

学术论文研究方法标注规范

1 数据标注背景

研究方法在人类的学术活动中发挥着核心作用，研究方法是解决科学问题的必由之路。了解和认识一个学科研究方法的使用和发展情况，对于加强该学科的学术研究以及该学科的教育教学工作具有重要意义。为了更好的认识一个学科内研究方法的使用情况，需要对学科内研究方法的使用情况进行调查分析。学术论文研究方法的人工标注是本研究的基础，在此基础上，本研究将探究单一语种中的研究方法自动标注和跨语言的研究方法自动标注。

2 概念界定

研究方法是指在科学研究中发现新现象、新事物，或提出新理论、新观点，揭示事物内在规律的工具和手段。本研究的标注工作主要集中于对图书情报领域学术论文的研究方法的标注。学术论文研究方法的标注就是识别一篇论文所使用的方法。一篇学术论文可能使用了多种研究方法，因此在学术论文的研究方法标注过程中需要标注出一篇学术论文所使用的所有研究方法。表 1 举例给出了使用一种或多种研究方法的论文。

表 1 使用不同数量研究方法的论文举例

论文标题	DOI	研究方法	方法数量
Semantic annotation of biosystematics literature without training examples	https://doi.org/10.1002/asi.21246	experiment	1
Factors affecting shapers of organizational wikis	https://doi.org/10.1002/asi.21266	interview, questionnaire	2
Analysis of participation in an online photo - sharing community: A multidimensional perspective	https://doi.org/10.1002/asi.21278	questionnaire, content analysis	2
An analysis of image queries in the field of art history	https://doi.org/10.1002/1532-2890(2000)9999:9999	experiment, questionnaire, interview, content analysis	4

3 任务描述和标注流程

(1) 标注任务描述

标注一篇学术论文所使用的研究方法是在理解研究方法分类体系的前提下,识别一篇学术论文所使用的研究方法。在本研究的方法分类体系下,标注人员需要从数据收集技术的角度来判别一篇学术论文所使用的研究方法。此外,一篇学术论文可能使用多种研究方法,在本研究的标注任务中需要标注出一篇学术论文所使用的所有的研究方法。

(2) 标注流程

标注流程如图 1 所示。首先制定初始标注规范,让所有标注人员认真学习研究方法编码体系,然后进行预标注。预标注过程将进行多轮,每一轮都在待标注数据中随机选取 20 篇学术论文(每一轮的标注数据不重复)。在每一轮的研究方法标注后,均需要计算两位标注人员的一致性(使用 Kappa 值度量),然后由两位标注人员对不一致的意见进行讨论。经过多轮预标注和讨论,直至达到可以接受的一致性,预标注的一致性记为 Kappa1。

在完成预标注后,进行正式标注,两位标注人员分别标注 500 篇学术论文。在两位标注人员的标注数据中有 100 篇的重复数据(该部分的重复数据将不会与预标注中的数据重复),用于计算在正式标注过程中两位标注人员的一致性(Kappa2)。

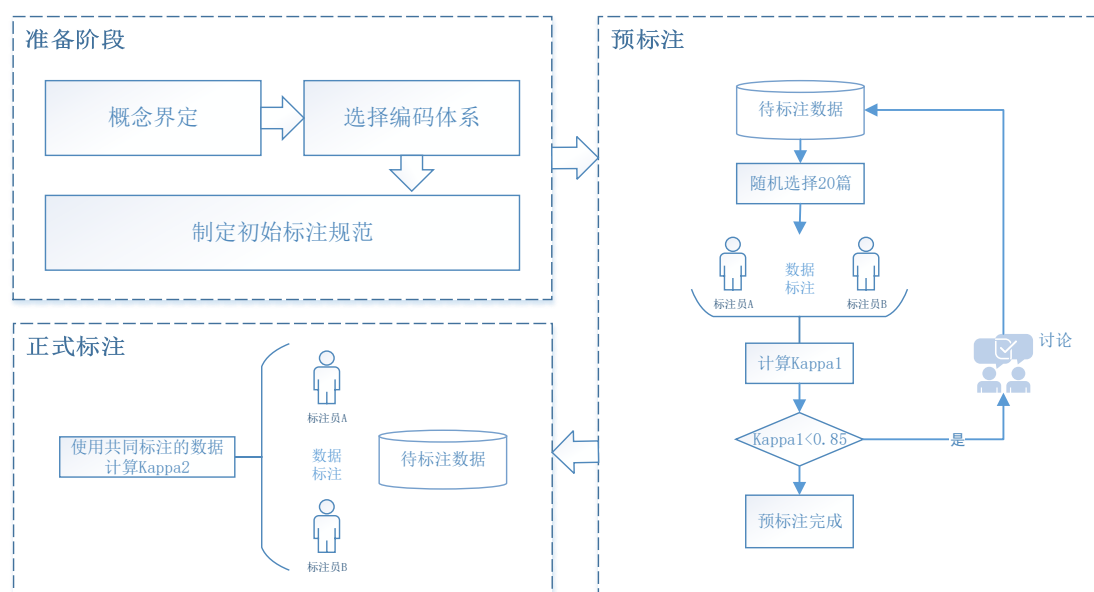


图 1 数据标注流程图

4 学术论文研究方法判别方法

对于学术论文研究方法的判别,需要通览摘要和全文。对于所使用的方法的判别,可以参考以下步骤。

(1) 在全文中查找论文研究方法章节

首先在全文中查找论文是否存在研究方法章节,如果存在则阅读该章节并通过判断论文的数据收集技术来判定论文所使用的研究方法。

(2) 判断论文所分析的数据来源

如果论文分析的数据是学术论文或专利文献，那么论文多半使用了文献计量方法。但是用了文献或专利数据，也不一定是文献计量方法，还得确切的判定文章是使用了计量方法，例如引文网络、共被引等。如果分析的数据是网络数据，那么可能使用了网络计量方法。同与文献计量方法一样，也要同时判断文章是否使用了网络计量的方法，比如一些频次分析等。如果数据源是历史文献等，可能是历史方法；如果分析的数据源是服务器等记录的用户访问记录等，那么可能使用的是事务日志分析法。其他研究方法可以通过类似的方式判别。

(3) 判断论文是否通过构建了理论模型或对某个概念进行了阐述和讨论

在早期的中文学术论文中，可能无法直接看出论文使用了什么数据集进行分析。早期的论文存在许多论文通过概念分析、理论检验或类似活动（如概念分析、建模及理论构建）来收集数据，这些论文多数是使用了理论方法。对于构建算法模型的论文，如果论文只进行了理论推导，没有做实验验证算法，那么该论文应当标注为理论方法。如果是对历史资料\信息进行分析讨论的理论方法，标注为历史方法。

5 标注平台的使用

本研究的研究方法标注任务使用 Doccano 开源标注平台完成。本研究在服务器上搭建了该平台，平台的访问链接如下：<http://121.36.2.198:8000/>。管理员将会通过管理员账户为不同标注员分配任务，标注员完成自己任务列表中的标注任务即可。

(1) 平台主页，右上使用管理员分配的账号点击登录

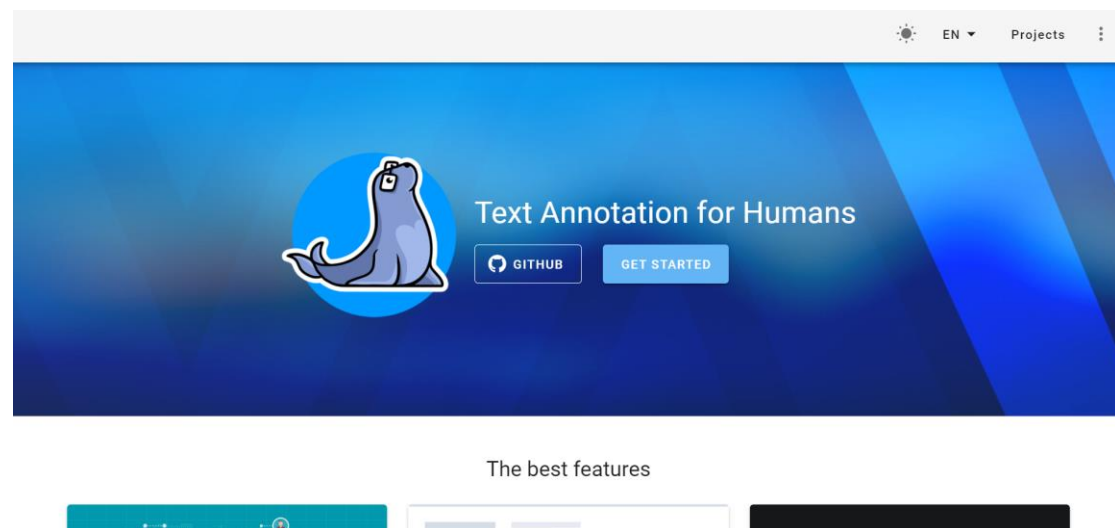


图2 标注平台主页

Login

Username

Password

Login

图 3 登录页

(2) 任务列表

<div><div>Create</div><div>Delete</div></div> <div>Search</div>				
<input type="checkbox"/>	Name	Description	Type	Updated
<input type="checkbox"/>	方法分类	学术论文研究方法标注	DocumentClassification	18/02/2022 03:34
<input type="checkbox"/>	跨语言中文标注 (去除综述书评)	跨语言中文标注 (去除综述书评)	DocumentClassification	14/03/2022 02:43
<input type="checkbox"/>	跨语言中文标注(预标注)	正式标注前的预标注	DocumentClassification	16/03/2022 09:22
Rows per Page 10 1-3 of 3 < < > >				

图 4 任务列表

(3) 进入任务，点击标注即可开始标注

跨语言中文标注(预标注)

Start Annotation

Home

Dataset

Labels

Members

Comments

Guideline

Statistics

Settings

Welcome to Doccano!

1 Import a dataset

图 5 标注任务主页

(4) 数据标注页面，直接按照标注规范标注即可；存在问题可以给管理员留言。



图 6 数据标注页面