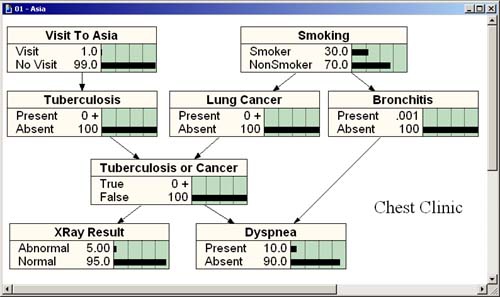
贝叶斯网络应用实例一：胸部疾病诊所（Chest Clinic）

假想你是Los Angeles一名新毕业的医生，专攻肺部疾病。你决定建立一个胸部疾病诊所，主治肺病及相关疾病。大学课本已经中告诉你了肺癌、肺结核和支气管炎的发生比率以及这些疾病典型的临床症状、病因等，于是你就可以根据课本里的理论知识建立自己的Bayes网。如根据如下数据信息：

* 美国有30%的人吸烟.
* 每10万人中就就有70人患有肺癌.
* 每10万人中就就有10人患有肺结核.
* 每10万人中就就有800人患有支气管炎.
* 10%人存在呼吸困难症状, 大部分人是哮喘、支气管炎和其他非肺结核、非肺癌性疾病引起.

根据上面的数据可以建立如下BN模型：

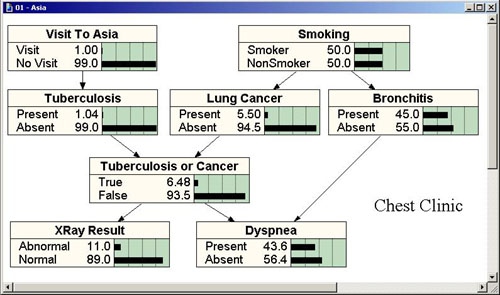
<img src="https://pic2.zhimg.com/21a569894683edbbcfcf39b0ecb46e99\_b.jpg" data-rawwidth="500" data-rawheight="297" class="origin\_image zh-lightbox-thumb" width="500" data-original="https://pic2.zhimg.com/21a569894683edbbcfcf39b0ecb46e99\_r.jpg">

这样的一个BN模型对你意义不大，因为它没有用到来你诊所病人的案例数据，不能反映真实病人的情况。当诊所诊治了数千病人后，会发现课本中所描述的北美的情况与实际诊所数据显示的情况是完全不同的，实际诊所数据显示：

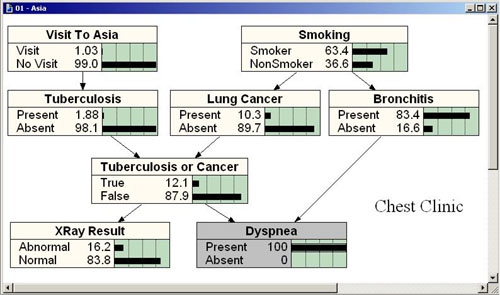
· 50%的病人吸烟.

· 1%患有肺结核.

· 5.5% 得了肺癌.

· 45% 患有不同程度支气管炎.  
将这些新数据输入到BN模型中，才真正的获得了对你有意义的实用BN模型：  
<img src="https://pic2.zhimg.com/0ef323c753f7a2a8b3f37a2134ab2319\_b.jpg" data-rawwidth="500" data-rawheight="295" class="origin\_image zh-lightbox-thumb" width="500" data-original="https://pic2.zhimg.com/0ef323c753f7a2a8b3f37a2134ab2319\_r.jpg">  
现在，看看如何在日常诊断中用该BN模型。

· 首先，应该注意到，上述模型反映了一个来诊所求医的新患者，为诊断之前我们没有这个患者的任何信息。而当我们向患者咨询信息时，BN网中的概率就会自动调整，这就是贝叶斯推理最完美、强大之处。贝叶斯网络最强大之处在于从每个阶段结果所获得的概率都是数学与科学的反映，换句话说，假设我们了解了患者的足够信息，根据这些信息获得统计知识，网络就会告诉我们合理的推断。

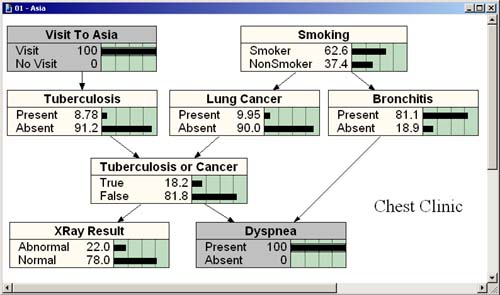
· 现在看看如何增加个别病人信息调节概率。一个女病人进入诊所，我们开始和她谈论。她告诉我们她呼吸困难。我们将这个信息输入到网络。我们相信病人的信息，认为其存在100%呼吸困难。  
<img src="https://pic1.zhimg.com/cbb64ffb6b65e847a420260f4ea93604\_b.jpg" data-rawwidth="500" data-rawheight="295" class="origin\_image zh-lightbox-thumb" width="500" data-original="https://pic1.zhimg.com/cbb64ffb6b65e847a420260f4ea93604\_r.jpg">可以观察到，一旦病人有呼吸困难症状，三种疾病的概率都增大了，因为这些疾病都有呼吸困难的症状。我们的病人存在这样的症状，某种程度上我们会推断这三种疾病可能性比较大，也增加了我们患者有严重疾病认识的信念。可以观察到，一旦病人有呼吸困难症状，三种疾病的概率都增大了，因为这些疾病都有呼吸困难的症状。我们的病人存在这样的症状，某种程度上我们会推断这三种疾病可能性比较大，也增加了我们患者有严重疾病认识的信念。

· 仔细看看推断的过程：

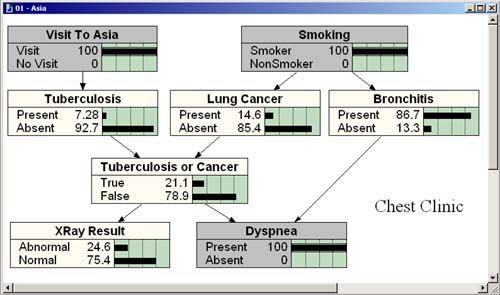
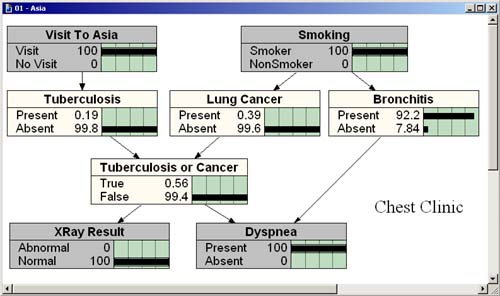
1. · 明显增大的是支气管炎，从 45% 到 83.4%. 为什么会有如此大的增长呢？因为支气管炎病比癌症和肺结核更常见. 只要我们相信患者有严重的肺部疾病，那最支气管炎的可能性会更大些。
2. 病人是抽烟者的几率也会随之增大，从50% 到63.4%.
3. 近期访问过亚洲的几率也会增大: 从1% 到1.03%, 显然是不重要的.
4. X光照片不正常的几率也会上涨，从11% 到16%.

知道现在我们还无法确认什么疾病困扰着我们的这个女患者，我们目前比较相信她患有支气管炎的可能性很大，但是，我们应该获得更多信息来确定我们的判断，如果我们现在就主观定了病症，她可能得的是癌症，那我们就是一个烂医生。这就需要更多信息来做最后的决定。

因此，我们按照流程依此问她一些问题，如她最近是不是去过亚洲国家，吃惊的是她回答了“是”。现在获得的信息就影响了BN模型。

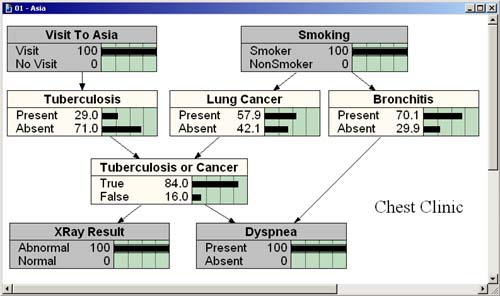
<img src="https://pic2.zhimg.com/d57be3a397a169b15e420da0f47a6729\_b.jpg" data-rawwidth="500" data-rawheight="295" class="origin\_image zh-lightbox-thumb" width="500" data-original="https://pic2.zhimg.com/d57be3a397a169b15e420da0f47a6729\_r.jpg">

· 患肺结核的几率显然增大，从 2%到 9%. 而患有癌症、支气管炎以及该患者是吸烟患者的几率都有所减少。为什么呢？因为此时呼吸困难的原因相对更倾向于肺结核。

· 继续问患者一些问题，假设患者是个吸烟者，则网络变为  
<img src="https://pic2.zhimg.com/a729b982b5dede08862a9101928dc5d5\_b.jpg" data-rawwidth="500" data-rawheight="295" class="origin\_image zh-lightbox-thumb" width="500" data-original="https://pic2.zhimg.com/a729b982b5dede08862a9101928dc5d5\_r.jpg">  
此时注意到最好的假设仍然是认为患者患有支气管炎。为了确认我们要求她做一个X光透视，结果显示其正常。结果如下：  
<img src="https://pic2.zhimg.com/eb1c17e85f7f4d4f1ac3d811ef1a9501\_b.jpg" data-rawwidth="500" data-rawheight="296" class="origin\_image zh-lightbox-thumb" width="500" data-original="https://pic2.zhimg.com/eb1c17e85f7f4d4f1ac3d811ef1a9501\_r.jpg">

· 这就更加肯定我们的推断她患有支气管炎。

· · 如果X光显示不正常的话，则结果将有很大不同：

· <img src="https://pic1.zhimg.com/4f79a2b89c9113dfbc3331862639c17c\_b.jpg" data-rawwidth="500" data-rawheight="296" class="origin\_image zh-lightbox-thumb" width="500" data-original="https://pic1.zhimg.com/4f79a2b89c9113dfbc3331862639c17c\_r.jpg">

Note the big difference. TB or Lung Cancer has shot up enormously in probability. Bronchitis is still the most probable of the three separate illnesses, but it is less than the combination hypothesis of TB or Lung Cancer. So, we would then decide to perform further tests, order blood tests, lung tissue biopsies, and so forth. Our current Bayes net does not cover those tests, but it would be easy to extend it by simply adding extra nodes as we acquire new statistics for those diagnostic procedures. And we do not need to throw away any part of the previous net. This is another powerful feature of Bayes nets. They are easily extended (or reduced, simplified) to suit your changing needs and your changing knowledge.

Summary

In this section we learned that a Bayesian network is a mathematically rigorous way to model a world, one which is flexible and adaptable to whatever degree of knowledge you have, and one which is computationally efficient.

注意最大的区别。结核病或肺癌增加的概率极大。支气管炎仍然是三个独立的疾病中最可能的一个,但它小于"结核或肺癌"这一组合的假设。所以,我们将决定进行进一步测试,血液测试,肺组织活检,等等。我们当前的贝叶斯网不包括这些测试,但它很容易扩展,只需添加额外的节点作为我们获得新的统计数据的诊断程序。我们不需要扔掉以前的任何部分。这是贝叶斯网的另一个强大的功能。他们很容易扩展(或减少,简化),以适应不断变化的需求和变化的知识。

总结

在本节中，我们了解到，贝叶斯网络是一个用严格的数学方法来模拟一个世界的方法，是灵活的，适应于任何你拥有的知识程度的方法，同时也是计算效率的方法。

本文引用地址：[http://blog.sciencenet.cn/blog-82650-255141.html](//link.zhihu.com/?target=http://blog.sciencenet.cn/blog-82650-255141.html" \t "_blank) 此文来自科学网邱士利博客

作者：小杰  
链接：http://www.zhihu.com/question/28006799/answer/38996563  
来源：知乎  
著作权归作者所有。商业转载请联系作者获得授权，非商业转载请注明出处。