

数字地形分析应用适配性知识的案例表达与推理方法

吴雪薇^{1,2}, 秦承志^{1,3*}, 朱阿兴^{3,4}

(1. 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, 北京 100101; 2. 中国科学院大学, 北京 100049; 3. 江苏省地理信息资源开发与利用协同创新中心, 南京师范大学地理科学学院, 南京 210023; 4. 威斯康星—麦迪逊大学地理系, 美国 麦迪逊 WI 53706)

摘要:数字地形分析(Digital Terrain Analysis, DTA)在应用时依赖于建模知识,尤其是关于所建的应用模型是否与研究区特点、数据等条件相适配的知识(称为“应用适配性知识”);由于这类知识难以形式化表达,现有的数字地形分析工具对此类知识缺乏利用,从而导致普通用户在设计应用数字地形分析时建模困难。针对该问题,设计了一套数字地形分析领域应用适配性知识的案例表达与相应的推理方法。以美国32个河网提取案例为例,通过交叉验证,初步表明案例及其推理应用方法适合于数字地形分析领域应用适配性知识的形式化表达与应用,该方法通过与建模环境的集成,可大幅降低数字地形分析应用建模难度。

关键词:数字地形分析;建模知识;案例;河网提取

1 引言

数字地形分析(DTA)基于数字高程模型(DEM)对地表形态进行地形属性计算及特征提取,为GIS核心功能空间分析的重要组成部分,已经在地理学及其他相关领域得到广泛应用(周启鸣等, 2006; Hengl et al, 2009)。随着应用领域扩展,用户群不断扩大,越来越多对DTA知识掌握不足的用户开始使用数字地形分析。

数字地形分析应用模型往往是由一系列的算法任务通过一定的逻辑关系组合而成。在实际的应用中,要建立一个适合特定研究区的DTA应用模型,往往涉及到很多DTA领域的建模知识(例如如何选择计算任务,计算任务如何建立连接,如何选择算法、设置参数等)。而现有的DTA应用建模工具在这方面尚难以帮助用户降低建模难度,要依赖

于用户自身的DTA应用建模知识(Qin et al, 2011),因此,DTA应用建模知识不足的非专家用户难以合理、便捷地应用数字地形分析。

DTA建模知识根据其内容可分为两类:模型工作流程构建知识和应用适配性知识(Lu et al, 2012; 卢岩君, 2012)。模型工作流程构建知识指的是建模过程中如何选择计算任务并建立连接。它不仅描述计算任务与其输入输出数据、各元数据值之间的二元关系,同时也描述算法与参数、元数据项之间的二元关系。由于这类知识具有结构性,较易于形式化表达及应用(江净超等, 2013)。Qin等(2011)使用语义网络(Russell et al, 2009)的知识形式化描述方法对模型工作流程构建知识进行表达和推理,获得了较好的效果。DTA应用适配性知识是用于决定建模过程中如何选择适合于应用上下文(如应用目的、数据特征、研究区特征等)的算法、设置合适参

收稿日期:2015-05;修订日期:2015-08。

基金项目:国家自然科学基金项目(41422109, 41431177) [Foundation: National Natural Science Foundation of China, No. 41422109, No. 41431177]。

作者简介:吴雪薇(1990-),女,湖北天门人,硕士研究生,主要从事数字地形分析研究, E-mail: wuxw@lreis.ac.cn。

通讯作者:秦承志(1977-),男,山东蒙阴人,研究员,主要研究方向为数字地形分析,研究兴趣还包括栅格地理计算并行化和地表过程综合建模环境, E-mail: qincz@lreis.ac.cn。

引用格式:吴雪薇, 秦承志, 朱阿兴. 2016. 数字地形分析应用适配性知识的案例表达与推理方法[J]. 地理科学进展, 35(1): 89-97. [Wu X W, Qin C Z, Zhu A X. 2016. Case-based formalization and inference method of application-matching knowledge on digital terrain analysis [J]. Progress in Geography, 35(1): 89-97.]. DOI: 10.18306/dlkxjz.2016.01.010

数的知识 (Lu et al, 2012)。应用适配性知识往往对 DTA 应用模型的实际结果有较大影响,因此在 DTA 应用建模中非常重要。目前,DTA 应用适配性知识属于一种经验知识,尚不够精确、系统,通常难以给出严格的数学表达,现有的应用适配性知识多存在于文献或记录中的应用实例、算法的评价结果、甚至是专家头脑中的经验等。由于 DTA 应用适配性知识尚缺乏形式化表达与推理应用方法,难以在现有的 DTA 工具中有效利用这一类知识。

针对以上问题,设计了一套基于案例的 DTA 应用适配性知识形式化表达与推理应用方法,使 DTA 工具可以利用这一类知识,以减轻用户进行 DTA 应用建模的难度。

2 DTA 应用适配性知识的案例表达及推理应用方法框架

2.1 基本思想

案例是人工智能领域中对非系统性知识的一种常用的表达方法。一个案例至少由两部分组成:问题和解决方案(Kaster et al, 2005)。问题是描述案例的目的以及一些约束条件;解决方案则是要达到这个目的所用的一套方法流程。相应的案例推理方法(Case-based Reasoning, CBR)是人工智能领域中的一种重要的基于案例知识的问题求解和学习方法(Schank, 1983),其基本思想是通过参考已有的相似案例的解决方案应用于新的问题 (Aamodt et al, 1994; Watson, Abdullah, 1994)。与传统的基于规则的知识表达和推理方法相比,基于案例的方法将知识的获取简化为案例的获取,并且无需显式的领域知识表达模型 (Watson, Maria, 1994),适合于缺乏系统性的经验知识表达。

在 DTA 领域,研究区地形条件是影响 DTA 算法选择和参数设置的主要因素。例如面积反映了研究区的区域尺度特征,面积差异较大的研究区可能在算法选择和参数设置上存在差异。目前这些知识多为经验性知识,难以用严格的数学表达式来

形式化表达。因此,基于案例的表达和推理方法在思路适合于 DTA 应用适配性知识 (卢岩君, 2012)。

基于上述思想,设计了 DTA 应用适配性知识基于案例的形式化表达及推理应用方法框架(图 1)。其中,案例形式化表达和相应的推理应用是该方法框架中两个主要阶段,以下分别进行设计。

2.2 基于案例的形式化表达

案例的形式化表达是将案例的属性进行提取并描述的过程。一般地,案例需要描述以下 3 个部分(Kolodner, 1993):案例问题,即案例发生时与案例有关因素的状态,该部分内容直接参与案例推理;解决方案,该部分内容作为案例推理的结果加以应用,不参与案例推理;案例输出,即案例发生后与案例相关因素的状态,该部分内容是可选的,不参与案例推理。

DTA 应用适配性知识的案例表达主要是对案例问题进行形式化描述(表 1),即确定影响 DTA 算法选择和参数设置的主要因素,相应确定适合于后续推理应用的定量指标,以体现 DTA 应用建模相关的应用上下文的状态。对于案例的解决方案,通过记录对应的 DTA 算法名称和所设置的参数值,即可完成形式化。由于 DTA 模型的运行不对其应用上下文产生影响,所以案例表达中不需考虑输出部分。

在形式化描述 DTA 应用适配性知识的案例问题部分时,根据 DTA 应用建模的特点,从应用目的、数据特征、研究区特征等三方面选取对 DTA 算法选择和参数设置有影响的主要因素进行指标化:

(1) 应用目的可形式化为枚举型变量。

(2) 数据特征主要包括 DEM 分辨率、数据源、数据生产方式等。特别是分辨率信息,通常作为 DTA 应用建模中需考虑的一个重要因素,在指标化时可直接对应为一个单一数值。

(3) 研究区特征,主要包括研究区的位置、面积等空间特征,坡度、起伏度等地形特征,其他环境条件(如气候、地质)等。这部分因素较为复杂,需要综合考虑信息完整性和实现上的简洁合理,选择影响

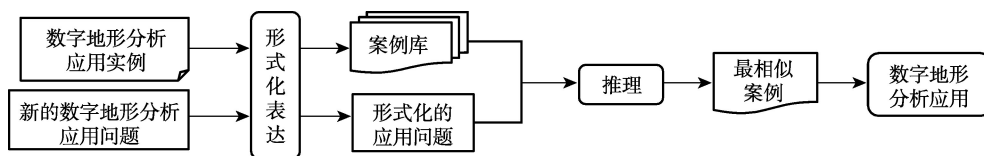


图 1 DTA 应用适配性知识基于案例的形式化表达及推理应用方法框架

Fig.1 Framework of case-based formalization and inference method of application-matching knowledge on DTA

表1 DTA应用适配性知识的组成与案例形式化
表达之间的对应关系

Tab.1 Corresponding relation between the composition
of DTA application-matching knowledge and case-based
formalization

案例组成	DTA应用适配性知识的组成
案例问题	应用目的 数据特征(分辨率、数据源等) 研究区特征(研究区位置、面积、地形特征、 其他环境条件等)
解决方案	DTA算法(及其参数值设置)
案例输出(可选)	(不需考虑)

DTA应用建模的主要因素分别进行指标化,除了枚举型量、单一数值量之外,还有可能用到某种频率分布或函数的形式进行指标化。

2.3 基于案例的推理应用

特征相似的应用问题,通常应选择相同或者相近的解决方案。根据这一思路,对于一个新的DTA应用建模问题,将其同样进行案例形式化之后,逐一计算其与案例库中每个案例间的相似度,取与其最相似的案例作为解决方案。

新的DTA应用建模问题与案例之间的相似度计算过程包括两个步骤:

(1) 单一指标的相似度计算

如前所述,DTA应用适配性知识在案例化时,所采用的指标可能有枚举型(如计算目的)、单一数值型(如DEM分辨率、研究区面积)、某种频率分布或函数的形式(如面积-高程积分曲线),对于每一种指标,需分别设计其对应的相似度算法,用于合理区分同一指标不同指标值之间的相似性程度。通常相似度的值域为[0,1]。

对于枚举型指标,一般采用布尔函数计算相似度值,即指标值匹配时相似度为1,不匹配则相似度为0。对于单一数值型指标,常用的相似度计算方法是简单求差值。如果已有DTA应用中体现出某个单一数值型指标上的数值差异不能直接反映相似性程度时,则需要进行某种数值变换,以合理计算相似度。例如DEM分辨率指标,数量级的差别较数值差别能更好体现相似程度,这时可先进行某种对数变换后,再求差值计算相似度。对于频率分布或函数形式的指标,相似度计算相对复杂,需根据具体的指标含义进行设计。

(2) 案例与新的DTA应用建模问题间的综合相似度计算

对案例与新的DTA应用建模问题间各单一指

标的相似度进行综合,得到一个综合相似度值。可选的常用方法有最小限制因子、加权平均等。

3 方法设计

在以上介绍的方法框架下,本节具体设计案例的指标及相似度计算方法,形成一套DTA应用适配性知识的案例表达及推理应用方法。

在描述DTA应用适配性知识的案例问题时,具体考虑的因素及相应的指标如表2所示:以DEM栅格分辨率来刻画数据特征(考虑到目前实际应用中主要采用栅格DEM,本方法目前设计为针对栅格DTA应用),暂不考虑数据源因素;由于研究区特征包含信息较多,且对于DTA应用建模有关键性的作用,因此设计以面积、地形起伏度、坡度分布、发育特征等4个指标来刻画研究区特征。考虑到空间位置对于DTA应用案例间差异的指示作用,可通过地形起伏度、坡度分布、发育特征等因素体现,因此未单独设置空间位置指标。为简洁起见,目前暂未考虑其他环境条件。

以下具体介绍各指标的设计以及相应的相似度计算方法。

(1) 计算目的

计算目的为枚举型指标,采用布尔函数计算相似度值,即指标值匹配时相似度为1,不匹配则相似度为0。

(2) 栅格分辨率

栅格分辨率是反映案例数据特征的一个重要指标,为单一数值型指标。根据案例与新的应用问题之间,所用的分辨率量级差异越大,相似性越低的认识,采用对数变换后按钟形相似度函数计算分辨率的相似性:

$$S_i = 2^{-(\lg R_{\text{new}} - \lg R_i)^{0.5}}$$

(1)

式中: S_i 是新应用问题与案例库中第 i 个案例的相似度值; R_{new} 、 R_i 分别是新应用问题、第 i 个案例所采用的栅格大小。分辨率相同时,相似度为1;分辨率相差1个数量级时,相似度为0.5。

(3) 面积

根据面积的量级差异越大,相似性越低的认识,计算案例与新应用问题之间面积值常用对数的绝对差,绝对差为0时相似度取为1,取各案例分别计算后的最大值设为相似度0值,线性变换出各案例与新应用问题之间在该指标上的相似度(表2)。

表2 DTA应用适配性知识形式化为案例时的指标设计
Tab.2 Case representation for DTA application-matching knowledge

应用上下文	考虑因素	指标	单一指标的相似度计算方法
应用目的	计算目的	计算目的名称	布尔函数
数据特征	分辨率	栅格分辨率/m	$S_i = 2^{-(\lg R_{\text{new}} - \lg R_i)^{0.5}}$
研究区特征	面积	面积/km ²	$S_i = 1 - S_i' / \max(S_i')$
			$S_i' = \lg(Area_{\text{new}}) - \lg(Area_i) $
	地形特征	地形起伏度/m	$S_i = 1 - S_i' / \max(S_i')$
			$S_i' = Relief_{\text{new}} - Relief_i $
		坡度分布(高程—坡度累积频率分布)	$S_i = \frac{\sum \min(SlpRlf_{\text{new}}, SlpRlf_i)}{\sum \max(SlpRlf_{\text{new}}, SlpRlf_i)}$
		发育特征(面积—高程积分曲线)	$S_i = 1 - S_i' / \max(1 - HI_{\text{new}}, HI_{\text{new}})$
			$S_i' = HI_{\text{new}} - HI_i $

注: S_i 为新应用问题与案例库中第 i 个案例的综合相似度(值域为[0,1]); R_{new} 、 R_i 分别为新应用问题、第 i 个案例的DEM分辨率(m); $Area_{\text{new}}$ 、 $Area_i$ 分别为新应用问题、第 i 个案例的面积(km²); $Relief_{\text{new}}$ 、 $Relief_i$ 分别为新应用问题、第 i 个案例的地形起伏度(m); $SlpRlf_{\text{new}}$ 、 $SlpRlf_i$ 分别为新应用问题、第 i 个案例的高程—坡度累积频率分布直方图; HI_{new} 、 HI_i 分别为新应用问题、第 i 个案例面积—高程积分值。

(4) 地形起伏度

根据地形起伏度差异越大,相似性越低的认识,计算案例与新应用问题之间地形起伏度的绝对差,绝对差为0时相似度取为1,取各案例分别计算后的最大值设为相似度0值,线性变换出各案例与新应用问题之间在该指标上的相似度(表2)。

(5) 坡度分布

为详细描述坡度在研究区中垂直结构的分布特征,基于高程—坡度分布频率进行定量刻画。其中,坡度分0~3°、3~8°、8~15°、15~25°、25~35°、35~45°及45°以上,共计7级(汤国安等, 2006; 朱梅等, 2009)。高程按等间隔分为10个等级,考虑到DEM分辨率对坡度的影响(Chang et al, 1991; Grohmann, 2015),在每个高程分级内,按坡度等级从低到高计算累积频率分布,生成高程—坡度累积频率分布表(为10×7的二维表),用于后续的相似度计算。此处高程采用10等分的方式,只考虑了研究区内部相对高程关系,而没有考虑研究区起伏度的绝对大小,这是由于在地形起伏度指标中已体现了这一信息,此处高程10等分的处理便于相似度算法自动化。

高程—坡度累积频率分布表可形象化为一个高程—坡度累积频率直方图,根据高程—坡度累积频率直方图越相近相似度越高的认识,计算案例与新应用问题高程—坡度累积频率直方图的交集与两直方图的并集的比例值作为相似度(表2)。

(6) 发育特征

面积—高程积分曲线是揭示流域发育特征的重要指标,将其用于刻画研究区发育特征(Strahler,

1952)。Strahler按面积—高程积分值将地貌发育分为3个阶段:幼年阶段(大于0.6),壮年阶段(大于0.35且小于0.6),老年阶段(小于0.35)。

进行相似度计算时,采用的是面积—高程积分值。根据面积—高程积分值差异越大相似度越低的认识,计算案例与新应用问题之间面积—高程积分值的绝对差,绝对差为0时相似度取为1,同时取新应用问题面积—高程积分值与其补集之间的较大者作为相似度0值所对应的指标绝对差,线性变换出各案例与新应用问题之间在该指标上的相似度(表2)。

对于案例综合相似度的计算,设计采用最小限制因子计算,即取应用问题和一个案例间所有单一指标相似度中的最小相似度值作为该应用问题与该案例间的综合相似度。

4 实现

将上述基于案例的推理方法编程实现后,作为知识推理引擎集成到一个基于B/S架构的启发式DTA建模环境软件原型系统中(Qin et al, 2013)。该原型系统此前已采用语义网络对DTA模型工作流程构建知识进行形式化表达和推理应用,以可视化的方式辅助用户完成DTA应用模型工作流的启发式构建(Qin et al, 2013; 江净超等, 2013)。

在原型系统原有的基于语义网络的DTA模型工作流程构建知识库的基础上,增加了DTA应用适配性知识的案例库和用于存储地形数据的数据库。

其中,案例库存储为XML文件格式。

用户进行DTA应用建模时,指定其目标计算任务,原型系统即以启发式的方式实现所需DTA工作流的交互构建。原型系统采用上述方法进行自动的算法选择和参数设置,对用户的应用问题进行与案例相应的形式化表达(即提取上述方法中各个指标的),然后调用案例知识推理引擎访问案例库,推理计算出与用户应用问题最相似的案例,将最相似案例所用的算法和参数作为解决方案,设置到DTA工作流程中,从而构建出适合于用户应用问题的DTA模型。该建模过程可减轻用户进行DTA应用建模的难度。

5 应用验证

基于栅格DEM提取河网为例,对上述方法进行应用验证。基于栅格DEM提取河网的通常步骤为:①对研究区栅格DEM进行填洼、去平区等预处理;②利用流向算法计算研究区的汇水面积空间分布;③设置汇水面积阈值,提取河网。其中,流向算法的选择(秦承志等, 2006)、提取河网时汇水面积阈值的设置,都会涉及DTA应用适配性知识。河网提取时通常采用的流向算法都是D8算法,而汇水面

积阈值则是一个经验性参数,直接影响最终河网提取效果。该阈值通常随研究区地形特征而变化,因此非专家用户难以快速确定一个合适的阈值。以下将通过河网提取时汇水面积阈值的自动确定表明DTA应用适配性知识案例化推理的有效性。

5.1 案例库准备

在河网提取案例准备时,主要基于SRTM DEM(分辨率为90 m),采用D8算法计算汇水面积,各案例区的汇水面积阈值是通过与参考河网进行比对来确定,以保证案例中所用的汇水面积阈值符合实际。用作参考的河网数据和流域边界数据为美国地质勘探局(USGS)发布的NHD(National Hydrography Dataset)数据和WBD(Watershed Boundary Dataset)数据。得到23个河网提取案例,这些案例分布在美国的华盛顿州(6个案例,编号为WA_01至WA_06)、阿肯色州(8个案例,编号为AR_01至AR_08)、缅因州(9个案例,编号为ME_01至ME_09)(图2)。

为体现案例中分辨率指标的作用,在9个案例研究区同时采用ASTER GDEM(分辨率30 m)数据进行了河网提取,得到一组30 m分辨率的案例(案例编号为对应90 m分辨率的案例编号加后缀“A”,例如WA_04A)。总计有32个河网提取案例存入案例库(图2),案例的分布区域具有明显差异,相应的

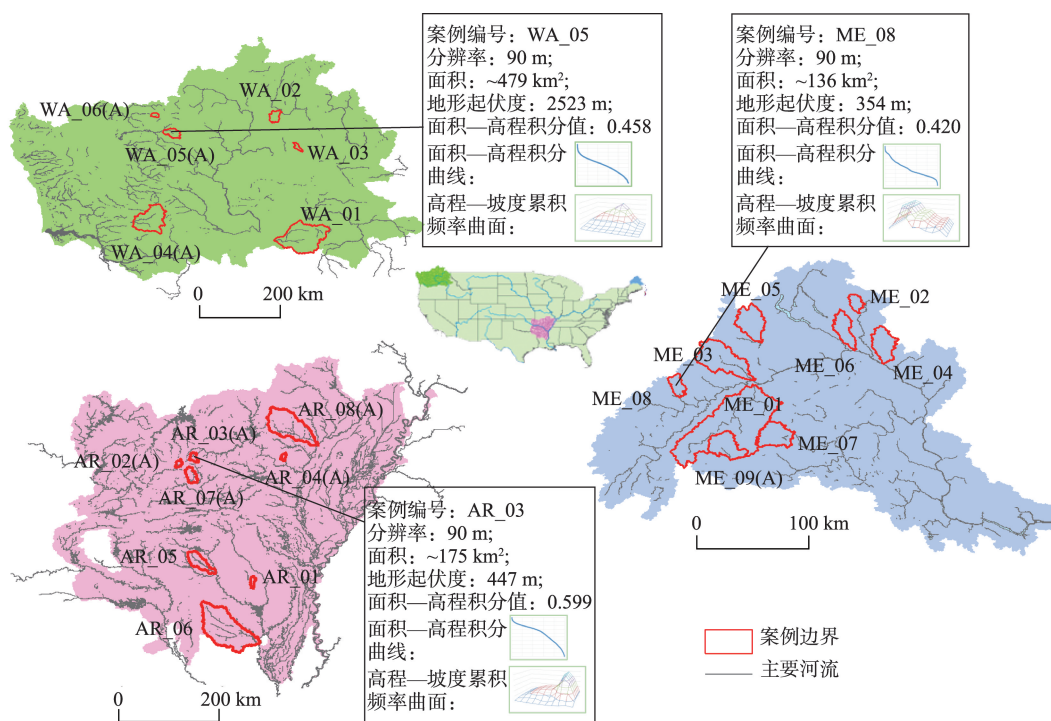


图2 案例分布图及形式化案例实例

Fig.2 Case distribution and formalized instance

汇水面积阈值也各有差异。例如案例WA_05、ME_08、AR_03分别来自于不同的州,其各属性也表现出一定的差异性,特别是高程—坡度累积频率曲面,可以看出三个案例差异较大;另外,面积—高程积分曲线有一定的差异,其对应的面积—高程积分值也各有不同,说明采用面积—高程积分值来计算面积—高程积分曲线的相似度有一定的合理性。

5.2 应用及评价方式

为对上述方法进行评价,采用交叉验证的方式,即每次以一个案例作为应用问题,案例库中其余的31个案例用于推理,计算各案例与应用问题间的综合相似度。比较分析最相似案例与应用问题的汇水面积阈值:如果与其他案例相比,应用问题与最相似案例两者阈值最接近,则可认为推理结果合理;如果有多个案例都与应用问题具有较高的相似性,推理出最相似案例对应阈值虽不是最接近,但是较为接近的(本实验中设为其在所有案例中的接近程度排在前10以内),也可以认为推理结果较为合理;否则为推理结果偏差较大。

5.3 结果及分析

交叉验证所得到的32个实验结果中,根据上述评价方式,有16个验证结果是合理的,7个验证结果是较为合理的,9个不满足合理条件(表3)。在推理结果合理的验证实验中,注意到以AR_06[17.42](表示AR_06案例所用汇水面积阈值为17.42 km²,下同)为验证案例时,得到的最相似案例是AR_08[7.96],虽然这两个阈值差异仍较大,但AR_08相对于其他案例而言,与AR_06阈值最接近,并且两者较低的相似度值也指示了最相似案例的不确定性较高。

以案例WA_01[1.60]的验证结果为例,推理得到的最相似案例为ME_03[1.27],将其阈值(1.27 km²)应用到案例WA_01上,得到的河网提取结果与案例WA_01原阈值(1.60 km²)对应的河网之间,空间分布接近(图3),表明了推理结果的合理性。

在推理结果较为合理的验证实验中,注意到AR_03A[0.08]作为应用案例进行验证时,得出其最相似案例是AR_02A[0.06],而非案例库中与AR_03A为同一研究区且阈值最接近的案例AR_03[0.08]。这是由于案例AR_03A、AR_03的分辨率分别为30 m、90 m,导致两者在分辨率指标上的相似度较低,并最终体现为两者间的综合相似度较低。与此同时,案例AR_02A是阈值与AR_03A第4接近的,因此可认为AR_03A的推理结果是较合理

表3 交叉验证结果统计

Tab.3 Results of cross validation

结果合理性	应用问题案例 [阈值/km ²]	最相似案例 [阈值/km ²]	案例综合相 似度
合理	WA_01[1.60]	ME_03[1.27]	0.43
	WA_03[0.26]	WA_02[0.74]	0.69
	WA_05[0.61]	WA_05A[0.61]	0.62
	AR_05[1.02]	ME_09[0.72]	0.62
	AR_06[17.42]	AR_08[7.96]	0.65
	ME_04[0.79]	ME_05[0.84]	0.94
	ME_05[0.84]	ME_04[0.79]	0.93
	ME_06[0.81]	ME_04[0.79]	0.82
	ME_07[0.78]	ME_04[0.79]	0.88
	ME_08[0.74]	ME_04[0.79]	0.75
	AR_08A[7.96]	AR_08[7.96]	0.62
	WA_05A[0.61]	WA_05[0.61]	0.62
	WA_06A[0.08]	WA_06[0.08]	0.62
	WA_04A[1.26]	WA_04[1.26]	0.62
	WA_04[1.26]	WA_04A[1.26]	0.62
	WA_06[0.08]	WA_06A[0.08]	0.62
较为合理	AR_07[2.16]	ME_06[0.81]	0.76
	AR_08[7.96]	ME_01[1.22]	0.87
	AR_03A[0.08]	AR_02A[0.06]	0.86
	AR_02A[0.06]	AR_03A[0.08]	0.87
	AR_02[0.06]	AR_03[0.08]	0.87
	AR_03[0.08]	AR_02[0.06]	0.86
	ME_09[0.72]	ME_05[0.84]	0.85
偏差较大	WA_02[0.74]	AR_03[0.08]	0.65
	AR_01[0.11]	ME_08[0.74]	0.64
	AR_04[0.03]	ME_02[0.28]	0.82
	ME_01[1.22]	AR_08[7.96]	0.87
	ME_02[0.28]	AR_04[0.03]	0.82
	ME_03[1.27]	AR_07[2.16]	0.65
	AR_03A[0.08]	AR_07A[2.16]	0.65
	ME_09A[0.72]	AR_07A[2.16]	0.62
	AR_04A[0.03]	ME_09A[0.72]	0.65

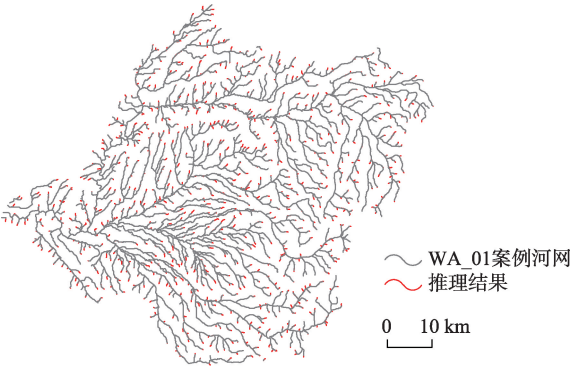


图3 WA_01案例的河网与推理结果对比图
Fig.3 Comparison between the river network and the reasoning result of the WA_01 case

的。其他推理结果较为合理的验证实验也大多是类似情况。

推理结果不太合理的情况出现的原因主要是验证案例与案例库中所有案例的相似度都较低,这就表示案例库中没有与验证案例非常相似的案例,导致选择的最相似案例阈值与验证案例不是很接近。进一步以 ME_03[1.27]为应用问题时的验证实验为例,该应用与其他大多数案例的综合相似度都取决于面积指标的相似度(表4),目前得到的最相似案例的综合相似度值不高,且两者阈值差别较大。而案例ME_05与ME_03在其他指标上均有很高的

相似度,且两者的汇水面积阈值相近,但受制于面积指标上相似度较低而未能成为合理的解决方案。注意到目前面积指标的相似度计算与案例库的大小有关,这类问题有可能通过扩大案例库而减少。

从表3还可看出,出现合理结果的交叉验证实验中有几个案例(如案例AR_05)与推理出的最相似案例并不在同一个州,表明常规认识中距离较近的流域往往具有较高的相似性不一定准确,从这个意义上来看,本文方法可以补充简单套用单一因素下经验知识的不足。

表4 案例ME_03交叉验证结果的相似度计算结果
Tab.4 Similarity values from cross validation of case ME_03

案例编号	分辨率相似度	面积相似度	面积高程积分曲 线相似度	地形起伏度	高程—坡度累积 频率相似度	综合相似度	汇水面积 阈值/km ²
AR_07	1	0.65	0.98	0.98	0.75	0.65	2.16
ME_05	1	0.62	0.94	0.97	0.94	0.62	0.84
AR_07A	0.62	0.65	0.98	0.98	0.71	0.62	2.16
ME_07	1	0.61	0.83	0.96	0.92	0.61	0.78
ME_09	1	0.61	0.89	0.94	0.84	0.61	0.72
ME_09A	0.62	0.60	0.93	0.96	0.90	0.60	0.72
ME_04	1	0.55	0.92	0.95	0.95	0.55	0.79
AR_05	1	0.98	0.84	0.91	0.54	0.54	1.02
AR_08	1	0.48	0.79	0.95	0.80	0.48	7.96
AR_08A	0.62	0.48	0.80	0.95	0.85	0.48	7.96
ME_06	1	0.46	0.94	0.96	0.81	0.46	0.81
ME_01	1	0.45	0.9	0.97	0.89	0.45	1.22
WA_02	1	0.66	0.79	0.82	0.35	0.35	0.74
WA_01	1	0.32	0.82	0.69	0.80	0.32	1.62
AR_06	1	0.28	0.88	0.89	0.56	0.28	17.42
AR_03A	0.62	0.28	0.66	0.98	0.41	0.28	0.08
AR_03	1	0.28	0.66	0.98	0.44	0.28	0.08
WA_03	1	0.19	0.94	0.82	0.31	0.19	0.26
ME_02	1	0.17	0.88	0.95	0.79	0.17	0.28
ME_08	1	0.17	0.96	0.95	0.79	0.17	0.75
AR_01	1	0.17	0.88	0.87	0.45	0.17	0.11
WA_05A	0.62	0.71	0.92	0.41	0.16	0.16	0.61
WA_05	1	0.71	0.90	0.43	0.16	0.16	0.61
AR_02	1	0.08	0.76	0.99	0.39	0.08	0.06
WA_06	1	0.08	0.91	0.56	0.22	0.08	0.08
WA_06A	0.62	0.08	0.91	0.55	0.22	0.08	0.08
AR_02A	0.62	0.08	0.76	0.99	0.38	0.08	0.06
WA_04	1	0.55	0.70	0.02	0.35	0.02	1.26
AR_04A	0.62	0	0.80	0.90	0.79	0	0.03
AR_04	1	0	0.82	0.90	0.85	0	0.03
WA_04A	0.62	0.55	0.70	0	0.34	0	1.26

应用结果初步验证了DTA应用适配性知识案例表达与推理应用方法的可行性。

6 结论与讨论

针对普通用户数字地形分析应用建模困难,现有DTA工具无法充分利用DTA应用建模知识的问题,探索了案例推理在数字地形分析领域应用适配性知识表达和应用上的可用性,设计了一套用于描述案例的指标及其相似度计算方法。以栅格DEM提取河网为例,选取了位于美国的32个案例进行交叉验证,结果初步反映出案例推理的方法适合于DTA领域应用适配性知识的形式化表达与推理应用,能够大幅度降低数字地形分析应用建模难度。

进一步的研究工作主要有以下几个方面:

一是推理方法的完善。方法在设计时尽量保证所考虑因素的合理性,但对于刻画各个因素具体指标的设计,还有待进一步的研究进行修订完善。目前的方法假定各案例的解决方案具有很高的可靠性,但在实际情况中,可能由于人为误差或者其他原因,导致案例的解决方案并不完全可靠(例如河网案例中的阈值在实际应用中并不能得到与真实河网完全一致的结果),后续研究可尝试在相似度计算中考虑案例的可信度因素。

二是解决方案的调整。目前方法中案例推理出的最相似案例,其解决方法直接应用于新的研究问题,没有经过修订和调整,如何根据新应用问题的特征对解决方案进行调整也是下一步的推理方法研究内容。

三是案例库的建立。目前的案例为人工获取,效率较低,且来源比较单一,今后可研发半自动的案例获取方法,以快速建立更大规模的案例库。

四是其他应用问题的应用及验证。目前仅对河网提取这一应用问题进行了应用验证,下一步将扩展到其他数字地形分析应用问题的应用验证。

参考文献(References)

江净超, 朱阿兴, 秦承志, 等. 2013. CyberSoLIM: 基于知识驱动的在线数字土壤制图原型系统[J]. 土壤学报, 50(6): 1216-1220. [Jiang J C, Zhu A X, Qin C Z, et al. 2013. CyberSoLIM: knowledge-driven on-line digital soil mapping prototype system[J]. Acta Pedologica Sinica, 50(6): 1216-1220.]

卢岩君. 2012. 知识驱动下的数字地形分析可视化建模环境

构建研究[D]. 北京: 北京师范大学. [Lu Y J. 2012. Construction of a knowledge-driven and visualized modelling environment for digital terrain analysis[D]. Beijing, China: Beijing Normal University.]

秦承志, 朱阿兴, 李宝林, 等. 2006. 基于栅格DEM的多流向算法述评[J]. 地学前缘, 13(3): 91-98. [Qin C Z, Zhu A X, Li B L, et al. 2006. Review of multiple flow direction algorithms based on gridded digital elevation models[J]. Earth Science Frontiers, 13(3): 91-98.]

汤国安, 宋佳. 2006. 基于DEM坡度图制图中坡度分级方法的比较研究[J]. 水土保持学报, 20(2): 157-160. [Tang G A, Song J. 2006. Comparison of slope classification methods in slope mapping from DEMs[J]. Journal of Soil and Water Conservation, 20(2): 157-160.]

周启鸣, 刘学军. 2006. 数字地形分析[M]. 北京: 科学出版社. [Zhou Q M, Liu X J. 2006. Digital terrain analysis [M]. Beijing, China: Science Press.]

朱梅, 李发源. 2009. 坡度分级对地面坡谱的影响研究[J]. 测绘科学, 34(6): 165-167. [Zhu M, Li F Y. 2009. Influence of slope classification on slope spectrum[J]. Science of Surveying and Mapping, 34(6): 165-167.]

Aamodt A, Plaza E. 1994. Case-based reasoning: foundational issues, methodological variations, and system approaches [J]. AI Communications, 7(1): 39-59.

Chang K T, Tsai B W. 1991. The effect of DEM resolution on slope and aspect mapping[J]. Cartography and Geographic Information Systems, 18(1): 69-77.

Grohmann C H. 2015. Effects of spatial resolution on slope and aspect derivation for regional-scale analysis[J]. Computers & Geosciences, 77: 111-117.

Hengl T, Reuter H I. 2009. Geomorphometry: concepts, software, applications[M]. Amsterdam, Netherlands: Elsevier.

Kaster D S, Medeiros C B, Rocha H V. 2005. Supporting modeling and problem solving from precedent experiences: the role of workflows and case-based reasoning[J]. Environmental Modelling & Software, 20(6): 689-704.

Kolodner J. 1993. Case-based reasoning[M]. San Mateo, CA: Morgan Kaufmann Publishers.

Lu Y, Qin C Z, Zhu A X, et al. 2012. Application-matching knowledge based engine for a modelling environment for digital terrain analysis[R]. GeoInformatics 2012, 15-17 June. Hong Kong, China: The Chinese University of Hong Kong.

Qin C Z, Jiang J C, Zhan L J, et al. 2013. A browser/server-based prototype of heuristic modelling environment for digital terrain analysis[R]. Geomorphometry 2013, 15-20 October. Nanjing, China: Nanjing Normal University.

Qin C Z, Lu Y J, Zhu A X, et al. 2011. Software prototyping of

- a heuristic and visualized modeling environment for digital terrain analysis[C]//Proceedings of the 11th International Conference on GeoComputation, 20-22 July. London, UK: University College London.
- Russell S, Norvig P. 2009. Artificial intelligence: a modern approach[M]. 3rd ed. New Jersey, NJ: Prentice Hall.
- Schank R C. 1983. Dynamic memory: a theory of reminding and learning in computers and people[M]. New York, NY: Cambridge University Press.
- Strahler A N. 1952. Hypsometric(area-altitude)analysis of erosional topography[J]. Geological Society of America Bulletin, 63(11): 1117-1142.
- Watson I, Abdullah S. 1994. Developing case-based reasoning systems: a case study in diagnosing building defects[C]//Case Based Reasoning: Prospects for Applications (Digest No. 1994/057). UK. IET: 1-3.
- Watson I, Marir F. 1994. Case-based reasoning: a review[J]. The Knowledge Engineering Review, 9(4): 327-354.

Case-based formalization and inference method of application-matching knowledge on digital terrain analysis

WU Xuewei^{1,2}, QIN Chengzhi^{1,3*}, ZHU Axing^{3,4}

(1. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China; 3. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, School of Geography, Nanjing Normal University, Nanjing 210023, China; 4. Department of Geography, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA)

Abstract: Application of digital terrain analysis (DTA) relies heavily on the DTA-domain knowledge on the match between the chosen algorithm (and its parameter-settings) and the application context (such as target task, terrain condition of the study area, and DEM resolution)—the so-called application-matching knowledge. This type of knowledge has a direct impact on the quality of DTA modelling when users of DTA, especially non-expert users, do not have sufficient amount of such knowledge to support their DTA applications. Existing DTA-assisted tools often cannot use application-matching knowledge because this type of knowledge has not been formalized in DTA to be available for inference in these tools. This is mainly because this type of DTA knowledge is currently inaccurate and non-systematic, and often exists in documents for specific case studies, or as personal knowledge of domain experts. This situation makes the DTA modelling process difficult for users, especially for non-expert users. Case-based reasoning method that originated from artificial intelligence is appropriate for formalization and inference of non-systematic knowledge. In this article, we propose a case-based formalization and inference method for the application-matching knowledge in DTA. The specific design of the proposed case-based method can be divided into two parts: formalization of the application-matching knowledge, and inference method. The case of this knowledge consists of a series of indices to formalize the DTA application-matching knowledge and the corresponding similarity calculation methods for inference based on the case. To evaluate the performance of the proposed method, we implemented it in a software prototype of DTA modelling environment and then applied it to a DTA application of river network extraction. In the experiment we prepared 32 cases of river network extraction in the USA. The results of cross validation preliminarily show that the proposed case-based method is suitable for using the application-matching knowledge in DTA. It reduced the modelling burden greatly for users.

Key words: Digital Terrain Analysis (DTA); modelling knowledge; case; river network extraction