

**О СТАТИСТИЧЕСКОМ ОЦЕНИВАНИИ ПАРАМЕТРОВ
РЕГРЕССИИ ПРИ НАЛИЧИИ СЛУЧАЙНОГО
ЦЕНЗУРИРОВАНИЯ**

Е. С. Агеева

МОДЕЛЬ

Рассмотрим модель множественной линейной регрессии, заданную уравнением [1]:

$$Y_t = \sum_{i=1}^M \theta_i X_t^i + \theta_0 + \xi_t, t = 1, \dots, n, \quad (1)$$

где $\{\xi_t\}$ – независимые в совокупности нормальные одинаково распределённые случайные величины с математическим ожиданием 0 и дисперсией $0 < \sigma^2 < \infty$; $\{\theta_0, \theta_1, \dots, \theta_M\}^T$ называются коэффициентами регрессии, $\{X_t^i\}$, $i = 1, \dots, M$ – регрессорами, $\{\xi_t\}$ – случайными величинами ошибок. Будем предполагать σ^2 известным. Нами будут наблюдаться значения регрессоров $\{X_t^i\}$, $i = 1, \dots, M$, $t = 1, \dots, n$ и события $Y_1 \in [a_1, b_1], \dots, Y_n \in [a_n, b_n]$ вместо точных значений Y_1, \dots, Y_n ; здесь $[a_t, b_t]$, $t = 1, \dots, n$ – интервал цензурирования, $a_t \leq b_t$.

Требуется построить оценки максимального правдоподобия для вектора регрессионных коэффициентов $\theta = (\theta_0, \theta_1, \dots, \theta_M)^T$ и получить асимптотическое выражение этой оценки в асимптотике $\Delta_t = b_t - a_t \rightarrow 0$, используя разложение в ряд по Δ_t . Для найденных оценок требуется получить выражения для смещения и матрицы вариаций.

ОЦЕНКИ И ИХ СВОЙСТВА

Пусть матрица $X = \begin{pmatrix} 1 & X_1^1 & \dots & X_1^M \\ 1 & X_2^1 & \dots & X_2^M \\ \dots & \dots & \dots & \dots \\ 1 & X_n^1 & \dots & X_n^M \end{pmatrix} = (X_{t,j})_{\substack{t=1, \dots, n \\ j=1, \dots, M+1}}$ – это

$n \times (M + 1)$ - матрица эксперимента, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$ – функция Лапласа.

Теорема 1. В случае множественной линейной регрессии (1) с интервальным цензурированием оценка максимального правдоподобия $\hat{\theta} = (\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_M)^T$ является решением системы $M + 1$ нелинейных уравнений:

$$\sum_{t=1}^n X_{t,j} \frac{\partial P_t}{\partial \theta_0} / P_t = 0, \quad j = 1, \dots, M + 1 \quad (2)$$

при условии, что $(M + 1) \times (M + 1)$ -матрица $A(\hat{\theta})$ отрицательно определена: $A(\hat{\theta}) \prec 0$. Здесь

$$P_t = \Phi((b_t - \theta_0 - \theta_1 X_t^1 - \dots - \theta_M X_t^M) / \sigma) - \Phi((a_t - \theta_0 - \theta_1 X_t^1 - \dots - \theta_M X_t^M) / \sigma),$$

$$A(\hat{\theta}) = \left(\sum_{t=1}^n X_{t,i} X_{t,j} \left(\frac{1}{\sigma} \frac{\partial^2 P_t}{\partial \theta_0^2} P_t - \left(\frac{\partial P_t}{\partial \theta_0} \right)^2 \right) \right) / P_t^2 \Big|_{i,j=1}^{M+1}.$$

Оценка максимального правдоподобия находится из условия [1]:

$$l(\theta) = l(\theta_0, \theta_1, \dots, \theta_M) = \sum_{t=1}^n \ln P(Y_t \in [a_t, b_t]) = \sum_{t=1}^n \ln P_t \xrightarrow{\theta_0, \theta_1, \dots, \theta_M} \max. \quad (3)$$

Раскладывая функцию $l(\theta)$ по степеням $\Delta_t = b_t - a_t$, получим 3 асимптотических выражения для оценки максимального правдоподобия, имеющих вид:

$$\hat{\theta} = C^{-1} F. \quad (4)$$

Здесь

$$C_{ij} = \frac{1}{n} \sum_{t=1}^n B_t X_{t,i} X_{t,j}, \quad F_i = \frac{1}{n} \sum_{t=1}^n B_t X_{t,i} \frac{a_t + b_t}{2}, \quad i, j = 1, \dots, M + 1, \quad (5)$$

где в случае $B_t = 1, t = 1, \dots, n$ порядок аппроксимации равен 1, а в случае

$$B_t = 1 - \frac{\Delta_t^2}{12\sigma^2} \quad \text{или} \quad B_t = \beta_t = \frac{\Delta_t e^{-\frac{\Delta_t^2}{8\sigma^2}}}{12\sqrt{2\pi}\sigma^2 (\Phi(\frac{\Delta_t}{2\sigma}) - \Phi(-\frac{\Delta_t}{2\sigma}))}, \quad t = 1, \dots, n$$

порядок аппроксимации равен 3.

Пусть рассматривается множество всевозможных интервалов на числовой прямой, имеющих фиксированную длину $\Delta > 0$:

$$S_\Delta \{(a, b) : a, b \in R, b - a = \Delta\} = \{(a, a + \Delta) : a \in R\} \subset B(R).$$

Пусть далее $F(x), x \in R$, – некоторая абсолютно непрерывная функция распределения вероятностей, задающая на $B(R)$ вероятностную меру $P\{\cdot\}$, абсолютно непрерывную относительно меры Лебега:

$$P\{(a, b)\} = P\{(a, a + \Delta)\} = F(a + \Delta) - F(a) \geq 0, a \in R. \quad (6)$$

Случайный интервал $(a, a + \Delta) \in S_\Delta$ полностью определяется заданием его левой границы a . Её плотность распределения вероятностей:

$$p_a(x; \Delta) = \frac{F(x + \Delta) - F(x)}{\int_{-\infty}^{\infty} (F(y + \Delta) - F(y)) dy} \geq 0. \quad (7)$$

Лемма 1. Пусть $F(x), x \in R$ – произвольная симметричная функция распределения вероятностей, такая что $\lim_{x \rightarrow +\infty} xF(-x) = 0$. Тогда функция

$$p(x; \Delta) = \frac{F(x + \Delta) - F(x)}{\Delta}, x \in R$$

является плотностью распределения вероятностей левой границы a случайного интервала $(a, a + \Delta) \in S_\Delta$. В частности, если $F(x) = \Phi(x), x \in R$ – функция распределения вероятностей стандартной нормальной случайной величины, то

$$p(x; \Delta) = \frac{\Phi(x + \Delta) - \Phi(x)}{\Delta}, x \in R.$$

По лемме 1 плотность распределения вероятностей нижней границы a_t равна

$$p_{a_t}(x; \Delta) = \frac{\Phi\left(\frac{x + \Delta_t - \theta_0 - \theta_1 X_t^1 - \dots - \theta_M X_t^M}{\sigma}\right) - \Phi\left(\frac{x - \theta_0 - \theta_1 X_t^1 - \dots - \theta_M X_t^M}{\sigma}\right)}{\Delta_t}. \quad (8)$$

Лемма 2. Если случайная величина a_t имеет плотность распределения вероятностей, заданную формулой (8), то её математическое ожидание и второй момент имеют следующий вид:

$$E\{a_t\} = \theta_0 + \theta_1 X_t^1 + \dots + \theta_M X_t^M - \frac{\Delta_t}{2},$$

$$E\{a_t^2\} = \sigma^2 + (\theta_0 + \theta_1 X_t^1 + \dots + \theta_M X_t^M)^2 - (\theta_0 + \theta_1 X_t^1 + \dots + \theta_M X_t^M) \Delta_t + \frac{\Delta_t^2}{3}.$$

Теорема 2. Если матрица C , определённая в (5), невырожденная, то оценки (4) являются несмещёнными, а матрица вариаций равна:

$$V(\hat{\theta}) = C^{-1} B C^{-1}, \quad (9)$$

где элементы матрицы B имеют вид:

$$B_{ij} = \frac{1}{n^2} \sum_{t=1}^n X_{t,i} X_{t,j} B_t^2 \left(\sigma^2 + \frac{\Delta_t^2}{12} \right), \quad i, j = 1, \dots, M+1. \quad (10)$$

Теорема 3. Пусть $|B_t| \leq D_1$, $\sigma^2 + \frac{\Delta_t^2}{12} \leq D_2$, $t = 1, \dots, n$, а ряды $\frac{1}{n} \sum_{t=1}^n |X_t^i X_t^j|$, $\frac{1}{n} \sum_{t=1}^n |X_t^i|$, $i, j = 1, \dots, M$ ограничены сверху константой D_3 , сразу для всех n , $i, j = 1, \dots, M$. Тогда элементы матрицы вариаций $V(\hat{\theta})$ будут стремиться к 0 при $n \rightarrow \infty$.

Следствие 1. В условиях теоремы 3 оценки (4) являются состоятельными по вероятности.

Теорема 4. Пусть $D = (\delta_{ij} \frac{\Delta_i^2}{12})_{i,j=1}^n$, $Y = (\frac{a_1 + b_1}{2}, \dots, \frac{a_n + b_n}{2})^T$, а матрица C , определённая в (5), невырожденная. Тогда несмещённой, состоятельной по вероятности оценкой дисперсии ошибок σ^2 будет статистика

$$\hat{\sigma}^2 = \frac{1}{n - M - 1} \left((Y - X\hat{\theta})'(Y - X\hat{\theta}) - \text{tr}((I_n - X(X'X)^{-1}X')D) \right) \quad (11)$$

ЗАКЛЮЧЕНИЕ

В работе рассмотрена модель множественной линейной регрессии при наличии случайного цензурирования. Получены три асимптотических выражения ОМП в асимптотике $\Delta_t \rightarrow 0$. Для нижней границы интервала цензурирования a_t , $t = 1, \dots, n$, найдена плотность распределения вероятностей $p_{a_t}(x)$. Доказана несмещённость и асимптотическая состоятельность предложенных оценок, а так же найдены матрицы вариаций для каждой из них. Для дисперсии ошибок σ^2 предложена несмещённая оценка.

Литература

1. Харин Ю. С., Жук Е. Е. Математическая и прикладная статистика: учеб. пособие. Мн., БГУ, 2005.
2. Gang Li, Cun-Hui Zhang. Linear regression with interval censored data // The Annals of Statistics. 1998. Vol. 26. N. 4. PP. 1306–1327.
3. Gomez G., Espinal A., Lagakos W. Inference for a linear model with an interval-censored covariate // Statistics in medicine. 2003. Vol. 22. P. 409–425.
4. Koul H., Susarla V., Ryzin V.J. Regression analysis with randomly right-censored data // The Annals of Statistics. 1981. Vol. 9. N. 6. P. 1276–1288.
5. Zhou L. A simple censored median regression estimator // Statistica Sinica. 2006. Vol. 16. P.1043–1058.

ОБ ОЦЕНКАХ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ ПАРАМЕТРОВ $AR(1)$ – ВРЕМЕННЫХ РЯДОВ ПРИ НАЛИЧИИ ЦЕНЗУРИРОВАНИЯ

И. А. Бодягин

1. ВВЕДЕНИЕ

Модель авторегрессии широко используется для описания случайных процессов во многих приложениях: в медицине, экономике, метеорологии, технике, астрономии [1; 2]. На практике значения временного ряда часто наблюдаются с искажениями, при наличии которых классические методы статистического оценивания модельных параметров дают смещенные и несостоятельные оценки. В связи с этим возникает необходимость построения новых методов, устойчивых к конкретным видам искажений [3].

В настоящей работе рассматривается такой вид искажения, как цензурирование временного ряда [4]. Оно заключается в том, что только часть наблюдений известна точно, а об остальных наблюдениях известно лишь, что они принадлежат некоторым числовым интервалам. Такая ситуация может возникать из-за наличия у приборов пределов измерения, разладки оборудования и других причин [4].

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Пусть временной ряд x_t описывается моделью $AR(1)$ авторегрессии первого порядка [1]:

$$x_t = \theta x_{t-1} + u_t, t \in Z, \quad (1)$$

где θ – коэффициент авторегрессии ($|\theta| < 1$), $\{u_t\}$ – независимые в совокупности одинаково распределенные случайные величины, имеющие нормальный закон распределения вероятностей: $L\{u_t\} = N(0, \sigma^2)$.

Пусть вместо значений временного ряда наблюдаются случайные события:

$$A_t^* = \{x_t \in A_t\}, t \in \{1, \dots, T\}, \quad (2)$$

где $\{A_t\}$ – заданные борелевские множества, T – длительность наблюдения. В данной работе рассматриваются два случая: 1) A_t состоит из одного элемента ($A_t = \{x_t\}$), тогда значение x_t известно точно; 2) $A_t = [c_t, +\infty)$ – числовой интервал, где c_t – известный уровень цензурирования, тогда имеет место цензурирование справа значения x_t .

Цензурированный справа временной ряд можно представить как последовательность полностью наблюдаемых и полностью цензурированных фрагментов. Пусть τ_i – продолжительность i -го цензурированного фрагмента ряда ($1 \leq i \leq M$), а t_i^* – начальный момент времени данного фрагмента; M – число цензурированных фрагментов.

Рассмотрим задачу построения робастных оценок для параметров модели (1) по наблюдаемым случайным событиям (2).

3. ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Введем обозначения:

$$\varphi(x) = (1/\sqrt{2\pi}) \exp(-x^2/2) -$$

плотность распределения вероятностей стандартного нормального закона $N(0,1)$;

$$I(v, y, m, s) = \int_y^{+\infty} (t-m)^v \frac{1}{s} \varphi\left(\frac{t-m}{s}\right) dt, \quad v \in \mathbb{N} \cup \{0\}, \quad y, m, s \in \mathbb{R};$$

$$v(l; k, i_1, \dots, i_k) = \sum_{j=1}^k (\delta(l, i_j) + \delta(l, i_j + 1)), \quad l, k, i_1, \dots, i_k \in \mathbb{N}; \quad \delta(i, j) = \begin{cases} 1, & i = j; \\ 0, & i \neq j; \end{cases}$$

$$d(l; A, \theta, \sigma) = \begin{cases} \sigma, & l \in A; \\ \sigma/\sqrt{1+\theta^2}, & l \notin A, \end{cases} \quad l \in \mathbb{N}, \quad A \subset \mathbb{N}, \quad \theta, \sigma \in \mathbb{R};$$

$$\bar{\mu}_l(t^*, \tau; \theta, \sigma) = \theta^l x_{t^*-1} + \theta^{\tau-1} \frac{1-\theta^{2l}}{1-\theta^{2\tau+2}} (x_{t^*+\tau} - \theta^{\tau+1} x_{t^*-1});$$

$$F_1(t^*, \tau; \theta, \sigma) = \frac{1}{\sigma} \varphi\left((x_{t^*+\tau} - \theta^{\tau+1} x_{t^*-1}) / \sigma \sqrt{\frac{1-\theta^{2\tau+2}}{1-\theta^2}} \right) (1+\theta^2)^{-\tau/2} \times$$

$$\times \left(\sum_{k=1}^{+\infty} \frac{\theta^k}{k!} \sum_{i_1, \dots, i_k=1}^{\tau-1} \prod_{l=1}^{\tau} I(v(l; k, i_1, \dots, i_k), c_{t^*+l-1}, \bar{\mu}_l(t^*, \tau; \theta, \sigma), d(l; \emptyset, \theta, \sigma)) + \right.$$

$$\left. + \prod_{l=1}^{\tau} I(0, c_{t^*+l-1}, \bar{\mu}_l(t^*, \tau; \theta, \sigma), d(l; \emptyset, \theta, \sigma)) \right), \quad (3)$$

$$F_2(t^*, \tau; \theta, \sigma) = \frac{\sqrt{1-\theta^2}}{\sigma} \varphi\left(x_{t^*+\tau} \sqrt{1-\theta^2} / \sigma \right) (1+\theta^2)^{(1-\tau)/2} \times$$

$$\times \left(\sum_{k=1}^{+\infty} \frac{\theta^k}{k!} \sum_{i_1, \dots, i_k=1}^{\tau-1} \prod_{l=1}^{\tau} I(v(l; k, i_1, \dots, i_k), c_{i^*+l-1}, \theta^{\tau-l+1} x_{i^*+\tau}, d(l; \{1\}, \theta, \sigma)) + \right. \\ \left. + \prod_{l=1}^{\tau} I(0, c_{i^*+l-1}, \theta^{\tau-l+1} x_{i^*+\tau}, d(l; \{1\}, \theta, \sigma)) \right); \quad (4)$$

$$F_3(t^*, \tau; \theta, \sigma) = (1 + \theta^2)^{(1-\tau)/2} \left(\prod_{l=1}^{\tau} I(0, c_{i^*+l-1}, \theta^l x_{i^*-1}, d(l; \{\tau\}, \theta, \sigma)) + \right. \\ \left. + \sum_{k=1}^{+\infty} \frac{\theta^k}{k!} \sum_{i_1, \dots, i_k=1}^{\tau-1} \prod_{l=1}^{\tau} I(v(l; k, i_1, \dots, i_k), c_{i^*+l-1}, \theta^l x_{i^*-1}, d(l; \{\tau\}, \theta, \sigma)) \right); \quad (5)$$

$$l^*(t_1, t_2; \theta, \sigma) = \sum_{t=t_1+1}^{t_2} \ln \frac{1}{\sigma} \varphi\left(\frac{x_t - \theta x_{t-1}}{\sigma}\right), 0 < t_1 < t_2 \leq T.$$

Теорема. Если для модели (1), (2) число фрагментов цензурированных данных $M > 1$, то логарифмическая функция правдоподобия равна:

$$l(\theta, \sigma) = \delta(t_1^*, 1) \ln F_2(t_1^*, \tau_1; \theta, \sigma) + (1 - \delta(t_1^*, 1)) \left(\ln \varphi(x_1 \sqrt{1 - \theta^2} / \sigma) + \right. \\ \left. + \ln(\sqrt{1 - \theta^2} / \sigma) + l^*(1, t_1^* - 1; \theta, \sigma) + \ln F_1(t_1^*, \tau_1; \theta, \sigma) \right) + \\ \left. + \sum_{i=2}^{M-1} \left(l^*(t_{i-1}^* + \tau_{i-1}, t_i^* - 1; \theta, \sigma) + \ln F_1(t_i^*, \tau_i; \theta, \sigma) \right) + \right. \\ \left. + l^*(t_{M-1}^* + \tau_{M-1}, t_M^* - 1; \theta, \sigma) + \delta(t_M^* + \tau_M, T) \ln F_3(t_M^*, \tau_M; \theta, \sigma) + \right. \\ \left. + (1 - \delta(t_M^* + \tau_M, T)) \left(\ln F_1(t_M^*, \tau_M; \theta, \sigma) + l^*(t_M^* + \tau_M, T; \theta, \sigma) \right) \right). \quad (6)$$

Оценки максимального правдоподобия находятся при решении экстремальной задачи: $(\hat{\theta}, \hat{\sigma}) = \arg \max_{\theta, \sigma} l(\theta, \sigma)$.

4. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Для оценки модельных параметров, часто на практике цензурированные значения временного ряда заменяют значением уровня цензурирования C_t , а затем к новому временному ряду применяют классические методы оценивания параметров, например метод наименьших квадратов (см., например, [4]).

Для сравнения оценок максимального правдоподобия с оценками наименьших квадратов, описанных выше, проведены компьютерные

эксперименты. Оценивается параметр θ , в предположении, что значение σ известно. Используются следующие значения параметров: $\theta = -0.3$, $\sigma = 1$, $c_t \equiv 0$, $T = 300$. При вычислении значений функции (3) – (5), входящих в (6), суммирование по k проводится от 1 до k_{\max} .

Таблица 1

Результаты компьютерных экспериментов

№	МНК	Ошибка	kmax=2		kmax=6	
			ОМП	Ошибка	ОМП	Ошибка
1	0.17639	0.47639	-0.345856	0.045856	-0.30225	0.002252
2	0.190799	0.490799	-0.333964	0.033964	-0.26261	0.037387
3	0.155476	0.455476	-0.389459	0.089459	-0.37559	0.075586
4	0.149382	0.449382	-0.427117	0.127117	-0.40333	0.103333
5	0.195661	0.495661	-0.383514	0.083514	-0.3518	0.051802
6	0.151674	0.451674	-0.417207	0.117207	-0.3855	0.085495
7	0.178645	0.478645	-0.365676	0.065676	-0.31018	0.01018
8	0.154852	0.454852	-0.433063	0.133063	-0.40333	0.103333
9	0.164298	0.464298	-0.341892	0.041892	-0.29036	0.00964
10	0.166816	0.466816	-0.294324	0.005676	-0.24874	0.051261
среднее	0.168399	0.468399	-0.37625	0.077508	-0.33337	0.053027

Результаты экспериментов приведены в таблице 1. Как видно из таблицы 1, метод максимального правдоподобия, в отличие от метода наименьших квадратов, используемого на практике, дает приемлемые результаты. Также отметим, что с ростом числа слагаемых в (3) – (5) точность оценивания по методу максимального правдоподобия увеличивается.

Литература

1. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. М. 1974.
2. Андерсон Т. Статистический анализ временных рядов. М. 1976.
3. Харин Ю. С. Оптимальность и робастность в статистическом прогнозировании. Мн., 2008.
4. Park J. W., Genton M. G., Ghosh S. K. Censored time series analysis with autoregressive moving average models // The Canadian journal of statistics. 2007. Vol. 35. № 1. P. 151–168.

**ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
В СТЕГАНОГРАФИИ**

Е. В. Вечерко

ВВЕДЕНИЕ

В настоящее время стеганография стремительно развивается и используется при решении задач защиты информации и авторских прав [1–5].

Вероятностно-статистические вопросы стеганографии являются недостаточно проработанными, в частности, актуальна проблема построения и анализа адекватных математических моделей контейнеров, используемых для скрытия (“вкрапления” [4]) защищаемой информации.

Метод LSB, заключающийся в использовании наименее значимых бит цифровых представлений контейнеров для встраивания защищаемой информации [1,3], является одним из наиболее распространенных в стеганографии методов, который применяется для различных форматов контейнеров. В [2] предложен метод, основанный на статистическом анализе пар значений, которые изменяются при встраивании информации.

В статье исследуются вероятностно-статистические свойства LSB-метода.

ВЕРОЯТНОСТНЫЕ МОДЕЛИ И СВОЙСТВА СТЕГОКОНТЕЙНЕРОВ

Введем обозначения: $V = \{0,1\}$, $V_N = \{J = (j_k) : j_k \in V, k = 1, \dots, N\}$ - множество 2^N двоичных N -векторов; $A = \{0, 1, \dots, 2^N - 1\}$ - множество из 2^N элементов; L - закон распределения вероятностей; $Bi(k, p)$ - биномиальный закон распределения с параметрами k, p ; $\delta_{i,j}$ - символ Кронекера. Математическую модель контейнера в достаточно общем случае можно представить последовательностью N -мерных двоичных случайных векторов-столбцов x_1, x_2, \dots, x_n :

$$x_t = (x_{t1}, x_{t2}, \dots, x_{tN})' \in V_N, \quad t = 1, \dots, n,$$

где n определяет размер контейнера. Двоичный вектор $x_t \in V_N$ будем отождествлять с числом $\langle x_t \rangle = x_{t1} + 2^1 x_{t2} + \dots + 2^{N-1} x_{tN}$, $\langle x_t \rangle \in A$.

Рассмотрим наиболее распространенный LSB-метод встраивания сообщения в наименее значимый 1-ый бит, причем механизм встраивания – “чисто случайный” [3,4]:

$$\tilde{x}_{t1} = \xi_t m_{\tau_t} + (1 - \xi_t) x_{t1} = \begin{cases} x_{t1}, \xi_t = 0, \\ m_{\tau_t}, \xi_t = 1 \end{cases}, \quad \tilde{x}_{tk} = x_{tk}, \quad k = 2, \dots, N, \quad (1)$$

$$L\{\xi_t\} = Bi(1, \beta), \quad \tau_t = \tau_t(\xi_1, \dots, \xi_t) = \sum_{i=1}^t \xi_i, \quad t = 1, \dots, n, \quad (2)$$

где $\tilde{x}_t = (\tilde{x}_{t1}, \dots, \tilde{x}_{tN})' \in V_N$ - контейнер, содержащий встроенное сообщение, называемый стегоконтейнером; ξ_t - последовательность независимых в совокупности случайных величин Бернулли, $P\{\xi_t = 1\} = 1 - P\{\xi_t = 0\} = \beta$,

определяющая механизм встраивания сообщения; $\beta \in [0,1]$ – доля встро-
енных в контейнер бит сообщения; $m_i \in V$ – встраиваемое сообщение.
Случайные последовательности $\{x_i\}$, $\{\xi_i\}$, $\{m_i\}$ являются взаимно незави-
симыми.

Введем случайную величину $\eta = \{\text{число бит, в которых } \tilde{x}_i \neq x_i\}$:

$$\eta = \sum_{i=1}^n I\{\tilde{x}_i \neq x_i\}, \quad \eta \in \{0, \dots, n\}.$$

Следующая теорема позволяет установить точное распределение ве-
роятностей стегоконтейнера x_i и случайной величины η .

Теорема 1. Пусть $x_i \in V_N$ - последовательность одинаково распределен-
ных двоичных случайных векторов с независимыми компонентами x_{i1} ,
 $m_i \in V$ - двоичная случайная последовательность, представляющая собой
встраиваемое сообщение, $P\{m_i = 1\} = 1 - P\{m_i = 0\} = p_1$, а стегоконтейнер \tilde{x}_i
строится согласно (1), (2). Тогда, если $\pi = (\pi_i)$ есть распределение веро-
ятностей x_i : $\pi_i = P\{x_i = i\}$, $i \in A$, то распределение вероятностей стего-
контейнера \tilde{x}_i и случайной величины η имеют вид:

$$\tilde{\pi}_{\langle J \rangle} = P\{\langle \tilde{x}_i \rangle = \langle J \rangle\} = P\{\tilde{x}_i = J\} = (1 - \beta)\pi_{\langle J \rangle} + \beta p_1^{j_1} (1 - p_1)^{1-j_1} \sum_{v=0}^1 \pi_{\langle J \rangle - j_1 + v},$$

$$L\{\eta\} = Bi(n, \beta(1 - p_1 - \sum_{i=0}^{2^{N-1}-1} \pi_{2i} + 2p_1 \sum_{i=0}^{2^{N-1}-1} \pi_{2i})). \quad (3)$$

На практике обычно имеют дело с большими выборками, поэтому
вместо точных формул удобно использовать их асимптотический при
 $n \rightarrow \infty$ вариант:

$$L\{\eta\} \xrightarrow{n \rightarrow \infty} \mathcal{N}(np, np(1 - p)),$$

где $p = \beta(1 - p_1 - \sum_{i=0}^{2^{N-1}-1} \pi_{2i} + 2p_1 \sum_{i=0}^{2^{N-1}-1} \pi_{2i})$; $\mathcal{N}(a, D)$ - нормальный закон рас-
пределения вероятностей с параметрами a, D .

Введем в рассмотрение функционал, характеризующий указанное
свойство:

$$\Delta_1(\tilde{\pi}) = \sum_{k=0}^{2^{N-1}-1} \sum_{v=0}^1 (\tilde{\pi}_{2k+v} - c_k)^2 \geq 0, \quad c_k = \frac{1}{2} \sum_{v=0}^1 \tilde{\pi}_{2k+v}.$$

Следствие 1. В условиях теоремы 1, если $p_1 = 1/2$, то

$$\Delta_1(\tilde{\pi}) = (1 - \beta)^2 \Delta_1(\pi).$$

Следствие 2. В условиях следствия 2, если $\beta = 1$, то $\Delta_1(\tilde{\pi}) = 0$, а если, вдобавок, $\{x_t\}$ – независимые случайные векторы, то

$$\widehat{\Delta}_1 = \Delta_1(\widehat{\tilde{\pi}}) \xrightarrow[n \rightarrow \infty]{P} 0,$$

где $\widehat{\tilde{\pi}}$ – статистическая оценка распределения вероятностей $\tilde{\pi}$ стегоконтейнера по наблюдаемой реализации $\tilde{x}_1, \dots, \tilde{x}_n$:

$$\widehat{\tilde{\pi}}_i = v_i / n, \quad v_i = \sum_{t=1}^n \delta_{\langle \tilde{x}_t, i \rangle}, \quad i \in A.$$

Рассмотрим более сложную модель контейнера, учитывающую марковскую зависимость в последовательности $\{x_t\}$ двоичных N -векторов.

Терма 2. Пусть x_t – стационарная цепь Маркова с пространством состояний V_N , стационарным распределением вероятностей $\pi = (\pi_i)$ и матрицей вероятностей одношаговых переходов $P = (p_{ij})$, $m_t \in V$ – последовательность независимых случайных величин, представляющая собой встраиваемое сообщение, $P\{m_t = 1\} = 1 - P\{m_t = 0\} = p_1$, а стегоконтейнер \tilde{x}_t строится согласно (1), (2). Тогда для двумерного распределения вероятностей $\tilde{\pi}_{\langle J, \langle K \rangle} = P\{\tilde{x}_{t-1} = J, \tilde{x}_t = K\}$, $J, K \in V_N$, стегоконтейнера \tilde{x}_t справедливо выражение:

$$\begin{aligned} \tilde{\pi}_{\langle J, \langle K \rangle} &= (1 - \beta)^2 \pi_{\langle J \rangle} p_{\langle J, \langle K \rangle} + \beta^2 p_1^{j_1+k_1} (1 - p_1)^{2-(j_1+k_1)} \sum_{v, h=0}^1 \pi_{\langle J \rangle - j_1 + v} p_{\langle J \rangle - j_1 + v, \langle K \rangle - k_1 + h} + \\ &+ \beta(1 - \beta) \left(p_1^{k_1} (1 - p_1)^{1-k_1} \sum_{v=0}^1 \pi_{\langle J \rangle} p_{\langle J, \langle K \rangle - k_1 + v} + p_1^{j_1} (1 - p_1)^{1-j_1} \sum_{v=0}^1 \pi_{\langle J \rangle - j_1 + v} p_{\langle J \rangle - j_1 + v, \langle K \rangle} \right). \end{aligned}$$

Следствие 4. Если $\beta = 1, p_1 = 1/2$, то

$$\widehat{\Delta}_2 = \Delta_2(\{\widehat{\tilde{\pi}}_{\langle J, \langle K \rangle}\}) = \sum_{k=0}^{2^{N-1}-1} \sum_{l=0}^{2^{N-1}-1} \sum_{v_1, v_2=0}^1 (\widehat{\tilde{\pi}}_{2k+v_1, 2l+v_2} - \widehat{c}_{k,l})^2 \xrightarrow[n \rightarrow \infty]{P} 0,$$

$$\text{где } \widehat{\tilde{\pi}}_{i,j} = \sum_{t=2}^n \delta_{\langle \tilde{x}_{t-1}, i \rangle} \cdot \delta_{\langle \tilde{x}_t, j \rangle} / (n-1), \quad i, j \in A, \quad \widehat{c}_{k,l} = \sum_{v_1, v_2=0}^1 \widehat{\tilde{\pi}}_{2k+v_1, 2l+v_2} / 4.$$

Рассмотрим более общий случай, когда сообщение встраивается в N_1 наименее значимых бит. Обозначим: $N = N_1 + N_2$, $1 \leq N_1 \leq N$, $m_\tau = (m_{\tau 1}, \dots, m_{\tau N_1}) \in V_{N_1}$, $J = (J_{(1)'}, J_{(2)'})' \in V_N, J_{(1)} \in V_{N_1}$, $\tilde{x}_t = (\tilde{x}_{t(1)'}, \tilde{x}_{t(2)'})'$, $\tilde{x}_{t(1)} \in V_{N_1}, \tilde{x}_{t(2)} \in V_{N-N_1}$. Стегоконтейнер \tilde{x}_t строится согласно обобщению (1):

$$\tilde{x}_{t(2)} = x_{t(2)}, \quad \tilde{x}_{t(1)} = \xi_t m_{\tau_t} + (1 - \xi_t) x_{t(1)} = \begin{cases} x_{t(1)}, \xi_t = 0, \\ m_{\tau_t}, \xi_t = 1 \end{cases}, \quad t = 1, \dots, n, \quad (4)$$

где случайные величины $\{\xi_t\}$ и функция τ_t определены в (2). Теорема 3 допускает следующее обобщение на случай $N_1 > 1$.

Терема 3. Пусть x_t - стационарная цепь Маркова с пространством состояний V_N , стационарным распределением вероятностей $\pi = (\pi_i)$ и матрицей вероятностей одношаговых переходов $P = (p_{ij})$, $m_t \in V_{N_1}$ - последовательность независимых случайных векторов, представляющая собой встраиваемое сообщение, $P\{m_t = J_{(1)}\} = p_{\langle J_{(1)} \rangle}$, $J_{(1)} \in V_{N_1}$, а стегоконтейнер строится согласно (4). Тогда для двумерного распределения вероятностей $\tilde{\pi}_{\langle J \rangle, \langle K \rangle} = P\{\tilde{x}_{t-1} = J, \tilde{x}_t = K\}$, $J, K \in V_N$, стегоконтейнера \tilde{x}_t справедливо выражение:

$$\begin{aligned} \tilde{\pi}_{\langle J \rangle, \langle K \rangle} = & (1 - \beta)^2 \pi_{\langle J \rangle} p_{\langle J \rangle, \langle K \rangle} + \beta^2 p_{\langle J_{(1)} \rangle} p_{\langle K_{(1)} \rangle} \sum_{v, h=0}^{2^{N_1-1}} \pi_{2^{N_1} \langle J_{(2)} \rangle + v} p_{2^{N_1} \langle J_{(2)} \rangle + v, 2^{N_1} \langle K_{(2)} \rangle + h} + \\ & + \beta(1 - \beta) \left(p_{\langle K_{(1)} \rangle} \sum_{v=0}^{2^{N_1-1}} \pi_{\langle J \rangle} p_{\langle J \rangle, 2^{N_1} \langle K_{(2)} \rangle + v} + p_{\langle J_{(1)} \rangle} \sum_{v=0}^{2^{N_1-1}} \pi_{2^{N_1} \langle J_{(2)} \rangle + v} p_{2^{N_1} \langle J_{(2)} \rangle + v, \langle K \rangle} \right). \end{aligned}$$

Литература

1. *Anderson R. J.* Stretching the Limits of Steganography. LNCS. Vol. 1174. London: Springer-Verlag, 1996.
2. *Westfeld, A., Pfitzmann A.* Attacks on Steganographic Systems. LNCS. Vol. 1768. Springer-Verlag, 2000.
3. *Грибунин В. Г.* Цифровая стеганография. М.: Солон-Пресс. 2002.
4. *Пономарев К. И.* // Параметрическая модель вкрапления и ее статистический анализ. Дискретная математика. 2009. Том 21.
5. *Вечерко Е. В., Харин Ю. С.* // О некоторых задачах статистической проверки гипотез в стеганографии. Материалы международной конференции Информационные системы и технологии. Минск. 2009. С. 14–18.

ВЫДЕЛЕНИЕ СРЕДНИХ ЛИНИЙ ОБЪЕКТОВ НА ТРЕХМЕРНЫХ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЯХ

Д. А. Гончаров

ВВЕДЕНИЕ

Выделение средних линий объектов является важным этапом в обработке изображений для последующих вычислений различных (линейных, площадных, объемных) характеристик объекта, а также описания и рас-

познавания объектов. Также оно может использоваться для мониторинга заболеваний путем контроля изменений на изображениях.

Существует ряд алгоритмов позволяющих построить бинарный 3D скелет. Но из-за особенностей медицинских изображений таких как зашумленность, большие расстояния между слоями, геометрические и оптические искажения, нечеткие границы объектов, динамическое изменение объектов, с их помощью нельзя построить устойчивые к шумам средние линии. К тому же в случае исследования лучевыми методами много информации получают на основе данных плотности ткани. Для решения этой задачи разработан алгоритм полутонного утоньшения, позволяющий построить образ сосудистой системы для дальнейшего анализа. В основе построения скелета сосудов лежит алгоритм псевдодистанционной карты [1], расширенный на трехмерный случай.

1. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ И ПОСТАНОВКА ЗАДАЧИ

Средние линии имеют несколько различных математических определений, содержащихся в технической литературе. Эти определения аналогичны для непрерывных пространств, но обычно приводят к разным результатам в дискретных пространствах.

В двумерном случае средняя линия фигуры это множество кривых, определяемых как геометрическое место точек, которые имеют, по крайней мере, две ближайшие точки границы фигуры [2]. В трехмерном случае, соответствующий объект называется средней (или медиальной) плоскостью.

Наиболее наглядное определение средней линии дается через точки гашения в модели распространения огня, где граница фигуры сделана полностью из сухой травы, которую подожгли и средняя линия состоит из множества точек, где огни разных фронтов встречаются и гасят друг друга [3].

Также средние линии могут определяться как геометрическое место центров максимальных шаров (или дисков в двухмерном случае) [4]. Более формально, положим $X \subset R^3$ – трехмерный объект. Шар радиуса r с центром в точке $x \in X$ определяется как $S_r(x) = \{y \in R^3, d(x, y) < r\}$, где $d(x, y)$ – расстояние между двумя точками x и y в R^3 . Шар $S_r(x) \subset X$ называется максимальным, если для него не существует шара большего радиуса, принадлежащего объекту, полностью содержащего $S_r(x)$.

Основное неудобство средних линий состоит в чувствительности к небольшим изменениям границ объекта.. Чем больше изменений на границе, тем больше ветвей в скелете. И большинство этих ветвей – беспо-

лезные, так как не несут никакой информации для анализа формы фигуры. Поэтому необходим процесс регуляризации – очистки полученного скелета от шума.

Таким образом, практическое использование скелетов в задачах обработки и анализа изображений требует разработки эффективных алгоритмов скелетизации и алгоритмов регуляризации скелетов путем очистки их от шумовых эффектов.

2. АЛГОРИТМ ПОСТРОЕНИЯ 3D СКЕЛЕТА

Из-за специфики получения изображений компьютерной томографии расстояние между срезами гораздо больше расстояния между вокселями, принадлежащими одному срезу. Возникает необходимость в восстановлении недостающих срезов.

Для восстановления использовался алгоритм интерполяции, основанный на морфинге изображений [5].

Для выделения топологических особенностей изображения использовалась расширенная для трехмерного случая псевдодистанционная карта (ПДК) [1]. Для оптимизации процесса построения ПДК использовалась чамферная метрика, которая позволяет получить оптимальное соотношение скорости и качества. Построение ПДК осуществляется за два прохода по изображению.

Затем с помощью анализа окрестности каждого вокселя находились топологические особенности (локальные максимумы, седловые точки, точки двойного седла), которые затем использовались для нахождения скелета (табл. 1).

Анализ на особость точек проводится вдоль плоскостей Oxy , Oyz , Oxz – для каждого случая анализировалась 8-окрестность вокселя. Также проверялись угловые элементы куба $3 \times 3 \times 3$. Таким образом, на особость воксель проверялся в 26-окрестности.

Здесь p – текущий пиксель, p_i – соседний пиксель, c_i условие $p > p_i$, $FindMax(i)$ – функция подсчета больших по величине пикселей в окружении пикселя p_i .

Все эти точки отмечались максимальным уровнем на псевдодистанционной карте и являлись стартовыми точками для роста скелета.

Затем в окрестности какой-либо стартовой точки искались остальные особые точки. В случае нахождения, строился кратчайший путь между двумя найденными точками с помощью алгоритма Дейкстры. Вновь найденные точки помещались в очередь, и алгоритм продолжал работу, уже считая эти точки стартовыми.

Топологические особенности, используемые для построения ПДК

Вид особой точки	Условия выделения
Точки седла	Если $(c_0 \wedge c_4 \wedge \overline{c_2} \wedge \overline{c_6}) \vee (\overline{c_0} \wedge \overline{c_4} \wedge c_2 \wedge c_6) \vee$ $(c_1 \wedge c_5 \wedge (\overline{c_2} \wedge c_3 \wedge c_4) \wedge (\overline{c_0} \wedge c_6 \wedge c_7)) \vee$ $(c_3 \wedge c_7 \wedge (\overline{c_0} \wedge c_1 \wedge c_2) \wedge (\overline{c_4} \wedge c_5 \wedge c_6))$
Точки двойного седла	Если $(c_5 \wedge n_1 = p \wedge \overline{c_0} \wedge \overline{c_2} \wedge FindMax(1) > 0) \vee$ $(c_1 \wedge n_5 = p \wedge \overline{c_4} \wedge \overline{c_6} \wedge FindMax(5) > 0) \vee$ $(c_3 \wedge n_7 = p \wedge \overline{c_0} \wedge \overline{c_6} \wedge FindMax(7) > 0) \vee$ $(c_7 \wedge n_3 = p \wedge \overline{c_2} \wedge \overline{c_4} \wedge FindMax(3) > 0) \vee$ $(c_4 \wedge n_0 = p \wedge \overline{c_2} \wedge \overline{c_6} \wedge FindMax(0) > 0) \vee$ $(c_0 \wedge n_4 = p \wedge \overline{c_2} \wedge \overline{c_6} \wedge FindMax(4) > 0) \vee$ $(c_6 \wedge n_2 = p \wedge \overline{c_0} \wedge \overline{c_4} \wedge FindMax(2) > 0) \vee$ $(c_2 \wedge n_6 = p \wedge \overline{c_0} \wedge \overline{c_4} \wedge FindMax(6) > 0)$
Локальные максимумы	Если $\sum c_i > 0$

ЗАКЛЮЧЕНИЕ

Предложенный алгоритм позволяет получить качественный бинарный скелет объекта и на его основе вычислить характеристики структуры объекта. Основным преимуществом данного метода является хорошая устойчивость к шумам, инвариантность относительно поворотов и масштабирования, скорость работы. Из недостатков стоит отметить большой объем используемой памяти, отсутствие возможности восстановления исходного объекта.

Данный алгоритм реализован в рамках программного комплекса анализа изображений компьютерной томографии.

Литература

1. Недзьведь А.М., Абламейко С.В., Белоцерковский А.М., Лемман Т. Быстрый метод выделения сети сосудов на полутоновых ангиографических изображениях // Искусственный интеллект. 2006. №2. С. 206–210.
2. W. Gong and G. Bertrand. A simple parallel 3D thinning algorithm // Proc. IEEE Pattern Recognition. 1990. P. 188–190
3. H. Blum. A Transformation for Extraction New Descriptors of Shape, Models for the Perception of Speech and Visual Form // MIT Press. 1967.
4. A. Lieutier. Any open bounded subset of R^n has the same homotopy type than its medial axis // Proc. ACM SMI, 2003.

5. *H. Atoui, S. Miguet and D. Sarrut* A fast morphing-based interpolation for medical images: application to conformal radiotherapy // *Image Analysis and Stereology*. 2006. №25. P. 95–103.

ПРОГРАММНАЯ РЕАЛИЗАЦИИ ЭМПИРИЧЕСКОЙ МОДЕЛИ РАСПРОСТРАНЕНИЯ ФРОНТА ПОЖАРА

Д. А. Горбацевич

Постановка задачи. Разработать, составить, запрограммировать и отладить алгоритм и программу, обеспечивающие реализацию эмпирической модели распространения фронта пожара Ротермела [1; 2]. Считается, что известны граница лесного массива, площадь которого может включать участки с разными характеристиками горимости; задаются контура: границы участков, начальная форма очага пожара. Требуется в заданные моменты времени рассчитать, сформировать и показать на экране изображения, иллюстрирующие прогнозируемое положение фронта пожара; для каждого момента времени рассчитать площадь выгоревшего леса, периметр фронта пожара.

Решение задачи. Традиционно применяемая полуэмпирическая модель основывается на описании геометрии на площади криволинейными границами, в настоящей же реализации все границы аппроксимируются на равномерной квадратной сетке. Дополнительным основанием введения сетки является то, что реальные данные о границах площади лесного массива, характеристиках горимости, параметрах площадей преобладающей породы и неоднородных включений, начального фронта возгорания имеют ограниченную точность. В настоящей методике площадь лесного массива покрывается совокупностью ячеек сетки, вводится понятие граничных и внутренних ячеек. Всем ячейкам присваивается код, характеризующий участок площади. Граничными называются ячейки, ограничивающие область выгоревшего леса, то есть перед ними лесной массив находится в первоначальном состоянии, за ними лес выгорел. Можно сказать, что такие ячейки (если размер ячейки мал) аппроксимируют фронт пожара. Внутренними называются ячейки, центр которых лежит внутри аппроксимированного контура. Частью приближения в реализации принятой модели распространения пожара также являются: каждый единичный элемент фронта за выбранный интервал времени порождает некоторую площадь, границы которой определяются формулами полуэмпирической модели; соответствующая форма заменяется многоугольником. Аппроксимированный (вычисленный по априорной методике) на принятую сетку графический шаблон в дальнейшем называется базовым элементом расширения.

Задача программной реализации эмпирической модели распространения фронта пожара включает следующие части:

1. Разработка интерфейса Загрузки / Создания / Редактирования контуров-границ лесного массива, однородных включений, очагов пожара.

2. Организация диалога выбора характеристик горимости всех однородных включений, в которых определяются базовые шаблоны расширения для площади массива и каждого отдельного включения.

3. Разработка интерфейса Загрузки / Создания / Редактирования требуемых шаблонов расширения в случае режима работы не по эмпирической модели, а по алгоритму пользователя.

4. Аппроксимация границ и контуров равномерной квадратной сеткой.

5. Нахождение граничных ячеек очага пожара.

6. Подсчет площади пожара и периметра фронта.

7. Расчет формы зоны, охваченной пожаром (с использованием базовых или введенных пользователем) шаблонов расширения.

Для работы программы необходимо, чтобы в каталоге, из которого она запускается, были наполненная файлами типа `x.x.x.x.x.***.ftm` папка `DB`. Файлы из папки `DB` формируются с учетом характеристик элементарной площадки – ячейки сетки [1]. В этих файлах хранятся соответствующие графические шаблоны (базовые элементы расширения). Обязательно наличие в папке `DB` файла `fuelTypes.ftl`, остальные могут отсутствовать – режим работы «с результатами наблюдений», когда пользователь не работает с автоматически формируемыми шаблонами, а самостоятельно прорисовывает и сохраняет шаблон. Также необходимо наличие в активном каталоге файла `config.cfg`. В этом файле задаются (и могут быть изменены) цветовые наборы – параметрами `r g b` компьютерной модели цвета `RGB` (каждый из упомянутых параметров задается целым числом из диапазона 0–255).

Реализована возможность создания, загрузки и редактирования контуров. Все вышеупомянутые функции доступны во вкладках меню «Площадь», «Контур». После задания контуров лесного массива, однородных включений и очагов пожара, для каждого из контуров-границ пожара определяется множество внутренних ячеек по следующему алгоритму:

1. Строится окаймляющий прямоугольник, который ограничивает площадь всего лесного массива. Считается, что ячейка принадлежит прямоугольнику, если ее центр лежит внутри данного прямоугольника.

2. Выделяется множество ячеек, принадлежащих построенному прямоугольнику.

3. Далее для каждой выделенной ячейки, проверяется ее принадлежность контуру. Если ячейка принадлежит контуру, то она помечается как внутренняя и заливается соответствующим цветом.

4. Если данный контур аппроксимирует границу очага пожара, то после определения внутренних ячеек необходимо найти и пометить граничные ячейки. Для этого для каждой внутренней ячейки проверяются ее соседние ячейки. Если среди них есть внешние ячейки для данного контура, то ячейка помечается как граничная и заливается соответствующим цветом.

Таким образом, решается задача аппроксимации на равномерной квадратной сетке контуров лесного массива, однородных включений и очагов лесного пожара, маркировки внутренних и внешних ячеек. Для получения прогнозного положения очага пожара через заданный промежуток времени необходимо загрузить из базы данных шаблоны для лесного массива и каждого однородного включения, которые находятся в папке DB и считываются в соответствии с типом лесной растительности данного однородного включения и заданными параметрами полуэмпирической модели (рассчитываются по методам [1, 2] и упомянутому алгоритму). В частности, учитываются: направление и скорость ветра, класс пожарной опасности, шаг по времени, шаг сетки. Параметры модели можно изменить, используя вкладку меню «Обработка» → «Параметры модели». Для загрузки шаблонов из базы данных необходимо выбрать «Шаблон» → «Загрузить из БД». После загрузки необходимых шаблонов можно приступить непосредственно к расчету прогнозируемого положения фронта пожара через заданный промежуток времени. Путем нажатия горячей клавиши «F4» или выбора пункта меню «Обработка» → «Наложить шаблон» рассчитываются прогнозируемое положение фронта пожара и параметры площади выгоревшего леса за заданный промежуток времени. Наложение шаблона происходит по следующему алгоритму:

1. На каждую ранее определенную граничную ячейку фронта пожара накладывается шаблон, соответствующий коду данной ячейки.
2. После наложения шаблона на каждую граничную ячейку фронта пожара необходимо определить новые граничные ячейки. Выделение граничных ячеек проводится по ранее описанному алгоритму.
3. При необходимости продолжения прогнозирования повторяют вышеописанный процесс.

Проиллюстрируем работу алгоритма на следующем примере. На рисунке 1 изображена эволюция распространения лесного пожара, который имеет включения, соответствующие лесопожарному заслону. Также показаны шаблоны расширения для каждого включения. Приведенные схемы получены для последовательных моментов времени: 0,195 минут соответственно.

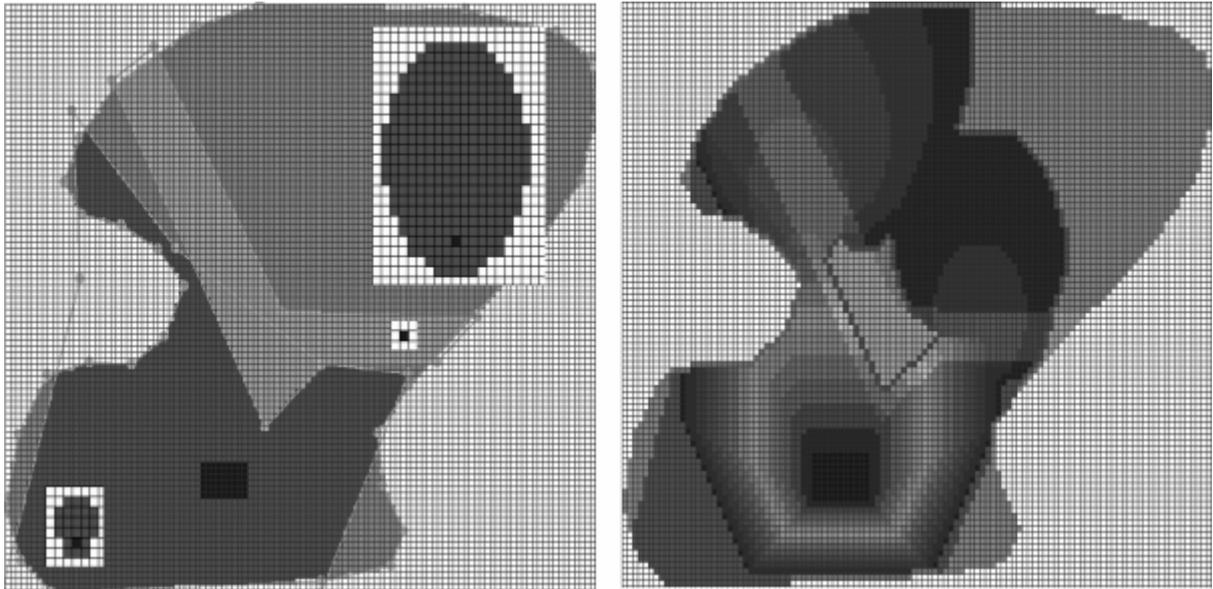


Рис. 1. Иллюстрации работы алгоритма на неоднородном лесном массиве.

Литература

1. Баровик Д. В. Об особенностях адаптации математических моделей вершинных верховых лесных пожаров / Д. В. Баровик, В. Б. Таранчук // Вестн. Белорус. ун-та. Сер. 1. 2010. № 1.
2. Rothermel R. C. A mathematical model for fire spread predictions in wildland fuels / R. C. Rothermel // USDA Forest Service. Res. Paper INT-115. Ogden. 1972.

СИСТЕМЫ МОДЕЛИРОВАНИЯ ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЕЙ

В. С. Горшунов

ВВЕДЕНИЕ

Развитие программных моделей, описывающих сложные процессы реального мира, стало актуальным вопросом сегодняшнего дня. Первоначально модели процессов были статичны, параметры системы не изменялись в течение всего процесса моделирования, что лишало гибкости и возможности быстро отреагировать на изменения реального мира. Дальнейшее развитие технологии привело к созданию динамических моделей, параметры которых могут изменяться в процессе функционирования. В связи с чем, появилась задача определения оптимальных параметров моделей. Для этого на первом этапе данные определялись посредством формирования отчетов функционирования, после чего эти отчеты обрабатывались вручную. Следующим этапом было создание интерактивных отчетов, позволяющих изменять представление накопленных данных. Этот этап получил название – исследования модели. С увеличением мощностей информа-

ционных систем стало возможным создавать отчеты в режиме реального времени. Естественным требованием для таких отчетов стала возможность вносить изменения в функционирующую модель. Однако системному аналисту, как и ранее, приходилось планировать и принимать конечное решение на основе собственных знаний и ранее полученного опыта. Таким образом, в связи с не идеальностью технологии возникла задача создания системы симуляции и прогнозирования поведения формальной модели на основе измененных входных данных в реальном времени. Что на данный момент является последним этапом развития систем моделирования и планирования сложных процессов.

ОБЩИЙ ПРОЦЕСС ФУНКЦИОНИРОВАНИЯ

Для функционирования системы описывающей сложные процессы реального мира необходимо, прежде всего, создать формальную модель, определяющую логику поведения системы. Формальная модель строится на формальном языке, для удобства возьмем подмножество XML. Она содержит входные и контрольные параметры. Все параметры могут быть как примитивными типами, такими как число, символ, булево значение и дата, так и массивами примитивных типов. Результатом работы формальной модели является набор предложений, требующих от пользователя ответов в диапазоне от 0 до 1, где 0 – отказ от предложения, 1 – принятия предложения, а промежуточные значения степень согласия, например 0.75 – да, но не сейчас, 0.25 – скорее нет, чем да и т. д. Ответы пользователей накапливаются в базе данных и классифицируются по параметрам, заданным как входные для формальной модели. На базе накопленных данных проводится процесс симуляции поведения пользователя в зависимости от изменения контрольных параметров. Процесс симуляции относится к так называемым хорошо распараллеливаемым процессам, поэтому процесс симуляции логично производить в вычислительном облаке. Облака предоставляют ресурсы в виде единой виртуальной машины, динамически распределяя процессы и данные по различным серверам системы для обработки, таким образом, удобно чтобы каждый вычислительный элемент в облаке производил симуляцию на своем подмножестве исторически накопленных данных. Виртуальные сервера вычислительных облаков легко масштабируются за счёт отсутствия необходимости в каких-либо дополнительных настройках. В данном случае элементы облака образуют кольцевую структуру, что упрощает процедуру распределения задач и получения симулированных результатов. Результаты симуляции аккумулируются с использованием метрик, задаваемых бизнес пользователем. На основе данных предоставляемых метриками принимается решения об изменении поведения модели через изменения значений контрольных параметров.

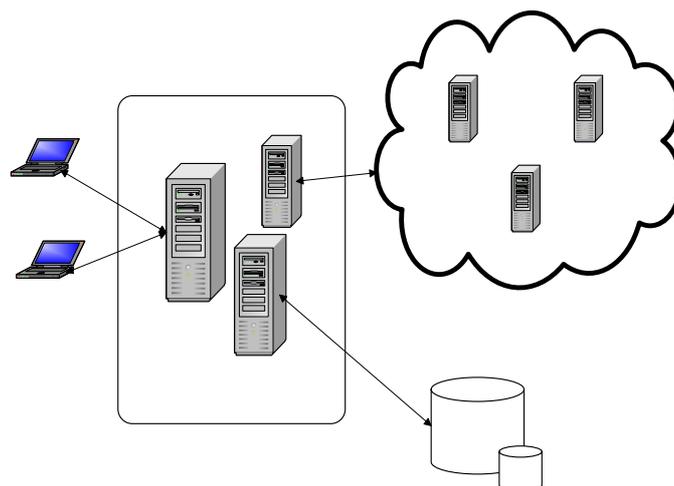


Рис.1. Общая архитектура системы планирования и прогнозирования поведения потребителей BI Client

АРХИТЕКТУРА СИСТЕМЫ

Как было описано ранее, система, описывающая сложные процессы реального мира, должна иметь возможность выполнять распределенные вычисления, хранить данные в кэше и иметь возможность обслуживать нескольких бизнес клиентов одновременно.

Архитектура подобного приложения изображена на рисунке 1), где:

- BI Client – бизнес пользователь
- App Servers Master Grid Node – управляющие сервера
- Grid Node – узлы вычислительного облака
- Data store – база данных

App Servers
Master Grid Node

ВЫВОД

Необходимо отметить широкий спектр применения систем, которые помогают описать сложные процессы реального мира. Подобные технологии, необходимы как в статистических системах, не преследующих коммерческих целей, так и в торговых или банковских системах, где вовремя принятое правильное решение помогает увеличить прибыль. Возможность оперативного прогнозирования и принятия обоснованного решения является одним из ключевых факторов современного мира.

Литература

1. *Pine B. J, Gilmore J.* Welcome to the Experience Economy. Harvard Business Review. 1998.
2. *Peppers D., Rogers M.* Return on Customer. – Doubleday, division of random House. Inc. ISBN 0-385-51030-6. 2005
3. *Smith S., Wheeler J.* Managing the Customer Experience: Turning customers into advocates. – Financial Times Press, ISBN 978-0273661955. 2002