

# 面向 SAR 数据处理流程的溯源方法研究

杜 莹<sup>1,2,3</sup> 林冰仙<sup>1,3</sup> 周良辰<sup>1,3</sup> 闫国年<sup>1,3</sup>

1 南京师范大学虚拟地理环境教育部重点实验室,江苏 南京,210023  
2 河南大学计算机与信息工程学院,河南 开封,475001  
3 江苏省地理信息资源开发与利用协同创新中心,江苏 南京,210023

**摘 要:**为了方便查询合成孔径雷达(synthetic aperture radar,SAR)数据产品的来源,评估其质量和可靠性,再现其产生过程,使用溯源技术记录其衍生过程并管理其处理流程。分析 SAR 数据处理流程的特点,提出了基于分层二部图的溯源模型(hierarchical bigraph provenance model,HBPM),设计了基于 HBPM 的溯源仓库及检索语言(provenance data query language,PDQL),并实现了流程再现、溯源信息可视化等应用。在江苏省地面沉降监测系统中,通过一个 SAR 数据处理流程溯源的案例——用基于 SAR 技术的短基线(small base-lines subset,SBAS)方式检测地面沉降情况,验证该溯源方法的可行性。

**关键词:**溯源模型;数据溯源;流程溯源;SAR 数据处理;空间信息处理

**中图法分类号:**P208 **文献标志码:**A

合成孔径雷达(synthetic aperture radar,SAR)应用广泛,不受气候和光照条件限制。通过 SAR 技术获取的原始影像需要经过一系列加工处理,才能得到可使用的产品数据。这些处理包括数据输入、多视、地理编码、配准、正射纠正、滤波以及分类等,相互关联又有一定次序关系,形成了特定的 SAR 数据处理流程。为保证产品数据的来源、质量和可靠性,通常采用溯源技术记录、管理和追踪 SAR 数据在处理流程中的衍生过程。

溯源是由“provenance”翻译而来,有追踪起源、重现历史的含义,不同领域有不同的提法<sup>[1-3]</sup>,最早起源于 GIS 领域。文献[4]定义了 GIS 中数据的世系是有关产生这个数据项的原始素材和演化过程的信息。目前,溯源技术在数据库和工作流领域研究最多<sup>[5-6]</sup>,在地理空间的研究也比较多<sup>[7-11]</sup>,大多采用 ISO19115 Lineage 模型、开放溯源模型(open provenance model,OPM)以及 W3C PROV 模型<sup>[5,12]</sup>。三者均可用于 SAR 数据处理流程溯源,但都存在缺陷。SAR 数据处理流程中处理计算量大,计算时间长,参数数目和类型多,根据操作环境不同具有较强的可变性,需要扩展现有模型才能准确表达溯源信息。ISO19115 Lineage 模型是针对地理信息领域的

数据溯源模

型,提供了一些简单的元素,用于描述空间数据生产过程中所涉及到的过程步骤、输入输出数据以及控制者等信息,需要映射到 PROV 模型才能扩展。OPM 模型表达范围比较窄,支持扩展能力不强,扩展后术语有重叠现象影响溯源信息表达。例如,同时要表达产品数据溯源信息以及流程本身溯源信息时,术语“Process”既表示执行又代表处理模板<sup>[13]</sup>。PROV 模型有较强的扩展能力,标准化及通用性方面也很强,但需要增加数量巨大的扩展节点才能做到溯源信息的准确表达<sup>[13]</sup>。

为了清楚记录溯源信息,常常需要对溯源数据进行多粒度划分。目前这方面的研究较少,文献[11]提出粗粒度和细粒度结合的溯源方法,但该方法对使用环境及操作人员要求较高。文献[14]针对不同粒度的地理数据溯源进行了探讨,为本文溯源粒度的划分提供了参考。

溯源是基于实体间不对称的关系,天生具备基于有向图的特性<sup>[9]</sup>,包括 OPM 和 PROV 在内的一些溯源模型都是基于有向图的。SAR 数据处理流程的特征符合特殊有向图——二部图的特点,能清晰地刻画数据处理流程溯源。本文针对 SAR 数据处理流程参数多、可变性强的特点,在二部图的基础上提炼出更合适的分层二部图,并

提出了基于分层二部图的溯源模型(hierarchical bigraph provenance model, HBPM),设计了基于 HBPM 的溯源数据仓库及检索语言(provenance data query language, PDQL),并实现了流程再现、溯源信息多粒度可视化应用。在实际案例中,实现了溯源中间件植入 SAR 数据处理平台,并在江苏省地面沉降监测系统中运行。

## 1 SAR 数据处理流程溯源模型

SAR 处理流程是以 SAR 数据为处理对象的一系列活动的偏序集合<sup>[15]</sup>。流程的溯源信息需要记录实体对象、实体间关系以及包括元数据在内的属性信息。SAR 数据处理流程溯源实体包括 SAR 流程、SAR 数据以及 SAR 处理。SAR 数据处理流程与二部图特性天然统一。二部图又称为二分图,顶点集分割为两个互不相交的子集,并且图中每条边依附的两个顶点都分属于这两个互不相交的子集。SAR 处理节点和 SAR 数据节点是互不相交的有限集合,二者的并集是所有顶点的集合,边的两端分别属于不同的节点集合。用二部图对 SAR 数据处理流程溯源的定义如下。

定义 1 SAR 数据处理流程溯源信息  $G_s = (V_A, V_D, E)$  是一个三元组,其中:

- 1)  $V_A$  表示 SAR 数据的集合,  $V_D$  中表示 SAR 处理的集合,  $E$  是连接的集合;
- 2)  $V_A \cap V_D = \emptyset$ , 且  $V_A \cup V_D = V$ ,  $V$  是所有顶点的有穷集合。  $E$  是边的有穷集合,  $E = \{(v_i, v_j) | v_i \in V_A, v_j \in V_D\}$ 。

但是,用二部图来表达 SAR 数据处理流程的溯源信息存在明显不足:SAR 数据处理中参数数目和类型都很多,而且操作环境不同还会发生较大变化。二部图没法表达环境参数,也不能直接表达处理之间或是数据之间的联系。因此,本文提出了分层二部图来解决这些问题。

将二部图分成若干个子图,每个子图都是以处理为中心、详细表达流程中的一步,包括输入输出数据、处理及相应参数以及其他元数据信息。多个分层子二部图嵌套在一起,顺利描述整个流程溯源信息。每个子图以一个处理为核心,图 1 中虚线框表示各个子图不同的运行环境。分层二部图定义如下:

定义 2 分层二部图  $G = \{G_1, G_2, \dots, G_n\}$ , 其中  $G_i = \{V_{i,1} \cup V_{i,2}, E_i\}$ ,  $n$  是子图的个数,  $1 \leq i \leq n$ 。

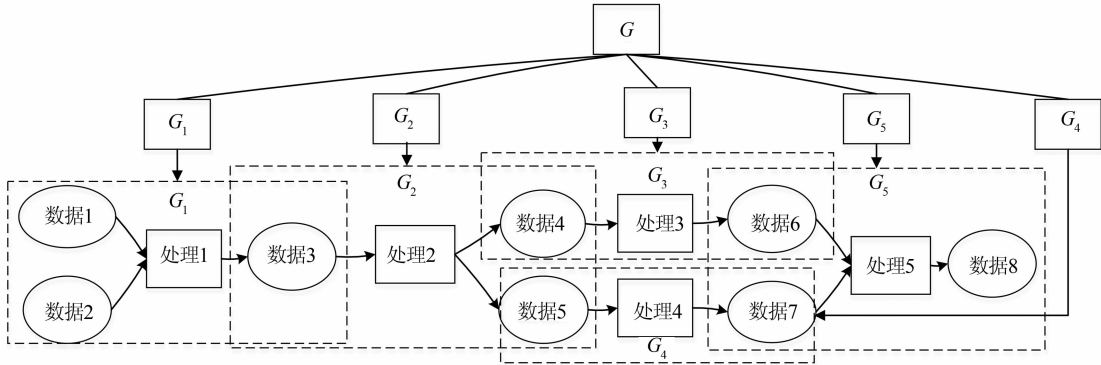


图 1 分层二部图  
Fig. 1 Hierarchical BiGraph

图 1 中子图  $G_1 = \{V_{1,1} \cup V_{1,2}, E_1\}$ ,  $V_{1,1} = \{\text{数据 1, 数据 2, 数据 3}\}$ ,  $V_{1,2} = \{\text{处理 1}\}$ ,  $E_1 = \{<\text{数据 1, 处理 1}>, <\text{数据 2, 处理 1}>, <\text{处理 1, 数据 3}>\}$ , 数据 3 节点既属于图  $G_1$ , 又属于  $G_2$ 。分层二部图由 3 部分组成:数据集(DataSet)、处理(Process)和关系(Relation)。本文借用 W3C PROV 模型中实体(Entity)与活动(Activity)之间的关系 Used 和 WasGeneratedBy, 细化并扩展得到关系 Relation 集合的 4 个基本元素:输入(Input)、被生产(beProduced)、输出(Output)、使用(Use)。其中,Input 和 Output 记录 SAR 数据处理正向流程,beProduced 和 Use 则是记录逆向

流程,方便正反两个方向溯源。用集合  $R_B$  来表示 4 种基本关系,记为  $R_B = \{\text{Input, Output, beProduced, Use}\}$ 。在此基础上,借助 PROV 模型中实体与实体之间的 WasDerivedFrom 关系以及活动与活动之间的 WasInformedBy 关系,得到扩充关系  $R_E = \{\text{Source, Result, HigherP, LowerP}\}$ 。Relation 集合各元素的推导规则为:

$$\begin{aligned} &\exists (D_i \xrightarrow{\text{Input}} P) \wedge (P \xrightarrow{\text{Output}} D_j) \Rightarrow D_i \xrightarrow{\text{Source}} D_j; \\ &\exists (D_i \xrightarrow{\text{beProduced}} P) \wedge (P \xrightarrow{\text{Use}} D_j) \Rightarrow D_i \xrightarrow{\text{Relation}} D_j; \\ &\exists (P_m \xrightarrow{\text{Output}} D) \wedge (D \xrightarrow{\text{Input}} P_n) \Rightarrow P_m \xrightarrow{\text{Higher}} P_n; \end{aligned}$$

$$\exists (P_m \xrightarrow{\text{use}} D) \wedge (D \xrightarrow{\text{beProduced}} P_n) \Rightarrow P_m \xrightarrow{\text{LowerP}} P_n;$$

$R_E$  关系具有传递性:

$$\exists (D_1 \xrightarrow{\text{Source}} D_2) \wedge (D_2 \xrightarrow{\text{Source}} D_3) \wedge \cdots \wedge (D_{n-1} \xrightarrow{\text{Source}} D_n) \Rightarrow D_1 \xrightarrow{\text{Source}} D_n;$$

$$\exists (D_n \xrightarrow{\text{Relation}} D_{n-1}) \wedge (D_{n-1} \xrightarrow{\text{Relation}} D_{n-2}) \wedge \cdots \wedge (D_2 \xrightarrow{\text{Relation}} D_1) \Rightarrow D_n \xrightarrow{\text{Relation}} D_1;$$

$$\exists (P_1 \xrightarrow{\text{Lower}} P_2) \wedge (P_2 \xrightarrow{\text{Lower}} P_3) \wedge \cdots \wedge (P_{n-1} \xrightarrow{\text{Lower}} P_n) \Rightarrow P_1 \xrightarrow{\text{Lower}} P_n;$$

$$\exists (P_n \xrightarrow{\text{Higher}} P_{n-1}) \wedge (P_{n-1} \xrightarrow{\text{Higher}} P_{n-2}) \wedge \cdots \wedge (P_2 \xrightarrow{\text{Higher}} P_1) \Rightarrow P_n \xrightarrow{\text{Higher}} P_1.$$

HBPM 溯源模型主要由 SAR 数据、SAR 处理以及它们之间的关系构成,还包括处理和数据集各自的属性,如图 2 所示。信息内容十分丰富,元数据以及数据处理的大量参数信息就包含在其中,还需进行扩展。

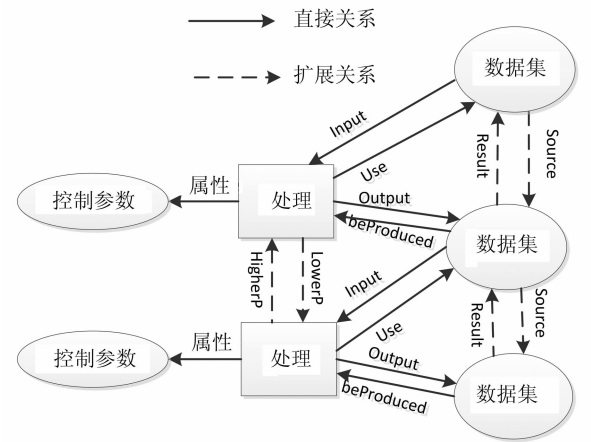


图 2 基于分层二部图的溯源模型

Fig. 1 Provenance Model Based on Hierarchical BiGraph

本文将溯源对象的粒度划分为流程级、数据级、处理级、关系级以及属性级 5 个层次。流程级针对整个流程,是粗粒度溯源;数据级、处理级和关系级分别针对数据、处理及二者间关系,是中粒度溯源;属性级是针对溯源对象的属性,是细粒度溯源。

## 2 基于 HBPM 的溯源方法

### 2.1 溯源数据收集与存储

溯源收集需要记录数据处理流程中流程步骤、执行环境、控制参数以及各对象属性等细节信

息,是较复杂的问题。目前已有许多不同的溯源收集机制,文献[6,16]中将溯源数据收集机制分成操作系统级、工作流级和处理级 3 个级别。操作系统级机制独立于流程管理系统,能追溯细粒度的溯源数据。工作流级机制通过获取工作流的规格说明书,派生流程中未发生的处理,受制于工作流管理系统,通用性差。处理级机制易于实现,收集细粒度的溯源数据,独立于流程管理系统,但只能收集发生过的流程执行和数据溯源信息<sup>[17]</sup>。

HBPM 模型中每个分层二部图的子图都是以一个处理节点为核心,与处理级收集机制相吻合。本文设计了基于处理级的溯源数据收集器,自动监听 SAR 数据的处理,记录输入输出数据、外界环境变量、控制参数及状态信息等较细粒度的溯源数据,并以 HBPM 支持的存储模式存入溯源数据仓库。

常用的溯源信息存储模式有普通文本方式、可扩展标记语言(extensible markup language, XML)方式、资源描述框架(resource description framework, RDF)以及关系模式。XML 或 RDF 等文本方式存储内容的冗余度高,检索效率低。关系数据库技术比较成熟,能有效克服上述缺点,适合溯源信息多粒度、多层次的存储与检索。本文以 HBPM 溯源模型为基础,设计了基于关系模型的溯源数据仓库,溯源仓库的关系数据模型如图 3 所示。图 3 中, ID 为身份标识(identity), URL 表示统一资源定位符(uniform resource locator)。

### 2.2 溯源数据检索

在关系数据模型的溯源仓库中,采用结构化查询语言(structured query language, SQL)检索溯源数据有一定缺点。首先,溯源数据仓库不同主题的溯源信息之间、主题与实际溯源信息表之间的关系无法用 SQL 语言直观简洁地表达;其次,SQL 面向数据集不支持步骤化,溯源关系库中还存储参与流程的处理节点和数据节点的元数据地址,需要嵌套很深的 SQL 语句才得到所需溯源数据。本文在 SQL 语言的基础上形成适合溯源数据检索的溯源查询语言 PDQL,能够多粒度、分层次地从溯源仓库检索溯源数据。主要修改内容如下。

1) 增加主题的定义和引用语句。

<主题定义> ::=

新建主题 < 主题名称>

表 <表名>{,<表名>}

主题属性 <属性名> ,

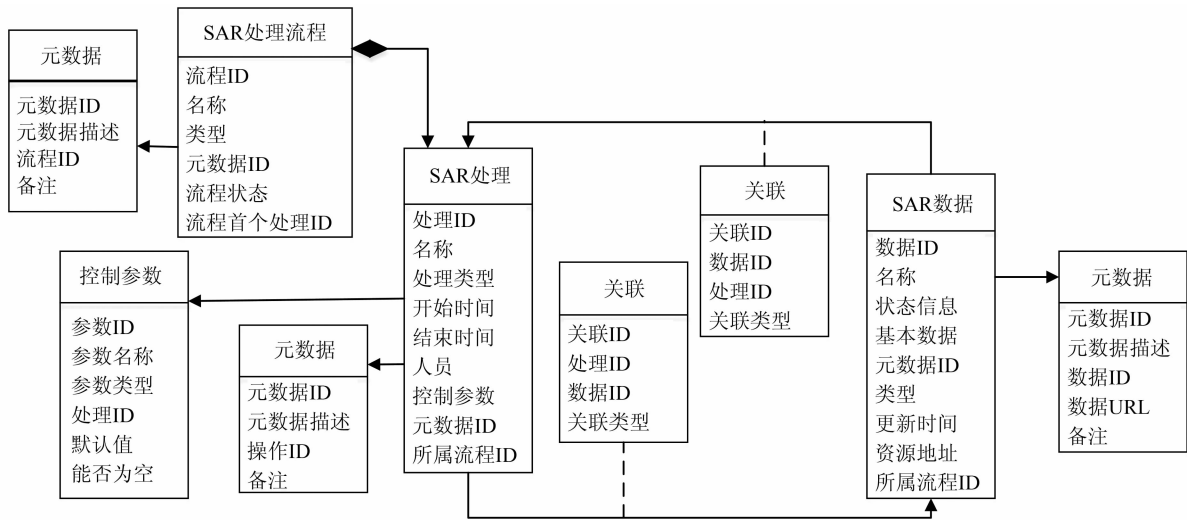


图 3 溯源仓库的关系数据模型

Fig. 3 Relational Data Model of Provenance Repository

描述属性 <属性名>

主题是溯源数据仓库非常重要的组成实体，数据仓库中所有溯源信息表对象都包含在某个主题中。该定义说明了主题由哪些表对象组成以及主题的两个重要属性：主题属性和描述属性。

2) 增加主题聚集表的指定语句。检索溯源数据时，需要返回某个主题包含所有表的聚集结果。增加 TABLES 语句，列出用户希望返回结果所在的主题中的表  $T.t$ ，其中  $T$  是主题， $t$  是主题中的表名，在新建主题 CREATE THEME 语句中定义。

3) PDQL 避免实际溯源信息表中复杂的关系以及关联查询，设计简洁的查询语句<sup>[18]</sup>。

<查询语句> ::=

查询(select) 表字段 从(from) 表  
(主题 主题子句  
(满足 条件子句  
(group by 分组子句  
(order by 排序子句))))

3 案例分析

地面沉降是一种缓慢变化的地质灾害<sup>[19]</sup>，地面沉降监测是防治质灾害的发生或重新利用土地资源的重要工作，是典型的 SAR 数据处理流程。江苏省地面沉降监测系统，利用基于 SAR 技术的短基线(small baselines subset, SBAS)干涉测量技术<sup>[20]</sup>对江苏省大面积地面缓慢地表形变进行高精度的监测。输入原始数据集—— $N+1$  副单视复数图像(single look complex image, SLC)的

SAR 图像，经过连接图生成、差分干涉生成、3D 解缠、相位编辑及轨道精炼等一系列处理，获取某地区的形变时间序列和平均沉降速率。以 SBAS 的方式监测某地区地面沉降处理流程呈现的分层二部图，如图 4 所示，包含有 7 个处理子图  $G_1-G_7$ 。每个子图围绕一个处理节点，输入数据集节点和输出数据集节点与之连接，处理节点和数据节点包含有各种属性信息，控制参数信息就包含其中。图 4 中，KML 是锁眼标记语言(keyhole markup language)，shp 是 shape 文件。

SAR 数据处理流程溯源中间件包括溯源数据的收集、存储、检索与应用功能。溯源数据收集器监听流程运行，收集溯源信息，存入溯源仓库；溯源数据处理分析器从溯源仓库中检索溯源信息，进行可视化、错误验证以及流程再现等应用，如图 5 所示。

下文用例子分别说明不同粒度的溯源信息检索方式。

示例 1 流程级溯源检索，查询主题名为“Nanjing\_SBAS”下某过程的运行时间。

关系表达式： $\Pi$  RunTime ( $\sigma$  FlowID = 'XXXX'

$\wedge$  theme = 'nanjing \_ sbas' ( SAR-FlowTable))

PDQL 伪代码：

查询 RunTime 字段 从 SARFlowTable 表

主题 nanjing \_ sbas 满足 FlowID = 'XXXX' 条件

示例2 属性级溯源检索，查询某一处理的控

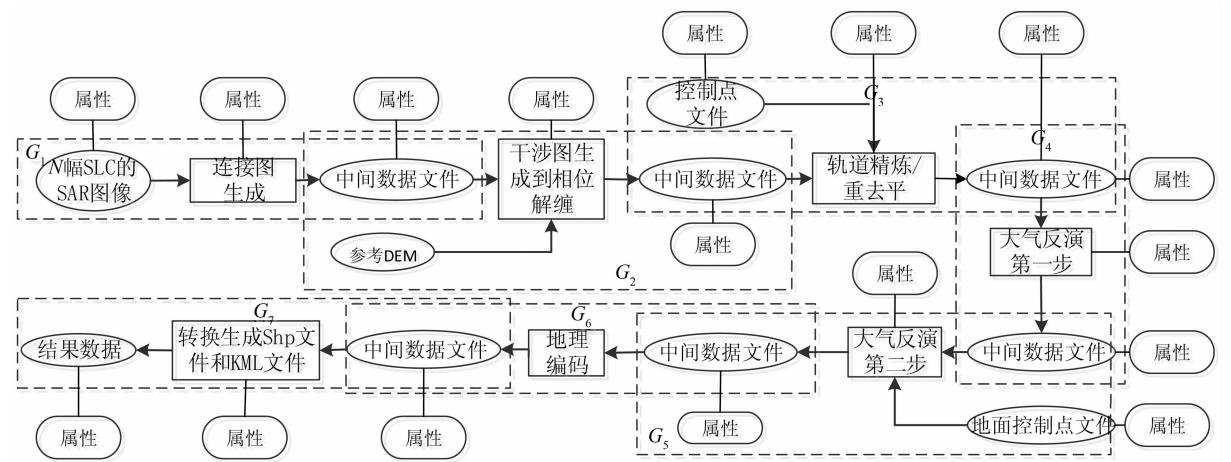


图 4 分层二部图呈现 SAR 流程

Fig. 4 SAR Data Processing Flow Presented by the Hierarchical Bigraph

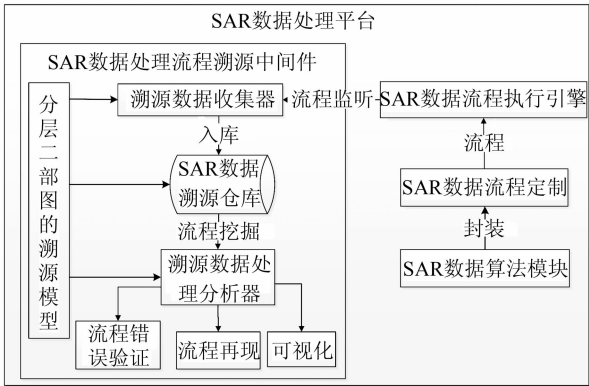


图 5 SAR 数据处理平台中溯源中间件

Fig. 5 Provenance Middleware in the Platform of SAR Data Processing

制参数信息。

关系表达式:  $\Pi$  ControlParaID, ParaName, ParaType, ParaValue ( ControlParaTable, ControlParaID( $\sigma$  ProcessName= 'XXXX'  $\wedge$  theme = ' nanjing \_ sbas ' ( SAR-FlowTable)))

PDQL 伪代码:

查询 所有字段 从 ControlParaTable 表  
主题 nanjing\_sbass

满足 SARProcessTable. ProcessName = 'XXXX' 条件

按 ControlParaTable . ControlParaID 排序

基于 HBPM 的溯源中间件实时收集产品的溯源信息,主要产生数据溯源和流程再现两个方面的溯源应用。数据溯源功能可以根据数据产品名称,将参与产生该产品的所有原始数据检索出,形成原始数据列表,可以点击查看任意一个数据的溯源信息,如图 6 所示;也可以生成由所有数据节

点组成的一个流程图,如图 7 所示,单击任意一个数据节点,弹出该节点溯源信息。流程再现功使用溯源数据重新生成已经执行过的流程,单击流程中任意节点,便可标注出该节点的溯源信息,如图 8 所示。

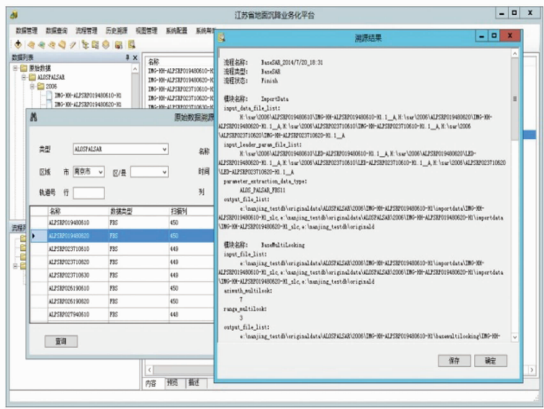


图 6 原始数据溯源

Fig. 6 Provenance about the Source Data

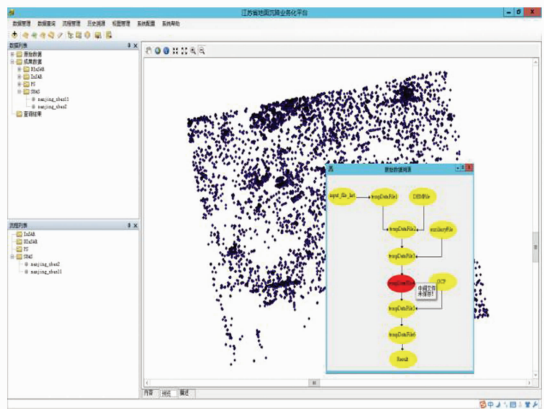


图 7 数据溯源流程图

Fig. 7 Flowchart of the Data Provenance

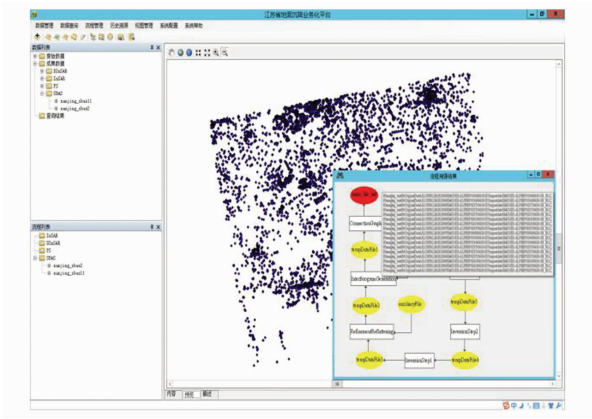


图 8 流程再现

Fig. 8 Reproduction Process

4 结 语

本文提出了基于 HBPM 模型的 SAR 数据溯源方法,在此基础上实现了溯源中间件,并应用于江苏省地面沉降监测系统中。该方法不仅适合 SAR 数据处理流程溯源,还适合以复杂高密度数据处理为主,且流程本身比较简单,处理参数复杂的流程溯源。

但 HBPM 模型的通用性不如 W3C PROV 模型,如果考虑互操作性,应进一步结合 PROV 模型进行应用扩展。本文中溯源信息的应用完成了数据及流程的历史溯源,能重现整个流程,但效果还比较单一,下一步可以研究历史溯源信息的三维可视化,以及更深入的流程挖掘等相关内容。

参 考 文 献

[1] Khanna P B S, Tan W C. Why and Where: A Characterization of Data Provenance[C]. The 8th International Conference on Database Theory, London, 2001

[2] Day M. Provenance and Data-Intensive Science, 2005[EB/OL]. <http://www.ukoln.ac.uk/projects/ebank-uk/provenance/report/draft-v01.pdf>, 2007

[3] Pancerella C, Myers J, Rahn L, et al. Data Provenance in the CMCS[C]. Workshop on International Provenance and Annotation Workshop Series (IPAW), Chicago, Illinois, USA, 2002

[4] Lanter D P. Design of a Lineage-Based Meta-Data Base for GIS[J]. *Cartography and Geographic Information Systems*, 1991, 18: 255-261

[5] Ming Hua, Zhang Yong, Fu Xiaohui. Survey of Data Provenance[J]. *Journal of Chinese Computer Systems*, 2012, 33(9): 1 917-1 923(明华,张勇,符

小辉. 数据溯源技术综述[J]. 小型微型计算机系统, 2012, 33(9): 1 917-1 923)

[6] Simmhan Y, Plale B, Gannon D. A Survey of Data Provenance in E-science [J]. *SIGMOD Record*, 2005, 34: 31-36

[7] Wang S, Padmanabhan A, Myers D J, et al. Towards Provenance-Aware Geographic Information Systems [C]. The 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, California, USA, 2008

[8] Yue Peng, He Lianlian. Geospatial Data Provenance in Cyberinfrastructure[C]. The 17th International Conference on Geoinformatics, Fairfax, USA, 2009

[9] Yue Peng, Peng Feifei, Gong Jianya. Research on SOA-Based Geospatial Data Provenance [J]. *Geography and Geo-Information Science*, 2010, 26(3): 6-10(乐鹏,彭飞飞,龚健雅. 基于 SOA 的空间数据起源研究[J]. 地理与地理信息科学, 2010, 26(3): 6-10)

[10] Yue Peng, Gong Jianya, Di Liping. A Linked Data Approach for Geospatial Data Provenance[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2013, 51(11): 5 105-5 112

[11] Huq M R, Apers P M G, Wombacher A. An Inference-Based Framework to Manage Data Provenance in Geoscience Applications [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2013, 51(11): 5 113-5 130

[12] Moreau L, Missier P. PROV-DM: The PROV Data Model, W3C Working Draft [EB/OL]. <http://www.w3.org/TR/2012/CR-prov-dm-20121211/>, 2012

[13] Missier P, Dey S, Belhajjame K, et al. D-PROV: Extending the PROV Provenance Model with Workflow Structure [C]. Workshop on the Theory and Practice of Provenance, Illinois, USA, 2013

[14] He Lianlian, Yue Peng. 2014. Adding Geospatial Data Provenance into SDI: A Service-Oriented Approach[J]. *IEEE Journal of Applied Earth Observations & Remote Sensing*, 2015, 8(2): 926-936

[15] Gao Yong, Wu Lun, Liu Yu. Spatial Process Modeling [J]. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2005, 40(6): 914-921(高勇, 邬伦, 刘瑜. 空间信息处理过程建模研究[J]. 北京大学学报(自然科学版), 2005, 40(6): 914-921)

[16] Freire J, Koop D, Santos E, et al. Provenance for Computational Tasks: A Survey[J]. *Computing in Science & Engineering*, 2008, 10(3): 11-21

[17] Marinho A, Murta L, Werner C, et al. ProvManager: A Provenance Management System for Scientific Workflows[J]. *Concurrency and Computation: Practice and*

*Experience*, 2012, 24(13): 1 513-1 530

[18] Jr Gadelha L M R, Wilde M, Mattoso M, et al. MTCProv: A Practical Provenance Query Framework for Many-Task Scientific Computing[J]. *Distributed and Parallel Databases*, 2012, 30(5/6): 351-370

[19] Luo Sanming, Du Kaifu, Wan Wenni, et al. Ground Subsidence Rate Inversion of Large Temporal and Spatial Scales Based on Extended PSInSAR Method [J]. *Geomatics and Information Science of Wuhan University*, 2014, 39(9): 1 128-1 134 (罗三明, 杜凯夫, 万文妮, 等. 利用 PSInSAR 方法反演大时空尺度地表沉降速率[J]. *武汉大学学报·信息科学版*, 2014, 39(9):)

[20] Yang Chengsheng, Zhang Qin, Zhao Chaoying, et al. Small Baseline Subset InSAR Technology Used in Datong Basin Ground Subsidence, Fissure and Fault Zone Monitoring[J]. *Geomatics and Information Science of Wuhan University*, 2014, 39(8): 945-950 (杨成生, 张勤, 赵超英, 等. 短基线集 InSAR 技术用于大同盆地地面沉降、地裂缝及断裂活动监测[J]. *武汉大学学报·信息科学版*, 2014, 39(8):945-950)

Provenance Method for SAR Data Processing Flow

DU Ying<sup>1,2,3</sup> LIN Bingxian<sup>1,3</sup> ZHOU Liangchen<sup>1,3</sup> LV Guonian<sup>1,3</sup>

1 Key Laboratory of Virtual Geographical Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023, China

2 College of Computer and Information Engineering, Henan University, Kaifeng 475001, China

3 Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

**Abstract:** Provenance technology can query the sources of Synthetic Aperture Radar (SAR) data products, assess the quality and the reliability and reproduce the production process by recording and managing the derivation history of SAR data products. A bi-graph describes the relationship between data and process in the SAR data processing flow. Hence, a Hierarchical BiGraph Provenance Model (HB-PM) is proposed in this paper that combines the characteristics of the SAR data processing flow with a bi-graph. Based on the HBPM, the provenance repository and the retrieval language PDQL (Provenance Data Query Language,PDQL) were designed and applied. Process reconstruction and the provenance information visualization were also implemented. The feasibility of the proposed method was verified through a case study on provenance of SAR Data Processing Flow---using Small Baselines Subset ( SBAS) technology to monitor surface deformation in the surface subsidence montitoring system of Jiangsu Province.

**Key words:** provenance model; data provenance; flow provenance; SAR data processing; spatial process

**First author:** DU Ying, PhD candidate, lecturer, specializes in virtual geographic environments. E-mail: hd\_duying@163.com

**Corresponding author:** LIN Bingxian, PhD, lecturer. E-mail: lbx1984@hotmail.com

**Foundation support:** The National Natural Science Foundation of China, No. 41301415; the Graduate Students' Research and Innovation Projects of the Education Department of Jiangsu Province, No. KYLX\_0702; the Priority Academic Program Development of Jiangsu Higher Education Institutions.