УДК 519.8+343:004
DOI 10.34822/1999-7604-2021-1-71-75

КОМПРОМИССНАЯ ПАРЕТОВСКАЯ ОЦЕНКА ПАРАМЕТРОВ РЕГРЕССИОННОЙ МОДЕЛИ УЩЕРБА ОТ КИБЕРПРЕСТУПЛЕНИЙ

С. И. Носков
Иркутский государственный университет путей сообщения, Иркутск, Россия
E-mail: sergey.noskov.57@mail.ru

В работе построена регрессионная модель влияния валового внутреннего продукта США, а также количества утечек данных и вскрытых записей на ущерб от киберпреступлений. Параметры этой модели представляют собой компромиссную паретовскую оценку в задаче минимизации векторной функции потерь. При идентификации параметров модели использованы методы наименьших квадратов и минимизации векторной функции потерь. С целью повышения адекватности модели задействована комбинированная переменная, обобщающая показатели, характеризующие похищенную информацию. В качестве информационной базы использованы статистические данные США за 2005–2019 гг.

Ключевые слова: утечки данных, вскрытые записи, ущерб от киберпреступлений, регрессионная модель, множественное оценивание параметров.

COMPROMISE PARETO'S EVALUATION OF PARAMETERS FOR REGRESSION MODEL OF DAMAGE BY CYBERCRIMES

S. I. Noskov
Irkutsk State Transport University, Irkutsk, Russia
E-mail: sergey.noskov.57@mail.ru

The paper builds a regression model of the influence of the US gross domestic product, as well as the number of data breaches and exposed records, on the damage from cybercrimes. The parameters of this model represent a compromise Pareto estimate in the problem of minimizing the vector loss function. In identifying the parameters of the model, the method of least squares and minimization of the vector loss function are used. To improve the adequacy of the model, a composite variable is used that summarizes indicators characterizing the stolen information. The US statistical data for 2005–2019 is used as an information base for the model.

Keywords: data breaches, exposed records, cybercrime damage, regression model, multiple parameter estimation.

Введение

В настоящее время кибератаки становятся все более серьезной угрозой безопасности не только для отдельных компаний и физических лиц, но и для государств. Так, в [1] отмечается, что за период с 2016 по 2018 г. в мире наблюдался рост доходов киберпреступников, который к 2020 г. достиг 2 трлн долл. В России постоянно возрастает количество киберпреступлений: с 66 000 в 2016 г. до 206 000 в 2018 г. Ущерб, нанесенный киберпреступностью в 2018 г. в России, оценивается в 2,2 млрд долл. Согласно проведенному в [2] исследованию, каждую секунду 18 пользователей старше 18 лет становятся жертвами киберпреступности. Средний урон, наносимый одной кибератакой на среднестатистического пользователя, составляет 197 долл. При этом осенью 2018 г. в интернете было зарегистрировано 73,4 млн пользователей старше 18 лет, что составляет более 56 % всего взрослого населения страны. Следует отметить неэффективность мер, предпринимаемых против киберпреступности на государственном уровне. Так, в соответствии с приведенными в [3] данными, несмотря на значи-

тельный рост числа уголовных дел в 2019 г. (248 тыс., т. е. на 60 % больше, чем в 2018 г.), расследование более 70 % из них приостановлено из-за неустановления лиц, подлежащих привлечению в ответственности.

Для анализа тенденций в области кибербезопасности, в частности влияния различных факторов на ущерб от преступлений в информационной сфере, применяются математические методы и методы моделирования, например: нелинейная математическая модель для изучения влияния вредоносного объекта на иммунный ответ компьютерной сети [4]; модель совершения киберпреступлений с использованием вредоносных кодов [5]; модель анализа потерь пользователя от киберпреступлений [6]; открытая регрессионная рекурсивная модель динамики структурных факторов киберугроз, позволяющая решать широкий круг проблем анализа и прогнозирования [7]; а также математическая модель динамики компьютерных преступлений, основанная на статистическом материале Иркутской области [8].

Идентификация параметров модели с помощью метода наименьших квадратов

В качестве подхода к моделированию влияния на величину ущерба от киберпреступлений факторов, характеризующих различные формы утечки информации, применим регрессионный анализ, основная задача которого [9–11] состоит в оценивании параметров линейного уравнения:

$$y_k = \sum_{i=1}^{m} a_i x_{ki} + \varepsilon_k \,,\ k = \overline{1,n}\,, \tag{1}$$

где y_k – значение зависимой переменной в k-ом наблюдении; x_{ki} – k-ое значение i-ой независимой переменной; a_i – i-ый, подлежащий оцениванию параметр; ε_k – ошибки аппроксимации; k – номер наблюдения; n – число наблюдений (длина выборки).

Уравнение (1) в векторной форме имеет вид:

$$y = Xa + \varepsilon\,, \tag{2}$$

где $y = (y_1, ..., y_n)^T$, $a = (a_1, ..., a_m)^T$, $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)^T$, $X - (n \times m)$ – матрица с компонентами x_{ki}.

Наиболее популярным и простым методом оценивания неизвестных параметров уравнения (2) является метод наименьших квадратов (МНК), состоящий в минимизации функции потерь:

$$I_2(\alpha) = \sum_{k=1}^{n} \varepsilon_k^2.$$

В качестве анализируемых в модели факторов введем следующие переменные:

y – ущерб от киберпреступлений в США (млн долл.);

x_1 – валовый внутренний продукт (ВВП) США (млрд долл.);

x_2 – количество утечек данных (data breaches) в США (млн);

x_3 – количество вскрытых записей (exposed records) в США (млн).

При этом будем использовать статистическую информацию по выделенным переменным за 2005–2019 гг. [12–13].

Попытка прямого использования МНК при построении уравнения (2) со свободным членом для данных переменных оказалась неуспешной вследствие низких значений критериев адекватности и несоответствия знаков параметров их смыслу. Поэтому введем комбинированную переменную x_{23} по правилу

$$x_{23} = x_2\, x_3.$$

Она представляет собой обобщенную характеристику похищенной информации. Использование мультипликативных конструкций, подобных x_{23} и представляющих собой

произведения исходных независимых переменных, является часто применяемым в регрессионном анализе приемом, направленным на улучшение качества модели [8]. В результате построенное с помощью МНК уравнение примет вид:

$$y = –4001{,}23 + 0{,}292\, x_1 + 0{,}0011\, x_{23} \qquad (3)$$

$$R = 0{,}86,\ F = 35{,}59,\ T = (–4{,}25,\ 4{,}79,\ 1{,}15).$$

Здесь R – коэффициент множественной детерминации; F – критерий Фишера; T – вектор значений критерия Стьюдента оценок параметров уравнения (3). Значения критериев адекватности уравнения указывают на его высокое качество. При этом, как следует из модели (3), величина ущерба от киберпреступлений растет с увеличением как ВВП, так и количества утечек данных и вскрытых записей, что вполне согласуется со смыслом этих переменных. Свободный член уравнения не несет смысловой нагрузки и выполняет сглаживающую функцию.

Вместе с тем первичный статистический анализ исходной информации показывает, что она содержит так называемые выбросы – наблюдения, слабо согласующиеся со всей выборкой в целом. Это обстоятельство предопределяет целесообразность применения к исследуемым данным методов оценивания параметров, реагирующих на подобные ситуации.

Идентификация параметров модели с помощью метода минимизации векторной функции потерь

В рамках регрессионного анализа разработаны методы, диаметрально противоположные по характеру реагирования на выбросы [14]. Первый – метод наименьших модулей (МНМ) с функцией потерь

$$I_1(\alpha) = \sum_{k=1}^{n} | \varepsilon_k |,$$

их игнорирует, второй – метод антиробастного оценивания (МАО) с функцией потерь

$$I_\infty(\alpha) = \lim_{\mu\to\infty} \sum_{k=1}^{n} | \varepsilon_k |^\mu = \max_{k=\overline{1,n}} | \varepsilon_k |,$$

к ним тяготеет.

В работах [14–15] описан метод множественного оценивания параметров уравнения (2) с векторной функцией потерь:

$$I(\alpha) = (I_1(\alpha),\ I_\infty(\alpha)).$$

Его реализация приводит к формированию многокритериальной задачи линейного программирования, решением которой является множество паретовских оценок, характеризующихся тем, что ни одну из них нельзя улучшить по одному критерию – $I_1(\alpha)$ или $I_\infty(\alpha)$, не ухудшив значение другого [14]. При этом все множество Парето представляет собой объединение выпуклых комбинаций паретовских вершин задаваемого указанной задачей многогранника. В таблице представлены оценки, соответствующие этим вершинам и полученные автором с помощью программы МОРМ [15].

Таблица

Паретовские оценки параметров модели

Номер уравнения	**Оценки параметров**	$I_1(\alpha)$	$I_\infty(\alpha)$
1	–6 067 0,434 –0,0013	5 869	583,6
2	–5 996 0,427 –0.0011	5 688	598,9

Окончание таблицы

Номер уравнения	**Оценки параметров**	$I_1(\alpha)$	$I_\infty(\alpha)$
3	–4 897 0,35 0,00068	4 294	735,5
4	–4 773 0,343 0,00067	4 271	741,3
5	–4 084 0,294 0,0012	3 551	978,7
6	–4 018 0,289 0,0013	3 497	1 002
7	–3 972 0,286 0,0014	3 486	1 006
8	–2 138 0,168 0,0023	3 129	1 461
9	–2 095 0,164 0,0024	3 121	1 476

Отметим, что, как следует из таблицы, МНК-оценка не является паретовской. Для выделения единственной из всего множества паретовских оценок воспользуемся приемом построения компромиссной по отношению к МНМ и МАО паретовской оценки [14]. В этом случае уравнение (3) преобразуется к виду:

$$y = –4156{,}66 + 0{,}299x_1 + 0{,}0012x_{23}.$$

Оно может быть эффективно использовано для решения задач анализа и прогнозирования ущерба от киберпреступлений.

Заключение

Дальнейшие исследования, посвященные анализу характера влияния различных факторов на ущерб от кибепреступлений математическими методами, следует проводить в направлении расширения круга этих факторов, а также расширения арсенала приемов моделирования.

В частности, вызывает интерес использование, наряду с линейными, различных нелинейных конструкций, значительно снижающих ошибки аппроксимации регрессионных моделей.

Литература

1. Дементьева М. А., Лихачева В. В., Козырев Т. Г. Киберпреступления в банковской сфере Российской Федерации: способы выявления и противодействия // Экономич. отношения. 2019. Т. 9, № 2. С. 1009–1020.

2. Анискин С. С., Селедцов В. Ю. Кибербезопасность как один из трендов цифровой экономики России // Образование и наука без границ: социал.-гуманитар. науки. 2019. № 12. С. 28–31.

3. Рогоза А. А., Сабиров В. Д., Лаптева А. В. Меры по борьбе с киберпреступностью в России // Экономич. исследования и разработки. 2020. № 12. С. 68–72.

4. Saini D. A Mathematical Model for the Effect of Malicious Object on Computer Network Immune System // Applied Mathematical Modelling. 2011. No. 8. P. 3777–3787.

5. Давыдов И. В., Шелупанов А. А. Формализация модели совершения киберпреступлений, совершаемых с использованием вредоносных кодов // Известия Томск. политех. ун-та. 2006. Т. 309, № 8. С. 126–129.

6. Швырев Б. А. Модель величины потерь пользователя от киберпреступлений // Финанс. экономика. 2018. № 3. С. 103–104.

7. Носков С. И., Вергасов А. С. Регрессионная модель структурных факторов киберугроз // Программная инженерия. 2020. № 4 (11). С. 251–256.

8. Глухов Н. И., Носков С. И., Попов П. Ю. Математическая модель динамики компьютерных преступлений // Информ. технологии и матем. моделирование в управлении сложными системами. 2020. № 1. С. 1–8.

9. Айвазян С. А., Бродский Б. Е. Макроэконометрическое моделирование: подходы, проблемы, пример эконометрической модели российской экономики // Прикладная эконометрика. 2006. № 2. С. 85–111.

10. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. М. : Диалектика, 2007. 911 с.

11. Носко В. П. Эконометрика. Элементарные методы и введение в регрессионный анализ временных рядов. М. : Фонд «Ин-т эконом. политики им. Е. Т. Гайдара», 2004. 501 с.

12. Annual number of data breaches and exposed records in the United States from 2005 to 1st half 2020. URL: https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/ (дата обращения: 18.02.2021).

13. ВВП (в текущих долларах США) – Russian Federation. Данные по национальным счетам Всемирного банка и файлы данных по национальным счетам ОЭСР. URL: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locale=ru&locations=RU (дата обращения: 18.02.2021).

14. Носков С. И. Компромиссные паретовские оценки параметров линейной регрессии // Матем. моделирование. 2020. Т. 32, № 11. С. 70–78.

15. Носков С. И., Базилевский М. П. Построение регрессионных моделей с использованием аппарата линейно-булевого программирования. Иркутск : ИрГУПС, 2018. 176 с.