Всероссийская научно-техническая конференция, посвященная 100-летию со дня рождения Б.И. Рамеева: «Информационно-управляющие и телекоммуникационные системы специального назначения». СЕКЦИЯ: «Биометрическая поддержка криптовалют и блокчейн реестров». Доклад состоится 16 мая, 2018 года, с 12^{45}-13^{00}, конференц-зал Технопарка «РАМЕЕВ», ул. Центральная, 1, г. Пенза.

Банных А.Г.

Номограмма регуляризации вычисления энтропии длинных кодов, полученная через описание бета распределением статистик расстояний Хэмминга

Аннотация.

Актуальность и цели. Целью работы является повышение устойчивости вычисления энтропии длинных кодов с зависимыми разрядами за счет возможности учета не только нормальных законов распределения значений расстояния Хэмминга.

Материалы и методы. Предложено использовать приближение реальных данных значений расстояний Хэмминга бета распределением. Показано, что наиболее устойчивым является вычисление математического ожидания и стандартного отклонения расстояний Хэмминга. В статье приводился связь этих двух статистических моментов с параметрами бета распределения. Решение системы двух нелинейных уравнений связывающих упомянутые выше параметры менее устойчиво. Программная реализация решения системы из двух нелинейных уравнений бета распределения не рационально.

Результаты. Построена номограмма, позволяющая переходить от младших статистических моментов распределения расстояний Хэмминга к параметрам бета распределения. Это позволяет вычислять вероятности ошибок второго рода для распределений расстояний Хэмминга сильно отличающихся от нормального.

Выводы. Так как по новому алгоритму нет необходимости отыскивать решения системы из двух нелинейных уравнений устойчивость вычислений в целом увеличилась. Произошла замена сложной плохо обусловленной программы итерационного поиска на простую устойчивую программу работу с таблицами и интерполяции результатов между элементами таблиц, заранее вычисленными с высокой точностью.

Ключевые слова: распределение расстояний Хэмминга, бета распределение, устойчивость вычислений, номограмма связи математического ожидания и стандартного отклонения с параметрами бета распределения.

Быстрый алгоритм вычисления энтропии в пространстве расстояний Хэмминга

Активные процессы информатизации современного общества приводят к необходимости массового использования средств криптографической защиты информации. В связи с тем, что пользователи не способны запоминать длинные пароли доступа и криптографические ключи в США и Евросоюзе развиваются технологии «нечетких экстракторов» преобразующих биометрический образ в короткий код доступа [1, 2, 3]. В России развивается технология нейросетевого преобразования биометрии в длинный код доступа или длинный код личного криптографического ключа [4, 5].

В связи с развитием нейросетевых технологий встает задача вычисления энтропии длинных кодов с зависимыми разрядами. Например, нейросеть среды моделирования «БиоНейроАвтограф» [6] дает код длинной 256 бит под применение алгоритмов отечественной криптографии. То есть актуальной является задача вычисления энтропии кодов длинной 256 бит.

Решить задачу вычисления энтропии кодов длинной 256 бит по Шеннону технически невозможно. По Шеннону придется вычислять сумму из 2^{256} слагаемых:

$$H("x_1, x_2, ..., x_{256}") = -\sum_{i=1}^{2^{256}} P_i \cdot \log_2(P_i) \qquad (1).$$

где P_i - вероятность появления одного из 2^{256} возможных состояний.

Проблема вычисления энтропии длинных кодов решается за счет перехода из пространства обычных кодов в пространство сверток Хэмминга [5, 7]:

$$h = 256 - \sum_{i-1}^{256} ("x_i") \oplus ("c_i") \qquad (2),$$

где "x_i" - значение i-го разряда свертываемого кода примеров образа «Чужой», "c_i"- значение i-го разряда кода примеров образа «Свой».

Принципиально важным является то, что свертка Хэмминга из исходных 2^{256} состояний длинного кода создает гораздо меньшее число всего в 256 состояний. Еще одним важным фактом является то, свертка Хэмминга эффективно нормализует данные. Суммирование 256 случайных слагаемых по центральной предельной теореме статистики является эффективной процедурой нормализации. Именно на эффекте нормализации построен ГОСТ Р 52633.3 [7]. По этому стандарту следует использовать порядка 100 случайно выбранных образов «Чужой». Схема численного эксперимента с использованием расстояний Хэмминга приведена на рисунке 1.

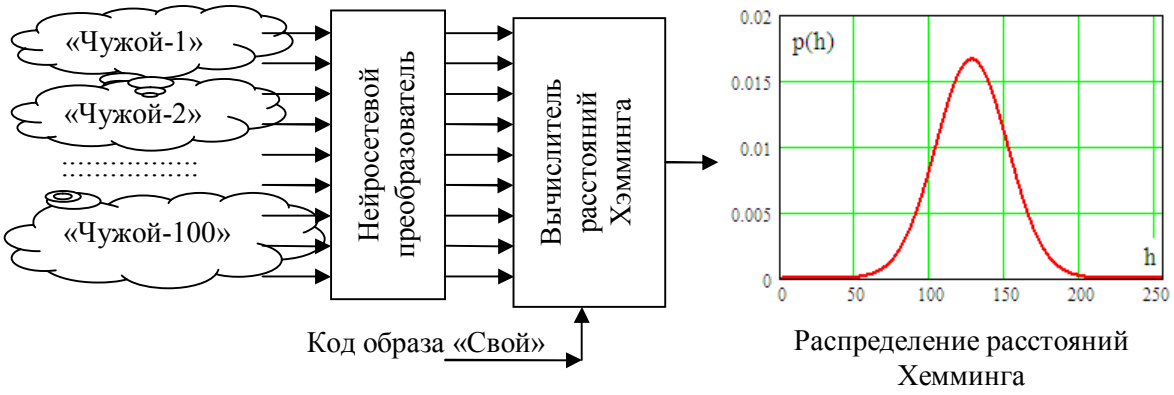

Рис. 1. Схема организации численного эксперимента по оценке вероятности ошибок второго рода (ошибочный пропуск «Чужого»)

Сотни расстояний Хэмминга вполне достаточно для вычисления достаточно надежной оценки математического ожидания - $E(h)$ и стандартного отклонения - $\sigma(h)$. По этим двум статистическим моментам мы можем оценить вероятность ошибок второго рода (ошибочное принятие «Чужого» как «Своего»).

$$P_2 = \frac{1}{\sigma(h)\sqrt{2\pi}} \int_{-\infty}^{1} \exp\left\{\frac{-(E(h)-u)^2}{2\cdot\sigma^2(h)}\right\} \cdot du \qquad (3).$$

Если бы нейросетевой преобразователь биометрия-код был «идеальным», то энтропия его выходных кодов была бы точно ровна 256 битам (по длине кода). При полностью независимых разрядах кода $E(h) = 128$, $\sigma(h) = 8$, $P_2 = 2^{-256}$. По мере увеличения корреляционных связей между разрядами кода стандартное отклонение распределения расстояний Хэмминга увеличивается. При этом энтропия кода падает, а вероятность ошибок второго рода возрастает $P_2 << 2^{-256}$. Оценка энтропии может быть выполнена через значение вероятности ошибок второго рода:

$$H("x_1, x_2, ..., x_{256}") \approx -\log_2(P_2) \qquad (4).$$

Использование аналитического описания бета распределения для обработки статистик расстояний Хэмминга

Известно, что бета распределение [8, 9, 10] может хорошо описывать не только нормальные распределения, но и иные асимметричные распределения. На рисунке 2 даны примеры нормального (симметричного) и нескольких асимметричных распределений.

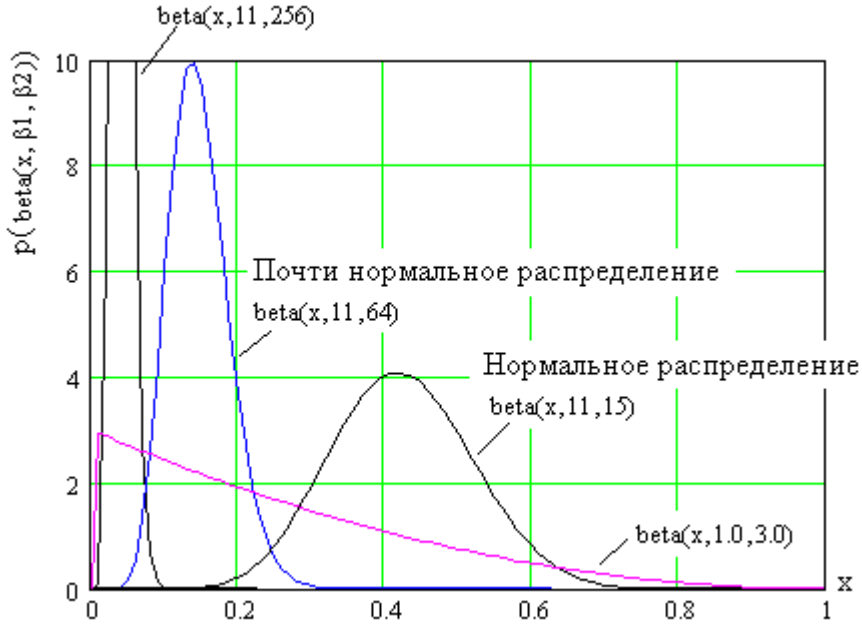

Рис. 2 Примеры симметричных и несимметричных распределений при разном соотношении двух параметрах бета распределений

Из рисунка 2 видно, что бета распределение применимо только для значений переменной в интервале от 0 до 1. В случае использования расстояний Хэмминга переменная меняется в интервале от 0 до 256. То есть нам необходимо выполнить нормировку переменных:

$$x = \frac{h}{\max(h)} \qquad (5).$$

В этом случае мы получим следующее аналитическое описание плотности распределения значений бета распределения:

$$p(beta(x,\beta 1,\beta 2)) = p(x,\beta 1,\beta 2) = \frac{(\beta 1+\beta 2+1)!}{\beta 1!\cdot\beta 2!}\cdot x^{\beta 1}\cdot(1-x)^{\beta 2} \qquad (6).$$

Из статистической теории известно, что математическое ожидание бета распределений связано с его параметрами следующим соотношением:

$$E(x) = \frac{\beta 1+1}{\beta 1+\beta 2+2} \qquad (7).$$

Стандартное отклонение бета распределения описывается иным уравнением:

$$\sigma^2(x) = \frac{(\beta 1+1)\cdot(\beta 2+1)}{(\beta 1+\beta 2+2)\cdot(\beta 1+\beta 2+3)} \qquad (8).$$

Воспользуемся уравнением (7), преобразуем его и выразим второй параметр бета распределения через первый параметр и математическое ожидание:

$$\beta 2 = \frac{\beta 1\cdot(1-E(x))-1}{E(x)} - 2 \qquad (9).$$

Подставив выражение (9) в (8) получим связь стандартного отклонения с первым параметром бета распределения:

$$\sigma(x)=\sqrt{\frac{(\beta 1+1)\cdot\left(\frac{\beta 1\cdot(1-E(x))-1}{E(x)}-1\right)}{\left(\beta 1+\frac{\beta 1\cdot(1-E(x))-1}{E(x)}\right)\cdot\left(\beta 1+\frac{\beta 1\cdot(1-E(x))-1}{E(x)}+1\right)}} \qquad (10).$$

По уравнению (10) может быть построена номограмма, связывающая между собой математическое ожидание и стандартное отклонение (рисунок 3).

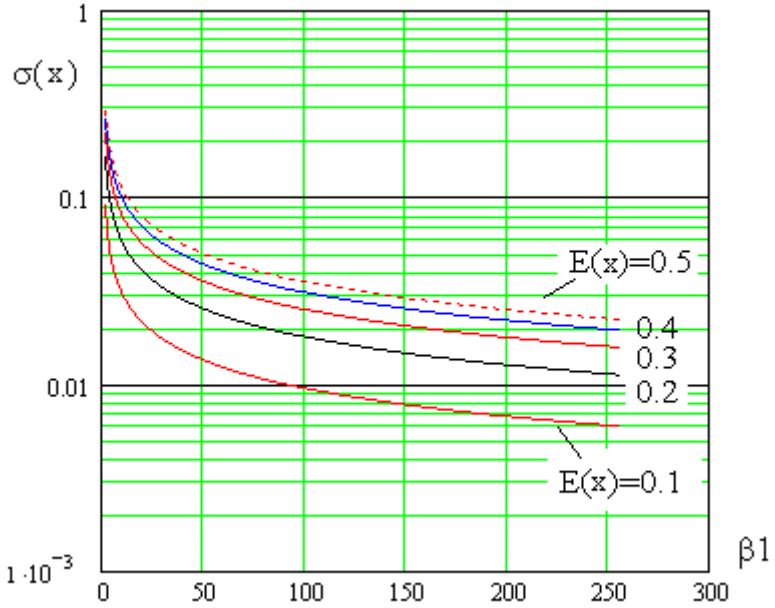

Рис. 3. Номограмма связи стандартного отклонения бета распределения с его математическим ожиданием.

Следует отметить, что номограмма рисунка 3 в реальном программном обеспечении должна быть представлена двухмерной таблицей значений первого параметра бета распределения - $\beta 1$. Второй параметр бета распределения вычисляется по формуле (9). После этого может быть оценена вероятность ошибок второго рода:

$$P_2 \approx \frac{(\beta 1+\beta 2+1)!}{\beta 1!\cdot\beta 2!}\cdot\int_0^{1/256} x^{\beta 1}\cdot(1-x)^{\beta 2}\cdot dx \qquad (11).$$

Таким образом, мы получили номограмму, пользуясь которой можно вычислять энтропию длинных кодов с зависимыми разрядами для широкого класса асимметричных распределений расстояний Хэмминга. Точность вычислений и их устойчивость может быть высокой, так как опирается на точность, заранее вычисленных таблиц (номограммы).

ЛИТЕРАТУРА:

1. Juels A., Wattenberg M. A Fuzzy Commitment Scheme // Proc. ACM Conf. Computer and Communications Security, Singapore — November 01 – 04, 1999, p. 28–36.

2. Ramírez-Ruiz J., Pfeiffer C., Nolazco-Flores J. Cryptographic Keys Generation Using FingerCodes. //Advances in Artificial Intelligence - IBERAMIA-SBIA 2006 (LNCS 4140), p. 178-187, 2006

3. Feng Hao, Ross Anderson, and John Daugman. Crypto with Biometrics Effectively, IEEE TRANSACTIONS ON COMPUTERS, VOL. 55, NO. 9, SEPTEMBER 2006.

4. ГОСТ Р 52633.0-2006 «Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации».

5. Язов Ю.К. и др. Нейросетевая защита персональных биометрических данных. //Ю.К.Язов (редактор и автор), соавторы В.И. Волчихин, А.И. Иванов, В.А. Фунтиков, И.Г. Назаров // М.: Радиотехника, 2012 г. 157 с. IBSN 978-5-88070-044-8.

6. Иванов А.И., Захаров О.С. Среда моделирования «БиоНейроАвтограф». Программный продукт создан лабораторией биометрических и нейросетевых технологий, размещен с 2009 г. на сайте АО «ПНИЭИ» http://пниэи.рф/activity/science/noc./bioneuroautugraph.zi для свободного использования университетами России, Белоруссии, Казахстана.

7. ГОСТ Р 52633.3-2011 «Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора».

8. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006 г., 816 с.

9. Королюк В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф. Справочник по теории вероятностей и математической статистике. — М.: Наука, 1985. — 640 с.

10. Иванов А.И., Безяев А.В., Елфимов А.В., Вятчанин С.Е. Корректное квантово-континуальное преобразование данных, многократно ускоряющее оценку вероятности ошибок биометрической аутентификации личности /«Специальная техника» 2017 г. № 1 с. 48-51.