International Core Journal of Engineering Volume 9 Issue 7, 2023
ISSN: 2414-1895 DOI: 10.6919/ICJE.202307_9(7).0006

Several ML Algorithms and Their Feature Vector Design for Gas Discrimination and Concentration Measurement with an Ultrasonically Catalyzed MOX Sensor

Jiufu Yan, Weimin Peng, Lianghao Zhu, Nan Pu, Mengda Jiang

Chongqing Water Resources and Electric Engineering College, China

Abstract

Because the change values of luminous signals and electrical signals of different gases are different, different gases can be identified by analyzing the characteristic signals of gases. The response curve of the gas sensor is divided into three stages: response stage, steady state and recovery stage. The article mainly introduces the severe ml algorithms and the infrared feature vector design for gas discrimination and concentration measurement with an ultrasonically catalyzed MOX sensor. Compared with optical sensors, Ultrasonically catalyzed MOX sensor, as a low-cost and low-power sensor, meets the requirements of wireless sensor networks used in home automation and environmental monitoring.The overall accuracy of KNN(K- nearest neighbor) algorithm under the third scheme is 59.84%, the overall accuracy of scheme 2 is close to that of scheme 3, and the overall accuracy of scheme 1 is the lowest. The average classification accuracy of RF (Random Forest) model for three gases with different concentrations is 0.84, 0.90, 0.91 and 0.92, respectively. Using ML technology can improve the classification efficiency of gases with different concentrations, which is the trend of gas sensor data classification in the future. The prediction accuracy of the CNN-LSTM (Convolutional Neural Network-Long Short Term Memory) algorithm in mixed gases is significantly better than that of the individual LSTM algorithm and other ML algorithms. Taking into account accuracy, efficiency, and complexity, the proposed algorithm is suitable for applications where concentration mutations may occur, such as continuous monitoring applications or chemical source localization, with a wider range of application scenarios.

Keywords

ML; Ultrasonically Catalyzed MOX Sensor; Gas Discrimination and Concentration.

1. Introduction

There are some problems in the actual detection process of gas sensors, such as nonlinear response, single selectivity and low information capacity. Therefore, researchers use the working mode of dynamic temperature modulation to collect the response information of gas sensors and combine it with neural network model to form an intelligent detection system to improve the accuracy and sensitivity of detection [1-2]. Compared with optical sensors, ultrasonically catalyzed MOX sensor, as a low-cost and low-power sensor, meets the requirements of wireless sensor networks used in home automation and environmental monitoring. Because the change values of luminous signals and electrical signals of different gases are different, different gases can be identified by analyzing their characteristic signals. However, considering the simplification, stability and miniaturization of the instrument, there is still a lot of room for the development of sensors for substance identification, and it is necessary to develop a sensor with fewer sensing units and detectors and rich information to

ISSN: 2414-1895 DOI: 10.6919/ICJE.202307_9(7).0006

identify analytes [3]. The article mainly introduces the severe ml algorithms and the infrared feature vector design for gas discrimination and concentration measurement with an ultrasonically catalyzed MOX sensor.

2. Application of KNN Algorithm in Gas Discrimination and Concentration Measurement

There are many kinds of gas sensors. At present, the commonly used gas sensors include electrochemical gas sensor, semiconductor ultrasonically catalyzed MOX sensor and optical fiber gas sensor. Fiber gas sensor has the characteristics of small volume, wide frequency band, low transmission loss, strong electromagnetic interference resistance and large amount of information, but it is not suitable for trace gas detection and the collected signal is weak [4]. In recent years, with the rapid development of electronic nose technology, people combine gas sensor array with sensor fusion technology, computer technology, applied mathematics and other fields for gas detection and pattern recognition.

In the sensing process of gas sensor array, there is often some noise interference, and some gases have weak response in the sensing process, so the gas sensor array obtains mainly noise signals, which will cause great interference to the subsequent gas identification and gas concentration regression using ML algorithm [5-6]. The gas classification task, like other ML classification tasks, includes feature engineering, model training and model application for classification. The accuracy of classifier is closely related to feature engineering and model training. Therefore, under specific tasks, we need to constantly compare experiments to get the best classification algorithm model.

We use three feature engineering scheme to carry out experiments with classification methods:

Scheme1: firstly, carry out Z-score standardization, and then use PCA (principal component analysis)to extract features;

Scheme2: first use polynomial-based data transformation, and then carry out Z-score standardization;

Scheme3: First, use polynomial-based data transformation, then use variance method to extract features, and then carry out Z-score standardization.

We will use KNN(K- nearest neighbor) algorithm with Euclidean distance to classify the gas sensor data in this paper. The calculation formula of Euclidean distance in gas sensor data is as follows:

$$L(a,b)=\left(\sum_{i=1}^{D}(a_i,b_i)^2\right)^{\frac{1}{2}} \tag{1}$$

Where a,b is the eigenvector of two gas samples whose distance needs to be calculated, a_i,b_i is the i eigenvalue of the two gas samples, and D is the characteristic number of the gas samples.

Next, we use the optimal k value of each scheme to carry out experiments, and the prediction results on 3 to 10 batches of data are shown in Figure 1.

It can be seen that the prediction accuracy of the three scheme for the third batch is above 70%, which is because the third batch is least affected by drift, so the accuracy is higher than other batches; Overall, the overall accuracy of KNN algorithm under the third scheme is 59.84%, the overall accuracy of scheme2 is close to scheme3, and the overall accuracy of scheme1 is the lowest.

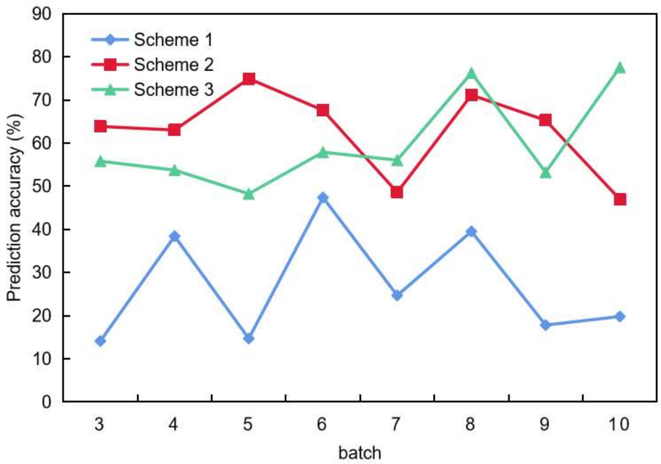

Figure 1. Prediction accuracy of KNN algorithm in 3 to 10 batches

3. Gas Discrimination and Concentration Measurement based on RF

Due to the cross-sensitivity of gas sensors and the phenomenon of gas drift, as well as the influence of external factors in the working environment, it is often difficult for a single sensor working in various gas environments to accurately measure the target gas [7], so how to accurately analyze and classify the monitoring data during data transmission is a difficult problem at present.

The essence of RF(Random forest) algorithm is a model composed of several decision trees, while RF algorithm takes all decision trees in the model to vote for the decision conclusion, and the one with the highest vote is the basic conclusion of classification [8-9]. Choose the data gain of data sample information to divide the category of optimal attribute into ID3 decision tree algorithm. Information entropy is an index reflecting the purity of statistical sample set, and information gain ed comes from information entropy, and the expression is:

$$E_d = -\sum_{k=1}^{|y|} P_k \log_2 P_k \tag{2}$$

Where: d is the data sample set; P_k is the percentage of k sample in the total sample.

The operation process of RF model is to use self-help sampling method to train the training set, and then generate several decision trees through the training results, and use the generated decision trees to classify the samples, and finally take the classification result of RF model with the most votes.

The essence of RF model is to combine multiple decision trees, each tree is independent and has the same distribution form, and the classification error of each tree is related to the correlation of each tree. In feature selection, each node is randomly segmented and the errors in different situations are compared. Using MATLAB platform, an RF algorithm model with 100 decision trees is established. After the model is built, the experimental system is selected to respond to four groups of output signals with different volume fractions and 100 eigenvector data, with a total of 300 groups of experimental data. The experimental data were randomly divided into training set and experimental set at the ratio of 7: 3.

The processed data is imported into the RF model for training, and the average value of each evaluation index of the model is obtained by many experiments and statistical classification results, as shown in Figure 2.

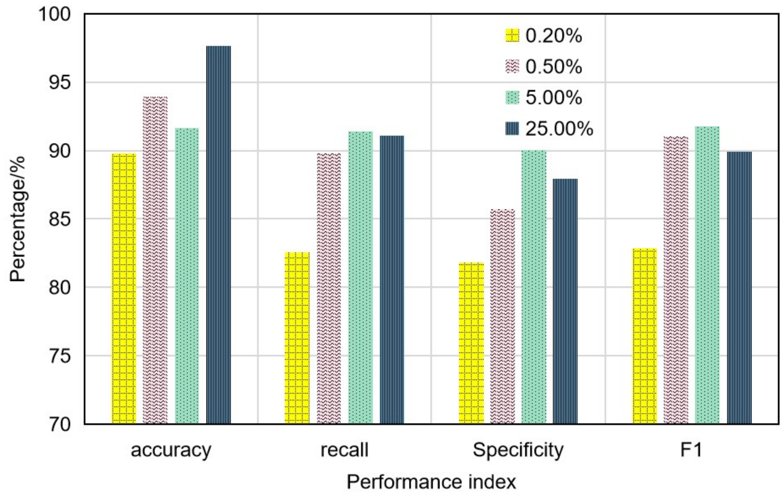

Figure 2. RF model performance index

It can be seen that the average classification accuracy of RF model for three gases with different concentrations is 0.84, 0.90, 0.91 and 0.92, respectively. That is to say, under ideal conditions, the average classification accuracy of RF model for different gases can reach 0.92, which greatly exceeds the accuracy of conventional methods. Using ML technology to classify gases can significantly reduce labor costs, shorten the classification time and improve the classification efficiency of gases with different concentrations, which is the trend of gas sensor data classification methods in the future.

4. Feature Vector Design

The response curve of gas sensor is divided into three stages: response stage, steady state and recovery stage. The response time is the time from the beginning of the reaction to the signal reaching the steady state after the measured gas enters the reaction chamber, and the recovery time is the time for the response signal to recover from the steady state value to near the baseline after the background gas is filled into the reaction chamber. When gas sensors are first applied, they tend to be relatively stable. However, with the increase of using time, the output signal in clean background gas will be different from before, that is, the baseline will drift. In order to achieve the effect of compensating baseline drift, gas detection experiments often carry out baseline processing on sensor data in preprocessing, such as difference, relative difference and fractional difference.

The detection and analysis of mixed gas is more complicated than that of single gas, and the response of ultrasonically catalyzed MOX sensor to mixed gas is lower than the simple addition value of the response to individual test of each gas. This is because the mixed gas itself has an influence on each other, and the gas with higher activity will have the privilege to consume the oxygen adsorbed on the oxide surface with lower activity, resulting in the subsequent reaction of the gas with lower reactivity. Therefore, compared with the gas in a single environment, the response of the sensor to the gas with lower activity becomes smaller [10]. Therefore, this paper puts forward a deep learning algorithm and shows the feasibility of this algorithm to predict the concentration of mixed gas quickly through subsequent experiments.

Because the data of ultrasonically catalyzed MOX sensor is a gas signal response curve based on time series, and it has nonlinear and high-dimensional characteristics, the information in the data can be learned more effectively by using the classical nonlinear calculation method of artificial neural network. The proposed CNN-LSTM (Convolutional Neural Network-Long Short Term Memory) algorithm consists of a convolutional layer, a recursive layer, and a fully connected layer. The convolution layer in the first stage is a feedforward neural network, which is added to preliminarily extract spatial features from response "fragments" with a fixed length of time. A convolution layer includes convolution operation, batch normalization operation and activation operation.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202307_9(7).0006

Gas concentration prediction is a classical regression problem. MSE (mean square error) is used as the loss function of the algorithm in the experiment, as shown in Formula (3).

$$Loss = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{3}$$

In the formula, y_i is the label value and $\hat{y}_i$ is the predicted value.

The algorithm proposed in this paper is implemented using Pytorch framework. In this paper, the experiment is based on public data sets to verify the effect of the proposed CNN-LSTM neural network algorithm, and the model is used to predict the concentration of mixed gas in a limited response time. Figure 3 shows the Loss curve of CNN-LSTM, and it can be seen that the model converges quickly.

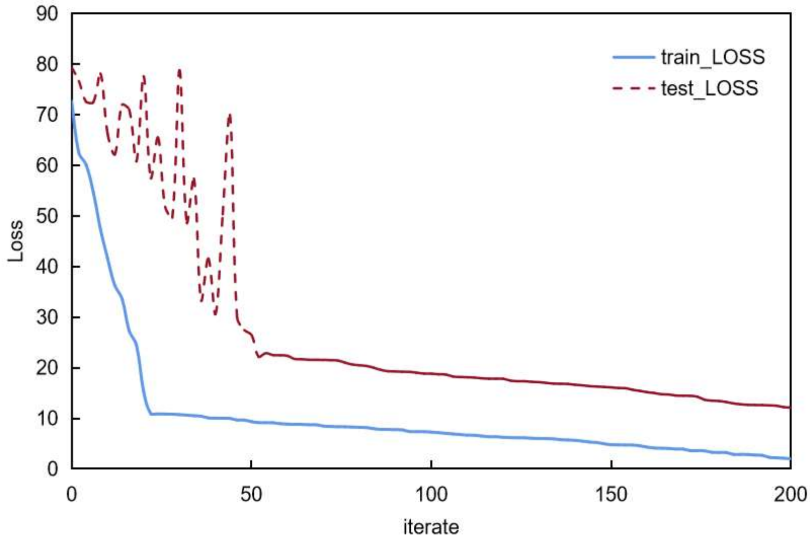

Figure 3. Loss curve of CNN-LSTM

Aiming at the low accuracy and efficiency of gas mixture concentration prediction caused by the nonlinear response characteristics of sensor array, a new algorithm combining CNN and LSTM is proposed in this chapter. Because of its nonlinearity and long-term memory, it overcomes the above limitations and successfully predicts the concentration in the mixture.

The prediction accuracy of CNN-LSTM algorithm in mixed gas is obviously better than that of LSTM algorithm alone and other ML algorithms. Therefore, considering the accuracy, efficiency and complexity, the proposed algorithm is suitable for applications where concentration mutation may occur, such as continuous monitoring applications or chemical source location, and the application scenarios are more extensive.

5. Conclusion

Compared with optical sensors, Ultrasonically catalyzed MOX sensor, as a low-cost and low-power sensor, meets the requirements of wireless sensor networks used in home automation and environmental monitoring. The overall accuracy of KNN algorithm under the third scheme is 59.84%, the overall accuracy of scheme2 is close to scheme3, and the overall accuracy of scheme1 is the lowest. The average accuracy of RF model for three gases with different concentrations is 0.84, 0.90, 0.91 and 0.92, respectively. Using ML technology to classify gases can significantly reduce labor costs, shorten the classification time and improve the classification efficiency of gases with different concentrations, which is the trend of gas sensor data classification methods in the future. The prediction accuracy of CNN-LSTM algorithm in mixed gas is obviously better than that of LSTM algorithm alone and other ML algorithms.

References

[1] Ederington J, Sandford J. Employer discrimination and market structure: Does more concentration mean more discrimination?[J]. International Journal of Industrial Organization, 2016, 48(9):1-33.

[2] CM Jerónimo, BR Álvaro, SA Ángel. Modelling of Biomass Concentration, Multi-Wavelength Absorption and Discrimination Method for Seven Important Marine Microalgae Species[J]. Energies, 2018, 11(5):1089.

[3] Yen T Y, Yao D J. Detection of the Freshness of Kiwifruit with a TD-GC-MS and a Gas-sensing Array Based on the Surface-acoustic-wave Technique[J]. IEEE Transactions on NanoBioscience, 2021, (99):1-1.

[4] Gabriele M, Alexandro C, Raj K, et al. Conductive Photo-Activated Porphyrin-ZnO Nanostructured Gas Sensor Array[J]. Sensors, 2017, 17(4):747.

[5] Abbatangelo M, Nunez-Carmona E, Sberveglieri V. Application of a novel S3 nanowire gas sensor device in parallel with GC-MS for the identification of Parmigiano Reggiano from US and European competitors[J]. Journal of Food Engineering, 2018, 236(10):36-43.

[6] Li H, Yu S, Yang W, et al. Flow Pattern Identification for Gas-Liquid Two-Phase Flow in Minichannel Based on a New Optical Array Sensor[J]. IEEE Sensors Journal, 2021, (99):1-1.

[7] Liu Q, Hu X, Mao Y, et al. Gas Recognition under Sensor Drift by Using Deep Learning[J]. International Journal of Intelligent Systems, 2015, 30(8):907-922.

[8] Matindoust S, Baghaei-Nejad M, Abadi M, et al. Food quality and safety monitoring using gas sensor array in intelligent packaging[J]. Sensor Review, 2016, 36(2):169-183.

[9] Iwaszenko S, Kalisz P, Sota M, et al. Detection of Natural Gas Leakages Using a Laser-Based Methane Sensor and UAV[J]. Remote Sensing, 2021, 13(3):510.

[10] Bang Y K, Lee C H. Design of a hierarchically structured gas identification system using fuzzy sets and rough sets[J]. Transactions of the Korean Institute of Electrical Engineers, 2018, 67(3):419-426.