Distributed Computing System for the Allocation of Terrigenous Reservoirs Using Machine Learning

S.S. Samboretskiy and I.A. Muravev
Tyumen State University
Terrigenous Reservoir Identification Problem

- It is necessary to identify terrigenous reservoirs along the depth of the wellbore, using the available WL data.
Terrigenous Reservoir Identification Problem

\[ y = f_{\text{name}}(d) \]
\[ y = f_{\text{name}}(d) \]
\[ y = f_{\text{name}}(d) \]

\[ F \]

\[ Z \]

\[ y = f_{\text{name}}(d) \]
\[ y = f_{\text{name}}(d) \]
\[ y = f_{\text{name}}(d) \]

\[ F_1 \]
\[ F_2 \]
\[ F_κ \]

\[ \sum \]

\[ Z' \]
Decomposition-Computation-Conjugation

• There are two types of calculation’s tasks:

  • $A = \{a; a_i | a_j, i, j \in 1 ... N\} – computations,$
  
  • $B = \{b; b_i | b_j, i, j \in 1 ... N, \forall b_i \exists A'_i \subset A\} – conjugations$
Decomposition-Computation-Conjugation

• 1. *Decomposition*. WL data is available for each well. The data is initially provided like this, there is no need for additional partitioning.

• 2. *Computation*. Data is processed by various machine learning algorithms.

• 3. *Conjugation*. The results of WL data processing by different algorithms for one well must be comprehensively evaluated.
Why not MapReduce?

• MapReduce is used to split datasets without internal links.

• Tasks that MapReduce does not solve:
  • Sector hydrodynamic modeling of the reservoir
  • Allocation of terrigenous reservoirs
  • large graphs handling (traffic flows modeling)
Software for DCC

• Requirements of software package:
  • Execution of parallel and distributed computations to the DCC process
  • Fault tolerance
  • Load Balancing
## Testing of Solution

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-layer neural network (5,3,2)</td>
<td>0.836004</td>
<td>0.121893</td>
</tr>
<tr>
<td>Extra Trees (Extra Tree)</td>
<td>0.802129</td>
<td>0.133246</td>
</tr>
<tr>
<td>k-nearest neighbors</td>
<td>0.80188</td>
<td>0.117094</td>
</tr>
<tr>
<td>Support vector machine</td>
<td>0.789811</td>
<td>0.172834</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.727287</td>
<td>0.173611</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>0.723322</td>
<td>0.154546</td>
</tr>
<tr>
<td>Extra Tree</td>
<td>0.699206</td>
<td>0.186389</td>
</tr>
</tbody>
</table>

The Results of WL Data Interpretation by Machine Learning Algorithms
Testing

Well Distribution by Interpretation Accuracy Using a Multi-Layer Neural Network Bar Chart

The Results of WL Data Interpretation by Various Machine Learning Algorithms and a Human
Testing

<table>
<thead>
<tr>
<th>M</th>
<th>np=1</th>
<th>np=6</th>
<th>S(6)</th>
<th>E(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>4m38.767s</td>
<td>1m55.954s</td>
<td>2.4041</td>
<td>0.4007</td>
</tr>
<tr>
<td>200</td>
<td>19m14.967s</td>
<td>8m08.926s</td>
<td>2.3617</td>
<td>0.3936</td>
</tr>
<tr>
<td>300</td>
<td>33m50.626s</td>
<td>14m11.451s</td>
<td>2.3849</td>
<td>0.3975</td>
</tr>
<tr>
<td>400</td>
<td>48m26.556s</td>
<td>20m24.071s</td>
<td>2.3745</td>
<td>0.3958</td>
</tr>
<tr>
<td>500</td>
<td>63m02.485s</td>
<td>26m46.355s</td>
<td>2.3547</td>
<td>0.3925</td>
</tr>
</tbody>
</table>

Runtime of Computations Using Purple Software Package (number of processors = 6)
Testing

Acceleration and Efficiency

<table>
<thead>
<tr>
<th>Number of processors</th>
<th>Acceleration</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.39602</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.34954</td>
<td>0.31662</td>
</tr>
<tr>
<td>24</td>
<td>7.59914</td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

• successful use of machine learning to identify terrigenous reservoirs

• successful application of decomposition-computation-conjugation (DCC) approach for this problem

• Our results allow assuming that in the near term the created software package will lead to automation and significant acceleration of a specific analysis procedure in well logging, thereby helping specialists make decisions faster.
Thank you for your attention