

КОГНИТИВНЫЙ ПОДХОД К АНАЛИЗУ МНОГОМЕРНОЙ МЕДИКО-ЭКОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

Актуальность.

В последние десятилетия произошла глобальная компьютеризация всей интеллектуальной деятельности, и процесс познания уже не мыслится вне связи с новыми компьютерными технологиями, в которых соединяется творческий потенциал человека и возможности компьютерных систем. Одна из ближайших задач этого направления – развитие методов визуализации данных на основе когнитивной компьютерной графики [1].

Использование когнитивных принципов особенно актуально при разработке интеллектуальных систем принятия решений, где выявление, систематизация и интеграция знаний о предметной области производится на основе информативных визуальных образов.

К визуализации можно отнести любые способы наглядного представления результатов эксперимента или теоретических исследований. Однако, наиболее известные инструменты визуализации – графики или диаграммы оказываются не достаточно эффективными, когда речь идет о наглядном представлении данных высокой размерности. Чтобы выявить наиболее важные свойства данных, тенденции их изменения и другие закономерности, исследователь должен перейти к более сжатому и удобному для анализа способу их представления.

В задачах анализа многомерной медико-экологической информации особый интерес представляет визуализация семантической структуры данных, способствующая выявлению опасных событий (идентификации опасности), оценке и прогнозированию рисков, связанных с изменением внешних условий среды, режима работы предприятий и т.п.

Постановка задачи.

Для заданной системы признаков таблица данных содержит описание состояния исследуемого объекта с максимальной полнотой. Чтобы извлечь пользу из этого описания, нужно перейти к более удобной форме его представления, т.е. обратиться к модели, которая реализует подходящий способ интерпретации данных. При анализе данных мониторинга чаще всего используется гипотеза о статистическом происхождении данных, которые рассматриваются как выборка из генеральной совокупности .

Анализ данных направлен на извлечение информации из таблиц, построенных в результате наблюдений (измерений), с целью получить наглядное, лаконичное и полезное описание этих данных [2]. При этом могут решаться разные задачи анализа, в том числе уменьшение размерности данных, разбиение на классы, выявление связей между параметрами и т.п. В

результате анализа исследователь получает определенное целостное представление о внутренней структуре данных – наглядный образ, который позволяет от количественных характеристик перейти к качественному описанию исследуемой совокупности таблиц.

В терминах теории вероятностей такие целостные образы (модели данных) можно описать законами их вероятностного распределения. Так, для описания данных социально-экологических исследований обычно применяют модель нормального распределения данных. (В пространстве признаков такие данные образуют один кластер с единственной точкой сгущения – точкой средних значений всех признаков. Более 60 процентов точек попадают в область эллипсоида с центром в точке сгущения и осями, равными собственным значениям *ковариационной матрицы*.)

Проблема визуализации заключается в том, чтобы представить многомерное облако данных в виде двумерной картинки, т.е. снизить размерность многомерного распределения до двух измерений. Согласно [2], *под визуализацией данных понимается такой способ представления многомерного распределения данных на двумерной плоскости, при котором качественно отражены основные закономерности, присущие исходному распределению – его кластерная структура, топологические особенности, внутренние связи между признаками и другие характеристики.*

Визуальный анализ данных в нашем понимании включает не только методы многомерного анализа информации, но и средства визуальной интерпретации полученных результатов. Некоторые популярные методы визуального анализа данных представлены в [3] и включают несколько типов информационных моделей, обеспечивающих наглядное представление и анализ многомерной информации на плоскости (2М-графики) или в трехмерном пространстве (3М-графики).

Для анализа данных экологического мониторинга, прогнозирования и управления качеством природной среды в современных условиях применяются геостатистические методы анализа данных, которые могут обеспечить преобразование многомерных таблиц, содержащих результаты наблюдений, в вероятностные характеристики этих данных, необходимые для оценки уровней экологического риска и прогнозирования опасных ситуаций. Полученные вероятностные характеристики с помощью ГИС-технологий можно представить в виде серии карт предполагаемых техногенных нагрузок или карт территориальных рисков.

Методы исследования.

К традиционным методам визуального анализа данных относится целенаправленное проецирование и многомерное шкалирование. Также в последнее время получили известность метод визуализации с помощью карт Кохонена и метод упругих карт, разработанный А.Ю. Зинovieвым [2].

Методы целенаправленного проецирования (projecting pursuit) осуществляют поиск такого отображения многомерного пространства на плоскость, которое оптимизирует заданный критерий. Достаточно часто

полагают, что отображение U является линейным оператором. При этом оптимизируемый функционал выступает в качестве *проекционного индекса* $Q(U, X)$, где X – исходный набор многомерных данных. К наиболее известным методам проецирования относится метод главных компонент, который можно использовать для построения комплексных оценок экологического состояния – экологических индексов [4].

Методы многомерного шкалирования предназначены для обработки данных о попарных расстояниях между объектами, где понятие расстояния в пространстве признаков рассматривается как аналогия сходства или различия. *Дистанционные модели различия* наиболее предпочтительны при визуальном анализе семантических пространств в социологических или психологических исследованиях [5].

В то же время, при анализе экологической информации в большинстве случаев речь идет о многомерных данных экологического мониторинга, привязанных к конкретным территориям. Поэтому содержательный анализ пространственно распределенных данных завершается картографическим представлением результатов анализа на основе ГИС-технологий.

Модели медико-экологических данных

Таблицу данных мы воспринимаем как непосредственное описание состояния исследуемой системы. Чтобы извлечь пользу из этого описания, необходимо указать метод представления его в более удобной для осмысления и оперирования форме. Собственно, речь идет о моделировании, т.е. создании информационной модели данных. Модель реализует определенный способ интерпретации исследуемых данных, сопоставляя объекту некоторое абстрактное его описание.

Из множества функций, которые выполняют модели, выделим две наиболее существенные. Во-первых, модель можно рассматривать как «аккумулятор знаний» об исследуемом фрагменте действительности, который позволяет прогнозировать поведение исследуемого объекта без реальных экспериментов. Во-вторых, модели играют определенную роль в образовании смыслов: подводят нас к выявлению новых знаний, понятий и терминов, способствуют упорядочиванию накопленной информации.

На начальных этапах создания модели рассматривается гипотеза о происхождении данных. В задаче анализа медико-экологических данных обычно используется два варианта гипотезы:

1. Модель основана на гипотезе о статистическом происхождении данных. Эта гипотеза состоит в том, что набор данных является выборкой из бесконечной генеральной совокупности объектов, распределение которой подчинено определенному вероятностному закону. Принятие этой гипотезы обеспечивает возможность применения к исследуемым данным теоретико-вероятностных подходов. Однако корректность этой гипотезы требует строгого доказательства, что далеко не всегда удается сделать.

2. Модель основана на гипотезе о динамическом законе. Можно предположить, что данные имеют не статистическую природу, а получены

как результат детерминированного воздействия определенных факторов, но с наложением определенных флуктуаций (шумов), которые могут быть описаны статистическими законами. При этом выбор закона основан на априорных соображениях или интуиции исследователя.

Большинство моделей прикладной статистики построено на основе гипотез о происхождении данных и привязывает данные к известным модельным схемам, что не всегда корректно обосновывается.

Визуальный анализ данных не использует подобных гипотез. Он направлен на исследование информационной модели данных без исследования внутренних механизмов, которые ее порождают (принцип «черного ящика» в кибернетике). Исследователь должен описать систему так, как она проявляет себя для внешнего наблюдателя. В то же время, модель должна быть полезной для прогнозирования новых событий и ситуаций.

Информационные модели медико-экологических данных могут использоваться для решения следующих задач:

- моделирование реакции системы на внешнее воздействие;
- классификация состояний системы (событий, ситуаций);
- прогноз динамических изменений состояния;
- анализ значимости используемых параметров;
- оптимизация состояния системы;
- управление поведением системы.

Для решения всех перечисленных задач может быть полезным создание визуальных образов исследуемых данных, с помощью которых удобно анализировать их структуру и свойства.

Когнитивный анализ многомерной информации.

Когнитивный анализ естественно определить как интеллектуальный анализ данных, ориентированный на выявление понятийной структуры исследуемой информации. Когнитивный анализ данных включает методы и компьютерные технологии, обеспечивающие представление данных в виде графических образов, которые отображают семантические характеристики этих данных. Такие информативно насыщенные образы будем называть когнитивными (понятийными) паттернами.

Большинство исследователей использует понятие о визуализации в узком смысле – как графическое представление данных или результатов анализа. Тем не менее, существует множество задач, где возможности интеллектуального анализа информации следует дополнить достижениями современной компьютерной графики. В более общем случае когнитивный подход к анализу многомерных данных должен включать не только сам процесс визуализации, но и весь комплекс задач, который с ним связан, т.е. методы математической статистики, анализ данных, визуализацию и интерпретацию полученных результатов.

Основная цель когнитивного подхода – это извлечение полезной информации (знаний). Однако в каком виде должны быть представлены эти

знания? Как соотносить информацию и знания?

Мы знаем два подхода к понятию информации. Первый из них лежит в основе *теории информации* и связан с именем К. Шеннона. Второй подход предложен Р. Фишером и относится к области математической статистики. Он использует понятие *достаточных статистик*, по которым можно восстановить исходное распределение данных. Так, для нормального распределения данных достаточными статистиками являются среднее арифметическое значение и выборочная дисперсия.

В общем случае извлечением информации считается такое описание данных, которое будет достаточно лаконичным, наглядным и полезным для решения поставленных задач. Таким образом, когнитивный анализ данных должен дать нам наглядное, сжатое и полезное их описание.

Рассмотрим более детально два направления когнитивного анализа, наиболее перспективных для задач медико-экологического мониторинга. К первому из них относятся методы, ориентированные на выявление структуры данных в пространстве исходных признаков и визуальное представление полученных результатов (методы классификации, кластерный анализ, распознавание образов). Второе направление когнитивного анализа – это поиск подходящих координат для визуализации пространства признаков. Здесь речь идет о выявлении или построении наиболее информативных признаков, с помощью которых можно восстановить значения всех остальных и определить зависимости, связывающие признаки.

Факторный анализ, например, позволяет представить наиболее информативные признаки (факторы) в виде линейных комбинаций из измеренных показателей. Для визуализации пространства признаков используется ортогональное проецирование данных на плоскость главных компонент (главных факторов).

Заметим, что наиболее важным результатом когнитивного анализа является качественное представление о внутренней структуре набора данных и о зависимостях, присущих этим данным. При этом количественные характеристики, извлеченные разными методами, играют второстепенную роль на фоне целостного визуального образа, который можно рассматривать как *информационный портрет* исследуемого объекта.

Когнитивный анализ рисков на основе данных мониторинга

Если в качестве случайных событий рассматриваются различные последствия воздействия вредного фактора, то значения вероятностей для разных исходов служат характеристиками риска, обусловленного данным фактором [6]. Для любой конкретной ситуации вероятности отдельных исходов будут удовлетворять известным свойствам вероятностной меры, т.е. вероятность любого события неотрицательна, вероятность полного события равна 1, вероятность суммы непересекающихся событий равна сумме вероятностей этих событий. Например, если в качестве возможных исходов принять степени тяжести конкретного заболевания, то для любой степени тяжести величина риска попадет в интервал от 0 до 1, а риск суммы двух

исходов равен сумме двух рисков.

Для анализа и содержательной интерпретации многомерных данных медико-экологического мониторинга предлагается воспользоваться схемой, представленной на рис. Эта схема включает основные этапы анализа, визуализации и вероятностной интерпретации исследуемой многомерной информации, которые описаны далее.

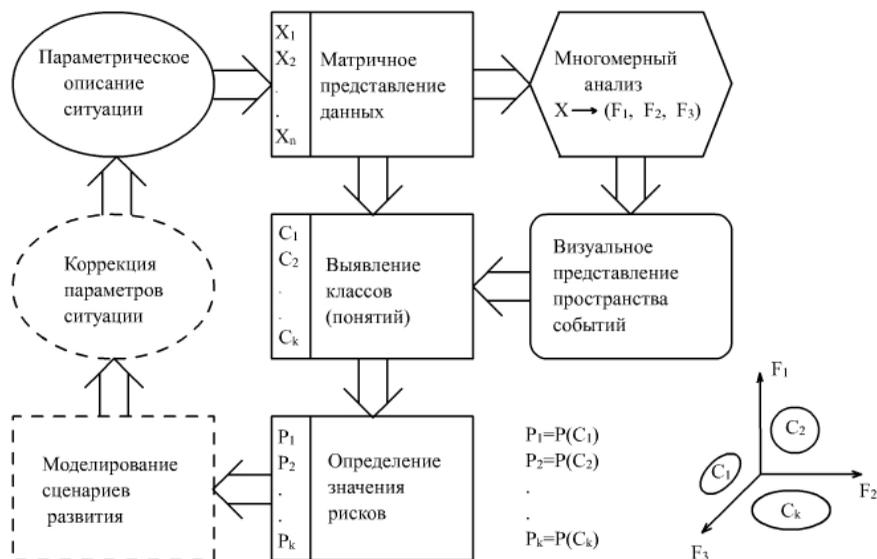


Рис. Схема когнитивного анализа данных мониторинга

1. Анализ пространства наблюдений

Если состояние исследуемой территории описывается множеством n наблюдений, каждое из которых включает m параметров, то для него можно построить пространство событий размерности m . Исходные наблюдения будут представлены в этом пространстве как множество точек с координатами, соответствующими числовым значениям указанных параметров. В качестве случайных событий можно рассматривать любые подмножества исходного множества наблюдений.

Если показатели, определяющие риск, известны заранее и являются параметрами пространства событий, то второй этап работы алгоритма упрощается. В этом случае для анализа неблагоприятных событий можно рассматривать отдельные подпространства (проекции размерности 2 или 3) пространства исходных параметров.

В противном случае необходимо понизить размерность пространства событий с помощью методов факторного анализа или проектирования на многомерную шкалу.

2. Визуальная интерпретация факторного пространства.

После снижения размерности пространство случайных событий приобретает визуальную интерпретацию (плоскую или трехмерную), удобную для семантического анализа на экране монитора. Множество событий может быть визуализировано как размытое облако, отдельные скопления точек или их произвольные конфигурации.

Напомним, что вопросы, связанные с использованием методов многомерного анализа данных для снижения размерности и визуальной интерпретации сложных экологических ситуаций, достаточно подробно рассматриваются в работе [4].

Выделенные факторы образуют семантические шкалы, с помощью которых можно оценить отдельные наблюдения [7]. Каждый из факторов объединяет несколько показателей как более сложный признак. Он может быть проинтерпретирован в соответствии с тем смыслом, который он обобщает в данной конкретной ситуации.

3. Качественная интерпретация отдельных событий.

Из всей совокупности возможных событий выделим те события, которые могут иметь качественную интерпретацию. Как правило, на этом этапе необходимо привлечение экспертных знаний о предметной области, чтобы установить связи между численными значениями показателей и теми классами, в которые попадают наблюдения. Каждая группа наблюдений может быть интерпретирована как отдельное понятие. В экологических исследованиях такие понятия чаще всего соответствуют разным уровням загрязнения территории или уровням экологического риска.

4. Оценка степени риска, связанного с каждым событием.

Каждому из выделенных классов наблюдений следует присвоить определенное значение риска. Степень риска, связанную со случайным событием, желательно оценивать в рамках определенной системы знаний (эталонных событий или прототипов). Примеры таких систем были разработаны в соответствии с методологией ЕРА [4, 6]. В большинстве случаев системы эталонов основаны на эмпирических данных, которые связывают уровни техногенных нагрузок (концентрации, дозы) с риском возникновения тех или иных заболеваний.

Предложенный алгоритм позволяет дать оценку потенциального риска, связанного с возможным изменением ситуации. Ситуация в данном контексте представлена как набор параметров, т.е. за счет изменения значений отдельных показателей существенно меняется качество ситуации в целом. Корректируя значения параметров в соответствии с возможными сценариями развития неблагоприятных ситуаций, можно построить оценки потенциального риска для наиболее опасных ситуаций.

Выводы

Обосновывается когнитивный подход к анализу многомерной информации, включающий методы интеллектуального анализа данных и

возможности когнитивной компьютерной графики

Задача когнитивного анализа многомерных данных формулируется как поиск наиболее подходящей содержательной интерпретации, которую может обеспечить их визуальное представление в пространстве обобщенных признаков (факторов риска).

Предложена методика структурирования экспериментальных данных, включающая многомерный анализ, выявление когнитивных паттернов и их содержательную интерпретацию в терминах опасных событий и рисков.

1. *Зенкин А. А.* Когнитивная компьютерная графика. М.: Наука, 1991. – 192 с.
2. *Зиновьев А. Ю.* Визуализация многомерных данных. Красноярск: ИПЦ КГТУ, 2000. – 168 с.
3. *Боровиков В.* STATISTIKA. Искусство анализа данных на компьютере: Для профессионалов. – СПб.: Питер, 2003. – 688 с.
4. *Сердюцкая Л.Ф., Каменева И.П.* Системный анализ и математическое моделирование медико-экологических последствий аварии на ЧАЭС и других техногенных воздействий. – К.: «Медэкол», 2000. – 173 с.
5. *Шмелев А.Г.* Введение в экспериментальную психосемантику: теоретико-методологические основания и психодиагностические возможности. М.: Изд-во МГУ, 1983. – 158 с.
6. *Большаков А.М., Крутько В.Н., Пуцилло Е.В.* Оценка и управление рисками влияния окружающей среды на здоровье населения. – М.: Эдиториал УРСС, 1999. – 256 с.
7. *Каменева И. П.* Просторово-семантичні моделі репрезентації знань в гео-екологічних дослідженнях // Геоінформатика. – 2005. – № 4. – С. 64 – 69.

Поступила 8.02.2010р.

УДК 004.056:004.274

Ю.М. Коростиль, А.Н. Давиденко, С.Я. Гильгурт, М.М. Панченко

АНАЛИЗ УГРОЗ И ОПАСНОСТЕЙ В КОМПЬЮТЕРНЫХ СИСТЕМАХ НА ПРЕДМЕТ ЗАЩИТЫ ЦИФРОВЫМИ РЕКОНФИГУРИРУЕМЫМИ УСТРОЙСТВАМИ

The utilization of the programmable logic for the computer security problems is investigated. A classification of threats and vulnerabilities in computer systems taking into account the potential of FPGA-based digital equipment when using for solving security tasks is proposed.

Реконфигурируемые вычисления, основанные на применении программируемых логических интегральных схем (ПЛИС), которые позволяют совместить быстрдействие специализированных аппаратных решений с