Extracting surface wave dispersion curves with deep learning

Introduction

Multi-channel Analysis of Surface Waves (MASW) is a well-established seismic method employed to obtain useful information about the near subsurface, specifically its shear-wave velocities (Park et al., 2007, Dal Moro, 2015). A fundamental processing step in the MASW methodology is represented by the extraction of dispersion curves from dispersion spectra. Although this process can be automated to some extent, it requires quality control by experts; this can be a time-consuming task and is inevitably affected by subjectivity, so it becomes highly inefficient when dealing with large datasets. To overcome such limitations, several methodologies have been proposed in recent years: Alyousuf and Colombo (2019) developed a deep belief network to pick fundamental modes in the phase velocity spectrum; Rovetta et al. (2020) developed a density-based spatial clustering algorithm for automatic picking of surface wave dispersion curves; finally, Kaul et al. (2021) proposed a machine learning model using a fully convolutional architecture with residual units to pick multi-mode dispersion curves through pixel-wise binary segmentation from dispersion spectra. A common limitation of these approaches is that the information from the dispersion spectra is being used and not the one from the seismic shot gathers; this can be counterproductive in the sense that there are many methods for obtaining such spectra; for example, when the technique developed by Park et al. (2007) is used, the maximum values in the dispersion panels do not always correlate with the fundamental mode (Dal Moro, 2015). Consequently, the output of such deep learning methods may be biased, causing poor inversion results (Zhang & Chan, 2003).

In this work, we propose a methodology to estimate dispersion curves directly from seismic shot gathers using a Convolutional Neural Network guided by geophysical theory. More specifically, given a site of interest, a set of 1D velocity models is created using prior knowledge of the geology of its near surface. The corresponding seismic shot gathers and their associated Rayleigh-wave phase dispersion curves are numerically modeled and used to train a simplified residual network to learn the direct mapping from such shot gathers to their associated dispersion curves. This approach is shown to achieve satisfactory predictions of dispersion curves on a synthetic test dataset and is ultimately deployed on a field dataset. The predicted dispersion curves are finally inverted, and the resulting shear-wave velocity model is plausible and consistent with prior geological knowledge of the area.

Methodology

As with any supervised machine learning algorithm, a labeled training data set is required to train a model and a validation set to evaluate the quality of the predictions. In our specific case, a large set of dispersion curves and associated seismic shot gathers that cover a variety of expected geological scenarios must be generated; this is done by choosing a suitable parametrization for the 1-D velocity and density profiles, which are then used to generate 1000 pairs of synthetic shot gathers and dispersion curves employing elastic finite-difference modeling (Thorbecke & Draganov, 2011) and the Dunkin’s matrix method (Dunkin, 1965), respectively. This is accomplished with the open-source tools Fdelmodc (Thorbecke & Draganov, 2011) and Disba (Luu, 2021a). Moreover, considering inevitable differences in the seismic domain between clean synthetic and field data, a data pre-processing sequence is introduced: the field data is initially low-pass filtered, and an average statistical wavelet is estimated from the entire dataset. This wavelet is then used to shape the spectrum of the synthetic dataset such that it more closely resembles the field data used at the inference stage. Similarly, the field dataset is normalized and convolved with the average wavelet from the training dataset. Finally, colored noise and linear events at early times are also added randomly to the modeled data.

For the neural network architecture, a deep residual convolutional neural network is chosen as this architecture is proven to be effective for many image classification tasks. More specifically, we use a lightweight ResNet architecture, namely ResNet-18, as it contains a relatively small number of trainable parameters compared to other popular ResNet architectures (He et al., 2016). Minor modifications are
applied to the original architecture to account for differences in the dimensions of our inputs and outputs: the input channel of the first layer is adjusted to a single channel, and the output dimension of the last layer is modified to produce an output of 500 samples corresponding to the frequency axis of the dispersion curves. Dropout with a probability of 20% is also added after every ReLU activation function to regularize the network; this strategy proved successful as predictions on the validation set show clear improvements over those produced by the same network without Dropout. The Huber norm between the true and predicted dispersion curves is chosen as the primary loss of the network training, and a secondary loss is introduced to enhance the smoothness of the predicted curves. From the previously described dataset, a randomly selected portion of 20% of the data is used as the validation dataset. As for the training, Adam is the optimizer with a learning rate of 0.001. Moreover, an early stopping checkpoint with a patience parameter of 128 epochs is employed to monitor whether the validation loss is decreasing through epochs. Training is automatically stopped after 1000 epochs.

**Numerical Results**

The proposed methodology is applied to extract dispersion curves from a field dataset provided by Fugro under an academic license. This dataset was acquired in 2018 for engineering purposes by firing 129 shots over a total length of 300 meters; each shot gather is composed of 24 traces at receiver intervals of 1 meter, with an initial offset of 15 meters. The seismic response of the medium is recorded using a 0.5 ms sampling interval for a total recording time of 1 second. To begin with, a synthetic dataset is created that closely resembles the characteristics of the field data, employing the techniques described in the previous section. This is vital for the neural network to learn the effective attributes of the data and perform a successful inference on the field data. The general distribution of the generated velocity and density profiles and related dispersion curves are shown in figure 1. The result of the inference for a selected pair of synthetic and field shot gathers is shown in figure 2.

**Figure 1.** a) Distribution of S-wave and P-wave velocity and density models and b) dispersion curves used for training on top of the average dispersion spectra obtained from the field dataset.

**CNN Visualization**

Feature attribution explains individual predictions by attributing each input feature according to how much it changed the prediction (positively, negatively, or with no influence). They can be classified into two groups: occlusion or perturbation-based, where the algorithm manipulates parts of the image to generate explanations in a model-agnostic way; or gradient-based, where the algorithm computes the gradient of the prediction with respect to the input features. The final output of both methods is a heatmap that quantifies each pixel's relevance to the image's prediction or classification (Christoph, 2022). Occlusion (Zeiler & Fergus, 2013) is a pixel attribution method, representing a particular case of a feature attribution method specifically designed for images. This model interpretability technique has been explored to assess the focus areas in the input shot gathers of our trained network. Occlusion maps proved valuable and reliable in providing a better understanding of the model behavior, ultimately showing that the network focuses on the dispersive components of the input shot gathers (figure 3).
Figure 2. a) Inference over a sample synthetic shot gather and b) Inference over a sample field shot gather. Dispersion spectra are only shown for visualization purposes.

Figure 3. Occlusion was applied to a sample field shot-gather. The first panel shows the model’s output on top of the dispersion spectra. Every other panel shows a heatmap of the importance of each pixel in the shot gather to the inference at the specified frequency.

Inversion results

After applying the proposed workflow to the entire field dataset, the retrieved phase velocity dispersion curves are inverted for isotropic 1D layered velocity models using the Competitive Particle Swarm Optimization algorithm provided in the open-source library evodcin (Luu, 2021). As shown in figure 4, the main geometry of the medium is well recovered, and the individual inversion results are continuous amongst the horizontal definition of layers. For a selected profile, inversion was achieved with a misfit below 0.02.

Conclusions

To accelerate the process of extracting dispersion curves from seismic data involved in the MASW workflow, a deep-learning-based methodology that directly maps seismic shot gathers to dispersion curves is proposed in this paper. Our methodology relies on the creation of realistic 1-D velocity and density profiles based on the geological knowledge of the area of interest and the creation of synthetic seismic datasets using a full-wavefield model engine, e.g., elastic finite-difference. Experiments carried out purely on a synthetic dataset show that the proposed neural network model has excellent performance, fast prediction time, and high accuracy. Several additional challenges arise during its application to field datasets: to overcome such limitations, various data pre-processing strategies are suggested, among which the convolution of both the synthetic and field data with the wavelet extracted from the opposite dataset and the addition of colored noise and linear events to the synthetic dataset since these features are prominent in the field dataset. Finally, the inversion of the extract dispersion curves for an S-wave velocity model proved consistent with the area’s previous knowledge. This work lays the foundations for a fully automated MASW workflow.
Figure 4. a) Inversion of the dispersion curves obtained by inference from the field dataset. b) inversion parameters and results for a sample dispersion curve compared with the misfit value.

Acknowledgments

The authors thank KAUST for supporting this research. DC and JZ performed this work as part of the Visiting Student Research Program (VSRP). We also thank Fugro for releasing the field dataset and permission to show the associated results.

References


