



Using AI/ML Approaches to Support Data Analysis Process Improvement

07/09/20

Austin Ruth

Introduction

- Research Engineer I
- With GTRI for 5 years (including 2 co-op years)
- Primary work has been Systems Engineering and Data Analysis
- Pursuing Masters of Computer Science and Interactive Intelligence (Fall 2020)

Current Process



Post Event Review



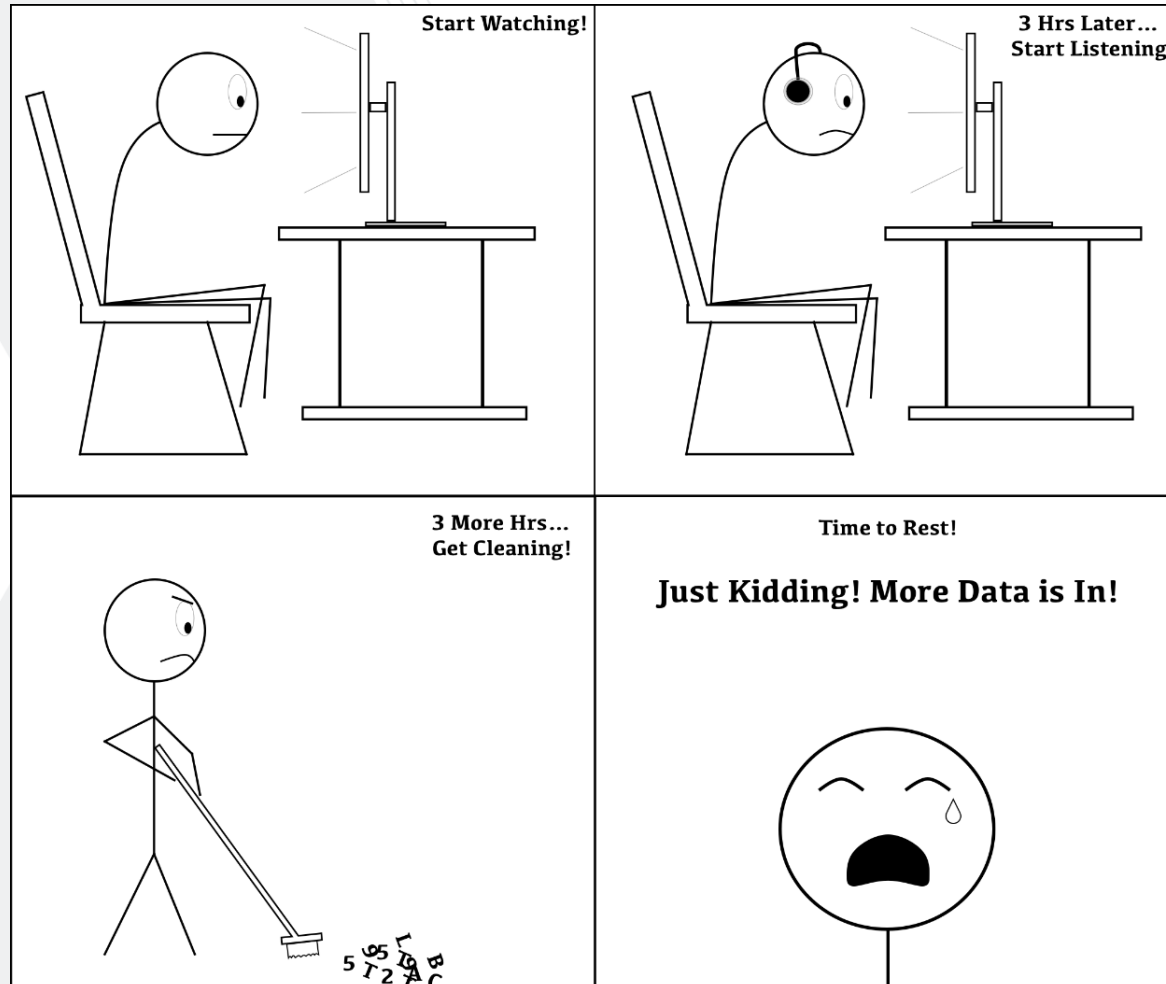
- Video and data review to answer immediate sponsor concerns or questions
- Review errors and possible fixes for the next event
- No capability for quick analysis or reporting to support the next event

What is the Problem?

- Large backlog of unprocessed data
- Constant stream of new data being added to the backlog
- Recent sponsor engagement revealed only 20% of data was being processed



What is the bottleneck?



- Data analysis generally is a labor intensive task
- Tasks include:
 - Video Review
 - Audio Review
 - Data preprocessing and culling to ensure good data
 - Data management (Where does this go?)
 - Data Correlation (What matches what?)
- For humans these tasks can take days to weeks!

Industry Techniques to Automate Analysis

The ability to quickly, accurately and intelligently process large sets of data makes current operations more effective and efficient.

Tools are available that can help decrease manual labor, and increase accuracy and efficiency

- Artificial Intelligence (AI)
- Machine Learning (ML)
- Distributed Processing

Techniques are constantly being developed by the open source community and in industry



Let's Automate

- Combine industry advanced computational techniques with distributed processing to automate data analysis.
- Our Solution: Automated Test & Logistics Analysis Support (ATLAS) framework
 - Uses Artificial Intelligence (AI) to automate labor intensive tasking
 - Distributes these tasks to decrease processing time
 - Can implement existing techniques into data flow easily



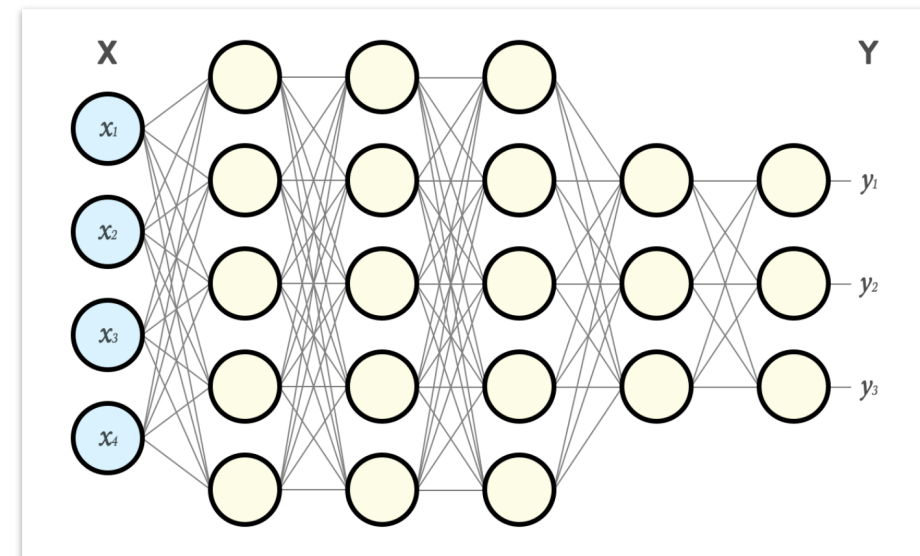
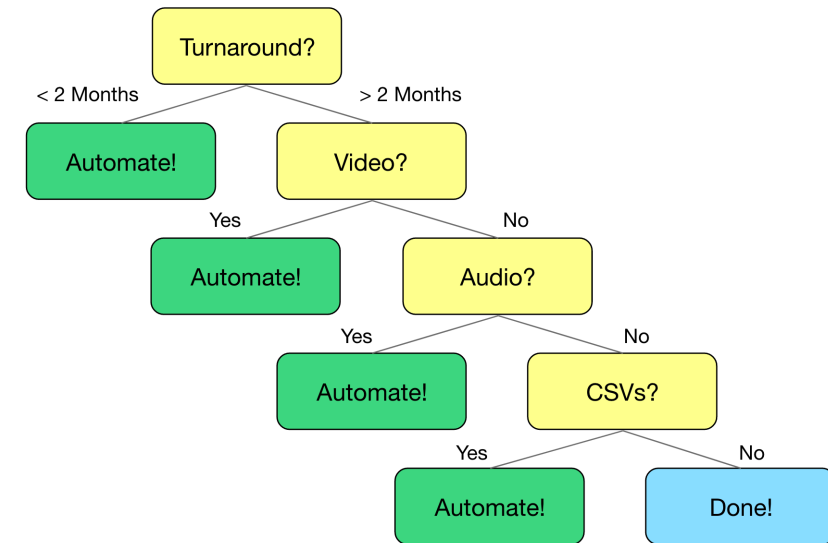
Speed of Automated Analysis

2 ½ Hours → 24 Minutes

Machine Learning in Data Analysis

- With large amounts of data, even simple tasks need automation
- Training of basic techniques like Decision Trees or Random Forests can decrease time needed to gather information.
- Advanced techniques such as Deep Neural Networks, can help model systems

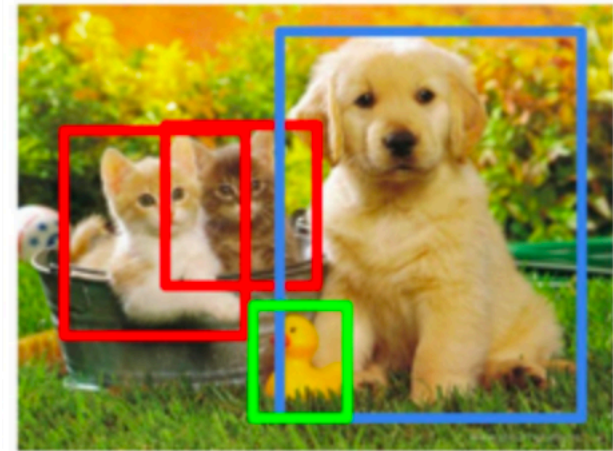
Should You Automate Data Analysis



Computer Vision: Object Detection & Tracking

- Computer Vision (CV) is an excellent technique to decrease needed manual labor
- Object Detection allows for multi-class detection of items in images and videos.
- Problems:
 - Camera Artifacts
 - Changing Video from events
 - Training is needed

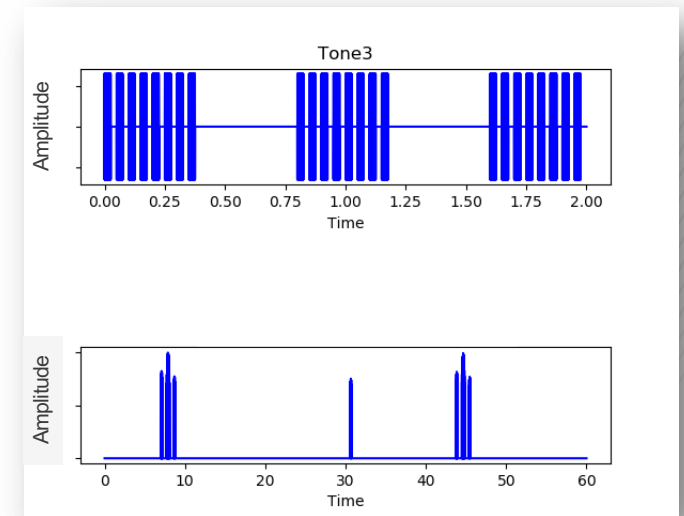
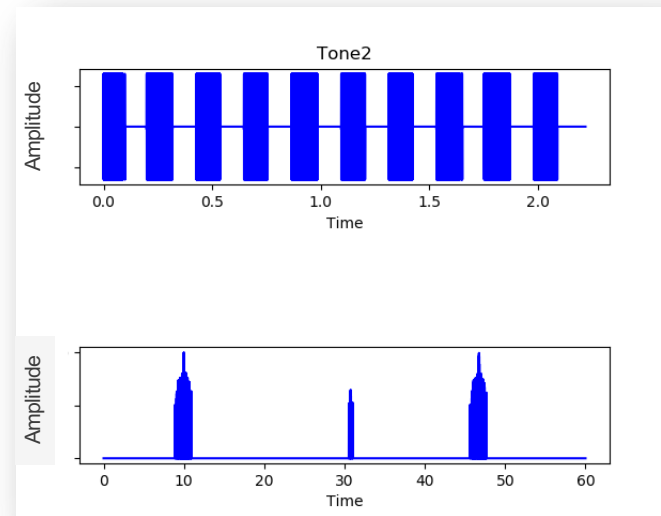
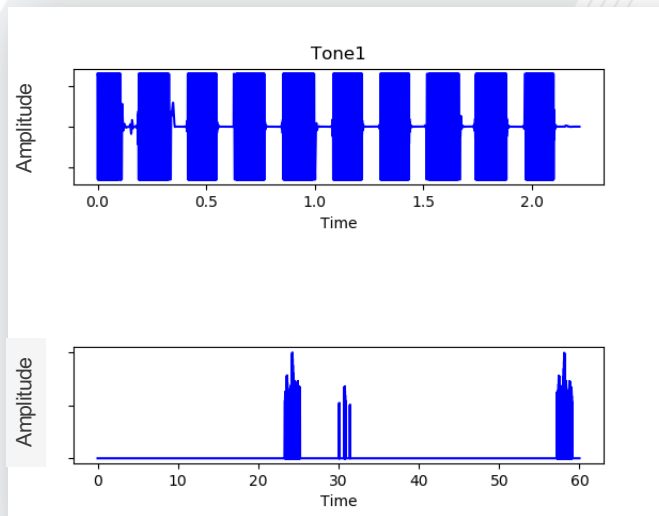
Object Detection



CAT, DOG, DUCK

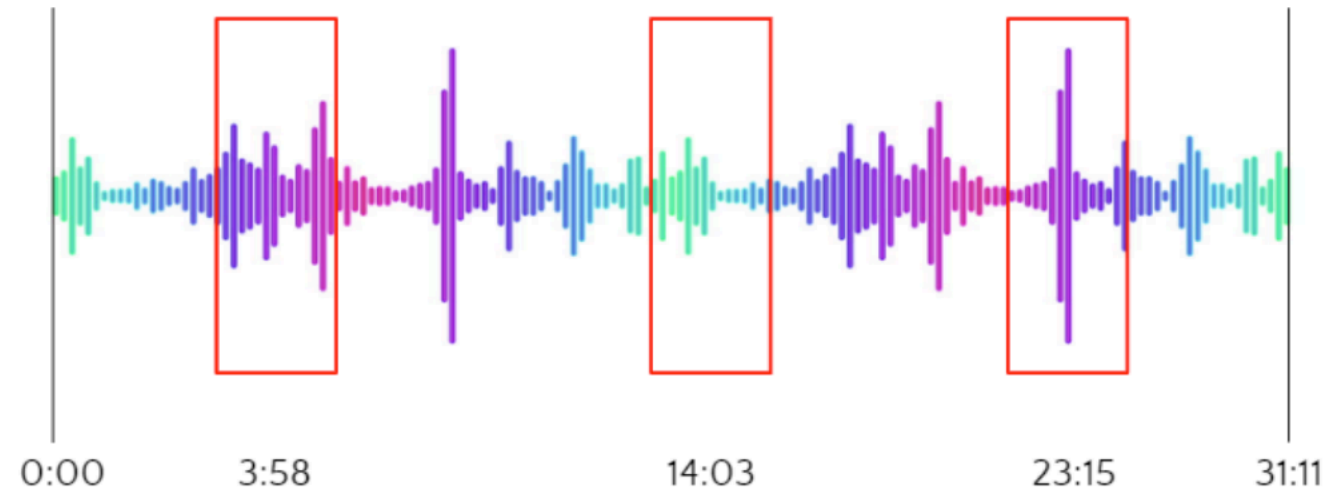
Audio Processing

- Audio tones can be difficult to log manually because of noise and timing
- Current detection is done using convolution algorithms to detect tones from noisy input audio



Automatic Speech Recognition

- Only a handful of words of interest (WOI)
- Isolate occurrences of WOI
- Log timestamps and cross-correlate
- Technique:
 - Sliding window methodology
 - Picovoice Porcupine
 - Wake Word Detection



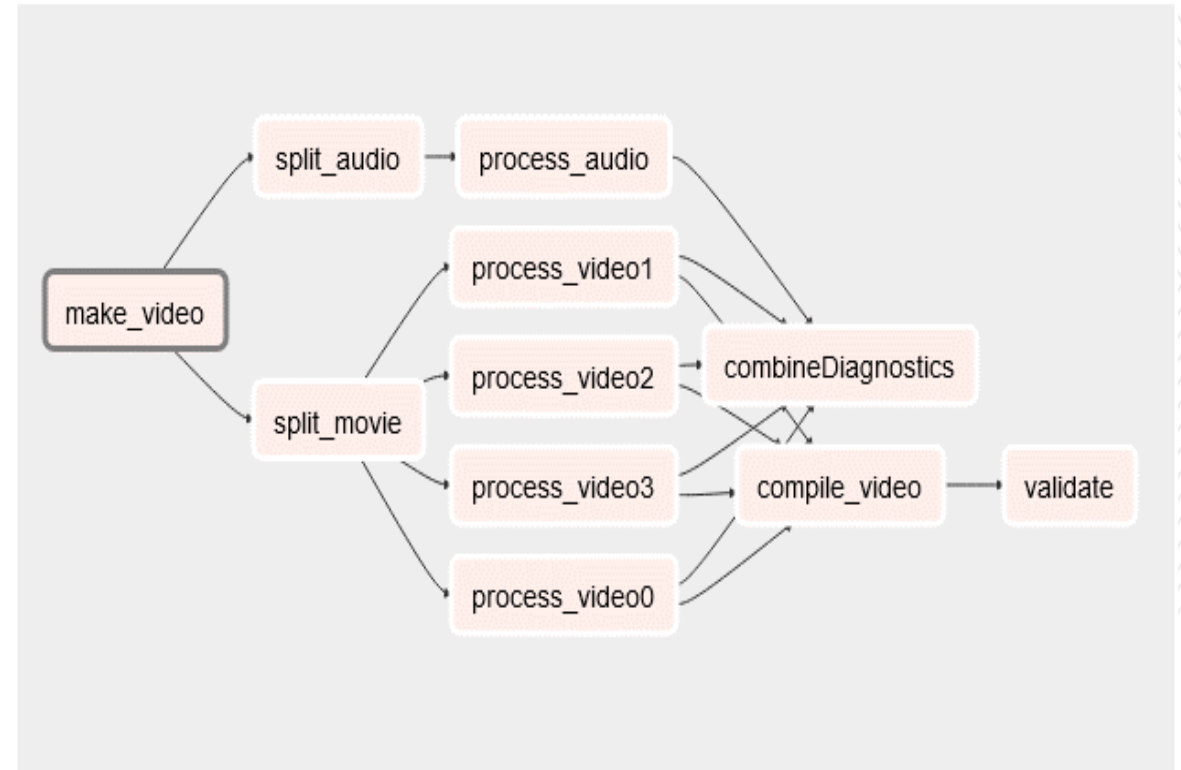
Distributed Processing Backend

ATLAS runs on a workflow scheduler, Airflow

Airflow triggers to manage analysis based on data availability

- Triggers also increase efficiency by allowing the system to continue working without human intervention

Parallel processing allows for faster analysis along with inputs of multiple data sources



Questions?