Introduction

Land seismic data processing and depth imaging in a desert environment is challenging due to complex near-surface conditions. Sand dunes, karsts features and variations in topography as well as alternating high and low velocity layers can dramatically distort seismic images (Keho and Kelamis, 2012). Characterization of the near surface is typically done by first break tomography of refracted waves. Sometimes it can be difficult to obtain good quality first break picks, for instance in case of poor coupling, noisy conditions or shingled guided waves. Also, first break tomography cannot detect velocity inversions. An alternative approach is to use surface waves to derive near surface velocity profiles. Surface waves are dispersive and dispersion curves can be inverted for 1D velocity profiles. Manually picking and editing dispersion curves is perhaps even more time consuming than first break picking. Automatic dispersion curve picking is therefore an active field of research.

The dispersion curve approach is inherently a 1D method and to obtain 2D or 3D earth models from surface waves one could attempt full waveform inversion (FWI). FWI can give tremendous results in achieving high-resolution velocity models, but, especially in the elastic case, has substantial computational cost, requires good starting models in case of poor low frequency content or missing long offsets and has convergence issues in case events or signals in the data that are not covered by the physical models.

Velocity estimation with deep neural networks

With the arrival of deep learning techniques, various authors have investigated whether estimation of a P-wave velocity model can be done using neural networks. Zheng et al. (2019) use supervised learning to train a CNN to perform prestack elastic inversion, estimating 1D Vp, Vs and density profiles from 2D shot gathers using a 1D earth model as labels. Biswas et al. (2019) apply a similar strategy to estimate 1D elastic profiles but use angle gathers instead of CMP gathers and propose explicit forward modelling from the estimated 1D profiles to ensure a physics driven neural network solution. Several authors have also attempted to estimate a 2D velocity model directly from gathers. Øye and Dahl (2019) use random 2.5D synthetic P-wave velocity models to generate acoustically modelled data, which is then used to train a network to directly estimate a velocity model from the shot gathers and conclude that the approach could be useful to generate smooth models. Duque et al. (2019) use a generative adversarial network to predict 2D P-wave velocity models from 2D acoustically modelled shot gathers. Yang et al. (2019) also aim to estimate a 2D (acoustic) velocity profile but use a collection of shots instead of a single shot to estimate one 2D profile for all the shots.

Method for estimating near-surface Vs profiles from shot gathers with Deep Learning

In this abstract we attempt to use deep neural networks to directly estimate velocity profiles from shot gathers. In no way do we expect this method to produce better results than physics based methods. The deep learning approach can play a role to provide good starting models or cost-effective alternatives to more elaborate solutions. Initially, our approach is 1D, analogous to dispersion curve inversion, but can be extended to estimate 2D velocity profiles from directly from shot-gathers. Similar to the referenced authors we aim to estimate velocity directly from shot gathers with deep learning, but now focus on surface waves instead of body waves. We aim to estimate the Vs profile in the first 200m from “raw” (synthetic) shot gathers. Since surface waves are the dominant energy in a gather, little preprocessing is required other than amplitude balancing (on our synthetic data).

One of the questions of concern is time-to-depth conversion. Algorithms such as traveltime tomography or full waveform inversion are based on the wave equation and explicitly have a concept of velocity and a sense of location built-in that allows for the domain transfer from time to depth. There is no requirement for seismic data to have particular dimensions, numbers of traces or time samples. Data does not even need to be on a regular grid, the main requirement with respect to the gathers is that their coordinate information is available. In training neural networks there is no concept of coordinates, velocities or the wave-equation. One image or array of numbers is simply transformed into another one. The authors of the papers referenced previously do not explicitly mention it, but their methods strictly
work only for the chosen range in offset-time and only output to a fixed depth range (in 1D) or location-depth range (in 2D). Deep learning methods are not flexible with respect to domain conversion, and this is (currently) a limitation to this approach to velocity estimation. That means if the training was done on shot gathers with 3000m offset and 4 seconds of data and one wants to re-use the trained network, then all new data must be sampled so that it represents exactly the same range. This same limitation applies to our approach. To introduce some flexibility on the input side, we focus on estimating Vs profiles from phase velocity vs. frequency (V-f) panels. Regardless of the offset range or number of time samples in a shot-gather, we can always transform it to a V-f domain with a fixed velocity and frequency range (although array forming and spatial sampling will have an impact). A limitation of this approach is that now the method has become inherently 1D, but that holds also for the conventional dispersion curve inversion approach. On the output side, we have no such transformation.

**Figure 1** Neural network architecture for prediction of Vs from shot V-f panels.

**Neural network architecture and training**

The training data is taken from the SEAM Arid modelling effort (3D elastic FD modelling). From the SEAM Arid model 3D survey we have extracted 90,000 2D shot gathers (3500m offset, positive and negative offset separated). The corresponding 2D earth models were averaged into 1D profiles for each shot gather. The gathers had $0.5$ gain applied and were transformed to the V-f domain with a 2D FK transform followed by scaling of the wavenumber axis. The frequency range is $[0, 40]$ Hz and phase velocity range $[0, 3000]$ m/s. This choice is important as all future panels must have the same range (regardless of sampling) if one wants to re-use the trained network. The amplitudes in the phase velocity panel are plotted on a dB scale and clipped to the range $[-60, 0]$ dB. Each plot is then converted to a neural network friendly amplitude range (zero mean and unit variance). The V-f panels are resampled to fixed input dimensions of 128x128 samples, although up to the Conv2D layer (indicated in orange in Figure 1) the network is fully convolutional so the input dimensions are flexible.

We have limited the “labels,” or Vs depth profiles, in our supervised training to cover a depth range of 200m (32 samples at 6.25m depth spacing). Analysis of the dispersion curves in our synthetic data indicated that the mode Rayleigh wave depth of penetration was roughly 100 m. Then the task of our neural network is to map this into a vector of 32 samples, representing velocities along a 200 m depth interval starting at surface (regardless of input dimensions). The neural network architecture that achieves this is shown in Figure 1 and consists of a sequence of pre-activated residual blocks, as proposed by He (2015). These residual blocks have been demonstrated to improve performance for deep neural networks and help to address the vanishing gradient problem.
Figure 2 Evaluation of the neural network on test data. The shot gathers are transformed to $V-f$ panels, which are then fed into the neural network. The right column show the true (black) and estimated (red) Vs profiles. The top three rows show good results, the bottom row a poor prediction.

There are three 2x2 pooling layers, which bring the image dimensions down from 128x128 to 16x16. The number of filters or channels starts at N and is increased to 3N the first two pooling layers and is then decreased all the way down to N again (N=64 in Figure 1, but like the input image size this is flexible). There are only two layers with a fixed dimensions in this network, which is done to ensure we have a column vector of length 32 at output. The first such layer is the 1x1 convolutional layer with 32 channels, which always produces an output tensor with 32 channels. The second layer with fixed output is the global max pooling layer which averages an entire activation map and produces a single scalar per channel, hence we are left with a column vector of length 32. The output of the global max pooling
is connected to a fully connected layer with linear activation (dark blue line in Figure 1). During training the mean squared error loss in minimized. The input is connected to known Vs velocity profiles (green line in Figure 1), so learn to predict Vs. We can actually use the same network to predict both Vp and Vs simultaneously, which is indicated by the dotted green and blue lines in Figure 1.

Results

The 90,000 shot gathers are split into 80% training 20% testing subsets. The test subset of the SEAM dataset is chosen to contain the southern part of the model and data, such that no part of the training will have seen the model beyond a y-coordinate of 3000 m. This test area was chosen because it contains velocity profiles with higher velocities than present in the training data (thus testing the generalization capability). All the testing examples in Figure 2 are from that area. We see numerous examples where the network produces good or acceptable Vs profile, defined as being very close to or capturing the trend of the true model. For some gathers the results are much poorer. The likely explanation is that the network can successfully predict the 1D Vs profile for those examples that occur also in the training subset, but does worse on the Vs profiles and shot-gathers not present in the training data.

Conclusions

The main point of critics of the deep learning approach to velocity estimation is that “there is no physics in it.” This is true. For our supervised learning we need for each shot gather the corresponding 1D earth model. This is impossible for real data and we therefore require synthetic data. Because that synthetic data is generated with our physics based models (elastic FD modelling) one could state that the “physics” enters the neural network via the generated data. We have shown that we can estimate 1D Vs profiles from 2D phase velocity-frequency gathers using a neural network trained with supervised learning. It is easy to over-interpret the significance of this result. We are confident that we can transform a set of 2D arrays into a set of 1D arrays, but this is not quite the same stating that we can now do dispersion curve inversion with deep learning. We are also confident that this does not generalize to unseen data with different velocity ranges and profiles, as was already indicated by the poor result on some test data. Our result is only applicable to V-f panels that resemble the training data. Despite this obvious limitation we are confident that this is are encouraging result. That is because when the new data is similar to our training data then also our predicted 1D velocity profiles will be similar to the true profile. The problem of generalization can then be solved by brute force. Once the range of near surface velocities in a certain area is known, one can forward model many variations (including noise, amplitude variations, etc.), train neural networks and run computationally cheap predictions. This takes the computational burden of advanced near-surface model building out of the project timeline and as such can have a positive business impact.

References