Раздел V. Биометрические и иммунологические методы защиты информации

УДК 004.065

Ю.А. Брюхомицкий

ПАРАМЕТРИЧЕСКИЕ МЕТОДЫ РАСПОЗНАВАНИЯ ОБРАЗОВ ДИНАМИЧЕСКОЙ БИОМЕТРИИ

Рассматриваются три параметрических метода распознавания биометрических параметров, предназначенные для использования в динамических биометрических системах аутентификации личности по рукописному и клавиатурному почеркам. Первый метод основан на вычислении квадратичной дискриминантной функции, коэффициенты которой определяются по характеристикам распределения биометрических параметров «своего» пользователя. Во втором методе дискриминантная функция представлена и вычисляется как многомерная функции плотности распределения биометрических параметров пользователя. Третий метод использует канонический вид квадратичной дискриминантной функции. Предлагаемые методы обладают повышенной точностью распознавания и простотой обучения.

Динамические биометрические системы аутентификации; многомерное нормальное распределение биометрических параметров; параметрические методы распознавания; дискриминантная функция; обучение; верификация.

Yu.A. Bryukhomitsky

PARAMETRIC METHODS OF RECOGNITION FORM A DYNAMIC BIOMETRICS

We consider three parametric method of pattern recognition biometrics for use in dynamic biometric authentication systems by individual handwriting and keyboard writing style. The first method is based on the computation of the quadratic discriminant function coefficients are determined by the characteristics of the distribution of biometrics "own" the user. In the second method discriminant function is presented and evaluated as a multivariate density function biometric user. The third method uses the canonical form of the quadratic discriminant function. The proposed methods have a high recognition accuracy and ease of learning.

Dynamic biometric authentication system; a multivariate normal distribution of biometric parameters; parametric methods of pattern recognition; discriminant function; training; verification.

В биометрических системах аутентификации личности (БСА), использующих в качестве идентификационных признаков динамику рукописного и клавиатурного почерков, большое влияние на конечные характеристики системы оказывают принципы сопоставления (классификации, верификации) биометрических признаков. Эти функции реализуются специальным блоком БСА – *мэтчером*, который осуществляет сопоставление машинных репрезентаций предъявляемых и эталонных биометрических признаков и по результатам сопоставления выносит соответствующее аутентификационное решение [1].

В теории распознавания известно множество методов классификации образов, которые могут быть положены в основу работы мэтчеров динамических БСА: геометрические, нейросетевые, параметрические, статистические и др. Каждый из них имеет свои характеристики, особенности и области применения.

В данной работе изложены принципы верификации образов на основе параметрических методов обучения и распознавания, которые в наибольшей степени отвечают требованиям, предъявляемым к мэтчерам динамических БСА по рукописному и клавиатурному почеркам.

В БСА задача идентификации входных биометрических данных пользователя реализуется последовательно, путем поочередного их сопоставления с имеющимися в биометрической базе данных (ББД) эталонами $k = \overline{1, M}$ зарегистрированных пользователей (сопоставление 1:*M*). Таким образом, в каждом сеансе сопоставления мэтчер БСА решает более узкую задачу – верификации предъявленных биометрических данных, сравнивая их с единственным эталоном (сопоставление 1:1). Положительный или отрицательный результат сопоставления свидетельствует, что предъявленные биометрические данные принадлежат «своему» или «чужому» пользователю соответственно.

Задача разделения (верификации) множества входных данных на «своих» и «чужих» в общем случае сводится выбору метода построения некоторой разграничительной (дискриминантной) функции $g(\mathbf{V})$, реализующей это разделение. В свою очередь метод построения $g(\boldsymbol{V})$ зависит от характера классифицируемых объектов.

Если решаются задачи классификации, для которых a priori известно, что каждый класс объектов характеризуется некоторой системой параметров, но значения этих параметров не известно, то такие задачи целесообразно решать с помощью параметрических методов. При этом дискриминантная функция $g(\boldsymbol{V})$ задается в явном виде с использованием *m* действительных параметров, называемых весами:

$$g(\boldsymbol{V}) = \boldsymbol{g}(\mathbf{V}, w_1, w_2, \dots, w_m)\,. \qquad (1)$$

Из множества различных функций $g(\boldsymbol{V})$, которые могут быть заданы выражением (1), выделяются определенные классы функций: линейные, кусочно-линейные, квадратичные и др. После чего, из каких-либо практических соображений (точность, быстродействие, простота и т.п.) выбирается один определенный класс функций. Последующее обучение классификатора в выбранном классе функций $g(\boldsymbol{V})$ сводится к некоторой процедуре подбора весов $w_1, w_2, \dots, w_m$ [2].

Таким образом, параметрические методы классификации – это такие методы, которые используют обучающее множество объектов для получения оценок величин параметров, которые затем используются для построения дискриминантных функций [2].

В динамических БСА биометрические параметры пользователей, в большинстве случаев, подчиняются нормальному закону распределения с неизвестными характеристиками математических ожиданий и центральных моментов. Следовательно, для классификации данных в таких БСА в наибольшей степени подходят параметрические методы.

Метод верификации на основе квадратичной дискриминантной функции. Рассмотрим метод параметрической верификации биометрических образцов, реализуемый на основе выражения (1), при условии, что биометрические параметры пользователей подчиняются нормальному закону распределения.

В соответствии с реальными условиями использования БСА, зададим область распределения биометрических параметров «своего» пользователя ограниченным множеством образцов, состоящим из векторов $\mathbf{V}_{ci},\ i = \overline{1, L}$, распределенных в пространстве E^N ортогональной системы координат:

$$\mathbf{V}_{ci} = \left(v_1, v_2, \dots, v_j, \dots, v_N\right),\ i = \overline{1, L},\ j = \overline{1, N}.$$

Центр распределения векторов $\mathbf{V}_{ci}, i = \overline{1, L}$ находится в точке $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_N)$, которая определяются математическими ожиданиями: $\mu(v_1) = \xi_1,\ \mu(v_2) = \xi_2, \dots,$

$\mu(v_N) = \xi_N$. Центральные моменты второго порядка распределения векторов $\mathbf{V}_{ci}, i = \overline{1,L}$ образуют квадратную матрицу (ковариационную матрицу):

$$\mathbf{Q} = \|\lambda_{jk}\| = \begin{Vmatrix} \sigma_{11}^2 & \lambda_{12} & \dots & \lambda_{1k} \\ \lambda_{21} & \sigma_{22}^2 & \dots & \lambda_{2k} \\ \dots & \dots & \dots & \dots \\ \lambda_{j1} & \lambda_{j2} & \dots & \sigma_{jk}^2 \end{Vmatrix},$$

где

$$\lambda_{jk} = \lambda_{kj} = M(v_j - \mu_j)(v_k - \mu_k) = \begin{cases} \sigma_j^2, & \text{при } j = k; \\ \operatorname{cov}(v_j, v_k), & \text{при } j \neq k. \end{cases}$$

Биометрические параметры в силу своей природы обладают внутренней корреляцией, и смешанные моменты в ковариационной матрице $\mathbf{Q} = \|\lambda_{jk}\|$ в общем случае не равны нулю.

Для построения оптимального классификатора воспользуемся симметричной функцией потерь [2]:

$$\eta(j|k) = 1 - \delta_{jk},$$

где δ_{jk} – дельта функция, равная

$$\delta_{jk} = \begin{cases} 1, \text{если } j = k; \\ 0, \text{если } j \neq k. \end{cases}$$

Указанная функция уменьшается на единицу при любой ошибке классификации (объект j-класса отнесен к k-классу и наоборот, объект k-класса отнесен к j-классу) и остается неизменной при правильной классификации.

Установлено [2], что при классификации нормально распределенных объектов с использованием симметричной функции потерь, оптимальный классификатор должен строиться на основе квадратичных дискриминантных функций.

В общем случае квадратичная дискриминантная функция имеет вид

$$g(\mathbf{V}) = \sum_{j=1}^{N} w_{jj} \cdot v_j^2 + \sum_{j=1}^{N-1} \sum_{k=j+1}^{N} w_{jk} \cdot v_j \cdot v_k + \sum_{j=1}^{N} w_j \cdot v_j + w_{N+1} \qquad (2)$$

или в матричной форме

$$g(\mathbf{V}) = \mathbf{V}^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{V} + \mathbf{V}^{\mathrm{T}}\mathbf{B} + C, \qquad (3)$$

где $\mathbf{V}$ – вектор-столбец;

$\mathbf{V}^{\mathrm{T}}$– транспонированный вектор $\mathbf{V}$ (вектор-строка);

$\mathbf{A}$ – матрица вида $\mathbf{A} = \|a_{jk}\|$ с элементами $a_{jj} = w_{jj}$, $a_{jk} = w_{jk}/2$, $j, k = \overline{1,N}$, $j \neq k$.

Член $\mathbf{V}^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{V}$ в (3) является квадратичной формой.

Функция (2) имеет $(N+1)(N+2)/2$ весов, в том числе N весов w_{jj} при квадратичных членах; N весов w_j при линейных членах; $N(N-1)/2$ весов w_{jk} при смешанных квадратичных членах ($j \neq k$) и один свободный член w_{N+1}.

Для построения классификатора на основе дискриминантной функции (2) представим ее M-мерным вектором

$$\mathbf{F} = \mathbf{F}(\mathbf{V}) = f_1, f_2, \dots, f_M,$$

компоненты которого являются функциями v_j, $j = \overline{1,N}$. Первые N компонент равны $v_j^2, j = \overline{1,N}$, следующие $N(N-1)/2$ компонент равны всевозможным парным сочетаниям v_j и v_k, последние N компонент равны v_j, $j = \overline{1,N}$. Очевидно, что $M = N(N+3)/2$.

$\mathbf{F(V)}$ является однозначным преобразованием, при котором для каждого $\mathbf{V} \in E^N$ существует единственный вектор $\mathbf{F} \in E^M$. Это позволяет представить дискриминантную функцию $g(\mathbf{V})$ в виде линейной функции компонент вектора $\mathbf{F}$:

$$g(\mathbf{V}) = w_1 f_1 + w_2 f_2 + \dots + w_M f_M + w_{M+1}. \qquad (4)$$

Выражение (4) позволяет реализовать оптимальный классификатор биометрических параметров.

Весовым коэффициентам в (4) необходимо поставить в соответствие параметры распределения векторов $\mathbf{V}_{ci}$. Очевидно, первые N коэффициентов $w_1, w_2, \dots, w_N$ соответствуют дисперсиям σ_j^2, следующие $N(N–1)/2$ коэффициентов $w_{N+1}, w_{N+2}, \dots, w_{N+N(N-1)/2}$ соответствуют смешанным моментам $\operatorname{cov}(v_j, v_k)$, следующие N коэффициентов $w_{N+1+N(N-1)/2}, w_{N+2+N(N-1)/2}, \dots, w_M$ соответствуют смещениям b_j, компонент v_j вектора $\mathbf{V}_{ci}$ относительно начала координат и определяются математическими ожиданиями $\mu_{i1}, \mu_{i2}, \dots, \mu_{iN}$. Последний коэффициент w_{M+1} является свободным членом.

Необходимые для получения весовых коэффициентов оценки математических ожиданий $\hat{\mu}_j$, дисперсий $\hat{\sigma}_j^2$ и смешанных моментов $\widehat{\operatorname{cov}}(v_j, v_k)$ вычисляются на имеющейся выборке примеров объема L по обычным статистическим формулам:

$$\hat{\mu}_j \approx \frac{1}{L}\sum_{i=1}^{L} v_{ij};$$

$$\hat{\sigma}_j^2 \approx \frac{1}{L-1}\sum_{i=1}^{L}\left[v_{ij} - \hat{\mu}(v_j)\right]^2;$$

$$\widehat{\operatorname{cov}}(v_j, v_k) \approx \frac{1}{L-1}\sum_{i=1}^{L}\left[v_{ij} - \hat{\mu}(v_j)\right]\left[v_{ik} - \hat{\mu}(v_k)\right].$$

При условии нормального распределения векторов $\mathbf{V}_{ci}$ в выражении (3) для квадратичной дискриминантной функции $g(\mathbf{V})$ матрица $\mathbf{A}$ становится положительно определенной, а член $\mathbf{V}^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{V}$ – соответственно положительно определенной квадратичной формой. Поверхность, описываемая функцией $g(\mathbf{V})$, представляют собой N-мерный эллипсоид равной плотности, осями которого являются собственные векторы матрицы $\mathbf{A}$:

$$\mathbf{V}^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{V} + \mathbf{V}^{\mathrm{T}}\mathbf{B} = C^2. \tag{5}$$

Константа C задает коэффициент пропорциональности между длинами главных полуосей эллипсоида и соответствующими среднеквадратическими отклонениями. Для учета рассеивания векторов $\mathbf{V}_{ci}$ можно ограничиться единичным эллипсоидом рассеивания, в котором $C = 1$.

В БСА при обработке биометрических данных рассматриваются распределения выборочных статистик, поэтому нормальный закон рассеивания векторов $\mathbf{V}_{ci}$ целесообразно задавать с учетом ошибки «своего» пользователя (ошибки первого рода). Для этого целесообразно применить t-распределение (Стьюдента), которое «расширяет» эллипсоид рассеивания векторов $\mathbf{V}_{ci}$ на коэффициент Стьюдента t, определяемый как

$$t = f[L, (1 - P_1)],$$

где P_1 – вероятность ложного отказа «своему» пользователю (ошибка первого рода);

L – число предъявленных образцов КП.

Для распределения Стьюдента выражение (5) можно представить в виде

$$\mathbf{V}^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{V} + \mathbf{V}^{\mathrm{T}}\mathbf{B} - t^2 = 0.$$

При классификации неизвестных биометрических признаков в виде векторов $\mathbf{V}$ по принципу «свой-чужой» уравнение $g(\mathbf{V}) = 0$ будет определять разделяющую поверхность между классами, а знак функции $g(\mathbf{V})$ – принадлежность предъявленного вектора $\mathbf{V}$ к одному из двух классов: «свой» или «чужой»:

$$\mathbf{V} \in \begin{cases} \mathbf{V}_{с}, \text{если } \operatorname{sign} g(\mathbf{V}) = 0; \\ \mathbf{V}_{ч}, \text{если } \operatorname{sign} g(\mathbf{V}) = 1. \end{cases}$$

Иллюстрация метода в пространстве E^2 приведена на рис. 1.

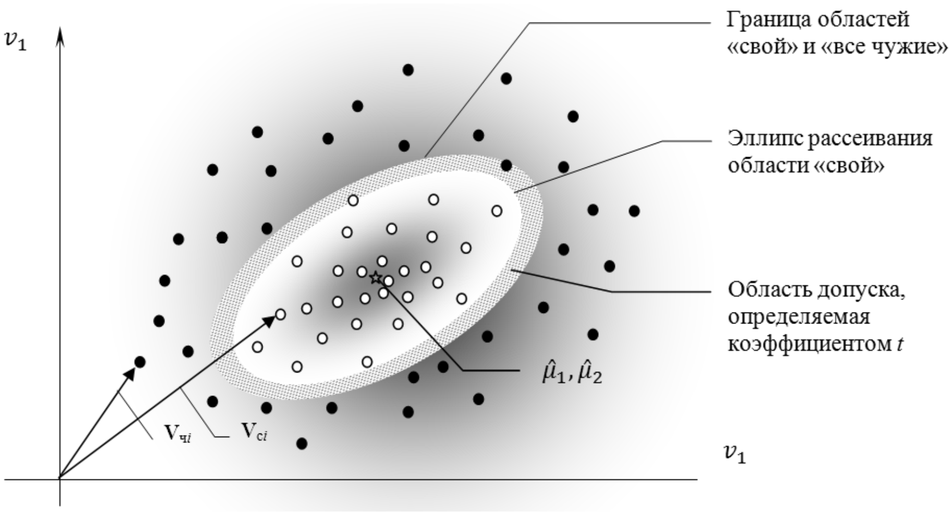

Рис. 1. Метод верификации на основе квадратичной дискриминантной функции в пространстве E^2

Мэтчер БСА, реализующий изложенный метод верификации, работает следующим образом.

В режиме *обучения* по биометрическим образцам векторов «своего» пользователя $\mathbf{V}_{сi}$, $i = \overline{1, L}$ из обучающей выборки определяются оценки математических ожиданий $\hat{\mu}_j$, дисперсий $\hat{\sigma}_j^2$ и смешанных моментов $\widehat{\mathrm{cov}}(v_j, v_k)$. Совокупность параметров $\hat{\mu}_j$, $\hat{\sigma}_j^2$, $\widehat{\mathrm{cov}}(v_j, v_k)$ представляют собой машинную репрезентацию биометрических признаков данного пользователя и выступают в роли его эталона. Процедура повторяется для всех $k = \overline{1, M}$ пользователей. Результатом обучения является набор из M эталонов, помещаемый в ББД БСА.

В режиме *верификации* пользователь, претендующий на доступ, предъявляет БСА свои биометрические данные в виде вектора $\mathbf{V}$, а также – дополнительный идентификатор своей личности (ID). По дополнительному идентификатору БСА извлекает из ББД единственный эталон, соответствующий данному пользователю с именем ID (если этот пользователь был ранее зарегистрирован). В мэтчере БСА на основе предъявленного вектора $\mathbf{V}$ и содержащихся в эталоне параметров $\hat{\mu}_j$, $\hat{\sigma}_j^2$, $\widehat{\mathrm{cov}}(v_j, v_k)$ вычисляется дискриминантная функция $g(\mathbf{V})$. Знак функции $g(\mathbf{V})$ определяет принадлежность предъявленного вектора $\mathbf{V}$ к одному из двух классов: «свой», если $\operatorname{sign} g(\mathbf{V}) = 0$ и «чужой», если $\operatorname{sign} g(\mathbf{V}) = 1$.

Структурная схема классификатора, реализующего верификацию пользователей по принципу «свой-чужой» на основе квадратичной дискриминантной функции (4) приведена на рис. 2.

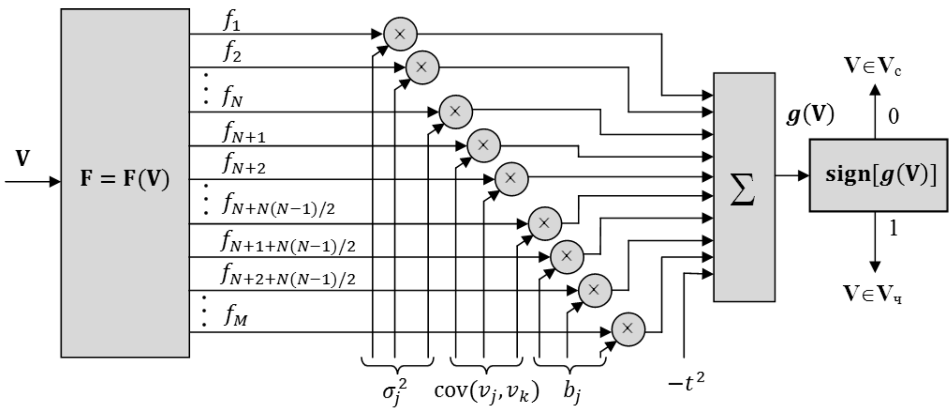

Рис. 2. Структурная схема классификатора, реализующего верификацию пользователей по принципу «свой-чужой» на основе квадратичной дискриминантной функции

Метод верификации на основе функции плотности распределения. Если закон распределения клавиатурных параметров известен, то задачу разделения (верификации) множества входных биометрических векторов **V** на «своих» (вектор $\mathbf{V}_с$) и «чужих» (вектор $\mathbf{V}_ч$) можно решить путем аналитического задания дискриминантной функции (1).

Полагая, что распределение клавиатурных параметров пользователей подчиняется нормальному закону, представим дискриминантную функцию (1) в виде функции плотности нормального распределения N-мерных векторов $\mathbf{V}_{ci}, i = \overline{1, L}$:

$$f(v_1, v_2, \dots, v_N) = \frac{1}{\sqrt{(2\pi)^N \cdot det\|\lambda_{jk}\|}} \exp\left[-\frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N} \Lambda_{jk}\,(v_j - \mu_j)(v_k - \mu_k)\right], \quad (6)$$

где $\det\|\lambda_{jk}\|$ – определитель ковариационной матрицы $\mathbf{Q} = \|\lambda_{jk}\|$.

Коэффициенты Λ_{jk} составляют матрицу $\boldsymbol{\Lambda} = \|\Lambda_{jk}\|$, обратную ковариационной матрице $\mathbf{Q} = \|\lambda_{jk}\|$**.** Для вычисления коэффициентов Λ_{jk} используется формула

$$\Lambda_{jk} = (-1)^{j+k}\frac{M_{jk}}{\det\|\lambda_{jk}\|},$$

где M_{jk} – минор определителя $\det\|\lambda_{jk}\|$, получаемый вычеркиванием j-строки и k-столбца.

Выражение, фигурирующее в показателе экспоненты функции плотности нормального распределения (6), является положительно определенной квадратичной формой. Поверхности, на которых эта форма постоянна, являются поверхностями равных плотностей вероятностей в пространстве E^N и представляют собой N-мерные эллипсоиды равной плотности, которые группируются вокруг центральной точки $(\mu_{i1}, \mu_{i2}, \dots, \mu_{iN})$. Обозначая указанную форму через C^2, получим

$$\frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N} \Lambda_{jk}\,(v_j - \mu_j)(v_k - \mu_k) = C^2. \quad (7)$$

Используя распределение Стьюдента, введем коррекцию на ошибки «своего» пользователя в условиях ограниченного объема выборки L. Выражение для дискриминантной функции $g(\mathbf{V})$ будет иметь вид

$$g(\mathbf{V}) = \frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N} \Lambda_{jk}\,(v_j - \mu_j)(v_k - \mu_k) - \{t[L_1, (1 - P_1)]\}^2. \quad (8)$$

Вычисления значений функции $g(\mathbf{V})$ производятся непосредственно по выражению (8). При этом уравнение $g(\mathbf{V})) = 0$ будет определять разделяющую поверхность между классами, а знак функции $g(\mathbf{V})$ – принадлежность предъявленного вектора $\mathbf{V}$ к одному из двух классов: «свой» или «чужой»:

$$\mathbf{V} \in \begin{cases} \mathbf{V}_{\text{с}}, \text{если } \operatorname{sign} g(\mathbf{V}) = 0; \\ \mathbf{V}_{\text{ч}}, \text{если } \operatorname{sign} g(\mathbf{V}) = 1. \end{cases}$$

Совокупность выражений, необходимых для верификации неизвестного входного вектора биометрических параметров $\mathbf{V}$:

$$\hat{\mu}_j \approx \frac{1}{L}\sum_{i=1}^{L} v_{ij},$$

$$\hat{\sigma}_j^2 \approx \frac{1}{L-1}\sum_{i=1}^{L}\left[v_{ij} - \hat{\mu}(v_j)\right]^2,$$

$$\widehat{\operatorname{cov}}(v_j, v_k) \approx \frac{1}{L-1}\sum_{i=1}^{L}\left[v_{ij} - \hat{\mu}(v_j)\right]\left[v_{ik} - \hat{\mu}(v_k)\right],$$

$$\hat{\lambda}_{jk} = \hat{\lambda}_{kj} = \hat{\mu}(v_j - \hat{\mu}_j)\,(v_k - \hat{\mu}_k) = \begin{cases} \hat{\sigma}_j^2 & \text{при } j = k; \\ \widehat{\operatorname{cov}}(v_j, v_k) & \text{при } j \neq k, \end{cases}$$

$$\Lambda_{jk} = (-1)^{j+k}\frac{M_{jk}}{\det\|\hat{\lambda}_{jk}\|},$$

$$\hat{g}(\mathbf{V}) = \frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N}\Lambda_{jk}\left(v_j - \hat{\mu}_j\right)(v_k - \hat{\mu}_k) - \{t[L,(1-P_1)]\}^2,$$

$$\mathbf{V} \in \begin{cases} \mathbf{V}_{\text{с}}, \text{если } \operatorname{sign} \hat{g}(\mathbf{V}) = 0; \\ \mathbf{V}_{\text{ч}}, \text{если } \operatorname{sign} \hat{g}(\mathbf{V}) = 1. \end{cases}$$

Структурная схема классификатора, реализующего верификацию пользователей по принципу «свой-чужой» на основе вычисления функции плотности распределения показана на рис. 3.

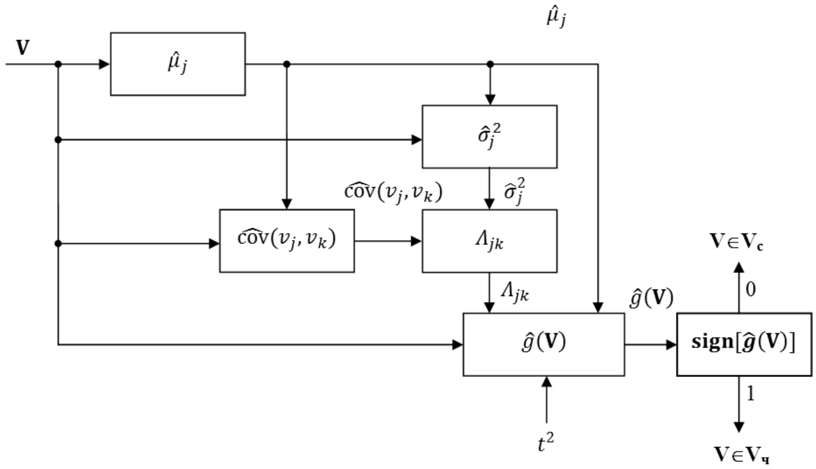

Рис. 3. Структурная схема классификатора, реализующего верификацию пользователей по принципу «свой-чужой» на основе вычисления функции плотности распределения

Метод верификации на основе канонического вида дискриминантной функции. Если биометрические параметры пользователей подчиняются нормальному закону распределения, возможен и другой подход к их верификации. Суть его заключается в использовании канонического вида квадратичной дискриминантной функции $g(\mathbf{V})$, который образуется из (2) при условии, что все веса при неквадратичных членах w_{jk} и w_j равны нулю:

$$g(\mathbf{V}) = \sum_{j=1}^{N} w_j \cdot v_j^2 + w_{N+1}. \tag{9}$$

Чтобы использовать такой вид дискриминантной функции для классификации биометрических данных, представленных векторами $\mathbf{V}$, необходимо предварительно преобразовать к каноническому виду их распределение. Для этого необходимо выполнить две последовательные операции: центрирования и декорреляции исходного нормального распределения. Применим указанные операции к распределению биометрических параметров «своего» пользователя, представленному векторами $\mathbf{V}_{ci}, i = \overline{1, L}$.

Центрирование осуществляется путем совмещения центра распределения векторов $\mathbf{V}_{ci}$ с началом координат:

$$\dot{\mathbf{V}}_{ci} = \mathbf{V}_{ci} - \mu_i,$$

где $\dot{\mathbf{V}}_{ci}$ – центрированные значения векторов $\mathbf{V}_{ci}$;

μ_i – точка математических ожиданий $(\mu_{i1}, \mu_{i2}, \dots, \mu_{iN})$ векторов $\mathbf{V}_{ci}$ $i = \overline{1, L}$.

Декоррелирующее преобразование в общем случае имеет вид

$$\tilde{\dot{\mathbf{V}}}_{ci} = \dot{\mathbf{V}}_{ci} \cdot \mathbf{D}^{-1},$$

где $\mathbf{D}^{-1}$ – матрица декоррелирующих преобразований.

Процедуру декорреляции векторов $\dot{\mathbf{V}}_{ci\,i}$ удобно проводить с использованием алгоритма ортогонализации Грама-Шмидта [3].

После указанных преобразований конфигурация области распределения векторов $\mathbf{V}_{ci}$, $i = \overline{1, L}$ представляет собой N-мерный эллипсоид, точка математических ожиданий которого совмещена с началом координат, а главные оси совмещены с осями координат. Каноническое уравнение эллипсоида соответствует дискриминантной функции (9):

$$\sum_{j=1}^{N} w_j \cdot \tilde{\dot{v}}_j^2 - w_{N+1} = 0, \tag{10}$$

где

$$w_j = \frac{1}{\tilde{\dot{\sigma}}_j}, \qquad w_{N+1} = C^2.$$

Здесь $\tilde{\dot{\sigma}}_j$ – дисперсии распределения центрированных декоррелированных векторов $\tilde{\dot{\mathbf{V}}}_{ci}$.

В результате дискриминантная функция для классификации векторов $\tilde{\dot{\mathbf{V}}}_{ci}$ будет иметь вид

$$g\left(\tilde{\dot{\mathbf{V}}}\right) = \sum_{j=1}^{N} w_j \cdot \tilde{\dot{v}}_j^2 - C^2. \tag{11}$$

Введем коррекцию на ошибки «своего» пользователя в условиях ограниченного объема выборки L. Для этого нормальное распределение заменим распределением Стьюдента. Выражение для дискриминантной функции преобразуется к виду

$$g\left(\tilde{\dot{\mathbf{V}}}\right) = \sum_{j=1}^{N} w_j \cdot \tilde{\dot{v}}_j^2 - \{t[L_1, (1 - P_1)]\}^2. \tag{12}$$

Весовые коэффициенты w_j в (12) определяются в процессе обучения БСА при вычислении смещенных и декоррелированных оценок математических ожиданий $\tilde{\mu}_j$ и дисперсий $\tilde{\sigma}_j^2$ распределения векторов $\mathbf{V}_{ci}$, а затем используются для построения дискриминантной функции (12). Для вычисления математических ожиданий $\tilde{\mu}_j$ и дисперсий $\tilde{\sigma}_j^2$ могут использоваться приведенные выше обычные оценки.

Иллюстрация метода в пространстве E^2 приведена на рис. 4.

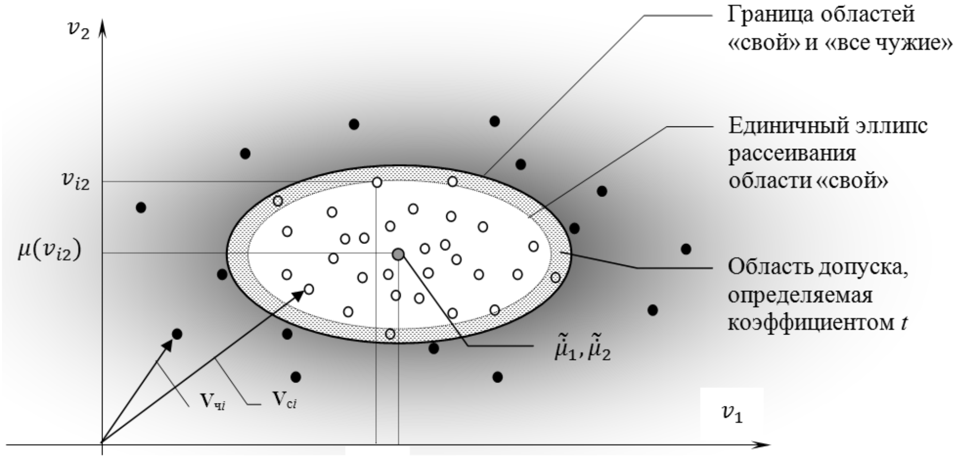

Рис. 4. Метод верификации на основе канонического вида дискриминантной функции

Мэтчер БСА, реализующий изложенный метод верификации, работает следующим образом.

В режиме *обучения* образцы векторов из обучающей «своего» пользователя вначале центрируются: $\mathbf{V}_{ci} \to \dot{\mathbf{V}}_{ci}$, $i = \overline{1, L}$ и декоррелируются: $\dot{\mathbf{V}}_{ci} \to \tilde{\dot{\mathbf{V}}}_{ci}$. Затем для распределения векторов $\tilde{\dot{\mathbf{V}}}_{ci}$ определяются оценки математических ожиданий $\tilde{\mu}_j$ и дисперсий $\tilde{\sigma}_j^2$. Совокупность параметров $\tilde{\mu}_j$ и $\tilde{\sigma}_j^2$ представляют собой машинную репрезентацию биометрических данных пользователя и выступают в роли его эталона. Процедура повторяется для всех $k = \overline{1, M}$ пользователей. Результатом обучения является набор из M эталонов, помещаемый в ББД БСА.

В режиме *верификации* пользователь, претендующий на доступ, предъявляет БСА свои биометрические данные в виде вектора $\mathbf{V}$, а также – дополнительный идентификатор своей личности (ID). По дополнительному идентификатору БСА извлекает из ББД единственный эталон, соответствующий данному пользователю с именем ID (если этот пользователь был ранее зарегистрирован). В мэтчере БСА на основе вектора $\tilde{\mathbf{V}}$ и содержащихся в эталоне параметров $\tilde{\sigma}_j^2$ вычисляется функция $g\left(\tilde{\mathbf{V}}\right)$. Знак этой функции определяет принадлежность предъявленного вектора $\mathbf{V}$ к одному из двух классов: «свой», если $\operatorname{sign} g\left(\tilde{\mathbf{V}}\right) = 0$ и «чужой», если $\operatorname{sign} g\left(\tilde{\mathbf{V}}\right) = 1$.

Структурная схема классификатора, реализующего верификацию пользователей по принципу «свой-чужой» на основе канонического вида квадратичной дискриминантной функции приведена на рис. 5.

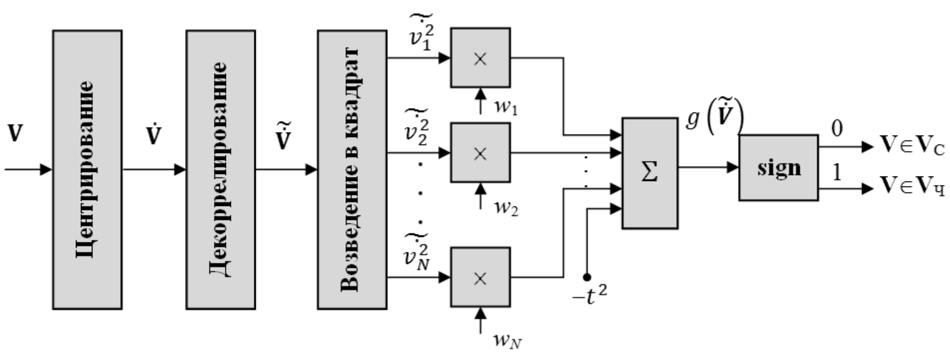

Рис. 5. Структурная схема классификатора, реализующего верификацию пользователей по принципу «свой-чужой» на основе квадратичной дискриминантной функции

Методы параметрической верификации обладает рядом преимуществ перед геометрическими и нейросетевыми методами.

По сравнению с геометрическими методами точность верификации существенно возрастает, вследствие более точной аппроксимации области распределения биометрических данных. Повышение точности при этом «оплачивается» дополнительным объемом вычислений, связанным с получением функции $g(\mathbf{V})$. Но, учитывая, что эти вычисления производятся по стандартным, фиксированным во времени процедурам, ощутимой потери быстродействия БСА не возникает.

По сравнению с нейросетевыми методами классификации исчезает необходимость неопределенно длительного обучения БСА в классическом его понимании (как подбор весов дискриминантной функции $g(\mathbf{V})$). Исчезают проблемы возникновения тупиков и «паралича» сети, а также проблема обучения на «чужих» пользователей.

Методы прошли экспериментальную проверку при создании БСА BioSing и BioKey [4], которые использовались в учебном процессе на кафедре безопасности информационных технологий Южного федерального университета. Проведенные на программах BioSing и BioKey экспериментальные исследования показали, что выигрыш в точности классификации при переходе от геометрического метода к параметрическому составил 30-40 раз. Проведенные исследования позволили одновременно установить оптимальные значения коэффициентов Стьюдента t при заданной мерности N пространства входных данных E^N.

Таким образом, параметрические методы эффективно сочетают в себе простоту реализации, высокую степень защиты от несанкционированного доступа, фиксированное малое время обучения и принятия решения. Эти качества позволяют широко использовать их в мэтчерах БСА самого различного назначения.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Болл Р.М., Коннел Дж.Х., Панканти Ш., Рахта Н.К., Сеньор Э.О.* Руководство по биометрии. – М.: Техносфера, 2007. – 386 с.
2. *Нильсон Н.* Обучающиеся машины: Пер. с англ. – М.: Мир, 1967. – 180 с.
3. *Беклемишев Д.В.* Курс аналитической геометрии и линейной алгебры. – М.: Наука, 1971. – 328 с.
4. *Брюхомицкий Ю.А., Казарин М.Н.* Учебные биометрические системы контроля доступа по рукописному и клавиатурному почеркам // Сборник научных трудов XIII Всероссийской научной конференции «Проблемы информационной безопасности в системе высшей школы». – М.: Изд-во МИФИ, 2006. – С. 33-34.

Статью рекомендовал к опубликованию к.т.н. М.Ю. Руденко.

Брюхомицкий Юрий Анатольевич
Технологический институт федерального государственного автономного образовательного учреждения высшего профессионального образования «Южный федеральный университет» в г. Таганроге.
E-mail: bya@tsure.ru.
347928, г. Таганрог, ул. Чехова, 2.
Тел.: 88634371905.
Кафедра безопасности информационных технологий; доцент.

Bryukhomitsky Yuriy Anatoly
Taganrog Institute of Technology – Federal State-Owned Autonomy Educational Establishment of Higher Vocational Education “Southern Federal University”.
E-mail: bya@tsure.ru.
2, Chekhov Street, Taganrog, 347928, Russia.
Phone: +78634371905.
The Department of Security in Data Processing Technologies; Associate Professor.

УДК 681.324

В.Д. Котов, В.И. Васильев

СИСТЕМА ОБНАРУЖЕНИЯ СЕТЕВЫХ ВТОРЖЕНИЙ НА ОСНОВЕ МЕХАНИЗМОВ ИММУННОЙ МОДЕЛИ

Системы обнаружения аномалий обладают большим потенциалом в области сетевой безопасности, однако на практике подобных систем реализовано мало. Хотя они способны обнаруживать атаки нулевого дня с приемлемым уровнем ложных срабатываний, существует проблема, связанная с необходимостью генерировать большое количество трафика, содержащего атаки. Подобные данные тяжело и дорого производить. В данной статье представлен адаптивный подход, основанный на иммунных механизмах. Поведение предлагаемой искусственной иммунной системы заимствует стратегию защиты у иммунной системы человека. В статье представлены результаты экспериментов, демонстрирующих перспективность технологии искусственных иммунных систем.

Система обнаружения вторжений; искусственные иммунные системы; адаптивные системы.

V.D. Kotov, V.I. Vasilyev

NETWORK ATTACKS DETECTION SYSTEM BASED ON THE MECHANISMS OF IMMUNE MODEL

The anomaly detection systems have big potential in the network security, but still too few of them are realized in practice. Although such systems can detect 0-day attacks with acceptable false alarm rate, the problem is that they have to be trained with the data, containing labeled attacks. And such data is hard and expensive to produce. This paper offers an adaptive solution based on the immunity mechanisms. The behavior of artificial immune system we proposed deploys the defense strategy of the human immunity. We show experimental results which demonstrate the efficiency of the artificial immune system technology.

Intrusion detection system; artificial immune systems; adaptive systems.

Введение. Система обнаружения вторжений (СОВ) является важным компонентом защиты компьютерных сетей. Её основная задача – это мониторинг сети или системы на предмет вредоносной активности. Несмотря на то, что проблема детектирования сетевых атак является довольно старой, она до сих пор актуальна. Не-