

每日小结

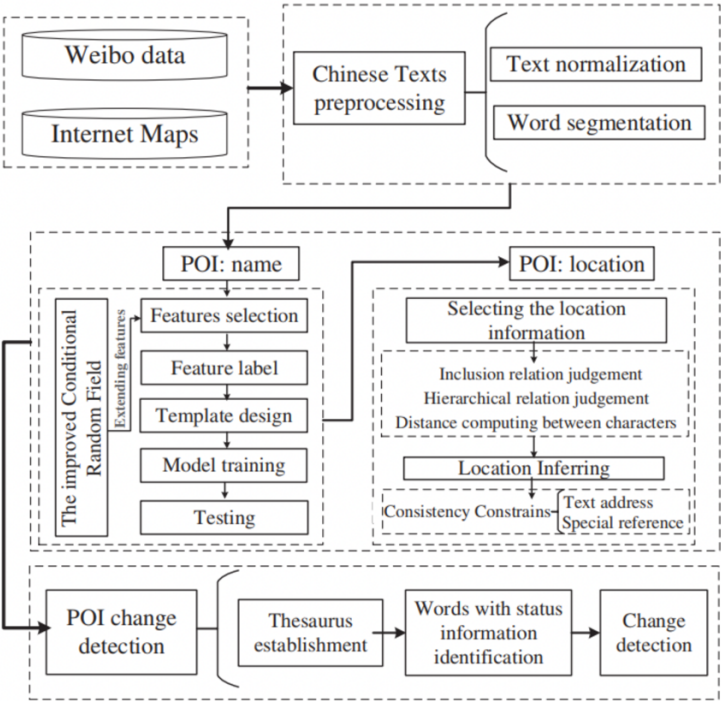
	周一	周二	周三	周四	周五
早	论文阅读，上课	上课	上课	新生讲课准备	上课
中	上课	论文阅读，上课	论文阅读	上课	上课，新生编程
晚	新生讲课准备	组会	新生编程	新生讲课准备	论文阅读

注：简单表述当前时间段工作，如看文献 1，整理数据等

科研详情

文献阅读

**文献1**  
**题目：**Extracting Disaster-Related Location Information through Social Media to Assist Remote Sensing for Disaster Analysis: The Case of the Flood Disaster in the Yangtze River Basin in China in 2020  
**作者：**Tengfei Yang, Jibo Xie, Guoqing Li, Lianchong Zhang, Naixia Mou,Huan Wang, Xiaohan Zhang, Xiaodong Wang  
**出处：**Remote Sensing  
**方法：**本研究利用社交平台文本数据与签到数据，提出了一种新的 POI 变化检测方法。该方法以国内微博社交平台提供的文本数据为信息源，通过 POI 名称识别、位置确认、变化检测，实现了大范围 POI 数据自动变化检测与更新。实验以武汉市为例，利用 2020 年获取的微博数据为信息源，对实验区域商业类 POI 进行了探测与分析。基于三种常用地图平台的验证、提供的数据和 55 个随机样本的人工实地调查，新增 POI、未改变 POI 和过期 POI 的识别准确率分别约为 100、95.8 和 91.7%。



- 启发：
1. 该研究中用到的多位置优化策略可以作为以后从社交文本数据中提取地理信息的相关工作的参考。
  2. 将 POI 融合遥感数据做土地利用分类，应该可以提取 POI 的密度及 POI 混合度等特征。但因为 POI 的更新周期慢难以提取时序特征。还需进一步阅读融合 POI 相关论文。

## 文献2

题目: Natural disasters detection in social media and satellite imagery: a survey

作者: Naina Said · Kashif Ahmad · Michael Riegler · Konstantin Pogorelov4 · Laiq Hassan1 · Nasir Ahmad1 · Nicola Conci

出处: Multimedia Tools and Applications

方法:

该综述将根据的相关多源社交媒体内容在灾害中应用的文献分为三类:

- (i) 文本语义中的灾害检测;
- (ii) 分析社交媒体与灾害有关的视觉内容 (视频、图片等);
- (iii) 卫星图像中的灾害检测。

此外, 综述还整理出了可用于目前可用于评估灾害检测框架的基准测试数据集。

Table 1. A brief summary of existing datasets.

Dataset	Types of Images	UAV imagery	Post Disaster	Resolution of Images	Classification	Semantic Segmentation	VQA
ImageNet [19]	Real-world images	No	No	average 400 × 350	✓	✗	✗
Cityscapes [48]	Real-world images	No	No	1280 × 720	✗	✓	✗
DAQUAR [51]	Real-world images	No	No	640 × 480	✗	✗	✓
COCO-QA [57]	Real-world images	No	No	640 × 480	✗	✗	✓
COCO-VQA [4]	Real world images, abstract cartoon images	No	No	640 × 480	✗	✗	✓
Visual Genome [40]	Real-world images	No	No	varies in size	✗	✗	✓
Visual7W [79]	Real-world images	No	No	varies in size	✗	✗	✓
TDIUC [37]	Real-world images	No	No	varies in size	✗	✗	✓
CLEVR [36]	Geometrical Shape	No	No	320 x 240 (in default settings)	✗	✗	✓
PATHVQA [33]	Medical Images	No	No	Varies in size	✗	✗	✓
VQA-MED [2]	Medical Images	No	No	Varies in size	✗	✗	✓
Nguyen et al. [53]	Post Disaster Images	No	Yes	Varies in size	✓	✗	✗
ABCD [25]	Pre and Post Disaster Images	No	Yes	Varies in size	✓	✗	✗
SpaceNet + Deepglobe [22]	Pre and Post Disaster Images	No	Yes	Varies in size	✗	✓	✗
Chen et al. [12]	Post Disaster Images	No	Yes	Varies in size	✗	✗	✗
OSCD [17]	Urban Change Images	No	No	Varies in size	✗	✗	✗
fMoW [14]	Pre and Post Disaster Images	No	Yes	Varies in size	✓	✗	✗
AIDER [44]	Post Disaster Images	Yes	Yes	Varies in size	✓	✗	✗
Rudner et al. [60]	Post Disaster Images	No	Yes	Varies in size	✗	✓	✗
xBD [29]	Pre and Post Disaster Images	No	Yes	Varies in size	✓	✓	✗
ISBDA [78]	Post Disaster Images	Yes	Yes	Varies in size	✗	✗	✗
FloodNet (Ours)	Post Disaster Images	Yes	Yes	4000x 3000	✓	<a href="https://github.com/nasirahmad1994/FloodNet">https://github.com/nasirahmad1994/FloodNet</a>	✓

启发:

1. 发现综述中所提到的所有相关多源社交媒体内容在灾害中应用的研究, 所使用的数据集几乎都为**自收集数据集**, 社交文本、图片、视频数据都通过爬虫等方式从公开社交平台 (推特、微博) 获取。

Table 1 Summary of some relevant works in disaster detection in Twitter text in terms of event types, modality (Single, multi-modal) of information, datasets used for the evaluations, and a brief description of the method

Refs.	Events	Modality	Dataset	Method
Shekhar and Setty [112]	Earthquake, droughts, floods and forest fires	S	Self-collected	The system checks the data for disaster distribution, Geo tagging and occurrence frequency. Location of the disaster is determined using Geo filter tag whereas distribution analysis is done using KNN.
To et al. [126]	Earthquake, flood and wildfire	S	Self-collected	A five step framework has been proposed to compare the performance of matching based and learning based approach for disaster detection. The steps include spam removal, data mapping to affected and unaffected regions, filtering of tweets, sentiment analysis and visualization.
Klein et al. [82]	Natural (flood, hurricane, fires (earthquake) and human disasters (epidemics)	S	Self-collected	Makes use of a state-of-the-art library namely Named Entity Recognizer (NER) [116], embedded with several NLP techniques, to extract emergency facts. The matchmaking approach is then used for assigning the posts to specific emergency clusters.
Truong et al. [127]	Hurricane	S	Self-collected	Relies on a Bayesian approach for the identification and classification of disaster related tweets. In contrast to [139], makes use of an effective reduced set of features consisting of only nine features. The system with reduced features set has been proved suitable for the hand held devices with less power capacity. However, a significant improvement in the performance has been reported by jointly using the newly extracted features with bag of words.
Cresci et al. [36]	Infrastructure or community damages	S	Publicly available Dataset [38]	Proposes a crisis mapping system relying on SVMs trained on a wide set of linguistic features including raw and lexical text features, morpho-syntactic features, syntactic features, lexical expansion features and sentiment analysis features [37].

2. 阅读了大量综述内研究论文, 发现仅从社交平台得到的文本、图片、视频数据等, 具有较大噪声和局限性 (例如地理标签不足和空间分布不足), 仍需要通过多种语义分析得到其地理编码、或者受灾严重地区本来就无人活动, 难以对遥感影像进行灾害监测问题具有较有效的信息补充, 因此还在寻找其他论文进一步思考。

## 文献3

题目: Urban sprawl and its impact on sustainable urban development: a combination of remote sensing and social media data

作者: Zhenfeng Shao, Neema S. Sumari, Aleksei Portnov, Fanan Ujoh, Walter Musakwa & Paulo J. Mandela

出处: Geo-spatial Information Science

## 方法:

该研究的重点是使用 Landsat 遥感数据, Twitter 平台的社交数据以及研究区域内的人口普查数据。在 2011 年至 2017 年的时间范围内, 使用该两种数据源来确定摩洛哥市政当局的城市蔓延的程度和模式。

主要框架提供了三个主要步骤: 1) 遥感数据和 Twitter 数据预处理。 2) 图像分类, 土地使用变更检测和验证。 3) 使用土地覆盖变化的分类结果确定的空间模式。在预处理中, 使用不同的分析方法对遥感和 Twitter 中提取的数据进行计算和分析, 以获得特定位置和时间城市的扩张用于估计“真实地面”, 然后开发了一个核密度映射模型分配得到扩张程度。最后, 通过将遥感结果与提取点进行比较来评估模型。预处理过程中产生的数据有助于解释主要数据以实现空间分布及其对可持续发展的影响。采用方法的详细步骤如图所述。

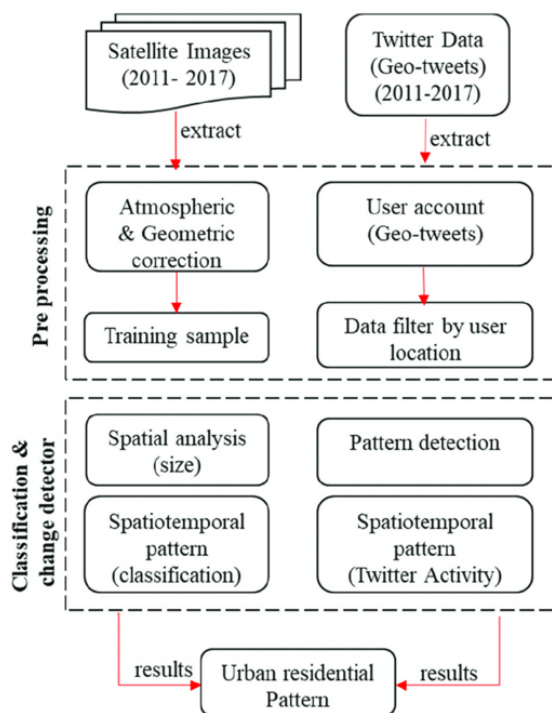


Figure 2. Overview of our research approach.

结果表明, 在研究期间, 摩洛哥城市的构建区域扩大了, 该区域蔓延与 2012 年至 2017 年之间的摩洛哥城市人口增加相对应。 Twitter 的年度增加帐户以及 2011 年至 2017 年之间的摩洛哥市政当局的 Twitter 用户的大量空间和数值增加相符合。

在时间和空间土地使用中, 使用遥感图像, 人口数据和 Twitter 使用数据的整合, 以模拟城市的模型在摩洛哥城市城市以外在半岛形式蔓延。尽管现有趋势显示农业土地覆盖范围以及生态系统服务的下降, 但该模式还意味着基础设施 (道路, 管道水域等) 和其他关键的市政服务 (固体废物管理, 污水管理, 污水处理, 污水和卫生, 安全等方面) 没有提供充分足够的服务。

## 启发:

1. 在本研究中, 将推特服务使用位置与城市贫富程度和扩张程度挂钩。因此, GPS 的精度是研究的一个重要影响因素。除此以外, 独立的社交媒体数据可能并不是城市扩张及贫富程度的可靠指标, 因为它可能只是代表互联网渗透需求, 智能手机使用之类的社交指标的增加。
2. 本研究中社交数据的数据量有限, 实验所使用在 2011 年至 2017 年之间的推特账户用户位置仅有 1579 个。
3. 本研究使用的社交数据为 2011 年至 2017 年间的推特用户账户位置, 存在一定的时序特征, 和文献 1 所使用的微博 POI 具有同样的局限性 (更新周期慢, 覆盖面不完整)。将其融合遥感数据做土地利用分类, 还需进一步阅读相关论文。

## 工作进展

- 1: 新生讲课: 两次新生讲课准备和补充修改;
- 2: 新生编程: 处理武汉出租车数据, 学习并正在尝试HGNN和HGNN+超图模型处理武汉出租车数据, 还未完成;
- 3: 了解电力数据类型, 以及电力数据融合遥感数据的不同应用场景。