基于 Attention 机制与 LRUA 模块的 ESports 行为模式预测模型

于 诚[1] 朱皖宁[1] 游 坤[1] 朱金付[2]
(金陵科技学院软件工程学院 南京 211169)[1] (中国传媒大学南广学院 南京 211172)[2]

摘 要 随着电子竞技产业的不断发展,对电子竞技比赛进行准确且快速的数据分析显得越来越重要。文中对电子竞技行为模式预测这一重要问题进行了研究。从度量学习的角度出发,通过引入修正余弦度量替代余弦度量的方法,改善了行为模式预测因为队伍评价尺度不同而导致模型不精确的问题。同时,为了进一步提高模型的精确度,从文中数据的特征出发,考虑到该问题较为注重数据的内容,因此引入 LRUA 模块进行内存的存取。实验表明,所提模型具有较高的准确率以及较低的波动性。

关键词 行为模式预测,电子竞技,元学习,度量学习,匹配网络
中图法分类号 TP18 **文献标识码** A

Prediction Model of E-sports Behavior Pattern Based on Attention Mechanism and LRUA Module

YU Cheng[1] ZHU Wan-ning[1] YOU Kun[1] ZHU Jin-fu[2]
(Department of Software Engineering, Jinling Institute of Technology, Nanjing 211169, China)[1]
(Communication University of China, Nanjing 211172, China)[2]

Abstract With the development of e-sports industry, it is more and more important to analyze data accurately and quickly. This paper studied the prediction of e-sports behavior pattern . From the perspective of measurement learning, the model inaccuracy of the prediction of e-sports behavior pattern caused by different evaluation scales of teams is improved by introducing modified cosine measure instead of cosine measure. Meanwhile, in order to further improve the accuracy of the model, this paper started from the characteristics of the data in this paper. Considering that this paper pays more attention to the content of data, LRUA module is introduced for memory access. Experiments show that the proposed model has high accuracy and low volatility.

Keywords Behavioral model prediction, Esports, Meta-learning, Metric learning, Matching networks

1 引言

近年来,电子竞技(Electronic Sports, ESports)产业不断发展,在政府态度转变与支持、雄厚资本的注入、网络新媒体的到来以及电子竞技设备的提升、电子竞技爱好者不断增加等条件下,电子竞技产业越来越壮大[1]。为了能够在电子竞技比赛中获得更大的优势,精确地分析电子竞技中的战略决策显得越来越重要[2]。由于相对较少的数据、不断变化的竞技状态、游戏的复杂程度高,人工进行这样的分析是较为困难且高成本的,因此各种电子竞技数据分析算法应运而生,以解决互联网时代背景下电子竞技比赛数据急需快速且精确分析的问题。

当前已有一些研究人员对电子竞技战略决策问题进行了相关研究。已有算法主要解决了以下问题:1)赛前数据预测,即英雄的选择与禁选(BAN AND PICK, BP)预测[3];2)比赛数据实时胜率预测[4-5];3)优势组合挖掘[6];4)MOBA 类角色的会话识别及数据获取[8-9]等。

本文经过研究发现,比赛中角色的行为模式(能否准确预测某个角色在某局比赛中的特定时间段内倾向于去往战场的某个区域做什么事情)对战术安排的影响极大。

本文以 MOBA 类电子竞技英雄联盟(League of Legends, LOL)为例。如图 1 所示,LOL 在比赛开始时,双方角色从基地出发,基地每隔一段时间会诞生一些小兵。小兵会沿着固定的道路行走,共有 3 条道路(红、蓝箭头所示),在基地和 3 条主要道路以外的区域统称为野区(黄色椭圆区域)。

图 1 LOL 赛场(电子版为彩色)

比赛时双方至少会有一名队员在野区活动(称为打野),

本文受江苏省高等学校自然科学研究面上项目(17KJB510054),云环境服务质量保障技术(jit-b-201705),金陵科技学院高层次人才科研启动基金资助(jit-b-201624),国家自然科学基金(61502101),江苏省品牌软件工程(40715108)资助。

于 诚(1997—),男,主要研究方向为机器学习,E-mail:17766006696@foxmail.com;朱皖宁(1983—),男,博士,主要研究方向为量子计算,E-mail:zhuwanning@jit.edu.cn(通信作者)。

因为野区缺乏视野,所以打野的行动往往可以出其不意的打击到对方(GANK),野区的怪物称为野怪,打野在比赛前期的主要活动是杀野怪(杀野),阻止对方打野杀野怪(反野)和GANK,因为打野成功的反野和 GANK 都可以在比赛前期为队伍获取极大的优势,所以对角色打野的行为模式分析尤其重要。而关于电子竞技角色行为模式分析的研究尚处于空白,本文就这一重要但至今无人解决的问题提出了一种解决方法。

因为电子竞技行为模式分析问题仍具有电子竞技战略决策的数据较少、类别多的特点,所以传统的深度学习网络应用于此问题的效果极差。因此,本文提出基于 Attention kernel[10]的网络,并采用最近最少使用存取策略(Least Recently UsedAccess,LRUA)的改进模型。实验结果表明,本文提出的方法对于电子竞技行为模式分析这一独立于序列的信息联合编码任务的效果比传统方法更优。

2　国内外的研究进展

在之前的研究中,我们提出了一种行为识别算法来获得角色的会话数据[7-8]。本文将基于会话数据进行行为预测并且巧妙地将问题转化为类文本预测问题(如机器翻译问题)。本节将介绍国内外的一些重要算法:最近最少存取模块、Attentionkernel 机制。

2.1　最近最少使用存取

LRUA 策略由 Santoro 等最初在 ICML2016 上提出[11]并被应用于元学习的神经图灵机(Neural Turing Machine,NTM)模型。LRUA 模块是一个纯粹的基于内容的内存写入器,它将内存写入最少使用的内存位置或最近使用的内存位置。本文模型中的 external memories 应用该模块后,模型预测本文问题(强调数据内容的序列预测的特征)的结果较好。

2.2　修正余弦度量

对于度量学习而言,是否能够准确地反映数据之间的相似性尤为重要。修正余弦度量这一相似性度量方法,考虑了用户间的不同评价尺度的问题。相比传统的余弦相似性度量,这一度量方法对存在评价尺度有分歧的问题有着较好的效果[13-14]。

2.3　注意内核

基于 Attention 机制的机器学习模型一经提出,就受到了广泛关注。Vinyals 等[10]最初将度量学习(metric learning)与外部储存器(external memories)相结合,提出基于 Attention 机制的匹配网络,如图 2 所示。图中,左边 4 张图片称为 supportexample,右下角图片称为 test example,全部图片称为一个 task。

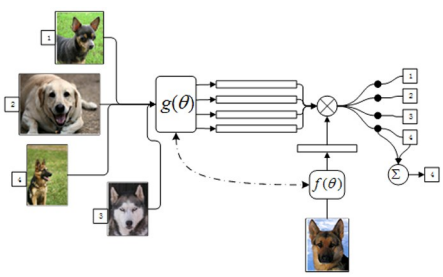


图 2　匹配网络结构

该网络能够快速并且以较高的准确率学习小样本数据,同时能够很好地处理 Seq2Seq 问题。Vaswani 等[12]于 2017 年提出 Attention 机制的 Transformer 结构,此结构大大减少了序列处理任务的计算量。由于本文所研究问题的样本数据并不复杂,数据长度并没有达到机器翻译水平并且数据量较少,导致计算量并不高,因此本文并没有采用 transformer,而对于模型如何提高预测准确性的问题进行了着重研究。

2.4　双向长短记忆网络(Bi-directional Long and Short Memory Network,Bi-LSTM)

传统记忆网络 LSTM 存在一个缺陷,它仅能够利用从正向序列来的以前的内容。对于行为预测问题来说,从反向序列而来的未来的内容对行为的准确预测也起着至关重要的作用。而 Bi-LSTM[15]的出现很好地解决了这一问题,如图 3 所示。

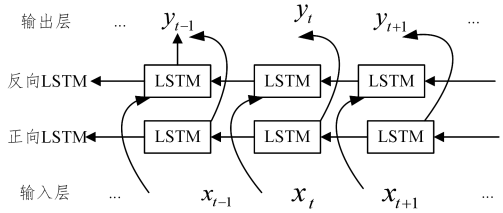


图 3　Bi-LSTM 的结构图

3　预备知识

本节主要介绍行为模式分析,讨论了这一问题的数据特征以及传统模型应用于本文问题上精度不高的原因。

如图 4 所示,本文经过研究发现角色的行为模式主要受 6 个因素影响:会话开始时间、会话 id、会话持续时间、队伍、用户、英雄(角色)id。文献[7]定义了会话,即表示某个角色在某段时间、某个地点所进行的行为,如蓝方打野在 3～4 min 在野区打野。

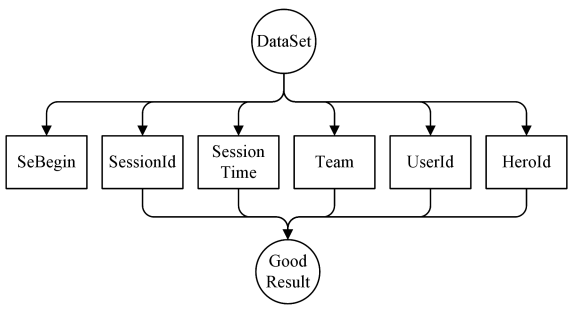


图 4　行为预测因素图

行为模式主要围绕角色的会话数据进行分析,其目的是得到选手使用某个英雄的行为趋向,即给定以往时刻会话数据集合,得到下一个时刻的决策,而每个时刻的会话数据类似机器翻译问题中所要翻译句子的单词数据,而下一个时刻的决策类似单词组成的句子的语义。因此,本文巧妙地将行为模式问题抽象化为 sequence to sequence(Seq2Seq)问题,提出利用能较好解决小样本 Seq 问题的 Attention 模型来解决该问题,并在传统模型上做出改进来提高精确度。

传统模型应用于本文问题时精确度并不高,原因如下:1)不同于传统的机器翻译问题,不同队伍对于战术有着不同的评价尺度,导致本文问题不同于传统的机器翻译问题,原模型的余弦度量并不能很好地反映数据间的相关性;2)传统模型中的外部存储器的存取方式对于本文这种强调独立于序列

的信息联合编码任务的应用效果并不理想。考虑到上文提到的传统模型的两点问题,本文提出应用最近最少使用策略来对外部存储器进行存取,采用修正余弦度量代替余弦度量来使机器评价更为公正。实验对比结果表明,此方法在本文问题中有着较好的效果。

4 问题定义

本文根据角色在一局游戏内的会话开始时间(session_begin)、会话 id(session_id)、会话持续时间(session_duration)、队伍(team)、用户(选手)(user)、英雄(角色)(hero)对角色下一时刻的行为进行预测,对本文问题给出如下的形式化定义。

首先,对于每个 task,有 support set 为 $S=\{(x_i,y_i)\}_{i=1}^{k}$,test set 为 $Te=\{(\hat{x}_j,\hat{y}_j)\}_{j=1}^{n}$。本文算法的目标为基于 S 使得概率 $P(\hat{x}_j\mid\hat{y}_j,S)$ 最大。其中 $i,j,k,n\in Z$,k 表示 support set 中 support example 的个数。n 表示 test set 中 test example 的个数。每个样本向量 $x_i=(x_{isession\_begin},x_{isession\_id},x_{isession\_duration},x_{iteam},x_{iuser},x_{ihero})$。$y_i$ 代表 x_i 的标签分类。

5 基于 Attention 与 LRUA 的元学习算法

本节主要介绍本文提出的基于 Attention 与 LRUA 模块的元学习模型。为了提高模型计算本文问题的准确性,我们从两个方面进行改进:1)用考虑用户评价尺度的修正余弦度量替代余弦度量;2)在 external memories 中使用 LRUA 模块,使得模型能够更好地考虑本文问题基于内容的特点。

5.1 基于 Attention 与 LRUA 的算法框架

本文在基于 Attention 模型的基础上,在 external memories 模块中利用 LRUA 模块进行存取来解决电子竞技行为模式分析问题。本文所提出的算法框架如图 5 所示。

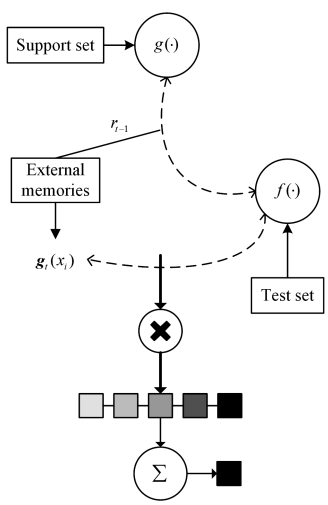


图 5 基于 Attention 与 LRUA 的算法框架

对于一个 task(Support Set 和 Test Set),分别利用 2 个嵌入函数 f 和 g 对数据集合进行处理,然后利用 LRUA 存取 kernel 向量,建立完整模型。

5.2 Attention 方法

本文算法的目的是让 $P(\hat{x}_j\mid\hat{y}_j,S)$ 最大,可以形式化为:

$$\hat{y}_j=\sum_{i=1}^{k}a(\hat{x}_j,x_i)y_i \tag{1}$$

其中 attention kernel 是 cosine 距离上的 softmax:

$$a(\hat{x}_j,x_i)=\frac{e^{\cos(f(\hat{x}_m),g(x_i))}}{\sum_{j=1}^{k}e^{\cos(f(\hat{x}_m),g(x_j))}} \tag{2}$$

其中,f 和 g 是 2 个嵌入函数,cos 为余弦度量公式。由于每个赛季比赛选手的次数有限导致数据集合较小,因此如何更准确地判断样本间的相关性尤为重要,具体到本文问题上,序列预测受到相似性准则的影响(cos)。对于电子竞技而言,每个团队对于每个电子竞技赛季的战术都有着不同的安排。不难看出,当他们进行战术决策时的评价准则是不同的,即每个团队有不同的评价方案。但是传统模型(2)中的余弦相似度很难体现数据的这一特征(余弦相似度计算方法仅限于不同用户的评分尺度基本相同的情况下,并未考虑用户之间不同的评价尺度)。

为了解决这一问题,本文提出利用修正余弦相似性度量,通过在余弦相似性的基础上减去队伍对数据的平均分数来弥补上述传统模型的缺陷,因此 $a(\hat{x}_j,x_i)=a(\hat{x}_m^{team},x_i^{team})$,令

$$simc=\sum(f(\hat{x}_m^{team})-\overline{g(x_i^{team})})(g(x_i^{team})-\overline{g(x_i^{team})})$$

$$simd=\sqrt{\sum_{team\in Team}(f(\hat{x}_m^{team})-\overline{g(x_i^{team})})^2}\sqrt{\sum_{team\in Team}(g(x_i^{team})-\overline{g(x_i^{team})})^2}$$

$$sim(f(\hat{x}_m^{team}),g(x_i^{team}))=\frac{simc}{simd} \tag{3}$$

其中,$\hat{x}_m^{team}$ 与 x_i^{team} 分别表示队伍为 team 的 test example m 和 support example i,$g(x_i^{team})$ 代表队伍 team 已评价项目的平均值。因此 attention 最终定义为:

$$a(\hat{x}_j,x_i)=\frac{e^{sim(f(\hat{x}_m),g(x_i))}}{\sum_{j=1}^{k}e^{sim(f(\hat{x}_m),g(x_j))}} \tag{4}$$

5.3 嵌入函数 Bi-LSTM

为了充分考虑过去与未来的特征,g 采用双向长短记忆网络(Bi-directional Long and Short Memory Network,Bi-LSTM)对每个 x_i 进行编码,每个维护 4 个变量:前向 $\overrightarrow{h}_i$,$\overrightarrow{c}_i$,后向 $\overleftarrow{h}_i$,$\overleftarrow{c}_i$。因此,Bi-LSTM 的前向变量可形式化为 $\overrightarrow{h}_i,\overrightarrow{c}_i=LSTM(x_i,\overrightarrow{h}_{i-1},\overrightarrow{c}_{i-1})$,后向变量为 $\overleftarrow{h}_i,\overleftarrow{c}_i=LSTM(x_i,\overleftarrow{h}_{i+1},\overleftarrow{c}_{i+1})$,其中,$\overrightarrow{h}_i$ 表示隐状态,$\overrightarrow{c}_i$ 表示记忆。test set 内样本特征由隐状态和原始归一化特征 $g'(x_i)$ 共同决定,$g(x_i)=\overrightarrow{h}_i+\overleftarrow{h}_i+g'(x_i)$,$i=1,2,\cdots,k$,此式对 test set 使用了 Bi-LSTM,使得每个样本不但和自身有关还间接和其他参考样本有关。对于每个 support example,通过 $l=1,2,\cdots,L$ 次迭代,使用一个 attentionLSTM 模型计算样本的特征,每个时步对 4 个状态变量进行维护:隐变量 h_l,$\hat{h}_l$,读数 r_l,记忆 c_l。在 l 时步,首先用 LSTM 模型计算原始隐变量和记忆:$\hat{h}_l,c_l=LSTM(f'(\hat{x}_j),[h_{l-1},r_{l-1}],c_{l-1})$,然后加上原始归一化特征 $f'(\hat{x}_j)$,得到本步的隐变量 $h_l=\hat{h}_l+f'(\hat{x}_j)$,读数是 support set 的加权和:

$$r_l=\sum_{i=1}^{k}a(h_{l-1},g(x_i))\cdot g(x_i) \tag{5}$$

其中,$a(h_{l-1},g(x_i))=softmax(h_{k-1}^{T}g(x_i))$,$T$ 表示为网络最大时步。

5.4 external memories 存取

为了能用最新更新和更相关的信息来更新内存,提出采用 LRUA 模块进行存取:将数据写入很少使用或最近写入的

位置。具体策略如下:引入 use attention 向量 $\boldsymbol{a}$,衰减在 t 时刻下过去的 use attention 向量 $\boldsymbol{a}$ 加上当前的读数与写数:

$$\boldsymbol{a}_t^u(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i)=\gamma\boldsymbol{a}_{t-1}^u(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i)+\boldsymbol{a}_t^r(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i)+\boldsymbol{a}_t^w(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i) \tag{6}$$

其中,$\boldsymbol{a}_t^r(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i)$如式 (4)所示。同时,引入 $m(\boldsymbol{v},n)$表示向量 $\boldsymbol{v}$ 中第 n 个最小的数,引入最少使用 attention 数 $a_t^{lu}(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i)$:

$$a_t^{lu}(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i)=\begin{cases}0, & \text{if } a_t^u(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i)>m(\boldsymbol{a}_t^u(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i),n)\\ 1, & \text{if } a_t^u(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i)\leqslant m(\boldsymbol{a}_t^u(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i),n)\end{cases} \tag{7}$$

其中,n 表示读取的 support example 的数量。同时利用 sigmoid 门参数来计算以往 attention 读数向量和最少使用数向量的一个凸组合来计算当前的写数向量 $\boldsymbol{a}_t^w(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i)$:

$$\boldsymbol{a}_t^w(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i)=\sigma(r_{t-1})\boldsymbol{a}_{t-1}^r(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i)+(1-\sigma(r_{t-1}))\boldsymbol{a}_{t-1}^{lu}(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i) \tag{8}$$

其中,$\sigma(\cdot)$表示 sigmoid 函数,r_{t-1}表示 $t-1$ 时刻的读数。在写入记忆之前,最少使用位置将从 $\boldsymbol{a}_t^u(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i)$中计算得到并置于 0,然后按照计算权值得到的向量进行写操作:

$$g_t(\boldsymbol{x}_i)=g_{t-1}(\boldsymbol{x}_i)+a_t^w(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i)f(\hat{\boldsymbol{x}}_i) \tag{9}$$

这样,内存就可以写入到零内存槽或之前使用的槽中;如果是后者,那么用得最少的记忆就会被抹去。如此便减少了不相关的 $g(\boldsymbol{x}_j)$,使得 $a(\hat{\boldsymbol{x}}_j,\boldsymbol{x}_i)=\dfrac{e^{\cos(f(\hat{\boldsymbol{x}}_t^m),g(\boldsymbol{x}_i))}}{\sum_{j=1}^{k}e^{\cos(f(\hat{\boldsymbol{x}}_m),g(\boldsymbol{x}_j))}}$ 中的 $\sum_{j=1}^{k}e^{\cos(f(\hat{\boldsymbol{x}}_m),g(\boldsymbol{x}_j))}$ 中的无关数据减少,提高了整体的精度。

6　实验设计

本节主要介绍本文选用的数据集合、数据的特征以及对比试验所用的模型。

6.1　数据集

为了验证模型的有效性,选用了文献[7]积累的 LOL 的 S8 比赛数据集合。该数据集合是由赛季钻石以上水平的玩家录像识别而成,共 5000 场比赛(已将消极比赛排除在外)。其中包括 100 个玩家(每 5 个玩家一个队伍进行灵活组排,共 20 个队伍)一个赛季的各 250 场比赛。

6.2　会话向量

每个会话向量包括会话开始时间、会话 id 、会话持续时间、队伍、用户、英雄(角色)id 等。各特征的未归一化整数范围如表 1 所列。

表 1　特征范围

No.	会话特征	特征范围
1	id	[1,9]
2	开始时间	[1,4200]
3	持续时间	[1,20]
4	队伍	[1,20]
5	用户	[1,100]
6	英雄	[1,140]

对特征的数值范围的解释如下:

(1)由文献[7]定义的 9 类会话。

(2)数据集合中最大游戏时长为 70 min,合计 4200 s。

(3)会话的持续时间最大为 20 s。

(4)由 20 个队伍组成。

(5)由于每个队伍的核心打野位会更换,因此范围设置为 1～100。

(6)S8 赛季英雄联盟共有 140 个英雄。

6.3　对比实验模型

为了验证模型的有效性,本文选用了擅长处理小样本的文本预测的 3 个模型和本文的 ALRUA-MN 结构从测试错误率均值、最小值、标准差这 3 方面进行对比。这 3 个模型分别为:

(1)文献[10]中基于 Attention 的 Matching Network(A-MN)结构。

(2)文献[11]中基于 LRUA 模块的(LRUA-NTM)结构。

(3)文献[12]中的 Transformer(TF)结构。

7　结果分析与结论

分析表 2 的实验统计结果,可得以下结论:

(1)对比模型的预测错误率的最小值和均值可以看出,本文算法拥有更好平均分类错误率和最小分类错误率,说明 ALRUA-MN 对于本文问题来说是一种有效且泛化性能较好的训练方法。

(2)对比分析,本文拥有更小的分类错误率标准差,说明 ALRUA-MN 模型的波动性更小,证明了本文模型的稳定性。

表 2　4 种模型分类错误率的比较结果

模型	A-MN	LRUA-NTM	TF	ALRUA-MN
test-Avg	15.79	26.88	20.77	13.10
test-Min	12.90	24.50	11.80	11.70
test-Std	2.86	1.38	2.95	1.02
train-Avg	13.03	22.00	18.20	11.85

结束语　电子竞技数据分析正受到越来越多研究人员的关注。本文主要讨论了电子竞技行为模式预测这一问题,从度量学习的角度出发,通过引入修正余弦度量替代余弦度量,改善了行为模式预测问题因为队伍评价尺度不同而导致模型不精确的问题。同时,本文为了进一步提高模型的精确度,从本文数据的特征出发,考虑到本文问题较为注重数据的内容,因此引入 LRUA 模块进行内存的存取。实验结果表明,本文提出模型拥有较高的模型准确率以及较低的波动性。下一步将考虑对更多特征的会话进行分析预测以应对不断更新的电子竞技游戏。

参 考 文 献

[1]　王竞成. 我国电子竞技发展的机遇和挑战[J]. 河南教育学院学报(自然科学版),2018(4).

[2]　SEMENOV A,ROMOV P,KOROLEV S,et al. Performance of Machine Learning Algorithms in Predicting Game Outcome from Drafts in Dota 2[C]// International Conference on Analysis of Images. 2016.

[3]　RIOULT F,MÉTIVIER J P,HELLEU B,et al. Mining Tracks of Competitive Video Games [J]. Aasri Procedia,2014,8:82-87.

[4]　WANG K,SHANG W. Outcome prediction of DOTA2 based on Naïve Bayes classifier[C]// IEEE/ACIS International Conference on Computer & Information Science. 2017.

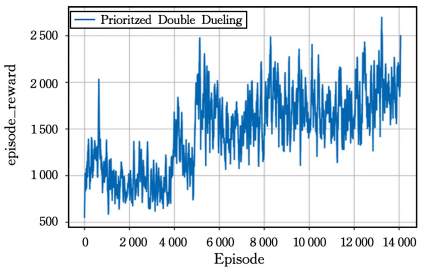
[5]　CHEN Z X,et al. Modeling Game Avatar Synergy and Opposition through Embedding in Multiplayer Online Battle Arena Games[J]. aarXiv:1803. 10402,2018.

[6]　朱益立. 关联规则算法研究及其在电子竞技中的应用[D]. 桂林:桂林电子科技大学,2017.

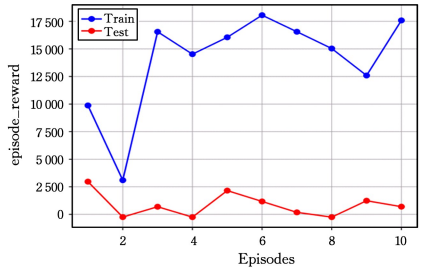
[7]　YU C,ZHU W N,LI L L,et al. Research on E-sports session identification[C]// MATEC Web of Conferences. EDP Sciences,2018:10011.

(下转第 97 页)

目标点，如处于障碍物正后方的目标点对于小车的跟随存在一定的难度。图 6(b)为环境 2 下训练集与测试集总奖励曲线，训练集中小车能达到较高的总奖励值，但个别点的奖励值过低，对于测试集则平均达到跟随 3～4 个随机目标点的结果。



(a)DDPG 总奖励曲线



(b)训练集和测试集下总奖励对比

图 6　Env2 实验数据图

结束语　本文将采用优先采样的 Vanilla DQN，Double Dueling DQN 和 DDPG 3 种算法作用于移动小车，分析了不同环境下小车对移动随机目标点的自主跟随能力。结果表明，在简单的环境 1 下，输出为离散动作的小车，仅仅利用较少的输入信息就能够学习到对移动目标的跟随能力，对于新的随机目标点也具有较好的泛化能力。在复杂的环境 2 下，相对于 DQN 和 DDPG，输出连续动作的 DDPG 算法，能够较快地学习到最优策略，较好地实现在部分环境信息未知下小车对随机移动目标点的跟随。但是在有障碍物的环境下，算法的泛化能力仍需要进一步提升。针对这个问题，未来的研究可以考虑改进小车避障能力和目标跟随能力的分开训练，或者对奖励值部分进行细分等方面。

参 考 文 献

[1] 王春颖，刘平，秦洪政. 移动机器人的智能路径规划算法综述[J]. 传感器与微系统，2018，37(8)：5-8.

[2] 刘全，翟建伟，章宗长，等. 深度强化学习综述[J]. 计算机学报，2018，41(1)：1-27.

[3] HASSELT H V，GUEZ A，SILVER D. Deep Reinforcement Learning with Double Q-learning[J]. Computer Science，2015.

[4] SILVER D，LEVER G，HEESS N，et al. Deterministic policy gradient algorithms[C]// InternationalConference on International Conference on Machine Learning. JMLR. org，2014：387-395.

[5] MNIH V，KAVUKCUOGLU K，SILVER D，et al. Playing Atari with Deep Reinforcement Learning[J]. Computer Science，2013.

[6] KONDA V. Actorcritic algorithms[J]. Siam Journal on Control & Optimization，2003，42(4)：1143-1166.

[7] LILLICRAP T P，HUNT J J，PRITZEL A，et al. Continuous-control with deep reinforcement learning[J]. Computer Science，2015，8(6)：A187.

[8] WANG Z，SCHAUL T，HESSEL M，et al. Dueling network architectures for deep reinforcement learning[J]. arXiv：1511.06581，2015.

[9] HASSELT H V，GUEZ A，SILVER D. Deep Reinforcement Learning with Double Q-learning[J]. Computer Science，2015.

[10] 郭宪，方勇纯. 深入浅出强化学习原理入门[M]. 北京：电子工业出版社，2018：125-141.

[11] SCHAUL T，QUAN J，ANTONOGLOU I，et al. Prioritized Experience Replay[J]. Computer Science，2015.

(上接第 79 页)

[8] LI L L，ZHU W N，YU C，et al. Esports analysis data acquisition algorithm based on convolutional neural network[C]// MATEC Web of Conferences. EDP Sciences，2018，189：03003.

[9] 于诚，朱皖宁. 基于战场热点图的 MOBA 类游戏战术分析研究[J]. 计算机科学，2018，45(S2)：149-151，175.

[10] VINYALS O，BLUNDELL C，LILLICRAP T，et al. Matching networks for one shot learning[C]// Advances in Neural Information Processing Systems. 2016：3630-3638.

[11] SANTORO A，BARTUNOV S，BOTVINICK M，et al. Meta-learning with memory-augmented neural networks[C]// International Conference on Machine Learning. 2016：1842-1850.

[12] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is all you need[C]// Advances in Neural Information Processing Systems. 2017：5998-6008.

[13] 李海林，郭崇慧. 时间序列数据挖掘中特征表示与相似性度量研究综述[J]. 计算机应用研究，2013，30(5)：1285-1291.

[14] 马宏伟，张光卫，李鹏. 协同过滤推荐算法综述[J]. 小型微型计算机系统，2009，30(7)：1282-1288.

[15] SAHU S K，ANAND A. Drug-Drug Interaction Extraction from Biomedical Text Using Long Short Term Memory Network[J]. Journal of Biomedical Informatics，2017：S1532046418301606.