

无处不在的贝叶斯网络

——读《数学之美》

小组成员：侯晓雯 叶青 陈琛 何禛 李林达

本文从一个经典的案例¹开始。

亲爱的读者你好。假如你是一名沉浸在爱情的甜蜜中的女性，某一天突然发现衣柜里有一件女性内衣，而该内衣不是你的。现在的问题是，你的男朋友出轨了吗。让我们用科学的手段解决问题——首先，请大概推测一下以下三个数字。

第一，如果你的男友出轨，这件内衣有多大可能出现。考虑到出轨需要隐藏痕迹，此处概率推断为 50%。第二，他没出轨的情况下这件内衣有多大可能出现——当然，万事皆有可能。假定这种概率为 5%。第三，你需要估计一下贝叶斯定理中所说的先验概率，即在发现内衣之前你认为男友出轨的概率有多大？我们引用的案例中，大数据告诉我们已婚夫妇任何一年的出轨概率在 4%左右，我们将其视为先验概率。

依据贝叶斯定理得出后验概率，如下(图表亦源自“大数据文摘”博客，2015)：

先验概率	概率数值
男友出轨的初始概率预测	4%
新事件：发现神秘内衣	
男友出轨的情况下，内衣出现的概率	50%
男友未出轨的情况下，内衣出现的概率	5%
后验概率	
在你发现内衣的情况下，你男友出轨的概率	29%

综上所述，男友出轨的概率并没有那么高。这个思路来自一个叫“贝叶斯网络”的数学模型。此处我们采用朱明敏（2016）的定义：

“贝叶斯网络 (BN) 是一个二元组，即 $BN = (G, P)$ ， $G = (V, E)$ 为有向无环图，其中 V 为节点集，与领域的随机变量一一对应； E 为有向边集，反映变量之

¹ 此处的案例出处见“参考文献与引用”第四条。

间的因果依赖关系，从节点 X 到节点 Y 的有向边表示 X 对 Y 有直接的因果影响， P 为节点的概率分布，表示节点之间因果影响强度，每个节点都有一个条件概率表，定量描述其所有父节点对该节点的作用效果。”

当然这个定义看起来并不说人话。而《数学之美》（吴军，2012）以及本篇读书笔记的目的，就是告诉大家这样一个看似不说人话的数学模型能有怎样的应用。

一、作者说了什么

首先，什么是贝叶斯网络。它是马尔科夫链的一个拓展模型，马尔科夫链是其特例。二者都满足马尔可夫假设：随机过程中各个状态 S_t 的概率分布，只与它前一个状态 S_{t-1} 有关。但马尔科夫链是条链，但贝叶斯网络是一张网，或者一个结构图，因此它可以更准确地描述相关性。试举一例，图 5.3 中可以从 m_2 到 m_3 到 m_4 或者从 m_3 到 m_3 到 m_4 ，但从 m_3 到 m_4 的概率永远是 0.3。²

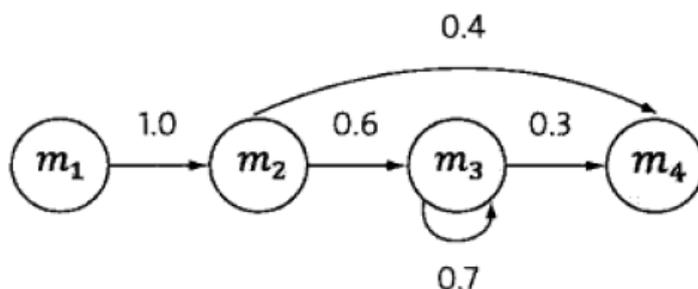


图 5.3 马尔可夫链



图 24.1 描述心血管疾病和成因的一个简单的贝叶斯网络

² 作为一篇读书笔记，文中各类案例图片出自《数学之美》第 22 章。

由上文中心血管疾病成因的图可见，贝叶斯网络包括三个部分。首先是圆圈中的状态或指标（运动量、吸烟与否等），然后是连线代表的因果关系（箭头指向原因）。第三是因果关系线上的数字，标识原因对结果的影响率。

这个模型主要被谷歌用来给文章分类。机智的工程师们主要干了三件事：在文本中寻找关键词、将关键词归类（一类称为一个概念）、将概念进行分类（一类称为一个主题）。通过这种叫主题模型的方式，他们制作出了如下的贝叶斯网络图，从文章里提炼出主题：

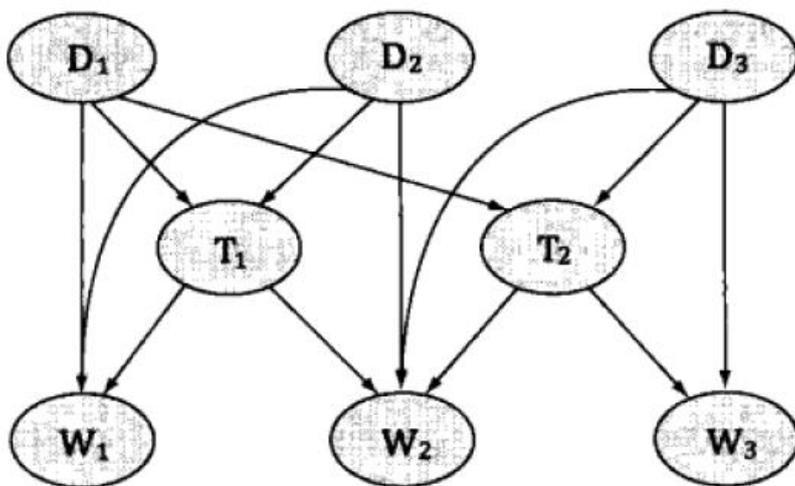


图 24.3 描述文章 (D)、主题 (T) 和关键词 (W) 的贝叶斯网络

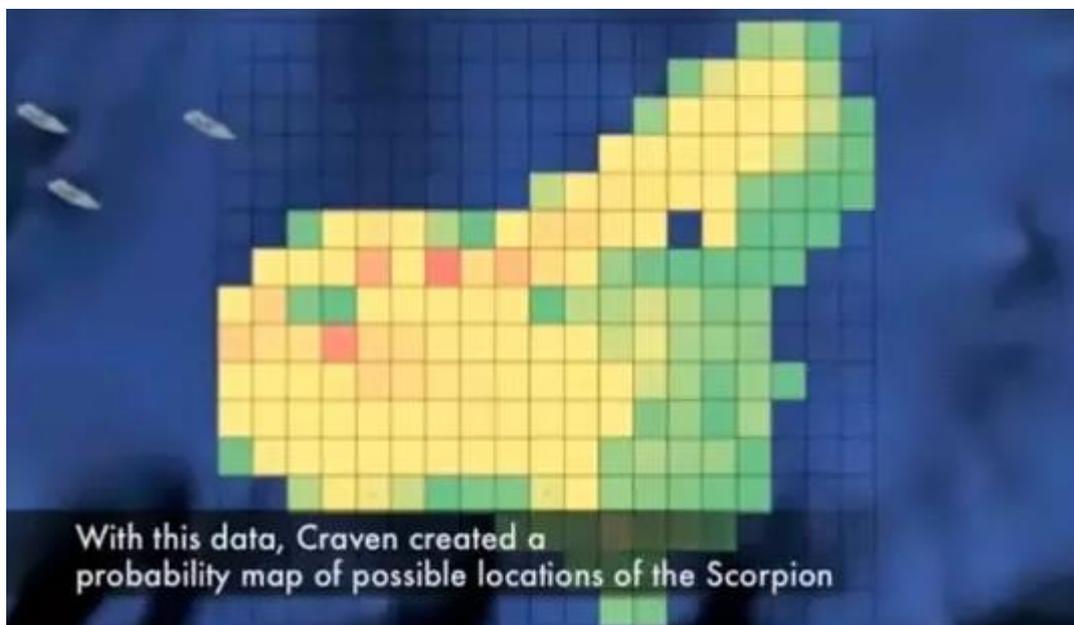
这种分类方法有一个问题。它只考虑了文章中的关键字而没有考虑语境，导致分类不合理——考虑进上下文的约束再次分类后，就获得了成功。

还有一个问题，就是如何训练贝叶斯网络，让这个模型更加合适。具体而言，就是如何确定其结构（结构训练）以及权重（参数训练）。面对成千上万的新闻，网络结构以及权重没法通过人工的方法预设，只能通过算法让机器自学。结构训练目标是让结构产生的序列从头走到尾的可能性最大。最直接的方法是计算每一条路径从头到尾的概率进行比较，但这样计算量太大。因此，工程师采用“贪心算法”简化计算。同时，通过最大熵模型的训练方法来确定参数，最后通过结构训练参数训练的交替进行做到模型的优化。

二、我们发现贝叶斯网络还可以做什么

贝叶斯网络看上去高深莫测，但在日常生活中，我们常常会运用它进行决策。比如我们要面对天字第一号难题：中午吃啥。在对餐厅不了解的情况下，就会通过经验判断：人多的餐厅食物会更美味一些。这就是通过先验知识（餐厅的上座率）进行的判断。而在真正尝试过以后，我们会对原来的判断进行修正，下次找吃的可能就容易一些。

当然应用不止中午吃什么。比如说，2014 年马航 MH370 失联事件牵动着许多人的神经。现在的问题是，怎么找到飞机——贝叶斯网络告诉你一种可能的方法。首先，我们把可能的范围划分成一个个小格子（如下图），每个小格子有两个概率值 p 和 q 。 p 是飞机在这个格子中的概率， q 是如果飞机在这里，我们找到它的机率。



我们先对 p 最大的格子进行搜寻，如果没有发现残骸，那么根据贝叶斯公式，这个格子中飞机存在的概率就会下降为

$$p' = \frac{p(1-q)}{(1-p) + p(1-q)} = p \frac{1-q}{1-pq} < p$$

则其他格子中飞机残骸存在的概率就会上升。在重新修正格子的概率分布后，再对修正后概率最大的格子进行搜救，如此往复，由此显著提高了搜救的效率（王晓峰，2015）。这种基于贝叶斯公式的方法被实践证明是行之有效的，现已成为海难空难的通用搜救方法。

在生物领域，贝叶斯公式也有一些用处。比如说，也许一种致病基因会导致多种疾病或多种基因导致了某种疾病，这时可以用贝叶斯网络来判断究竟是哪种或哪几种基因会致病，从而研究对应的治疗方法。在金融界，基金经理用贝叶斯公式找到投资策略：首先根据自己的预期收益率构造一个投资组合，在实际盈亏中不断修正投资的权数，调整头寸，最终达到风险最小收益最大的平衡点。在 IT 领域，互联网公司用贝叶斯公式改进搜索功能，帮助用户过滤垃圾邮件……大数据、人工智能和自然语言处理中都大量用到贝叶斯网络，可见贝叶斯网络的巨大力量。

也许我们会问，贝叶斯网络对我个人来说有什么用呢？正如我们开头所举的例子，衣柜

里发现的别人的内衣, 你的伴侣有没有出轨? 这条街上这么多餐厅我该选择哪一家? 我应该考研还是直接找工作才能使我的未来收益最大化? 我们常常要在信息不充分的情况下作出决策, 而贝叶斯公式则提供了一种思维方式, 即对每一个影响因素分配一个概率值, 通过实践或观测获得的样本数据重新代入修正概率值, 从而使模型达到最优化, 为我们的预测或决策提供较为准确的方向和信息。

贝叶斯网络也为我们的行为决策提供了一些启示。当我们在面临选择的时候, 往往要通过自己的经验或预期得出一个接近事实的先验概率, 这就要求我们在平时不断思考并积累经验, 这样在纠错时只需要进行少量的迭代循环就可以达到趋于真实的结果。另外, 要勇于对自己的行为决策不断的调整, 这样在迭代循环中才能快速地试错和改善。

三、我们还想到了什么

以上描述了本书中的一个具体章节。然, 透过它可以折射出本书的一个中心论点: 可以用数学模型解读、模拟、预测这个世界。当然, 这个原则作者从开篇就不断强调起: 本书主要关注自然语言处理, 而作者在首章中就将信息传递化约为“编码——译码”的过程, 并基于这个过程实现了分词和新闻聚类等技术。最初的尝试中人们试图穷尽所有的语言规律——作者将这种繁杂的做法批判了一番, 进而引入了使用数学模型的概念。

数学模型极简, 几个公式; 世界极繁, 从昨天日元的汇率波动到村口二大爷家儿子的情感纠纷。以极简建构极繁, 可能这就是数学之美, 也是经济学之美, 以及量化方法之美。对于经济学(以及其他社会科学)而言, 数学的引入是运用这样一套简明的体系化约人的行为。

然, 这种方法能否穷尽一切?

讨论中, 我们提到了计划经济。它为人诟病主要因为预先计划的方式容易僵化。即使当今, 量化的方式依然难以穷尽整个经济市场——那么, 假设一切都可测量, 计划经济是不是就能实现了? 同理, 在社会学中, 质性研究者批评定量研究忽略了大样本中的个体, 难以研究微观视角下事件发生的具体机理。但随着数据获取手段的升级, 数学模型是否有一天能纳入世界上的一切变量?

我们也没有答案。

参考文献及引用

1. 吴军. 数学之美[M]. 北京：人民邮电出版社. 2012
2. 朱明敏. 贝叶斯网络结构学习与推理研究[D]. 西安：西安电子科技大学, 2013.
3. 王晓峰. 大数据背后的神秘公式：贝叶斯公式[EB/OL].
<http://it.sohu.com/20160415/n444284325.shtml>. 2016
4. 大数据文摘. HELP！我发现他有其他女人内衣，他出轨的可能性有多大？[EB/OL].
http://blog.sina.com.cn/s/blog_591ab0eb0102vcl6.html. 2015