

6. Metodika ocenki kachestva uslug podvizhnoj radiotelefonnoj svyazi [The methodology for the quality of service assessment in telecommunication networks]. Available at: <http://minsvyaz.ru/ru/documents/4473> (accessed 01.12.2016).
7. Kachestvo svyazi v Tomske [The quality of service in telecommunication networks in Tomsk]. Available at: <http://качествосвязи.рф/sfo/tomsk.htm> (accessed 01.12.2016).
8. Shmakov D.B. Analiz kachestva pokrytiya setej svyazi standarta LTE v Tomske [The analysis of coverage quality in LTE networks in Tomsk]. *Materialy 22 MNPК "Prirodnye i intellektual'nye resursy Sibiri"*, 2016, pp. 139-142.
9. Shmakov D.B. Kompleksnaya ocenka kachestva uslug sotovoj svyazi v Tomske [The integrated assessment of quality of service in telecommunication networks in Tomsk]. *Materialy 22 MNPК "Electronic means and control systems"*, 2016, part 1, pp. 40-42.
10. Srinivasan K., Levis P. *RSSI is Under Appreciated*. Available at: <https://sing.stanford.edu/pubs/rssi-emnets06.pdf> (accessed 01.12.2016).

*Received 10.03.2017*

## ЭЛЕКТРОМАГНИТНАЯ СОВМЕСТИМОСТЬ И БЕЗОПАСНОСТЬ ОБОРУДОВАНИЯ

УДК 519.24; 519.7; 57.017

### СНИЖЕНИЕ ТРЕБОВАНИЙ К РАЗМЕРУ ТЕСТОВОЙ ВЫБОРКИ БИОМЕТРИЧЕСКИХ ДАННЫХ ПРИ ПЕРЕХОДЕ К ИСПОЛЬЗОВАНИЮ МНОГОМЕРНЫХ КОРРЕЛЯЦИОННЫХ ФУНКЦИОНАЛОВ БАЙЕСА

*Иванов А.И.<sup>1</sup>, Ложников П.С.<sup>2</sup>, Сулашко А.Е.<sup>2</sup>, Серикова Ю.И.<sup>3</sup>*

<sup>1</sup>*Пензенский научно-исследовательский электротехнический институт, Пенза, РФ.*

<sup>2</sup>*Омский государственный технический университет, Омск, РФ.*

<sup>3</sup>*Пензенский государственный университет, Пенза, РФ.*

*E-mail: ivan@pniei.penza.ru*

Рассмотрена проблема повышения надежности процедур биометрической аутентификации путем настройки сетей квадратичных форм и функционалов Байеса с использованием малой тестовой выборки. На малых тестовых выборках коэффициенты корреляции биометрических данных имеют значительную погрешность. Это препятствует их использованию при обучении (настройке) классических квадратичных форм и сетей Байеса. Показано, что многомерные корреляционные функционалы Байеса могут быть модифицированы под использование одинаково коррелированных по модулю биометрических данных. Предложено воспользоваться приемом симметризации корреляционных связей. Доказано, что в этом случае требования к объему биометрических данных существенно снижаются. Как следствие, настройка (обучение) квадратичных форм и сетей наибольшего правдоподобия Байеса становятся гораздо более устойчивыми задачами. Приведены графики, позволяющие оценить, насколько удастся снизить размеры выборки биометрических данных.

**Ключевые слова:** биометрическая идентификация, квадратичные формы, многомерные функционалы Байеса, сети квадратичных форм, сети функционалов Байеса

#### **Введение**

Одной из важнейших проблем современной науки и техники является проблема плохой обусловленности большинства вычислительных алгоритмов. Иногда эту проблему образно называют «проклятием размерности». Наиболее глубоко проблема «проклятия размерности» изучена применительно к задаче решения систем линейных уравнений или обращения матриц [1-2]. Введен

и широко используется специальный параметр контроля «качества матрицы»: число обусловленности  $\text{cond}[A]$ . Параметр  $\text{cond}[A]$  может изменяться в пределах от 1 до  $\infty$  и по своей сути является коэффициентом усиления погрешности исходных данных. При  $\text{cond}[A] = 1$  погрешность исходных данных не усиливается, при  $\text{cond}[A] > 1$  погрешность конечного результата увеличивается пропорционально числу обусловленности.

Обычно перед решением системы линейных уравнений  $[A] \cdot \bar{x} = \bar{y}$  оценивают число обусловленности матрицы  $[A]$  и вектор ошибок входных данных –  $\Delta \bar{x}$ . Это позволяет заранее оценить значение вектора ошибок конечного результата  $\Delta \bar{y} = \text{cond}[A] \cdot \Delta \bar{x}$ .

Если результирующая погрешность  $\Delta \bar{y}$  оказывается велика, существует два пути возможного ее понижения. Во-первых, можно снизить погрешность исходных данных за счет их накопления (увеличения объема выборки в  $k$  раз). Применяв метод наименьших квадратов [3], можно снизить результирующую ошибку в  $\sqrt{k}$  раз. Во-вторых, мы можем регуляризовать задачу по Тихонову [4], добившись необходимых значений вектора ошибок вычислений.

Следует отметить, что приведенные выше два пути повышения стабильности результатов матричных вычислений актуальны для любых вычислений. Любые вычисления можно улучшить, вводя избыточность (увеличивая выборку) или применив какой-либо метод регуляризации вычислений.

Особенно ярко все это видно при решении задач биометрической идентификации личности. Если при решении идти по пути классической линейной алгебры, то придется иметь дело с квадратичными формами и обращением корреляционных (ковариационных) матриц очень высокой размерности от 400 до 1000:

$$e^2 = (E(\bar{x}) - \bar{x})^T [R]^{-1} (E(\bar{x}) - \bar{x}), \quad (1)$$

где  $E(\bar{x})$  – математическое ожидание вектора контролируемых биометрических параметров в нормированной системе координат;  $[R]^{-1}$  – корреляционная матрица контролируемых биометрических параметров.

Поскольку по техническим причинам обращать корреляционные матрицы  $[R]^{-1}$  размерности более 8 не удастся, приходится вместо квадратичных форм использовать искусственные нейронные сети. Применение искусственных нейронных сетей вместо классических квадратичных форм [5-7] следует рассматривать как один из методов топологической регуляризации нерешаемой задачи. Производится топологическая замена технически нерешаемой задачи обращения корреляционных матриц высокой размерности на независимое обучение 256 искусственных нейронов. Следует отметить, что обучение сети из 256 нейронов стандартным алгоритмом [8] обладает рекордной устойчивостью и имеет линейную вычислительную сложность. Стандартизован-

ный алгоритм [8] абсолютно устойчив, так как не является итерационным (в нем исключен направленный итерационный поиск весовых коэффициентов нейронов). Фактически стандартный алгоритм обучения является полностью детерминированным, так как он однозначно вычисляет знаки и значения весовых коэффициентов нейронов, опираясь на знание вектора математических ожиданий биометрических параметров  $E(\bar{x})$  образа «Свой», вектора стандартных отклонений биометрических параметров  $\sigma(\bar{x})$  образа «Свой», а также вектора стандартных отклонений образов «Чужие».

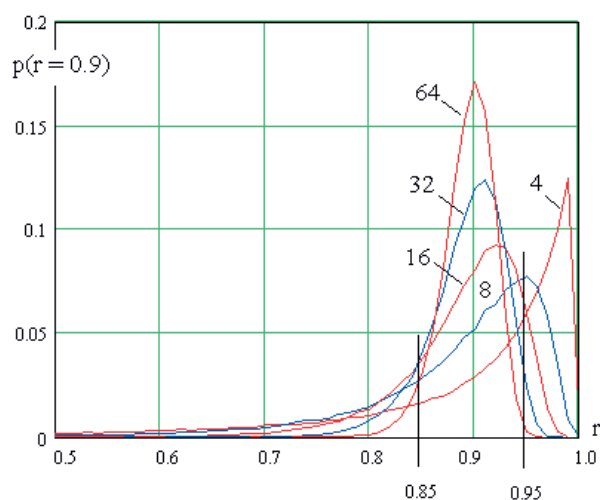


Рис. 1. Плотности распределения значений коэффициентов парной корреляции для сильно зависимых данных  $E(r) = 0,9$  при  $n = 4, 8, 16, 32, 64$

### Проблема точного вычисления коэффициентов корреляции на малых выборках

Классический коэффициент парной корреляции двух биометрических параметров вычисляется по следующей формуле:

$$r(x_1, x_2) = \frac{1}{n} \sum_{i=1}^n \frac{(E(x_1) - x_{1,i}) (E(x_2) - x_{2,i})}{\sigma(x_1) \sigma(x_2)}, \quad (2)$$

где  $E(\cdot)$  – операция вычисления математического ожидания;  $\sigma(\cdot)$  – операция вычисления стандартного отклонения;  $n$  – размер выборки.

На малых выборках биометрических данных возникают значительные ошибки вычисления математических ожиданий  $\Delta E(x_1)$ ,  $\Delta E(x_2)$  и стандартных отклонений  $\Delta \sigma(x_1)$ ,  $\Delta \sigma(x_2)$ . Совершенно так же, как и при решении систем линейных уравнений возникает эффект накопления ошибок. Оценить возможные значения ошибок вычисления коэффициентов парной корреляции можно по распределениям, приведенным на рис. 1.

Из рис. 1 видно, что при размере выборки  $n = 4$  в значения коэффициента корреляции попадают в интервал от 0,85 до 0,95 с вероятностью 0,35. Однако если размер тестовой выборки увеличить до 8; 16; 32; 64, происходит рост вероятности попадания вычисляемых значений в заданный интервал до величин 0,59; 0,77; 0,89; 0,96.

### Симметризация корреляционных связей как способ получения эквивалентной корреляционной матрицы с заведомо плохой обусловленностью

Известно [7; 9], что многомерные биометрические данные «Свой» всегда можно симметризовать, то есть заменить иными многомерными данными с той же самой энтропией. После симметризации корреляционная матрица биометрических данных вне диагонали будет иметь одинаковые коэффициенты парных корреляционных связей. Если биометрические данные сильно коррелированы, то мы получим вне диагонали коэффициенты парной корреляции близкие к единице. Ситуация, когда все коэффициенты корреляционной матрицы вне диагонали близки к единице, приводит к очень плохой обусловленности матриц квадратичных форм. В частности, для данных, соответствующих рис. 1, пятимерная корреляционная матрица будет иметь большое число обусловленности:

$$\text{cond} \begin{bmatrix} 1 & 0,9 & 0,9 & 0,9 & 0,9 \\ 0,9 & 1 & 0,9 & 0,9 & 0,9 \\ 0,9 & 0,9 & 1 & 0,9 & 0,9 \\ 0,9 & 0,9 & 0,9 & 1 & 0,9 \\ 0,9 & 0,9 & 0,9 & 0,9 & 1 \end{bmatrix} = 46; \quad (3)$$

$$\bar{\lambda} = \begin{bmatrix} 4,6 \\ 0,1 \\ 0,1 \\ 0,1 \\ 0,1 \end{bmatrix}. \quad (4)$$

Ошибки, возникающие при вычислении коэффициентов корреляции при попытках применения квадратичных форм (1), будут усиливаться примерно в 46 раз, что недопустимо для практики. Неприятности, порождаемые симметричными корреляционными матрицами, обусловлены тем, что число обусловленности определяется как отношение максимального и минимального собственных чисел  $\lambda_i$  используемой матрицы

$$\text{cond}[R] = \frac{\max(\lambda_i)}{\min(\lambda_i)}. \quad (5)$$

Как видно из выражения (4), собственные числа симметричной корреляционной матрицы (3) имеют одно большое значение  $\lambda_1 = 4,6$  и четыре одинаково малые значения собственных чисел  $\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 0,1$ . Симметричные корреляционные матрицы являются наихудшим вариантом для квадратичных форм (1), так как они не позволяют выделить и отбросить пару наиболее зависимых биометрических параметров.

### Многомерные симметричные корреляционные функционалы Байеса

Полная симметрия корреляционных связей, являющаяся наихудшей для квадратичных форм, оказывается наилучшим соотношением данных для функционалов наибольшего правдоподобия Байеса [10, 11]. Следует отметить, что классический двухмерный коэффициент корреляции (2) по своей сути есть не что иное, как одна из форм записи правила Байеса:

$$P(x_2/x_1) \cdot P(x_1) = P(x_1/x_2) \cdot P(x_2). \quad (6)$$

В этом легко убедиться, рассматривая предельные значения коэффициентов корреляции:

$$\begin{cases} r(x_1, x_2) = 1 \Rightarrow P(x_1) = \\ = P(x_2) \vee P(x_1/x_2) = P(x_2/x_1) = 1; \\ r(x_1, x_2) = 0 \Rightarrow P(x_1, x_2) = P(x_1) \cdot P(x_2). \end{cases}$$

Для интервала корреляционных связей от 0,7 до 1,0 корреляционный вариант формулы Байеса можно записать следующим образом:

$$r(x_2, x_1) \cdot P(x_1) \approx r(x_1, x_2) \cdot P(x_2). \quad (7)$$

Приближение (7) становится точным равенством только при предельно высоких значениях коэффициентов парной корреляции. В принципе, можно построить некоторые корреляционные функции  $F(r(x_1, x_2); P(x_2))$ ;  $F(r(x_2, x_1); P(x_1))$ , которые приближение (7) сделают равенством при всех больших значениях модулей коэффициентов корреляции. Для нас важно то, что коэффициенты парных корреляций функционально связаны с правилом Байеса и, соответственно, на их базе может быть создано множество решающих правил Байеса разной размерности.

Одним из самых простых правил является использование равнокоррелированных биометрических параметров с суммированием близких значений, вычисленных на реальных данных ко-

эффициентов парной корреляции. В частности, может быть использован симметричный корреляционный функционал Байеса третьего порядка:

$$R(x_1, x_2, x_3) = \frac{r(x_1, x_2) + r(x_1, x_3) + r(x_2, x_3)}{3}. \quad (8)$$

Решающее правило для такого функционала строится в виде правого и левого порога для его допустимых значений от  $\min(R(x_1, x_2, x_3))$  до  $\max(R(x_1, x_2, x_3))$ . Определение на реальных биометрических данных образа «Свой» допустимых порогов  $\min(R(x_1, x_2, x_3))$ ,  $\max(R(x_1, x_2, x_3))$  является настройкой (обучением) трехмерных решающих правил Байеса.

Очевидно, что по аналогии с трехмерным корреляционным функционалом Байеса может быть использован четырехмерный симметричный корреляционный функционал Байеса

$$R(x_1, x_2, x_3, x_4) = \frac{r(x_1, x_2) + r(x_1, x_3) + r(x_1, x_4)}{6} + \frac{r(x_2, x_3) + r(x_2, x_4)}{6} + \frac{r(x_3, x_4)}{6}. \quad (9)$$

В общем случае  $n$ -мерный симметричный корреляционный функционал Байеса имеет следующее формальное описание:

$$R(x_1, x_2, \dots, x_n) = \{r(x_1, x_2) + r(x_1, x_3) + \dots + r(x_1, x_n) + r(x_2, x_3) + \dots + r(x_2, x_n) + \dots + r(x_{n-1}, x_n)\} \cdot \frac{1}{(n-1) + (n-2) + \dots + (n-n)}. \quad (10)$$

Нетрудно заметить, что корреляционные функционалы Байеса третьего порядка построены на усреднении трех коэффициентов парной корреляции, каждый из которых имеет случайную ошибку, обусловленную малой выборкой биометрических данных [12]. Так как происходит усреднение трех коэффициентов парной корреляции, случайные составляющие ошибок каждого из частных коэффициентов корреляции ослабляются в  $\sqrt{3}$  раз в значениях результирующего трехмерного корреляционного функционала (8).

Четырехмерный корреляционный функционал подавляет случайную составляющую в  $\sqrt{6}$  раз. При применении  $n$ -мерного корреляци-

онного функционала происходит подавление случайных составляющих ошибки примерно в  $\sqrt{(n^2 - n)/2}$  раз. С ростом размерности решающих правил мы наблюдаем почти линейный рост подавления случайных ошибок, возникающих из-за малого размера обучающей (тестовой) выборки.

### Численный эксперимент по оценке эффекта подавления случайных погрешностей, вычисления коэффициентов парных корреляции на малых выборках

Эффект подавления случайных составляющих погрешности с ростом размерности корреляционных функционалов Байеса можно оценить численным моделированием. Для этой цели необходимо использовать равнокоррелированные данные, получаемые умножением вектора псевдослучайных независимых чисел  $\xi$  на симметричную связывающую матрицу [7; 9] с единичной диагональю и одинаковые элементы вне диагонали:

$$\begin{bmatrix} 1 & a & \dots & a \\ a & 1 & \dots & a \\ \dots & \dots & \dots & \dots \\ a & a & \dots & 1 \end{bmatrix} \times \begin{bmatrix} \xi_{1,i} \\ \xi_{2,i} \\ \dots \\ \xi_{m,i} \end{bmatrix} = \begin{bmatrix} y_{1,i} \\ y_{2,i} \\ \dots \\ y_{m,i} \end{bmatrix} \Rightarrow R_m = \begin{bmatrix} 1 & r & \dots & r \\ r & 1 & \dots & r \\ \dots & \dots & \dots & \dots \\ r & r & \dots & 1 \end{bmatrix}.$$

В итоге мы получаем выходные данные с близкими коэффициентами корреляции (см. рис. 2). Из рис. 2 видно, что по мере увеличения размерности корреляционных функционалов Байеса монотонно падает их стандартное отклонение, что приводит к росту вероятности попадания вычисленного функционала в интервал значений от 0,85 до 0,95. Так, для классического двухмерного коэффициента корреляции выборка из 16 примеров дает вероятность 0,77 попадания значений в заданный интервал. Переход к использованию четырехмерных корреляционных функционалов позволяет повысить вероятность попадания в интервал от 0,85 до 0,95 до величины 0,84. Это эквивалентно увеличению тестовой выборки с 16 до 24 примеров (рост объема выборки на 50%).

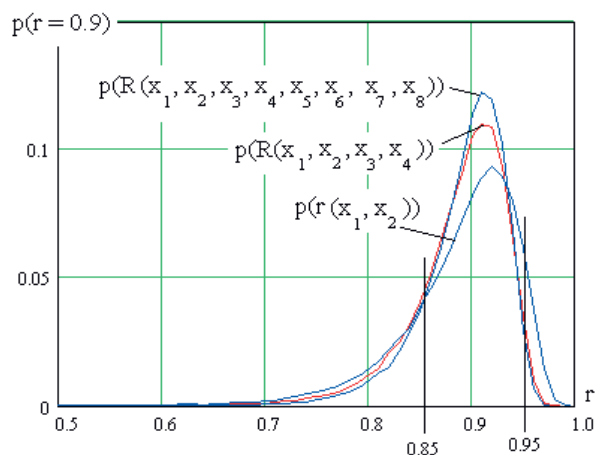


Рис. 2. Распределения значений классических двумерных коэффициентов корреляции, а также четырехмерных и восьмимерных корреляционных функционалов Байеса, вычисленных на выборках из 16 примеров

Еще больший рост объема выборки будет наблюдаться при переходе от двумерных коэффициентов корреляции к восьмимерным корреляционным функционалам. В этом случае в заданный интервал значения функционалов попадают с вероятностью 0,88, что эквивалентно росту обучающей выборки с 16 до 31 примера (рост объема выборки на 94%). При дальнейшем увеличении роста размерности вычисляемых корреляционных функционалов монотонно будет падать случайная составляющая погрешности вычислений. Это эквивалентно некоторой топологической регуляризации вычислений, осуществляемых сетями Байеса-Хэмминга [10-11].

### Практически реализуемый алгоритм симметризации задачи многомерного корреляционного анализа Байеса

Изложенные выше теоретические обоснования многомерной корреляционной обработки биометрических данных приложим к конкретной технологии распознавания рукописных образов. Ориентация именно на технологию распознавания рукописных образов обусловлена тем, что наши данные может проверить любой желающий. Только для рукописных образов любой желающий может получить самостоятельно достоверные биометрические данные в любом требуемом объеме.

Для того чтобы это было возможно, в Лаборатории биометрических и нейросетевых технологий «АО Пензенского научно-исследовательского электротехнического института» для университетов России, Казахстана и Беларуси создан сво-

бодно распространяемый программный продукт «БиоНейроАвтограф» [13]. Этот продукт анализирует 416 биометрических параметров почерка, являющихся некоторыми эквивалентами коэффициентов двумерного преобразования Фурье от пары колебаний пера  $X(t)$ ,  $Y(t)$ , возникающих во время воспроизведения человеком рукописных символов. Эти 416 параметров доступны для наблюдения, так как среда моделирования «БиоНейроАвтограф» [13] сохраняет их в файле «param.txt» после каждой инициации режима «проверить».

На рис. 3 дана гистограмма распределения значений 86 320 коэффициентов парной корреляции между 416 наблюдаемыми параметрами.

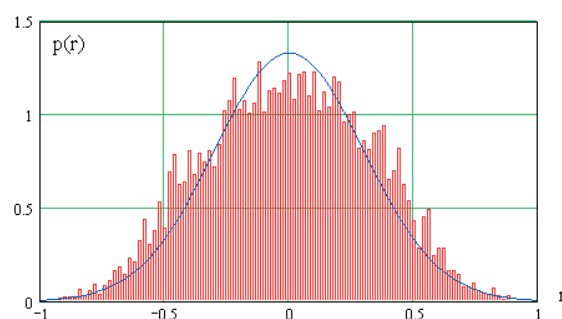


Рис. 3. Гистограмма распределения значений коэффициентов парной корреляции биометрических данных, порождаемых рукописным образом «Пенза»

Из рис. 3 видно, что гистограмма близка к нормальному закону распределения значений (сплошная тонкая линия). В рамках гипотезы нормальности распределения с нулевым математическим ожиданием и стандартным отклонением  $\sigma = 0.34$  в интервал значений от 0,85 до 0,95 коэффициенты корреляции попадают с вероятностью 0,005. Это означает, что мы можем обнаружить 430 высоких значений положительных коэффициентов корреляции и столько же отрицательных значений коэффициентов корреляции. Используя 860 значительных корреляционных связей, можно построить несколько сотен 32-мерных корреляционных функционалов Байеса [10-11].

### Заключение

Если анализировать устойчивость процедур вычисления коэффициентов корреляции и вводить для них некоторый аналог числа обусловленности, то при тестовой выборке в 16 примеров биометрического образа мы получим некоторое условное значение  $\text{cond}(r(x_1, x_2))_{16}$ . Если далее мы будем пользоваться двумерными корреляционными функционалами Байеса, то мы получим некоторое решающее правило, дополняя-

ющее классические квадратичные формы и нейросетевые преобразователи [10-11]. Как следует из материалов данной статьи, увеличение размерности корреляционных функционалов Байеса эквивалентно формальному увеличению тестовой выборки.

Применение четырехмерных корреляционных функционалов Байеса эквивалентно увеличению на 50% размеров тестовой выборки. Этот эффект можно рассматривать как значительное снижение числа обусловленности  $(r(x_1, x_2))_{16} > \text{cond}(r(x_1, x_2))_{24}$  или как значительное повышение устойчивости вычислений по отношению к случайным ошибкам, порождаемым малыми объемами тестовых выборок.

Отмеченный выше эффект повышения устойчивости вычислений наблюдается для любых биометрических данных, однако для получения этого эффекта необходимо использовать только одинаково коррелированные по модулю биометрические данные. Главное – это близость модулей коэффициентов корреляции. Многомерные корреляционные функционалы Байеса (10) легко могут быть модифицированы под использование одинаково коррелированных по модулю биометрических данных. Работа выполнена при финансовой поддержке РФФИ (грант № 16-07-01204)

### Литература

1. Форсайт Дж., Молер К. Численное решение систем линейных алгебраических уравнений. М.: Мир, 1969. – 168 с.
2. Райс Д. Матричные вычисления и математическое обеспечение. М.: Мир, 1984. – 412 с.
3. Лоусен Ч., Хенсон Р. Численное решение задач методом наименьших квадратов. М.: Наука, 1966. – 230 с.
4. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. М.: Наука, 1979. – 248 с.
5. Качайкин Е.И., Иванов А.И. Идентификация авторства рукописных образов с использованием нейросетевого эмулятора квадратичных форм высокой размерности // Вопросы кибербезопасности. № 4(12), 2015. – С. 42-47.
6. Волчихин В.И., Ахметов Б.Б., Иванов А.И. Быстрый алгоритм симметризации корреляционных связей биометрических данных высокой размерности // Известия высших учебных заведений. Поволжский регион. Технические науки. №1, 2016. – С. 3-7.
7. Нейросетевая защита персональных биометрических данных. Под ред. Ю.К. Язова. М.: Радиотехника, 2012. – 157 с.
8. ГОСТ Р 52633.5-2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа.
9. Ахметов Б.С., Надеев Д.Н., Фунтиков В.А., Иванов А.И., Малыгин А.Ю. Оценка рисков высоконадежной биометрии. Алматы: Изд-во КазНТУ им. К.И. Сатпаева, 2014. – 108 с.
10. Иванов А.И., Ложников П.С., Качайкин Е.И. Идентификация подлинности рукописных автографов сетями Байеса-Хэмминга и сетями квадратичных форм // Вопросы защиты информации. №2 (109)? 2015. – С. 28-34.
11. Иванов А.И., Качайкин Е.И., Ложников П.С., Сулавко А.Е. Биометрическая идентификация рукописных образов с использованием корреляционного аналога правила Байеса // Вопросы защиты информации. №3 (110), 2015. – С.48-54.
12. Волчихин В.И., Иванов А.И., Серикова Ю.И. Компенсация методических погрешностей вычисления стандартных отклонений и коэффициентов корреляции, возникающих из-за малого объема выборок // Известия высших учебных заведений. Поволжский регион. Технические науки. №1, 2016. – С. 45-49.
13. Иванов А.И., Захаров О.С. Среда моделирования «БиоНейроАвтограф» // URL: <http://пниэи.рф/activity/science/noc.htm> (с 2009 г. для свободного использования вузами России, Белоруссии, Казахстана).

*Получено 14.02. 2017*

**Иванов Александр Иванович**, д.т.н., начальник Лаборатории биометрических и нейросетевых технологий Пензенского научно-исследовательского электротехнического института. Тел. (8-841) 239-10-78. E-mail: [ivan@pniei.penza.ru](mailto:ivan@pniei.penza.ru).

**Ложников Павел Сергеевич**, к.т.н., заведующий Кафедрой комплексной защиты информации (КЗИ) Омского государственного технического университета (ОмГТУ). Тел. (8-381) 221-77-02. E-mail: [lozhnikov@gmail.com](mailto:lozhnikov@gmail.com)

**Сулавко Алексей Евгеньевич**, к.т.н., старший преподаватель Кафедры КЗИ ОмГТУ. Тел. (8-381) 221-77-02. E-mail: sulavich@mail.ru

**Серикова Юлия Игоревна**, студентка Пензенского государственного университета. Тел. (8-841) 239-10-78, E-mail: julia-ska@yandex.ru

## ON BIOMETRIC DATA SAMPLE SIZE REDUCTION WHEN USING BAYESIAN MULTIDIMENSIONAL CORRELATION FUNCTION

*Ivanov A.I.<sup>1</sup>, Lozhnikov P.S.<sup>2</sup>, Sulavko A.E.<sup>2</sup>, Serikova Y.I.<sup>3</sup>*

*<sup>1</sup>Scientific and Research Electrotechnical Institute of Penza, Penza, Russian Federation*

*<sup>2</sup>Omsk State Technical University, Omsk, Russian Federation*

*<sup>3</sup>Penza State University, Penza, Russian Federation*

*E-mail: ivan@pniei.penza.ru*

The problem of improving the reliability of procedures for biometric authentication is studied using small training sample for configuring of recognition machines. The present article shows that the Bayesian multidimensional correlation functionals can be modified using the identically correlated absolute values of biometric data. The correlation coefficients of biometric data have a significant error at a small test samples. It impedes their use when configuring the classical quadratic form and Bayesian networks. The authors suggest using the method of symmetrization of biometric data, i.e., replace biometric data by other multidimensional data with the same entropy. After symmetrization of correlation matrix, it may have the same coefficients of pair correlations outside the diagonal. Our findings indicate that the required sample size of biometric data is significantly reduced when using the aforementioned method. Computational error of calculating correlation coefficients is also reduced. The effect of increasing the stability of calculations is observed for any biometric data. As a result, training process of classical quadratic forms and Bayesian maximum likelihood networks becomes more stable. The research provides the data that estimate how much the size of biometric data sample can be reduced.

**Keywords:** biometric identification, quadratic form, Bayesian multidimensional functionals, quadratic forms neural network, Bayesian functionals neural network

**DOI:** 10.18469/ikt.2017.15.2.12

**Ivanov Alexander Ivanovich.** Scientific and Research Electrotechnical Institute of Penza, 440000, Russian Federation, Penza, ul. Soviet, 9, Doctor of Technical Sciences, Chief of Laboratory biometric and neural network technologies, phone: (841-2) 39-10-78, E-mail: ivan@pniei.penza.ru

**Lozhnikov Pavel Sergeevich.** Omsk State Technical University, 644050, Russian Federation, Omsk, Mira ave., 11, PhD in Technical Sciences, Head of Complex Information Protection Department, phone: 8 (3812) 21-77-02; E-mail: lozhnikov@gmail.com

**Sulavko Alexey Evgenjevich.** Omsk State Technical University, 644050, Russian Federation, Omsk, Mira ave., 11, PhD in Technical Sciences, Senior Lecturer of Complex Information Protection Department, phone: 8 (3812) 21-77-02; E-mail: sulavich@mail.ru

**Serikova Yulia Igorevna.** Penza State University, 440026, Russian Federation, Penza, ul. Red, 40, Student, phone: 8 (841-2) 39-10-78; E-mail: julia-ska@yandex.ru

### References

1. Forsythe G., Moler C. *Computer solution of linear algebraic systems*. NJ: Prentice-Hall, Englewood Cliffs, 1967. 148 p. (Russ. ed.: Il'ina V. P., Kuznetsova Y. I. Chislennoe reshenie sistem lineynykh algebraicheskikh uravnenij Moscow, Mir Publ., 1969. 168 p.).
2. Rice J. *Matrix Computations and Mathematical Software*. New York, McGraw-Hill, 1981. 248 p. (Russ. ed.: Voevodin V.V., Arushanyan O.B. Matrichnye vychisleniya i matematicheskoe obespechenie. Moscow, Mir Publ., 1984. 264 p.).

3. Lawson Ch., Hanson R. *Solving least squares problems*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1974, 230 p. (Rus. ed.: Tyrtshnikova E.E. Chislennoe reshenie zadach metodom naimen'shikh kvadratov. Moscow, Nauka Publ., 1986. 230 p.).
4. Tikhonov A.N., Arsenin V.Y. *Metody resheniya nekorrektnykh zadach* [Methods of solution of some incorrect problems]. Moscow, Nauka Publ., 1979. 248 p.
5. Kachajkin E.I., Ivanov A.I. Identifikatsiya avtorstva rukopisnykh obrazov s ispol'zovaniem nejrosetevogo ehmulyatora kvadraticnykh form vysokoj razmernosti. [Identification of authorship of handwritten Images using neural network emulator of quadratic forms high dimension]. *Voprosy kiberbezopasnosti*, 2015, no. 4, pp. 42-47.
6. Volchikhin V.I., Akhmetov B.B., Ivanov A.I. Bystryj algoritm simmetrizatsii korrelyatsionnykh svyazej biometricheskikh dannyykh vysokoj razmernosti. Izvestiya vysshikh uchebnykh zavedenij [Fast symmetrization algorithm for correlations of biometric data of high dimension]. *Povolzhskij region. Tekhnicheskie nauki*, Penza, PGU, 2016, no 1, pp. 3-7.
7. Yazov Y.K., Volchikhin V.I., Ivanov A.I., Funtikov V.A., Nazarov I.G. *Nejrosetevaya zashhita personal'nykh biometricheskikh dannyykh* [Neural protection of personal biometric data]. Moscow, Radiotekhnika Publ., 2012, 157 p. ISBN 978-5-88070-044-8.
8. GOST R 52633.5-2011 «Zashhita informatsii. Tekhnika zashhity informatsii. Avtomaticheskoe obuchenie nejrosetevykh preobrazovatelej biometriya-kod dostupa».
9. Akhmetov B.S., Nadeev D.N., Funtikov V.A., Ivanov A.I., Malygin A.Y. *Otsenka riskov vysokonadezhnoj biometrii* [Risk assessment of highly reliable biometrics]. Almaty, KazNTU im. K.I. Satpaeva Publ., 2014, 108 p.
10. Ivanov A.I., Lozhnikov P.S., Kachajkin E.I. Identifikatsiya podlinnosti rukopisnykh avtografov setyami Bajesa-Khehminga i setyami kvadraticnykh form [Verification of authenticity for handwritten signatures using Bayesian-Hamming networks and quadric form networks]. *Voprosy zashhity informatsii*, 2015, no. 2, pp. 28-34.
11. Ivanov A.I., Kachajkin E.I., Lozhnikov P.S., Sulavko A.E., Biometricheskaya identifikatsiya rukopisnykh obrazov s ispol'zovaniem korrelyatsionnogo analoga pravila Bajesa [Biometric identification of handwritten images via correlation analog of Bayes rule]. *Voprosy zashhity informatsii*, 2015, no. 3, pp. 48-54.
12. Volchikhin V.I., Ivanov A.I., Serikova Y.I. Kompensatsiya metodicheskikh pogreshnostej vychisleniya standartnykh otklonenij i koehffitsientov korrelyatsii, vznikayushhikh iz-za malogo ob»ema vyborok. Izvestiya vysshikh uchebnykh zavedenij [Compensation of methodological errors in calculations of standard deviations and correlation coefficients occurring due to small sample sizes]. *Povolzhskij region. Tekhnicheskie nauki*, Penza, PGU, 2016, no. 1, pp. 45-49.
13. Ivanov A.I., Zakharov O.S. *Sreda modelirovaniya «BioNeuroAvtograf»: arkhiv s programmnyy obespecheniyem razmeshchen na sayte OAO «Penzenskiy nauchno-issledovatel'skiy elektrotekhnicheskiy institut»* [The modeling environment «BioNeuroAvtograf». Developed by «Penza research Electrotechnical Institute»]. Available at: <http://pniehi.rf/activity/science/noc.htm>.

*Received 14.02.2017*

## **УПРАВЛЕНИЕ И ПОДГОТОВКА КАДРОВ ДЛЯ ОТРАСЛИ ИНФОКОММУНИКАЦИЙ**

УДК 331.461

### **СЕТЕВОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССА ОБРАБОТКИ ИНФОРМАЦИИ В ОРГАНИЗАЦИОННЫХ СИСТЕМАХ УПРАВЛЕНИЯ ПРОИЗВОДСТВОМ**

*Алекина Е.В., Яговкина Е.Н., Яговкин Н.Г.*

*Самарский государственный технический университет, Самара, РФ*

*E-mail: bjd@list.ru*

Предложено иерархическое представление организационных систем управления на трех уровнях: микроструктура, макро-структура и мегаструктура. На первом уровне производится идентификация фактора, на втором – обстановки в информационной среде, на третьем – составление прогноза ее развития с максимальной достоверностью и определение вероятности того, какой она станет через фиксированный интервал времени в будущем. Составляется сетевой граф событий, определяются статистические характеристики времени завершения ситуации – среднее значение и дисперсия. Рассчитывается ожидаемое время свершения определенной ситуации при требуемом значении вероятности ее свершения; оцениваются особенности поведения