APPLYING BIG DATA ANALYTICS TO SEISMIC INTERPRETATION

Lei Huang, Chao Chen, Yuzhong Yan
Prairie View A & M University

Ted Clee
TEC Applications Analysis

Rice University Oil & Gas HPC Conference
March 2nd, 2016
**THE BIG DATA CHALLENGE IN SEISMIC EXPLORATION**

- **Dataset sizes for a seismic survey 50 km square:**
  - **2D ~1975**
    - 50 lines, spaced 1 km apart, each line 50 km w/ 50 m bins
    - 20-fold stack, 6 sec @ 4 msec sampling
    - 1 M traces, 6 GB
  - **3D ~2000**
    - 25 x 25 m bins, 2000 x 2000 lines, 100-fold stack
    - 400 M traces, 1.2 TB
  - **Advanced 3D ~2015**
    - Double bin resolution, double fold, 8 azimuths
    - 8 sec recording for longer offsets
    - 25,600 M traces, >100 TB
  - **Time-Lapse (4D) exploration through production**
    - 1 TB /km² - Eric Green, BP (Mar 2015)
    - 10,000 M traces (multiple passes), 2,500 TB
Seismic interpretation increasingly seeks to incorporate analysis of a great variety of attribute measurements.
- Yields more data volumes

Limited time with more data volumes to explore
- Need more computation power

Problem goes beyond existing desktop solutions
- Require scalable platforms
Big data analytics platforms may help
- Scalable solution for data storage and analytics
- Machine learning techniques to support interpretation process

Project: seismic analytics platform
- Distributed and scalable platform
- Machine learning + signal/image processing
- User-friendly programming interface
High level languages: Scala, Java and Python
- Keep data in distributed memory, and move computation to data
- A common data representation (RDD) for parallel transformations
- An open and integrated platform with a variety of tools/packages
- Balance of performance and productivity
Seismic Analytics SDK

SeismicVolume

- loadFromFile (HDFS)
- repartition
- aggregate
- overlap
- transpose
- save (HDFS)

- get
  - sample
  - trace
  - line

- applyMap (usrFunc)
  - sample
  - trace
  - line
SEISMIC ATTRIBUTE PARALLEL PROCESSING TEMPLATES

(a) 1 -> 1

(b) N -> 1

(c) 1 -> N

(d) Overlap
SEISMIC ATTRIBUTE PARALLEL TEMPLATES PERFORMANCE

- Hilbert Transform

![Graph showing speedup times for different split sizes and core counts.](Image)
USE CASE: GEOLOGICAL FAULT DETECTION

- Apply the big data analytics approach to detecting geological faults from a collection of seismic attribute volumes
SEISMIC DATA ANALYTICS PROCESS

1. Training:

- Seismic volume

  ➔ Seismic Attributes

  ➔ Feature Vectors

  ➔ Label data

  ➔ Machine Learning Algorithms

  ➔ Evaluation Metrics

  ➔ Model

  ➔ Predicted Results

2. Application:

- Seismic volume

  ➔ Seismic Attributes

  ➔ Feature Vectors

  ➔ Model

  ➔ Predicted Results

Seismic Data Analytics SDK
Spark Distributed Processing Engine
Each seismic dataset yields a suite of attributes that are possible faulting indicators:

- Thinned fault likelihood* from amplitude envelope
- Dip-direction curvature from amplitude envelope

* Many thanks to the “fault likelihood” attribute computed by Dave Hale’s IPF package from Colorado School of Mines.
Machine learning process requires a reasonable number of non-redundant attributes...

- PSTM volumes for near, middle, and far offsets
- Raw data, amplitude envelope, and instantaneous phase attributes for each volume
- Slopes, gradient, planarity, semblance, fault likelihood for each attribute
- Several curvature attributes (based on second derivatives) for each volume
CLASSIFICATION EXPERIMENTS

Labeled Faults

Predicted Faults

Area under ROC = 0.9265, Accuracy = 95.6%  F-Measure = 47.5%
CONCLUSIONS AND FUTURE WORK

- Spark + Hadoop provide a big data storage, seismic data processing and analytics solution
- PVAMU seismic analytics platform provides a scalable and productive SDK for seismic interpretation applications
- Geological feature detection use case demonstrated encouraging results
- Will further improve scalability and explore advanced machine learning use cases
ACKNOWLEDGEMENT

- National Science Foundation
- Dave Hale’s IPF package from Colorado School of Mines