

医疗语义众包标注平台设计和应用研究^{*}

郑文婕 张敬谊 李 静 余 盼

(万达信息股份有限公司 上海 201112)

〔摘要〕 详细阐述医疗语义众包标注平台的标注流程与方法,包括平台角色分配、任务动态分配、用户能力评估、质量控制与检查等方面,介绍平台技术架构、结构与功能,通过实验对比传统人工标注与众包标注平台的应用效果,结果证明后者在标注速度、耗时、正确率上均优于前者。

〔关键词〕 语义标注;标注平台;众包

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2020.07.011

Study on Design and Application of Medical Semantic Crowdsourcing Annotation Platform ZHENG Wenjie, ZHANG Jingyi, LI Jing, SHE Pan, Wonders Information Co., Ltd, Shanghai 201112, China

〔Abstract〕 The paper expounds the annotation process and methods of medical semantic crowdsourcing annotation platform in detail, including platform role allocation, dynamic task allocation, user capability assessment, quality control and inspection, etc., introduces the technical architecture, structure and functions of the platform, and compares the application effect of traditional manual annotation and crowdsourcing annotation platform by experiments. The results show that the latter is superior to the former in annotation speed, time consuming and accuracy.

〔Keywords〕 semantic annotation; annotation platform; crowdsourcing

1 引言

自然语言处理研究通常依赖规模庞大、经过标注的语料库。通过语义标注为文本添加规范化的知识标签,使机器能够理解文本内容,有助于对文本信息进行深度挖掘。语义标注方法主要包括机器标注和人工标注。需要进行语义标注的文本通常专业

性较强、蕴藏大量隐性关联知识,而限于机器学习技术水平,机器语义标注水平较低,无法满足实际使用需求。目前多数情况仍采用传统人工标注方式,需要耗费大量人力和时间成本。近年来以医疗文本数据为基础的智能应用场景和方法层出不穷,如电子病历文本症状自动识别^[1]、医疗事件识别^[2]、在线医疗文本实体识别^[3]等。而语义标注是进行有监督模型训练学习的先决条件,标注的效率和质量是重要考量标准。传统人工标注方式容易受到标注人员医学专业水平、工作态度及精神状态等因素影响,标注结果一致性较差、质量难以保证且进度难以把控。在互联网时代背景下通过众包平台完成海量医疗标注任务成为一种新型的语义标注方式。当前基于群体智慧^[4-5]理论的众包语义标注平

〔收稿日期〕 2019-12-27

〔作者简介〕 郑文婕,工程师,发表论文1篇;通讯作者:余盼,中级工程师。

〔基金项目〕 上海市经济信息化委人工智能创新发展专项“面向复杂传染病智能诊疗的数据共享平台建设”(项目编号:2018-RGZN-02058)。

台在军事语料提取^[6]、地理空间信息研究^[7]、图像分类^[8]、语音语料库设计^[9]等领域得到广泛应用,其标注结果优于个体人工标注方式。本研究构建基于众包的互联网语义标注平台,将大规模标注任务转化成若干小任务分配给互联网用户,降低单个用户完成任务的难度,从而缩短标注工作时间、提升整体效率。

2 标注流程与方法

2.1 标注流程

为构建多方参与的互联网语义标注平台,保证众包协作标注的质量,设计平台整体标注流程,见图1。任务发起人将需标注文本发布至标注平台。标注人在通过平台能力评估后可参与相应标注任务。当多个标注人完成同一个标注任务后,平台通过质量检查机制评估标注结果,将有争议标注文本传由检查人员进行裁决。任务结束后平台将标注和质量评估结果反馈给发起人,重新评估参与任务者的标注能力。

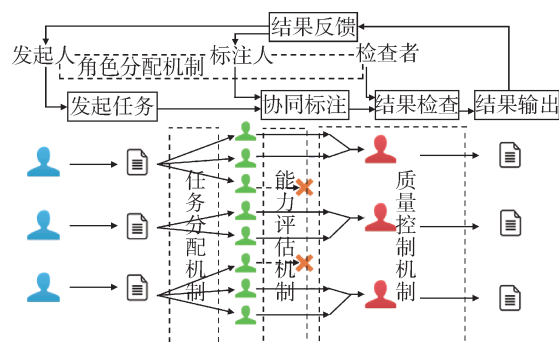


图1 众包标注流程

2.2 平台角色分配

包括发起人、标注人和检查者。发起人是拥有待标注文本、有标注需求的用户。标注人拥有某一领域专业知识技能,愿意参与完成众包标注任务并获得一定报酬或奖励。检查者通常是某一领域专家,当标注结果不一致时提供裁决和意见,确定最终结果。在一个标注任务中标注人不能兼任检查者。

2.3 任务动态分配

所有用户可报名参与任务,人数不限,尽量缩短标注任务花费时间。可设置使多个用户标注同一道题目,发挥群体智慧优势,通过平台质量控制与检查机制得到最终结果。

2.4 用户能力评估

对不同领域、难度的标注任务设置相应门槛,选择符合条件的标注人参与,以提高标注效率。为此平台引入用户能力评估机制,考察用户标注能力,主要分为标注能力测试和标注结果评分两部分。在标注任务创建阶段,发起人可对标注任务设置用户能力等级要求和测试题要求,两项均满足的用户可报名参与任务。任务结束后,平台根据所有标注人的结果以及检查人员的最终裁定结果,通过算法模型及时计算出每个标注人的正确率,更新标注人的用户等级得分,该得分会在标注人参与其他标注任务时产生影响。

2.5 质量控制与检查

平台引入质量控制与检查机制,主要通过标注工具、质量检查算法^[10]和人工裁决3方面实现。在标注阶段平台提供一套标注工具,整合各种标注类型(分类标注、关键词标注、实体标注等)规范,帮助标注人规避不符合要求的标注,提升标注质量。当标注完成后平台根据质量检查算法对当前标注结果进行评估,找出结果存在争议的标注题目,交由检查者进行裁决。平台将集体标注结果和裁决结果优化汇总后得到标注任务最终结果。

3 语义标注平台实现

3.1 技术架构

采用B/S架构,基于Django框架(版本1.11)实现系统各项功能,开发系统为Linux,数据库版本为MySQL 5.7.16。Django是基于Python语言的

Web 开发框架，通过其独特的模型 - 视图 - 模板 (Model - View - Template, MVT) 模式提升开发效率和设计灵活性。Django 具备用户登录、后台管理等多种 Web 开发通用模块和第 3 方插件，避免重复性编程，支持动态数据库访问，大大简化数据库编程。

3.2 结构与功能 (图 2)

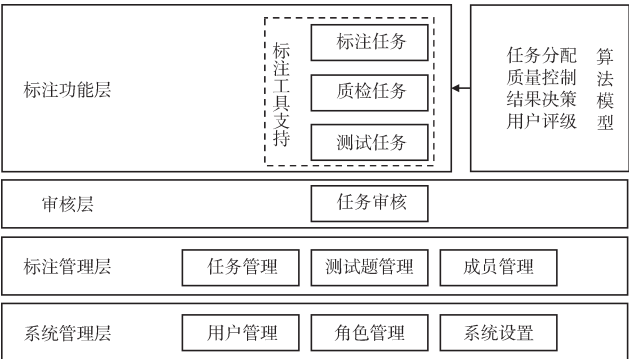


图 2 标注平台结构

- 3.2.1 系统管理层 由管理员负责进行管理和维护，包含用户管理、角色管理、系统设置等基础功能。
- 3.2.2 标注管理层 便于发起人创建和管理标注任务及管理参与任务的各个成员。发起人可创建和编辑测试题，作为标注人参与某个标注任务的先决条件，评估标注人对于该领域文本的标注能力。
- 3.2.3 审核层 对标注任务内容进行必要审核，由平台自动审核和人工审核两种方式共同实现。
- 3.2.4 标注功能层 标注人和检查人员可利用平台集成的标注工具完成标注、测试及检查任务。对于不同类型的标注任务 (分类标注、实体标注、关键词标注等)，标注工具提供不同标注界面和操作方式辅助用户更快更好地完成标注工作。

4 实验案例与应用效果

4.1 实验案例

- 4.1.1 实验描述 现利用众包标注平台完成一项

文本分类标注任务，以考察众包标注平台实际应用效果，与传统人工标注方法进行对比。采用 6 000 条某三甲医院某科室超声文本作为标注对象，要求对每条文本进行分类。实验数据有真实标签，用于最终计算标注正确率。实验邀请两组标注人，均为来自该医院某科室的医学研究生。第 1 组采用传统人工标注方法，共有 3 名标注人参与，各自标注 2 000 条文本，在 Excel 文件中为每条文本打上标签。第 2 组利用众包标注平台进行标注，共有 9 名标注人和 1 名检查人员，在图形化操作环境下利用标注工具进行标注。文本重复率为 3，即每条文本有 3 个标注人打上的标签，可算得每个标注人平均答题量为 2 000，与第 1 组相同。标注题目由平台随机分配，各标注人答题量由答题速度决定。标注完成后检查者对不一致结果进行分析、重新标注，得到最终结果。记录实验过程，考察两组人完成标注任务所耗费的时间，标注人平均速度以及标注正确率。

4.1.2 实验结果 (表 1) (1) 平均标注速度。造成标注速度差异的原因，一是标注平台向用户提供友好的标注界面和便捷的标注工具，使用户看题更清楚、操作更灵活，标注速度较快；二是众包标注平台任务分配机制下，工作效率高的标注人会多答题，有助提升整体标注速度。(2) 标注花费时间。传统人工标注方法为约 48 小时；众包标注为 40 小时，标注和检查总耗时为 42 小时，低于传统人工标注。众包标注平台标注人的平均速度较快，完成任务时间缩短。同时众包标注过程依靠多人标注的群体智慧并受到平台质量控制机制影响，最终不一致的标注结果较少，减轻检查人工作量，检查耗时较短。(3) 标注正确率。众包标注高于传统人工标注方法。两组标注人员能力相近，标注正确率存在差异的原因在于两方面：一是众包标注结果体现集体智慧优势，生成优于个人标注的集体性标注成果；二是利用平台质量控制与检查机制对不一致的标注结果进行汇总和裁决，最终生成科学性、一致性较高的标注成果。

表 1 众包标注平台与传统人工标注对比

标注方法	参与人数	标注花费时间 (小时)	检查花费时间 (小时)	平均速度 (条/每人每小时)	正确率 (%)
传统人工标注	3 名标注人	48	—	42	92.0
众包标注平台	9 名标注人, 1 名检查者	40	2	50	99.8

4.2 应用效果

相比于传统人工标注方法,应用众包标注平台不仅能提升标注速度、缩短标注任务耗时,还能大幅提高标注正确率。平台提供的标注工具和操作界面能够强化标注人个体标注能力;平台质量检查机制能够控制和加强标注人群的群体智慧。

5 结语

本研究提出众包标注的流程规范与方法,设计多种算法模型实现标注平台的高效运作和质量控制,在此基础上构建基于众包的互联网语义标注平台。实验案例证明相比于传统人工标注方法,利用众包标注平台效率和正确率较高。众包语义标注平台是解决语义标注问题的有效途径,对于构建大规模语料库、助力自然语言处理研究具有一定价值。目前本研究建立的标注平台尚有较大发展空间,例如用户能力评估模型和质量控制机制仍需改进和完善,集体标注结果需要更复杂的算法模型进行汇总优化等。随着标注平台实际运营数据的增长,将对平台数据进一步深入分析,加入更多模型和算法,不断提升众包标注工作效率、优化集体标注成果。

参考文献

1 龚凡,王梦婕,阮彤,等. 电子病历文本症状自动识别方法 [J]. 医学信息学杂志, 2016, 37 (7): 7-14.

2 侯伟涛,姬东鸿. 基于 Bi-LSTM 的医疗事件识别研究 [J]. 计算机应用研究, 2018, 35 (7): 1974-1977.

3 苏娅,刘杰,黄亚楼. 在线医疗文本中的实体识别研究 [J]. 北京大学学报 (自然科学版), 2016, 52 (1): 1-9.

4 柯永红,俞士汶,穗志方,等. 基于群体智慧的语料标注方法研究 [J]. 中文信息学报, 2017, 31 (4): 108-113.

5 Bonabeau E. Decisions 2.0: the power of collective intelligence [J]. Mit Sloan Management Review, 2009, 50 (50): 45-52.

6 周彬彬,张宏军,张睿,等. 军事语料实体标注系统的设计与实现 [J]. 信息系统工程, 2018 (8): 56-60, 62.

7 张春菊,张雪英,王曙,等. 中文文本的事件时空信息标注 [J]. 中文信息学报, 2016, 30 (3): 213-222.

8 陆寄远,黄承慧,侯昉,等. 基于 Hadoop 云计算平台的图像分类与标注 [J]. 电信科学, 2014, 30 (2): 65-69.

9 徐健,热依曼·吐尔逊,吾守尔·斯拉木. 在线多语种语音语料库平台的研究与实现 [J]. 信息通信, 2018 (4): 150-153.

10 逢居升. 众包质量控制策略及评估算法研究 [J]. 计算机学报, 2013, 36 (8): 1636-1649.

关于《医学信息学杂志》启用
“科技期刊学术不端文献检测系统”的启事

为了提高编辑部对于学术不端文献的辨别能力,端正学风,维护作者权益,《医学信息学杂志》已正式启用“科技期刊学术不端文献检测系统”,对来稿进行逐篇检查。该系统以《中国学术文献网络出版总库》为全文比对数据库,可检测抄袭与剽窃、伪造、篡改、不当署名、一稿多投等学术不端文献。如查出作者所投稿件存在上述学术不端行为,本刊将立即做退稿处理并予以警告。希望广大作者在论文撰写中保持严谨、谨慎、端正的态度,自觉抵制任何有损学术声誉的行为。

《医学信息学杂志》编辑部