



基于卷积神经网络的极化码译码算法

郭锐, 冉凡春

(杭州电子科技大学, 浙江 杭州 310018)

摘要: 针对现有基于神经网络的极化码译码算法仅能对短码进行译码的问题(码长 $N \leq 64$), 提出针对极化码长码($N \geq 512$)的卷积神经网络译码算法, 利用神经网络 epoch 和 batch 参数来调控神经网络的输入, 而不是固定从数据样本中抽取一定比例用作训练集和测试集, 解决了由码字过长造成的数据爆炸问题。此外, 研究了 batch 和 epoch 参数对卷积神经网络译码性能的影响及神经网络使用不同激活函数的性能差异。仿真结果表明, 与传统 SCL (successive cancellation list, 串行抵消列表) 译码算法比较, 卷积神经网络在低信噪比下取得略优于 SCL ($L=2$) 的性能, 在高信噪比下取得与 SCL ($L=2$) 算法相近的性能, 并且训练数据集越大, 神经网络译码性能越好。

关键词: 深度学习; 神经网络; 信道译码; 极化码译码

中图分类号: TN911

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2020130

Polar codes decoding algorithm based on convolutional neural network

GUO Rui, RAN Fanchun

Hangzhou Dianzi University, Hangzhou 310018, China

Abstract: In order to solve the problem that the existing Polar code decoding algorithm based on neural network can only decode short codewords (codewords length $N \leq 64$), a new decoding algorithm using convolution neural network for long codewords ($N \geq 512$) was put forward. Instead of using a fixed drawn from the data sample proportion as training set and test set, the neural network parameters epoch and batch were used to control the neural network input, and the data explosion problem caused by the long codewords was solved. In addition, the influence of batch and epoch parameters on the decoding performance of convolution neural network was explored and the performance difference of neural network using different activation functions were investigated. Simulation results show that with the traditional SCL (successive cancellation list, $L=2$) decoding algorithm, convolution neural network in low signal-to-noise ratio on the performance is better than that of SCL ($L=2$), in high signal-to-noise ratio and SCL ($L=2$) algorithm of similar performance, and the larger the training data set, the better the decoding performance of neural network.

Key words: deep learning, neural network, channel decoding, Polar code decoding



1 引言

5G 已经在许多国家被广泛推广^[1], 极化码作为 5G 控制信道短码的编码技术, 广泛应用于无线通信领域^[2]。在极化码译码算法研究方面, 深度学习的发展为极化码译码提供新的途径, 研究人员试图利用神经网络来解决极化码的译码问题^[3]。

参考文献[4]提出了使用神经网络来代替传统 BP (belief propagation, 置信度传播) 译码算法的子译码器, 改进了极化码 BP 迭代译码算法, 在降低算法复杂度的同时, 译码性能也有所提升。参考文献[5]提出了使用循环神经网络进行线性码译码的译码算法, 在减少参数的前提下得出了与全连接神经网络相近的误码率性能。参考文献[6]研究了全连接神经网络在不同隐藏层数、每个层的节点数和激活函数条件下的译码性能。最近, 参考文献[7]发现极化码比随机码更容易被学习, 将极化码和随机码分别用全连接神经网络进行译码, 给出了极化码短码和随机码短码之间的译码差别。

尽管上述文献已经对基于神经网络的极化码译码进行了充分的研究, 但这些算法都是针对短码 ($N \leq 64$) 译码, 对于极化码长码数据集 (2^N , N 为码长) 来说, 当 N 增大时, 数据集呈爆炸式增长, 会导致神经网络数据输入阻塞。针对这一问题, 本文提出了基于卷积神经网络极化码长码译码, 解决了数据集过大的问题, 同时, 取得与 SCL ($L=2$) 译码算法相近的性能。

2 极化码译码及卷积神经网络简介

2.1 极化码译码简介

在信息传输过程中, 干扰和噪声会影响信息传输的质量, 可以使用信道编码技术提高信息传输可靠性^[8]。2009 年, Arikan E 提出了极化码的概念, 并证明极化码在二进制离散无记忆信道下,

当码长无限长时, 能够达到香农极限^[9-10]。极化码译码算法主要是串行抵消 (successive cancellation, SC) 译码算法: 当码长无限长时, 极化码的 SC 译码性能接近香农极限; 而对于中短码长极化码, 信道极化不完全是会导致 SC 译码性能不足^[11]。极化码译码器的作用是根据译码器接收序列, 通过一系列迭代运算得到信息序列的估计值。SC 译码器通过已知序列 (y_1^N, u_{A^c}) 产生对 u_1^N 的估计值 $\hat{u}_1^N$ 。SC 译码器包含 N 个译码单元, 每个译码单元对应一个信息源, SC 译码是按位译码。其中冻结位对于译码端是已知的, 对于信息位, 译码器需要进行似然比的迭代运算, 并且需要借助先前译码估计 $\hat{u}_1^{i-1}$, 对于第 i 位进行译码判决, 可以通过式 (1) 计算似然比:

$$L_N^{(i)}(y_1^N, \hat{u}_1^{i-1}) \cong \frac{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1} | 0)}{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1} | 1)} \quad (1)$$

译码判决式为:

$$\hat{u}_i = \begin{cases} 0 & L_N^{(i)}(y_1^N, \hat{u}_1^{i-1}) \geq 1 \\ 1 & \text{其他} \end{cases} \quad (2)$$

SC 译码算法在每次译码判决时会直接给出当前位的估计值, 为“0”或者“1”, 这种方式仅能保证局部最优化, 对于整个译码过程来说, 如果能保证全局最优化, 那么就可以改善 SC 译码的缺陷。SCL 译码算法可以保留 L 条译码路径, 与 SC 相比, 它的译码性能更好, 但需要更多译码时延和存储空间^[12]。

2.2 卷积神经网络简介

近年, 卷积神经网络在图像处理方面取得了优异的性能, 更在降噪上独树一帜^[13-15]。这主要表现在隐藏层中的卷积层和池化层极大地减少了网络所需的训练参数, 减少了算法复杂度, 进而加快了训练速度, 也有利于抗过拟合和提高算法的稳健性。与全连接神经网络不同, 卷积神经网络的特征是先从矩阵的某个局部区域进行特征提取, 再用相同的方法在另一个区域进行特征提取,

即用同一个卷积核对全部的矩阵区域进行卷积计算, 得到整张矩阵图对应于这个卷积核的特征图。在特征提取中, 需要使用激活函数来进行区别, 如图 1 所示, 使用卷积核对输入层矩阵提取特征, 通过激活函数递归, 最后通过输出层输出, 表 1 中介绍了几种不同类型的激活函数。

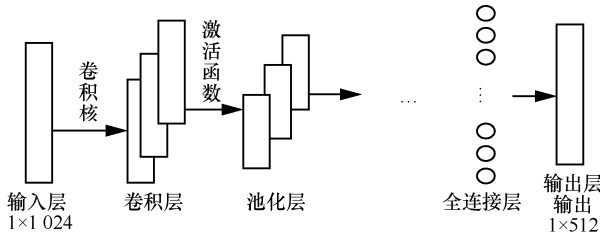


图 1 CNN 结构

表 1 激活函数

激活函数	表达式	激活函数	表达式
linear	x	Sigmoid	$\frac{1}{e^x + 1}$
ReLU	$\max(0, x)$	tanh	$\frac{e^{2x} - 1}{e^{2x} + 1}$

3 基于 CNN 的极化码长码译码

3.1 基于卷积神经网络的极化码译码

CNN (convolution neural network, 卷积神经网络) 译码算法架构如图 2 所示, 在信号发射端, 信息序列为 X , 长度为 K , 然后, 将 X 编码为码率为 1/2 的极化码 (N, K) , 其中 N 为码长, K 为信息位长度, 编码后的码字为 U , 通过 BPSK (binary phase shift keying, 二进制相移键控) 调制为适合在信道传输的符号 S , 本文假设信道为加性高斯白噪声信道。在信号接收端, 接收符号 Y 可以表示为: $Y=S+n$ (n 为噪声)。

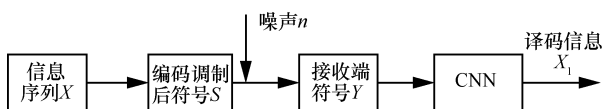


图 2 CNN 译码算法架构

噪声服从高斯分布, 是 $1 \times N$ 的向量, 传输符

号 Y 通过卷积神经网络得到译码符号 X_1 。损失函数是实际神经网络输出与期望输出之间的差值。在本文中, 使用 Sigmoid 作为激活函数, 使用均方误差作为损失函数, 其定义为:

$$L_{\text{MSE}} = \frac{1}{K} \sum_{i=0}^{K-1} (X_i - \hat{X}_{I_i})^2 \quad (3)$$

通过神经网络反向传播算法 (梯度下降) 优化神经网络参数, 达到最小化损失函数的目的, 得到最佳的译码性能^[16-17]。在实验过程中, 对于 CNN 参数定义见表 2。

表 2 CNN 参数定义

隐藏层类型	卷积核大小/步长	输入尺寸
卷积层一	$3 \times 1/1$	$N \times 1 \times 1$
池化层一	$2 \times 1/2$	$N \times 1 \times 8$
卷积层二	$3 \times 1/1$	$N/2 \times 1 \times 8$
池化层二	$2 \times 1/2$	$N/2 \times 1 \times 16$
卷积层三	$3 \times 1/1$	$N/4 \times 1 \times 16$
池化层三	$2 \times 1/2$	$N/4 \times 1 \times 32$
卷积层四	$N/8 \times 1/1$	$N/8 \times 1 \times 32$
squeeze	降低维度	$1 \times 1 \times K$
Sigmoid	分类输出	$1 \times K$

3.2 数据处理

基于 CNN 的极化码译码可以视为一种有监督的学习方法, 其构建分为两个阶段, 即训练阶段和测试阶段^[18-19]。在训练阶段, 利用生成的多组训练样本对神经网络中的权值和偏差进行校正, 使损失函数达到最小。测试阶段, 也就是实际的译码阶段, 只是对接收到的信息序列进行估计。在神经网络的训练阶段有两大难题: 首先是实验数据的生成, 其次, 由于在实际解码过程中的信噪比不确定, 因此需要找到一个最佳的训练信噪比, 使得在实际解码过程 (测试) 中达到最优译码性能。

数据生成可以这样描述: 从信息比特集合 $M(2^N)$ (码长为 N , 码率为 1/2) 中随机抽取 X



(128×10^6), 然后通过信道编码、BPSK 调制, 然后通过信道得到向量 Y (128×10^6), 测试集也是从 M 中随机选取的, 数据大小为 128×10^5 。在实验过程中, 1 024 码长所生成的数据量过于巨大, 导致数据无法在训练阶段完成输入, 为了解决这一问题, 本文去掉中间数据的存储, 直接将生成的数据作为输入, 通过 batch 和 epoch 这两大参数来限制输入数据的大小, 而在测试过程采用同样的数据处理方式。这样, 训练阶段及测试阶段的数据都是从 2^N 组数据中随机抽取的, 在训练阶段的数据几乎不可能出现在测试阶段, 因此神经网络的过程更像是学习, 而不是阅读和记录。同时也可以评估神经网络的泛化能力。关于训练过程中信噪比的设定, 由于实际解码 (测试阶段) 的信噪比未知且时变, 因此神经网络译码性能在很大程度上取决于训练样本的信噪比。本文的训练信噪比设置方法称为归一化验证误差^[7]。用 ρ_t 和 ρ_v 表示训练集和验证集的信噪比, 相应的用 $BER_{NND}(\rho_t, \rho_v)$ 表示在 ρ_t 训练后的 CNN 在 ρ_v 上的误比特率。用 $BER_{MAP}(\rho_v)$ 表示 MAP 译码在信噪比为 ρ_v 的误比特率, 对于 S 组不同的验证数据集, 其信噪比表示为 $\rho_{v,1}, \rho_{v,2}, \dots, \rho_{v,S}$ 。

$$NVE(\rho_t) = \frac{1}{S} \sum_{i=1}^S \frac{BER_{NND}(\rho_t, \rho_{v,i})}{BER_{MAP}(\rho_{v,i})} \quad (4)$$

其中, 当 $\rho_t \rightarrow +\infty$ 时, 训练阶段没有噪声, 神经网络不能译码; 当 $\rho_t \rightarrow -\infty$ 时, 训练阶段只有噪声, 神经网络不能学习码字之间的结构。这清楚地表明在这两种之间有一个最优值, 因此, 使用不同的信噪比 ρ_t 的数据集来训练神经网络, 本文选择最优的 ρ_t , 使得 NVE 最小。

4 仿真结果与分析

在实验中, 极化码 (N, K) 的码率为 1/2, 其中 N 表示极化码码字长度, K 为信息位长度, 极化码编码信息位采用高斯近似选取。训练阶段输入是从 $M(2^N, N \text{ 为码长})$ 中随机抽取 128×10^6

组 1024×1 个序列, 测试阶段输入从 M 中随机抽取 128×10^5 个序列, batch 设置为 128, epoch 设置为 10^6 , 在训练阶段的信噪比设置为 0~20 dB, 测试阶段的信噪比为 0~4 dB, 间隔为 0.5 dB。图 3、图 4 为 CNN、SC、SCL 译码算法在码长为 1 024 和 512 时的译码曲线。图 5、图 6 为在 batch 不变的情况下, epoch 分别为 10^4 、 10^5 、 10^6 时的卷积神经网络的误码曲线。图 7、图 8 为在总训练集为 128×10^6 不变的情况下, batch 为别为 64、128、256 下的卷积神经网络的误码率曲线。图 9 为不同激活函数下的卷积神经网络译码曲线。

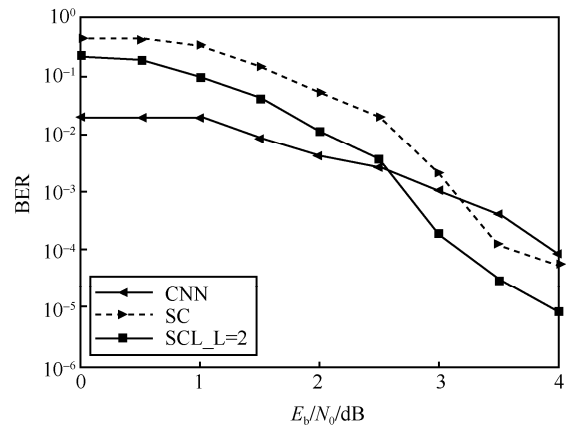


图 3 不同译码算法误码比特性能 ($N=1024$)

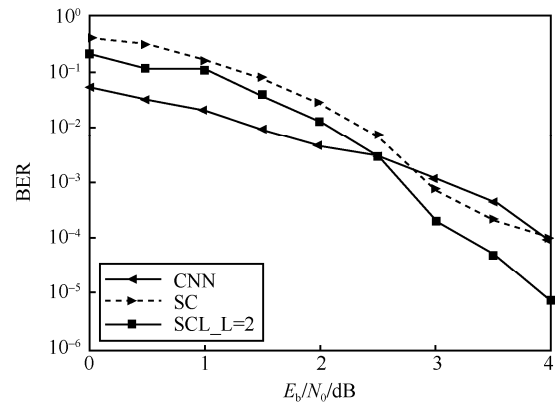
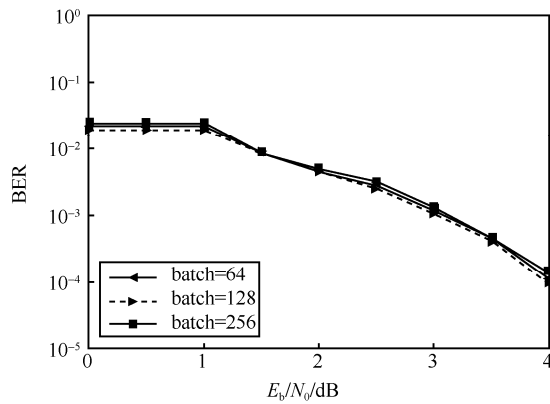
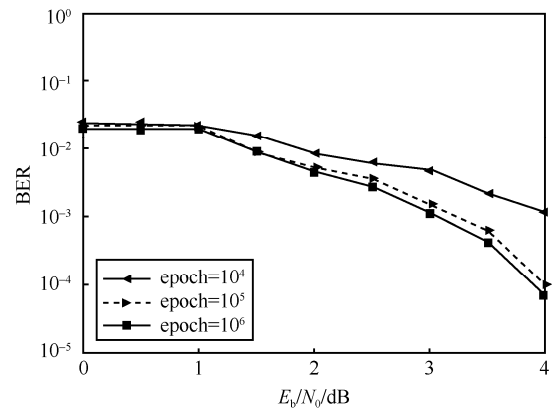
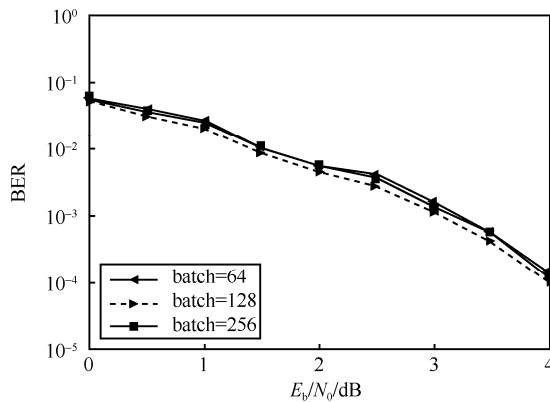
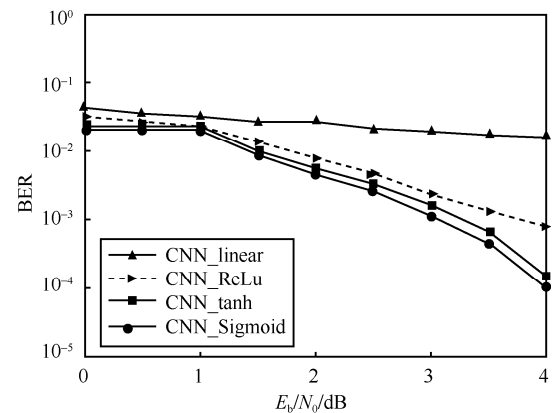
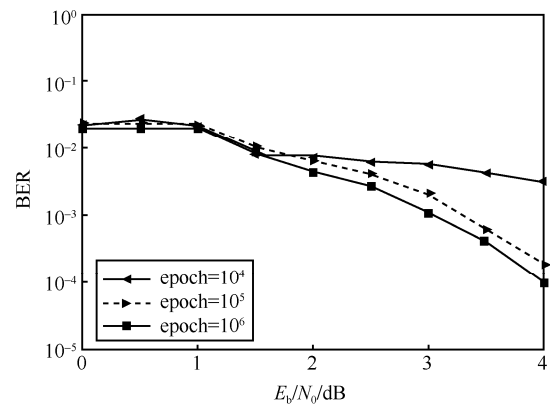


图 4 不同译码算法误码比特性能 ($N=512$)

从图 3 和图 4 看出, 对 SC、SCL 译码算法来说, 码长越长, 译码性能越好, 卷积神经网络在信噪比较低的情况下性能优于传统译码算法, 但是, 随着信噪比的增加, 曲线下滑幅度明显减小,

图5 不同 batch 下的神经网络译码性能 ($N=1024$)图8 不同 epoch 下的神经网络译码性能 (训练集不变, $N=512$)图6 不同 batch 下的神经网络译码性能 ($N=512$)图9 不同激活函数下的神经网络译码性能
(batch=128, $N=1024$, epoch = 10^6)图7 不同 epoch 下的神经网络译码性能 (训练集不变, $N=1024$)

卷积神经网络在高信噪比下取得与 SCL ($L=2$) 算法相近的性能。从图 5~图 8 中看出, epoch 越大, 卷积神经网络的译码性能越突出, 在 batch 不变的情况下, epoch 指数级的改变影响了训练集的大小, 而在训练集大小不变的情况下, 改变 batch 大小对译码性能的影响几乎可以忽略, 这说明在

一定范围内改变 batch 的大小几乎不会对神经网络的译码性能产生影响。从图 9 中可以看出, 基于卷积神经网络的极化码译码算法中 linear 激活函数的性能较差, Sigmoid 最好。

5 结束语

本文提出了基于卷积神经网络的极化码长码译码算法, 基于卷积神经网络对极化码码字结构和噪声特征进行提取, 通过误差函数达到对神经网络参数的调控, 以此得到最佳译码性能, 通过对 epoch 和 batch 参数的调控来达到对神经网络的输入进行限制, 解决了数据量过大产生的输入阻塞。虽然本文研究了基于卷积神经网络的极化码长码的译码性能, 但由此可以以点带面, 基于不同神经网络的极化码译码算法、神经网络的其他



因素等都有较大的研究价值和实际作用,而极化码译码算法的研究会在未来相当长的一段时间成为大众的焦点。

参考文献:

- [1] BOCCARDI F, HEATH R W, LOZANO A, et al. Five disruptive technology directions for 5G[J]. IEEE Communications Magazine, 2014, 52(2): 74-80.
- [2] ARIKAN E. Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels[J]. IEEE Transactions on Information Theory, 2009, 55(7): 3051-3073.
- [3] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536.
- [4] CAMMERER S, GRUBER T, HOYDIS J, et al. Scaling deep learning-based decoding of Polar codes via partitioning[C]//Proceedings of GLOBECOM 2017-2017 IEEE Global Communications Conference. Piscataway: IEEE Press, 2017: 1-6.
- [5] NACHMANI E, MARCIANO E, BURSHTEN D. RNN decoding of linear block codes[Z]. 2017.
- [6] SEO J, LEE J, KIM K. Decoding of Polar code by using deep feed-forward neural networks[C]//Proceedings of 2018 International Conference on Computing, Networking and Communications (ICNC). Piscataway: IEEE Press, 2018.
- [7] GRUBER T, CAMMERER S, HOYDIS J, et al. On deep learning-based channel decoding[C]//Proceedings of 2017 51st Annual Conference on Information Sciences and Systems (CISS). [S.l.:s.n.], 2017: 1-6.
- [8] WANG X, TAN Z H. Development of channel coding technology under wireless channel[J]. Journal of Beijing Jiaotong University, 2004, 28(3): 82-87.
- [9] SHANNON C E. A mathematical theory of communication[J]. Bell Labs Technical Journal, 1948, 27(4): 379-423.
- [10] MORI R, TANAKA T. Performance and construction of Polar codes on symmetric binary-input memory less channels[C]//Proceedings of IEEE International Symposium on Information Theory. Piscataway: IEEE Press, 2009.
- [11] NIU K, LIN J R, CHEN K. List successive cancellation decoding of Polar codes[J]. Electronics Letters, 2012, 48(9): 500-501.
- [12] BALATSOUKAS-STIMMING A, PARIZI M B, BURG A. On metric sorting for successive cancellation list decoding of Polar codes[C]//Proceedings of IEEE International Symposium on Circuits & Systems. Piscataway: IEEE Press, 2015.
- [13] LI Y D, HAO Z B, LEI H. Research review of convolutional neural networks [J]. Computer Applications, 2016, 36(9): 2508-2515.
- [14] 贾川民, 赵政辉, 王苦社, 等. 基于神经网络的图像视频编码[J]. 电信科学, 2019, 35(5): 38-48.
- JIA C M, ZHAO Z H, WANG J S, et al. Image and video coding based on neural network [J]. Telecommunications Science, 2019, 35(5): 38-48.
- [15] 苏赋, 吕沁, 罗仁泽. 基于深度学习的图像分类研究综述[J]. 电信科学, 2019, 35(11): 58-74.
- SU F, LV Q, LUO R Z. Review of image classification based on deep learning[J]. Telecommunications Science, 2019, 35(11): 58-74.
- [16] SHELKE S, APTE S. A novel multi-feature multi-classifier scheme for unconstrained handwritten Devanagari character recognition[C]//Proceedings of International Conference on Frontiers in Handwriting Recognitions. Piscataway: IEEE Press, 2010.
- [17] KOLEN J F, KREMER S C. Gradient flow in recurrent nets: the difficulty of learning longterm dependencies[M]//A field guide to dynamical recurrent networks. Piscataway: IEEE Press, 2001: 237-243.
- [18] FANG X N. Principles of four deep learning methods from supervised learning to intensive learning[J]. Robotics Industry, 2017(4): 50-55.
- [19] LYU W, ZHANG Z, JIAO C, et al. Performance evaluation of channel decoding with deep neural networks[C]//Proceedings of 2018 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2018: 1-6.

[作者简介]



郭锐(1980—),男,博士,杭州电子科技大学副教授、硕士生导师,主要从事无线通信、信道编码等方面的研究工作。



冉凡春(1995—),男,杭州电子科技大学硕士生,主要研究方向为信道编码及深度学习。