中国科学院自动化研究所

算法研究阶段实验报告

实验：面向WCST的智能体认知、记忆、推理的算法研究

汇报人： 郝洺，梁靖旖，任卓明

时 间： 2018年11月29日

修订记录

| 日期 | 修订版本 | 修改章节 | 修改描述 | 作者 |
| --- | --- | --- | --- | --- |
| 2018.11.24 | 智能体V1 |  |  |  |
| 2018.11.29 | 智能体V2 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# 实验目的

本实验的目的在于通过神经网络拟合出人类认知、记忆与推理的智能体，研究出一套面向认知、记忆与推理的类人脑算法。我们将设计好的网络接入到WCST平台，以验证网络的认知、记忆与推理能力。

WCST平台是针对于人类大脑前额叶是否有损伤而进行测试的一个平台。该平台由5张卡片组成，分别是：一张裁判卡，四张决策卡。每轮游戏基于裁判卡的颜色，形状，数量这三个条件推测平台本回合的选择决策，在4张卡片中选出符合该决策的卡片，平台将告诉你选择是否正确。每种策略连续出现10次，一共进行60轮选择。分类结束后，根据测量指标对实验结果，对神经网络的拟合能力进行分析。

# 算法原理和流程

设置一个智能体与平台交互，通过平台传入的当前轮卡牌状态（state）以及上一轮卡牌选择正确与否来进行当前轮卡牌地选择。该智能体由两个网络组成：策略网和执行网。策略网是通过前几轮的选择正确与否来预测当前轮的选择策略；执行网是根据当前桌面卡牌态势及策略网做出的本轮选择策略来进行卡牌的选择。的具体流程如图1：

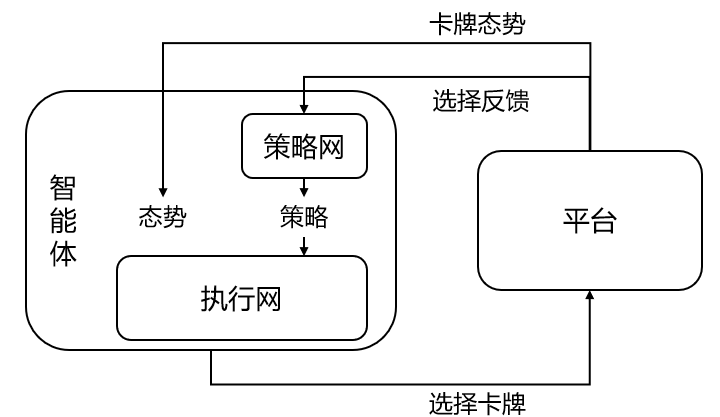


图1：智能体与平台交互示意图

1. **策略网**

策略网是由一个单层LSTM与一个全连接层组成，通过输入由历史决策决策Dt-1以及该决策的判定Rt-1的组成的二元组序列（在本实验中，我们默认输入前6轮历史信息，也就是说，序列长度为6），预测出t时刻的决策。策略网预测流程如下：

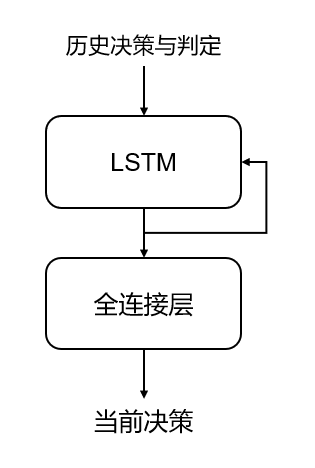


图2：决策网网络结构

1. **执行网**

执行网是由一个单层MLP网络组成，通过执行网，将态势以及策略网生成的决策（颜色，形状，数量策略中的一个策略）输入，选择出对应的卡牌，将选择出的卡牌位置发送给平台。图3为执行网选择卡牌的流程：

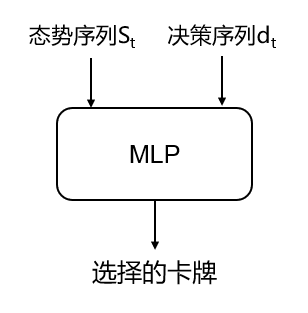


图3：MLP执行网流程图

输入:

态势特征St：五张卡牌的态势（位置，数目，颜色，形状）

决策特征dt：本次选择的策略

输出：本回合游戏应当选择的卡牌的位置（0，1，2，3）

# 实验设置

1. **训练数据说明**
2. 策略网

* 数据集：训练集根据规则人工生成，共生成规则85条，每条规则有10条相同的样本，共计850条样本；测试集是从训练集中随机抽取20%条样本组成。训练集包含测试集。

历史信息：智能体在前六轮做出的决策和平台的反馈，包含四种状态（数目，颜色，形状，空），分别对应成数字，2维。

标签数据：智能体在当前轮做出的决策，转为one-hot编码，3维。

* 数据集格式示例：

[[[1, 1], [1, 1], [1, 1], [1, 1], [1, 1], [1, 0]], [1,0,0]]

[[[3, 3], [3, 3], [3, 3], [1, 0], [2, 0], [0, 0]], [0,0,0]]

其中，三种不同决策分别为{数目:0,颜色:1,形状:2,空：3}。

1. 执行网

* 数据集：训练数据和测试数据由态势特征与策略特征组成。为了保证训练结果的准确性及有效性，我们刺激卡的数量进行了扩充。按照只有一张刺激卡符合选择策略的原则去生成数据集，共生成了10000条无重复的数据，并将其随机分为7000条训练数据与3000条测试数据。

态势特征：五张卡牌的态势（位置，数目，颜色，形状），共20维

策略特征：本次选择的策略:从三种已知策略中选择适合本轮的策略（数目，颜色，形状），1维。

标签数据:在应用本次选择的策略中应选中的正确的卡牌号码（或图案），转为one-hot编码，4维。

* 数据集格式示例：

[(0,1,2,3), (0,0,0,0), (1,1,1,1), (2,2,2,2), (3,3,3,3),0]

[(0,1,1,1), (0,0,0,0), (1,1,1,1), (2,2,2,2), (3,3,3,3),2]

其中，三种不同决策分别为{数目:0,颜色:1,形状:2},四种不同颜色分别为{红色:0,绿色:1,蓝色:2,黄色:3},四种不同形状分别为{圆:0,方块:1,三角:2,星形:3},四种不同数目分别为{一个:0,两个:1,三个:2,四个:3}

1. **模型参数情况**

|  |  |  |
| --- | --- | --- |
| 参数名 | 策略网LSTM训练参数 | 执行网MLP训练参数 |
| 隐层单元 | 150 | 100 |
| 激活函数 | - | ReLU |
| 优化器 | Adam | Adagrad |
| 学习速率 | 5e-3 | 0.3 |
| Dropout概率 | 0.7 | 0.76 |
| 训练轮数 | 10000 | 10000 |
| 测试集准确率 | 94.57% | 97.8 |

# 实验结果

1. **WCST平台模拟结果**

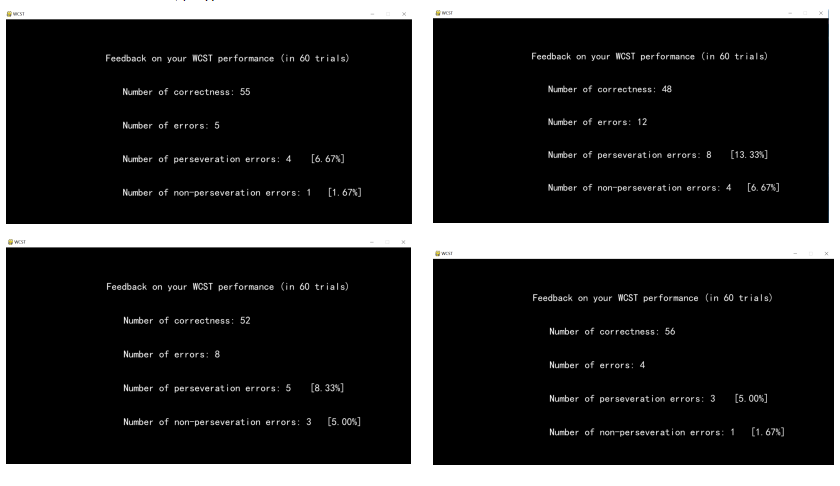
我们随机进行10次实验，连续型错误平均值为7.4，错误百分比为8.13%（低于人类病态水平19%），非连续行错误平均值为2.7（低于人类病态水平≤24），我们随机挑选了四次实验结果，如下图所示：

图4：随机挑选四次实验结果

# 实验结果分析和总结

1. **实验结果分析**

通过第一版实验可知，智能体可以在得到态势信息和决策信息的情况下，选出正确的卡牌，即第一版智能体已经具备了该任务的选择分类功能。然而，第一版智能体的决策错误率极高，体现在：

* 经常犯游戏中持续性应答错误，即明知道该策略是错的，依然按照该策略去选择；
* 回答正确后，并没有顺着正确的策略继续去选，而是选择了其他策略，即智能体不具备记忆功能；
* 经常在错误的答案中徘徊，不去选择正确的答案，即智能体缺乏推理功能。

经过优化LSTM策略网络，第二版智能体在该游戏环境下已经完全具备的人类水准，但我们仍不能得出“该智能体已具备认知、记忆与推理能力”的结论，原因为训练智能体决策网络的数据是根据策略规则生成的，该数据覆盖了所有可能出现的情况。在训练的时候数据集拟合程度极高，因此该智能体就像去执行程序员写好的程序一般，自己并没有做决策。

1. **未来工作**

未来我们工作的将基于该架构，优化决策网络的架构与功能，使其在不依靠规则化数据的情况下具备记忆与推理的能力。然后可以将该成果在其他场景中进行验证。