

МЕТОДЫ И ИНСТРУМЕНТЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА И ИНТЕРПРЕТАЦИИ ЦИФРОВЫХ ПОЛЕЙ

Таранчук В. Б., Ходос Ю. О.

Кафедра компьютерных технологий и систем, Факультет прикладной математики и информатики,
Белорусский государственный университет
Минск, Республика Беларусь
E-mail: taranchuk@bsu.by, fpm.hodos@bsu.by

Обсуждаются методические и алгоритмические вопросы разработки специализированных программных модулей аппроксимации системы ГеоБазаДанных для адекватного формирования разработчиками и экспертами цифровых полей параметров компьютерных моделей. Отмечены новые функциональные возможности, обеспеченные включением в ГБД реализующих методы кластерного анализа данных исполняемых модулей системы компьютерной алгебры Wolfram Mathematica.

ВВЕДЕНИЕ

При компьютерном моделировании программный инструментарий для цифрового описания распределений изучаемых параметров должен обеспечивать сбор и обработку данных наблюдений, оценок и прогнозов изменений состояния объекта и среды окружения, ответных реакций на воздействия, поступление в среду веществ, энергий. Многие исследователи подчеркивают, что целью разработки компьютерных экспертных систем должны быть не только программы расчетов для простой констатации фактов. Соответствующие системы (комплексы, пакеты программ) должны включать инструментарий для выполнения экспертных оценок и моделирования как основы прогноза состояния объектов, среды, средства и инструменты для оценок точности принимаемых интерпретаций, моделей. В силу этого эффективная система подготовки данных для компьютерных моделей невозможна без следующего: обоснованной схемы размещения точек контроля, фиксированных или перемещающихся пунктов наблюдений, которые обеспечены необходимыми техническими средствами для замеров, визуальных наблюдений и накопления данных; сформированных и постоянно дополняемых баз атрибутивных и пространственных данных, географических информационных систем (ГИС), обеспечивающих систематизацию, хранение, анализ результатов наблюдений и измерений, визуализацию этих результатов; цифровых (математических, численных) моделей, работающих с информационной основой, обеспечивающих планирование наблюдений и экспериментов, построение и оценку прогнозных сценариев, поддержку принятия управленческих решений.

I. ОСНОВНЫЕ ВОЗМОЖНОСТИ СИСТЕМЫ ГЕОБАЗАДААННЫХ

При проектировании и реализации ГеоБазаДанных (ГБД) в качестве определяющих были приняты требования, чтобы создаваемая компьютерная система, обеспечивала сбор, хране-

ние, обработку и визуализацию любых видов географически привязанной, пространственно-координированной информации [1]. Разрабатываемая система должна быть основой, сервисом автоматизированного рабочего места специалиста для решения типовых задач, связанных с инвентаризацией, анализом, моделированием, прогнозированием и управлением; инструменты, программные модули должны обеспечивать возможность оперировать с атрибутивной (описывающей свойства объекта) и пространственной (описывающей местоположение объекта) информацией. В системе следует предусмотреть возможности хранения пространственной информации в векторном и/или в растровом представлении; возможности организации графической информации в виде тематических слоев, соответствующих либо ее конкретному содержанию, либо решаемым задачам, когда наложение разных слоев позволит формировать тематические карты.

В настоящее время разработаны и применяются большое количество разнообразных как инструментальных ГИС, так и систем, адаптированных для решения задач в конкретных отраслях. ГеоБазаДанных – программная платформа, интерактивная компьютерная система, объединяющая комплекс интеллектуальных вычислительных подсистем, математического, алгоритмического и программного обеспечения. ГБД предназначена и используется (более 20 лет) для наполнения, сопровождения и визуализации различных баз данных, в том числе: геометрических, геофизических, техногенных, геолого-экологических параметров геологических объектов, экосистем; характеристик тепло-, влаго-, массопереноса в породах, грунтах и атмосфере; цифровых и регламентных описаний источников и видов воздействия на природную среду; входных данных для имитационных и математических моделей.

Инструментарий ГБД позволяет выполнять предобработку и визуализацию исходных данных, характеризующих физико-географические

и гидрографические особенности исследуемой территории. Средствами системы можно формировать и визуализировать цифровые описания пространственных распределений данных об источниках загрязнения, о геологическом строении изучаемых объектов; графически иллюстрировать решения задач, описывающих динамические процессы миграции флюидов, переноса тепла, влаги, минеральных водорастворимых соединений в толщах пород; конструировать и реализовывать интерактивные сценарии визуализации и обработки результатов вычислительных экспериментов. Специализированные подсистемы ГБД позволяют рассчитывать и выполнять в разных приближениях экспертные оценки локальных и интегральных характеристик экосистем, выполнять расчет распределений концентраций и массовых балансов загрязняющих веществ; формировать и выводить на твердые копии тематические карты.

II. КОНЦЕПЦИЯ, ОСНОВЫ ГБД, ДОПОЛНЕННАЯ ИНТЕЛЛЕКТУАЛЬНОСТЬ

Ядром и теоретической основой для построения цифровых моделей эко- и геосистем в концепции, реализованной в системе ГБД, является цифровое описание распределения изучаемого параметра на выбранном пространственном слое. Таковым параметром может быть, например, концентрация загрязняющего вещества в приповерхностном слое почвы, уровень дефолиации деревьев лесного массива на конкретном участке, насыщенность нефтью пропластка или пласта, значение абсолютной отметки уровня ограничивающей пласт (слой почвы, территории) поверхности. Считается, что исходными данными при этом являются *цифровые значения наблюдаемого параметра в точках с известными геометрическими координатами*, причем исходные точки могут быть размещены на площади в плане нерегулярно (например, данные наблюдений неравномерно расположенной сети станций), а в цифровом описании ГБД – на равномерной прямоугольной сетке (также употребляют названия – правильной, регулярной). Подобные цифровые поля системы – не что иное, как сеточные функции, а с ними можно работать средствами численного анализа. В частности, определять экстремумы, дифференцировать, интегрировать, выполнять арифметические действия с несколькими полями, анализировать корреляцию. Подобные цифровые поля визуализируются в режимах одномерной, двумерной и трехмерной графики. Другими словами основной задачей моделирования в системе ГБД является получение адекватного цифрового описания каждого параметра на слоях.

Как создавать, адаптировать цифровые модели, используя инструментарий системы ГБД, описано в [1]. Следует отметить отличительные особенности реализованных в ГеоБазаДанных алгоритмов аппроксимации, восстановления цифровых распределений, из которых важнейшими являются методы математического описания, названные в системе терминами: *Включение, Граница, Граничное значение, Зона относительной однородности, Краевые условия, Подобласть, Разрыв*. Перечисленное реализовано авторскими алгоритмами, запрограммировано в соответствующих модулях ГБД, и при проведении расчетов в системе выполняется интерактивно пользователем. Часто, так как спектр возможных действий очень широк, пользователю при подборе желаемого варианта, отвечающего запросам эксперта, приходилось сопоставлять большое число результатов вычислительных экспериментов. В текущей конфигурации ГБД добавлены модули реализации инструментов кластерного анализа [2], искусственных нейронных сетей [3], что существенно ускоряет выбор предпочтительных вариантов адаптации цифровых распределений.

ЗАКЛЮЧЕНИЕ

В докладе будут обсуждены методические и алгоритмические решения выбора предпочтительного варианта адаптации цифровых полей, отвечающих априорным требованиям экспертов. Специальными расчетами иллюстрируются особенности результатов, получаемых в вариантах настройки искусственной нейронной сети при использовании методов ADAM, RMSProp. В сериях результатов, когда применяются инструменты кластерного анализа, подготовлены и будут интерпретированы решения, полученные разными методами кластеризации, в частности, когда используется функция FindClusters с критериями CriterionFunction и нормами вычисляемыми по метрикам DistanceFunction, EuclideanDistance, Manhattan Distance, Chessboard Distance, Bray Curtis Distance.

III. СПИСОК ЛИТЕРАТУРЫ

1. Таранчук, В. Б. Компьютерные модели подземной гидродинамики / В. Б. Таранчук // Минск : БГУ, 2020. – 235 с.
2. Taranchuk, V. Methodological and Technical Solutions for the Implementation of Clustering Algorithms in the GeoBazaDannych System / V. B. Taranchuk // Communications in Computer and Information Science book series (CCIS. Springer, Cham. International Conference on Open Semantic Technologies for Intelligent Systems. OSTIS 2021). – 1625. – 2022. – P. 349–360.
3. Таранчук, В. Б. Средства и примеры интеллектуальной обработки данных для геологических моделей / В. Б. Таранчук // Проблемы физики, математики и техники. –2019. –№ 3 (40). –С. 117–122.