Research on Safety Early Warning Technology for Road Sections with Poor Sight Distance based on Acoustic Signals

Xiaonan Cheng, Xin Chen*, Jian Li, Jielong Song and Xiankang Tang

School of Automation, Nanjing University of Science and Technology, Nanjing, Jiangsu, 210094 China

*Corresponding author's e-mail: 1094279526@qq.com

Abstract. In order to study the technical issues of traffic safety early warning on roads with poor sight distance, we first collect the sounds of vehicles running on the road, select representative vehicle sounds as recognition objects, pre-emphasize the vehicle sound signals, add windows into frames, and calculate the power spectrum. Input the Mel filter bank to obtain the MFCC of the vehicle sound signal and construct a feature vector with its characteristic values; then the BP neural network algorithm is improved to classify and identify the vehicle signal, thereby achieving the purpose of vehicle identification. Experiments show that the accuracy of the proposed voice recognition technology reaches more than 90%. This technology can be applied to road sections with poor sight distance to identify passing vehicles.

Keywords: Traffic safety 1; Vehicle identification 2; MFCC 3; Improved BP algorithm 4

1 Introduction

Transportation systems have become the basis for economic growth in all countries. However, in the current road traffic environment, dangerous road sections such as blind spots have become a major safety hazard. Complex and changeable terrain conditions make it difficult for drivers to predict road conditions, and unpredictable safety problems are very likely to occur. Therefore, for this type of special road section, there is an urgent need for a vehicle classification technology that can accurately remind drivers, have low-cost sensors, can be deployed in large quantities, and issue early warnings to drivers in a timely manner. It is a key vehicle classification technology to improve road traffic safety [1].

In recent years, video image recognition technology has been widely used in the field of vehicle detection, and its significant advantages lie in the easy installation of equipment, rich information acquisition and high real-time performance [2-3]. However, this method also faces some challenges, such as the relatively cumbersome recognition process, and the detection effect is easily interfered by a variety of external factors, such

G. Zhao et al. (eds.), *Proceedings of the 2024 7th International Symposium on Traffic Transportation and Civil Architecture (ISTTCA 2024)*, Advances in Engineering Research 241,
https://doi.org/10.2991/978-94-6463-514-0_39

as weather and environment [4]. Comparatively, the vehicle recognition method based on acoustic signals has gradually received attention from researchers due to its unique advantages. This method utilizes the acoustic sensing network to collect data, and realizes the classification and recognition of vehicles through the comprehensive application of acoustic classification and recognition technology. It is not limited by external factors such as weather and environment, while it does not require real-time acquisition, which significantly reduces the cost of signal acquisition, storage and processing. Therefore, the vehicle classification and recognition method based on acoustic signals has become a hot spot of current research at home and abroad. For example, a variety of machine learning algorithms have demonstrated significant success rates in the task of vehicle classification with acoustic signals [5]. Sharma et al [6] utilized the methods of spectral statistics and wavelet transform to analyze the vibration signals of vehicles in the time-frequency domain; Kandpal et al [7] combined the Fourier transform and the time-domain waveform analysis to study the vehicle acoustic signals, and applied the neural network for classification; Yang et al [8] used discrete spectrum analysis to extract the features of vehicle sound signals, and proposed that in recent years, sound recognition technology has been widely used in the field of vehicle detection, and its significant advantage lies in the easy installation of equipment, rich information acquisition and high real-time.

Applied sound detection technology has shown strong application potential in various fields in recent years, from environmental monitoring to equipment maintenance, and has achieved remarkable results [9]. Due to its powerful nonlinear mapping ability and good learning performance, BP neural network has a small model and high accuracy. Therefore, this paper uses the improved BP neural network to identify the sound signals of passing vehicles. By training the neural network model, it can be obtained from Extract valuable features from sound signals and perform accurate classification or identification.

This research aims to explore the application of traffic safety based on sound detection and BP neural network recognition technology on dangerous road sections such as blind spots. It is expected to improve road traffic safety and protect people's lives and property through timely and accurate early warning systems. Especially in the field of transportation, the sound signals generated by passing vehicles contain rich information, such as vehicle speed, vehicle type, etc. If this sound information can be effectively used, it will be of far-reaching significance for improving road safety. Especially in adverse weather conditions such as heavy rain, strong winds, storms or fog, which cause reduced visibility and worsened road conditions, the risk of such an accident nearly doubles. When an accident occurs in such an environment, prompt and effective notification of approaching vehicles becomes critical. Failure to notify in time will most likely trigger a series of chain reactions, namely a multi-vehicle collision (MVC) accident. Therefore, the research and development of safety early warning system technology for road sections with poor sight distance based on sound detection is particularly important. It can provide timely warning through sound detection when the sight distance is limited, thereby significantly reducing the probability and severity of accidents.

2 Method Introduction

2.1 Extraction of Sound Signal Features

The most commonly used speech feature for sound feature extraction is the Mel Falling Spectral Coefficient (MFCC). According to the research of human ear hearing mechanism, it is found that the human ear has different hearing sensitivity to sound waves of different frequencies. Speech signals from 200Hz to 5000Hz have a large impact on the clarity of speech. The specific operation steps are as follows [10-12]:

(1) Preprocessing: Audio signal is a kind of short-term stability but long-term change of complex signals. In 10-30ms, the "frame" is approximately stable, so the recognition system is based on its short-term analysis. Before extracting the features, it is necessary to do pre-emphasis and window frame processing. The transfer function of pre-emphasis is:

$$H(z) = 1 - \mu z^{-1} \tag{1}$$

In Eq. (1), $\mu \in [0.9,1]$ and in general $\mu = 0.9375$.

Pre-emphasis reduces the overall amplitude of the signal at the expense of enhancing the amplitude of the high frequency components.

(2) Short-time energy: short-time energy is one of the commonly used features in audio analysis. The short-time energy of the audio signal changes with time relatively obviously, which can show the characteristics of audio better. The short-time energy at the beginning of the nth frame when the window is added is defined as:

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 \tag{2}$$

(3) Fast Fourier Transform (FFT): Since it is difficult to observe the signal characteristics in the time domain, it is usually converted into a frequency domain energy distribution. Different speech characteristics correspond to different energy distributions. Thus, after multiplying the Hamming window, each frame also needs to go through the Fast Fourier Transform (FFT) to get the spectral energy distribution. The spectrum of the speech signal after taking the mode square is the power spectrum. The process can be expressed as:

$$X_a(k) = \sum \quad _{N-1}^{n=0} x(n) e^{-\frac{2\pi ik}{N}}, 0 \le k \le N \tag{3}$$

Eq. (3) where is the input speech signal and N denotes the number of points of Fourier transform.

(4) Mel Frequency Cepstrum Coefficients (MFCC): The frame signal is FFT transformed to obtain the spectrum and the discrete power spectrum is computed and then delivered to the Mel filter bank. In essence, the Mel filter bank is a series of triangular bandpass filters, which are characterized spectrally by having many filters in the low frequency region and fewer filters in the high frequency region. This reflects the linear distribution characteristic over the Mel spectrum. The conversion between frequencies and Mel frequencies is shown below:

$$f(M) = 700\left(e^{\frac{M}{1125}} - 1\right) \tag{4}$$

After taking the logarithm, it is again converted to cepstrum by Discrete Cosine Transform (DCT) to output the desired Mel cepstrum coefficients. Eventually, we can extract the MFCC feature coefficients from the converted Mel cepstrum.

2.2 BP Neural Network

Artificial neural network is a mathematical model that simulates the information processing of biological neural networks, in particular, BP neural network (error back propagation neural network) is one of them, which adopts multi-layer feed-forward structure and BP algorithm to adjust the weights in order to realize the nonlinear mapping from the input to the output. Its network structure is shown in Figure 1. The BP algorithm is divided into the learning and training stage and the target recognition stage, which adjusts the parameters of the network by inputting the training samples, so that the output is close to the desired value; in the recognition phase, the network then determines the class of the input signal based on the trained weights.

The original BP neural network is based on the idea of error back propagation, by calculating the partial derivatives (i.e., gradients) corresponding to each weight, and then updating the weights according to the principle of gradient descent. A key characteristic of this method is that the update direction of the weights is completely determined by the current gradient, which only takes into account the local information and does not utilize the parameter to update the historical information, and thus may oscillate or even fall into a local minimum during the training process, which makes the training efficiency and performance of the model limited.

In contrast, the BP neural network with momentum introduces a momentum term to improve this problem. The design idea of the momentum term is to refer to the concept of momentum or velocity in physics, retaining the direction and magnitude of the previous weight update, so that the update of the weights is not only affected by the current gradient, but also by the historical update. This can make the weight update smoother and reduce the oscillations in the training process, and also overcome the gradient vanishing and local minimum problems to a certain extent, accelerate the convergence of the model, and improve the performance of the model [13]-[14].

The method of dynamically adjusting the learning rate can be specifically realized as the following steps:

(1) Initialize the learning rate: first, we set an initial learning rate value for each weight parameter. Suppose we have n parameters, and use $a(0) = [a_1, a_2, \cdots a_n]^T$ to denote the initial learning rate vector.

(2) Calculate the gradient and gradient change rate: in each iteration, we not only need to calculate the gradient of each parameter, but also need to calculate the gradient change rate.

Assuming that $g(t) = [g_1(t), g_2(t), \cdots g_n(t)]^T$ is the gradient vector of the tth iteration, then the rate of change of the gradient r(t) can be defined as:

$$r(t) = \frac{|g(t)-g(t-1)|}{|g(t)-1|} \tag{5}$$

where "|" denotes the second paradigm of the vector, i.e., the square root of the sum of squares of the vector elements.

(3) Dynamically adjusting the learning rate: next, we can adjust the learning rate of each parameter according to the rate of change of the gradient. A simple strategy is to increase the corresponding learning rate if the gradient change rate is less than a preset threshold δ (e.g., $\delta = 0.01$); otherwise, decrease the corresponding learning rate. This can be expressed as:

$$a_i(t) = a_i(t-1) * (1+\beta), r_i(t) \prec \delta \tag{6}$$

$$a_i(t) = \frac{a_i(t-1)}{1+\beta}, r_i(t) \geq \delta \tag{7}$$

Where $a_i(t)$ is the learning rate of the tth iteration parameter i, $r_i(t)$ is the gradient change rate of the tth iteration parameter i, and β is a hyperparameter regulating the change amplitude of the learning rate.

(4) Update parameters: finally, each parameter is updated using a new learning rate vector $a(t) = [a_1(t), a_2(t), \cdots a_n(t)]^T$:

$$w(t) = w(t-1) - diag\big(a(t)\big) * g(t) \tag{8}$$

Where $w(t) = [w_1(t), w_2(t), \cdots w_n(t)]^T$ is the parameter vector of the first t iteration, $diag(a(t))$ denotes the conversion of the vector $a(t)$ to a diagonal matrix.

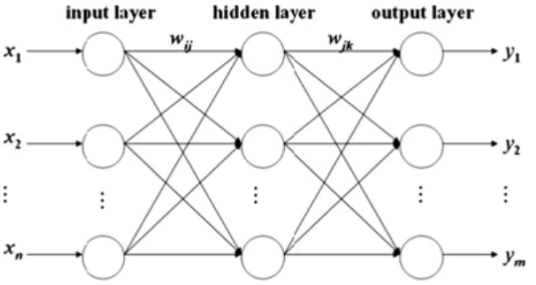

Fig. 1. Neural network structure diagram.

3 Experiment ang Result

This experiment collected several acoustic scenarios representing the sounds produced by vehicles traveling on the road. The sounds produced by several typical vehicles passing by the roadway were recorded: small vehicles, medium-sized vehicles, and large vehicles. Subsequently, common noises on the road were recorded: vehicle honking, wind, and human speech. In addition, combinations of these noises were recorded, with 100 samples of each sound, in order to better reflect the real situation. Also, to ensure the integrity of the data and the quality of the collected recordings, recordings were

made in wav format only when a single vehicle was passing by, with a set sampling frequency of 44.1 kHz.

Vehicle driving sounds are complex, multiple noise sources and non-smooth signals, whereas the Fourier transform is suitable for smooth signals. Therefore, it is common to treat such signals as smooth for a short period of time (e.g., 10-30ms) and process them into multiple "smooth" signals by splitting them into frames. The number of frames depends on the frame length and the frame shift, and the frame shift is set to 1/2 the frame length to prevent signal loss, which is also part of the short-time Fourier transform. The Hamming window is added to the audio signal, and the length of the window function is taken as 15-30 ms. All the feature parameters are obtained after the above steps, and in order to carry out the supervised learning, it is necessary to label the speech signals, i.e., to give each speech sample a corresponding classification label.

Next, we will use MATLAB to use the improved BP neural network algorithm to design a classifier for identifying three different vehicle acoustic signals. First, we collected 100 sets of each type of vehicle acoustic signal and noise, for a total of 400 sets of data. In order to ensure the generalization ability of the model, we divide the data set into training samples and test samples, of which 70% or 280 sets of data are used as training samples, and the remaining 30% or 120 sets of data are used as test samples. There is no intersection.

Since the speech feature input signal has 24 dimensions, and there are 4 categories of speech signals that need to be classified (including three vehicle sound signals and one possible other category), we set the structure of the BP neural network to 24-25-4, that is, the input The layer contains 24 nodes to correspond to the dimensions of the input features, the hidden layer contains 25 nodes to process nonlinear features, and the output layer contains 4 nodes to correspond to 4 possible classification results.

During the training process, we use training samples to train the BP neural network. Since there are a total of 280 sets of training samples, we ensure that these samples are fully utilized to train the neural network. After the training is completed, we will use the remaining 120 sets of test samples to test the trained neural network to evaluate its classification ability.

Through the improved BP neural network algorithm, we expect to be able to accurately distinguish these three different vehicle acoustic signals. Figure 2 and Table 1 list the results of two different algorithms (including traditional BP neural network and improved BP neural network) in identifying these vehicle acoustic signals. Through comparison, it can be seen that the improved algorithm has better classification accuracy. Whether there is any improvement.

Table 1. Recognition accuracy table

Type	BP algorithm	Improved BP algorithm
Small car	83%	91%
Mid-size car	84%	92%
Large car	76%	93
Noise	15%	10%

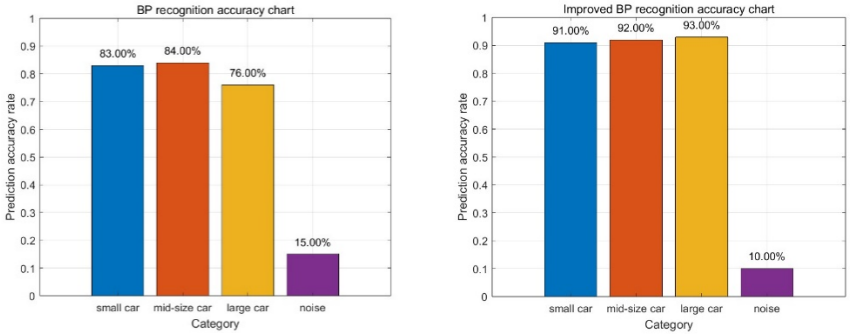

Fig. 2. Original bp recognition accuracy chart and improved bp recognition accuracy chart

4 Conclusion

This study explores the application of sound recognition technology in traffic safety early warning on waterfront and cliff sections. By collecting, processing and analyzing vehicle sound signals, an improved BP neural network algorithm is used to achieve high-accuracy classification and recognition of vehicle sounds (90 %above). However, the generalization ability of sound samples and testing in actual road environments requires further research and verification. Nonetheless, this research provides new perspectives and tools for road traffic safety warning, which has potential application value. Future research will be dedicated to improving existing methods and exploring more applications of sound recognition technology in vehicle identification and traffic safety warning.

References

1. Shokravi, H.; Shokravi, H.; Bakhary, N.; Heidarrezaei, M.; Rahimian Koloor, S.S.; Petrů, M. A Review on Vehicle Classification and Potential Use of Smart Vehicle-Assisted Techniques. Sensors 2020, 20, 3274. https://doi.org/10.3390/s20113274.
2. Shobha, B. S., and R. Deepu. "A review on video based vehicle detection, recognition and tracking." 2018 3rd International conference on computational systems and information technology for sustainable solutions (CSITSS). IEEE, 2018.
3. Ali, Mohsin, Muhammad Atif Tahir, and Muhammad Nouman Durrani. "Vehicle images dataset for make and model recognition." Data in brief 42 (2022).
4. Yu, Ye, et al. "Embedding pose information for multiview vehicle model recognition." IEEE Transactions on Circuits and Systems for Video Technology 32.8 (2022): 5467-5480.
5. Sreenivas Sremath Tirumala, Seyed Reza Shahamiri, Abhimanyu Singh Garhwal, Ruili Wang, Speaker identification features extraction methods: A systematic review, Expert Systems with Applications, Volume 90, 2017, Pages 250-271.
6. Sharma, Navdeep, Anoop Kumar Jairath, Bhopendra Singh, and Ashutosh Gupta. "Detection of various vehicles using wireless seismic sensor network." In 2012 International Conference on Advances in Mobile Network, Communication and Its Applications, pp. 149-155. IEEE, 2012.

7. Kandpal, Manisha, Varun Kumar Kakar, and Gaurav Verma. "Classification of ground vehicles using acoustic signal processing and neural network classifier." In 2013 international conference on signal processing and communication (ICSC), pp. 512-518. IEEE, 2013.
8. Yang, Seung S., Yoon G. Kim, and Hongsik Choi. "Vehicle Identification using Discrete Spectrums in Wireless Sensor Networks." J. Networks 3, no. 4 (2008): 51-63.
9. Sammarco, M. and Detyniecki, M. Crashzam: Sound-based Car Crash Detection. [C] In Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems - VEHITS; ISBN 2018:978-989-758-293-6;
10. Mateen, A.; Hanif, M.Z.; Khatri, N.; Lee, S.; Nam, S.Y. Smart Roads for Autonomous Accident Detection and Warnings. Sensors 2022, 22, 2077. https://doi.org/10.3390/s22062077
11. S. Sathruhan, O. K. Herath, T. Sivakumar and A. Thibbotuwawa, "Emergency Vehicle Detection using Vehicle Sound Classification: A Deep Learning Approach," 2022 6th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI), Colombo, Sri Lanka, 2022, pp. 1-6.
12. Arpitha, Y., Madhumathi, G.L. & Balaji, N. Spectrogram analysis of ECG signal and classification efficiency using MFCC feature extraction technique. J Ambient Intell Human Comput 2022 13, 757–767.
13. Lin, Y.J., Chen, X.J. BP Neural Network Learning Algorithm and its Software Implementation. AMM 2014 513–517, 738–741.
14. A. B. Gumelar, E. M. Yuniarno, D. P. Adi, A. G. Sooai, I. Sugiarto and M. H. Purnomo, "BiLSTM-CNN Hyperparameter Optimization for Speech Emotion and Stress Recognition," 2021 International Electronics Symposium (IES), Surabaya, Indonesia, 2021, pp. 156-161.