

关联索引式知识表示及其正反混合推理

朱 浩 (解放军第92医院微机室)

邓德顺 (厦门大学计算机科学系)

摘 要

尽管基于规则的专家系统已有不少,但这种知识表示法在表达大范围临床医学知识时却有一定困难和不足。本文提出一种新的关联索引式知识表示方法,可以很方便地表示这种大范围的临床医学知识,且表达效率较高。也讨论了以这种知识表示方法为基础的正反混合推理,这种推理方式具有并行推理的特性,同时还考虑和处理了每个证据的全部意义,从而能动态地产生和处理多个暂时目标。这较为符合临床专家的诊断过程。

一、引 言

医疗诊断是专家系统应用领域之一。继MYCIN之后,相继推出了许多其它系统^[1],它们大多数都使用基于规则的知识表示法来诊断某一系统范围内的疾病,即整个疾病树的某一小子树,如细菌感染性疾病、风湿病、肿瘤等。之所以采用基于规则的知识表示,是因为一般认为临床医学知识主要是经验性的,这种知识较宜用规则表示。但是,临床上各种疾病及其症状数量颇为繁多,形成一个多因多果的巨大因果网络,若要构造一个具有较全面专业知识的专家系统,真正体现出专家系统的实用价值,那么基于规则的知识表示法便显得力不从心,这表现在:①一条规则的规模有限,难以表述一个症状或疾病的全部有关知识,②规则是固定不变的,而医

疗诊断是证据不充分推理,所缺省的证据因人而异,随机不定,具有很大的自由性,③医疗诊断是一种并行推理,对每一个症状都应同时考虑到与该症状相关的所有疾病,必须并行处理每个症状的全部意义,这在诊断过程的初期阶段尤为重要,因此,在每一推理阶段都动态地存在着多个可能目标。而基于规则的专家系统其推理方法一般是假定单一目标,逐条启动规则线性地推进的,无法做到全局并行推理。对于上述不足,一些晚些发表的系统如INTERNIST-1, CASNET改用了网络式知识表示。虽然知识库性能有了较大改进,但仍不是全局并行推理^[2,3,4],本文介绍笔者设想并用的一种新的知识表示方法及相应的正反混合并行推理,其效果较好。

架”和“一般规则”三种知识表示方法,这三种方法对断言的描述除“一般规则”中的前提部分以外,都采用框架表示。其中“规则/框架”表示方法是以规则为骨架,规则的前提与结论都是由断言框架或它们的逻辑组合而成;而本文所论述的“类框架”表示法其结构与框架一致,但是所描述的各属性之间还可根据需要具有某些逻辑关系;“一般规则”表示法中规则的前提为字符串,而结论部分与“规则/框架”表示相同。这三种知识表示方法都是对现有的一些知识表示方法作进一步改进与扩充后得到的,目前

尚未见有这类表示方法的报导。系统还提供了这几种知识表示方法的编辑、维护功能。这三种知识表示方法可以十分简炼与确切地表达地震预报领域的各类知识,实际使用效果良好。

地震预报专家系统ESEP/PC由安徽省地震局与吉林大学计算机系合作完成,已于1989年11月通过国家地震局组织的专家技术鉴定。赵瑞清教授审阅了本文并提出了宝贵意见,借此表示衷心感谢。

参考文献

[1]赵瑞清,专家系统原理,气象出版社,1987。

二、关联索引式知识表示

假设症状与疾病组成一个 $m:n$ 型因果网络,用 F 表示全部症状, D 表示全部疾病,某一症状 f_i 与 n_i 种疾病相关,由 n_i 种疾病构成疾病组 $D_i \in D$,对于不同的症状有不同的 n_i 及 D_i ,且 D_i 之间可有非空交集。同样,某一疾病 d_i 可能有 m_i 个症状,由 m_i 个症状构成症状组 $F_i \in F$,对于不同的疾病有不同的 m_i 及 F_i ,且 F_i 之间可能有非空交集。必须注意的是,疾病 d_i 可能有 m_i 个症状,而 m_i 个症状又只是这个疾病 d_i 的症状谱。在实际情况中 d_i 只表现为 m^* 个症状,由 m^* 个症状构成症状组 F^*_i , $m^* < m_i$, $F^*_i \subset F_i$,而 m^*_i 和 F^*_i 又是因人(同样患疾病 d_i)而异的,即患有相同疾病的人症状表现不完全相同,这就是证据缺省的随机性。把一个症状及与之相关的全部疾病并联在一起组成一条症状知识, m 个症状就有 m 条症状知识;把一个疾病及与之相关的全部症状关联在一起组成一条疾病知识, n 个疾病就有 n 条疾病知识;所以整个知识库共有 $m+n$ 条知识。把症状知识和疾病知识称为知识概念(**concept**),把症状知识和疾病知识称为知识单元(**unit**),并给每个知识单元赋以一个地址。如果每个知识单元中与该知识概念相关的全部疾病(或症状)都用字符显式地表示,则一条知识单元的体积将极为庞大,为使知识库不显得臃肿,我们在每条知识单元中将与该知识概念相关的全部症状或疾病用相应症状知识或疾病知识的知识单元地址来表示,由这些地址指向相应的各症状或疾病,从而使全部症状知识与疾病知识相互间建立起双向地址索引关系,这就是所谓的知识单元地址索引表示法。这种[知识单元地址+知识概念+相关知识地址索引表]的结构与神经元的轴突+胞体+树突的结构有很大的相似性。

一个疾病所表现出的许多症状的重要性是不尽相同的,因此必须规定每一症状对各相关疾病的诊断意义值,这个值可由领域专家确定,并与相应的疾病知识单元一起组成地址-意义值对,作为相关知识地址索引表的表单元。

我们在汉化的Turbo Prolog 1.1版上使用这种知识表示法设计了一个医疗诊断专家系统MED1,其知识单元的结构为

```
unit(integer, integer, string, list)
```

list即为带意义值的相关知识地址索引表,定义为

```
list = item•
```

```
item = integer•
```

症状知识与疾病知识可分别表示为

```
unit(f-address, ask-no, finding, diseaseslist)
```

```
unit(d-address, —, disease, findingslist)
```

三、正反混合并行性推理

医疗诊断的前件一般具有模糊性与多义性,模糊性是指许多疾病诊断的前件因诊者而异,且无一致标准。对具有某组症状的一个病人,有的医师认为满足某病的诊断,有的则认为尚不满足。多义性是指,假设 P 个人患有疾病 d_i ,则这些人分别有症状组 $F^*_i, F^*_i, \dots, F^*_i, F^*_i \subset F_i, 1 \leq j \leq p$,且一般有 $F^*_i \subset F^*_i, 1 \leq j, k \leq p$,即这些患者所表现出的症状是不尽相同的,因为个体差异及发病条件不同等多种原因。所以固定的IF-THEN-结构不适应这种模糊多义的诊断标准,且一个人可能并发多种疾病,症状交织,这也与一般的单一原因源问题的诊断有很大不同,因此我们采用一种动态并行累计全部相关疾病意义值的办法,并行处理每个症状的全局意义。若某疾病的意义值累计超过临界值,则该病诊断成立,其诊断推理过程如下:

A.首先建立2个黑板结构:黑板1与黑板2,在黑板1上存放并运算推理过程中所产生的所有相关疾病的地址-意义值对,黑板2上存放推理过程中获得的全部症状(这里把主诉症状、体检体征、检查结果统称为症状)的地址。

B.获取第1个症状,将该症状的diseaseslist写到黑板1上,将该症状的地址写到黑板2上。

C.获取症状,在黑板1上处理该症状的diseaseslist,这分二种情况:①若某疾病在黑板1上已存在,则累计其意义值,②若该病在黑板1上不存在,则把该病的地址-意义值对写到黑板1,把该症状单元的地址写到黑板2,记下已处理过的症状,即推理过程。

D.重复过程C,直至处理完病人的现有症状。

E.对黑板1上的全部疾病按意义值大小排序,得到意义值累计最高的 s 个疾病,作为初步目标。

F.处理初步目标的findingslist,对每一症状地址-意义值对,先查看黑板2,若黑板2上已有,则跳过;若黑板2上没有,就由专家系统提问该症状,获取病人的回答。回答可能有三种情况:①肯定,这时执行过程C,②否定,直接将症状地址记上黑板2,③不确定,直接跳过。

G.重复过程F,直至处理完 s 个初步目标。

H. 对黑板1上的全部疾病重新按意义值大小排序, 意义值超过临界值的疾病即为成立的诊断结论。若无一疾病意义值超过临界值, 可中止诊断推理或再执行过程F。

可以看出, 这个推理过程与临床诊断过程十分相似, 首先由数据驱动正向推理产生一系列初步目标, 然后选取可能性最大的数个初步目标, 驱动反向推理来详细检验这些目标并同时考虑与处理随之可能产生的新目标。显然, 在推理过程中所产生的目标总数可能是很多的, 对这些目标能否形成有效的约束是值得注意的。实际上, 约束是自动产生的, 由于每种疾病内在的解剖病理生理原因, 其表现的一组症状有一定的必然性与相关性, 这种核心症状是主要诊断依据。诊断标准的多义性是由非核心症状造成的, 一个疾病与一组核心症状相对应, 核心症状组起了主要的约束作用。若随意输入若干毫无关联的症状, 则约束可能失败, 从而各初步目标的意义值无明显差异。

四、应用

一般认为知识获取与建库是建造专家系统的瓶颈。因为把专家提供的庞杂粗糙的原始知识整理加工组织成一条条具有内在联系的规则需耗费大量的工作; 使用本方法情况大为改善, 因知识单元的结构简明规整, 无需象规则那样注意相互间连接, 易于直接从领域专家获得符合结构要求的知识。其次本法具有较好的通用性, 因为与基于规则的代表法

不同, 知识单元的结构不因知识内容的改变而改变, 易于扩充或更换知识库。关于学习我们将另文讨论。此外我们也正在考虑意义值累计算法的改进、多层次关联索引式知识表示及关联索引式知识表示与基于规则的知识表示的结合等问题。

参考文献

- [1] Waterman, D. A., A guide to expert system, Addison-wesley Publishing Co. 1986
- [2] Randolph, A. M., et al, Internist-1, an experimental computer-based diagnostic consultant for general internal medicine, The New England Journal of Medicine, 1982, Aug.
- [3] Alty, J. L. and Coombs, M. T., Associative and causal approaches to diagnosis-INTERNIST-1 and CASNET, Expert system concepts and examples, Manchester, NCC Publications, 1984
- [4] Buchanan, B. and Shortliffe, E., Rule-based expert systems, Addison-wesley, 1984
- [5] Fagan, L. M., et al, Representation of dynamic clinical knowledge, measurement interpretation in the intensive care unit, Proceeding IJCAI-79, 1979