****

人工智能与机器学习

实 验 报 告

|  |  |
| --- | --- |
| **学 院** | 网络空间安全学院 |
| **专 业** | 网络空间安全 |
| **班 级** | 23270312 |
| **学 号** | 23280343 |
| **学生姓名** | 王书利 |
| **教师姓名** | 黄娜娜 |
| **完成日期** | 2025/05/19 |
| **成 绩** |  |
| **实验四 贝叶斯推理** | | |
| 1. **实验目的**   1. 掌握贝叶斯网络的基本概念和构建方法；  2. 理解条件概率表(CPT)的定义和使用；  3. 学习如何将实际问题转化为贝叶斯网络表示；  4. 掌握贝叶斯网络的概率推理方法。 | | |
| 1. **实验内容**   贝叶斯网络是一种概率图模型，用于表示一组随机变量及其条件依赖关系。它由节点（代表随机变量）和有向边（表示依赖关系）组成。本实验将通过构造一个简单的贝叶斯网络来模拟吸烟、感冒、气管炎、咳嗽、气喘和发烧之间的关系，并进行推理分析。  根据给定的医学知识构建一个贝叶斯网络，具体要求如下：  1.确定网络中的节点及其相互关系；  2.建立正确的网络结构（有向无环图）；  3.为每个节点构建条件概率表(CPT)；  4.实现基本的概率推理计算。  **实验数据**  **给定的医学知识：**   * 吸烟可能会患气管炎； * 感冒也会引起气管发炎，并伴有发烧、头痛等症状； * 气管炎又会有咳嗽或气喘等症状。   **给定的概率数据：**   * 吸烟 P(吸烟)=0.6，P(不吸烟)=0.4 * 感冒 P(感冒)=0.8，P(没感冒)=0.2 * 气管炎 P(气管炎∣吸烟,感冒)=0.35，P(气管炎∣不吸烟,感冒)=0.25，P(气管炎∣吸烟,没感冒)=0.011，P(气管炎∣不吸烟,没感冒)=0.002 * 咳嗽 P(咳嗽∣气管炎)=0.85，P(咳嗽∣非气管炎)=0.15 * 气喘 P(气喘∣气管炎)=0.50，P(气喘∣非气管炎)=0.10 | | |
| 1. **实验环境**   硬件配置：  CPU：AMD Ryzen 9 7945HX with Radeon Graphics @ 2.50 GHz  RAM：16GB  存储：1T SSD  显卡：NVIDIA GeForce RTX 4060 Laptop GPU  软件配置：  操作系统：Windows 11  编程语言：python 3.10.16  IDE：vscode | | |
| 1. **主要操作步骤及实验结果记录**    1. 网络结构图   Smoking Cold  \ /  \ /  v v  Bronchitis  / \  v v  Cough Asthma  图1.有向无环图(DAG)  1. Smoking -> Bronchitis：吸烟会增加气管炎发生的风险  2. Cold -> Bronchitis：感冒也可能引起气管炎  3. Bronchitis -> Cough：气管炎会导致咳嗽  4. Bronchitis -> Asthma：气管炎会引起气喘   * 1. 条件概率表   | Smoking | 概率 |  | ----------- | ----------- |  | 是（Yes） | 0.6 |  | 否（No） | 0.4 |  | Cold | 概率 |  | ---------- | --------- |  | 是（Yes） | 0.8 |  | 否（No） | 0.2 |  | Smoking | Cold | 气管炎 = 是 | 气管炎 = 否 |  | ------------------ | -------------- | ------------ | ---------- |  | 是（Yes） | 是（Yes） | 0.35 | 0.65 |  | 否（No） | 是（Yes） | 0.25 | 0.75 |  | 是（Yes） | 否（No） | 0.011 | 0.989 |  | 否（No） | 否（No） | 0.002 | 0.998 |  | 气管炎 | 咳嗽 = 是 | 咳嗽 = 否 |  | --------------- | ---------- | --------- |  | 是（Yes） | 0.85 | 0.15 |  | 否（No） | 0.15 | 0.85 |  | 气管炎 | 气喘 = 是 | 气喘 = 否 |  | ---------------- | --------- | --------- |  | 是（Yes） | 0.50 | 0.50 |  | 否（No） | 0.10 | 0.90 |   * 1. 代码实现   1.定义网络结构  model = DiscreteBayesianNetwork([      ('Smoking', 'Bronchitis'),      ('Cold',    'Bronchitis'),      ('Bronchitis', 'Cough'),      ('Bronchitis', 'Asthma')  ])  2.定义各节点的CPT  cpd\_S = TabularCPD(      variable='Smoking',      variable\_card=2,      values=[[0.4], [0.6]],      state\_names={'Smoking': ['No','Yes']}  )  cpd\_C = TabularCPD(      variable='Cold',      variable\_card=2,      values=[[0.2], [0.8]],      state\_names={'Cold': ['No','Yes']}  )  cpd\_B = TabularCPD(      variable='Bronchitis',      variable\_card=2,      values=[          [1-0.002, 1-0.011, 1-0.25, 1-0.35],          [   0.002,    0.011,    0.25,    0.35]      ],      evidence=['Smoking','Cold'],      evidence\_card=[2,2],      state\_names={          'Bronchitis': ['No','Yes'],          'Smoking':     ['No','Yes'],          'Cold':        ['No','Yes']      }  )  cpd\_G = TabularCPD(      variable='Cough',      variable\_card=2,      values=[[1-0.15, 1-0.85], [0.15, 0.85]],      evidence=['Bronchitis'],      evidence\_card=[2],      state\_names={'Cough': ['No','Yes'], 'Bronchitis': ['No','Yes']}  )  cpd\_A = TabularCPD(      variable='Asthma',      variable\_card=2,      values=[[1-0.10, 1-0.50], [0.10, 0.50]],      evidence=['Bronchitis'],      evidence\_card=[2],      state\_names={'Asthma': ['No','Yes'], 'Bronchitis': ['No','Yes']}  )  3.将CPD加入模型并检查  model.add\_cpds(cpd\_S, cpd\_C, cpd\_B, cpd\_G, cpd\_A)  model.check\_model()  4.进行推理  infer = VariableElimination(model)  result = infer.query(      variables=['Bronchitis'],      evidence={'Smoking':'Yes','Cold':'Yes'}  )  print(result)   * 1. 推理计算     即   * 1. 结果分析   在已知“吸烟=是”和“感冒=是”的条件下，后验概率 P(气管炎=是∣吸烟=是,  感冒=是)与我们在 CPT 中直接给出的 0.35 完全一致，这说明当所有上游条件都确定时，查询结果不会被其它变量边缘化所改变。整个网络结构充分体现了吸烟和感冒对气管炎的影响，以及气管炎对咳嗽和气喘的直接作用，参数值也符合医学常识：双重诱因下的气管炎风险最高，而在既不吸烟也未感冒时几乎为零。若后续需要在仅观测到咳嗽或气喘时反推吸烟或感冒状态，可利用 VariableElimination 自动完成多次边缘化与归一化计算。 | | |
| 1. **实验分析总结及心得**   通过本次贝叶斯网络的构建与推理实验，我深入理解了贝叶斯网络的基本概念以及它在表示和处理不确定性问题时的优势。首先，在搭建网络结构的过程中，我体会到如何根据领域知识判断变量之间的因果关系——比如明确吸烟和感冒会共同影响气管炎，再由气管炎推导出咳嗽和气喘这两个症状。这样的有向无环图不仅直观，而且为后续的概率计算提供了清晰的依赖路径。  在编写 CPT 并用 pgmpy 库实现推理时，我对“证据变量”和“查询变量”之间的关系有了更深刻的认识。尤其是当给定“吸烟=是，感冒=是”这两个证据时，结果直接对应 CPT 中的 0.35，这让我意识到在所有父节点都被观测的情况下，后验分布无需额外边缘化就能直接读取。这一特点大大简化了计算流程，也让我初步体会到贝叶斯网络在实际应用中的高效性。  当然，实验中也遇到了一些小问题：库版本兼容。不同版本的 pgmpy 对类名和模块路径有调整，需要在调试时多加留意  总的来说，这次实验让我既巩固了理论知识，又提升了动手能力。今后若要处理更复杂的真实场景，比如加入更多症状节点或对缺失数据进行推断，我会更加得心应手，也愿意尝试不同的推理算法（如采样法）来比较性能与精度。贝叶斯网络作为一种强大的图模型，已经为我打开了概率推理的新视野，后续学习中我会继续深入探索它在医学诊断、故障检测等领域中的应用。 | | |