USGS Community for Data Integration Session 1 [1]

  Submitted by erinmr on Fri, 2014-06-06 16:57   Tuesday, July 8, 2014 - 09:00 to 10:30
Event: Summer Meeting 2014 [2]
Session Type: Workshop [3]
Abstract/Agenda:
Presenter #1: Michelle Guy, USGS National Earthquake Information Center (303.273.8650; mguy@usgs.gov [4])

Title: Characterization of Earthquake Damage and Effects Using Social Media Data

Abstract:

The U.S. Geological Survey (USGS) operates a real time system that detects felt earthquakes, using only data from Twitter—a service for sending and reading public text-based messages, known as tweets. The detector algorithm scans for significant increases in tweets containing the word "earthquake" or its counterpart in several non-English languages and then sends internal alerts with the detection time, representative tweet texts, and the location of the population center where most of the tweets originated. The system has been running in real-time for over two years and finds, on average, three felt events per day with a false detection rate of 8%. The main benefit of the tweet-based detections is speed, with most detections occurring between 20 and 120 seconds after the earthquake origin time. This is considerably faster than seismic detections in poorly instrumented regions of the world. The detections have reasonable coverage of populated, earthquake prone, areas globally. The number of tweet-based detections is small compared to the number of earthquakes detected seismically, and only a rough location and qualitative assessment of shaking can be determined based on tweet data alone. However, the tweet-based detections are generally caused by widely felt events that are of more immediate interest than those with no human impact. We will provide a technical overview of the system and investigate the potential for rapid characterization of earthquake damage and effects using the 22 million "earthquake" tweets that the system has so far amassed. We will also investigate the potential use of other social media sources such as Instagram for rapid impact assessment. Additionally, this effort looks towards establishing a data feed of the tweet-based detections for sharing with collaborators and integrating derived products, such as event characterization, with seismically derived solutions for sharing with collaborators and USGS seismic monitoring and analysis systems.

From: SOCIAL MEDIA BASED EARTHQUAKE DETECTION AND CHARACTERIZATION: SYSTEM ASSESSMENT AND FUTURE DIRECTIONS. EARLE, P. S., USGS, Golden, CO, USA, pearle@usgs.gov [5]; GUY, M. R., USGS, Golden, CO, USA, mguy@usgs.gov [4]; HORVATH, S. R., USGS, Reston, VA, USA, shorvath@usgs.gov [6]; TURNER, J. S., USGS, Golden, CO, USA, jturner@usgs.gov [7]; BAUSCH, D., FEMA, Denver, CO, USA, Douglas.Bausch@dhs.gov [8]; Smoczyk, G. USGS, Golden, CO, USA, gsmoczyk@usgs.gov [9]

Presenter #2: Joseph Long, USGS St. Petersburg Coastal and Marine Science Center (727.502.8024; jwlong@usgs.gov [10])
Title: Online, on demand access to coastal digital elevation models
Abstract:

Scientists working in the coastal environment use bathymetric and topographic data to evaluate storm-induced coastal change, shoreline change, and ecosystem vulnerability. Moreover, forecast models for waves and water levels require gridded elevation surfaces that seamlessly span the land-water interface. The USGS performs copious amounts of lidar surveys and conducts geophysical surveys of the nearshore bathymetry to help address these needs. However, inadequate tools to merge these data which are collected at varying temporal and spatial resolutions, and from a variety of instrument platforms, limit the availability and applicability of the data. The focus of this work is to address the immediate need of integrating land and water-based elevation data sources so they are

readily accessible to coastal scientists and decision makers requiring a seamless data surface that spans the terrestrial-marine boundary. The two primary products will be 1) a geoprocessing service that merges and interpolates the data sources; and 2) a map-based web interface and associated information management platform that allows users to identify an area of interest to implement the interpolation algorithms, access the final gridded data derivative, and save and document the configuration parameters for future reference. The resulting products and tools could be adapted to future data sources and projects beyond the coastal environment.

Presenter #3: Matt Neilson, USGS Southeast Ecological Science Center (352.264.3519; mneilson@usgs.gov [11])
Title: NASWeb API – Web Services Access to the Nonindigenous Aquatic Species Database
Abstract:
The wealth of biodiversity data, including georeferenced records of species occurrences, held by public and private institutions represents an important resource to researchers, natural resource managers, and policy makers. Despite the growing prevalence of major distributed database efforts such as the Global Biodiversity Information Facility (GBIF) and Biodiversity Information Serving Our Nation (BISON) projects, providing and obtaining access to these biodiversity data stores represent a major challenge for both data stewards and users alike. The USGS Nonindigenous Aquatic Species (NAS) Database represents a unique biodiversity dataset containing over 150,000 georeferenced occurrence records from ~1,000 introduced aquatic species in the United States. The work proposed herein addresses access constraints to the NAS database by creating a publically-accessible Web Services API, providing an automated, machine-readable access point to NAS occurrence data. This open API (aligned with the recent focus on open access to government data) will increase exposure for this integral USGS data resource, enhance its integration with other internal and external data sources, and stimulate novel applications and analysis of NAS occurrence data.

Notes:
Opening remarks from Kevin Gallagher; USGS Associate Director for Core Science Systems

1)Michelle Guy: Characterization of Earthquake Damage and Effects Using Social Media Data

TED: Tweet Earthquake Detection & Social Media Based Event Characterization

How to detect earthquakes w/Twitter? Metadata: timestamp, lat/lon and text

Tweet for "Earthquake" is searched in several languages

Method: Collect, geocode, into database, select tweets, create timeseries, detect/locate and generate outputs (text messages, emails, etc)

Can tune out "noise". 9.5% false detection rate

System detects ~3 "felt" events/day

90% of detections occur in <2min from earthquake time

Location accuracy: 65% within 100km; 90% within 200km of event

Wide range of detected magnitudes: M1.4 to M8.2

This TED is often NEIC's first indication of a widely felt event

Detection of small events in sparsely populated or poorly instrumented parts of the world

Can tweets be used to characterize shaking?
-look at punctuation and wordiness of tweets

Must account for languages where multiple words mean earthquake. ex: Spanish language: terremoto (big) vs temblor (minor)

In testing: Event characterization
-key words, word patterns, punctuation usage
-asses density and spatial extent of early tweets
-harvest data from other social media sources like Instagram

Derive Characterization:
-Sum of weighted characterization metric and a simple symbol to represent a range/estimate

Twitter system augments but does not replace current USGS monitoring systems.
-Only provides qualitative indication of shaking level.
-Poor geocoding of many observations

The USGS TED is for internal use only.

Public can follow @USGSted for examples and alerts

2)Joseph Long: Online, on demand access to coastal digital elevation models

Motivation: seamless topographic/bathymetric DEMs that span the land-water interface

Challenges: Spatial sampling and accuracy.
-Inconsistent spatial resolution, data type and accuracy

Challenges: Temporal sampling.
-Sampling can be months to even years apart

Challenges: Aligning features
-To integrate land and nearshore data w/larger scale DEMs, myst align submerged features with sub-aerial islands measured a different times and different scales

Objective: develop web processing service that will initiate geospatial interpolation algorithms to generate coastal DEMs using user-specified parameters

Example using CoNED data: Coastal National Elevation Dataset (CoNED) for dedicated regions by compling topo and bathymetric elevation data from multiple systems. This project will use CoNED datasets as base DEMs for future updating

Web Processing Service
-Parameters defined through the user interface

Data Query:
-US Interagency Elevation Inventory: provides raw data
-NOAA Coastal DEMs: Provides background grid
-User Supplied Data!

Geospatial Interpolation Task
-Initial focus on Cartesian DEM
-Convert Matlab code to Python
-Optimization of code and reasonable computation time

Variable Smoothing Scales
-Cross-shore length scales (10s of meters) vs along-shore length scales (100s of meters)

Outputs:
-New elevation product

-Error estimates
-Output and metadata in Bathymetry Attributed Grid or netCDF
-Resulting grids catalogued in ScienceBase for future use
-User-defined parameters also catalogued for future use

Pilot study: Northern Gulf of Mexico

Several ideas provided for future expansion of the tool

3)Matt Neilson: NASWeb API - Web Services Access to the Nonindigenous Aquatic Species Database

Nonindigenous Aquatic Species Database is central repository for spatially referenced biogeographic records of introduced aquatic species from 1970-present

Terms:
Nonindigenous: any species introduced outside of its native range

Organisms tracked:
-aquatic organisms introduced to fresh, brackish and marine waters of the USA
-High-profile species

NAS Database Design
-Relational databases (SQL Server)
-9 major tables w/ 60+ fields

NAS Specimen Records capture basic info about NAS species sightings (What? Where? When?)

Database services/capabilities
-Species-specific fact sheets
-Point mapping
-Web-based queries

Query Types:
-Text-based
-Watershed
-State/County

Discussion of Problems/Limitations
-Query forms somewhat restrictive
-Cannot do user-defined polygon searches

Web Services API
-Goal: Web-services access to the NAS database
-Discussion of web services API benefits

Target Users
-Users wanting easier access to NAS data within a range of interest (temporal and spatial)
-Natural resource managers, NGOs, academia
-USFWS, USACOE

API Design
-Modeled after GBIF web services
-RESTful web service w/results via JSON

Discussion of Timeline. Public release in September

Session Leads: **Name:** Jennifer Carlino [12]

Organization(s): USGS [13]

Presenters:

Name: Michelle Guy [14]
Organization(s): USGS [13]

Name: Joseph Long [15]
Organization(s): USGS [13]

Name: Matt Neilson [16]
Organization(s): USGS [13]

Notes takers:

Name: Kyle Nelson [17]
Organization(s): University of Wisconsin Madison [18]
Email: wxkylenelson@gmail.com [19]

Creative Common License: Creative Commons Attribution 3.0 License
Teaser: Showcasing the USGS CDI projects.
Accepted:

Source URL: https://commons.esipfed.org/node/2482

Links
[1] https://commons.esipfed.org/node/2482
[2] https://commons.esipfed.org/2014SummerMeeting
[3] https://commons.esipfed.org/session-type/workshop
[4] mailto:mguy@usgs.gov
[5] mailto:pearle@usgs.gov
[6] mailto:shorvath@usgs.gov
[7] mailto:jturner@usgs.gov
[8] mailto:Douglas.Bausch@dhs.gov
[9] mailto:gsmoczyk@usgs.gov
[10] mailto:jwlong@usgs.gov
[11] mailto:mneilson@usgs.gov
[12] https://commons.esipfed.org/node/2483
[13] https://commons.esipfed.org/taxonomy/term/768
[14] https://commons.esipfed.org/node/2610
[15] https://commons.esipfed.org/node/2611
[16] https://commons.esipfed.org/node/2612
[17] https://commons.esipfed.org/node/1936
[18] https://commons.esipfed.org/taxonomy/term/222
[19] mailto:wxkylenelson@gmail.com