

# “领域本体七步法”在中医辨证推理 知识库构建中的应用\*

郭梦莹, 周璐, 孙燕\*\*

(北京中医药大学中医学院 北京 100029)

**摘要:** 中医辨证知识间高度的关联性导致在中医辨证智能化研究中,探索构建的中医辨证模型临床实用性不高。故探索一种可使辨证知识形成高度关联性的表达方法成为现阶段的一大研究方向。“领域本体七步法”是将零散的知识结构化的一种方法,本体、语义网和知识图谱均可在七步法的指导下形成结构化的知识表达体系。故本研究以“领域本体七步法”为指导方法,设计中医辨证知识图谱,使中医辨证知识之间形成一种结构化的连接,以加强知识之间的关联性,以期进一步构建高辨证性的中医辨证推理知识库。

**关键词:** 知识库 领域本体七步法 知识图谱 辨证

doi: 10.11842/wst.20190110004 中图分类号: R33 文献标识码: A

知识库是一种基于知识的高度结构化、具有推理规则的系统,主要由两部分组成,一是数据库,由高度结构化的知识组成,又可以看作是高度结构化的数据符号,使其能够被有效地存储、搜索和应用。二是规则库,是支配知识的规则集合,用来决定知识的使用方式,规则的制定与具体的领域相关,不同的规则适用于不同的领域。由于中医辨证的特殊思维方式与中医辨证知识之间高度的关联性,使非结构化的中医辨证推理模型的辨证推理功能具有较低的弹性与效率,所以探索一种适用于中医辨证知识的关联结构对于中医辨证的智能化发展非常关键。而“领域本体七步法”是将零散的知识结构化的一种方法,本体、语义网和知识图谱均可在七步法的指导下形成结构化的知识表达体系。故本研究以“领域本体七步法”为指导方法,设计适用于中医思维的中医辨证知识图谱,使中医辨证知识之间形成一种结构化的连接,以加强知识之间的关联性,并探索适用于此知识结构的推理

算法与规则。以期为构建具有模拟人类辨证思维的中医辨证智能系统提供结构基础。

## 1 知识表示

知识表示是将知识以计算机可理解的方式存储并使用,是形成相应知识库的前提,传统的知识表示方法主要包括框架表示法、一阶谓词逻辑表示法和产生式表示法等。此类方法很难表示复杂知识<sup>[1]</sup>。随着信息的爆炸,本体、语义网和知识图谱等被相继提出,用于组织管理庞杂的信息,作为改进的知识表示方法被广泛应用于各个领域,如知识工程、人工智能、图书情报和信息检索等。

本体原是一个哲学概念,后应用于人工智能领域,Neches等<sup>[2]</sup>最先将本体定义为定义了包含主题区域的词汇表的基本术语和关系,以及用这些术语和关系规定词汇表扩展的规则。Gruber<sup>[3]</sup>认为本体论是概念化的明确规范。邓志鸿等<sup>[4]</sup>认为本体是描述概念及

收稿日期:XXXX-XX-XX

修回日期:XXXX-XX-XX

\* 中华人民共和国科学技术部国家重点研发计划中医药现代化研究专项(2017YFC1700303):基于名老中医学术经验的辨证论治临床辅助决策系统,负责人:李宇航;中华人民共和国科学技术部国家重点研发计划中医药现代化研究专项(2017YFC1700300):面向名老中医学术经验传承的关键技术和应用平台的系统化研究,负责人:徐凤芹。

\*\* 通讯作者:孙燕,副研究员,主要研究方向:仲景学术思想研究。

概念间关系的一种概念模型,通过概念之间的关系来描述概念的语义。通过以上定义可看出本体是将其具有共同特点的事物抽象为一类概念,并对概念之间关系进行描述的模型。其形成的知识结构化表达模型实质上是一组概念框架,包括概念及概念间关系,所以本体可看作为一种侧重概念的知识表达模型,而语义网则多用于对具体事物的表达,是本体下的具体实例,如本体概念是方剂,桂枝汤则为其下的一个实例。

语义网是1998年由万维网联盟 Tim Berners-Lee 及其合作者提出的,最早可追溯至20世纪50年代末和60年代初,代表人物有 Ross Quillian 和 Robert F. Simmons,前者在1966年将其作为人类联想记忆的一个心理模型首次提出了语义网络的概念,后者在1969年提出了模式推理的思想,对语义网的发展具有重大意义。语义网的主要原则是采用逻辑语言表达万维网中信息的含义(即语义)<sup>[6]</sup>,可见相比于本体,语义网更侧重语义关系的表达,用语义关系连接各种知识信息,从而实现对信息的智能化处理、分析及应用。

随着数据间的关系日趋复杂,为了建立更智能的搜索引擎,2012年谷歌正式提出知识图谱,2013年后被引用到各个领域。知识图谱<sup>[9]</sup>本质上是一种叫做语义网络的知识库,具有有向图结构,是以(节点,边,节点)连接形成的关系网络结构图,节点代表真实世界中存在的各种概念和实体,边代表实体、概念间的语义关系<sup>[7]</sup>。其对实体与概念间的关系描述是按实体与概念的对应关系,将(点,边,点)的结构具体划分为(概念1,关系,概念2)“三元组”,从实体层与概念层两个层面对事物进行描述。如概念层是(疾病,具有,证),实体层则有(心悸,具有,心虚胆怯证)。

以上三者均可作为知识结构化表达的模型,但本体侧重于概念的表达,如方剂、中药是中医中的概念,二者的关系是中药组成方剂,可描述为一个三元组(中药,组成,方剂)。语义网则多用于对具体事物的表达,是本体下的具体实例,如(桂枝,组成,桂枝汤)。而知识图谱融合了语义网与本体对知识的表示形式,可描述更复杂的逻辑关系,构建更为复杂的知识模型。故本研究采用知识图谱作为中医辨证知识知识库中数据库的模型,以图的形式对知识进行描述和储存,可分为模型层与数据层<sup>[8]</sup>,模型层又称概念层,是对概念解释和层次关系的说明,由本体构成,因为本体可以作为知识图谱表示的概念模型和逻辑基础<sup>[9]</sup>,

旨在描述概念及概念间的关系;数据层又称实例层,管理特定实例的说明,是与概念层对应的具体实例,旨在描述概念对应的具体实例以及这些实例间的关系,只有创建类的具体实例后,才能形成完整的领域知识库。

## 2 知识库结构化的指导方法

### 2.1 领域本体七步法

知识图谱实际上是一种图数据,可作为知识结构化表达的图模型,构建其模型需要一定的指导方法。在图计算中,基本的数据结构表达式为 $G = (V, E)$ ,  $V = \text{vertex}$ (节点),  $E = \text{edge}$ (边),与本体中三元组语言描述(概念,关系,概念)模型相同<sup>[10]</sup>。而在医学领域中,“领域本体七步法”<sup>[11-12]</sup>由斯坦福大学医学院开发,用于医学领域本体的构建,是指导领域知识结构化表达的方法,被广泛应用于构建本体或基于本体的知识库,如中医证候本体表达模式构建、电子病历知识库构建、某疾病本体构建、中医症状疾病表示模型构建等。

“领域本体七步法”的七个步骤是:①确定本体的专业领域和范畴;②考查复用现有本体的可能性;③列出本体中的重要术语;④定义类和类的等级体系;⑤定义类的属性;⑥定义属性的分面;⑦创建实例。通过步骤可见其连接了概念层和实例层,既能指导构建本体,又可指导构建知识图谱。前六步梳理明确中医辨证知识中的关键术语、概念、实例、属性及其中的各种关系,使中医辨证知识形成结构化的表达,完成概念层的构建。最后一步创建概念下的实例,从而完成实例层的构建。

### 2.2 构建知识库的工具

确定了知识库结构化表达的模型与方法,下一步需要使用工具将其储存到计算机中,建立知识库中的数据库。Neo4j是由德国 Neo4j 公司推出的一种图数据库管理平台,是一种高性能的 NoSQL 图数据库<sup>[13]</sup>,使用图相关的概念来描述数据模型,把数据保存为图中的节点以及节点之间的关系,其具有完全支持 ACID 事务、基于磁盘的持久存储、支持海量数据、高可用的分布式集群和迅速的图查询等特点,可以存储和查询大规模、复杂的图数据,有效缓解数据存储方面的压力。与关系型数据库相比,在高维度关联查询的效率明显提高,故本研究选用 Neo4j 对知识进行结构化表达,将辨证知识以图数据的形式进行储存,形成知识

图谱进而完成知识库的构建。

### 3 基于“领域本体七步法”的中医辨证知识结构化

应用“领域本体七步法”使辨证知识结构化的过程即是设计辨证知识图谱结构的过程,主要通过以下的七个步骤进行。

#### 3.1 确定本体的专业领域和范畴

包括领域、目的、系统能力等问题。领域即为中医辨证领域。目的是使中医辨证知识结构化,构建小型中医辨证知识图谱。此知识图谱的能力是可以有向连接进行由“症状→证”的推理。

因为构建中医辨证本体需要辨证领域的知识数据,故本研究构建知识库的数据源为朱文峰主编的高等教育“十五”国家级规划教材《中医诊断学》,作为《伤寒论》辨证内容的继承与创新,辨证知识更为系统、完整与准确,其知识数据可作为辨证知识库的基础数据。

#### 3.2 考查复用现有本体的可能性

为了提高知识的复用率,减少不必要的资源浪费,有些本体或知识图谱的构建是可以建立在以往本体研究基础上的。考虑到本研究构建的知识图谱主要功能为辨证推理,而在进行了相关文献调研后,目前的中医辨证领域或证候领域的本体结构尚不能满足此能力,故无可复用的本体。

#### 3.3 列出本体中的重要术语或概念

重要术语主要包含概念术语、关系术语及属性术语。因梳理的是中医辨证知识,要完成的是从症状到证的推理,所以重要的概念术语主要是证和症状。证主要具有与其相关的症状,包括体征,所以证与症状间存在“具有”关系。症状反过来同样可以反映证,临床辨证即是以此为基础用症状去推理证的,是中医司外揣内辨证思想的体现。而症状通常反映的是以八纲、气血津液病性为主的证的本质,如“发热”对“风寒犯表证”的本质反映为表,对“痰热壅肺证”的本质反映为里热,因此症状与证候间又存在着“本质反映”的关系,其取值范围是以八纲和气血津液为主的证的本质。此外有些症状对病位具有较高的诊断性,如心悸、胸闷通常反映病位在心胸,所以症状到证还有“病位反映”关系。

证为疾病某阶段的病理概括,包括病因、病位、病性、病势等。朱文峰等<sup>[14]</sup>在进行了相关研究后认为证

表1 类及类间关系

类(概念)	关系	类(概念)
证(Syndrome)	具有(Has)	症状(Symptom)
症状(Symptom)	本质反映(Essential reflection)	证(Syndrome)
症状(Symptom)	病位反映(Location reflection)	证(Syndrome)

的内容主要可归纳为病性和病位,这类描述在领域本体中称为“属性”。症状是临床者望闻问切后收集的四诊资料,望诊包含望全身、局部、排出物、小儿指纹和舌诊,问诊主要问寒热、汗、疼痛、头身等,闻诊主要涉及声音、气味等,切诊又称脉诊,我们将得到症状的方式定义为症状的归类属性,如发热的归类属性为寒热,心悸的归类属性为胸腹。综上所述,中医辨证的重要术语具体表述如下:

- ①概念术语:证、症状。
- ②关系术语:具有、本质反映、病位反映。
- ③属性术语:病性、病位、归类属性。

#### 3.4 定义类和类的等级体系

类是领域本体的核心,由领域内的术语来定义,也是概念层的核心构成,因为类与类的等级体系表达的就是概念与概念之间的等级体系,而本研究中医辨证领域的概念术语主要为证与症状,故只需定义两个类,即证与症状,层次较为简单明了,重要的是对类的描述及类之间的关系,关系到中医辨证知识图谱的知识结构。

#### 3.5 定义类的属性和关系

类的属性:以上讨论中,从中医辨证概念层提取出证与症状两个类。证的属性是对证的描述,有以八纲、气血津液证候为主的对证的本质的描述,即病性,将其定为证的证候本质属性;有以具有较高诊断性的症状对病位的描述,将其定为证的症状要点属性。因此本研究中证的属性为“证候本质”与“症状要点”。如心阳虚证的证候本质是阳虚,症状要点为心悸和胸闷。症状的属性设置一个为“归类属性”,如心悸的归类属性为心胸。

类之间的关系有“具有”、“本质反映”、“病位反映”,故可列出类及类间的关系见表1所示。

#### 3.6 定义属性的分面

属性的分面是指属性的属性,如属性的类型、容许值、取值个数等<sup>[15]</sup>。在计算机领域中,用来说明属性是字符型、数值型、日期、对象实例、对象类、文件路径等类型。在中医辨证领域则主要涉字符型数据,属性

“证候本质”的取值范围是八纲、气血津液等基础性;属性“症状要点”的取值范围即症状。总结以上分析可列出类的属性及其属性分面见表2。

以上即是用“领域本体七步法”中的前六步提取了中医辨证中重要概念术语、类即类间关系、属性及属性分面,确立中医辨证知识图谱中概念层的设计,以便创建实例及实例间关系的填充。

### 3.7 创建实例

以知识图谱的三元组模式对实例进行表示,用Neo4j 构建工具进行储存及后续应用,以期完成以《中医诊断学》辨证内容为数据源的中医辨证小型知识图谱。

已知概念层三元组模式有(证,具有,症状)、(症状,本质反映,证)与(症状,病位反映,证)。以心阳虚证为例构建实例,则概念证下的实例即为心阳虚证,参考高等教育“十五”国家级规划教材《中医诊断学》<sup>[6]</sup>心阳虚证相关知识,可知其具有的症状有心悸、怔忡、心胸憋闷、心胸痛、气短、自汗、畏寒、肢冷、神疲、乏力、面色恍白、面唇青紫、舌淡胖、苔白滑、脉弱、脉结代,其辨证要点为以心悸、胸闷和以阳虚证为主的症状,故有证候本质属性:阳虚;症状要点:心悸,怔忡,心胸憋闷,心胸痛。以知识图谱“三元组”模式进行描述即有(心阳虚证,具有,心悸),(心阳虚证,具有,畏寒),(心悸,心,心阳虚证),(畏寒,阳虚,心阳虚证),具体构建的概念层与实例层见图1。

### 4 基于辨证知识图谱的推理设计

通过上述步骤,中医辨证推理知识库的数据库设计已基本完成,是以知识图谱为知识结构化表达模型,可通过知识图谱的有向连接与Neo4j 工具的Cypher 语言实行推理设计,通过症状对证的本质反映与病位反映关系进行证候本质与症状要点属性的匹配。以证候本质匹配为例,通过Neo4j 图数据平台中“(节点)→(关系)→(节点)”的“字符串”匹配模式,设计(症状)→[本质反映]→(证)的匹配模板。以实现症状的传入,并进一步与可能证之间形成具有方向性的映射,模拟人的有向认知思维,即发挥计算机的搜索优势,对输入症状可能对应的证进行全面检索,找到所有的可能。具体而言,以Input\_Symptom表示输入症状的集合,并根据表2中各个概念与属性的英文表示,设计有向连接推理模板为:

表2 类的属性及属性的分面

类(概念)	属性	属性的分面
证(Syndrome)	证候本质(Essence)	{阳虚,阴虚,...}
	症状要点(Key point)	{心悸,胸闷,...}
	名称(Name)	{心阳虚证,...}
症状(Symptom)	归类属性(Class)	{寒热,心胸,...}
	名称(Name)	{心悸,胸闷,...}

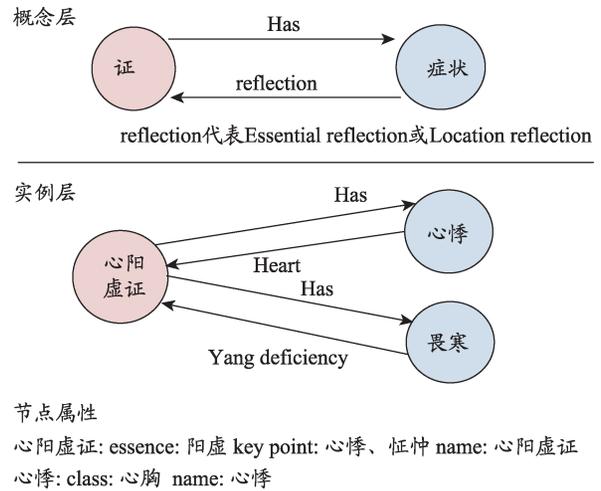


图1 概念层与实例层(部分)

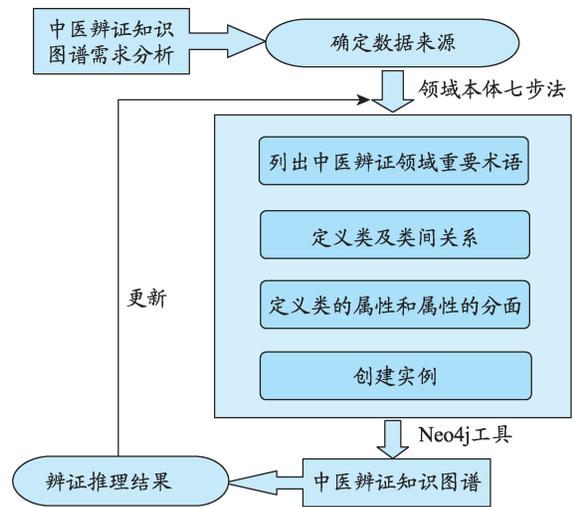


图2 中医辨证推理知识库构建流程

Match (Symptom) → [ES: Essential reflection] → (Syndrome) where Symptom. Name in Input\_Symptom Return Syndrome, ES

症状要点的匹配模板与证候本质的匹配模板类似。对其中匹配度最高的证进行输出,完成从症状到证的推理,对推理结果进行分析,分析现有知识库存在的问题并进行完善,整个辨证推理知识库构建过程见图2。

由“领域本体七步法”除可指导辨证知识形成结构化表达外,其在知识库中的作用相当于“知识的抽取与处理”这一步骤,既可整理原始数据,同时可发现知识之间的关联,并且可以通过不断的测试对原有数据结构进行相应更改。

## 5 讨论

自上世纪80年代起,中医界就已经开始了将计算机等现代技术与中医结合进行中医现代化与智能化研究,使中医得到了突飞猛进的发展。在中医智能化辨证模型研究中也取得了诸多可喜的进展,如基于随机森林的中医“内生五邪”辨证模型<sup>[7]</sup>、基于贝叶斯网络的“证素”辨证模型<sup>[8]</sup>,以及基于深度学习的中医脾胃病辨证模型<sup>[9]</sup>,但这些模型均是基于机器学习的方法进行的研究,存在着医案样本需求大但质量不高的问题,限制了这些模型的应用效果。而本研究开展的基于公认辨证知识的辨证推理模型研究,是小样本的

基础知识研究,可降低样本不足或质量不高对辨证模型的影响,以期为进一步的智能化辨证研究提供基线模型。

但本研究目前尚面临以下几个主要的问题:①症状的拆分与组合问题,如对脉象的描述,脉弦细是否可以拆分为脉弦与脉细;②证与证之间具有高度的耦合性,本研究虽然在证的基础上设计的证候本质和症状要点的推理,在一定程度上减轻了此类问题,但证候本质之间仍然具有较高的耦合性。故对证候本质的解耦合研究以及对证候术语的规范化研究将成为本课题的另一重点研究内容。

此外,知识的完备性对整个模型的应用至关重要,如果知识的完备性差,则难以完成准确有效的辨证推理。因此,需不断对知识库中的辨证知识进行补充,还应严格的审查每一条录入的知识是否严谨,避免知识间的冲突、错误,所以辨证推理知识库的构建尚需要中医领域专家的帮助与指导。

## 参考文献

- 李智,徐展琦,丁喆,等.基于本体的业务源模型知识库构建及应用研究.计算机应用与软件,2018,35(2):16-21+29.
- Neches R, Fikes R, Finin T, et al. Enabling technology for knowledge sharing. *Ai Mag*, 1991, 12(3): 36-56.
- Gruber T R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 1993, 5(2): 199-220.
- 邓志鸿,唐世渭,张铭,等. Ontology 研究综述. 北京大学学报(自然科学版), 2002, 38(5): 730-738.
- Bemerslee T, Hendler J, Lassila O. The semantic web. *Sci Ame*, 2001, 284(5): 34-43.
- 漆桂林,高桓,吴天星.知识图谱研究进展.情报工程,2017,3(1):4-25.
- Singhal A. Introducing the knowledge graph: things, not strings. Official Google Blog, 2012. <https://googleblog.blogspot.co.za/2012/05/introducing-knowledge-graph-things-not.html>.
- 高莹,侯凌燕,刘秀磊.基于本体的煤矿瓦斯知识库建设.北京信息科技大学学报(自然科学版),2017,32(5):50-55.
- 郝伟学.中医健康知识图谱的构建研究.北京:北京交通大学硕士学位论文,2017.
- 肖庆都,屈亮亮,侯霞.基于Neo4j图数据库的课程体系知识图谱系统设计与实现.电脑知识与技术,2017,13(36):130-132.
- Swartout B, Ramesh P, Knight K, et al. Toward distributed use of largescale ontologies. *Spring Symposium on Ontological Engineering of AAAI*, 1997.
- Noy NF, McGuinness DL. Ontology development 101: a guide to creating your first ontology. *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05* and *Stanford Medical Informatics Technical Report SMI-2001-0880*, 2001.
- Neo4j. 2016-05-17[2016-05-21]. <http://neo4j.org/>.
- 朱文锋,张毕敏.“证素”的基本特征.中国中医基础医学杂志,2005,11(1):17-18.
- 李景,孟连生.构建知识本体方法体系的比较研究.现代图书情报技术,2014,(7):17-22.
- 朱文锋.中医诊断学.北京:中国中医药出版社,2002:178.
- 王华珍,胡雪琴.中医“内生五邪”的智能证型分类.计算机工程与应用,2011,47(6):156-160+163.
- 朱咏华,朱文锋.基于贝叶斯网络的中医辨证系统.湖南大学学报(自然科学版),2006,33(4):123-125.
- 颜建军,刘章鹏,刘国萍,等.基于深度森林算法的慢性胃炎中医证候分类.华东理工大学学报(自然科学版),2019,45(4):593-599.

## Application of “Domain Ontology Seven-step Method” in the Knowledge

### Base Construction of TCM Differentiation

Guo Mengying, Zhou Lu, Sun Yan

(School of Traditional Chinese Medicine, Beijing University of traditional  
Chinese Medicine, Beijing 100029, China)

**Abstract:** The high correlation between TCM syndrome differentiation knowledge leads to the low clinical practicability of constructing TCM syndrome differentiation model in the research of intellectualization of TCM syndrome differentiation. Therefore, exploring an expression method that can make differentiation knowledge highly relevant has become a major research direction at this stage. “Domain ontology seven-step method” is a method of structuring scattered knowledge. Ontology, semantic web and knowledge map can form a structured knowledge expression system under the guidance of seven-step method. Therefore, this study takes the “Domain ontology seven-step method” as the guiding method to design the knowledge map of TCM syndrome differentiation, so as to form a structured connection between TCM syndrome differentiation knowledge. Then strengthen the correlation between knowledge, so as to further construct a knowledge base of TCM syndrome differentiation with high differentiation.

**Keywords:** Knowledge base, domain ontology seven-step method, knowledge map, differentiation

(责任编辑: 闫 群, 责任译审: 钱灵姝)