data that are generated by physical entities. The use of coal mine safety management data further acts on the entity of coal mine safety management and increases the efficiency of safety management (Figure 2). In other words, coal mine safety management in the context of big data does not abandon traditional safety management entities but adds big data on coal mine safety management for the realization of concrete and concise traditional safety management entities.

2.2.2. Changes in coal mine safety management modes

Safety management modes can be divided into empirical safety management, institutional safety management, precontrol safety management and safety data culture management (Van Dyke et al., 2020). Big data safety management is a newly proposed concept that represents the embodiment of big data thinking in safety management. Under the background of big data...
data, coal mine safety management modes have changed from empirical safety management and institutional safety management to precontrol safety management and safety data culture management and finally realize big data security management (Figure 3). Wang (2021) put forward the concept of safety 4.0, pointed out the development process of the safety field from experience—technology—risk—intelligence, and gave the question of safety 5.0. Big data safety management will pay more attention to the objectivity, timeliness, foresight and relevance of safety management. The data source of safety management from the perspective of big data is full sample data to reduce human error in the simulation process. The results obtained are based on the current safety data obtained under a certain algorithm to avoid the subjective defect brought by traditional safety management.

3. Features of CMSBD

3.1. Basic characteristics of CMSBD

No unified definition is available for big data. Among the many definitions, the most accepted is the definition from the perspective of big data features, which include the “6Vs” characteristics, namely, volume, variety, velocity, value, variability and veracity characteristics (Amanullah et al., 2020; Sun et al., 2018). Hence, big data refers to high-value, low-density, real-world datasets. With the development of coal mine safety technology and management information, coal mine enterprises have also accumulated large amounts of safety big data. The CMSBD also conform to the “6Vs” characteristics of big data. The characteristics are presented in Figure 4.

3.1.1. Massive scale of CMSBD

The coal mine production data include not only data on unsafe human behaviour but also data on the safety status of equipment and the environment. In addition, there are safety data that are related to the overall scheduling and production operations of coal mining enterprises. According to statistics, a medium-sized coal mine can produce approximately 10 GB of safety monitoring data every day (Qi, 2020). The multifaceted data collection has increased the amount of available coal mine safety management data, which has exceeded the capabilities of manual processing.

3.1.2. Wide variety of CMSBD

In China, CMSBD are diverse, and there is no unified standard. The data include both long-term static data and changing dynamic data. At the same time, the data also include structured data with regular forms and unstructured data with messy forms. The collection form of coal mine safety management data is also different. Safety management data are automatically generated by sensors, as well as manual safety data. In addition, there are numerical safety management data, as well as a large amount of textual data. Therefore, coal mine safety big data presents multsource and heterogeneous characteristics. To realize the integration of different data, it is necessary to establish a unified data warehouse to integrate coal mine safety management data.

3.1.3. Low value density of CMSBD

With the increased availability of big data on coal mine safety, the amount of valuable and invisible information is also increasing, but the growth rate is far lower than the growth rate of the data volume, leading to a lower proportion of safety information in safety data. The production conditions in most places underground will not change frequently. Therefore, numerous steady-state time series data are generated. The utilization value of these data is low, and the amount of data involving disasters and accidents is relatively small. In addition, unstructured data account for a large proportion of coal mine production. A large amount of content has no analytical value, which is also an important factor causing the low value density of CMSBD.

3.1.4. Fast processing speed of CMSBD

For a long time, coal mines in China have been regarded as nontechnical industries with high labour participation and low-tech operations. However, with the development of science and technology, the
The degree of automation of coal mine production is increasing, and the participation of personnel is decreasing. This scenario has brought about an improvement in the level of safety management, but it has also induced a rapid increase in the amount of safety data. The fast-growing amount of CMSBD requires that the speed of data processing be correspondingly improved to make effective use of a large amount of data. Otherwise, the increasing data will not bring advantages to solve problems but will become a burden to solve problems quickly.

3.1.5. Veracity of CMSBD

CMSBD can comprehensively and carefully depict the trajectory of changes in the safety level of coal mine entities. Although the CMSBD is more complete, true, detailed and solid, there are still some noise and false safety data. Even some coal mine managers illegally modify safety data to avoid legal responsibility, which will lead to the deviation of superior leaders’ safety decisions. Therefore, it is necessary to clean the CMSBD and eliminate all adverse factors that cause the false phenomenon of big data. Only by fine filtering of coal mine safety big data can we present the objective relationship between safety data and risk events and ensure the quality and veracity of coal mine safety big data.

3.1.6. Continuous variation of CMSBD

CMSBD shows irregular changes, and the variability of these data will appear in humans, equipment and the environment in coal mine safety activities. Specifically, different miners have different psychological or physiological characteristics, so the unsafe behaviour rules obtained from big data should consider personal behaviour differences. In addition, the upgrading of coal mine equipment will also affect the variability of CMSBD. In addition, the underground environment of coal mines is complex and changeable, and the degree of safety of the underground environment of different coal mines is also inconsistent. Some coal mines are low gas mines, and the risk of gas explosion is low, while some coal mines are high gas mines, so the attention to gas data is higher. The variability of CMSBD in these aspects brings trouble to safety decision-making. Mine managers need to make appropriate decisions according to the actual situation of the coal mine to reduce the information confusion caused by variability.

3.2. Advantages and limitations of CMSBD compared with CMSSD

In recent years, the explosive development of big data has led some people to think that big data are omnipotent (Lang et al., 2018; Wang et al., 2018). However, in addition to its advantages, the use of big data in coal mines has several limitations compared with coal mine safety small data (CMSSD). The different features of CMSBD and CMSSD are shown in Table 1.

3.2.1. Advantages of CMSBD compared with CMSSD

One of the main advantages is the convenience of safety index selection via CMSBD. In traditional small data analysis, the largest headache for scholars is the selection of indicators. Many factors affect coal mine safety management. The selection and quantification of indicators are challenging if using small data. However, if using the big data of coal mine safety management, it is not necessary to consider the problem of data sampling because all the safety data can be used for analysis and, facilitated by a mining algorithm, for data processing.

The second advantage is the high-speed information processing that is realized with CMSBD. The main objective in using big data is to process a large amount of data and to identify effective association rules. Coal mining enterprises produce large amounts of safety data, which cannot
be processed by manpower every day, every hour or even every minute. By using computer technology that is related to big data, we can identify potentially useful association information from a large “data sea” to provide a basis for safety decision-making of managers. In addition, high-speed information processing enables researchers to identify characteristic groups that are rare and cannot be judged by the human brain, such as unsafe behaviour characteristic groups in coal mining enterprises.

Third, resource integration is possible with CMSBD. Big data bring the advantage of resource integration to coal mine safety management. The traditional safety management method is linear, and the safety management method under big data is a nonlinear grid. Thus, the integration of various resources is realized, and various safety problems that could not be solved originally can be solved.

3.2.2. Limitations of CMSBD compared with CMSSD

One limitation is the algorithm lag of CMSBD. According to scholars, human beings have entered the era of big data, and the amount of data on the internet will double every two years (Wang et al., 2019). However, the update speed of big data algorithms is far behind the growth speed of big data volume. Especially in the theory of coal mine safety management, few data mining algorithms are available for coal mine safety management.

The second limitation is the heterogeneity of CMSBD. Machine learning can only deal with data that have the same structure, and heterogeneous data must be structured. However, there is a considerable amount of heterogeneous data in coal mining enterprises. Substantial amounts of time and experience are required for processing data that amount of heterogeneous data in coal mining enterprises. Substantial data processing has pertinence, causality and maturity. Small data analysis has been applied in the field of coal mine safety for a long time and is in the mature stage but lacks innovation. Coal mine safety under big data analysis is in the rapid growth stage, but lacks stability. Stability is required when facing a coal mine safety management problem. Scholars should comprehensively consider both big data and small data analyses (Zhang, 2016). The combination of the two promotes the innovative development of the theoretical system of coal mine safety.

The third limitation is the privacy of CMSBD. An important reason for the rapid development of big data is that data acquisition has become easier. Coal mining enterprises are typically unwilling to share their safety management data with other enterprises. One reason is that the number of accidental deaths in these data is sensitive and not subject to public opinion. Another reason is that the data also relate to the production and operation status of the company. The disclosure of these data may cause problems such as plunging stock and strict supervision by government departments. Finally, the relevant conclusions that are obtained by using private coal mine safety big data cannot provide data for verification and sharing; hence, the research conclusions lack authority.

Fourth, there is a lack of causality in CMSBD. In the field of coal mine safety, small mistakes can lead to severe accidents. Therefore, after using the associations that are obtained from CMSBD, the severe accidents should also be explained causally. Causes are always accompanied by results, and the results are always due to causes (H Wu et al., 2019). Therefore, the premise of research on big data for coal mine safety management is to acknowledge the universality and objectivity of causality.

3.3. Combination of CMSBD and CMSSD

In the study discussed above, the CMSBD is found to have advantages and limitations. The limitations of big data are often the advantages of small data (Wang and Wu, 2020). Therefore, by attempting to combine the CMSBD and CMSSD, the combination of data-driven and model-driven analysis is realized. The details are as follows.

1. The combination of big data analysis and small data analysis. Big data analysis is more dynamic, relevant, forward-looking and comprehensive. Small data analysis has pertinence, causality and maturity. Small data analysis has been applied in the field of coal mine safety for a long time and is in the mature stage but lacks innovation. Coal mine safety under big data analysis is in the rapid growth stage, but lacks stability. Stability is required when facing a coal mine safety management problem. Scholars should comprehensively consider both big data and small data analyses (Zhang, 2016). The combination of the two promotes the innovative development of the theoretical system of coal mine safety.

2. The combination of big data association and small data causality. Big data theory is useful for identifying the relationships that small data theory cannot identify and for building a series of hypotheses that are based on these relationships. According to the requirements of safety management, researchers can design and conduct a series of experiments, observations, and simulations to evaluate the accuracy of these relationships to develop a basis for coal mine safety science theory.

3. The combination of group analysis of big data and individual analysis of small data. The prediction accuracy of big data for group behaviour is much higher than the prediction accuracy for individual behaviour, due mainly to the larger amount of data of the group and more effective relationships. Small data are useful for the analysis of individual characteristics, and the accuracy of the analysis results for a single post, accident or individual behaviour is higher than the accuracy of a single analysis of big data. Therefore, large data group analysis and small data individual analysis can be combined.

4. The combination of a big data machine learning algorithm and a small data probabilistic method. Although big data have a high signal-to-noise ratio, which reflects mainly the correlations between the variables, involvement of researchers is very low. The result is more objective. In addition, small data statistical methods have various framework limitations; hence, it is difficult to break through the original framework for innovation. However, computer technology and big data algorithms only provide the possibility for secondary innovation of coal mine safety management theory. Via the combination of big data and small data, it is possible to realize the innovation of the coal mine safety management theory, model, algorithm and data processing.

4. Research process and paradigm of CMSBD

4.1. Basic process of CMSBD research

In the application of CMSBD, both general research processes and research content are related to the data characteristics. The data mining process of CMSBD includes five stages: (1) determination of safety issues, (2) data collection, cleaning and conversion, (3) construction of a data mining model, (4) analysis and interpretation of the mining results, and (5) application and optimization of the model (Ahmed et al., 2017; Glaeser et al., 2018). Figure 5 presents a basic flow chart of CMSBD research. The analysis steps are as follow.
4.1. Determination of safety issues

When applying big data to coal mine safety, we should first determine the background and development status of the safety issues that are being studied. Then, we should identify the cause of the problem to provide a factual basis for the selection of subsequent data and the construction of the model.

4.1.2. Data collection, cleaning and conversion

The structure of CMSBD is intricate and complex, and the data include structured, unstructured and semistructured data. Therefore, data cleaning, integration and transformation must be completed prior to data mining (Dos Santos et al., 2019; Mondal, 2015). Anomalous data can be replaced via data analysis methods such as moving averages or autoregressive models; noise data should be processed using filtering or wavelet denoising; and missing data can be supplemented via smooth processing methods.

4.1.3. Construction of a data mining model

Data mining models have many functions, such as classification, clustering, association, and prediction. The algorithms differ according to the functions. Artificial neural networks are considered the most commonly used machine learning prediction method (Shah et al., 2021). In addition, the utilization of support vector machine and random forest methods is also increasing (Osarogiagbon et al., 2021). Text mining is used in numerous accident reports to find the key causes of accidents (Qiu et al., 2021). Data visualization technology is also the current research hotspot of safety big data. Data visualization realizes dynamic risk monitoring by converting data into graphics (Wu et al., 2019). Therefore, when constructing a coal mine safety data mining model, it is necessary to choose a suitable identifier for the operation of the data mining algorithm based on the characteristics of the data and the expected target.

4.1.4. Analysis and interpretation of the mining results

The use of data mining models to analyse CMSBD can yield useful results, and these results focus on the correlations between the data and the lack of causal analysis. Therefore, data mining results must be evaluated, analysed, and interpreted to eliminate meaningless or low-value associations. Finally, via a comprehensive analysis of safety issues, researchers have proposed targeted countermeasures and suggestions.

4.1.5. Application and optimization of the model

The application and optimization of the data mining model constitute a feedback process. The application of the model can drive changes in the safety level of the coal mine, and the improvement of the safety level will also improve the data mining model. A safety model that is established based on available safety data can address current safety issues. However, with changes in the coal mine safety level and safety issues, the data mining model is no longer applicable to the new safety environment. At this time, it is necessary to reorganize the safety issues and to build a new data mining model.

4.2. Conversion model of CMSBD, safety information and safety laws

Coal mine safety driven by big data requires the reorganization of the relationship between data and laws so that it can explain the role of big data in coal mine safety (Moradpour and Long, 2019). This section attempts to build the relationship and transformation model among CMSBD, safety information, safety knowledge and safety laws (Figure 6).

According to Figure 6, the CMSBD, coal mine safety information and coal mine safety laws are not simple straight-line models but triangular models that involve four safety transformation paths. All four models are transformed via safety knowledge as an intermediary variable (Huang et al., 2019). The analysis process is as follows:

4.2.1. Transformation from CMSBD to safety information

Data carry information, and information is the external manifestation of data. CMSBD are transformed into safety information via the use of data processing. In the process of converting data into information, safety knowledge also plays a supporting role. Data processing for coal mine safety should use safety knowledge as a guide to realize the transformation of safety information. Comprehensively extracting safety
4.3. Basic paradigm system of CMSBD research

The research logic of CMSBD that is driven by big data focuses on safety issues, safety data, safety information, safety laws, and safety practices (Figure 7). CMSBD combines the characteristics and basic processes of big data to build a basic paradigm system for CMSBD.

According to Figure 7, the CMSBD research paradigm is a combination of the basic process of big data analysis and the relationships among the safety elements. The paradigm aims to describe the “panoramic” decision framework for coal mine safety that is driven by big data. The connotation of the research paradigm system is as follows.

1. Emphasize the intersection of multiple research lines. The paradigm of CMSBD that is driven by big data includes three cross-cutting research lines: First is the main line of safety issues—safety data-safety information-safety laws-safety practice logic, and the second is the conventional data mining subline of safety issues-data analysis-model construction-result evaluation-model optimization. The third is the system thinking subline of safety status quo-safety interpretation-safety prediction-safety decision. The three research lines are crossed to realize the innovation of the research paradigm of coal mine safety that is driven by big data.

2. Emphasize the relevance of safety thinking. In terms of safety information-safety laws, this process requires the completion of a rough safety association. Safety laws, principles, models, and techniques are used mainly to analyse information to derive safety laws (Guo et al., 2016; Lozada et al., 2019). In terms of safety law-safety practice, this process requires the completion of safety fine association. According to the specified safety requirements, the resulting safety laws are targeted to safety practices.

3. Emphasize the importance of safety big data processing. In the aspect of safety data analysis, we must perceive and collect data and be able to decompose and aggregate data in various dimensions and levels; in the aspect of cross-border association, we must capture data relationships and their dynamic changes and be able to integrate internal and external data that are multisource and heterogeneous. In view of the safety issues studied, the data collection process must be as comprehensive as possible. At the same time, some safety data contain noise or suffer from missing data, and the safety data must be cleaned for direct application.

4. Emphasize the integration of CMSBD and CMSSD. Most traditional small data theoretical models are not suitable for processing big safety data (Montaños et al., 2019; Zhou et al., 2019). The locally weighted linear regression (LWLR) model was proposed to provide early warning of screen out events (Hu et al., 2020). Therefore, the safety model can be suitably modified to enable it to complete the processing of CMSBD so that the improved small data model can be combined with CMSBD to realize decision-making improvements. Similarly, some big data models are not suitable for small sample data. Enrichment of the available small data theoretical models by improving the big data model is also an innovative method.

5. Discussion

The types and scope of CMSBD are expanding. The CMSBD is expanded from the original numerical data to text, picture, video, audio and other data, and the emergence of these data drives the rapid development of big data mining methods. In the field of coal mine safety, the research scope of big data is often larger than the research scope of small data. The coal mine safety method for small sample data focuses more on the pertinence and concentration of the problem, while the coal mine safety method under large sample data focuses more on the comprehensiveness and inclusiveness of the problem (Ham and Park, 2020). Therefore, the big data of coal mine safety are not limited to the internal safety data of coal mine enterprises and the safety data of regulatory agencies but have been expanded to include external big data, such as public opinion data on coal mine safety, safety laws and regulations.

The core objective of CMSBD is prediction. Coal mine data have the 6V characteristics of big data. Big data technology can be used to analyse and predict the likelihood of coal mine accidents. By collecting all data that are related to safety phenomena and using big data technology to process the mathematical relationships between data values, we derive their correlations and predict the safety status of the
production process. Correlation analysis on big data provides a new perspective for safety management and decision-making that is based on data analysis results, thereby effectively reducing errors in human intuition, facilitating the establishment of a safety management system that is centred on data, technology, and analysis.

Scholars have attempted to compare the advantages and disadvantages of the social science research paradigm from the perspectives of big data and small data (Ouyang et al., 2018; Perrons and McAuley, 2015). The advent of the big data era does not correspond to the end of the use of traditional small data safety methods. The background of big data can compensate for the shortcomings of traditional safety theory. Traditional safety theory focuses more on the causal relationships between safety elements, while safety big data focuses more on the relationships between data (Bilal et al., 2016). By combining causality and association, it is possible to mine coal mine safety information more comprehensively and efficiently.

The use of big data to improve the current theoretical model of coal mine safety and to realize the new paradigm of "model driven + data driven" has innovative significance for current safety research (Iqbal et al., 2018). The effective combination of big data theory and coal mine safety theory not only broadens the scope of current safety theory but also improves the depth of research on coal mine safety theory.

Although researchers focus mainly on the advantages of big data in safety science, inaccurate safety data and invalid safety laws may lead to the occurrence of coal mine accidents, thereby reducing the efficiency of coal mine safety. Therefore, to investigate coal mine safety against the background of big data, scholars should conduct research with a questioning and careful attitude.
6. Conclusions

This article describes the changes in CMSBD from three aspects: the connotation, characteristics and research paradigm. The conclusions are as follows.

(1) The management connotation of CMSBD focuses more on the role of big data in coal mine safety. The main changes are reflected in two aspects: the transformation of coal mine safety big data from physical entities (such as departments, teams, miners, equipment and the environment) to safety data (structured, semistructured and unstructured). This process accelerates the maturity of safety data culture and finally reaches the stage of big data safety in coal mines.

(2) CMSBD still conforms to the 6 V characteristics of big data, but its characteristics have several shortcomings. By analyzing the advantages and limitations of CMSBD compared with CMMSSD, we attempt to combine big data and small data to realize both big data and small data thinking, to determine the relationship between big data and small data causality and to conduct big data group analysis and small data analysis. The fusion of individual analysis, big data machine learning algorithms and small data probabilistic methods promotes the rapid development of coal mine safety research.

(3) By integrating the two aspects of the basic process of big data analysis and the relationship model of safety elements, the basic research paradigm of CMSBD is constructed. The research paradigm emphasizes the intersection of the main lines of research, the relevance of safety thinking, the importance of safety data collection and analysis, and the fusion of big data with traditional small data models.

In summary, the arrival of big data does not mean the end of traditional safety theory. The background of big data can compensate for the shortcomings of traditional safety theory. Traditional safety theory focuses more on the causal relationships between safety elements, while big data focuses more on the relationships among the data. By combining causality and association, it is possible to more comprehensively and efficiently mine deep-level coal mine safety information.

Declarations

Author contribution statement

Wanguan Qiao: Conceived and designed the analysis; Analyzed and interpreted the data; Contributed analysis tools or data; Wrote the paper.

Xue Chen: Conceived and designed the analysis; Contributed analysis tools or data; Wrote the paper.

Funding statement

Dr wanguan qiao was supported by Jiangsu Construction System Science and Technology Project [2019DDB085].

Data availability statement

No data was used for the research described in the article.

Declaration of interest’s statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

Ahmed, E., Yaqoob, I., Hashem, I.A.T., Khan, I., Ahmed, A.A., Imran, M., Vanlakos, A.V., 2017. The role of big data analytics in Internet of Things. Comput. Network. 129, 459–471.

Amanullah, M.A., Habaeb, R.A.A., Nasaruddin, F.H., Gani, A., Ahmed, E., Naisar, A.S.M., Akim, N.M., Imran, M., 2020. Deep learning and big data technologies for IoT security. Comput. Commun. 151, 495–517.

Bilal, M., Oyedele, L.O., Qadiri, J., Mumir, K., Ajayi, S.O., Akinade, O.O., Owolabi, H.A., Alaka, H.A., Pasha, M., 2016. Big Data in the construction industry: a review of present status, opportunities, and future trends. Adv. Eng. Inf. 30, 500–521.

Cheng, J., Yang, S., 2012. Data mining applications in evaluating mine ventilation system. Saf. Sci. 50, 918–922.

Dos Santos, B.S., Steiner, M.T.A., Fenerich, A.T., Lima, R.H.P., 2019. Data mining and machine learning techniques applied to public health problems: a bibliometric analysis from 2009 to 2018. Comput. Ind. Eng. 138, 106120.

Glaeser, E.L., Kominers, S.D., Luca, M., Naik, N., 2018. Big data and big cities: the promises and limitations of improved measures of urban life. Econ. Inq. 56, 114–137.

Guo, S.Y., Ding, L.Y., Luo, H.B., Jiang, X.Y., 2016. A Big-Data-based platform of workers’ behavior observations from the field. Accid. Anal. Prev. 93, 299–305.

Ham, D.H., Park, J., 2020. Use of a big data analysis technique for extracting HRA data from event investigation reports based on the Safety-II concept. Reliab. Eng. Syst. Saf. 194, 106222.

He, X., Song, L., 2012. Status and future tasks of coal mining safety in China. Saf. Sci. 50, 894–898.

He, P., Gu, L., Wang, T., Zhang, Z., 2017. The synthetic geo-environmental ecological evaluation of a coastal coal-mining city using spatiotemporal big data: a case study in Longzhou, China. J. Clean Prod. 142, 854–866.

Hu, J., Khan, F., Zhang, L., Tian, S., 2020. Data-driven early warning model for screenout scenarios in shale gas fracturing operation. Comput. Chem. Eng. 143.

Huang, L., Wu, C., Wang, B., Ouyang, Q., 2018. Big data-driven safety decision-making: a conceptual framework and its influencing factors. Saf. Sci. 109, 46–56.

Huang, L., Wu, C., Wang, B., 2019. Challenges, opportunities and paradigm of applying big data to production safety management: from a theoretical perspective. J. Clean. Prod. 231, 592–599.

Iqbal, R., Doctor, F., More, B., Mahmud, S., Youssuf, U., 2018. Big data analytics: computational intelligence techniques and application areas. Technol. Forecast. Soc. Change 119,253.

Lang, H., Chao, W., Bing, W., Ouyang, Q., 2018. A new paradigm for accident investigation and analysis in the era of big data. Process Saf. Prog. 37.

Liu, Q., Li, X., Hassall, M., 2019a. Regulatory regime on coal mine safety in china and australia: comparative analysis and overall findings. Resour. Policy. 101,454.

Liu, Q., Meng, X., Li, X., Luo, X., 2019b. Risk precontrol continuum and risk gradient control in underground coal mining. Process Saf. Environ. Protect. 129, 210–219.

Liu, Z.G., Li, X.Y., Zhu, X.H., 2022. Scenario modeling for government big data governance decision-making: Chinese experience with public safety services. Inf. Manage. 59, 103022.

Lozada, N., Arias-Pérez, J., Perdomo-Charry, G., 2019. Big data analytics capability and co-innovation: an empirical study. Heliyon 5, e02541.

Manenning, F., Ibit, C.R., Shanksr, V., Abdel-Aty, M., 2020. Big data: traditional data and the tradeoffs between prediction and causality in highway-safety analysis. Anal. Methods Accid. Res. 25, 100113.

Mondal, K., 2015. Big data parallelism: issues in different X-information paradigms. Process Saf. Environ. Protect. 147, 367–384.

Ouyang, Q., Wu, C., Huang, L., 2018. Methodologies, principles and prospects of applying big data to production safety management: a case study. Case Stud. Transport Policy 7, 178–184.

Ouorougbo, A.U., Khan, F., Venkatesan, R., Gillard, P., 2021. Review and analysis of supervised machine learning algorithms for hazardous events in drilling operations. Process Saf. Environ. Protect. 147, 367–384.

Ouyang, Q., Wu, C., Huang, L., 2018. Methodologies, principles and prospects of applying big data in safety science research. Saf. Sci. 101, 60–71.

Perrons, R.K., McAuley, D., 2015. The case for “n-all”: Why the Big Data revolution will probably happen differently in the mining sector. Resour. Pol. 46, 234–238.

Qi, C.C., 2020. Big data management in the mining industry. Int. J. Miner., Metall. Mater. 27 (2), 131–139.

Qiao, W., 2021. Analysis and measurement of multifactor risk in underground coal mine accidents based on coupling theory. Reliab. Eng. Syst. Saf. 208.

Qiao, W., Liu, Q., Li, X., Luo, X., Wan, Y., 2018. Using data mining techniques to analyze the influencing factor of unsafe behaviors in Chinese underground coal mines. Resour. Pol. 59, 210–216.

River, T., Paz, M., Martin, J.E., Mattas, J.M., García, J.F., Taboada, J., 2011. Explaining and predicting workplace accidents using data-mining techniques. Reliab. Eng. Syst. Saf. 96, 739–747.

Summerfield, R., Rosell, J.M., Vintror, C., 2015. Study of Spanish mining accidents using data mining techniques. Saf. Sci. 75, 49–55.

Shah, M.I., Javed, M.F., Alqahtani, A., Aldrees, A., 2021. Environmental assessment based surface water quality prediction using hyper-parameter optimized machine learning models based on consistent big data. Process Saf. Environ. Protect. 151, 324–340.

Sun, G., Chang, Y., Guan, S., Ramachandran, M., Li, J., Liao, Y., 2018. Big data and Internet of things—fusion for different services and its impacts. Future Gener. Comput. Syst. 86, 1368–1370.
Talari, G., Cummins, E., McNamara, C., O’Brien, J., 2021. State of the art review of Big Data and web-based Decision Support Systems (DSS) for food safety risk assessment with respect to climate change. Trends Food Sci. Technol. In Press.
Van Dyke, M., Klemetti, T., Wickline, J., 2020. Geologic data collection and assessment techniques in coal mining for ground control. Int. J. Min. Sci. Technol.
Wang, B., 2021. Safety intelligence as an essential perspective for safety management in the era of Safety 4.0: from a theoretical to a practical framework. Process Saf. Environ. Protect. 148, 189–199.
Wang, B., Wang, Y., 2021. Big data in safety management: an overview. Saf. Sci. 143, 105414.
Wang, B., Wu, C., 2019. Demystifying safety-related intelligence in safety management: some key questions answered from a theoretical perspective. Saf. Sci. 120, 932–940.
Wang, B., Wu, C., 2020. Safety informatics as a new, promising and sustainable area of safety science in the information age. J. Clean. Prod. 252, 119852.
Wang, B., Wu, C., Reniers, G., Huang, L., Kang, L., Zhang, L., 2018. The future of hazardous chemical safety in China: opportunities, problems, challenges and tasks. Sci. Total Environ. 643, 1–11.

Wang, B., Wu, C., Huang, L., Kang, L., 2019. Using data-driven safety decision-making to realize smart safety management in the era of big data: a theoretical perspective on basic questions and their answers. J. Clean. Prod. 210, 1595–1604.
Wu, H., Wu, D., Zhao, J., 2019. An intelligent fire detection approach through cameras based on computer vision methods. Process Saf. Environ. Protect. 127, 245–256.
Wu, Y., Chen, M., Wang, K., Fu, G., 2019. A dynamic information platform for underground coal mine safety based on internet of things. Saf. Sci. 113, 9–18.
Yang, M., Khan, F., Lye, L., Amyotte, P., 2015. Risk assessment of rare events. Process Saf. Environ. Protect. 98, 102–108.
Zhang, C., 2016. Study on big data processing and knowledge discovery analysis method for safety hazard in coal mine. J. Saf. Sci. Technol. 9, 176–181.
Zhang, W., Wang, M., Zhu, Y.-c., 2020. Does government information release really matter in regulating contagion-evolution of negative emotion during public emergencies? From the perspective of cognitive big data analytics. Int. J. Inf. Manag. 50, 498–514.
Zhou, W., Zhang, P., Wu, R., Hu, X., 2019. Dynamic monitoring the deformation and failure of extra-thick coal seam floor in deep mining. J. Appl. Geophys. 163, 132–138.