

神经网络在足球比赛中的胜负预测

付 裕

陕西国防工业职业技术学院 陕西西安 710300

摘 要:因人工神经网络具有极强的非线性逼近能力,所以在诸多领域中得到了广泛的应用。足球作为世界第一大运动,相关专家及球迷一直热衷于对其比赛的结果进行预测。本文提出了一种使用神经网络来预测比赛胜负的新方法,并用 R 语言进行实现并在相关数据集上进行测试,通过测试发现所提出的新预测方法具有一定的可靠性。

关键词:神经网络;R 语言;球赛结果预测

人们一般用足球是圆的来形容足球比赛结果的变换莫测。但是随着足球运动的快速发展,对足球比赛结果的预测已经成为一个热门的研究方向。通常的预测方法是相关专家结合对战双方的历史交战记录,以及两支球队当前的状态来进行预测。这种预测方法过于依赖历史交战记录,而球队当前的状态对将要进行的比赛的结果的影响程度也无法定量的描述。

一场足球比赛中能够体现出若干的数据特征,诸如射正球门的次数、前场任意球的个数、角球个数、进球数,可以将一场球赛的结果看成是这些特征共同的作用,每一个特征与结果之间都有一定的相关性,这种相关性并非一种简单的线性关系,而是一种非线性关系。神经网络事实上是一种由输入到输出的非线性映射。所以应用神经网络的方法非常适合处理这种足球比赛结果预测的非线性问题。

本文为了克服传统预测方法的不准确性,采用神经网络的方法对足球比赛的结果进行预测。以射正球门的次数、前场任意球的个数、角球个数、进球数这四个特征作为神经网络的输入,以比赛结果作为神经网络的输出。本文通过对西班牙甲级联赛巴塞罗那俱乐部在 2015-2016、2016-2017 两个赛季的数据为训练集训练神经网络,并用所训练的网络对球队在 17-18 赛季的成绩进行了预测。通过试验发现,所训练的神经网络模型具有可靠性。

神经网络模型的训练

1 输入数据

本文选用数据来自于专业足球数据查询平台创冰 data 中西甲联赛足球数据,所选用的数据具有可靠性。本文统计了 2015-2016、2016-2017 两个赛季中巴塞罗那俱乐部 24 场比赛中的射正球门的次数、前场任意球的个数、角球个数、进球数及比赛结果(胜、平、负)等特征数据。部分数据集如表 1 所示。

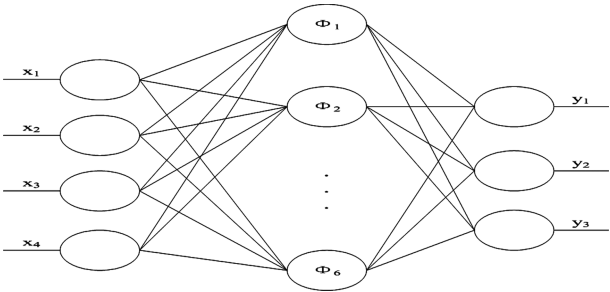
表 1 巴塞罗那球队比赛部分特征数据

射正	任意球	角球	进球数	胜负	类别标签
4	8	4	1	胜	0 0 1
12	5	15	1	胜	0 0 1
5	8	6	1	平	0 1 0
4	1	2	2	平	0 1 0
7	12	10	1	负	1 0 0
3	7	4	1	负	1 0 0

2 神经网络模型

本文所要训练的 mlp 神经网络[1]模型具有 3 层结构,分别是输入层、隐层及输出层。其中输入层具有四个神经元,每一个神经元分别对应着射正球门次数、前场任意球的个数、角球个数、进球数这四个特征;输出层具有三个神经元分别对应着比赛结果的胜、平、负;由于隐层神经元的个数到目前为止还

没有一种选取的标准只能在训练中根据模型的拟合程度进行调整,本文为隐藏选取 6 个神经元。网络结构如下图所示,其中 x1-x4 为输入层,y1-y3 为输出层。



本文应用的神经网络结构图

3 R 语言编程

确定好输入数据与神经网络模型之后,需要编程实现测试,本文使用 R 语言进行编程。使用 read.csv() 函数读入数据;使用 decodeClassLabels() 函数将数据集中的第 5 列即结果列进行类别标签。输出的 3 种结果,胜、平、负对应的类别标签分别为 0 0 1,0 1 0,1 0 0;使用 splitForTrainingAndTest() 函数将数据集按比例划分为训练集与测试集,本文把比例设置为 0.25,splitForTrainingAndTest() 函数返回的四个元素分别是训练的评价集、结果集、测试的评价集、结果集;使用 normTrainingAndTestSet() 函数将训练集与测试集的数据进行标准化到[0, 1]。使用 mlp() 对网络进行训练。最后利用模型预测,使用 confusionMatrix() 函数生成混淆矩阵,观察预测精度。

4 结果分析及结论

通过编程测试,最后得到的混淆矩阵如表 2 所示。通过分析矩阵的第一行可知,有两场球赛的结果原本是胜,但是模型将一场比赛的结果预测成了负;分析第二行可知有六场球赛的结果原本是平,模型将一场结果预测成了胜,一场预测成了负;分析第三行可知对于比赛结果为负,模型的预测是相当准确的。混淆矩阵反映了模型的整体预测精度是可靠的。

表 2 混淆矩阵

	胜	平	负
胜	1	0	1
平	1	4	1
负	0	0	1

参考文献:

[1] 蒋宗礼. 人工神经网络导论[M]. 北京: 高等教育出版社,2001.8(2013.5 重印).

作者简介:付裕(1987-),陕西西安人,助教,研究方向:数据挖掘,电气自动化。