

类方衍化关系内涵分析及形式化表达

朱彦 高博 崔蒙

(中国中医科学院中医药信息研究所,北京,100700)

摘要 以前述类方概念研究为基础,结合目前方剂相似度研究的不足,分析了类方衍化关系自动发现的难点。然后,对类方衍化关系的内涵进行了剖析,提出其应包含变化、制方时间、药物组成和相似度等多方面的特征和因素。最后,对中药、方剂等概念和关系进行形式化定义的基础上,完成了类方衍化关系的形式化表达,为后续类方衍化关系的判别模型的构建奠定了基础。

关键词 类方;类方衍化关系;形式化表达

Investigation and Formal Representation of Analogous Formulae Derivation Relationship

Zhu Yan, Gao Bo, Cui Meng

(Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China)

Abstract Following the previous research on analogous formulae, this paper analyzes the difficulties to automatically discover the analogous formulae derivation relationship (AFDR) based on the review of current formulae similarity models. Furthermore, the paper carries on thorough analysis on connotation of AFDR and suggests it should comprise of many features and factors such as transformation, time of prescription, drug compositions and similarity. At last, on the basis of the formal representations for such concepts as Chinese materia medica and formulae as well as their relations, formal representation for AFDR is completed, which lays a stable foundation for the further research on AFDR classification model.

Key Words Analogous formulae; Analogous formulae derivation relationship (AFDR); Formal representation

中图分类号:R289;O23 文献标识码:A doi:10.3969/j.issn.1673-7202.2017.11.061

方剂是中药应用的基本形式,是按一定目的和规律组成的药物组合。其目的即方剂治法的体现,其规律即配伍原则及临床经验。

类方是具有组成结构相近、功效治法相类特点的一类方剂,一般有一个制方较早的基础方剂和在此基础上加减衍化出的一系列方剂。因为类方有这种特点,所以在方剂学习过程中,类方是一条学习的捷径,只要熟练掌握了祖方及加减变化规律,在临床中就能够应对自如;而类方组成结构和功效方面的相似性,也对配伍规律研究有着重要意义。

然而中医数千年临床实践中积累下来的成方,以及临床实际应用中不断产生新的衍化方数目庞大,单纯靠手工查询、整理和分析,难免错讹疏漏,如果能基于计算机技术,对类方衍化关系进行自动分析研究,将大大提高研究效率。而要实现计算机的自动分析,形式化表达是基础。本文将对类方衍化关系的内涵进行剖析,并在业内共识和适合计算机处理之间进行取舍和平衡,完成类方衍化关系的形

式化表达,为后续类方衍化关系的判别模型的构建奠定基础。

1 类方研究现状

1.1 类方相关概念研究 在前述研究^[1]中,通过文献研究和专家咨询相结合的方法,梳理了类方相关的一系列概念,并进行了定义,界定了广义类方、狭义类方、事实类方、理论类方等相关概念,与近似概念如衍生方、广义方剂衍化、加减方等作出辨析和区别,并对类方的衍化关系从最基本的三元组到高级形式类方树进行了形式化表达,以使之更易为计算机所识别,为使用计算机完成类方衍化关系的自动发现奠定基础。

本研究将基于前述研究,进一步分析类方衍化关系的内涵,进而对其进行形式化表达。

1.2 现有方剂相似度研究及不足 在对大量方剂数据处理过程中,已有一些研究者发现中医方剂相似度模型在处理冗余数据和重复数据、梳理鉴别新方剂数据和整理研究古代方剂演变规律等方面的重

基金项目:国家自然科学基金青年科学项目(81503494);北京市自然科学基金资助项目(7174328);中国中医科学院基本科研业务费自主选题项目(ZZ090311);中国中医科学院博士研究生创新人才培养基金项目(CX201510)

作者简介:朱彦(1983.04—),男,博士,助理研究员,研究方向:中医药信息学,E-mail:zhuyan166@126.com

通信作者:高博(1981.02—),女,博士,助理研究员,研究方向:中医基础理论,E-mail:gaobo_cat@126.com

要性,并做出了有益的尝试。

操牡丹等^[2]分别从组成药、功效、主治功能和名称等方面,基于集合相似度和字符串相似度来进行计算相似性。顾铮^[3]基于文本分类算法,从组成药物名称、描述和用量3方面计算方剂的相似度。李园白等^[4]以中药为基本单元,从中药组成、中药组成剂量、中药功效3个层次进行方剂之间的相似程度计算,其中中药组成相似度是利用集合相似度比较法计算,中药组成剂量的相似程度及方剂功效的相似度利用夹角余弦比较法计算;并在随后的研究中使用该算法将现代临床上的哮喘处方与古籍中记载的经典处方进行比较,得出了大多数处方都是在前人基础上形成,独创性的处方比较少的结论^[5]。

而类方具有组成相似的特点,从方剂相似度入手进行类方的辨析,是一个良好的切入点。方剂相似度模型的构建,将是类方衍化关系发现研究的重要一环。然而,分析已有相似度模型研究,发现并不能满足类方衍化关系的判别,主要体现在:1)具有衍化关系的方剂间必然存在相似性,但相似度是一个甚为含糊的概念^[6],而且,不同研究团队对于相似的对象及相似程度的认识也不相同。现有的相似度模型,要么简单的从药物组成、剂量和字符串角度进行计算;要么分别从药物组成、剂量、功效等不同侧面分别计算,得出相似度结果。目前为止,尚没有一个综合指标和有意义的阈值,也无法对其进行结果优劣的评价。2)现有相似度模型中,虽然也有研究考虑到了功效要素^[4],但只将组成方剂的药物功效进行了简单叠加。一方面,没有解决中药功效集合的量化整合问题;另一方面,没有考虑各组成药功效之间的非独立性和相互作用过程。夹角余弦计算方法的前提作为特征向量是正交的,直接进行计算,而不考虑不同功效间复杂而不容忽视的关系,计算结果的准确性值得推敲;另外,方剂的衍化过程更加复杂,不但需要考虑功效这个维度,而且还要体现出功效的改变。也就是要考虑随着药物和药量的增减,药物间的配伍关系发生了哪些相应变化,从而导致方剂整体功效发生变化的过程。3)衍化关系可能还需要考虑方剂的其他维度,如方剂整体的性、味、归经倾向间的相似性等。

1.3 类方衍化关系的自动发现的难点 基于已有的文献调研来看,使用计算机来完成类方衍化关系的自动发现尚未见报道。归纳其研究难点主要包括以下几方面:1)“类方”作为中医方剂学中一个十分重要的概念,不但广泛出现在诸多方剂著述中,也已

成为一个特定的专题研究领域。然而,从权威著述、教材及已公开发表的研究成果来看,类方的内涵和外延仍存在着多种不同理解,类方衍化关系更是一个比较模糊的概念,需要构建一个适合计算机进行分析处理的类方衍化关系模型。2)类方衍化关系涉及到方剂的相似度计算,而已有的相似度模型主要从方剂的组成药物和剂量上来进行计算,无法满足对类方衍化关系的判别需求。需要从方剂的时间源流关系、性、味、归经、功效相似性等方面综合考虑,来实现相似度的度量。3)除了考虑以上维度外,还要体现出功效的改变。也就是要考虑随着药物和药量的增减,药物间的配伍关系随之变化,从而导致方剂整体功效变化的过程。这不仅需要实现基于方剂整体功效的自动计算,还要完成衍化方相对于祖方的整体功效变化程度度量。

2 “类方衍化关系”内涵

根据对前述研究^[1]中衍化类方的分析和定义,结合方剂学理论与临床方面专家的咨询结果,下面从数据可得性、研究的实际需求、计算机分析的特点等方面,对类方衍化关系的概念进行分析和讨论。

2.1 变化 “衍”在《辞海》中释为“水广布或长流^[7]”之义,引申为延展;在《现代汉语词典》中主要释为“开展、发挥^[8]”;“衍化”一词,则是动词,有“发展变化”之义。

可见从字面上来说,“类方衍化”就意味着方剂要变化。如果2个完全相同的方剂,则不能存在“衍化关系”。这种变化除了包括药物组成、剂量上的加減以外,还可以是剂型、用法(内服或外用)等方面的变动。

本研究主要基于“药物组成是变化形式,功效主治是变化结果”的思路,对其他变化因素(如剂型)不做探讨。

2.2 制方时间 对于满足类方衍化关系的2个方剂,在制方时间上,“被衍化方”应该早于“衍化方”。

2.3 药物组成 类方衍化关系是母方与子方之间的“不变”与“变化”的有机结合的体现,而决定这种关系的最直接的原因是药物组成的“不变”与“变化”。然而,这种“不变”与“变化”的程度没有一个统一的标准和共识,更难以量化。其中,主要的争议在于,子方是否要包含母方所有的药物组成?

这个问题在已有的关于“类方”概念的著述中讨论得不多。本研究主要通过专家咨询的方式来获得业内的共识和意见。目前能达成的共识是子方应该包含母方的主药(包括君药和臣药),而佐使

药是否应该包括则有分歧。本研究采纳的观点为“衍化方应该包含祖方所有的药物组成”，原因有以下几点：1) 君臣佐使理论存在着一定的局限性^[9]：a. 君臣佐使理论的界限有时比较模糊，尤其是臣药与佐助药之间；b. 有些方剂（尤其是药物组成较多的合方）用君臣佐使理论来解释并不太适用。因而本研究不基于君臣佐使理论对方剂的构成进行分析。2) “祖方”作为一组类方中的源头的“基本方”，除了制方时间最早，还应包含“基本”的含义。也就暗示了祖方应该是一个“高聚合体”，很难再进行简化。如四君子汤，甘草，即便为使药，也有益气和中的作用；如果能简化掉甘草，“四君子汤”很可能就不适合作为一个类方的祖方，而应该是“三君子汤类方”了。3) 在前述研究^[1]中，对衍化方和加减方也进行了区分，二者最大的区别在于：加减方可以改变原方的药物组成。4) 本研究采用“衍化方应该包含祖方所有的药物组成”的观点，更符合类方衍化关系的涵义，也便于计算机处理。

2.4 相似度 已有的研究基本认同将药物组成和主治及功效上的相似性作为类方判断的主要依据，可见，类方中衍化方与被衍化方之间的相似度是一个综合的概念和指标。

药物组成的相似度是方剂整体相似度的物质基础，然而单纯基于药物组成的相似度并不足以衡量方剂整体的相似度。一个典型的例子是：基于对方剂数据库进行检索，检出包含人参、白术、茯苓和甘草 4 种药物的方剂一共有 3 000 多首，通过随机抽样 185 首对其进行专家评判，发现只有 106 首与四君子汤“是衍化关系”，占 57%。

因为“药有个性之专长，方有合群之妙用”，组方是在辨证立法的基础上，按照一定的组成原则，选择适当的药物，规定适当的剂量而组成，故而医家对于类方的判断依据，除了药物组成，更多基于方剂的功效与治疗目的。

本研究基于以下考虑，将方剂功效作为重点研究对象。1) 功效涉及中医基础理论的多个层次，包括基础物质、脏腑、病证等，且与主治具有一定的锁定关系；2) 功效包含的概念相对较少，比较容易实现标准化和规范化；3) 已有研究表明，通过药物的功效来推导方剂整体功效是可行的。另外，本研究还将考虑性、味、归经等因素，即通过药物的性、味、归经属性，来计算方剂的性、味、归经度量，从而计算方剂间的相似度。

3 形式化表达及判别模型公式

3.1 概念建模 本研究基于中医药一体化数据模型框架^[10]进行设计，复用了已有的核心实体和关系，并针对本研究的实际需要进行了取舍和扩充。

模型核心实体(entity)包括：药材、饮片、方剂、功效、药性、药味、经络等。实体间的主要关系(Relationship)有炮制、药物组成、文献出处、衍化、饮片药效、方剂药效、功效间作用等。

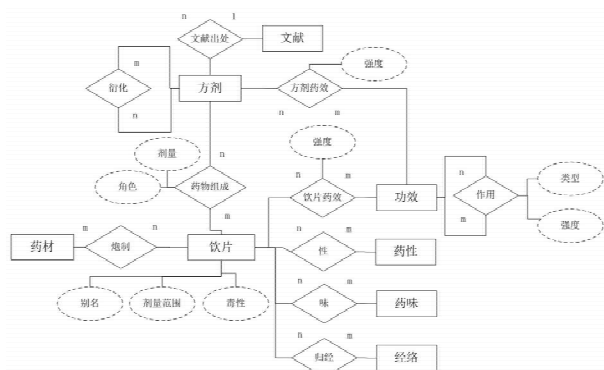


图1 概念模型 E-R

3.2 基础概念和关系的形式化表达 为了便于后续的分析讨论，基于上面的概念模型，先对饮片、方剂等实体和关系进行形式化定义。

3.2.1 饮片相关概念 药材 (DrugMaterial) = < DMID, DMName >

药材 ID: DMID

药材名称: DMName

功效 (Efficacy) = < EID, EName >

功效 ID: EID

功效名称: EName

药性 (DrugNature) = { 大寒, 寒, 微寒, 大凉, 凉, 平, 微温, 温, 热, 大热 }

药味 (DrugFlavor) = { 酸, 苦, 甘, 辛, 咸, 淡 }

归经 (Channel) = { 胆, 肝, 肺, 大肠, 胃; 脾; 心; 小肠; 膀胱; 肾; 心包; 三焦 }

饮片 (PreparedDrug): < PDID, PDName, PDAlias, DrugMaterial, PDE, PDN, PDF, PDC, PDRegularMinDosage, PDRegularMaxDosage >

饮片的标识符: PDID

饮片名称: PDName

饮片名称异名集合: PDAlias = { alias1: String, alias2: String, ..., aliasn: String }

药材来源: DrugMaterial

功效集合 PDE = { efficacy1: Efficacy, efficacy2: Efficacy, ..., efficacyn: Efficacy }

药性集合 PDN = { nature1: DrugNature, na-

ture2: DrugNature, ..., naturen: DrugNature}

药味集合 PDF = { flavor1: DrugFlavor, flavor2: DrugFlavor, ..., flavorn: DrugFlavor }

归经集合 PDC = { channel1: Channel, channel2: Channel, ..., channeln: Channel }

常用最小剂量 PDRRegularMinDosage

常用最大剂量 PDRRegularMaxDosage

定义关系: 饮片的药材来源相同 FromSameMaterial (prepareddrug1: PreparedDrug, prepareddrug2: PreparedDrug) $\equiv$ prepareddrug1. DrugMaterial = prepareddrug2. DrugMaterial

3. 2. 2 方剂相关概念 文献出处 LQ (Formula literaturequote) = < LQID, LQName, LQYear, Position >

文献标识符 LQID: 用来识别该文献的身份;

文献名称 LQName

文献产生年代: LQYear

在文献中出现的位置 Position

定义关系: 文献来源出现晚 LQ_IsLater (lq1: LQ, lq2: LQ) $\equiv$ (lq1. LQYear > lq2. LQYear) $\vee$ ((lq1. LQID = lq2. LQID) $\wedge$ (lq1. Position > lq2. Position)) 。

配伍角色 (Formula Ingredient Role) FIR = ;

药物组成单元 (Formula Composition Unit) FCU = < PreparedDrug, Dosage, DosageUnit, Role >

PreparedDrug: 饮片

Dosage: 剂量

DosageUnit: 剂量单位

Role: 该药物组成在方剂中的配伍角色

定义关系: 药物组成单元 fcu2 和药物组成单元 fcu1 相同, 即 FCUSameAs (fcu1: FCU, fcu2: FCU) $\equiv$ fcu1. PreparedDrug = fcu2. PreparedDrug $\wedge$ fcu1. Dosage = fcu2. Dosage $\wedge$ fcu1. DosageUnit = fcu2. DosageUnit

方剂 (Formula) 的形式化描述为: Formula = < FID, FName, FC, FE, LQ >

方剂标识符 FID: 用来识别该方剂的身份;

方剂名称 FName (Formula Name): 文献记载的方剂名称;

药物组成集合 FC (Formula Compositions) = { fcu1: FCU, fcu2: FCU, ..., feun: FCU }

功效集合 FE = { efficacy1: Efficacy, efficacy2: Efficacy, ..., efficacyn: Efficacy }

文献来源出处 LQ

定义关系: 方剂 f2 和方剂 f1 组成相同, 即 For-

mulaCompositionIsSame (f2: Formula, f1: Formula) $\equiv$ ($\forall$ fcu1 $\in$ f1. FC, $\exists$ fcu2 $\in$ f2. FC, 满足 FCUSameAs (fcu1, fcu2)) $\wedge$ ($\forall$ fcu2 $\in$ f2. FC, $\exists$ fcu1 $\in$ f1. FC, 满足 FCUSameAs (fcu2, fcu1))

定义关系: 药物组成集合 FC 含有饮片 prepareddrug, 即 FC_ContainDrug (fc: FC, prepareddrug: PreparedDrug) $\equiv \forall$ fcu $\in$ fc $\wedge$ fcu. PreparedDrug = prepareddrug

定义关系: 方剂 f 含有饮片 prepareddrug, 即 Formula_ContainDrug (f: Formula, prepareddrug: PreparedDrug) $\equiv$ FC_ContainDrug (f. FC, prepareddrug)

定义关系: 方剂 f2 药物组成包含方剂 f1, 即 FormulaInclude (f2: Formula, f1: Formula) $\equiv \forall$ fcu1 $\in$ f1. FC, Formula_ContainDrug (f2, fcu1. PreparedDrug)

定义关系: 方剂 f2 真包含方剂 f1, 即 FormulaProperInclude (f2: Formula, f1: Formula) $\equiv$ FormulaInclude (f2, f1) $\wedge$ (NotFormulaCompositionIsSame (f2, f1))

如果包含关系也考虑来自相同药材的饮片的话, 可以类似地定义一下函数:

定义关系: 药物组成集合 FC 扩展含有饮片 prepareddrug, 即 FC_ExContainDrug (fc: FC, prepareddrug: PreparedDrug) $\equiv \forall$ fcu $\in$ fc $\wedge$ (fcu. PreparedDrug = prepareddrug $\vee$ FromSameMaterial (fcu. PreparedDrug, prepareddrug)

定义关系: 方剂 f 扩展含有饮片 prepareddrug, 即 Formula_ExContainDrug (f: Formula, prepareddrug: PreparedDrug) $\equiv$ FC_ExContainDrug (f. FC, prepareddrug)

定义关系: 方剂 f2 扩展包含方剂 f1, 即 FormulaExInclude (f2: Formula, f1: Formula) $\equiv \forall$ fcu1 $\in$ f1. FC, Formula_ExContainDrug (f2, fcu1. PreparedDrug)

定义关系: 方剂 f2 扩展真包含方剂 f1, 即 FormulaProperExInclude (f2: Formula, f1: Formula) $\equiv$ FormulaExInclude (f2, f1) $\wedge$ (NotFormulaCompositionIsSame (f2, f1))

3. 2. 3 类方衍化关系的定义和形式化表达 经过以上分析讨论, 对于 2 个方剂是否存在类方衍化关系 AFDR 的充分必要条件为: 1) 变化上: 2 个完全相同的方剂, 没有“变化”, 故不能称为“衍化”。2) 产生的时间先后上: 衍化方应该比被衍化方出现晚; 3) 药物组成上: 衍化方应该包含祖方的所有药物; 4) 衍化方的功效应该和被衍化方保持一定程度的相似

性,尤其是功效上的相似度。

类方衍化关系的形式化表达如下:类方衍化关系 AFDR,是二元关系,即 $\text{Formula} \times \text{Formula}$ 上的一个子集。对于 $f1: \text{Formula}, f2: \text{Formula}$,充分必要条件包括:1)方剂 $f2$ 出现晚于方剂 $f1$;即满足文献来源出现晚函数 $\text{LQ_IsLater}(f2, \text{LQ}, f1: \text{LQ})$;2)方剂 $f2$ 扩展真包含方剂 $f1$ 。即满足扩展真包含函数: $(f2, f1)$;3)功效主治上:衍化方的功效主治应该保持和被衍化方功效主治一定程度的相似性。

3.2.4 类方衍化关系判别模型 于是,得出以下类方衍化关系的判别模型:

模型计算的结果 $\text{DP}(f1, f2)$,来衡量输入的方剂 $f1$ 和 $f2$ 存在类方衍化关系的可能性。如果 $\text{DP}(f1, f2) \geq 0.5$ 则表示存在,否则表示不存在。该模型公式体现了类方衍化关系的内涵;而类方衍化关系的外延是一个比较模糊的集合,即某些方剂之间是否存在类方衍化关系是见仁见智的。本研究将通过领域专家判别结果来作为金标准,通过机器学习方法来获得类方衍化关系的外延边界。

根据前述章节的形式化定义,和是可以直接计算的。在同时满足条件,和的情况下,方剂 $f1$ 和 $f2$ 的相似度 $\text{SimP}(f1, f2)$ 则直接决定了 $\text{DP}(f1, f2)$ 的取值;否则 $\text{DP}(f1, f2) = 0$ 。然而,方剂 $f1$ 和 $f2$ 的相似度 $\text{SimP}(f1, f2)$ 是一个模糊的概念,是集合了方剂的药物组成、功效、性味归经等方面的一个综合指标。

4 小结及展望

通过对类方相关概念、方剂相似度、类方衍化关系自动发现现有研究的分析可见,现有方剂相似度模型无法满足类方衍化关系的自动判别。类方衍化关系涉及的方剂相似度计算应综合考虑到方剂的时间源流关系、性味归经相似性、功效改变等方面。要考虑随着药物和药量的增减,药物间的配伍关系随之变化,从而导致方剂整体功效变化的过程。

通过上述分析,本研究确立了“类方衍化关系”的内涵应包括以下内容:1)衍化方必须在被衍化方基础上有所变化,包括药物组成、剂量上的加减,以及剂型、用法(内服或外用)等方面的变动;2)衍化方制方时间必须晚于被衍化方;3)衍化方必须包含被衍化方的所有药物组成;4)判断是否存在衍化关

系的方剂相似度指标,是集合了方剂的药物组成、功效、性味归经等方面的一个综合指标。

在“类方衍化关系”内涵确立的基础上,对基础概念和关系进行了形式化表达,并建立了“类方衍化关系判别模型”。

未来研究将对方剂 $f1$ 和 $f2$ 的综合相似度 $\text{SimP}(f1, f2)$ 进行量化建模。该指标是集合了方剂的药物组成、功效、性味归经等方面的信息,其中药物组成的相似度计算最为直接,方剂的性味归经的量化也可以根据各药物组成的性味归经计算获得,最大的挑战来自于方剂整体功效的自动计算。下一步将基于课题组前期研究基础^[11],通过构建中药功效的语义网络表达和量化功效间的语义关系,实现基于方剂的药物组成,自动计算推测方剂的整体功效,再进行方剂间整体功效相似度的计算。完成药物组成、功效、性味归经各指标计算后,再通过机器学习的方法,获得方剂 $f1$ 和 $f2$ 的综合相似度 $\text{SimP}(f1, f2)$ 公式的各计算参数,最终完成类方衍化关系自动计算模型。

参考文献

- [1] 朱彦,刘静,刘丽红,等.类方概念探析及形式化表达[J].中国中医药图书情报杂志,2016,40(3):1-5.
- [2] 操牡丹,何前锋,王柏.中医药方剂相似度模型[J].计算机工程,2009,35(16):275-277.
- [3] 顾铮.基于文本分类技术计算中医方剂相似度[J].微计算机信息,2010,26(12):199-201.
- [4] 李园白,崔蒙,杨阳,等.方剂相似度探析[J].中华中医药学刊,2012,30(5):1096-1097.
- [5] 李园白,崔蒙,杨阳,等.哮喘处方相似度研究[J].北京中医药大学学报,2012,35(9):588-590.
- [6] 冯石强,谢鸣.“类方”之内涵探析[J].北京中医药大学学报,2011,34(12):800-803.
- [7] 夏征农,陈至立.辞海(第6版缩印本)[M].上海:上海辞书出版社,2010.
- [8] 中国社会科学院语言研究所词典编辑室.现代汉语词典[M].6版.北京:商务印书馆,2012.
- [9] 陈冠林.基于“药性-功效知识元语义分析”探析中药复方的整体功效[C].海口:中华中医药学会脾胃病分会第二十三次全国脾胃病学术交流会,2011.
- [10] 刘丽红,朱彦,李海燕,等.中医药一体化数据模型的构建[J].中国数字医学,2015,10(10):70-72,91.
- [11] 朱彦,朱玲,高博,等.中药功效语义网络的构建及应用[J].中华中医药杂志,2016,31(5):1710-1715.

(2016-11-14 收稿 责任编辑:杨觉雄)