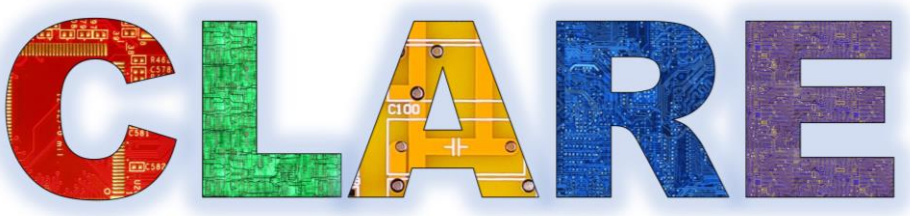


Computer Learning Algorithm for Records Evaluation

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

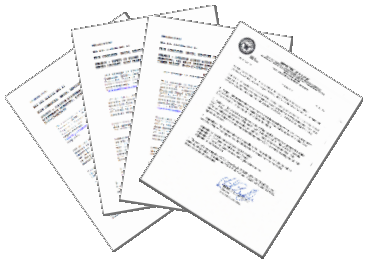
Supervised Learning

Missing Labels ~90% of data

Logbook Labeling

Manual Method

Convert *aviation maintenance records* to *engineering reliability data*



into

Records

> 40M exist

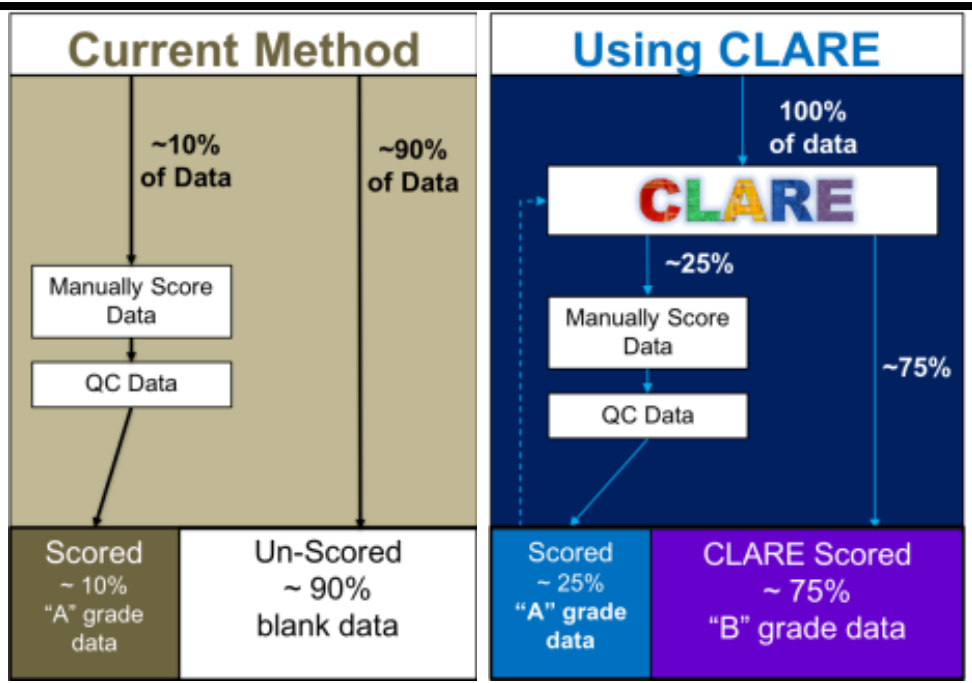
Reliability data

~ 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%**

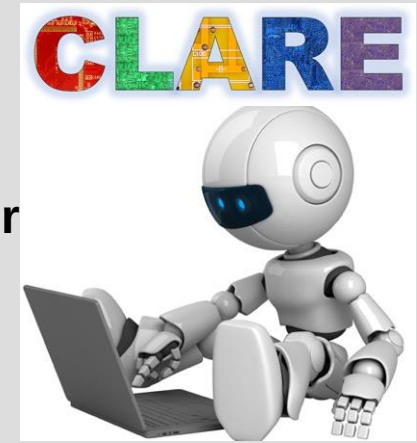


DISTRIBUTION A: Approved for public release

CLARE

Computer Learning Algorithm for Records Evaluation

Computer
Learning
Algorithm for
Records
Evaluation



Natural Language Text Fields **Labels to predict**

NARR	CORR_NARR
WATER LEAK CHECK REQUIRED FOR REPLACEME...	INSP OK
OPERATIONAL CHECK OF WINDSHIELD ANTI-ICE ...	MOC COMPLETE
PMD	INSP OK
2ND 9-11 HR TQ CHECK REQ ON RED M/R DAM...	COMPLETED PASSED AT 1320 IN LBS
INSP A020 - PERFORM M/R SPINDLE LUGS, M/R ...	INSP OK COMPLETED WITH PMD
PMD	COMPLETED
PMD	PMD COMPLETED 4MAY SEE PMD E...
INSP A100 - #1 ENGINE HISTORY RECORDER REA...	COMPLETED
INSP A101 - #2 ENGINE HISTORY RECORDER REA...	COMPLETED
MED #2 FAILCAUTION LIGHT ILLUMINATED 3 TI	CHECK FOUND OK COULD NOT DU

SCD1	SCD2	RFG
C	N	02A01A
C	N	02A01A
S	N	02_PMD
C	N	05A01Z01
S	N	05A01I
S	N	02_PMD
S	N	02_PMD
S	N	04A068
S	N	04A068
U	U	08B17
U	H	04A06

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

DISTRIBUTION A: Approved for public release

CLARE

Feature Selection and Natural Language Processing

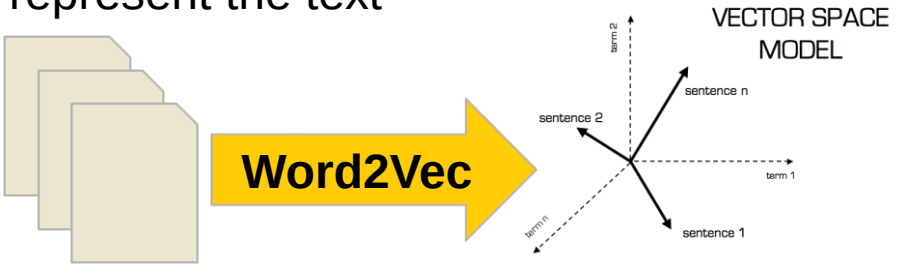
Feature Selection

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

1	-0.08039	-0.08039	-0.01862	0.016829	0.017333
-0.08039	1	0.999995	0.822677	0.026849	0.021802
-0.08039	0.999995	1	0.822695	0.026963	0.021703
-0.01862	0.822677	0.822695	1	0.023438	0.022398
0.016829	0.026849	0.026963	0.023438	1	-0.00712
0.017333	0.021802	0.021703	0.022398	-0.00712	1
0.106338	-0.04879	-0.04869	-0.05827	-0.07077	-0.10712
0.106336	-0.04877	-0.04867	-0.05825	-0.07075	-0.10713
0.024497	0.040425	0.04045	0.029496	0.150539	0.019357

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



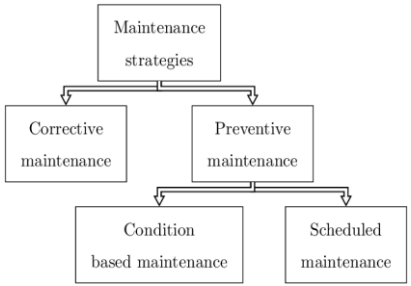
Logbook text

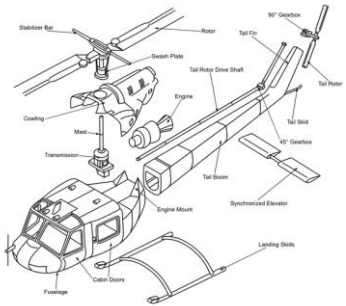
CLARE

Label Dependencies

Maintenance Cause

Maintenance Type

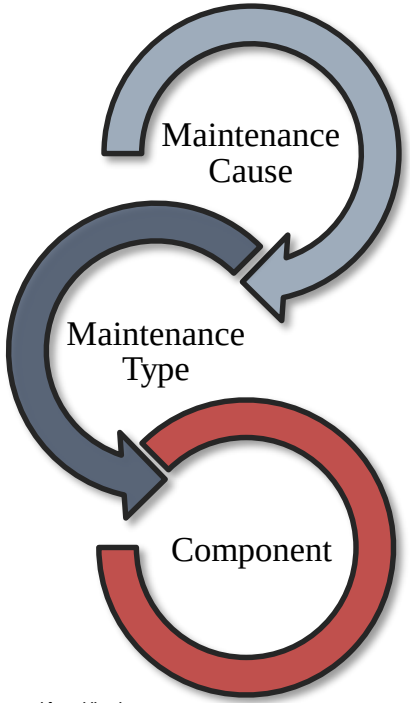




Component Label

Over 1200 unique labels

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

CLARE

Machine Learning Techniques

Distributed Random Forest (DRF) used for label predictions

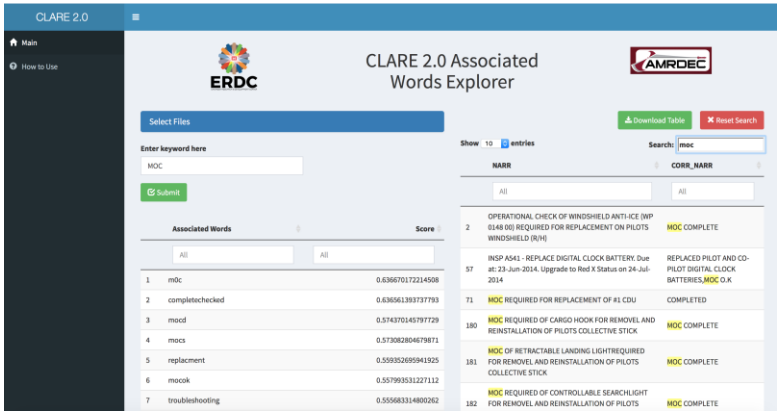
- Classification and regression
- Good for large, complex data
- Reduces overfitting
- Computationally simple
- Easily distributable
- Average prediction over all trees creates final prediction

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?

DISTRIBUTION A: Approved for public release