Frac Hits Detection Using Deep Learning and Fiber Optic

1. Problem Statement

The advance in horizontal drilling and hydraulic fracturing technologies have accelerated "Shale Revolution" in the United States. These technologies significantly increase oil and gas production in the U.S by allowing the wellbore to have more contact with the producing formation. To efficiently and economically develop unconventional reservoirs, operators try to drill their horizontal wells as close to each other as possible. With such well spacing pattern, it becomes norm to observe induced fractures to propagate from the well being fracked to adjacent wells. The impacts of frac hits were widely reported in the industry with mixed results. In several cases, frac hits can reduce the productivity of the adjacent wells. Nonetheless, frac hits can give feedback about completion parameters such as fracture length, width, and height if the project is carefully designed to do so. Optimizing completion parameter is critical to improve hydraulic-fracture efficiency and unconventional production performance. Therefore, it is critical to detect the frac hits and pinpoint the depth and time when they occurred.

The objective of this project is to develop a deep learning model to detect frac hits given the distributed acoustic sensor (DAS) as the input. So the operator can take this information to incorporate to their simulation to improve their efficiently and production enhancement.

2. Data Acquisition and Cleaning

The fiber-optic data used in this project is confidential and internal within my company. The whole dataset is about 14GB, which consist of 68,333 LAS files. Each LAS file is the slow strain measurement along the wellbore from the surface to the end of the wellbore at every second. All slow strain data was loaded into a Pandas dataframe. Fiber optic cable was setup in the configuration that it is connected to the IDAS interrogator unit within the acquisition trailer and runs from there to the wireline unit nearby and to the wellhead and down to subsurface. Because the zero-depth reference was conventionally setup at the wellhead, the depth of fiber from the wellhead back to acquisition trailer is showed as negative values. Usually, we are only interested in the wellhead to subsurface section; therefore, I filtered the negative depth out of the dataframe. Next, the negative and positive slow strain measurements at each depth are separated into negative and positive slow strain dataframe.

There were 79 hydraulic stages for this project. The start time and stop time of each stage is defined by when the pumps were turned on and off respectively. We usually started recording 5 minutes before the stage starts and continued recording for another hour after the pumps were shut off. However, because the pumping company was able to start the subsequent stage within 10-15 minutes after the previous stages ended, the data was the previous stage was recorded until 5 minutes before the next stage started. The "Pre_frac', 'During frac' and 'Post frac' cumulative slow strain were computed. Furthermore, more engineer features were created at this stage for machine learning model later, such as 'RMS', 'FFT' for both negative and positive slow strain prefrac, during frac and post frac. In addition, 'Total_Strain_Duringfrac' and 'Total_Strain_Afterfrac' were computed as well as the 'Delta_SSPS' and 'Delta_SSNS' during and post frac. Finally, I combined all features for each stage into a dataframe, which were later export out as csv files.

Next step, I need to provide the label for my data to indicate at which depth the frac hits occurs. Normally, field engineers within my company will start label the frac hits as soon as the treatment stage ends. Nevertheless, after I QC the provided labels, I discovered that the process to label the frac hits were subjective. For the same stage treatment, the labels will be different depending on who labels it. This adds more noise to the data. After I spent some time for literature review plus my observations, I decided to use different algorithm to label the frac hits. The algorithm allows me to pick the same frac hit depth regardless who performs it. There were 79 frac hits csv files that corresponding to 79 slow strain csv files. After loading each slow strain and frac hit csv files into separated dataframe, they are merged by depth column.

3. Data Exploration

First, I loaded all the csv files into the dataframe 'data'. Then, I want to take a quick glance into the data.

data.head()

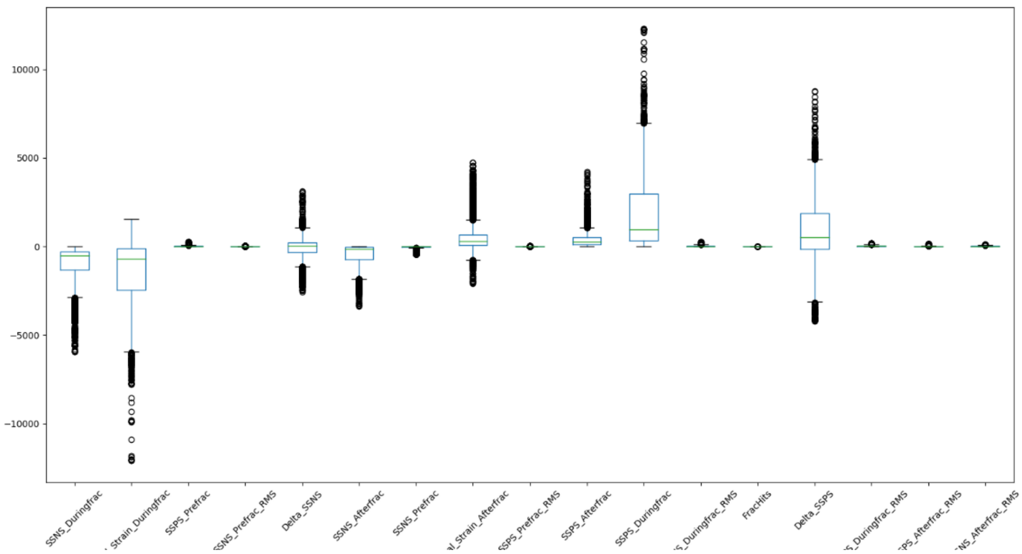
```
           SSNS_Duringfrac  Total_Strain_Duringfrac  SSPS_Prefrac  \
Depth
20258.791 -393.92          80.11                    2.64
20262.146 -392.62          75.02                    2.37
20265.502 -392.14          74.63                    2.24
20268.857 -385.57          73.90                    1.89
20272.215 -386.79          72.50                    2.14

           SSNS_Prefrac_RMS  Delta_SSNS  SSNS_Afterfrac  SSNS_Prefrac  \
Depth
20258.791  0.070640          85.85      -33.04          -1.11
20262.146  0.081158          80.17      -35.21          -1.26
20265.502  0.090646          77.96      -36.47          -1.39
20268.857  0.097074          81.75      -32.66          -1.49
20272.215  0.085538          75.79      -37.55          -1.27

           Total_Strain_Afterfrac  SSPS_Prefrac_RMS  SSPS_Afterfrac  \
Depth
20258.791  360.88                  0.127410          118.89
20262.146  357.41                  0.115686          115.38
20265.502  355.67                  0.111086          114.43
20268.857  352.91                  0.100615          114.41
20272.215  349.24                  0.108873          113.34

           SSPS_Duringfrac  SSNS_Duringfrac_RMS  FracHits  Delta_SSPS  \
Depth
20258.791  38.78            4.104243             0         -355.14
20262.146  40.36            4.108679             0         -352.26
20265.502  39.80            4.086344             0         -352.34
20268.857  40.51            4.048087             0         -345.06
20272.215  40.84            4.054293             0         -345.95

           SSPS_Duringfrac_RMS  SSPS_Afterfrac_RMS  SSNS_Afterfrac_RMS
Depth
20258.791  0.724665             1.776064            0.705337
20262.146  0.758402             1.726569            0.729098
20265.502  0.777089             1.720084            0.746282
20268.857  0.788190             1.713045            0.698112
20272.215  0.798294             1.714032            0.759866
```

Let take a closer look into each column of the dataframe. As we can see from the figure below, there are 6122 observations. Most of the columns are float data, except for column 'FracHits' which is the labels for this project and has integer values. Null values are not observed from this dataframe.

```
In [63]: data.info()
<class 'pandas.core.frame.DataFrame'>
Float64Index: 6122 entries, 20258.791 to 9078.459
Data columns (total 17 columns):
SSNS_Duringfrac            6122 non-null float64
Total_Strain_Duringfrac    6122 non-null float64
SSPS_Prefrac               6122 non-null float64
SSNS_Prefrac_RMS           6122 non-null float64
Delta_SSNS                 6122 non-null float64
SSNS_Afterfrac             6122 non-null float64
SSNS_Prefrac               6122 non-null float64
Total_Strain_Afterfrac     6122 non-null float64
SSPS_Prefrac_RMS           6122 non-null float64
SSPS_Afterfrac             6122 non-null float64
SSPS_Duringfrac            6122 non-null float64
SSNS_Duringfrac_RMS        6122 non-null float64
FracHits                   6122 non-null int32
Delta_SSPS                 6122 non-null float64
SSPS_Duringfrac_RMS        6122 non-null float64
SSPS_Afterfrac_RMS         6122 non-null float64
SSNS_Afterfrac_RMS         6122 non-null float64
dtypes: float64(16), int32(1)
memory usage: 997.0 KB
```

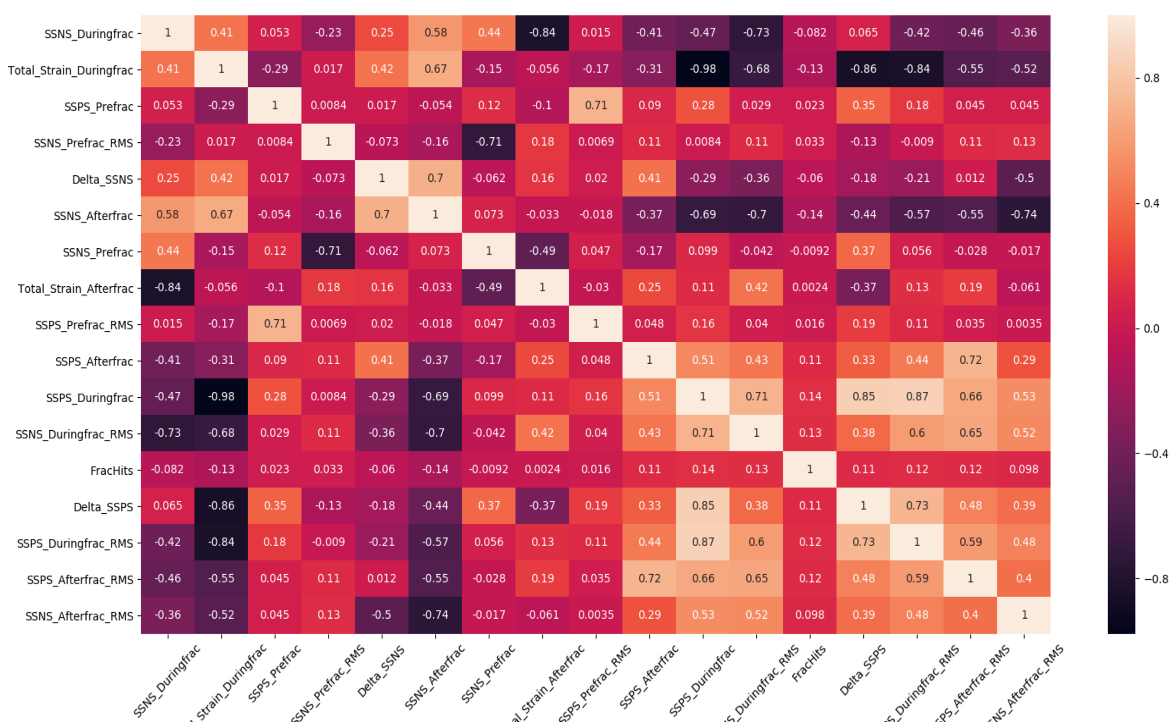
Next, I want to look at descriptive statistic of the data. It’s easier to look at the box-plot.

```
       SSNS_Duringfrac  Total_Strain_Duringfrac  SSPS_Prefrac  \
count  6122.000000      6122.000000              6122.000000
mean  -972.211643      -1451.836241              24.091351
std    1050.515400      1812.950636              41.835338
min   -5935.040000     -12088.850000             0.000000
25%   -1317.292500     -2454.120000              0.000000
50%   -517.785000      -702.260000               7.440000
75%   -277.437500      -118.545000               30.055000
max   -0.050000         1552.900000              250.470000

       SSNS_Prefrac_RMS   Delta_SSNS  SSNS_Afterfrac  SSNS_Prefrac  \
count  6122.000000       6122.000000  6122.000000     6122.000000
mean   2.199767         -45.164913   -452.004670     -36.360420
std    5.838411          588.703189   577.942085      76.845282
min    0.000000         -2565.210000 -3369.550000    -433.180000
25%    0.000000         -310.970000  -741.940000     -34.430000
50%    0.160984          51.685000   -159.230000     -3.220000
75%    1.288943          240.640000  -25.215000       0.000000
max    47.693855         3128.140000  0.000000        0.000000

       Total_Strain_Afterfrac  SSPS_Prefrac_RMS  SSPS_Afterfrac  \
count  6122.000000             6122.000000       6122.000000
mean   520.206973              1.066433          406.839757
std    858.577009              2.605551          454.146102
min   -2092.130000             0.000000          0.000000
25%    85.630000               0.000000          129.300000
50%    305.095000              0.380103          263.920000
75%    653.967500              1.122730          501.505000
max    4745.900000             36.798000         4228.370000

       SSPS_Duringfrac  SSNS_Duringfrac_RMS     FracHits   Delta_SSPS  \
count  6122.000000      6122.000000          6122.000000  6122.000000
mean   1858.675998      39.731490            0.009311      886.464355
std    2000.283960      49.524605            0.096049      1771.451903
min    0.000000         0.009129             0.000000    -4205.000000
25%    338.062500       5.679909             0.000000    -134.307500
50%    967.160000       16.199486            0.000000     506.855000
75%    2992.250000      55.832395            0.000000     1883.562500
max    12301.040000     254.994588           1.000000     8778.930000

       SSPS_Duringfrac_RMS  SSPS_Afterfrac_RMS  SSNS_Afterfrac_RMS
count  6122.000000          6122.000000         6122.000000
mean   33.350053            17.411284           15.572593
std    28.624948            23.920166           19.570420
min    0.000000             0.000000            0.000000
25%    7.615124             2.764001            1.223586
50%    27.326357            6.673329            6.033174
75%    52.878354            21.525894           25.292452
max    188.028913           169.310345          128.567894
```



As we can see, there are lots of outliers in each features of my data. It's necessary to investigate on these outliers. However, for this study, I will leave the outliers as they are in my data, as I suspect it might reflect the natural of our measurement data.

The next step is to investigate the correlation between each feature in the data.

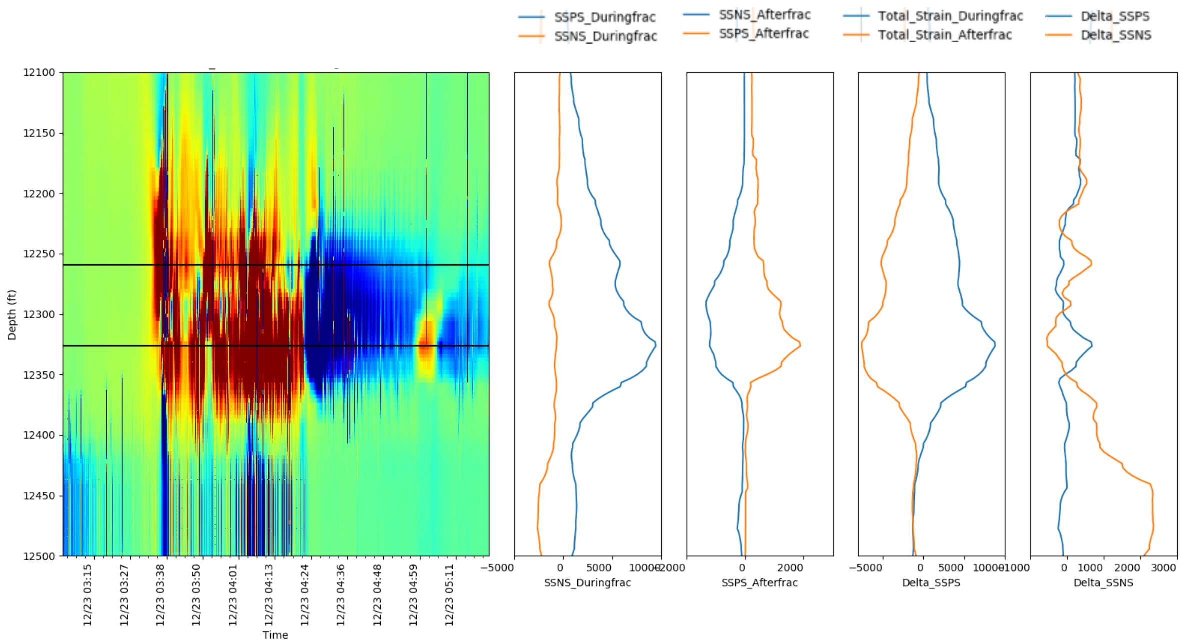


Several important correlations are observed from the heatmap above. These are critical because this is what I was searching for in order to train my model to pick the frac hits.

- 'SSNS_Afterfrac' exhibits negative linear relationship with 'SSPS_Duringfrac'.

- ‘Total_Strain_Duringfrac’ show strong linear relationship with ‘SSNS_Afterfrac’ and inverse relationship with ‘SSPS_Duringfrac’ and ‘Delta_SSPS’

After observing the correlation between features, I’m wondering if these correlation will help to pick the frac hits. Therefore, I plot the slow strain measurement together with the features, to see if I can pick out any correlation.



The first plot on the left shows the slow strain measurement along the time and depth. The two solid black lines indicates the depth where the frac hits occurred. The remaining plots on the right show the features used to model the frac hit pick. There appears some peak correlation between the curves where the frac hits happen.