

图像识别在互联网“问题地图”监控中的应用

程维中

绥化学院 信息工程学院, 黑龙江 绥化 152000

【摘要】: 分析地图发展历程及互联网“问题地图”识别依据, 阐述互联网“问题地图”监控发展史及“问题地图”监控主流架构。结合“问题地图”监控的不同应用场景, 论述国家节点、省级节点支持下的互联网“问题地图”监控部署。引入图像识别技术, 分析卷积神经网络图片识别与样本库图片识别在互联网“问题地图”监控的应用。着重阐述卷积神经网络图片识别的人工干预流程, 简述卷积层、池化层、激活函数三大模块作用机理, 并指出自适应算法优化方案。结合当下成功案例, 在结论处对未来发展前景进行展望。

【关键词】: 图像识别; “问题地图”; 监控系统; 卷积神经网络

互联网以便捷和高速传播著称, 但因受面广且传播速度快等特点, 易引发舆论热潮。“问题地图”事件作为社会恶性事件之一, 易被媒体利用而大肆宣传。如迪奥宣讲会“问题地图”事件、苹果内置地图将台湾单列一国事件, 危害国家社会治安, 对国家利益造成一定影响。为科学利用互联网技术, 需对互联网地图全局布控、全天候监测, 结合图像识别技术, 对“问题地图”进行实时监控, 将问题扼杀在源头。

1 互联网“问题地图”简述

1.1 地图发展历程

地图是人类描述各地区地貌绘制而成的二维图像, 依照数学原理或投影技术概括或表示自然现状或社会现象。早在时期时代, 古人便开始制作地图。我国地图最早出现在春秋战国时期。约1900年前, 古希腊地理学家撰写《地理学指南》, 描述地图绘制方法。我国地域学家裴秀基于前人经验创立《制图六体》理论, 提出地图尺寸、方位、距离等概念, 实现了地图的精准绘制。

在互联网技术的高速发展下, 地图绘制及测绘方式发生巨大变化, 地图产品朝向多样化发展, 基本实现普及化及全面化。目前, 电子地图、多媒体地图、互联网地图成为地图的主要表现形式。面对地图产品及呈现形式的变化, 地图的合法性与合规性问题逐渐凸显。“问题地图”的概念应运而生。

1.2 互联网“问题地图”识别

我国关于地图管理的条例包括:《中华人民共和国测绘法》、《地图管理条例》、《公开地图内容表示若干规定》、《公开地图内容表示补充规定》、《地域审核管理规定》。依照上述法律条例, 互联网“问题地图”的界定包括:

(1) 危害国家统一及主权领土完整等严重问题的地图, 以台湾省为例, 将台湾单独列出直接关乎我国领土完整, 严重违背“一个中国”原则; 漏绘或错绘, 以钓鱼岛为例, 钓鱼岛是我国不可分割的一部分, 在钓鱼岛事件爆发期, 存在明显的地图漏绘问题, 是最典型的“问题地图”事件。

(2) 危害国家利益及安全的地图。法律要求地图上不可标识军事禁区, 一旦将军事禁区绘制到地区上将暴露我国军事部署, 影响国家布防及国家安全。

(3) 有悖地图管理条例的地图也归类为“问题地图”, 通过互联网传播的不符合我国规定的地图或无资质绘制网络地图的单位, 将依法处置。

2 互联网“问题地图”监控

2.1 互联网“问题地图”监控发展史

参照上文论述, 初步认识到互联网“问题地图”的危害, 为解决“问题地图”, 我国学者在2005年开始布控地图监控系统。地图监控的目的是减少“问题地图”。起初地图监控由人力操作, 即通过群众反馈更正“问题地图”。在地图监控工作开始后, 先后经历了图片监控、POI兴趣点监控。2014年, 我国正式推行地图网络服务, 实现了

“问题地图”的网络全方位监控, 有效解决了信息泄露及机密信息交易问题。

2.2 互联网“问题地图”监控

目前, 我国互联网“问题地图”监控建立了国家主节点监控室, 实现了全国范围内的地图监控, 在原有横向拓展的基础上实现了纵向延伸。国家主节点负责国家机关网站、新闻媒体、互联网地图服务网站、中大型企业及商业网站、社会热点及国家大事等监控工作, 同时对省级节点运行情况进行实时把控。省级节点负责监控本地政府机关、新闻媒体、地图互联网网站、省内大事及热点、本地地图服务网站等。互联网“问题地图”监控部署以谁主管谁负责为原则, 各部门各司其职, 协同办公, 在自然资源系统内搭建国家级与省级节点数据连通机制, 可对违法、违规信息及时截断, 避免其进一步扩散。

3 图像识别在互联网“问题地图”监控中的应用

3.1 图像识别技术简述

图像识别技术是基于互联网衍生的识别不同目标对象的技术, 通过比对存储信息与当前信息是否一致, 实现对图像的管理。图像识别技术是基于深度学习算法的拓展应用, 包括卷积神经网络、深度置信神经网络、堆叠自动编码器三大模块。该技术适用卫星遥感、交通监测、公安监控等众多领域。

3.2 图像识别在互联网“问题地图”监控的应用

图像识别技术在“问题地图”监控的应用最显著的优点是释放人力劳动, 减轻人工检索工作量。“问题地图”监控中, 图像识别技术有两个适用场景, 首先是图片筛选, 即从大量的图片中抽取有问题的图片, 并标记为“问题图片”; 其次是发现“问题地图”图片后寻找成因, 并挖掘根源, 通过追溯IP等方式调查问题成因。

3.2.1 卷积神经网络支持下的地图图片智能识别

(1) 卷积神经网络架构

卷积神经网络可分为卷积层、池化层、激活函数三部分。首先, 卷积层主要负责处理输入的图像信息, 对数据进行特征识别, 评估地图中的各类图片是否归类“问题地图”。由于卷积核数量不同, 特征提取阶段个特征图的神经元权重有所差异。卷积神经网络在地图图片识别应用时会根据该网站潜发的风险类型进行权重分级, 如将“危害领域完整”的图片权重最高, 分析过程中更偏重此类图片。起初池化层包括最大池化和平均池化两种, 其对卷积层传递的特征采样时有一定缺陷。因此, 当前池化层处理特征信息的方式包括: 基于值、等级、概率、转换域四种。上述传递的“问题地图”图片经过卷积计算机非线性映射, 将映射后的特征送入池化层。经池化层处理将输入的特征向量进行采样, 降低其维度, 为后续评估奠定基础。常规条件下, 卷积层提取特征信息时将产生大量冗余信息, 此时特征参数维度较高, 存在传播速率慢且不准确等问题, 池化层负责对特征信息降维处理。激活函数是评估降维后的特征信息是否属于“问题地图”, 通过激活函数映射计算, 进一步强化非线性能力。互联网“问题地图”监控中,

分类数据的拟合均与线性密不可分,卷积层特征变化方式为线性变化,拟合数据的能力较弱,激活函数的作用是增强拟合的非线性。

(2) 卷积神经网络支持下的地图图片智能识别

卷积神经网络支持下的地图图片智能识别主要通过训练网络输出预测数据的方式实现。预测数据和实际参数的误差需利用损失函数计算。受网络性能影响,实际监测出的“问题地图”可能与真实参数存在误差,此时需对数据的预测值和真实值进行比较,模型最终训练结果的优劣由网络参数决定。

互联网“问题地图”监控主要应用主动学习样本扩容算法及卷积神经网络自动识别,对具有“问题地图”潜发风险的网站仅深度搜索,凭借图像识别的拟人化优势,对各类图片进行分类,并表示有悖国家法律的问题地图图片(流程图见图1)。

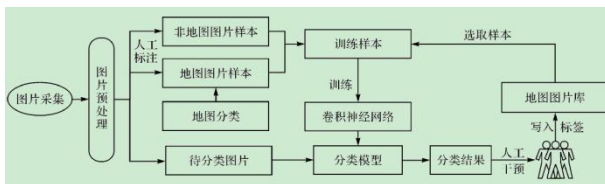


图1 基于卷积神经网络的图片识别流程图

根据图1,找到目标网站,对网站图片进行预处理,结合人工方式对地图图片和非地图图片样本进行分类。分类后的图片由卷积神经网络处理,结合内置算法建构分类模型,分析图片内容,显示分类结果,并将其纳入图片库。纳入图片库的地图图片再次回传给训练样本,存入卷积神经网络内部,丰富算法数据库,进一步提升算法精准率。上述流程多应用海量地图图片的自动识别,尚无法识别“问题地图”,对人工判定的要求较高。

目前,该方法布控于全国监控系统,对中国政府网、经济网、新浪网、中国共青团在线网进行图片搜索与识别。该算法对四个网站图片检索准确率约为80%,随着样本数量的增加,算法准确率将进一步提升。

(3) 处理方法优化

(2)中列示的内容虽然取得广泛应用,但对人工判定要求较高,从图1可见,多个流程均需要人工判定。对此,可采用自适应学习率优化算法,依照算法在学习率设定上的差异,选取不同优化目标,对特征信息进行自动存储。Adagrad提出卷积神经网络自适应图片识别方案,着重对低频与高频参数应用大小学习率的方式进行调参,可在目标数据稀疏时提高准确率,鲁棒性有所增强,缺点在于预设了全局学习率,很容易导致网络未收敛时就已经停止,同时训练梯度较大,迭代到最小学习率时,训练也会终止。由于“互联网“问题地图”监控

参考文献:

- [1] 吕非彼, 杨洁艳, 吴浩然,等. 图像识别技术在5G网络覆盖率评估中的应用探讨[J]. 邮电设计技术, 2020(6):5.
- [2] 罗曦. 浅谈深度学习在图像识别领域的应用现状与优势[J]. 科技资讯, 2020, 18(3):2.
- [3] 于乃功, 魏雅乾, 王林. 基于卷积神经网络的仿鼠脑海马结构认知地图构建方法[J]. 模式识别与人工智能, 2020, 33(1):9.
- [4] 徐振磊, 曾懿辉, 郭圣,等. 基于图像识别技术的输电线路智能监控系统应用[J]. 计算机系统应用, 2020, 29(1):6.
- [5] 张凯悦, 张鸿. 基于注意力机制网络的航运监控图像识别模型[J]. 计算机应用, 2021, 41(10):7.
- [6] 杨冬, 章耀辉, 张忠. 基于图像识别技术的输电线路智能监控系统[J]. 信息技术, 2021, 45(9):6.
- [7] 王玉茜, 张雪涛, 闫飞,等. 基于视觉的无人机大范围室外道路检测及拓扑地图构建[J]. 无人系统技术, 2021, 4(4):10.
- [8] 王威, 唐权. 一种基于多特征及BP神经网络的高分辨率遥感影像道路提取方法[J]. 现代测绘, 2020(2):3.
- [9] 武张静, 刘敏, 史禹龙,等. 语音示教+自主巡航智能垃圾分类机器人的研究与设计[J]. 科技创新与应用, 2020(32):3.
- [10] 蹇锐, 苏卫华, 张世月. 搜救环境下动态物体消除方法及其在SLAM中的应用研究[J]. 医疗卫生装备, 2020, 41(6):6.

需从全流程评估,相对“问题地图”筛选有一定缺陷。对此,基于Adagrad算法进行优化。基于Adagrad提出的算法,着眼于大小固定的项,并对所有项的均值进行近似计算,以此加快损失值下降速度,对学习率约束调整为动态边界约束,将自适应算法过渡到SGD算法,解决自适应算法及随机梯度下降算法之间的泛化问题,在释放人力劳动的同时,加快初期学习效率,为互联网“问题地图”监控提供保障。

3.2.2 “问题地图”样本库快速挖掘

“库”是互联网技术常见术语,是分析数据、存储数据、处理数据的基础。“问题地图”样本库就是汇集近年“问题地图”案例,抽取共性元素、特异元素,并在处理过程中持续累积数据,以此丰富样本库,为后续“问题地图”筛选与处理提供保障。“问题地图”样本库是以图搜图自动识别实现流程见图2。

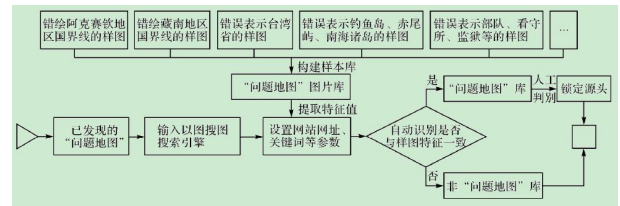


图2 基于样本库的以图搜图自动识别流程图

将阿克赛钦地区国界线、藏南地区国界线的错绘、台湾省错误标识、钓鱼岛和赤尾屿、南海诸岛的错误表示、部队及监狱错误表示等作为样本,构建样本库,即“问题地图”错误族群。由样本库提供“问题地图”图片素材。根据大数检索结果,提供已发现的“问题地图”,输入以图搜图搜索引擎,根据检索关键字或关键词自动识别图片内容,将其与样本库中的问题图片进行特征比对,比对相似的标记为“问题地图”;比对完全不同的标记为非“问题地图”。标记为“问题地图”的图片由人工再次筛选,并锁定问题来源。最后,将“问题地图”回传给样本库。该方法多用于寻找问题源头,检索效果与样本库素材丰富度相关,因此样本库越丰富筛选准确率越高。目前,该方法更多地应用在互联网“问题地图”监控部门搜索源图片。

4 结束语:

图像识别技术在互联网“问题地图”监控的应用将随着族群库丰富、算法优化等支持下愈加深入,实现空间维度、特征维度、特征传输的三方互联,让卷积层、池化层、激活函数发挥各自功用。同时,图像识别的自适应特性将在人工智能与神经网络的高速发展下更具实效,弥补人工评估工作量大、准确率低等不足。在未来互联网“问题地图”识别中,将实现多尺度特征融合,健全双线性卷积神经网络,并实现图像识别技术的集成化发展,将其与态势感知模式融合。