- генерация набора случайных данных с помощью встроенного генератора на основе описанных в документации синтаксических конструкций и копирование готового кода;
- графическое отображение полученного набора данных на основе выбора встроенных объектов из перечня, вычисление и копирование готового кода;
- символьные вычисления на основе выбора нужных функций и условий синтаксиса;
- визуализация и копирование готового кода;
- добавление в графическое отображение различных элементов;
- преобразование графического отображения в интерактивную модель на основе определения элементов управления, набора данных, изменяемых переменных, начальных значений, диапазонов изменений значений; вычисление и копирование готового кода;
- оформление интерактивной модели и добавление опции инициализации;
- преобразование интерактивной модели в формат CDF.

В Интернете представлены сотни интерактивных моделей, полученных в результате символьных вычислений в Mathematica [6]. Данные модели могут быть использованы в иллюстративном качестве для исследовательской или образовательной деятельности, они также могут представлять интерес в качестве объектов изучения и основы для собственного компьютерного моделирования на основе символьных вычислений.

В июне 2014 г открыт сетевой ресур Wolfram Programming Cloud (облако программирования Wolfram), который позволяет в любом браузере и с любого устройства создавать готовые CDF-документы, приложения, работать с прямым API, создавать автоматически генерируемые отчеты, отсроченные задания, веб-страницы и многое другое [7].

Хотя инструментарий основан на использовании английского языка, в Интернете имеется достаточно ресурсов для изучения Wolfram Language и Mathematica на русском языке [8].

Список использованных источников

1. Stephen Wolfram A. New Kind of Science / [Электронный ресурс]: Книга. – Электрон. изд. – Режим доступа: https://www.wolframscience.com/
2. Wolfram Language / [Электронный ресурс] – Режим доступа: http://www.wolfram.com/language/
3. Wolfram|Alpha / [Электронный ресурс] – Режим доступа: http://www.wolframalpha.com/
4. Computable Document Format (CDF) for Interactive Content / [Электронный ресурс] – Режим доступа: http://www.wolfram.com/cdf/
5. Wolfram CDF Player for Interactive Computable Document / [Электронный ресурс]: – Режим доступа: http://www.wolfram.com/cdf-player
6. Wolfram Demonstrations Project & Contributors / [Электронный ресурс] – Режим доступа: http://demonstrations.wolfram.com/
7. Wolfram Programming Cloud: Introducing a Programming / [Электронный ресурс] – Режим доступа: www.wolfram.com/programming-cloud/
8. Ресурсы для изучения Wolfram Language (Mathematica) на русском языке / [Электронный ресурс]: Статья. – Режим доступа: http://habrahabr.ru/post/244451/

УДК 519.2: 519.6

СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ МНОГОМЕРНОЙ ЭНТРОПИИ ШЕННОНА

Асп. Палуха В.Ю., д.ф.-м.н., чл.-корр. НАНБ, проф. Харин Ю.С.

Белорусский государственный университет

ВВЕДЕНИЕ

Современные средства криптографической защиты информации используют генераторы псевдослучайных последовательностей. Стойкость криптосистем зависит от того, насколько близка генерируемая последовательность по своим свойствам к равномерно распределённой случайной последовательности (РРСП). Одним из подходов к оценке качества генератора является статистическое оценивание энтропии и сравнение полученной оценки с ожидаемым значением для РРСП. В данном докладе описываются методы построения статистической оценки энтропии, а также приводятся её вероятностные свойства. Кроме того, приведён алгоритм вычисления параметров распределения вероятностей оценки.

СТАТИСТИЧЕСКАЯ ОЦЕНКА ЭНТРОПИИ

Будем рассматривать стационарные в узком смысле двоичные последовательности $\{x_t\} \in V = \{0,1\}$ на некотором вероятностном пространстве (Ω, *F*, *P*). Пусть $p_{i_1,\dots,i_s} = P\{x_{t+1} = i_1, \dots, x_{t+s} = i_s\}$ – распределение вероятностей *s*-граммы $(x_{t+1},\dots,x_{t+s}) \in V_s$, которое предполагается не зависящим от $t \in \mathbb{Z}_0 = \mathbb{Z} \cup \{0\}$. Многомерная (*s*-мерная) энтропия Шеннона для фрагмента длины $s \geq 1$ равна:

Витебский государственный технологический университет

$$H\{x_1,\dots,x_s\} = h(s) = -\sum_{i_1,\dots,i_s \in V} p_{i_1,\dots,i_s} \ln p_{i_1,\dots,i_s}. \quad (1)$$

Обозначим: $i = \sum_{j=1}^{s} 2^{j-1} i_j$ – представление числа $i \in \{0,1,\dots,2^s-1\}$ в двоичной системе счисления,

$p_i(s) = P\{\sum_{j=1}^{s} 2^{j-1} x_j = i\} = p_{i_1,\dots,i_s}, \quad i = 0,\dots,2^s-1$. Тогда формула (1) примет вид

$$h(s) = -\sum_{i=0}^{2^s-1} p_i(s) \ln p_i(s). \quad (2)$$

Пусть наблюдается n фрагментов длины s: $X^{(k)} = (x_1^{(k)},\dots,x_s^{(k)}) \in V_s, \quad k = \overline{1,n}$, сформированных на основе последовательности $\{x_t\}$. Построим частотные оценки распределения вероятностей $\{p_i(s)\}, i = 0,\dots,2^s-1$:

$$\hat{p}_i(s) = \frac{v_i}{n}, \quad v_i = \sum_{k=1}^{n} \delta_{\bar{X}^{(k)},i}, \quad \bar{X}^{(k)} = \sum_{j=1}^{s} 2^{j-1} x_j^{(k)}, \quad \delta_{\bar{X}^{(k)},i} = \begin{cases} 1, \bar{X}^{(k)} = i; \\ 0, \bar{X}^{(k)} \neq i. \end{cases} \quad (3)$$

Используя подстановочный принцип, построим статистическую оценку энтропии (2) с использованием оценок (3):

$$\hat{h}(n,s) = -\sum_{i=0}^{2^s-1} \hat{p}_i(s) \ln \hat{p}_i(s). \quad (4)$$

АЛГОРИТМ ПОСТРОЕНИЯ СТАТИСТИЧЕСКОЙ ОЦЕНКИ

Как следует из вышесказанного, вычисление статистической оценки энтропии состоит из двух частей:

1) построение частотных оценок вероятностей (3);

2) вычисление оценки энтропии по формуле (4).

Опишем, как программно реализуется вычисление оценки.

На этапе построения частотных оценок вероятностей создаётся массив для хранения частот v_i объёма 2^s. Этот массив можно заполнить за один проход наблюдаемой последовательности. Временная сложность в таком случае имеет порядок O(sn). Однако необходимо затратить память порядка O(2^s). При больших s такой подход требует доработки. Для решения проблемы нехватки памяти предлагается зафиксировать некоторый порядок s_0. Вычисление частотных оценок вероятностей потребует 2^{s-s_0} проходов последовательности, каждый из которых будет состоять в следующем:

1) Для $\bar{i} \in 0,\dots,2^{s-s_0}-1$ фиксируем (s–s_0)-грамму $(i_{s_0+1}\dots i_s), \quad \bar{i} = \sum_{j=s_0+1}^{s} 2^{j-s_0-1} i_j$.

2) Вычисляем частоты

$$v_{i\bar{i}} = \sum_{k=1}^{n} \delta_{\bar{X}^{(k)},i} \delta_{\bar{Y}^{(k)},\bar{i}}, \quad \bar{X}^{(k)} = \sum_{j=1}^{s_0} 2^{j-1} x_j^{(k)}, \quad \bar{Y}^{(k)} = \sum_{j=s_0+1}^{s} 2^{j-s_0-1} x_j^{(k)}.$$

3) Вычисляем частную сумму

$$\hat{h}_{\bar{i}}(n,s) = -\sum_{i=0}^{2^{s_0}-1} \hat{p}_{i\bar{i}}(s) \ln \hat{p}_{i\bar{i}}(s), \hat{p}_{i\bar{i}}(s) = \frac{v_{i\bar{i}}}{n}. \quad (5)$$

После завершения всех проходов мы суммируем полученные частные суммы:

$$\hat{h}(n,s) = \sum_{i=0}^{2^{s-s_0}-1} \hat{h}_{\bar{i}}(n,s).$$

В итоге временная сложность пунктов 1) – 2), заключающихся в вычислении частотных оценок (3), будет иметь порядок $O(2^{s-s_0} sn)$, затрачиваемая память будет иметь порядок $O(2^{s_0})$.

В пункте 3) мы преобразуем частоты v_i в оценки вероятностей. Кроме того, нам требуется вычисление логарифма для каждой частоты. Последовательное вычисление каждого слагаемого суммы (5) требует $O(2^{s_0})$ операций. Нами предлагается вычислять слагаемые параллельно на графическом процессоре при помощи технологии CUDA [3]. Для этого необходимо скопировать в память видеокарты массив, в котором хранятся частоты v_i. Затем функция ядра графического процессора вычисляет каждое слагаемое и сохраняет их в переменную типа вектор, для чего необходимо использование библиотеки thrust [3]. Там же, на видеокарте, производится суммирование элементов с помощью операции reduce. Затем полученный

Витебский государственный технологический университет

результат передаётся на процессор. Таким образом, число операций является константным, временные затраты определяются объёмом массива частот, который копируется на видеокарту.

ВЕРОЯТНОСТНЫЕ СВОЙСТВА ОЦЕНКИ ЭНТРОПИИ

Для того, чтобы оценить качество генератора в смысле близости его по энтропийным свойствам к РРСП, необходимо знание вероятностных свойств статистической оценки. Распределение вероятностей статистической оценки многомерной энтропии описывается следующей теоремой.

Теорема. При истинной гипотезе H_* = {{x_t} есть РРСП} статистическая оценка *s*-мерной энтропии (4), построенная по подстановочному принципу, в специальной асимптотике: $n, N = 2^s \to \infty, n/N \to \lambda, 0 < \lambda < \infty$, имеет асимптотически нормальное распределение $\hat{h}(n,s) \square N(\mu, \sigma^2)$, где для параметров асимптотического распределения справедливы следующие формулы

$$\mu = \ln n - e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\ln(k+1)\lambda^k}{k!}, \tag{6}$$

$$\sigma^2 = \frac{e^{-\lambda}}{n} \sum_{k=1}^{+\infty} \frac{(k+1)\lambda^k}{k!} \ln^2(k+1) - \frac{e^{-2\lambda}}{2^s} \left(\sum_{k=1}^{+\infty} \frac{\ln(k+1)\lambda^k}{k!} \right)^2 - \frac{e^{-2\lambda}}{n} \left(\sum_{k=1}^{+\infty} \ln(k+1) \frac{\lambda^k}{k!} (k+1-\lambda) \right)^2. \tag{7}$$

ВЫЧИСЛЕНИЕ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ ОЦЕНКИ

Приведём алгоритм вычисления математического ожидания оценки по формуле (6).

Зададим максимальное число итераций *K*, вспомогательные переменные *S*, *a*, *b*.

Присвоим значения $S \leftarrow 0, b \leftarrow 1$.

$$\text{Для } k = 1, \ldots, K: \; b \leftarrow b\frac{\lambda}{k}; \; a \leftarrow b\ln(k+1); \; S \leftarrow S + a.$$

Результатом работы алгоритма является величина $\ln n - e^{-\lambda} S$.

Аналогично вычисляется дисперсия по формуле (7).

Зададим максимальное число итераций *K*, вспомогательные переменные *S*, *A*, *B*, *a*, *b*.

Присвоим значения $S, A, B \leftarrow 0, b \leftarrow 1$.

Для $k = 1, \ldots, K$:

$$b \leftarrow b\frac{\lambda}{k}; \; a \leftarrow b\ln(k+1); \; S \leftarrow S + a; \; A \leftarrow A + a(k+1)\ln(k+1); \; B \leftarrow B + a(k+1-\lambda).$$

Результатом работы алгоритма является величина $\frac{e^{-\lambda}}{n}\left(A - e^{-\lambda}\left(S^2\lambda + B^2\right)\right)$.

Выбор максимального числа итераций в обоих алгоритмах определяется параметром λ. Кратко зависимость описывается так: для малого λ необходимо меньшее число итераций.

ЗАКЛЮЧЕНИЕ

Для решения задач криптографии часто возникает необходимость статистического оценивания энтропии. В данной работе приведён алгоритм построения статистической оценки многомерной энтропии Шеннона, а также приведено асимптотическое распределение вероятностей этой оценки в случае специальной асимптотики, когда количество наблюдаемых данных и количество оцененных параметров бесконечно возрастает. Даны описания алгоритмов вычисления математического ожидания и дисперсии статистической оценки энтропии в случае справедливости гипотезы о том, что наблюдаемая последовательность является РРСП.

Список использованных источников

1. Криптология / Ю.С. Харин [и др.]. – Минск: БГУ, 2013. – 512 с.
2. Holst, L. Asymptotic normality and efficiency for certain goodness-of-fit tests / L. Holst // Biometrika. – №59, 1972. – P. 137–145.
3. CUDA Toolkit / https://developer.nvidia.com/cuda-toolkit

Витебский государственный технологический университет