

疾病诊断专家系统项目说明

14307130122 戴嘉润

一、 项目设计.....	2
1.1 问题分析与解决方案.....	2
1.2 系统结构图解.....	3
1.3 知识表示与相关计算.....	4
1.3.1 疾病（科属）编码.....	4
1.3.2 症状编码.....	4
1.3.3 科属归属度规则编码与计算方式.....	5
1.3.4 规则库编码与确信因子计算.....	6
1.3.5 结果推理.....	7
二、 项目实施.....	8
2.1 开发环境与运行方式.....	8
2.2 代码注解.....	8
2.2.1 函数注解.....	8
2.2.2 变量与数据结构注解.....	9
2.3 外部编码库和规则库.....	10
三、 项目展示.....	10
3.1 专家入口.....	11
3.2 用户入口.....	12
四、 讨论.....	14
五、 参考资料.....	15

一、项目设计

1.1 问题分析与解决方案

病情的诊断作为疾病治疗的前提与基础，在医疗事业中具有十分重要的意义。本项目通过模拟医学专家诊疗的过程，结合人工智能专家系统技术，实现了一个层次化的疾病诊断系统，用于辅助医生诊疗和帮助病人自我诊断病情。

在实现上，主要是通过询问各症状以及其严重程度和确信程度，以此为依据推断病情。对于病人使用者，两者皆为主观感觉。对于医生使用者，各症状确信程度为 1，即确认病人的确存在该症状，严重程度基于量化标准或从业经验。

通常来说，医学诊疗是一个比较复杂的过程，特别是医生主导的问询式诊断。医生需要对可能存在的症状一一排查。对于专家系统来说，就需要对每一个症状都进行判断。当可能的疾病和症状的数量上升时，计算的复杂度会快速提升。

现实生活中解决该问题的方案是对先疾病进行树状分类，再进行有针对性的诊疗。如医院会将疾病分为外科和内科，对于内科会进一步细分为呼吸内科，消化内科等。最后，对于隶属于呼吸内科的各类疾病只需要问询呼吸系统症状和部分内科综合症状即可完成诊断。

将该方案进行建模时需要注意的是，该专家系统是一个综合的诊疗系统，不是为某一具体的科属而建立的。如当病人同时患有呼吸系统疾病和消化系统疾病时，系统也能进行有效的判断。同时，考虑到有无科属辨别能力的平民使用者，系统需要先从症状出发，模糊判别疾病所在科属。

1.2 系统结构图解

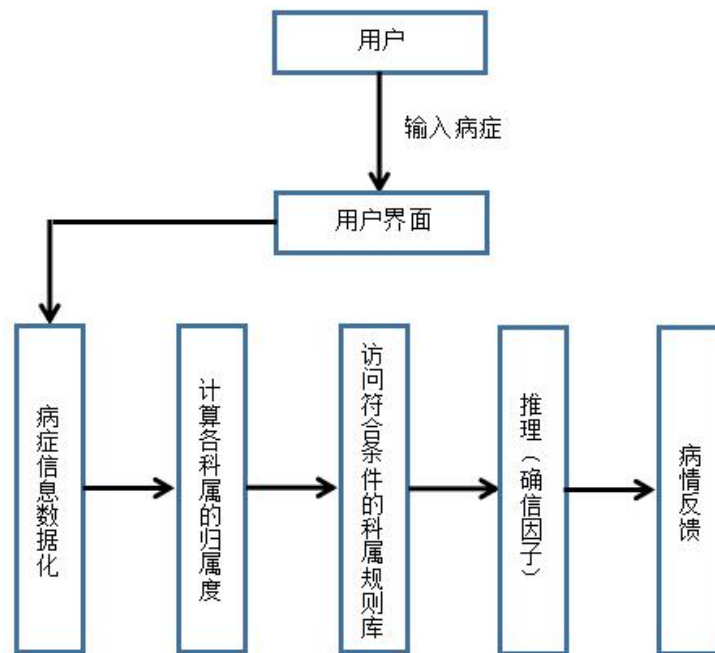


图 1.疾病诊断系统——用户接口

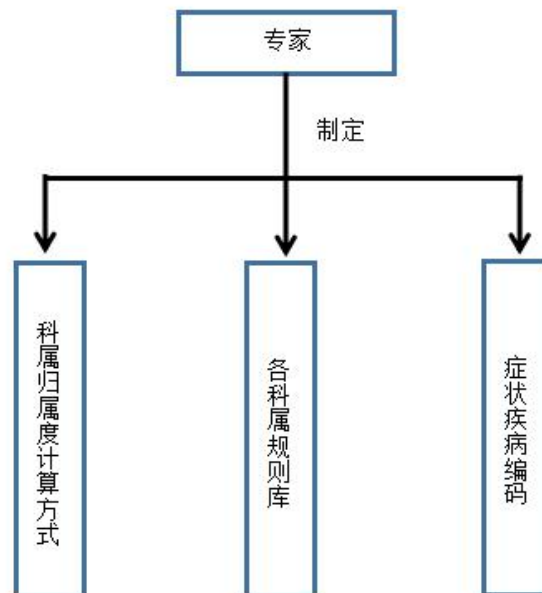


图 2.疾病诊断系统——专家接口

1.3 知识表示与相关计算

1.3.1 疾病（科属）编码

考虑到该系统层次化的特点，对于疾病的编码需要携带科属信息，即树形结构中的父亲节点。（如图）

疾病名称	疾病 ID	所属父类	父类 ID
呼吸系统疾病	1	内科疾病	0
感冒	9	呼吸系统疾病	1

图 3.疾病（科属）编码条目

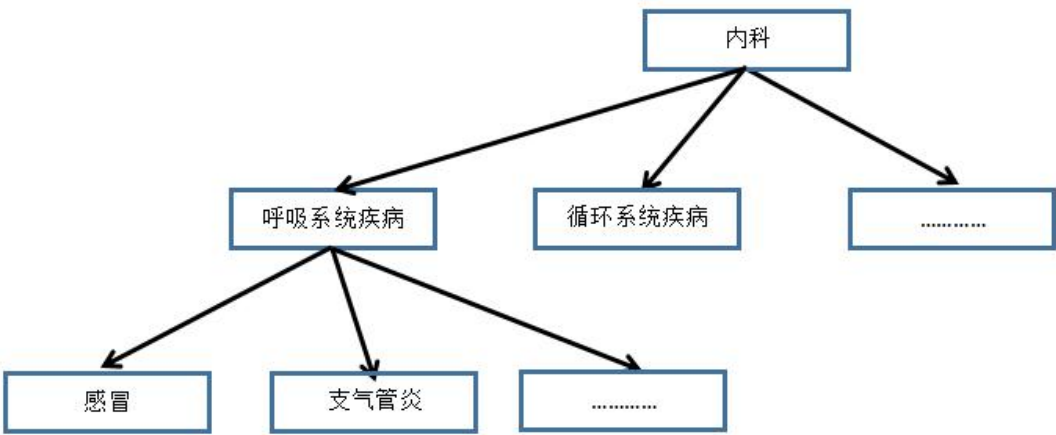


图 4.疾病（科属）编码树

1.3.2 症状编码

如上一节所说，对于用户接口，输入病症后首先要判断其可能的科属，故对于某一具体的症状，需要记录其父类症状来完成这一功能。

注：症状类型一栏中，A 表示症状树的根节点，S 代表叶子节点。即每出现一个 A 就多一棵以 A 为根节点的症状子树。（如图）

症状	症状 ID	症状类型	症状隶属
四肢异常	4	A	无
四肢酸痛	19	S	4
乏力	20	S	4
肢体麻木	33	S	4

图 5.症状编码条目

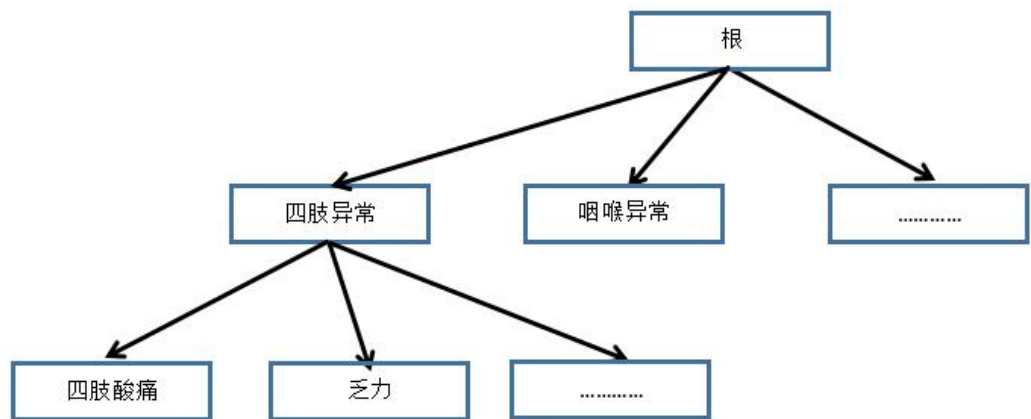


图 6.症状编码树

1.3.3 科属归属度规则编码与计算方式

当用户输入症状时，系统首先要确定其可能的科属。

用户界面输入的是具体的症状（S 症状，见 1.3.2），通过这些繁杂的症状来判断归属的话会降低计算性能，可靠性也不高。故在得到这些症状后，首先将其转换为父类症状（A 症状，见 1.3.2），并对每一父类症状设置一个特征值来表示其重要程度，若特征值为 0，则表示未出现该父类症状。

式子如下：

$U = [E(A_{(0)}), E(A_{(1)}), \dots, E(A_{(n)})]$ ， U 表示用户特征向量， $E(A_{(i)})$ 表示父类编码 ID 都第 i 大的父类症状所代表的特征值，父类症状数为 n （远远小于子类症状数）。

$E(A_{(i)}) = \sum_{S_j \in A_{(i)}} CF(S_j) * W(S_j)$ ， S_j 隶属于父类症状 $A_{(i)}$ ， $CF(S_j)$ 表示子症状 S_j 的确信程度， $W(S_j)$ 代表子症状的严重程度。

以上为根据用户信息得到的特征向量 U 。

系统中对于每一个科属都保存了其归属标准。如呼吸系统疾病（疾病 ID 为 1）。则从内科被分配到呼吸疾病科的归属标准（ST）被编码为：

$$ST_1 = [0.2, 0.7, 0.6, 0.4, 0.3, 0.5, 0.5, 0.3, 0.6, 0, 0, 0, 0]$$

注：在 Demo 中向量各维代表各 A 类症状：【长期发作，呼吸异常，咽喉异常，头晕头痛，四肢异常，咳嗽喷嚏，畏寒发热，胸腔异常，鼻腔异常，颈部异常，心绪异常，排尿异常，视力异常，出汗异常】，此处 $n=14$ 。

接下来引入归属度（Sim）的概念。

$$Sim_i = Sim(U, ST_i) = \frac{\sum_{E \in U, E' \in ST_i} E(A_{(j)}) * E'(A_{(j)})}{\sqrt{\sum_{E \in U} E^2(A_{(j)})} * \sqrt{\sum_{E' \in ST_i} E'^2(A_{(i)})}} : \text{表示用户输入对于科属 } i \text{ 的归属度}$$

属度，即向量 U 和 ST_i 的余弦相似度，归属度越大，则用户所患疾病越可能属于该科属。

在对每一个科属都计算完归属度后，设置阈值或者取 Top-k 决定用户所在科属，并启用属于该科属的规则库。

1.3.4 规则库编码与确信因子计算

考虑到疾病针对的特殊性，即患有该疾病的人不一定拥有该疾病可能出现的所有症状，并且当患者拥有更多与该疾病相符的症状时，其患该病的可能性提高。并且，对于每一种疾病，都有所谓的“关键症状”，它在疾病推断中占比较大的作用。

因此，在各科属的规则库中引入确信因子，并确定规则模式如下：

(具体的规则见附件，报告中不再列出)

IF 【Evidence 1】——【CF of Evidence 1】

AND 【Evidence 2】——【CF of Evidence 2】

AND ……

AND 【Evidence P】——【CF of Evidence P】

THEN 【Disease】——【 $CF = 0.5 + 0.5 * (P/N)$ 】

则:

$$CF(Disease, Evidence_1 \cap Evidence_2 \dots \cap Evidence_P) = \min[cf_1, cf_2, \dots, cf_P] * (0.5 + 0.5 * P/N)$$

对于结果的确信度是考虑到条件数目的不同，直接影响到结果的推断。

例如，同时患有喘息并且长期发作的用户肯定比只患有喘息的用户得哮喘的概率大。

至于结果 CF 值的确定，可以设置其他可行的格式，甚至可以对不同 Evidence 的存在设置不同的 cf，目前的 $0.5 + 0.5 * P/N$ 只是一种为了方便理解，没有科学依据的直观表达。

为了方便读写，将以上规则进行编码。

#疾病名称#各 Evidence 的 cf 值

#哮喘#0.8,1,1,1,0.3,0,0,0,0,0,0,0,0,0,0

其 cf 对应的 Evidence 分别为：【反复发作,喘息,气促,胸闷,咳嗽,寒气受激咳嗽,咳痰,畏寒,发热,头痛,四肢酸痛,乏力,咽痒,胸骨后疼痛,喷嚏,鼻塞流涕，咽痛】

因为哮喘隶属于呼吸系统疾病，故以上 Evidence 实际上是所有呼吸系统疾病共用的。

但此时即使用户有排尿异常的父类症状，也不会作为 Evidence，因为排尿在该科属（呼吸系统疾病）的标准归属向量中值为 0。

1.3.5 结果推理

综合上述所有操作，对所有的可能的科属中的所有疾病计算一个结果值（Res）。

$$Res(Disease) = CF(Disease, Evidence_1 \cap Evidence_2 \dots \cap Evidence_P) * Sim(U, ST_i)$$

设置阈值后对以上结果值进行排序，值越大，患该病的可能性越大。

在最后的实现上，对 Res 的值赋予了含义，设置数个阈值得出“很有可能”，“可能”，“几乎不可能”，“证据不足”，来描述患病可能性。

二、项目实施

2.1 开发环境与运行方式

开发环境（Windows/vs2015）：项目逻辑采用 C++ 语言编写，调用 BOOST 库。前端用 QT 编写。

运行方式：将 QT 程序打包为 ExpertSystem.exe，双击运行即可。具体的使用方式会在下一节中细谈。

2.2 代码注解

2.2.1 函数注解

```
//功能函数 类型转换
string wstr2str(const wstring &wstr); //wstr to str
wstring str2wstr(const string& s); //str to wstr

//前端接口
void ui_map_change(string symptom, double cf, double degree); //
前端症状输入监听
void ui_map_output(); //查看所有输入症状

//专家系统准备，读取各规则库和编码库
bool read_symptomID(string filepath, string filename);
bool read_diseaseID(string filepath, string filename);
bool read_diseaseType(string filepath, string filename);
bool read_rule_A(string filepath, string filename);
bool read_rule_S(string filepath, string filename);
int system_prepare();

//结果推断
void user_interface();
```



```

void view_result();

//变量重置
void reset_environment();

```

2.2.2 变量与数据结构注解

```

map<string, double> inputcf;// UI 接口
map<string, double> inputdegree;
set<string> inputstr;
priority_queue<UIRes> ui_result;

map<string, diseaseIDInfo> diseaseID;//疾病编码
map<string, set<string>> diseaseType;//疾病类别树

map<string, symptomIDInfo> symptomID;//symptomName -
(symptomType(A, S), symptomID)

map<string, EvidenceSpecific> secondlevel;//各科属规则库

int evidence_abstract_size = 0;//主病类数量
int evidence_abstract[MAX_SYMPTOM_TYPE] = { 0 };//第一层次分类
主病症指标
map<string, SystemCfVector> firstlevelRule;//第一层分类的各项
CF 如呼吸系统疾病- [0.8, 0.7……]

double FirstLevelThreshold;//归属度阈值
double SecondLevelThreshold;//结果值阈值
priority_queue<FirstLevelRes> firstlevelres;//科属归属度
priority_queue<SecondLevelRes> secondlevelres;//疾病推断结果值

```

2.3 外部编码库和规则库

（详情见附件，略）

disease_symptom_A.txt: 父类症状编码

disease_symptom_S.txt: 子类症状编码（包括其所属父类）

diseaseID.txt: 疾病（科属）编码

diseaseType.txt: 各疾病所属父类

symptom.txt: 症状编码

三、项目展示

最终实现的疾病诊断专家系统包括专家入口和用户入口。

专家入口用于修改各种编码和规则。

用户入口用于进行疾病诊断。



图 7.系统主界面

3.1 专家入口

进入专家入口后，可以在文本编辑框右侧选择需要修改的编码库或规则库，修改结束后点击 **Save** 即可保存修改。

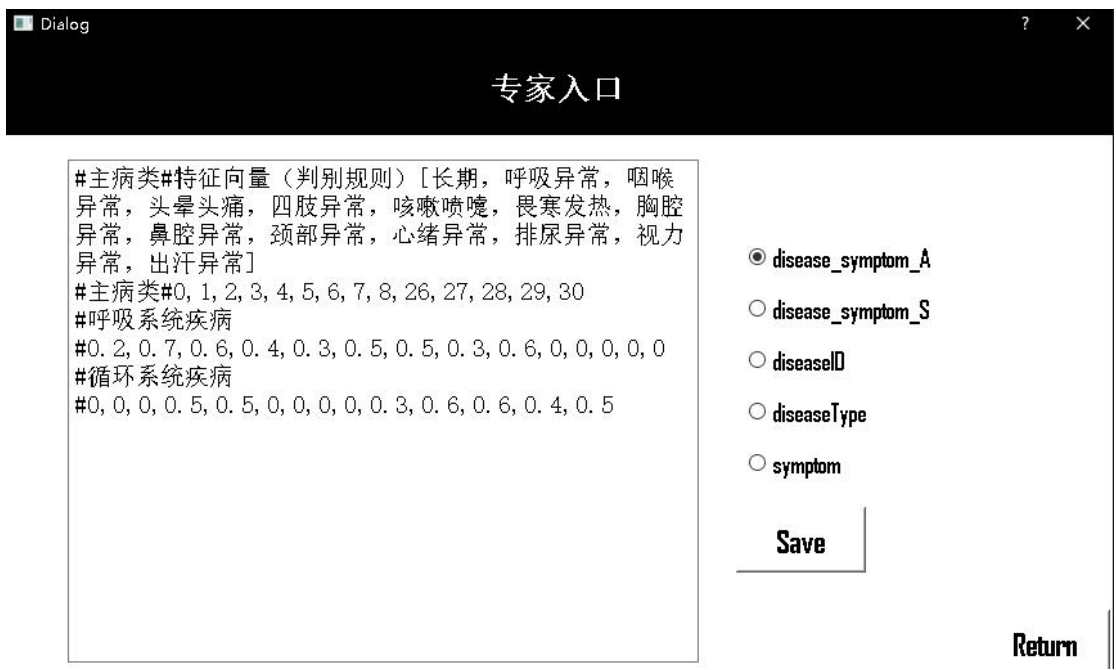


图 8.专家入口



图 9.完成库修改

3.2 用户入口

在该 Demo 中，只保留了部分症状用于实际展示。

Dialog
?
X

用户入口

格式：症状 确信程度[0,1] 症状严重程度[0,1]

长期

反复发作
0.00
0.00

呼吸异常

喘息
0.00
0.00

气促
0.00
0.00

咽喉异常

咽痛
0.00
0.00

咽痒
0.00
0.00

头晕头痛

头晕
0.00
0.00

头痛
0.00
0.00

四肢异常

乏力
0.00
0.00

四肢酸痛
0.00
0.00

肢软
0.00
0.00

肢体麻木
0.00
0.00

咳嗽喷嚏

咳嗽
0.00
0.00

喷嚏
0.00
0.00

寒气受激咳嗽
0.00
0.00

畏寒发热

畏寒
0.00
0.00

发热
0.00
0.00

胸腔异常

胸闷
0.00
0.00

胸骨后疼痛
0.00
0.00

鼻腔异常

鼻塞流涕
0.00
0.00

颈部异常

颈项板紧
0.00
0.00

心绪异常

心悸
0.00
0.00

排尿异常

夜尿增多
0.00
0.00

少尿
0.00
0.00

视力异常

眼黑
0.00
0.00

视力模糊
0.00
0.00

出汗异常

冷汗
0.00
0.00

身体检测

Return

图 10.用户入口

用户对于出现的症状，分别确定其确信度和严重度（注意，若存在该症状，需保证两者都不为 0，否则视为无该症状）。完成输入后，单击“身体检测”按钮进行测试。

例如，输入：

【咽痛，1，0.8】，【头晕，1，0.4】，【鼻塞流涕 1，1】。

则结果为：

Dialog
?
X

测试结果

很有可能：
感冒

长期

反复发作
0.00
0.00

呼吸异常

喘息
0.00
0.00

气促
0.00
0.00

咽喉异常

咽痛
0.00
0.00

咽痒
0.00
0.00

头晕头痛

头晕
0.00
0.00

头痛
0.00
0.00

四肢异常

乏力
0.00
0.00

四肢酸痛
0.00
0.00

肢软
0.00
0.00

肢体麻木
0.00
0.00

咳嗽喷嚏

咳嗽
0.00
0.00

喷嚏
0.00
0.00

寒气受激咳嗽
0.00
0.00

畏寒发热

畏寒
0.00
0.00

发热
0.00
0.00

胸腔异常

胸闷
0.00
0.00

胸骨后疼痛
0.00
0.00

鼻腔异常

鼻塞流涕
0.00
0.00

颈部异常

颈项板紧
0.00
0.00

心绪异常

心悸
0.00
0.00

排尿异常

夜尿增多
0.00
0.00

少尿
0.00
0.00

视力异常

眼黑
0.00
0.00

视力模糊
0.00
0.00

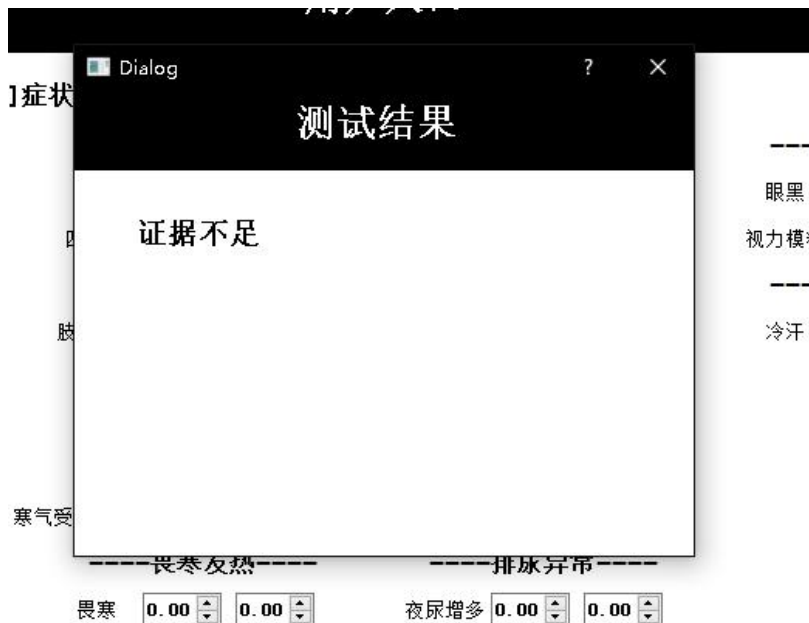
出汗异常

冷汗
0.00
0.00

例如，输入：

【反复发作，0.6，0.6】

则结果为：



四、讨论

总的来说，这是一个有效降低了计算复杂性的病情诊断专家系统。在应用上，可以辅助医生治疗，快速得到模糊评估，也可以帮助无医学知识的人自我诊断。

在实现的亮点上，树状的编码格式，层次化的计算结构为其提供了很好的扩展空间和计算性能。

当然也有尚待完善的地方：

（1）如其规则与确信度设置需要大量的病历数据和医学知识的支持，但因为开发条件有限，这一部分到目前为止是没有达到应有标准的。

（2）对于归属度的计算，比较简单的使用了余弦相似度，设置了一个各维度平等的比较空间。但是，即便是父类病症，对于每个科属也是具有特殊性的。这部分可以通过权重的思想解决。

五、参考资料

- 【1】一个仿人疾病诊断专家系统模型_张红梅
- 【2】基于 C 语言建造专家系统的研究_李月娥
- 【3】用 C 语言开发地基基础专家系统_娄常青