Research on Error Correction Coding Closed Set Recognition Technology Based on Deep Learning Method

Jianjun Chen
University of Electronic Science and Technology of China, Chengdu, China
429334539@qq.com

Abstract. It is very difficult to obtain information in non-cooperative areas. Error correction coding analysis is one of the key and difficult links. The general analysis is highly demanding and highly dependent on the level of personal analysis. For this reason, we mainly study various deep learning methods, which are applied to scrambling codes, short constraint length convolutional codes, and short constraint length block codes and we also I want to rely on the development of deep learning now, to study and realize intelligence and automated analysis.

1. Introduction
In the digital communication system, in order to improve the reliability of data transmission, error correction coding technology is usually adopted, and some redundancy is artificially added, so that the system has the capability of automatic error detection or error correction. In the field of non-cooperative communication, in order to obtain more effective information, it is necessary to identify the coding mode of intercepted data and estimate its coding parameters. Therefore, the blind recognition technology of error correction coding has important practical application value. The traditional recognition method then introduces the recognition method based on solving the equation pattern.

2. Based on solving equations and pattern recognition
2.1. Traditional identification method
So far, researchers have proposed a large number of automatic code recognition methods. Traditional recognition algorithms mainly include matrix analysis method, Euclidean analysis method, Walsh Hadamard transformation analysis method, finite field Fourier transform analysis method, dual spatial analysis method and so on. Matrix analysis is a more general method. It is applicable to BCH codes, RS codes, LDPC codes, convolutional codes, TURBO codes [1], etc. However, due to the large data demand and poor error-resistance performance, it is mainly used at present. Analysis of short block length linear block codes and convolutional codes; Euclidean analysis method is based on the special algebraic structure of codes, using algebraic methods to solve coding parameters, suitable for RS code, BCH code and other coding types with obvious algebraic structure However, the error-resistance performance of the method is weak; the Walsh Hadamard transform analysis method has the characteristics of simple principle and good error-resistance performance, but the solution process is an NP-hard problem, and the calculation amount varies with the code length index. Growth, generally only
used in the analysis of short code length block code and short constraint length convolutional code; finite field Fourier transform analysis method, which identifies the coding parameters according to the spectral features of the code, is the mainstream method for identifying RS code and BCH code. At present, the identification of RS code and BCH code mostly adopt this method; the dual space method is also a relatively common method, which is the most important method for analyzing LDPC and TURBO codes in the traditional method. It is also commonly used in the analysis of convolutional codes and general linear block codes.

2.2. Based on CNN to solve equation pattern recognition

The first question we have to solve first is what the data input looks like. For the scrambling code, it is essentially a pseudo-random number; at the same time, we can't determine what content is transmitted in the channel coded source. If it is encrypted data, it is a random number, so if we send these random data directly Machine learning, training, recognition, the effect will be bad, so the team based on its own communication background and computer background, studied various channel coding analysis methods combined with neural network, initially proposed a route based on solving equations recognition technology. In this way, in fact, in the field of channel coding analysis, combined with the field of image recognition in the neural network, the team finally used the data experiment and obtained good preliminary experimental results. The specific steps are as follows.

The following is based on solving the equation pattern identification technology route. Taking the RS code as an example, we know that the RS code is a cyclic code, and each codeword c in the same code is generated by the same generator polynomial g(x), that is, each The codeword polynomial corresponding to a codeword c is a multiple of the generator polynomial g(x). We can express it as c(x)=m(x)g(x). Suppose we get two correct RS code words c1 and c2, and they belong to the same RS code. Then we can get the following two equations:

\[ c_1(x)=m_1(x)g(x) \]  \[ c_2(x)=m_2(x)g(x) \]

We find that the right side of the above two equations has the same factor g(x), then we can use the Euclidean algorithm (transfer division method) to find the generator polynomial

\[ g(x)=\gcd(c_1(x),c_2(x)) \]

Through the above method, we can effectively obtain the generator polynomial of the RS code. Then, the problem of this method in this project is also very obvious. The first is that in the presence of error conditions, the polynomial cannot be effectively obtained, and the convolutional code and the scrambling code generally do not use this method, so the data I want Can not be obtained by this method. At the same time, the project team considers other statistics including 0, 1 run range statistics, code weight statistics, root distribution solution, etc., either for data quality requirements, or cannot be effective at the same time in the scrambling code, convolutional code, block code three The class played a role, so the project team finally determined the results of the three types of data to solve the equation, as a data processing method based on solving the technical idea of solving the equation, to realize the automatic identification and classification of channel coding[2].

We have already seen the block code. We will briefly introduce the pseudo-random sequence, linear feedback shift register and convolutional code to illustrate the theoretical basis of our method.
The number of stages of the linear feedback shift register is set to \( n \), and the feedback coefficient of the linear feedback shift register can be considered to be taken from the binary domain \( \mathbb{F}_2 \),

\[
an = -(c_1a_{n-1} + c_2a_{n-2} + \ldots + c_na_0), \quad \text{where } c_1, c_2, \ldots, c_n \in \mathbb{F}_2
\]  

(4)

It can be seen that it has a linear relationship, so for the scrambling code sequence, the linear relationship matrix can be solved by solving the equation, for example, using the Gaussian elimination method.

For convolutional codes, we consider a convolutional code with a rate of 1/2, and let the convolutional code be

\[
c = \{[u(D)g_1(D), u(D)g_2(D)] | u(D) \in \mathbb{F}[D]\}.
\]

Where, \( u(D) = u_0 + u_1D + u_2D^2 + \ldots \) is the input information sequence, \( g_1(D) \) and \( g_2(D) \) are generator polynomials. Let \( c_i(D) = (c_{i,1}(D), c_{i,2}(D)) \), where

\[
c_{i,1}(D) = c_{0,1} + c_{1,1}D + c_{2,1}D^2 + \ldots,
\]

\[
c_{i,2}(D) = c_{0,2} + c_{1,2}D + c_{2,2}D^2 + \ldots,
\]

Solving \( g_i(D) \) is to solve the following linear equations:

\[
\begin{bmatrix}
    e_{i,1} & e_{i,2} & \cdots & e_{i,3} & e_{i,4} & \cdots & e_{i,12} & \cdots & e_{i,32} & e_{i,33} \\
    e_{i,1} & e_{i,2} & \cdots & e_{i,3} & e_{i,4} & \cdots & e_{i,12} & \cdots & e_{i,32} & e_{i,33} \\
    \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\
    e_{i,1} & e_{i,2} & \cdots & e_{i,3} & e_{i,4} & \cdots & e_{i,12} & \cdots & e_{i,32} & e_{i,33} \\
\end{bmatrix}
\begin{bmatrix}
g_{i,0} \\
g_{i,1} \\
\vdots \\
g_{i,32} \\
\end{bmatrix} = 0
\]

So far, we already know that whether it is scrambling code or cyclic code, convolutional code, we can obtain the generator polynomial or linear law by solving the equation.

So far, we already know that whether it is scrambling code or cyclic code, convolutional code, we can obtain the generator polynomial or linear law by solving the equation.
Especially in the case of error, the general solution of the equation can not be directly applied or requires another expert to analyze, and we can see that several solutions to the equation result are indeed distinguishable if they are understood as a bitmap. The CNN neural network is used to learn the features in the form of picture recognition to achieve the purpose of distinguishing different codes.

3. Error correction coding based on RNN and CNN
With the above theoretical support, We have realized the error correction coding recognition based on the equations of RNN and CNN [3]. The algorithm block diagram and some performances are as follows. The recognition rate is 86.66% without error, and the recognition rate is 83.77% when there is error (bit error rate). From 0.001 to 0.01), the data set contains convolutional codes (CV (3,2,9), CV (3,1,5), CV (4,1,8)), BCH code (BCH (15,11) ), BCH (15, 7), BCH (15, 5)) and RS code (RS (7, 5), RS (7, 4), RS (7, 6)),

![Figure 3. RNN-based error correction coding recognition framework](image)

We can get the confusion matrix

![Figure 4. confusion matrix (RNN)](image)
Figure 5. Error Correction Coding Recognition Performance Based on RNN

Then we use a convolutional neural network consisting of the input layer, convolutional layer 1, convolutional layer 2, average pooling layer, fully connected layer, and output [4].

Figure 6. confusion matrix (CNN)

Figure 7. Error Correction Coding Recognition Performance Based on RNN

4. Summary
In this paper, the problem of error correction coding closed set recognition is mapped from the traditional communication field problem to the field of image recognition based on deep learning, and its theoretical basis is provided. It effectively solves the problem of low efficiency of analysis and identification, and relies heavily on labor. The experimental data has achieved good results, and the
recognition effect has been verified, which can be applied to the field of error correction coding closed set recognition.

References
[1] Wang Xinmei, Xiao Guozhen. Error Correction Code—Principles and Methods “M”. Xi'an: Xi'an University of Electronic Science and Technology Press, 2001
[2] Liu Yujun. Channel Coding [M]. Zhengzhou: Henan Science and Technology Press. 2001
[3] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv: 1409. 1556, 2014.
[4] Zhang C L, Luo J H, Wei X S, et al. In Defense of Fully Connected Layers in Visual Representation Transfer [C]. Advances in Multimedia Information Processing, Harbin, China, 2017: 807-817