

武汉大学计算机学院

2018-2019 学年度第一学期 2018 级

《认知过程的信息处理》期末考试试卷（A）答案（开卷）

试题 1 解答：

(a) $AD \perp BC$ 蕴含 $y_3(y_2 - y_1) + x_3(x_2 - x_1) = 0$

$|AB| = |AC|$ 蕴含 $x_2^2 + y_2^2 = x_1^2 + y_1^2$

$BD//BC$ 蕴含 $(y_3 - y_1)(x_2 - x_1) = (y_2 - y_1)(x_3 - x_1)$

为了证明 $|BD| = |CD|$ 等价于

$(y_3 - y_1)^2 + (x_3 - x_1)^2 = (y_3 - y_2)^2 + (x_3 - x_2)^2$

(b) 该问题有三个自由变元， x_1, y_1, x_2 （或者 $x_1, y_1, y_2; x_2, y_2, x_1; x_2, y_2, y_1$ ）

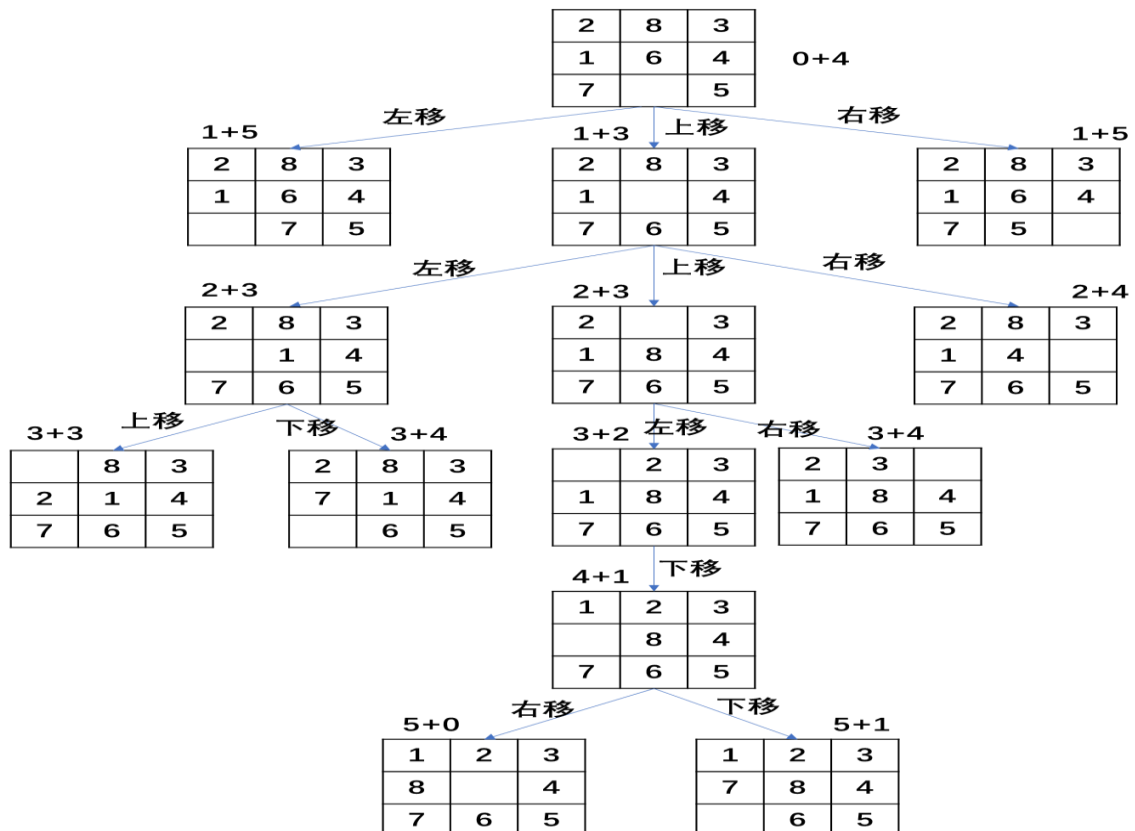
试题 2 解答：

(a) $F3 = \text{FRem}(F2, F1, x_2) = 1 + x_2^2 + x_2^2 - x_1x_2 - x_2^2 = 1 + x_2^2 - x_1x_2$

(b) $F4 = \text{FRem}(F3, F1, x_2) = (x_1 + x_2)x_1 + 1 + x_2^2 - x_1x_2 = 1 + 2x_1^2$

试题 3 解答：

假设 m, n 分别表示棋子移动的步数和位置不正确数字的个数，则代价函数定义为 $m+n$ 。
该问题的搜索树如下：



试题 4 解答:

“吴方法”的基本思想是将几何问题代数化,即用多项式来表达几何问题的条件以及结论,通过证明条件所组成的多项式交集的零点是结论对应多项式的子集来完成证明。主要步骤如下:

- 将条件和结论用代数多项式表达;
- 确定自由变元和约束变元,对约束变元排序,确定消元的次序;
- 将条件所对应的多项式三角化,设三角化后的多项式为 $F_1, F_2, \dots, F_k, \dots, F_n$, 保证 F_k 只包含前面 k 个约束变元;
- 将结论所对应的多项式从三角形底部到顶部依次消除最后的约束变元,如果最后所得到的剩余多项式为 0 则表明命题为真, 否则命题为假。

试题 5 解答:

该问题用遗传算法求解的具体步骤如下:

(1) 设定种群规模,编码染色体,产生初始种群。

将种群规模设定为 4; 用 4 位二进制数编码染色体; 取下列个体组成初始种群 S_1

$$s_1 = 6 (0110), \quad s_2 = 13 (1101)$$

$$s_3 = 4 (0100), \quad s_4 = 9 (1001)$$

(2) 定义适应度函数

$$f(x) = x^2$$

(3) 计算各代种群中的各个体的**适应度**, 并对其染色体进行遗传操作,直到适应度最高的个体,即 15 (1111) 出现为止。

- 比例选择:

$$P(x_i) = \frac{f(x_i)}{\sum_{j=1}^N f(x_j)}$$
$$q_i = \sum_{j=1}^i P(x_j)$$

选择-复制: 设从区间 $[0, 1]$ 中产生 4 个如下随机数:

$$r1 = 0.450126, \quad r2 = 0.110347$$

$$r3 = 0.572496, \quad r4 = 0.985031$$

染色体	适应度	选择概率	积累概率	选中次数
$s_1=0110$	36	0.12	0.12	1
$s_2=1101$	169	0.56	0.68	2
$s_3=0100$	16	0.05	0.73	0
$s_4=1001$	81	0.27	1.00	1

于是，经复制得群体：

$$s_1' = 13 (1101), \quad s_2' = 6 (0110)$$

$$s_3' = 13 (1101), \quad s_4' = 9 (1001)$$

● 交叉

设交叉率 $p_c=100\%$ ，即 s_1 中的全体染色体都参加交叉运算。

设 s_1' 与 s_2' 配对， s_3' 与 s_4' 配对，假定交换后两位基因，得新染色体：

$$s_1'' = 1110 (14), \quad s_2'' = 0101 (5)$$

$$s_3'' = 1101 (13), \quad s_4'' = 1001 (9)$$

● 变异

设变异率 $p_m=0.01$ 。这样，群体 s_1 中共有 $4 \times 4 \times 0.01 = 0.16$ 位基因可以变异，故本轮不进行变异。

于是，得到第二代种群 s_2 ：

$$s_1=1110 (14), \quad s_2=0101 (5)$$

$$s_3=1101 (13), \quad s_4=1001 (9)$$

算法循环，直至满足终止条件。

试题 6 解答：

Google Duplex 三大主要模块为自动语言识别系统(ASR)、循环神经网络(RNN)、文本转语音系统(TTS)。

自动语言识别系统：该模块功能为将听到的声音转换为文本信息。

循环神经网络：该模块功能为理解输入的文本信息，并产生对话内容(回答)。

文本转语音系统：将文本信息转换为语音，并决定语音在语调、语气以及一些语言习惯上的特征，使之更加自然。

试题 7 解答：

DeepQA 的核心设计准则是：（1）大规模并行处理；（2）概率问题和内容分析的整合；（3）可行度评估；（4）浅层与深层知识的整合。

试题 8 解答：

策略网络是一个 13 层的深度卷积神经网络，它从当前已有的棋局中进行学习，根据输入的棋局状态，预测所有合法的下一步的概率，选择概率较大的位置落子，模拟人类棋手的落子棋感，属于局部行为。估值网络是一个 12 层的深度卷积神经网络，它侧重于全局形势判断，模拟人类棋手的胜负棋感，输出单个预测结果。

试题 9 解答：

结构化数据表示具有定义好的长度和格式，其元数据、视图和词汇语义是明确定义的，大部分结构化数据存储在传统的关系型数据库和数据仓库中。非结构化和半结构化数据没有特定的格式和明确的语义定义，其语义需要发掘和提取，收集、存储、管理、分析非结构化和半结构化数据需要专门方法，它们一般和非关系型数据库一起使用。

试题 10 解答：

监督机器学习：通过标注样本实现学习过程完成学习任务，例如 K-近邻、决策树等算法；

无监督机器学习：通过无标注样本抽取通用规则完成学习任务，例如层次聚类、主成分分析等算法；

半监督机器学习：通过部分标注样本和未标注样本完成学习任务，例如自训练、生成式模型等算法。

试题 11 解答：

知识图谱 (Knowledge Graph) 的概念由谷歌 2012 年提出，旨在实现更智能的搜索引擎。知识图谱本质上是一种叫做语义网络 (semantic network) 的知识库，即具有有向图结构的一个知识库，其中图的结点代表实体或者属性，而图的边代表实体与实体或者实体与属性之间的各种语义关系。

基于知识图谱的检索和基于关键词检索的主要区别在于：找到最想要的信息；提供全面的摘要；让搜索更有深度和广度。例如，当用户输入 “Jackie Chan”，基于知识图谱的搜索引擎可以识别出 Jackie Chan 其实就是成龙，而且会给出成龙的各种属性信息，比如说出生日期、国籍、配偶等。基于关键词的检索则不能做到这些。