A new P-wave reconstruction method for VSP data using conditional generative adversarial network

Yanwen Wei and Haohuan Fu, Tsinghua University and China National Supercomputer Center in Wuxi; Yunyue Elita Li, Jizhong Yang, National University of Singapore*

Summary

Wave-mode separation is an essential step of multicomponent seismic data processing. Due to limited acquisition at sparse receiver locations, conventional physics-based wave-mode decomposition methods are only feasible based on many strong assumptions such as homogeneous subsurface velocity and flat structures. We propose a new P-wave reconstruction method based on VSP data, aiming to learn the domain transformation directly from full waveform elastic VSP data to their P-wave components. This method is based on optimal transportation theory and implemented by Generative Adversarial Network. We train network on 22180 pairs of synthetic data produced on 2D subsurface model and test on 40596 pairs. The network outputs achieve an average accuracy rate of 97.87%. The results tested on the 2D synthetic data show the network is capable of learning the waveform of target separated data beyond phase recognition and wave classification.

Introduction

Vertical Seismic Profile (VSP) data are one kind of seismic data received by the geophones that clamp to the wall of a drilled well. Seismometers distributed in the well directly gets the signals in the inner of the earth, which contain more accurate depth information of the stratigraphy, and can form the image near the drilled well with high resolution (Kennett et al., 1980; Balch et al., 1982; Hardage et al., 1983). To excavate valid information in the VSP data, wave-mode separation is an essential step of data processing.

Conventional wave decomposition methods for VSP data can be divided into three categories, divergence and curl based methods (Devaney and Oristaglio, 1986; Dellinger and Etgen, 1990; Huang et al., 2007), domain transform methods (Dankbaar, 1987; Wang et al., 2002) and polarization analysis methods (DiSiena et al., 1984; Lei, 2005).

The wave-equation methods, based on divergence and curl, cannot be applied direct to VSP data. They are commonly exploited in elastic inversion and migration algorithms, where two (in 2D) or three (in 3D)-dimensional signals can be derived from snapshots to calculate the separated part of the waveform, whereas practical vertical seismic data only have one dimension distribution of the receiver along the drilled well. When applying the wave-equation method to practical data by approximating the nearby record may lead to inaccurate results (Huang et al., 2007), especially in the cases of complicated subsurface and large source spacing.

The domain transform methods can be divided into F-K domain and τ-P domain method. They transform the wave equations to the target domain and utilize the domain properties to separate the P- and S-wave. Both of these two domain transform methods have to deal with manual parameters aiming for different situations, and these conventional methods usually rely on extra knowledge of velocity.

We aim to find a sufficient neural network to learn about the optimal transmission path to transform the three-component elastic VSP data to their compressional wave-field. The idea of utilizing machine learning methods for traditional geophysical problems is gaining popularity recently. Before a substantial breaking-out development in deep learning (LeCun et al., 2015) the networks that have been used are single and fundamental (Roethe and Tarantola, 1991) Limited by the progress of computer vision and the computation ability, they could only deal with 1D or small 2D problems. When deep learning shows its potential in the various domain such as visual object recognition, object detection, speech recognition, and natural language processing, through convolutional neural networks with very deep layers (LeCun et al., 2015), a large sum of researches begins to seek the potential of machine learning in geophysics again.

Machine learning methods have shown some successful and potential usage in the domain of geophysics (Jia and Ma, 2017) including works that can be defined as classification problems (Dowla et al., 1994; Ramirez Jr and Meyer, 2011) and regression problems (Araya-Polo et al., 2018; Wang et al., 2018; Zhang et al.,2018; Richardson, 2018) which can be hard to solve by using conventional inversion methods. For the challenging problem of wave-mode decomposition issue, machine learning has the potential to extract targeted part of the wave out of the mixed waveform. Unfortunately, most of the existing works remain in the domain of classification for the different wave modes (Barak, 2017). When the mixture of the waveform becomes more complicated, detection and recognition cannot satisfy the need of the seismic data processing.

In this abstract, we propose an optimal transportation based neural network method, aiming to learn the domain mapping directly from full waveform elastic VSP data to P-wave data. This method learns the potential relationship between the two domains through the prepared, separated data, which we can produce directly base on the divergence and curl method on a massive amount of synthetic data. The appropriate neural network helps us find out the sufficient coefficients and acts as a wave decomposition filter to the VSP data. This artificial filter does not need set parameters manually and has the potential to perform better than the conventional domain transform methods.

Theory

Wave-mode separation by divergence and curl

The proposed data-driven machine learning method extracts implicit relationship from the data produced directly based on divergence and curl operator of wave theory.

According to Lame theorem, displacement u can be represented by Helmholtz potential (Aki and Richard, 1980),

$$\vec{u} = \nabla\phi + \nabla\times\psi \quad (1)$$

where ϕ is a scalar that has $\nabla \times \phi = 0$, and ψ is a vector which satisfies $\nabla \cdot \psi = 0$.

Applying divergence and curl operator to the displacement, the scalar and vector potential separate and we get P- and S-potential fields,

$$P = \nabla \cdot \vec{u} = \nabla^2\phi \quad (2)$$

$$S = \nabla\times\vec{u} = \nabla \times (\nabla \times \psi) = -\nabla^2\psi \quad (3)$$

Optimal transportation theory and convergence

Generative Adversarial Network (GAN) (Goodfellow et al., 2014) is a neural network framework for learning to generate data which obey the target distribution. A conditional GAN (Mirza and Osindero, 2014) provides additional input data of different distribution from the target, and take the place of the Gaussian noise layer in the original GAN. It leads GAN theoretical do an optimal transportation mapping from one domain to another. Unfortunately, when lacking diversity of generated samples, training a GAN cannot always get converged. This phenomenon is called mode collapse, and from the perspective of optimal transportation, it is caused by the contradiction between continuous local chart representation and discontinuous transportation map (Guo et al., 2019).

To avoid the mode collapse problem, we apply three kinds of strategy from the supports of theoretical and practical.

- Assure sufficient high sample rate of the data, to guarantee a reconstructed discrete describe of the distribution. This is relatively easy for our wave decomposition problem in geophysics when pre-training on synthetic data.
- Modify the normalization layers, distances (Arjovsky et al., 2017; Gulrajani et al., 2017; Salimans et al., 2018) and optimization strategy (Arora et al., 2017; Mescheder et al., 2018; Lucic et al., 2018). This is more empirical than theoretical and instance-dependent.
- Search for appropriate neural network based on the Brenier theory of optimal transportation, where neural networks have the potential to resolve conflicts theoretically by simulating the gradient of a convex function instead of the discontinuous transportation map (Guo et al., 2019).

Method

Training data generation

We use the finite difference numerical method to simulate the elastic wave signals recorded by a common VSP geometry. The subsurface velocity and density model are composed of two, three and four plane layers. A case of a five-layer subsurface velocity model is shown in Figure 2. For each geology model, 51 sources are set at five meters below than the surface. The velocity of P-wave, the velocity of S-wave and density is ranging from 1800 m/s to 5000 m/s, 1040 m/s to 2890 m/s, and 2.2 kg/m3 to 2.8 kg/m3 respectively.

Then based on the character of P- and S-waves, we use di-

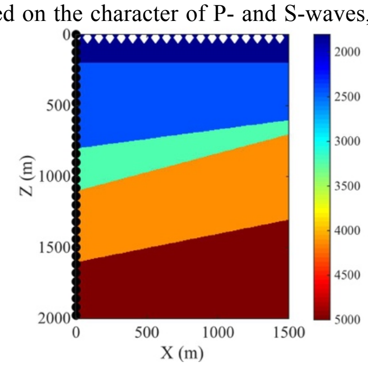

Figure 1: The P wave velocity (m/s) of a five-layered model and the distribution of the sources and receivers. Sources: white point near the surface; receivers: black point near x=0.

vergence and curl to calculate the ideal P-wave and S-wave during the simulation. This method takes 2D (3D in a three-dimensional simulation) spatial information into account, while the conventional method designed for real VSP data can only make use of 1D information along the well.

Number of layers	Number of models	Number of shot gathers	Number of training datasets
2	90	4590	2040
3	286	14586	4080
4	420	21420	8160
total	796	40596	22180

Table 1: Number of synthetic datasets.

We use the pairs of simulated VSP data to train the networks. The number of subsurface models and shot gathers are listed in Table 1. We use no more than 60% of the shot gathers of different types of layers. For the most complicated data produced from the four-plane-layers model, we use 8000 of these data for training, the rest for testing.

Network building

To customize networks that are suitable for this separation problem, we extract the feature maps of the hidden layers in the neural networks, to determine appropriate number and size of convolutional kernels in each layer. First, we choose U-net as a generator to train on a certain proportion of the synthetic data. U-net has been proven to perform well by using only a small number of training data, in Biomedical image segmentation (Ronneberger et al., 2015) and image-to-image translation (Isola et al., 2017). From our visualization of the hidden layers of the pre-trained U-net on our data, some of the layers do not show a positive effect. Because of the convolution with large kernels in more than eight layers, the size of the most inner hidden layer is rather tiny compared to the original input data. Meanwhile, the number of these kernels are too small to carry the adequate information to recover to the target P-wave field. According to the above analysis of the U-net, we rebuild our customized convolutional network in the generator of the cGAN with smaller, but a larger number of different kernels.

Training strategy

The final aim of this study is to make accurate P- and S-wave decomposition on real VSP data. Considering the noises and wave diversity and the limited quantity in real data, we begin training with the synthetic data. To train the separation network on a massive number of synthetic data, we employ a multi-scale strategy, in which simple waveforms are fed in first before more complicated waveforms. This strategy not only gives us a look at the learning capabilities of simple data, but also get to more complicated data step by step. After sufficient cycles of training and testing on synthetic data, we test the results on a synthetic dataset that the network has not seen before. The results show us that the performance of the network makes a relatively good prediction on near-offset shot gather. Then we continue to train the networks using this near-offset data pair and predict the far-offset wave decomposition results.

Evaluation method

To measure the performance of the separation results in a volume of more than 10 thousand pairs of data, efficient evaluation method is required. The value of the mean square error, which is commonly used to measure the difference the prediction and true results, requires an experiential threshold and is not intuitive for the vast amounts of data. As the separation work is not a typical classification or regression problem, we integrate six effective evaluation methods inheriting from that of both classification and regression to make an all-sided measurement. The six types of evaluation approaches are listed in Table 2. The class and envelope of the data is made by using activation function and Hilbert transform respectively.

Analysis of the results

After training GAN on 22180 pairs of synthetic data, we got a trained network to test on two types of data to show the capability of separating P-waves.

One type of testing datasets is generated under the flat layered model, with a total number of 40596. The performance of the trained neural network is measured by the evaluation method listed in Table 2. After searching for better-performed networks and cycles of training and testing, the average value of the six evaluations are high enough to reconstruct P-waves from elastic VSP data.

Evaluation type	Average	Maximal	Minimal
Accuracy (class)	0.9787	0.9925	0.9389
Precision (class)	0.8147	0.9720	0.5133
Recall (class)	0.9001	0.9996	0.5105
F1 score (class)	0.8488	0.9457	0.6581
R^2 for envelope	0.8044	0.9329	0.4065
R^2 for raw data	0.6582	0.7777	0.2335

Table 2: The evaluation of the testing results on four plane layer datasets.

Another type of the testing dataset is more complicated than any data in the datasets listed in Table 1. One of the subsurface models is shown in Figure 1, where the number of layers is more than four, and inclined layers take the place of flat layers. We choose the predicted results on the near-offset shot gather (x=0m) and the far-offset shot gather (x=1500m). Here we show the original elastic z component of the synthetic VSP data, the predicted P-wave made by GAN, the ideal P-wave made by applying divergence operator in Figure 2, where results on near-offset are in the top row, and results on far-offset are in the bottom row. To check the details of predicted waveforms, we extract the time series from the receiver at depth of 500m, 1000m and 1500m both in predicted P-wave by GAN and divergence operator. Shown in Figure 3, the contrast on the near-offset is in the left column, and far-offset is on the right column.

In the prediction of the near-offset shot gather, GAN gives out a close skeleton of the P-waves owning an over 90% accuracy rate from the P-wave extracted by applying divergence operator. Most of the waveforms made by GAN have the same positive or negative phase. However, the neural network is vulnerable to the high amplitude 'noises', which are not commonly seen in the training data. In the far-offset shot gather, the predicted phase of waveform performs worse than that in the near-offset shot gathers. This case may be caused by the dense data and large azimuth of incidence.

Conclusions

P-wave reconstruction for VSP data via cGAN

We propose a new P-wave reconstruction method for VSP data based on the optimal transportation theory, mapping elastic recorded VSP data to P-waves through the deep neural network. In the architecture of the Generative Adversarial Network, we set up an appropriate neural network for this problem with the help of visualizing the hidden layers of the pre-trained U-net. The training and testing data are synthetically produced directly from the wave equation and the divergence and curl operator. Compared to the separation approaches using domain transform, the trained networks also can be regarded as a separated filter, which is free from subjective parameters and strong assumptions. From our training and testing on 22180, 40596 pairs of synthetic data, we demonstrate the capability of the machine learning method in predicting waveform signals beyond conventional phase recognition or wave classification.

Acknowledgments

This work was supported in part by the National Key R&D Program of China (Grant No. 2017YFA0604500), by the National Natural Science Foundation of China (Grant No. 51761135015), and by Center for High Performance Computing and System Simulation, Pilot National Laboratory for Marine Science and Technology (Qingdao). Dr. Yunyue Elita Li and Dr. Jizhong Yang are supported by the MOE Tier 1 Grant R-302-000-165-133 and R-302-000-182-114.

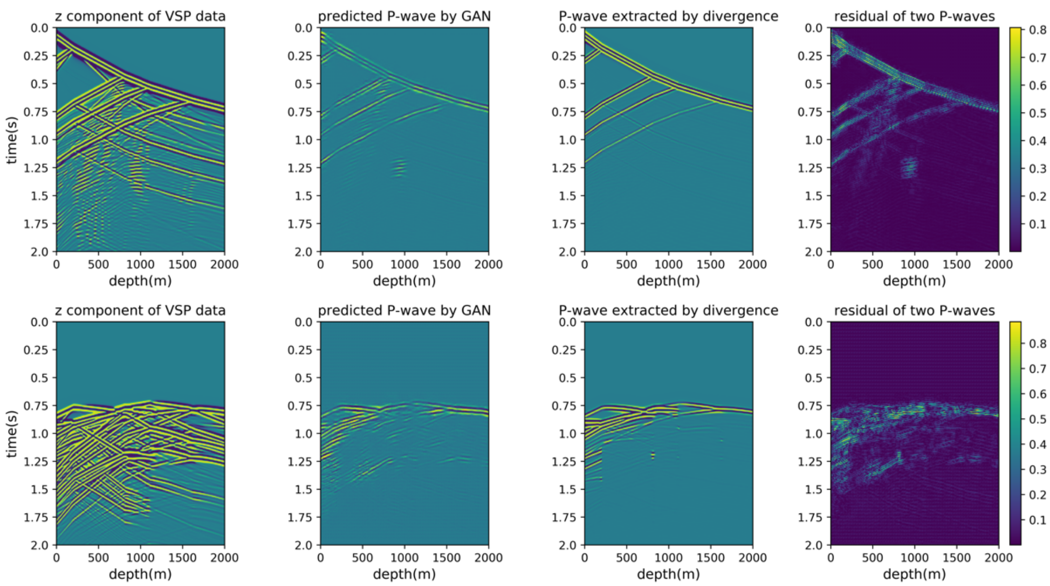

Figure 2: The results testing on synthetic data by GAN and contrast with the P-waves extracted by divergence operator.

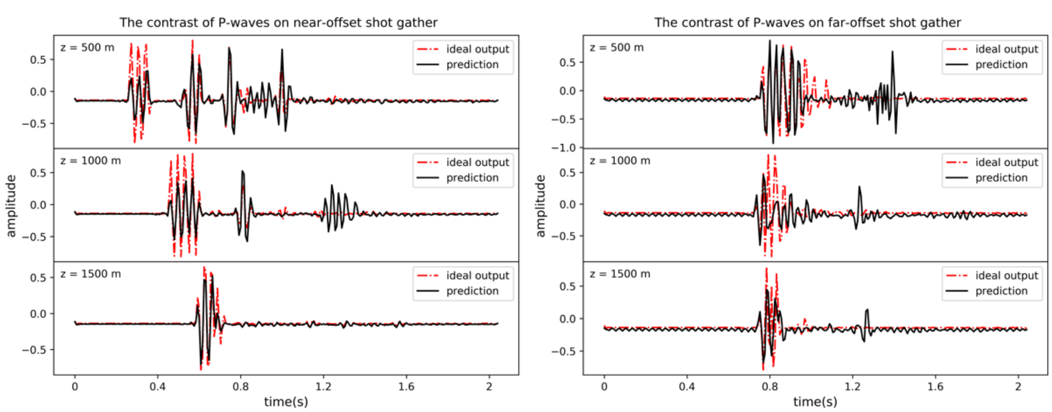

Figure 3: The signal contrast between P-waves extracted by divergence operator and GAN in the depth of z = 500m, 1000m, 1500m.