**方案：队形设计部分**



**一、建立数据库中的数据训练神经网络：**

该神经网络分为3个子网络：1 环境感知网络，2无人机队形神经网络,3无人机内部参数神经网络。其架构如下图所示。

**1）子神经网络1：**

读取周围环境图片，用卷积神经网络判断周围战场环境。

**2）子神经网络2：**

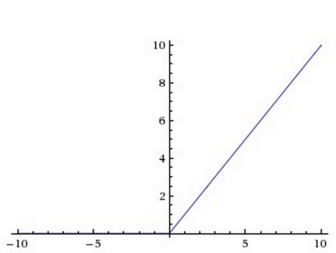
其中，无人机队形神经网络的目标是通过学习以往经验从而决策无人机队形。该神经网络的输入为**任务类型、无人机数目、战场环境**；输出为最优队形。

神经网络的输出为一个9维列向量，其中所相对应的无人机阵型的输出为1，其余阵型对应的输出为零。例，如采用第4中无人机阵型，V队，则相对应的输出为[ 0 0 0 1 0 0 0 0 0 ]T

该神经网络结构：

**隐藏层**采用整流线性单元，即ReLu函数，其函数形式如下：





该函数优点：此函数是被推荐用于大多数前馈神经网络的默认激活函数。将此函数用于线性变换的输出将产生非线性变换。然而，函数仍然非常接近线性，在这种意义上它是具有两个线性部分的分段线性函数。由于整流线性单元几乎是线性的，因此它们保留了许多使得线性模型易于使用基于梯度的方法进行优化的属性。它们还保留了许多使得线性模型能够良好泛化的属性

**代价函数**使用训练数据和模型预测间的交叉熵作为代价函数，交叉熵刻画了两个概率分布之间的距离，它是分类问题中使用比较广泛的一种损失函数。给定两个概率分布*p*和*q*通过*q*来表示*p*的交叉熵为：



**输出单元**采用softmax函数。Softmax最常用作为分类器的输出，来表示n个不同类上的概率分布。从神经科学的角度来看，认为softmax是一种在参与其中的单元之间形成竞争的方式：softmax输出总和为1，所以一个单元的值增加必然对应着其他的值减少，这与被认为存在于皮质中相邻神经元之间的侧抑制效果。其形式如下：



**3）子神经网络3：**

①无人机内部参数神经网络的目标是通过学习数据库中的经验从而计算无人机之间内部参数

该神经网络输入为**无人机队形输出**，**无人机*i*的价值、攻击防御能力、通信功率、机动能力、最小安全距离**；**任务类型**；**风速、降雨量、气压**；输出为无人机*i*的重要程度、无人机*i*与邻机的最优距离*di*。

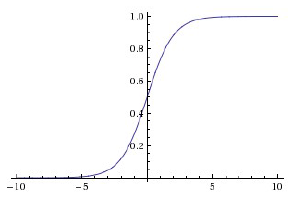
**隐藏层**采用整流线性单元，即ReLu函数，或者其拓展形式（绝对值整流，leaky ReLu，maxout等）。

**代价函数**采用均方误差函数（MSE），该函数是一种普遍的代价函数。



**输出单元**采用sigmoid函数。





②参数调整

由于*di*与*dj*不同，首先根据相邻无人机*i*和*j*的硬性约束条件（机动能力、最小安全距离）进行约束，再根据无人机*i*和*j*的重要程度进行加权平均得到两机间的最优距离。

**二、无人机在执行任务过程中根据变化的实时环境在线优化两机之间的相对距离。**



（1）问题建模：将两机之间相对距离的优化问题转化为马尔科夫决策过程

①状态

选取状态为。其中为无人机性能，包括无人机的价值、攻击防御能力、任务类型、通信功率、机动能力、最小安全距离；为两机距离；为战场环境，包括风速、降雨量、气压；为任务类型。综合考虑有人/无人机机动能力约束及最小安全距离约束，确定两机之间的最小距离为（必须大于最小转弯半径与最小安全距离），两机之间的最大距离由通信功率约束确定（必须保证两机之间能够进行数据传输）。以0.1m的距离为单位，将两机之间的距离离散化，即两机之间的距离现为有限个状态量。

②动作

取为距离调整方向或，则两机之间的实际调整距离为，其中表示调整的单位距离，即0.1m。

③回报函数

**拟选取优化问题中的性能指标函数作为回报函数**，即在当前两机距离下，选取调整方向后的收益。

④策略

将策略选取为**贪婪策略**

⑤折扣因子

选取折扣因子

因此，对应的状态价值函数

****

其中N为人为设定参数。

所以，最优状态价值函数

****

根据贝尔曼最优方程：

****

对应的最优策略：



（2）求解

考虑到由于动作个数比较少，因此选取下面的输入输出表示方法。



Value网络：

* 输入：无人机*i*的价值、攻击防御能力、任务类型、通信功率、机动能力、最小安全距离；任务类型；风速、降雨量、气压；无人机间距离，即状态s。
* 输出：、、
* 训练

①根据贝尔曼最优方程



所以损失函数

****



②由于神经网络进行训练时，假设输入数据是独立同分布的，而通过强化学习采集到的数据之间存在关联性，利用这些数据进行训练，很容易出现不收敛的情况，而且标签中与输出中的网络参数相同，这样就容易使得数据间存在关联性，训练不稳定。

引入记忆机制：DeepMind的创始人Hassabis利用人的海马体在睡觉时候把一天的记忆重放给大脑皮层这一机制，构造了神经网络的训练方法：经验回放（experience replay）。

将大量经历数据以存储在内存里，作为D大块，从D大块中随机抽取小块数据作为输入进行训练。