Th_R11_09

The First-Break Detection For Real Seismic Data With Use of Convolutional Neural Network

G. Loginov[1,2]*, D. Anton[1,2], D. Litvichenko[3], S. Alyamkin[4]

[1] Novosibirsk State University; [2] Institute of Petroleum Geology & Geophysics SB RAS; [3] Gazpromneft NTC Ltd.; [4] Expasoft Ltd.

Summary

In this study, we appraise a convolutional neural network for the detection of the first breaks on the real 3D seismic data set. The use of convolution as a learning kernel is followed by an assumption that the seismic trace can be considered as a convolution of source signal with the reflectivity function. The investigation area includes mixed elevations, floodplains of the rivers and the regions of strong permafrost, where the shingling effect is observed. We consider the first-break detection for each trace independently to preserve the complicated structure of the arrival times. The proposed approach was apprised on real exploration 3D seismic data set with size over 4.5 million traces. This test showed that the error between the original and predicted first breaks is not more than 3 samples for 95 percents of data set. The final quality control of picking results was established by the calculation of static corrections and computing seismic stacks, which showed that the proposed approach provides better results.

Introduction

The first breaks are the basic data to restore the near-surface velocity model and calculate static corrections. The static corrections highly depend on the first break picking reliability and are essential for the effectiveness of reflection- and refraction-based methods of seismic processing (Yilmaz, 2001). Conventionally, refraction tomography is used to build the velocity model from the first breaks. Due to a large volume of seismic data the detection of first breaks of the wave signal should be precise, rapid and automated (Sabbione and Velis, 2010). The wave field of first arrivals usually consists of direct, head and refracted waves. First breaks detection accuracy is influenced by the quality of seismic records and the complexity on the velocity model. Due to the complexity of the wavefield and high uncertainty of the detection of the first breaks, any picking algorithm provides uncertainty, which can be resolved only while further inversion procedures (Akram and Eaton, 2016).

The majority of algorithms for first-break picking are based on signal energy analysis in a sliding window (Allen, 1978). Maity et al. (2014) proposed the fully-connected neural network with use of many features extracted by traditional algorithms of detection. The convolution network introduced in (Yuan et al., 2018) and suited for waveforms classification for the exploration seismic images. In (Zhu and Beroza, 2018) authors aimed to classify the waveforms of seismological and microseismic events. The majority of of proposed approaches relies on deep learning that implies huge training set and big number of hidden layers in a neural network.

In this study, we apprise convolution neural network for the picking of the first breaks on the real 3D seismic data set. We aimed to use a simple architecture of the neural network that could be reproducible in a production. The total set size is about 4.5 million traces. The exploration area is complicated by strong heterogeneities, associated with the presence of watering, permafrost and thawing. The wave types identification is sophisticated. In some regions of strong permafrost the shingling effect is observed. In addition, the survey area includes mixed elevations and floodplains of the rivers. To estimate the performance, we calculated the static corrections and evaluate final stacks quality.

Method

We consider the convolution neural network (CNN) approach for automatic first break picking. The use of convolution as a learning kernel is followed by an assumption that the seismic trace can be considered as a convolution of source signal with the reflectively function. The profiles of first arrivals can be rather complicated due to the complexity of the near-surface velocity model. We pick first breaks independently for each trace to preserve complicated structure of the arrival times. Further in this section we discuss main steps of proper CNN model development: the input data processing and markup, the choice of architecture and loss function, introducing regularization procedures, CNN output processing.

Picking of the first breaks as a classification problem. The original data markup is a binary matrix with number of rows corresponding to number of seismic trace samples and number of rows is number of classes. The element of a matrix equals 1, where the trace sample belongs to certain class. The standard machine learning approach would be to introduce two classes (binary classification): first break and non-first break. This binary classification provides a strong imbalance of true (first break) and false samples per seismic trace. A standard technique to overcome this problem is the weighting of classes in the loss function while the CNN training. But, such a strategy would demand an additional experimental research of optimal weights. We propose an alternative approach and consider 3 classes: noise, the first break and the signal. According to the wide range of traces with different offsets used for training CNN, we have a well balance of at least two classes: noise and signal. The results of numerical tests showed that such approach provides more robust detection of first break. Since the classes are balanced, the cross entropy loss function is used in its original form.

Input data processing. The seismic data processing accounts dozens of procedures to increase the signal to noise level. Basically, the procedures are bandpass filtering, gain correction, traces stacking, moveout correction, etc. We put the effort to minimize the set of processing procedures for first break detection.

Thus, in the following sections we consider only mean value of trace subtraction (detrend) and division by difference between minimum and maximum value per trace (normalization). As to the processed trace values belong the interval of (-1, 1), the learning of CNN is simplified and speed of convergence is improved. In addition, the proposed processing does not strongly affect on the wavefield consistence and avoid artifacts origination.

Architecture. For the majority of cases, the CNN model can be developed only while experimental testing of different architectures. Conventionally, the architecture term mainly implies the set of hidden layers. In this paper, we consider following procedures of the layer: convolution, batch normalization, activation and dropout. Each procedure has the set of its hyper parameters, selected while the numerical experiments. The parameters of convolution are: number of filters, its length and padding type. The batch normalization procedure is used to provide zero mean and unit variance of the input data. It is performed by adjusting and scaling factors, that are learned while CNN training. Conventionally, for the hidden layers is recommended to use *ReLU* (rectified linear unit) activation function. Finally, the dropout technique is used to reduce overfitting, with default parameter (rate) to be 0.5.

Model evaluation and QC. The CNN model application to input seismic trace is the matrix (predicted) with same size as the original markup matrix. The matrix element is the probability value of the trace sample to be one of three classes: noise, first break, signal. The first break pick is the *argmax* of a matrix column that belongs to first break class. The CNN model metrics is the mean absolute error (MAE) between original and predicted markup matrices. To estimate the accuracy of the proposed approach we measure the MAE between original and predicted first breaks. It is generally accepted that the problem of first break picking has a big uncertainty due to complexity of signal function and velocity model uncertainties. In addition, the original data markup can contain the errors artifacts. The zero error between original and predicted picks would not describe the performance of CNN model precisely. Thus, we set a window of acceptable discrepancy between original and predicted picks and calculate the percentage of errors that belong to this window. Finally, we calculated the static correction by the predicted picks and performed the data stacking to illustrate the CNN model performance.

Experiments

The training set is a real 3D exploration seismic data with explosive source. The seismic survey contained offsets up to 1200 m over 30000 sources on the area of a 1000 km^2. The total number of traces is about 4.5 millions with the sampling step of 2 ms. The original picks of the first breaks were obtained using the "CGG Geovation" software. The original first breaks are verified by internal production QC metrics and further processing workflow was successfully performed: near-surface tomography, static corrections, stacking, migration etc. We aimed to develop a CNN model that would be suitable for different seismic sets, at least for the same source type and similar survey conditions. We aimed to develop the CNN model with minimal set of layers to support the ability of retraining and reproducing the CNN.

We tested different size of training set and number of CNN layers. The trained CNN was evaluated over the full data set (4.5 million traces). The final CNN model architecture is presented in table 1. The CNN hidden layers contains set of operations. First, convolution with 32 filters of 32 sample length is applied. The convolution output is activated by *ReLU* function, that chousen experimentally. Then, the batch normalization and dropout techniques are performed. The output layer of CNN consider convolution layer with 3 filters with *sigmoid* activation function, that provides classification of input trace samples into 3 classes: noise, first break and wavefield. The evaluation results of trained CNN with a different number of layers and the size of the training set are presented in table 2. The table 2 presents percentage of picks detected with absolute error of not more than 3 samples. It can be seen that starting from 4 hidden layers and 5 thousands train examples, the accuracy almost stable. According to the test results, the CNN with 4 hidden layers was chosen.

The most adequate way for picking quality control is static corrections calculation and data stacking. In the figure 1 the data stacks are presented: on the left - stacking with original first breaks and right - the CNN results; upper palette presents results for the training set and the bottom one - for the test

Layer	Procedure	input size	output size
	Input	$n \times 1$	
1..k	Convolution (32 filters of 32 samples length) Activation ReLU Batch normalization (with scaling and shifting) Dropout (with rate of 0.5)	$n \times 1$	$n \times 32$
k+1	Convolution (3 filters of 32 samples length) Activation *sigmoid*	$n \times 32$	$n \times 3$
	Output		$n \times 3$

Table 1 *The proposed architecture of convolution neural network, where n - number of input trace samples, k - number of hidden layers.*

	Training set size				
No. of CNN Layers	**5000**	**10 000**	**25 000**	**50 000**	**100 000**
1	83.5	83.6	N/A	N/A	N/A
2	91.0	91.3	N/A	N/A	N/A
3	93.0	94.7	N/A	N/A	N/A
4	94.3	95.0	95.3	95.6	96.0
5	94.5	95.7	95.7	95.9	96.1
6	94.1	95.8	95.7	96.0	96.0
7	95.3	95.1	95.7	95.9	95.8

Table 2 *The percent of first-break picks with accepted error (not more then 3 samples) predicted by CNN models, trained with a different size of the training set and the number of hidden layers.*

set. One can see, that the stack calculated by the predicted first breaks provides more flattened horizons, higher contrasts of the reflectors and highlights more details of the image. This example illustrates the applicability of the proposed approach and proves successful apprising on the real seismic data.

Conclusions

The first break picking algorithm, based on convolution neural network is proposed. Different depth of neural network and training set is tested. As a results of this tests we showed that 4 hidden layers and train set of 5 thousands traces is enough for first-break picking sufficient quality. The proposed approach was apprised on real exploration 3D seismic data set with size over 4.5 million traces. This test showed that the error between original and predicted first breaks is not more then 3 samples for 95% of data set. The final quality control of picking results was established by the calculation of static corrections and computing seismic stacks, which showed that the proposed approach provides better results.

Acknowledgements

The reported study was funded by RFBR according to the research project 18-35-00412. The authors thank ”GAZPROMNEFT NTC LLC“ for the opportunity to test the proposed approach on real data and publish the research results.

References

Akram, J. and Eaton, D.W. [2016] A review and appraisal of arrival-time picking methods for downhole microseismic dataArrival-time picking methods. *Geophysics*, **81**(2), KS71–KS91.

Allen, R.V. [1978] Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, **68**(5), 1521–1532.

Maity, D., Aminzadeh, F. and Karrenbach, M. [2014] Novel hybrid artificial neural network based autopicking workflow for passive seismic data. *Geophysical Prospecting*, **62**(4), 834–847.

Sabbione, J.I. and Velis, D. [2010] Automatic first-breaks picking: New strategies and algorithms. *Geophysics*, **75**(4), V67–V76.

Yilmaz, Öz. [2001] *Seismic data analysis*, 1. Society of Exploration Geophysicists, Tulsa.

Yuan, S., Liu, J., Wang, S., Wang, T. and Shi, P. [2018] Seismic waveform classification and first-break picking using convolution neural networks. *IEEE Geoscience and Remote Sensing Letters*, **15**(2), 272–276.

Zhu, W. and Beroza, G.C. [2018] PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method. *arXiv preprint arXiv:1803.03211*.

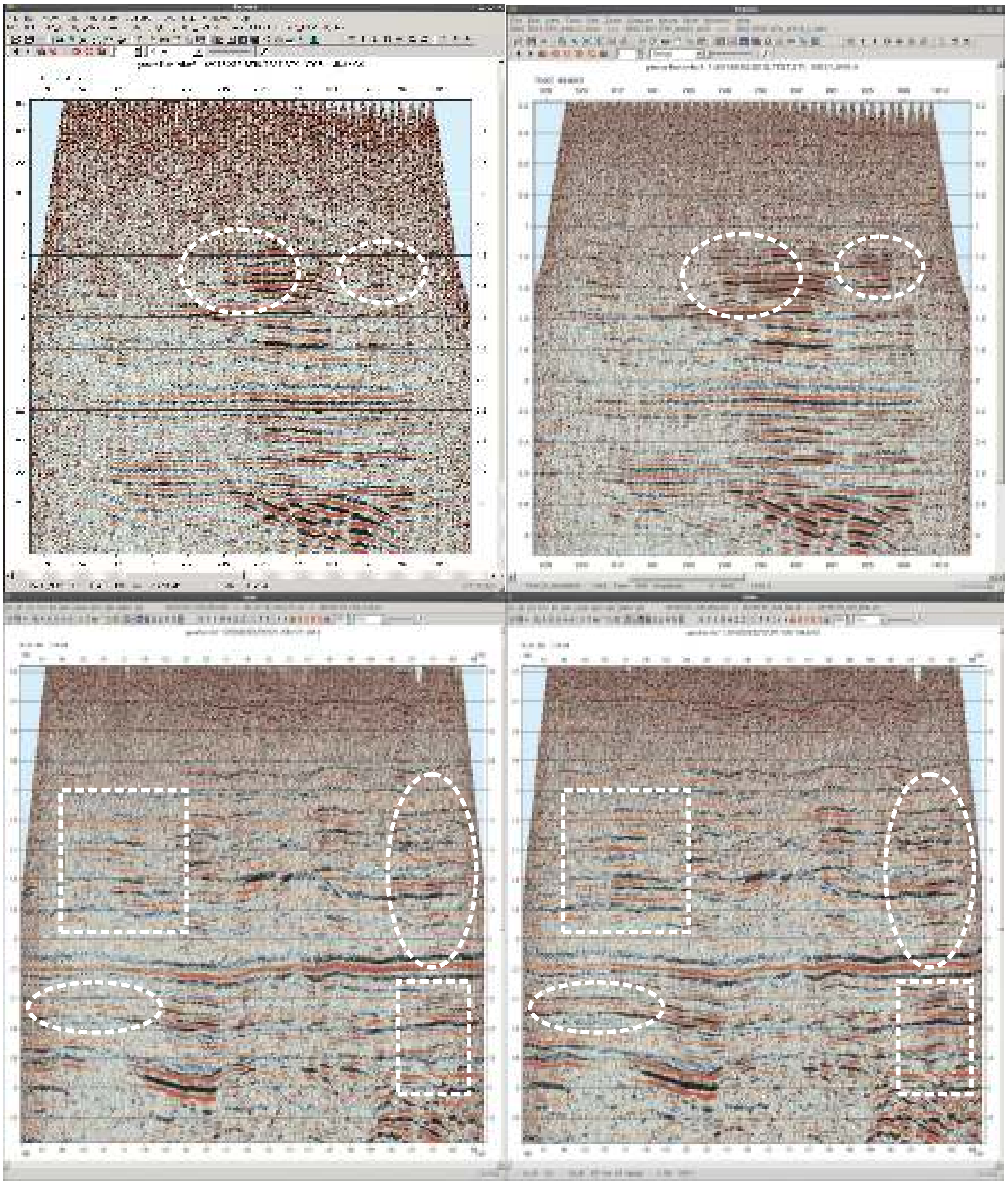

Figure 1 *Data stacks calculated with different static corrections: left — original first breaks; right — predicted. White shapes highlights the most differing futures of stack image. The upper palette — for the training seismic set; bottom — test set.*