EAGE Conference on Reservoir Geoscience

3-5 December 2018 • Kuala Lumpur, Malaysia

www.eage.org
Elastic Docs As An Automated Information Retrieval Platform For Unstructured Reservoir Data Utilizing a Sequence of Smart Machine Learning Algorithms within a Hybrid Cloud Container

N.M. Hernandez, P.J Lucanas, I. Panganiban, C. Mamador, C.Yu  
Iraya Energies
Outline of Presentation

• Problem Statement
  ▪ Dealing with Unstructured Reservoir Data

• Methodology
  ▪ Machine Learning
  ▪ Database & Infrastructure

• Results
  ▪ Supervised and Unsupervised
  ▪ NLP & Reservoir Image Processing
  ▪ Metrics
  ▪ API Demonstration

• Conclusion
“Where oil is first found is in the minds of men”
-Wallace E. Pratt
Pioneer Petroleum Geologist (1959)

“Good intuition requires years of practice”
-Malcolm Gladwell
Author, Blink
Symbiosis between ML and Data Mining

“PRACTICE” OR “FAMILIARITY”

DATA

Machine Learning

Deep Neural Networks

Enables

Employs

Data Mining

Insights

“INTUITION”

Artificial Intelligence

Big Data

Statistics

Math

Human Experience

Enables

Employs

Analytics

Visuals

Technical Intelligence

Business Intelligence

2018, NMH modified from LinkedIn Source
Problem Statement: ElasticDocs

Highly advanced platforms for well and seismic data

Unstructured documents are inaccessible

Technical Memory is Lost

- Reservoir Properties
- Geological Interpretations
Problem Statement

Using the latest advances in Machine Learning:

• How to gain reservoir experience leveraging from existing data
• How to maintain or recover corporate memory
Methodology

ElasticDocs ML Pipeline
Methodology

IT Infrastructure
- Cloud-Based Or Hybrid Infrastructure
- Database SQL NoSQL
- Microservices

ML Core

User-Facing API
- Web Layer
- GIS Layers

CURATED, OPEN SOURCE LIBRARIES
- Python, Tensorflow, Leaflet, Elasticsearch, etc
Named Entity Recognition

Geology Identification

(“the glauconitic claystone sample is barren of foraminifera. no definitive environment interpretation is possible.”,{'entities':[4,25,'GEOL']}),

(“massive claystone interbedded with silty claystone and thin argillaceous siltstone”,{'entities':[8,17,'GEOL'],35,50,'GEOL'],60,82,'GEOL'}),

(“chevron australia pty ltd acme 1”,”entities’:[[]]), (example of no geological entity to be detected)

Hydrocarbon show Identification

(“STAIN N-D FLUORESCENCE 3699-3709m This interval constitutes the upper part of the”,’entities’:[23,27,’DEPTH’],28,32,’DEPTH’))
Natural Language Processing

Word Clouds

Unsupervised Clustering with Topic Models
Image Recognition

Multi-Format Images

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Map</td>
<td>0.83</td>
<td>0.96</td>
<td>0.89</td>
</tr>
<tr>
<td>Seismic</td>
<td>1.00</td>
<td>0.95</td>
<td>0.97</td>
</tr>
<tr>
<td>Core</td>
<td>0.89</td>
<td>0.98</td>
<td>0.94</td>
</tr>
<tr>
<td>SEM</td>
<td>0.95</td>
<td>0.93</td>
<td>0.94</td>
</tr>
<tr>
<td>Others</td>
<td>0.91</td>
<td>0.73</td>
<td>0.81</td>
</tr>
</tbody>
</table>
Image Recognition

Well Twinning

Twin Set A

Twin Set B

Applications in:
- Analog Search
- Anomaly detection
- Quality control
Scene Detection

Train Station Signage

Seismic Signage

Cross-section of Seismic Line Showing Planned Drill Holes
Scene Detection

- thrust
- km
- on
- (not
- footwall
- or
- listric
- velocity
- 'with
- 'real
- ms
- 'horse
- are
- by
- crest
- fault
- 500
- splay?
- 'increasing
- deformation
- "enhanced"
- structure
- of
- <trands
- > shape
- loss
- 'amplitude
- seismic
- pull-up
- seismic
- depth?
- f amplitude
- of
- between
- artefacts'
- in
- these
- 'synthetic
- hints
- thrusts
- =DHI?"
Metrics

**Speed**

![Training Loss Graph]

**Accuracy**

- Precision: proportion of positive identification is correct
  \[
  \text{Precision} = \frac{T.P.}{T.P. + F.P.}
  \]
- Recall: proportion of actual positives is correct
  \[
  \text{Recall} = \frac{T.P.}{T.P. + F.N.}
  \]
- F1 score: harmonic mean of precision and recall
  \[
  \text{F1 score} = \frac{2(\text{precision} \times \text{recall})}{\text{precision} + \text{recall}}
  \]

**Precision and Recall Table**

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>T.N.</td>
</tr>
<tr>
<td>1</td>
<td>F.N.</td>
</tr>
</tbody>
</table>
# Metrics

<table>
<thead>
<tr>
<th>MT Application</th>
<th>Task</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCR</td>
<td>Text Extraction only, excluding Image Classification</td>
<td>150,000 pages 13 hrs</td>
</tr>
<tr>
<td></td>
<td>Text Extraction and Image Extraction</td>
<td>4,542 pages 6.31 GB 25 Final Well Reports 10 hrs</td>
</tr>
<tr>
<td>NLP</td>
<td>Lithology / Geology Indicator Frequency Analysis (i.e. Carbonates, Sandstone, etc)</td>
<td>4 hours</td>
</tr>
<tr>
<td></td>
<td>Well Cataloguing</td>
<td>1,500+ input las files 66, 515 curves identified 5,681 top log curves (cali, gr, neu, por, ...) 2 hrs 33.66 min</td>
</tr>
<tr>
<td>DCNN</td>
<td>Imag Classification</td>
<td>2,598 tagged images input 10% Tables 6% Figures 19% Map 24% Charts 33% Noise 20-30 mins during training &lt;s after training</td>
</tr>
<tr>
<td></td>
<td>Currently includes 8 classes: thin section SEM, seismic, stratigraphic chart, cores, map and general classes such as chart, figure, table</td>
<td></td>
</tr>
</tbody>
</table>
ElasticDocs

- ElasticSearch
- Geolocation
- Metadata extraction
- AutoImage Recognition

Supports geoscientists’

- Knowledge
- Intuition
- Experience

through accessible, verifiable big data

Tying it all together with an API
Dealing with huge amount of unstructured reservoir dataset is made more effective in ElasticDocs by:

- Curation and thorough investigation of appropriate machine learning algorithms
- Creating both structured and non-structured database to host and properly standardize reservoir data as input to machine learning algorithm
- Apply appropriate compute infrastructure, leveraging on availability of compute resources, either with on-prem or cloud
- Design a user-friendly API that all geoscientists can access and analyze their own data

There is huge potential in application of machine learning, and we are barely scratching the surface

- complex networks
- Increasing granularity in object identification