

基于神经网络的 RoboCup 进攻策略

刘 亮 李龙澍

(安徽大学计算机科学与技术学院,合肥 230039)

(安徽大学 计算智能与信号处理教育部重点实验室,合肥 230039)

E-mail: alexdream@163.com

摘 要 机器人足球赛 Robocup(Robot World Cup)是国际上规模最大且影响最为广泛的机器人足球赛事。机器人足球比赛已经成为当前人工智能研究的热点之一,作为多 Agent 系统的一个理想的试验平台,它涉及到了多个技术领域。文章应用神经网络技术解决 Robocup 仿真组比赛的进攻策略问题,对射门底层技术进行了有效优化。我们把这种策略与以往的基于逻辑准则的策略进行了比较。基于神经网络的优化射门策略在很多情况下会做出更好的决定。

关键词 神经网络 策略优化 RoboCup 射门技巧

文章编号 1002-8331-(2005)36-0044-04 文献标识码 A 中图分类号 TP183

A Offensive Strategy Based on Neural Network in RoboCup

Liu Liang Li Longshu

(School of Computer Science and Technology, Anhui University, Hefei 230039)

(Key Lab of IC&SP at Anhui University, Ministry of Education, Hefei 230039)

Abstract: At now, RoboCup (Robot World Cup) is the most important Robot Soccer game in the world. The Robot Soccer has been become the research center of many Artificial Intelligence researchers. As the ideal testing platform of Multi-Agent Systems, it is involved in many technical domains. We use neural network to make the offensive strategy in RoboCup and optimize shooting skill, and compared the neural network based strategy and an usual rule based strategy. The neural network based strategy makes better decisions in most cases than the rule based strategy.

Keywords: Neural Network, strategy optimization, RoboCup, shooting skill

1 引言

机器人足球的最初想法由加拿大不列颠哥伦比亚大学的 Alan Mackworth 教授于 1992 年正式提出。1993 年, Minoru Asada(浅田埴)、Hiroaki Kitano(北野宏明)和 Yasuo Kuniyoshi 等著名学者创办了机器人足球世界杯赛(Robot world cup soccer games, 简称 RoboCup)。RoboCup 机器人足球赛最重要的目的是检验信息自动化前沿研究、特别是多主体系统^[1]研究的最新成果, 交流新思想和新进展, 从而更好的推动基础研究和应用基础研究及其成果转化。它的最终梦想是能在 2050 年发展出能打败人类足球运动员的机器人足球队。

我们的研究是基于 RoboCup 仿真组平台上的, 在 RoboCup 仿真组比赛中, 每个队由 11 个全自治的智能体组成, 它们都连接到服务器上, 在仿真组比赛中, 没有真正的机器人参与, 比赛场面可以通过播放器即时播放或做成记录文件随时播放, 播放的场面效果很像一场真实的足球赛或是电子游戏中的足球。每个仿真程序可以有自己独立的策略和技术特征^[2]。一个队由 11 个这样的程序组成, 在由网络连接的数台计算机上进行比赛。

机器人足球作为仿真的实验平台, 其运行时的场上状态千变万化, 也就是说由于状态空间的过于庞大以及反馈延迟过大, 不融入任何先验知识来设计决策几乎不可能实现。许多非常简单的概念如: 带球、截球、射门、传球等要让智能体从与环

境的交互当中总结出来, 而且组织成知识都是十分困难的。所以智能体本身具有的学习和适应能力, 能够自己从环境中获取知识, 成为决策系统是否成功的关键所在。目前流行的自适应学习的算法有 BP、动态规划、Q-learning、遗传算法等^[3]。而神经网络在优化如射门、传球等底层技术中显现了巨大的潜力^[4,5]。下面的章节将详细地介绍在实验中所应用的技术。第二节论述了 BP 神经网络的基本知识。第三节论述了有关 BP 神经网络在底层技术中应用方法及性能对比。

2 BP 神经网络

2.1 神经网络理论及其典型网络模型

人工神经网络是由大量简单的基本元件—神经元相互联结, 模拟人的大脑神经处理信息的方式, 进行信息并行处理和非线性转换的复杂网络系统^[6,7]。人工神经网络处理信息是通过信息样本对神经网络的训练, 使其具有人的大脑的记忆、辨识能力, 完成各种信息处理功能。人工神经网络具有良好的自学习、自适应、联想记忆、并行处理和非线性转换的能力, 避免了复杂数学推导, 在样本缺损和参数漂移的情况下, 仍能保证稳定的输出。

按照网络的拓扑结构和运行方式, 神经网络模型分为前馈多层式网络模型、反馈递归式网络模型、随机网络模型等。目

基金项目: 国家自然科学基金资助项目(编号: 60273043); 安徽省高校学科拔尖人才基金(编号: 05025102); 安徽省自然科学基金(编号: 050420204)

作者简介: 刘亮(1981-), 男, 安徽合肥人, 安徽大学硕士研究生, 主要研究方向为智能体协作和机器人足球。李龙澍(1956-), 男, 安徽大学教授, 博士生导师, 主要研究方向为智能软件 and 知识工程。

前在模式识别中应用成熟较多的模型是前馈多层式网络中的BP反向传播模型,它也是向前网络的核心部分,并体现了人工神经网络最精华的部分。其模型结构如图1。

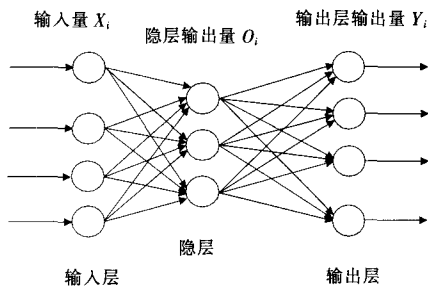


图1 典型BP网络结构模型

2.2 BP神经网络基本原理

BP网络模型处理信息的基本原理是:输入信号 X_i 通过中间节点(隐层点)作用于输出节点,经过非线性变换,产生输出信号 Y_k ,网络训练的每个样本包括输入向量 X 和期望输出量 t ,网络输出值 Y 与期望输出值 t 之间的偏差,通过调整输入节点与隐层节点的联接强度取值 W_{ij} 和隐层节点与输出节点之间的联接强度 T_{jk} 以及阈值,使误差沿梯度方向下降,经过反复学习训练,确定与最小误差相对应的网络参数(权值和阈值),训练即告停止。此时经过训练的神经网络即能对类似样本的输入信息,自行处理输出误差最小的经过非线性转换的信息。

BP网络主要应用于:

- (1)函数逼近 用输入矢量和相应的输出矢量训练一个网络逼近一个函数。
- (2)模式识别 用一个特定的输出矢量将它与输入矢量联系起来。
- (3)分类比 把输入矢量以所定义的合适方式进行分类。
- (4)数据压缩 减少输出矢量维数以便于传输或储存。

2.3 BP神经网络模型

BP网络模型包括其输入输出模型、作用函数模型、误差计算模型和自学习模型^[6]。

(1)节点输出模型

隐节点输出模型: $O_j = f(\sum W_{ij} \times X_i - \theta_j)$

输出节点输出模型: $Y_k = f(\sum T_{jk} \times O_j - \theta_k)$

f -非线性作用函数; θ -神经单元阈值。

(2)作用函数模型

作用函数是反映下层输入对上层节点刺激脉冲强度的函数又称刺激函数,一般取为(0,1)内连续取值 Sigmoid 函数:

$$f(x) = \frac{1}{1 + e^{-x}}$$

(3)误差计算模型

误差计算模型是反映神经网络期望输出与计算输出之间误差大小的函数:

$$E_p = \frac{1}{2} \times \sum (t_{pi} - O_{pi})^2$$

t_{pi} - i 节点的期望输出值; O_{pi} - i 节点计算输出值。

(4)自学习模型

神经网络的学习过程,即连接下层节点和上层节点之间的权重矩阵 W_{ij} 的设定和误差修正过程。BP网络有师学习方式-需要设定期望值和无师学习方式-只需输入模式之分。自学习

模型为:

$$\Delta W_{ij}(n+1) = \eta \times \Phi_i \times O_j + \alpha \times \Delta W_{ij}(n)$$

η -学习因子; Φ_i -输出节点 i 的计算误差; O_j -输出节点 j 的计算输出; α -动量因子。

2.4 BP网络的训练过程

为了应用神经网络,在选定所要设计的神经网络的结构之后(其中包括的内容有网络的层数、每层所含神经元的个数和神经元的激活函数),首先应考虑神经网络的训练过程。下面用两层神经网络为例来叙述BP网络的训练步骤。

步骤1 用小的随机数对每一层的权值 w 和偏差 b 初始化,以保证网络不被大的加权输入饱和,同时还要进行以下参数的设定或初始化:

- (1)设定期望误差最小值:err_goal。
- (2)设定最大循环次数:max_epoch。
- (3)设置修正权值的学习速率:一般选取 lr=0.01~0.7。
- (4)从1开始的循环训练:for_epoch=1;max_epoch。

步骤2 计算网络各层输出矢量A1和A2,以及网络误差E:

$$A1 = \text{tansig}(w1 * p, b1);$$

$$A2 = \text{purelin}(w2 * A1, b2);$$

$$E = T - A;$$

步骤3 计算各层反向传播的误差变化D2和D1,并计算各层权值的修正值及新的权值:

$$D2 = \text{deltalin}(A2, E);$$

$$D1 = \text{deltalan}(A1, D2, w2);$$

$$[dw1, db1] = \text{learnbp}(p, D1, lr);$$

$$[dw2, db2] = \text{learnbp}(A1, D2, lr);$$

$$w1 = w1 + dw1; b1 = b1 + db1;$$

$$w2 = w2 + dw2; b2 = b2 + db2;$$

步骤4 再次计算权值修正后的误差平方和:

$$SSE = \text{sumspr}(T - \text{purelin}(w2 * \text{tansig}(w1 * p, b1), b2))$$

步骤5 检查 SSE 是否小于 err_goal,若是,则训练结束;否则继续。

以上就是BP网络利用MATLAB神经网络工具箱训练的过程。以上所有的学习规则与训练的全过程,还可以用函数 trainbp 来代替。它的使用同样需要定义有关参数:显示间隔次数、最大循环次数、目标误差和学习速率,在调用 trainbp 函数后,返回训练后的权值、循环训练的总数和最终误差。

2.5 算法优化

由于BP算法采用的是梯度下降法,因而易陷于局部最小并且训练时间较长。在实际应用中,原始的BP算法很难胜任,因此出现了很多的改进算法。BP算法的改进主要有两种途径,一种是采用启发式学习方法,另一种则是采用更有效的优化算法。在神经网络工具箱中,函数 trainbpx()采用动量法和自适应学习速率自适应调整两种策略,从而提高了学习速度并增加了算法的可靠性,动量法降低了网络对于误差曲面局部细节的敏感性,有效地抑制了网络陷于极小;自适应调整学习速率有利于缩短学习时间。

3 BP神经网络在底层技术中应用及效能

3.1 设计理念

在RoboCup仿真组比赛中,底层技术的效能往往直接影响到高层策略能否实现。虽然比赛平台中给我们提供了基本动

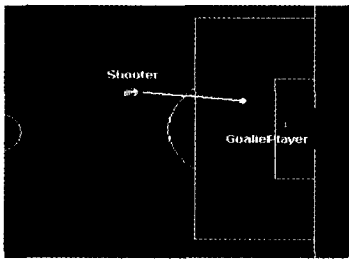


图2 无防守队员时

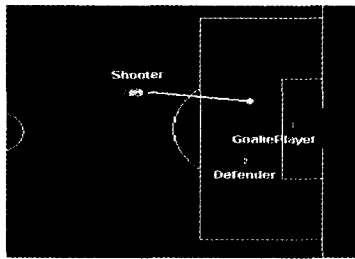


图3 有一个防守队员时

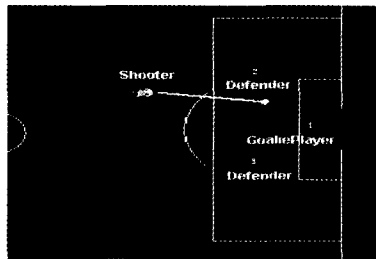


图4 有两个防守队员时

作指令,但由于存在误差,使得直接使用这些指令时无法取得期望的效果。而且比赛中的情况往往十分复杂,无法通过手工编码实现。所以在传球、截球、射门等底层技术中用人工智能技术改进其效率就成为必然。

我们的实验就是基于 BP 网可以对非线性函数逼近的基础上,在射门技术中应用 BP 网络,并与手工编码的效能进行比较,从而提高射门的效率。

实验将分为如图 2、3、4。

在这三种场景中,射门时的情况越来越复杂,球员在决定射门方向时所考虑的因素也越来越多。我们将先采用手工编码的方法,进行离线的射门训练,并把每次射门的相关数据收集起来。之后用这些数据训练神经网络,并把训练后所得的函数写入程序中,再进行射门训练同时采集同等规模的数据。通过对两组数据的比较,观察运用神经网络以后射门效率是否得以提高。

3.2 实验过程

首先,我们用手工编码,在 DoBallMove() 中让球员的射门点随机产生(在+7 与-7 之间)。因为采集的数据将用于神经网络的训练,所以选择的射门点要尽可能的均匀覆盖样本空间。为此我们首先把每种场景再划分为三类,既射门动作分别由 9、10、11 号来完成,这就保证了我们采集的样本将覆盖比赛中前锋射门的所有区域。其次在训练采样时发任意球的位置要尽量做到在所选球员的射门区域中均匀分布。

我们让每个前锋进行十场比赛,这样每个场景有 180 000 个周期可以采样,采集到的射门数据将送入编写好的 MATLAB 程序,进行 BP 网络的训练。在这一步骤中,对所采集的射门数据进行筛选。首先我们将剔除射门失败的数据,其次将输入向量中不发生变化的删除从而减少输入向量的个数(如射门点的 X 坐标始终为 52.5)最后是删除脏数据,这些脏数据可以分为三类:

(1)不完整数据:所采集的数据中缺伐必要的输入向量,如:一些射门成功的样本中由于球员未看见守门员,而缺少守门员站位的数据。

(2)不合理数据:某些时候,一些不应该进的球却射进,这样的数据也应剔除。

(3)相近输入不同输出:当进攻球员于防守球员站位基本相同时(X、Y 坐标小于 0.5),却有很多射门点,应将数据合并取平均值。

删除第一类脏数据的原因显而易见,我们更要强调的是删除第二类 and 第三类脏数据,它们的存在将直接影响到 BP 网络的训练。我们会再网络训练章节中详细阐明此点。

3.3 网络训练

以无人防守场景为例。在手工编码的程序下,记录了 30 场

比赛的数据。共射门 4 181 次,记录中进球 821 次,平均每场进球 27.4 次,射门成功率 19.6%。把射门成功的数据提取出来进行网络训练。

首先去除了无用的输入向量,使输入向量的个数减为 4 个既射门球员与守门员的 X、Y 坐标。输出向量为射门点的 Y 坐标。接着剔除了 46 个不完整的记录(守门员坐标无效),把剩下的 775 个数据用来训练神经网络。

我们建立了一个三层的 BP 网络,在隐含层中设计 10 个神经元单元。训练进行后我们发现,无法达到预订目标,函数不收敛(如图 5)。

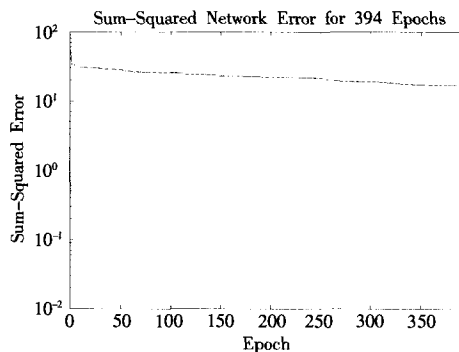


图5 函数不收敛

在调整了隐层的神经元数及训练时间后,依然无法使函数收敛,这使我们意识到可能输入的数据依然有问题。我们又剔除了 37 个不合理的进球数据,可函数还是无法收敛。通过对数据的观察我们终于发现原始数据中存在不少相近输入不同输出的数据,这就无法建立函数。在合并了 127 个这样的数据后,训练结果如图 6。

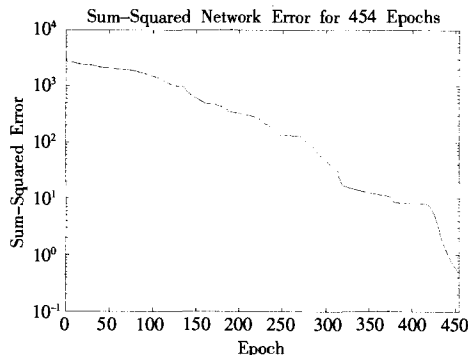


图6 函数收敛

当训练了 454 个周期后,达到预期目标(0.5),此时隐层有 32 个神经元单元,训练结束。我们把训练得出的函数代回原程序。

3.4 测试

我们用经过 BP 神经网络改进后的程序在设定的三个场景中同样数量的射门训练。对比结果如下:

表 1 无后卫

	平均射门次数	平均进球数	进球率
手工编码	139	27.4	19.6%
BP 网络	148	50.3	34.0%

表 2 一后卫

	平均射门次数	平均进球数	进球率
手工编码	132	24.6	18.6%
BP 网络	152	48.4	31.8%

表 3 两后卫

	平均射门次数	平均进球数	进球率
手工编码	135	24	17.7%
BP 网络	149	47.1	31.6%

从表中可知进球率有了很大的提高(约 15%),而平均进球数提高了 90%以上。由此可知应用神经网络之后,射门效能有了很大的提高。

4 结论

通过 BP 神经网络在射门技术改进中的应用,可以看到神经网络技术在改进 RoboCup 底层技术领域中的巨大前景。但 BP 神经网络的训练需要大量的数据,并且对所采集的数据有很高的要求,这就使的采集数据的工作会花费大量的时间,目前只适宜进行离线训练,而无法进行有效的在线训练。

未来要做的工作还很多:首先我们希望用 RBF 网络替代

BP 网络。众所周知,BP 网络用于函数逼近时,权值的调节采用的是负梯度下降法,这种调节权值的方法有它的局限性,即存在着收敛速度慢和局部极小等缺点。而径向基函数(Radial Basis Function,RBF)网络无论在逼近能力、分类能力和学习速度等方面均优于 BP 网络。其次我们希望将这一方法更广泛地应用于各种底层技术中,如传球和抢断。

(收稿日期:2005 年 7 月)

参考文献

1. Peter Stone. Layered learning in Multi-Agent Systems[D]. Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, 1998-12
2. Kraus S. Negotiation and Cooperation in Multi-Agent Environments[J]. Artificial Intelligence, 1997; (9): 79~97
3. 李实, 陈江, 孙增圻. 清华机器人足球队的结构设计与实现[J]. 清华大学学报(自然科学版), 2001; 41(7): 94~97
4. Yang Zeng-guan, Li Long-shu et al. Study on Shooting Skill in RoboCup Simulator League[C]. In: Proceedings of the Second International Conference on Machine Learning and Cybernetics, IEEE Systems, Man and Cybernetics Technical Committee on Cybernetics, IEEE Catalog Number: 03EX693, ISBN: 0-7803-7865-2, 2003-11; 2089~2092
5. Mihal Badjonski, Kay Schroter, Jan Wendler et al. Learning of Kick in Artificial soccer[C]. In: RoboCup-2000, 2000
6. Simon Haykin 著. 叶世伟, 史忠植译. 神经网络原理[M]. 北京: 机械工业出版社, 2004
7. 丛爽著. 面向 MATLAB 工具箱的神经网络理论与应用[M]. 中国科学技术大学出版社, 2003
8. 飞思科技产品研发中心. MATLAB 6.5 辅助神经网络分析与设计[M]. 电子工业出版社, 2003

(上接 2 页)

设敌手用任意的自变量 $x_0 \in Z_{p-1} - X_n$ 按照解密步骤试图解密, 因 $k \in Z_{p-1}^*$ 是可逆元(从而不是零因子), 从上述零因子攻击成功的原因及 $e \equiv kf(x_0) \equiv 0 \pmod{p-1} \Leftrightarrow f(x_0) \equiv 0 \pmod{p-1}$ 可知: 与原取法相比, $k \in Z_{p-1}^*$ 的取法可提高系统抵抗零因子攻击的能力。获证。

还可类似研究基于 (Z_p^*, g) 上 DLP 的公钥加密算法(q 为 $p-1$ 的素因子, p 为 Z_p^* 的一个生成元)。

4 结论

YVZ 算法及其改进算法的基本安全性都是基于 DLP 的难解性, 但后者通过省略不必要的强单向散列函数提高了计算效率, 并通过选择明文 $M \in Z_p^*$, 随机数 $k \in Z_{p-1}^*$ 及 $k' \in Z_{p-1}$ 保持了理想的语意特性。鉴于 Z_{p-1} ($p \geq 5$) 为零因子环, 二者均可遭受零因子攻击, 但后者成功的概率较小, 从而系统具有较高的安全性。

利用改进的加密算法构造安全的组密钥管理方案是我们下一步的研究工作。(收稿日期: 2005 年 9 月)

参考文献

1. Rafaeli S, Hutchison D. A Survey of Key Management for Secure Group Communication[J]. ACM Computing Survey, 2003; 35(3): 309~329
2. Dondeti L R, Mukherjee S, Samal A. Scalable Secure One-to-Many Group Communication Using Dual Encryption[J]. Computer Communications, 2000; 23(7): 1681~1701
3. Hardjono T, Dondeti L R. Multicast and Group Security[M]. Norwood, MA: Artech House, INC, 2003
4. 许勇. 可缩放安全多播通信系统的研究和实现[J]. 计算机工程与应用, 2005; 41(19): 116~122
5. Yi M, Varadharajan V, Zhao W. A Robust and Secure Broadcasting Scheme[C]. In: Proceedings of IndoCrypt 2001, Lecture Notes in Computer Science, Springer Verlag LNCS Series, 2001
6. Chaddoud G, Varadharajan V. Efficient secure group management for SSM[C]. In: 2004 IEEE International Conference on Communications, Piscataway, NJ, USA, 2004
7. Stinson D R 著. 冯登国译. 密码学原理与实践[M]. 第 2 版, 北京: 电子工业出版社, 2003
8. Menezes A J, Oorschot P van, Vanstone S. Handbook of Applied Cryptography[M]. Boca Raton: CRC Press, 1997
9. Mao W 著. 王继林等译. 现代密码学理论与实践[M]. 北京: 电子工业出版社, 2004
10. 柯召, 孙琦. 数论讲义(上册)[M]. 第 2 版, 北京: 高等教育出版社, 2001