Automatic log analysis with NLP for the CMS workflow handling

CHEP 2019

Daniel Robert Abercrombie¹, Hamed Bakhshiansohi², Sharad Agarwal³, Jennifer Adelman-McCarthy⁴, Andres Vargas Hernandez⁵, Weinan Si⁶, Lukas Layer⁷, Jean-Roch Vlimant⁸

on behalf of the CMS collaboration

Massachusetts Institute of Technology¹, DESY², CERN³, FNAL⁴, Catholic University of America⁵, University of California Riverside⁶, Università e sezione INFN di Napoli⁷, California Institute of Technology⁸
Automatization of failing workflow handling

- Central MC production in CMS utilizes the **LHC computing grid**

- Thousands of **workflow tasks** each with thousands of jobs on more than 100 sites worldwide

- A **certain rate fails** and has to be handled manually by **Computing Operators**: e.g. resubmit, kill, change memory, change splitting, etc.

- **Common errors** are missing/corrupt input files, high memory usage, etc.

Operational Intelligence: Automatize the failure handling using Machine Learning - CMS Tools & Integration group
Strategy and Dataset

- Information of ~ **33,000 failing workflow tasks** collected since 2017

- **Actions** of the operators stored

- For each **task** we know on which **sites** how many times an **error code** was thrown

  - Build a sparse matrix for each workflow

  - Goal: Predict the operator’s action

- **Challenges:** small data, sparse input, class imbalance

- **Tried:** Feed-Forward NN, CNN, Embedding Models

- **Similar results** for different models

  - Idea for further improvement:
    - Addition of Error Logs using Natural Language Processing
# Pipeline

## WTC Console $\leftrightarrow$ Operator

### Store actions

<table>
<thead>
<tr>
<th>task_name</th>
<th>errors</th>
<th>sites</th>
<th>site_state</th>
<th>count</th>
<th>action</th>
<th>splitting</th>
<th>memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>/pdmvserv_1...</td>
<td>[82, -1]</td>
<td>[T2_US,Wisc...]</td>
<td>[good, good]</td>
<td>[1, 1]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>/pdmvserv_1...</td>
<td>[80, 64, 65]</td>
<td>[T2_CH,CERN,...]</td>
<td>[good, bad,...]</td>
<td>[57, 2, 1,...]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>/pdmvserv_1...</td>
<td>[-1, 71104]</td>
<td>[T2_UK,Lond,...]</td>
<td>[bad, bad,...]</td>
<td>[1, 5, 1]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>/pdmvserv_1...</td>
<td>[1, 92]</td>
<td>[T1_IT,CNAF,...]</td>
<td>[bad, good,...]</td>
<td>[1, 1, 1, 1,...]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>/pdmvserv_1...</td>
<td>[-1]</td>
<td>[T2_US,Cat,...]</td>
<td>[good]</td>
<td>[1]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>
 Pipeline

WTC Console  Operator

Store actions

Errors + Sites

<table>
<thead>
<tr>
<th>task_name</th>
<th>errors</th>
<th>sites</th>
<th>site_state</th>
<th>count</th>
<th>action</th>
<th>splitting</th>
<th>memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 /pdmvserv_1...</td>
<td>92, 1</td>
<td>[T2_US_Wisc...</td>
<td>[good, good]</td>
<td>[1, 1]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>1 /vimant_task...</td>
<td>50964, 85,...</td>
<td>[T2_CH_CERN...</td>
<td>[good, bad,...</td>
<td>[57, 3, 1,...</td>
<td>acde</td>
<td>10x</td>
<td>None</td>
</tr>
<tr>
<td>2 /pdmvserv_1...</td>
<td>-1, 71104,...</td>
<td>[T2_UK_Lond...</td>
<td>[bad, bad,...</td>
<td>[1, 15, 1]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>3 /pdmvserv_1...</td>
<td>-1, 92, -1...</td>
<td>[T1_IT_CNAF...</td>
<td>[bad, good,...</td>
<td>[1, 1, 1, 1,...</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>4 /prabeo_R...</td>
<td>[-1]</td>
<td>[T2_US_Cat...</td>
<td>[good]</td>
<td>[1]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

Counts

Actions

CERN SWAN
Platform for interactive analysis in the cloud

Large Scale Analytics

Filter + Map

Error log

<table>
<thead>
<tr>
<th>task_name</th>
<th>error</th>
<th>site</th>
<th>error_msg</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 /pdmvserv_task_HIG-Ru...</td>
<td>8001</td>
<td>T2 DE_RWTH</td>
<td>Failed to find a step...</td>
</tr>
<tr>
<td>1 /pdmvserv_task_HIG-Ru...</td>
<td>50660</td>
<td>T2 US Nebraska</td>
<td>Error in CMSSW step c...</td>
</tr>
<tr>
<td>2 /vimant_task_TOP-Run...</td>
<td>84</td>
<td>T2 CH CERN_HLT</td>
<td>Adding last 25 line...</td>
</tr>
<tr>
<td>3 /pdmvserv_task_HIG-Ru...</td>
<td>50660</td>
<td>T1 RU JINR</td>
<td>Error in CMSSW step c...</td>
</tr>
<tr>
<td>4 /pdmvserv_task_HIG-Ru...</td>
<td>50660</td>
<td>T2 CH CERN_HLT</td>
<td>Error in CMSSW step c...</td>
</tr>
</tbody>
</table>

Lukas Layer | CHEP 2019
**Pipeline**

**WTC Console**  
**Operator**

1. **Store actions**
2. **Counts**
3. **Actions**

<table>
<thead>
<tr>
<th>task_name</th>
<th>errors</th>
<th>sites</th>
<th>site_state</th>
<th>count</th>
<th>action</th>
<th>splitting</th>
<th>memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 /pdmvserv_task_R...</td>
<td>92, -1</td>
<td>T2_US_Wisc...</td>
<td>good, good</td>
<td>[1, 1]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>1 /litmap_task...</td>
<td>50964, 85...</td>
<td>T2_CH_CERN...</td>
<td>good, bad...</td>
<td>[57, 3, 1,...]</td>
<td>acdc</td>
<td>10x</td>
<td>None</td>
</tr>
<tr>
<td>2 /pdmvserv_task...</td>
<td>-1, 71104...</td>
<td>T2_UK_Lond...</td>
<td>bad, bad...</td>
<td>[1, 15, 1]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>3 /pdmvserv_task...</td>
<td>-1, 92, -1...</td>
<td>T1_IT_CNAF...</td>
<td>bad, good...</td>
<td>[1, 1, 1, 1,...]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>4 /pmbku_R...</td>
<td>-1</td>
<td>T2_US_Cali...</td>
<td>good</td>
<td>[1]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

**Join**  
**Aggregate**  
**Preprocess**

**CERN SWAN**  
Platform for interactive analysis in the cloud

**Large Scale Analytics**

**Filter + Map**  
**Error log**

<table>
<thead>
<tr>
<th>task_name</th>
<th>error</th>
<th>site</th>
<th>site_state</th>
<th>count</th>
<th>msg_encoded</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 /pdmvserv_task_HIG-Ru...</td>
<td>8001</td>
<td>T2_DE_RWTH</td>
<td>Unknown, T...</td>
<td>102, 1</td>
<td>nan, nan</td>
</tr>
<tr>
<td>1 /pdmvserv_task_HIG-Ru...</td>
<td>50660</td>
<td>T2_US_Nebraska</td>
<td>bad, bad, ...</td>
<td>1, 1, 2,...</td>
<td>nan, [102, ...]</td>
</tr>
<tr>
<td>2 /litmap_task_TOP-Ru...</td>
<td>84</td>
<td>T2_CH_CERN_HLT</td>
<td>good, good...</td>
<td>1, 1, 1, 1,...</td>
<td>[35, 12, 1,...]</td>
</tr>
<tr>
<td>3 /pdmvserv_task_HIG-Ru...</td>
<td>50660</td>
<td>T1_RU_JINR</td>
<td>bad, bad...</td>
<td>1, 1, 1, 1,...</td>
<td>[ nan, nan, ...]</td>
</tr>
<tr>
<td>4 /pdmvserv_task_HIG-Ru...</td>
<td>50660</td>
<td>T2_CH_CERN_HLT</td>
<td>bad, bad...</td>
<td>1, 1, 1, 1,...</td>
<td>[35, 12, 1,...]</td>
</tr>
</tbody>
</table>

**Actions**

**Counts**

**Preprocess**

**Operator**  
**WTC Console**

**CERN SWAN**  
Platform for interactive analysis in the cloud

**Large Scale Analytics**

**Filter + Map**  
**Error log**
Pipeline

WTC Console

Operator

Store actions

Counts

Actions

<table>
<thead>
<tr>
<th>task_name</th>
<th>errors</th>
<th>sites</th>
<th>site_state</th>
<th>count</th>
<th>action</th>
<th>splitting</th>
<th>memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 /pdmserv_task_hi...</td>
<td>[92, -1]</td>
<td>[T2_US_Wisc...</td>
<td>[good, good]</td>
<td>[1, 1]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>1 /mlmt_task...</td>
<td>[50964, 85,...</td>
<td>[T2_CH_CERN...</td>
<td>[good, bad,...</td>
<td>[57, 3, 1, ...</td>
<td>acdc</td>
<td>10x</td>
<td>None</td>
</tr>
<tr>
<td>2 /pdmserv_task...</td>
<td>[-1, 71104,...</td>
<td>[T2_US_Lond...</td>
<td>[bad, bad,...</td>
<td>[1, 10, 1]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>3 /pdmserv_task...</td>
<td>[-1, 92, -1,...</td>
<td>[T1_IT_CNAF...</td>
<td>[bad, good,...</td>
<td>[1, 1, 1, 1,...</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>4 /mabukuro_R...</td>
<td>[-1]</td>
<td>[T2_US_Cati...</td>
<td>[good]</td>
<td>[1]</td>
<td>acdc</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

Error log

<table>
<thead>
<tr>
<th>task_name</th>
<th>error</th>
<th>site</th>
<th>msg_encoded</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 /pdmserv_task_hi...</td>
<td>8001</td>
<td>T2_DE_RWTH</td>
<td>Failed to find a step...</td>
</tr>
<tr>
<td>1 /pdmserv_task_hi...</td>
<td>50660</td>
<td>T2_US_Nebraska</td>
<td>Error in CMSSW step c...</td>
</tr>
<tr>
<td>2 /mlmt_task...</td>
<td>84</td>
<td>T2_CH_CERN_HILT</td>
<td>Adding last 25 line...</td>
</tr>
<tr>
<td>3 /pdmserv_task_hi...</td>
<td>50660</td>
<td>T1_RU_JINR</td>
<td>Error in CMSSW step c...</td>
</tr>
<tr>
<td>4 /pdmserv_task_hi...</td>
<td>50660</td>
<td>T2_CH_CERN_HILT</td>
<td>Error in CMSSW step c...</td>
</tr>
</tbody>
</table>

CERN SWAN
Platform for interactive analysis in the cloud

Large Scale Analytics

Filter + Map

Join

Aggregate

Preprocess

Keras

Caltech GPU cluster

NumPy

Pandas

Python

Natural Language Analyses

with NLTK

Lukas Layer | CHEP 2019
NLP Workflow

Log snippet selection → Tokenization + Cleaning → Indexing

An exception of category 'FileReadError' occurred while [0] Calling InputSource:getNextItemType [1] Reading branch EventAuxiliary [2] Calling Xr
core/data/RUN2018A/EGamma/RM/v1/000/315/363/00000/
76CD1546-8548-EB11-9B3C-FA163EBA2BF6.root\trie=cc
xrl1003.in2p3.fr,dev21.khartov.uw",flags=Ox10,
permissions=0660) => error ['FATAL'] Connection error (errno=0, code=108) Additional Info: [a] Active source: ccxrl1003.in2p3.fr:1095 [site T1_FR_CCCIN293]

Tokenization +
Cleaning

Indexing

Word
Embeddings +
Message
Representation

[35, 12, 10, 37, 186, 34, 25, 3, 11, 291, 532, 4, 188, 183, 273, 16, 11, 83, 251, 33, 11, 38, 77, 33, 7, 17, 24, 23, 68, 19, 48, 41, 15875, 44, 47, 6, 6, 6, 39, 847, 6, 107, 2210, 65, 66, 531, 19, 867, 58, 1 68]
Unsupervised word embeddings

- Map **similar words** to **nearby points** in a high-dimensional vector space → **Capture relations and reduce dimensionality**

- Train with **Word2vec Skip-Gram** algorithm: model to predict context given a word

- Average the word vectors in each message → **input for ML**
Supervised word embeddings: RNN

- Averaging does not exploit the **sequential nature of language** → use a RNN (LSTM / GRU)

- GRU **encodes** the embedding vectors of the error log words in a **vector**

![Diagram](image-url)

Message padded with zeros

Supervised word embeddings: RNN

- Averaging does not exploit the **sequential nature of language** → use a RNN (LSTM / GRU)

- GRU **encodes** the embedding vectors of the error log words in a **vector**

![Diagram](image-url)

Supervised word embeddings: RNN

- Averaging does not exploit the **sequential nature of language** → use a RNN (LSTM / GRU)

- GRU **encodes** the embedding vectors of the error log words in a **vector**

![Diagram](image-url)
NLP model for error logs

Matrix of Error Logs

Embedding + GRU for each Error Log

Matrix of Error Counts

Concatenate

Dense

Classification
NLP model for error logs

Errors
Sites
Words

InputLayer
- input: (None, 77, 81, 500)
- output: (None, 77, 81, 500)

Reshape
- input: (None, 77, 81, 500)
- output: (None, 6237, 500)

TimeDistributed(Model)
- input: (None, 6237, 500)
- output: (None, 6237, 10)

Encoded Error Log

Errors Sites Words
NLP model for error logs

Errors Sites Counts

Combined information

Arb. depth + Dim. Reduction e.g. Shared Dense Layer

Binary Classification

InputLayer

Reshape

Embedding

GRU

Reshape

InputLayer

Concatenate

Flatten

Dense

Dropout

Dense

Concatenate

input: ([None, 77, 81, 10], [None, 77, 81, 2])

output: (None, 77, 81, 12)

InputLayer

Reshape

output: (None, 77, 81, 10)

Flatten

input: (None, 77, 81, 12)

output: (None, 74844)

Dense

input: (None, 74844)

output: (None, 50)

Dropout

input: (None, 50)

output: (None, 50)

Dense

input: (None, 50)

output: (None, 1)

Lukas Layer | CHEP 2019
Training and target

acdc w.o. modification
Retry only failed jobs
No modification of memory or splitting

vs. all other actions

• Simplest target → future: multiclass classification for more actions

• Small data + imbalanced classes → Cross-validation (3-Folds)

• Sparse input: batch size = 4 → $O(1h)$ / epoch on a NVidia GeForce GTX 1080

Hyper-parameter Optimization

• Bayesian Optimization with scikit-optimize

• 2-8 GPUs on Caltech GPU cluster with NVidia GeForce GTX 1080 / Titan X

• Experimental: distributed training with the NNLO framework

https://github.com/vlimant/NNLO

→ Talk at CHEP

• Experimental: Training models in parallel with spark_sklearn on SWAN with the help of CERN IT
Results

ROC AUC for acdc w.o. modification vs. other

Baseline: FF with counts only
AVG: FF for averaged w2v + counts
RNN: RNN for embeddings + counts

Successful training of first complex NLP models

Similar results as baseline - performance improving with more data

Work ongoing → Full potential not yet exploited
Summary

• Implementation of a **pipeline for DAQ and ML** of error logs using big data analysis tools → [https://github.com/llayer/AIErrorLogAnalysis](https://github.com/llayer/AIErrorLogAnalysis)

• Development of a prototype **NLP model** in Keras

Outlook

• **Further investigations** of data and model → Reimplement in **pytorch / tensorflow** and exploit the **NNLO framework**

• Get involved with the **NLP community** and learn from their experiences

• **RUCIO** (→ [Talk at CHEP](https://www.chep.org)): Development of a **common system** to collect and categorize errors, provide shifters with actions and collect feedback on the suggestion → could also contain **error log snippets**
BACKUP
Word embeddings for log messages

- Vocabulary: ~30,000 words after filtering
- Unsupervised training with word2vec
- Standard parameters with embedding dimensions 10, 20, 50
- Visualization: non-linear dimensionality reduction with the t-sne algorithm in 2D
Message representation: RNN + Attention

- Not all words are equally important
- Return all RNN states and pass through a shallow network to decide weight corresponding to each vector
- The weighted sum of each vector embodies the meaning of those vectors combined

Figure 1: Schematic of our proposed “feed-forward” attention mechanism (cf. (Cho, 2015) Figure 1). Vectors in the hidden state sequence $h_t$ are fed into the learnable function $\alpha(h_t)$ to produce a probability vector $\alpha$. The vector $c$ is computed as a weighted average of $h_t$, with weighting given by $\alpha$.

Model for averaged w2v vectors
Best hyperparameters for the RNN model

Batch size = 4  Maximum number of words = 400

```
skot_dim_dimred = [
    Real( low=1e-5, high=1e-3, prior='log-uniform', name='learning_rate' ),
    Real( low=1e-3, high=0.1, prior='log-uniform', name='dropout' ),
    Integer( low=5, high=32, name='embedding' ),
    Integer( low=2, high=20, name='rnn_units' ),
    Integer( low=10, high = 100, name = 'units_site' ),
    Integer( low=1, high = 5, name='dense_layers' ),
    Integer( low=10, high=50, name='dense_units' ),
    Integer( low=0, high=1, name='encode_sites' ),
]
```

<table>
<thead>
<tr>
<th>learning_rate</th>
<th>dropout</th>
<th>embedding</th>
<th>rnn_units</th>
<th>units_site</th>
<th>dense_layers</th>
<th>dense_units</th>
<th>encode_sites</th>
<th>result</th>
<th>std</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000010</td>
<td>0.001</td>
<td>5.0</td>
<td>20.0</td>
<td>100.0</td>
<td>5.0</td>
<td>50.0</td>
<td>1.0</td>
<td>-0.833228</td>
<td>0.010969</td>
</tr>
<tr>
<td>0.000037</td>
<td>0.100</td>
<td>20.0</td>
<td>14.0</td>
<td>62.0</td>
<td>2.0</td>
<td>25.0</td>
<td>1.0</td>
<td>-0.831024</td>
<td>0.010876</td>
</tr>
<tr>
<td>0.000018</td>
<td>0.100</td>
<td>22.0</td>
<td>15.0</td>
<td>54.0</td>
<td>4.0</td>
<td>21.0</td>
<td>1.0</td>
<td>-0.828022</td>
<td>0.010423</td>
</tr>
<tr>
<td>0.000019</td>
<td>0.100</td>
<td>21.0</td>
<td>15.0</td>
<td>54.0</td>
<td>3.0</td>
<td>21.0</td>
<td>1.0</td>
<td>-0.826678</td>
<td>0.011404</td>
</tr>
<tr>
<td>0.000010</td>
<td>0.001</td>
<td>5.0</td>
<td>20.0</td>
<td>100.0</td>
<td>1.0</td>
<td>50.0</td>
<td>1.0</td>
<td>-0.824910</td>
<td>0.012229</td>
</tr>
</tbody>
</table>
## Actions

<table>
<thead>
<tr>
<th>action</th>
<th>splitting</th>
<th>Site list</th>
<th>memory</th>
<th>rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACDC</td>
<td>None</td>
<td>Modified</td>
<td>Not set</td>
<td>87.53%</td>
</tr>
<tr>
<td>clone</td>
<td>None</td>
<td>Not modified</td>
<td>Not set</td>
<td>4.61%</td>
</tr>
<tr>
<td>ACDC</td>
<td>None</td>
<td>Modified</td>
<td>&gt; 20GB</td>
<td>3.28%</td>
</tr>
<tr>
<td>ACDC</td>
<td>None</td>
<td>Modified</td>
<td>&gt; 10GB</td>
<td>1.12%</td>
</tr>
<tr>
<td>ACDC</td>
<td>None</td>
<td>Modified</td>
<td>&lt; 10GB</td>
<td>0.75%</td>
</tr>
<tr>
<td>ACDC</td>
<td>10x</td>
<td>Modified</td>
<td>Not set</td>
<td>0.74%</td>
</tr>
<tr>
<td>Other actions (27 more rows)</td>
<td></td>
<td></td>
<td></td>
<td>&lt; 2%</td>
</tr>
</tbody>
</table>