

Error_Analysis

基于VerilogEval benchmark（共包括156个编程问题），我们分别收集了Qwen2.5-Coder-32B-Instruct, GPT-3.5-turbo, GPT-4-turbo（不同参数量，以及代码专用/通用）大模型的83、89、65个错误编程样本（设置temperature = 0.0，保证输出的稳定性），共计237个错误样本，研究LLM RTL生成错误的原因和类型。

如下分析了不同的错误样本（以问题编号 编码）对应的错误类型。

（需要注意的是，单一的错误代码样本可能存在多种不同类型的错误，我们将其同时划分为不同的类中）

1. Qwen-Coder-32B-Instruction模型

Type I: Insufficient knowledge of specialized RTL programming

- 数字处理逻辑错误：009, 141
- 向量处理逻辑错误：092（反了），148（位选错位，越界）
- 赋值类型不匹配：068, 079, 088, 100, 119, 138, 140, 144, 156

Type II: Misinterpretation of design specification

1. 缺乏知识：

- 时序知识：035, 037, 038, 040, 041, 046, 060, 067, 068, 073, 088, 096, 107, 109, 111, 120, 136, 137, 138, 139, 148（同步/异步逻辑），066（in前递信号记录），080（错误样本输出较于golden会慢一个周期，always块和assign块的区别）
- -固定的特殊设计编程知识：070（SOP和POS），078（DualEdge），082（LFSR32），086（LFSR5）
- 数字和向量处理知识：063, 084（most-significant-bit理解），071, 146（least significant），112（Case条件向量设置），115（符号位设置）
- 复杂场景下FSM处理知识：089, 096, 128, 129, 133, 137, 139, 140, 142（状态丢失或转移逻辑错误），143, 146, 149, 151, 152, 154, 155, 156
由知识缺乏导致的错误设计：45

2. 描述模糊：

028, 034, 053 (缺乏初始化), 045, 062, 074, 104 (设计功能描述模糊/错误), 079, 143, 150 (State信号的描述模糊)
由描述模糊导致的错误设计: 10

3. 多模态信息处理:

- 表格: 069, 121, 124, 134
 - KMAP: 050, 057, 093, 113, 116, 122, 125
 - Waveform: 098, 101, 102, 103, 117, 145, 147, 154
 - 状态转换图: 091, 099, 109, 135, 136, 138, 143, 150
- 由多模态数据导致的错误设计: 27

4. 过长上下文理解失误细节丢失问题:

093, 139, 140, 142, 150, 151, 152, 153, 155, 156
长上下文导致的错误设计: 10

2. GPT-3.5 Turbo模型

Type I: Insufficient knowledge of specialized RTL programming

- 不完备代码: 016 (缺失submodule), 018, 021 (缺失位选逻辑), 048 (代码功能不完整)
- 变量重复定义: 041, 047, 073, 121
- 向量处理逻辑错误: 068, 089, 109 (位选错位), 105 (反了)
- 数字处理逻辑错误: 071
- 赋值类型不匹配: 079, 100, 107, 111, 139, 140, 143, 146
- 阻塞非阻塞赋值混用: 108
- 错误使用Generate语句: 092
- 使用未定义变量: 093
- integer变量使用错误: 144

Type II: Misinterpretation of design specification

1. 缺乏知识:

- 时序知识: 037, 060, 066, 073, 088, 111, 120, 121, 128, 133, 136, 137, 138, 139, 140, 146, 148, 149, 151, 154, 156 (同步/异步逻辑) 054, 066, 139 (prev_in记录), 049 (posedge/negedge逻辑)

- 固定的基础编程知识：070（SOP与POS知识），078（DualEdge），082（LFSR32），086（LFSR5），
- 数字和向量处理知识：033（有符号数加法溢出），071，134（case条件设置），105（右移错误），112（最低位数字理解），115（符号位处理），141（时间处理），146（数字记录）
- 复杂场景下FSM处理知识：089，091，095，096，127，128，129，133，137，139，140，142，146，148，149，151，152，154，155，156
由知识缺乏导致的错误设计：43

2. 描述模糊：

028，034，053（缺乏初始化），045，062，074，104（设计功能描述模糊/错误），

028（ena信号未描述清楚），079，143，150（state信号未描述清楚）

由描述模糊导致的错误设计：11

3. 多模态信息处理：

- 表格：069，124，134
- KMAP：050，057，093，113，116，122，125
- Waveform：098，101，102，103，117，126，130，131，145，147，150，154
- 状态转换图：088，091，099，135，138，143，150
由多模态理解导致的错误设计：29

4. 过长上下文理解失误细节丢失问题：

092，093，094，142，150，151，152，153，154，155，156

长上下文导致的错误设计：11

2. GPT-4 Turbo模型

Type I: Insufficient knowledge of specialized RTL programming

- 不完备代码：016（缺失submodule），037，048（缺少end），
- 赋值类型不匹配：068
- 错误使用Generate语句：108
- 使用未定义变量：060
- integer变量使用错误：144，153

Type II: Misinterpretation of design specification

1. 缺乏知识:

- 时序知识: 120, 133 (同步/异步逻辑), 066 (prev_in记录)
 - 固定的基础编程知识: 070 (SOP与POS知识), 078 (DualEdge), 082 (LFSR32), 086 (LFSR5),
 - 数字和向量处理知识: 033 (加法溢出判断), 112 (case), 141 (时间处理)
 - 复杂场景下FSM处理知识: 089, 095, 128, 133, 137, 139, 140, 142, 146, 149, 151, 154, 155, 156
- 知识缺乏导致的错误设计: 23

2. 描述模糊:

028, 034, 053 (缺乏初始化), 045, 062, 074 (设计功能描述模糊/错误), 079, 150 (state信号未描述清楚)

描述模糊导致的错误设计: 8

3. 多模态信息处理:

- 表格: 134
 - KMAP: 050, 057, 093, 113, 116, 122, 125
 - Waveform: 098, 101, 102, 103, 117, 131, 145, 147, 154
 - 状态转换图: 088, 091, 099, 135, 143, 150
- 多模态信息理解导致的错误设计: 23

4. 过长上下文理解失误细节丢失问题:

093, 094, 124, 142, 150, 151, 153, 154, 155, 156

长上下文导致的错误设计: 10

我们希望这样的分类工作为日后LLM生成RTL的工作提供包括训练, 推理使用等方面的一些insight, 辅助LLM生成功能更完善, 质量更高的RTL代码。

Note: 需要注意的是, 由于分类是基于人工评判的, 同时我们仅在Qwen系列模型和GPT系列模型上进行评测, 可能存在由于主观判断导致不准确和不完善, 我们期待您的进一步建议