



(12) 发明专利申请

(10) 申请公布号 CN 104899413 A

(43) 申请公布日 2015. 09. 09

(21) 申请号 201510174490. 1

(22) 申请日 2015. 04. 14

(71) 申请人 湘潭大学

地址 411105 湖南省湘潭市雨湖区羊牯塘
27 号湘潭大学

(72) 发明人 程戈 张振宇 李强 李聪 张云
何春辉

(51) Int. Cl.

G06F 19/00(2011. 01)

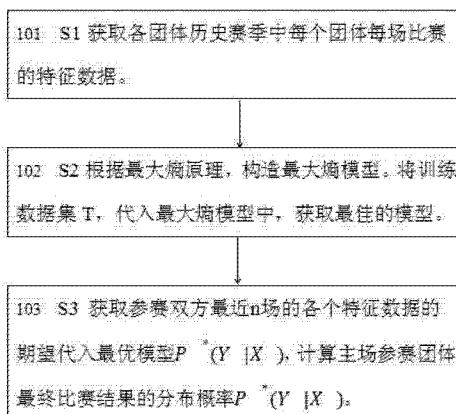
权利要求书1页 说明书3页 附图1页

(54) 发明名称

基于最大熵模型的团体比赛结果预测方法

(57) 摘要

本发明涉及计算机数据挖掘领域,尤其涉及一种基于最大熵模型的团体比赛结果预测方法。该方法的具体步骤如下:S1 获取各团体历史赛季中每场比赛参赛双方的特征数据。S2 根据最大熵原理,构造最大熵模型。将训练数据集 T,代入最大熵模型中,获取最佳的模型。S3 获取参赛双方最近 n 场的各个特征数据的期望代入最优模型 $P^*(Y|X)$, 计算主场参赛团体最终比赛结果的分布概率 $P^*(Y|X)$ 。本发明技术方案给赛事关注者提供一个科学的比赛结果预测。



1. 基于最大熵模型的团体比赛结果预测方法,其特征在于,该方法包括以下步骤:

S1 获取各团体历史赛季中每场比赛参赛双方的特征数据;

S2 根据最大熵原理,构造最大熵模型;

将训练数据集 T,代入最大熵模型中,获取最佳的模型;

S3 获取参赛双方最近 n 场的各个特征数据的期望代入最优模型 $P^*(Y|X)$,计算主场参赛团体最终比赛结果的分布概率 $P^*(Y|X)$ 。

2. 根据权利要求 1 的方法,其中所述步骤 S1 所述的特征数据,其特征在于,特征数据包括两种:1、影响比赛最终结果的内在因素的数据和外在因素的数据,内在因素是指团体自身内部的因素,例如,团体主教练的战术体系,外在因素是指非团体自身内部的因素,2、主场团体最终比赛结果的数据,将特征数据的总体作为训练数据集 T。

3. 根据权利要求 1 或 2 的方法,其中所述步骤 S2 中所述的构造最大熵模型并求解获得最佳模型,其特征在于,将每场历史比赛参赛的两个团体,主场团体的最终比赛结果作为随机事件 Y,两个团体各自的内在数据和外在数据特征作为随机事件 X,构造条件概率分布 $P(Y|X)$;条件概率 $P(Y|X)$ 的熵为:

$H(p) = -\sum_{x,y} \tilde{P}(x) P(y|x) \log P(y|x)$;根据最大熵原理,将训练数据集 T 代入模型,求出最优的模型 $P^*(Y|X)$ 。

基于最大熵模型的团体比赛结果预测方法

技术领域

[0001] 本发明涉及计算机数据挖掘领域,尤其涉及一种基于最大熵模型的团体比赛结果预测方法。

背景技术

[0002] 随着互联网的发展和人们对于各种体育竞技赛事的关注度的提高,像欧洲足球五大联赛、美国职业橄榄球联赛、美国职业篮球联赛等团体体育的竞技赛事已经不只受到当地居民的关注,这些赛事已经普及全球。各大型博彩公司、大型网站纷纷对赛事的结果进行分析与预测,比赛关注者根据自己的主观偏好和网站给出的分析对一场比赛的输赢预测。但是由于影响团体实力的因素众多,而且每一个因素影响的程度又不尽相同,因此预测的正确率非常低。处于大数据时代的今天,我们更应该根据已有的历史比赛数据来进行科学的合理的推导,而不能仅仅靠我们的情感和少量的数据对一场比赛的输赢预测。

[0003] “熵”是信息量的度量方法,它表示某一事件出现的消息越多,事件发生的可能性就越小,数学上就是概率越小。最大熵原理指出,当我们需要对一个随机事件的概率分布进行预测时,我们的预测应当满足全部已知的条件,而对未知的情况不要做任何主观假设。在这种情况下,概率分布最均匀,预测的风险最小,因为这时概率分布的信息熵最大,所以人们称这种模型叫“最大熵模型”。我们知道,决定比赛最终结果的因素可能有几十甚至上百种,我们对体育竞技中已经比赛过的赛事作为已知的数据进行统计分析,然后通过利用最大熵方法找到一个同时满足成千上万种不同条件的最优模型,从而对比赛最终结果进行预测。

发明内容

[0004] 本鉴于此,本发明的主要目的在于提供了一种基于最大熵模型的团体比赛结果预测方法。

[0005] 为了得到上述的目的,本发明的技术方案实现的步骤:

S1 获取各团体历史赛季中每场比赛参赛双方的特征数据。

[0006] S2 根据最大熵原理,构造最大熵模型。将训练数据集 T,代入最大熵模型中,获取最佳的模型。

[0007] S3 获取参赛双方最近 n 场的各个特征数据的期望代入最优模型 $P^*(Y|X)$, 计算主场参赛团体最终比赛结果的分布概率 $P^*(Y|X)$ 。

[0008] 进一步,作为一种优选方案,步骤 S1 所述的特征数据包括两种:

1、影响比赛最终结果的内在因素的数据和外在因素的数据,内在因素是指团体自身内部的因素,例如,团体主教练的战术体系。外在因素是指非团体自身内部的因素。2、主场团体最终比赛结果的数据,将特征数据的总体作为训练数据集 T。

[0009] 进一步,作为一种优选方案,步骤 S2 中所述的构造最大熵模型,并求解获得最佳

模型的具体方法：

将每场历史比赛参赛的两个团体，主场团体的最终比赛结果作为随机事件 Y ，两个团体各自的内在数据和外在数据特征作为随机事件 X ，构造条件概率分布 $P(Y|X)$ 。条件概率 $P(Y|X)$ 的熵为： $H(p) = -\sum_{x,y} \widetilde{P}(x)P(y|x) \log P(y|x)$ ；根据最大熵原理，将训练数据集 T 代入模型，求出最优的模型 $P^*(Y|X)$ 。

[0010] 本发明所提供了一种基于最大熵模型的团体比赛结果预测方法具有以下优点：

本发明主要利用比赛的历史数据，通过最大熵原理求出影响团体比赛结果的内在因素和外在因素的比重。建立符合团体比赛的最优条件概率模型，然后综合比赛团体双方最近的竞技状态，将最近的特征数据代入到利用最大熵模型解出的最优模型中，计算出参赛团体获胜的概率。给赛事关注者一个科学比赛预测。

附图说明

[0011] 图 1 为本发明的流程图；

101 S1 获取各团体历史赛季中每场比赛参赛双方的特征数据。

[0012] 102 S2 根据最大熵原理，构造最大熵模型。将训练数据集 T ，代入最大熵模型中，获取最佳的模型。

[0013] 103 S3 获取参赛双方最近 n 场的各个特征数据的期望代入最优模型 $P^*(Y|X)$ ，计算主场参赛团体最终比赛结果的分布概率 $P^*(Y|X)$ 。

具体实施方式

[0014] 为了使上述目的的特征和优点能够更加明显易懂，下面结合附图和具体实施方案给发明做出进一步详细的说明。

[0015] 如图 1 所示，本发明所提供了一种基于最大熵模型的比赛结果预测方法，其特征在于如下步骤：

S1 获取各团体历史赛季中每场比赛参赛双方的特征数据。

[0016] S2 根据最大熵原理，构造最大熵模型。将训练数据集 T ，代入最大熵模型中，获取最佳的模型。

[0017] S3 获取参赛双方最近 n 场的各个特征数据的期望代入最优模型 $P^*(Y|X)$ ，计算主场参赛团体最终比赛结果的分布概率 $P^*(Y|X)$ 。

[0018] 在步骤 S1 中 团体竞赛选取美国男子篮球职业联赛(NBA)，其中的历史赛季为在 NBA 官网上获取 2000-2014 赛季的 30 个球队整个常规赛的比赛数据。其中的特征数据：包括每场比赛参赛球队 A、B 各自的内在因素两分球、三分球、助攻、篮板、盖帽、抢断、球员伤病、主教练，外在因素：球队两场比赛之间休息的天数、天气状况，和最终主场的胜负。

[0019] 在步骤 S2 中 $X=(A \text{ 两分球}、A \text{ 三分球}、A \text{ 助攻}、A \text{ 篮板}、A \text{ 盖帽}、A \text{ 抢断}、A \text{ 球员伤病}、A \text{ 主教练})$ ，

断、A 球员伤病、A 主教练、A 球队两场比赛之间休息的天数、A 天气状况、B 两分球、B 三分球、B 助攻、B 篮板、B 盖帽、B 抢断、B 球员伤病、B 主教练、B 球队两场比赛之间休息的天数、B 天气状况), $Y=(\text{胜、负})$; 构成训练集

$T=\{(x_1, y_1)(x_2, y_2), \dots, (x_N, y_N)\}$ 。则条件概率分布 $P(Y|X)$ 的条

件熵为: $H(p)=-\sum_{x,y} \widetilde{P}(x)P(y|x)\log P(y|x)$; 将训练集代入模

型, 经过训练得出最优的模型 $P^*(Y|X)$ 。

[0020] 在步骤 S3 中 提取即将比赛的参赛双方 A、B 最近 6 场比赛的两分球、三分球、助攻、篮板、盖帽、抢断、球员伤病、主教练、球队两场比赛之间休息的天数、天气状况等特征数据的期望代入最优的模型 $P^*(Y|X)$ 。由于篮球比赛只有胜负之分, 所以只需要计算即

将比赛中主场球队的胜率 $P(Y=\text{胜}|X)$ 。

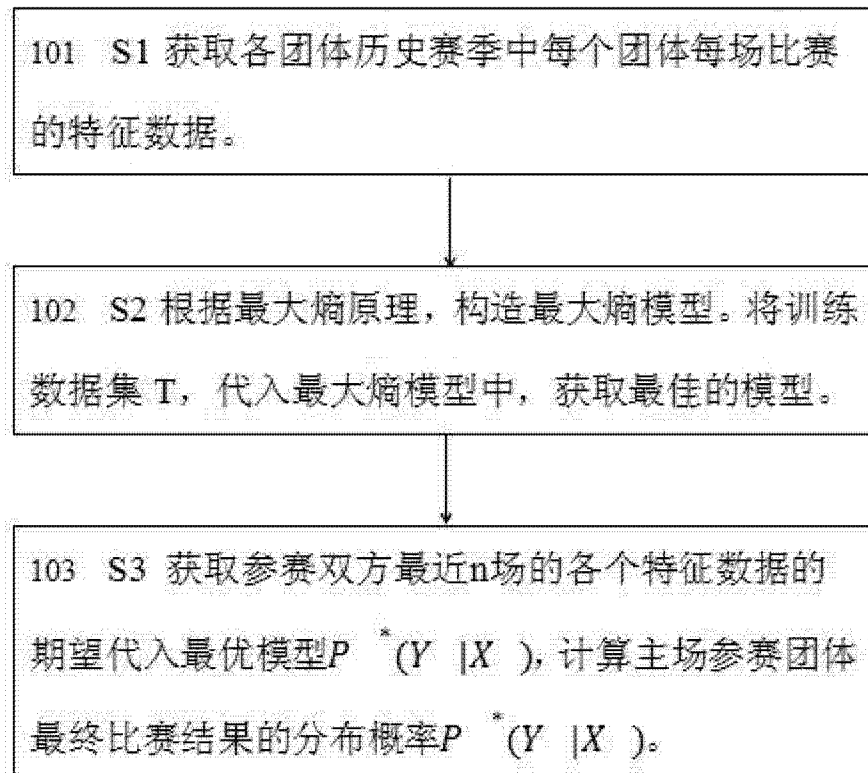


图 1