

叙事型图像语义标注模型研究*

徐 雷 王晓光

摘 要 由于叙事型图像对时空语义信息表达的特殊性,目前叙事相关的语义模型并不适合直接用于该类型图像的语义标注。针对这一问题,本文设计了一个适用于叙事型图像的深度语义标注本体模型。该模型借助模块化本体设计思想,结合开放标注协同框架(OAC),以情节、实体、活动、情境为核心,解决了叙事型图像时空信息组织表达以及叙事模型和标注信息映射的问题。使用嵌入该标注模型的图像语义标注工具,对敦煌壁画进行语义标注实验,验证了该模型的可靠性和可用性。但是叙事型图像中情节的划分存在不同的见解,给情节提取和标注带来了不确定性,这是图像深度语义标注面临的挑战。图 8。表 1。参考文献 22。

关键词 叙事型图像 故事情节 语义描述 图像标注

分类号 G254

Narrative Image Annotation Ontology for Semantic Web

XU Lei & WANG Xiaoguang

ABSTRACT

Narrative image is one type of visual literature resources that intends to tell stories. In face of its obscure and abstract nature, existing narrative-relevant semantic models appear inadequate for its semantic annotation and description. To address the issue, this paper designs an ontology model for deep semantic annotation applicable for narrative images.

The model resorts to the design thought as in modular ontology, integrates with Open Annotation Collaboration (OAC) framework, and takes plot, entity, dynamics and context as its core. The model addresses issues such as the organization and representation of spatio-temporal information in narrative images, and the mapping between narrative model and annotation information.

Plot lies at the core of the framework. Each plot takes place in a specific context and involves some entities, while different plots are semantically related in terms of sequence or overlap. In a specific plot annotation area within an image, a dynamics is normally used to present dynamic features of the plot. The descriptive framework, as a whole, is an abstract semantic annotation model and accomplishes the semantic

* 本文是教育部人文社科重点研究基地重大项目“大数据资源语义表示与组织——面向文化遗产领域”(项目编号:16JJD870002)和教育部“新世纪优秀人才”基金项目“文献信息结构与语义组织方法研究”资助成果之一。(This article is an outcome of the project “Semantic Representation and Organization of Resource in Big Data Era—Towards Cultural Heritage Field” (No. 16JJD870002) supported by Foundation of Key Research Institute of Humanities and Social Science at Universities, Chinese Ministry of Education and the project “Semantic Publication Structures and Models” supported by Program for New Century Excellent Talents in University of Ministry of Education of China.)

通信作者:王晓光,Email:whu_wxg@126.com,ORCID:0000-0003-1284-7164(Correspondence should be addressed to WANG Xiaoguang,Email:whu_wxg@126.com,ORCID:0000-0003-1284-7164)

description and annotation by mapping annotated objects onto the abstract model. Plot constitutes the core of the annotation framework and corresponds to the Body of OAC in image annotation.

In detail, temporal semantic relations between plots contain nextplot, lastplot, overlaps, meets, starts, during, finishes, equal. In addition to the above relations, other obscure and abstract spatio-temporal semantic relations can also be identified in plots. This paper places TemporalSpatialEntity under the category of Entity. To meet the need for annotation of spatio-temporal information in narrative images, this paper divides TemporalSpatialEntity into Place and TemporalThing, which are respectively further divided into two types-Concrete (ConcretePlace, ConcreteTemporalThing) and Abstract (AbstractPlace, AbstractTemporalThing). Dynamics is another indispensable element in the model. As a core sub-category in the process of annotation, Action has two object properties, namely hasActor and hasRecipient; the former being the agent of the action while the latter the patient. At the same time, Dynamics can also have a type (DynamicsType). Context is a special item in the plot annotation of narrative images, and it provides the annotation model with additional extension ability. Instead of being directly accessible from image, information involved in context is hidden behind the image, yet helpful in understanding the image.

To represent the complex relations of overlap between different annotated areas within an image, this paper puts forward OAC extended vocabularies, including no-overlap, complete-overlap, inclusiveness, and partial-overlap. In the end, this paper performs a semantic annotation experiment on Dunhuang narrative murals, using an image semantic annotation tool embedded with the model, and the annotation data is converted and stored in the form of RDF triples. The experiment verifies the model's reliability and applicability. 8 figs. 1 tab. 22 refs.

KEY WORDS

Narrative image. Story plot. Semantic description. Image annotation.

0 引言

图像是一种传递信息、知识和思想的视觉媒介,能够表示文本难以表达的复杂信息。图像的具体表现形式十分多样,载体也数不胜数。常见的图像包括照片、绘画、图表、草图等,它们包含的丰富信息对于人眼较为容易辨识,但对于机器却相当难于理解和查询。目前,基于元数据的图像搜索和浏览较为常见和成熟,但针对图像内部特定知识的搜索和浏览还远不成熟,特别是文化遗产领域的图像,如壁画、油画、织锦画,因为其包含深刻的艺术内涵和复杂的时空场景,图像的理解和获取都还面临很多困难,严重制约了其传播、理解和学术研究。

近年来,随着文化遗产数字化的发展,数字

文化遗产资源激增,数字图像资源的管理与开发需求也日益突出。图像的语义描述与标注是管理和开发数字图像资源的基础。如何突破传统的元数据描述的限制和不足,揭示和描述数字图像内的丰富语义知识,是解决海量图像资源管理的关键。传统的视觉资源元数据和表示主题的标签,难于揭示文化遗产图像内丰富的语义内涵,比如人物和器物的具体名称、相对位置、故事情节、场景情感等等。借助已有的语义技术和本体思想构建新的图像语义描述框架,实现图像视觉知识的文本化和数据化转换势在必行。

叙事型图像(Narrative Image)是一类以诉说故事为目的的视觉文献资源,包括单场景图像和多场景图像两类。作为人类最早使用的媒介,图像比文字更适合向普通大众传播特定的

精神和信念知识,在宗教传播和艺术表现领域十分常见,是重要的文化遗产和传播载体,可以说是一种特殊的“视觉语言”。本文针对叙事型数字图像资源设计了一个用于图像标注的本体模型(Narrative Image Annotation Ontology, NIAO),对图像的叙事性知识进行语义表示,并以敦煌壁画中的叙事型图像为例,借助已开发的图像标注工具,验证了该标注模型的适用性。

1 相关研究

1.1 研究基础及进展

数字图像的描述与标注是当前的热门研究主题,尤其是在计算机视觉研究领域,图像的自动识别研究及应用已经取得了巨大进展,这主要得益于图像大数据的建设、深度学习技术的发展以及计算机硬件性能的提升。知名的计算机图像识别研究项目包括美国斯坦福大学的ImageNet^[1]项目、微软公司的计算视觉API^[2]项目等。借助计算机图像识别技术,对图像进行自动的描述与标注逐渐由简单的拍摄时间、地点、体积大小、上传时间等基本的图像元数据标注,扩展到诸如图像内的人物、物体、情感等更为精细的语义信息标注。

文化遗产领域的数字图像是艺术作品和文化遗产数字化加工生成的一类文献资源,主要来源包括油画、壁画、瓷器和陶器上的绘画、织锦上的图案等等。相对于现实生活中的数字摄影而言,文化遗产数字图像资源分散,而且普遍缺乏高质量的标注信息,标准数据集建设滞后,这给机器学习带来了极大困难。同时,由于该类图像的特殊性,采用传统元数据的方式进行标引,会丢失大量的语义信息,造成严重的漏标,甚至错标。叙事型图像的标注问题尤其严重,少数几个图像标签(Tag)或关键词虽可在一定程度上描述图像的主题>Aboutness),但对于图像内容关涉物(Ofness)的描述却是远远不足。如何理解叙事型图像的故事性内涵,并对其进行规范的和形式化的标注与表示,是理解图像

潜在语义、丰富图像检索方式和提升图像资源发现能力的关键,同时也是数字信息资源管理领域的新兴前沿课题。

叙事,简单来说就是描述一个故事,可以借助口述、文字、图像、动画、声音等一种方式或多种方式的混合来实现。传统的图像叙事研究主要集中在图像学和艺术史领域。近年来,随着图像检索的发展需要,信息科学领域的图像叙事研究日渐增多。2016年,微软研究院发布了一个面向图像叙事的数据集VIST^[3],其目的是为了研究视觉叙事(Visual Storytelling)过程,以实现人工智能理解视觉资源内的时间结构与主题表达过程;Tuffield等给出了童话故事领域的相关叙事性词汇^[4]。叙事的形式化研究也已经出现,如故事本体Stories Ontology^[5],它是一个RDF格式的数据模型,核心类属性结构包括Story、EventSlot、slot、sub_story等,但其未明确提供故事参与的主客体接口,直接用于图像标注的数据建模将缺乏图像对象识别这一核心需求;Nakasone、Ishizuka^[6]给出了一个OWL格式的叙事本体(Storytelling Ontology),其将故事视为一系列连续的事件,虽然定义了故事参与的主体Agent及角色Role,但采用面向文本组织的修辞结构理论^[7](Rhetorical Structure Theory),引入Act使得该本体模型并不适合描述叙事型图像;Ciotti^[8]提出了一个OWL格式的叙事本体,包含了Actant、Action、Actor、Event、Quality(用于表征Actor的特性)、Object等核心要素,但它并不能够很好地表述叙事型图像连续性特征;Damiano^[9]提出了一个顶层叙事与动作本体(Narrative和Action),其结构中的Dynamics类用于表征action、processes和state,Entity用于表征character和object,Story用于表征各种stories,Description Templates用于表述role模式,同时还有Geographical Place和Temporal Collocation空间和类型,这项工作十分出色,基本包含了叙事对象的所有信息,但同样没有解决叙事型图像中stories之间的关联语义的表述问题,以及与标注图像映射的问题;ABC Ontology^[10]与

之类似,也定义了 Event、Action、Temporality、Actuality 等核心要素,并区分了 existential Actuality 和 universal Actuality,但同样具有 Rossana Damiano 的模型的局限;BBC 的 Storyline Ontology^[11]用于组织新闻事件,包含了 Storyline、StorylineSlot、Event、Topic 等要素,并通过 follows 来表示 StorylineSlot 之间的时序关系,但在抽象的艺术创作的叙事型图像中,不仅包含这种基本的关系类型,而且该模型面向的是新闻领域。阿姆斯特丹自由大学开发的 Simple Event Model (SEM)^[12]包含了 Event、Actor、Place、Time 以及 Type 和 Role 等核心要素,并通过 hasSubEvent 属性来表示 Event 之间的整体部分关系,在描述叙事型图像方面和 Storyline Ontology 具有同样的局限,另外叙事型图像在绘制的过程中一般并不重视故事发生的具体地理位置和时间信息,这与大多数的事件模型较看重事件的时空信息有所不同;Activity Ontology^[13]和 SEM 类似,不过并不具有角色 (Role) 信息;Krisnathi^[14]使用模块化的开发方法设计的事件模型更简洁,给出了事件要素的形式定义;CIDOC CRM^[15]也是类似的框架,提供了对时间和地理位置更为详尽的建模方案;Event-model F^[16]也提供了对事件、地点、时间、参与者等类型的建模能力,并采取了模块化开发的方法来增加其可扩展性;LODE^[17]提供了一个现有事件模型之间的交互功能,其目标不是建立一个事件模型,而是辅助各种事件模型数据之间的信息访问存取。

现有的叙事模型在各自领域和应用中具有各自的优势和特点,但直接应用在叙事型图像的标注与描述上却存在明显的不足,主要是叙事型图像中的时空概念较为抽象,尤其是艺术图像中,时间与空间的表达是模糊的,由此导致现有的叙事模型难于表达叙事型图像情节语义的时空属性。另一局限主要是现有模型是纯粹的叙事模型,面向事件或叙事本身,而不像叙事型图像这样利用图像的形式来间接叙事,由于没有面向图像标注的接口,这些叙事模型不能直接应用在图像语义标注过程中。因此,有必

要设计融合时空信息的情节语义描述方法,来规范表达叙事型图像的叙事性内涵。

1.2 叙事型图像深度语义标注需求

图像深度语义标注 (Deep Semantic Annotation)^[18]的目标是深入到图像内部 (Ofness),侧重于表达图像内容中具体的对象及其语义关系。深度语义标注不仅仅要标注出图像中的事物属类,如人、动物、车辆,还要标注出它们具体的名称,如释迦摩尼、九色鹿、步辇。此外,对于叙事型图像,深度语义标注还应该标注清楚事件的时空信息、情节信息及其复杂的语义关系。

叙事型图像深度语义标注不应该利用一段式文字或者标签列表进行图像标注,而应该将语义信息与图像的内容片段及其语义进行精准的对应,同时将文字型信息结构与图像自身表达的叙事内涵结构相对应。比如图像中包含一个故事的多个情节,每个情节的参与人物不同,那么图像的深度语义描述信息就应该规范地说明清楚每个人物属于哪个情节。这一要求,以往的图像标注模型都难以满足。

图像深度语义标注数据的应用空间十分广泛,可预见的应用包括图像检索、图像识别、图像自动说明 (Image Caption)、图像合成等。图像检索指利用图像的元数据和标注信息检索图像,以及按照样本图像检索相似的图像;图像识别属于计算机视觉范畴,识别对象包括图像中的人脸、物体、文字符号、个体情感、人物年龄等;图像自动说明指对图像的内容使用自然语言句子进行描述;图像合成是根据用户的语义描述自动由多张图像合成一张符合需要的图像。叙事型图像的语义标注可以应用于叙事型图像的检索,也可以作为训练数据集辅助图像自动语义标注、图像识别、图像自动说明和图像合成。

2 叙事型图像标注模型构建

叙事型图像主要利用图像而非文字进行叙

事,既包括一张图像描述故事一个情节的情况,也包括一张图像描述故事多个情节的情况。图1是位于莫高窟257号洞窟的一幅壁画的数字图像,其内容描绘的是一个关于九色鹿的本生故事画,它是一个典型的叙事型图像。该图画中讲述的故事可通过如下文字来简单描述:“九

色鹿在河边救起一位落水人调达之后,被这位落水人向国王告密其行踪,国王下令追捕九色鹿,最后九色鹿和国王对质,说明了事情的原委,国王责备了落水人,放走了九色鹿,而落水人则满身生疮,得到了报应”。这幅壁画显然描述了一个故事的多个情节,跨越了多个时空。



图1 九色鹿本生故事画

该数字图像是摄影师通过拍照设备对洞窟中的壁画进行拍摄而得的人造物(Man-made Thing),它的类型是数字图像(Digital Image),更细的类型划分是叙事型数字图像(Narrative Digital Image),具有图像大小、分辨率、拍摄者、拍摄时间、拍摄地点等基本的元数据信息。数字图像是对洞窟中的壁画实物对象(Man-made object)的反映,这一实物对象的类型是壁画(Mural),该实物壁画的描述信息可以通过洞窟位置、长度、高度、绘制时间、绘制者、绘制技法等维度来刻画。如果该实物对象可以移动,并且收藏在博物馆内,描述信息还可以包含收藏机构名称、入馆时间等信息。

数字图像作为一种视觉文献资源除了具备自身的元数据以外,还记录和反映了壁画本身的内容信息,包括壁画的主题、故事类型,包含的人物、动物、环境、动作、情节等一系列抽象的概念对象(Conceptual Object),这些信息是图像内容分析和图像学研究的对象,是数字图像作为一种数字文献存在所承载的价值核心。

图1这幅本生故事画描述的是神话动物、落水人、国王等对象之间的故事,这些概念隶属于特定的知识系统,传达了特定的佛教寓意和文化内涵,是佛教文化传播与研究领域的常见主题。数字图像与其表达对象之间的这种关系可以通过图2来表示。

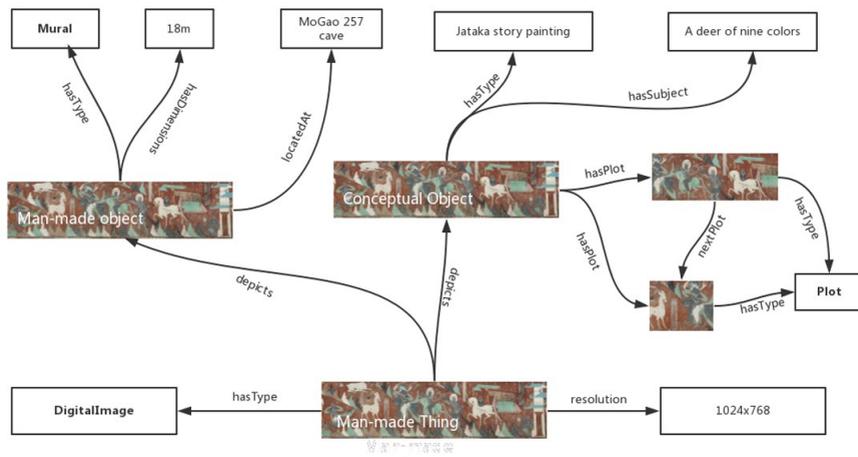


图2 数字图像的顶层本体框架

2.1 叙事型图像内容的构成要素及情节定义

本文研究的图像语义描述与标注对象是类似于图1的叙事型图像,核心是对图像所表达的概念对象(Conceptual Object)的描述与标注。Xu和Wang^[19]提出了数字图像内容语义描述的主要构成,包括对象(Object)、空间关系(Object space)、场景(Scene)、行为动作(Behaviors/Activities)、情感(Sentiment)等语义信息。对于叙事型图像而言,其内容除了包含以上语义信息外,还包含情节(Plot)信息。通过对叙事型图像中情节语义进行数据建模,可以实现该类图像的语义标注与检索应用。

在图2框架中,情节可看作一种概念对象(Conceptual Object)集合的语义单元,情节之间有时序的先后关系。在叙事型图像中,将情节定义为叙事图像反映的故事或事件的一个子变化过程,且各个子变化过程通过有机的时空联系组合成叙事图像表达的故事。这里的情节,类似于话剧中的一幕(Act)或电视剧中的一集(Episode),主要的区别在于图像中的情节语义粒度要更小、更简单。毕竟一个图像中的情节对应着一个指定的图像区域,不像一幕话剧或一集电视剧这种粗粒度的内容流。该定义指出情节是一个变化的过程,因此情节具有特定的时空情境,并通过时空信息来联系其他情节。广泛意义上讲,没有时空关联的叙事可看作是单情节叙事。情节在叙事型图像中的对应部分通常表现为一个动作、行为等动态的元素,如奔跑、歌唱等动作,以及自然现象中的水流动、下雨、地震等外界触发性元素,这给叙事型图像的情节标注提供了操作指示和标注要素来源。

2.2 叙事型图像标注模型构成

本文的主要工作是通过叙事型图像的情节建模,来辅助该类型图像的情节语义标注。因此需要在图像标注和图像情节模型之间搭建起连接的桥梁。这个桥梁就是OAC(Open Annotation Collaboration)^[20],这是一个关于资源标注的元数据协议,它方便了标注数据之间的

互操作,进而提升了资源用户之间交流的便利性。在OAC中,被标注的资源是一个Annotation类型,具体的Annotation连接一个Body和一个Target对象,它们是实际的标注过程中使用的词汇。Target侧重于被标注的资源的物理特征的描述,如图像资源被标注区域的大小、在整个图像中的位置关系等,Body侧重于被标注内容信息的描述,可以和图2中的概念对象对接,因此叙事型图像的情节标注主要对Body进行了扩展。整个叙事型图像情节标注模型如图3所示,该框架将情节(Plot)作为Body的一种类型(type),和Plot相连的是几个核心概念。

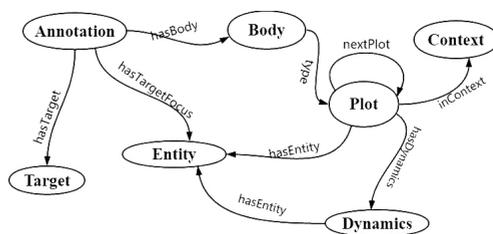


图3 以情节为核心的叙事型图像语义标注框架

该框架中以情节为核心,情节之间具有时序先后及重叠的语义关系,情节发生在特定的情境(Context)下,并涉及到一些实体(Entity),如情节参与的人、物体等。图像中的一个情节标注区域,通常通过一个动态元素(Dynamics)来呈现情节的动态发展特征,如参与者的一个动作(Action),该动态元素是表征实体(Entity)之间关系的关键。整个描述框架是一个抽象的语义描述模型,通过和OAC标注框架结合,将具体的标注对象和该抽象模型融合映射,完成对叙事型图像的语义描述标注。下面对该框架的核心概念进行阐述。

(1) 情节 Plot:

叙事型图像的情节反映故事或事件的一个子变化过程,且各个子变化过程通过有机的时空联系组合成叙事图像反映的叙事内容。情节是该标注框架的核心,在图像标注中对应一个OAC中Annotation的Body。由于叙事型图像通常对时空信息处理得比较抽象,因此这种抽象

的时空信息就被赋予在了情节上,情节之间具有时序和空间上的语义关系。情节之间的时序语义关系可以通过如下的对象属性来表达。其中, *nextPlot* 和 *lastPlot* 是比较常用的两个情节间时序关系,其余的几个时序关系,参考了 Allen^[21] 对时间的时序关系的表述,它们对情节之间的起止时间的要求更为严格。实际的情节标注可根据实际情况来选择这些属性以表达更为精确的时序关系。

nextPlot: 下一个情节,表示情节之间的时间先后顺序,与 *lastPlot* 为互逆属性,不受时间精确度约束;

lastPlot: 上一个情节,表示情节之间的时间先后顺序,与 *nextPlot* 为互逆属性,不受时间精确度约束;

overlaps: 表示情节之间在时序上是部分重叠的,且该情节早于另一情节开始并提前结束,和 *overlappedby* 为互逆属性;

meets: 表示一个情节结束时,另一个情节正好开始,和 *metby* 为互逆属性;

starts: 表示该情节和另一个情节同时开始,但提前结束,和 *startedby* 为互逆属性;

during: 表示情节的发生时间段位于另一情节之中,和 *contains* 为互逆属性;

finishes: 表示该情节晚于另一个情节开始,但同时结束,和 *finishedby* 为互逆属性;

equal: 表示两个情节发生时段完全相同。

时序关系是一种抽象的和隐含的语义关系,只有两个情节存在才能体现这种关系,并且在叙事型图像中很难自动提取。图像情节的时

序关系在图像表现上不一定存在顺序,比如图 1 的故事情节就是从图像的两侧向中间发展的,反映了图像的艺术绘制手法。

情节除了在时序上具有上述语义关系外,还具有空间上的语义关系。相对于时间关系,情节之间的空间语义关系在叙事型图像中更为抽象。情节之间的空间关系是指情节发生的地理位置之间的关系。地理位置之间的关系较为复杂,比如地理位置在区域方位上的关系有包含关系、同一关系、交错关系、地理方向关系(如“东方”)等,以及地理位置基础上的行政区域的上下隶属关系,地理位置的演变关系等。本文并不试图解决地点空间关系的建模问题,一方面这部分工作可以通过已有的地理空间数据来完成,另一方面在叙事型图像中很难比较情节发生的地理位置之间的这种语义关系。相对而言,将地理位置作为一种实体(Entity),来反映情节的空间信息及其之间的空间语义关系则更为合适,具体见 Entity 部分。

(2) 实体 Entity:

情节的发展需有相应实体的参与,这是促成情节发展的必要元素。实体包括动态的主体,主要是人物(Person)为主,也包括动物、植物,以及相关的其他各种对象(Object)。

图 4 是实体的一个简单分类体系,根据不同的领域也可以使用领域特有的实体分类体系,该分类体系对图像中的各种可辨别的实体进行分类。Entity 的类型和具体领域有关,比如对于敦煌壁画领域,人物实体就有飞天、菩萨等多种

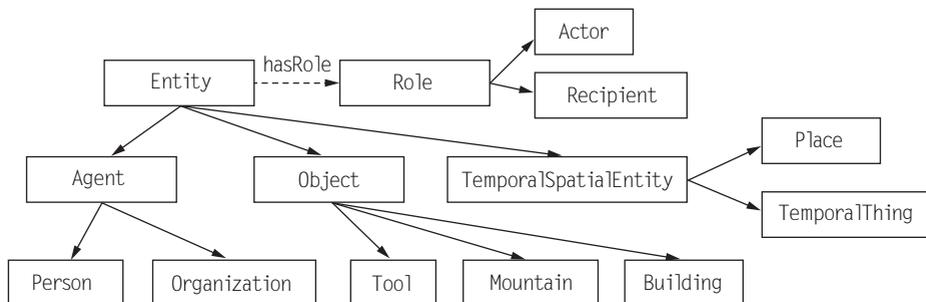


图 4 实体的分类体系示例

类型,这也是常见的图像内容标注任务的重点,将图像中的实体和相应领域中的概念相关联。实体都可以被赋予一个角色 Role,比如图像中的一个人是国王(King),国王就是一个角色,动作发起方(Actor)和接受方(Recipient)也是角色。

在图像中,尤其是艺术作品中,地点通常表现为山脉、河流、宫殿等对象(Object),而不是具体的地理位置信息,如具体的省市街道或者具体的经纬度等,这些具体的地理信息可能需要从其他的资源中去获取。这也是图3的标注框架中情节的空间语义关系没有扩展的原因。

叙事型图像中的时间信息同样比较抽象,难以通过图像来表达,比如情节发生在某年月、某朝代,同样需要参考诸如其他文字类型的资源来获取。图像中可以识别的时间要素通常是通过一定的绘画元素和技术来实现的,如图像中有太阳就是白天,日落就是傍晚等等,这些都是抽象的时间对象。因此,本文将时空实体(TemporalSpatialEntity)放在了实体类别下,并针对叙事型图像标注任务需求,将时空实体划分为 Place 和 TemporalThing 两类,Place 和 TemporalThing 各分为具体(ConcretePlace、ConcreteTemporalThing)和抽象(AbstractPlace、AbstractTemporalThing)两类,以满足叙事型图像时空信息标注的需求。

AbstractTemporalThing 是指时间不确定不具体,如朝代、上午、周末、早晨、春夏秋冬等时间对象,经常用在艺术创作中。ConcreteTemporalThing 则是常见的时间本体建模的对象,如年月日时分秒,Time Ontology^[22]就属此类。需要注意,现实中的时间具有延续性,但在叙事型图画中这种延续性可能不存在,它是零碎割裂的,比如,一个情节是在上午,另一情节可能是下一年的某一个下午。通常情况下,这种时间的表达在图画作品中可能并不明显,因此不能通过时间要素来体现情节的发展。相反,通过情节本身的发展顺序却可以确认时间要素的先后顺序,这也是通过情节本身而不是通过时间要素

来表达情节时序关系的一个原因。

(3) 动态元素 Dynamics:

动态元素是标注模型另一个必不可少的要素。本文沿用了 Rossana Damiano^[9]对 Dynamics 的建模方式,如图5所示,不同之处在于,将原始 Action 的分类使用 DynamicsType 来表示。

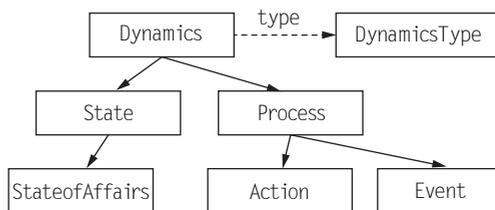


图5 Dynamics 类型

其中 Event 类是常见的事件建模方式,如 Simple Event Model^[12]本体等。如果可以确定事件发生的具体时间、地点,以及比较关注事件的起因、经过、结果等方面可以使用 Event 来建模。其他类在叙事型图像标注中可以暂时不考虑。图像是静止的,图像中的动态元素比较适合用动作(Action)来表征。Action 有两个对象属性 hasActor 和 hasRecipient,hasActor 表示该动作的施事者,hasRecipient 表示该动作的受事者。动态元素 Dynamics 可以有一个类型(DynamicsType),如图5中,原始的 Action 有 KillingAction 和 HeroAction 等子类型。

(4) 情境 Context:

情境(Context)是叙事型图像情节标注中的一类特殊的内容,为该标注模型提供了额外的扩展能力,比如情节发生的天气环境、时代背景等。这些信息并不直接反映在图像中,而是隐藏在图像背后,但对图像的理解很有帮助。在潘诺夫斯基的图像学三层架构中,这部分信息属于第三层面。

整个情节标注框架使用 OWL 进行形式化,元素间语义关系如表1所示。在实际的标注任务中可能存在差异的地方在于,叙事型图像中情节粒度对应的区域大小的划分问题。一个情节被标注在图像的一个被划分的区域,这些被

划分的区域之间可以存在重叠交叉,可理解为图像中不同情节之间在图像区域上存在相同的内容元素,比如图像中某一个特定位置上的一个人物实体可以在多个情节中被共享通用。虽然不同的图像划分区域之间的重叠关系可以通过像素位置的几何运算来精确得到,但通过在OAC的Target上扩展这种区域划分关系来实现更直观方便。图像不同区域间的空间关系包括不重叠(nooverlap)、完全重叠(completeoverlaps)、

包含(contains)、部分重叠(partialoverlaps)。不重叠好理解,完全重叠是指一个图像标注区域所表示的情节有多个,包含关系在图像区域上是指一个大面积的标注区域中包含有小面积的标注区域,部分重叠是指情节所对应的标注区域之间有部分重合。在四种重叠关系中,不重叠、部分重叠较为常见,其中部分重叠也是绘画艺术中常见的缩略手法。

表1 情节标注模型核心属性关系

属性名称	Domain	Range	说明
nextPlot, lastPlot, overlaps, overlappedby, meets, metby, starts, startedby, during, contains, finishes, finishedby, equal	Plot	Plot	情节之间的时序关系
hasEntity	Plot/ Dynamics	Entity	hasEntity 具有子属性: hasObject (Plot, Object), hasParticant (Plot, Agent), hasPlace (Plot, Place), hasTemporalThing (Plot, TemporalThing), hasActor (Action, Entity), hasRecipient (Action, Entity) 等
hasType	Action	ActionType	动作的类型, ActionType 为 DynamicType 的子类
hasRole	Entity	Role	实体的角色
inPlot	Role	Plot	角色所处的情节
hasDynamics	Plot	Dynamics	情节和动作的关系, 具有子属性 hasAction (Plot, Action)
nooverlap, completeoverlaps, contains, partialoverlaps	Seclector	Seclector	标注区域重叠关系, Seclector 为 OAC 中的元素
hasTargetFocus	Annotation	Entity	表示标注区域所指示的现实世界的实体概念
inContext	Plot	Context	情节发生的情境

3 叙事型图像标注模型的应用

本节使用上述叙事型图像标注模型对图1进行情节标注来说明该标注框架的使用情况。通过图1并结合相关的叙事文本可以发现,

图1叙事型图像的叙事顺序是从图像的左右两侧向中心靠拢来体现图像反映的叙事情节的。图6给出了对图1进行标注结果的其中3个情节内容,图7使用RDF三元组的形式呈现。由于版面关系,图6显示的并不是完整的标注结果。

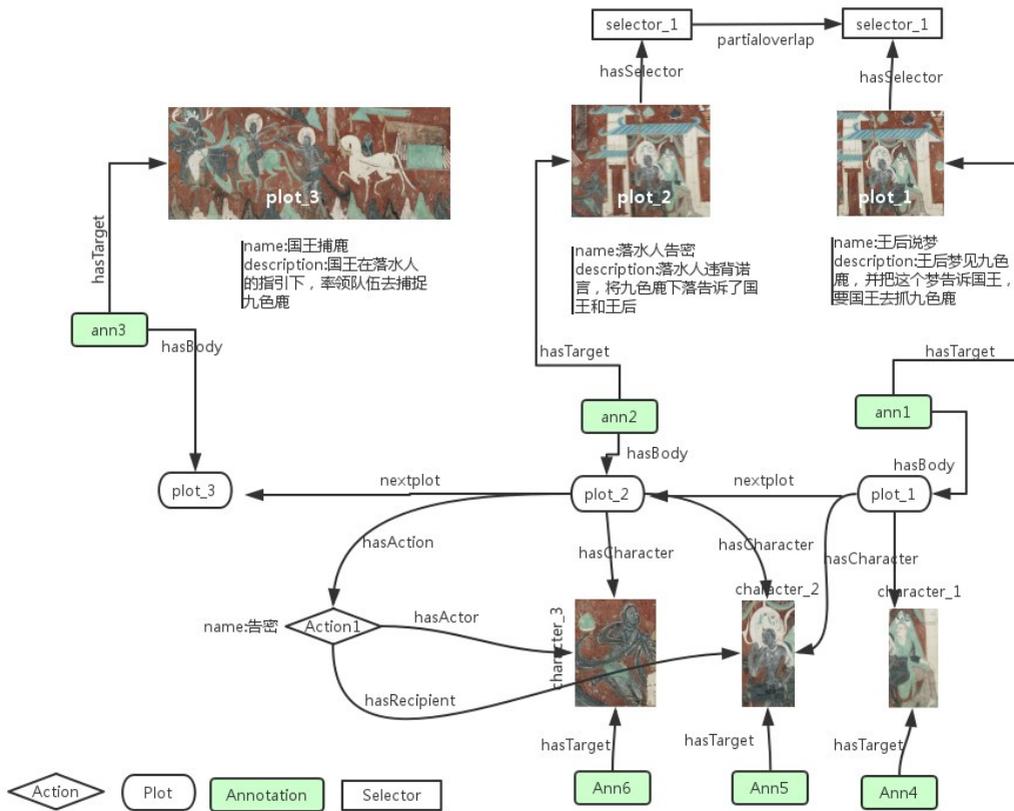


图 6 对图 1 的情节标注示例

情节标注模型中应用到的词汇前缀

```

@prefix base: <http://niao.www.whu.edu.cn/#>.
@prefix oa: <http://www.w3.org/ns/oa#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix dc: <http://purl.org/dc/elements/1.1/>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
    
```

情节 1: plot_1 王后说梦 (带有下划线_前缀的元素是 blank node)

```

<ann1 a oa:Annotation>. #标注 ann1 是一个 oa:Annotation 类型
<ann1 oa:hasBody plot_1>. #标注 ann1 的 Body 是 plot_1
<plot_1 a base:Plot>. # plot_1 是一个情节类型
<ann1 oa:hasTarget _:target_1>. #标注 ann1 的 Target 是 target_1, 指定标注区域
<plot_1 base:name "王后说梦">. # plot_1 的名称
<plot_1 dc:description "王后梦见九色鹿, 并把这个梦告诉国王, 要国王去抓九色鹿">.
<plot_1 base:hasParticant character_1>. # plot_1 情节的参与者
<plot_1 base:hasParticant character_2>. # plot_1 情节的参与者
<character_1 a foaf:Person>. # character_1 是一个人物类型
<character_2 a foaf:Person>. # character_2 是一个人物类型
<character_1 base:hasRole _:role1>. #人物的角色
<character_2 base:hasRole _:role2>. #人物的角色
<_:role1 a base:Role>. # role1 是一个 Role 类型
<_:role2 a base:Role>. # role2 是一个 Role 类型
<_:role1 base:name "王后">.
    
```

```

<_:role2 base:name "国王">.
<_:role1 base:inPlot plot_1>. #角色 role1 在 plot_1 情节
<_:role2 base:inPlot plot_1>.
<ann4 a oa:Annotation>.
<ann4 oa:hasBody _:ann4body>.
<ann4 oa:hasTarget _:ann4target>.
<ann4 base:hasTargetFocus character_1>.# 标注区域 ann4 的目标聚焦于 character_1
<ann5 a oa:Annotation>.
<ann5 oa:hasBody _:ann5body>.
<ann5 oa:hasTarget _:ann5target>.
<ann5 base:hasTargetFocus character_2>.# 标注区域 ann4 的目标聚焦于 character_2
<plot_1 base:hasPlace place_1>. # plot_1 发生的地点是 place_1
<place_1 a base:Place>. # place_1 是一个 Place 类型
<place_1 base:name "宫殿">.
<plot_1 base:nextPlot plot_2>. # plot_1 的下一个情节是 plot_2
<plot_1 base:hasAction _:action_1>. # plot_1 有一个动作是 action_1
<_:action_1 a base:Action>.
<_:action_1 base:hasType _:actiontype1>. #动作 action_1 的类型是 actiontype1
<_:action_1 base:name "说梦">. #该动作 action_1 具体名称是 "说梦"
<_:actiontype1 base:name "语言动作">. # 该动作类型的名称是语言类型
<_:action_1 base:closeNameInWordNet "tel#1">. #和 WordNet 关联
<_:action_1 base:closeNameInWordNet "whisper#1">. #和 WordNet 关联
<_:action_1 dc:description "王后侧身向国王耳语梦境">.
<_:action_1 base:hasActor character_1>. # action_1 的施事者是 character_1
<_:action_1 base:hasRecipient character_2>. # action_1 的受事者是 character_2

```

情节 2 plot_2 落水人告密（部分标注结果）

```

<ann_2 oa:hasBody plot_2>.
<ann_2 oa:hasTarget _:target_2>.
<plot_2 base:name "落水人告密">.
<plot_2 dc:description "落水人违背诺言，将九色鹿下落告诉了国王和王后">.
<plot_2 base:hasParticant character_1>.
<plot_2 base:hasParticant character_2>.
<plot_2 base:hasParticant character_3>.
<character_3 base:name "落水人">.
<plot_2 base:hasPlace place_1>. # plot_2 发生的地点是 place_1，和 plot_1 一样
<plot_2 base:lastPlot plot_1>. # plot_2 的上一个情节是 plot_1
<plot_2 base:hasAction _:action_2>. # plot_2 有一个动作是 action_2
<_:action_2 base:hasType actiontype1>. # action_2 和 action_1 具有一样的类型
<_:action_2 base:name "告密">.
<plot_2 base:hasAction _:action_3>. # plot_2 有一个动作是 action_3
<_:action_3 base:hasType _:actiontype2>.
<_:actiontype2 base:name "肢体动作">.
<_:action_3 base:name "跪着">.
<_:action_2 base:hasActor character_3>.
<_:action_3 base:hasActor character_3>.
<_:action_2 base:hasRecipient character_1>.
<_:action_2 base:hasRecipient character_2>.

```

情节 3 plot_3 国王捕鹿（部分标注结果）

```

<ann_3 oa:hasBody plot_3>.
<ann_3 oa:hasTarget _:target_3>.
<plot_3 base:name "国王捕鹿">.
<plot_3 base:lastPlot plot_2>.
<plot_3 base:hasParticant character_2>.
<plot_3 base:hasParticant character_3>.

```

图 7 图 1 情节标注示例的 RDF 三元组形式

上述三元组可通过如图 8 所示的图像标注工具获得,该工具基于 HTML5 开发,融入了本文设计的叙事模型,每一个标注的区域都会在工具下部的表格中生成一行,表头字段全部来源于本文叙事本体的属性关系,同时用户也可以在标注过程中动态添加自己的属性。所有的标注结果都会记录标注区域的像素值,连同用户填写的其他语义标注信息,一起转换为 RDF 三元组的形式,便于后期基于这些标注数据进行图像语义查询的应用。

上述 3 个情节标注结果很清楚地展现了该标注模型在标注任务中的作用,通过和 OAC 标注框架的结合,叙事型图像的情节得到了有效的关联。需要注意的几点是:

(1) 情节 3 中的国王和落水人虽然在图片上表现的姿态、动作、所处的几何位置和情节 1 以及情节 2 中国王和落水人不同,但在高层语义上,即情节上指的是同一个人物;

(2) 情节 plot_1 和情节 plot_2 的标注目标区域 target_1 和 target_2 之间存在交叉重叠关系,即两个情节在图像的情节标注区域有交叉现象,标注区域之间有共用的区域,表示共享的语义对象。两个情节中重叠的区域对应的实体是国王、王后以及宫殿这几个实体。对应的核

心标注结果如下,该结果是通过标注工具后台自动转换而得:

```
<_:target_1 oa:hasSelector _:selector_1>.
<_:selector_1 oa:region "100,100,300,300">
<_:target_2 oa:hasSelector _:selector_2>.
<_:selector_2 oa:region "50,100,300,300">
<_:selector_1 base:partialoverlap _:selector_1>.
```

(3) 角色 Role 需要进行约束,比如情节 plot_1 中人物 character_1 是“王后”“施事者”(说梦动作的发起人)角色,但在情节 plot_2 中就变成了“王后”“受事者”(告密动作的接收者)角色,通过 inPlot(Role,Plot)关系来进行约束,指定角色所处的特定情节。

(4) 由于采用 RDF 三元组语义进行标注数据的形式化与发布,和已有的受控术语或关联数据进行链接是一个最佳实践,可以极大地促进用户对标注结果的检索与浏览。在该标注框架中扩展的 hasTargetFocus(Annotation,Entity)语义关系,将标注的部分和其对应的概念实体相关联,而实体具有很强的领域特征,可以引入领域概念体系来支撑这种关联。另一个外部链接可以通过 closeNameInWordNet(Action,word#sensenumber)来实现动作术语和 WordNet 词汇的关联,以支撑更为智能的语义检索。

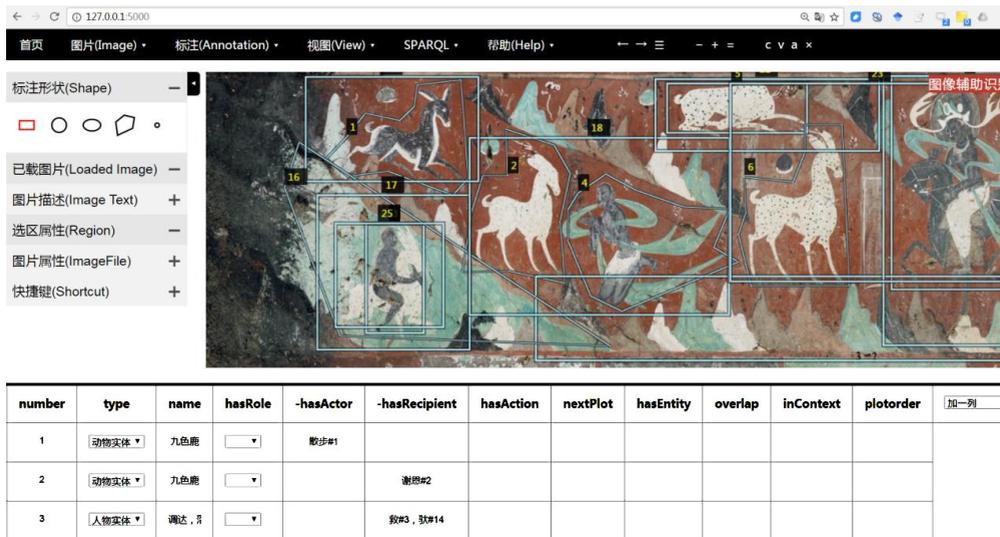


图 8 叙事图像语义标注工具界面

4 讨论与结论

本文对叙事型图像的语义内容标注进行了研究,针对叙事型图像的特点提出了结合 OAC 的叙事型图像的情节标注模型,该模型以情节为核心,构筑 Plot、Dynamics、Entity、Context 等相关的核心词汇,并对各个概念进行模块化分析与设计,以九色鹿本生故事画为标注对象,形成 RDF 三元组的情节标注结果,并提供和领域主题词表等外部关联数据的连接接口。通过设计该情节标注框架,为叙事型图像的检索提供便利,尤其对于叙事型艺术图像作品的情节标注有很大意义。通过使用该情节标注模型,用户可以实现以下检索需求:某情节的下个情节是什么;哪些情节是并行发生的;该情节由哪些人参与,还有哪些对象,有哪些动作,谁是主动的,谁是被动的;哪些人物/对象在多个情节中出现;他们的角色有无变化;他们对应的图像区域有无重叠;相似动作类型的情节有哪些。

上述只列举了一部分显示该模型特性的能力问题,已经可以很好地体现出它对于叙事型图像内容研究的意义,各种需求组合以及推理规则都可以被进一步提出来。当然在具体的标注应用中仍有大量的工作需要做,比如如何使用计算机视觉技术来辅助识别叙事型图像的情节和对象。目前,图像自动识别研究主要针对一般性日常图像,针对叙事型艺术图像的研究十分稀少。此外,叙事型图像作品本身就带有原始作者强烈的主观认识,后人的解读也各有差异,因此叙事型图像中情节的划分存在不同的见解,这给情节提取和标注带来了不确定性。为了反映这种认识上的差异,OAC 提供了标注信息创作者这样的元数据,这种做法解决同样一个图像作品分散式标注的差异性冗余问题,但未来还需考虑如何对不同标注者的标注信息进行映射;最后,该情节标注模型和领域性主题词表,以及通用主题词表 AAT、ICONCLASS 等外部数据的关联问题,也是实际的标注工作需要解决的,而并非是图像情节标注工作所独有。

参考文献

- [1] Stanford Vision Lab. ImageNet [EB/OL]. [2017-06-01]. <http://www.image-net.org/>.
- [2] Microsoft Research. Microsoft cognitive services-computer vision API [OL]. [2017-06-01]. <https://www.microsoft.com/cognitive-services/en-us/computer-vision-api/>.
- [3] Microsoft Research. Visual storytelling dataset (VIST) [EB/OL]. [2017-06-01]. <http://visionandlanguage.net/VIST/>.
- [4] Tuffield M M, Millard D E, Shadbolt N R. Ontological approaches to modelling narrative [EB/OL]. [2017-06-01]. <http://eprints.soton.ac.uk/id/eprint/261962>.
- [5] Jewell M O, Rissen P, Harris T. The stories ontology [EB/OL]. [2017-06-01]. <http://www.contextus.net/stories/>.
- [6] Nakasone A, Ishizuka M. Storytelling ontology model using rst [C]//Proceedings of the IEEE/WIC/ACM international conference on Intelligent Agent Technology. IEEE Computer Society, 2006: 163-169.
- [7] Mann W C, Thompson S A. Rhetorical structure theory: a theory of text organization [M]. University of Southern California, Information Sciences Institute, 1987.
- [8] Ciotti F. Toward a formal ontology for narrative [J]. MATLIT: Materialidades da Literatura, 2016, 4(1): 29-44.
- [9] Damiano R, Lieto A. Ontological representations of narratives: a case study on stories and actions [C]//OASIS - Open Access Series in Informatics. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2013: 76-93.

- [10] Lagoze C, Hunter J. The ABC ontology and model[J]. Journal of Digital Information, 2006, 2(2): 1-18.
- [11] Rissen P, Lippell H, Chadburn M, et al. Storyline ontology [EB/OL]. [2017-06-01]. <http://www.bbc.co.uk/ontologies/storyline/>.
- [12] van Hage W, Malaisé V, Segers R, et al. Simple event model[EB/OL]. [2017-06-01]. <http://semanticweb.cs.vu.nl/2009/11/sem/>.
- [13] Meditskos G, Dasiopoulou S, Efstathiou V, et al. Ontology patterns for complex activity modelling [C]// International Workshop on Rules and Rule Markup Languages for the Semantic Web. Springer Berlin Heidelberg, 2013: 144-157.
- [14] Krisnadhi A A, Hitzler P. A core pattern for events[EB/OL]. [2017-06-01]. http://ontologydesignpatterns.org/wiki/images/5/52/WOP2016_paper_04.pdf.
- [15] Doerr M. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata [J]. AI Magazine, 2003, 24(3): 75-92.
- [16] Scherp A, Franz T, Saathoff C, et al. F—a model of events based on the foundational ontology dolce+DnS ultralight [C]// Proceedings of the Fifth International Conference On Knowledge Capture. ACM, 2009: 137-144.
- [17] Shaw R, Troncy R, Hardman L. Lode: linking open descriptions of events [C]// Asian Semantic Web Conference. Springer Berlin Heidelberg, 2009: 153-167.
- [18] Clarke, D. Deep Image annotation: making a difference in knowledge organization [EB/OL]. [2017-06-01]. <http://www.iskouk.org/sites/default/files/ClarkePaperISKO-UK2015.pdf>.
- [19] Xu L, Wang X. Semantic description of cultural digital images: using a hierarchical model and controlled vocabulary [J/OL]. D-Lib Magazine, 2015, 21(5/6) [2017-06-01]. <http://www.dlib.org/dlib/may15/xu/05xu.html>.
- [20] Open Annotation Community Group. Open annotation collaboration [EB/OL]. [2017-06-01]. <http://www.open-annotation.org/>.
- [21] Allen J F. Maintaining knowledge about temporal intervals [J]. Communications of the ACM, 1983, 26(11): 832-843.
- [22] Cox S, Little C. Time ontology in OWL [EB/OL]. [2017-06-01]. <https://www.w3.org/TR/owl-time/>.

徐 雷 武汉大学信息资源研究中心讲师。湖北 武汉 430072。

王晓光 武汉大学信息资源研究中心教授。湖北 武汉 430072。

(收稿日期: 2017-04-25; 修回日期: 2017-06-07)