



北京大学

本科生毕业论文

题目: 基于神经网络的极化码译码探
究

A Study on Polar Decoding Based on Neural
Networks

姓 名: 张峻伟
学 号: 1400012820
院 系: 信息科学技术学院
本科专业: 电子信息科学与技术
指导教师: 赵玉萍

二〇一八年五月

北京大学本科毕业论文导师评阅表

学生姓名	张峻伟	学生学号	1400012820	论文成绩	
学院 (系)	信息科学技术学院			学生所在专业	电子系
导师姓名	赵玉萍	导师单位/ 所在研究所	信息科学技术学院 现代通信研究所	导师职称	教授
论文题目 (中、英文)	基于神经网络的极化码译码探究 A Study on Polar Decoding Based on Neural Networks				
<div>导师评语</div> <div>(包含对论文的性质、难度、分量、综合训练等是否符合培养目标的目的等评价)</div> <p>张峻伟同学撰写的毕业论文选题较新，将神经网络这一当前的研究热点与信道译码相结合。论文选题符合电子学系专业培养目标要求，研究结果体现出作者较强的独立思考和实践探究能力。</p> <p>全文结构合理，思路清晰，语言通顺、简洁，层次分明。绪论部分介绍了研究背景和研究意义，系统地介绍了神经网络和极化码的基本概念和工作原理，为论文主体工作的展开做了铺垫。在此基础上给出了基于神经网络的极化码译码模型，以及模型训练和测试的详细过程。仿真和结论部分系统地呈现了仿真结果、结论以及自己的思考，整体较为全面，完成度较高。</p> <p>本论文工作难度较高，仿真数据详实，结论分析到位，具有较强的创新性和启发性。论文结构合理，逻辑清晰，参考的文献资料与论题和论文内容结合紧密，符合培养目标的要求。</p> <div>导师签名： 年 月 日</div>					

版权声明

任何收存和保管本论文各种版本的单位和个人，未经本论文作者同意，不得将本论文转借他人，亦不得随意复制、抄录、拍照或以其他方式传播。否则，引起有碍作者著作权之问题，将可能承担法律责任。

摘要

人工神经网络作为模拟人脑神经元连接的一种计算模型，在机器学习、信号处理等领域拥有广泛的应用前景。目前，将神经网络模型用于无线通信领域信道译码环节是一个较新的课题。

本文以极化码为例，探究基于神经网络的极化码译码的普遍特征和规律，比较了神经网络与传统译码算法在性能上的差异。本文通过大量的仿真数据，验证了训练信噪比、神经网络规模和极化码码率三个因素对神经网络译码性能的影响，并总结出了有关神经网络译码的一系列具有实际指导意义的经验性结论。

主要结论概括如下：

- (1) 在神经网络的训练过程中，训练信噪比越低，所需训练周期数越多，平均训练周期数随着训练信噪比的降低呈指数上升趋势；
- (2) 在极化码的解码过程中，基于低信噪比训练的神经网络的误码率特性优于基于高信噪比训练的神经网络。
- (3) 神经网络规模越大，所需训练周期数越少，训练结果的误码率越低。

关键词：机器学习 神经网络 信道译码 极化码

Abstract

As a computing model imitating the actual neuron connections inside human brains, artificial neural networks possess broad applications in domains such as machine learning and signal processing. It is a relatively new topic recently to apply neural network models to channel decoding in the field of wireless communications.

Taking the polar code as an example, this essay investigates common features and patterns underlying neural network-based polar decoding, and evaluates its performance compared to traditional decoding algorithms. The essay explores, through mass simulations, the influence of training signal-to-noise ratio (SNR), neural network scales, and code rates on the decoding performance of neural networks, and summarizes a series of empirical conclusions regarding neural network-based channel decoding.

Main conclusions are listed below:

1. In the training process of neural networks, the number of required training epochs increases exponentially as the training SNR decreases;
2. In the polar decoding process, the bit error rate (BER) of neural networks trained with lower training SNRs outperforms those trained with higher training SNRs;
3. As the neural network becomes larger in scale, the number of required training epochs decreases, and the BER of the neural networks decreases, too.

Key Words: machine learning, neural networks, channel decoding, polar code

全文目录

摘要.....	I
Abstract.....	II
第一章 绪论.....	1
1. 研究背景.....	1
2. 基本概念.....	2
2.1 神经网络与万能逼近特性.....	2
2.2 极化码.....	2
2.3 基于神经网络的信道译码.....	3
3. 研究意义.....	3
第二章 概念介绍.....	5
1. 神经网络.....	5
A. 前向传播.....	6
B. 反向传播.....	7
2. 极化码.....	8
A. 极化码的基本概念.....	8
B. 极化码的编码与译码.....	9
第三章 网络模型.....	11
1. 网络模型.....	11
2. 模型训练与测试.....	12
A. 训练过程.....	12
B. 测试过程.....	13
3. 研究方向.....	14
第四章 仿真结果与讨论.....	15
1. 训练 SNR 对训练结果的影响.....	15
A. 训练 SNR 影响的定性分析.....	15
B. 结论验证.....	18
C. 结论分析.....	19
D. 训练 SNR 与学习速度的定量分析.....	20
2. 神经网络规模对训练结果的影响.....	22
3. 极化码码率对训练结果的影响.....	25
第五章 总结与展望.....	28
1. 结论总结.....	28
2. 研究展望.....	29
参考文献.....	30
本科期间的主要工作和成果.....	31
致谢.....	32

第一章 绪论

1. 研究背景

人工神经网络的计算模型最早由 W. S. McCulloch 和 W. Pitts 在 1943 年提出，用于解决一系列实际的计算问题^[1]。直到上世纪 80 年代反向传播算法的提出解决了非线性分类和学习的问题^[2]，神经网络模型才开始真正有了实际的应用，比如手写体的识别^[3]。将神经网络用于信道译码也开始有了最初的探索。1989 年，J. Bruck 与 M. Blaum 指出 Hopfield 神经网络可用于模拟线性分组码的最大似然译码过程^[4]。1990 年，人们发现神经网络模型可以直接用于学习带噪声的输入码字与输出信息位的映射关系^[5]。这一想法的提出为之后该领域的研究奠定了基本思路。90 年代间，人们开始将神经网络用于不同种类信道编码方案的译码环节中。比如，1995 年 L. G. Tallini 等提出用神经网络寻找接收码字的校正子与错误比特位置的映射关系，从而实现了汉明码的译码^[6]。1996 年 X. A. Wang 和 S. B. Wicker 将神经网络用于卷积码的译码，并发现其性能与维特比译码性能相当^[7]。后来，A. Hamalainen 等人采用了循环神经网络，提升了基于神经网络的卷积码译码性能^[8]。

由于训练方法的落后和硬件计算能力的限制，神经网络的发展经历了一个漫长的停滞期。直到 2006 年 G. Hinton 等提出无监督预训练、梯度下降微调的训练方法^[9]，才使较大规模神经网络的训练成为可能。随着计算机硬件能力的提升，尤其是 GPU 的快速普及，神经网络的训练过程大大加快，成本显著降低。自此之后，神经网络被广泛应用于机器学习的各个领域，比如计算机视觉、自然语言处理等。但是，自从神经网络复苏以来，基于神经网络的信道译码还未得到进一步的研究。2017 年，T. Gruber 等重新审视了基于神经网络的信道译码问题，并通过仿真总结出了一些经验性结论^[10]。本文是在 T. Gruber 等工作的基础上，通过仿真数据进一步揭示神经网络在极化码译码时表现出来的内在规律与共同特征，并试图建立对神经网络译码的一个直观、感性的认识，为今后将神经网络应用在信道译码领域提供经验性指导。

2. 基本概念

2.1 神经网络与万能逼近特性

神经网络(Neural Networks, NN)是受生物神经网络启发而形成的计算模型,用于估计或近似函数。神经网络是由多层相互连接的神经元构成的。工作时,信息从输入层输入,经过逐层传播,最后从输出层输出,这就实现了一个输入到输出的函数映射。

神经网络的万能逼近特性指明,含有两个及以上隐层的神经网络能够以任何精度逼近任何函数^[11]。这表示只要神经网络规模足够大,理论上可以无限逼近一切可能的函数映射关系。增加隐层的数量可以大大减少每层所需要的神经元个数。神经网络的万能逼近特性是其能够广泛应用于机器学习各领域的重要原因。

决定神经网络映射关系的主要因素是连接每两个神经元的边的权重。在使用前寻找并优化每条边上权重的过程就称为训练。具体来说,神经网络的训练是通过已知输入和输出的训练集合来优化网络权重,调整网络输入输出的映射关系,使之满足实际需要的过程。神经网络的训练是网络使用前不可缺少的步骤。

2.2 极化码

在无线通信领域,信道编码是一种通过增加信息位冗余提高传输可靠性的方法。极化码是 2009 年 Erdal Arıkan 提出的新型信道编码方法之一,理论上证明能够达到二进制输入离散无记忆信道(B-DMC)的香农极限^[12]。

极化码的基本原理是将多个完全相同的实际信道进行组合、分离操作,生成同等数量的、信噪比各不相同的子信道(bit-channel),这个过程称为极化。而后,从这些子信道中选择信噪比最高、噪声最小的部分子信道传信息(码率为 1),其余的信噪比低、噪声高的子信道不传信息(码率为 0)。不同的信道条件下,信道极化结果也不同,子信道的选取方法也有所不同。因此极化码的最佳构造方式随信道参数变化而变化,这是极化码区别于其它信道编码方案的所在。

Arıkan 提出的 Successive Cancellation (SC)译码是极化码的经典译码算法,复杂度是 $O(N\log N)$ 。本论文将以 SC 译码的结果作为神经网络译码结果的参考。

2.3 基于神经网络的信道译码

基于神经网络的信道译码模型如下图 1 所示：

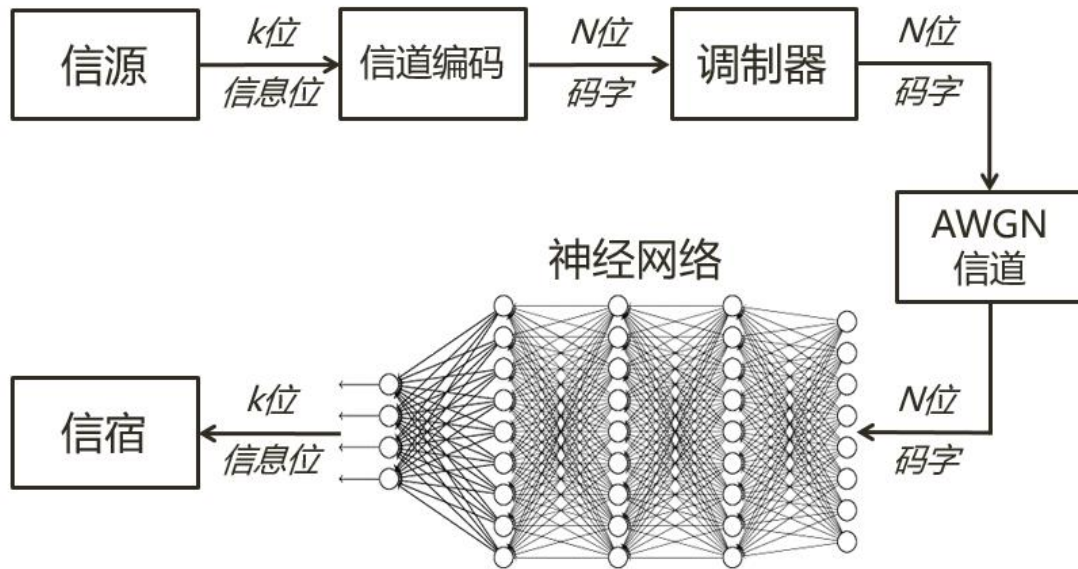


图 1：基于神经网络的信道译码模型

可见，在图 1 所示信道译码模型中，神经网络取代了传统通信系统中信道译码器的位置。信道输出的带噪声的 N 位码字直接作为神经网络的输入，神经网络的输出就是信宿接收到的 k 位译码结果。

3. 研究意义

将神经网络用于信道译码既是对神经网络应用领域的拓展，也是对信道译码方案的新尝试。

用神经网络进行信道译码主要有以下三方面的意义：

(1) 复杂度低，延迟小。传统信道译码算法，比如卷积码的维特比译码、极化码的 SC 译码算法等，大都基于迭代的思路，往往需要不断尝试、寻找幸存路径，因此算法复杂度高，相对耗时。尤其当码长较长时，译码过程计算量很大。用神经网络进行信道译码时，一旦训练完成，网络参数固定，神经网络译码器就变成了一个近似的线性系统，译码复杂度降为 $O(N)$ 。信号从网络输入后只需要进行简单的线性运算即可输出，速度快、延迟低。

(2) 可塑性高, 可用于不同的信道编码方案。根据万能逼近特性, 神经网络可以无限逼近任何函数, 也就可以适用于任何信道编码方案。同一个神经网络, 用极化码训练后可以解极化码, 用卷积码训练后可以解卷积码, 适用范围广。

(3) 从根本上规避了机器学习过程中的过拟合问题。大多数有监督学习模型的训练集是有限的, 容易产生过拟合现象, 并且需要人为对训练集进行逐个标注, 十分耗时。相比之下, 用于信道译码的神经网络的训练集是无限的, 且无需人为标注(具体原因见第三章模型介绍部分), 这从根本上避免了过拟合问题的出现。这一特征是神经网络应用于信道译码的优势所在。

本论文探究基于神经网络的极化码译码的意义主要有以下两点:

(1) 寻找合适的训练信噪比。神经网络的训练过程中, 训练 SNR 的选取不仅影响神经网络的训练时间长短, 而且影响最终神经网络的译码效果。本论文给出的训练信噪比的最佳选取方法, 对神经网络译码器的实际应用具有指导意义。

(2) 寻找合适的神经网络规模。神经网络规模的选取也对训练和译码过程产生直接影响。本论文给出的神经网络规模与训练时间长短、最终译码性能之间的定性关系能够指导实际应用中网络规模的恰当选取。

第二章 概念介绍

本章主要介绍神经网络与极化码的基本概念和工作原理。极化码部分还将介绍其编码方法和 SC 译码算法。

1. 神经网络

神经网络（又称人工神经网络）是受生物神经网络启发而形成的计算模型，用于估计或近似函数。构成神经网络的基本单元叫做感知器。感知器包含 n 个输入，1 个输出，其结构如下图 2 所示：

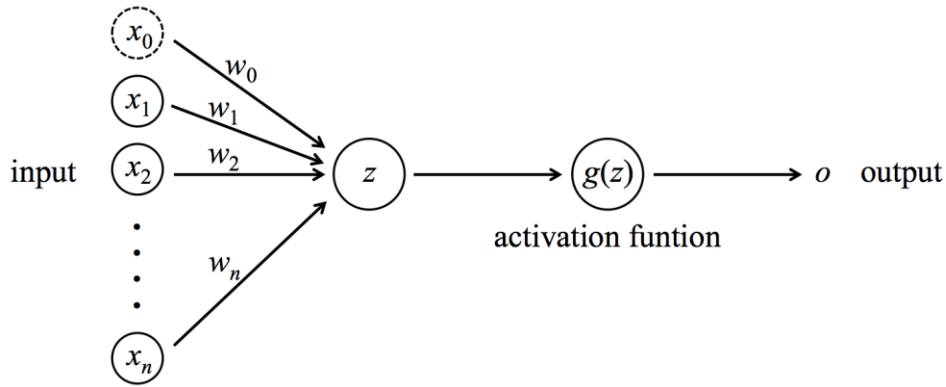


图 2：神经网络感知器模型

图 2 中， x_1 到 x_n 代表 n 个输入。其中 $x_0 = 1$ ，称为偏置项。感知器首先对输入求线性加权和

$$z = \sum_{i=0}^n w_i x_i$$

其中， w_i 称为权重。 z 经过激活函数 $g(z)$ 变换得到输出 o 。常用的非线性激活函数有很多，本论文采用的激活函数有 ReLU 函数和 Sigmoid 函数，它们分别定义为

$$g_{\text{sigmoid}}(z) = \frac{1}{1 + e^{-z}}, \quad \mathbb{R} \rightarrow (0, 1)$$

$$g_{\text{relu}}(z) = \max\{0, z\}, \quad \mathbb{R} \rightarrow [0, +\infty)$$

因此感知器的输入输出关系可以表示为

$$o = g\left(\sum_{i=0}^n w_i x_i\right)$$

通过把多个感知器整合在一起，就形成了基本的神经网络构成形式。具体来讲，把 m 个感知器并列排在一起，共享输入节点 x_i ，就得到了一层 n 输入、 m 输出的神经网络。将多层神经网络按照先后次序排列在一起，第 i 层网络的输出 m_i 作为第 $(i+1)$ 层网络的输入 n_{i+1} ，这就得到了多层神经网络，也叫多层感知器，其结构如下图 3 所示：

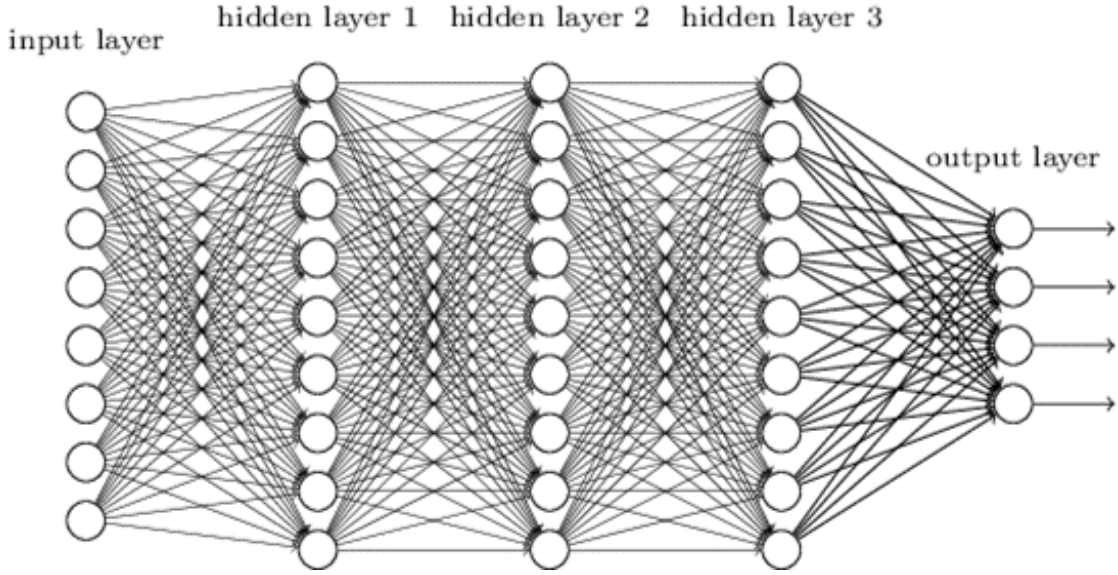


图 3：神经网络模型

由图可见，神经网络是有多层相互连接的神经元构成，最左边是输入层，中间包含不定数量的隐层，最右边是输出层。

A. 前向传播

前向传播是神经网络工作时的过程。此时，信息从输入层输入，经过多层隐层的识别和分析，逐层传递，最终从输出层输出。设第 i 层包含 n_i 个输入和 m_i 个输出，则第 i 层实现了一个非线性映射

$$f^{(i)}: \mathbb{R}^{n_i} \rightarrow \mathbb{R}^{m_i}$$

其中，映射过程包含线性加权和激活函数两个步骤，映射包含的参数是第 i 层每条边上的权重。记神经网络的输入为 \mathbf{x} ，输出为 \mathbf{o} ，则神经网络工作时的输入输出关系可以表示为

$$\mathbf{o} = f(\mathbf{x}; \boldsymbol{\theta}) = f^{(L-1)} \left(f^{(L-2)} \left(\dots \left(f^{(0)}(\mathbf{x}) \right) \right) \right)$$

其中, L 是神经网络隐层 (包含输出层) 的个数, 称为神经网络深度。 θ 表示神经网络中所有权重参数构成的向量。

B. 反向传播

显然, 神经网络输入、输出的映射关系由网络中的权重参数向量 θ 决定。为了寻找特定应用场景下最佳的权重参数, 需要通过已知输入、目标输出的训练集对神经网络进行训练。反向传播是神经网络训练时的一种计算网络权重参数梯度的技术。

设训练样本的输入为 \mathbf{x} , 输出为 \mathbf{t} 。神经网络对输入 \mathbf{x} 产生一个输出 \mathbf{o} 。我们用损失函数 $L(\theta)$ 来度量神经网络输出 \mathbf{o} 与目标输出 \mathbf{t} 之间的误差。损失函数反映了当前网络对训练样本 (\mathbf{x}, \mathbf{t}) 的拟合程度。损失函数越小, 拟合程度越高。损失函数通常定义为 \mathbf{o} 与 \mathbf{t} 之间的均方差(MSE)

$$L_{MSE} = \frac{1}{k} \sum_i (t_i - o_i)^2$$

或者二进制交叉熵(BCE):

$$L_{BCE} = -\frac{1}{k} \sum_i [t_i \ln(o_i) + (1 - t_i) \ln(1 - o_i)]$$

其中, k 代表神经网络输出层的神经元个数, 即 \mathbf{o} 与 \mathbf{t} 的维度。对于二进制交叉熵损失函数而言, $t_i \in \{0, 1\}$, 即训练集目标输出是二进制序列。训练过程就是通过寻找并优化网络参数 θ , 使得网络在训练集上的损失函数 $L(\theta)$ 最小化, 即尽可能让神经网络拟合所有的训练样本。神经网络的目标就是在测试集上也能正确地预测输出, 这是神经网络泛化能力的体现。

反向传播过程把误差项从输出层逐层反向传导到参数需要修改的层, 然后用标准的梯度下降方法去修改各层参数。参数优化过程可以表示为

$$\theta \leftarrow \theta - \alpha \nabla L(\theta)$$

其中,

$$\nabla L(\theta) \equiv \left[\frac{\partial L}{\partial \theta_0}, \frac{\partial L}{\partial \theta_1}, \dots, \frac{\partial L}{\partial \theta_n} \right]$$

表示权重参数的梯度, 代表了权重空间中损失函数 L 的最快增长方向。 α 是学习率, 表示每一次参数更新的步长, 是神经网络需要预先指定的超参数之一。

除了标准的梯度下降算法以外，神经网络的优化方法还包括 AdaGrad 算法、RMSProp 算法、Adam 算法等。它们都基于梯度下降的基本原理，但引入了自适应学习率的概念，能够根据优化过程中梯度的变化等因素自动调整学习率。经验表明，如果需要更快的收敛，或者训练更深更复杂的神经网络，需要用一种自适应的算法。其中 Adam 算法是目前较为主流的优化算法。

2. 极化码

A. 极化码的基本概念

极化码是 Erdal Arıkan 于 2009 年提出的一种基于信道极化效应的新型信道编码方案。考虑一个二进制输入离散无记忆信道 B-DMC $W: \mathcal{X} \rightarrow \mathcal{Y}$ ，其中 $\mathcal{X} = \{0, 1\}$ 代表输入字母表， \mathcal{Y} 代表输出字母表，记 $W(y|x)$ ， $x \in \mathcal{X}$ ， $y \in \mathcal{Y}$ 为信道转移概率。定义信道 W 的对称容量

$$I(W) \triangleq \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \frac{1}{2} W(y|x) \log \frac{W(y|x)}{\frac{1}{2} W(y|0) + \frac{1}{2} W(y|1)}$$

和巴氏参数

$$Z(W) \triangleq \sum_{y \in \mathcal{Y}} \sqrt{W(y|0) W(y|1)}$$

巴氏参数 $Z(W) \in [0, 1]$ ，描述了信道的可靠性。 $Z(W)$ 越低，信道越可靠（对称容量越大）， $Z(W)$ 越高，信道越不可靠（对称容量越小）。

N 个完全相同且相互独立的 B-DMC W ，经过组合、分离操作后，可以得到 N 个容量互不相同的子信道 $\{W_N^{(i)}: 1 \leq i \leq N\}$ 。Arıkan 发现，随着 N 不断变大，子信道的对称信道容量 $I(W_N^{(i)})$ 和巴氏参数 $Z(W_N^{(i)})$ 均趋向 0 或 1。这个现象称为信道极化。对于极化码而言，只选择 $I(W_N^{(i)}) \rightarrow 1$ ，或等价于 $Z(W_N^{(i)}) \rightarrow 0$ 的可靠子信道传有效信息，其余子信道传无效信息。

具体来讲，给定一个 B-DMC W ，极化码可以表示为 $(N, k, \mathcal{A}, u_{\mathcal{A}^c})$ 。其中， N 是极化码码长，且总是 2 的指数次幂，即 $N = 2^n, n \geq 0$ 。我们考虑极化码中的

第 i 个比特位是通过子信道 $W_N^{(i)}$ 传输的。 k 是信息位个数，表示 N 个子信道中用于实际传输有效信息的子信道个数。注意到 $k \leq N$ ，且码率 $R = \frac{k}{N}$ 。 $\mathcal{A} \subseteq \{1, 2, \dots, N\}$ 是一个 k 元子集，代表上述 k 个子信道的集合。比如， $\mathcal{A} = \{1, 3, 5, 7\}$ 表示第 1、3、5、7 个子信道传有效信息，其余信道传无效信息。记 \mathcal{A}^c 为 \mathcal{A} 在全集 $\{1, 2, \dots, N\}$ 上的补集。为了达到对称信道容量 $I(W)$ ， \mathcal{A} 的选取要使得

$$Z(W_N^{(i)}) \leq Z(W_N^{(j)}), \quad \text{for all } i \in \mathcal{A}, j \in \mathcal{A}^c,$$

这保证了选取的 k 个子信道是所有子信道中可靠性最高的。最后，我们记 $u_{\mathcal{A}} \in \mathcal{X}^k$ 为传输的 k 位有效信息位， $u_{\mathcal{A}^c} \in \mathcal{X}^{N-k}$ 为传输的无效信息位，称为冻结位。冻结位 $u_{\mathcal{A}^c}$ 通常是通信前双方约定的一串固定的比特序列，不携带任何有效信息。极化码的误码率表现与冻结位 $u_{\mathcal{A}^c}$ 的选取方式无关。

B. 极化码的编码与译码

极化码的 N 阶生成矩阵表示为

$$G_N = B_N \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes n}$$

其中， $A^{\otimes n}$ 是矩阵 A 的 Kronecker 积， B_N 是位反转 (bit-reversal) 置换矩阵。极化码的编码方法如下

$$x_1^N = u_{\mathcal{A}} G_N(\mathcal{A}) + u_{\mathcal{A}^c} G_N(\mathcal{A}^c)$$

其中， $G_N(\mathcal{A})$ 表示生成矩阵 G_N 中由集合 \mathcal{A} 中元素指定的行所构成的子矩阵。注意到，如果将冻结位 $u_{\mathcal{A}^c}$ 视为不变，则上式中第二项

$$c = u_{\mathcal{A}^c} G_N(\mathcal{A}^c)$$

为一个常向量。

SC 译码是极化码经典的译码算法。设 x_1^N 为信道输入， y_1^N 为信道输出。在 \mathcal{A} ， $u_{\mathcal{A}^c}$ ， y_1^N 已知的情况下，SC 译码器预测信息位 u_i 的函数关系为

$$\hat{u}_i = \begin{cases} u_i, & \text{if } i \in \mathcal{A}^c \\ h_i(y_1^N, \hat{u}_1^{i-1}), & \text{if } i \in \mathcal{A} \end{cases}$$

其中，决策函数 $h_i: \mathcal{Y}^N \times \mathcal{X}^{i-1} \rightarrow \mathcal{X}$ 定义如下

$$f(x) = \begin{cases} 0, & \text{if } \frac{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|0)}{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|1)} \leq 1 \\ 1, & \text{otherwise} \end{cases}, \quad \text{for all } y_1^N \in \mathcal{Y}^N, \hat{u}_1^{i-1} \in \mathcal{X}^{i-1}$$

注意到，SC 译码器从 $i = 1$ 到 $i = N$ 逐个对 \hat{u}_i 进行预测。对于有效信息位 $i \in \mathcal{A}$ 而言， \hat{u}_i 的预测依赖于整个信道输出 y_1^N 和之前预测的比特位 \hat{u}_1^{i-1} 。对于无效信息位 $i \in \mathcal{A}^c$ 而言， u_i 表示冻结位，由于通信前 $u_{\mathcal{A}^c}$ 已知， u_i 可以直接得出。

第三章 网络模型

本章主要介绍本论文采用的基于神经网络的极化码译码模型及其工作机制，以及在实际仿真中神经网络的训练过程和测试过程。本章最后会在模型建立的基础上提出本论文的主要研究方向。

1. 网络模型

本论文采用的基于神经网络的极化码译码模型如下图 4 所示：

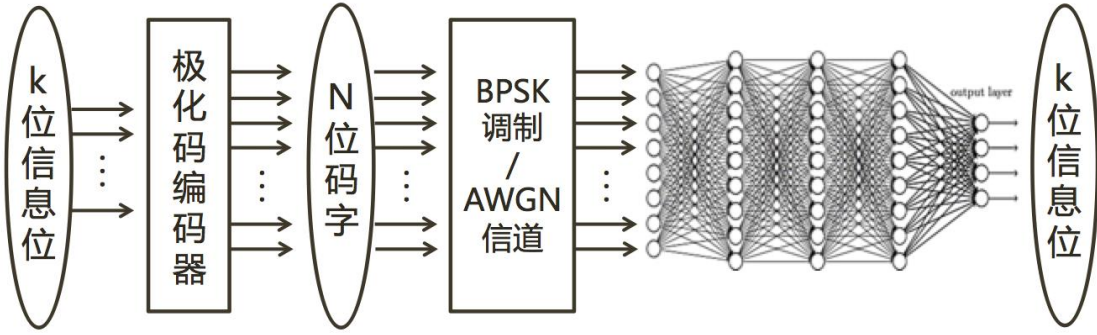


图 4：极化码译码模型

由图可见，通信发送方生成一个长度为 k 的信息位，作为将要传输的信息。这 k 位信息首先经过极化码编码器编码成长度为 N 位的码字，而后经过 BPSK 调制后送入 AWGN 信道。在通信接收方，信道输出的带噪声的 N 比特信息作为神经网络输入，网络经计算输出的 k 比特信息就是神经网络的译码结果。

如图所示，本论文所采用的神经网络是最基本的全连接式神经网络，信息从左往右单向传播。网络输出层包含 k 个神经元，代表发端发送的 k 个信息位。本论文采用含有三个隐层的神经网络。在后文中我们将用 $n1-n2-n3$ 表示从左至右三个隐层各自包含的神经元个数。显然，神经网络输入层包含 N 个神经元。

本论文中，由于输出层表示译码结果，因此输出层的激活函数选 Sigmoid 函数，网络输出范围为 $(0, 1)$ ，表示译码结果为 1 的置信度，是一个软判决值。最终判决时，网络输出在 $(0, 0.5)$ 区间内的判决为 0， $(0.5, 1)$ 区间内的判决为 1。其余隐层的激活函数均选 ReLU 函数。

本论文中极化码的冻结位 $u_{\mathcal{A}^c}$ 一律取全零向量，即无效信息位上全部传 0。

2. 模型训练与测试

基于神经网络的极化码译码实现过程分为训练过程和测试过程。下面分别进行介绍。

A. 训练过程

上文提到，神经网络的训练是通过已知输入和输出的训练集来进行的。在以上模型中，神经网络输入是带噪声的 N 位极化码码字，正确输出是最初生成该码字 k 位信息位。因此，训练集可以通过随机生成码字来产生。训练过程如图 5 所示。

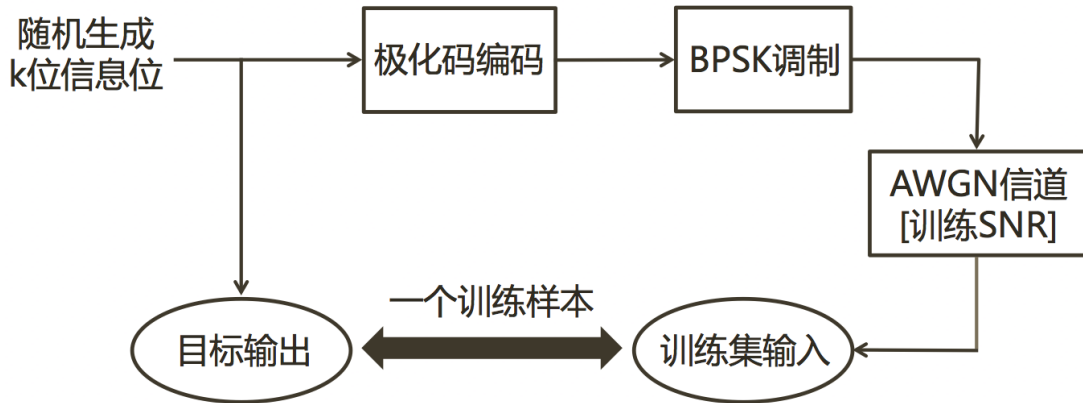


图 5：神经网络训练过程

如图 5 所示，训练时，首先随机生成 k 位信息位作为待传的信息。而后依次通过极化码编码器、BPSK 调制器和 AWGN 信道，得到 N 位带噪声的码字，这就是训练集的输入。该输入对应的正确输出就是最开始随机生成的 k 位信息位。这一对输入和输出就构成了一个训练样本。按照这个方法可以产生无数个不同的训练样本。注意到，虽然 k 位信息位仅对应有限的 2^k 种不同的码字，但是经过 AWGN 后带噪声的码字有无数种。因此可以认为，在训练过程中，神经网络输入各不相同，训练集数量无限，这就排除了训练时过拟合的可能性。另一方面，由于训练集输出就是最开始随机生成的信息位，这就免去了人为标注训练样本的过程，大大简化了训练成本。这是将神经网络用于信道译码的优势所在。

在实际训练过程中，我们往往将多个码字组合在一起形成一个矩阵，将该矩阵作为神经网络训练集的输入，同时将训练集输出按照同样的方法组合，这样可

以实现多个码字的同时训练。我们定义，每同时训练 2^k 个随机码字称为一个**训练周期(Training epoch)**。由于可能码字的总数为 2^k ，从统计意义上看，每一个训练周期中都把所有的码字进行了一次训练。显然，随着信息位个数 k 的增加，可能的码字总数呈指数增加，每一个训练周期训练的码字个数也呈指数增加。这大大降低了训练速度，也一定程度上限制了神经网络译码最大可能的信息位个数。

训练过程中，码字经过 BPSK 调制后进入 AWGN 信道。AWGN 信道存在一个信噪比的参数。我们定义，训练时信道的信噪比为**训练信噪比(Training SNR)**。显然，训练 SNR 高时，网络输入的噪声小；训练 SNR 低时，网络输入噪声大。不同的训练 SNR 所训练出来的神经网络在性能上存在差异。本次论文研究的重点之一，就是探究训练 SNR 对神经网络学习过程的影响，为实际训练 SNR 的选择提供指导性建议。

B. 测试过程

训练完成之后就是神经网络译码的测试过程。测试过程如图 6 所示。

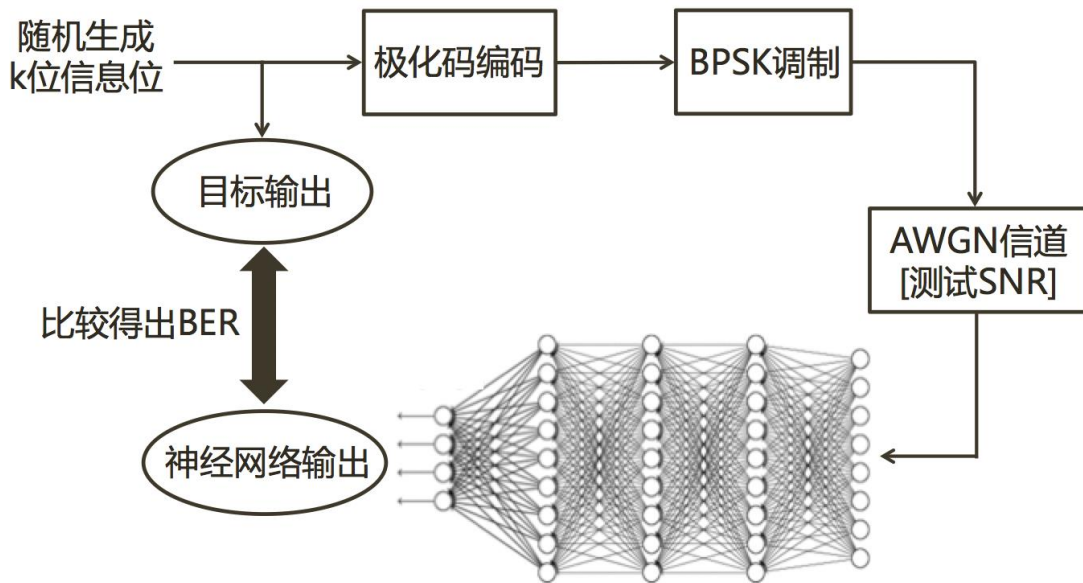


图 6: 神经网络测试过程

测试过程中，神经网络的所有参数都是固定的，目的是为了测试网络在不同的信道条件下的误码率曲线。测试过程与训练过程相似。首先在发送端随机生成大量码字，而后经过 BPSK 调制和 AWGN 信道后输入神经网络。此时 AWGN 信道的信噪比称为**测试信噪比(Test SNR)**，是测试时模拟的信道参数值。神经网络

的输出就是网络的译码结果。只需要将此结果与正确结果，即最开始随机生成的码字进行一一比对，即可得到神经网络在该测试 SNR 下的误码率数值。调整测试 SNR 的数值就可以得到神经网络在不同信道环境下的误码率曲线。

3. 研究方向

基于以上模型，本次论文的研究方向主要有以下三个方面：

(1) 训练 SNR 的选取对训练结果的影响。正如上一节提到，训练 SNR 是神经网络训练过程中的一个关键的参数。本次论文将探究在不同的训练 SNR 下神经网络学习过程的差异，主要将从学习速度和误码率两个方面进行评估。

(2) 神经网络规模对训练结果的影响。神经网络规模也是深度学习需要预先指定的超参数之一。本论文将探究在不同神经网络规模下学习过程的差异，也将从学习速度和误码率两个方面进行比较。

(3) 极化码码率对训练结果的影响。在码长 N 不变的情况下，极化码的码率 R 决定了信息位的个数 k ，也决定了可能的码字总数。本论文将探究在码长 N 不变时，随着信息位 k 个数的变化，神经网络译码性能的变化规律。

第四章 仿真结果与讨论

基于第二章介绍的神经网络信道译码模型与三个主要的研究方向，本章将依次给出相应的仿真结果，并根据仿真数据归纳、总结出一些经验性结论。

本论文中，除了极化码 SC 译码的误码率曲线仿真在 MATLAB 平台上进行以外，其余所有仿真在 Python 中进行。神经网络框架借助 TensorFlow 模块实现。训练时的损失函数定义为标准输出与实际输出的均方误差值(MSE)。训练优化时选择 Adam 优化器，初始学习率取 0.0001。

1. 训练 SNR 对训练结果的影响

在对用于信道译码的神经网络进行训练时，训练 SNR 的选取是自由的。但是，不同的训练 SNR 训练出来的神经网络性能也是不同的。本节的内容就是探究训练 SNR 对训练过程和结果的双重影响，为实际训练 SNR 的选取提供经验性指导。

A. 训练 SNR 影响的定性分析

在这一部分中，首先选取极化码码长 $N=16$ ，信息位 $k=8$ ，规定编号为 9—16 的八个子信道传信息。神经网络规模选为 128-64-32。为了探究不同训练 SNR 下神经网络学习过程的差异，我们分别选取训练 SNR=8, 10, 12, 14 四个值，记录它们在训练过程中误码率的变化曲线，如图 7 所示。

图 7 中，横坐标是训练周期，代表训练过程从左往右逐步推进；纵坐标是神经网络在测试 SNR=10dB 条件下的误码率数值。四条曲线分别代表四种训练 SNR 下网络的学习过程。可以看到，随着训练过程的不断进行，四条曲线的误码率数值均呈下降趋势，这就是网络通过训练集学习极化码译码算法的过程。

如图 7 中所示，四条曲线的误码率在下降到 0.06 附近时都经历了一段时间的停滞。在这段时间内，网络的误码率维持在 0.06 附近，不随训练周期的增加而下降。我们将这样的点定义为**驻点**。从之后的仿真数据中可以看到，驻点是神经网络学习信道译码算法过程中一个普遍存在的现象。通过观察曲线特征，结合

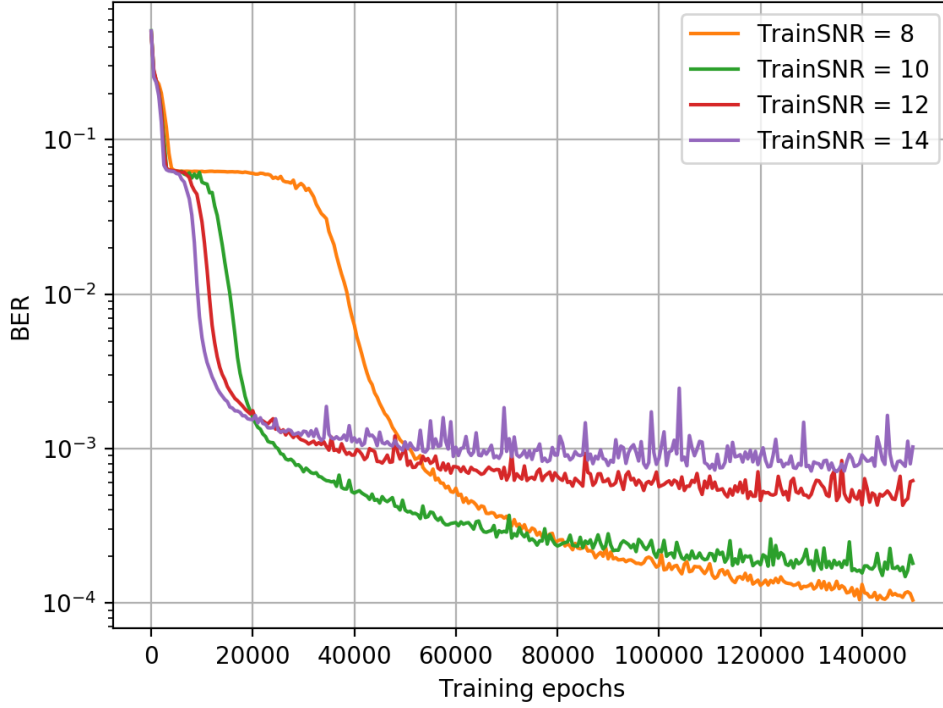


图 7: 不同训练 SNR 条件下的训练过程图

驻点的定义，我们将神经网络的学习过程划分为四个阶段：**准备期、停滞期、学习期、收敛期**。准备期是神经网络从刚初始化后的状态（误码率约为 0.5）经训练快速下降到驻点附近的阶段，对应图 7 中左上角的一小部分；停滞期是网络误码率维持在驻点附近的一段时期；学习期对应图 7 中四条曲线主要下降的阶段，是神经网络离开驻点，真正开始学习的时期；收敛期是神经网络学习基本完成的时期，对应图 7 中误码率趋于平稳的阶段。

观察图 7 发现，对于不同的训练 SNR，网络学习过程是不同的，下面从两个层面分析。

(1) 从最终误码率数值上看，经过 150000 次训练周期后，训练 SNR=8dB 对应的神经网络在测试 SNR=10dB 信道下的误码率最低，接近 1×10^{-4} 。接下来分别是训练 SNR=10dB 对应的神经网络，测试误码率约为 1.8×10^{-4} ；训练 SNR=12dB 的神经网络，测试误码率约为 6.2×10^{-4} ；训练 SNR=14dB 的神经网络，测试误码率约为 1×10^{-3} 。可以发现，神经网络的训练 SNR 越低，测试误码率也越低。值得注意的是，按照对机器学习规律的常规理解，测试集数据分布越接近训练集数据分布，模型在测试集上的表现应该越好，因为这时模型能够更好地将训练时学习到的映射规律应用到测试集上。此次仿真中，由于测试

SNR=10dB，按常理推断应该是训练 SNR=10dB 的神经网络表现最佳。但仿真结果表明，训练 SNR=8dB 的神经网络的测试误码率优于前者。这是神经网络应用在信道译码领域上的特殊性。

(2) 从停滞时长上看，训练 SNR=8dB 的神经网络的停滞期时间最长，大约持续了 20000 个训练周期。另外，训练 SNR 越高，神经网络停滞时间越短，网络越早从停滞期进入学习期。为了描述方便，本论文将用神经网络停滞期的长短作为衡量网络学习速度的指标。停滞时间越短，我们称之为学习速度越快；反之，学习速度越慢。需要说明的是，神经网络优化器的学习率（决定了每次参数优化的步幅）也对学习速度有所影响。为了控制变量，本论文中所有曲线的仿真都是在同一个 Adam Optimizer 学习率 0.0001 下完成的。从图 7 可以得出，在学习率相同的前提下，训练 SNR 越高，网络学习速度越快；反之，学习速度越慢。

图 7 反映的是测试 SNR=10dB 时的误码率随训练周期的变化曲线。为了验证神经网络在不同测试 SNR 下的误码率，我们按照图 7 继续对神经网络训练至 200000 个周期，并对训练完成后的网络进行了误码率仿真，如下图 8 所示：

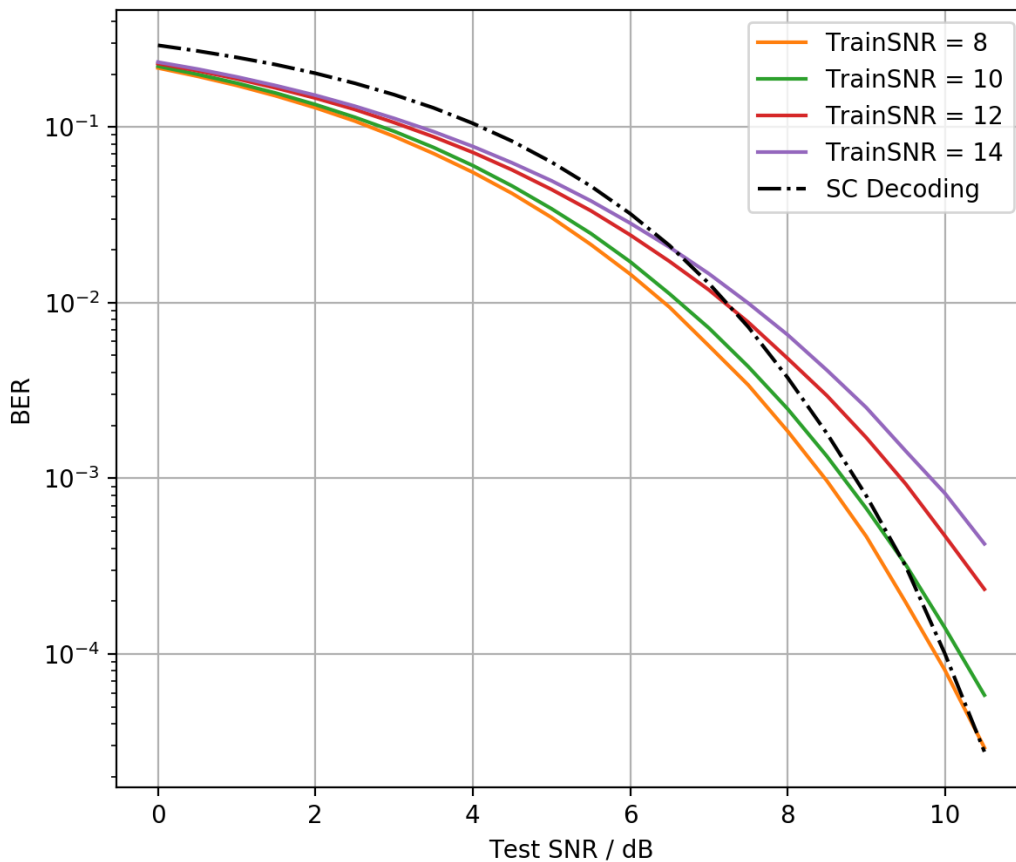


图 8：不同训练 SNR 条件下的误码率曲线

上图中，横坐标是测试 SNR 值，纵坐标是误码率，不同颜色的曲线代表不同训练 SNR 下神经网络的训练结果。其中，黑线表示极化码 SC 译码算法的误码率曲线，以供参考。通过图中可以看到，经过 200000 个训练周期后，训练 SNR=8dB 对应的神经网络的误码率最低，且在测试 SNR 取 0dB 至 11dB 的整个范围内均优于其他训练 SNR 对应的神经网络。从图中还可以看出，误码率曲线从低往高的次序依次对应训练 SNR=8, 10, 12, 14dB 的神经网络。这进一步验证了神经网络的训练 SNR 越低，测试误码率也越低的结论。

B. 结论验证

为了进一步验证以上得出的初步结论，我们调整了一些基本参数：极化码码长 $N=16$ 不变，信息位 k 从 8 变为 10，规定编号为 7—16 的十个子信道传信息。神经网络规模从 128-64-32 变为 256-128-64。训练 SNR 分别取 6, 8, 10, 12, 14dB，测试 SNR 同样取 10dB，仿照图 7 给出神经网络训练过程图如下图 9 所示：

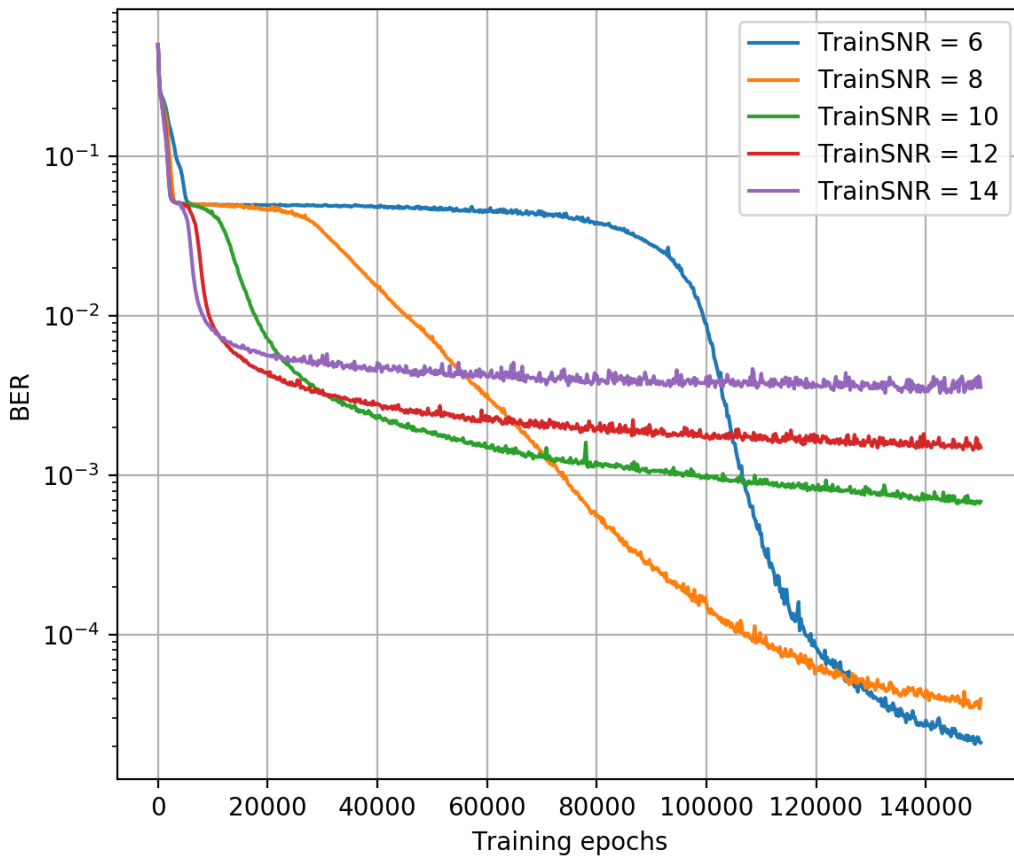


图 9：不同训练 SNR 条件下的训练过程图

观察图 9 发现,神经网络学习过程仍然可以划分为准备期、停滞期、学习期和收敛期四个阶段,驻点现象依然存在,位置大致在 $BER=0.05$ 处。从最终误码率水平看,经过 150000 个训练周期后,训练 $SNR=6dB$ 对应的神经网络在测试 $SNR=10dB$ 下的误码率最低,大约为 2.1×10^{-5} 。往上依次对应的是训练 $SNR=8, 10, 12, 14dB$ 的神经网络。这印证了训练 SNR 越低,神经网络误码率性能越好的结论。从学习速度上看,训练 $SNR=6dB$ 对应的神经网络学习速度最慢,停滞期持续了接近 80000 个训练周期。接下来依次是训练 $SNR=8, 10, 12, 14dB$ 对应的神经网络。这印证了在学习率相同的前提下,训练 SNR 越高,神经网络学习速度越快的结论。

至此,我们总结出两个基本结论:

结论一:神经网络在极化码译码训练过程中,会经历准备期、停滞期、学习期、收敛期四个阶段;

结论二:在一定范围内,训练 SNR 越高,神经网络学习速度越快,但最终的误码率越高;训练 SNR 越低,神经网络学习速度越慢,但最终的误码率也越低。

C. 结论分析

对于结论二,可以用以下思路理解。神经网络训练极化码译码的过程,实际上等价于在训练集上寻找一个映射关系,使训练集中每个码字都被映射到正确的译码结果上去。训练 SNR 高时,训练集的噪声小,不同训练码字间的汉明距离大,映射关系更明显,因此神经网络的学习难度更小,速度越快。但是,因为网络是在低噪声环境下训练出来的,不具备处理噪声的能力,因此一旦测试集的噪声较高,神经网络的译码准确率就会大大降低。相反,训练 SNR 低时,训练集本身噪声较高,码字间的界限更加模糊,码字到译码结果的映射关系也更复杂,因此神经网络的学习难度更大,速度越慢。但是,高噪声环境下训练出来的神经网络不仅可以处理低噪声的信号,还能够适应高噪声的测试环境,因此相比于低噪声环境下训练的神经网络而言适用范围更广、译码准确率越高。换言之,降低训练 SNR 的过程其实是提高神经网络译码能力的过程,而非使神经网络适用于不同信道条件的过程。这个过程的代价是学习速度变慢,学习时间变长。当然,如果训练 SNR 过低,训练噪声太大,输入、输出的映射关系被完全淹没在噪声

中,那么神经网络将理论上无法学习到译码算法,或者说,学习时间无限长。因此,准确的说,结论二是在一定范围的训练 SNR 内成立的。

事实上,结论二揭示了人工神经网络与人类学习过程间奇妙的相似性。结合生活实际,当人们通过大量练题来学习一个新的领域时,如果题目非常浅显易懂,我们能够很快速的理解一些基本道理。但是凭借这些知识,我们能够解决的实际问题很有限,因为一旦问题变得复杂,我们将无能为力。这对应了神经网络模型中训练 SNR 太高的情况。反之,如果训练的题目本身很复杂,那么人们学习起来会感到吃力、费时,这对应了低训练 SNR 曲线上较长的停滞期。但是一旦我们掌握了更深奥的难题,就能够解决实际中更多的问题,包括简单的和复杂的。这就是高噪声环境下训练出来的神经网络误码性能更好的原因。通过这个类比,我们可以对结论二建立一个更加直观和感性的认识。同时,这一类比也一定程度上说明了人工神经网络与真实人脑之间除了结构上相似以外,功能上也存在某种相似性。

D. 训练 SNR 与学习速度的定量分析

接下来探究学习速度与训练 SNR 之间定量关系的问题。根据结论二,为了得到更低的误码率,训练神经网络时应该选择尽可能小的训练 SNR。另一方面,训练 SNR 越小,学习速度越慢,训练时间越长,而在实际操作中无限期训练下去是不现实的。因此,我们希望能够根据训练 SNR 来预计训练时间,或者说,确定训练时间与训练 SNR 之间的定量关系,以便在实际中指导我们对训练 SNR 的选取。

本论文中,我们定义**开始学习时间(Learning Start Time, LST)**为神经网络结束停滞期、进入学习期时所经历的训练周期数。我们将开始学习时间作为对神经网络学习速度的定量描述。显然,学习速度越快,停滞期越短,开始学习时间越小。神经网络进入学习期的标志可通过误码率下降到某一临界值来确定。根据停滞期和学习期的定义,这个临界值应略低于训练时的驻点。由于不同的参数设定下驻点的位置是不确定的,因此临界值的选取也要根据具体情况分析。事实上,误码率临界值的选取不影响我们对开始学习时间随训练 SNR 变化关系的判断。

我们首先回到图 7 的参数设定:极化码码长 $N=16$, 信息位 $k=8$, 编号 9—16

的子信道传信息，神经网络规模为 128-64-32，测试 SNR=10dB。观察图 7 的仿真结果知，驻点位置大致在 BER=0.06 处。我们规定临界点为 BER=0.05，开始学习时间为曲线误码率下降到临界点 BER=0.05 时所经历的训练周期数。下图 10 记录的是不同的训练 SNR 所对应的开始学习时间。横轴是训练 SNR，纵轴是开始学习时间。由于训练过程的随机性，每次训练的开始学习时间会在小范围内波动，因此图 10 中的每个点都是训练十次所记录的开始学习时间的平均值。

图 10 进一步印证了结论二关于学习速度的论断：在的一定范围内，训练 SNR 越低，开始学习时间越长，学习速度越慢。同时，我们注意到，随着训练 SNR 的降低，开始学习时间增加的速度越来越快，呈现类似指数增长的趋势。另一方面，我们注意到，训练 SNR 不能过低，否则神经网络将理论上无法学习成功，这时将不存在开始学习时间，因此曲线的自变量有下界。

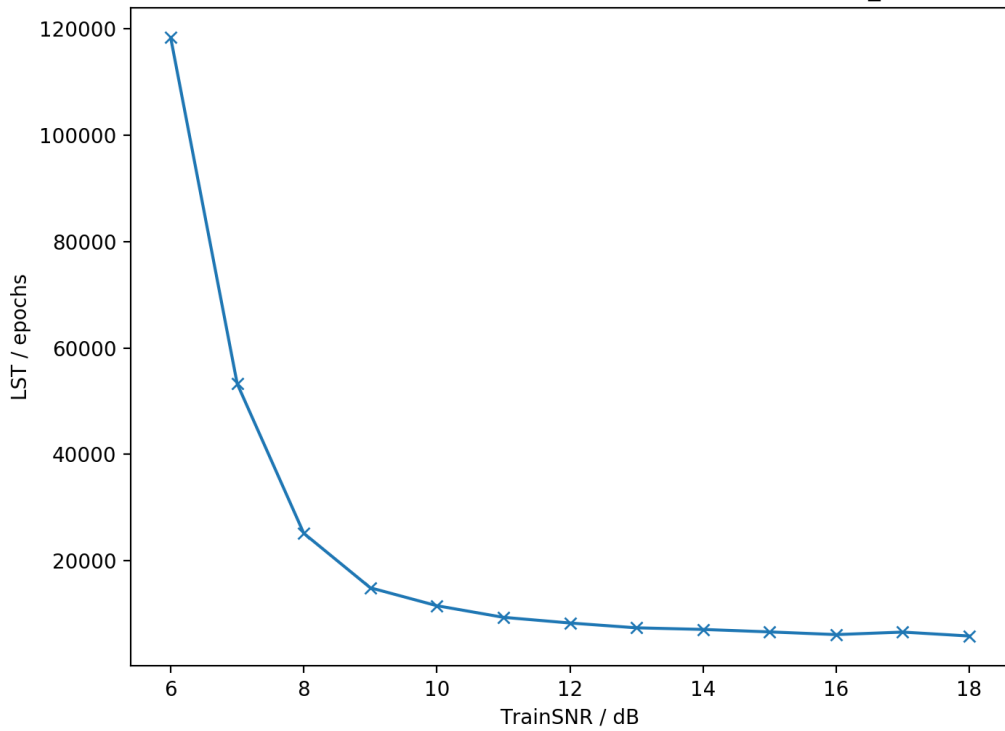


图 10: 开始学习与训练 SNR 关系图

为了进一步验证以上得出的规律，我们再次考虑图 9 所示情形：极化码码长 $N=16$ ，信息位 $k=10$ ，编号 7—16 的十个子信道传信息，神经网络规模 256-128-64。测试 SNR 取 10dB。图 9 中驻点的位置大致在 BER=0.05，因此我们规定临界值 BER=0.035。图 11 记录了这种情形下开始学习与训练 SNR 的关系。

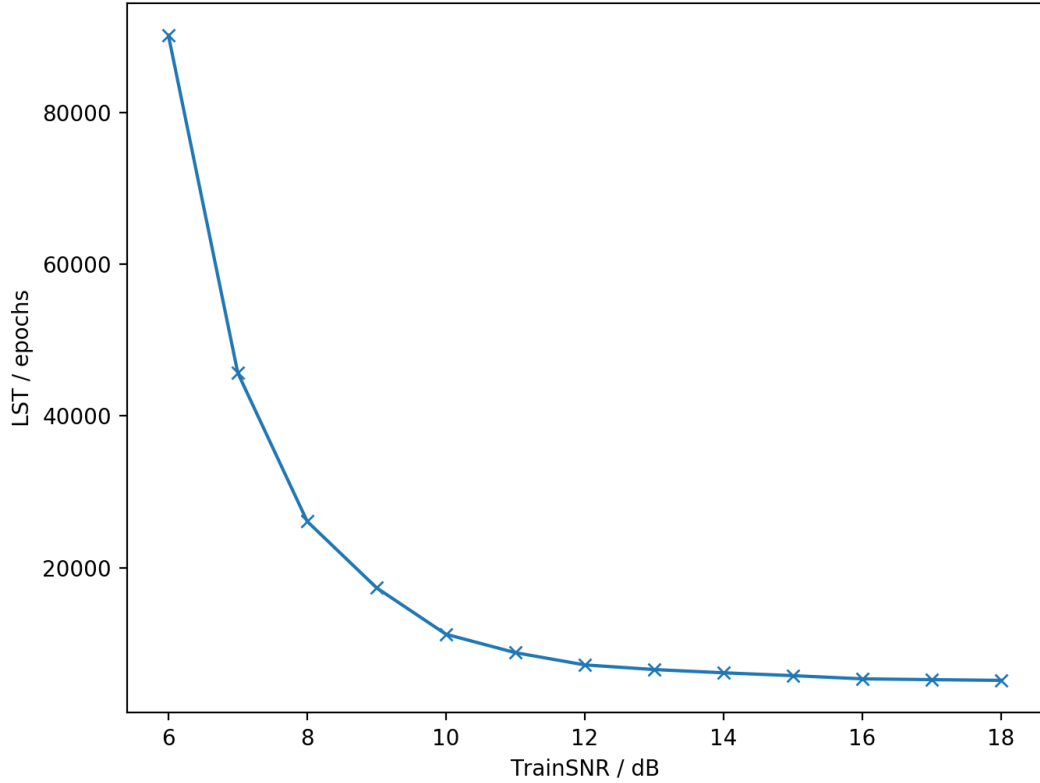


图 11: 开始学习与训练 SNR 关系图

比较图 10 和图 11 可以看出, 开始学习与训练 SNR 之间均满足类似指数函数的关系。

由此, 我们可以得出以下结论:

结论三: 在一定范围内, 神经网络开始学习时间随训练 SNR 的减小指数增长。

由结论三我们可以总结出, 虽然在一定范围内降低训练 SNR 可以提高神经网络最终的译码性能, 但是随着训练 SNR 的减小, 训练时间的边际成本是不断增加的, 因此选取训练 SNR 其实是一个权衡训练时间与最终性能的过程。在实际操作中, 我们往往有一个最长训练时间的限制。根据图 10、图 11, 我们可以很容易的找到开始学习与训练 SNR 的对应关系, 从而能够估计出可能的最低训练 SNR。这个最低的训练 SNR 就是实际中训练 SNR 的最佳取值。这是图 10、图 11 在实际操作中的意义所在。

2. 神经网络规模对训练结果的影响

在这一部分, 我们将探究神经网络的规模对极化码译码的影响。我们将主要

比较 128-64-32 结构的神经网络（记为“小规模神经网络”）与 256-128-64 结构的神经网络（记为“大规模神经网络”）在学习速度和最终误码率上的区别。

取极化码码长 $N=16$, $k=8$, 编号 9—16 的子信道传信息, 我们对两种规模的神经网络均进行 100000 个周期的训练, 训练 SNR 分别取 8, 12dB, 得到各自的误码率曲线如下图 12 所示。

图 12 中, 实线代表小规模神经网络的误码率曲线, 虚线代表大规模神经网络的误码率曲线, 不同颜色的曲线代表不同的训练 SNR。同样地, 黑线代表极化码 SC 译码下的误码率曲线。从图中可以看出, 相同的训练 SNR 下, 大规模神经网络的误码率低于小规模神经网络的误码率。这一点相对容易解释。神经网络规模越大, 网络参数就越多, 就能够更快、更精确地拟合任何一个函数, 学习过程也更加容易。在传统深度学习训练过程中, 增加网络规模往往伴随着更高的过拟合风险, 因此需要控制网络参数总量。但前面提到, 信道译码的训练中不存在过拟合的风险, 因此增加神经网络规模几乎能够保证最终的译码性能更佳。

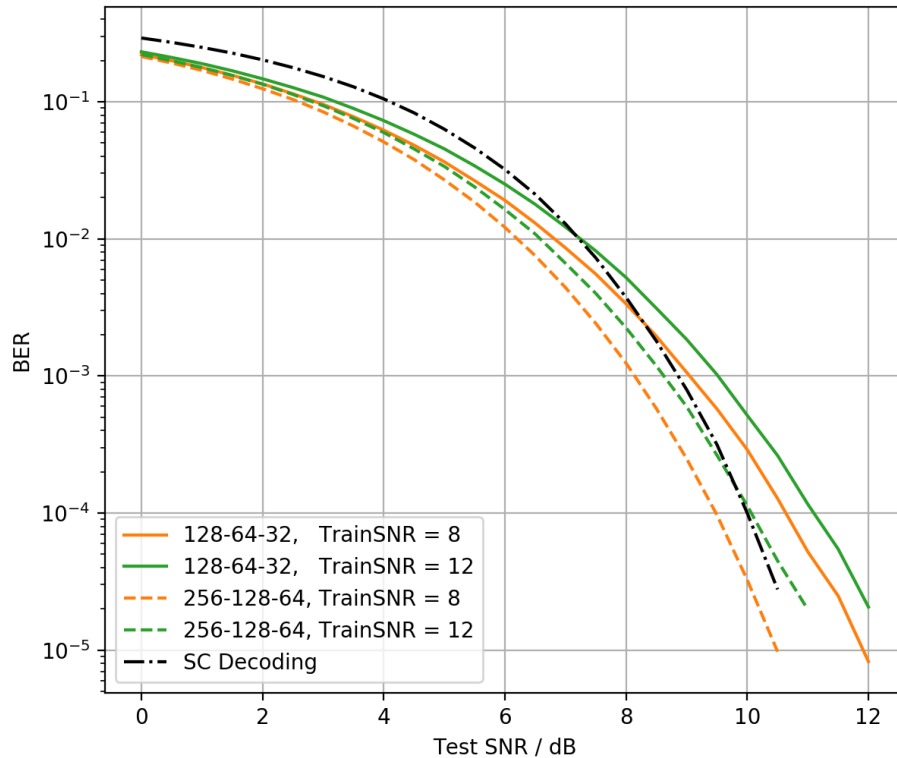


图 12: 不同规模神经网络的误码率曲线

图 12 直观的展示了大规模神经网络在最终误码率上的优势。为了探究不同规模神经网络在学习速度上的差异, 我们给出如图 13 所示的训练过程图的对照。

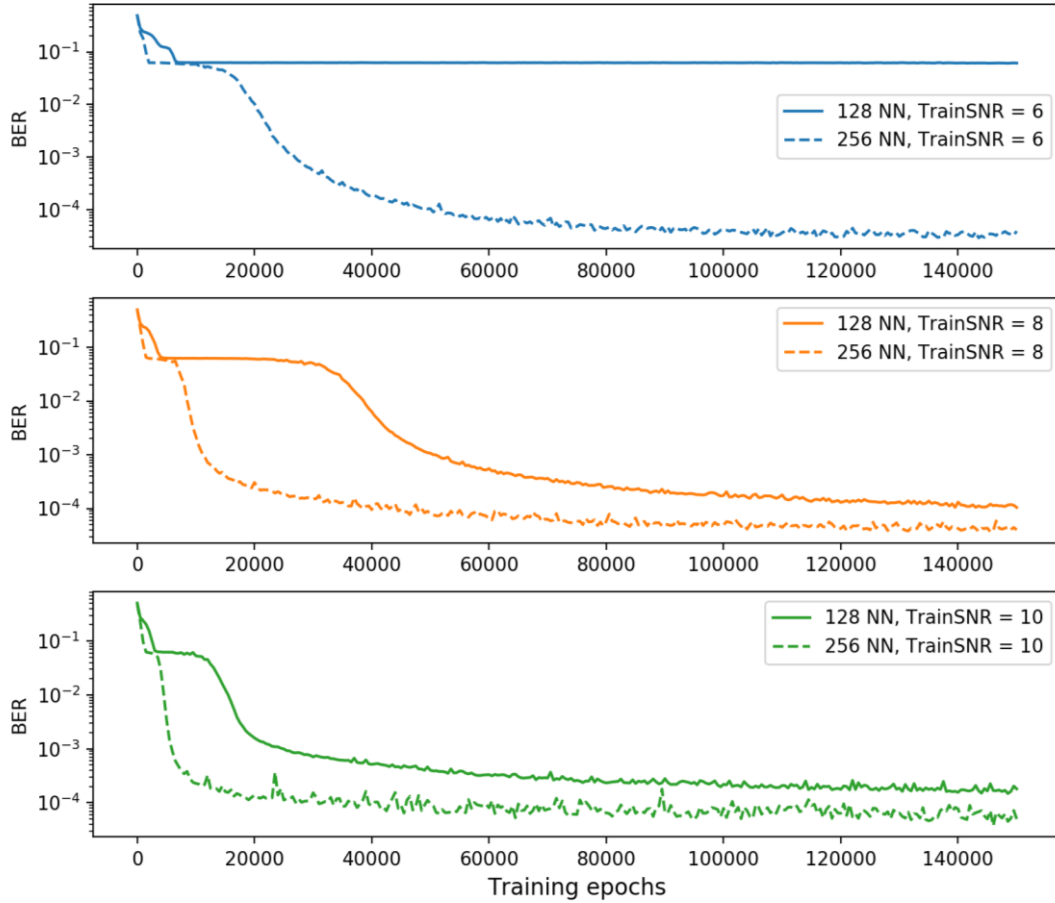


图 13: 不同规模神经网络的误码率曲线

图 13 中, 横轴是训练周期, 纵轴代表神经网络在测试 $\text{SNR}=10\text{dB}$ 下的误码率。三张图分别代表训练 $\text{SNR}=6, 8, 10\text{dB}$ 时的训练过程图。每张图中, 实线代表小规模神经网络, 虚线代表大规模神经网络。综合三张图的信息发现, 大规模神经网络学习过程中的停滞期明显短于小规模神经网络, 说明大规模神经网络的学习速度更快。注意到, 在第一张图的训练周期内, 小规模神经网络由于学习速度过慢, 始终处于停滞期, 尚未进入学习期, 所以曲线一直保持水平。

从图 13 中第二、三张图还可以看到, 大规模神经网络不仅学习速度比小规模神经网络更快, 最终的误码率也更低。这印证了图 12 中观察得出的结论。由此, 我们将该结论总结如下:

结论四: 其他条件不变, 神经网络规模越大, 学习速度越快, 且最终的误码率越低。

结论四事实上体现了许多领域中用空间换时间的思想。神经网络规模越大, 参数越多, 需要的存储空间必然更大。但是更大的存储空间可以带来更快的学习

速度和更低的误码率。

另外需要指出的是，神经网络规模越大，需要更新的参数数量就更多，单次训练消耗时间也越长。当神经网络规模足够大时，继续增加网络规模不仅不能节省训练时间，反而会使训练时间变长。因此，在实际操作中，需要权衡训练周期数和每个训练周期所耗时间。神经网络规模不是越大越好，需要通过大量仿真测试来合理选择。

3. 极化码码率对训练结果的影响

这一部分，我们将探究极化码码率对神经网络译码的影响。我们固定极化码长 $N=16$ ，比较信息位个数 $k=8, 10$ 两种情况下学习速度和最终误码率上的区别。

取神经网络规模为 256-128-64，训练 SNR 分别取 6, 8, 10dB，比较极化码信息位 $k=8$ 和 10 两种情况下的训练过程图如下所示：

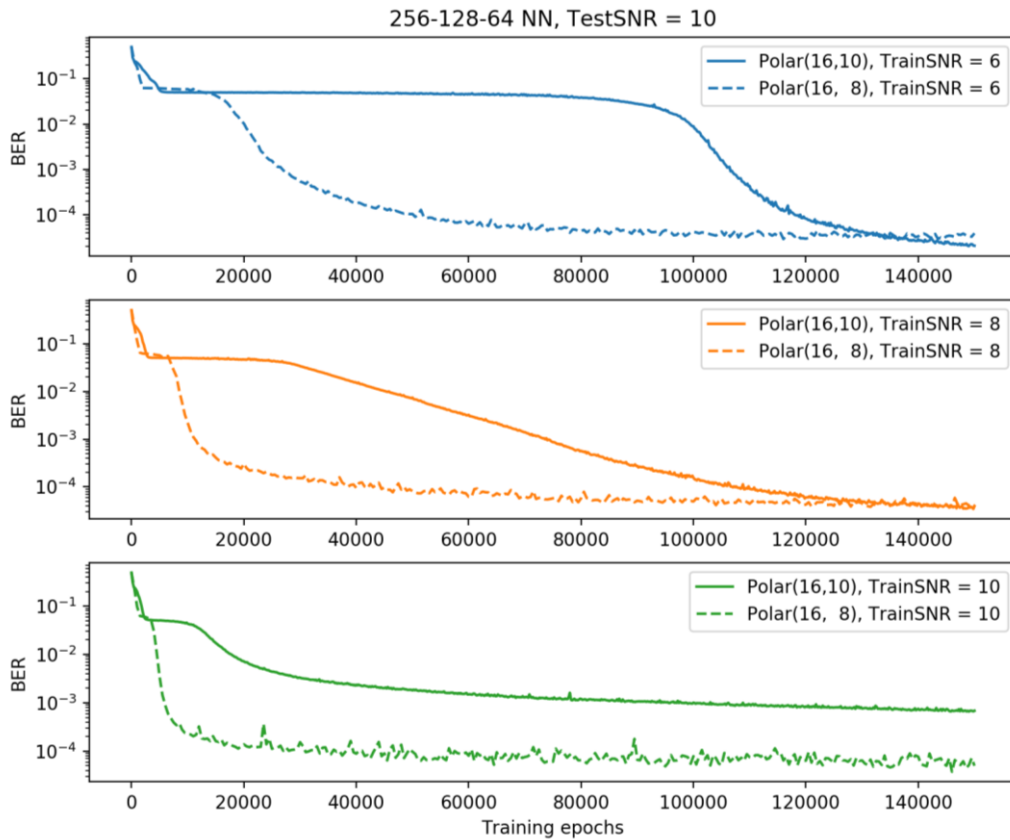


图 14: 不同码率下的训练过程图

图 14 中，横轴表示训练周期，纵轴是在测试 SNR=10dB 时测得的神经网络

误码率。实线表示信息位 $k=10$ 时神经网络的训练过程，虚线表示信息位 $k=8$ 时神经网络的训练过程。三张图分别表示训练 SNR=6, 8, 10dB 时对应的训练过程。从学习速度上看，信息位个数越少，停滞期时间越短，学习速度越快。这是因为信息位个数 k 决定了可能的码字总数 2^k ，从而决定了映射的复杂程度。信息位越少，码字总数越少，映射越简单，学习速度越快；反之，学习速度越慢。另一方面，由于我们定义同时训练 2^k 个码字为一个训练周期，因此随着 k 的增大，每一个训练周期需要训练的码字数量呈指数增加，训练时间也呈指数增加。总之，随着极化码信息位的增多，不仅单个训练周期所耗时间显著变长，总共需要的训练周期数也变多。因此，信息位个数成为了影响神经网络训练时间的关键因素。事实上，这也正是目前神经网络译码只能适用于短码的主要原因。

从最终误码率上看，图 14 中出现了一个很有意思的现象。对于传统的信道译码算法而言，提高码率的后果往往是误码率上升。这是因为，码率越高，码字间的汉明距离更小，或者说，比特间的相互联系更弱，因此译码器的纠错能力也降低，误码率上升。但从图 14 第二张图来看，信息位 $k=8$ 和 10 最终对应的误码率几乎相同；从第一张图看，甚至 $k=10$ 的误码率低于 $k=8$ 的误码率。为了验证第一张图的“反常”现象在其他测试 SNR 下是否存在，我们取第一张图中两条曲线各自训练 150000 个周期的结果，验证它们在 0—10dB 范围内的误码率曲线，如下图 15 所示：

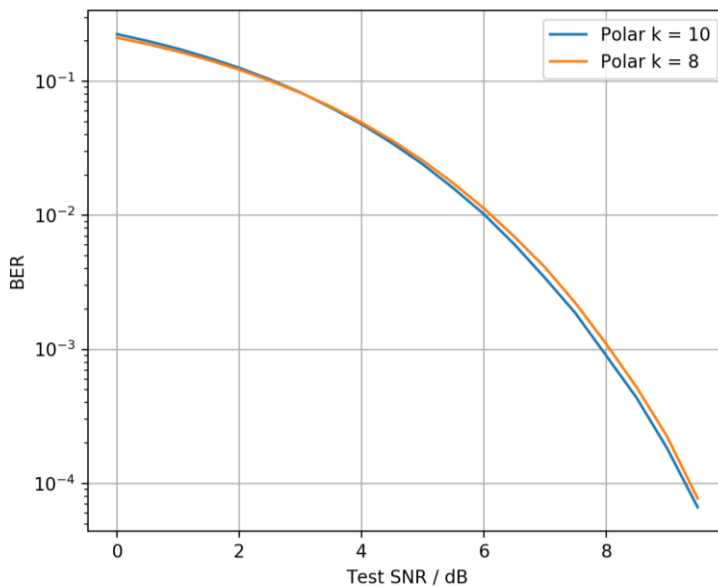


图 15：不同码率下的误码率曲线

从图 15 可以看出,在大部分测试 SNR 下,信息位 $k=10$ 对应的误码率的确略低于 $k=8$ 时的误码率。这说明极化码码率越高,误码率反而越低的“反常”现象确实存在。这种现象产生的原因可能与神经网络译码与传统译码算法基于的原理不同有关。与传统译码算法不同,神经网络译码不是基于比特间的相互联系,因此码率越高,误码率一定越高的规律不适用于神经网络译码。对于神经网络而言,任何学习过程,包括信道译码,都可以归结为一个函数拟合的过程。在训练过程中,神经网络通过不断优化参数一步步逼近最佳的映射关系,而并未考虑译码过程的具体细节和原理。因此,真正的译码算法对神经网络是不可见的。这种译码原理上的差异必然会使基于神经网络的信道译码拥有一系列独有的规律。当然,“反常”现象也不总是存在的,比如训练 SNR=10dB 时,信息位个数越多,误码率也越高(见图 14 第三张图);另外,设想信息位个数 k 过高,如 $k=N=16$,码字间汉明距离均为 1,相当于没有信道译码过程,此时误码率应该会高于 $k=8$, 10 时的误码率。因此,“反常”现象的出现是不确定的。

至此,我们总结出下面的结论:

结论五:极化码码率越高,神经网络学习速度越慢,但最终误码率的变化还取决于训练 SNR 的选取、神经网络规模等其他因素。

由于随着信息位增多,训练时间指数增加,仿真难度过大,因此本论文没有进一步探究神经网络误码率随极化码码率的变化规律。作者猜想,对于不同的极化码码率,最佳训练 SNR 的选取是不同的,因此图 14 在相同训练 SNR 下比较不同码率的极化码的误码率存在不公平的因素。关于此问题的探究,可作为下一步的研究方向。

第五章 总结与展望

1. 结论总结

经过大量仿真，本论文总结出以下五条关于用神经网络进行极化码译码的经验性结论：

结论一：神经网络在极化码译码训练过程中，会经历准备期、停滞期、学习期、收敛期四个阶段；

结论二：在一定范围内，训练 SNR 越高，神经网络学习速度越快，但最终的误码率越高；训练 SNR 越低，神经网络学习速度越慢，但最终的误码率也越低；

结论三：在一定范围内，神经网络开始学习时间随训练 SNR 的减小指数增长；

结论四：其他条件不变，神经网络规模越大，学习速度越快，且最终的误码率越低；

结论五：极化码码率越高，神经网络学习速度越慢，但最终误码率的变化还取决于其他因素。

以上结论均通过仿真结果归纳、总结得出，没有理论推导和依据。事实上，关于神经网络、深度学习背后的实质，学术界仍没有一个公认的答案。目前，人们使用神经网络时仍更多地将其视为一个“黑匣子”，为达到特定的目的而使用它。本论文也只是站在使用者的角度，通过仿真探索基于神经网络的信道译码的基本特点，并总结出一系列神经网络工作时的规律，为该领域今后的进一步研究打下基础。即使如此，本论文也试图在给出每一个结论的同时给出其直观解释，以说明结论的合理性，同时帮助我们更好地理解神经网络本身。

值得说明的是，本论文虽然以极化码为例，但是结论应该适用于其他信道编码方法，如 BCH 码、卷积码等。原因如前所述，对于神经网络而言，译码算法是不可见的；换言之，不同的译码算法在神经网络看来属于同一个问题，即建立一个从 N 比特输入码字到 k 比特输出信息位的最佳映射，从而屏蔽了码字构造方式、译码算法等具体细节。因此我们有理由认为，以上结论在基于神经网络的信道译码范畴内具有一定的普适性。

2. 研究展望

正如第一章提到，基于神经网络的信道译码具有诸多优点。然而，目前阻碍该领域进一步发展的主要原因在于训练码字数量随着信息位 k 的增多而指数增长^[7]，比如当 $k=100$ 时，码字总数达到了 2^{100} ，这在实际训练过程中是不可能完成的。本论文也仅讨论到了极化码码长 $N=16$ ，信息位个数 $k=10$ 的情况。这极大地限制了基于神经网络的信道译码方法投入应用的可能性。

针对这个问题，T. Gruber 等提出了一种可能的基于 zero-shot learning 的解决方案^[10]：仅将码字总量的一小部分用于神经网络的训练，如果神经网络能够学习到更深层次的译码规律，就能够预测所有码字的译码方法。如果该方案可实现，训练成本就可以大大降低，神经网络也就能够应用于更长的码字范围。但是，基于论文给出的仿真结果，虽然神经网络能够成功译码一些训练集之外的码字，但是误码率水平与传统译码算法相比差距仍然较大。今后的研究可以基于这个思路，专注于如何选取小部分有代表性的训练码字的问题。因为如果在小部分训练集上学习到的映射关系足够好，神经网络是有能力处理未训练过的数据的。

另外，全连接式的神经网络也不是唯一的选择方案。今后的研究还可以专注于其他类型的神经网络在信道译码问题上的表现，比如循环神经网络等。A. Santoro 等提出了一种叫做记忆增强型神经网络的模型，在 one-shot learning 的表现上具有很大潜力^[13]，也是值得探究的方向之一。在机器学习范畴，zero-shot learning 和 one-shot learning 均属于迁移学习，要点在于使神经网络能够预测训练集中没有出现过，或者出现次数很少的类别。这是解决神经网络信道译码时训练集指数增长问题的有效方法。

参考文献

- [1] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.
- [2] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1." Cambridge, MA, USA: MIT Press, 1986, pp. 318–362.
- [3] Y. LeCun, B. Boser, et al. "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, Dec. 1989.
- [4] J. Bruck and M. Blaum, "Neural networks, error-correcting codes, and polynomials over the binary n -cube," *IEEE Trans. Inform. Theory*, vol. 35, no. 5, pp. 976–987, Sept. 1989.
- [5] W. R. Caid and R. W. Means, "Neural network error correcting decoders for block and convolutional codes," in *Proc. IEEE Globecom Conf.*, vol. 2, Dec. 1990, pp. 1028–1031.
- [6] L. G. Tallini and P. Cull, "Neural nets for decoding error-correcting codes," in *Proc. IEEE Tech. Applicat. Conf. and Workshops Northcon95*, Oct. 1995, pp. 89-.
- [7] X.-A. Wang and S. B. Wicker, "An artificial neural net Viterbi decoder," *IEEE Trans. Commun.*, vol. 44, no. 2, pp. 165–171, Feb. 1996.
- [8] A. Hamalainen and J. Henriksson, "A recurrent neural decoder for convolutional codes," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, vol. 2, 1999, pp. 1305-1309.
- [9] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, July 2006.
- [10] Gruber, Tobias et al. "On deep learning-based channel decoding." 2017 51st Annual Conference on Information Sciences and Systems (CISS) (2017): 1-6.
- [11] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.
- [12] E. Arıkan, "Channel polarization: A method for constructing capacity- achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051 –3073, 2009.
- [13] Santoro, A., Bartunov, S. et al. "One-shot Learning with Memory-Augmented Neural Networks," *Proc. 33rd International Conference on Machine Learning*.

本科期间的主要工作和成果

本科期间参加的主要科研项目有：

1. 2017 年 1 月至 2017 年 12 月，在北京大学无线通信实验室赵玉萍教授的指导下开展了极化码在加密系统中应用的课题研究；

2. 2017 年 7 月至 2017 年 9 月，在哥伦比亚大学数据科学研究所 Xiaodong Wang 教授的指导下开展了关于低秩矩阵补全的课题研究。

已完成的论文有：

- [1] Junwei Zhang, Xuan Guo, Yuping Zhao. Reliable Cryptosystem based on Polar Codes, submitted to IEEE Communication Letters.
- [2] Morteza Ashraphijuo, Xiaodong Wang, Junwei Zhang. Low-Rank Data Completion With Very Low Sampling Rate Using Newtons Method, submitted to IEEE Transactions on Signal Processing.

致谢

我首先要感谢我的论文指导老师、北京大学无线通信实验室赵玉萍老师。赵老师对我论文的研究方向做出了指导性的意见，在论文撰写过程中及时对我遇到的困难和疑惑给予悉心指点，提出了许多有益的改善性意见。在此对赵老师给予我的帮助和关怀表示诚挚的谢意！同时，还要感谢无线通信实验室李斗老师、杨延军老师，在论文完成过程中提出宝贵的改进建议。感谢实验室郭璇师兄，一直以来带领我在实验室开展各项研究工作，提供论文选题的建议，并随时解答我的疑惑。感谢实验室所有的师兄师姐和同学们，大家在一起互相学习、互相帮忙，共同度过了一段难忘而美好的时光。

此外，还要感谢朋友以及同学们在论文编写中带给的大力支持和帮忙，给我带来极大的启发。也要感谢参考文献中的作者们，透过他们的研究文章，使我对研究课题有了很好的出发点。

最后，谢谢论文评阅老师们的辛苦工作。衷心感谢我的家人、朋友，以及同学们，在他们的鼓励和支持下我才得以顺利完成此论文。