

УДК 62.12.14

А.Р. КОРСУНОВ

Украинская инженерно-педагогическая академия, Украина

РАЗРАБОТКА АВТОМАТИЗИРОВАННЫХ СРЕДСТВ ИНТЕЛЛЕКТУАЛИЗАЦИИ СУБД КОМПЛЕКСА ЭЛЕКТРОМАГНИТНОГО СОЧЕТАННОГО ВОЗДЕЙСТВИЯ НА БИООБЪЕКТЫ

В статье показано, что традиционные методы интеллектуального анализа обладают некоторыми ограничениями, что выдвигает в определённых случаях на первый план поиск специализированных средств интеллектуализации систем обработки данных.

интеллектуальный анализ, база данных, паттерн, модель пользователя, тиражирование, экспертная оценка

Введение

В качестве модели данных в теории баз данных используются различные алгебраические структуры, обобщающие классические модели данных: реляционную, иерархическую и сетевую [1]. Основная задача интеллектуального анализа структурированных данных состоит в выявлении скрытых правил и закономерностей в наборах данных [2]. Методы традиционной математической статистики оказываются полезными главным образом для проверки предварительно сформулированных гипотез (*VDDM – verification driven data maining*), поскольку она оперирует усреднёнными характеристиками [3], которые, в свою очередь, часто являются виртуальными.

1. Формулирование проблемы

Развитие технологий интеллектуального анализа данных связано с необходимостью аналитической обработки информации. Сложность и разнообразие методов интеллектуального анализа данных требуют создания специализированных средств конечного пользователя для решения типовых задач анализа информации в конкретных разделах исследовательской деятельности и обработки данных.

1.1. Анализ публикаций

Оперативная аналитическая обработка данных (*online analytical processing, OLAP*) помогает исследователю сформировать показатели, характеризующие тот или иной фрактал. Она представляет собой несколько десятков методов прогноза динамики показателей. Однако основанием её являются различные эмпирические модели [4]. Основным недостатком нейросетевой парадигмы является необходимость иметь очень большой объём обучающей выборки [5]. Система рассуждений на основе аналогичных случаев (*case based reasoning – CBR*) имеет основной недостаток в том, что при этом не создаются какие-либо модели, обобщающие предыдущий опыт [6]. Генетические алгоритмы имеют недостаток в самой своей основе, поскольку критерий отбора хромосом и сама процедура являются эвристическими. Подобный критерий не гарантирует нахождения «лучшего» решения, так как эволюция может оказаться тупиковой из-за выбора непродуктивной ветви [7].

1.2. Цель статьи

Создать специализированные средства автоматического интеллектуального анализа данных, адапти-

рующиеся к свойствам структуры БД комплекса сочетанного электромагнитного воздействия на био-структуры для решения диагностических задач успешности указанного сочетанного воздействия. Это позволит переложить бремя поиска паттернов, характерных для каких-либо фрагментов неоднородных многомерных данных и физиологов, биологов и врачей на компьютер.

2. Решение проблемы

Встроенная модель биообъекта в БД используется в режимах формирования параметров диагностических наборов и представляется в следующем виде:

$$\forall_z \in [1 \dots n]: S^z \subset S, \quad (1)$$

где \forall_z – набор диагностики;

S – множество параметров, состоящее из отдельных разделов S^z ;

z – общее число разделов (показатели температуры, показатели кровеносной системы и т.д.).

Раздел представляет совокупность подразделов (систолическое давление, диастолическое давление и т.п.), для каждого из которых задаётся мера значимости:

$$S^z = \left\{ \left[S_1^z \right]^{C_1^z}; \left[S_1^z \right]^{C_2^z}, \dots, \left[S_1^z \right]^{C_n^z} \right\}; \quad C_v^z \in [0,1], v \in \overline{1, n}, \quad (2)$$

где v – число подразделов в разделе S^z ; C_v^z – величина, характеризующая степень важности при диагностике специалистом по исследованию из подраздела S_v^z .

Если $S_1^z, S_2^z, \dots, S_v^z$ – подмножества (не обязательно непересекающиеся) диагностических задач одного раздела, то запишем:

$$\forall_v \in \overline{1, n}: S_v^z \subset S^z; S_v^z = \left[S_{v1}^z, S_{v2}^z, \dots, S_{vm_v}^z \right], \quad (3)$$

где $S_{v1}^z, S_{v2}^z, \dots, S_{vm_v}^z$ – диагностические задания в подразделе S_v^z ;

m_{vm} – число диагностических заданий в подразделе S_v^z .

Учитывая выражения (3) и (4), записываем полный набор параметров, необходимых при диагностике по заданному разделу

$$S^z = \left\{ \left[S_{11}^z, S_{12}^z, \dots, S_{1m_1}^z \right]^{C_1^z}, \left[S_{21}^z, S_{22}^z, \dots, S_{2m_2}^z \right]^{C_2^z}, \dots, \left[S_{v1}^z, S_{v2}^z, \dots, S_{vm_m}^z \right]^{C_v^z} \right\}. \quad (4)$$

Полные наборы диагностических заданий для каждого из разделов формируются группой экспертов и заносятся в базу данных.

2.1. Модель пользователя

Для обеспечения эффективности процесса электромагнитного воздействия и получения адекватной оценки диагностики в систему вводятся модель пользователя, формально представляемая в виде объекта

$$MS = \left\{ ID, (a_1, \langle p.a_1 \rangle), (a_2, \langle p.a_2 \rangle), \dots, \left[S_{v1}^z, S_{v2}^z, \dots, S_{vm_v}^z \right]^{C_v^z} \right\}, \quad (5)$$

где ID – идентификатор;

a_1, a_2, a_i – имена полей;

$p.a_1, p.a_2, p.a_i$ – значения полей.

В качестве имен полей выступают статические характеристики пользователя, а значения полей – шкалированные оценки этих характеристик.

В базе данных системы содержится набор функций преобразования, обеспечивающий формиро-

ние множества требований предъявляемых при диагностике:

$$S \xrightarrow{f_z(\{p-a_i\})} S^z;$$

$$S^z = \left\{ \left[S_{11}^z, S_{12}^z, \dots, S_{1m_1}^z \right]^{C_1^z}, \left[S_{21}^z, S_{22}^z, \dots, S_{2m_2}^z \right]^{C_2^z}, \dots, \left[S_{v1}^z, S_{v2}^z, \dots, S_{vm_v}^z \right]^{C_v^z} \right\}. \quad (6)$$

2.2. Критерии экспертной оценки

Оценка уровня достигнутых параметров при электромагнитном взаимодействии в пределах одного раздела выполняется в соответствии с критерием

$$\theta^z = \frac{1}{V} \sum_{v=1}^V \frac{C_v}{m_v} \sum_{m=1}^{m_v} T_{mv} \cdot S_{mv}, \quad (7)$$

где V – число подразделов;

C_v – показатель значимости для подраздела;

m_v – число заданий в подразделе;

T_{mv} – мера трудности задания;

S_{mv} – булева переменная, равная 1 в случае достижения положительного эффекта при электромагнитном взаимодействии и 0 – в противном случае.

Условием успешного проведения сеанса электромагнитного взаимодействия является выполнение следующего неравенства

$$\theta_{\min}^z \leq \theta^z, \quad (8)$$

где θ_{\min}^z – заданное в режиме формирования требований минимально допустимое значение критерия для данного раздела.

Исходные данные для оценки представляется в виде:

- фактов;
- правил;
- дерева параметров;
- дерева контекстов;
- функционированной иерархии разделов.

Данные, используемые системой, делятся на статические и динамические. Статические данные хранятся в базе данных и не зависят от интерактивного режима. Они включают в себя правила сравнения, вывода и сведения об иерархии объектов. Динамические знания включают в себя дерево целей, активные факты и правила, используемые в интерактивном режиме.

В системе хранится фиксированная иерархия объектов:

- 1) биоструктура;
- 2) пациент;
- 3) форма взаимодействия с ЭМП;
- 4) методика взаимодействия.

2.3. Адаптивное тиражирование данных

Существующие методы разработки реляционных БД используют так называемую статическую схему тиражирования данных [8]. Это требует предпроектного исследования интенсивности операций чтения и записи в системе на основании статических данных, изменение указанных параметров в процессе эксплуатации требует модификации прикладного программного обеспечения [9].

В связи с этим в СУБД используем метод адаптивного тиражирования репликаций со следующим алгоритмом, учитывая при этом, что информационная сеть комплекса содержит 6 узлов.

Итак, пусть $S = \{A, B, C, D, E, F\}$ – сеть с топологией, заданной на рис. 1.

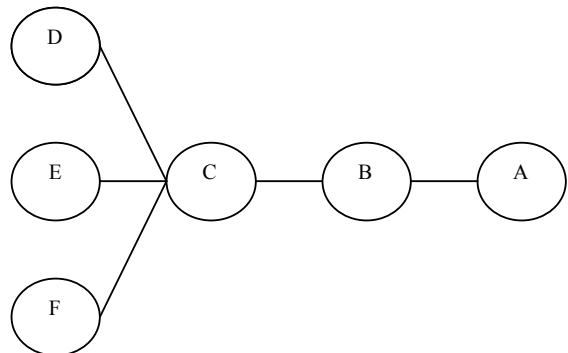


Рис. 1. Фрагмент заданной топологии информационной сети

Репликационная схема R состоит из узлов C, D, E и $R = \{C, D, E\}$. В общем случае, сеть представляет собой дерево с двунаправленными ребрами $S = (U, E)$, где U – множество узлов, а E – множество рёбер. Репликационная схема Q – непустое подмножество S . Определим алгоритм модели «чтение-запись» объекта для узла i заданной сети S и репликационной схемы Q в виде пары $A = \{(R_i, W_i) | I = 1, n\}$, где R_i – количество операций чтения, W_i – количество операций записи, I – узел, куда тиражируются данные, а n – число узлов в сети. Обозначим r_i – вес операции чтения объекта, а w_i – вес операции записи. Отсюда для заданной репликационной схемы Q модели «чтение-запись» A вес репликационной схемы $C(Q, A)$:

$$\sum W_i \cdot w_i + \sum R_i \cdot r_i, \quad (9)$$

для всех i , принадлежащих U .

При данном алгоритме формирования репликационной схемы минимизируются стоимость коммуникационных затрат и время тиражирования данных.

Заключение

Разработанные в работе специализированные средства для СУБД комплекса сочетанного электромагнитного воздействия на биообъекты (КЭМВБ) основаны на концептуально новых принципах. В результате информационная система КЭМВБ не просто хранилище информации, а благодаря встроенным моделям диагностики, модели пользователя в виде биообъекта, адаптивному тиражированию данных ей приданы свойства интеллектуального анализа данных. Это порождает процесс управления решением диагностических и инструментальных задач

и, в конечном счёте, приводит к продуцированию знаний.

Литература

1. Жуковский В.Д. Автоматизированная обработка данных клинических функциональных исследований. – М.: Медицина, 1985. – 352 с.
2. Интеллектуальные технологии ввода и обработки информации / Под ред. В.Л. Арлазорова, Н.Е. Емельянова. – М.: УРСС, 2001. – 256 с.
3. Ульман Дж. Основы систем баз данных. – М.: Мир, 1983. – 405 с.
4. Кардиомониторы. Аппаратура непрерывного контроля ЭКГ: Учебное пособие для вузов / Под ред. А.Л. Барановского и А.П. Немирко. – М.: Радиосвязь, 1993. – 422 с.
5. Муха Ю.П., Скворцов Н.Г., Авдеюк О.А. Принципы построения медицинских диагностических комплексов на базе нейросетевых технологий // Биомедицинская радиоэлектроника. – 2001. – № 4. – С. 42 – 47.
6. Баевский Р.М., Кириллов О.И., Клецкин С.М. Математический анализ изменений сердечного ритма при стрессе. – М.: Наука, 1984. – 194 с.
7. Левыкин В.М., Чалый С.Ф. Особенности моделирования сложных информационных систем на основе генетических алгоритмов // Проблемы бионики. – 2000. – Вып. 52. – С. 38 – 39.
8. Дейтл Г. Введение в операционные системы. В 2-х томах. – М.: Мир, 1987. – Т. 1. – 359 с. – Т. 2. – 400 с.
9. Керниган Б., Пайк Р. Unix: Универсальная среда программирования. – М.: Финансы и статистика, 1992. – 302 с.

Поступила в редакцию 20.12.2004

Рецензент: д-р техн. наук, проф. В.Д. Сахацкий,
Украинская инженерно-педагогическая академия,
Харьков.