



www.volsu.ru

ИННОВАЦИИ В ИНФОРМАТИКЕ, ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКЕ И УПРАВЛЕНИИ

DOI: <https://doi.org/10.15688/jvolsu10.2017.4.1>

УДК 004.056

ББК 32.973.26

УСОВЕРШЕНСТВОВАНИЕ ИДЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ ПО КЛАВИАТУРНОМУ ПОЧЕРКУ

Константин Владимирович Стародубов

Старший преподаватель кафедры информационных систем и защиты информации,
Тамбовский государственный технический университет
gromov@is.tstu.ru
ул. Советская, 106, 392000 г. Тамбов, Российская Федерация

Юрий Юрьевич Громов

Доктор технических наук, профессор,
директор института автоматизации и информационных технологий,
Тамбовский государственный технический университет
gromov@is.tstu.ru
ул. Советская, 106, 392000 г. Тамбов, Российская Федерация

Аннотация. В качестве объекта усиления была выбрана парольная аутентификация. В процессе работы были проанализированы существующие методы классификации пользователей при использовании метода биометрической аутентификации по клавиатурному почерку. Также был рассмотрен подход к выявлению ключевых параметров характеристики клавиатурного почерка пользователя. Проведено исследование с целью определения оптимального значения коэффициента Стьюдента, отвечающего за уменьшение ошибки первого рода.

Ключевые слова: идентификация, клавиатурный почерк, параметрическое обучение, ковариационная матрица биометрической аутентификации.

В задаче идентификации пользователя путем использования клавиатурного почерка

важным этапом является обработка входных (первичных) данных. Входные данные можно

классифицировать как характеризующие те или иные качества аутентифицируемых пользователей. Признаки, подвергаясь обработке, позволяют получить эталонные характеристики пользователя.

Задача биометрической идентификации пользователя по особенностям клавиатурного почерка – это выделение идентифицирующих компонентов $\gamma(t)$ из функции $\vartheta(t)$, которая определяет исходные данные для идентификации в процессе измерения временных характеристик клавиатурного почерка пользователя, выделяя $\theta(t)$ и $\lambda(t)$ из исходной функции $\vartheta(t)$. Очевидно, что из-за невозможности построения модели набора текста единственным приемлемым решением является анализ и хранение статистических данных о клавиатурном почерке большого числа пользователей и создание эмпирических зависимостей $\theta(t)$ и $\lambda(t)$.

Вероятность идентификации пользователя по времени удержания клавиш клавиатуры в зависимости от длины ключевой фразы является более стабильной характеристикой клавиатурного почерка, чем время между нажатиями клавиш (пауз) клавиатуры. Процесс нажатия клавиши на клавиатуре является подсознательным для пользователя. Характер входной функции мало меняется для большинства пользователей независимо от их квалификации и опыта работы с клавиатурой. Составляющая $\gamma(t)$ характеризуется временем удержания клавиш идентифицирующим лицом при наборе текста.

Время между нажатиями на клавиатуру можно считать менее стабильной характеристикой клавиатурного почерка, чем время удержания клавиш. Вероятность правильной идентификации от пауз между нажатиями клавиш имеет максимум своего значения при длине ключевой фразы в 8–10 символов. Ключевые фразы малой длины, максимум до двух слов, пользователь набирает подсознательно. Подсознательные движения стабильны до тех пор, пока в них не вмешивается сознательный уровень набора текста. У квалифицированных работников на клавиатуре наблюдается эффект включения сознательного мышления и остановок в наборе текста для принятия решения. В соответствии с этим можно утверждать о том, что в системах аутентификации

пользователя по клавиатурному почерку не рекомендуется использовать слишком длинные выражения в качестве ключевой фразы, так как это приводит к тому, что пользователь начинает «осмысленно» выполнять набор текста, что может создать больше ошибок идентификации. Таким образом, можно утверждать, что время между нажатиями клавиш наиболее полно передает характер составляющей $\theta(t)$ функции клавиатурного почерка $\vartheta(t)$. Исключение составляющей $\theta(t)$ из функции $\vartheta(t)$ осуществляется введением корреляционной функции $\Delta(t)$, эмпирически получаемой из анализа эталонных статистических данных. Так как характер корреляционной функции $\Delta(t)$ должен выбираться в зависимости от опыта работы пользователя с клавиатурой, необходимо ввести параметр L , который является корреляционной функцией и составляющей $\theta(t)$.

Для работы системы биометрической динамической идентификации по клавиатурному почерку легальный пользователь должен предъявить L эталонных подписей, то есть пройти обучение программного комплекса. Эталонные значения пользователя соответствуют L реализациям вектора биометрических параметров $V = \{V_1, V_2, \dots, V_L\}$.

В динамических биометрических системах идентификации, основанных на клавиатурном почерке, распределение векторов биометрических параметров V_i в N -мерном пространстве можно рассматривать равным нормальным [2], полученные векторы $V_i, i = \overline{1, L}$ можно задать в виде функции плотности нормального распределения векторов V_i с неизвестными средними, которая при $L \rightarrow \infty$ в ортогональной системе координат описывается гиперэллипсоидом рассеивания. Компоненты биометрических векторов $V_i, i = \overline{1, L}$ коррелированы между собой, оси гиперэллипсоида рассеивания не параллельны осям координат. Формула гиперэллипсоида идентификации пользователя – это контроль попадания вектора его биометрических параметров V внутрь N -мерной области, описываемой гиперэллипсоидом рассеивания [1].

Метод обучения состоит из трех последовательных этапов:

Этап 1. Выявляется зависимость дискриминантных функций $g(V)$ от параметров,

характеризующих функцию плотности нормального распределения векторов V_i .

Этап 2. Множество векторов V_i ; оцениваются величины этих параметров.

Этап 3. Принимается вероятным, что оценки являются правильными значениями параметров, и они подставляются в выражение для $g(V)$, полученное на этапе 1.

Идентификацию пользователя, набравшего тест на клавиатуре и получившего параметры в виде вектора V , можно рассматривать как задачу классификации вектора V на $(M + 1)$ классов – по числу M зарегистрированных в системе пользователей («свои») плюс один класс для всех остальных не зарегистрированных в системе пользователей («чужие»). До идентификации пользователь аутентифицирует себя парольной системой аутентификации. При этом предварительно зарегистрированные в системе биометрические параметры данного пользователя могут использоваться как параметры классификатора. Это обстоятельство позволяет без нарушения общности свести задачу классификации вектора V на $(M + 1)$ классов к задаче классификации только на два класса: «свой» – вектор V_C и «чужой» – вектор $V_ч$. При наличии в системе многих зарегистрированных пользователей после аутентификации каждого из них достаточно поочередно внести в классификатор в качестве эталона соответствующие ему параметры (один из M зарегистрированных векторов V_C).

Распределение пользователей на две области – «свой» (вектор V_C) и «чужой» (вектор $V_ч$) – параметрический классификатор может реализовать с использованием только одной дискриминантной функции $g(V)$, знак которой будет определять принадлежность предъявленного вектора V к одному из двух классов: V_C и $V_ч$. При этом области распределения биометрических параметров всевозможных «чужих» в совокупности можно рассматривать как интегральную область «все чужие», расположенную вокруг компактной области «свой».

Пусть в общем случае область распределения биометрических параметров «своего» пользователя задана множеством образцов Ψ_C , состоящим из L векторов V_{Ci} , $i = 1, \dots, L$, нормально распределенных в N -мерном пространстве ортогональной системы координат,

а каждый вектор V_{Ci} , $i = 1, \dots, L$, представлен своими N компонентами

$$V_{Ci} = \{v_1, v_1, \dots, v_j, \dots, v_N\}, j = \overline{1, N}. \quad (1)$$

Центр распределения векторов V_{Ci} находится в точке $(\xi_1, \xi_2, \dots, \xi_N)$, которая определяется N математическими ожиданиями $m_{v_1} = \xi_1, m_{v_2} = \xi_2, \dots, m_{v_N} = \xi_N$. Центральные моменты второго порядка распределения векторов V_{Ci} образуют квадратную матрицу моментов (ковариационную матрицу):

$$Q = \|\lambda_{jk}\| = \begin{vmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1N} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2N} \\ \dots & \dots & \dots & \dots \\ \lambda_{N1} & \lambda_{N2} & \dots & \lambda_{NN} \end{vmatrix}, \quad (2)$$

где

$$\lambda_{jk} = \lambda_{kj} = M(v_j - \xi_j)(v_k - \xi_k) = \begin{cases} j & \text{при } j = k \\ [(\text{cov}(v))]_{j,k} & \text{при } j \neq k \end{cases}, j, k = \overline{1, N}.$$

Для нормального закона распределения векторов V_{Ci} , $i = 1, \dots, L$, функция плотности распределения имеет вид:

$$f(v_1, v_2, \dots, v_N) = \frac{1}{\sqrt{(2\pi)^N \det\|\lambda_{jk}\|}} \exp\left[-\frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N \lambda_{jk} (v_j - \xi_j)(v_k - \xi_k)\right], \quad (3)$$

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

Коэффициенты λ_{jk} составляют матрицу $\Lambda = \|\Lambda_{jk}\|$, обратную ковариационной матрице $Q = \|\lambda_{jk}\|$.

Для вычисления коэффициентов Λ_{jk} используется стандартная формула

$$\Lambda_{jk} = (-1)^{j+k} \frac{M_{jk}}{\det\|\lambda_{jk}\|}, \quad (4)$$

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

Выражение, фигурирующее в показателе экспоненты функции плотности нормального распределения векторов V_{Ci} , является по-

ложительно-определенной квадратичной формой. Поверхности, на которых эта квадратичная форма постоянна

$$\frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N \Lambda_{jk} (v_j - \xi_j)(v_k - \xi_k) = \text{const}, \quad (5)$$

являются поверхностями равных плотностей вероятностей в N -мерном пространстве и представляют собой гиперэллипсоиды, которые группируются вокруг точки $(\xi_1, \xi_2, \dots, \xi_N)$.

Обозначая константу в правой части выражения через k^2 , получим:

$$\frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N \Lambda_{jk} (v_j - \xi_j)(v_k - \xi_k) = k^2. \quad (6)$$

Константа k задает коэффициент пропорциональности между длинами a_j главных полуосей гиперэллипсоида и соответствующими среднеквадратическими отклонениями s_j :

$$a = k \cdot a; \quad a = k \cdot a; \quad \dots, \quad a_N = k \cdot a_N. \quad (7)$$

Для оптимального решения задачи классификации из всех поверхностей равных плотностей вероятностей целесообразно выбрать ту, которая характеризует рассеивание векторов V_{Ci} относительно точки $(\xi_1, \xi_2, \dots, \xi_N)$. Эта поверхность соответствует так называемому единичному гиперэллипсоиду, у которого главные полуоси равны соответствующим среднеквадратическим отклонениям $\sigma_1, \sigma_2, \dots, \sigma_N$, то есть для единичного гиперэллипсоида $k = 1$, и преобразуется к виду

$$\frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N \Lambda_{jk} \cdot k (v_j - \xi_j)(v_k - \xi_k) = 1. \quad (8)$$

Из-за ограниченной статистики биометрических образцов, предъявляемых на стадии регистрации «своим» пользователем, всегда остается вероятность того, что образец, предъявленный этим же пользователем при аутентификации, выйдет за пределы зафиксированного в эталоне диапазона. Для уменьшения этой вероятности дополнительно задается величина допуска между областями «свой» и «все чужие» в виде коэффициента Стьюдента $C[L, (1-P)]$, исходя из заданной ошибки первого рода (вероятность P_i ложного отказа «своему» пользователю) и числа L предъявленных на стадии регистра-

ции образцов. Введение указанного допуска в уравнение:

$$\frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N \Lambda_{jk} (v_j - \xi_j)(v_k - \xi_k) = \{C[L, (1-P)]\}^2. \quad (9)$$

Используем полученное выражение (20) для формирования дискриминантной функции $g(V)$, разделяющей области «свой» и «все чужие»:

$$g(V) = \frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N \Lambda_{jk} \cdot k (v_j - \xi_j)(v_k - \xi_k) - \{C[L, (1-P)]\}^2. \quad (10)$$

Уравнение $g(V) = 0$ в этом случае будет определять искомую разделяющую поверхность, а знак функции $g(V)$ – принадлежность входного вектора V к одному из двух классов: «свой» или «чужой» (попадание в область «все чужие»):

$$g(V) < 0, \text{ если } V \in V_C, \quad (11)$$

$$g(V) > 0, \text{ если } V \in V_{\text{ч}}. \quad (12)$$

Идентификация пользователей сводится к проверке: попадает ли предъявленный пользователем вектор биометрических параметров V в эталонную область сформированных на основе эталонных параметров V .

В качестве основы для усиления был взят метод парольной аутентификации, который обладает рядом положительных качеств, описанных ранее. В качестве «усилителя» предлагается идентификация по клавиатурному почерку, являющаяся одним из методов динамической аутентификации по биометрическим данным. Усиления аутентификации возможно использовать в качестве самостоятельного модуля, который будет выполнять идентификацию по клавиатурному почерку методом параметрического обучения классификатора, содержащего необходимые данные из пользовательского приложения, так называемый вектор входных параметров.

Для использования возможностей библиотеки нужно в пользовательское приложение внедрить модуль подключения библиотеки, состоящий из процедур сбора данных и двух функций, одна из которых будет передавать

данные для регистрации нового пользователя, а вторая – для идентификации пользователей.

Само пользовательское приложение может содержать свой механизм парольной аутентификации и использовать тот, который был реализован в процессе выполнения данной работы. Данная возможность библиотеки позволяет как усиливать уже существующие системы парольной аутентификации, так и внедрять системы аутентификации в приложения, которые до этого не имели такой системы.

СПИСОК ЛИТЕРАТУРЫ

1. Брюхомицкий, Ю. А. Метод обучения нейросетевых биометрических систем на основе копирования областей / Ю. А. Брюхомицкий, М. Н. Казарин // Перспективные информационные технологии

и интеллектуальные системы. – 2003. – № 3 (15). – С. 17–23. – Электрон. текстовые дан. – Режим доступа: <http://pitis.tsure.ru>. – Загл. с экрана.

2. Иванов, А. И. Биометрическая идентификация личности по динамике подсознательных движений / А. И. Иванов. – Пенза : Изд-во Пенз. гос. ун-та, 2000. – 188 с.

REFERENCES

1. Bryukhomitskiy Yu.A., Kazarin M.N. Metod obucheniya neyrosetevykh biometricheskikh sistem na osnove kopirovaniya oblastey [Method of Training Neural Network Biometric Systems based on Copying Domains]. *Perspektivnye informatsionnye tekhnologii i intellektualnye sistemy*, 2003, no. 3 (15), pp. 17-23. URL: <http://pitis.tsure.ru>.

2. Ivanov A.I. *Biometricheskaya identifikatsiya lichnosti po dinamike podsoznatelnykh dvizheniy* [Biometric Identification of the Personality According to the Dynamics of Subconscious Movements]. Penza, Izd-vo Penz. gos. un-ta, 2000. 188 p.

IMPROVEMENT OF USER IDENTIFICATION BY KEYBOARD RHYTHM

Konstantin Vladimirovich Starodubov

Senior Lecturer, Department of Information Systems and Information Protection,
Tambov State Technical University
gromov@is.tstu.ru
Sovetskaya St., 106, 392000 Tambov, Russian Federation

Yuriy Yuryevich Gromov

Doctor of Technical Sciences, Professor,
Director of the Institute of Automation and Information Technologies,
Tambov State Technical University
gromov@is.tstu.ru
Sovetskaya St., 106, 392000 Tambov, Russian Federation

Abstract. Keyboard rhythm is an important element of processing the input (primary) data at user identification. The input data can be classified as the data characterizing certain qualities of authenticated users. Signs, being processed, allow receiving standard characteristics of the user.

Password authentication is chosen as the object of system protection. We analyze the existing methods of user classification when using the method of biometric authentication by keyboard rhythm. We also discuss the approach to identifying the key parameters of user's keyboard rhythm. The research purpose consists in determining the optimal value of Student's t-test, which is responsible for reducing the error of the first kind.

Key words: identification, keyboard rhythm, parametric training, covariance matrix of biometric authentication.