

SHANGHAI JIAO TONG UNIVERSITY

本科生毕业设计（论文）中期检查报告



论文题目： 基于数值型输入输出样例的程序合成器

学生姓名: 王绍宇

学生学号: 5080379072

专 业: 软件工程

指导教师: 赵建军

学院 (系): 软件学院

教务处制表

**填表说明**

1. 请每位学生根据学校及院（系）检查的要求认真进行自查，及时发现课题研究过程中存在的问题，分析原因，并提出解决思路和措施，明确下一阶段任务。
2. 每位学生应根据项目实施情况认真、实事求是填写。填写字体请用宋体小四号，并用A4纸打印，于左侧装订成册。
3. 毕业设计（论文）中期检查报告总字数应满足院（系）要求。
4. 该表填写完毕后，须请指导教师审核，并签署意见。
5. 《上海交通大学本科生毕业设计（论文）中期检查报告》将作为答辩资格审查的主要材料之一。
6. 本表格不够可自行扩页。

|  |
| --- |
| 课题进展情况：  根据用户的输入和输出样例进行函数的合成，需要找到用户提供的输入输出对儿之间的联系。  我现已设计了程序合成所需的简单语言，语言利用到相邻（一项或两项）输入之间的递推关系，包含了加减乘除的简单运算。  已完成了对如f(n) = 2n, f(n) = n \* n \* n – 1这类简单n次函数的合成。理论上讲，对任何给定的输入输出对都能找到一个满足条件的简单n次函数，但这并不一定是用户所需要的。  能够使用一种简单方法来评估合成的函数与给定输入输出的匹配的自信度。  详细内容请参考下一节。  总体来看，课题进展比较顺利。 |

|  |
| --- |
| 课题研究已取得的阶段性成果：  已对搜索语言进行了较明确的设计和规定   1. 运算操作：（表示为 \* ）定义为PLUS,MINUS,R\_MINUS,MUL,DIV,R\_DIV六种。   其中R\_MINUS, R\_DIV 是第二操作数减去/除以第一操作数（即逆向运算）   1. 基本变换：   [ f ] := f \* g \* h ...  即f(n)的计算结果与g(n)的计算结果进行第一个运算操作，其结果在于h(n)进行第二个运算操作。其中，g，h等的定义同f，见下文。   1. 函数的搜索空间定义：   f(n) := C0  或 f(n) := [ f(n-1) ], f(n0) := C0  或 f(n) := [ f(n-2) ], f(n0) := C0, f(n0 + 1) := C1  或 f(n) := [ [ f(n-2) ] \* [ f(n-1) ] ], f(n0) := C0, f(n0 + 1) := C1  第一行是f退化为常数函数，每一项都等于C0；第二行是只和前一项相关的递推定义，（需要已知某一项的特殊值定义）；第三行是只和再前一项相关的递推定义，（需要已知某连续两项的特殊值定义）；第四行是和前两项都相关的递推定义（需要已知某连续两项的特殊值定义）。   1. 另一方面，要计算比给定的n0小的输入的函数值，需要将函数的定义转换成如下形式：   f(n) := C0  或 f(n) := { f(n+1) }, f(n0) := C0  或 f(n) := { f(n+2) }, f(n0) := C0, f(n0 + 1) := C1  或 f(n) := { { f(n+2) } \*\* [ f(n+1) ] }, f(n0) := C0, f(n0 + 1) := C1  其中{ } 为对应[ ]的逆运算，\*\*为\*对应的逆运算。  例如：若[ f(n) ]表示f(n)×3 + 1，则{ f(n+1) }表示(f(n+1) - 1) / 3；  若f(n) = f(n-2) \* f(n-1)，则f(n-2) = f(n) \*\* f(n-1),将n+2替换n，即有f(n) = f(n+2) \*\* f(n+1)。  即利用[ ]和\*运算对应的逆运算{ }和\*\*，以及f(n+2)，f(n+1)来进行逆向推算。   1. 现阶段已可以对函数的表示进行toString()操作，即可以将最终合成出的“程序”用表达式表示出来。   toString()的实现中，注意到了每个子式是否需要加括号的问题。如果某一个运算符的两个操作数子式的优先级比这个运算符优先级低，那么子式需要添加括号。   1. 还对合成的函数可信度（confidence degree, CD）进行了粗略的定义（后面会调整）。现阶段即定义CD为除去推算函数所必须的输入输出对儿以外，每多给一个符合函数的输入输出对儿，CD就加1。 |

|  |
| --- |
| 存在的问题及解决思路：   1. 非连续输入的问题   用户给定的输入输出对儿中的输入可能是不连续的，可是，通过递推关系确定函数时，却需要有连续的数据点（输入输出对儿中的输入，需要是连续的整数）。  通过搜索来合成函数，确定良好的搜索策略，并且将可信度（CD）计算的定义合理化应该可以解决这个问题。   1. 搜索策略   目前的推导简单n次函数合成的过程其实没有分支，并不能称得上是搜索。而我最终的合成，是要在所有可能空间中去搜索，而盲目搜索的代价又过于庞大，很可能无法在可被接受的时间内给出合适的解，这样好的启发式搜索策略就至关重要。  近期我需要研究一下人工智能中关于搜索策略和机器学习的相关知识。初步设想给合成的各个片段和片段组合赋一个分值，使频繁使用的合成片段被优先地搜索，以此来提升搜索的效率。   1. 自定义函数的问题   根据最初的设想，在合成的过程中，用户可以提供一些没法用一般表达式表示的自定义函数，如质数数列构成的函数等。通过这样的指定，用户可以有目的性地指导合成。但自定义函数的输入方式成了最大的困难。  出于与“用户给定输入输出来合成程序”保持一致的想法，现在想到的解决方法是：用户应给定函数可能用到的输入对应的每一项输出。比如，最终需要合成的函数有20项，每一项是质数数列加1，那么用户需要将质数数列的前二十项依次给出。  由于这种功能是否有价值还有待商榷，并且实现起来存在诸多问题，可能会根据进度情况决定是否将此功能删除。   1. 噪音点的排除问题   如果允许用户给定的输入输出中存在噪音数据点，那么就要增加一种模式，让合成函数的过程中，屏蔽掉某些数据点（一到两个），以获得更高的可信度（CD）。 |

|  |
| --- |
| 下一阶段的工作计划和研究内容：  解决现在已有的问题，并研究人工智能中关于搜索策略和机器学习的相关知识，实现启发式搜索算法。  优化算法，提升程序合成的效率。  将输入输出的接口变得更加友好。  如果进度来得及，我会将合成器包装一个图形界面。 |

|  |
| --- |
| 指导教师意见：  指导教师签名：  年 月 日 |
| 学院（系）意见：    审查结果： □ 同 意 □ 不 同 意  学院（系）负责人签名：  年 月 日 |