

Отчет по проекту "Применение рекуррентных нейронных сетей для дебита нефти при газовой закачке в пласт"

1 Введение

Закачка смешивающегося углеводородного газа в пласт — один из газовых методов увеличения нефтеотдачи. Обычно процессы 3D фильтрации моделируются в гидродинамических симуляторах (tNavigator, Eclipse, STARS, INTERSECT и прочие). Однако такое полномасштабное моделирование требует большого количества ресурсов как денежных (стоимость лицензии на ПО), так и временных (детальная модель может рассчитываться несколько дней). Как результат, появляется необходимость построить упрощенную модель, которая будет оценивать суммарную добычу в более короткие сроки.

В прошлом семестре мы рассмотрели задачу линейной регрессии с несколькими переменными. Было необходимо построить алгоритм машинного обучения, который будет предсказывать накопленную добычу пласта после 126 месяцев разработки. Были применены следующие алгоритмы: линейная регрессия `LinearRegression`, линейная регрессия с регуляризацией `LassoRegression`, градиентный бустинг на деревьях `XGBoost` и метод опорных векторов `SVR`. В этом семестре мы обучим рекуррентную нейронную сеть (RNN) для предсказания добычи для последовательности месяцев.

2 Постановка задачи и описание датасета

Будем рассматривать такой же пласт с теми же свойствами. Схема пласта представлена на рисунке 1.

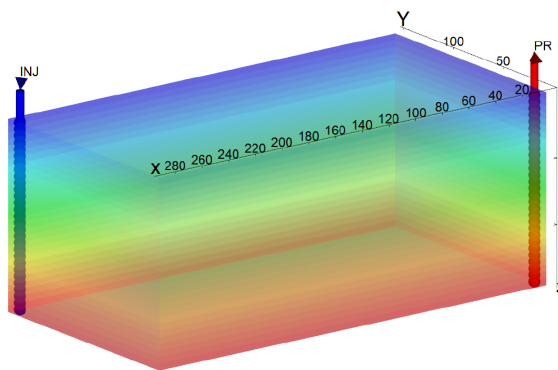


Рис. 1: Схема пласта

2.1 Задача А.

Необходимо построить RNN, которая по данным о дебите за первый год будет предсказывать дебит в каждом месяце второго года. В качестве датасета используем данные о первых 24 месяцах добычи нефти для 200 скважин. На каждом шагу RNN по

данным за месяца с i -го по $(i+12)$ -ый будет предсказывать добычу нефти в $(i+1)$ — $(i+13)$ -ых месяцах. В итоге мы получим предсказание дебита на второй год разработки по реальной истории за первый год. Входные и выходные данные для обучения модели имеют размер $(1650, 12, 1)$: 1650 примеров, данные за 12 месяцев, 1 значение дебита в каждом месяце.

2.2 Задача В.

Задача В принципиально не отличается от **А**, за исключением одного момента: в датасете к данным о дебите мы добавим данные о среднем давлении по пласту и об объеме закачки газа в каждом месяце. Таким образом по дебиту, давлению и закачке за первый год RNN предскажет дебит, давление и закачку за второй год. Размер входных и выходных данных — $(1650, 12, 3)$: 1650 примеров, данные за 12 месяцев, 3 значения (дебит, давление, закачка) в каждом месяце.

2.3 Задача С.

Необходимо обучить RNN, которая будет использовать данные о дебите в каждом месяце за год и мета-данные из прошлого семестра (давление насыщения, начальное давление пласта, забойное давление в нагнетательной скважине, температура пласта и расстояние между скважинами), чтобы предсказать дебит на втором году разработки. Входные и выходные данные этой задачи имеют размер $(1650, 12, 6)$ и $(1650, 12, 1)$ соответственно: 1650 примеров, данные за 12 месяцев, в каждом месяце задаем 6 значений на входе (5 мета-параметров и 1 значение дебита в i -том месяце) и одно на выходе (дебит в $(i+1)$ -ом).

2.4 Задача D.

В этой задаче мы обучим алгоритм XGBoost предсказывать накопленную добычу за второй год разработки. Будем использовать тот же датасет, однако входные данные будут иметь размер $(200, 17)$, а выходные — $(200, 1)$: 200 скважин, для которых на входе задаем 17 значений (5 мета-параметров и последовательность значений дебита за первые 12 месяцев разработки), на выходе получаем одно значение для каждой скважины — суммарную добычу за 13-24 месяца.

После вычисления накопленной добычи за второй год по реальным данным и предсказанным, мы найдем средние абсолютные ошибки (MAE) между этими величинами для задач **А**—**D** и сравним их.

2.5 Задача Е.

В этой задаче мы обучим RNN, которая будет по данным о дебите в предыдущем месяце и мета-данным предсказывать дебит в следующем. В качестве входных и выходных данных возьмем следующие. Входные данные будут иметь размер $(200, 125, 6)$, выходные — размер $(200, 125, 1)$: 200 скважин, для каждой из которых на входе задаем по 6 значений для 125 месяцев (5 мета-параметров и добыча за i -ый месяц), а на выходе 1 значение (добыча за $(i+1)$ -ый месяц). Таким образом RNN будет восстанавливать всю историю разработки за 126 месяцев.

3 Описание подхода

Для решения задач **A**, **B**, **C**, **E** мы применили рекуррентные нейронные сети, поскольку их преимущество в том, что они могут обрабатывать последовательность событий во времени. Мы использовали сети с долгой краткосрочной памятью LSTM, которые улавливают долгосрочные зависимости. **Задача D** - это задача линейной регрессии, для решения которой мы использовали XGBRegressor.

4 Эксперименты и результаты

На рисунке 2 представлены графики реального дебита (синяя кривая) и предсказанного для **задач A** (оранжевая), **B** (зеленая) и **C** (красная) для шести тестовых примеров. Видно, что для разных скважин лучший результат демонстрируют разные алгоритмы. Например, для первой скважины более близкий дебит к реальным данным получается при решении задачи с учетом данных о давлении и закачке (**B**). Однако для пятой скважины этот алгоритм дает худший результат. Чтобы оценить точность всех алгоритмов, мы посчитали MAE между реальным дебитом на втором году разработки и предсказанным. Результаты этих вычислений представлены в таблице 1.

Алгоритм	MAE
Задача A (дебит)	8481.61
Задача B (дебит, давление, закачка)	9480.11
Задача C (дебит, мета-данные)	8821.90
Задача D (XGBoost)	6275.38

Таблица 1: Средняя абсолютная ошибка

Видно, что задача **B** дает результат хуже, чем задача **A**, то есть при добавлении данных о давлении и закачке, точность предсказаний не увеличивается, а наоборот, падает. Это объясняется тем, алгоритму приходится обучаться предсказывать также давление и закачку газа, то есть 3 величины, а не одну, как в задаче **A**. Задача **C** также дает результат хуже, чем **A**. Можем сделать вывод, что добавление мета-данных может сказаться негативно на точности. Самый лучший результат продемонстрировал XGBoost. Это можно объяснить тем, что при предсказании в RNN мы используем предсказанные данные о прошлом месяце, таким образом ошибка копится на каждом месяце второго года, в то время как XGBoost сразу предсказывает накопленную добычу за второй год.

На рисунке 3 представлены результаты решения задачи **E**. Видно, что в большинстве случаев RNN удалось качественно восстановить график дебита. Однако для некоторых примеров (4 и 6) можем увидеть всплеск вначале предсказанного дебита, в то время как в реальной истории его нет. Это объясняется тем, что такой всплеск — это распространенный эффект в первых месяцах разработки, и RNN запомнила эту закономерность и повторила на этих примерах. MAE между реальным дебитом и предсказанным равна 16334.13, что существенно выше, чем в **задачах A—D**, поскольку промежуток, за который мы делаем предсказание, больше.

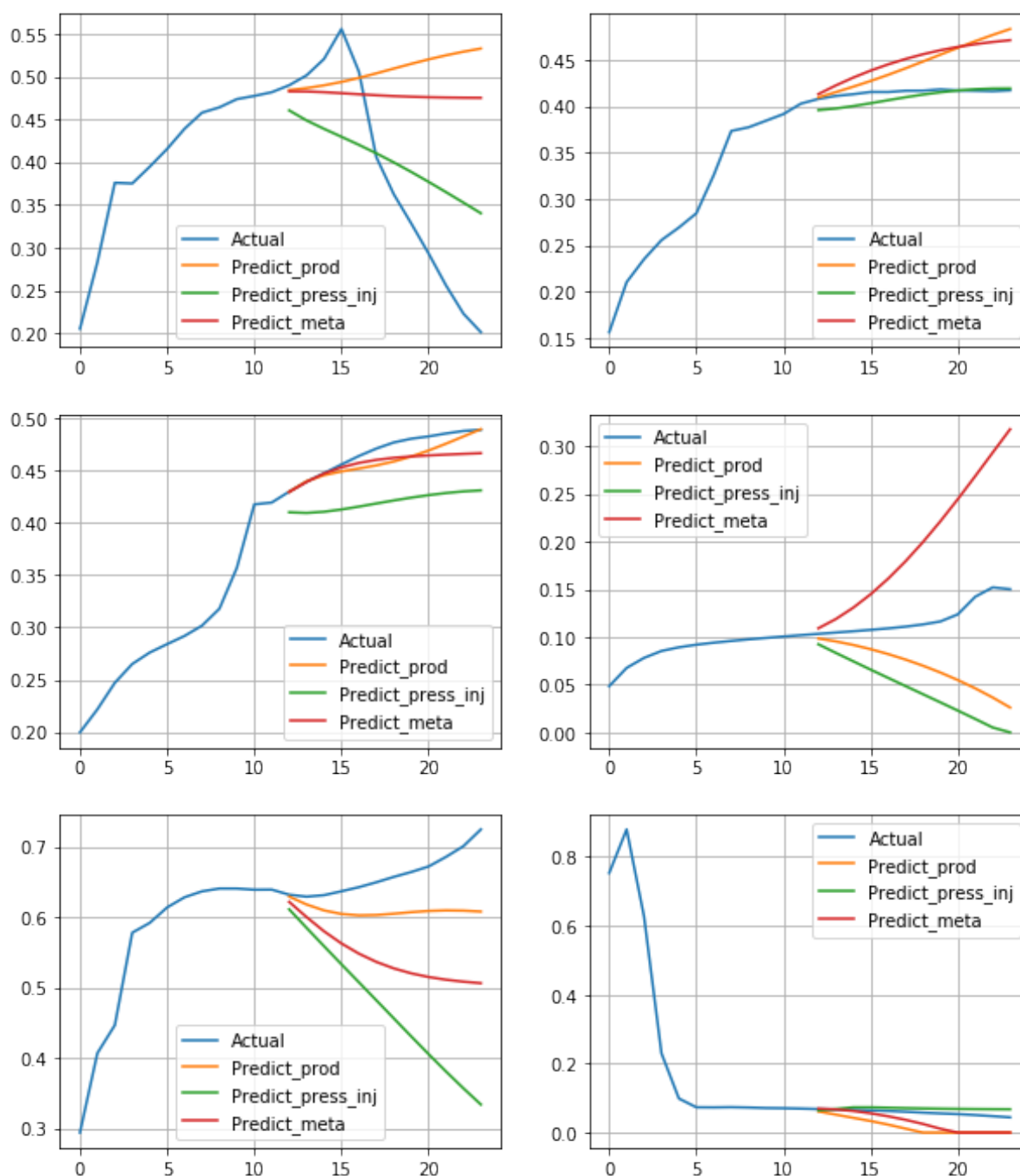


Рис. 2: Результаты решения задач А, В, С.

5 Заключение

Были решены пять разных задач, и использованы два алгоритма: RNN и XGBoost. С предсказанием накопленной добычи нефти за указанный промежуток лучше справился XGBoost, поскольку RNN накапливает ошибку, предсказывая последовательность значений и используя для предсказания уже предсказанный ранее результат. Однако RNN показала хорошие результаты для предсказания последовательности значений добычи нефти. Также RNN продемонстрировала способность запоминать

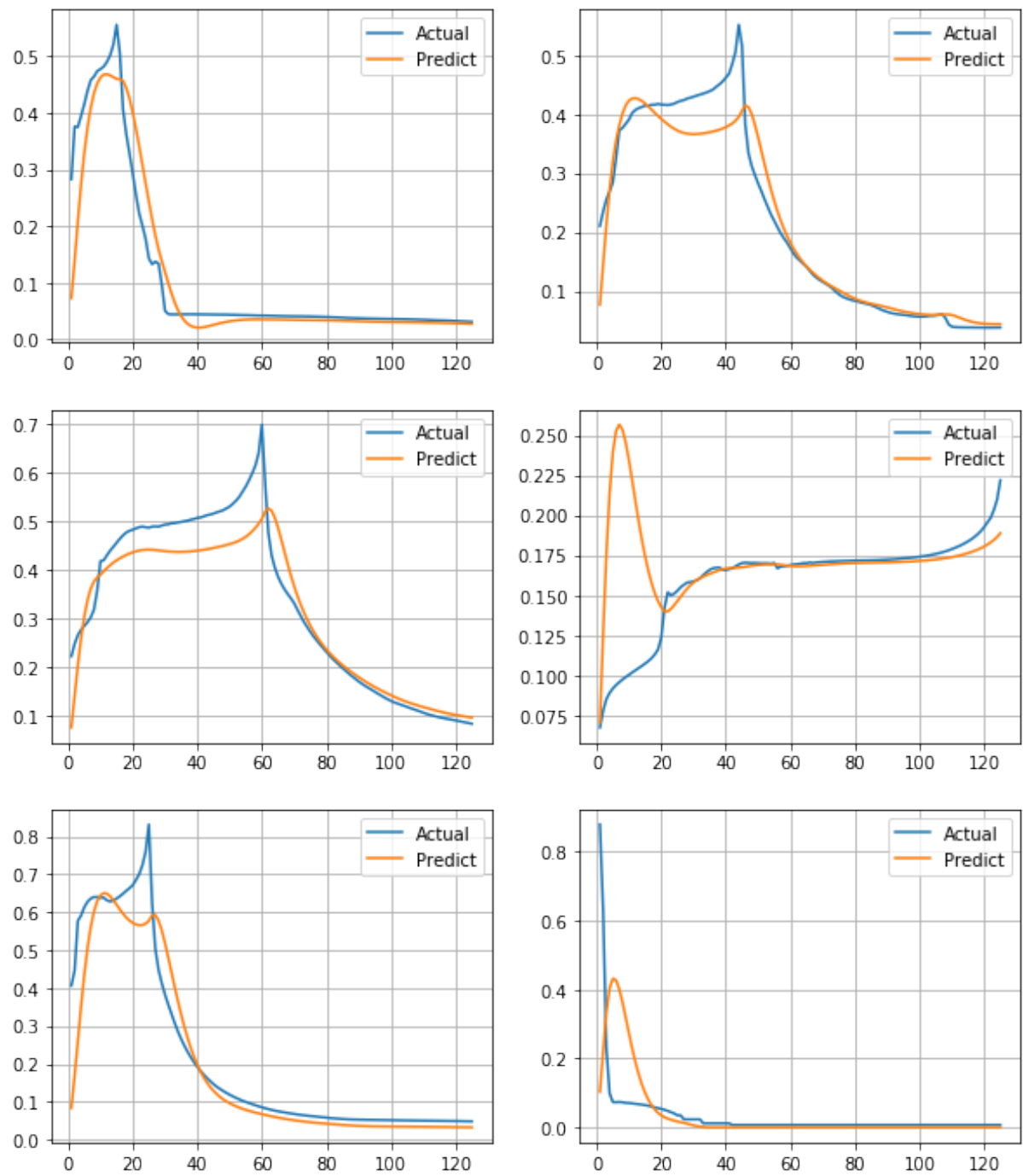


Рис. 3: Результаты решения задач Е.

общие закономерности последовательней, на которых она обучается.