Computer Learning Algorithm for Records Evaluation

NIST STANDARDS REQUIREMENTS GATHERING WORKSHOP FOR NATURAL LANGUAGE ANALYSIS
May 21, 2019

ERDC Information Technology Laboratory: Maria Seale, Alicia Ruvinsky, Joshua Church, Grace Nabholz, Brandon Hansen
CCDC Aviation and Missile Center: Nathan Rigoni
Topics

1. Logbook Labeling
2. CLARE

Natural Language Processing
Machine Learning

3. Next Steps
4. Questions
Logbook Labeling Problem

Maintenance Logs

Ram Metrics

Logbook Labels
~10% of Logs

- Maintenance Cause
- Maintenance Type
- Component Label

Supervised Learning

Missing Labels
~ 90% of data
Logbook Labeling

Manual Method

Convert aviation maintenance records to engineering reliability data

Records

> 40M exist

Reliability data

~ 4M scored

Maintenance Events

Logbook Reports

Manual Scoring

ONLY

10% scored

> 4M scored

~ 4M scored
Logbook Labeling
Automated Method

- Reduce burden on analysts
- Enable 100% of logbook data to be used for analysis
- Increase analyst-scored data to 25%
- Provide machine-labeled data for remaining 75%
CLARE
Computer Learning Algorithm for Records Evaluation

- Operational on 3 platforms for 10 labels
- > 90% per record accuracy
Feature Selection

- Based on SME guidance
- Correlation analysis reduced original feature set

NLP

- Two fields are free-form text
- Both are important to scoring logbook data
- Machine Learning algorithms can’t use text in its original form
- Word2Vec produces numeric vectors that represent the text
CLARE
Label Dependencies

Predicting labels individually revealed dependencies

Learning Using Privileged Information (LUPI) model

The cause of maintenance helps predict the type of maintenance, which helps predict the component involved

Maintenance Cause

Maintenance Type

Component Label

Over 1200 unique labels

Maintenance strategies
Corrective maintenance
Preventive maintenance
Condition based maintenance
Scheduled maintenance

Component
CLARE

Machine Learning Techniques

Distributed Random Forest (DRF) used for label predictions

- Classification and regression
- Good for large, complex data
- Reduces overfitting
- Computationally simple
- Easily distributable
- Average prediction over all trees creates final prediction

DRF label predictions used in LUPI strategy to produce final results
Next Steps

CLARE and its enabling technologies will allow Army maintenance data to become a reliable, significant factor in providing guidance for increasing RAM of Army platforms.

- Bridge multiple maintenance data sets
- Correlate logs with sensor data
- Develop cross-service capabilities
- Generalize to other platforms
Thank you!

Questions?