

## **ВЫБОР ВЕРОЯТНОСТНЫХ ПОКАЗАТЕЛЕЙ КАЧЕСТВА ИНФОРМАЦИИ БАЗ ДАННЫХ АВТОМАТИЗИРОВАННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ**

С.Н. Звиглянич

(Объединенный научно-исследовательский институт ВС Украины, Харьков)

*Обоснован выбор вероятностных показателей качества информации базы данных автоматизированной информационной системы. Качество информации базы данных предлагается оценивать с учетом времени ее обновления. Данные показатели отражают суть протекающих процессов, являются критичными и чувствительными по отношению к качеству информации, циркулирующей в автоматизированной информационной системе.*

***качество информации, автоматизированная информационная система***

**Постановка проблемы в общем виде.** Принятие решений при управлении сложными системами предопределяет необходимость целенаправленной переработки значительных объемов информации. В этом аспекте системообразующим элементом в современных АСУ является автоматизированные информационные системы (АИС). На АИС возлагаются функции сбора, хранения, поиска и обработки информации. Требования интеграции информации, циркулирующей в АИС, приводят к необходимости создания баз данных (БД). Решение ряда информационно-расчетных задач, принятие решений с использованием хранящейся в БД информации обуславливают требования учета качества данной информации. Таким образом, возникает необходимость введения показателей, отражающих качество используемой информации БД автоматизированной информационной системы.

**Анализ последних исследований и публикаций.** Проблема оценки качества функционирования автоматизированных информационных систем приобретает все большую актуальность в связи с повышением уровня автоматизации процесса принятия решений. Это связано с увеличением объемов обрабатываемой информации и сокращением времени на принятие решения [1]. Задачи анализа и синтеза средств обработки информации, рассмотренные в [2], в основном освещают проблему эффективного использования каналов связи, применения оптимизационных алгоритмов маршрутизации. В [3] рассматриваются подходы к постро-

нию БД, дается общая оценка различных моделей представления данных. В [4] используются довольно известные подходы к оцениванию степени риска при принятии решений. Данные подходы ориентируются, прежде всего, на безусловное знание вероятностей состояний обстановки. При этом информационная составляющая процесса принятия решения не выделяется и не рассматривается.

Можно отметить, что на сегодняшний день остается актуальным вопрос проведения количественной оценки качества информации, циркулирующей в АИС. При этом необходимо использовать легко вычисляемые, доступные показатели, отражающих цель функционирования данных систем.

**Постановка задачи.** Современные системы управления базами данных (СУБД) ориентированы на реляционную модель данных, которая предусматривает построение баз данных в виде таблиц. Каждому реальному объекту соответствует объект (запись) баз данных как некая его информационная модель. Соответственно, каждому параметру реального объекта поставлено в соответствие поле записи БД. Каждый параметр объекта характеризуется вполне определенным временем – периодом обновления информации. Отметим, что использование значения параметра, у которого период обновления превышен, крайне не желательно. Устаревшая информация является источником ошибок различных вычислений, а также увеличивает степень риска при принятии решений. Исходя из выше сказанного, с учетом периодов обновления записей баз данных, поставим цель – обосновать показатели, характеризующие качество информации БД АИС.

**Обобщенное теоретическое решение задачи.** Рассмотрим базу данных. Записи базы данных имеют  $N$  полей. Рассмотрим  $i$ -ое поле записи. Представим интервалы обновления этого поля как случайные величины  $T_{oi}$  (рис. 1). Сделаем допущение, интервалы обновления образуют простейший поток событий с интенсивностью  $\lambda_i$ . Время решения задачи с использованием информации из БД строго фиксировано, т.е. интервалы  $T_p$  постоянные. С учетом того, что поток  $T_{oi}$  простейший [5] вероятность хотя бы одного обновления на интервале  $T_p$

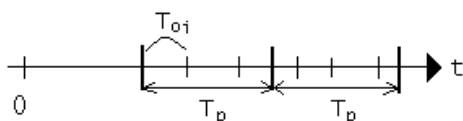


Рис. 1. Шкала времени,  $T_p$  фиксированы

Сделаем допущение, интервалы обновления образуют простейший поток событий с интенсивностью  $\lambda_i$ . Время решения задачи с использованием информации из БД строго фиксировано, т.е. интервалы  $T_p$  постоянные. С учетом того, что поток  $T_{oi}$  простейший [5] вероятность хотя бы одного обновления на интервале  $T_p$

$$P_{i(>1)} = 1 - e^{-\lambda_i T_p}. \quad (1)$$

Тогда для записи в целом можно ввести обобщенный показатель – вероятность обновления записи

$$P_Z = \prod_{i=1}^N P_{i(>1)}. \quad (2)$$

Отметим, что конъюнктивная форма обобщенного показателя при достаточно большом  $N$  имеет низкую чувствительность.

При большом количестве полей в записи имеет смысл обобщенный показатель представить в дизъюнктивной форме. Для этого рассмотрим простейший поток  $T_{oi}$ . Математическое ожидание  $T_{oi}$  равно

$$M[T_{oi}] = 1/\lambda_i.$$

Введем характеристическую переменную

$$x_i = \begin{cases} 1, & \text{если } M[T_{oi}] < T_p; \\ 0, & \text{в противном случае.} \end{cases}$$

Математическое ожидание  $x_i$  равно

$$x_i^* = P_{i(>1)}.$$

Введем обобщенный показатель – приведенный уровень обновления записи, который отражает уровень согласованности периодов обновления полей записи

$$X_Z = \frac{\sum_{i=1}^N x_i^*}{N}. \quad (3)$$

Обобщенный показатель  $X_Z$  лежит в пределах  $0 - 1$ , где единице соответствует полная согласованность периодов обновления, т.е., к моменту времени решения задачи все поля обновлены.

Рассмотрим случай, когда  $T_p$  являются по своей природе случайными (рис. 2). Выше было сделано допущение, что интервалы обновления полей записи есть случайные и образуют простейший поток – как вид потока Пальма [5]. Тогда, интервалы обновления поля записи  $T_{oi}^*$ , на которые попадает случайная точка  $t_p$  (момент решения задачи), являются случайными и имеют математическое ожидание [5]

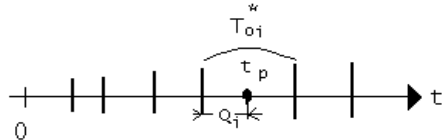


Рис. 2. Шкала времени,  $T_p$  случайны

$$M[T_{oi}^*] = M[T_{oi}] + \frac{D[T_{oi}]}{M[T_{oi}]}, \quad (4)$$

где  $M[T_{oi}] = 1/\lambda_i$ ;  $D[T_{oi}] = 1/\lambda^2$ .

В свою очередь математическое ожидание интервалов  $Q_i$  выражается как

$$M[Q_i] = M[T_{oi}^*]/2.$$

Заменим в приведенных выше выражениях (1), (3) фиксированные интервалы между решениями задачи  $T_p$  на  $M[Q_i]$ . Такая замена позволяет получить введенные обобщенные показатели в конъюнктивной (2) и дизъюнктивной (4) формах для варианта решения задач в случайные моменты времени.

Введем в рассмотрение индекс рассогласования для двух записей БД как

$$\Delta x_{ij} = |X_{zi} - X_{zj}|.$$

Составим матрицу рассогласования для БД.

$$S = \begin{pmatrix} 0 & \Delta x_{12} & \dots & \Delta x_{1M} \\ \Delta x_{21} & 0 & \dots & \Delta x_{2M} \\ \dots & \dots & \dots & \dots \\ \Delta x_{M1} & \Delta x_{M2} & \dots & 0 \end{pmatrix},$$

где  $M$  – число записей БД.

Запрос на использование информации из БД представим в виде вектора

$$R^T = \{a_1, a_2, \dots, a_i, \dots, a_M\}, \quad (5)$$

где  $a_i = \begin{cases} 1, & \text{если используется } i\text{-я запись;} \\ 0, & \text{в противном случае.} \end{cases}$

Рассмотрим произведение  $S \times R = Z$ . Полученный вектор  $Z$  характеризует уровень рассогласования по времени обновления информации для запроса, представленного в виде вектора  $R$ .

Для практического использования уровень рассогласования будем оценивать модулем вектора  $Z$

$$|Z| = \sqrt{Z_1^2 + Z_2^2 + \dots + Z_M^2}. \quad (6)$$

Для получения интегральной оценки уровня рассогласования БД можно взять среднее значение  $|Z|$  для всех возможных комбинаций вектора  $R$ . Однако, для реальных БД с большим количеством записей такое решение в виде прямого перебора всех комбинаций вектора  $R$  вряд ли уместно. Используем метод имитационного моделирования. Будем формировать вектор  $R$  случайным образом. Представим  $a_j$  (5) как случайную величину с равномерным законом распределения. После  $M$  опытов методом жребия сформируем вектор  $R'_j$ . Затем найдем  $Z'_j = S \times R'_j$ , после чего определим  $|Z'_j|$ . Повторим эту процедуру  $K$  раз и представим уровень рассогласования БД как оценку математического ожидания модуля вектора  $|Z|$

$$M^* [|Z|] = \frac{\sum_{j=1}^K Z'_j}{K}. \quad (7)$$

**Выводы.** Введенные вероятностные показатели качества информации (2), (4), (6), (7) позволяют количественно оценить состояние БД АИС, прогнозировать возможные риски при проведении расчетов, принятии решений с использованием автоматизированных информационных систем.

## ЛИТЕРАТУРА

1. Ларичев О.И. Теория и методы принятия решений. – М.: Логос, 2002. – 392 с.
2. Королев А.И., Кучук Г.А., Пашинев А.А. Адаптивная маршрутизация в корпоративных сетях. – Х.: ХВУ, 2003. – 224 с.
3. Конолли Т., Бегг К., Страчан А. Базы данных: проектирование, реализация и сопровождение. Теория и практика. – М.: Вильямс, 2000. – 1120 с.
4. Системный анализ в управлении: Учеб. пособие / В.С. Анфилатов, А.А. Емельянов, А.А. Кукушкин // Под ред. А.А. Емельянова. – М.: Финансы и статистика, 2002. – 368 с.
5. Вентцель Е.С., Овчаров Л.А. Теория случайных процессов и ее инженерные приложения. – М.: Наука, 1991. – 384 с.

Поступила 5.01.2006

**Рецензент:** доктор технических наук, профессор В.А. Краснобаев,  
Харьковский национальный технический университет сельского хозяйства.