

ELECTROENCEPHALOGRAM SIGNAL PREPROCESSING METHOD, SYSTEM AND TERMINAL DEVICE

TECHNICAL FIELD

5 **[0001]** The present disclosure relates to the field of electroencephalogram data preprocessing technologies, and particularly relates to an electroencephalogram signal preprocessing method, system and terminal device.

BACKGROUND

10 **[0002]** Electroencephalogram (EEG) data mainly refer to EEG waves. EEG waves mainly appear in the temporal and occipital lobes. For example, δ rhythm can be observed in the deep-sleep people, infants, and patients with severe organic brain disorders, the waveform is a sinusoidal wave of 30-50 μ V, and the amplitude decreases along the occipital-parietal-frontal-temporal direction with a significant amplitude modulation phenomenon under normal circumstances. The δ rhythm is related to visual activities and presents the largest quantity and
15 the highest amplitude when people are awake, quiet and with their eyes closed. When people fall asleep, α rhythm completely disappears. Factors such as the frequency, amplitude, and spatial distribution of the α rhythm are important indicators reflecting the functional state of the brain. With the maturation of the brain or increase in age, the quantity of the α rhythm gradually increases and its frequency gradually rises. The α rhythm tends to be stable in
20 adulthood, and in old age gradually slows down. In the sleep state, the background rhythm in the light sleep stage is 5-6Hz (θ wave), and the background rhythm in the deep sleep stage is 2-3Hz (δ wave).

25 **[0003]** The original EEG data is a finite discrete time series data composed of many sample points. The number of sample points is determined by the sampling rate. For example, if the sampling rate is 1000Hz, then there are 1000 data sample points per second.

[0004] Each sample point data represents the amplitude of the EEG waves, which is called

voltage value in physics, and the unit is volt (V). Since EEG signals are usually weak, the more commonly used unit is microvolt (μV).

[0005] The process of EEG data preprocessing refers to performing a series of processing on the original EEG data, so as to extract EEG activities from noise for subsequent EEG data analysis.

[0006] In the related process of EEG data preprocessing, it is often necessary for professionals to conduct specialized programming processing for the EEG signal preprocessing. There is a lack of an EEG data preprocessing system that can integrally adapt to the preprocessing requirements of EEG data under different circumstances..

SUMMARY

[0007] In view of this, embodiments of the present disclosure provide an electroencephalogram (EEG) signal preprocessing method, system and terminal device to eliminate or improve one or more defects existing in the related art.

[0008] A first aspect of the present disclosure provides an electroencephalogram (EEG) signal preprocessing method, and the method includes the following steps: receiving original EEG signals collected by an EEG collection device, wherein a data structure of the original EEG signals corresponds to pre-positioned EEG signal channel locations, and each channel corresponds to a coordinate location on a scalp of a subject; performing artifact removal operation and noise removal operation on the collected original EEG signals to obtain pure EEG signals; obtaining event marks of the original EEG signals, and dividing the pure EEG signals into multiple segments, wherein each of the multiple segments contains EEG signals within a preset length of time range before and after an event occurs, and each of the event marks refers to one event; and obtaining, for the segments divided from the pure EEG signals, preset analysis indicators to be extracted, respectively performing time-domain analysis to extract time-domain analysis indicators and respectively performing frequency-domain analysis to extract frequency-domain analysis indicators, and taking the extracted time-domain analysis indicators and frequency-domain analysis indicators as output results of the EEG signal preprocessing.

[0009] In some embodiments of the present disclosure, for the received original EEG signals collected by the EEG collection device, the method further comprises: deleting original EEG signals from bilateral mastoid point locations on the scalp of the subject and/or from electrooculogram channels.

5 **[0010]** In some embodiments of the present disclosure, subsequent to the step of performing artifact removal operation and noise removal operation on the collected original EEG signals, the method further comprises: obtaining target features to be analyzed of EEG data, and selecting a reference electrode of the EEG collection device and a reference electrode calculation method based on the target features to be analyzed, wherein the reference electrode
10 calculation method comprises average reference, bilateral mastoid reference and chain reference.

[0011] In some embodiments of the present disclosure, the method further comprises: during the process of receiving the original EEG signals, calculating an amount of data of the original EEG signals collected by the EEG collection device received within a unit time, and
15 sending a sampling rate reduction instruction to the EEG collection device when the amount of data exceeds a preset threshold, so as to load pressure of receiving the original EEG signals.

[0012] In some embodiments of the present disclosure, the method further comprises: when an abnormal channel with EEG signal quality lower than a preset threshold or with missing EEG signals is detected, selecting EEG signals from a channel at an adjacent location of the
20 abnormal channel to interpolate and generate a new EEG signal, and using the new EEG signal as the EEG signal from the abnormal channel.

[0013] In some embodiments of the present disclosure, the method further comprises: screening the segments divided from the pure EEG signals based on a preset threshold, and finding and deleting segments containing artifacts or outliers beyond the preset threshold.

25 **[0014]** In some embodiments of the present disclosure, the artifact removal operation is based on blind source separation technology, and the performing artifact removal operation on the collected original EEG signals comprises: decomposing the original EEG signals into multiple independent artifacts or neurons based on the blind source separation technology, using independent component analysis (ICA) technology to identify independent variation
30 sources in the EEG signals, and removing the identified artifacts and/or independent variation

sources.

[0015] In some embodiments of the present disclosure, the noise removal operation is based on canonical correlation analysis technology, the performing noise removal operation on the collected original EEG signals comprises: using autocorrelation in a given time series to describe signal components in the original EEG signals, finding the highest linear correlation and autocorrelation among the original EEG signals based on the canonical correlation analysis technology, and distinguishing and removing noise from EEG signals based on autocorrelation lower than a threshold.

[0016] In some embodiments of the present disclosure, the noise removal operation is based on filtering processing and mathematical algorithm to decompose EEG signals and noise, wherein the filtering processing comprises one or more of high-pass filtering, low-pass filtering, band-pass filtering and notch filtering.

[0017] In some embodiments of the present disclosure, the artifact removal operation and noise removal operation are based on a deep learning model, the performing artifact removal operation and noise removal operation on the collected original EEG signals comprises: using EEG data containing artifacts and noise as a training set to train the deep learning model, and using the trained deep learning model for automatic detection and automatic removal of artifacts and noise.

[0018] In some embodiments of the present disclosure, the performing artifact removal operation and noise removal operation on the collected original EEG signals further comprises: using any one of signal noise separation (SNS) processing, random sample consensus (RANSAC) processing and cross-validation processing to perform artifact removal operation and noise removal operation on the collected original EEG signals.

[0019] In some embodiments of the present disclosure, the time-domain analysis indicators comprise one or more of average voltage, peak voltage, peak-to-peak, standard deviation, root mean square and event-related potential, and the frequency-domain analysis indicators comprise one or more of spectral density, frequency band energy, frequency band ratio, spectral peak, spectral entropy, spectral power density and power spectrum.

[0020] A second aspect of the present disclosure provides an EEG signal preprocessing system, including a processor and a memory, computer instructions are stored in the memory,

and the processor is configured to execute the computer instructions stored in the memory, and when the computer instructions are executed by the processor, the system implements the steps of the method in any of the above embodiments.

[0021] A third aspect of the present disclosure provides a terminal device, and the terminal
5 device is configured to implement the steps of the method in any of the above embodiments.

[0022] A forth aspect of the present disclosure provides a computer-readable storage medium, on which a computer program is stored, and when the program is executed by a processor, the steps of the method in any of the above embodiments are implemented.

[0023] The EEG signal preprocessing method, system and terminal device according to the
10 present disclosure can perform artifact removal and noise removal on the collected EEG signals to obtain relatively pure EEG data, the pure EEG data is divided into multiple segments, and time-domain analysis and frequency-domain analysis are performed on each segment, thereby completing the EEG signal preprocessing. The EEG signal preprocessing method according to the present disclosure can integrally adapt to the preprocessing requirements of EEG data under
15 different circumstances, and can improve the processing efficiency of removing noise and artifacts from the original EEG data, thereby improving the quality of EEG data and retaining the frequency bands and events of interest.

[0024] The additional advantages, purposes, and features of the present disclosure will be partially elaborated in the following description, and will partially become obvious to those
20 ordinary skilled in the art after studying the following text, or can be learned according to the practice of the present disclosure. The purposes and other advantages of the present disclosure can be realized and obtained through the structures specifically pointed out in the specification and the accompanying drawings.

[0025] Those skilled in the art will understand that the purposes and advantages that can
25 be achieved by the present disclosure are not limited to the above specific descriptions, and the above and other purposes that the present disclosure can achieve will be more clearly understood based on the following detailed description.

BRIEF DESCRIPTION OF DRAWINGS

[0026] The accompanying drawings described herein are provided to further understand the present disclosure, which constitute a part of the present disclosure, and do not constitute any limitation to the present disclosure. In the drawings:

5 [0027] FIG. 1 shows an electroencephalogram signal preprocessing method according to an embodiment of the present disclosure.

[0028] FIG. 2 shows an example diagram of artifact removal operation and noise removal operation according to an embodiment of the present disclosure.

DESCRIPTION OF EMBODIMENTS

10 [0029] To make the purpose, technical solutions and advantages of the present disclosure clearer, a further detailed description of the present disclosure is provided with reference to the following embodiments and drawings. Here, the explanations of the present disclosure are used to explain the present disclosure, but do not limit the present disclosure.

[0030] It should also be noted that in order to avoid obscuring the present disclosure due to unnecessary details, only the structures and/or processing steps closely related to the solutions of the present disclosure are shown in the drawings, and other details that are not significant to the present disclosure are omitted.

[0031] It should be emphasized that the term "include/comprise" when used in the context refers to the existence of features, elements, steps or components, but does not exclude the existence or addition of one or more other features, elements, steps or components.

[0032] It should also be noted that, unless otherwise specified, the term "connection" in the context refers to direct connection, or indicate indirect connection with an intermediate component.

[0033] Below, embodiments of the present disclosure will be described with reference to the drawings. In the drawings, the same reference signs represent the same or similar components, or the same or similar steps.

[0034] In order to solve the problems existing in the related EEG signal preprocessing methods, the present disclosure provides an EEG signal preprocessing method, system and

terminal device, which improves the processing efficiency of removing noise and artifacts from the original EEG data. Electroencephalography (EEG) data, also known as electroencephalogram, is a method of recording brain activities using electrophysiological indicators.

5 **[0035]** FIG. 1 shows an EEG signal preprocessing method according to an embodiment of the present disclosure. The method includes the following steps:

[0036] Step S110, receiving original EEG signals collected by an EEG collection device, data structure of the original EEG signals corresponds to pre-positioned EEG signal channel locations, and each channel corresponds to a coordinate location on a scalp of a subject.

10 **[0037]** In some embodiments, positioning channel locations refers to allocating a coordinate position on the scalp for each channel based on the electrode arrangement system, so as to facilitate spatial analysis and visualization. Positioning the EEG signal channel locations is mainly used to determine the locations of the electrodes on the scalp. In neuroelectrophysiological experiments such as electroencephalogram (EEG) or event-related
15 potential (ERP) experiments, multiple electrodes usually need to be placed on the scalp to record the electrical activities of the brain. The locations of these electrodes need to be precisely positioned to accurately measure and record the electrical activities of the brain. The electrode arrangement system of the EEG signals usually adopts the standard electrode placement manner stipulated by the International Federation of Clinical Neurophysiology, namely the 10-
20 20 system. In the 10-20 system, the distance from the midpoint of the frontal pole to the nasion and the distance from the occipital point to the external occipital protuberance each account for 10% of the total length of the connection line, and the remaining points are separated by 20% of the total length of this connection line. Therefore, it is named as the 10-20 system. By positioning the EEG signal channel locations, the specific location of each electrode on the
25 scalp may be determined to ensure the accuracy and reliability of the experimental results. This is of great significance for researching issues in the field of neuroscience such as brain function, cognitive processes, emotions, and attention.

[0038] Step S120: performing artifact removal operation and noise removal operation on the collected original EEG signals to obtain pure EEG signals.

30 **[0039]** Step S130: obtaining event marks of the original EEG signals, and dividing the pure

EEG signals into multiple segments, each of the multiple segments contains EEG signals within a preset length of time range before and after an event occurs, and each of the event marks refers to one event.

[0040] In some embodiments, segmentation refers to dividing a long segment of continuous data into segments based on the event marks in the experiment. Each segment contains data within a certain time range before and after an event occurs.

[0041] Step S140: for the segments divided from the pure EEG signals, obtaining preset analysis indicators to be extracted, respectively performing time-domain analysis to extract time-domain analysis indicators and respectively performing frequency-domain analysis to extract frequency-domain analysis indicators, and taking the extracted time-domain analysis indicators and frequency-domain analysis indicators as output results of EEG signal preprocessing.

[0042] In embodiments of the present disclosure, the processing efficiency of removing noise and artifacts from the original EEG data can be improved, thereby improving the quality of EEG data and retaining the frequency bands and events of interest.

[0043] In an embodiment of the present disclosure, for the received original EEG signals collected by the EEG collection device, the method further includes: deleting original EEG signals from bilateral mastoid point locations on the scalp of the subject and/or from electrooculogram channels. Specifically, according to the purpose and design of the experiment, some unnecessary channels may be removed, such as bilateral mastoid points and electrooculogram channels, etc. combining with other multimodal biological data, artifact data in EEG may be detected and distinguished. For example, combining with electrooculogram and eye movement measurements, blinks and eye movements in EEG signals may be identified. Combining with electrocardiogram signal measurements, electrocardiogram artifacts in EEG may be identified. Combining with acceleration signal measurements, head movement artifacts in EEG may be identified.

[0044] In embodiments of the present disclosure, the EEG signals of some unnecessary channels can be deleted, avoiding interference with the feature extraction and use of EEG signals.

[0045] In some embodiments of the present disclosure, after the step of performing artifact

removal operation and noise removal operation on the collected original EEG signals, the method further includes: obtaining target features to be analyzed of EEG data; and selecting a reference electrode of the EEG collection device and a reference electrode calculation method based on the target features to be analyzed, the reference electrode calculation method includes
5 average reference, bilateral mastoid reference and chain reference. During specific implementation, re-reference refers to choosing an appropriate reference electrode or method based on different analysis targets, such as average reference, bilateral mastoid reference, chain reference, etc., to change the voltage value and spatial distribution of data. There are online reference electrodes in the EEG system, such as the middle location on the top of the head and
10 bilateral earlobes. Reference conversion is a kind of linear conversion, and any activity of the reference electrode will affect the signals of other electrodes. Therefore, the reference electrode should be placed accurately and ensure good contact, otherwise noise will be generated.

[0046] In embodiments of the present disclosure, during the process of EEG data processing, re-reference can be performed during offline analysis to reduce the noise data
15 brought by the reference electrode.

[0047] In some embodiments of the present disclosure, the method further includes: during the process of receiving the original EEG signals, calculating the amount of data of the original EEG signals collected by the EEG collection device received within a unit time, and sending a sampling rate reduction instruction to the EEG collection device when the amount of data
20 exceeds a preset threshold, thereby reducing load pressure of receiving the original EEG signals. During specific implementation, reducing the sampling rate refers to selecting an appropriate sampling rate based on the Nyquist sampling theorem of signals.

[0048] In embodiments of the disclosure, the amount of data and calculation can be reduced, thereby avoiding redundant information brought by excessive sampling.

[0049] In some embodiments of the present disclosure, the method further includes: when an abnormal channel with EEG signal quality lower than a preset threshold or with missing EEG signals is detected, selecting EEG signals from a channel at an adjacent location of the abnormal channel to interpolate and generate a new EEG signal, and using the new EEG signal as EEG signal from the abnormal channel. During specific implementation, the channel that
25
30 needs interpolation refers to a channel with poor data quality or missing data due to poor

electrode contact or other reasons.

[0050] In embodiments of the present disclosure, it is possible to utilize data of surrounding channels for interpolation supplementation, so as to restore the integrity of data.

[0051] In some embodiments of the present disclosure, the method further includes:
5 screening the segments divided from the pure EEG signals based on a preset threshold, and obtaining and deleting segments containing artifacts or outliers beyond the preset threshold. During specific implementation, some segments containing excessive artifacts or outliers may be found and deleted based on some preset thresholds or manual checks to improve data quality. The program can set the threshold, and the abnormal bands can be found, a reasonable filter is
10 selected based on the frequency range of the artifacts, which can effectively reduce artifacts of the original EEG data. For example, the high-pass filtering is set to 0.1Hz, and the low-pass filtering is set to 30Hz. In addition, in order to eliminate interference from the mains supply, notch filtering may be selected to remove interference at 50Hz. At the same time, when interested in a certain signal frequency range, band-pass filtering may be used. For example, if
15 it is interested in the α wave, band-pass filtering may be used to retain the signal frequency band of 8-13Hz for analysis.

[0052] In embodiments of the present disclosure, some segments containing excessive artifacts or outliers can be found and deleted to improve data quality.

[0053] In some embodiments of the present disclosure, the artifact removal operation is
20 based on blind source separation technology, and the performing artifact removal operation on the collected original EEG signals includes: decomposing the original EEG signals into multiple independent artifacts or neurons based on the blind source separation technology, using Independent Component Analysis (ICA) technology to identify independent variation sources in EEG signals, and removing the identified artifacts and/or independent variation
25 sources. Here, the ICA is a mathematical method.

[0054] During specific implementation, the Blind Source Separation (BSS) technology is based on the distinguishable spatiotemporal observation patterns of artifacts and neural signals, effectively decomposing multi-channel EEG data into multiple independent artifacts or neural sources. In EEG analysis, the ICA attempts to identify independent variation sources in EEG
30 data. Each row of the original data matrix represents the situation in which the potential

difference between each channel and the reference channel changes over time. After performing ICA decomposition, each row of data represents the situation where the independent components obtained after spatial filtering of the channel data change over time. Therefore, the ICA decomposition provides the temporal and spatial attributes of the independent components. The ICA is an implementation of the BSS. Typical abnormal patterns include: a) energy is concentrated only in the frontal lobe (such as electrooculogram artifacts); b) the topography is discontinuous (such as noise artifacts); and c) the topography is limited to a single electrode point (such as electrode artifacts). The above artifacts are only examples, and the present disclosure is not limited to this.

[0055] In embodiments of the present disclosure, mathematical methods may be used to decompose multiple mixed signals into several mutually independent and statistically uncorrelated components to identify and eliminate some artifact components, such as electrooculogram, electromyogram, electrocardiogram, etc.

[0056] In some embodiments of the present disclosure, the noise removal operation is based on canonical correlation analysis technology, and the performing noise removal operation on the collected original EEG signals includes: using autocorrelation in a given time series to describe signal components in the original EEG signals, obtaining the highest linear correlation and autocorrelation between the original EEG signals based on the canonical correlation analysis technology, and distinguishing and removing noise from EEG signals based on autocorrelation lower than a threshold.

[0057] During specific implementation, Canonical Correlation Analysis (CCA) is a multivariate statistical analysis method used to study the correlation between two groups of variables. It can reveal the intrinsic connection between two groups of variables. In the CCA, the autocorrelation in a given time series is used to describe the signal components in the uncleaned dataset. Autocorrelation describes the similarity between the observed data as a time delay function between them, that is, data points that are close in time usually have strong autocorrelation. The CCA means that the cross-covariance matrix is used to find the highest linear correlation and autocorrelation between datasets. To sum up, this technique can distinguish noise from EEG signals based on relatively low autocorrelation values. Especially for the study of ERP, CCA is considered to create a linear model between the stimuli and the

brain responses.

[0058] In embodiments of the present disclosure, noise can be distinguished from EEG signals based on relatively low autocorrelation values, thereby filtering out the noise contained in the EEG signals to obtain purer EEG signals.

5 **[0059]** In some embodiments of the present disclosure, the artifact removal and noise removal operations are based on a deep learning model, and the performing artifact removal and noise removal operations on the collected original EEG signals includes: using EEG data containing artifacts and noise as a training set to train the deep learning model, and using the trained deep learning model for automatic detection and automatic removal of artifacts and
10 noise. During specific implementation, based on the deep learning method, existing datasets are used to train the AI algorithm to automatically detect and reject artifacts in the record (ideally in real time). Mashhadi et al have successfully achieved reliable accuracy to reduce the mean square error between the target signal and the predicted signal.

[0060] Furthermore, the performing artifact removal and noise removal operations on the
15 collected original EEG signals further includes: using any one of a Signal - Noise Separation (SNS) method processing, Random Sample Consensus (RANSAC) method processing and Cross-Validation method processing to perform artifact removal operation and noise removal operation on the collected original EEG signals. Through methods such as the SNS method, the RANSAC method, and the Cross-Validation method, the Signal-to-Noise Ratio (SNR) of
20 the EEG signals can be improved to a certain extent.

[0061] In some embodiments of the present disclosure, the time-domain analysis indicators include a variety of average voltage, peak voltage, peak-to-peak, standard deviation, root mean square and event-related potential, and the frequency-domain analysis indicators include a
25 variety of spectral density, frequency band energy, frequency band ratio, spectral peak, spectral entropy, spectral power density and power spectrum.

[0062] The time-domain analysis indicators of EEG data refer to some parameters used to describe and quantify the characteristics of EEG data in the time domain, such as: (1) average voltage: the arithmetic average of the voltage values of EEG signals within a certain time range, reflecting the overall level of signals; (2) peak voltage: the maximum or minimum voltage
30 value of EEG signals within a certain time range, reflecting the extreme value of signals; (3)

peak-to-peak: the difference between the maximum voltage value and the minimum voltage value of EEG signals within a certain time range, reflecting the amplitude range of signals; (4) standard deviation: the average of the square roots of the differences between the voltage values of EEG signals within a certain time range and the average voltage, reflecting the fluctuation degree of signals; (5) root mean square: the square root of the average of the squares of the voltage values of EEG signals within a certain time range, reflecting the effective amplitude of signals; (6) event-related potential (ERP): after aligning EEG signals with an event (such as a stimulus or response), a waveform obtained by averaging all trials at each time point, reflecting the changes in brain electrical activities caused by the event.

[0063] The frequency-domain analysis indicators of EEG data refer to some parameters used to describe and quantify the characteristics of EEG data in the frequency domain, such as: (1) spectral density: the energy values of the EEG signals at each frequency, reflecting the frequency distribution and intensity of signals; (2) frequency band energy: the total energy value of EEG signals within a certain frequency band (such as δ , θ , α , β , γ , etc.), reflecting the activity level of signals in this frequency band; (3) frequency band ratio: the ratio of the energy values of EEG signals within two or more frequency bands, reflecting the relative activity level of signals in different frequency bands; (4) spectral peak: the maximum energy value of EEG signals at a certain frequency, reflecting the dominant frequency and rhythm of signals; (5) spectral entropy: the uncertainty or complexity of the energy values of EEG signals at all frequencies, reflecting the randomness and regularity of signals.

[0064] Spectral power density is an indicator that describes the distribution of the power of signals (here referring to EEG signals) in the frequency domain. The physical meaning is the signal power within a unit frequency bandwidth, with the unit watts per hertz (W/Hz). Spectral power density can reveal the energy distribution of signals at different frequencies and plays an important role in analyzing and designing signal processing systems, communication systems, etc.

[0065] Power Spectral Density (PSD) is obtained by averaging the data of each channel of over time:

$$result_{ch,i} = \frac{1}{durationInSec} \sum_{startTimeInSec}^{endTimeInSec} psdValues_{ch,j \times segmentLength + i} ; (1)$$

[0066] The input parameters in the above formula (1) include: (1) psdValues: a two-dimensional array representing PSD values of each channel, each row corresponds to one channel and each column corresponds to a time point; (2) segmentLength: an integer representing the frequency segmentation length of the PSD values; (3) startTimeInSec: an integer representing the starting time point for calculating the average of the PSD values, with the unit of seconds; (4) durationInSec: an integer representing the continuous time period for calculating the average of the PSD values, with the unit of seconds. The output results in the above formula (1) include: result: a two-dimensional array representing the averages of the PSD values of each channel, each row corresponds to one channel and each column corresponds to a frequency segmentation. The specific steps in the above formula (1) include: initializing the result array, to make the number of rows same as the psdValues array and the number of columns equal to the segmentLength.

[0067] In each channel (chIndex), the following operation is performed: initializing the result[chIndex] array to have the same length as the psdValues[chIndex] array. For each frequency segmentation (i), the following operation is performed: initializing the variable sum to 0. For each time point (j) from startTimeInSec to endTimeInSec (durationInSec + startTimeInSec), the following operations are performed: adding psdValues[chIndex][j * segmentLength + i] to the variable sum, and dividing the variable sum by durationInSec to obtain result[chIndex][i].

[0068] The power spectrum is an indicator used to characterize the energy distribution of signals, and is generally represented by the corresponding power spectral density function. For continuous-time signals, the power spectrum is the Fourier transform of the time-domain signals, representing the amplitudes and phases of different frequency components. For discrete-time signals, the power spectrum is the Fourier transform of the discrete-frequency-domain signals, also representing the amplitudes and phases of different frequency components. The power spectrum estimation belongs to the frequency domain. It means that the total energy of the time series is decomposed into components of different frequencies through the Fourier transform to obtain the power spectral density function of signals.

[0069] The power spectrum may be calculated based on the Welch method or the Bartlett method. The Bartlett method is a frequency spectrum estimation method, and the Welch method is also a spectrum estimation method and an improvement of the Bartlett method.

[0070] The steps for calculating the power spectrum based on the Welch method are as follows.

[0071] First, the data $x(n), n=0,1,\dots,N-1$ with a length of N is divided into L segments, each segment has M data, the data of the i -th segment is expressed as:

$$x_i(n) = x(n + iM - M), 0 \leq n \leq M, 1 \leq i \leq L; \quad (2)$$

[0072] Then, the window function $w(n)$ is applied to the data of each segment and the periodograms of each segment are calculated. The periodogram of the i -th segment is:

$$I_i(\omega) = \frac{1}{U} \left| \sum_{n=0}^{M-1} x_i(n) w(n) e^{-j\omega n} \right|^2, i = 1, 2, \dots, M-1; \quad (3)$$

[0073] In the above formula, U is called the normalization factor and the formula is expressed as:

$$U = \frac{1}{M} \sum_{n=0}^{M-1} w^2(n); \quad (4)$$

[0074] The periodograms of each segment are approximately regarded as uncorrelated with each other. Finally, the power spectrum estimation is:

$$P_{xx}(e^{j\omega}) = \frac{1}{L} \sum_{i=1}^L I_i(\omega); \quad (5)$$

[0075] The signals are divided into overlapping segments: the original data is divided into L data segments each with a length of M , and there are D overlapping points between adjacent segments. If $D = M/2$, the overlap is 50%; if $D = 0$, the overlap is 0%, which is the same as the Bartlett method.

[0076] In the above steps of calculating the power spectrum based on the Welch method, windowing may be applied to each segment: each data segment is multiplied by a window function (in the time domain), the window function usually provides greater weight to the central part of the data and less weight to the edge parts, which means there is a loss of

information. In order to reduce this loss, the data segments can have a certain overlap in time (as in the previous step). The periodograms of each segment are calculated: a discrete Fourier transform is performed on each data segment after windowing and then their squared amplitudes are calculated. The periodograms of all segments are averaged, thereby reducing the variance in power spectrum estimation.

[0077] An example of algorithm code:

[0078] Comment 1: the number of segments and the frequency vector are calculated.

[0079] `int nfft = (int)Math.Pow(2, Math.Ceiling(Math.Log(nperseg, 2)));`

[0080] Comment 2: the number of FFT points is the smallest integer power of 2 that is greater than or equal to nperseg.

[0081] `int nseg = (signal.Length - noverlap) / (nperseg - noverlap);`

[0082] Comment 3: the number of segments is the signal length minus the overlap length divided by the length of each segment minus the overlap length.

[0083] `freq = new double[nfft / 2 + 1];`

[0084] Comment 4: the frequency vector consists of nfft/2+1 points evenly distributed from 0 to half of the sampling frequency.

[0085] FIG. 2 is an example diagram of artifact removal and noise removal operations according to an embodiment of the present disclosure. After considering artifact removal, there may still be a lot of noise in the collected EEG signals. The signals and noise may be decomposed through post-filtering and mathematical algorithms to improve the signal-to-noise ratio.

[0086] It is worth noting that filtering cannot completely filter out the desired frequency band. For example, for a low-pass filtering of 30Hz, it does not mean that all signals outside 30Hz can be completely filtered out. Instead, with 30Hz as the cutoff frequency, signals higher than the cutoff frequency may be gradually attenuated.

[0087] In an embodiment of the present disclosure, in a coordinate system, with the horizontal axis representing the frequency and the vertical axis representing the gain coefficient, on the left side of the cutoff frequency f , the initial gain coefficients are all 1, which means that the signals are multiplied by 1 and are completely retained. In a certain range close to the cutoff frequency, the gain coefficient begins to gradually decrease until at the cutoff

frequency f , the gain coefficient is exactly 0.5. That is, after low-pass filtering with a cutoff frequency of f , 50% of the data at the f frequency band is filtered out and 50% is retained. Subsequently, the gain coefficient gradually decreases and the signals are gradually attenuated until the gain coefficient approaches 0.

5 **[0088]** Therefore, in some embodiments of the present disclosure, after performing a low-pass filtering of 30Hz on the EEG signals, a notch filtering of 50Hz may be performed again. This is because performing a low-pass filtering of 30Hz does not completely filter out all signals above 30Hz, and the interference at 50Hz is relatively strong, so it is best to perform another notch filtering of 50Hz.

10 **[0089]** The introduction to filtering methods is as follows.

[0090] (1) High-pass filtering and low-pass filtering: high-pass filtering means that signals higher than a certain frequency can pass through, while signals lower than this frequency may be attenuated and filtered out. On the contrary, low-pass filtering means that signals lower than a certain frequency can pass through, while signals higher than this frequency may be
15 attenuated and filtered out. Therefore, when inputting values, a smaller value should be input for the high-pass filtering, while a larger value should be input for the low-pass filtering.

[0091] (2) Band-pass filtering: band-pass filtering means that signals within a certain frequency range can pass through, while signals outside this frequency range may be attenuated and filtered out, which is equivalent to performing high-pass filtering and low-pass filtering
20 simultaneously.

[0092] (3) Notch filtering: notch filtering means that signals within a certain frequency range can be attenuated and filtered out, while signals outside this frequency range may be retained. This operation is usually used to remove the interference of mains electricity at 50Hz. In China, the frequency of the mains electricity is 50Hz, so there will be a very strong signal at
25 the frequency band of 50Hz in the collected signals, and notch filtering may be used to remove it.

[0093] In the EEG signal preprocessing method, system, and terminal device according to the present disclosure, artifacts and noise from the collected EEG signals can be removed, thereby obtaining relatively pure EEG data. The pure EEG data is divided into multiple
30 segments, and time-domain analysis and frequency-domain analysis are performed on each

segment, thereby completing the preprocessing of the EEG signals. The EEG signal preprocessing method according to the present disclosure can integrally adapt to the preprocessing requirements of EEG data under different circumstances, which can improve the processing efficiency of removing noise and artifacts from the original EEG data, thereby
5 improving the quality of EEG data and retaining the frequency bands and events of interest.

[0094] Furthermore, the EEG signal preprocessing method and system according to the present disclosure can realize the real-time and uninterrupted preprocessing of EEG signals. The real-time and uninterrupted preprocessing of EEG signals plays an important role in the subsequent artificial intelligence analysis of EEG signals because it can affect the accuracy and
10 reliability of the analysis results. For example, in fields such as EEG-based brain-computer interfaces, diagnosis of neurological diseases, and cognitive function assessment, the selection of EEG data preprocessing methods and parameters will have an impact on the classification or recognition effect.

[0095] Corresponding to the above methods, the present disclosure also provides an EEG
15 signal preprocessing system, which includes a processor and a memory. Computer instructions are stored in the memory, and the processor is configured to execute the computer instructions stored in the memory. When the computer instructions are executed by the processor, the system implements the steps of the aforementioned methods.

[0096] The present disclosure also provides a terminal device, and the terminal device is
20 configured to implement the steps of the aforementioned methods.

[0097] The present disclosure also provides an edge computing terminal device, which may include an intelligent sensor and a programmable logic controller (PLC). The edge computing terminal device is configured to collect data of the steps of any of the aforementioned methods.

[0098] An embodiment of the present disclosure also provides a computer-readable storage
25 medium on which a computer program is stored. When the computer program is executed by a processor, the steps of the aforementioned methods are implemented. The computer-readable storage medium may be a tangible storage medium, such as random access memory (RAM), internal memory, read-only memory (ROM), electrically programmable ROM, electrically erasable programmable ROM, registers, floppy disks, hard disks, removable storage disks, CD-
30 ROMs, or any other form of storage medium known in the technical field.

[0099] Those of ordinary skill in the art should understand that the exemplary components, systems, and methods described in combination with the disclosed embodiments in the document may be implemented in hardware, software, or a combination of both. Whether it is implemented in hardware or software specifically depends on the specific application and design constraints of the technical solutions. Professional technicians may use different methods to implement the described functions for each specific application, but such implementation should not be considered to be beyond the scope of the present disclosure. When implemented in hardware, it may be, for example, an electronic circuit, an application-specific integrated circuit (ASIC), appropriate firmware, plug-ins, function cards, and so on. When implemented in software, the elements of the present disclosure are programs or code segments used to perform the required tasks. Programs or code segments can be stored in a machine-readable medium or transmitted on a transmission medium or communication link through a data signal carried in a carrier wave.

[00100] It should be clear that the present disclosure is not limited to the specific configurations and processing described above and shown in the drawings. For the sake of conciseness, the detailed description of known methods is omitted here. In the above embodiments, several specific steps are described and shown as examples. However, the method process of the present disclosure is not limited to the specific steps described and shown. Those skilled in the art can make various changes, modifications, and additions or change the order of steps after understanding the spirit of the present disclosure.

[00101] In the present disclosure, features described and/or illustrated for certain embodiment can be used in the same way or in a similar way in one or more other embodiments, and/or combined with features of other embodiments or replace features of other embodiments. The above is only the preferred embodiments of the present disclosure and is not used to limit the present disclosure. For those skilled in the art, various changes and modifications can be made to the embodiments of the present disclosure. Any modifications, equivalent replacements, improvements, etc. made within the spirit and principles of the present disclosure shall be included in the protection scope of the present disclosure.

CLAIMS

What is claimed is:

1. An electroencephalogram (EEG) signal preprocessing method, comprising:
 - receiving original EEG signals collected by an EEG collection device, wherein a data structure of the original EEG signals corresponds to pre-positioned EEG signal channel locations, and each channel corresponds to a coordinate location on a scalp of a subject;
 - performing artifact removal operation and noise removal operation on the collected original EEG signals to obtain pure EEG signals;
 - obtaining event marks of the original EEG signals, and dividing the pure EEG signals into multiple segments, wherein each of the multiple segments contains EEG signals within a preset length of time range before and after an event occurs, and each of the event marks refers to one event; and
 - obtaining, for the segments divided from the pure EEG signals, preset analysis indicators to be extracted, respectively performing time-domain analysis to extract time-domain analysis indicators and respectively performing frequency-domain analysis to extract frequency-domain analysis indicators, and taking the extracted time-domain analysis indicators and frequency-domain analysis indicators as output results of the EEG signal preprocessing.
2. The method according to claim 1, wherein for the received original EEG signals collected by the EEG collection device, the method further comprises:
 - deleting original EEG signals from bilateral mastoid point locations on the scalp of the subject and/or from electrooculogram channels.
3. The method according to claim 1, wherein subsequent to the performing artifact removal operation and noise removal operation on the collected original EEG signals, the method further comprises:
 - obtaining target features to be analyzed of EEG data, and selecting a reference electrode of the EEG collection device and a reference electrode calculation method based on the target

features to be analyzed, wherein the reference electrode calculation method comprises average reference, bilateral mastoid reference and chain reference.

4. The method according to claim 1, further comprising:

during the process of receiving the original EEG signals, calculating an amount of data of the original EEG signals collected by the EEG collection device received within a unit time, and sending a sampling rate reduction instruction to the EEG collection device when the amount of data exceeds a preset threshold, so as to load pressure of receiving the original EEG signals.

5. The method according to claim 1, further comprising:

when an abnormal channel with EEG signal quality lower than a preset threshold or with missing EEG signals is detected, selecting EEG signals from a channel at an adjacent location of the abnormal channel to interpolate and generate a new EEG signal, and using the new EEG signal as the EEG signal from the abnormal channel.

6. The method according to claim 1, further comprising:

screening the segments divided from the pure EEG signals based on a preset threshold, and finding and deleting segments containing artifacts or outliers beyond the preset threshold.

7. The method according to claim 1, wherein the artifact removal operation is based on blind source separation technology, and the performing artifact removal operation on the collected original EEG signals comprises:

decomposing the original EEG signals into multiple independent artifacts or neurons based on the blind source separation technology, using independent component analysis (ICA) technology to identify independent variation sources in the EEG signals, and removing the identified artifacts and/or independent variation sources.

8. The method according to claim 1, wherein the noise removal operation is based on canonical correlation analysis technology, the performing noise removal operation on the

collected original EEG signals comprises:

using autocorrelation in a given time series to describe signal components in the original EEG signals, finding the highest linear correlation and autocorrelation among the original EEG signals based on the canonical correlation analysis technology, and distinguishing and removing noise from EEG signals based on autocorrelation lower than a threshold.

9. The method according to claim 1, wherein the noise removal operation is based on filtering processing and mathematical algorithm to decompose EEG signals and noise, wherein the filtering processing comprises one or more of high-pass filtering, low-pass filtering, band-pass filtering and notch filtering.

10. The method according to claim 1, wherein the artifact removal operation and noise removal operation are based on a deep learning model, the performing artifact removal operation and noise removal operation on the collected original EEG signals comprises:

using EEG data containing artifacts and noise as a training set to train the deep learning model, and using the trained deep learning model for automatic detection and automatic removal of artifacts and noise.

11. The method according to claim 10, wherein the performing artifact removal operation and noise removal operation on the collected original EEG signals further comprises:

using any one of signal noise separation (SNS) processing, random sample consensus (RANSAC) processing and cross-validation processing to perform artifact removal operation and noise removal operation on the collected original EEG signals.

12. The method according to claim 1, wherein the time-domain analysis indicators comprise one or more of average voltage, peak voltage, peak-to-peak, standard deviation, root mean square and event-related potential, and the frequency-domain analysis indicators comprise one or more of spectral density, frequency band energy, frequency band ratio, spectral peak, spectral entropy, spectral power density and power spectrum.

13. An electroencephalogram (EEG) signal preprocessing system, comprising a processor and a memory, wherein computer instructions are stored in the memory, the processor is configured to execute the computer instructions stored in the memory, and when the computer instructions are executed by the processor, the system implements the steps of the method according to any one of claims 1 to 12.

14. A terminal device, wherein the terminal device is configured to implement the steps of the method according to any one of claims 1 to 12.

15. An edge computing terminal device, comprising an intelligent sensor and a programmable logic controller (PLC), wherein the edge computing terminal device is configured to collect data of steps of the method according to any one of claims 1 to 12.

16. A computer-readable storage medium, on which a computer program is stored, wherein when the program is executed by a processor, steps of the method according to any one of claims 1 to 12 are implemented.