Swung Hackathon 2019 – Velocity Modelling Project our consultants review...

Who are SWUNG?

The Software Underground (SWUNG https://softwareunderground.org/) is an open geoscience community with over 1500 members.

This year there was an opportunity to organise a hackathon the weekend before the EAGE conference in London to get together and code.

There was a lot of support for the event, we had around 50 people attending from various Oil majors, service companies and academia.

We would like to thank our sponsors: Sword Venture, BP, Earth Science Analytics, Equinor and Agile Scientific. The venue was Work.Life at Clerkenwell in London. It usually serves as a co-working environment, a great venue with good facilities

and a few breakout meeting rooms. The food was amazing it was provided by Elysia Catering and most of it was prepared fresh at the location.

First Day

Saturday morning, we kicked off our Hackathon by announcing the schedule, projects and forming teams. We had a packed agenda and attendees could choose and set their own schedule.

There were tutorials running throughout the two days on Pytorch, Pangeo and Dask.

Another option was to participate in working on open source projects like Segio, Devito and Pylops, the core developers for these projects were at the Hackathon and they were leading sprints.

A third option was to participate in projects and again we had several options: a project on getting data from google earth into an xarray format, getting it ready for remote sensing workflows and a project on analyzing velocity and depth conversion models.

I also announced a couple of ideas around predicting Non-Productive Time (NPT) which got some attention and other ideas around data quality and scraping. In the end I decided to try and help the Velocity Model project. My motivation was two fold it was a problem I haven't encountered before so an opportunity to learn something new and we had real data, which meant we did not have to spend time on collecting datasets, which can be time consuming as there are limited open datasets out there.

Velocity Modelling Project (Play by Play)

Saturday morning Doug stood up and announced that he has a project idea around velocity modelling, he brought a research paper which explains the problem and he also managed to bring data from his company, Tullow Oil. When I heard he had real data I became very interested in the project. I also haven't done much with sonic logs and seismic so wanted to learn something new, step out of my comfort zone a little bit. Doug a geophysicist clearly knew the domain but he had little experience with Python, so I thought it would a good fit, I can help him

with the coding and he can help me understand the problem; we can both learn from each other.

The Problem

It was about finding good models which explain velocity through the well at each formation, the model should explain velocity in terms of the initial velocity (V0) at the top of the formation and rate of change of velocity (K) through depth.

Here is a useful resource if you're not familiar with sonic logs https://www.spec2000.net/06-velocity.htmand the link to the research paper we were following https://library.seg.org/doi/10.1190/1.1444203We spent the morning locked in a small, cozy room, getting the data ready, reading through the research paper and discussing potential solutions to the problem. It took us a couple of hours to get to the heart of the problem, Doug was patiently walking me through the issue and we bounced a few ideas around. Meanwhile the others were doing the Pytorch tutorial lead by Lukas.

It turned out this was a real problem he faces in his day job; he spends up to two weeks manually fitting models in spreadsheets and then loading them into Petrel for checking, he can only do a few models. At this point we were thinking if we can solve this, it would save geophysicists up to 1-2 weeks of time in the future and potentially reduce the risk of surprises during drilling.

We started hacking on the problem and I started to look for people who would be interested in working on this. We quickly got a team together which included Monica, Evgeniy, Ian and Gerald. We went into one of the meeting rooms and started occupying one half of the large table in there, the other half was occupied with the people working on the Devito library.

During lunch I was going around talking to people about our project and quickly got a suggestion from Steve to go and talk to Lukas who is an expert on this stuff and he might be able to help. That's the good thing about a hackathon, there are clever people in the room within shouting distance.

Here is a look at one of the sonic logs, depth is on the X-

axis and velocity in m/s is on the y axis.

During lunch we managed to convince Lukas to come and take a look at the issue, his maths and statistics knowledge proved valuable. By the time Lukas joined us we managed to break the problem into smaller pieces and isolate a minimum solution which would also provide value, an MVP. We setup a Github report for the project and started loading the data. Things started moving quickly at this point we brainstormed a lot, Lukas was our whiteboard person, I was coding and projecting to large screen in the room so everyone could see what I was doing, Doug was walking everyone through the problem.

The team settled into a nice rhythm of brainstorming, running their own version of the code. I was committing regularly to github and people were experimenting around me, shouting there ideas out, googling error messages for me, it was a team effort.

The solution

We came up with a nice solution, it seems very simple in hindsight, but we explored lots of options before ending up here. We simulated lots of linear models through a section of the sonic log (VS) velocity measurement, the section represented a formation as it was picked by a geologist. We generated linear models, we selected the best models by calculated the root mean squared error (RMSE). The research paper contained a nice contour plot of the parameters and RMSE we spent some time on this and by the end of the day we managed to reproduce the plot from the paper.

We then spent some time on clustering, just for fun we wanted to see if we run a couple of clustering algorithms to pick the formation intervals would they pick something sensible. In the first stage of the project we just used an interval as it was picked by the geologist. Turns out our first model using K-Means did really well, we tried a few parameters and a few other algorithms, in the end we landed on GaussianMixtures, we used scikit-learn for the clustering work.

We then finished the day talking about next steps and what else we can try on Sunday. We were very pleased with our work, we managed to get to the heart of the problem and crack it in the first day.

Day 2

On Sunday morning we spent some time reviewing the whole solution. Saturday afternoon had passed very quickly, with multiple ideas and code so we just wanted to make sure the whole team was on the same page. We also got some new members Khushal and Max joined us, doing a recap was a good way to get them onboard.

During the review several ideas came up, we worked a little bit on our plots eventually settling on the plot below which shows the solution clearly. We didn't have this the first day, it was only in our head, and some people were struggling to see the final solution. The plot describes the data through the chosen interval and our best models our shown. From red to yellow we show 500 models with decreasing RMSE scores, red is low RMSE, yellow high RMSE. We also show the best model in green, and then we randomly sample 10 models which are shown in blue. These models are exported to a csv file, and this is what Doug 'our customer' needs to load into Petrel and to continue his work there, matching them up with seismic data.After getting this plot everyone was excited, what else can we do? We eventually landed on making the plots more interactive, so we spent the rest of the day cleaning up the solution and adding interactive sliders to the plots. This was again a team effort, I was the main coder but people were googling errors, sending me code when I got stuck. A few team members tried out some other experiments, Lukas looked into refining the clustering algorithms, Monica tried to get the plots working in Power BI and Evgeniy was playing around with denoising the sonic log.

I was still in cleanup mode, which essentially means I have broken our code, when we realized it was time to present our solution, it was a tense 30 minutes to get everything working again and I think there was complete silence in the last 5 minutes when we rushed to get the code finalised for our presentation, it all worked out in the end and our presentation was well received.Written by: Attila Balazs – Principal Consultant at Sword venture

Attila would like to thank you everyone for their authorisation to use photos from the event and his co-organisers especially Fillippo Broggini.

Photos courtesy of Jesper Dramsch (dramsch.net), Evgeniy Markin and Douglas McClymont.

If you would like more information about the solution please follow the link below:

https://github.com/attibalazs/velocityhack

Or call us to arrange a call with our consultant…

Home