Оценка параметров модели случайного процесса нагрузки сервера, обрабатывающего заявки

Владимир Руцкий, 5057/12

Санкт-Петербургский государственный политехнический университет

31 мая 2011

План презентации

1. Постановка задачи
2. Решение в случае бесконечного времени обработки заявки
   - Итеративный метод
   - Оценивание EM-алгоритмом
3. Решение в случае конечного времени обработки заявки
   - Метод наименьших квадратов
4. Результаты работы

Постановка задачи

Сервер

- Сервер обрабатывает приходящие заявки
- Для обработки заявки используются ресурсы сервера
  - Количество используемых ресурсов сервера — загрузка сервера — скалярная величина, например процессорное время

Задача

Дан лог загрузки сервера. Необходимо:

1. идентифицировать моменты прихода заявок,
2. оценить:
   - интенсивность прихода заявок,
   - загрузку сервера в фоновом режиме,
   - использование ресурсов сервера для обработки одной заявки

Математическая модель (1)

Загрузка сервера — случайный процесс $X(t)\colon \mathbb{R} \to \mathbb{R}$

Фоновая загрузка сервера

Сумма постоянной загрузки и винеровского процесса (шум):

$$B(t) = m + \sigma \mathcal{W}(t)$$

Загрузка сервера при обработке заявки

Заявка, пришедшая в $t = t_c$, увеличивает загрузку сервера на:

$$K_{t_c}(t) = \mathcal{N}(m_c, \sigma_c^2) \cdot \mathrm{I}(t - t_c) \cdot e^{-\lambda_c (t - t_c)}$$

$\mathrm{I}(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geqslant 0 \end{cases}$ — Функция Хевисайда

Математическая модель (2)

$T_c = \{t_{c_1}, \ldots, t_{c_R}\}$ — множество моментов времени, когда поступили заявки (всего R заявок за время наблюдения за сервером)

Интенсивность поступления заявок

Интервал времени между поступлением двух последовательных заявок распределён экспоненциально:

$$(t_{c_i} - t_{c_{i-1}}) \sim \mathrm{Exp}(\lambda)$$

Общая загрузка сервера

$$X(t) = B(t) + \sum_{t_c \in T_c} K_{t_c}(t)$$

Перейдём к дискретному случайному процессу

$$X(t_i), \quad t_i = t_0 + i \cdot \Delta t \quad (t_0 \text{ и } \Delta t \text{ даны})$$

Входные данные

Лог загрузки сервера

Траектория $X(t_i)$: $\{x_i \mid i = 1, \ldots, N\}$

Считаем, что Δt достаточно мало и $t_{c_j} = t_0 + i \cdot \Delta t$ — заявки пришли в некоторые наблюдаемые моменты времени t_i.
Без ограничения общности будем рассматривать случай $t_0 = 0$

Задача

Задача

По траектории $X(t_i)$

1. идентифицировать моменты поступления заявок,
2. оценить параметры модели:
   - m, σ — параметры фоновой загрузки,
   - λ — интенсивность поступления заявок,
   - m_c, σ_c, λ_c — параметры загрузки сервера при обработке заявок

Случай бесконечного времени обработки заявки

$\lambda_c \approx 0$

При поступлении заявки в момент времени t_c загрузка сервера увеличивается на $\Delta x \sim \mathcal{N}(m_c, \sigma_c)$

$$K_{t_c}(t) = \mathcal{N}(m_c, \sigma_c^2) \cdot \mathrm{I}(t - t_c)$$

Разностный аналог производной

$\mathrm{d}X(t)$

Рассмотрим ненормированный разностный аналог производной:

$$\mathrm{d}X(t) = X(t) - X(t - \Delta t)$$

$\mathrm{d}X(t)$ в момент времени поступления заявки t_c

$$\mathrm{d}X(t_c) = \mathcal{N}(m_c, \sigma^2 \Delta t + \sigma_c^2)$$

(при условии, что в момент времени $(t_c - \Delta t)$ заявки не было)

$\mathrm{d}X(t)$ в момент времени отсутствия заявок t

$$\mathrm{d}X(t) = \mathcal{N}(0, \sigma^2 \Delta t)$$

(при условии, что в момент времени $(t - \Delta t)$ заявки не было)

Итеративный метод идентификации поступления заявок

Предположим, что в отрезке времени $[t_k, t_{k+n}]$ не пришло ни одной заявки.
Тогда $\mathrm{d}x_k, \ldots, \mathrm{d}x_{k+n}$ — наблюдения $\mathcal{N}(0, \sigma^2 \Delta t) = \mathrm{d}X(t)$.
Оценим σ^2 по $[x_k, x_{k+n}]$ (ММП):

$$\Delta t \cdot \widehat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} ((x_{k+i} - x_{k+i-1}) - 0)^2$$

Гипотеза H_0

В отрезке времени $[t_{k+n}, t_{k+n} + \Delta t]$ не поступило ни одной заявки

Критерий принятия H_0 с уровнем значимости α

Разность значений наблюдений $(x_{k+n+1} - x_{k+n})$ лежит в $(1-\alpha)$ квантиле нормального распределения $\mathcal{N}(0, \widehat{\sigma}^2 \Delta t)$:

$$H_0 \text{ принимается } \iff (x_{k+n+1} - x_{k+n}) < \mathcal{N}_{1-\alpha}$$

EM-алгоритм

Метод наименьших квадратов

Результаты работы (1)

Выводы