УДК 004.056.5

РОБАСТНОСТЬ ОЦЕНОК ОТНЕСЕНИЯ ОБЪЕКТОВ К КРИТИЧЕСКИ ВАЖНЫМ

Д.А. КОМЛИКОВ, А.В. ПОТАПОВИЧ*

Военная академия Республики Беларусь
Минск–57,220057, Беларусь

**Белорусский государственный университет информатики и радиоэлектроники*
П. Бровки, 6, Минск, 220013, Беларусь

Поступила в редакцию 27 сентября 2011

Дан анализ параметров и показателей, по которым объекты относятся к критически важным. Для метода учета остаточных рисков исследуется устойчивость и робастность при отнесении объектов к критически важным. Получены расчетные выражения для вычисления функции влияния.

Ключевые слова: критически важные объекты, информационная безопасность, функция влияния.

Введение

Ограниченные ресурсы, которые государство может направить на защиту информации, требуют новых организационных решений и подходов в области информационной безопасности. Таким образом, приоритетной является задача выделения из всей совокупности объектов тех, которые по тем или иным условиям требуют специального, отличного от других подхода. Такие объекты будем считать критически важными объектами (КВО).

Характеризуя особенности обеспечения информационной безопасности на КВО, необходимо учитывать следующие обстоятельства:

- основные принципы обеспечения информационной безопасности КВО аналогичны принципам, применяемым для обычных объектов;
- государственный контроль за мерами обеспечения информационной безопасности таких объектов должен осуществляться на всех уровнях управления.

При обеспечении информационной безопасности основное внимание следует уделить заданной надежности функционирования (т.е. обеспечения доступности услуг и информации уполномоченным пользователям, а также целостности системы и информации), и затем – конфиденциальности информации. Данные ресурсы могут защищаться в рамках правовых режимов, например, режима государственных секретов, или организационных режимов, обеспечивающих безопасность информации, циркулирующей в рамках информационной инфраструктуры. Как организационный режим, обеспечивающий безопасность информации, циркулирующей в рамках информационной инфраструктуры, может рассматриваться режим обеспечения безопасности информации КВО, в котором основным охраняемым качеством является не конфиденциальность, а доступность и целостность информации.

Одним из подходов в отнесении объектов к критически важным, является метод учета остаточных рисков, использующий статистические оценки рисков на объекте [1]. Для этих оценок одним из важнейших параметров является их устойчивость, робастность.

Робастность в статистике предоставляет подходы, направленные на снижение влияния выбросов и других отклонений в исследуемой величине от моделей, используемых в классических методах статистики. На практике наличие в выборках даже небольшого числа резко выделяющихся наблюдений способно фатально повлиять на результат статистического исследования (примером может служить метод наименьших квадратов или метод максимального правдоподобия), а характеристики, получаемые в результате статистических исследований не будут отражать фактическое положение при проведении исследования. Для того, чтобы избежать подобных неприятностей, необходимо сформировать оптимальную модель, в которой необходимо максимально учесть влияние «плохих» наблюдений, либо вовсе исключить их.

Для того чтобы ограничить влияние несущественных отклонений или выбросов, либо вовсе исключить это влияние, существует множество различных подходов. Среди них можно выделить два основных:

- группировать данные на интервалах различной длины, не отбраковывая отдельные наблюдения: таким образом значительно снизить возможность ухудшения выборки отдельными выпадами.

- отслеживать выбросы непосредственно в процессе анализа и для определения параметров закона распределения использовать итерационную процедуру с усеченными или th-сниженными М-оценками.

Теоретический анализ

Группирование данных как метод робастной статистики.

Посредством группирования выборки можно резко снизить влияние отдельных наблюдений, не отбрасывая их. Разбиение на интервалы не представляет особых трудностей и дает весьма ощутимый результат. Существует три наиболее распространенных способа разбиения.

1. Разбиение на интервалы равной длины

Наиболее простой и потому распространенный способ. Разбиение на интервалы равной вероятности, также называемое равночастотным группированием, что отражает практическую реализацию этого метода. В результате такого группирования выборки осуществляется максимизация величины информационной энтропии

$$\sum -P_i \ln ,$$

где и достигается наибольшая асимптотическая мощность критерия

$$P_i = \int_{x_i - 1}^{x_i} f(x)dx$$

критерия согласия, либо критерия отношения правдоподобия [2].

2. Разбиение на асимптотически оптимальные интервалы.

При таком разбиении минимизируются потери информации в результате группирования, то есть максимизируется фишеровская информация

$$\sum \left(\frac{\partial \ln P_i}{\partial \theta} \right)^2 ,$$

где θ – оцениваемый параметр закона. Для многих законов распределения удалось получить инвариантные относительно параметров границы интервалов, и были составлены соответствующие таблицы. Такое разбиение позволяет максимизировать мощность критерия.

3. Подход, основанный на функции влияния.

Рассмотрим аспекты, касающиеся оценивания параметров закона распределения по «засоренной» выборке с использованием подхода, предложенного Хампелем [3]. Для того, чтобы изучить влияние отдельно взятого наблюдения на оценку (рассматриваемую статистику) того или иного параметра закона распределения Хампелем вводится так называемая функция влияния (influence function), которая представляет собой ни что иное, как производную этой статистики.

Введем функционал *T* как функцию от некоторой выборки $X = \left(X_1 ... X_n \right) \in \mathbf{X}$ из распределения *F* с параметром $\theta \in \Theta$ (оно же F_0). *T* зависит от $X : F_0$. Значит *T* является функцией от закона F и от параметра θ. Пусть T также удовлетворяет некоторым условиям состоятельности и регулярности:

$$T\left(F\right)=\theta,\ \int TdF=0.$$

Определим производную этого функционала T в точке с распределением F следующим образом:

$$\exists a:\lim_{t\to 0}\frac{T\left(\left(1-t\right)F+tG\right)-T\left(F\right)}{t}=\int adG,$$

где a – некая функция, смысл которой прояснится на следующем шаге, а G – некий закон распределения, отличный от F.

Подставим Δ_x, приписывающую единичную массу событию $X = x$, вместо G, в результате чего от интеграла в правой части выражения останется только $a\left(x\right)$ и перепишем получившийся результат в следующем виде:

$$IF=\lim_{t\to 0}\frac{T\left(\left(1-t\right)F+t\Delta_x\right)-T\left(F\right)}{t}.$$

Эту функцию и называют функцией влияния.

Чтобы пояснить смысл введенного понятия, подставим $\frac{1}{n}$ вместо t, заменив предел. В результате выражение $F_{t,x}=\left(1-t\right)F+t\Delta_x$ преобразуется в $F_{\frac{1}{n},x}=\frac{\left(n-1\right)F+\Delta_x}{n}$ что соответствует ситуации, когда в выборку, состоящую из n – наблюдения, подчиняющихся распределению F, добавляют еще одно новое. Таким образом, IF отслеживает реакцию используемого функционала T на внесенное добавление, показывая влияние от вклада отдельного наблюдения x на оценку по всей совокупности данных.

Для характеристики влияния отдельных наблюдений также вводят понятие чувствительности к большой ошибке γ:

$$\gamma=\sup_{x\in X}\left|IF\left(x\right)\right|$$

Если функция влияния ограничена, то соответствующую оценку называют B(бэ)-робастной.

4. М-оценки.

Наиболее эффективными и широко используемыми оценками параметров законов распределений являются оценки максимального правдоподобия (ОМП), которые определяются одним из следующих условий:

$$\sum_i \ln P_i\to\max_{\theta\in\Theta},\quad \sum_i\frac{\partial\ln P_i}{\partial\theta}=0,\quad \sum_i\frac{P_i'}{P_i}=0,$$

где в случае негруппированной выборки $P_i=f\left(x_i,\ \theta\right)$, а в случае группированной – $P_i=\left(\int\limits_{x_i-1}^{x_i}f\left(x,\theta\right)dx\right)$

М-оценки – есть некое обобщение ОМП. Они определяются аналогично одним из соотношений:

$$\sum_{i=1}^{N}\rho\left(x_i,\theta\right)\to\max_{\theta\in\Theta},\quad \sum_{i=1}^{N}\Phi\left(x_i,\theta\right)=0.$$

Если наложить условие регулярности в подстановке

$$F_{t,x} = (1-t)F + t\Delta_x$$

и продифференцировать его по t в 0:

$$0 = \frac{\partial}{\partial t}\int \Phi\left(x, T\left(F_{t,x}\right)\right) dF_{t,x} ,$$

$$0 = \int \frac{\partial \Phi\left(x, T\left(F_{t,x}\right)\right) dF_{t,x}}{\partial \theta} IFdF_{t,x} + \int \Phi\left(x, T\left(F_{t,x}\right)\right) d\, \frac{\partial\left(\left(1-t\right)F + t\Delta_x\right)}{\partial t},$$

$$0 = IF\int \frac{\partial \Phi\left(x, T\left(F_{t,x}\right)\right)}{\partial \theta} dF_{t,x} + \Phi\left(x, T\left(F_{t,x}\right)\right),$$

то не представляет большого труда получить выражение функции влияния для М-оценок:

$$IF = \frac{-\Phi(x)}{\int \Phi_{\theta}^{'}\left(x\right) dF}.$$

Указанное выражение позволяет сделать вывод о том, что М-оценки эквивалентны с точностью до ненулевого множителя-константы.

Исследования

Пример функций влияния для усеченных ОМП параметров сдвига (кривая – 1) и параметра масштаба (кривая – 2) стандартного нормального закона распределения показан на рисунке.

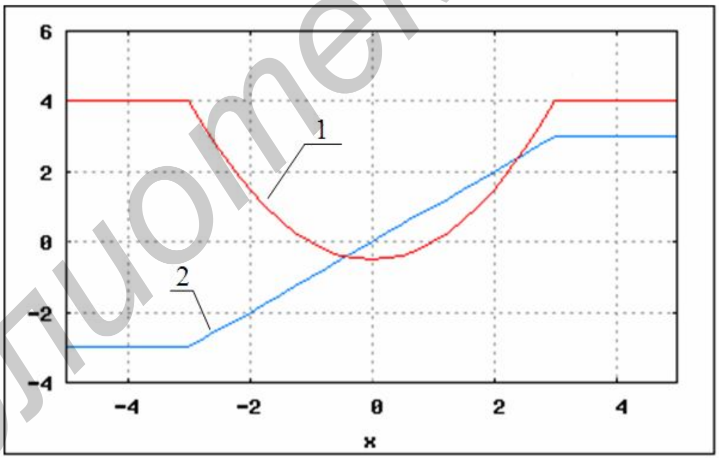

Рис. 1. Функции влияния для усеченных ОМП параметров сдвига

Несложно проверить, что для ОМП стандартного нормального закона распределения $N(0,1)$ функции влияния IF параметра сдвига и параметра масштаба выглядят соответственно:

$$IF = x, \quad IF = \frac{1}{2}x^2 - \frac{1}{2}.$$

Эти функции не ограничены, а это значит, что ОМП не является робастной в терминах В-робастности.

Для того чтобы это исправить, М-оценки искусственно ограничивают, а значит и ограничивают ее IF (см. выражение IF для М-оценок), устанавливая верхний барьер на влияние резко выделяющихся (далеко отстоящих от предполагаемых значений параметров) наблюдений. Делается это введением так называемых усеченных М-оценок, определяемых выражением:

$$\Phi_b(z)=\begin{cases}\Phi(b), & b< z\\ \Phi(z), & -b<z\le b\\ \Phi(-b), & z\le -b\end{cases}$$

где $z=\dfrac{x-\theta}{S}$ и S – оценки параметров сдвига и масштаба соответственно.

Среди усеченных М-оценок оптимальными с точки зрения B-робастности являются усеченные ОМП [2].

Процедура оценивания параметров

Чтобы решить уравнение $\sum_{i=1}^{n}\Phi(x_i,\theta)=0$, необходимо воспользоваться каким-либо численным методом. Для этого понадобится выбрать начальные приближения. Нулевым параметром сдвига обычно служит медиана, параметром масштаба – значение, кратное медиане отклонений от медианы.

Например, если необходимо оценить параметр сдвига, скажем, нормального закона распределения, можно воспользоваться методом Ньютона численного нахождения корней уравнения. В результате вся процедура нахождения параметра сводится к итеративному вычислению выражения:

$$\theta_{k+1}=\theta_k-\frac{\sum_{i=1}^{N}\Phi(x_{i,}\ \theta_k)}{\sum_{i=1}^{N}\Phi'_\theta(x_{i,}\ \theta_k)}=\theta_k-\frac{\sum_{i=1}^{N}\Phi((x_{i,}\ \theta_k)/S)}{\sum_{i=1}^{N}\Phi'_\theta((x_{i,}\ \theta_k)/S)}=\theta_k+S\frac{\sum_{i=1}^{N}\Phi(z)}{\sum_{i=1}^{N}\Phi'_z(z)},$$

где S – некоторая оценка параметра масштаба, которая нужна для того, чтобы уравнять распределения с разными параметрами.

Заключение

Используя вышеописанные методы, можно повысить робастность оценок отнесения объектов к числу КВО, что существенно может уменьшить применение избыточного количества технических средств и объема капитальных вложений при создании КВО.

FACILITY ESTIMATE ROBUSTNESS OF THE ASSIGNMENT THEM TO CRITICAL IMPORTANT ONES

D.A. KOMLIKOV, A.V. POTAPOVICH

Abstract

The analysis of parameters and indicators by which the facilities can be assigned to the critical ones is given. The stability and robustness of residual risks accounting for the critical facilities are investigated. The expressions for calculation of the influence function are obtained.

Литература

1. *Макаров О.С., Орлов А.В., Тепляков А.А.* Формирование требований к нормативному регулированию параметров критически важных объектов. Управление защитой информации. Минск, 2009.
2. *Кендалл М., Стьюарт А.* Статистические выводы и связи. М., 1973.
3. *Хампель Ф., Рончетти Э., Рауссеу П. и др.* Робастность в статистике. Подход на основе функций влияния. = Robust statistics: the approach based on influence functions. M., 1989.
4. *Кендалл М., Стьюарт А.* Статистические выводы и связи. М., 1973.