

И. М. ЛЕВКИН

МОДЕЛЬ ОБРАБОТКИ ДОКУМЕНТАЛЬНЫХ ИСТОЧНИКОВ ИНФОРМАЦИИ ДЕЛОВОЙ РАЗВЕДКИ

Предлагается модель обработки текстовых документов, позволяющая определить вероятность решения информационной задачи в информационных контурах ситуационных центров деловой разведки при ограничениях по времени, а также требуемый состав средств обработки.

Ключевые слова: деловая разведка, информационный документ, элементарный фрагмент информации, эффективность обработки информации, входной поток информационных документов.

Важнейшей особенностью деловой разведки является необходимость решения ряда информационных задач в условиях ограничения по времени. Сложность решения этих задач связана, во-первых, с необходимостью обработки большого числа источников информации в целях выявления информационных признаков, относящихся к конкретной задаче, и, во-вторых, с ограниченными возможностями сотрудников информационно-аналитической структуры деловой разведки по семантической обработке источников информации [1, 2].

В связи с этим возникает необходимость оценки вероятности решения соответствующей информационной задачи в зависимости от интенсивности входного информационного потока, состава информационного контура (числа автоматизированных рабочих мест) и требований по времени решения задачи.

В основу формирования этой оценки может быть положена операционно-временная модель процесса обработки текстовых отчетно-информационных документов.

Последовательность агрегированных действий при обработке документальных источников информации представлена на рис. 1. При этом учитывается, что информационный документ состоит их элементарных фрагментов информации (ЭФИ), каждый из которых представляет собой законченную совокупность предложений, характеризующую некую сущность — служебную информацию (название, выходные данные и т.п.), описание информационного признака, поясняющие фрагменты и т.п. [1].

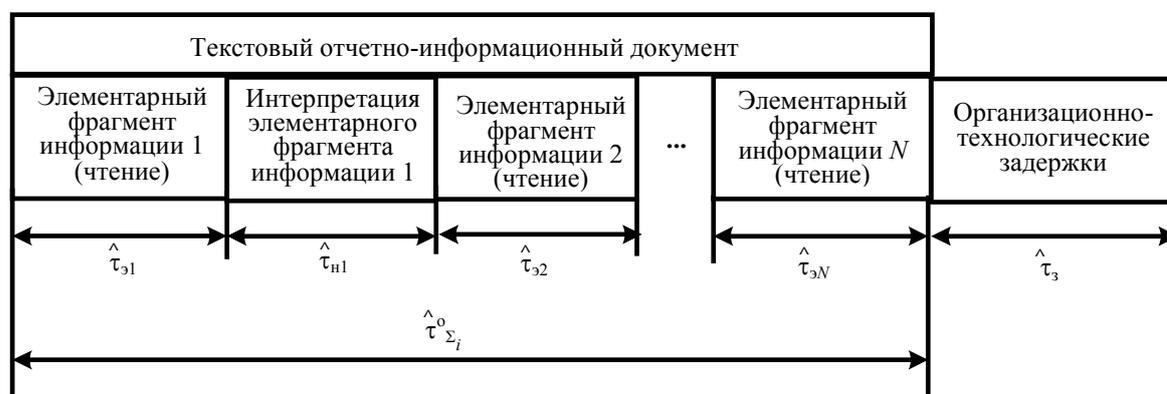


Рис. 1

Объем каждого текстового отчетно-информационного документа (число ЭФИ) зависит от таких случайных факторов, как вид и качество источника информации, число описываемых объектов деловой разведки, опыт и квалификация составителя документа и др. В связи с этим

общую продолжительность обработки i -го текстового информационного документа (операционное время) $\hat{\tau}_{\Sigma_i}^0$, $i = \overline{1, N}$, следует рассматривать как случайную величину (символ „ $\hat{\tau}$ “ — знак случайной величины).

Величина $\hat{\tau}_{\Sigma_i}^0$ (см. рис. 1) формируется из следующих составных частей:

— суммарной продолжительности выполнения операций ознакомления с элементарными фрагментами информации (чтение): $\hat{\tau}_{\Sigma_i} = \hat{\tau}_{\Sigma_{i1}} + \hat{\tau}_{\Sigma_{i2}} + \dots + \hat{\tau}_{\Sigma_{iN}}$;

— суммарной продолжительности выполнения операций интерпретации элементарного фрагмента (объекта): $\hat{\tau}_{\Sigma_i} = \hat{\tau}_{\Sigma_{i1}} + \dots + \hat{\tau}_{\Sigma_{ik}}$, где $k \in [1, N]$.

Исследования показывают, что число терминов, содержащихся в ЭФИ, также является случайной величиной \hat{n}_3 и подчиняется нормальному закону распределения с параметрами $\sigma_{\hat{n}_3}$, $m_{\hat{n}_3}$. Это позволяет предположить, что продолжительность обработки $\hat{\tau}_3$ фрагмента также подчиняется нормальному закону распределения с плотностью $\varphi_{\hat{\tau}_3}(\tau)$ и параметрами $\sigma_{\hat{\tau}_3}$, $m_{\hat{\tau}_3}$. Такой же характер носят и случайные величины $\hat{\tau}_H$ и $\hat{\tau}_3$.

При данных предположениях модель $\varphi_{\hat{\tau}_{\Sigma_i}}(\tau)$ продолжительности выполнения операций изучения всех ЭФИ i -го текстового отчетно-информационного документа будет описываться $m_{\hat{n}_3}$ -кратной композицией нормальных законов распределения $\varphi_{\hat{\tau}_3}(\tau)$ (для упрощения записи индекс „ Σ “ у символа „ Σ “ здесь и далее опущен):

$$\varphi_{\hat{\tau}_{\Sigma}}(\tau) = \underbrace{\varphi_{\hat{\tau}_3}(\tau) \otimes \dots \otimes \varphi_{\hat{\tau}_3}(\tau)}_{m_{\hat{n}_3}} = \frac{1}{\sigma_{\hat{\tau}_{\Sigma}} \sqrt{2\pi}} \exp \left\{ -(\tau - m_{\hat{\tau}_{\Sigma}})^2 / 2\sigma_{\hat{\tau}_{\Sigma}}^2 \right\} \prod (\tau; m_{\hat{\tau}_{\Sigma}} - 3\sigma_{\hat{\tau}_{\Sigma}}; m_{\hat{\tau}_{\Sigma}} + 3\sigma_{\hat{\tau}_{\Sigma}}),$$

где $\sigma_{\hat{\tau}_{\Sigma}} = \sqrt{m_{\hat{n}_3} \sigma_{\hat{\tau}_3}^2}$, $m_{\hat{\tau}_{\Sigma}} = m_{\hat{n}_3} m_{\hat{\tau}_3}$.

Используя аналогичные рассуждения для интерпретации полученной информации, по результатам обработки фрагментов i -го текстового отчетно-информационного документа $\hat{\tau}_{H\Sigma}$ можно записать:

$$\varphi_{\hat{\tau}_{H\Sigma}}(\tau) = \underbrace{\varphi_{\hat{\tau}_H}(\tau) \otimes \dots \otimes \varphi_{\hat{\tau}_H}(\tau)}_{m_{\hat{n}_H}} = \frac{1}{\sigma_{\hat{\tau}_{H\Sigma}} \sqrt{2\pi}} \exp \left\{ -(\tau - m_{\hat{\tau}_{H\Sigma}})^2 / 2\sigma_{\hat{\tau}_{H\Sigma}}^2 \right\} \prod (\tau; m_{\hat{\tau}_{H\Sigma}} - 3\sigma_{\hat{\tau}_{H\Sigma}}; m_{\hat{\tau}_{H\Sigma}} + 3\sigma_{\hat{\tau}_{H\Sigma}}),$$

где $\sigma_{\hat{\tau}_{H\Sigma}} = \sqrt{m_{\hat{n}_H} \sigma_{\hat{\tau}_H}^2}$, $m_{\hat{\tau}_{H\Sigma}} = m_{\hat{n}_H} m_{\hat{\tau}_H}$, $\hat{n}_H < \hat{n}_3$.

Следовательно, модель операционного времени $\hat{\tau}_{\Sigma}^0$, необходимого для достижения целевого эффекта — обработки текстового отчетно-информационного документа, определяется композицией законов распределения $\varphi_{\hat{\tau}_{\Sigma}}(\tau)$ и $\varphi_{\hat{\tau}_{H\Sigma}}(\tau)$, а учет продолжительности организационно-технологической задержки $\hat{\tau}_3$ предполагает дополнительную свертку с моделью $\varphi_{\hat{\tau}_3}(\tau)$ [3]. В случае представления модели организационно-технологической задержки в виде нормального закона распределения модель общих временных затрат на обработку i -го текстового документа может быть представлена следующим образом (для упрощения записи индекс „ $\hat{\tau}_{\Sigma_i}^0$ “ обозначим как q):

$$\varphi_q(\tau) = \varphi_{\hat{\tau}_{\Sigma}}(\tau) \otimes \varphi_{\hat{\tau}_{\text{НС}}}(\tau) \otimes \varphi_{\hat{\tau}_3}(\tau) = \frac{1}{\sigma_q \sqrt{2\pi}} \exp\left\{-\frac{(\tau - m_q)^2}{2\sigma_q^2}\right\} \prod(\tau; m_q - 3\sigma_q, m_q + 3\sigma_q),$$

где $\sigma_q = \sqrt{\sigma_{\hat{\tau}_{\Sigma}}^2 + \sigma_{\hat{\tau}_{\text{НС}}}^2 + \sigma_{\hat{\tau}_3}^2}$, $m_q = m_{\hat{\tau}_{\Sigma}} + m_{\hat{\tau}_{\text{НС}}} + m_{\hat{\tau}_3}$.

В связи с тем, что достижение целевого эффекта обеспечивается интеллектуальным потенциалом сотрудника информационного центра деловой разведки, в качестве функции связности случайных величин результативности \hat{v}_i и оперативности $\hat{\tau}_i$ обработки следует выбрать нелинейную функцию вида [1]

$$\hat{v}_i = 1 - \exp\left\{-(\lambda_i (\hat{\tau} - \hat{\tau}_i - \hat{\tau}_3))\right\},$$

где λ_i — производительность (квалификация) сотрудника, $\hat{\tau}_i = \hat{\tau}_{\Sigma} + \hat{\tau}_{\text{НС}}$.

Совместная плотность вероятности функционально связанных случайных величин \hat{v}_i и $\hat{\tau}_i$ определяется как

$$\begin{aligned} \varphi_{\langle \hat{v}_i, \hat{\tau}_i \rangle}(v, \tau) &= \varphi_{\hat{v}_i}(\tau) \varphi_{\hat{\tau}_i / \hat{v}_i}(\tau; v_i); \\ \varphi_{\hat{v}_i}(v) &= \frac{1 - v_i}{\lambda_i \sigma_{\hat{\tau}_i} \sqrt{2\pi}} \exp\left\{-\frac{\left(\hat{\lambda}_i^{-1} \ln(1 - v_i)^{-1} - m_{\hat{\tau}_i}\right)^2}{2\sigma_{\hat{\tau}_i}^2}\right\}; \\ \varphi_{\hat{\tau}_i / \hat{v}_i}(\tau; v) &= \frac{1}{\hat{\sigma}_{\hat{\tau}_3} \sqrt{2\pi}} \exp\left\{-\frac{\left(\tau - (m_{\hat{\tau}_i} - 3\sigma_{\hat{\tau}_i}) - \lambda_i^{-1} \ln(1 - v_i)^{-1} - m_{\hat{\tau}_3}\right)^2}{2\sigma_{\hat{\tau}_3}^2}\right\}, \end{aligned}$$

а эффективность процесса обработки отчетно-информационного документа, определяемая вероятностью достижения цели $P_{\text{д.ц}}$, вычисляется по формуле

$$P_{\text{д.ц}} = P_{\text{д.ц}} \left[(\hat{v}_i \geq v^*) \cap (\hat{\tau}_i \leq \tau^*) \right] = \int_0^v \int_0^{\tau} \varphi_{\hat{v}_i}(v) \varphi_{\hat{\tau}_i / \hat{v}_i}(\tau; v) dv d\tau,$$

где v^* — требуемое значение результативности обработки, τ^* — требуемое значение ее оперативности.

График, характеризующий зависимость эффективности процесса обработки текстового отчетно-информационного документа от ее результативности \hat{v}_i и оперативности $\hat{\tau}_i$, представлен на рис. 2.

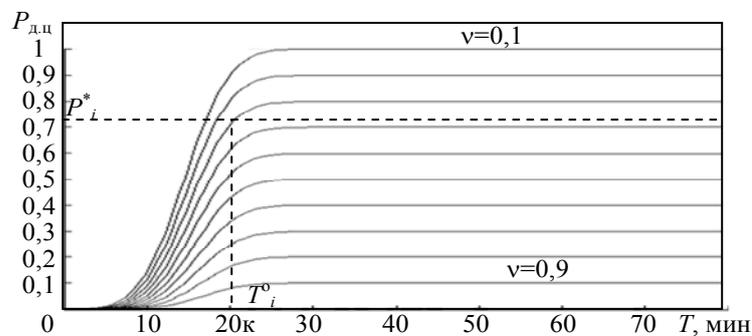


Рис. 2

Данные зависимости позволяют:

— определить время обработки рассматриваемого источника информации: $T_i^0 = \arg(P_{\text{д.ц}}(t) | P_{\text{д.ц}} = P_i^*)$, где P_i^* — требуемая вероятность решения информационной задачи;

— определить интенсивность обработки информации сотрудником деловой разведки на одном автоматизированном рабочем месте: $\mu_j = 1/\bar{T}^0$, где \bar{T}^0 — средняя продолжительность обработки информационного документа на j -м рабочем месте;

— определить требуемое число автоматизированных рабочих мест J в информационной структуре ситуационного центра деловой разведки, обеспечивающее непрерывную обработку потока информации, поступающего с интенсивностью $\omega_{\text{вх}}$, исходя из условия [4]

$$\omega_{\text{вх}} \leq \sum_{j=1}^J \mu_j.$$

СПИСОК ЛИТЕРАТУРЫ

1. Левкин И. М. Теория и практика информационно-аналитической работы. Курск: НАУКОМ, 2011. 389 с.
2. Доронин А. И. Бизнес-разведка. М.: Ось-89, 2006. 496 с.
3. Петухов Г. Б., Якунин В. И. Методологические основы внешнего проектирования целенаправленных процессов и целеустремленных систем. М.: АСТ, 2006.
4. Двухуровневая модель информационного взаимодействия / Б. Я. Советов, М. О. Колбанёв, Т. М. Татарникова // Материалы VIII Санкт-Петербург. межрегион. конф. „Информационная безопасность регионов России“, 23—25 окт. 2013 г. СПб: СПОЙСУ, 2013. С. 184—185.

Сведения об авторе

Игорь Михайлович Левкин — д-р воен. наук, профессор; Санкт-Петербургский государственный экономический университет, кафедра прикладных информационных технологий;
E-mail: lev.kin@yandex.ru

Рекомендована кафедрой
прикладных информационных
технологий

Поступила в редакцию
28.04.14 г.