

基于深度学习的实用 HDPC 码译码方法研究^①



郭军军, 白硕栋, 王 乐

(西安工业大学 计算机科学与工程学院, 西安 710021)

通讯作者: 郭军军, E-mail: guojunjun@xatu.edu.cn

摘 要: 针对现有高密度校验码量化译码性能问题, 本文提出了一种基于深度学习的量化最小和译码算法-QMSND. 借助深度神经网络, 通过对神经最小和译码信道输入向量和每轮迭代过程中节点更新信息进行非均匀间隔量化, 动态调整 Tanner 图边的权重参数, 改善消息传播效能. 计算机仿真实验结果表明, 本文提出的方法在对 BCH 码进行译码时仅需要 8 比特表示信息即可接近未经量化的浮点译码性能. 因此, 所提出的 QMSND 译码方法便于硬件实现, 具有一定的实用性.

关键词: 深度学习; 信道译码; 加性高斯白噪声; BCH 码

引用格式: 郭军军, 白硕栋, 王乐. 基于深度学习的实用 HDPC 码译码方法研究. 计算机系统应用, 2019, 28(4): 247-251. <http://www.c-s-a.org.cn/1003-3254/6786.html>

Research on Practical Method for Decoding to HDPC Codes via Deep Learning

GUO Jun-Jun, BAI Shuo-Dong, WANG Le

(School of Computer Science and Engineering, Xi'an Technological University, Xi'an 710021, China)

Abstract: Aiming at the problem of quantitative decoding performance of high-density parity-check codes, this study proposes a quantitative min-sum decoding approach based on deep learning, referenced as QMSND. In order to improve the decoding performance and efficiency, the proposed decoder can adjust the weight parameters of a Tanner graph dynamically, and quantize the input vector and the message of nodes at every iteration in nonuniform fashion via deep neural network. Computer simulation results show that the performance of proposed QMSND decoding with 8-bits presentation is very close to that of the float neural min-sum decoding without quantization. Therefore, the proposed decoding approach is easy to implement by hardware and has some practicability.

Key words: deep learning; channel decoding; AWGN; BCH codes

深度学习是机器学习的重要分支之一, 近年来由于其突出的性能而备受工业界和学术界的广泛关注^[1]. 深度学习的基本原理是模仿人脑中多层神经节点堆叠在一起构成的深度神经网络完成认知学习. 基于数据驱动, 深度学习技术能够自动抽取特征进行由浅入深学习, 试图求解难以用传统机器学习方法解决的复杂问题, 例如游戏竞赛^[2]、图像分类^[3]、语音识别^[4]和机

器翻译^[5]等.

最近, 国内外专家学者纷纷将深度学习引入到信道译码领域. 借助深度神经网络, 一些中短长度的高密度校验码 (High-Density Parity-Check, HDPC) 译码性能大幅提高, 且译码复杂度相对较低^[6-11]. 鉴于构造译码训练集的困难性, Eliya 等利用译码概率的独立性特点, 通过训练 HDPC 码对应 Tanner 图中消息交换边的权

① 基金项目: 陕西省重点研发计划项目 (2018GY-023); 西安工业大学校长基金 (XAGDXJJ16016)

Foundation item: Key Research and Development Program of Shaanxi Province (2018GY-023); President Fund of Xi'an Technological University (XAGDXJJ16016)

收稿时间: 2018-08-26; 修改时间: 2018-09-20; 采用时间: 2018-09-29; csa 在线出版时间: 2019-03-28

重,分别提出了前馈神经网络译码算法和循环神经网络译码算法^[9,10].中科大梁飞等提出了一种基于迭代的置信传播与卷积神经网络的组合译码方法,该译码方法便于并行实现,可广泛用于各类线性分组码^[11].但是,目前所提出的深度神经网络译码器由于时间和空间开销较大,从而制约了在一些诸如空间受限、低延时、低功耗的物联网和存储芯片电路系统等领域的应用.

针对现有译码方法的不足,受到文献^[12]启发,提出了针对加性高斯白噪声信道下 HDPC 码的基于深度神经网络的量化最小和译码算法—QMSND.仿真实验结果表明,所提出的方法仅需要 256 个量化阶(8 比特表示)即可达到与现有浮点型神经网络译码器相同的译码性能,因此便于硬件实现,具有一定的实用价值.

1 基于深度神经网络的最小和译码

1.1 深度神经网络模型

一个神经网络是由多个相连的神经元组成的.每个神经元就是一个最小的计算单元,通过累加其输入权重和可选的偏置量并利用非线性激活函数将结果直接输出或传递给其它神经元.这些神经元分层排列构成了输入层、隐层和输出层.如果神经网络包括多个隐层,如卷积层、池化层、循环层和激活层等,则称这类神经网络为深度神经网络(Deep Neural Network, DNN)^[1].

令 r 和 s 分别表示 DNN 的输入和输出, L 为层数或深度, ρ 为相关参数,则深度神经网络的计算模型可用式(1)来表示.

$$r = f(s, \rho) = f^{(L-1)}(f^{(L-2)}(\dots(f^{(0)}(s)))) \quad (1)$$

从式(1)中可知,每一层神经网络的功能都是将上一次产生的结果作为输入,函数经逐级复合,将问题化繁为简.使用大量测试数据,在损失函数的约束下,对目标函数进行最优化处理,最终训练出合适的深度神经网络模型.

1.2 HDPC 码神经最小和译码

设一个二元 HDPC 码的校验矩阵对应 Tanner 图模型为 $G = (VUC, E)$, 图 G 中变量节点集 $V = \{v_1, v_2, \dots, v_n\}$, 校验节点集 $C = \{c_1, c_2, \dots, c_r\}$, n 和 r 分别表示校验矩阵的列数(码长)和行数, E 是变量节点与校验节点之间相连的边集, $E = V \times C$. 如果校验矩阵的第 i 列第 j 行元素是非零的,则 Tanner 图 G 的节点 v_i 与 c_j 之间存在一条

边 $v_i, c_j \in E$. 在不产生歧义的情况下,本文后面表示变量节点和校验节点时省略其下标值.在加性高斯白噪声信道下,变量节点 v 的对数似然比 LLR 的计算见式(2).

$$l_v = \log \frac{\Pr(c_v = 1|y_v)}{\Pr(c_v = 0|y_v)} = \frac{2y_v}{\sigma^2} \quad (2)$$

式(2)中 y_v 为码字位 c_v 对应的信道接收值, σ 信道噪声方差.在 HDPC 码最小和译码算法中,消息沿着变量节点和校验节点之间的边进行迭代传递.令 $m_{c \rightarrow v}^{(l)}$ 表示在第 l 次迭代中由变量节点 v 传递到校验节点 c 的消息,同理可定义 $m_{c \rightarrow v}^{(l)}$ 为第 l 次迭代中由校验节点 c 传递到变量节点 v 的消息.在第 0 次迭代中, $m_{v \rightarrow c}^{(0)}$ 为变量节点 v 在其信道观察值对应的消息,与校验节点 c 无关,用 m_v 来表示.当 $l > 0$ 时,变量节点的消息更新规则如下所示:

$$m_{v \rightarrow c}^{(l)} = m_v + \sum_{c' \in N(v) \setminus \{c\}} m_{c' \rightarrow v}^{(l)} \quad (3)$$

校验节点的消息更新是由上一次进入校验节点消息的函数决定的.

$$m_{c \rightarrow v}^{(l)} = \left(\prod_{v' \in N(c) \setminus \{v\}} \text{sign}(m_{v' \rightarrow c}^{(l-1)}) \right) \min_{v' \in N(c) \setminus \{v\}} |m_{v' \rightarrow c}^{(l-1)}| \quad (4)$$

在式(3)和(4)中, $N(v)$ 是 Tanner 图中变量节点 v 的邻居校验节点集合, $N(c)$ 是校验节点 c 的邻居变量节点集合.若 $x > 0$,符号函数 $\text{sign}(x) = 1$, 否则 $\text{sign}(x) = -1$.

若采用基于深度学习的最小和译码时,校验节点的消息更新如式(5)所示.

$$m_{c \rightarrow v}^{(l)} = \text{ReLU} \left(\min_{v' \in N(c) \setminus \{v\}} \beta_{v' \rightarrow c}^{(l)} |m_{v' \rightarrow c}^{(l-1)}| \right) \prod_{v' \in N(c) \setminus \{v\}} \text{sign}(m_{v' \rightarrow c}^{(l-1)}) \quad (5)$$

式(5)中,激活函数 $\text{ReLU}(x) = \max(x, 0)$, 调优参数 $\beta_{v' \rightarrow c}^{(l)}$ 是通过小批次随机梯度下降法学习得到的非负实数.

2 量化神经最小和译码

2.1 量化方法

为了便于硬件电路实现,本文提出了一种基于多阶量化的 HDPC 码最小和译码算法.由于采用了事先训练好的深度神经网络进行一次性译码,因此译码速度大幅提高.信道噪声和译码消息经非均匀量化器过滤,采用有限比特表示量化阶,可以有效节约存储空间.量化激活函数 $\text{QReLU}(x) = \max(Q(x), 0)$, 其中量化函数 $Q(x)$ 见式(6).

$$Q(x) = \begin{cases} L_i, & T_i \leq x < T_{i+1} \\ 0, & \text{其它情况} \end{cases} \quad (6)$$

式(6)中, 门限集合 $T = \{T_1, T_2, \dots, T_M\}$, 量化阶 L_i 是正整数, 量化阶的数量为 $M+1$. 量化阶的选择会影响译码的性能, 且与门限之间存在着非线性关系, 因此需要针对不同的码通过学习来确定量化阶及其门限值的大小.

2.2 译码框架

假设码长为 n 的二元高密度 BCH 码 $C \in \{0, 1\}^n$, 一个码字 $\mathbf{c} \in C$ 经无记忆信道传输后, 其对应的信道接收向量 $\mathbf{y} = (y_1, y_2, \dots, y_n)$, 译码器输出的估计码字向量 $\hat{\mathbf{c}} = (\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n)$, 所提出的基于深度学习的量化最小和译码 QMSND 算法描述如下:

(1) 变量节点更新

通过式(3)计算由校验节点(除本节点外)传入到变量节点的消息. 当 $l=0$ 时, $m_{v \rightarrow c}^{(0)} = m_v$.

(2) 校验节点更新

由变量节点(除本节点外)进入到变量节点的消息由式(7)计算.

$$m_{c \rightarrow v}^{(l)} = \text{QReLU} \left(\min_{v' \in N(c) \setminus \{v\}} \mathcal{Q} \left(\beta_{v' \rightarrow c}^{(l)} \left| m_{v' \rightarrow c}^{(l-1)} \right| \right) \right) \prod_{v' \in N(c) \setminus \{v\}} \text{sign}(m_{v' \rightarrow c}^{(l-1)}) \quad (7)$$

由于每次计算参数 $\beta_{v' \rightarrow c}^{(l)}$ 值不尽相同, 一般采用离线方式事先进行基于小批量随机梯度下降的学习训练.

(3) 译码判决

$$m_{v \rightarrow c}^{(l)} = m_v + \sum_{c \in N(v)} m_{c \rightarrow v}^{(l)} \quad (8)$$

如果 $m_{v \rightarrow c}^{(l)} > 0$, 则对应码位的判决值 $\hat{c}_v = 0$; 否则 $\hat{c}_v = 1$.

(4) 停止条件

若 $\hat{\mathbf{c}}^T \mathbf{H} = \mathbf{0}$ 或超过最大迭代次数宣告译码结束, 否则跳转到(1)继续执行.

QMSND 译码算法的最大迭代次数与码长、码的结构以及信道噪声大小有关. 一般而言, 码字结构越复杂、码长越长或信道质量越差, 译码最大迭代次数的设值也越大. 通常, 为了平衡译码性能和速度, 对于中短长度的 HDPC 码, 如 BCH 或 Polar 码的最大迭代次数设置在 5~30 之间.

3 实验结果及分析

3.1 实验环境

本文实验所采用的神经网络是构建在基于

TensorFlow 的 Keras 框架之上. 使用 Python 语言编写仿真程序, 操作系统平台为 Windows 10 专业版, 采用蒙特卡洛法进行试验并比较高密度 BCH 码在不同信噪比下的性能. 为了加快训练速度, 应用了英伟达公司的 GeForce GTX 1080 Ti 系列图形加速 GPU.

3.2 网络模型训练

本文所采用的神经网络译码器是对 Tanner 图边的权重和量化阶进行训练. 交叉熵函数定义为:

$$L(o, y) = -\frac{1}{n} \sum_{v=1}^n y_v \log(o_v) + (1 - y_v) \log(1 - o_v) \quad (9)$$

式(9)中, o_v 是深度神经网络的非线性输出 sigmoid 函数, 定义为 $o_v(x) = \frac{1}{1 + e^{-x}}$.

所构建的量化最小和神经译码器 QMSND 是基于深度循环神经网络 (Recurrent Neural Network, RNN) 的, 包括 1 个输入层、4 个循环层和 1 个输出层. 其中输入层又细化为调制层、噪声层和 LLR 层, 每个循环层对应一次完整的量化最小和译码迭代过程, 循环层内进一步细化为变量节点计算层、校验节点计算层 (QReLU 量化激活层). 变量节点计算层完成式(3)的计算任务, 校验节点计算层对进入节点的信息进行量化激活, 上次校验节点的计算结果作为下一次循环变量节点的输入, 循环层的训练目标是最小化交叉熵函数式(9). 循环层最后输出经式(7)计算后得到译码的最终结果.

训练数据来自于 BCH 码的全零码字在加性高斯白噪声信道的随机构造的数据集, 信噪比范围从 1 dB 到 8 dB. 在不同信噪比下, 采用小批量随机梯度下降方式, 每次训练批次提交 20 个码字, 共使用 1600 个训练样例. 量化阶采用非均匀间隔, 间隔大小基本服从正态分布, 通过学习确定不同间隔大小. 由于循环神经网络前后层之间的存在相关性约束, 每层学习时的参数都会发生变化, 从而使得训练速度减慢. 因此, 在训练过程中学习率 r 设置较小. 本次训练中, 我们将 r 设置为 0.001.

3.3 实验结果分析

为了验证本文提出方法的正确性和有效性, 选取了三个典型的 BCH 码作为测试码, 分别是 BCH(63, 36), BCH(63, 45) 和 BCH(127, 99). 本次实验中量化阶数选择了 128 和 256 两种, 不同 BCH 码的译码性能比较分别见图 1、图 2 和图 3.

图1是采用传统和积译码SPA、文献[12]提出的DNNMS和本文提出的7比特量化QMSND-128以及8比特量化QMSND-256译码算法对BCH(63, 36)码的误比特率BER进行数值仿真结果比较。从图1中可以看出,当信噪比大于3 dB, QMSND译码性能逐步改善, QMSND-256译码器性能逼近未经量化处理的DNNMS译码器;在信噪比为7 dB时, SPA误码率为 9.7×10^{-4} , DNNMS误码率是 1.2×10^{-4} , QMSND-128和QMSND-256译码器的误码率分别是 6.8×10^{-4} 和 2×10^{-4} 。

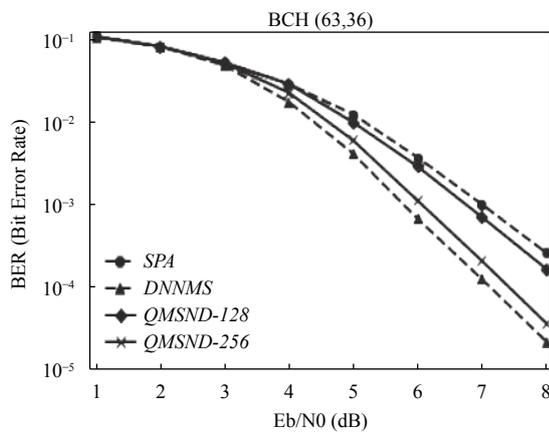


图1 BCH(63, 36)性能仿真图

图2是对BCH(63, 45)码进行性能数值比较。在误码率为 2.5×10^{-4} 时, SPA译码器信噪比是8 dB, QMSND-256译码器的信噪比为6.45 dB, 本文提出的8比特量化译码性能提高1.55 dB, 与DNNMS性能差距仅0.18 dB。

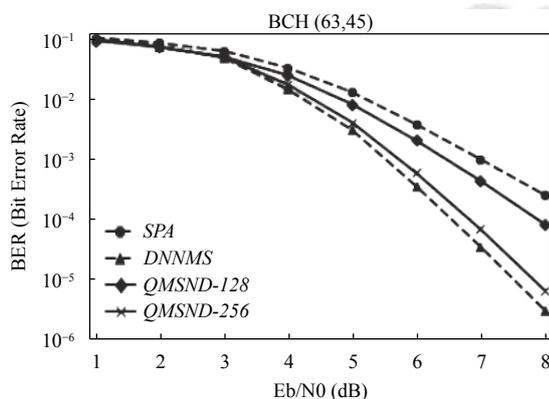


图2 BCH(63, 45)性能仿真图

图3是对BCH(127, 99)长码进行仿真试验得到的译码性能比较,其仿真结果与前两个码类似。但QMSND

量化译码性能改进没有前两个码那样显著,这是因为随着码长增加,样本训练复杂度急剧上升导致量化精度降低。

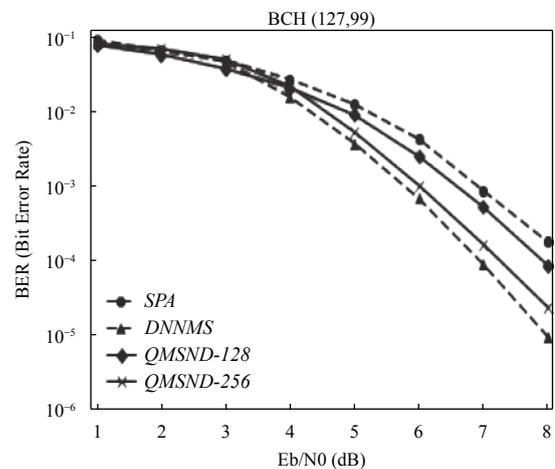


图3 BCH(127, 99)性能仿真图

从以上三个译码性能比较图中可以看出,在低信噪比区域,传统的和积译码SPA、基于深度神经网络的最小和译码DNNMS以及本文提出的128和256阶量化神经最小和译码QMSND的译码性能差距很小,但在中高信噪比区域,SPA译码性能明显较差,DNNMS性能至少比SPA高1 dB,而QMSND-128虽然没有DNNMS的译码性能优异,但总体而言比SPA要好,特别是QMSND-256译码性能与DNNMS十分接近。

4 结论

为了提高加性高斯白噪声信道下高密度BCH码译码的高效性和实用性,本文提出了基于深度神经网络的量化最小和译码方法—QMSND。仿真实验结果表明,所提出的方法仅需要256个量化阶(8比特)即可达到与现有浮点型神经网络译码器相近的译码性能。所提出的方法译码性能好,便于硬件实现,具有一定的应用前景。今后将对进一步优化译码算法,研究单轮迭代时能够自适应量化阶区间的学习方法,提高译码性能和硬件实现效率。

参考文献

- 1 孙志军, 薛磊, 许阳明, 等. 深度学习研究综述. 计算机应用研究, 2012, 29(8): 2806-2810. [doi: 10.3969/j.issn.1001-3695.2012.08.002]

- 2 郭潇道, 李程, 梅俏竹. 深度学习在游戏中的应用. 自动化学报, 2016, 42(5): 676–684.
- 3 胡二雷, 冯瑞. 基于深度学习的图像检索系统. 计算机系统应用, 2017, 26(3): 8–19.
- 4 侯一民, 周慧琼, 王政一. 深度学习在语音识别中的研究进展综述. 计算机应用研究, 2017, 34(8): 2241–2246. [doi: [10.3969/j.issn.1001-3695.2017.08.001](https://doi.org/10.3969/j.issn.1001-3695.2017.08.001)]
- 5 Luong MT, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. eprint arXiv:1508.04025, 2015.
- 6 Völker M, Hammer J, Schirrmeyer RT, *et al.* Intracranial error detection via deep learning. eprint arXiv:1805.01667, 2018.
- 7 Ye H, Li GY. Initial results on deep learning for joint channel equalization and decoding. 2017 IEEE 86th Vehicular Technology Conference. Toronto, ON, Canada, 2018. 1–5.
- 8 Cammerer S, Gruber T, Hoydis J, *et al.* Scaling deep learning-based decoding of polar codes via partitioning. eprint arXiv:1702.06901, 2017.
- 9 Nachmani E, Beery Y, Burshtein D. Learning to decode linear codes using deep learning. eprint arXiv:1607.04793, 2016: 341–346.
- 10 Nachmani E, Marciano E, Lugosch L, *et al.* Deep learning methods for improved decoding of linear codes. IEEE Journal of Selected Topics in Signal Processing, 2018, 12(1): 119–131. [doi: [10.1109/JSTSP.2017.2788405](https://doi.org/10.1109/JSTSP.2017.2788405)]
- 11 Liang F, Shen C, Wu F. An iterative BP-CNN architecture for channel decoding. IEEE Journal of Selected Topics in Signal Processing, 2018, 12(1): 144–159. [doi: [10.1109/JSTSP.2018.2794062](https://doi.org/10.1109/JSTSP.2018.2794062)]
- 12 Lugosch L, Gross WJ. Neural offset min-sum decoding. eprint arXiv:1701.05931, 2017: 1361–1365.