Вестник КазНТУ им.К.И.Сатпаева. - 2011. - № 6(88). - С. 36-40

УДК 519.7+621.394.67

Б.С. Ахметов, А.И. Иванов, В.А. Фунтиков

СТАТИСТИЧЕСКОЕ ОПИСАНИЕ ВЫХОДНЫХ СОСТОЯНИЙ НЕЙРОСЕТЕВЫХ ПРЕОБРАЗОВАТЕЛЕЙ БИОМЕТРИЯ-КОД[1]

Введение. Преобразователь биометрия-код (ПБК) вектор из N непрерывных биометрических параметров образа «Свой» - $\overline{v}_i$ должен однозначно преобразовывать в вектор выходного кода - $\overline{c}$, состоящий из n двоичных разрядов. Если же на вход преобразователя биометрия-код поступает вектор из N непрерывных биометрических параметров - $\overline{\xi}_i$ примера образа «Чужой», то на выходе должен появиться случайный код - $\overline{z}_i$, состоящий из n двоичных разрядов. Это означает, что преобразователь биометрия-код для входных данных «Свой» и «Чужой» описываться двумя разными векторно-функциональными уравнениями дискретизации:

$$\begin{bmatrix} \text{ПБК} \\ N \times n \end{bmatrix} \cdot \overline{v}_i = \overline{c} \qquad (1)$$

$$\begin{bmatrix} \text{ПБК} \\ N \times n \end{bmatrix} \cdot \overline{\xi}_i = \overline{z}_i \qquad (2)$$

где $\begin{bmatrix} \text{ПБК} \\ N \times n \end{bmatrix}$ - матрица нелинейных нечетких (нейросетевых) функционалов, осуществляющих обогащение входных данных и их дискретизацию (преобразующих вектор входных непрерывных биометрических данных плохого качества в выходной код).

Уравнение (1) является тривиальным по сравнению с уравнением (2), так описывает единственное состояние выходного кода - $\overline{c}$. Уравнение (2) в статистическом плане является гораздо более сложным, так как описывает ПБК при воздействии на него случайных векторов биометрических параметров - $\overline{\xi}_i$, которые дают случайные выходные коды $\overline{z}_i$.

Статистическое описание идеального преобразователя биометрия-код. ГОСТ Р 52633.0-2006 формулирует общие требования к выходным кодам $\overline{z}_i$, в частности идеальный преобразователь биометрия-код должен обеспечивать полную парную независимость разрядов выходных кодов:

$$r_{m,j} \equiv 0.0 \qquad (3)$$

где m, j – номера разрядов потока выходных кодов - $\overline{z}_i$ образов «Чужой», $r_{m,j}$ - коэффициент корреляции между исследуемыми разрядами потока выходных биометрических кодов.

Кроме того, ГОСТ Р 52633.0-2006 требует обеспечить равновероятные состояния «0» и «1» в каждом разряде кодов «Чужой»:

$$P_i(\text{"0"}) = P_i(\text{"1"}) = 0.5 \qquad (4)$$

Проведенные исследования показали, что выполнить условия (3) и (4) даже по одиночке для реальных преобразователей биометрия-код технически невозможно. Для реальных преобразователей биометрия-код условия (3) и (4) выполняются приближенно и необходимо учитывать реальные значения ошибок выполнения условий (3), (4).

В том случае, если условия (3) и (4) выполняются для всех разрядов биометрических кодов преобразователь биометрия-код будем считать идеальным. Идеальные преобразователи биометрия-код наиболее просто описываются, если от вероятностей появления того или иного

[1] Статья подготовлена в рамках выполнения комплексного проекта «Разработка и подготовка производства телекоммуникационного оборудования, разработка программного сетевого, прикладного и специального обеспечения для создания цифровых сетей связи с персонализированным доступом» в соответствии с Постановлением Правительства № 218 от 09.04.2010 г.

кодового состояния перейти к вероятности появления расстояний Хэмминга от кода $\overline{c}$ до кодов $\overline{z}_i$:

$$h_i = \sum_{m=1}^{n} c_m \oplus z_{m,i} \qquad (1)$$

где i – номера кода в исследуемой последовательности;

m – номер разряда в исследуемых кодах $\overline{z}_i$;

h – расстояние Хэмминга;

$\oplus$ - операция сложения по модулю двух сравниваемых разрядов.

Вероятность появления расстояний Хэмминга «Свой»/«Чужой» для выходных биометрические коды идеального преобразователя (выполняются условия (3), (4)) описываются классическим биномиальным законом:

$$P(h) = \frac{n!}{h!(n-h)!} \cdot \{P("0")\}^h \cdot \{1-(P("0"))\}^{n-h} \qquad (5)$$

где h = 0, 1, 2, 3, …., n – возможные значения расстояний Хэмминга.

Подчеркнем, что в формуле (5) величина - P(«0») – это вероятность выпадения состояний «0» для одной и той же подбрасываемой «монеты» при численном эксперименте по схеме Бернулли. В этом смысле для идеального преобразователя биометрия-код следует воспользоваться подстановкой P(«0»)=0.5.

В том случае, если преобразователь биометрия-код является не идеальным по балансу состояний «0» и «1» в каждом выходном разряде, но все эта не идеальность одинакова для всех разрядов:

$$P_1(«0»)=P_2(«0»)= P_3(«0»)=, \quad , P_m(«0»), \ldots.., = P_n(«0») \neq 0.5 \qquad (6)$$

классическая формула (5) остается работоспособной. Это обусловлено тем, что речь идет по прежнему о реализации схемы Бернулли подбрасывания одной «монеты» с вполне определенной асимметричностью. Для вычислений достаточно найти и подставить в (5) нужное значение асимметричности P(«0»)≠0.5.

Учет индивидуальных небалансов состояний «0» и «1» в каждом из разрядов биометрического кода. Практика показывает, что каждый разряд биометрического кода имеет свой небаланс состояний:

$$P_1(«0»)=0.45; \quad P_2(«0»)=0.56; \quad , P_m(«0») = 0.3, \ldots.. ,= P_{256}(«0») =0.5 \qquad (7)$$

Для учета небалансов необходимо найти среднее геометрическое отклонение модулей от состояния P(«0»)=0.5:

$$\tilde{P}("0") = \sqrt[n]{\prod_{m=1}^{n} \{0.5 + |0.5 - P_m("0")|\}} \qquad (8)$$

Особое внимание следует обратить на то, что среднее геометрическое отклонений от состояния P(«0»)=0.5 в большую и меньшую сторону, например $P_1(«0»)=0.45$ и $P_2(«0»)=0.55$ при вычислении среднего геометрического не компенсируют друг друга. То есть любой небаланс вероятности появления состояний «0» и «1» однозначно приводит к ухудшению качества преобразователя биометрия код (к снижению энтропии выходных кодов).

Заметим, что выражение (8) может быть получено, если мы откажемся от эксплуатации классической схемы Бернулли и воспользуемся ее модификацией, учитывающей использование n разных монет с разным показателем асимметрии. При этом монеты должны быть пронумерованы. При численном эксперименте последовательность бросания монет строго соответствует их номерам. Бросание каждой монеты производится из одного положения, однако каждый бросок осуществляется со случайной высоты и при случайном значении момента закручивания монеты (при большом числе бросаний вероятности $P_m(«0»)$ и $P_m(«1»)$ дополняют друг друга $P_m(«0») + P_m(«1») = 1.0$)).

Учет влияния одинаково корреляционных связей между разрядами биометрических кодов. Еще одной проблемой является учет корреляционных связей существующих между разрядами биометрических кодов. В самом простом случае следует рассматривать значения

равной коррелированности между всеми парами разрядов кода. Эта ситуация наиболее просто моделируется программно [2]. При исследованиях моделировался преобразователь биометрия-код с 256 выходами (личный ключ формирования ЭЦП по ГОСТ Р 34.10-94 имеет длину 256 бит), в качестве единственного регулируемого параметра использовалось одинаковое значение парных коэффициентов корреляции между разрядами выходных кодов – «r». На рис. 1 приведены распределения вероятностей появления значений расстояний Хэмминга при разных значениях регулируемого параметра r = 0.05, 0.10, …., 0.48.

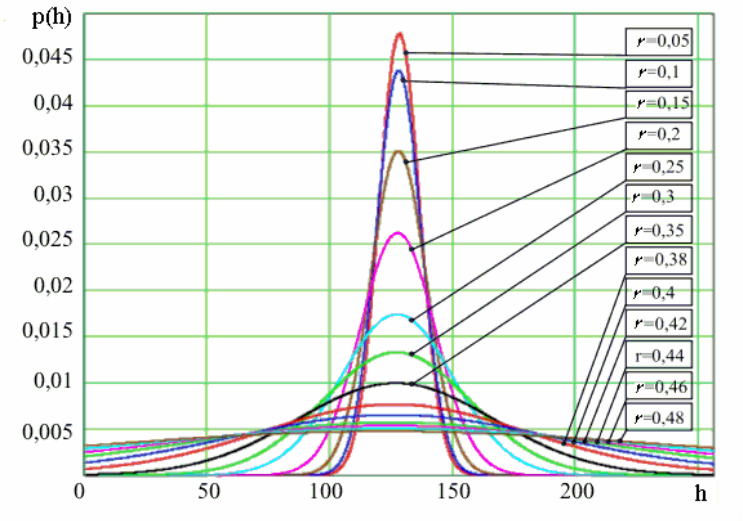

Рис. 1. Номограмма распределения расстояния Хэмминга для кодов длинной 256 бит и разных значений корреляционных связей между разрядами биометрических кодов

Точного аналитического описания двухмерной функции P(h, r) на данный момент нет, однако можно предположить, что создать такое аналитическое описание возможно и его можно построить путем введения в классический биномиальный закон (5) трех разных функций параметра – r:

$$P(h,r) = \frac{a_1(r)\cdot n!}{h!(n-h)!}\cdot\{P("0")\}^{a_2(r)\cdot h}\cdot\{1-(P("0"))\}^{a_3(r)\cdot(n-h)} \qquad (9)$$

Предположительно функции $a_1(r)$, $a_2(r)$, $a_3(r)$ являются аналитическими по крайней мере для хорошо исследованной точки длины кодов n=256.

Заметим, что для схемы получения кодов с равнокоррелированными разрядами возможно указать соответствующую модификацию схемы Бернулли. Схема сводится к расположению всех монет на противне в заданном порядке и в заданном положении «0» или «1», а также к последующему подъему листа на случайную высоту над поверхностью падения монет. Подбрасывание всех монет осуществляется одновременно (зависимо) путем резкого удара по листу снизу. Траектории движения множества монет близки и потому их конечные состояния зависимы между собой. Случайность остается только в неопределенности высоты, неопределенности силы удара по противню и неопределенности точки удара по противню.

Учет вариаций коэффициентов парной корреляции между случайно выбранными разрядами биометрических кодов. Еще одной особенностью реальных преобразователей биометрия-код является то, что значения парных коэффициентов корреляции между разрядами кодов имеют симметричное распределение относительно точки r = 0.0. Однако сам закон распределения значений коэффициентов парной корреляции существенно не нормальный (распределение имеет «тяжелые хвосты»). На рис. 2 приведен пример распределения коэффициентов корреляции реального преобразователя биометрия-код длиной 256 бит.

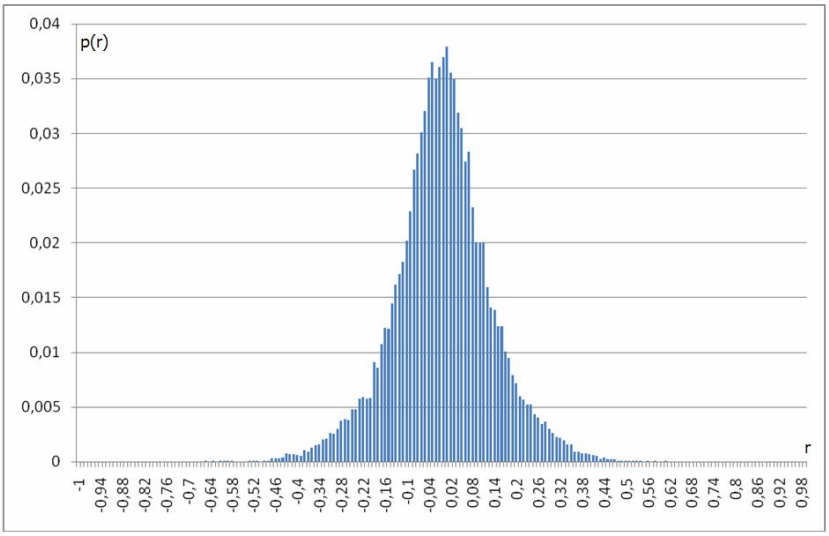

Рис. 2. Пример распределения коэффициентов парной корреляции случайно выбранных разрядов выходного кода длиной 256 бит

Из рис. 2 видно, что наиболее вероятное значение коэффициентов корреляции нулевое, все обнаруженные значения коэффициентов корреляции находятся в интервале от ±0.5. Простое усреднение коэффициентов корреляции недопустимо. Практика показала, что хорошее совпадение результатов моделирования с действительностью получается, если в качестве регулируемого параметра – r использовать математическое ожидание модулей коэффициентов парной корреляции:

$$r = E(|r_{m,j}|), \tag{10}$$

где $m \neq j$, а их значения выбираются случайно в интервале от 1 до 256,

$E(|r_{m,j}|)$ - операция вычисления математического ожидания модулей коэффициентов парной корреляции.

Заключение. Таким образом, корректное статистическое описание биометрических кодов вполне возможно. Для корректного статистического описания необходимо как минимум вычислять среднее геометрическое отклонение небалансов вероятностей состояний «0» и «1» в каждом разряде и среднее значение модулей коэффициентов парной корреляции разрядов исследуемой последовательности кодов. При применении этих усредненных параметров размерность задачи статистического описания состояний выходных кодов может быть существенно снижена. При этом для длины ключа n=256 статистическое описание в пространстве расстояний Хэмминга уже получено в виде соответствующих таблиц с изменяемыми параметрами r (шаг 0.01, интервал от 0.00 до 0.99) и P(«0») (шаг 0.01 интервал от 0.01 до 0.5).

ЛИТЕРАТУРА

Малыгин А.Ю., Волчихин В.И., Иванов А.И., Фунтиков В.А. Быстрые алгоритмы тестирования нейросетевых механизмов биометрико-криптографической защиты информации /Пенза-2006 г., Издательство Пензенского государственного университета, 161 с.

Резюме

Өзгертуші биометрия – кодының нейрожүйелік санақтық сипаттау есебі қаралады. Классикалық биномиалды бөлу заңы, тек қана жақсы түрлендіруші биометрия кодына қолданылатыны көрсетілген. Биометрияның нақты түрлендіргіштері үшін, Бернулли кестесін түрлендіріп және мәндердің биометрия-кодының дәрежесі күйлерінің ықтимал небалансы мен олардың дәрежесінің түзегендігін есепке алынуға қабілетті бөлінуінің биномиалды заңының түрлендіруін құру керек.

Summary

The problem of the statistical description of neuronet converter a biometry-code is viewed. It is shown that the classical binomial law of distribution is applicable only to ideal converters a biometry-code. For the real

converters a biometry-code it is necessary to modify the scheme of Bernulli and to create updatings of the binomial law of distribution of the values, capable to consider imbalance of the probabilities of conditions of categories of biometric codes and correlation of their categories.

Keywords. Biometric image, biometrics-code converter, modeling of dependent codes

КазНТУ им. К.И. Сатпаева *Поступила 08.09.11 г.*