

ПРИМЕРЫ ИНТЕГРАЦИИ СРЕДСТВ ИНТЕЛЛЕКТУАЛЬНЫХ ВЫЧИСЛЕНИЙ WOLFRAM MATHEMATICA И СИСТЕМЫ ГЕОБАЗАДАННЫХ

В. Б. Таранчук

*Белорусский государственный университет,
пр. Независимости, 4, 220030, г. Минск, Беларусь, taranchuk@bsu.by*

Обсуждаются методические и технические решения интеграции модулей интеллектуальных вычислений системы Wolfram Mathematica и инструментов программного комплекса ГеоБазаДанных в задачах формирования, интерпретации, обработки, визуализации цифровых полей при компьютерном моделировании объектов геологии, геоэкологии. Предложена модификация типового способа кластеризации, расчетами на представительных данных подтверждены преимущества.

Ключевые слова: система ГеоБазаДанных; компьютерная модель; цифровые поля; интерактивная визуализация; интеллектуальная адаптация; система Wolfram Mathematica; кластеризация.

Введение

Компьютерное моделирование включает развитие математических методов и алгоритмов; разработку программных средств, проведение вычислительных экспериментов. Одной из важнейших составляющих при разработке и внедрении, в частности, постоянно действующих компьютерных геологических, геоэкологических моделей является задача оценки адекватности и точности предлагаемых цифровых описаний. Ключевыми являются вопросы автоматизации формирования, адаптации моделей с учетом постоянно поступающих дополнительных данных, а также ревизии результатов обработки исходной информации на основе новых интерпретаций. Актуально повышать качество предобработки, интерпретации исходных данных и получаемых разными способами результатов, причем, из-за их все возрастающего объема принципиально важно автоматизировать все этапы анализа, обработки, визуализации информации, внедрять компьютерные интеллектуальные инструменты. При этом интеллектуальный анализ данных не исключает человеческое участие в обработке и анализе. Соответственно, программное обеспечение должно быть интерактивным, интерфейсы программ дружелюбными и предсказательными – тогда реально значительно ускорить процессы обоснования экспертных заключений, делая их содержание и рекомендации доступными для широкого круга аналитиков, не являющихся специалистами в статистике, математике или программировании.

Ряд вопросов по анализу, оценкам качества пространственных данных можно эффективно решать с использованием компьютерного комплекса ГеоБазаДанных [1]. Ниже упомянуты возможные варианты, методические решения и программные инструменты, которые позволяют с использованием средств искусственного интеллекта подтвердить обоснованность интерпретаций, а также визуализировать и получить числовые значения погрешностей, рассчитываемых разными методами результатов обработки данных, включаемых и используемых в компьютерных геологических, геоэкологических моделях. Для иллюстраций выбрана ключевая задача подготовки цифровых полей, используемых в компьютерных моделях. В частности, обсуждаются предложенные, апробированные при решении разных прикладных задач методы, реализованные в интерактивном компьютерном комплексе ГеоБазаДанных авторские алгоритмы расчета аппроксимирующих цифровых полей.

Общие сведения о комплексе ГеоБазаДанных

ГеоБазаДанных – комплекс интеллектуальных компьютерных подсистем, математического, алгоритмического и программного обеспечения наполнения, сопровождения и визуализации

зации данных для имитационных и математических моделей, средств обеспечения вычислительных экспериментов, алгоритмических и программных средств создания постоянно действующих компьютерных моделей. В ГеоБазаДанных можно формировать и визуализировать цифровые описания пространственных распределений; графически иллюстрировать решения задач, описывающих динамические процессы многофазной фильтрации, миграции флюидов, переноса тепла, влаги, минеральных водорастворимых соединений в толщах пород; конструировать и реализовывать интерактивные сценарии визуализации и обработки результатов вычислительных экспериментов. Подсистемы ГеоБазаДанных позволяют рассчитывать и выполнять в разных приближениях экспертные оценки локальных и интегральных характеристик экосистем; создавать постоянно действующие модели объектов нефтедобычи; формировать и выводить на твердые копии тематические карты. Основные компоненты, примеры применения ГеоБазаДанных описаны в [1].

Последние годы активность использования инструментов искусственного интеллекта при решении задач геологии, геоэкологии стремительно повышается. В частности, по рассматриваемым в настоящей работе алгоритмам и методам кластерного анализа каждый год публикуются десятки статей. В некоторых приводятся решения практических задач, изложены методы предобработки и интерпретации геофизических данных, анализа результатов, экспертные заключения (например, [2, 3].)

Ниже отмечены основные дополнения комплекса ГеоБазаДанных, включающие методические аспекты и программные компоненты, реализованные с использованием инструментов искусственного интеллекта. При этом, чтобы было понимание этапов разработки и взаимосвязи составных частей, приведены ссылки основных публикаций с примерами использования обновленных компонент комплекса.

Так, в [4, 5] опубликованы методические и технические решения, описаны программные инструменты, результаты и примеры обработки данных типичных для геофизических методов изучения геологических объектов на профилях наблюдений. Показаны, подтверждены примерами эффективность применения искусственных нейронных сетей для устранения в результатах замеров шумов и искажений, выполнения необходимой для математических моделей предобработки данных сглаживанием с целью подготовки регулярных цифровых распределений.

В [6] рассмотрены и приведены примеры интерактивного формирования в вычислительных экспериментах цифровых моделей геологических, геоэкологических объектов, отвечающих интуитивным требованиям эксперта. Отмечены методические и алгоритмические решения эффективные при обработке данных дистанционного экологического мониторинга, специальные инструменты системы ГеоБазаДанных, приведены результаты интерактивной адаптации, сопоставления с типовыми эталонными решениями в подсистеме «Генератор геологической модели залежи». Рассматриваются и поясняются результаты разных способов аппроксимации и воспроизведения цифрового поля, его интерактивной адаптации. В частности, отмечено, что эффективным является метод восстановления цифровых полей на регулярную сетку по данным замеров из нерегулярно расположенных пунктов наблюдений (рассеянного множества точек), если применяется методика ГеоБазаДанных “выделено”. В примерах показано, что так можно значительно повысить качество, адекватность цифрового описания. Но нужно понимать, что на том этапе состояния алгоритмического и программного обеспечения комплекса ГеоБазаДанных само выделение реализуется по интуитивным предложениям эксперта.

Вопросы автоматического определения участков типа “выделено” с использованием интегрированных в систему ГеоБазаДанных инструментов кластерного анализа обсуждались на приведенных конференциях, в научных трудах [7, 8] опубликованы некоторые положительные результаты. В частности, представлены иллюстрирующие эффекты выбора и подтверждения лучших алгоритмов кластеризации результаты, сопоставлены варианты использования разных методов кластеризации, причем, для разных способов задания метрического расстояния.

В упомянутых опубликованных материалах, наряду с положительными, отмечены недостатки описанных программных реализаций. Новые дополнения в комплекс ГеоБазаДанных, уточнения настроек инструментов системы *Wolfram Mathematica* для кластеризации данных апробированы в расчетах, описаны и обсуждаются.

Исходные данные для вычислительных экспериментов

Результаты представляемого исследования являются показательными и адекватными благодаря правилам их подготовки. С одной стороны, исходные данные формируются генераторами случайных чисел и уточняются экспертом с целью придать им характер подобных данным натурных наблюдений в практике геофизических методов изучения геологических, геоэкологических объектов. С другой стороны, для значений замеров в наблюдениях используются принятые математические выражения. По факту, для каждого конкретного алгоритма обработки известно, можно рассчитать с нужной точностью базовое (эталонное) цифровое распределение – сопоставляя расчеты с эталоном можно судить о достоинствах и недостатках используемого метода. Ниже эталонное распределение (в принятой ГеоБазаДанных терминологии – поверхность) несколько отличается от рассмотренного в [7], изменены местоположение фрагментов возмущений и у некоторых их форма и размеры. Вообще говоря, необходимость изменений отсутствовала, т.к. комплект фрагментов типовых элементов рельефа в [7] был достаточно представительным, но специально расчеты выполнялись для эталонной поверхности другой формы и масштаба, чтобы убедиться в устойчивом воспроизведении получаемых результатов. Все описанные ниже варианты расчетов были выполнены для всех эталонных распределений упомянутых публикаций, они подтвердили, что параметры точности одинаковы, не меняются при вариациях размеров, положения, ориентации возмущений. Приведенные в этой работе результаты получены с использованием рассчитываемых по формуле (1) числовых значений отметок уровней (эталонной) поверхности (возмущения $fHillN$ всегда типовые):

$$zSurfH[x_, y_] = fOriginF[x, y] + 400 * fHill6[0.005 * (x - 250), 0.007 * (y - 400) + 600 * fHill3[0.01 * (x - 150), 0.01 * (y - 150) - 200 * fHill[0.01 * (x - 880), 0.015 * (y - 500) - 150 * fHill[0.02 * (x - 920), 0.004 * (y - 100) + 200 * fHill5[0.006 * (x - 450), 0.01 * (y - 150); \quad (1)$$

Визуализация эталонной поверхности $zSurfH$ приведена на рис. 1 и 2. Показаны виды 3D в вариантах поверхность и объем; на рис. 3 приведена карта изолиний, на ней дополнительно в изображение добавлены (в скобках синим цветом) номера фрагментов возмущений.

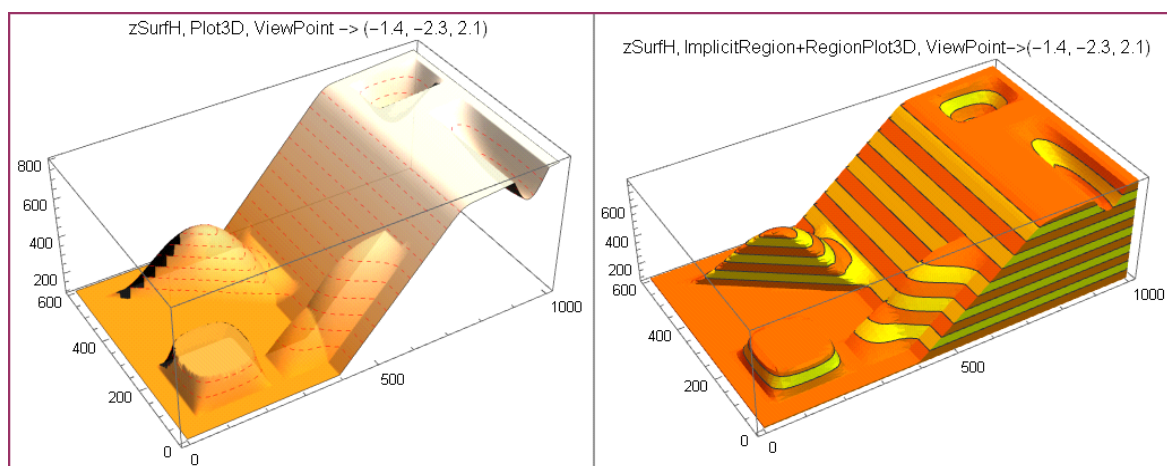


Рис. 1. 3D виды эталонной поверхности $zSurfH$

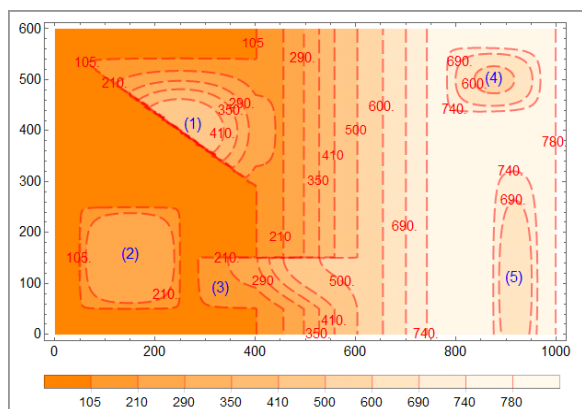


Рис. 2. Карта изолиний
эталонной поверхности zSurfH

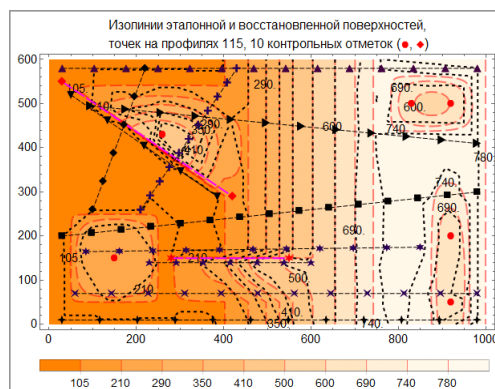


Рис. 3. Схема точек замеров уровней, карта
изолиний эталонной и восстановленной по-
верхностей

Исходные данные для оценок в этом исследовании методов и алгоритмов кластеризации получены с использованием генератора случайных чисел. Задавались: число профилей наблюдений, точек на каждом профиле, и в указанных диапазонах значений генерировались координаты начала и конца профиля. Значения в точках на профилях рассчитывались по формуле (1) – имитация замеров уровня восстанавливаемой поверхности. Заметим, что по факту имеем рассеянное множество точек. Итоговая, принятая к обработке схема размещения точек с замерами уровня показана на рис. 3, где также даны изолинии эталонной поверхности и восстановленной в *Wolfram Mathematica* (метод *Interpolation*, *InterpolationOrder* = 1). Заметим, что на карте изолиний рис. 3 четко видны участки, где нет воспроизведения эталона, что объясняется отсутствием замеров. Понятно, что на частях площади, где восстановленное цифровое поле отлично от эталона классификация / привязка к фрагменту-возмущению маловероятна, и не следует ожидать пользы от кластеризации.

Иллюстрации результатов уточняемых методов кластеризации

Кластерный анализ допускает много различных типов методов/алгоритмов кластеризации. Отметим развитие, новый способ обработки, представления. Не вдаваясь в детали алгоритмической, программной реализации, напомним, что в [7, 8] кластеризация выполнялась по двум параметрам, группировка осуществлялась по критерию близости точек с замерами. В представленных уточненных результатах группировка выполняется по совокупности трех значений – для каждой точки учитываются их координаты и значение уровня поверхности.

Приведем несколько интерпретаций представленных на рис. 4 результатов расчетов с использованием функции *Mathematica* *FindClusters*, поясним иллюстрации.

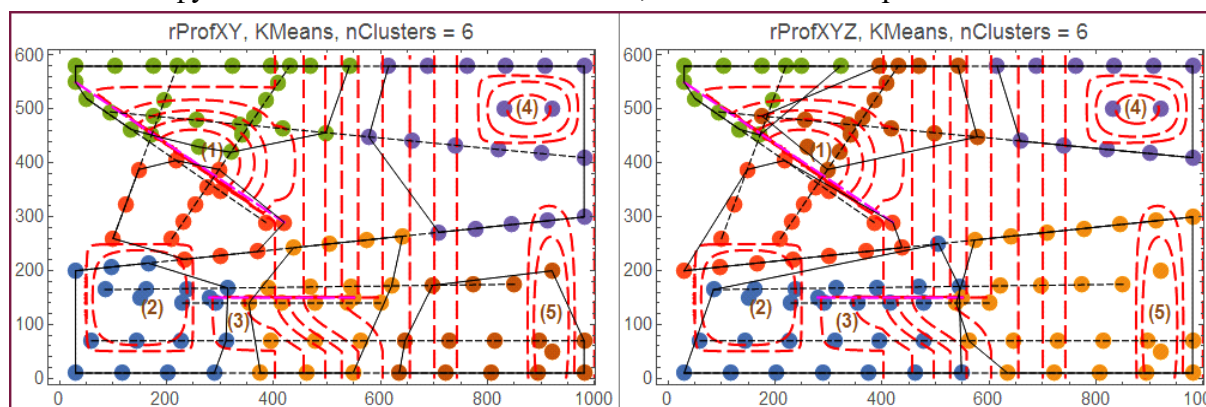


Рис. 4. Результаты кластеризации по двум (rProfXY)
и трем (rProfXYZ) параметрам

На рис. 4 показаны результаты кластеризации методом KMeans, слева – по двум параметрам (rProfXY), справа – по трем (rProfXYZ). Левая иллюстрация – результаты расчетов с настройками, как в работе [8]. Предлагаемый метод кластеризации по трем параметрам явно предпочтительней метода по двум, очевидно в части локализации фрагментов 1, 4 и 5. Результаты вблизи фрагментов 2 и 3 не являются показательными из-за разрывного распределения с разных сторон от горизонтальной пунктирной пурпурной линии (в геологии – «разлом»). Подобные разрывы вовсе не идентифицируются в обычных системах восстановления цифровых полей без специальных априорных дополнительных условий. В ГеоБазаДанных такие условия воспроизводятся, корректируя интерактивно исходную информацию на карте ([1]).

Отметим, что приведены результаты варианта, когда задано число кластеров равное 6. Почему. Определение количества кластеров – это одна из важнейших проблем сегментации. В более широком смысле – это проблема инициализации алгоритма. Рассчитывались и сопоставлялись результаты для числа кластеров от 4 до 9. Предпочтительным по итогам сравнения представляется вариант 6 кластеров, и объяснением может быть то, что воспроизводится распределение, когда есть базовая непрерывная поверхность-лента и 5 фрагментов-возмущений, т.е. 6 разных форм.

При обсуждении результатов в докладе предполагается демонстрация коллекции полученных результатов (28 карт) методами: Automatic, Agglomerate, DBSCAN, GaussianMixture, JarvisPatrick, KMeans, KMedoids, MeanShift, NeighborhoodContraction, Optimize, SpanningTree, Spectral с интерпретациями и сопоставлением кластеров для предпочтительных из них, когда в методе сопоставляются и оцениваются эффекты метрик – апробированы: EuclideanDistance, NormalizedSquaredEuclideanDistance, SquaredEuclideanDistance, BinaryDistance, BrayCurtisDistance, CanberraDistance, CanonicalWarpingDistance, ChessboardDistance, CorrelationDistance, CosineDistance, ManhattanDistance, WarpingDistance.

Библиографические ссылки

1. Таранчук В.Б. Компьютерные модели подземной гидродинамики. Минск : БГУ, 2020. 235 с.
2. Abdideh M., Ameri A. Cluster Analysis of Petrophysical and Geological Parameters for Separating the Electrofacies of a Gas Carbonate Reservoir Sequence // Nat Resour Res 2020. N 29, P. 1843-1856. URL: <https://doi.org/10.1007/s11053-019-09533-1>
3. Ткаченко М. А., Карелина Е. В. Кластерный анализ как метод выявления потенциального хромитового оруденения в пределах Войкаро-Сыньинского массива // Международный научно-исследовательский журнал. 2023. №10 (136). С. 1-7.
4. Таранчук В. Б. Методы и примеры интеллектуальной обработки данных для геологических моделей // Научные ведомости Белгородского государственного университета. Серия: Экономика. Информатика. 2019. № 3 том 46. С. 511-522.
5. Taranchuk V. Tools and examples of intelligent processing, visualization and interpretation of GEODATA // Modelling and Methods of Structural Analysis. IOP Publishing. IOP Conf. Series: Journal of Physics: Conf. Series Vol. 1425 (2020) 012160. P. 9. doi:10.1088/1742-6596/1425/1/012160
6. Taranchuk V. B. Interactive Adaptation of Digital Fields in the System GeoBazaDannych // Communications in Computer and Information Science. Book series Springer (CCIS, volume 1282): 2020. P. 222-233. https://doi.org/10.1007/978-3-030-60447-9_14
7. Taranchuk V. Methodological and Technical Solutions for the Implementation of Clustering Algorithms in the GeoBazaDannych System // Communications in Computer and Information Science // 2022, vol 1625. Springer, P. 1-13
8. Taranchuk V. B. Features and examples of clustering in the system GeoBazaDannych // AIP Conference Proceedings 2497, 030025 (2023): Volume 2497, Issue 1, 2023, P. 1-9. doi: <https://doi.org/10.1063/5.0103603>.