

基于概率直方图的场景预测

向宇, 刘惟一

(云南大学 信息学院, 云南 昆明 650091)

摘要: 给出一种基于概率直方图来进行场景预测的方法, 即用概率直方图描述场景发生的概率分布, 进而通过计算不同概率区间内的均值向量来完成对场景的预测与分析. 实践表明, 该方法容易实现, 具有很强的适应性, 对避免确切信息的丢失也能起到较好的效果.

关键词: 概率直方图; 场景预测; 智能信息处理

中图分类号: TP181 **文献标识码:** A **文章编号:** 1007 - 855X(2005)05 - 0061 - 04

Scenario - Forecast Based on Probability Histogram

XIANG Yu, LU Wei-yi

(Computer Science Department, Yunnan University, Kunming 650091, China)

Abstract: The probability histogram is used to solve the scenario - forecast problem in this paper, i.e. the probability of scenario is represented with probability histogram, and scenario vectors in different intervals of probability is calculated in order to forecast and analyze scenes. It is shown that the method is easy to carry out and has strong adaptability. Good effect can be gotten on avoiding the information losing.

Key words: probability histogram; scenario - forecast; intelligent information processing

0 引言

场景预测大多是基于筛选“驱动力”(driving force)的方法^[1],并在广泛运用中体现出巨大的优越性^[2].但人们并没有停止对场景预测理论的发展与完善:Best^[3]提出概率分布对于场景集合划分的积极意义;Duong^[4]提出如何运用不精确概率理论来帮助决策者构造决策场景集合.为克服传统预测方法难以给出场景集合发生概率的缺陷,我们提出一种基于概率直方图并运用直方图上的相应操作来进行场景预测的方法.

1 概率直方图在场景预测中的运用

直方图提供了对事物的全局描述,可看作函数 $y = f(x)$ 的离散表示.从概率的观点来理解,事件出现的频率可看作其出现的概率.如果用直方图记录事件发生的频度,那么直方图就对应于概率密度函数,而概率分布函数就是直方图的累积和,即概率密度函数的积分.如果直方图的柱高 n_{xi} 表示事件 x_i 发生的概率,那么这样的直方图就叫做概率直方图^[5-7].

某个系统可能发生的未来用集合 S 表示,场景 s 就是对这个系统未来状况的详细描述,是 S 中的一个元素,其对应着一个 m 维向量 $N = (n_1, n_2, \dots, n_m)$,向量每个维度 n_i 都是一个数值参数.有了这个概念,下面介绍如何运用概率直方图来对场景进行预测.我们的预测方法共分 5 个步骤:

1.1 建立场景向量和统计向量

将要比较的 p 个场景转换为等价的场景向量,并构造一个统计向量 N_s :

收稿日期: 2005 - 04 - 10. 基金项目: 国家自然科学基金 (项目编号: 60263003); 云南省自然科学基金 (项目编号: 2002F0011M); 科学院智能信息处理开放实验室课题资助 (项目编号: IIP2002 - 2).

第一作者简介: 向宇 (1978 ~), 男, 在读硕士研究生. 主要研究方向: 智能信息处理. E-mail: lionx@sina.com

$$N_s = (n_{si} / n_i = n_{ji} / n_j, N_j, 1 \leq i \leq m, j = 1, 2, \dots, p) \quad (1)$$

为了在比较中提供一个统一的衡量尺度,需要将所有向量标准化^[8],我们构造一个向量标准化模型:

$$\begin{aligned} \text{Vec}(n_i) &= [E(ep \ n_i) \ \times \text{bel}(ep)] \\ g_i &= \frac{\text{Vec}(n_i)}{\sum_{1 \leq j \leq m} \text{Vec}(n_j)}, \quad n_i \in N, \quad n_j \in N \end{aligned} \quad (2)$$

其中 $E(ep \ n_i)$ 表示专家组中某位成员 ep 对维度 n_i 影响力的估计值, $\text{bel}(ep)$ 表示决策者对专家 ep 的信赖度, $0 < \text{bel}(ep) < 1$. g_i 为标准化后的维度 n_i 的值.

1.2 求出各 N_i 与 N_s 之间的带权距离,并建立距离直方图

用一个带权的距离来描述两个向量 N_i 与 N_s 间的相似程度:

$$\begin{aligned} d_i &= k \left| \frac{\text{Vec}(n_{ix})}{\sum_{1 \leq j \leq m} \text{Vec}(n_{ij})} - \frac{\text{Vec}(n_{sx})}{\sum_{1 \leq j \leq m} \text{Vec}(n_{sj})} \right|, \quad n_i \in N_i, \quad n_s \in N_s \\ d_i &= k |g_{ix} - g_{sx}|, \quad g_i \in N_i, \quad g_s \in N_s \end{aligned} \quad (3)$$

k 是场景向量的影响因子,表示场景的影响力.在求出所有场景向量与统计向量间的带权距离后,与直方图某个柱 b_j 的距离低于阈值 (即 $|d_i - b_j|$) 的带权距离归于同一个柱 b_j ,于是便可构造出表示距离的概率直方图 H .

1.3 计算 μ , 得到预测区间^[8]

假设上一步构造的概率直方图共有 q 个柱,于是可在 H 上构造一个概率期望 μ

$$\mu = \frac{\sum_{1 \leq j \leq q} b_j n_{bj}}{\sum_{1 \leq j \leq q} n_{bj}} \quad (4)$$

其中 $b_j n_{bj}$ 为柱 b_j 和柱高 n_{bj} 的乘积, n_{bj} 是柱 b_j 的高度.如果给直方图上赋予质量,那么 μ 可以理解为整个直方图质量取得平衡的位置,即概率分布的均衡点.进一步求出直方图 H 上各柱点 b_j 到概率均衡点 μ 的概率距离:

$$= \sqrt{\sum_{1 \leq j \leq q} \left[(b_j - \mu)^2 \frac{n_{bj}}{\sum_{1 \leq j \leq q} n_{bj}} \right]} \quad (5)$$

得到概率区间 $[\mu - , \mu +]$.可以这样认为:若事件再次发生(场景再次出现),其与统计向量的距离极有可能落入到该区间内,结论的可信度为:

$$= \frac{\sum_{1 \leq j \leq q} n_{bj}}{\sum_{1 \leq j \leq q} n_{bj}}, \quad b_j \in [\mu - , \mu +] \quad (6)$$

1.4 分区间求均值场景向量并讨论结果

得到概率区间 $[\mu - , \mu +]$ 后,可将整个距离直方图 H 划分为 3 个区间 $[0, \mu -)$, $[\mu - , \mu +]$, $(\mu + ,$

$$N_- = (n_i / n_i = \frac{n_{ji}}{C_-}, n_{ji} \in N, d_{nj} \in [0, \mu -), 1 \leq i \leq m) \quad (7)$$

$$N = (n_i / n_i = \frac{n_{ji}}{C}, n_{ji} \in N, d_{nj} \in [\mu - , \mu +], 1 \leq i \leq m) \quad (8)$$

$$N_+ = (n_i / n_i = \frac{n_{ji}}{C_+}, n_{ji} \in N, d_{nj} \in [\mu + ,), 1 \leq i \leq m) \quad (9)$$

其中, C_-, C, C_+ 分别为落入区间 $[0, \mu_-)$, $[\mu_-, \mu_+]$, (μ_+, ∞) 的场景向量个数. 运用公式 (6) 计算概率区间 $[\mu_-, \mu_+]$ 的可信度, 若结果达到期望, 直接用 N 作为预测结果, 否则还需经步骤 1.5 作进一步分析.

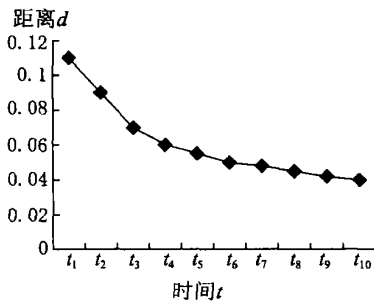


图1 下降的变化

Fig.1 Downward movement

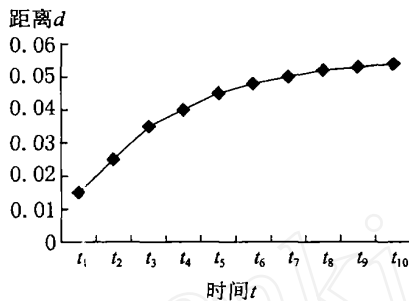


图2 上升的变化

Fig.2 Upward movement

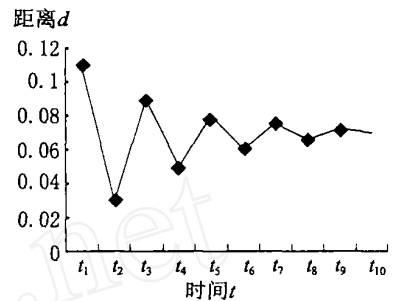


图3 波动的变化

Fig.3 Fluctuant movement

1.5 验证

每隔一段时间对待预测系统进行一次场景取样, 建立即时的技术统计向量 R , 并求 R 与 N_s 的距离 d_R . 观察 d_R 的变化情况, 推测其将落入哪个区间内. d_R 可能的变化情况如图 1 ~ 3 所示. 图中 $t_1 \sim t_{10}$ 表示取样的时间点. 当到达某一时刻 t_i ($1 < i < 10$), 发现 d_R 的变化已趋于稳定. 这时检查 d_R 进入了哪个概率区间, 便以该区间的均值向量作为预测结果. 当然, 参与构造直方图的场景越多, 那么划分出来的概率区间也越符合预测需要.

2 实例

每场篮球比赛后都能得到参赛两支球队的技术统计数据, 根据这些数据可建立起场景向量 N (前篮板, 后篮板, 助攻, 犯规, 抢断, 失误, 盖帽, 得分). 现在希望通过对两支球队过去若干场比赛的技术数据进行场景分析, 来预测他们之间进行比赛的结果.

1) 创建火箭队过去 14 场比赛各自对应的技术统计向量, 分别记为 N_1, N_2, \dots, N_{14} . 如 $N_1 = (11, 27, 20, 18, 7, 11, 5, 85)$. 再根据公式 (1) 构造出 14 场比赛总的技术统计向量 N_s . 根据公式 (2), 将这 15 个向量标准化.

2) 根据 (3) 计算 N_i 与 N_s 的距离 ($1 \leq i \leq 14$), 影响因子 k 可根据比赛的重要程度来确定, 如季前赛 $k = 0.8$, 常规赛 $k = 1$, 季后赛 $k = 1.2$. 根据得到的带权距离构建距离直方图: 柱高 n_{b_i} 表示与 N_s 距离为 b_i (或接近 b_i , 差距低于阈值) 的向量个数. 该距离直方图记为 H . 例如, N_1 与 N_s 的距离 $d_1 = 0.09$, 于是 H 的柱 $h_{0.09}$ 的高度加 1 (阈值 = 0.05). 最终构造的距离直方图如图 4 所示.

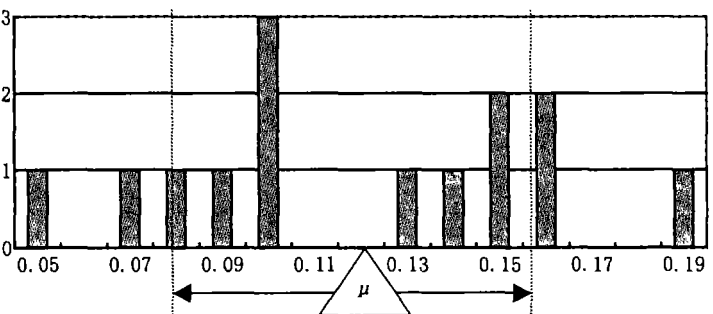


图4 预测结果

Fig.4 Result of forecasting

3) 对图 4 所示的距离直方图进行计算, 根据公式 (4) 和 (5), $\mu = 0.11929$, $\sigma = 0.0404$, 于是求得三个概率区间 $[0, \mu - \sigma) = [0, 0.07889)$ 、 $[\mu - \sigma, \mu + \sigma] = [0.07889, 0.15969]$ 、 $(\mu + \sigma, \infty) = (0.15969, 0.2]$, 其中距离上界 = 0.2.

4) 对于区间 $[0, 0.07889)$, 有 2 场比赛落入该区间内, 根据公式 (6) 计算可信度为 $2/14 = 14\%$, 根据公式 (7), 该区间对应均值向量 $N = (11, 32, 21, 23, 5, 15, 5, 82)$. 对于区间 $[0.07889, 0.15969]$, 有 9 场比赛落入该区间内, 可信度为 $9/14 = 64\%$, 根据公式 (8), 该区间对应均值向量 $N = (13, 31, 19, 23, 6, 16, 6,$

89). 对于区间 $(0.15969, 0.2]$, 有 3 场比赛落入该区间, 可信度为 $3/14 \approx 21\%$, 根据公式 (9) 该区间对应均值向量 N_+ (9, 33, 16, 23, 10, 18, 8, 89).

同理, 根据湖人队过去若干场比赛的技术统计, 也可求出相应距离直方图上的三个概率区间, 以及对各概率区间的可信度和均值向量. 如区间 $[0, \mu_+)$ 的可信度为 37.5%, 对应均值向量 N_+ (10, 28, 28, 20, 7, 13, 3, 110). 区间 $[\mu_-, \mu_+)$ 的可信度为 50%, 均值向量为 N_+ (7, 29, 28, 22, 10, 14, 3, 107). 区间 $(\mu_+, 1]$ 的可信度为 12.5%, 均值向量为 N_+ (11, 32, 20, 22, 10, 23, 8, 95). 比较后得知, 如果火箭队与湖人队进行一场对抗, 那么无论比赛结果落入那个区间内, 湖人队都将获胜.

3 总结

提出一种基于概率直方图并运用直方图上的操作来进行场景预测的方法. 实践证明, 该方法容易操作, 可应用于大多数场景预测任务; 此外, 该方法对避免确切信息的损失也能起到较好的效果. 在未来工作中, 我们计划寻找更好的方法构造决策场景集合; 另外, 如何更加准确、高效地对场景集合进行划分也是一项很有意义的工作.

参考文献:

- [1] Bannis A, Sideris. The Use of Scenario Planning Methodology in Mobile Business as a Learning Tool[R]. ELTRUN, The e-Business Center Athens University of Economics and Business, 2000. 1~8.
- [2] Foster R N, Kaplan S. Managing the Uncertain Future—The Value of Scenario Planning[C]. Creative Destruction. BM Advanced Business Institute. OFDA Dealer Strategies Conference, 2002. 1~5.
- [3] Eric Best. Probabilities - Help or Hindrance in Scenario Planning? [R]. Technical Report The Global Business Network, 1991, 2(4): 1~9.
- [4] Minh Ha Duong. CRED, CNRS and Carnegie Mellon University[Z]. USA, Imprecise Probability Bridges Scenario - Forecast Gap. June 2, 2003. 11~7.
- [5] Alan M. Polansky. The Normal Approximation for Probability Histograms[M]. Devision of Statistics Northern Illinois University, 1996. 3~10.
- [6] Mendenhall W, Wackerly D, Scheaffer R. Mathematical Statistics with Applications[M]. Chapter 2. Section 2-1 and 2-2. Duxbury Press, Fourth Edition, 1989. 18~25.
- [7] MAR D F, TR DLA. Probability Distributions[J]. Chapter 4. Section 4-1 and 4-2. Triola, Elementary Statistics, Eighth Edition. Addison Wesley Longman, 2001. 43~45.
- [8] Martina Mudrová. 3rd Lesson in Image Processing[R]. Histogram and its Processing, Teaching Materials October, 2002. 1~14.