

绮算法-基于加权知识图谱的宠物智能问诊系统

数学原理报告

本系统旨在通过多轮问答交互，结合宠物主人提供的主诉信息，逐步推断宠物可能患有的疾病。为实现推理过程的可解释性、结构化和可扩展性，我们采用了**加权知识图谱**作为核心数据结构，统一建模疾病、症状、属性之间的概率关系，并在此基础上构建贝叶斯风格的诊断机制。

一、图谱结构设计

1.1 节点集合

1) 疾病节点 Disease Node

- 每个疾病对应一个图节点 D_i
- 属性：`name`: 疾病名称; `prior_prob`: 疾病先验概率 $P(D_i)$

2) 症状节点 Symptom Node

- 每个症状（含具体属性）为图节点 S_j
- 属性：

`description`: 症状描述

`prior_prob`: 症状先验概率 $P(S_j)$

`category`, `aspects`, `attribute_values`: 元信息

1.2 图谱边

1) 疾病→症状:

- 从疾病 D_i 指向症状 S_j
- 表示疾病导致某症状的条件概率
- 边权重:

$$w_{D_i \rightarrow S_j} = P(S_j | D_i)$$

2) 症状↔症状:

- 两个有共现关系症状之间添加**无向边**
- 边权重为联合概率:

$$w_{S_i \leftrightarrow S_j} = P(S_i | S_j)$$

二、基于贝叶斯框架的推理过程数学建模

在问诊过程中，核心是计算疾病的后验概率：

$$P(D \mid \text{observed symptoms})$$

采用多轮贝叶斯更新的形式，结合症状之间的关系建模如下。

2.1 初始后验估计

如果已知某症状 S_1 ，则对每个疾病 D 计算：

$$P(D \mid S_1) = \frac{P(D) \cdot P(S_1 \mid D)}{P(S_1)}$$

其中：

$P(D)$ ：疾病的先验概率，由疾病节点属性获得

$P(S_1 \mid D_i)$ ：每个疾病下该症状概率，由边 $D_i \rightarrow S_j$ 的权重获得

$P(S_1)$ ：症状概率，由症状节点属性获得

2.2 后续症状的迭代更新

对于多个症状 S_1, \dots, S_n ，使用联合症状建模：

$$P(D \mid S_1, S_2, \dots, S_n) \propto P(D) \cdot \prod_{i=1}^n \left[\frac{P(S_i \mid D)}{P(S_i \mid S_{<i})} \right]$$

其中：

$P(S_i \mid D)$ ：由图谱边 $D \rightarrow S_i$ 提供

$P(S_i \mid S_{<i})$ ：建模为：

$$\max_{j < i} \left[\frac{P(S_i, S_j)}{P(S_j)} \right] = \max_{j < i} P(S_i \mid S_j)$$

三、症状推荐建模

传统策略以**疾病后验概率提升最大化**为依据进行推荐，等价于寻找具有最大信息增益的症状。本报告基于该原则，进一步结合**症状-症状联合概率图谱结构**，提出一种图谱增强的推荐方法。

3.1 基础推荐逻辑：信息增益驱动

基于信息增益的推荐函数如下：

$$S^* = \arg \max_{S_j \in \text{未问}} \Delta P(D \mid S_{\text{known}} \cup S_j)$$

即：在当前 top-k 疾病集合 $D \in D_{\text{top}}$ 中，选择一个尚未确认的症状 S_j ，使得将其加入到现有信息后，可以**最大提升某个疾病的后验概率**。

数学表达：

对于每个候选疾病 D 和候选症状 S_j

$$P(D | S_{\text{known}} \cup S_j) \propto P(D) \cdot \prod_{S_i \in S_{\text{known}} \cup S_j} P(S_i | D)$$

3.2 增强策略：症状共现驱动（图谱驱动）

为克服疾病标签稀疏、先验不准等问题，我们引入**症状-症状共现图谱**，其边权表示 $P(S_j | S_i)$ ，即症状之间的联合概率关系。

当前确认症状集合 S_{known} 。候选症状集合 $S_{\text{candidate}} = S \setminus (S_{\text{known}} \cup S_{\text{not}})$

我们定义候选症状的共现得分如下：

$$\text{score}(S_j) = \sum_{S_i \in S_{\text{known}}} P(S_j | S_i)$$

最终推荐为：

$$S^* = \arg \max_{S_j \in S_{\text{candidate}}} \text{score}(S_j)$$

3.3 混合策略：症状共现驱动（图谱驱动）

为同时结合“疾病后验概率提升”与“共现强度”，我们定义一个**混合推荐评分函数**：

$$\text{final_score}(S_j) = \lambda \cdot \underbrace{\Delta P(D | S_j)}_{\text{信息增益}} + (1 - \lambda) \cdot \underbrace{\sum_{S_i \in S_{\text{known}}} P(S_j | S_i)}_{\text{共现关系}}$$

四、负症状推荐建模

如果症状 S 被用户否认：

- 将其对后验概率的影响建模为：

$$\text{cond_factor} = \frac{1 - P(S)}{1 - P(S | D)} \in (0, 1)$$

用于削弱疾病概率，但不会完全否定。

五、系统终止条件

系统满足以下任一条件即停止问诊：

- 当前后验概率最高的疾病 D^* 满足：

$$P(D^* | E) > \theta$$

- 已询问次数达到最大限制 T