

7. Кауров Б.А. Живой организм как имитационная система // Биометрические аспекты изучения целостного организма. – М.: МГУ, 1987. – С. 82-91.
8. Кауров Б.А. Возрастная биоинформатика человека – новое направление в геронтологии // VI Всероссийская научно-практ. конф. "Общество, государство и медицина для пожилых". – М., 2009. – С. 33-34.

E-mail: bokar@mail.ru.

УДК 001.891.34:681.51.012

О.Г. Берестнева, д-р техн. наук, **Е.А. Муратова**, канд. техн. наук
(Томский политехнический университет)

ПРОБЛЕМЫ УНИФИКАЦИИ ДАННЫХ В НАУЧНЫХ ПСИХОЛОГИЧЕСКИХ И МЕДИЦИНСКИХ ИССЛЕДОВАНИЯХ¹

Представлен обзор приемов унификации разнотипных данных. Предложен эффективный алгоритм адаптивного кодирования разнотипной информации, позволяющий обрабатывать малые выборки, строить локальные диагностические шкалы, формировать новые знания на основе экспериментального материала.

Ключевые слова: информационные интеллектуальные системы, унификация данных и знаний, алгоритм адаптивного кодирования разнотипной информации.

Введение

В настоящее время использование информационных интеллектуальных систем (ИИС), в том числе и автоматизированных рабочих мест (АРМ), выполняющих роль эксперта и консультанта, бесспорно актуально. При их разработке решаются проблемы, связанные с представлением данных и знаний, выявлением закономерностей, формированием решающих правил, принятием и обоснованием решений. Выявление закономерностей из имеющихся массивов данных на начальных этапах проектирования ИИС и АРМ – один из наиболее сложных и трудоемких процессов, и он не всегда заканчивается успешно, поскольку базы данных содержат неполную, нечеткую, разнотипную и противоречивую информацию.

Например, специалисту в области медицинской психодиагностики придется оперировать следующими видами информации:

- 1) *психодиагностической* (данные психодиагностического эксперимента, измеренные в количественных, порядковых, номинальных шкалах);
- 2) *медицинской* (результаты параклинических методов исследования, имеющие количественный и интервальный характер);
- 3) *консультационно-биологической* (пол, возраст, образование, профессия, семейное положение – качественные шкалы);

¹ Работа выполнена при финансовой поддержке РФФИ (проект №08-06-00313а).

4) *социально-демографической* (образование, профессия, семейное положение – качественные шкалы).

Каждой шкале измерения соответствует свой способ математической обработки. Существуют процедуры для случаев, когда одни признаки измерены в одной шкале, а другие – в другой: например, дисперсионный анализ. Тем не менее большинство разработанных в настоящее время методов обработки экспериментальных данных рассчитано на анализ информации, представленной в количественной форме. Эти методы основаны на усреднении характеристик исследуемого объекта, вследствие чего исчезают индивидуальные качества объектов, приводящие к снижению эффективности математического анализа. Чтобы иметь возможность применять эти методы в случае разнотипных данных, исследователь сталкивается с проблемами унификации данных.

Унификации данных в научных психологических и медицинских исследованиях

Анализируемое многомерное признаковое пространство $\{x_1, x_2, \dots, x_i, \dots, x_p\}$ содержит признаки, измеренные в разных шкалах: количественных x_i^e ($i=\overline{1, q}$), порядковых x_i^i ($i=\overline{1, l}$), номинальных x_i^i ($i=\overline{1, m}$), где $q+l+m=p$ (p – общее количество признаков).

Предложены различные приемы унификации данных: 1) сведение количественных значений к двоичным признакам; 2) оцифровка номинальных и порядковых признаков [1]; 3) декомпозиция признакового пространства на подзадачи, в которые попадают признаки одинаковой физической природы с последующим объединением полученных решений [2].

В основе первого подхода лежит введение вместо каждого исходного признака x_i серии признаков $y_1, y_2, \dots, y_j, \dots, y_k$, где k – число градаций, если признак x_i , измерен в порядковой шкале, и количество интервалов для признака x_i , измеренного в количественной шкале, а сами значения признака y_j принимают два значения: 0 или 1. При таком способе унификации используется разбиение, предлагаемое экспертом, или область значений признака разбивается на равные по длине интервалы, число которых либо задается экспертом, либо определяется по соответствующим формулам. Основные недостатки данного подхода: 1) увеличение признакового пространства, 2) преобразование связано с закруглением признака, что может привести к потере содержащейся во всей статистике информации и снижению качества решения задачи.

Во втором подходе номинальные x^i и порядковые признаки x^i усиливают (оцифровывают) до уровня количественных x^e путем приписывания их градациям числовых значений [1]. Данный метод основан на условии максимальной корреляции между признаком, измеренным в более сильной шкале x^c , и оцифрованным признаком x^{*c} . В целом оцифровка качественных признаков является задачей сложной как в вычислительном, так и в теоретическом плане. Главным требованием в этом случае является сохранение тех или иных отношений, которые

априорно существуют между исследуемыми объектами. Недостатком данного подхода унификации является искажение имеющейся информации: ее добавление или ее потеря. А преимуществом оцифровки слабых шкал является то, что эта процедура позволяет использовать математические и статистические методы, разработанные для анализа экспериментальных данных с признаками, измеренными в количественных шкалах.

Третий подход унификации основан на декомпозиции общего множества признаков $\{x\}$ на подмножества [2]: количественных признаков $\{x^k\}$, порядковых признаков $\{x^n\}$, номинальных признаков $\{x^h\}$, бинарных признаков $\{x^b\}$. Такое разбиение достаточно условно, поскольку порядковые признаки могут рассматриваться как числовые, а бинарные – как номинальные и т.д. Следовательно, здесь также присутствует субъективизм экспертов, которые задают данную декомпозицию многомерного признакового пространства. Преимуществом этого разбиения является то, что при использовании методов теории распознавания образов для каждого такого подмножества могут быть использованы определенные метрики.

Проведенный анализ методов преобразования (унификации) разнотипной информации показал, что в настоящий момент не существует универсального способа унификации, и поскольку в исследованиях, проводимых авторами (моделирование механизмов эффективной интеллектуальной самореализации субъекта; выявление специфики когнитивного развития субъектов с высоким и сверхвысоким уровнем интеллекта; изучение влияния типов социальных взаимодействий на интеллектуальную самореализацию и социальную адаптацию одаренных студентов в технических вузах; исследование феноменологии, закономерностей динамики, факторов трансформации коммуникативного стресса студентов; выявление психологических особенностей и условий социализации в формировании репродуктивного поведения женщин) используются логические модели, то для их построения предложен алгоритм адаптивного кодирования разнотипной информации, что обеспечивает сохранность априорно имеющихся соотношений между результатами психологических и медицинских исследований [3, 4]. Данный алгоритм выделяет информативные интервалы и тем самым восстанавливает процедуру принятия решения экспертом в исследуемой предметной области, а также позволяет формировать новые знания на основе имеющегося экспериментального материала.

Кроме того, предложенный алгоритм адаптивного кодирования разнотипной информации используется при конструировании локальных диагностических или специфических норм (шкал) для обследования определенных групп людей. Эти исследования особенно актуальны в задачах диагностики слабоструктурированных предметных областей – таких как биомедицина, биоэкология, восстановительная медицина.

Заключение

Проведен анализ приемов унификации разнотипных данных. Предложен эффективный алгоритм адаптивного кодирования разнотипной информации. Но-

визна предложенного алгоритма заключается в возможности работать с малыми выборками, строить локальные диагностические шкалы, формировать новые знания на основе имеющегося экспериментального материала.

ЛИТЕРАТУРА

1. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. – Новосибирск: Изд-во Ин-та математики, 1999.
2. Вълев В. Задачи распознавания с большими массивами обучающей информации – метод выбора существенных подзадач // Сб. докладов юбилейной научной сессии. – Шумен, 1982. – С. 195-204.
3. Янковская А.Е., Муратова Е.А., Берестнева О.Г. Извлечение знаний с применением алгоритма адаптивного кодирования разнотипной информации // Искусственный интеллект (Донецк). – 2002. – №2. – С.315-322.
4. Муратова Е.А., Берестнева О.Г., Янковская А.Е. Эффективный алгоритм адаптивного кодирования разнотипной информации // Труды конгресса «Искусственный интеллект в XXI веке». – М.: ФИЗМАТЛИТ, 2001. – С. 155-167.

E-mail: cfpd@amur.ru.

УДК 614.1+314.484+57.087.1

Т.М. Смирнова

(Институт системного анализа РАН, Москва)

НОВЫЙ МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА СМЕРТНОСТИ

Представлен метод сравнительного анализа смертности для группы популяций с учетом возрастной динамики рисков смерти, основанный на ранжировании популяций по уровню смертности в возрастных интервалах таблиц смертности.

Ключевые слова: смертность, возрастные коэффициенты, метод анализа.

Выбор адекватного показателя смертности – одно из необходимых условий, обеспечивающих системность эпидемиологических, демографических и популяционных исследований. Этот выбор нередко достаточно сложен из-за сильной зависимости риска смерти от возраста. Например, для человека, достигшего 80 лет, вероятность смерти в течение следующего года в сотни раз выше, чем для 10-летнего. Поэтому при необходимости сравнивать смертность в разных популяциях или населенных пунктах наиболее простая ее характеристика – общий коэффициент смертности (*ОКС*), т.е. отношение числа умерших за определенный период времени к среднему числу живущих, пригодна лишь в случае, когда их возрастная структура примерно одинакова. В целях исключения погрешностей, обусловленных различиями возрастной структуры, при сравнении смертности населений с неодинаковой структурой используют стандартизованные коэффициенты смертности (*СКС*), для вычисления которых от структуры реальных населений переходят к структуре специально выбранного стандартного населения, или табличные