Research on CART Model of Mass Concrete Temperature Prediction Based on Big Data Processing Technology

Li Baoyu\textsuperscript{1}, Li Guoxing\textsuperscript{2}, Wang Guiyu\textsuperscript{2}, Zhang Guofeng\textsuperscript{2}, and Yang Man\textsuperscript{3}

\textsuperscript{1}Henan Vocational College of Water Conservancy and Environment, Zhengzhou, China
\textsuperscript{2}Henan No.2 Hydraulic Engineering Bureau, Zhengzhou, China
\textsuperscript{3}Zhengzhou University, University of Colorado, Zhengzhou, China

Corresponding author: Zhang Guofeng (e-mail: 272217640@qq.com).

“\textit{This work was supported in part by the Henan province water conservancy science and technology research project.}”

\textbf{ABSTRACT} Due to the influence of temperature changes or temperature gradients in the construction process of mass concrete, temperature cracks will occur in the concrete. In order to achieve a reasonable prediction of the temperature change of the concrete during the construction process and accurately obtain the temperature change trend, this paper attempts to construct a CART prediction model based on big data processing technology based on the characteristics of the temperature change of the mass concrete. Data processing methods such as missing value filling and moving average and the construction method of CART prediction model, combined with engineering examples, demonstrate the feasibility of the model method. The results show that the model and method can better predict the temperature change of mass concrete. It has high prediction accuracy and can provide necessary guidance for practical engineering.

\textbf{INDEX TERMS} Mass concrete temperature; Big data processing technology; CART prediction model

\section{I. INTRODUCTION}
Mass concrete is a common structural form in modern engineering structures. Mass concrete will generate a lot of hydration heat during the pouring process. Due to the large size of concrete structure, poor thermal conductivity of concrete, construction constraints, and temperature changes, temperature cracks usually occur during construction\cite{1-6}. The mechanism of temperature cracks is mainly: the mass concrete structure generates stress and strain due to the temperature difference and temperature gradient between the inside and outside. The internal and external constraints of the structure prevent the development of this stress and strain. When the stress and strain exceed the limit value, it will cause concrete structural damage, resulting in cracks. Therefore, in order to control concrete cracking, it is necessary to control the maximum tensile stress or the maximum tensile strain not to exceed the limit of the corresponding concrete. The existence of temperature cracks not only affects the appearance of the concrete structure, but also has a great impact on the safety of the structure, such as reducing the bearing capacity and safety of the structure, affecting the waterproof performance of the structure, and reducing the durability and service life of the structure.

Necessary temperature control of mass concrete is one of the key means to ensure the quality of mass concrete pouring, such as controlling temperature changes during constructions to avoid large temperature changes or large temperature gradients during concrete construction\cite{7-11}. Predicting the temperature in the process of concrete pouring, predicting the changing process of concrete temperature in advance, and taking corresponding temperature control measures can ensure the construction quality of mass concrete\cite{12, 13}.

At present, the main analysis methods about temperature change of mass concrete include empirical formula method, finite difference method, finite element method and so on\cite{14}. In 1933, measures such as low-heat cement, split joints, and intermittent pouring were adopted for the first time in the construction of Hoover Dam; around 1960, some concrete dams in the United States adopted measures such as reducing the proportion of cement and pre-buying cooling water pipes, and achieved good temperature control and crack prevention.; In 1968, Professor EL Wilson of the University of California first introduced the finite element time history analysis method into the temperature stress analysis of concrete dams; in 1990, Enrique Mirambell et al established a two-dimensional finite difference analysis model; in 1992,
Barrett PK et al. creatively applies the cracking model to the calculation and analysis of 3D temperature stress; in 2004, Yunus Ballim introduced the development and application of a finite difference thermal model, which can predict the time-varying temperature distribution in mass concrete members; in 2008, Denzil Lokuliy studied the relationship between the generation, development and temperature stress of temperature cracks in mass concrete structures, and systematically analyzed the relationship between the development of mass concrete cracks and other internal effects; in 2014, Lawrence.AM et al. The element method was used to analyze the effect of early strength and cracking of concrete with different auxiliary cementitious materials. Since the temperature changes process of mass concrete is relatively complex, there are many influencing factors, and the above conventional physical analysis methods have simplified the heat diffusion equation, boundary conditions and external environment, which can reflect the thermal change law of concrete, but the above analysis is based on the physical diffusion process of temperature change, which requires high temperature parameters and will have a large deviation from the actual temperature change of concrete[15-25].

Big data processing technology is a new type of data analysis and processing method, which can make full use of existing data resources, conduct in-depth mining processing, and extract potential data laws and useful knowledge information[26-30]. During the pouring process of mass concrete, the temperature change data can be collected by burying the temperature sensor, and various data of concrete pouring can also be obtained[31-34]. These data provide a certain data basis for the analysis and application of big data technology. For example[35-40], with the help of mass concrete temperature data, Norris et al. proposed an embedded nano- and micro-electromechanical sensing system, and successfully applied it to the temperature and humidity monitoring inside the concrete structure; Kim et al. proposed a surface acoustic wave monitoring system, It can be used for non-destructive monitoring of concrete temperature. However, there is still a lack of relevant research on temperature prediction considering the real-time change of concrete temperature with time, temperature measurement points with spatial distribution, sensor anomaly and failure and other factors. There is also less research, especially the lack of practical methods that can guide the actual engineering of mass concrete temperature monitoring and predictive analysis.

Aiming at the temperature change of mass concrete[41-44], this paper attempts to introduce big data processing technology by analyzing the temperature data that has occurred to provide a new analysis model for the temperature prediction of mass concrete.

II. MATERIALS AND METHODS

A. Research Materials

1) INSTANCE SELECTION

The engineering example is selected as the construction of the left-side joint bottom plate of the intake pool of a booster pumping station in Henan Province. The left joint bottom plate has a pouring length of 25 m, a width of 12.2 m, a thickness of 3 m (the thickness of the tooth groove is 4.5 m), and the distance of the tooth groove is 2 m from the rear boundary. Figure 1 is a schematic diagram of the left-linked bottom plate project, showing the structure of the project.
According to the construction schedule, the pouring of the left-linked bottom plate of the inlet pool will end at 10:30 on March 11, 2021. Use the blue line to represent the minimum temperature, the red line to represent the maximum temperature, and set the abscissa as time and ordinate as the temperature value. Draw a schematic diagram of the outside temperature change during pouring in Figure 2, showing the process of outside temperature change during pouring.

![Figure 2. Schematic diagram of changes in outside air temperature.](image)

2) MASS CONCRETE TEMPERATURE DATA

There are 21 temperature sensors embedded in the concrete of the left-sided bottom plate, 7 on each floor, 3 floors, among which the top sensor is embedded in the concrete 0.1 m away from the concrete surface, the middle sensor is 1.3 m away from the concrete surface, and the bottom sensor is located 2.9 m away from the concrete surface, and all sensors are 0.5 m away from the concrete side. The schematic diagram of the layout position of the temperature sensors of each layer are shown in Figure 3, which respectively indicate the position, number and arrangement position of the cooling water pipes of the temperature sensors on the upper, middle and lower layers.

![Figure 3. Schematic diagram of the coordinate position of the temperature sensor.](image)

A summary of the temperature data coordinates of each measuring point in Table 2.

| Location       | Number | Coordinate/Distance |
|----------------|--------|---------------------|
| Bottom point   | 1      | (12.15 12.5 0.05)   |
|                | 2      | (6.1 24.95 1.65)    |
|                | 3      | (6.1 18.725 0.05)   |
|                | 4      | (6.1 6.275 0.05)    |
|                | 5      | (6.1 0.05 0.05)     |
|                | 6      | (0.05 12.5 0.05)    |
|                | 7      | (6.1 12.5 0.05)     |
| Distance of lower water pipe |      | 0.95                |
| Middle point   | 1      | (12.15 12.5 1.3)    |
|                | 2      | (6.1 18.725 1.3)    |
|                | 3      | (6.1 24.95 2.9)     |
|                | 4      | (6.1 6.275 1.3)     |
|                | 5      | (6.1 0.05 1.3)      |
|                | 6      | (0.05 12.5 1.3)     |
|                | 7      | (6.1 12.5 1.3)      |
| Distance of lower water pipe |      | 0.3                 |
| Upper water pipe distance |      | 0.7                 |
| Top point      | 1      | (0.05 12.5 2.95)    |
|                | 2      | (6.1 24.95 4.55)    |
|                | 3      | (6.1 18.725 2.95)   |
|                | 4      | (6.1 6.275 2.95)    |
|                | 5      | (6.1 0.05 2.95)     |
|                | 6      | (12.15 12.5 2.95)   |
|                | 7      | (6.1 12.5 2.95)     |
| Distance of lower water pipe |      | 0.95                |

The embedded temperature sensor adopts a digital temperature sensor, which is saved and stored according to the set time. Considering that the concrete temperature changes process is relatively smooth, in order to ensure the accuracy and quantity requirements of the temperature data, the storage time of the selected data is 10min/group, that is, once an hour. A temperature sensor stores 6 temperature data.

In the process of collecting temperature data of large-volume concrete, due to the influence of factors such as construction conditions, weather environment, instrument failure, and the amount of temperature collection data from different data sources (temperature sensors), the consistency of data attributes cannot be obtained. It is guaranteed that the original temperature data collected may include abnormal points, missing values, and inconsistent attribute format data, and the data quality is difficult to meet the requirements.

If the original data is used directly, the outliers and missing values that appear will seriously affect the data effect. Therefore, it is necessary to introduce the relevant technology of big data analysis to deal with the outliers and missing values that appear to improve the quality of the original data; And for the accidental error of the original data[45, 46], moving average processing is performed to eliminate the error[47, 48].
B. Mass Concrete Temperature Prediction Model

1) MODEL SELECTION

Decision tree is one of the most classic predictive models in the field of data science, and it also is a very widely used classification method. Compared with conventional methods, decision tree has the advantages of low computational time and space complexity, easy to understand output results, insensitivity to missing features, and no feature processing. A large number of researchers have proposed different decision tree generation algorithms in terms of feature types, impurity measures or variables, and the selection method of segmentation and the number of sub-node splits, such as the common ID3, C4.5, conditional inference tree and CART.

The ID3 (Iterative Dichotomiser 3) algorithm pioneered decision trees. This algorithm uses information entropy to measure node impurity, and uses information gain to evaluate node splitting. It can only handle discrete features, and for continuous features, enumerate all values are used to split nodes. The algorithm tends to select features with more values and tends to split many leaf nodes with a few samples, which is easy to cause over fitting. On the basis of improving ID3, the C4.5 algorithm takes the sample size of each split node into consideration, and uses the information gain rate instead of the information gain as the evaluation criterion for node splitting, avoiding splitting more sub-nodes with a small sample size and selecting those For features with more values, continuous features are processed by splitting thresholds. However, both ID3 and C4.5 can only solve the classification problem, but cannot solve the regression problem, so they are not suitable for reasonable prediction of the temperature change trend in the later stage of mass concrete construction.

Conditional inference trees are similar to classic decision trees, but determine the choice of independent variables and split points based on statistical tests, rather than measuring impurity. The significance test is a permutation test. Since the conditional inference tree is only split into two child nodes each time, the conditional inference tree does not need pruning, and the threshold determines the complexity of the model, so how to determine the threshold parameter is very important. Since the temperature change of mass concrete construction is affected by many factors such as the concrete material itself, the external temperature, and the construction process, these factors interact with each other. When using the conditional inference tree, it is necessary to distinguish independent variables and dependent variables independently, and it is difficult to reasonably determine the threshold parameters.

Classification and Regression Tree (CART) can solve both classification and regression problems. The decision tree prediction model can directly reflect the characteristics of the data, and simultaneously process data-type and general-type attributes, and can produce feasible and effective results for large data sources in a relatively short period. When solving the classification problem, the node impurity is measured by the Gini index, and the node splitting evaluation uses the Gini index drop value. To solve the regression problem, the node impurity is measured by the variance of the target feature, and the node splits evaluation standard using the variance drop value. It acts as a tree structure, where each internal node represents a test on an attribute, each branch represents a test output, and each leaf node represents a category, each split into only two child nodes. For continuous features, they are split into two child nodes by a split threshold. For discrete features, CART no longer splits child nodes according to all the values of the feature, but selects one value at a time, and splits nodes according to whether it is the value. When the CART model is learned, the training data is used to establish a decision tree model according to the principle of minimizing the loss function. Therefore, this paper recommends the use of CART algorithm to predict the temperature of mass concrete.

The CART algorithm consists of the following two steps:

1) Decision tree model generation: Generate a decision tree based on the training data set, and the generated decision tree should be as large as possible; 2) Decision tree model pruning: Pruning the generated tree with the verification data set and selecting the optimal subtree. The minimum loss function can be selected as the criterion for pruning.

2) CART MODEL BUILDING

The generation of the CART model is a process of recursively constructing a binary decision tree, which uses the Gini index as the basis for selecting the optimal partition feature[49-52]. The Gini value refers to the probability that two samples are randomly selected from a sample set, and the two samples do not belong to the same class. The main function of the Gini value is to measure the "impurity" of the data division. The smaller the Gini value, the higher the "purity" of the sample after the division. The formula for calculating the Gini value is Formula 1.

\[
Gini(p) = \sum_{k=1}^{K} p_k (1 - p_k) = 1 - \sum_{k=1}^{K} p_k^2
\]

(1)

\(K\) is number of categories; \(p_k\) is probability of getting a subset of the \(k\) class in the sample set.

For the regression tree, the method of minimizing the square error is used to determine the division method. In the input space, each area is recursively divided into two sub-areas, and the output value is determined. Formula 2 is the calculation formula with the smallest square calculation error amount.

\[
\text{Error} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

(2)
\[ \min_{j,s} \left[ \min_{c_1 \in R(j,s)} \sum_{x_i \in R(j,s)} (y_i - c_1)^2 + \min_{c_2 \in R(j,s)} \sum_{x_i \in R(j,s)} (y_i - c_2)^2 \right] \]

\[ R_1(j,s) = \{ x \mid x^{(j)} \leq s \} \]

\[ R_2(j,s) = \{ x \mid x^{(j)} > s \} \]

\[ c_1 = \frac{1}{N_1} \sum_{x_i \in R(j,s)} y_i \]

\[ c_2 = \frac{1}{N_2} \sum_{x_i \in R(j,s)} y_i \]

(2)

\( x_i \) is the sample attribute characteristics; \( y_i \) is the sample value; \( j \) and \( s \) respectively represent the division \( j \) point of the variable \( s \); \( x \) represents a collection of sample attribute feature values in the region; \( x^{(j)} \) represents the characteristic value of the sample attribute in the segmentation variable \( j \); \( R_1(j,s), R_2(j,s) \) indicate the area to which it belongs after splitting variable \( j \) and splitting point \( s \); \( N_1, N_2 \) represents the number of samples in the region; \( c_1, c_2 \) represents the sample mean value of the region, the smaller the value of the expression, the better the prediction effect. The attribute with the smallest square error can be selected as the splitting scheme.

3) CART MODEL PRUNING

The classic decision tree is the core of the CART model. The pruning of the regression tree is to prevent the model from overfitting. The CART algorithm performs pruning by calculating the error gain rate on the training set. Formula 3 is the calculation formula of Gini value.

\[ g(t) = \frac{C(t) - C(T_i)}{|T_i| - 1} \]  

(3)

\( g(t) \) is error gain ratio; \( C(t) \) indicates the error of the node after the subtree of node \( t \) is pruned; \( C(T_i) \) indicates the error of subtree \( T_i \) when the subtree of node \( t \) has not been pruned; \( |T_i| \) indicates the number of \( T_i \) leaf nodes.

The pruning strategy is to take out the node corresponding to the smallest index \( g(t) \), cut it out and generate the first subtree, repeat this process until only the root node is left, use it as the last subtree, and then use the verification set to verify all subtrees, and take the tree with the smallest error.

C. Big Data Processing Technology And Methods

The big data analysis ideas for temperature changes of mass concrete, including data collection, data processing, data analysis and data application, are briefly described as follows:

Data collection: It is necessary to obtain the corresponding raw data first, as the data basis of big data analysis, which includes the temperature data of each measuring point of the concrete, the position data of the measuring point, the concrete pouring data, the outside air temperature data, the outside weather data, and the outside wind data.

Data cleaning: For the acquired raw data, due to problems such as collection equipment and human records, part of the data will be wrong or missing. Therefore, the raw data needs to be cleaned and sorted to meet the data usage requirements and improve the application effect of big data.

Data analysis: through data mining on the processed data, the application of machine learning, intelligent algorithms, mathematical modeling and other means to achieve in-depth analysis of the data, and obtain the internal laws and predictive models of the data;

Data application: Using the prediction model obtained by big data analysis, according to different working conditions, inputting the corresponding working condition parameters, calling the analysis model, it can carry out reasonable prediction analysis on the temperature change of the mass concrete.

1) OUTLIER IDENTIFICATION

Outliers mainly refer to data samples that deviate significantly from other data and are unreasonable, also known as outliers. The processing of outliers usually requires detection, screening and processing. The more commonly used monitoring methods include boxplots, simple statistics (such as observing maximum or minimum values), the 3σ principle, etc. The more commonly used processing methods include deletion method, imputation method (regression imputation, multiple imputation) and substitution method (replacement of the mean of continuous variables, and substitution of mode and median for discrete variables), etc.

Since there are many data sources for mass concrete temperature acquisition, the data samples are large, and the number of outliers is generally small. Therefore, this paper recommends using a boxplot to identify outliers, and directly remove the identified outliers, so that it can make the processing of outliers very simple, reduce the work intensity of data detection, screening, interpolation, replacement, etc., and obtain ideal data in the fastest way, and at the same time, it has less impact than the huge data source.

Boxplot is a method commonly used in statistics to display the distribution of data without any restrictions on the data. The boxplot is a graph drawn with the maximum, minimum, 0.25 quantile, median, and 0.75 quantile in a set of data. With the help of the boxplot, the degree of dispersion of the data, the overall distribution and the identification of samples can be intuitively reflected. Singular values in the data. For the vertical boxplot, the upper and lower bounds of the box are the 0.75th quantile and the 0.25th quantile, respectively, and the red notch in the middle is the sample median.

For box plots, outliers (points beyond the upper and lower edges) are empirical to identify the boundary. Usually the lower boundary of the box is taken to extend the dotted line to the position of “0.25 quantile - 1.5 × (0.75 quantile - 0.25 quantile)” , and the upper boundary is derived from the dotted line to the position of “0.75 quantile+1.5×(0.75 quantile - 0.25 quantile)”.
quantile-0.25 quantile)”, the sample points outside the boundary are regarded as abnormal points.

The data points that fall into the upper and lower boundaries of the box are normal points, and the data points outside the boundary of the box are abnormal points. For abnormal points, they need to be eliminated.

2) MISSING VALUE FILL

Due to the working failure of data flow measurement equipment, missing records, outlier elimination, etc., there are certain missing values in the measurement data. For the processing of missing values, filling methods such as mean, mode filling and linear regression filling are usually used, or the setting of It is a dummy variable, or it is directly eliminated. Using mean and mode filling, the method is simple, but the filling results are rough, which even has a negative impact on model training; using linear regression filling is relatively complex, and can obtain better collinearity, but the selection of filling basis is particularly important; when missing values When the number is small, if the impact on the entire data analysis is small, the method of setting dummy variables or eliminating them can be used for direct processing.

For the missing values in the mass concrete temperature data, because the data acquisition devices such as thermometers are embedded in the concrete, it is difficult to repair or replace the thermometers when damaged, which will inevitably cause some continuous data to be missing. In order to reflect the different positions inside the mass concrete The temperature change trend, this paper proposes to use the nearest neighbor algorithm model for filling [53-56].

The nearest neighbor algorithm is a clustering algorithm based on a similarity measurement strategy. It judges the relationship between two data items according to the similarity between different data items, and then obtains it according to the "distance" between different data items Estimates of missing values for related data items. Suppose the two data samples are respectively i, j. Formula 4 is the expression of the proximity value between two samples.

\[
d(i, j) = \left( |x_1 - y_1| + |x_2 - y_2| + \cdots + |x_p - y_p| \right)^{1/p}
\]

(4)

\(x_1, \ldots, x_p\) is the sample i, \(y_1, \ldots, y_p\) is the sample j, p is order, if \(p = 1\), near value is manhattan distance; if \(p = 2\), the proximity value is euclidean distance.

The nearest neighbor algorithm flow is as follows:

1) Divide the data into a data \(D_w\) set with missing items and a data \(D_{-w}\) set without missing items;
2) Select a data \(S\) sample from \(D_{-w}\);
3) Calculate the adjacent value of each sample \(S\) in the sample \(S\) and the data without missing items according to formula (1);
4) According to the calculated neighbor values, arrange them from largest to smallest, and select the largest samples as similar sets of target samples;
5) Select the mean of similar samples in similar sets as the missing value;
6) Repeat steps 2-5 until the set \(D_w\) is empty.

3) RANDOM ERROR ELIMINATION

Many random factors in the process of data measurement cause the magnitude and direction of the measured value to be difficult to predict, resulting in measurement errors. Due to the universality of measurement errors, it is determined that any measurement will have random errors. In the process of data trend monitoring, the fluctuation of random errors, that is, the burr phenomenon in the data curve. The commonly used methods for random error processing in the data monitoring process include: median filtering, arithmetic mean filtering, moving average filtering, exponential smoothing filtering, and linear quadratic exponential smoothing filtering. Median value filtering is a simple and rough data processing method. It has better processing effect on static measurement signals and is suitable for dealing with coarse errors; arithmetic mean filtering is suitable for filtering signals with random interference in general, but it does not have the Anti-error performance; moving average filtering is a predictive filtering method, suitable for systems with high real-time requirements but slow processing speed; exponential smoothing filtering and linear quadratic exponential smoothing filtering are used as weighted moving average filtering. In the process of predictive filtering, the latest measurement value should have a larger weight than the earlier measurement value, which effectively solves the lag error problem of the latter.

Since the mass concrete temperature data is continuous monitoring data, its change law and trend prediction need to be analyzed, and the error accuracy and tolerance performance are high. At the same time, the weight of the temperature data is suitable for equal treatment in the whole monitoring process. Therefore, this paper recommends using the moving average. The filtering method is a widely used random error processing and analysis and prediction method. Moving average is a commonly used technical method for data analysis. The basic idea is to eliminate errors caused by dynamic changes of the averaged object over time, thereby eliminating errors caused by random fluctuations of uncertain factors during the measurement process. Formula 5 is the model calculation formula for moving average processing [57-60].

\[
T(t) = \frac{\sum_{i=1}^{n(t)} T_i(t)}{n(t)} \quad t \in \Delta(t)
\]

(5)

\(T(t)\) is temperature value after moving average, \(\Delta(t)\) is the time interval containing the moment, \(n(t)\) is the number of temperature values collected by the monitoring object in the time interval, \(T_i(t)\) is raw temperature monitoring data.

III. RESULTS AND DISCUSSION

A. Temperature data processing
1) OUTLIER IDENTIFICATION
Identify the abnormal value of the collected raw temperature data according to the box plot, and determine the normal value range according to the distribution of the box plot, as shown in Table 3 below.

| Threshold     | Temperature value |
|---------------|-------------------|
| Q3            | 46.2              |
| Q1            | 28.3              |
| Upper boundary| 73.1              |
| Lower boundary| 1.4               |

Through box plot processing, it is considered that the collected temperature data is not in the interval [1.4, 73.1] as outlier points and needs to be deleted, improve the overall quality of sample data by identifying and removing outliers.

2) MISSING VALUE FILL
Due to the removal of abnormal values and collection failures, certain missing values will appear in the temperature data. In this example, due to the measurement failure of the upper temperature sensor, the data is from 16:20 on March 20 to March 22, 2021. No temperature data is collected between 0:20, and there are continuous vacancies. Figure 4 is the original distribution of temperature data at the upper level measurement points.

According to the nearest neighbor algorithm constructed in this article, the above missing values are filled, Figure 5 is the distribution of the missing values of temperature data at the upper level measurement points after filling in.

By filling the missing data, the continuity of the data is ensured, the law of filling the data is consistent with the temperature change law of the temperature measuring point where it is located, and the data filling has a certain scientific nature.

3) MOVING AVERAGE PROCESSING
Based on the abnormal value processing and missing value filling of the collected temperature data, the mathematical method of moving average is used to process the temperature data to eliminate the regular systematic errors in the data and improve the data quality.

Take the 3# point in the temperature sensor of the upper layer as an example, the characteristic temperature obtained is obtained by moving and averaging 5 data points including this point (2 points forward and backward). Figure 6 is data distribution after moving average of point 3# on the upper layer, Figure 7 is a partial enlarged view of the upper layer 3# point after moving average.
Figure 7. A partial enlarged view of the upper layer 3# point after moving average.

The error ranges between the characteristic temperature and the original temperature after moving average processing is between [-0.3, 0.3]. The average value of the single-point absolute difference between the original temperature and the characteristic temperature is calculated to be 0.046, and the fluctuation of the temperature value becomes smaller. The temperature value is a smooth curve over time, eliminating the effect of errors caused by high-frequency fluctuations.

B Temperature prediction model training

The temperature data collected by the upper temperature sensor 1#-7# and the middle temperature sensor 1#, 2#, 3#, 6# and 7# arranged on the left side of the inlet pool bottom plate, the temperature data collected by the temperature sensor location, collection time, outside air temperature, wind, weather conditions, and height information from the temperature sensor to the cooling water pipe are input into the CART regression model for model training.

Through cross-validation, it is determined that the hyperparameters under the CART algorithm are set to 10 for max_depth, 4 for min_samples_split, and 3 for min_samples_leaf. The mean square error of the training set is 0.047, and the mean square error of the test set is 0.086. At this time, the training set and the test set are The mean square error is the smallest.

C Temperature prediction model verification

Using the above temperature data training model to verify the model of the middle temperature sensor 4# and 5# arranged on the left side of the inlet tank bottom and the lower temperature sensor 1#-7# arranged on the left side of the inlet tank floor, Figure 8 is a comparison diagram of temperature prediction and actual measurement at points 4# and 5# in the middle layer, Figure 9 shows the measured temperature and predicted temperature of the lower temperature sensor from 1# to 7# on the left side of the bottom of the inlet tank:

Through the above model verification, the trend changes of the predicted temperature curve and the measured temperature curve are kept consistent, which basically reflects the change law of the concrete temperature during the construction period. The prediction error is larger at the inflection point of the temperature curve and at the local points in the later stage of temperature change, but it is basically maintained within 1 °C. Compared with the measured curve, the overall prediction curve fluctuates. Basically, they are concentrated around 0°C.

In summary, the temperature prediction CRAT model trained by using big data processing technology can better reflect the law of concrete temperature changes during the construction period of mass concrete projects, and realize the prediction of temperature changes, and the prediction accuracy is high, with strong generalization ability and stability.

IV. Conclusion

In this paper, combined with the characteristics of temperature change of mass concrete, on the basis of sorting out the prediction models and methods, a CART model suitable for temperature prediction of mass concrete is constructed. Value identification, missing value filling, and random error elimination, improving the quality of the data. Combined with engineering examples, the collected
temperature data is processed and analyzed to verify the feasibility of temperature change prediction and analysis based on big data processing technology and CART prediction model. The results show that the model and method can better predict the temperature change of mass concrete.

Because the model method proposed in this paper is based on the acquisition and analysis of a large amount of data during the construction of mass concrete, the data is rationalized with the technology of big data processing, and the temperature control model is trained on this basis. Targeted, higher prediction accuracy; the model and method can also be modified in real time according to different working conditions in the construction process, which is more applicable than the traditional method; at the same time, the model method is easy to understand and easy to use, suitable for Mass concrete construction enterprises analyze and process a large amount of temperature data during the construction process and predict the temperature change trend, which is convenient for the construction personnel to adjust the concrete construction process in time according to the prediction results, optimize the engineering measures, save the engineering investment, reduce the occurrence of concrete temperature cracks, and improve the concrete construction quality.

REFERENCES

1. Keun, L.C., Thermal Cracking Control of Mass Concrete by Vertical Pipe Cooling Method. Journal of the Korean Recycled Construction Resources Institute, 2014. 2(3): p. 233-238.
2. Sargam, Y., et al., Predicting thermal performance of a mass concrete foundation - A field monitoring case study. Case Studies in Construction Materials, 2019. 11. DOI10.1016/j.cscm.2019.e00289
3. Tankasala, A. and A.K. Schindler, Early-Age Cracking of Lightweight Mass Concrete. Aci Materials Journal, 2020. 117(1): p. 223-232.
4. Tankasala, A. and A.K. Schindler and K.A. Riding, Risk of Thermal Cracking from Use of Lightweight Aggregate in Mass Concrete. Transportation Research Record, 2017(2629): p. 42-50.
5. Timmers, M. and A.T. Jacobs, Concrete apartment tower in Los Angeles reimagined in mass timber. Engineering Structures, 2018. 167: p. 716-724.
6. Trong Chuc, N., et al., FINITE ELEMENT ANALYSIS OF TEMPERATURE AND STRESS FIELDS IN THE CONCRETE MASS WITH PIPE-COOLING. Structural Integrity and Life-Integritet I Vek Konstrukcija, 2020. 20(2): p. 131-135.
7. Abeha, H., S. Agyeman, and M. Adom-Asamoah, Thermal effect of mass concrete structures in the tropics: Experimental, modelling and parametric studies. Cogent Engineering, 2017. 4(1): DOI10.1080/23319196.2016.1278297
8. Cuo, W., et al., Experimental study on controlling temperature cracks of mass foundation concrete by circulating water. Journal of Natural Disasters, 2016. 25(1): p. 97-102.
9. Chen, G., et al., A review of recent advances in controlling temperature cracks in mass concrete construction. Journal of Natural Disasters, 2016. 25(3): p. 159-65.
10. Guo, W. and S. Yang, The control Study of Mass Concrete Floor Temperature Crack. in 2nd International Conference on Structures and Building Materials (ICSBM). 2012. Hangzhou, PEOPLES R CHINA.
11. Ouyang, J., et al., Application of distributed temperature sensing for cracking control of mass concrete. Construction and Building Materials, 2019. 197: p. 778-791.
12. Liu, D., et al., Prediction of Hydration Heat of Mass Concrete Based on the SVR Model. Ieee Access, 2021. 9: p. 62935-62945.
13. Jeong, D.J., et al., Analytical model to parameterize the adiabatic temperature rise of concrete. Construction and Building Materials, 2021. 268. DOI10.1016/j.conbuildmat.2020.121656
14. Kumari, M. and G. Sahoo, Cost-Effective Resource Provisioning in Cloud Using Cooperative Coevolutionary Genetic Algorithm. Proceedings of the Fourth International Conference on Microelectronics, Computing and Communication Systems. MCCS 2019. Lecture Notes in Electrical Engineering, ed. V. Nath and J.K. Mandal. 2021. 391-401.
15. Bai, Y. and S.N. Amirikhian, KNOWLEDGE-BASED EXPERT-SYSTEM FOR CONCRETE MIX DESIGN. Journal of Construction Engineering and Management-Asce, 1994. 120(2): p. 357-373.
16. Ballim, Y., A numerical model and associated calorimeter for predicting temperature profiles in mass concrete. Cement & Concrete Composites, 2004. 26(6): p. 695-703.
17. Sok, T., Y.K. Kim, and S.W. Lee, Numerical evaluation of built-in temperature distribution effects on stress development in concrete pavements. Road Materials and Pavement Design, 2021. 22(4): p. 871-93.
18. Nishizawa, T., et al., Finite Element Model Analysis of Thermal Stresses of Thick Airport Concrete Pavement Slabs. Transportation Research Record, 2009/2095: p. 3-12.
19. Epps, A., Design and analysis system for thermal cracking in asphalt concrete. Journal of Transportation Engineering, 2000. 126(4): p. 300-307.
20. Cervera, M., J. Oliver, and T. Prato, Thermo-chemo-mechanical model for concrete. II: Damage and creep. Journal of Engineering Mechanics-Asce, 1999. 125(9): p. 1028-1039.
21. Wu, Y. and R. Luna, Numerical implementation of temperature and creep in mass concrete. Finite Elements in Analysis and Design, 2001. 37(2): p. 97-106.
22. Lucas, J.M., A. Berred, and C. Louis, Thermal actions on steel box girder bridge. Proceedings of the Institution of Civil Engineers-Structures and Buildings, 2003. 156(2): p. 175-182.
23. Deng, Z.Y., et al., Model analysis of multicrack mechanisms in ceramic/superplastic laminates. Journal of the American Ceramic Society, 2002. 85(7): p. 1793-1803.
24. Fransplass, H., M. Langseth, and O.S. Hopperstad, Experimental and numerical study of threaded steel fasteners under combined tension and shear at elevated loading rates. International Journal of Impact Engineering, 2015. 76: p. 118-125.
25. Liang, J.-F., et al., Influence of high temperature on mechanical properties of concrete containing recycled fine aggregate. Computers and Concrete, 2018. 21(1): p. 87-94.
26. Ahlawat, K. and A.P. Singh, Understanding Concepts, Future Trends, and Case Studies of Big Data Technologies. Proceedings of 2nd International Conference on Communication, Computing and Networking. ICCCN 2018. Lecture Notes in Networks and Systems, ed. C.R. Krishna, M. Dutta, and R. Kumar. 2019. 979-87.
27. Huang, J., et al., Technical Aspects and Case Study of Big Data based Condition Monitoring of Power Apparatuses, in IEEE PES Asia-Pacific Power and Energy Engineering Conference (IEEE PES APPEEC). 2014. IEEE Hong Kong Joint Chapter Power & Energy Ind Applicat Power Elect Ind E. Hong Kong, PEOPLES R CHINA.
28. Nowakowski, K. and W. Nowakowski, Some problems of BIG DATA technology, in Elektroinformatika, 2016. 57(1): p. 84-90.
29. Sankari, S. and S. Bose, Elaborative survey on storage technologies for XML big data: A real-time approach. 2016 Fifth International Conference on Recent Trends in Information Technology. 2016. DOI10.1109/ICRTIT.2016.7569569
Sankari, S., S. Bose, and Iee. Elaborative Survey on Storage Technologies for XML Big Data: A Real-time Approach. In 5th International Conference on Recent Trends in Information Technology (ICRTIT), 2016. Chennai, INDIA.

Shen, P., et al., Research on hydration heat temperature field in mass concrete side wall of large water pump station. 2011 International Conference on Consumer Electronics, Communications and Networks. 2011. 3436-9. DOI:10.1109/CECNET.2011.5769375

Wang, J., Y. Zhang, and X. Zhu. Experimental Study on Temperature Data of Mass Concrete Construction in Hot Weather Conditions. In 2nd International Conference on Civil Engineering, Architecture and Building Materials (CEABM 2012). 2012. Yantai, PEOPLES R. CHINA.

Wei, D., et al., Temperature field of mass concrete foundation with pipe-cooling. Journal of Tianjin University (Science and Technology), 2016. 49(4): p. 437-42.

Lin, P., Q. Li, and P. Jia, A Real-Time Temperature Data Transmission Approach for Intelligent Cooling Control of Mass Concrete. Mathematical Problems in Engineering, 2014. 2014. DOI:10.1155/2014/514006

Kim, J., et al., Concrete temperature monitoring using passive wireless surface acoustic wave sensor system. Sensors and Actuators a-Physical, 2015. 224: p. 131-139.

Norris, A., M. Saafi, and P. Romine, Temperature and moisture monitoring in concrete structures using embedded nanotechnology/microelectromechanical systems (MEMS) sensors. Construction and Building Materials, 2008. 22(2): p. 111-120.

Riding, K.A., et al., Evaluation of temperature prediction methods for mass concrete members. ACM Materials Journal, 2006. 103(5): p. 357-365.

Bobko, C.R., V.Z. Zadeh, and R. Seracino, Improved Schmidt Method for Predicting Temperature Development in Mass Concrete. ACM Materials Journal, 2015. 112(4): p. 579-586.

Riding, K.A., et al., Temperature boundary condition models for concrete bridge members. ACM Materials Journal, 2007. 104(4): p. 379-387.

Wasala, W.M.T.D and H.D. Yapa, Prediction of Temperature Development in Concrete using Semi-adiabatic Temperature Measurements. EngineerJournal of the Institution of Engineers Sri Lanka, 2017. 50(3): p. 1-8.

Barbhuiya, S. and D. Kumala, Behaviour of a Sustainable Concrete in Acidic Environment. Sustainability, 2017. 9(9). DOI:10.3390/su9091556

Mora, B.A. and M. Hilpert, Differences in Infiltration and Evaporation of Diesel and Gasoline Droplets Spilled on Concrete Pavement. Sustainability, 2017. 9(7). DOI:10.3390/su9091271

Kodur, V., S. Banerji, and R. Solimhinziaei, Effect of Temperature on Thermal Properties of Ultra-High-Performance Concrete. Journal of Materials in Civil Engineering, 2020. 32(8). DOI:10.1061/(ASCE)MT.1943-5533.0003286

Yang, J., et al., Thermal analysis of mass concrete embedded with double-layer staggered heterogeneous cooling water pipes. Applied Thermal Engineering, 2012. 35: p. 145-156.

Bantel, M., Uncertainty of measurement. Instruments India, 1971. 6(10): p. 11-26.

Schwertfeger, H.J., Correction of accidental errors by cyclic codes in the form of direct decoding by cyclic permutation. Nachrichtentechnik, 1972. 22(5): p. 139-143.

Merigo, J.M. and R.R. Yager, Aggregation operators with moving averages. Soft Computing, 2019. 23(21): p. 10601-10615.

Mohammadpour, M. and A.R. Soltani, Forward Moving Average Representation in Multivariate MA(1) Processes. Communications in Statistics-Theory and Methods, 2010. 39(4): p. 729-737.

Gugaliya, J.K., R.D. Gudi, and S. Lakshminarayanan, Multimodel decomposition of nonlinear dynamics using a fuzzy-CART approach. Journal of Process Control, 2005. 15(4): p. 417-434.

Li Baoyu was born in December 1963 , Associate Professor. In July 1986, she graduated from North China University of Water Conservancy and Hydropower, majoring in engineering geology and hydrogeology. From July 1986 to now, she has taught at Henan Vocational College of Water Conservancy and Environment (formerly Zhengzhou Water Conservancy School). She has successively served as a lecturer on engineering geology and hydrogeology, soil mechanics, foundation foundation, building materials and other courses for higher vocational, secondary vocational and TV students, with an average of more than 600 class hours per year, published many papers, and participated in the compilation of many books. In 2001, she was rated as an excellent teacher in the "Striving for Three Excellences" activity. In 2003, the course "Engineering Geology and Rock and Soil Mechanics" was rated as the Teaching Excellence Award. In the "Three Education People" activity in the 2003/2004 school year, he was rated as an advanced individual. ; In 2001, she was rated as an excellent teacher in the "Striving for Three Excellences" activity, and in 2003, the course "Engineering Geology and Rock and Soil Mechanics" was rated as the Teaching Excellence Award. She has successively won the outstanding tutors of the 2nd National Water Conservancy Secondary Vocational School "Zhongyuan Cup" Skills Competition, the excellent referee and proposition expert of the 3rd National Water Conservancy Secondary Vocational School "Dunhuang Cup" Skills Competition, and the 4th National Water Conservancy
Secondary Vocational School "Gan" Excellent referee and proposition expert in Po Cup” skills competition.

**Li Guoxing**, was born in October 1976, is a professor-level senior engineer. He is currently the director of the Engineering Technology Department of the Second Engineering Bureau of Henan Province, mainly engaged in the construction management of water conservancy projects. Professor Li has been engaged in water conservancy work for more than 20 years, and has participated in a number of national and provincial key engineering construction projects. As the project leader, he has completed project construction and assessment, and has won many notifications and awards. He has published more than 10 academic papers and applied for more than 20 patents.

**Wang Guiyu** was born in May 1989, is an engineer. In 2015, he graduated from Zhengzhou University majoring in water conservancy and hydropower engineering with a master's degree. He is currently the deputy director of the Engineering Technology Department of the Second Engineering Bureau of Henan Province, mainly engaged in the construction management of water conservancy projects. As the head of the technology department, she led the company’s water conservancy technology innovation, won a number of scientific and technological achievements awards, published 3 academic papers, and applied for more than 10 patents.

**Zhang Guofeng** was born in May 1967, is a professor-level senior engineer. He is currently the director of the Henan Provincial Water Conservancy Second Engineering Bureau, in charge of the overall work, mainly engaged in water conservancy project construction management and enterprise management. Professor Li has been engaged in water conservancy work for more than 30 years, and has led and participated in more than ten national, provincial and ministerial key construction projects and won more than 20 honors. He has published more than 10 academic papers and applied for more than 10 patents.

**Yang Man** was born in Henan, China, in 1987. She received the B.S. and M.S. degree in water conservancy and hydropower engineering from Zheng Zhou University, Zheng Zhou China, in 2010 and 2013, respectively and the Ph.D. degree in high-way and railway engineering at Zheng Zhou University, Zheng Zhou, China, in 2020. In 2020, she joined the Zheng Zhou University China. Her research interests include nondestructive detection technique in engineering and numerical simulation methods (e.g., finite-difference time-domain method, finite-element method, symplectic methods).