

LDPC 码陷阱集知识图谱构建方法研究

郭军军 王 乐 荆 心 李桂萍

(西安工业大学 计算机学院, 陕西 西安 710021)

摘 要: 陷阱集是影响 LDPC 码在错误平台区域译码性能的直接原因。笔者提出了一种基于 Tanner 子图构建陷阱集知识图谱的方法, 通过对基础陷阱集迭代地耗尽搜索所有可能扩展路径, 生成新的陷阱集, 并建立陷阱集及译码特征之间的关系, 最后通过分析比较, 证实了陷阱集知识图谱的有效性和应用价值。

关键词: LDPC 码; 错误平台; 陷阱集; 知识图谱

中图分类号: TN911.22 **文献标识码:** A **文章编号:** 1003-9767 (2020) 04-052-03

Research on the Construction of Knowledge Map of LDPC Code Trap Set

Guo Junjun, Wang Le, Jing Xin, Li Guiping

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

Abstract: The trap set is the direct reason that affects the decoding performance of LDPC code in the wrong platform area. In this paper, a method of constructing knowledge map of trap set based on Tanner subgraph is proposed. By iteratively searching all possible extension paths of the basic trap set, a new trap set is generated, and the relationship between trap set and decoding features is established. Finally, through the analysis and comparison, the effectiveness and application value of the trap set knowledge map are confirmed.

Key words: LDPC code; error platform; trap set; knowledge map

0 引言

低密度校验 (Low-Density Parity-Check, LDPC) 码是 Gallager 在 1962 提出的一种逼近香农容量限的信道编码技术^[1]。在加性高斯白噪声信道下, 采用消息传递迭代译码算法时, 码长为 10^7 、码率为 0.5 的非规则 LDPC 码译码门限距香农容量限仅差 0.004 5 dB, 是目前发现的最实用好码之一。与其他现代编码体系相比: 首先, LDPC 码译码算法复杂度低, 特别是采用比特翻转译码 (BF) 算法和有限字母表迭代译码 (FAID) 算法硬件实现非常容易, 便于并行计算, 因而译码速度快延时小; 其次, 编码构造简单, 采用稀疏的二部图 Tanner 表示 LDPC 码字结构, 译码过程易于理解, 便于改进^[2]。LDPC 码的诸多优越性使得在 5G 通信、数字存储系统和深空通信等领域具有广阔的应用前景。

然而, 中短长度的 LDPC 码在高信噪比区域存在错误平台现象, 从而造成译码性能急剧下降^[3-4]。在加性高斯白噪声信道和二元对称信道下, 陷阱集 (一些特殊的 Tanner 子图结构) 是导致 LDPC 码译码错误平台问题的主要原因。目

前, 国内外关于 LDPC 码的陷阱集研究主要集中在以下两方面。第一, 陷阱集搜索。由于 LDPC 码的陷阱集是由短环及其相邻的路径组成的, Karimi 等人提出了一种针对非规则 LDPC 码的陷阱集暴力搜索方法^[5]。该方法的基本思路是首先从 Tanner 图中找出一个度数较小的变量节点作为基础, 然后通过单边、路径或其他特殊子图扩展来搜索非规则 LDPC 码消息传递译码的陷阱集。此外, 还有基于随机抽样等方法搜索陷阱集, 这类方法不但适用于规则 LDPC 码, 也同样适用于非规则 LDPC 码。第二, 陷阱集应用。由于陷阱集不完全依赖于具体的 LDPC 码字, 因此, 利用陷阱集知识可以对 LDPC 码进行优化设计和译码性能评估。Vasic 等人提出了陷阱集本体 (Trapping Set Ontology) 概念, 并利用陷阱集本体进行枚举各类衍生陷阱集^[6]。陷阱集本体刻画了陷阱集以及陷阱集之间的拓扑关系, 它是设计高性能 LDPC 码和评估其译码性能的重要手段。

知识图谱已广泛应用于金融投资、生物医药、AI 教育、社交等领域, 但在通信领域特别是信道编译码方面还未见应

基金项目: 陕西省重点研发计划项目 (项目编号: 2018GY-023)。

作者简介: 郭军军 (1977—), 男, 陕西横山人, 博士研究生, 副教授。研究方向: 信息论与信道编码、机器学习研究。

王乐 (1992—), 男, 陕西安塞人, 硕士研究生。研究方向: 信息论与信道编码、信息安全领域研究。

用。本文提出的陷阱集知识图谱是 LDPC 码智能编译码的关键技术,通过构建陷阱集知识图谱,进一步揭示陷阱集之间的结构特征、陷阱集与 LDPC 码字关系,为设计高效编译码方案奠定坚实的基础。

针对陷阱集本体数据库构建效率低、复杂度高,不能全面反映陷阱集之间深层关系和潜在价值的不足,本文提出了基于子图扩展的陷阱集知识图谱构建算法,并将该方法产生的知识图谱应用到译码算法的改进和性能评估等方面,仿真实验结果证明了所提出方法的有效性。

1 预备知识

令 $G=(V, C, E)$ 是 LDPC 码 \square 对应的 Tanner 图。 G 是一种特殊的无向二部图,其顶点是由变量节点和校验节点组成的,变量节点集 $V=\{v_1, v_2, \dots, v_n\}$, 校验节点集 $C=\{c_1, c_2, \dots, c_m\}$, n 和 m 分别表示变量节点和校验节点数, E 是校验节点与变量节点之间相连的边集, $E=\{<v, c>: v \in V, c \in C\}$ 。变量节点 v 的邻居校验节点集 $N(v)=\{c: c \in C, <v, c> \in E\}$, 相应地, 校验节点 c 的邻居变量集 $N(c)=\{v: v \in V, <v, c> \in E\}$ 。由于 LDPC 码具有稀疏性特点, 故图 G 中边数较少。若码 \square 的 Tanner 图中所有变量节点和校验节点的度分布相同, 则称该码为规则 LDPC 码, 否则称之为非规则 LDPC 码。

定义 1: 陷阱集 (Trapping Set) 是 LDPC 码经多轮迭代译码后不能成功译出的 Tanner 图中对应的变量节点集合。一个 $\tau(t, s)$ 陷阱集的 Tanner 子图是由 t 个变量节点以及 s 个度为奇数的校验节点组成, 其中 t 表示陷阱集的大小。

图 1 是一个典型的 $\tau(4,2)$ 陷阱集, 图中黑色圆圈表示陷阱集中变量节点, 空心方框表示陷阱集导出子图中与变量节点相连的度为偶数的校验节点, 而黑色实心方框表示与子图中变量节点相连的度为奇数的校验节点。

一般情况下, 陷阱集与信道质量和具体码字结构有关, 陷阱集大小呈固定、震荡和随机分布模式, 其中绝大多数为固定模式, 陷阱集越小, 其危害性越大, 错误平台问题越突出。

定义 2: 关键数 (Critical Number) 被定义为陷阱集中变量节点对应的初始信道接收信息有误且经译码器最终不能成功译出的变量节点最小数。这些变量节点也称之为关键节点。

例如, 在图 1 所示的一个 $(4,2)$ 陷阱集中, 变量节点集 $\{v_1, v_2, v_3, v_4\}$ 对应的信道接收信息中 v_1, v_2, v_3 有误, v_4 对应接收信息正确, 经多轮迭代译码后, 而将 v_1, v_2, v_3 和 v_4 全部译错, 若对应接收信息少于 3 个有误时, 译码器均可成功译码, 则该陷阱集的关键数为 3。

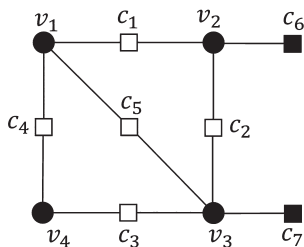


图 1 $(4,2)$ 陷阱集的 Tanner 导出子图

定义 3: 基础陷阱集 (Fundamental Trapping Set) 是指陷阱集导出子图中变量节点和校验节点数目很小的陷阱集。

基础陷阱集是一类最基本的小陷阱集, 其危害性是所有陷阱集类型中最大的。对于任意 (t,s) 基础陷阱集, 在中高信噪比下, 若 $t \leq 10$ 且 $s \leq 2$, 特别是 $s=0$ 或 $s=1$ 时, 由于陷阱集 Tanner 子图相对封闭, 内部消息交换形成环路, 外界强迭代消息难以进入, 容易造成持续的震荡或固定错误模式, 从而会形成错误平台, 进而会影响 LDPC 码译码质量。

2 陷阱集知识图谱构建方法

2.1 基本思想

陷阱集知识图谱揭示了陷阱集之间的层次结构关系, 反映了不同码字的结构特征、危害性、信道质量等方面的语义信息。建立陷阱集 Tanner 子图拓扑关系是陷阱集知识图谱构建的关键技术, 以高危基础陷阱集为出发点, 其通过不断增减节点或边形成较大规模的陷阱集, 再利用图的相似性比较方法, 确定不同陷阱集之间的层次结构关系。

图 2 是列重为 3 的规则 LDPC 码在加性高斯白噪声信道下采用消息迭代译码算法的四个基础陷阱集的 Tanner 导出子图。它们在拓扑结构上具有强关联关系, 其中图 2(a) 是 $(4,0)$ 陷阱集, 通过增加 c_6 和 c_7 两个校验节点后分别与 v_2 和 v_4 相连, 构成了如图 2(b) 所示的 $(4,2)$ 陷阱集, 增加校验节点 c_8 与 v_3 相连并阻断 c_5 与 v_3 之间的边后演化为图 2(c) 所示的 $(4,4)$ 陷阱集, 继续增加变量节点 v_5 与 c_6 和 c_8 相连, 再与新增的 c_9 校验节点连接构成了图 2(d) 所示的 $(5,3)$ 陷阱集。依照上述方式可以不断延展构建更庞大的陷阱集拓扑结构。

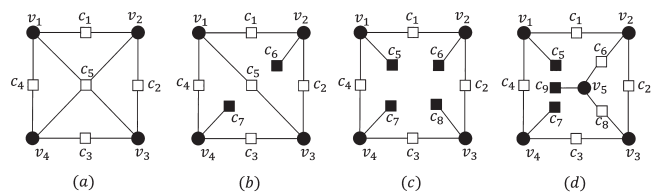


图 2 基础陷阱集拓扑结构

2.2 构建算法

结合上述陷阱集拓扑结构关系演化过程, 考虑到陷阱集危害程度、关键数、码字结构、译码算法和信道质量等因素, 陷阱集知识图谱构建算法描述如下。

步骤 1: 初始化陷阱集知识图谱集合, $G = \{\varphi\}$ 。

步骤 2: 从基础陷阱集导出子图集 T 中取出一个陷阱集导出子图 s , 并将此作为图 G 的节点。

步骤 3: 通过对子图 s 中增加变量节点或校验节点和扩展连接边, 形成新的陷阱集导出子图 s' , 利用图的相似性检测陷阱集导出子图同构, 并确定同构数。

步骤 4: 建立 s 与 s' 之间的关系, 生成子图模型 g , 并关联对应的语义特征节点 $\ell \in L$, 扩展图 G 。

步骤5: 继续对 s' 施加步骤3到步骤4的操作, 直至陷阱集大小达到最大阈值 M 。

步骤6: 直到枚举完集合 T 中所有子图结构后结束算法, 否则跳转到步骤2。

在上述构建算法的输入中, 陷阱集语义特征包括所关联陷阱集的码字结构特征, 如码长、码率、度分布、是否规则等, 译码算法 (BP、LP、FAID 等) 的选择、信道类型 (AWGN、BEC、BSC 等) 以及信噪比范围等信息。

2.3 图谱结果

采用2.1节提出的算法对列重为3的码长为512规则LDPC码在加性高斯白噪声信道下BP译码构建的部分陷阱集知识图谱如图3所示。

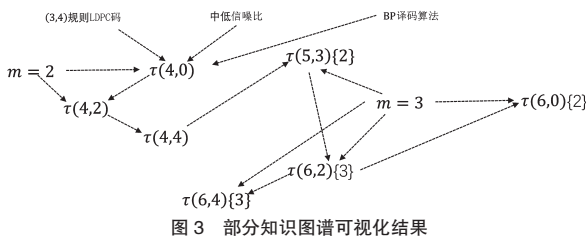


图3 部分知识图谱可视化结果

图3中陷阱集后面的花括号里数字表示同构数, 不同特征或陷阱集实体用图中的节点表示, 它们之间的边表示隶属、父子、扩展、继承等关系。

3 陷阱集知识图谱应用

陷阱集知识图谱可以用在基于语义的陷阱集图模型检索、性能评估与推理、LDPC 码字构造等方面, 典型应用场景如下。

3.1 应用1: 译码性能评估

在陷阱集知识图谱中, 根据译码算法、信道质量、关键数和危害性等特征对陷阱集进行分类。利用陷阱集的分类结果可以预测译码性能, 假定一个码长为 n 的 LDPC 码包含某类陷阱集, 且 T 中有 k 个陷阱集, 其关键数为 m , 最大陷阱集

的大小为 M , 危害程度系数为 s , 误帧率 $\Pr\{\mathcal{R}\} = \sum_{r=m}^M \frac{sk \binom{r}{m}}{\binom{n}{m}}$ 。

在二元对称信道下, (155.64) Tanner 码^[7]和 (900, 450) QC 码仿真与预测比较见表1。

表1 译码性能仿真与预测结果比较

转移概率		0.025	0.02	0.015	0.01	0.005
Tanner 码	仿真	8.1×10^{-2}	1.2×10^{-2}	5.3×10^{-3}	9.2×10^{-4}	9.1×10^{-5}
	预测	9.0×10^{-2}	1.1×10^{-2}	6.0×10^{-3}	9.2×10^{-4}	8.8×10^{-5}
QC 码	仿真	6.5×10^{-3}	7.3×10^{-3}	9.0×10^{-4}	5.3×10^{-4}	7.1×10^{-5}
	预测	7.0×10^{-3}	7.4×10^{-3}	9.0×10^{-4}	5.2×10^{-4}	7.1×10^{-5}

从表1可以看出, 利用陷阱集知识图谱预测译码性能与

实际仿真结果差距很小, 特别是码长越长越接近蒙特卡洛仿真结果。因此, 一旦拥有了陷阱集图谱可以方便地评估这类 LDPC 码字的译码性能, 从而省去大量的仿真测试时间。

3.2 应用2: 码字构造

对于一个特定的译码算法和 LDPC 码字, 通过改进码字结构, 有效地避免高危陷阱集, 可以显著改善错误平台问题。就二元对称信道而言, 错误平台性能主要受到关键数的影响, 因此避免关键数值小的陷阱集可以保证译码器良好的纠错性能。

利用陷阱集知识图谱中父子关系可以简化 LDPC 码字的构造算法。假定欲构造围长为8, 避免出现 (5,3) 或更小陷阱集的规则码, 译码算法为 Gallager-A, 从图谱中可以看出所有 (6,2) (6,0) 和 (7,1) 陷阱集都是 (5,3) 陷阱集的后代, 若避免了 (5,3) 陷阱集就自然避免了上述这些危险陷阱集, 因此使得基于边交换改进码字构造方法更加简洁高效。

4 结语

陷阱集知识图谱是评价和设计高效 LDPC 码的基础, 本文提出了一种基于 Tanner 子图扩展的方法构建陷阱集图谱数据库, 并结合 LDPC 码的结构特征、信道质量和译码算法, 使得陷阱集拓扑关系具体更加丰富的语义信息。今后将进一步改进陷阱集知识图谱的构建方法, 深入挖掘陷阱集知识图谱的实用价值。

参考文献

- [1] Gallager R. Low-density parity-check codes[J]. Journal of Circuits & Systems, 2008, 8(1): 3-26.
- [2] 郭军军, 白硕栋, 慕建君, 等. 一种有效的 LDPC 码伪码字搜索算法[J]. 西安电子科技大学学报: 自然科学版, 2018, 45(6): 162-166.
- [3] Richardson T. Error floor of LDPC codes[C]//42th Annual Allerton Conference, 2003: 648.
- [4] Tao X, Li Y, Liu Y, et al. On the Construction of LDPC Codes Free of Small Trapping Sets by Controlling Cycles[J]. IEEE Communications Letters, 2018, 22(1): 9-12.
- [5] Karimi M, Banihashemi A H. Efficient Algorithm for Finding Dominant Trapping Sets of LDPC Codes[J]. Information Theory, 2012, 58(11): 6942-6958.
- [6] Vasic B, Chilappagari S K, Nguyen D V, et al. Trapping Set Ontology[C]//47th Annual Allerton Conference, 2009: 489.
- [7] Lu J, Moura J M F, Niesen U. A class of structured LDPC codes with large girth[C]//IEEE International Conference, 2004: 64.