**机器类脑认知与推理研究**

**研究计划书**

修订记录

| 日期 | 修订版本 | 修改章节 | 修改描述 | 作者 |
| --- | --- | --- | --- | --- |
| 2018.5.2 | 观察实验v1 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# 研究目的

总体目标：研究出一套面向认知与推理的类人脑算法

研究阶段1：验证当前的通用算法在认知与推理的场景下的推理能力

研究阶段2：研究当前算法在场景内的局限性并对其进行修正，以实现突破

研究阶段2是在研究阶段1的基础上进行研究，因此本任务书只探讨研究阶段1。

# 研究阶段1的研究方案

## 验证环境

目前神经学、心理学等领域均有很多测试人类智力、认知、推理能力的方案，其中威斯康星的卡牌测试可以通过测试人类的记忆与推理来评估人脑前额叶功能是否完善。因此本研究以通过威斯康星卡牌测试为目标，模拟人脑前额叶的认知推理功能

## 研究方案

由于威斯康星卡牌测试是基于人类具备先验知识的情况下进行测试，因此在研究阶段1中需要寻找出一个具备先验知识的智能体才能评估其认知与推理水平。

基于此，我们提出两种研究方案：

方案一：基于推理环境下的智能体推理能力测试

1. 先验知识获取环境：从各个渠道收集可用于推理的视觉类题目，如图形推理题，即通过三个图形，推断出第四个图形的样子。将这些题目进行整理、清洗，将其构建成一个类强化学习的游戏环境，通过该环境训练智能体。该环境是训练智能体认知、记忆、推理的主要环境。
2. 训练特征提取器：通过基于CNN的自编码技术，训练一套特征提取器。用于智能体在面向图片时的特征提取。该特征提取器的训练数据可采用多种渠道：先验知识获取环境中的题目，也可以是类似于CIFAR10等多种图像识别中的图片。

同时在训练特征提取器的同时，值得注意的是，训练提取器的数据集也可以看做是智能体的先验知识，因此可以再后期实验中对比两种训练特征提取器的效果。

1. 严格按照心理学、医学上投入使用的标准去设计威斯康星验证平台，并可以准确生成评价指标。
2. 将训练好的智能体接入验证平台中，验证该智能体的认知与推理水平。同时可以通过该结果验证智能体是否模拟出人类前额叶功能。

方案二：基于特定场景下的智能体推理能力测试：

1. 选择合适的如推箱子、走迷宫等游戏环境，该环境满足以下条件：
   1. 益智类游戏
   2. 已有开源的强化学习的智能体进行游戏
   3. 该智能体的游戏水平要高于平均人类水平
2. 实现智能体接口，使其能在方案一种已实现的威斯康星验证平台上进行验证实验。如果该实验比正常人类高，可以得出机器已拟合出人类认知与推理的功能；如果验证结果比人类低，且差距很大，则可以得出结论：智能体只是在特定任务下进行特定的训练能超过人类，并不具备认知与推理的功能。
3. 研究该智能体在其游戏中的细节，找出该智能体在其游戏环境中能得到高分值的原因，并分析该原因，看对下一阶段生成类脑智能体有无帮助。

## 实验分工：

收集智力测试数据集，并将其清洗

制作特征提取器，并通过验证可以准确提取图片特征

实现威斯康星验证平台

了解强化学习框架，并实现智能体训练环境。

查找已开源益智类游戏的智能体，并研究其work的原理

## 实验进展：

第一阶段：构建lstm和mlp网络，经过训练，使lstm网络达到91%分类正确率；使mlp网络达到97%分类正确率。参数：mlp hidden\_size=180；lstm hidden\_size =100，word step=3；

第二阶段：将word\_step 加长至10。