

УДК: 519.2, 612.087, 621.319.7

Вятчанин С.Е., Иванов А.И.

Симметризация многомерной статистической модели сети квадратичных форм при использовании малого объема исходных биометрических данных образа «Свой».

В настоящее время нейросетевые преобразователи биометрия-код [1, 2] строятся исходя из применения в нейроне входного сумматора (линейного элемента) для обогащения входных данных перед их квантованием. Это прежде всего связано с относительной простотой и абсолютной устойчивостью используемого алгоритма обучений нейронов с линейными функционалами предварительного обогащения входных данных [3].

Однако из теории известно, что квадратичные функционалы повышения качества данных так же имеют простые и устойчивые алгоритмы обучения [4, 5, 6]. То есть в ряде биометрических приложений вместо сетей нейронов вполне могут быть использованы сети квадратичных форм.

В свою очередь, перспектива применения в будущем сетей квадратичных форм ставит вопрос о их корректном статистическом описании, построенном на малом объеме данных из 10-16 примеров образа «Свой».

Рассмотрим ситуацию, когда используется простейшая квадратичная форма с 8 входами, выбранными случайно из 416 контролируемых биометрических параметров [7]:

$$y_j = \sum_{i=1}^8 \frac{(E(v_i) - v_{i,j})^2}{(\sigma(v_i))^2} \quad (1),$$

где v_i - значения контролируемого биометрического параметра примеров образа «Свой» с нумерацией, соответствующей нумерации входов сумматора (связь номера входа сумматора – i с номером, контролируемого параметра - k с номерами от 1 до 416, задана таблицей связей); j – номер биометрического примера в обучающей или тестовой выборке.

Очевидно, что описание статистических свойств функционала (1) является тривиальным, если биометрические параметры являются независимыми. Положение меняется, когда связь, контролируемых биометрических параметров, оказывается существенной. Для зависимых биометрических данных происходит смещение моды распределения значений в левую сторону, как это показано на рисунке 1.

Отметим, что смещение вершины гистограммы (моды) в левую сторону можно наблюдать, анализируя экспериментальные данные. Так, если мы имеем сеть из 256 квадратичных форм и образ «Свой» будет представлен 16 примерами, то общий массив данных статистического анализа составит $256 \times 16 = 4096$ опытов. Этого объема вполне достаточно для гистограммы с единичными интервалами, которая приведена на рисунке 1 ($\tau \neq 0.0$).

Аналогичную гистограмму мы можем построить, привлекая имитационное моделирование. Так как мы знаем математические ожидания - $E(v_i)$ и стандартные отклонения - $\sigma(v_i)$, мы можем создать вектор независимых данных

и для него построить гистограмму $r = 0.0$. Естественно, что эти две гистограммы должны иметь существенное расхождение.

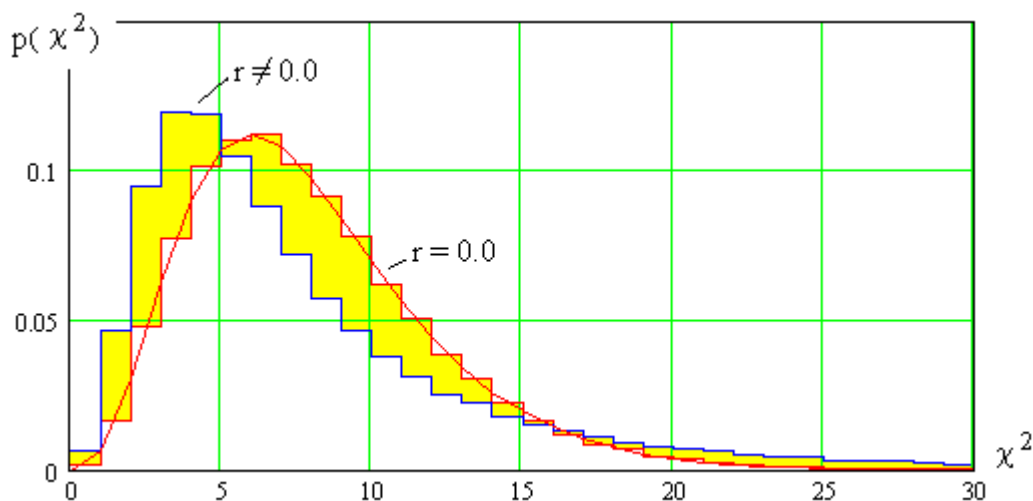


Рис. 1. Различие (отмечено заливкой) между идеальными данными без корреляции и реальными биометрическими данными со значительными корреляционными связями

На рисунке 1 расхождение гистограммы зависимых и независимых данных помечено заливкой.

Очевидно, что увеличивая корреляционные связи синтезируемых данных, мы будем получать уменьшение ошибки между площадями двух гистограмм. Добиться роста корреляционных связей между синтезированными данными удастся, если умножить вектор псевдослучайных данных на связывающую матрицу:

$$R := \begin{pmatrix} 1 & a & a & a & a & a & a & a \\ a & 1 & a & a & a & a & a & a \\ a & a & 1 & a & a & a & a & a \\ a & a & a & 1 & a & a & a & a \\ a & a & a & a & 1 & a & a & a \\ a & a & a & a & a & 1 & a & a \\ a & a & a & a & a & a & 1 & a \\ a & a & a & a & a & a & a & 1 \end{pmatrix}$$

Вне диагонали матрицы [1, 6] находятся одинаковые элементы «а», что обеспечивает равную коррелированность вектора случайных данных после умножения его на связывающую матрицу - R. В случае, когда параметр регуляризации оказывается нулевым выходные данные оказываются независимыми. Если же регулируемый параметр «а» отличен от нуля, то выходные данные связывающего преобразования $\bar{z} = R \cdot \bar{x}$ оказываются равно коррелированными.

Для осуществления многомерной статистической регуляризации биометрических данных нам достаточно осуществлять монотонное увеличение параметра регуляризации «а», добиваясь минимума ошибки расхождения гистограммы распределения реальных биометрических данных и гистограммы распределения синтезированных, равнокоррелированных данных. Точка

минимума ошибки (минимума площади заливки) даст значение коэффициента коррелированности многомерной статистической модели образа «Свой».

Можно показать, что перечисленные выше условия всегда корректны в контексте минимизации модуля ошибки расхождения двух гистограмм или дифференциального критерия Джини [8, 9]. В этом отношении можно говорить о корректности регуляризационных вычислений по дифференциальному критерию Джини.

ЛИТЕРАТУРА:

1. 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. ISBN 978-5-88070-044-8.
2. ГОСТ Р 52633.0-2006 «Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации».
3. ГОСТ Р 52633.5-2011 «Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа».
4. Саймон Хайкин. Нейронные сети: полный курс. М.: «Вильямс», 2006. — С. 1104.
5. Галушкин А.И., Цыпкин Я.З. Нейронные сети: история развития. М. Радиотехника, 2001 г., 840 с.
6. Ахметов Б.Б., Иванов А.И. Многомерные статистики существенно зависимых биометрических данных, порождаемые нейросетевыми эмуляторами квадратичных форм: Монография. Казахстан – Алматы. Из-во LEM, 2016. 86 с.
7. Иванов А.И., Захаров О.С. Среда моделирования «БиНейроАвтограф». Программный продукт создан лабораторией биометрических и нейросетевых технологий, размещен с 2009 г. на сайте АО «ПНИЭИ» <http://пниэи.рф/activity/science/noc.htm> для свободного использования университетами России, Белоруссии, Казахстана.
8. Серикова. Н.И., Иванов А.И., Качалин С.В. Биометрическая статистика: сглаживание гистограмм, построенных на малой обучающей выборке. /Вестник СибГАУ 2014 № 3(55) с.146-150
9. Серикова, Н.И. Оценка правдоподобия гипотезы о нормальном распределении по критерию Джини для сглаженных гистограмм, построенных на малых тестовых выборках/ Н.И. Серикова, А.И. Иванов, Ю. И. Серикова //Вопросы радиоэлектроники – М.: ЦНИИ «Электроника», сер. СОИУ, 2015, вып. 1, с. 85 – 94.

Статья поступила 20.08.2016, опубликована 03.09.2016 по положительной рецензии д.т.н. Малыгина А.Ю.