

基于中医药数据仓库智能查询的研究

蔡永铭 周怡

广东药学院 医药信息工程教研室 510006

摘要: 为使中医药数据仓库更好地为决策支持服务, 本文提出了一种面向中医药知识的智能查询方法, 增强了中药数据仓库查询的智能化程度, 使用户能够通过单味药、中药复方、疾病、症状等数据库的互相关联对中药信息进行管理, 输入任一关键字, 即刻显示出中药复方、功能主治、组成和成分等, 并能将具有某一功效的单味药按密集度和功效权重量化进行排序。

关键词: 知识库 数据仓库 智能查询

1 引言

目前国内外已经建立了许多中医药数据仓库, 但所有这些数据库和数据管理系统缺乏对中药信息的综合分析和智能检索能力, 不能对中药及其有效部位或有效成分的活性和功能主治进行定量和权重排序, 对支持中药的科学研究和产品的理性开发能力有限。数据仓库的建立为用户决策提供了坚实的数据查询和分析平台, 现在用户主要由数据仓库的附属工具从数据仓库中获得这些信息, 如综合查询工具、OLAP分析工具和数据挖掘工具, 但这些工具一般不具备智能化支持, 而且需要用户具有相应的领域知识, 要掌握存取和检索数据仓库中数据的方法, 还要理解数据仓库中检索出来的数据的语义, 并且有能力将数据转换成事实和信息, 再把事实转换成知识, 最后再用知识来进行决策或提供建议。然而, 对于非计算机专业用户, 使用上述方法进行查询是比较困难的, 而且用户查询的概念需在数据仓库中存在, 否则无法获得结果, 如对于一些如“高血压”在数据仓库中没有直接存储数据对象与其对应的概念, 系统无法理解和获取。而且用户不止关注数据仓库中查询出相关数据信息, 对其中隐含的信息和知识更感兴趣。智能查询就是在用户不具备太多的计算机知识的情况下能够对自己感兴趣的方面进行查询并给出相应的信息。国内外很多学者对智能查询进行过卓有成效的研究, 本文在现有的数据仓库相关查询应答智能查询的设计思路基础上提出了一种应用于中药数据仓库面向中药知识的智能查询方法, 使用户能够基于领域知识进行动态查询, 并且还可与 OLAP、数据挖掘工具结合, 为用户的查询提供更多的相关信息及决策支持。

2 中医药知识的结构化表示

要对数据仓库进行有效的智能查询, 必须有结构合理的领域专家知识库的支持。目前, 知识

库的开发过程已被视为一个知识建模活动。建立知识库意味着建立一个具有专家能力的求解问题的计算机模型，使知识具有结构化。需要对不同层次的知识进行分类、整理，并有条理的组织起来，使其便于利用、共享、重用和维护。系统采用的领域知识库由一组关于这一主题的任务组成，每个任务代表一个知识源，存放一组知识，用于求解某个特定的问题。由解决这一问题所需的概念、公式、规则组成。而一个任务又可以被分解为不同的子任务，这种分解直到每个子任务可以直接用已知知识求解为止。通过主题分析，我们把涉及的领域概念分为两类：即基本概念和高层概念。基本概念就是在数据仓库中有具体存储数据对象（表中属性）相对应的概念（如药品名称、药品编号等）；而没有具体存储数据对象与其对应的概念，如降血压称为高层概念。对于高层概念，可以通过定义它与其他概念间的层次关系，这种关系可以通过规则、公式等知识形式来体现，以便在查询过程中通过问题的分解和层层映射，最终映射到能够具体求解的基本概念对象，从而获取相应概念对象的取值。

在中医药数据仓库中建立了单味药、中药复方、化学成分、方剂、剂型、疾病、证候、症状、病理生理、效应部位、药理分类、临床药理分类、毒理分类、指标等十四张标准库表。中医药数据仓库在设计和实现时，不仅有一般意义的查询结果的浏览，而且由于对数据进行了主题解析和符合中医知识逻辑关系的存储，加上对各种信息的主题标引，使系统在查询之余，还可对查询结果进行关联分析、数据总结、数据分类、聚类分析、序列模式分析等。由于知识库是有结构的，这样就可以将整个知识库存放在关系数据库中，使知识的获取、修改、管理更加容易，实现多用户共享。有了这种结构化的知识表示，在查询过程中，推理机就可以利用领域知识库中所提供的领域知识进行有效地推理求解。

3 面向主题的智能查询

3.1 设计思想

在查询过程中，查询引擎基于知识库中提供的领域知识，对用户的查询内容进行分析，获取查询中涉及的所有概念，若涉及到的概念均为基本概念对象，直接转换成可执行的数据库查询 SQL 语句，通过访问数据仓库获取查询结果，从而结束查询；若存在高层概念，需要通过领域知识库中的定义，提交给推理机，推理机通过运用领域知识库中提供的相应的专家知识及数据仓库的元数据，通过推理计算获取高层概念的取值，再对初始结果集进行过滤，最终完成用户查询。

在智能查询模块中查询引擎和推理机是其中的核心部分。查询引擎和推理机的功能有机的结合，相互渗透。查询引擎主要负责对查询内容进行词法分析，获取查询中所涉及的概念，对查询条件语法检查，概念转换和查询初始数据集生成；推理机主要负责知识推理，完成高层概念的求解。图 1 所示。

在查询过程中用户无需关注所涉及的数据和概念所对应的数据仓库中的具体存储数据对象，以及在数据仓库中的存储位置，系统通过利用知识库中存储的领域知识以及数据仓库中的元数据，通过智能查询程序自动的完成用户的动态查询需求，从而有效地提高了查询的智能化程度。

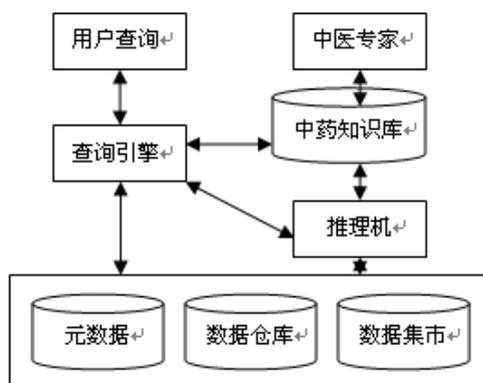


图 1 智能查询结构

3.2 算法描述

面向中医药的查询智能过程非常复杂，涉及到知识库、推理机制、词法分析等。通过对查询主题的选择，查询目标、查询条件的分析，获得查询的概念集，主要查询过程算法描述如下：

(1) 确定主题、词法分析获得概念集

```
Selectsubject () //选择主题
```

```
Conceptset //分析查询内容(包括查询条件、查询目标)获得的概念集
```

```
Condition Conceptset //作为查询条件的概念集
```

(2) 形成本次查询的基本概念集和高层概念集

```
While (Conceptset) { //将所得概念分类
```

```
For each Concept C ∈ Conceptset {
```

```
  If C ∈ SBconceptset then {
```

```
    //SBconceptset 为所选主题的基本概念集
```

```
    Insert (Bconceptset , C) ;
```

```
    //将 C 加入本次查询的基本概念集中，Bconceptset 的初值为空
```

```
    Delete (Conceptset , C) ; //将 C 从 Conceptset 移去
```

```
  }
```

```
  Else if C ∈ SAconceptset then {
```

```
    //SAconceptset 为所选主题的高层概念集
```

```
    Insert (Aconceptset, C) ;
```

```

        //将 C 加入本次查询的高层概念集中，Aconceptset 的初值为空
        Delete (Conceptset, C) ;
    }
Else
    exit () ; //存在未知概念，查询结束，交给领域专家处理
End if
}
}

```

(3) 形成初始结果集

```

If (!Aconceptset) then{
    SQLexp(); //将基本概念用数据仓库中的数据对象替换，自动生成 SQL 查询，返回查询结果，
    并结束查询
Return;
}
Else
    Recordset; //将基本概念用对应的数据库对象代替，据主题生成 SQL 语句，获得的初始查
    询数据集
End if

```

(4) 高层概念求值、获得最终结果集

```

While (Aconceptset) { //当还存在高层概念没有求得值
    for each AC ∈ Aconceptset {
        InsertColumn(Recordset, AC);
        //将 AC 作为 Recordset 的一个属性插入
    for each r ∈ Recordset
        Recordset.AC = Reason_Compute (AC, Recordset);
        //对数据集中的每个元组，通过规则推理或公式计算，获得高层概念 AC 的取值
    if AC ∈ Condition Conceptset then{
        Filter (Recordset, AC); //概念 AC 如果在查询条件中存在，利用 AC 取值对 Recordset
        进行过滤
        Delete (Aconceptset, AC); //将概念 AC 从 Aconceptset 中删除
    }
}
}

```

```
}  
}  
}
```

本算法主要有上述四部分组成。另外，还可以与 OLAP、数据挖掘工具相结合，挖掘出更多的中医药的规则、知识，为医生提供更好的支持。

3.3 实例分析

结合实例分析一下上述算法。

例 查询治疗高血压的中药处方？

涉及主题：治疗方案问题

查询目标：方剂名称，方剂类型，功能主治

查询条件：方剂类型=‘单味药’ and 功能主治 like %高血压%’ or 方剂类型=‘复方药’ and 功能主治 like ‘%高血压%’

查询过程如下：

(1)确定查询主题：治疗方案问题

涉及的表有单味药、化学成分、方剂、剂型、疾病、证候、症状、效应部位等。

(2)获取初始查询数据集

根据前面的查询条件生成相应标准的 SQL 表达式，获取初始查询数据集。

(3)对高层概念求解，并对数据集进行过滤

涉及的高层概念有疾病症状、效应部位等，利用中医药知识库中给出的规则、公式等进行计算、推理取得其值。其中在中医药治疗方案查询条件中，需用它作为参数对初始数据集进行过滤。

(4)将查询到治疗高血压的中医药治疗方案输出。

4 结束语

为了提高中医药数据仓库前端工具的智能化程度和查询分析能力，提出了一种面向中医药知识的智能查询分析方法，在此系统中结合了数据仓库、人工智能、数据挖掘等领域的相关技术，对中医药数据仓库深层研究进行尝试，以后将进一步往数据深层挖掘及实现迈进。

参考文献

1. 蔡湘；徐东平；赵玉伟. 基于知识智能查询的通用数据挖掘系统研究 《计算机工程与设计》，2006，(8)
2. 孙瑜；李志平； 基于多领域本体的智能查询系统模型 计算机工程，2005年 13期

3. 彭勇, 梁少伟 . 国内医药信息数据库简介 中国中医药信息杂志, 1999 ;6(1) :73 -5
4. 任廷革, 刘晓峰 . “中医药基础数据库系统”介绍 中国中医药信息杂志, 2001 ;8(11) :90 -1
5. 陈峰, 崔蒙 . 中药信息系统建立初探 中草药, 2000 ;31

作者简介（蔡永铭、广东药学院、讲师、13719326834、Bruce_cai@163.com）