Pipeline Monitoring

Employing Geospatial Big Data in ArcGIS
Pipeline Monitoring Employing Geospatial Big Data in ArcGIS

A treasure trove of new data offers the potential to identify leading indicators of human activity near critical energy assets. New breakthroughs in algorithmic feature extraction and machine learning can bring greater awareness of significant changes around expensive oil and gas infrastructure, enabling oil and gas companies to mitigate risks and reduce costs.

Harsh Govind, Manager of Strategic Products
John-Isaac Clark, Director of Product Management
Overview

What problem are we trying to solve?
• Can we effectively leverage machine learning algorithms on raster data along with open social vector data to identify changes along pipeline routes better and more efficiently than typical methods?

What advantage did GIS provide us in solving it?
• Traditional basic spatial operators and functionality within ESRI tools, combined with cloud-based top-level processing at scale = best of both worlds!

What types of GIS Analysis did we use?
• Intersection queries, metadata filtering, area reduction, and aggregation

What were our findings/results?
• Integrating access to these capabilities within ArcMap around corporate GIS assets will enable very exciting new methods for Pipeline Monitoring (and similar) applications [site selection, R-o-W assessment,...]

What can we learn from this?
• Making advances in remote sensing, “big” vector data, and machine learning that are easily consumable within every-day tools can improve outcomes by increasing efficiency.
Monitoring dynamic changes along pipeline infrastructure is complex, costly, and time consuming.

Automated methods to narrow the focus of field/inspection staff on areas that warrant investigation = opportunity for efficiency.

Using on-demand raster analysis of any new satellite data collected over infrastructure has historically been impossible or too expensive.

Aggregating, collecting, and managing all the vector open data sources that could be used in this endeavor is a large undertaking for any GIS team.

But if you COULD do the above, GIS users and their stakeholders would greatly benefit.

Problem Statement
Vector Big Data Problem

OpenStreetMap

Report run at Apr, 27th 2016 1:03 am CET
Statistics of the free wiki world map (OpenStreetMap.org) created In Central European Time (CET)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of OSM Nodes in the database</td>
<td>3338646203</td>
</tr>
<tr>
<td>Number of members who are the last modifier of at least one node</td>
<td>433616</td>
</tr>
<tr>
<td>Max. OSM Node ID</td>
<td>4149979395</td>
</tr>
<tr>
<td>Yesterday's number of created nodes</td>
<td>1892397</td>
</tr>
<tr>
<td>Yesterday's number of modified nodes</td>
<td>270919</td>
</tr>
<tr>
<td>Yesterday's number of deleted nodes</td>
<td>341569</td>
</tr>
</tbody>
</table>

OpenStreetMap stats report run at 2016-04-27

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of users</td>
<td>2618722</td>
</tr>
<tr>
<td>Number of uploaded GPS points</td>
<td>5171741556</td>
</tr>
<tr>
<td>Number of nodes</td>
<td>3336836604</td>
</tr>
<tr>
<td>Number of ways</td>
<td>344702535</td>
</tr>
<tr>
<td>Number of relations</td>
<td>4178242</td>
</tr>
</tbody>
</table>

Twitter

Just Twitter (all time)
2,664,928,875,000,000 (2.6 Trillion)

Just OSM (all time)
3,336,646,203 (3.3 Billion)

Every second, on average, around 6,000 tweets are tweeted on Twitter (visualize them here), which corresponds to over 350,000 tweets sent per minute, 500 million tweets per day and around 200 billion tweets per year. The chart below shows the number of tweets per day throughout Twitter's history.
Raster Big Data Problem

DigitalGlobe Satellite Data Alone...

3,500,000 km² collected
13,200,000,000,000 pixels
~70 Terabytes

EVERY DAY

90 Petabyte Archive
But it CAN be processed

Crowd-sourced Analysis

Deep Learning (Machine Learning/AI)


Around my pipeline, show me...

New Construction

Changes in soil or land use

New roads

New Open Streetmap Features

New indicators of human activity (Twitter)
DigitalGlobe recently released a capability which allows analysis of raster and vector data at scale, and on demand, on DigitalGlobe (and other’s) content via cloud services and RESTful web service APIs.

We wanted to explore using existing customer data within ArcMap (pipelines/buffers/ROWs/etc) to leverage these cloud-based capabilities in addressing this use case. In order to do this, we identified the following requirements for an ArcMap add-in.

Given selected vectors or buffers that exist in ArcMap...

• Let me select the type of changes I want to detect within past or future Satellite imagery
  - Building footprints, new roads, human activity, change in land use, etc.
• Let me select the type of changes I want to detect within recent or future vector data sources
  - Increase in human activity (Tweets near assets), new OpenStreetMap features near my assets, etc.
• Run my selections every day/week/month/whenever new imagery or vector data is available!

When you get results I care about....

• Let me easily see “where” there is an increase in change and what the change is/may be directly in ArcMap.
• Give me an easy way to show/share that with my organization so I can take action/inform others.
But we needed a connection!
Solution - connect ArcMap to Cloud Geobigdata!

Comparing Open Street Map Vector updates (RED) to road feature extraction (BLUE) from most recent satellite imagery (ground truthing).
Demonstration – Inspect & Compare

- Identify changes, events, and activities manually across assets
- Past activity history
- New updates on activity
- Back-correlate to prior events
- Other types of analysis
Demonstration - Share

• How do we share with other organizational stakeholders?

• How can organizational users request information that we can use within GIS team?

• Can other organizational elements benefit from this capability?
Findings

- We were able to identify multiple types of potential benefits for this use case.
- We are exploring making this an automated process versus manually performed.
- Better to deliver “results” or “insights” without the noise (WHERE is there change? WHAT is the change?)
- Allow analyst to view/see pixels and output vectors used in the analysis for high-probability interest areas (man-in-loop validation) as WMS.
- What if we add aerial/UAS data that customers may already have?
- Can we integrate classification determination as well?
Our Vision

• At DigitalGlobe, we see a future in which operators of pipelines and other infrastructure can be kept dynamically aware of physical changes to and nearby their assets

• Adopters of these new technologies will enjoy significant cost savings in monitoring, analysis and compliance reporting relative to those using traditional methods

• The marriage of machine learning and human verification can be an incredibly cost effective way to monitor change at scale on an ongoing basis

• We’re engaged in robust dialogue with our clients and industry partners about these technologies and their numerous use cases in real-world applications, and welcome all feedback and suggestions
Q&A

• Open Q&A Session
Backup