Reverse recognition of LDPC codes based on log-likelihood ratio

LIU Ren-xin¹, ZHANG li-min¹, ZHONG Zhao-gen², SUN Xue-li²

¹Naval Aviation University Research Institute of Information Fusion, Yan’Tai Shandong 264001, China
²Naval Aviation University Department of Electronic Basis, Yan’Tai Shandong 264001, China

Author: LIU Ren-xin; email: iam1rxin@163.com; phone: 17854287773;

Abstract. With the development of Adaptive Modulation and Coding (AMC) technology, blind identification of channel coding parameters has recently attracted more attention. To reduce the computational complexity of the closed set recognition of problem low-density parity-check (LDPC) codes, this paper improves a closed set reverse recognition algorithm based on the maximum mean log-likelihood ratio. Before using soft decision information to make a decision, the check vector is first filtered to select some check vectors to reduce the computational complexity. The experimental results show that this method can complete the inverse identification of closed sets of LDPC codes in the case of low signal-to-noise ratio. Compared with the previous algorithm, the time complexity of the proposed algorithm is significantly reduced.

1. Introduction

In recent years, adaptive coding and modulation technology has developed rapidly¹, and the analysis and identification of related channel coding parameters has become a hot issue in research²-³. At present, there is relatively little research on low-density parity-check (LDPC) blind recognition. Imad et al.⁴-⁵ proposed to calculate the log likelihood ratio (LLR) cumulative and minimum values to estimate the starting point of the codeword of the LDPC code under the constant parameter channel. This method uses soft-decision starting point recognition with limited parameters. Moosavi R et al.⁶-⁷ proposed in the article that soft decision information can be used to quickly determine whether the received sequence is a certain encoded codeword. Compared with the Chabot algorithm, the performance of the algorithm has been significantly improved. At the same time, they applied the algorithm to In the closed set identification of channel coding, the threshold setting of the algorithm is uncertain, and a certain threshold cannot be given, because the exact distribution of the check relationship is unknown, and the relevant parameters can only be obtained through statistics and simulation, but these parameters are easy affected by the signal-to-noise ratio. Yu Ming et al.⁸ proposed a method for identifying the code length and code rate of LDPC codes. This method performs traversal search on the candidate code set to find the code length and code word combination with the highest degree of compliance. The algorithm uses hard decision information, which results in poor fault tolerance of the algorithm. Xia T et al.⁹-¹¹ proposed in the literature to use cumulative log-likelihood ratio for identification in binary coding and non-binary coding, but the recognition performance of the algorithm under low signal-to-noise ratio needs to be improved. The current closed-
set recognition algorithm for LDPC codes generally has the problem of high computational complexity.

This paper improves a closed set reverse recognition algorithm for LDPC codes. First, it uses the received information sequence to make soft decisions, and defines the log likelihood ratio of the check relationship, and analyzes the log likelihood of different sequences under the check relationship. According to the characteristics of the ratio value, part of the check vectors are selected in advance to reduce the computational complexity of the algorithm when verifying the check relationship. Finally, the maximum mean likelihood ratio is used to reversely identify the closed set.

2. Reverse recognition problem model

2.1. Log likelihood ratio

Compared with the hard decision output, the soft decision output can be used to measure the reliability information of the received data. Using soft decision can improve the recognition performance. In this paper, the LLR is used to measure the reliability. Suppose it is a random variable in a finite field, then its LLR is defined as

\[ L_x(x) = \ln \frac{P_x(x=0)}{P_x(x=1)} \]  

(1)

In equation (1), \( P_x \) is the probability that \( X \) takes the value \( x \). If \( X \) is constrained by another random variable \( Y \), then according to the Bayes formula, the conditional log-likelihood ratio \( L_{xy}(x \mid y) \) can be expressed as

\[ L_{xy}(x \mid y) = \ln \frac{P_{xy}(x=0 \mid y)}{P_{xy}(x=1 \mid y)} = \ln \frac{P(x=0)P_{xy}(y \mid x=0)}{P(x=1)P_{xy}(y \mid x=1)} + \log \frac{P_x(x=0)}{P_x(x=1)} \]

(2)

In particular, when the probability distribution of \( X \) and \( Y \) is equal, there are \( L_{xy}(x \mid y) = L_{yx}(y \mid x) \). On the premise of not causing confusion, the subscripts of probability and LLR value in the above formula can be ignored, then

\[ P(x = 0 \mid y) = \frac{e^{L_{xy}(y)}}{1 + e^{L_{xy}(y)}}, \quad P(x = 1 \mid y) = \frac{1}{1 + e^{L_{xy}(y)}} \]  

(3)

Let the symbol \( \oplus \) represent the addition of binary fields, then for any independent random variables \( X_1 \) and \( X_2 \), there is

\[ P(x_1 \oplus x_2 = 0 \mid y_1, y_2) = P(x_1 = 0 \mid y_1) \cdot P(x_2 = 0 \mid y_2) + P(x_1 = 1 \mid y_1) \cdot P(x_2 = 1 \mid y_2) \]

\[ = \frac{e^{L_{xy}(y)}}{1 + e^{L_{xy}(y)}} \cdot \frac{1}{1 + e^{L_{xy}(y)}} \cdot \frac{1}{1 + e^{L_{xy}(y)}} \cdot \frac{1}{2} \sum_{i=1}^{2} e^{L_{xy}(y)_i} \cdot \frac{1}{2} \sum_{i=1}^{2} e^{L_{xy}(y)_i} + \frac{1}{2} \sum_{i=1}^{2} \tanh \left( \frac{L_{xy}(y)_i}{2} \right) \]

(4)

\[ P(x_1 \oplus x_2 = 1 \mid y_1, y_2) = 1 - P(x_1 \oplus x_2 = 0 \mid y_1, y_2) \]

\[ = \frac{1}{2} - \frac{1}{2} \sum_{i=1}^{2} \tanh \left( \frac{L_{xy}(y)_i}{2} \right) \]

(5)

So,

\[ L(x_1 \oplus x_2 \mid y_1, y_2) = \ln \frac{P(x_1 \oplus x_2 = 0 \mid y_1, y_2)}{P(x_1 \oplus x_2 = 1 \mid y_1, y_2)} = \ln \frac{1 + \sum_{i=1}^{2} \tanh \left( \frac{L_{xy}(y)_i}{2} \right)}{1 - \sum_{i=1}^{2} \tanh \left( \frac{L_{xy}(y)_i}{2} \right)} \]

(6)

If the \( \boxplus \) operation is defined as

\[ L(x_1) \boxplus L(x_1) = L(x_1 \oplus x_1) \]

It can be proved according to mathematical induction.
\[
\sum_{j=1}^{L} \ln L(x_j | y_i) = L(x_1 \oplus x_2 \oplus \cdots \oplus x_L | y_1, y_2, \cdots, y_L) = \ln \frac{1 + \prod_{j=1}^{L} \tanh \left(e^{L(x_j | y_i) / 2} \right)}{1 - \prod_{j=1}^{L} \tanh \left(e^{L(x_j | y_i) / 2} \right)}
\]  

(8)

There is a hyperbolic tangent function \( \tanh(\bullet) \) in equation (8), the calculation process is more complicated, and it is not realized by engineering. In practical applications, equation (8) can be approximated as \[12\]

\[
\sum_{j=1}^{L} \ln L(x_j | y_i) \approx \left( \prod_{j=1}^{L} \text{sign}(L(x_j | y_i)) \right) \min_{i=1, \ldots, L} |L(x_i | y_i)|
\]

(9)

2.2. Recognition algorithm

Suppose the information sequence sent by the sending end is \( c^i = (c_{i,1}', \cdots, c_{i,k}', \cdots, c_{i,T}')^T \), and the information sequence received by the receiving end is \( r^i = (r_{a,1}', \cdots, r_{a,i}', \cdots, r_{a,T}')^T \). From equation (2), the log likelihood ratio of the received information sequence can be obtained as

\[
L(c_i | r_i) = \ln \frac{P(c_i = 0 | r_i)}{P(c_i = 1 | r_i)}
\]

(10)

The symbols in the LDPC codeword are 0 and 1 with equal probability, so the posterior probability likelihood ratio of the information sequence at the sending end can be obtained by (5)(6)

\[
\sum_{j=1}^{L} \ln L(x_j | r) = \ln \frac{1 + \prod_{j=1}^{L} \tanh \left(e^{L(x_j | r) / 2} \right)}{1 - \prod_{j=1}^{L} \tanh \left(e^{L(x_j | r) / 2} \right)}
\]

(11)

among the

\[
L(c_i | r_i) = \ln \frac{P(c_i = 0 | r_i)}{P(c_i = 1 | r_i)} = \ln \left( \frac{1}{\sqrt{2\pi \sigma_0}} \exp \left( \frac{r_i - a}{2\sigma_0^2} \right) \right) - \ln \left( \frac{1}{\sqrt{2\pi \sigma_0}} \exp \left( \frac{r_i + a}{2\sigma_0^2} \right) \right) = \frac{2ar_i}{\sigma_0^2}
\]

(12)

a represents the signal amplitude, further simplifying (11), so that can be obtained

\[
r_{i, \theta}^0 \approx \left( \prod_{j=1}^{L} \text{sign} \left( \frac{2ar_{i,j}}{\sigma_0^2} \right) \right) \min_{i=1, \ldots, L} \left| 2ar_{i,j} \right| \sigma_0^2
\]

(13)

Select \( \theta^0 \) from the set of candidate codewords \( \Theta \). From the log-likelihood ratio and the check relationship, if \( \theta = \theta^0 \), the log-likelihood ratio is positive, and if \( \theta \neq \theta^0 \), the check-digit log-likelihood ratio is positive or negative. Not sure, if all the check digits are averaged, the positive and negative will cancel each other out and get a number with a smaller average value.

Suppose the average likelihood ratio of the received i-th sequence is

\[
\bar{r}_{i, \theta} = \frac{1}{n-k} \sum_{i=1}^{n-k} r_{i, \theta}
\]

(14)

The selection of the coding matrix can be set as

\[
Y = \arg \max_{\theta} (\bar{r}_{i, \theta})^2
\]

(15)

The specific steps for identification are as follows

Step 1: Set the candidate codeword for closed set recognition to \( \Theta = \{ \theta_1, \theta_2, \theta_3, \cdots, \theta_N \} \), which contains the check vector \( H \) of the codeword.

Step 2: Group the received sequence according to the code length \( n \) of the alternative code word, and divide it into N code groups \( N = \lceil L / n \rceil \) (L indicates the length of the information sequence and \( \lceil \bullet \rceil \) indicates rounding down).
Step 3: Calculate the log-likelihood ratio of each group according to the above description, use equations (14) and (15) as the decision function, calculate the column Hamming weight of the check matrix, and select the column with the smallest Hamming weight. The check vector is calculated to reduce the complexity of the calculation, all the calculation results are stored in the constant set, the corresponding number that meets the judgment value is selected from the set, and the closed set is selected from the candidate code set according to the corresponding number. Identify the resulting codeword.

3. Simulation
In order to verify the effectiveness of the algorithm in this paper, this section uses MATLAB simulation software for experiments, uses the LDPC code standard in the IEEE802.16e protocol to encode the information sequence, and uses the correct recognition rate as a measure of the effectiveness of the algorithm.

3.1. Algorithm validity verification
According to the IEEE802.16e protocol, set the code length \( L = 576 \), code rate \( R = 1/2 \), code block length \( M = 100 \), SNR = -1dB, the alternative code set contains 30 different codeword types. The experimental results are shown in figure 1.

![Figure 1 Correctly identify the schematic](image)

It can be seen from figure 1 that the average likelihood ratio has a distinct peak. The peak is defined as the codeword corresponding to the identification sequence. It can also be seen that there are four distinct secondary peaks in figure 1. The secondary peaks have the same code length and different codes. The average likelihood ratio calculated by the rate check matrix and the soft decision information can be used to identify the code length and code rate corresponding to the code word by selecting the position of the peak, so the algorithm can complete the reverse identification of the closed set of the LDPC code.

3.2. Influence of code length on recognition algorithm
In order to test the effect of code length on the algorithm, this section conducts simulation experiments under different code lengths. Set code length \( L = 576, 672, 768 \) and 864, code rate \( R = 1/2 \), code block length \( M = 100 \), SNR = -3dB~5dB interval 0.25dB, Carry out 1000 times of Monte Carlo simulation experiments, the candidate code set contains 30 different codeword types. The experimental results are shown in figure 2.
Figure 2 Impact of code length on recognition algorithm

It can be seen from figure 2 that under the same signal-to-noise ratio, the recognition effect of the LDPC code with the code length of 576 is the worst. As the code length becomes longer, the recognition effect becomes better, and the recognition effect of the code length 768 and the code length 864 are very close, when the signal-to-noise ratio is greater than -2dB, the recognition rate of all code lengths is close to 1.

3.3. Influence of code rate on recognition algorithm

Code length L=576, code rate R=1/2, 2/3A,3/4A,5/6, code block length M=100, signal-to-noise ratio SNR=-3dB~5dB interval 0.25dB, Carry out 1000 times of Monte Carlo simulation experiments, the alternative code set contains 30 different codeword types. The experimental results are shown in figure 3.

Figure 3 Impact of code rate on recognition algorithm

It can be seen from figure 3 that under the condition of code length L=576, when the code rate R=1/2, the recognition effect is the best, when the signal-to-noise ratio is greater than -1.5dB, the recognition rate reaches 1, and the code rate R=2/3A, when the signal-to-noise ratio is greater than
1dB, the recognition rate reaches 1, when the code rate $R=3/4$, when the signal-to-noise ratio is greater than 3.25dB, the recognition rate reaches 1, when the code rate $R = 5/6$ When the signal-to-noise ratio is greater than 4.25dB, the recognition rate reaches 1. It can be concluded that the lower the code rate, the better the recognition effect, because in the case of a low code rate, there are more parity digits, and in the case of a higher code rate, there are fewer parity digits, it is difficult to detect the constraint relationship, and the recognition rate A sharp decline.

3.4. Comparison between this algorithm and other algorithms
In order to verify the differences of other algorithms in this paper, figure 5 shows the performance of this algorithm and the literature [11] under different code lengths.

![Figure 4 Comparison between the algorithm in this paper and the algorithm in [11]](image)

It can be seen from figure 4 that when the signal-to-noise ratio is -6 ~ 3.5dB, compared with the experimental algorithm in [11], the recognition probability of this algorithm is slightly lower than that in the experimental algorithm in [11]. Selecting part of the check vectors, without traversing the check matrix, the recognition rate drops somewhat when the signal-to-noise ratio is low. When the signal-to-noise ratio is greater than -3.5dB, it can be seen that the recognition effect of the algorithm in this paper is basically consistent with the experimental algorithm in [11].

4. Conclusion
In this paper, for the closed set identification of LDPC codes under low signal-to-noise ratio, the maximum likelihood difference estimation identification algorithm is improved. This algorithm uses the received sequence to make a soft decision to obtain the log-likelihood ratio and analyse its statistical characteristics. It is concluded that the average likelihood ratio of the correct codeword will have a peak, and the identification of the LDPC code can be completed according to the peak. In order to verify the effectiveness of the algorithm in this paper, simulation experiments were carried out on the LDPC code under the IEEE802.16e protocol. The experimental results show that the algorithm in this paper still has a high recognition rate under low signal-to-noise ratio.

References
[1] XIE Hui,HUANG Zhitao,WANG Fenghua .Research Progress of Blind Recognition of Channel Coding[J].Acta Electronica Sinica,2013,41(06) :1166-1176.
[2] Huang L, Chen W, Chen E, et al. Blind recognition of k/n rate convolutional encoders from noisy observation[J]. Journal of Systems Engineering and Electronics, 2017, 28(2): 235-243.
[3] ZHONG Zhaogen, WU Zhaojun, LIU Jie, et al Blind Identification of Frame Structure of Turbo Codes on Trellis Termination at High Bit Error[J]. Acta Electronica Sinica,2019,47(10):2089-2097.
[4] Imad R, Houcke S, Douillard C. Blind frame synchronization on gaussian channel[C]//2007 15th
European Signal Processing Conference. IEEE, 2007: 555-559.

[5] Imad R, Sicot G, Houcke S. Blind frame synchronization for error correcting codes having a sparse parity check matrix[J]. IEEE transactions on communications, 2009, 57(6): 1574-1577.

[6] Moosavi R, Larsson E G. A fast scheme for blind identification of channel codes[C]/2011 IEEE Global Telecommunications Conference-GLOBECOM 2011. IEEE, 2011: 1-5.

[7] Moosavi R, Larsson E G. Fast blind recognition of channel codes[J]. IEEE Transactions on Communications, 2014, 62(5): 1393-1405.

[8] YU Ming,GAO Xinhua,DU Yufeng. A error-toleration code length and rate identification algorithm for LDPC codes[P].Liaoning: CN102916774A,2013-02-06.

[9] Xia T, Wu H C. Novel blind identification of LDPC codes using average LLR of syndrome a posteriori probability[P]. ITS Telecommunications (ITST), 2012 12th International Conference on,2012.

[10] Xia T, Wu H C, Chang S Y, et al. Blind identification of binary LDPC codes for M-QAM signals[C]/2014 IEEE Global Communications Conference. IEEE, 2014: 3532-3536.

[11] Xia T, Wu H C. Joint blind frame synchronization and encoder identification for low-density parity-check codes[J]. IEEE Communications Letters, 2014, 18(2): 352-355

[12] Hagenauer J, Offer E, Papke L. Iterative decoding of binary block and convolutional codes[J]. IEEE Transactions on information theory, 1996, 42(2): 429-44