Uncertainty prediction of mining safety production situation

Menglong Wu · Yicheng Ye · Nanyan Hu · Qihu Wang · Wenkan Tan

Received: 16 September 2021 / Accepted: 11 April 2022 / Published online: 27 April 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
In order to explore the occurrence and development law of mining safety production accidents, analyze its future change trends, and aim at the ambiguity, non-stationarity, and randomness of mining safety production accidents, an uncertainty prediction model for mining safety production situation is proposed. Firstly, the time series effect evaluation function is introduced to determine the optimal time granularity, which is used as the window width of fuzzy information granulation (FIG), and the time series of mining safety production situation is mapped to Low, R, and Up three granular parameter sequences, according to the triangular fuzzy number; then, the mean value of the intrinsic mode function (IMF) is maintained in the normal dynamic filtering range. After the ensemble empirical mode decomposition (EEMD), the three non-stationary granulation parameter sequences of Low, R, and Up are decomposed into the intrinsic mode function components representing the detail information and the trend components representing the overall change, and then the sub-sequences are reconstructed according to the sample entropy to highlight the correlation among the sub-sequences; finally, the cloud model language rules of mining safety production situation prediction are created. Through time series discretization, cloud transformation, concept jump, time series set division, association rule mining, and uncertain reasoning, the reconstructed component sequence is modeled and predicted by uncertainty information extraction. The accuracy of the uncertainty prediction model was verified by 21 sets of test samples. The average relative errors of Low, R, and Up sequences were 9.472 %, 16.671 %, and 3.625 %, respectively. The research shows that the uncertainty prediction model of mining safety production situation overcomes the fuzziness, non-stationarity, and uncertainty of safety production accidents, and provides theoretical reference and practical guidance for mining safety management and decision-making.

Keywords Mining safety prediction · Uncertainty · Time granularity · Fuzzy information granulation · Ensemble empirical mode decomposition · Cloud model

Introduction
Mining is an important basic industry of the national economy, supporting the rapid development of social economy, but mining is also an industry with high casualty rate in industrial production process (Carvalho 2017). At present, China is in the process of rapid development of modernization. In general, it is still in a special period of prone and frequent production safety accidents (Li et al. 2020a, b, 2021a). Accurate prediction of mining safety production situation is the premise of providing mining safety decision-making and ensuring mining safety production. Therefore, it is of great practical significance to carry out research on the occurrence and development laws of the mining safety production situation and analyze its future change trends.

The mining safety system is a complex circulatory system, which is affected by many factors such as natural conditions, laws and regulations, social environment, enterprise environment, and operating environment, and has the characteristics of non-linearity, non-stationarity, and uncertainty (Haas and Yorio 2016). The existence of accidental factors, sudden factors, and random factors in the mining system has brought difficulties and challenges to the safety accident
prediction. At present, domestic and foreign scholars have carried out a lot of research on mining system accident prediction.

Barman and Choudhury (2019) optimized the parameter selection of SVM by Firefly algorithm (FA) to improve the prediction accuracy of SVM model. Huang and Zhou (2019) carried out the modeling and simulation of large mine accident data analysis and prediction based on decision tree-BP neural network. Luo et al. (2020) based on cusp mutation model and DS fusion evidence theory, the instability warning of underground rock mass in mining area was carried out. These methods often require a large amount of historical data and clear risk indicators. However, the development of China’s mining safety accident database is not perfect, the accurate statistical data that can be obtained is limited, and some data are greatly affected by random factors; there is a certain statistical error, unable to correctly identify its distribution. Although the occurrence of accidents is disorderly and irregular, the surface phenomenon of contingency of mining safety accident system is always dominated by its internal laws. A large number of statistical data show that the accident situation and its influencing factors are a closely linked whole, and this whole has relative stability and sustainability and is essentially causal and inevitable (Stemn et al. 2019). And although individual accidents have random uncertainty, but in a large sample of the overall still show some regularity, still can understand and grasp the future law and trend of the accident to a certain extent. Therefore, the detailed analysis of various influencing factors can be abandoned, and the trend of mining accidents can be predicted on the basis of scientific statistical data. As the situation of mine casualties changes over time, there will be a certain statistical law (Joshi et al. 2021). Therefore, time series prediction method can be used to predict mining accidents.

For the time series prediction of mining accidents, Wang et al. (2017) combined auto regressive moving average (ARMA) model with Markov prediction model to construct an ARMA-Markov combined prediction model that can take into account the randomness and trend of mining safety production accidents. Verma and Chaudhari (2017) developed a system based on accident prediction fuzzy inference method (FRA) to predict the possibility of mine accidents in India. However, these methods are subject to their own inherent defects, and the robustness of the prediction results is poor. Therefore, many scholars have made corresponding improvements and improvements on the basis of these models. Suhermi and Prastyo (2018) aiming at the problem of low prediction accuracy of ARMA model for nonlinear factors, an ARMA-ANN prediction model is established by introducing artificial neural network (ANN). Li et al. (2021b) decomposed the acoustic emission waveform into several intrinsic mode components (IMF) by Hilbert-Huang transform (HHT) and obtained the structural characteristics of coal and rock mass at different loading stages, which provided a basis for the monitoring of coal and rock dynamic disasters. Wu et al. (2020) smoothed the time series of mining safety production situation through empirical mode decomposition, and then modeled and predicted the stationary series separately according to ARMA. The prediction accuracy of these improved methods is improved. However, these models still have the following two problems. One is that only point prediction of time series can be obtained without considering the fuzziness of time series. There are many uncertain factors and statistical errors in mining safety production situation. If the statistical error of mining accidents and some random factors of mining accidents can be considered, the quantitative time series of mining safety production situation will be internalized, and the prediction results will be more reliable. Therefore, it is of more important value in practical application to accurately predict the change trend and change space of mining safety production situation and master its fluctuation range. Second, most of these methods do not consider the important factor in time series data — time granularity. If the granularity is too large, it is easy to lose important or even decisive information. On the contrary, if the particle size is too small, the prediction will fall into details, resulting in too large data changes and unpredictable. Different granularity of time data plays different roles in time series prediction (Lin and Li 2020). Appropriate granularity is a necessary condition for correctly understanding knowledge and rules.

Although in the past few decades, scholars have done a lot of research and exploration on the safety accident prediction of mining system, but the summary and analysis of the relevant literature of the accident prediction model for a long time shows that in the prediction of mining safety accident system, there are still the following difficulties: in the process of mining safety production situation prediction, less consideration of time granularity leads to the prediction into details or unpredictable; smooth processing is easy to lose the original sequence feature information or feature performance is not obvious; mining safety production situation is fuzzy and uncertain. In order to carry out the mining safety management work in a targeted and targeted manner, it is necessary to further explore the relevant factors and characteristics of the occurrence and development of the accident, excavate the internal information of the mining safety production situation time series, establish the most suitable system prediction model combined with the characteristics of the mining safety accident system, scientifically and accurately predict the occurrence and development trend of mining accidents, and reduce the probability of accidents, and conduct targeted research on these three problems.

In order to better excavate the regularity of mining safety production situation and reveal the characteristics of mining
Safety production situation, the best time granularity of mining safety production situation should be determined firstly, take the best time granularity as the window width. The mining safety production situation is granulated according to the best time granularity, and the information granules that can reflect the characteristics of the original data are obtained. Some statistical errors and uncertainty factors of non-coal mine safety production situation are reflected in the granulation interval. The error of mining safety production situation is mapped in three windowed time series of Low, R, and Up, highlighting the inherent characteristics of mining safety production situation time series.

The time series of mining safety production situation is a dynamic stochastic process, which has non-stationary characteristics and is difficult to accurately identify its distribution. The ensemble empirical mode decomposition can discrete the random non-stationary time series into the high-frequency signal sequence that represents the overall trend and the low-frequency signal sequence that represents the random disturbance. It can decompose the hidden information of the original time series, characterize the characteristics of the original time series more clearly, and preserve its dynamic characteristics. Then, the characteristics of the original sequence can be revealed by studying these components. Compared with differential operation and empirical mode decomposition, aggregated empirical mode decomposition can completely preserve the information of the original time series and can well suppress the phenomenon of modal aliasing. For the stationary sequence after ensemble empirical mode decomposition, the sample entropy is used to reconstruct it, which can take into account the correlation in the sub-sequence and better reflect the original characteristics of the non-coal mine safety production situation time series.

Qualitative concepts and uncertainty reasoning reflect human intelligence and are one of the most significant differences between human and machine at the same time. Cloud model is a model that combines randomness and fuzziness to realize the uncertain transformation between qualitative and quantitative knowledge. How to embody the soft reasoning ability in language thinking? How to reflect uncertainty in natural language, especially fuzziness and randomness? How to realize the conversion between qualitative and quantitative knowledge? How to express the qualitative knowledge expressed in natural language? Therefore, we propose a qualitative and quantitative uncertainty conversion model which called cloud model, to realize the uncertainty conversion between qualitative concepts and quantitative values based on cloud model.

Based on this, the analysis of time series characteristics based on mining safety production situation is studied. Firstly, the optimal time granularity of mining safety production situation is determined, and it is used as a window width. Then, the triangular fuzzy number is used to granulate the fuzzy information of mining safety production situation time series. Then, the windowed time series are smoothed by ensemble empirical mode decomposition and sample entropy reconstruction, which highlights the characteristics of the original sequence. According to the characteristics of the reconstructed sequence, the uncertainty prediction of the reconstructed sequence is carried out based on cloud theory, and the uncertainty prediction model of mining safety production situation is constructed. The development trend of mining safety production situation is predicted, which provides a novel idea for the prediction of mining safety production situation.

The rest of this paper is organized as follows: “Establishment of the model” section provides a detailed introduction to the modeling principle and process of the uncertainty prediction model. “Model application” section explains the source of the data and validates the model. The comparison of prediction accuracy of the models is performed in the “Model verification” section. “Future work” section discusses and elaborates the future research directions while the conclusions drawn from this research are summarized in the “Conclusions” section.

**Establishment of the model**

**Fuzzy information granulation of time series**

The accuracy of time series prediction is closely related to the inherent law of the original time series. Since there
is a certain statistical error in the statistics of accidents, the inherent law of time series will be weakened and the accuracy of the model will be affected. Therefore, considering the fuzzification of statistical accident information and the granulation of fuzzy information, the time series of mining safety production situation is mapped to triangular fuzzy information particles including low boundary value Low, median value $R$, and high boundary value $Up$, and the statistical error is reflected in the original time series of mining safety production situation.

The triangular fuzzy information particle is used to granulate the time series of mining safety production situation. Firstly, the original time series is divided into some sub-sequences according to a certain window $w$. The original time series window takes one season as a unit, namely, the window size is 3. Then fuzzy granulation is performed on each window to obtain three time series after windowing. The membership function of fuzzy particles after granulation is (Akram and Luqman 2020):

$$A(x, a, m, b) = \begin{cases} 
0 & x < a \\
\frac{x-a}{m-a} & a \leq x \leq m \\
\frac{b-x}{b-m} & m \leq x \leq b \\
0 & x > b 
\end{cases} \tag{1}$$

where $x$ is a variable in the domain. $a$, $m$, and $b$ are parameters, corresponding to the three variables Low, $R$, and $Up$ after window fuzzification. Low describes the minimum change of the fuzzy granulation window; $R$ describes the general average level of the fuzzy granulation window change; and $Up$ describes the maximum value of the fuzzy granulation window change.

**Decomposition and recombination of granulated sequences**

The time series of mining safety production situation is a dynamic stochastic process, which has non-stationary characteristics and is difficult to accurately identify its law of distribution. Aggregated empirical mode decomposition can decompose the fluctuations or trends in different scales of the signal step by step, and produces a series of data sequences with different characteristic scales, namely, the intrinsic mode function (IMF) (Stalnone et al. 2020). EMD realizes the decomposition of the original signal through the process of “sieving” and shows the changing trend of the original signal more clearly through transformation. The algorithm process is as follows:

**Step 1:** For a given signal, determine all extreme points. The upper and lower envelopes are obtained by connecting all the maximum and minimum points with cubic spline curves. Order:

$$h_i = X(t) - (E^+(t) + E^-(t))/2 \tag{2}$$

**Step 2:** Separate it and get a difference time series that removes the IMF1 component:

$$RES_1 = X(t) - IMF_1 \tag{3}$$

Repeat the screening process of step 1 for the new time series until the residual sequence of the $n$th order becomes a monotonic sequence and cannot continue to decompose the IMF components.

$$RES_n = RES_{n-1} - IMF_n \tag{4}$$

**Step 3:** The original signal in mathematics can be expressed as the sum of $n$ IMF components and the last residual RES:

$$X(t) = \sum_{i=1}^{n} IMF_i + RES_n \tag{5}$$

where IMF$_i$ represents the detail component. RES$_n$ represents the approximate component.

In the practical application of EMD, the characteristic components of different time scales are decomposed into a characteristic modal function component, or the same time scale component appears in different characteristic modal functions, then, modal aliasing. Ensemble empirical mode decomposition can effectively make up for the deficiency of EMD. It defines the mean value of a whole test as the real IMF component, and each component contains signal and white noise with limited amplitude, which can clearly separate each time scale. The process of EEMD is as follows:

1. Adding white noise signals to the target data;
2. Decompose the added white noise signal into IMF;
3. Each time adding different white noise sequence, repeated steps (1) and (2);
4. Take the mean value of each IMF as the final result.
Sample entropy recombination component

For multiple sub-sequences obtained by ensemble empirical mode decomposition, if they are separately built to predict, not only the calculation amount is greatly increased, but the correlation between the sub-sequences is ignored. The sample entropy algorithm is used to reconstruct the sub-sequences, which can better highlight the typical characteristics of similar sequences. Sample entropy reflects the complexity of time series. The more obvious the periodicity of time series is, the smaller the signal noise interference is, the smaller the sample entropy value is. On the contrary, the greater the signal noise interference is, the greater the sample entropy value is. This algorithm contains the characteristics of approximate entropy and is, the greater the sample entropy value is. This algorithm is, the greater the sample entropy value is.

The calculation process of sample entropy is as follows:

1. For the sequence after aggregated empirical mode decomposition, an n-dimensional vector group $X_m(1), ..., X_m(N - m + 1)$ is formed according to the sequence number, where $X_m(i) = \{x(i), x(i + 1), ..., x(i + m - 1)\}, 1 \le i \le N - m + 1$.

2. For a given $X_m(i)$, calculate the number of distances between $X_m(i)$ and $X_m(j)$ less than or equal to $r$, which is the number of template matching $B_m$, and calculate the ratio of $B_m$ to the total distance $N - m + 1$, $B_m(r)$:

$$B_m(r) = \frac{1}{N - m + 1} B_m$$

3. Define $B_m(r)$ as:

$$B_m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} B_m(r)$$

4. Increase the dimension $m$ to $m + 1$ dimension, repeat steps (1)–(3) for the $m + 1$ dimension vector to obtain $B_{m+1}(r)$.

5. The sample entropy of this sequence is:

$$SampEn (m, r, N) = -\ln \left[ \frac{B_{m+1}(r)}{B_m(r)} \right]$$

where $m$ is the reconstruction dimension, generally $m = 2$, $N$ is the data length, $r$ is similar tolerance, generally 0.1–0.25 times of standard deviation.

Prediction of granulation time series based on cloud theory

Time series prediction based on cloud theory is mainly based on the following two principles (Gu et al. 2021):

1. Fourier transform in physics can often deal with complex situations in one space simply in another.

2. Any probability distribution can be decomposed into several normal distribution combinations.

Time series analysis mainly studies the random dynamic data from the perspective of frequency domain and time domain and conducts uncertain data modeling for dynamic data. There are few studies on the prediction of time series based on cloud model. In this paper, the method of time series prediction based on cloud theory is mainly based on the normal cloud model generator. The prediction of granulation time series based on cloud theory is as follows.

Discretization of time series

The purpose of discretization of time series data is automatic generation of concept level. The common methods for discretization of time series data are mainly the equal-distance interval method and the equal-frequency interval method, that is, dividing the time series into several subintervals subjectively with equal width (or according to the frequency of data value, equal frequency). Both methods have strong subjective factors, which cannot reflect the uncertainty of extracting concepts from continuous time series data. Moreover, the concept tree generated by these methods is rigid, and the membership relationship between hierarchical concepts cannot be reflected, and the actual data distribution is not considered.

Fourier transform in physics can transform the time-domain space and frequency-domain space, and often simply handle the complex situation of one space in another space. Any probability distribution can be decomposed into several normal distribution combinations. Based on the above two points, the cloud transform algorithm is used to discretize the time series data.

Given the frequency distribution function $f(x)$ of time series, it is converted into the superposition of multiple normal cloud models according to a specific algorithm. The mathematical expression is:

$$F(x) \rightarrow \sum_{i=1}^{n} c_i \ast C(c_i \ast C(Ex_i, En_i, He_i))$$

where $c_i$ is the coefficient and $n$ is the number of discrete qualitative concepts generated.
Concept hierarchy

The cloud transformation method automatically generates a series of basic concepts represented by the cloud model according to the distribution of data values in the attribute domain and realizes the soft partition of the domain. In the process of knowledge discovery, these basic concepts can be regarded as the leaf nodes of the pan-concept tree for concept jump. The so-called concept leap is to raise the concept directly to the required concept granularity or concept level on the basis of the concept leaf node. Common types are user-defined jump, intelligent jump, and interactive jump. According to human cognitive psychological characteristics, generally 7 ± 2 concepts are the most appropriate.

Set two adjacent cloud models \( C_1(Ex_1, En_1, He_1), C_2(Ex_2, En_2, He_2) \), and \( Ex_1 < Ex_2 \). Let the merged cloud model be \( C(Ex, En, He) \), then:

\[
C = C_1 \cup C_2 \Leftrightarrow \begin{cases} 
Ex = \frac{2Ex_1Ex_2 + En_1 + En_2}{4} \\
En = \frac{2En_1En_2 + Ex_1 + Ex_2}{4} \\
He = \max(He_1, He_2)
\end{cases}
\]

(10)

Through the above methods, the basic concepts of mining safety production situation can be upgraded to the required conceptual level.

Time series set partitioning

When the concept is jumped to the appropriate granularity or level, the maximum value judgment method is used to determine the values of each time in the time series.

Given a time series \( X_t \), \( t \in N \), the concept set of a concept granularity corresponding to \( X_t \) is:

\[
C = \{C_1(Ex_1, En_1, He_1), C_2(Ex_2, En_2, He_2), \ldots, C_n(En_n, Ex_n, He_n)\}
\]

(12)

Then each \( x_t \) as the corresponding value to be determined, the idea of maximum determination method is shown in Fig. 1. The antecedent cloud generator \( CG_1, CG_2, \ldots, CG_n \) corresponds to each concept in \( C \), and the membership degree of \( x_t \) to each concept is calculated as \( \mu_1, \mu_2, \ldots, \mu_n \) respectively. If \( \mu_k = \max(\mu_k) \), \( k = 1, 2, \ldots, n \), then the concept \( C_i \) is the subordinate concept of \( x_t \). If there are two or more maximum membership degrees, one of them is randomly selected as the membership concept of \( x_t \). According to the definition of cloud, \( \mu_1, \mu_2, \ldots, \mu_n \) are a set of random variables with stable tendency. Therefore, \( x_t \) may belong to different concepts in overlapping regions under different conditions, which is consistent with people’s psychological cognitive characteristics.

Association rule mining based on cloud

The basic idea of cloud association knowledge mining is to use cloud to represent the "soft threshold" of support and confidence in association rule mining. The positive cloud generator is used to realize the conversion from qualitative linguistic value to quantitative value. Cloud transformation is used to divide the universe of discourse of numerical attributes into multiple qualitative concepts represented by clouds, so as to realize the discretization of numerical attributes. On this basis, Apriori algorithm is finally used to realize the mining of different types of association rules in the database. These rules are composed of the concepts represented by the cloud model. Due to the fuzziness and randomness of the cloud model, the obtained association rules are more suitable for human thinking. The support and confidence represented by the cloud model make the threshold no longer fixed, but because of the fuzziness of its unique boundary, more uncertain mining results can be obtained in the process of mining. Furthermore, it is easier to obtain association rules at different conceptual levels by means of concept jumping strategy. According to the requirements of association mining algorithm, this kind of conceptual sequence represented by normal cloud model can be converted into Boolean data table.

Uncertainty prediction

Based on concept, the qualitative knowledge is constructed into rule generator according to cloud theory, which is called uncertainty reasoning based on cloud model. With the help of normal cloud model to replace the division of the universe of quantitative attributes, the concept of normal cloud association rules on quantitative attributes is proposed, and the method of mining normal cloud association rules is given. Since the normal cloud model better softens the dividing boundary of the quantitative attribute universe, it can make
the mined normal cloud association rules and prediction results more easily understood.

If A then B, where A and B are concept cloud objects. This is a simple form of qualitative rules. A \(X\)-condition cloud generator and a \(Y\)-condition cloud generator are connected to a single-condition rule generator, as shown in Fig. 2. When inputting a specific value \(x_0\), \(CG_A\) will be excited to generate a stable random membership value \(\mu_i\), which will be used as the input of \(CG_B\). At the same time, \(CG_B\) will be excited to generate a cloud droplet \((x, \mu)\) is a random cloud droplet generated by the rule generator. Because of the uncertainty of the cloud model in the algorithm, the quantitative values \(x\) and \(\mu\) of this group of cloud droplets are not unique or determined. This illustrates the transmission of uncertainty in the reasoning process of cloud rule generator. Regular generator is more common: one-dimensional or multidimensional antecedent cloud (multidimensional structure usually through a number of one-dimensional antecedent cloud splicing) and one-dimensional consequent cloud. Time series prediction based on cloud model can generate a single-condition single-rule generator through the qualitative knowledge obtained by the final mining, and then construct a prediction inference engine of multiple qualitative rules to achieve time series prediction.

**Model application**

Taking the mortality rate of 10,000 people from mining accidents in mining industry across the country from January 2005 to June 2020 as the research object. The data comes from the statistics of domestic mining safety production accidents and the statistical analysis of mining employees on the “National Data” website. Based on the sample characteristics, the K-fold cross validation method is used to divide the training set and test set, taking into account the accuracy and calculation amount (Wu et al. 2021). One hundred seventy-five sets of data on the mortality rate of 10,000 people from mining accidents from January 2005 to September 2018 were used as the training set to build the model. The 21 sets of data on the death rate of 10,000 people from mining accidents from October 2018 to June 2020 are used as the test set to verify the reliability of the model. Fig. 3 shows the time series of the death rate of 10,000 people from mining safety accidents across the country from 2005 to 2020.

The time series of mining safety production situation is the waveform of time, and the time series of mining safety production situation has the characteristics of random fluctuation. China’s mining safety accident database development is not perfect, can get accurate statistical data that is limited, and has certain fuzziness. In order to solve the fuzziness of mining safety production situation, the fuzzy information of the recombinant sub-sequence is first granularized according to a certain window to obtain the information granules that can reflect the characteristics of the recombinant subsequence, and the fuzziness is reflected in the information interval (Fig. 4).

In the process of fuzzy information granulation, the selection of windowing width plays a crucial role in the results of fuzzy information granulation. Previous studies often determine the window width through subjective experience, which has strong subjectivity and will affect the actual prediction. In order to reduce the subjectivity of window width selection, the best time granularity of mining safety production situation is considered as the window width. According to the time series similarity of mining safety production situation with different

![Fig. 2](image)

![Fig. 3](image)
time granularity, the time series mining evaluation effect function $F(a_i \in A_T)$ is constructed. This function represents the accuracy of mining safety production situation prediction based on the extraction of time granularity $a_i$. Based on this function, the best time granularity of mining safety production situation can be found. The best time granularity determination method of mining safety production situation is as follows.

Assuming that the domain $U_T$ and its corresponding time granularity set $A_T$ are given, the optimal time granularity $a_i$ is defined as $\forall a_u \wedge a_u \neq a_i \wedge F(a_i) > F(a_u)$. The acquisition of optimal time granularity is of great significance for mining safety production situation prediction. In general, the best time granularity is neither atomic nor domain-based (Luo et al. 2019). Due to the complexity of mining accident data, the optimal time granularity is often impossible to obtain at one time, and the prediction accuracy reaches a certain degree before it has application significance. Based on the prediction accuracy threshold $\lambda$, an algorithm for solving the optimal time granularity from coarse to fine according to the time granularity is given.

Input: Time domain $U_T$, alternative $n$ equivalent partition time series based on time granular, time series data mining evaluation effect function $F(a_i \in A_T)$, prediction accuracy threshold $\lambda$.

Output: Optimal time granule shape OPT.

From $n$ partitions, according to $F(a_i \in A_T)$, a partition is selected to get the time granular set $T_i$ under $a_i$. The prediction effect under each $T_i$ is evaluated in turn. If the accuracy reaches the preset threshold, the algorithm ends. Otherwise, according to each time granule of the current $T_i$, the time granule expression sequence with the best $F(a_i \in A_T)$ effect is selected. Further, the time granule under the current granularity is tried to be divided, and the divided results are merged according to the precursor and successor relationship of the time granule.

We continue to evaluate the prediction effect under the current particle size until the accuracy reaches the preset threshold or is less than a certain time particle size, and obtain the optimal time particle size and its expression, denoted as OPT.

If $n = \{n_1, n_2, n_3\} = \{1, 2, 3\}$ respectively represents three optional methods of time granularity ($1$, $2$, $3$ represent the 1st, 2nd, and 3rd fractions of the given time granularity), then the next method for solving the optimal time granularity under a given $U_T$ is shown in Fig. 5.

Taking every 3 months (1 quarter) as an information granulation window, the mining safety production situation time series fuzzy granulation is converted into $Low$, $R$, and $Up$ three parameters. The fuzzy information granulation results are shown in Fig. 4, where $Low$ represents the minimum monthly change of the mortality rate per 10,000 people in the mining safety production situation, $R$ represents the general average monthly change of the mortality rate per 10,000 people in the mining safety production situation, and $Up$ represents the maximum monthly change of the mortality rate per 10,000 people in the mining safety production situation. It can be seen from Fig. 4 that the three parameters of $Low$, $R$, and $Up$ after fuzzy information granulation are non-stationary. In order to solve this problem, the original data are decomposed by aggregated empirical mode decomposition to obtain multiple intrinsic mode functions and a monotonic sequence that can reflect the change trend of original data, which highlights the general law of mining safety production situation. The mining safety production situation curve is shown in Fig. 6 after aggregated empirical mode decomposition.

It can be seen from Fig. 6 that at each decomposition level, the trend sequence represents the main fluctuation/ fluctuation of the mining safety production situation, and the detail sequence mainly represents the small-scale fluctuation with large frequency. With the increase of decomposition level, the details are gradually extracted. After 6-level aggregate empirical mode decomposition, the trend has lost a lot of detail information contained in the original mining safety production situation curve, and the detail part has appeared large-scale fluctuations only at low frequencies. This shows that the six decomposition levels can meet the requirements of mining safety production situation curve decomposition, and the original mining safety production situation curve can be decomposed into low-frequency detail sequence and high-frequency trend sequence.
The complexity of the original time series can be reduced by decomposing the curve of the granulation parameter series into several sub-sequences. If each sub-sequence is modeled separately, not only the computation is increased, but also the correlation between each sub-sequence is ignored. Therefore, the sample entropy is used to determine the complexity of each component and reconstruct it, which highlights the characteristics of similar sequences. The sample entropy distribution of sub-sequences is shown in Fig. 7. The sample entropy reflects the complexity of the sequence. The higher the auto-correlation of the sequence is, the smaller the sample entropy is. On the contrary, the sample entropy is larger. The sample entropy values of the given sequence and sub-sequence are calculated, and the sample entropy value is significantly lower than that of the sub-sequence of the given sequence, which can constitute the trend component. The sample entropy is significantly higher than the sub-sequence of a given sequence can constitute a random component. Sub-sequences whose sample entropy values are within the range of sample entropy threshold $\lambda$ of a given sequence constitute the detail components. The value of $\lambda$ is determined according to the sample entropy distribution of the sub-sequence, $\lambda = 0.40$. The recombinant sequence is shown in Fig. 8.

Analysis of the characteristics of Fig. 8 shows that trend component, detail component, and random component have their own typical characteristics. The trend component roughly reflects the overall fluctuation trend of the original
mining safety production situation. The detail component can represent the detail fluctuation of mining safety production situation. The random component represents the fluctuation caused by other random factors. The above classification also basically meets the composition of mining safety production situation.

Some data of mining safety production situation fluctuate greatly under the influence of random factors, which have strong randomness and cannot correctly identify its distribution law. To solve this problem, the cloud model is used to predict the reconstructed sub-sequence. The cloud model can realize the mutual transformation from qualitative concept to quantitative value, characterize the randomness and fuzziness in time series and the correlation between them, realize the qualitative concept representation of the predicted quantitative value, and reduce the influence of randomness and fuzziness of mining safety production situation.

In fact, cloud transformation is a process of concept induction and learning from the perspective of data mining, and the results are often not unique. Two points that cloud transformation need to consider frequently are as follows: the higher the frequency is, the more concentrated the data distribution is, and it can be regarded as the convergence center of concept; extraction concept is an iterative process. Before the next search, the data part represented by the last extracted concept is removed from the original frequency distribution. Taking the detail sequence as an example, the discrete results of the window detail sequence are shown in Fig. 9. Then, according to the method in the “Concept hierarchy” section, the basic concept of discrete mining safety production situation can be upgraded to the required conceptual level, as shown in Fig. 10.

In Fig. 10, the cloud drop represents the actual data in the database, the x-axis represents the domain of attribute $i$, and the y-axis represents the membership of each cloud drop to its respective concepts. In this way, through the cloud transformation, an arbitrary irregular distribution is mathematically transformed according to a certain law, so that it becomes the superposition of several clouds with different sizes. The more superimposed clouds are, the more they can reflect the actual data distribution, and the smaller the error is. It can be seen from the experiment that the definition domain of attribute $i$ in the database can be divided into four parts: low, lower, higher, and high.
The digital characteristics of cloud models in concepts after concept jump are shown in Table 1. For the frequency distribution function \( f(x) \) of the maximum and minimum, the cloud transformation algorithm is used to extract the concept. The new data distribution function \( f(x)' \) is obtained by subtraction the fitted distribution function \( f(x) \) from the frequency distribution function \( f(x) \). On this basis, the above steps are repeated until the fitting error meets the threshold requirements of the cloud transformation algorithm. According to the obtained eigenvalues \( E_x \) and \( E_n \) of the cloud model and the range represented by the cloud model, \( H_e \) was calculated according to the reverse cloud algorithm, and finally the qualitative concept represented by the cloud model was extracted from the frequency distribution function \( f(x) \). The maximal judgment method is used to select a concept with a higher degree of membership as its membership concept. By taking the mortality rate of all mining safety accidents as the input value, the membership degree \( \mu_i \) of each value for each concept can be obtained, from which the maximum membership degree is selected. The corresponding concept is the concept belonging to the value. After the soft division of attributes, the amount of data is greatly reduced, and the internal relationship between attributes is also highlighted.

For the division of time, drawing on the method of 24 solar terms in China, see Fig. 11. In this paper, the 12 months of a year are divided into four cloud models: “spring (January to March),” “summer (April to June),” “autumn...
“(July to September),” and “winter (October to December).” Their entropy is 0.50 and their super entropy is 0.03.

Then, the maximum judgment method is used to determine the membership concept of each data value including time in the time series. The determined results are taken as the input data set, and the Apriori algorithm based on cloud model is used for association rule mining. With the concept as the basic representation, the mined qualitative knowledge is constructed into a rule generator according to the cloud theory, namely, the so-called uncertainty reasoning based on cloud model. With the help of normal cloud model to replace the division of the universe of quantitative attributes, the concept of normal cloud association rules on quantitative attributes is proposed, and the method of mining normal cloud association rules is given. Since the normal cloud model better softens the dividing boundary of the quantitative attribute universe, it can make the mined normal cloud association rules and prediction results more easily understood.

**Model verification**

**Model accuracy test**

The predicted values of windowed time series can be obtained by combining the predicted values of reconstructed detail series, random series, and trend series. Using the model established by 165 groups of samples in the training set, 21 groups of data from October 2018 to June 2020 in the test set are used as the test set. Based on the randomness and fuzziness of cloud reasoning, 2000 cloud droplets are generated by each input, and their expected values are used as the output of the prediction results. The prediction results and relative errors of the detailed sequence and random sequence are shown in Table 2.

The confidence interval is 95%, and the prediction residual of the granulation sequence is tested by normal distribution. Fig. 12 shows that the prediction residual of the granulation sequence is evenly distributed in the 95% confidence interval, so the prediction residual of the granulation sequence obeys the normal distribution. The model extracts the internal information of the granulation sequence well, indicating that the prediction error is a comprehensive reflection of various random phenomena, and the model has high prediction accuracy.

**Comparative analysis of models**

The mining safety production situation interval prediction model based on fuzzy information granulation is compared with ARIMA, wavelet neural network, and particle swarm optimization (PSO)-SVM prediction model. The results are shown in Table 3. The prediction n effect of each model is shown in Fig. 13.

It can be seen from Table 3 and Fig. 13 that the relative error of uncertainty prediction model based on cloud theory is significantly lower than that of ARIMA model, wavelet neural network model, and PSO-SVM model. The relative error fluctuation of the uncertainty prediction model based on cloud theory is relatively stable, indicating that the model has better prediction stability. The prediction results of ARIMA and wavelet neural network models fluctuate greatly beyond the interval range, because the ARIMA model will lose the information of the original sequence in the process of smooth processing of the original sequence by difference, and the more the difference times are, the more the information will be lost. EEMD can effectively retain the information of the original time series. Ahmadi et al. (2020) also found a similar phenomenon in the study. Differential processing will lose the internal information of the original time series, and it is difficult to mine this part of...
the information lost in the prediction, resulting in a decline in the prediction accuracy. Therefore, the prediction accuracy of stationary time series after EEMD decomposition is significantly higher than that after difference processing, and the prediction stability of the model is also high. Compared with wavelet neural network, wavelet analysis is developed for the shortcomings of Fourier transform, which improves the shortcoming of Fourier transform that has no resolution ability in time domain. However, compared with EEMD, the selection of wavelet basis for wavelet neural network prediction has strong subjectivity, which affects the prediction accuracy of the model, and similar phenomena are also found in the study. Moreover, the neural network often needs a large number of training samples, and the uneven distribution of fewer training samples will lead to an increase in error, while the time series samples of mining safety production situation are limited. When the data are few, the performance of deep learning algorithm is not good. This is because deep learning algorithm requires a lot of sample data, so the prediction effect of wavelet neural network is not ideal.

Compared with the support vector machine model optimized by PSO algorithm, the predicted value of PSO-SVM model is generally in the interval range, but the fitting
effect with the original time series is not good. It can only predict the overall trend, and the prediction accuracy of the detailed part is not high. It can be seen that the single-point prediction model is difficult to accurately predict the time series of nonlinear and non-stationary mining safety production situation. And in the short-term prediction, the average relative error of the particle swarm optimization support vector machine model is small. When the long-term prediction is carried out, the fitting effect of the particle swarm optimization support vector machine model is significantly lower than that of the uncertainty prediction model based on cloud theory. Although the machine learning method of particle swarm optimization support vector machine has good prediction effect for short-term prediction of small samples and nonlinear sequences, due to the lack of consideration of sequence timing, the prediction accuracy in medium- and long-term prediction is lower than that of uncertainty prediction model based on cloud theory. From the average relative error, the average relative error of the PSO-SVM model is close to the uncertainty prediction model based on cloud theory when the prediction step is 5, and the average relative error is 7.538 % and 9.299 %, respectively. When the prediction step is more than 5, the fitting effect of the PSO-SVM model is significantly lower than that of the uncertainty prediction model based on cloud theory. And in the short-term prediction, the fluctuation is large, so the stability of the short-term prediction is not as good as the uncertainty prediction model based on cloud theory. When using SVM built on wavelet transform...
model for prediction, Dhiman et al. (2019) also found the subjectivity of wavelet transform and the low and unstable accuracy of SVM in medium- and long-term prediction.

According to Fig. 13, it can be intuitively seen that the prediction results of the uncertainty prediction model based on cloud theory are consistent with the real change trend. The uncertainty prediction based on cloud theory not only considers the trend characteristics of mining safety production situation fluctuation, but also considers the periodic characteristics of mining safety production situation fluctuation, that is, the influence of time series and nonlinear factors of mining safety production situation is considered in long-term prediction. Therefore, good prediction results are obtained in long-term mining safety production situation prediction. It can be fully proved that the uncertainty prediction model based on cloud theory has good practical guiding significance in the prediction of mining safety production, which is conducive to helping safety decision makers accurately grasp the changes in mining safety production and make correct safety decisions.

Future work

The relevant research results have clearly shown that the proposed method can obtain high prediction accuracy in the prediction of mining safety production situation time series and can effectively solve the three core problems of time granularity selection, non-stationarity, and uncertainty in time series prediction. However, in order to obtain better prediction results in future prediction, the next step will continue to study in the following two aspects:

(1) It can be seen from the relevant results that the feature extraction method of time series based on ensemble empirical mode decomposition-sample entropy can better retain the information of the original time series and highlight the characteristics of the original sequence. EEMD keeps the mean value of IMF within the normal dynamic filtering range, so that the white noise is offset against each other, so as to suppress the mode mixing phenomenon and further improve the prediction accuracy. According to the chaotic prediction theory, if the maximum Lyapunov exponent of the system is $\lambda$, the maximum predictable time is about $1/\lambda$. The response frequency component reconstructed by ensemble empirical mode decomposition-sample entropy decomposition is only a part of the original signal. At this time, the maximum Lyapunov exponent of each decomposed signal is less than that of the original signal, which is of great significance to improve the maximum predictable time. Future research can be carried out in the direction of improving the maximum predictable time.

(2) The accuracy of time series prediction is closely related to the inherent law of the original time series. Due to the statistical error of accident statistics, the inherent law of time series will be weakened, which affects the accuracy of the model. In the time series prediction based on cloud theory, a certain range can be obtained according to the $3En$ principle. The confidence interval $[Ex - 3En, Ex + 3En]$ can be considered as a fuzzy interval to further improve the reliability of the prediction results. Therefore, on the basis of existing research, interval prediction based on cloud theory can be carried out, and the prediction error is reflected in the prediction interval.

Conclusions

(1) Time series mining evaluation effect function is introduced to evaluate the prediction effect based on different time particles in a given time domain to determine the optimal time granularity. The optimal time granularity is taken as the window width of fuzzy information granulation, and then the time series of non-coal mine safety production situation are mapped into fuzzy information particle series including low boundary value $Low$, median value $R$, and high boundary value $Up$ by triangular fuzzy number. The fuzzy information granulation method with the best time granularity as the window width can intervalize the quantitative mining safety production situation time series and take into account the statistical error of mining accidents and some random factors. The prediction results are more reliable and have more reference significance in practical application.

(2) Based on EEMD, the fluctuation or change trend in random non-stationary granulation time series is decomposed step by step, and a series of stationary sequences with different characteristic scales are generated. The characteristics of the original time series are characterized more clearly, and the dynamic characteristics are preserved. The stability of each sub-sequence is more obvious, and it also shows stronger regularity. Aiming at the correlation between sub-sequences, the complexity of each component after EEMD is analyzed according to the sample entropy theory, and the components after EEMD are reconstructed according to the sample entropy value. A time series feature extraction method based on ensemble empirical mode decomposition and sample entropy is proposed. The feature extraction of time series based on ensemble empirical mode decomposition-sample entropy highlights the characteristics of similar series, which is more conducive to grasping
various local characteristics of mining safety production situation.

(3) According to the characteristics of each component after the feature extraction of the time series of ensemble empirical mode decomposition-sample entropy, the random dynamic data are studied from the perspective of frequency domain and time domain, and the uncertain data are modeled for the dynamic data. Through cloud transformation, concept jump, time series set partition, and association rule mining, the original numerical sequence is converted into an abstract conceptual sequence represented by a certain time and information granularity. The average relative error of the predicted value of Low parameter based on cloud theory is 10.83457%, the average relative error of the predicted value of R parameter is 20.20790%, and the average relative error of the predicted value of Up parameter is 0.65197%. The prediction accuracy of Low and Up parameter values is significantly higher than that of R value, indicating that the upper and lower boundary values after fuzzy granulation can accurately describe the variation law of the time series of mining safety production. When the actual value is higher than the high boundary value expressed by cloud theory, it indicates that practical measures should be taken to reduce the occurrence of mining safety accidents.

Acknowledgements The authors would like to thank editor and reviewers cordially for their positive and constructive suggestions.

Author contribution WM is the executor of the modeling design and theoretical analysis of this study and is responsible for the writing of the first draft. YY, HN, and WQ completed data analysis and guided the writing and revision of the paper; TW participates in the modeling process and results analysis. All authors read and approved the final manuscript.

Funding This work is supported by the National Natural Science Foundation of China (Grant No. 51704213) and the Key R & D projects in Hubei Province, China (No. 2020BCA082).

Data availability The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

Akram M, Luqman A (2020) Granulation of ecological networks under fuzzy soft environment. Soft Comput 24(16):11867–11892
Ahmadi R, Shahrabi J, Aminshahidy B (2020) Forecasting multiple-well flow rates using a novel space-time modeling approach. J Petrol Sci Eng 191. https://doi.org/10.1016/j.petrol.2020.107027
Barmann M, Choudhury NBD (2019) Season specific approach for short-term load forecasting based on hybrid FA-SVM and similarity concept. Energy 174:886–896
Bauer A, Zülfie M, Herbst N, Zehe A, Hotho A, Kounov S (2020) Time series forecasting for self-aware systems. P IEEE 108(7):1068–1093
Carvalho FP (2017) Mining industry and sustainable development: time for change. Food Energy Secur 6(2):61–77
Dhiman HS, Deb D, Guerrero JM (2019) Hybrid machine intelligent SVR variants for wind forecasting and ramp events. Renew Sust Energy Rev 108:369–379
Gu B, Zhang T, Meng H, Zhang J (2021) Short-term forecasting and uncertainty analysis of wind power based on long short-term memory, cloud model and non-parametric kernel density estimation. Renew Energy 164:687–708
Haas EJ, Yorio P (2016) Exploring the state of health and safety management system performance measurement in mining organizations. Saf Sci 83:48–58
Huang Y, Zhou QY (2019) Mine accident prediction and analysis based on multimedia big data. Multimed Tools Appl. https://doi.org/10.1007/s11042-019-7175-6
Hassen OA, Darwish SM, Abu NA, Abidin ZZ (2020) Application of cloud model in qualitative forecasting for stock market trends. Entropy 22(9). https://doi.org/10.3390/e22090991
Joshi D, Sharma I, Gupta S, Singh TG, Dhiman S, Prashar A, Gulati M, Kumar B, Vishwas S, Chellappan DK (2021) A global comparison of implementation and effectiveness of materiovigilance program: overview of regulations. Environ Sci Pollut R 28(42):59608–59629
Li W, Ye YC, Wang QH, Wang XH, Hu NY (2020) Fuzzy risk prediction of roof fall and rib spalling: based on FFTA–DFCE and risk matrix methods. Environ Sci Pollut Res 27(8):8535–8547
Li XL, Cao ZY, Xu YL (2020) Characteristics and trends of coal mine safety development. Energy Source Part A. https://doi.org/10.1080/15567036.2020.1852339
Li XL, Chen SJ, Liu SM, Li ZH (2021) AE waveform characteristics of rock mass under uniaxial loading based on Hilbert-Huang transform. J Cent S Univ 28(6):1843–1856
Li XL, Chen SJ, Zhang QM, Gao X, Feng F (2021) Research on theory, simulation and measurement of stress behavior under regenerated roof condition. Geomech Eng 26(1):49–61
Lin YD, Li RM (2020) Real-time traffic accidents post-impact prediction: based on crowdsourcing data. Accid Anal Prev 145. https://doi.org/10.1016/j.aap.2020.105696
Luo C, Tan CH, Zheng YJ (2019) Long-term prediction of time series based on stepwise linear division algorithm and time-variant zonary fuzzy information granules. Int J Approx Reason 108:38–61
Luo QZ, Li YY, Qin YG, Wen L (2020) A method developed for early warning of underground rock mass instability in mining area based on cusp catastrophe model and DS fusion evidence theory. Chin J Geo Hazard Control 31(5):60–69 [In Chinese] https://doi.org/10.16031/j.cnki.issn.1003-8035.2020.05.09
Moureñas D, Artemyev AV, Zhang XJ (2020) Dynamical properties of peak and time-integrated geomagnetic events inferred from sample entropy. J Geophys Res-Space 125(2). https://doi.org/10.1029/2019JA027599
Suhermi N, Prastyo DD (2018) Roll motion prediction using a hybrid deep learning and ARIMA model. Procedia Comput Sci 144:251–258
Stallone A, Cicone A, Materassi M (2020) New insights and best practices for the successful use of empirical mode decomposition, iterative filtering and derived algorithms. Sci Rep-UK 10(1) https://doi.org/10.1038/s41598-020-72193-2
Stenn E, Botinger C, Cliff D, Hassall ME (2019) Examining the relationship between safety culture maturity and safety performance of the mining industry. Saf Sci 113:345–355
Verma S, Chaudhari S (2017) Safety of workers in Indian mines: study, analysis, and prediction. Saf Health Work 8(3):267–275
Wang YL, Li YM, Li C (2017) Prediction of coal mine gas accidents based on time series Markov model. Chin Min 26(12):179–183 ([in Chinese])
Wu ML, Ye YC, Hu NY, Wang QH, Li W, Jiang HM (2021) Interval prediction of mining work safety situation based on fuzzy information granulation. Chin Saf Sci J 31(09):119–127 ([in Chinese])
Wu ML, Ye YC, Hu NY, Wang QH, Jiang HM, Li W (2020) EMD-GM-ARMA model for mining safety production situation prediction. Complexity, 2020. https://doi.org/10.1155/2020/1341047

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.