

# 数据标注师 人工智能背后的人工力量



目前人工智能落地场景不断丰富,智能化应用正改变着我们的生活。而在AI产业高速发展的背后,数据标注师这个新职业的从业人数也正在壮大。数据标注行业流行着一句话,“有多少智能,就有多少人工”。目前AI算法能学习的数据,必须通过人力逐一标注,这些人力为AI产业提供养料,构建了AI金字塔的基础。

## 数据标注需求持续增加

“现在科研界研究的都是无监督、小样本的深度学习,通过三维合成数据,用虚实结合的数据生成方式来训练机器,尽量减少数据的采集和标注,让机器自主学习、自主进化。”王金桥说,但由于缺乏理论上的突破性技术,所以虽然技术增长速度很快,但整体水平还比较低,目前的深度学习还是依赖基于统计意义的大数据模型,这要求数据足够多、足够均衡,基本满足真实世界的分布。

因此,标注这项工作会一直存在。

但王金桥也表示,随着无监督、小样本深度学习的进步,重复性标注的工作量会越来越小。“机器的识别和人一样,人经过几千年的进化,用语言用文字记录和存储几千年的文明,所以看到桌子就知道是桌子,看到灵芝知道是灵芝。机器也需要不断理解更多的内容,有数据标签,它才能学习,才会有智能。数据的加工是一个长期存在的过程,由画框到基础词汇,慢慢形成自己的知识图谱,才能自我推理和思考。”

目前的数据标注公司基本采取“计件付费”的模式,标注员的待遇与任务量和难度直接相关,熟练工一天能标几千张图片,月收入最高过万。这项工作也有一定专业性,受过培训才知道怎么标、标得清楚,人也要认真细心。“每天产生的数据量太大了,数据量持续增加,对标注的需求也持续增加。”王金桥说。

据阿里巴巴集团副总裁、阿里巴巴人工智能实验室总经理陈丽娟介绍,未来项目的整体规划将聚焦贫困地区,寻找更多更适合发展“AI标注”产业的地区来落地。同时,也希望更多的人工智能企业加入,把AI标注的订单定向输送给贫困地区。

## 机器学习必需数据标注

AI数据标注员被称作“人工智能背后的人工”。“数据是人工智能的血液。当下是大数据基础上的人工智能,是数据智能的深度学习时代,可以说谁掌握了数据,谁就有可能做好。”中科院自动化所研究员、视语科技创始人王金桥说,当前的人工智能也被称作

数据智能,在这个发展阶段,神经网络的层数越多,神经网络越深,需要用于训练的数据量越大,“比如目前人脸识别做得好的是中青年人脸识别系统,因为年轻人坐车住酒店,采集的数据量大,小孩和老年人数据相对较少。”

但同时,只有数据是没用的。

对于深度学习来讲,数据只有加上标签才有意义,才能用于机器的学习和进化。“标注是一个必须的工作。”王金桥说。

王金桥介绍,从数据的收集、清洗、标注到校验都离不开人工。数据标注最基本的就是画框,比如检测目标是

车,标注员就需要把一张图上的所有车都标出来,画框要完全卡住车的外接矩形,框得不准确机器就可能“学坏”。再比如人的姿态识别,就包括18个关键点,经过训练的标注员才能掌握这些关键点的标注,标注完成的数据也才能符合机器学习的标准。

## 人工标注帮助AI快速落地

随着人工智能的发展,数据的训练量非常大,数据标注公司应运而生,这些公司以网络方式运作,一个平台有产品经理和项目经理,接到一个任务就找人来做,大家通过网络群组报名后,由产品经理来培训,之后各自领取自己的任务,登录账号进行标注,检验经理校验合格后就付钱,不合格则需要重新修正。

“目前已经形成庞大的数据加工队伍,仅北京就有一百多家专门从事数据标注的公司,全国从事这项工作的人大概超过千万,很多头部的互联网技术企业都有自己的数据标注公司。”王金桥说,“这个阶段数据对性能的贡献是最大的,数据越多越丰富、代表性越强、模型效果越好,算法的健壮性和鲁棒性就越强。目前情况是大

部分AI公司都还没有实现盈利,但标注公司除外。”

据王金桥介绍,国外也是一样,无人零售、无人驾驶等都需要大量的人力,基于用工成本的问题,除了隐私数据之外,他们会把标注工作放在第三世界国家完成,马来西亚、泰国、印度等国家都有数据标注分公司。

他直言,目前这种大量的人工标注是有价值的,因为理

论上解决问题很难,但有了大量数据,设计深度学习网络,可以在特定场景特定应用中用数据训练神经网络,从而在很多场景中可以让AI快速落地占领市场、驱动行业应用、促进行业升级和迭代。

王金桥表示,目前人工智能的智能性虽然比较弱,但在各行各业都会带来改变,这是AI推动产业革命的机会。



大开眼界

## “智能背包”

近日,谷歌与奢侈品品牌伊夫·圣罗兰合作,推出一款智能背包。该产品主打特色是触控式肩带,可用于控制、连接你的智能手机。在背包内,有一枚可拆卸的“标签”,内置电池和内部硬件。一旦与手机配对,用户就可以使用左肩带正面触敏区域上的四个手势来控制设备:向上滑动,向下滑动,双击并覆盖它。



## 瑞典推出求职面试机器人

据俄罗斯卫星网近日报道,为了消除意识和潜意识产生的偏见,瑞典斯德哥尔摩市下辖的厄普兰德斯—布罗市政府决定让一个名为“天盖”(Tengai)的机器人来对求职者进行面试。来自该市的安德斯·奥赫德成为瑞典第一个被所谓的“无偏见”人工智能招聘程序录用的人。

Tengai由瑞典科技公司Furhat Robotics与瑞典最大招聘公司之一TNG联手研制,他们称,这款机器人能在面试时

保持中立,不受求职者年龄、性别、种族或外貌的影响。

据瑞典广播电台报道,在与Tengai进行了富有成效的面试后,安德斯获得了市政府数字协调员的工作,成为第一个被Tengai招聘的人。

安德斯说:“被Tengai面试和被人面试时的感觉迥然不同。比如,你无法通过观察Tengai的表情来判断它是否满意你的答案。”

TNG公司首席执行官阿萨·埃德曼·卡尔斯特罗默在此前的

一次采访中解释道:“人类倾向于寻找并喜欢和自己有相同兴趣爱好的人,这通常通过私人问题来完成,而这些问题会让我们投入情感,所以,每名招聘人员都携带着某种潜意识的偏见。”

在Furhat Robotics发布的一则广告中,一位戴着头巾、拥有学位的女性被三位招聘经理拒绝,最终却被一个“没有偏见”的机器人录用为最佳候选人。

(本版综合)