

40th Asian Conference on Remote Sensing, Daejeon, Korea, 2019

Knowledge Discovery and Data Analytics from Earth Observation, Meteorological and Social Media Data



Presenter: Ilias Gialampoukidis CERTH





Overview

- \triangleright Introduction
- Change detection in EO data
- Road passability in EO data
- Concept detection in social media images
- Event detection in social media
- > Similarity fusion from multiple sources
- Clustering of social media text
- Community detection in social media accounts



F





Introduction

> What is available?

Large and daily availability of EO data (satellite imagery) and non-EO data (social media)

▷ What is our goal?

- Climate monitoring
- Land monitoring
- Emergency management

▷ How do we achieve it?

Exploiting EO data, non-EO data and their fusion to discover relevant knowledge





Water bodies Detection in EO data

- > Automatically generate water bodies masks:
 - Trained a DNN model based on Italian lakes
 - Uses Sentinel-1 VV and VH processed bands
 - Paired with Elevation information (DEM)
- ▷ Increased performance:
 - Minimizes holes (false negatives) in extended water surfaces
 - Filters out steep areas (false positives)
 - Trasimeno lake
 - (a) Sentinel-2 Natural
 - (b) Sentinel-1DNN water mask









Change Detection in EO data

- > Detects flood incidents within a time-series of optical imagery
- ▷ Two approaches:
 - DCNN model using all combinations of the differences of RGB images → Allows detection of changed differences on new unseen data
 - Remote sensing: Outlier detection on Image differencing technique between consecutive spectral indices relative to water (MNDWI)

DCNN approach:

(a) RGB non-flooded day(b) RGB flooded day(c) Differences image

Change detected





6



Road Passability in EO data

> Infers if a road is passable or non-passable due to water

- Deep learning network for determining road passability (i.e Fine-tuned pre-trained on ImageNet VGG-19)
- Splits initial image in segments and predicts the passability for each tile







Concept Detection in Social Media Images

- ▷ Extracts high-level concepts from visual low-level information to mark content as relevant or not to a target event
- Concept examples: animal, boat ship, clouds, waterscape waterfront
- Approach: \triangleright
 - Fine-tune pre-trained 22-layer GoogleNet DCNN network to recognize the 345 **TRECVID INS concepts**
 - Thresholding to keep concepts with higher probability



		Image 1:				Outdoor venicle frees River
	Fine - Tuned GoogleNet	Outdoor Nighttime Vehicle Waterscape Moonlight	0,9 0,8 0,5 0,4 0,01	Concept Thresholding	Image 1: Outdoor Nighttime Vehicle Waterscape	ge 1: door nttilme ticle terscape
		Image 2:			Moonlight	
1.97		Furniture	0,8		Image 2:	
		Chair	0,5 0,5		Indoor	
		Person	0,3		Chair	
		Laboratory	0,01		Person Doorway	
		Image 3:			Laboratory	Fother 12-year-old san saved after car sweat away by fload
		Outdoor	0,9		Image 3:	in Greene County posted by W/89Lf
3-		Daytime	0,8	V	Outdoor	Inu, 0/ Feo 2019 13:3/
		Tent	0,4		Outdoor	Outdoor Doytime Outdoor Suburban Trees
<		Trees	0,2		Tent	Eukaryatic Organism Natural-Disaster Waterscope Waterfront
Innut Images		Sunny	0,1		River	Walking Running Rocky Ground Man Made Thing
input inlages	List of Concepts and probabilities				Sunny Final List of	Sunny inal List of
					Concepts	



8



Event Detection in Social Media

- Prerequisite: daily collection of tweets that match predefined criteria, such as keywords, bounding boxes, relevant accounts
- An outlier in the fluctuation of the number of collected tweets per day is considered an event

$$|z - score| = \left| \frac{x_t - \mu_t}{\sigma_t} \right| > constant = 3 \text{ then}$$
outlier

▷ Additional metric:
$$change = \frac{x_t - \mu_t}{\mu_t} * 100$$



```
{
    "events": [
    {
        "useCase": "GreekSnow",
        "score": 15.05,
        "change": "1369%",
        "epoch_timestamp": 1557051013592
    },
    {
        "useCase": "ItalianFloods",
        "score": 3.93,
        "change": "220%",
        "epoch_timestamp": 1557051017312
    }
}
```





Similarity Fusion from Multiple Sources

- Retrieval of top-k related EO products or social media posts in response to multimodal query
- EO products & posts are equipped with several modalities (visual, textual, spatiotemporal)
- ▷ Social media posts:
 - Textual representation using word2vec
 - Visual features using DCNN-based feature
 - Visual concepts
- Unsupervised fusion of modalities realized in scalable way considering memory & computational complexity
 - Use a non-linear graph-based approach
 - Base modality keeps only top-l posts
 - Reranking of top-l posts using other modalities







Clustering of Social Media Text

- Grouping tweets according to the relevancy of their text
- Density-based clustering with LDA topic model
- Most frequent words of a textual cluster are visualized as word clouds
- Each word cloud comprises the tweets that are grouped into that cluster







Community detection in Social Media Accounts

- Detect and visualize user communities through their interaction
- ▷ Create a network of social media accounts that are linked when one mentions the other → Visualized as graph
- ▷ Modularity maximization (Louvain) community detection → Visualized as different colors of nodes
- ▷ Find the key-players in these communities → Visualized as a top-10 list



Thanks!

Any questions?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 776019

