

Машинное обучение при обработке сейсмических данных

Команда:

Няго Ю. А.

Радеев Н. А.

Квашин М. Б.

Мокроусов А. С.

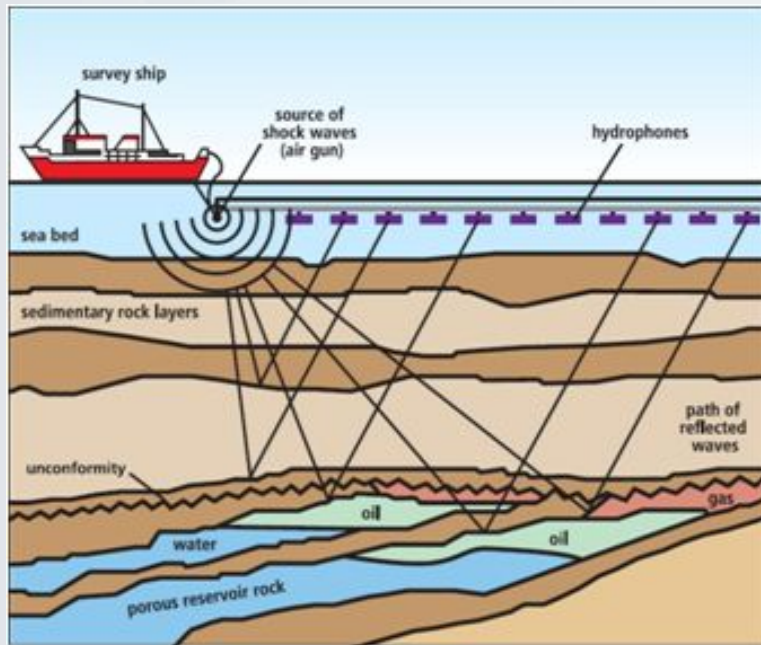
Иванишкин Д. С.

Солодовников М. А.

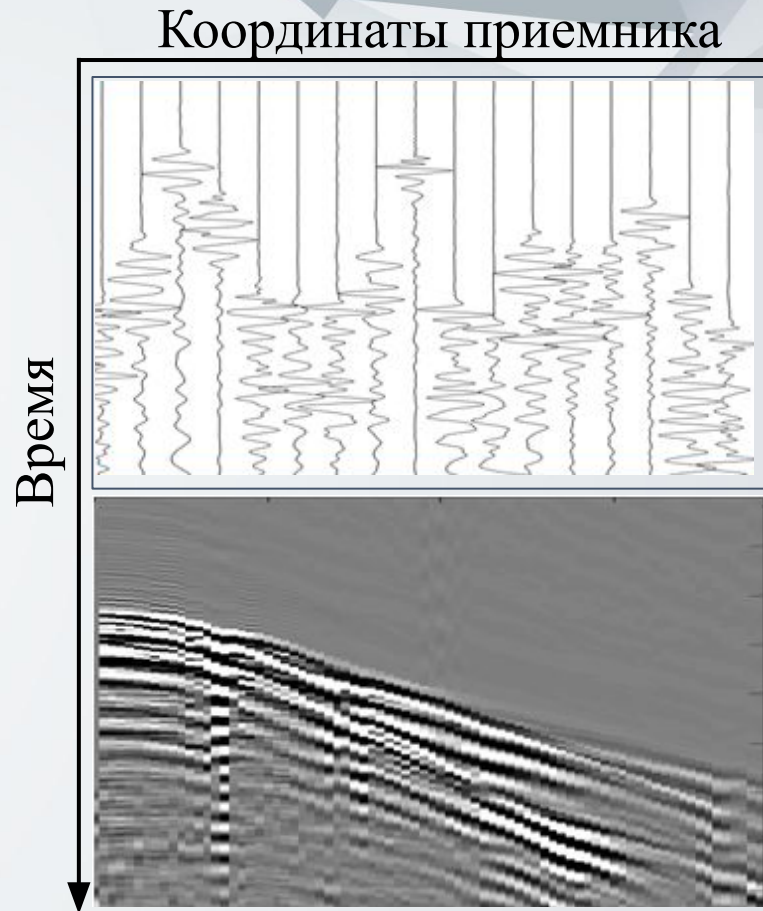
Руководитель:

Дучков А. А.

Исследование структуры Земли



- Большие объемы данных
- Необходимость автоматизации для оперативной обработки



Цель

- Автоматическое снятие времени вступления сейсмической волны с помощью нейросети

Задачи

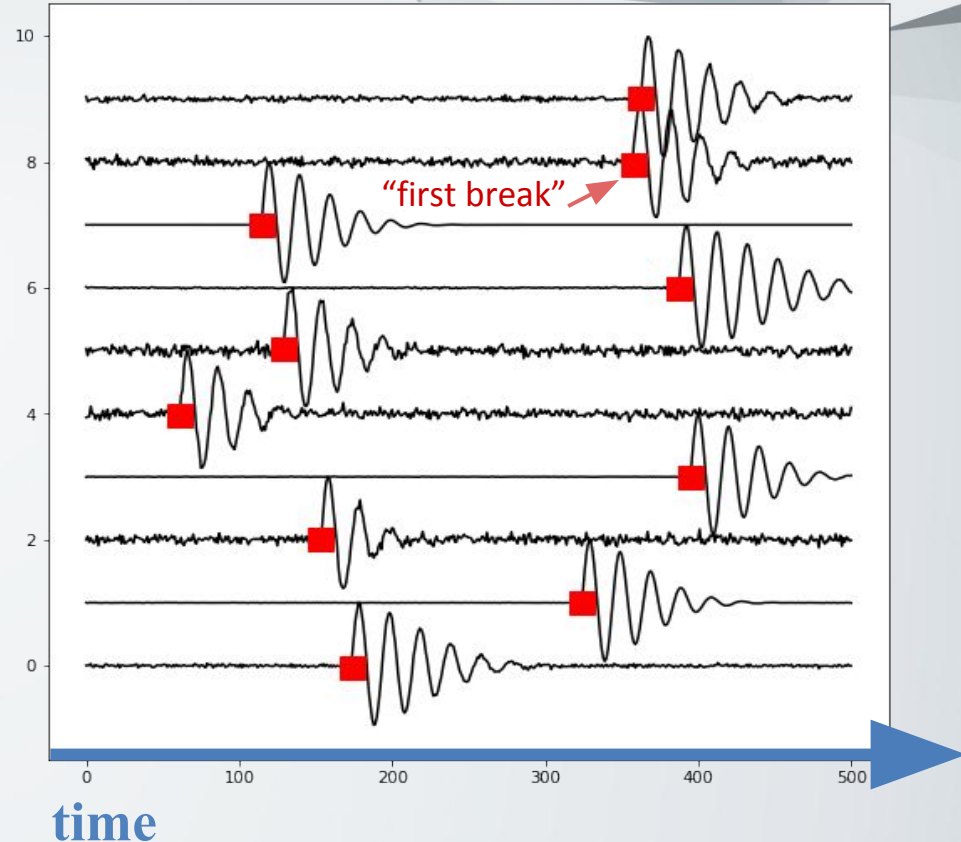
- Проверить эффективность нейросетей, обученных с помощью синтетических данных
- Подобрать наилучшую конфигурацию нейросети
- Минимизировать объем вручную размечаемых данных обучающей выборки

Методы

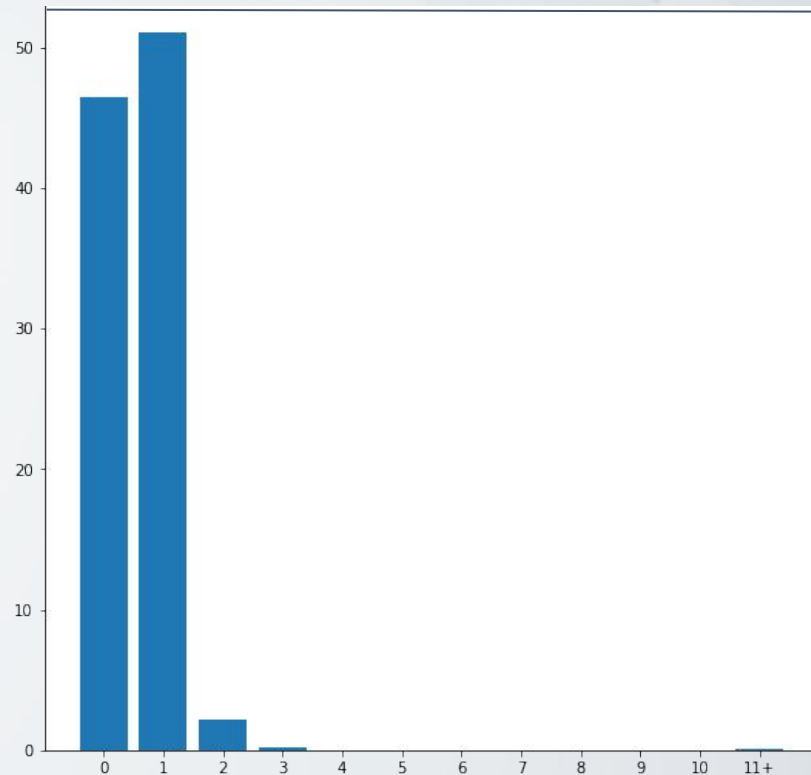
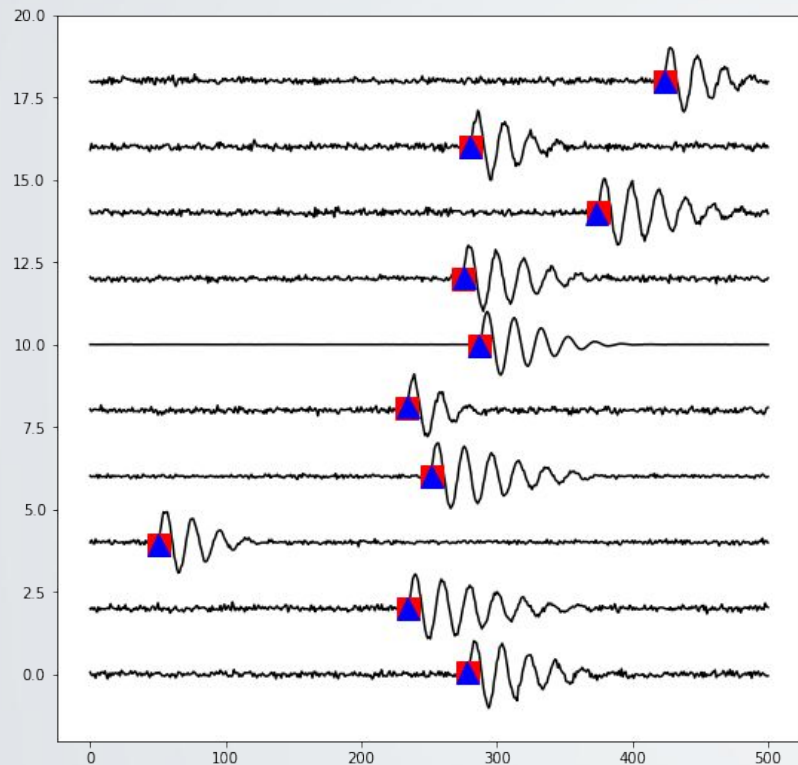
1. Рассмотреть полносвязные, рекуррентные и сверточные нейронные сети в различной конфигурации.
2. Провести их обучение и тестирование на синтетических и реальных сейсмических данных:
 - Синтетические данные получены с помощью генератора с регулируемыми параметрами: частота колебаний, скорость затухания, шум.
 - Реальные данные: 3366 трасс длиной 501 отсчет.

Метод машинного обучения

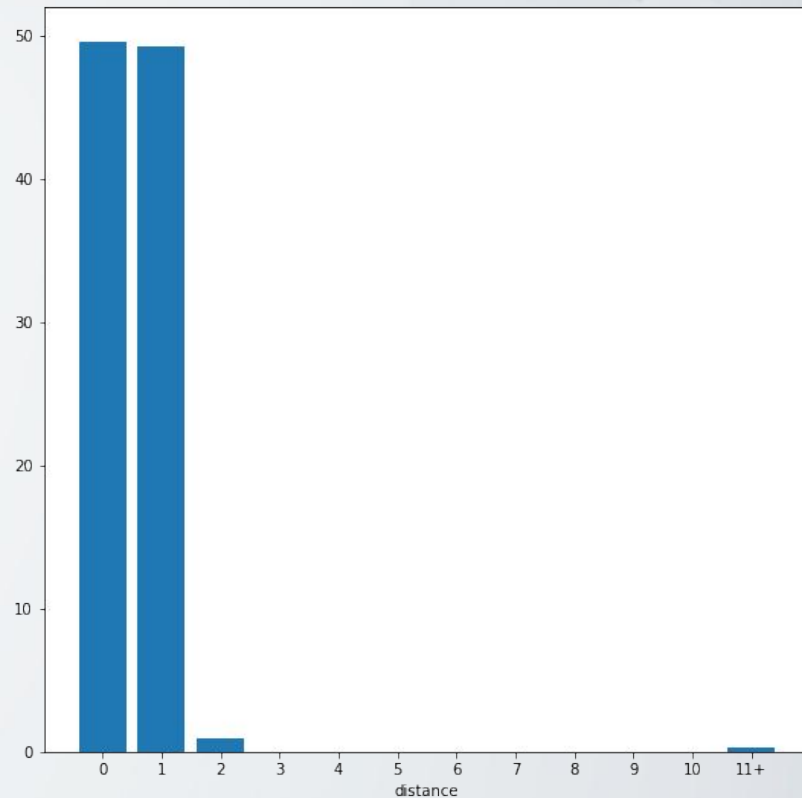
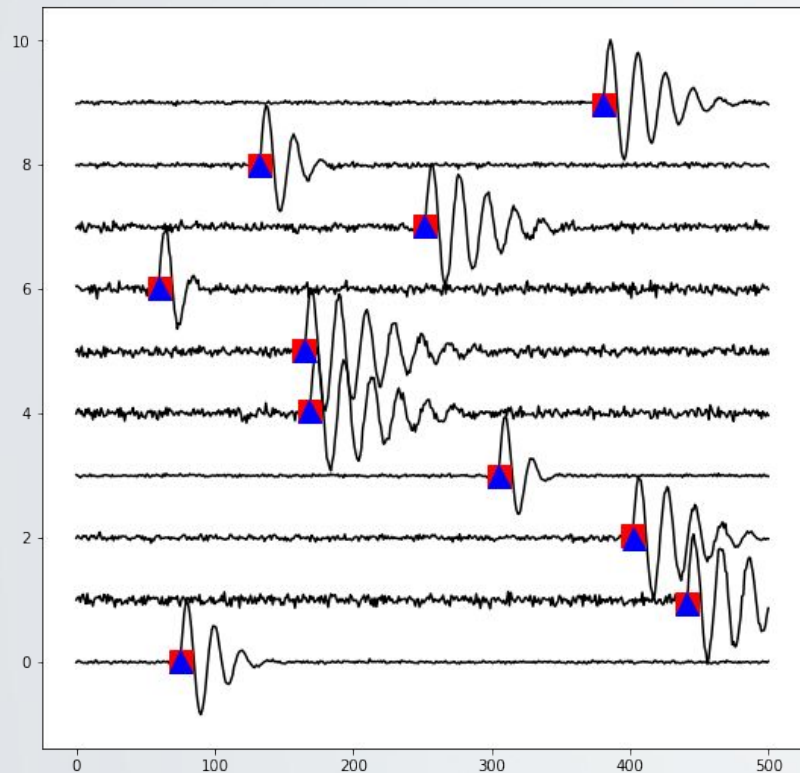
- Подготовить обучающую выборку.
- Обосновать выбор нейронной сети.
- Обучить и протестировать нейронную сеть.



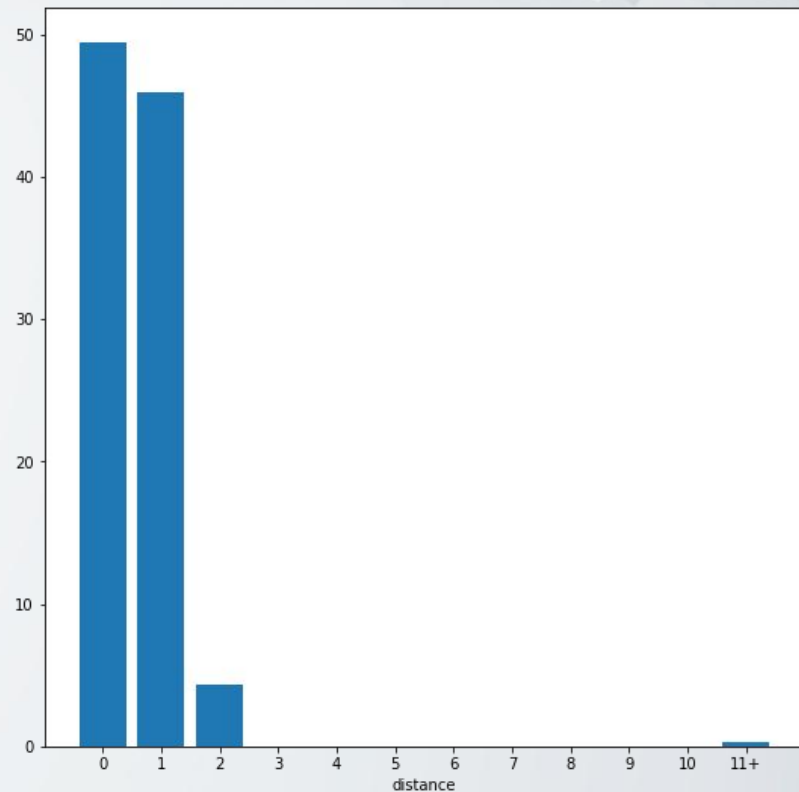
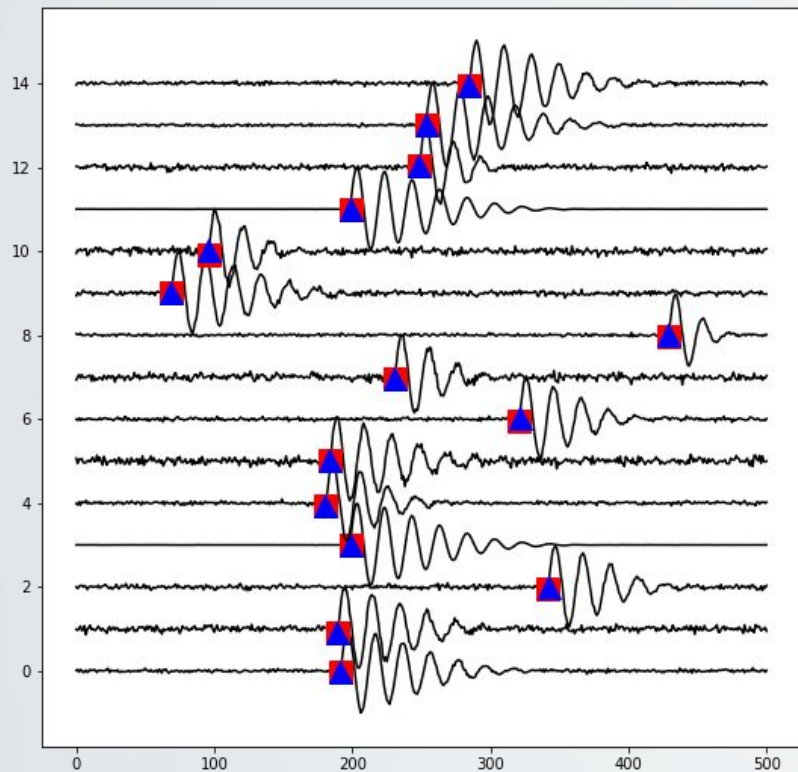
Синтетические данные: полносвязная сеть (99%)



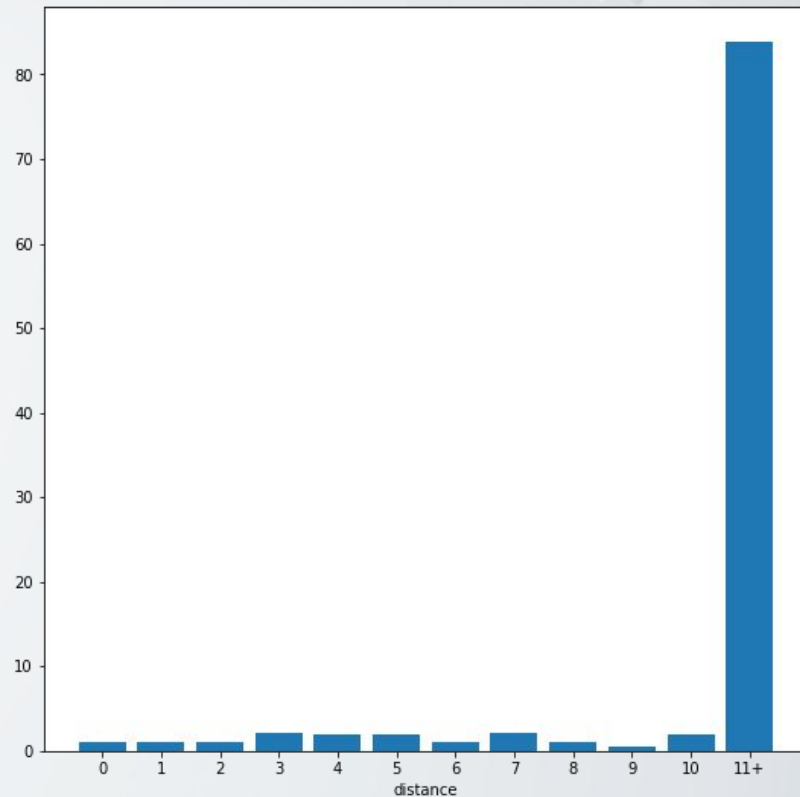
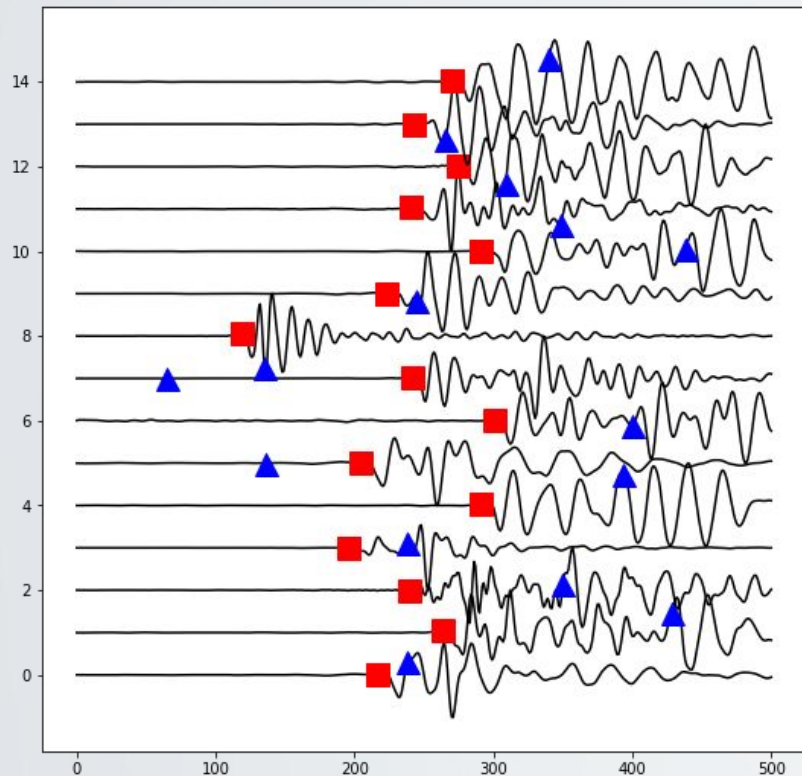
Синтетические данные: сверточная сеть (98%)



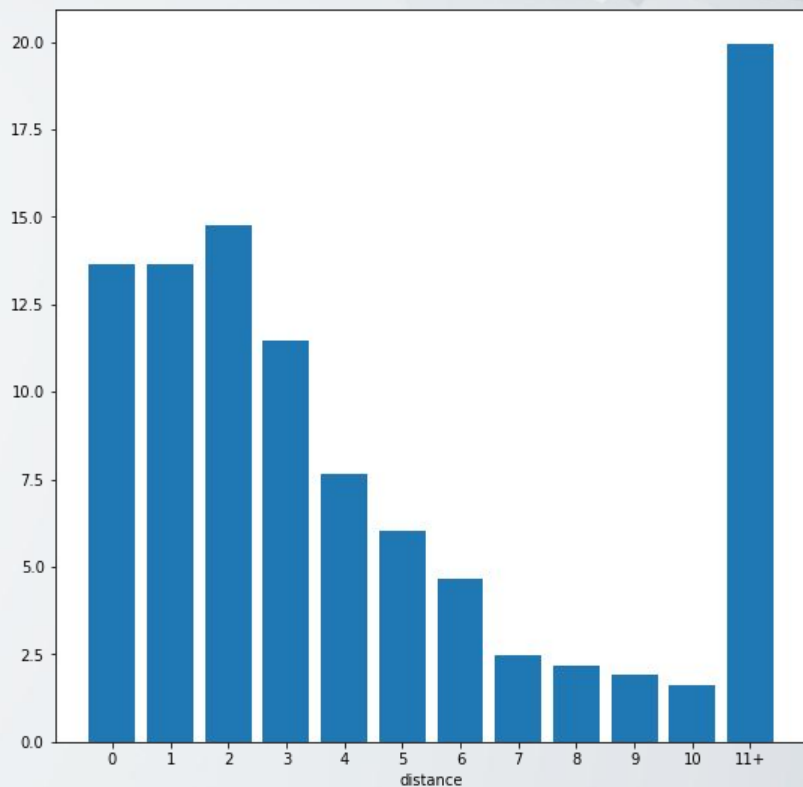
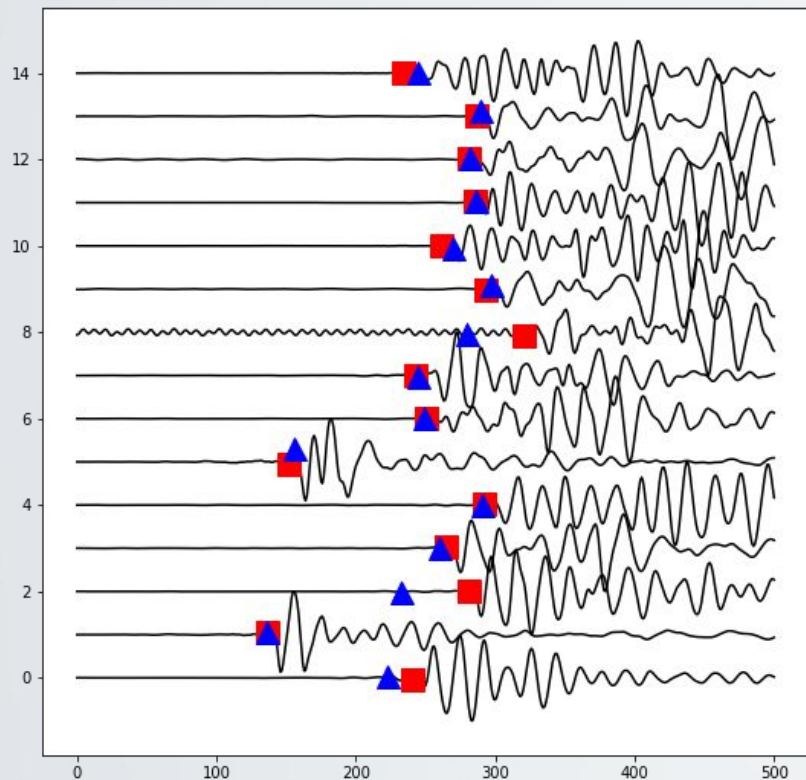
Синтетические данные: рекуррентная сеть (99%)



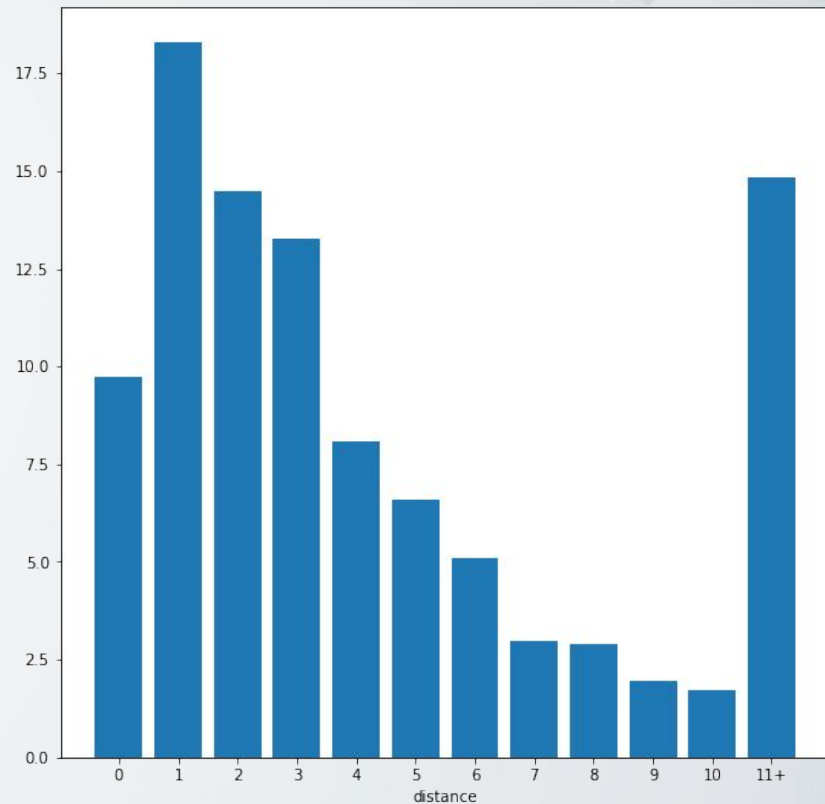
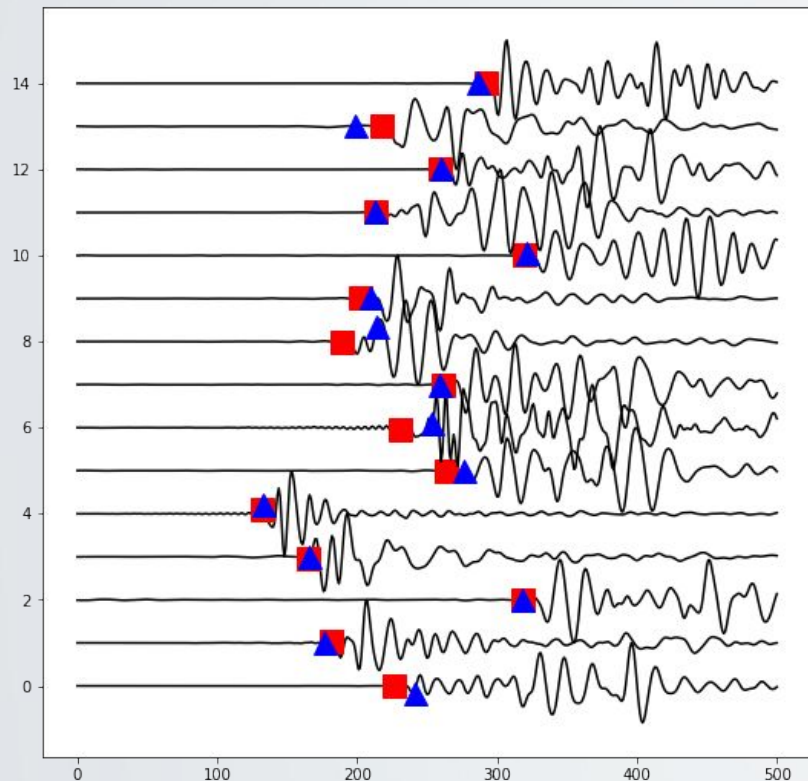
Обучение на синтетике, тестирование на реальных данных: рекуррентная сеть (16%)



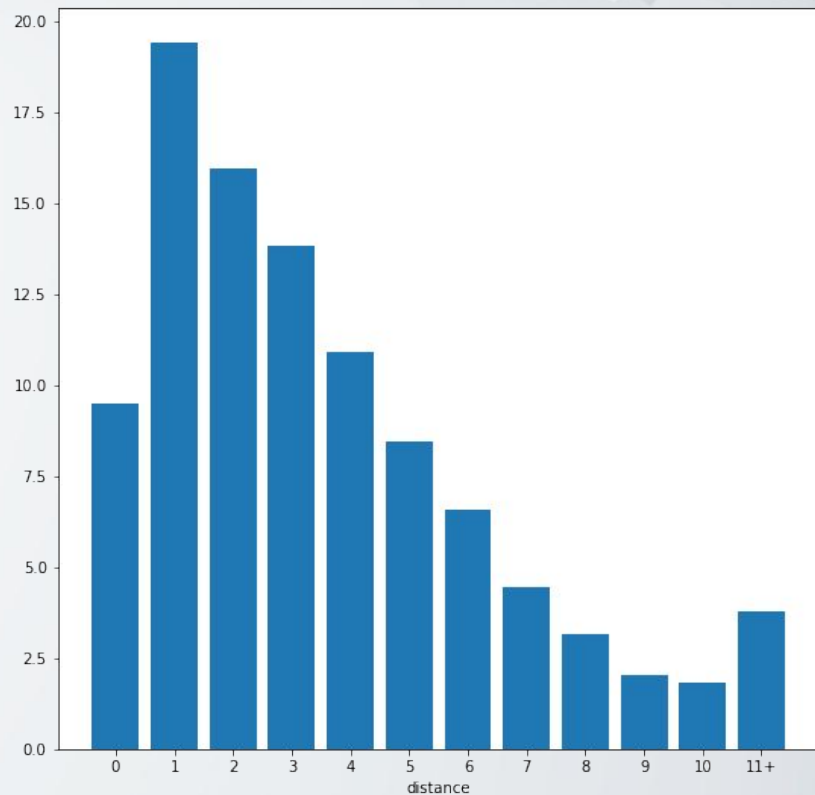
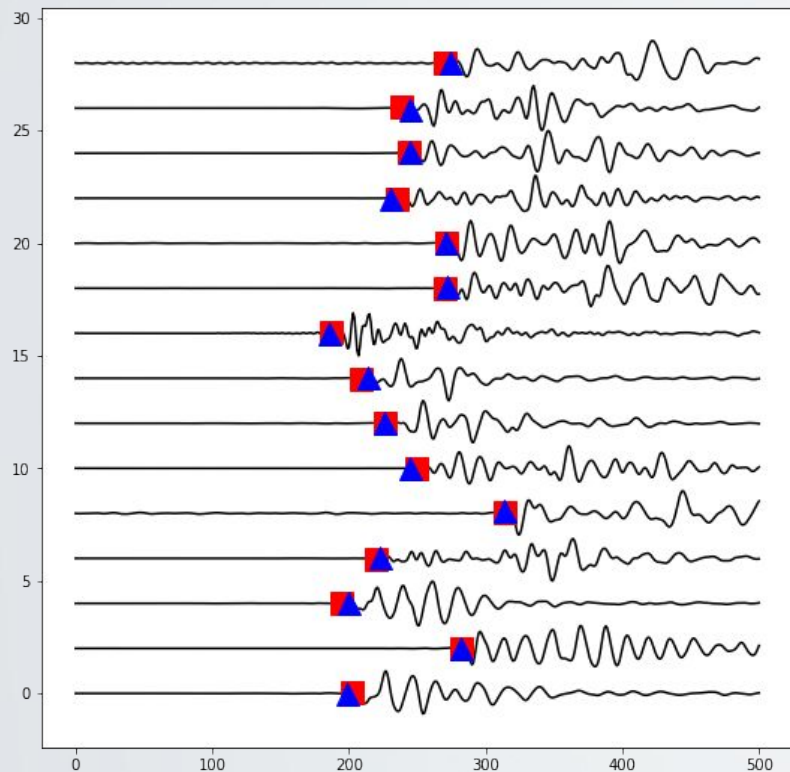
Реальные данные: рекуррентная сеть (80%)



Реальные данные: сверточная сеть, 1 слой (85%)



Реальные данные: сверточная сеть, 4 слоя (97%)



Оптимизация размера выборки обучающих данных

Размер обучающей выборки	Точность
1500	97%
1250	92%
1000	89%
750	83%

Выводы

- Сети с любой конфигурацией показывают хорошую точность при работе с синтетическими данными (98-99 %)
- Нейросеть, обученная на синтетических данных, некорректно работает с реальными данными (точность 16%)
- При работе с реальными данными наилучший результат показала сверточная нейросеть с 4 слоями (97%)
- Оптимальный размер выборки обучающих данных: 1500 треков

Как улучшить решение

- Решить проблемы генератора синтетических данных:
 - Распределение шума не соответствует реальным данным
 - Фиксированная начальная фаза сигнала
 - Форма сигнала далека от реальной
- Продолжить эксперименты с макропараметрами сети:
 - Повышение точности
 - Минимизация необходимого размера обучающей выборки

Ссылка на Github

Здесь лежит информация, требуемая для продолжения работы над проектом и его улучшением.

А также минимально необходимая литература на русском языке.

<https://github.com/dark0mike/Project-Geophysics/tree/draft>

Спасибо за внимание!