

ВЛАДИМИРОВ Д. Н., КОЖЕВНИКОВА П. В.
КЛАСТЕРНЫЙ АНАЛИЗ ПРИ ОЦЕНКЕ ДОСТОВЕРНОСТИ ПРОГНОЗА
ФИЗИКО-ГЕОЛОГИЧЕСКИХ ПАРАМЕТРОВ

УДК 001.891.57:53, ГРНТИ 28.17.23

Кластерный анализ при оценке
достоверности прогноза физико-
геологических параметров

Cluster analysis in assessing the
reliability of forecasting physico-
geological parameters

Д. Н. Владимиров,
П. В. Кожевникова

D. N. Vladimirov,
P. V. Kozhevnikova

Ухтинский государственный
технический университет, г. Ухта

Ukhta State Technical University,
Ukhta

В статье описано исследование, в рамках которого рассмотрено применение кластерного анализа данных в качестве одного из потенциальных инструментов, необходимых для определения областей различной плотности данных при оценке достоверности подсчёта запасов углеводородов.

The article describes a study that examines the application of cluster analysis of data as one of the potential tools necessary for determining areas of different data densities when assessing the reliability of hydrocarbon reserve calculations.

Ключевые слова: запасы углеводородов, петрофизическая модель, оценка достоверности прогноза, плотность данных, кластеризация данных

Keywords: hydrocarbon reserves, petrophysical model, forecast reliability assessment, data density, data clustering

Введение

В процессе анализа отношений между физико-геологическими параметрами петрофизических моделей одной из актуальных проблем является выявление неоднородности плотности этих параметров на различных участках зависимостей, что в значительной степени влияет на достоверность результатов подсчета запасов углеводородов [1].

С целью исследования данной проблемы была выполнена научно-исследовательская работа, задачей которой является кластерный анализ выборки данных. В рамках реализации поставленной задачи разработано программное средство для кластеризации данных различными методами с пользовательским интерфейсом с использованием фреймворка AvaloniaUI + C#.

В качестве исследуемой выборки данных взяты отношения таких физико-геологических параметров как:

- пористость по керну (получены на основе анализа кернового материала);
- пористость по ГИС (получены на основе интерпретации результатов геофизических исследований скважин);
- остаточная водонасыщенность.

Кластерный анализ данных – это один из ключевых методов машинного обучения, который может использоваться в условиях почти полного отсутствия информации о законах распределения данных [2] и позволяет автоматически группировать данные, принадлежащие к различным категориям или классам. Этот подход позволяет найти скрытые структуры, закономерности и паттерны в данных, что в контексте дифференцирования по достоверности физико-геологических зависимостей поможет выделить области высокой плотности обеспеченности данными и определить области низкой плотности. Что в совокупности позволяет учитывать различный уровень достоверности данных на разных участках зависимости и расширяет возможности прогнозирования.

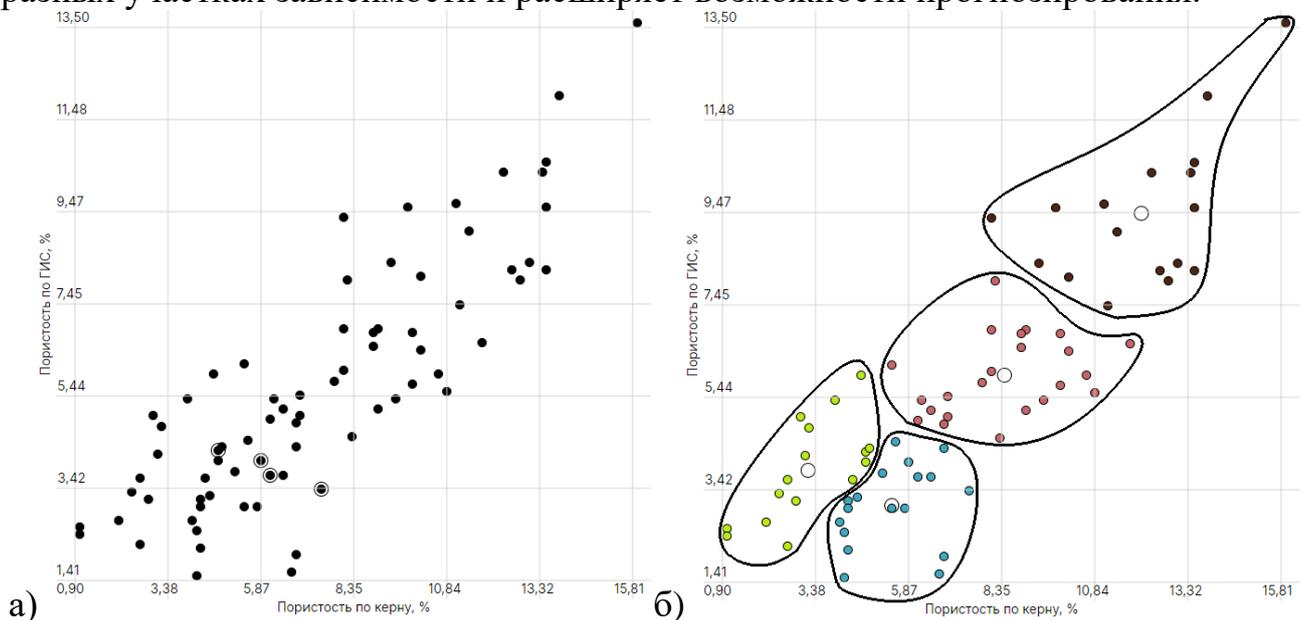


Рисунок 1. Эксперимент кластеризации
а) начальные условия, б) результат кластеризации

В рамках настоящего исследования использовался метод кластеризации – алгоритм *k*-средних. Он весьма распространён как нечеткий метод кластеризации количественных данных [2]. Одним из недостатков данного алгоритма является, то что необходимо вручную заранее указать число кластеров. Вместе с этим данный параметр играет решающую роль в процессе кластеризации данных. Однако в нашем случае такое ограничение не является критическим, так как в ходе экспериментов будет подбираться необходимое количество кластеров на основании визуальной оценки каждой зависимости для конкретной петрофизической модели.

Ещё одним важным параметром кластерного анализа является выбор начальных центроидов, которые целесообразно выбирать случайным образом. Однако, в результате ряда экспериментов выявлено, что не любое расположение центроидов приводит к предсказуемому распределению кластеров,

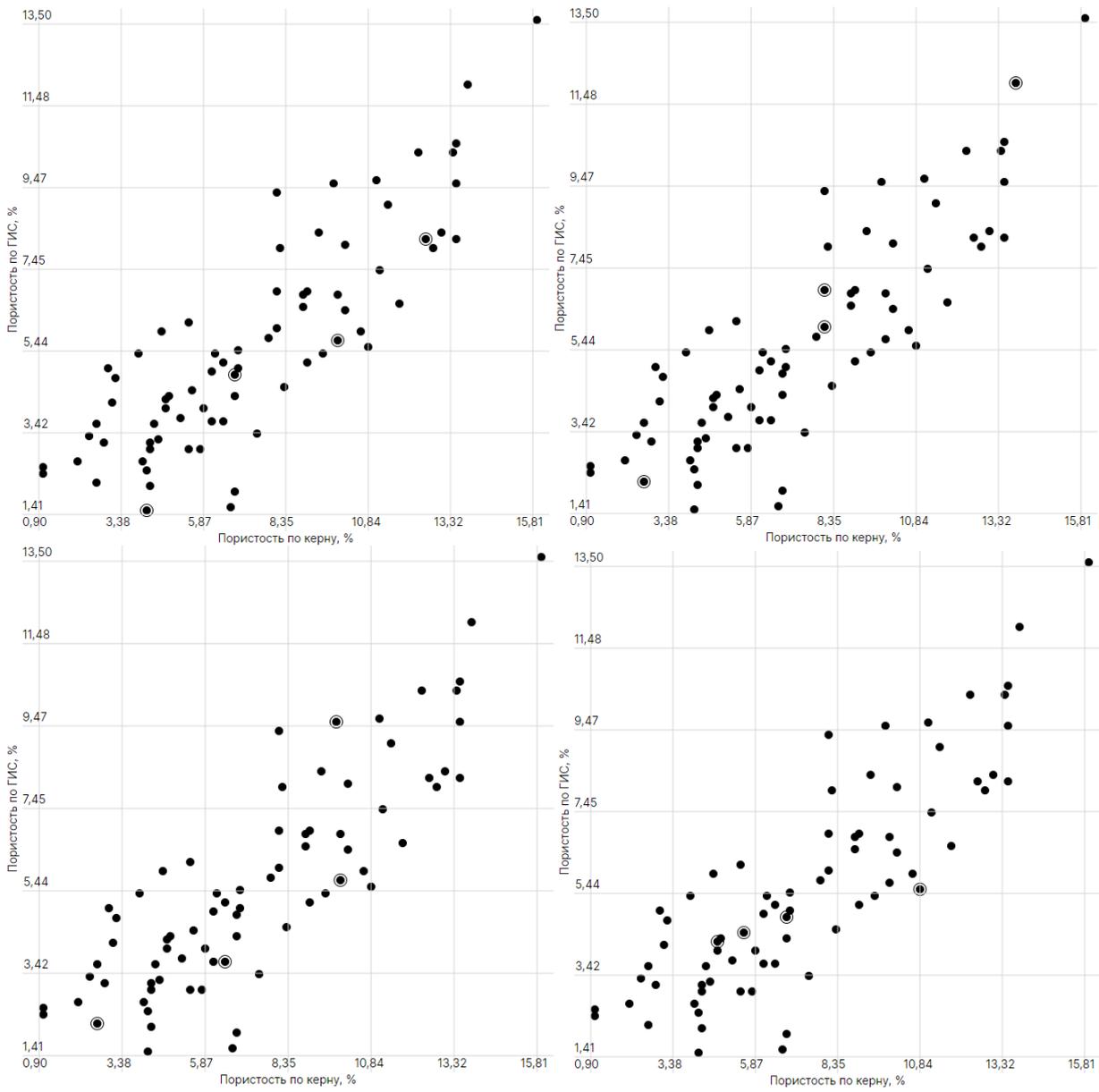


Рисунок 2. Начальные условия ряда экспериментов

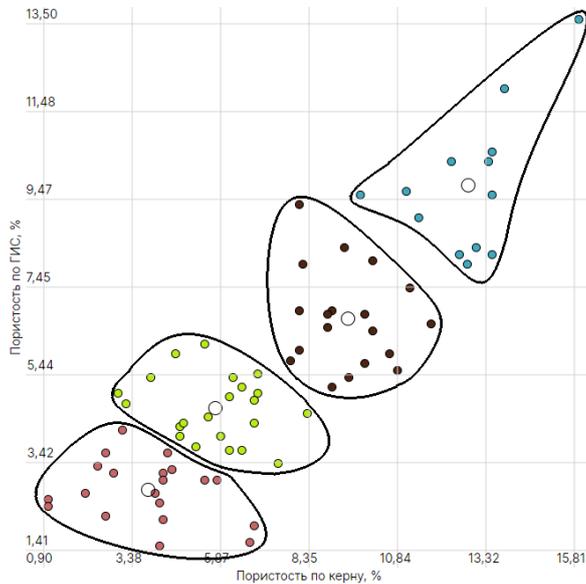


Рисунок 3. Результат кластеризации

удовлетворяющему целям и задачам исследования. Так на Рисунке 1 изображен эксперимент, при котором кластеры распределились не вдоль оси линейной зависимости исследуемых параметров. В свою очередь на Рисунке 2 изображены начальные условия ряда экспериментов, которые дали один и тот же предсказуемый результат кластеризации, изображённый на Рисунке 3.

Заключение

Исходя из результатов экспериментов можно сделать вывод, что наряду с выбором числа кластеров, также решающую роль в процессе кластеризации играет выбор начальных центров кластеров. Учитывая, что исследуемые наборы данных являются линейными зависимостями различных физико-геологических параметров, следует иметь ввиду, что все кластеры конкретной выборки должны лежать приблизительно на одной прямой параллельной оси линейной зависимости. В связи с чем, в рамках текущего исследования наиболее целесообразным распределением начальных центроидов является их расположение вдоль прямой, параллельной оси линейной зависимости. Такое расположение приводит к предсказуемому и ожидаемому распределению кластеров.

Кластерный анализ может оказаться мощным инструментом при оценке достоверности прогнозов физико-геологических параметров. Он позволит выделить области с разной плотностью данных, поможет выявить закономерности и может улучшить точность прогнозов. При этом необходимо учитывать особенности кластеризации, такие как выбор метрики расстояния, определение числа кластеров и интерпретация результатов. Важно продолжить исследования в данной области и рассмотреть другие методы кластеризации, которые могут оказаться более эффективными и подходящими в данной предметной области.

Список использованных источников и литературы

1. Кожевникова П. В. Разработка математических моделей информационно-экспертной оценки достоверности промыслово-геологических моделей на основе нечеткого моделирования. – Петрозаводск.: 2018
2. Чубукова И. А. Data Mining. – М.: Интернет-Университет информационных технологий; БИНОМ . Лаборатория знаний, 2006
3. Кластеризация: алгоритмы k-means и c-means [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/articles/67078/> , (дата обращения: 13.02.2024).
4. Обзор алгоритмов кластеризации данных [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/articles/101338/>(дата обращения: 13.02.2024).
5. Обзор методов кластеризации [Электронный ресурс]. – Режим доступа: <https://scikit-learn.ru/clustering/> (дата обращения: 15.02.2024).

List of references

1. Kozhevnikova P. V. Development of mathematical models for information and expert assessment of the reliability of field geological models based on fuzzy modeling. – Petrozavodsk: 2018
2. Chubukova I. A. Data Mining. – М.: Internet University of Information Technologies; BINOM. Knowledge Laboratory, 2006
3. Clustering: k-means and c-means algorithms, <https://habr.com/ru/articles/67078/>, access date: 02/13/2024.
4. Review of data clustering algorithms, <https://habr.com/ru/articles/101338/> , access date: 02/13/2024.
5. Review of clustering methods, <https://scikit-learn.ru/clustering/>, access date: 02/15/2024.