

XV Международная научно-практическая конференция студентов аспирантов и молодых учёных
«Молодёжь и современные информационные технологии»

ОЦЕНКА ИНФОРМАТИВНОСТИ ВЫБОРКИ ДАННЫХ НЕЙТРОН-НЕЙТРОННОГО КАРОТАЖА

В.В. Шестаков, А. А. Аржаник, А. А. Шевченко
Научный руководитель: Д. Ю. Степанов
Томский политехнический университет
arzh_sasha@mail.ru

Введение

Информация о параметрах геологической среды представляет собой ограниченный набор данных, измеряемый преимущественно в стволах скважин. Эти параметры несут важную информацию о местоположении и миграции флюидов, и задача построения их трехмерных моделей на текущий момент является актуальной.

Один из методов решения данной задачи разрабатывается на кафедре программной инженерии Томского политехнического университета и в качестве исходных данных использует атрибуты наземной сейсморазведки и геофизических исследований скважин (ГИС) [1]. Одним из факторов, влияющих на точность работы метода, является информативность исходной выборки данных. Она должна максимально описывать строение исследуемого месторождения и при этом не нести избыточную информацию. Идеальной является ситуация, в которой каждый элемент выборки описывает уникальный участок геологической среды. На реальном месторождении такая ситуация невозможна, из-за неоднородности среды, однако выборку можно приблизить к желаемой путем исключения из нее одинаковых либо очень схожих элементов.

Описание исходных данных

Исходные данные представлены в виде таблицы EXCEL, в которой записаны измерения импульсно нейтрон-нейтронного каротажа (ИННК) шести скважин. Данные ИННК привязаны к относительной временной шкале. Нулевая отметка шкалы в рассматриваемом случае является границей залегания Баженовской свиты - нефтегазоносного объекта Томской области.

ИННК основан на облучении земных пород вдоль скважины импульсами быстрых нейтронов высокой энергии и последующим контролем спада их плотности из-за эффектов рассеивания. Диаграммы ИННК позволяют определять местоположение нефтяных коллекторов, отличать водяные коллектора от нефтяных [2].

Постановка задачи

Имеется выборка кривых ИННК, измеренных в стволах пяти скважин методами ГИС. Данная выборка использовалась при расчете модели некоторого месторождения. Также имеется кривая ИННК шестой скважины, добавление которой в исходную выборку приведет к необходимости пересчета модели. Требуется ответить на вопрос: увеличится ли информативность исходной

выборки при добавлении в неё данных шестой скважины?

Метод решения задачи

Для определения наличия или отсутствия связи между данными двух скважин могут быть использованы различные меры сходства, такие как Евклидово расстояние и коэффициент корреляции Пирсона. Проанализируем преимущества и недостатки обеих рассматриваемых мер [3].

Коэффициент корреляции Пирсона характеризует существование линейной зависимости между двумя величинами.

Один из главных недостатков коэффициента корреляции как меры сходства - он чувствителен к форме за счет снижения чувствительности к величине различий между переменными. Переменные, у которых одновременно велики абсолютные значения и стандартные отклонения, могут подавить влияние переменных с меньшими абсолютными размерами и стандартными отклонениями [3].

Для наглядности недостатков рассмотрим пример на рисунке 1:

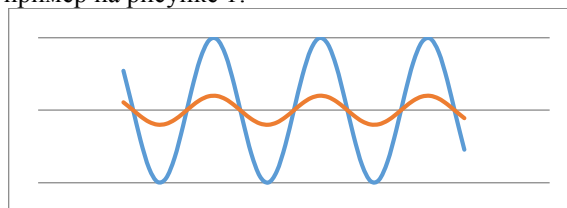


Рис. 1. Синусоидальные кривые с одинаковым периодом и различными амплитудами

На нем изображены две синусоидальные кривые, которые, визуально, существенно отличаются друг от друга. Однако коэффициент корреляции равен единице и говорит о том, что между ними существует однозначная линейная связь. Из примера очевидно, что делать выводы о схожести кривых, основываясь только на значении коэффициента корреляции, нельзя. Расстояние Евклида (17,56) в данном случае позволяет сделать более правильные выводы.

Евклидово расстояние – расстояние между двумя точками евклидова пространства. Основной недостаток Евклидова расстояния - оно чувствительно к величине различий между переменными и мало восприимчиво к форме [3]. Для наглядности рассмотрим два графика (рис. 2, рис. 3):

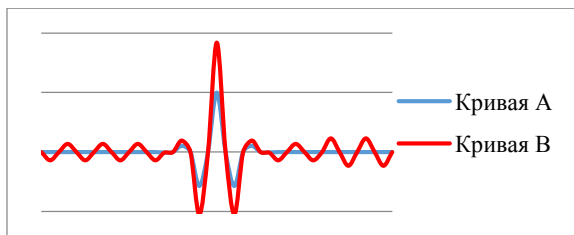


Рис. 2. Кривые А и В

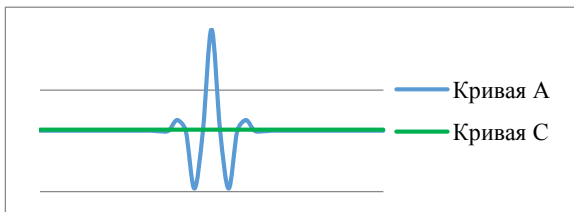


Рис. 3. Кривые А и С

На рисунках 2 и 3 изображены графики кривых: С - прямая линия, А и В - импульсы Пузырева ($S(t) = e^{-\beta^2 t^2} \cos(2\pi f_0 t + \varphi)$). Сравнивая по метрике Евклида кривые А и В (расстояние 1,26) и кривые А и С (расстояние 1,29), можно прийти к выводу, что кривая А одинаково похожа на С и на В, что неверно. В данном случае критерий Пирсона несет больше информации о несхожести кривых.

Из-за перечисленных недостатков данные метрики нельзя использовать для решения данной задачи по отдельности. Таким образом, выводы о схожести кривых будем принимать на основании одновременно обеих метрик.

Ход работы

Для определения информативности скважины №6 она попарно сравнивалась с имеющимися пятью. Оценка схожести данных осуществлялась с помощью скользящего временного окна, перемещающегося в пределах всего временного интервала. Выбор размера окна в 20 мс (период колебаний сейсмической волны) обусловлен тем, что в алгоритме построения модели геологической среды ковариационные свойства данных скважин аналогично рассчитываются в скользящем окне.

На первом этапе были измерены коэффициенты корреляции. В качестве выборки X выбирались по 20 последовательных значений параметра скважины №6, в качестве выборки Y – соответствующие значения из скважин №1-5. На втором этапе были измерены расстояния Евклида между точками, координаты которых – 20 последовательных значений из скважин.

Так как параметр ИННК мало информативен начиная со второго знака после запятой, для евклидова расстояния принимаем порог равным 2 ($0,1 * \text{количество отсчетов}$), но для надежности расчетов уменьшаем до 1. Принято считать, что коэффициент корреляции 0,7 и более говорит о высокой степени зависимости кривых, зададим порог чуть меньше этого значения, то есть 0,6.

Определим точки, в которых одновременно выполняются оба условия, то есть коэффициент корреляции больше 0,6, евклидово расстояние

меньше 1. Построим график зависимости от времени t (мс), на котором функция принятия решения будет принимать значение 1, если выполнены оба условия, и 0 в других случаях (рис. 4).

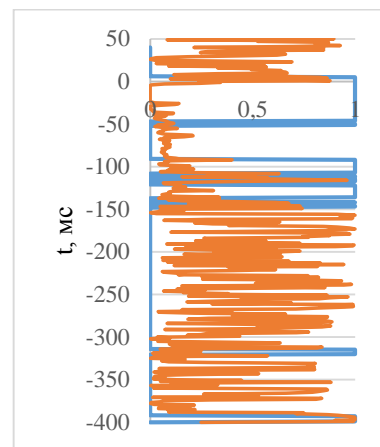


Рис. 4. Функция принятия решения и кривая ИННК 6-й скважины

На графике выделяются четыре зоны с единицами: [-389; 383], [-310; -305], [-137; -82], [-41; 15] – в этих зонах шестая скважина похожа на одну из пяти предыдущих. Данные скважин во всех прочих зонах - различны, следовательно, их нужно добавить в выборку.

Выводы

Для выполнения поставленной задачи было решено использовать меры схожести: коэффициент корреляции и евклидово расстояние.

Так как обе имеют недостатки (коэффициент корреляции Пирсона сравнивает только форму, а евклидово расстояние – суммарное значение амплитуд), было решено учитывать обе метрики.

В работе были рассчитаны коэффициент корреляции и евклидово расстояние на примере данных 6-ти реальных скважин. Было выявлено, что информативность исходной выборки увеличится при добавлении в неё данных скважины №6 за исключением 4-х определенных зон.

Список использованных источников

- 1) Шестаков, В.В. Построение трехмерных моделей параметров геологических сред / В.В. Шестаков, Д.Ю. Степанов, Г.А. Сысолятина. – Томск, 2016. – 5 с.
- 2) Косков, В. Н. Геофизические исследования скважин и интерпретация данных ГИС / В. Н. Косков, Б. В. Косков. – Пермь.: Изд-во ПГТУ, 2007. – 317 с.
- 3) Мьюллер, У. Р. Клетка и др.; Под ред. И. С. Енюкова. Факторный, дискриминантный и кластерный анализ: Пер. с англ. / Дж.-О. Ким, Ч. У. — М.: Финансы и статистика, 1989. — 215 с.