Состоятельность долгосрочных профилей

в системах выявления вторжений

Ермохин М. А.

Кафедра МаТИС
МГУ им. М.В. Ломоносова

Интеллектуальные системы и компьютерные науки,
2016

Содержание

Содержание

Цели

Задачи систем активного аудита:

- Выявление злоумышленной активности (на основе регулярных выражений).
- Выявление нетипичной активности.

Выявление злоумышленной активности

Плюсы:

- Гарантированность работы

Выявление злоумышленной активности

Плюсы:

- Гарантированность работы

Минусы:

- Экспоненциальный взрыв числа состояний.
- Невозможность выявлений атак, не заданных в базе сигнатур.

Неизвестные атаки

Один из способов выявления неизвестных атак — выявление ситуаций, нетипичных в том или ином смысле.

Неизвестные атаки

Один из способов выявления неизвестных атак — выявление ситуаций, нетипичных в том или ином смысле. Причины появления нетипичных ситуаций:

- проводимая атака;
- успешно проведенная атака;
- программный сбой;
- аппаратный сбой.

Содержание

Формализация задачи

Задача выявления нетипичного поведения может быть сведена к двум подзадачам:

- описание типичного поведения;
- сравнение текущего поведения с типичным.

Здесь рассматривается первая подзадача.

Описание типичного поведения

- Типичное поведение с течением времени может меняться.
- Более поздние события желательно учитывать с бóльшим весом.

Предложенные модели:

- скользящее окно (учитываются события за фиксированный промежуток времени);
- усреднение с экспоненциально убывающими весами.

Содержание

Случайные величины и их распределение

- $\{x_t\}_{t=1}^{\infty}$ — последовательность независимых дискретных случайных величин, принимающих значения из множества $\{1, \ldots, M\}$.
- $\mathbb{P}^I = \{(t_1, \ldots, t_I) \in \mathbb{R}^I \mid \forall i \in \overline{1, I}\ t_i > 0, \sum_{i=1}^{I} t_i = 1\}$.
- Задано семейство: $P(t) = (p_1(t), \ldots, p_M(t)) \in \mathbb{P}^M$.
- Для всех t случайная величина x_t имеет распределение $P(t)$.
- Положим $q_m(t) = 1 - p_m(t)$.

Усреднение

- Положим $\omega(t) = (\omega_1(t), \ldots, \omega_t(t))$ — некоторый t-мерный весовой вектор $(\omega_\tau(t) \geq 0,\ \tau = 1, \ldots, t)$, заданный для каждого момента времени t.

Всюду далее считаем, что конечные суммы весов равномерно ограничены константой $B > 0$ и отделены от нуля.

- Положим $\tilde{p}_m(t) = \sum_{\tau=1}^{t} \omega_\tau(t) I_{\{x_\tau = m\}}$.
- Вектор частот: $\tilde{P}(t) = (\tilde{p}_1(t), \ldots, \tilde{p}_M(t))$.

Содержание

Примеры

Пример

Случай $\omega_\tau(t) = ab^{-c(t-\tau)}$, где a, b и c — неотрицательные константы, задает усреднение с экспоненциально убывающими весами.

Пример

Случай $\omega(t) = \left(\underbrace{0, \ldots, 0}_{t-n}, \underbrace{\frac{1}{n}, \ldots, \frac{1}{n}}_{n}\right)$ соответствует схеме «скользящее окно».

Содержание

Определение

Профиль называется состоятельным, если он сходится к своему математическому ожиданию.

Теорема

Для того, чтобы $\forall \varepsilon > 0 \lim\limits_{t\to\infty} \mathrm{P}\{|\tilde{p}_m(t) - \mathrm{E}\tilde{p}_m(t)| \geq \varepsilon\} = 0$, *необходимо и достаточно, чтобы*

$$\lim_{t\to\infty} \sum_{\tau=1}^{t} \omega_\tau^2(t) p_m(\tau) q_m(\tau) = 0.$$

Для доказательства теоремы понадобится следующая лемма:

Лемма

Пусть последовательность дискретных случайных величин $\{\sigma_t\}_{t=1}^{\infty}$ *равномерно ограничена константой* $B > 0$, *неотрицательна и сходится по вероятности к константе* C. *Тогда дисперсия величины* $(\sigma_t - C)$ *стремится к* 0 *при* $t \to \infty$.

Содержание

Доказательство леммы

- Значения σ_t — множество $A_t = \{a_{1t}, \ldots, a_{n_t t}\}$.
- Необходимо для произвольного $\varepsilon > 0$ найти такое $T \in \mathbb{N}$, что $\forall t > T$ $\mathtt{D}\sigma_t < \varepsilon$.
- Рассматриваем $\varepsilon' > 0$ и $\delta' > 0$.
- По определению дисперсии: $\mathtt{D}(\sigma_t - C) = \sum_{\tau=1}^{n_t} (a_{\tau t} - \mathtt{E}\sigma_t)^2 \mathtt{P}\{\sigma_t = a_{\tau t}\} = D_1 + D_2$, где
- $D_1 = \sum\limits_{\tau : |a_{\tau t} - C| \geq \varepsilon'} (a_{\tau t} - \mathtt{E}\sigma_t)^2 \mathtt{P}\{\sigma_t = a_{\tau t}\}$,
- $D_2 = \sum\limits_{\tau : |a_{\tau t} - C| < \varepsilon'} (a_{\tau t} - \mathtt{E}\sigma_t)^2 \mathtt{P}\{\sigma_t = a_{\tau t}\}$.

Доказательство леммы

- Оцениваем D_1:
  - $D_1 \leq 4B^2 \sum_{\tau:|a_{\tau t}-C|\geq\varepsilon'} \mathrm{P}\{\sigma_t = a_{\tau t}\}$.
  - σ_t сходится по вероятности к C, следовательно, $D_1 < 4\delta' B^2$.

Доказательство леммы

- Оцениваем D_1:
  - $D_1 \leq 4B^2 \sum_{\tau:|a_{\tau t}-C|\geq\varepsilon'} \mathrm{P}\{\sigma_t = a_{\tau t}\}$.
  - σ_t сходится по вероятности к C, следовательно, $D_1 < 4\delta' B^2$.
- Оцениваем D_2:
  - $D_2 < (\varepsilon')^2 \sum_{\tau:|a_{\tau t}-C|<\varepsilon'} \mathrm{P}\{\sigma_t = a_{\tau t}\} < (\varepsilon')^2$

Доказательство леммы

- Оцениваем D_1:
  - $D_1 \leq 4B^2 \sum\limits_{\tau:|a_{\tau t}-C|\geq\varepsilon'} \mathrm{P}\{\sigma_t = a_{\tau t}\}$.
  - σ_t сходится по вероятности к C, следовательно, $D_1 < 4\delta' B^2$.
- Оцениваем D_2:
  - $D_2 < (\varepsilon')^2 \sum\limits_{\tau:|a_{\tau t}-C|<\varepsilon'} \mathrm{P}\{\sigma_t = a_{\tau t}\} < (\varepsilon')^2$
- Таким образом, для всех $t > T'$ справедливо $\mathrm{D}(\sigma_t - C) < 4\delta' B^2 + (\varepsilon')^2$. Если $(\varepsilon')^2 < \varepsilon$ и $\delta' < \frac{\varepsilon-(\varepsilon')^2}{4B^2}$, то при $T = T'$ и любом $t > T$ имеем $\mathrm{D}(\sigma_t - C) < \varepsilon$.

Доказательство теоремы

- Необходимость:
  - В силу независимости x_t:
    $$\sum_{\tau=1}^{t} \omega_\tau^2(t) q_m(\tau) p_m(\tau) = \mathtt{D}\tilde{p}_m(t).$$
  - Необходимость условия является следствием леммы.

Доказательство теоремы

- Необходимость:
  - В силу независимости x_t:
    $$\sum_{\tau=1}^{t} \omega_\tau^2(t) q_m(\tau) p_m(\tau) = \mathtt{D}\tilde{p}_m(t).$$
  - Необходимость условия является следствием леммы.
- Достаточность:
  - Следует из неравенства Чебышёва:
    $$\mathtt{P}\{|\tilde{p}_m(t) - \mathtt{E}\tilde{p}_m(t)| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2}\mathtt{D}\tilde{p}_m(t) \xrightarrow[t\to\infty]{} 0$$

Следствия

Следствие

Если в условиях теоремы для любых m и t справедливы неравенства $p_m(t) \geq \delta > 0$ для некоторого δ, то сходимость $\tilde{p}_m(t)$ по вероятности эквивалентна условию

$$\lim_{t\to\infty} \sum_{\tau=1}^{t} \omega_\tau^2(t) = 0$$

Следствие

При усреднении с экспоненциально убывающими весами и использовании «скользящего окна» последовательность $\tilde{p}_m(t)$ не сходится по вероятности