

基于影响图的传球训练策略的实现

张润梅

(安徽建筑工业学院 电子与信息工程学院, 安徽 合肥 230022)

摘要:影响图是决策问题的图形表示,利用影响图可以有效地解决智能 Agent 的行为选择,而传球问题就是球员 Agent 选择传球速度和角度的问题。建立了一个基于影响图的传球模型,并根据该模型创建一个评价传球效果的效用函数;然后利用由教练程序改写的记录员程序来记录数据,并进行统计得到效用函数中的概率值;最后通过训练得到效用函数中系数的值,并利用此效用函数作传球训练。训练结果表明该效用函数是可信的。

关键词:机器人足球;影响图;训练;传球

中图分类号: TP18

文献标识码: A

文章编号: 1673-629X(2008)12-0245-03

Implementation of Strategy of Passing Based on Influence Diagrams

ZHANG Run-mei

(School of Computer and Information Engineering, Anhui Institute of Architecture & Industry, Hefei 230022, China)

Abstract: Influence diagram is a diagram of decision-making, using influence diagrams can solve the behavior choice of intelligence agent effectively, while the question of passing help the agent choose the speed and angle. In this paper, first build a model of passing and a utility function which can evaluate passing decision. Second, data are recorded by recorder program which is recomposed by coach program and then calculate these data to obtain the probability of utility function. In the end, the values of coefficients in utility function gain by training, and then use the utility function to train the team. The training result indicates that the utility function is credible.

Key words: robocup; influence diagrams; training; passing

0 引言

在 Robocup 比赛中有很多因素都是不确定的,这是由于赛场上环境的不确定性和 Agent 的自身状态的不确定性造成的。目前,贝叶斯网络被认为是解决不确定性问题的有效方法,而影响图是对贝叶斯网络的扩展,是在贝叶斯网络的基础上添加了决策节点和效用节点^[1]。利用影响图能有效地解决智能 Agent 建模和决策问题,而传球问题就是球员 Agent 选择传球速度和角度的问题。

1 基于影响图的决策

影响图 (Influence Diagrams, IDs) 是 Howard 和 Matheson 于 1984 年提出的,它是一种将节点以树的形式表示的有向无环图^[2],可以表示为: $G = \{N, A\}$, 其中 N 为节点集, A 为弧集。一般情况下,影响图由

以下三种类型的节点构成:决策节点、随机节点、效用节点。其中,决策节点表示决策者能够选择的所有可能的行为;随机节点表示和决策相关的随机变量,它和一个条件概率表相联系;效用节点表示最大化的效用,它和一个效用函数相联系。

影响图中指向不同类型节点的边也具有不同含义:指向随机节点的边,表示概率依赖关系;指向效用节点的边,表示非概率的函数依赖关系;指向决策节点的边,表示一种偏序关系(表示决策之前必须知道的那些变量)。设 $A = \{a_1, a_2, \dots, a_n\}$ 是一个相互排斥的行为集合, H 是决策行变量。效用函数 $U(A, H)$ 为每个行为变量和决策变量的组合状态产生的效用。基于影响图的决策是为每个决策节点寻找一个最优决策,就可以通过计算使期望效用最大的行为来实现^[3],即:

$$EU(a) = \sum_H U(a, H) * p(H | a).$$

2 基于影响图的传球模型的建立

2.1 传球问题的模型表示

在 Robocup 比赛中,场上有 11 个我方球员和 11 个对方球员,这些球员的状态和球的状态构成了球场

收稿日期:2008-03-28

基金项目:安徽省自然科学基金(2006KJ036B),安徽建筑工业学院青年基金(2005jq1144)

作者简介:张润梅(1971-),女,副教授,博士研究生,研究方向为 Agent 技术和机器学习。

上的全部决策信息^[4]。但是,在进行传球决策时这些信息中有很大一部分影响不大,比如离传球球员较远的球员的信息。为了简化问题,暂且不考虑对手的信息以及和传球动作无关的队友的信息。因此,球场上只剩下传球球员、接球球员和球,那么传球的场景可以如图 1 所示。

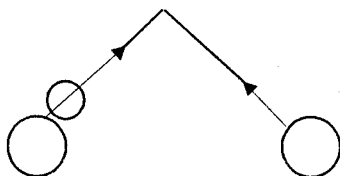


图 1 传球场景

由图 1 的传球的场景可以看出,球在传球球员的脚下,因此,可以将球(ball)和传球球员(passer)视为同一点。无论球和接球球员在场上的什么位置,都可以以球(或传球球员)为坐标原点,以球和接球球员的连线为横坐标建立直角坐标系^[5-7]。

在该图中,随机节点是:球速、球的角度、接球球员 x 方向的速度、接球球员 y 方向的速度、球与接球球员之间的距离以及球与球员之间的夹角;决策节点仍然是:传球决策,取值为直传、传点和保持; U 为传球决策的效用函数。

在做传球决策时,并不需要知道是直传,还是传点,而只需要知道以什么样的速度和角度将球踢出去,因此该图中的决策节点也需要修改。这里将传球决策的取值改为由传球速度和角度组成的二元组:(传球速度,传球角度)。

2.2 效用函数的建立

效用函数是一个度量函数,用来表示 Agents 选取某种行为的偏好程度,只要在评价时使用同一个标准,可以采用灵活的形式,可以是线性的,也可以是非线性的。为简单起见,文中选择线性函数。

效用函数与 Agents 可能选取行为的影响因素的概率相结合,得到每个可能行为的效用,然后选择这些效用中的最大效用或较大效用所对应的行为为决策行为。由图 2 可知影响因素有:球速(Vb)、球的角度(Ab)、接球球员 x 方向的速度($Velx$)、接球球员 y 方向的速度($Vely$)、球与接球球员之间的距离($dist$)以及球与球员之间的夹角(Abr)。那么可以建立效用函数:

$$U = K_1 * Vb * P(Vb) + K_2 * Ab * P(Ab) + K_3 * Velx * P(Velx) + K_4 * Vely * P(Vely) + K_5 * dist * P(dist) + K_6 * Abr * P(Abr)$$

其中, $P(Vb)$ 、 $P(Ab)$ 、 $P(Velx)$ 、 $P(Vely)$ 、 $P(dist)$ 、 $P(Abr)$ 分别为球速、球的角度、接球球员 x 方向的速

度、接球球员 y 方向的速度、球与接球球员之间的距离的概率; K_1 、 K_2 、 \dots 、 K_6 分别为其系数,代表它们所对应的项对传球决策的影响程度。

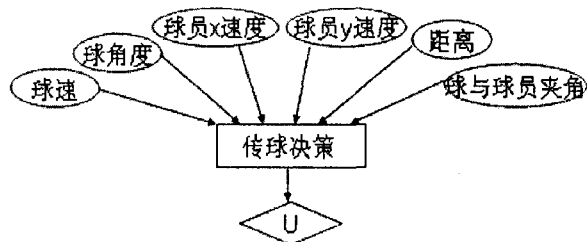


图 2 基于影响图的传球模型的建立

3 实验数据的收集和统计

设计思想是首先得到在某传球时刻的球的速率、球的方向、接球球员的速度(含 x 和 y 两个方向的速度)、球到接球队员的距离和接球球员和球之间的夹角。如果能得到 n 组这样的数据,再将数据分段,统计出 n 组数据中每个段有 m 个数据($m < n$),那么出现在此段中的概率可以表示为 m/n 。

文中以教练程序底层作为基础编写一个类似教练的程序,专门用来记录比赛中传球时球的速率、球的角度、接球球员的速度、接球球员的角度、接球球员到球的距离以及接球球员和球之间的夹角。

程序中,以 BasicCoach 为基类创建 TMPass 类,添加了变量 logBall(记录球的信息以便观察)、logOwnTeam(记录我方球队每个球员在每个周期的信息以便观察)、logBallfile(记录球的信息以供后面的统计程序使用)、logOwnfile(记录我方球队每个球员在每个周期的信息以供后面的统计程序使用)。修改成员函数 logPerformance、training、setNewScenario 等。其中,最主要的是 logPerformance,它是用来记录数据的,其核心内容如下:

```
logBall << "epoch: " << m_ iCurrentEpoch << endl;
logBall << "BallSpeed: " << WM->getBallSpeed() << " "
<< "BallAng: " << WM->getBallDirection() << " "
<< "BallPosition: " << WM->getBallPos() << endl;
logBallfile << m_ iCurrentEpoch << " " << WM->
getBallSpeed() << " " << WM->getBallDirection() << " "
<< WM->getBallPos() << endl;
logOwnTeam << "epoch: " << m_ iCurrentEpoch <<
endl;
for( i = OBJECT_ TEAMMATE_ 1 ; i <= OBJECT_
TEAMMATE_ 11 ; i++ )
{
logOwnTeam << "OBJECT_ TEAMMATE_ " << i - 60
<< ":" << " ";
logOwnTeam << "Speed: " << WM->getGlobalVelocity
```

```

((ObjectT)i) << " " << "Position: " << WM->getGlobal-
Position((ObjectT)i) << endl;
logOwnfile << m_iCurrentEpoch << " " << i-60 <<
" " << WM->getGlobalVelocity((ObjectT)i) << " " <<
WM->getGlobalPosition((ObjectT)i) << endl;
}

```

文中用 Visual C++ 6.0 编写了一个统计程序。该统计程序有如下功能:能够设定球的速率等因素的步长,并根据步长初始化列表;从文件(Ball_file 和 Own_file)中读取数据;统计数据个数并记录数据总数;计算数据的概率;将统计好的概率输出到文件以备观察和使用。

4 训练

在一场训练中,首先随机产生接球球员的速度(x 和 y 两个方向的速度)和位置以及接球球员到球的距离和角度^[8],然后在这几个量不变的情况下,改变球的速率和角度,记录下在不同速率和角度的情况下接球球员接到球所需要的周期。

这里规定接球球员在 50 个周期内如果没有接到球就认为失败,用 50 减去接球所需要的周期数为该状态下的效用值。对比前面建立的效用函数,已经知道了概率和效用值,只要随机训练六次就可以得到六个不同状态下的方程,就能求出一组系数。但随着训练的次数的增加就会得到很多组系数值,再取这些值的方差(或协方差)作为最后系数的值。

文中仍然以 UVA 球队程序为底层为基础来编写程序,图 3 是训练的流程图。

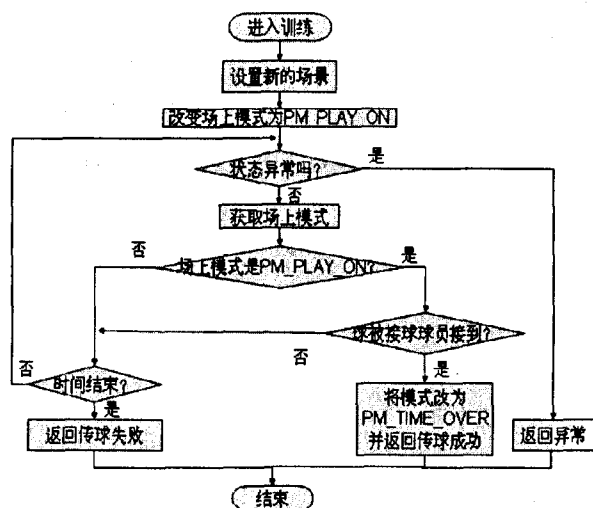


图 3 训练的流程图

在 TMPass 类中,由于要读入和保存前面已经得到的数据段和概率,在此类中设置了六个二维数组(第一维长度均为 100,第二维长度均为 2),并设置 6 个整

型变量来保存数据段和概率的长度(例如,当球的速率的步长设置为 0.1 时,可分成 27 段,这里只用了 100 位中的前 27 位)。当然还要定义在效用函数中涉及到的球的速率(m_BallInitSpeed)、球的角度(m_BallInitAng)、接球球员 x 方向的速度(m_ReceiverxSpd)、接球球员 y 方向的速度(m_Receiveryspd)、接球球员到球的距离(m_DistToReceiver)和接球球员和球之间的夹角(m_AngleToReceiver)。另外,为了方便在设置场景时使用,还定义了球和接球球员的位置(m_PosBall、m_PosReceiver);为了判断接球球员在 50 个周期内是否接到了球,定义 bool 型变量 m_bPassSuccess;为了记录在一次传球训练中接球球员接到球所花费的周期数,定义变量 m_iInterceptCycles。

对该类的改造除了上面变量的增加之外,还增加了以下成员函数:initArray、getMaxUa、getBallSpeed,修改了以下成员函数:training、setNewScenario、logHeader、logPerformance,并修改了构造函数 TMPass。

经过 500 次训练后,得到系数 K 的值分别为 $K_1 = 1046.82$ 、 $K_2 = 4234.29$ 、 $K_3 = -852.668$ 、 $K_4 = -1597.29$ 、 $K_5 = -14.6409$ 、 $K_6 = -2961.87$ 。再将这六个值代入 getMaxUa 函数里的效用函数中,并将文件 isUseUa.conf 中的值改为 yes。这样运行训练程序来观察得到的效用函数的决策效果,部分如表 1。

表 1 效用函数的决策效果

场 次	随机产生的接球球员的状态				决策信息		所花周
	Velx	Vely	dist	Abr	Vb	Ab	期数
1	0.394157	-0.526166	9.82061	-57.3633	2.1	-45	1
2	-0.678025	-0.23722	6.16524	-46.4812	1.6	-60	1
3	-0.607863	-0.594301	15.7683	11.2819	2.1	20	1
4	-0.425247	-0.311052	18.7315	38.4954	1.8	20	5
5	-0.676972	-0.162583	11.0058	-46.6265	2.3	-50	3
6	-0.40893	0.00142406	14.3089	-51.2114	1.6	-40	2
7	-0.346748	-0.0701235	18.4716	-18.1179	1.8	-70	20
8	-0.104668	-0.57321	14.7651	55.7778	2.1	75	1
9	0.344264	0.3991	12.8178	27.7377	1.6	70	12
10	-0.0248758	-0.31214	17.0034	49.2851	2.4	55	1

5 结束语

从上面的系数 K 值中发现:

① 凡是传球球员所能够改变的项的系数均为正数(如 K_1 和 K_2),凡是传球球员无法改变的项的系数均为负数(如 $K_3 \sim K_6$)。

② 从训练程序记录下来的数据来看,对传球决策的效果(接球球员接到球的周期数)影响比较大的项是角度;这一点可以从系数值上体现出来,角度项系数(K_2 和 K_6)的绝对值比其它项系数的绝对值大。

从上面两点来看,得到的系数值是可信的,但同时

(下转第 250 页)

0.9, $C_0 = 0.95$, 初始的正负样本集的大小均为 50 组。为了便于比较, 采用同样的训练样本和测试样本分别对 SVDD 算法和文中算法进行仿真, 试验结果见表 1。

表 1 试验结果的比较

算法	σ^2 值	0.0005	0.001	0.008	0.01	0.05	0.1	0.3	0.6	1
SVDD	检测率(%)	92.86	75.00	57.14	50.00	39.29	39.29	39.29	-	-
	误报率(%)	62.86	57.14	18.57	22.86	38.86	32.86	32.86	-	-
	操作时间(s)	110.4	104.1	135.3	321.9	186.9	177.9	156.2	-	-
文中算法	检测率(%)	-	-	-	42.86	71.43	96.43	96.43	96.43	1
	误报率(%)	-	-	-	22.86	12.86	5.71	2.86	4.29	12.86
	操作时间(s)	-	-	-	42.39	24.23	21.48	18.4	19.1	19.2

试验结果表明无论 σ^2 取何值, 采用文中算法的操作时间都不到 SVDD 的 1/8, 且当 σ^2 取 0.3 时, 取得了较高的检测率和较低的误报率, 表明此算法的可行性和有效性。而采用 SVDD 方法, 当 σ^2 取比较小的值时, 虽然可以达到不错的检测率, 但这种效果是建立在较多的训练样本都作为支持向量的基础上, 训练时间比较长, 误报率也比较高。

利用文中算法对该检测点每天 14:30~18:30 时间段进行建模, 共 24 天的数据, 选取前 19 天的数据作为训练样本, 采样间隔为 5min, 即训练样本为 931 组, 其中有事件样本为 49 组; 测试样本为 245 组, 其中有事件样本为 20 组。试验结果证明, 当取 $C = 1$, $C_0 = 0.9$, $\sigma^2 = 0.3$ 时, 分时段建模的检测率为 1, 误报率为 0, 操作时间为 314.8s, 达到最好的检测效果。

多次试验证明, 多时间段建模的检测结果大多优于不分时间段的整体建模检测结果, 但是多时间段建模所需的模型比较多, 比较适合于复杂路段, 以提高检测效果。

(上接第 247 页)

也发现系数中的不和谐之处。 K_3 和 K_4 是接球球员在 x 和 y 两个方向上的速度, 它们对传球的影响应该不分伯仲, 但从数据上来看, 它们的值有一定的差别。

文中得到的效用函数是在没有对手的情况下得到的, 因此还有很大的局限性。下一步所要做的工作首先是要对手进行建模, 而建模所需要的对手数据是很难得到的^[9], 可以尝试从 log 文件中分析对手信息, 然后进行效用函数的计算。

参考文献:

- [1] 张润梅, 王浩, 姚宏亮, 等. 影响图及其在 Robocup 中的应用[J]. 系统仿真学报, 2005, 17(1): 134-137.
- [2] 姚宏亮. 动态多智能体决策问题研究[D]. 合肥: 合肥工业大学, 2006: 257-289.
- [3] 张润梅. 基于影响图的智能 Agent 学习及其在 Robocup 中

3 结束语

高速公路事件自动检测是高速公路管理的一个重要环节, 文中提出把支持向量数据描述应用到高速公路的事件检测上, 这种算法不受输入数据顺序的影响, 训练速度快, 泛化能力强, 试验结果表明该算法具有较好的检测结果。另外对一些复杂路段进行分时段建模也是一个值得进一步研究的方向。对于 C_0 的确定可以设为动态值, 使其随着样本数目的增加而增大。也可对不同原因引起的交通事件进行细分类, 即多类分类识别问题, 可采用一对一、一对多方法进行实现。

参考文献:

- [1] 徐吉谦. 交通工程总论[M]. 北京: 人民交通出版社, 1996.
- [2] 姜紫峰, 刘小坤. 基于神经网络的交通事件检测算法[J]. 西安公路交通大学学报, 2000, 20(3): 67-69.
- [3] 周伟, 罗石贵. 基于模糊综合识别的事件检测算法[J]. 西安公路交通大学学报, 2001, 21(2): 70-73.
- [4] Tax T M J, Duin R P W. Support Vector Domain Description[J]. Pattern Recognition Letters, 1999, 20(11-13): 1191-1199.
- [5] 肖健华. 智能模式与识别方法[M]. 广州: 华南理工大学出版社, 2006.
- [6] 郑晓星, 吴今培. 基于支持向量数据描述的数据约简[J]. 计算机应用, 2007, 20(2): 74-76.
- [7] Klaus-Robert M. Intelligent Data Analysis Group[EB/OL]. [1998-08]. <http://ida.fhg.de/projects/bench/benchmarks.htm>.
- [8] Petty K. The Freeway Service Patrol Project and the I-880 Database[EB/OL]. [1995-08-14]. <http://ipa.eecs.berkeley.edu/pettyk/FSP/>.

的应用[D]. 合肥: 合肥工业大学, 2004: 257-289.

- [4] Tuyls K, Maes S, Manderick B. Q-learning in Simulated Robotic Soccer Large State Spaces and Incomplete Information[DB/OL]. 2002. <http://como.vub.ac.be:8080/Publications/uploads/1/icmla02.ps>.
- [5] 张晓勇, 彭军. RoboCup 中传球策略的实现[J]. 计算机工程, 2004, 30(23): 123-124.
- [6] 郭博, 程家兴. RoboCup 仿真组的传球策略[J]. 计算机技术与发展, 2006, 16(2): 129-131.
- [7] 于磊, 王浩, 王骋. RoboCup 中传球策略研究[J]. 计算机工程与应用, 2004(28): 59-61.
- [8] 张家旺, 韩光胜, 张伟. C5.0 算法在 RoboCup 传球训练中的应用研究[J]. 计算机仿真, 2006, 23(4): 132-134.
- [9] Ledezma A, Aler R, Sanchis A, et al. Predicting opponent actions by observation[DB/OL]. 2004. <http://www.springer-link.com/content/ec4rq5k3vd278p62/fulltext.pdf>.