

万能的贝叶斯决策——应用总结

学完《模式识别》一课之后，收获颇多。说实话，这门课要想学好不简单，但是老师教会我们要掌握方法，不要拘泥于大堆的公式。方法的思想掌握了，遇到问题以后就可以开阔思路，直接拿来用了。课上主要讲了四大块，Beyes 决策，概率密度函数估计，线性判别以及聚类和 Fuzzy 模式识别。下面就其中的 Beyes 判别一项做一下应用方面的总结，所选材料均来自学校图书馆 CNKI 中国学术期刊全文总库。

众所周知，Beyes 公式是统计学里一个非常重要的公式，而 Beyes 决策理论方法则是统计模式识别中的一个基本方法。根据 Beyes 决策设计的分类器理论上性能最优，经常被用来作为衡量其他分类器优劣的标准。

当然，要想使用 Beyes 理论进行决策，还必须满足几个条件：（1）对象的所有特征观察量，我们设为 d 维特征空间，记为 $d = [x_1, x_2, \dots, x_d]$ ；（2）要决策分类的类别数，我们设为 c 类，用 ω_i 来表示， $\omega \in \Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ ；（3）各类别总体的概率分布，即 ω_i 出现的先验概率 $p(\omega_i)$ ；（4）类条件概率密度 $p(x|\omega_i)$ 。知道以上几个条件以后，给定一个观测值 x ，我们就可以根据需要利用相应的 Beyes 决策规则把它分到相应的类去。几种决策规则包括：基于最小错误率的 Beyes 决策、基于最小风险的 Beyes 决策、最小最大决策以及序贯分类方法等。

Beyes 决策理论是模式识别中的一个比较基础的决策方法，应用十分广泛，几乎涉及到了方方面面。

1. 医学方面

Beyes 决策在医学方面有非常重要的地位，主要应用在医疗诊断中。比如我们模式识别经典课本中所列举的癌细胞判别的例子。在医疗诊断中，许多疾病的症状比较相似，即使同一种病，病情的严重程度不同，症状更复杂（如：阑尾炎是慢性，急性还是穿孔；胃癌的早期，中期与晚期等），这就给医生的诊断带来了一定的困难。利用 Beyes 统计决策就可以很好的解决这一问题。

例如：诊断阑尾炎的例子^[1]

设有三种疾病状态： A_1 表示慢性阑尾炎， A_2 表示急性阑尾炎， A_3 表示阑尾炎穿孔，根据以往的统计经验先验概率已知。又设疾病的症状可分为 n 类，表示为 B_1, B_2, \dots, B_n 。疾病 $A_i (i = 1, 2, 3)$ 涉及到症状 $B_j (j = 1, 2, \dots, n)$ 的概率为 $p(B_j | A_i)$ 。

假定各症状的出现是相互独立的。则有公式：

$$P(B | A_i) = P(B_{j_1} | A_i) \cdot P(B_{j_2} | A_i) \cdots P(B_{j_k} | A_i)$$

其中，k 表示出现症状的数量。

现有一病例，晚上 6:00 突发全腹痛，拌恶心呕吐，大便正常，体温 38.6℃，右下腹固定压痛，肌紧张（+），白细胞：18500。那么应该把他判为哪种疾病呢？根据 Bayes 公式：

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{\sum_{i=1}^m P(A_i)P(B | A_i)}$$

我们可以求出 $P(A_1 | B)$ 、 $P(A_2 | B)$ 、 $P(A_3 | B)$ ，然后根据最小错误率决策进行判决。上例中，可以求出该患者患穿孔性阑尾炎的概率最大（73.63%），应尽快作手术。

Bayes 决策在医学上的应用还有很多，文献 [2] 利用 Bayes 公式精确计算出遗传病再发风险；文献 [3] 通过引入 Bayes 理论，从 DNA 分子标记的表现型推断其基因型，也就是推断 DNA 的来源；文献 [4] 引入了 Bayes 多类线性判别分析，利用左心室节段功能参数对 67 例左心室造影图像进行了分析，最后利用节段功能参数 Bayes 多类线性判别分析计算机辅助的方法诊断冠心病，效果显著。

2. 经济方面

Bayes 决策在经济方面也有很多的应用。例如在房地产开发方面^[5]，如何在不确定环境下进行房地产项目的方案比选，已经成为房地产开发中亟待解决的问题。那么，我们就可以把影响项目开发的关键因素作为分析对象，对不同方案采用 Bayes 方法进行决策分析，从而选出合理科学的项目方案。

该方法进行房地产项目方案比选的步骤如下：

（1）识别关键因素，估计其发生的概率（先验概率）。选择具有不确定性的关键因素（往往是外部因素）进行分析，并根据关键因素的发生概率和组合情况，计算不同外部环境 θ 发生的概率 $P(\theta_i)$ ；

（2）提出待选方案，根据市场预测相应的方案收益值 $R(a_i, \theta_j)$ ($i = 1, 2, 3, \dots, n; j = 1, 2, 3, \dots, m$)；

（3）根据历史统计资料取得条件概率，利用 Bayes 公式计算后验概率；

（4）利用后验概率计算决策函数的 Bayes 收益值，并选择最优决策方案：

$$S_k = \sum_{i=1}^m (\sum_{j=1}^m R((\delta_k X_j), \theta_j) P(\theta_j | X_i)) P(\theta_i) (k = 1, 2, \dots, n^m)。$$

S_k 中的最大值对应的方案就是最优决策方案，相应的决策函数为最优决策函数。

当然，除了在房地产中的应用外，Beyes决策还可以用在许多方面，比如：超市产品的销售，工厂产品的合理规划等。文献 [6] 则提出了一种新的想法，把Beyes方法应用到个人信用评估上，当然，这也涉及到了Beyes网络的问题，再次不加以讨论。

3. 工业方面

这里所说的工业主要指的是矿产等重工业产业。例如数据挖掘和人工智能都喜欢举的一个例子——探矿工程。已知某块山地表层开有兰花，那么这块山地地下有铜矿的可能性是多大？这就要用到贝叶斯决策了。

文献[7]提出，把Beyes决策应用到地下工程岩爆发生及烈度分级预测上。综合分析影响岩爆主要因素，选取最大切向应力、岩石抗压强度、岩石抗拉强度和弹性能量指数作为判别因子建立岩爆预测的Bayes判别分析模型。文献 [8] 则是在混凝土结构碳化深度预测上应用了Beyes决策，并能够根据不同时期的工程检测数据，实现贝叶斯动态信息更新，其更新结果更加接近工程结构的实际情况。

由于工业方面的分析比较复杂，设计的参数以及理论知识比较多，在此，就不加以叙述了。

4. 军事方面

自有人类存在以来，军事问题是一个一直都存在的话题。这涉及到了一个国家的安全稳定。那么，我们万能的Beyes决策理论同样也在军事问题上找到了自己的用武之地。充分表现在我们的军事情报的定量分析上 [9] 等等。

军事情报分析是一个复杂的过程，不完全性、不确定性、不精确性及涉及人的认知领域，这些特点一直成为军事情报量化和自动化分析发展的“瓶颈”。尤其是当前，各种出版物、媒体的高开放性，使我们获取情报素材的手段有了很大的提高。

军事情报分析过程划分为两个阶段：（1）情报汇集；（2）情况分析判断。做出关于其战略意图、战场态势、作战能力或行动意图的判断。假设事件 $h_1, h_2, \dots, h_i, \dots, h_n$ 互不相容且构成一个完全时间集合，已知他们的概率

$p(h_i), i=1,2,\dots,n$ 。观测到某种证据 d 与 $h_1, h_2, \dots, h_i, \dots, h_n$ 相伴随而出现，且已知条件概率 $p(d|h_i)$ ，求在新证据 d 出现之后事件 h_i 的概率，即 $p(h_i|d)$ ：

$$p(h_i|d) = \frac{p(h_i)p(d|h_i)}{p(h_1)p(d|h_1) + \dots + p(h_i)p(d|h_i) + \dots + p(h_n)p(d|h_n)}$$

这个提法还是比较新颖的，这是第一次把数学方法和计算机技术注入到传统上依靠人的认知进行定性分析的军事情报分析领域，提高了分析问题的深入程度和准确性，较大程度地避免了情报主观分析时认知领域的缺陷，开辟了军事情报分析的新途径。

不仅是在情报的分析上，在军事的其他方面，Beyes决策同样有很大的应用。例如：文献[10][11][12]都提到了导弹的问题，其中包括导弹的射程、精度鉴定以及贮存可靠性问题，几乎涉及到了方方面面。

5. 其他

除了以上几个方面，Beyes决策理论在其他方面也同样有大展身手的余地。例如：现在讨论比较火的人脸识别问题、图像标注问题、车用汽油的红外光谱识别问题[13]、垃圾邮件的过滤问题[14]等等。甚至有人专门研究起来了Beyes决策理论的哲学问题[15]。

综上所述，Beyes决策理论这样一个小小的理论在我们日常的生活中应用是很广泛的，几乎涉及到了方方面面。我们要做的就是要掌握这种方法，尤其是思路，在遇到问题以后可以把这个方法拿过来加以利用。同时也可以引申到其他的方法上。毕竟，会想会用了，我们才算是真正掌握了一种方法。

参考文献：

- [1] 张改英,张辉.医生诊断中的 Bayes 决策模型.第四军医大学学报,1998,19:137,138
- [2] 申本昌,章远志,谢天炽.利用 Bayes 公式计算单基因遗传病再发风险的教学策略.中国优生与遗传杂志,2009,17(1):9,10,102
- [3] 莫惠栋,姜长鉴.应用 Bayes 理论推断 DNA 分子标记基因型.遗传学报,2001,28(6):568-574
- [4] 冯全洲,白敦衍,陆振刚.以 X 线造影左心室节段功能为参数计算机辅助的冠心病诊断方法初探.第一军医大学学报,1996,16(2):82,83

- [5] 温海珍,贾生华,杨志威.BAYES 方法在房地产项目方案比选中的应用 .土木工程学报,2006,39(9):108-111
- [6] 郭春香,李旭升.贝叶斯网络个人信用评估模型.系统管理学报,2009,18(3):249-260
- [7] 宫凤强,李夕兵,张伟.基于 Bayes 判别分析方法的地下工程岩爆发生及烈度分级预测.岩土力学,2010,31(1):370-377
- [8] 吴本英,周锡武.基于贝叶斯方法的混凝土结构碳化深度预测研究.武汉理工大学学报,2011,33(3):103-107
- [9] 程英,高庆德,尹照武.一种军事情报定量分析系统研究.微计算机信息,2009,25:24,25
- [10] 李刚,鲜勇,王明海.弹道导弹最大射程评定的 Bayes 方法.弹道学报,2009,21(3):74-76
- [11] 查亚兵,谢莉萍.导弹射击精度鉴定中 Bayes 方法的运用.系统仿真学报,2002,14(6):812-814
- [12] 刘飞,窦毅芳,张为华.基于成败试验的导弹贮存可靠性 Bayes 分析模型.航空动力学报,2007,22(1):102-107
- [13] 田高友,熊春华,刘慧颖.近红外光谱技术识别车用汽油的应用研究.Modern Scientific Instruments,2006,5:83-86
- [14] 李惠娟,高峰,管晓宏,黄亮.基于贝叶斯神经网络的垃圾邮件过滤方法.微电子学与计算机,2005,22(4):107-111
- [15] 熊卫.贝叶斯决策理论的理性观及其相对性.科学技术与辩证法,2006,23(5):35-39

万能的贝叶斯决策——应用总结

2011 年 6 月 8 日