П. 4 Комплекс моделей процессов сбора объектов от источников (информации, составных элементов и др.)

Модели сбора каких-либо объектов от источников (например, информации, составных элементов при комплексировании системы и др.) поясним на примере информационной системы (ИС). То есть в качестве собираемых объектов выступает информация.

Должностные лица ИС используют в своей деятельности «анонимные» выходные документы, получаемые ими в результате обработки их запросов на ЭВМ. Было бы идеальным, если бы выходной документ сопровождался датой, на которую содержащиеся в нем значения элементов информации являлись бы по умолчанию актуальными. Это было бы полностью аналогично продаваемым продуктам питания – «срок годности указан на упаковке». Однако в реальности это оказывается далеко не так. На практике реальные изменения состояния объектов учета отображаются в ИС с некоторой задержкой, причем для различных данных эти задержки различны. Более того, одна и та же информация может быть актуальной для решения одной задачи и оказаться полностью неактуальной при выполнении другой. Иными словами, в выходном документе под одной датой его формирования могут отражаться данные различной степени актуальности. Таким образом, для решения задач используется выходная информация, в общем случае отличная от реальной в силу ее изменения во времени. Если это отличие существенно, то использование такой информации может привести к ошибкам.

Под актуальностью на момент t выходной информации, отражающей с допустимой погрешностью $\Delta(t-\tau)$ состояние объектов учета и явлений на момент $t-\tau$, понимается ее свойство отражать с заданной погрешностью $\Delta(t)$ реальное состояние на момент t. Так, если через $x(t)$ обозначить реальное значение какого-либо количественного параметра в момент t, а через $\tilde{x}(t)$ – его значение, отраженное в ИС в момент τ $(0\le\tau<t)$, то значение $\tilde{x}(t)$ будет актуальным на момент t, если для наперед заданных границ Δ^- и Δ^+

$$\Delta^- \le \tilde{x}(t) - x(t-\tau) \le \Delta^+,$$

где Δ^- и Δ^+ – соответственно нижняя и верхняя границы допустимых переделов, задаваемых пользователями информации из соображений обоснованного решения конкретной функциональной задачи, причем в общем случае $\Delta^- = \Delta^-(x,t,\tau)$, $\Delta^+ = \Delta^+(x,t,\tau)$. Если $\Delta^- = \Delta^+ = 0$, то это означает, что должностным лицам требуется для решения задач 100%-но точная информация о состоянии объектов и явлений.

Введение границ допустимых пределов изменения значений используемой информации тесно связано с понятием значимого для пользователя ИС изменения реальной характеристики объекта учета или явления. Изменение характеристики объекта учета будем называть значимым для решения конкретной задачи, если своевременное представление безошибочной информации об этом изменении пользователю ИС изменяет логику или результат решения.

На актуальность информации в ИС оказывают влияние не только технические характеристики используемых компьютерных средств, но и применяемые технологии сбора данных в системе. Следовательно, при обосновании технических решений по поддержанию используемой информации в актуальном состоянии неизбежно возникает вопрос количественной оценки достигаемой актуальности для различных технологий сбора и доведения данных до пользователей.

Ниже предлагаются математические модели для оценки актуальности используемой информации в зависимости от:

технологии сбора данных в ИС и ее параметров;

законов объективного значимого изменения состояний объектов учета;

затрат времени с момента начала подготовки документов у конечных источников до ввода соответствующей информации в ИС.

Суть формализации отражена на рис.П.4.

Для оценки вероятности сохранения актуальности используемой информации предлагается следующее утверждение П.4.

Утверждение П.4 [1]. При условии существования стационарных распределений времени подготовки, передачи и ввода в ИС входной информации, времени между значимыми изменениями состояний объектов учета и интервалов времени между соседними обновлениями данных в ИС и их независимости предельная вероятность сохранения актуальности информации на момент ее использования существует и равна:

а) для дисциплины D_1 выдачи информации от источника сразу по происшествии значимого изменения текущего состояния объектов учета:

$$P_{акт} = \frac{1}{\xi_i}\int_0^\infty B(t)[1-C(t)]dt, \qquad (П.4.1)$$

б) для дисциплины D_2 обновления данных в ИС вне зависимости от наличия или отсутствия изменения текущего состояния объектов учета (в т.ч., при сборе по регламенту):

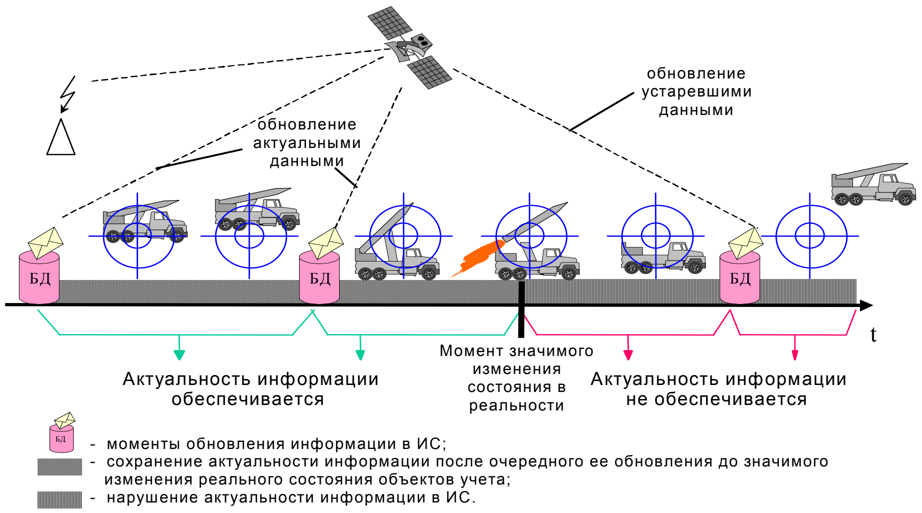

Рис. П.4. Иллюстрация случаев сохранения и утраты актуальности

$$P_{акт} = \frac{1}{q_i}\int_0^\infty \left\{ [1-Q(t)][1-\int_0^\infty C(t+\tau)dB(\tau)] \right\} dt, \qquad (П.4.2)$$

где $C(t)$ – ФР времени значимого изменения состояний объекта учета относительно информации, хранимой в ИС, ξ_i – среднее;

$B(t)$ – ФР времени подготовки, передачи и ввода данных в ИС;

$$B(t)=B_1 * B_2 * B_3(t),\ B_1(t)=1-exp(-t/\omega_{i1}),\ B_2(t)= 1-exp(-t/\delta_{i1}),\ B_3(t)= 1-exp(-t/\beta_{i1});$$

$Q(t)$ – ФР длительности интервалов между соседними обновлениями данных в ИС, q_i –среднее.

Доказательство. Доказательство утверждения проведено для одного отдельно взятого элемента информации. Рассмотрим случайный процесс $\eta_{акт.}(t)$, характеризующий состояние используемых данных с точки зрения их актуальности для решения конкретной задачи. Величина $\eta_{акт.}(t)$ может принимать одно из двух значений, характеризующих пространство элементарных событий:

$$\eta_{акт.}(t) = \begin{cases} \text{«Информация актуальна для решения задачи», если значения, отраженные в ИС,} \\ \text{являются достаточно актуальными и на момент } t \text{ их использования;} \\ \\ \text{«Информация недостаточно актуальна для решения задачи» – в противном случае} \end{cases}$$

Построим случайный процесс $\eta(t)$, $0\le t<\infty$ следующим образом:

$$\eta(t) = \begin{cases} \eta_1(t) \text{ при } 0 \le t_1 < z_1, \\ \eta_2(t) \text{ при } z_1 \le t < z_1 + z_2, \\ \dots \\ \eta_k(t) \text{ при } t_{k-1} \le t < t_k, \end{cases}$$

где $t_0=0,\ t_k=z_1+z_2+\ldots+z_k,\ k\ge 1$;

z_k – случайная величина, определяющая интервал времени между (k-1) и k-м последовательными значимыми изменениями состояний объектов учета и явлений и имеющая ФР $C(t)$;

$\eta_k(t)$ – случайная функция, определенная на k-м интервале и принимающая на нем те же значения, что и $\eta_{акт.}(t)$.

Определенный таким образом процесс $\eta(t)$ является регенерирующим, моменты t_k – моментами регенерации, а пара (z_k, η_k) – циклами регенерации.

Рассмотрим функцию $\mu(t)=P\{\eta(t)=$«Информация актуальна для решения задачи», $z_k>t\}$, определяющую вероятность того, что информация на k-м цикле к моменту времени t функционирования системы не потеряет своей актуальности и очередное значимое изменение в реальности произойдет после момента t. Поскольку время подготовки, передачи и ввода данных в ИС и время значимого изменения количественных параметров объектов учета являются независимыми, то

$$\mu(t) = \int_t^\infty dC(\tau)B(t).$$

Далее введем функцию $M(t) = \int_0^t \mu(t-y)d\,C^{*n}(y),$

где $C^{*n}(y)$, $n\ge 0$ – n-я свертка ФР $C(t)$.

Функция $M(t)$ непосредственно интегрируема по Риману на $[0,\infty)$ при n=0, поскольку $M(t)=\mu(t)$. Это условие является необходимым для применения предельной теоремы для регенерирующих процессов, согласно которой для дисциплины D_1

$$P_{акт.i} = \lim_{t\to\infty} P\left\{ \begin{matrix} \eta(t) = \text{"Информация актуальна} \\ \text{для решения задачи"}, z_k > t \end{matrix} \right\} = \int_0^\infty \mu(t)dt \Big/ \int_0^\infty t dC(t).$$

Отсюда и следует справедливость утверждения для дисциплины D_1.

Для дисциплины 2 рассмотрим два случая.

Случай 1 (вспомогательный). Пусть время доведения информации от источника до ИС $\tau_{дов.}=0$, т.е. данные доводятся до ИС мгновенно. Рассмотрим случайный процесс $\eta(t)$, определенный выше, но с той разницей, что z_k – случайная величина, определяющая интервал времени между (k-1) и k-м последовательными обновлениями данных в ИС и имеющая ФР $Q(t)$;

$\eta(t)$ – случайная функция, определенная на k -м интервале и принимающая на нем те же значения, что $\eta_{акт.}(t)$.

Определенный таким образом процесс $\eta(t)$ является регенерирующим, моменты t_k – моментами регенерации, а пара (z_k, η_k) – циклами регенерации.

Рассмотрим функцию $\mu(t)=P\{\eta(t)=$«Информация актуальна для решения задачи», $z_k>t\}$, определяющую вероятность того, что информация на k-м цикле является актуальной на момент ее использования t и следующее обновление данных в ИС произойдет позднее t. Поскольку времена между соседними значимыми изменениями объекта учета и обновления данных в ИС независимы, то $\mu(t)=[1-Q(t)][1-C(t)]$. Рассмотрим функцию $M(t) = \int_0^t \mu(t-y)d\,Q^{*n}(y),$ где $Q^{*n}(t)$, $n\ge 0$ – n-я свертка ФР $Q(t)$.

$M(t)$ непосредственно интегрирует по Риману на $[0,\infty)$ при $n=0$, поскольку $M(t)=\mu(t)$. Это условие является необходимым для применения предельной теоремы для регенерирующим процессов, согласно которой

$$P_{акт.i} = \lim_{t\to\infty} P\left\{ \begin{matrix} \eta(t) = \text{"Информация актуальна} \\ \text{для решения задачи"}, z_k > t \end{matrix} \right\} = \int_0^\infty \mu(t)dt \Big/ \int_0^\infty t dQ(t).$$

Отсюда $P_{акт.i} = \int_0^\infty [1-Q(t)][1-C(t)]\,dt \Big/ q_i.$

Случай 1 назван вспомогательным, поскольку в реальности $\tau_{дов.}\ne 0$, оно всегда является положительной величиной, но результаты случая 1 позволяют сравнительно просто получить искомые результаты для общего случая 2, когда $\tau_{дов.}>0$.

Случай 2. Пусть $\tau_{дов.}>0$, т.е. задается ФР времени подготовки, передачи и ввода данных в ИС. Этот случай сводится к предыдущему путем введения новой ФР времени значимого изменения состояний объекта учета $\widetilde{C}(t) = P\{max(0,\ \tau_{изм.}-\tau_{дов})<t\}$, где $\tau_{изм.}$ распределено по закону $C(t)$, а $\tau_{дов.}$– по закону $B(t)$.

Учитывая, что распределение максимума двух независимых случайных величин равно произведению их распределений, получаем, что

$$\widetilde{C}(t) = \int_0^\infty C(\tau + t) dB(\tau).$$

Тем самым задача свелась к случаю 1 с заменой $C(t)$ на $\widetilde{C}(t)$. На этом *доказательство утверждения П.4 завершено*.

В инструментариях в приложении к информации i-го типа реализованы варианты:

а) $Q(t) = \begin{cases} 0,\ t \le q_i\,, \\ 0,\ t > q_i \end{cases}$ для строго регламентного режима сбора информации, когда длительность интервалов между соседними обновлениями данных q_i фиксирована (расчетные значения обозначаются как $P_{акт.строг.}$);

б) $Q(t) = 1 - exp(-t/q_i)$ для нефиксированной длительности интервалов между соседними обновлениями данных в ИС, равной в среднем q_i (расчетные значения обозначаются как $P_{акт.нестрог.}$);

в) $C(t)=1\text{-}exp(\text{-}t/\xi_i)$.

Явные аналитические формулы, реализованные в инструментариях, получаются на основе интегрирования обычными методами полученных выражений (П.4.1) и (П.4.2).

Необходимые для моделирования пределы исходных значений ξ задают в постановках функциональных задач, значения ω, δ, β устанавливают в результате натурных испытаний, экспериментов, дополнительного моделирования или сравнения с аналогами, дисциплину обновления информации в ИС и значения q указывают в эксплуатационной документации.