Выделение первых вступлений на сейсморазведочных данных с помощью нейронных сетей

А.А. Вязниковцев\* (МГУ имени М.В. Ломоносова, ООО «НЕОГЕН»), А.Н. Ошкин (МГУ имени М.В. Ломоносова, ООО «НЕОГЕН»), С.В. Колпинский (Некоммерческий Фонд развития науки и образования «Интеллект»)

Введение

В настоящей работе приведены первичные результаты научного исследования, посвященного автоматизации процесса выделения первых вступлений на сейсморазведочных данных. Сейсморазведка является обслуживающей дисциплиной в отрасли поиска, разведки и добычи полезных ископаемых и относится к блоку геологических дисциплин.

Актуальность данной работы обусловлена тем фактом, что выделение первых вступлений является первоочередной и неотъемлемой задачей в рамках обработки сейсморазведочных данных любого типа (акваторная, наземная, скважинная), отрасли (нефтяная, инженерная) и объема. В свою очередь, объем данных, получаемых в ходе сейсморазведочных работ даже на сравнительно небольшом участке настолько велик, что он не может быть полноценно проанализирован обработчиком в устанавливаемые на производственном процессе сроки. Именно поэтому процесс выделения первых вступлений при производстве сейсморазведочных работ проводится лишь на небольшой части данных, что, безусловно, способствует скорости обработки, но потенциально может ухудшать итоговый результат.

Целью данной работы является создание нейронной сети, позволяющей полностью автоматизировать процесс выделения первых вступлений, уменьшить трудо-; и времязатраты в производстве, а также улучшить качество сейсморазведочных результатов.

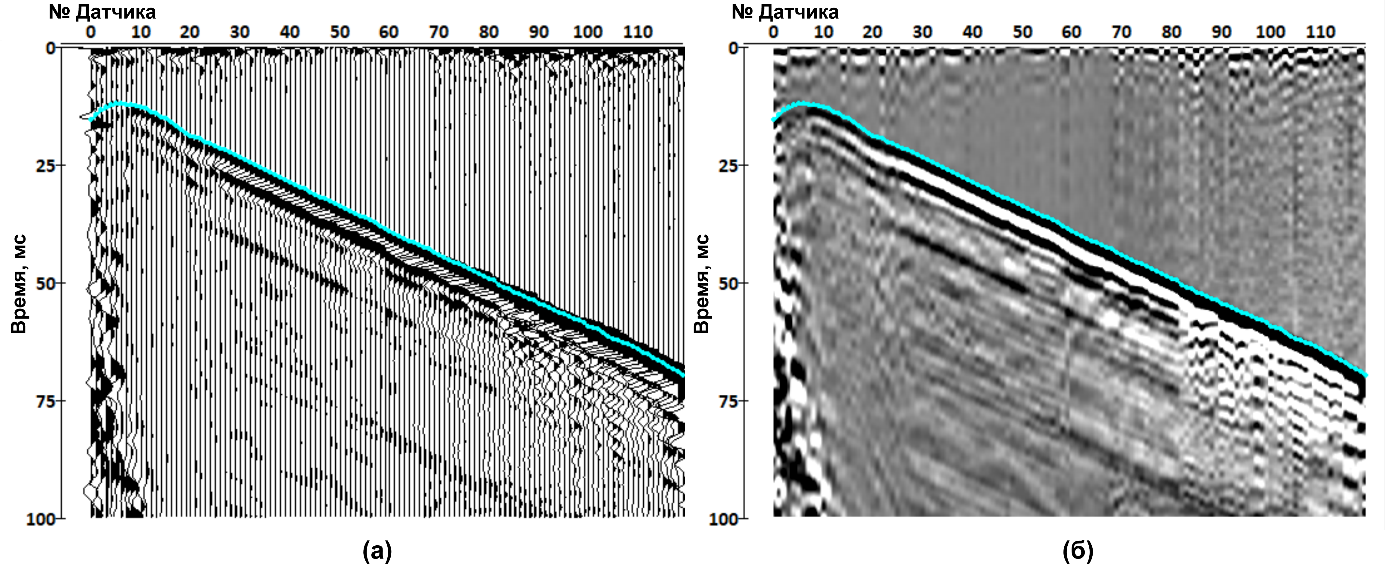
Устройство сейсморазведочных данных

Элементарной единицей информации в сейсморазведке служит временной ряд, соответствующий записи амплитуд некоторого искусственно возбуждаемого волнового процесса с помощью датчика скорости смещения грунта в каждый момент времени. Подобный ряд именуется сейсмической трассой (или сейсмотрассой):

,

Однако более распространенной единицей информации чаще всего служит не отдельная сейсмотрасса, а их совокупность, соответствующая некоторому ансамблю близко расположенных и упорядоченных датчиков:

Таким образом, единичная сейсмограмма чаще всего аналогична по конфигурации одноканальному изображению и представляет собой двумерный массив 32-битных чисел. Типичный пример сейсмограммы приведен на рис. 1.



1. Пример типичной сейсмограммы в потрассном (а) и непрерывном режиме визуализации (б). Вертикальная ось соответствует времени записи от начала возбуждения волнового процесса, горизонтальная ось – номерам датчиков. Голубой линией отмечены моменты первого вступления.

Постановка задачи

Как правило, начальный момент времени на сейсмограмме соответствует моменту возбуждения искусственного волнового процесса. В таком случае задачу выделения первых вступлений можно сформулировать как задачу определения моментов времен , соответствующих времени наискорейшего распространения волнового процесса от пункта возбуждения до датчика . Необходимо отметить, что по принципу Ферма, существует единственный подобный момент для каждого .

В работе [3] данная задача рассматривалась в качестве задачи бинарной классификации (ответ на вопрос, успел ли распространиться процесс за время в рассматриваемом отсчете?). Также существует [2] модель, решающая данную задачу как задачу сегментации.

Формирование обучающей выборки

На начальном этапе работы с целью выбора архитектуры модели и отлаживания процесса обучения, предложено использовать синтетические данные, полученные с помощью конечно-разностного моделирования [1] и аугментаций, включающих в себя искажения сейсмограмм посредством добавления белого шума различного уровня, волн-помех с моделью электромагнитной природы (Э/М наводки), линейный шумов (случайное семлирование точечных Фурье-спектров), а также семплирование шумов с реальных данных. На текущий момент сгенерировано и размечено порядка 50 тысяч сейсмограмм (около 50 Гб).

Формирование тестовой выборки

Хотя и продуманные аугментации способствуют формированию синтетических сейсмограмм, в высокой степени схожих с реальными, тестовая выборка должна полностью состоять из реальных данных. На текущем этапе работы автором накоплен существенный объем реальных данных, полученных при выполнении производственных задач (около 5 Гб). Накопленные данные, однако, не имеют полноценной и качественной разметки. Таким образом, разметка реальных данных послужит одной из постоянных задач в данной работе.

Метрики качества

Наиболее подходящей метрикой при работе с прогнозированием моментов времен может послужить точность (accuracy), рассчитанная при допустимой погрешности в N отсчетов. В рамках данной метрики момент времени считается предсказанным верно, если предсказание отстоит от истинной отметки не далее, чем на N отсчетов. Например, в работе [3] с применением CNN-архитектуры для решения аналогичной задачи была достигнута 87% точность при допустимой погрешности в 12 отсчетов.

Архитектура модели

Первичную модель предложено построить на основе архитектуры ResNeXt. Однако в отличие от классической архитектуры, представленной в работе [4] и предназначенной классификации изображений, необходимо выбрать параметры слоев таким образом, чтобы в процессе прохождения через модель, входная сейсмограмма сохраняла пространственные размеры (использовать сжатие только по оси времен) и заменить головной модуль модели.

Планирование дальнейшего исследования

Ближайшими задачами в рамках работы над исследованием послужат доработка первичной архитектуры модели и отладка базового процесса обучения. Указанные выше задачи планируется полностью выполнить к воркшопу № 2.

По причине доступности достаточного количества данных, задачи пополнения обучающей выборки и составления тестовой выборки являются второстепенными.

Библиография

1. Louboutin, M. and Lange, M. and Luporini, F. and Kukreja, N. and Witte, P. A. and Herrmann, F. J. and Velesko, P. and Gorman, G. J. (2019) Devito (v3.1.0): an embedded domain-specific language for finite differences and geophysical exploration. Geoscientific Model Development, 12(3), 1165-1187. doi: 10.5194/gmd-12-1165-2019
2. Tarasov Aleksei. DaloroAT/first\_breaks\_picking [электронный ресурс]. URL: <https://github.com/DaloroAT/first_breaks_picking> (дата обращения 20.06.2023)
3. Yuan, Sanyi; Liu, Jiwei; Wang, Shangxu; Wang, Tieyi; Shi, Peidong (2018). Seismic Waveform Classification and First-Break Picking Using Convolution Neural Networks. IEEE Geoscience and Remote Sensing Letters, 15(2), 272–276. doi:10.1109/LGRS.2017.2785834
4. Zhuang Liu and Hanzi Mao and Chao-Yuan Wu and Christoph Feichtenhofer and Trevor Darrell and Saining Xie. (2022) A ConvNet for the 2020s. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.48550/arXiv.2201.03545