基于 FME 的地理国情监测

变化信息通用属性项检查

陈兴波 宁夏回族自治区基础测绘院

1 绪论

1.1 背景

基础性地理国情监测引入的变化信息通用属性项,通过人机交互检查容易出 现错漏,且速度慢,因此需要设计对应的自动化检查流程,提高检查正确率和效 率。

1.2 难点

基础性地理国情监测引入了变化信息通用属性项,需要对比更新数据(下简称"新数据")与本底数据(下简称"旧数据")后进行填写,且图形结点数量变 化不视为变化。

(1) 地表覆盖分类变化信息通用属性项

地表覆盖分类变化信息通用属性项含有 ChangeType 一项,用于标识地表覆盖图斑的不同变化类型,取值范围见表 1-1^[1]。

取值	含义
1	一般型伸缩
2	新增
4	细化
-3	合并型伸缩
-2	精细型伸缩
9	改错
-9999	无变化

表 1-1 地表覆盖 Change Type 取值及含义

地表覆盖 ChangeType 存在特殊情况,即本底数据原有图斑灭失,但监测期 在原图斑相邻区域出现了与共边或共点的新图斑, ChangeType 赋值为"2"^[2]。 ChangeType 的填写与图形变化对应,较为复杂。

(2) 地理国情要素变化信息通用属性项

地理国情要素变化信息通用属性项含有 ChangeType 和 ChangeAtt,分别用 于标识国情要素图形的不同变化类型和专有属性项的修改情况,取值范围见表 1-2^[2]。

取值	含义			
1	一般型伸缩			
2	新增			
3	删除			
0	属性变化			
-1	打断、分割			
-2	精细型伸缩			
9	改错			
-9999	无变化			

表 1-2 国情要素 Change Type 取值及含义

ChangeAtt 填写需要对比两期要素间的属性项,未变化的填写缺省值,有变化的填写变化字段名称,人工填写、检查容易出现错漏。

1.3 现有质检软件

(1) 基础性地理国情监测数据入库检查工具

该工具由国家基础地理信息中心与吉威时代公司联合开发,主要用于基础性 地理国情监测入库检查。针对技术难点(1)进行检查,运行速度慢,容易出现 漏报。

(2) 地理国情监测检验系统 2018

该系统由四川省测绘产品质量监督检验站开发,针对基础性地理国情监测进 行检查,可以检查技术难点(1)和(2),运行速度较快,但存在漏报。

(3) ChangeType 检查方法^[3]

该方法由中国矿业大学资源与测绘学院、河北省第三测绘院与山东省国土测 绘院联合设计,针对 ChangeType 属性字段进行检查。检查效果和效率较好,但 以面积容差作为检查指标,与规范中的容差要求不一致。且检查思路为优先判断 ChangeType 赋值情况,再根据赋值情况进行检查,若属性填写错误,易造成误 判,较为依赖作业员属性填写正确性。

2. 检查流程

2.1 检查规则

(1) 使用 FME 中的转换器,组合判断图形的变化情况,对 ChangeTpye 赋 值情况进行检查。

ChangeType 取值	判断规则	比对原则	错误提示
1, -2, -3	更新图斑在本底数据中存在 CC 相	赋值与枚举值不	伸缩图斑,
	同,位置相交的图斑,且两者容差	一致,提示错误	ChangeType

表 2-1 地表覆盖数据 ChangeType 枚举值判断规则。

	大于 0.01 米 ^[2]	赋值错误
2	更新图斑在本底数据中存在 CC 不 相同,位置相交的图斑	新增图斑, ChangeType 赋值错误
4	更新图斑在本底数据中存在 CC 不 相同的图斑,但更新图斑被本底数 据中 CC 上级类图斑所包含	M值相获 细化图斑, ChangeType 赋值错误
9	更新图斑在本底数据中存在 CC 不 相同,但图形相同图斑	纠错图斑, ChangeType 赋值错误
-9999	更新图斑在本底数据中存在 CC 相同,图形相同的图斑	无变化图斑, ChangeType 赋值错误

表 2-2 国情要素数据 ChangeType 枚举值判断规则

ChangeType 取值	判断规则	对比原则	错误提示				
-1	更新要素在本底数据中存在包含其 的要素		打断要素, ChangeType 填写错误				
0	更新要素在本底数据中存在图形相 同的要素,但非通用属性项发生了 变化,赋值与枚举值不一致,提示 错误	新要素在本底数据中存在图形相 的要素,但非通用属性项发生了 化,赋值与枚举值不一致,提示 吴 赋值与枚举值不					
1, -2, 9	更新图斑在本底数据中存在图形不 相同,但位置相交的要素,且两者 容差大于 0.01 米 ^[2]	一致,旋小锥族	伸缩要素, ChangeType 填写错误				
2	更新要素在本底数据中存在图形不 相同,且位置不相交的要素		新增要素, ChangeType 填写错误				
3	更新要素在本底数据中存在图形相	图形、属性与本 底不一致	删除要素,图 形和属性不 应发生变化				
	同,非通用属性项相同的图斑	ChangeAtt 不为 "_"	删除要素, ChangeAtt填 写错误				
-9999	更新要素在本底数据中存在图形相 同,非通用属性项相同的图斑,赋 值与枚举值不一致,提示错误	赋值与枚举值不 一致,提示错误	无变化要素, ChangeType 填写错误				

(2)使用 FME 中的转换器,组合提取伸缩和属性变化要素的变化属性项, 再对 ChangeAtt 赋值情况进行检查。 2.2 功能实现

2.2.1 地表覆盖分类变化信息通用属性项检查

(1) 通过 ChangeDetector 转换器比较新旧数据中的地表覆盖图层的图形和 CC 值,得到部分无变化图斑;

(2)输出比较后各自不一致的部分,再次比较图形,得到图形无变化,CC 值变化的图斑,即改错图斑;

(3) 遍历(2) 中输出的新图斑,查找旧数据中有无与新图斑 CC 值相同, 空间相交的旧图斑,若有,则为伸缩图斑,若无,则为新增图斑;

(4)根据九交模型^{[4][5]},遍历(3)中输出的伸缩图斑,查找旧数据中有无 与新图斑 CC 值相同,共边或共点的图斑,若有,则为伪伸缩图斑,若无,则为 伸缩图斑;

(5) 遍历(4) 中输出的伪伸缩图斑,将新图斑与 CC 值相同的旧图斑做面相交,容差设置为 0.01 米,判断两者有无重叠部分,若有,则为伸缩图斑,若无,则为新增图斑;

(6) 遍历(4)和(5)中输出的伸缩图斑,将新图斑与旧图斑做面相交,容差设置为0.01米,判断两者重叠部分数量,若等于1,则为无变化图斑,若大于1,则为伸缩图斑;

(7) 遍历(3) 和(5) 中输出的新增图斑,查找旧数据中有无包含新图斑 且 CC 值前两位相同,且新图斑 CC 值大于旧图斑 CC 值的图斑,若有,则为细 化图斑,若无,则为伸缩图斑;

(8)将(1)至(7)输出结果汇总,得到无变化图斑、伸缩图斑、新增图 斑、细化图斑与纠错图斑,比较 ChangeType 赋值与分析结果,输出不一致图斑, 填写错误说明,辅助修改。



检查流程如图 2-1 所示,图中红色部分即为最终输出的图斑检查结果。

图 2-1 地表覆盖分类变化信息通用属性项检查流程示意图

by b

图 2-2 地表覆盖分类变化信息通用属性项检查模板截图

2.2.2 地理国情要素变化信息通用属性项检查

(1) 从新要素中,根据填写的 ChangeType 值,筛选出删除要素与非删除要素,将删除要素与空间位置对应的旧要素进行图形和属性项比较,若有变化,则提示错误,若无变化,则为删除要素,检查 ChangeAtt 是否为"-",若不为"-",则提示错误;

(2)将(1)中输出的非删除要素通过 ChangeDetector 转换器与旧要素进行 比较,判断图形和属性项变化情况,得到无变化要素;

(3)将(2)中输出的比较后不一致部分,再次进行比较,得到图形无变化, 属性变化要素;

(4)遍历(3)中输出的旧要素,判断新要素是否被其包含,若被包含,则为打断(分割)要素,若不被包含,则为伸缩要素或新增要素;

(5) 遍历(3) 中输出的图形无变化,属性变化要素,筛选出与其对应的旧要素,使用 Matcher 转换器关联 FEATID,比较新、旧要素属性项异同,提取变化的属性项名称,与 ChangeAtt 比较,若不一致,则提示错误,并提示应填写的正确内容;

(6) 遍历(3) 中输出的旧要素,判断与(4) 中输出的伸缩要素或新增要素是否相交,若相交,则为伸缩要素,若不相交,则为新增要素;

(7) 遍历(6) 中输出的伸缩要素,判断其 FEATID 赋值情况,若为"-" 或空值,则为新增要素,若存在 FEATID,则为伸缩要素,并使用 Matcher 转换 器关联旧要素 FEATID,比较新、旧要素属性项异同,提取变化的属性项名称, 与 ChangeAtt 比较,若不一致,则提示错误,并提示应填写的正确内容;

(8)将(1)至(7)输出结果汇总,得到删除要素、无变化要素、打断(分割)要素、属性变化要素、伸缩要素与新增要素,比较 ChangeType 赋值与要素分析结果,输出不一致要素,并填写错误说明,辅助修改。

检查流程如图 2-2 所示,图中红色部分即为最终输出的要素检查结果。

5



图 2-3 地理国情要素变化信息通用属性项检查流程示意图



图 2-4 地理国情要素变化信息通用属性项检查模板截图



图 2-5 ChangeAtt 对比模板截图

3 检查案例与评价

3.1 评价方法

检查结果直接关乎数据质量的好坏,因此需做到"零漏报与低误报"。将本

方法的检查结果,与现有质检软件的检查结果进行比较。在此前提下,建立漏报 率与误报率的评价方法[6]。

$$FNR = \frac{FN}{TP + FN}$$

式中, FNR 代表漏报率, FN 代表漏报个数, TP 代表正确个数。

$$FPR = \frac{FP}{TN + FP}$$

гр

式中, FPR 代表误报率, FP 代表误报个数, TN 代表无错误个数。

$$PR = \frac{TP}{TP + FP}$$

式中,PR 代表准确率,TP 代表正确个数,FP 代表误报个数。

其中将本方法检查结果,与现有质检软件的检查结果逐个排查,取全部正确 检查结果的合集,作为错误总量。

3.2 案例分析

以 2018 年基础性地理国情监测宁夏回族自治区银川市永宁县的地表覆盖与 部分点、线、面国情要素数据作为实验数据,进行实验验证。

表 3-1 实验数据情况表

监测区 名称	监测区		图斑、雪	要素数量		错误总数			
	面积 (km ²)	LCRA	LRDL	HYDA	SFCP	LCRA	LRDL	HYDA	SFCP
永宁县	934.28	35813	1032	910	192	50	8	12	1

实验电脑使用 Intel i7 6700 处理器、64G DDR4 内存、4TB 机械硬盘、Windows 7 旗舰版操作系统, FME 2018.1。实验结果见表 3-2。

表 3-2 实验情况统计表

检查方法	要素类型	漏报数	漏报率	误报数	误报率	准确率	耗时
基础性地理 国情监测数 据入库检查 工具	LCRA	37	74.00%	0	0.00%	100%	41'11"
地理国情监 测检验系统 2018	LCRA	33	66.00%	30	0.08%	36.17%	15'48"
	LRDL	3	37.50%	4	0.39%	55.56%	
	HYDA	7	58.33%	0	0%	100%	
	SFCP	1	100.00%	0	0%	0%	
本文方法	LCRA	0	0%	1	0.003%	98.04%	
	LRDL	0	0%	3	0.29%	72.73%	2202
	HYDA	0	0%	2	0.22%	85.71%	3.6
	SFCP	0	0%	0	0%	100%	

注:基础性地理国情监测数据入库检查工具和地理国情监测检验系统 2018 漏报多为伪伸 缩图斑 ChangeType 赋值错误,可能是检查规则未设置。

从表 3-2 可以看出,本文提出的方法,与现有检查软件比较,能做到无漏报, 低误报,准确率高,且相对耗时较短,能够满足基础性地理国情监测检查要求。

4 总结与展望

4.1 总结

(1)本文中使用了 ChangeDetector 作为初始数据比较工具,后续地表覆盖 比较时,经过比较与实验,确定使用 SpatialFilter 代替 SpatialRelator 查找比较相 同 CC 值的图斑,获取新增和伸缩图斑,大幅提高了处理效率。

(2)由于 FME 需要处理多种格式的数据,不能使用数据自身的容差,导致数据比较、选择时,容易出现错漏^[7]。后经查询,并与 FME 中国的徐珩交流后,采用了添加 Bufferer 解决这类问题,但一定程度上降低了数据处理效率。

(3)在从伸缩图斑中排除伪伸缩图斑和结点数量变化导致未变化图斑时, 使用了 AreaOnAreaOverlayer,设置容差后,可以较好的得到想要的结果。判断 是否重叠和重叠数量,可以很好的解决问题。

(4) 在要素属性比较时,使用 Matcher 关联属性,使用 Tester 和 AttributeCreator 比较属性项并创建列表,最后将列表写入属性,解决了属性比较 的问题。

(5) 很多情况下,数据筛选需要经过复杂的步骤,容易改变原始数据,可以使用 Counter 和 FeatureMerger 组合的方式进行操作。

(5)本文根据基础性地理国情监测有关规定,结合实际生产作业情况,制 定了详细的检查规则,基于 FME 设计了高效的检查流程,目前该检查工具已运 用于本单位基础性地理国情监测成果质量检查当中,大幅提高了工作效率与数据 质量。基于 FME 的检查流程,逻辑清晰,可控性强,随时可以根据检查要求进 行调整,以适应和提高检查效率与正确性。

4.2 展望

(1) 截止本文完成时, SAFE Software 已经推出了 FME 2019 试用版。对本 文所用到的关键转换器 ChangeDetector 和 Matcher 进行了升级,两者都针对几何 图形比较引入了容差参数选项,省去了因为需要额外使用多个转换器的麻烦,提 高了工作效率。同时 ChangeDetector 引入了更新(Updated)和插入(Inserted) 输出项,可以更好的帮助判断要素变化情况^[8]。

(2)目前针对线状要素 ChangeType 的判断仍有一定难度,伸缩要素可以首和(或)尾结点不变,而其他结点发生变化,从而导致图形伸缩,但该情况与新增一条线状要素与原现状要素相交情况一样,难以区分。虽可以通过 FEATID 赋

8

值进行判断,但由于国情要素属性项多,作业员采集时多采用复制、粘贴方式赋 属性,容易造成误判。希望有新的思路或者转换器可以解决问题。

(3) ChangeAtt 赋值正确性检查需要对比每个非通用属性项,流程较长,因此将其放入了自定义转化器中,目前除了 FeatureReader 没有其他转换器可以读取数据结构,但其无法读取处理过程中的数据,因此无法用循环来对比属性项。希望有新的转换器可以解决问题。

(4)虽然检查速度已经够快,但面对面积更大、图斑数量更多的监测区时,运行速度仍会下降。后期将对 Workbench 进行优化,以加快检查速度^[9]。

(5)理论上该 FME 模板可以用来直接进行赋值操作,但由于存在误报,导 致需要人工排查错误,进行修改。对误报项进行分析后,发现是转换器容差所致。 希望可以针对不同格式的数据,提供不同选项的容差设置。 参考文献

[1] GQJC 01-2018 基础性地理国情监测数据技术规定[S]. 北京: 国家基础地理信息中心, 2018.

[2] 国家基础地理信息中心. 关于印发《2018 年基础性地理国情监测有关问题解答》的通知 [Z]. 北京: 2018, 11.

[3] 刁丰华, 李燕燕, 丁莹莹. 基于规则约束的地理国情监测变化信息属性正确性自动检查 方法及实践[J]. 工程勘察, 2019(1): 51-55.

[4] Clementini E , Di Felice P , Van Oosterom P . A small set of formal topological relationships suitable for end-user interaction[C]// International Symposium on Advances in Spatial Databases. Springer-Verlag, 1993.

[5] Clementini E , Sharma J , Egenhofer M J . Modelling topological spatial relations: Strategies for query processing[J]. Computers & Graphics, 1994, 18(6):0-822.

[6] Fawcett T. An introduction to ROC analysis[J]. Pattern Recognition Letters, 2006, 27(8): 861-874.

[7] Mark Ireland. FME 2018 Infinity War: How Automatic Tolerance Defeats Infinite Precision without a Snap – but with Anchored Vertices[EB/OL]. 2018-09-06. https://www.safe.com/blog/2018/09/fme2018-tolerance-evangelist179.

[8] Mark Ireland. Upserts and Dog-Walking: What's New with Change Detection in FME 2019[EB/OL]. 2019-01-31.

https://www.safe.com/blog/2019/01/change-detection-2019-fmeevangelist184.

[9] Rylan Maschak, Mark Stoakes. Performance Tuning FME[EB/OL]. 2018-08-29. https://knowledge.safe.com/content/kbentry/579/performance-tuning-fme.html.