**基于机器学习的RTS强化学习系统的应用**

## 摘要

电子游戏从出现开始就推动着计算机技术的发展，在过去的十年里，人工智能（AI）技术的发展也离不开游戏的研究。作为AI研究绝佳的测试平台，在人工智能方面从古老的棋盘游戏、之后又到经典的街机游戏、再到2016年AlphaGo战胜李世石，人工智能Agent都表现卓越。MMORPG游戏作为网络游戏的一支很重要的大军，势必要搭上人工智能的车重新占领游戏市场。本文通过深入了解游戏人工智能历史，基于近期DeepMind团队与暴雪新开源星际争霸2机器学习平台，结合机器学习在RTS策略游戏上的研究成果——在星际争霸中多兵种协同作战的实现，完善该强化学习系统，基于暴雪魔兽世界插件库，实现多人作战插件，运用到战斗系统插件中，从而提升MMORPG游戏用户体验。

## 关键词: 人工智能、MMORPG、RTS星际争霸、游戏人工智能应用

## 1.研究背景与意义

在过去的十年里，人工智能以迅猛的发展之势横扫计算机界，而在有监督学习的情况下，机器在图像识别和语音识别能力上，已经有预期的超越了人类的认知水平。前不久，DeepMind在A3C上发表的论文中提出了一套UNREAL系统，在Atari游戏合集上的表现得分基本与人类玩家持平，而且有的还能超过人类。而阿里的新研究则通过BiCNet在RTS（《星际争霸》）上实现了多兵种协同作战，该实验证明了这一系统在RTS游戏上各兵种协调全作的有效性，以及指定奖励和学习策略存在相关性。谷歌DeepMind团队与暴雪开源了星际争霸2机器学习平台，提供了游戏输入输出API，DeepMind又做了基于Python的封装，持续发布基于replay的增强学习训练数据，此外两家公司还发布了上百万帧来自天梯专业选手的游戏记录，这将有助于基于平台算法应用到不同应用上。

MMORPG游戏在过去的十多年发展到现在，已经处在一个比较尴尬的状态，处于下滑态势。一方面由于游戏快餐化，娱乐模式呈现井喷，MMORPG又大体复制从出生就有的游戏模式，使得游戏在创新方面已无法满足玩家游戏乐趣。然而MMORPG停滞不前的十年正是人工智能发展的十年，随着人工智能不断的发展，游戏需要由低智能向高智能发展。首先以遗传算法和神经网络为代表的强人工智能将大大提高游戏的智能性。此外，前不久来自阿里巴巴和UCL的研究者们通过提出双向协调网络（BiCNet）实现了多兵种协调作战。这为MMORPG游戏多Boss作战系统，Boss战玩家与Agent体协作系统，以及玩家战斗协作系统提供了思路。以魔兽世界为例，基于玩家个人操作数据，完全可以通过训练实现一套智能插件辅助系统。

## 2 本课题的研究历史与现状

#### 2.1 发展历史

2.1.1 MMORPG发展历史

1980年第一款MMORPG（大型多人联机角色扮演）游戏诞生——《凯斯迈之岛》，后该类游戏逐渐进入玩家视野，从上世纪90年代末《网络创世纪》、《无尽的任务》到2004年在全球发售运营的《魔兽世界》。



图1 MMORPG发展历程

但是由于游戏市场的更迭以及各种类游戏对市场的逐渐分成，使得MMORPG在近年光环褪去，早在2010年，以《魔兽世界》为例，从曾一度拥有1200w续费玩家降到了780w，而且由于MMORPG游戏研发费用高昂，普通小型公司无法应付资金的巨大压力，而且大型公司则是由于多年市场经验不愿承担高额风险来换取该类型游戏的新鲜血液，导致MMORPG运转状况不佳，难以继续吸引大量新的玩家。

 图2 MMORPG的黄金时代

图3 MMORPG的黄金时代

图4 MMORPG的黄金时代

Trion Worlds的《时空裂隙》和EA的《星球大战：旧共和国》都曾希望复制《魔兽世界》的成功模式，然而创新的玩法以及更高的智能体系统在没出现之前，旧模式始终无法玩家的游戏性。

《魔兽世界》截止目前仍然是最成功的MMORPG，2010年付费用户和收入到达峰值后，标志着欧美网游市场MMORPG开始走下坡路。

 图5 MMORPG的黄金时代

2.1.2 人工智能在电子游戏上的发展

电子游戏一直是人工智能研究和测试的重要组成部分。在二十世纪九十年代中期，IBM 的超级计算机「深蓝」数次击败了国际象棋世界冠军卡斯帕罗夫。后来到了 2011 年，IBM 的沃森人工智能在游戏《危险边缘》就击败了最优秀的人类玩家，并展示了 IBM 在自然语言处理的进展。2015 年，DeepMind 就开始使用机器学习来训练人工智能玩一些复古的雅达利游戏（Atari games），并使其至少能做得和人类一样好。后来在 2016 年的三月份，DeepMind 通过另一种方法训练了 Alpha Go，并击败了围棋世界冠军李世乭。

DeepMind在2015年提出了 UNREAL（无监督强化和辅助学习/UNsupervised REinforcement and Auxiliary Learning）代理。在一套 57 个 Atari 游戏合集和拥有 13 个级别的 3D 迷宫游戏 Labyrinth 中测试了这一新系统。在所有游戏中 UNREAL 代理被用同样的方式训练，系统只接收屏幕图像的信息，试图在游戏中获得最多的得分和奖励。在不同游戏中，得分的方式各不相同，从玩 3D 迷宫到《Space Invaders》——同样的 UNREAL 算法学会了所有这些游戏，得分几乎与人类玩家持平，有些甚至超过了人类。

去年第一季度,暴雪和 DeepMind开源了星际争霸2的机器学习平台，这个平台对于state-of-the-art的深度强化学习算法是极好的测试平台，此外，两家公司还发布了上百万帧的来自天梯上专业选手的游戏记录。在新界面里，《星际争霸 2》的图形被简化以便于机器学习系统进行识别，同时他们开放 API，允许系统读取游戏中的数据，实现原先电脑玩家的部分功能。暴雪还发布游戏 replay 文件数据集以供机器学习。DeepMind 的最终目标是让人工智能系统和人类玩家一样，通过处理视觉信息理解游戏。

#### 2.2 国内外研究最新动态

最近，阿里巴巴和UCL的研究者们新提出了双向协调网络BiCNet来相互交流，通过评估-决策方式来完成星际争霸战斗任务。以 RTS 游戏《星际争霸》为测试场景，设定任务为多个智能体互相协作试图击败敌人。为了保证沟通方式有效且可扩展，他们引入了多智能体双向协调网络（BiCNet），它具有向量化扩展评价器（actor-critic）形式。验证了 BiCNet 可以协调不同兵种，在不同的场景和两方智能体数量任意的情况下正常工作。分析证明，在没有手动标记数据进行监督学习的情况下，BiCNet 可以学会多种有经验的人类玩家展示出的协调策略。而且，BiCNet 能够轻松适应异构智能体任务。在实验中，在不同的场景下用该新方法与不同的基准进行了对比；BiCNet 展现出了最先进的性能，它具有在现实世界大规模应用的潜在价值。

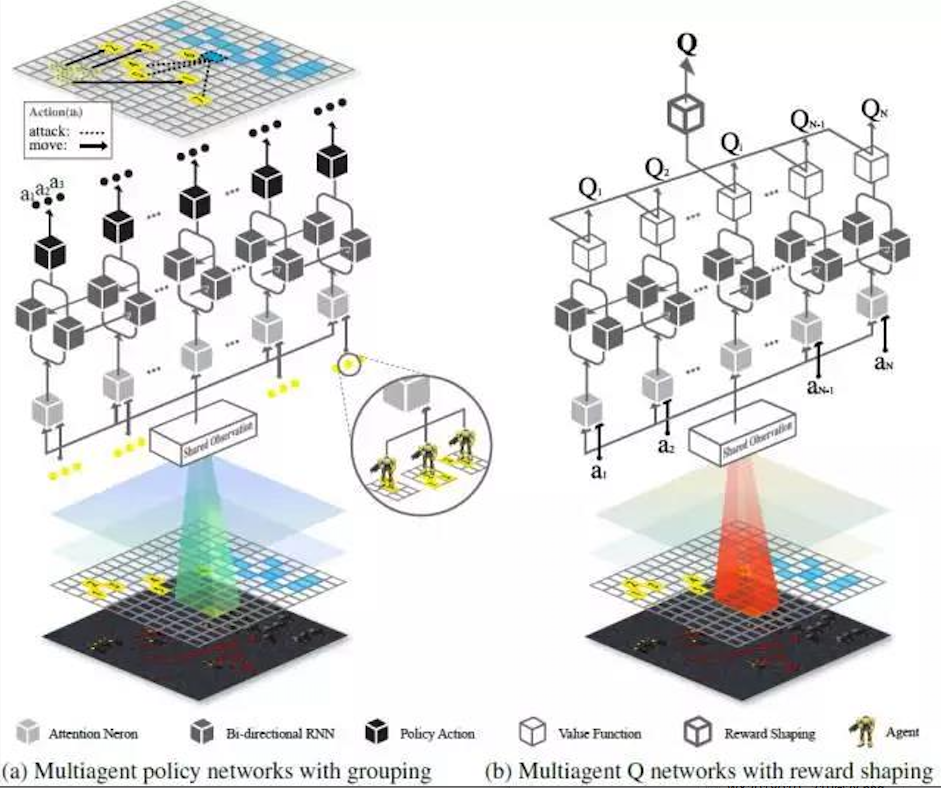
然而，中国科学院自动化所在近期提出的一种强化学习+课程迁移学习的方法，可以让AI Agent在组队作战过程中掌握微操作能力，控制多个单元。该研究定义了一种高效的状态表征，提出一个参数共享多智能体梯度下降Sarsa()(PS-MAGDS)算法训练单元，该方法使用一个神经网络作为函数近似器，以评估动作价值函数，还提出了一个帮助单元平衡智能体的移动和攻击。通过强化学习和课程迁移学习，智能体Agent可以在星际争霸的微操场景中学习合适的策略。

## 3. 我文要实现的内容

3.1 研究基于星际争霸的RL基本实现原理

3.1.1 BiCNet是一种利用双向神经网络的多智能体强化学习框架。通过构建矢量化的评估-决策方式，其中每个维度对应一个智能体Agent。Agent之前的协调通过内部双向通信。基于端到端学习，BiCNet可以学会多种有效的协同作战策略。BiCNet提出了一种强化学习与迁移学习的方法，从而可以解决星际争霸中多Agent控制问题。

讨论多Agent在RL上的研究

图6 双向协调网络示意图

通过对代码分析，掌握基本实现原理，使用算法处理用户操作。

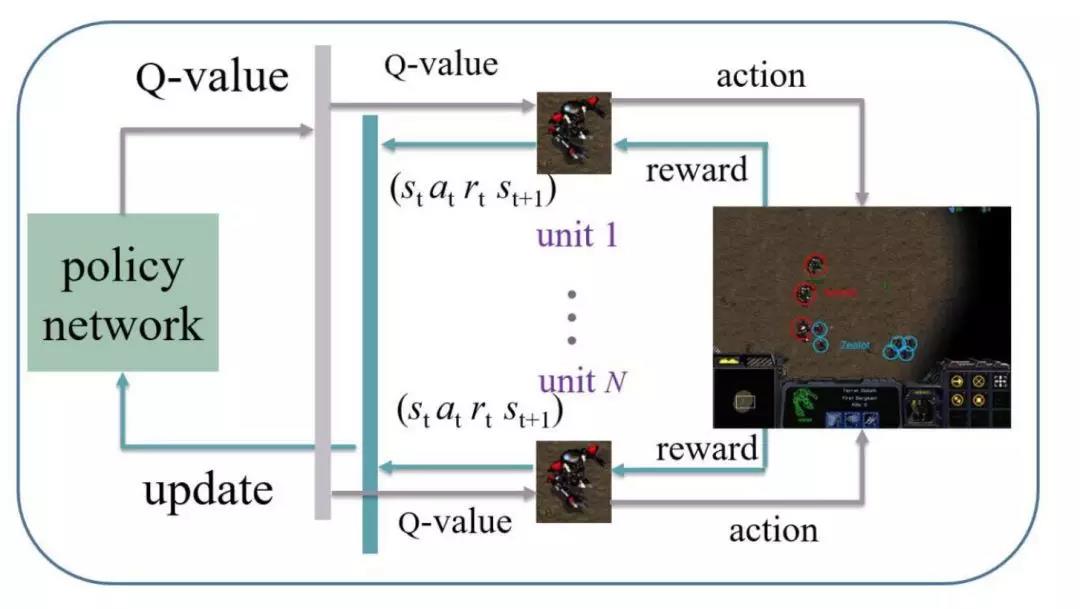
3.1.2 在中科院的研究中，星际争霸微操被定义为多智能体强化学习模型，整个强化学习范式如图7：

图7 StarCraft微操场景中的PS-MAGDS强化学习图示

该学习模型提出了一种高效的状态表征方法用来处理星际微操的大型状态空间。这种方法考虑了单位的属性和距离，并允许双方单位数量任意。使用PSMAGDS算法训练单位，鼓励合作，并在RL模型中提供奖励函数，以提高相互协作能力。同时，用一种迁移学习的方法来扩展，使得在大规模情况下，一组单位有更好的学习性能。

3.2分析该系统应用的可行性

首先World of Warcraft API是Blizzard WoW客户端提供的一套Lua函数与工具集，可以使用户与游戏客户端进行交互，同时用户操作的数据可以通过接口获取到；

其次，Lua作为一门轻量而快速的脚本语言，功能在高级动态语言中十分完备，对C API以及线程安全的VM支持，非常适合于处理业务逻辑；

本应用的几个难点在于：

1. 如何通过WoW API收集用户操作以产生可用的训练数据集；
2. 如何将可用的训练的数据运用到算法中，从而产生训练结果以及结果回传；
3. 如何进行训练算法的改进，提供完整的适用于玩家插件需求的结果数据；

最后，训练算法基于C++实现，Lua在C/C++上的数据传递以及封装上都比较方便，又魔兽世界又有开放的Lua脚本API，这将为该强化学习的系统在其上应用提供了保障。

3.3 实现通过脚本来使用机器学习算法训练数集的结果

1. 收集用户训练数据集；
2. 将收集的用户数据整理成可用的格式；

(3) 通过训练算法得出训练结果数据；

(4) 训练算法回传，得到结果；

## 4．总结

基于星际争霸的机器学习算法，使得星际争霸的多智能体能够进行合作，以达到人类玩家微操的效果，这种强化学习以及迁移学习的方法，在RTS战略游戏上的可行性得到了初步的证明。

而MMORPG游戏，特别是魔兽世界，在即时战略系统上也可以使用相同的原理，基于魔兽世界插件的开放接口，可以将该训练模型用于魔兽团队战略插件系统中。

这样的一次尝试将机器学习训练系统运用到MMORPG游戏中，这将进一步肯定AI算法在游戏中的应用，将有利于创造更完善的游戏AI系统，增加更智能的Boss体，以及更加智能的战斗游戏策略，以此来提高游戏体验。

## 参 考 文 献

[1] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, et al., "Human-level control through deep reinforcement learning," Nature, vol. 518(7540), pp. 529-533, 2015.

[2] Peng Peng, Quan Yuan, Ying Wen, Yaodong Yang, Zhenkun Tang, Haitao Long, Jun Wang, “Multiagent Bidirectionally-Coordinated Nets: Emergence of Human-level Coordination in Learning to Play StarCraft Combat Games,”, (1) Alibaba Group, (2) University College London, 2017.

[3] Kun Shao, Yuanheng Zhu, Dongbin Zhao, “StarCraft Micromanagement with Reinforcement Learning and Curriculum Transfer Learning,”,IEEE, 2018.

[4] Oriol Vinyals, Timo Ewalds, Sergey Bartunov, Petko Georgiev, Alexander Sasha Vezhnevets, Michelle Yeo, Alireza Makhzani, Heinrich Küttler, John Agapiou, Julian Schrittwieser, John Quan, Stephen Gaffney, Stig Petersen, Karen Simonyan, Tom Schaul, Hado van Hasselt, David Silver, Timothy Lillicrap, Kevin Calderone, Paul Keet, Anthony Brunasso, David Lawrence, Anders Ekermo, Jacob Repp, Rodney Tsing, “StarCraft II: A New Challenge for Reinforcement Learning,”, 2017.

[4] 陈钇似.基于机器学习的游戏智能系统研究与应用[D].电子科技大学.2017

[5] 郑其宝.人工智能影响下数字游戏智能化发展探究[D].南京艺术学院.2016

[6] 陈汉伟.基于人工智能的虚拟角色行为决策能力的研究[D]. 浙江理工大学. 2015

[7] 常迎春.基于分层时序记忆算法的游戏人工智能的异常检测[D].河北工业大学.2015

[8] 姜文军.网络游戏中人工智能的研究及应用[D].上海交通大学.2012

[9] 王淑琴.神经网络和遗传算法在游戏设计中的应用研究[D].东西师范大学.2004

[10] <http://www.infoq.com/cn/articles/atari-reinforcement-learning>

[11] <https://blog.csdn.net/mydear_11000/article/details/51488118>

[12] <http://baijiahao.baidu.com/s?id=1566325283820033&wfr=spider&for=pc>

[13] <https://www.nextbigfuture.com/2014/12/deep-learning-and-deep-reinforcement.html>