

потока лампы подсветки, работой механической части сканера, аддитивными шумами сенсоров ПЗС-линейки.

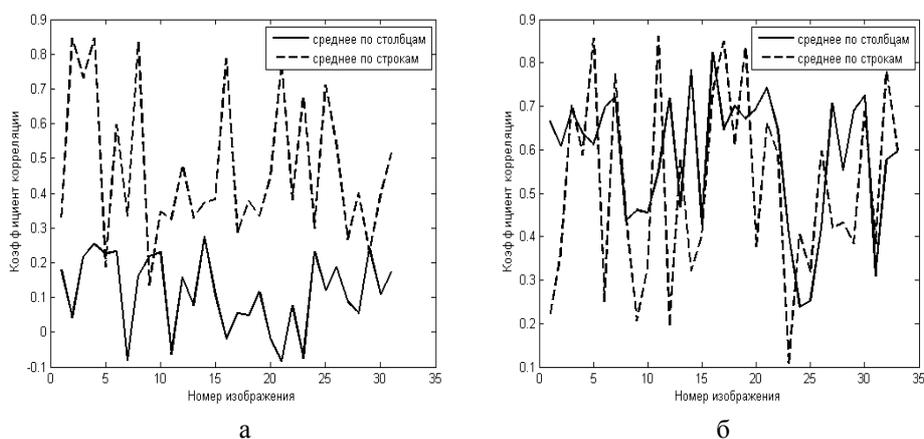


Рис. 3. Корреляция отпечатка сканера №2 и отпечатков изображений из выборки, полученной при помощи сканера №2(а) и сканера №1 (б)

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Jan Lukáš, Jessica Fridrich, and Miroslav Goljan Determining Digital Image Origin Using Sensor Imperfections. Proceedings of the SPIE-2005, Vol. 5685, pp. 249-260.
2. Mehdi, K.L. Sencar, H.T. Memon, N. Blind source camera identification. International Conference on Image Processing, 2004, Vol. 1, pp. 709- 712.
3. Lukáš J., Fridrich J., and Goljan M.: “Determining Digital Image Origin Using Sensor Imperfections”, Proc. SPIE Electronic Imaging, Image and Video Communication and Processing, San Jose, California, January 16–20, 2005, pp. 249–260.
4. Kharrazi, M., Sencar, H. T., and Memon, N.: “Blind Source Camera Identification”, Proc. ICIP’ 04, Singapore, October 24–27, 2004. pp. 312-317.
5. Рублёв Д.П., Чумаченко А.Б. Идентификация цифровых фотокамер по карте светочувствительности матрицы // XIII Всероссийская научно-практическая конференция “Проблемы информационной безопасности в системе высшей школы”. – М.: МИФИ, 2007. – С. 78-79.
6. Рублёв Д.П., Чумаченко А.Б., Макаревич О.Б., Фёдоров В.М. Идентификация цифровых микрофонов по неидеальностям тракта записи // Известия ЮФУ. Технические науки. Тематический выпуск “Информационная безопасность”. – Таганрог: Изд-во ТТИ ЮФУ, 2007. – С. 84-92. №1(76).

УДК 681.324

Ю.А. Брюхомицкий

КЛАССИФИКАЦИЯ НЕСТАЦИОНАРНЫХ ВЕРОЯТНОСТНЫХ БИОМЕТРИЧЕСКИХ ПАРАМЕТРОВ ЛИЧНОСТИ*

Применяемые в настоящее время динамические биометрические системы идентификации (БСИ) личности (по голосу, рукописному и клавиатурному почеркам) основаны на анализе индивидуальных особенностей динамики подсознатель-

*Работа выполнена при поддержке грантов РФФИ: № 08-07-00117а; № 06-07-96609-р_юг_а, 06-07-89010-а.

ных движений, которые в общем случае могут быть представлены N -мерными векторами биометрических признаков $\mathbf{V} \in R^N$. В математической постановке такие системы решают задачу классификации векторов биометрических признаков \mathbf{V} на классы: «свой» – \mathbf{V}^+ и «чужой» – \mathbf{V}^- . Построение решающих правил для классификации основано на сравнении векторов \mathbf{V} лиц, претендующих на доступ, с эталонными векторами \mathbf{V}_Ω^k , $k = \overline{1, M}$ всех M зарегистрированных и хранящихся в памяти БСИ «своих» пользователей.

Вероятностный характер векторов биометрических признаков \mathbf{V} приводит к необходимости построения среднестатистических эталонов \mathbf{V}_Ω на основе серии образцов, предъявленных системе в режиме обучения. Построение хороших решающих правил для классификации биометрических признаков \mathbf{V} обычно затруднено по следующим причинам:

- отсутствуют сведения о законах и числовых характеристиках распределения векторов \mathbf{V} ;
- имеет место статистическая нестабильность распределения векторов \mathbf{V} ;
- отсутствует представительная выборка обучающих векторов \mathbf{V}^- «чужих» пользователей;
- ограничен размер представительной выборки обучающих векторов \mathbf{V}^+ «своих» пользователей (во избежание ухудшения потребительских качеств системы);
- велика размерность пространства входных данных (векторов \mathbf{V}).

Векторы \mathbf{V} можно рассматривать как значения N -мерной случайной величины ξ , распределение которой описывает статистическую изменчивость биометрических параметров пользователя и характеризуется плотностью $p(\mathbf{V})$. Соответственно распределения векторов биометрических признаков \mathbf{V}^+ «своих» и \mathbf{V}^- «чужих» пользователей будут характеризоваться плотностями $p(\mathbf{V}^+)$ и $p(\mathbf{V}^-)$.

В реальных условиях распределения $p(\mathbf{V}^+)$ и $p(\mathbf{V}^-)$ неизвестны, поэтому возникает необходимость в построении их оценок $\hat{p}(\mathbf{V}^+)$ и $\hat{p}(\mathbf{V}^-)$ на основе обучающих выборок векторов \mathbf{V}^+ и \mathbf{V}^- . Однако если формирование ограниченной обучающей выборки векторов \mathbf{V}^+ не вызывает принципиальных трудностей, то формирование даже сильно ограниченных обучающих выборок векторов \mathbf{V}^- «всевозможных чужих» является серьезной проблемой [1, 2]. В связи с этим возникает задача построения хороших решающих правил для классификации векторов \mathbf{V}^+ и \mathbf{V}^- в условиях наличия только ограниченной обучающей выборки векторов \mathbf{V}^+ : $\Psi^+ = (\mathbf{V}_1^+, \mathbf{V}_2^+, \dots, \mathbf{V}_L^+)$.

В общем случае решающее правило должно выполнять классификацию векторов биометрических признаков \mathbf{V} на M классов (по числу зарегистрированных в БСИ пользователей). Однако в БСИ задачу классификации векторов \mathbf{V} на M классов без нарушения общности всегда можно свести к задаче классификации векторов \mathbf{V} только на два класса: вектор \mathbf{V}^+ – «свой» и вектор \mathbf{V}^- – «чужой». Это становится возможным благодаря принятой в БСИ двухэтапной процедуре идентификации/аутентификации. На первом этапе – идентификации – претендующий на доступ k -пользователь ($k = \overline{1, M}$) в рамках стандартной процедуры ОС предъявляет свой символьный идентификатор, инициируя внесение в структуру классификатора своего персонального биометрического эталона \mathbf{V}_Ω^k . На втором этапе – собственно аутентификации – k -пользователь предъявляет свои биометрические харак-

теристики в виде вектора \mathbf{V} , классификатор сравнивает предъявленный вектор \mathbf{V} с эталоном \mathbf{V}_3^k этого пользователя и выносит решение о допуске.

Учитывая отмеченную выше особенность аутентификации в БСИ, решающее правило может строиться в расчете на использование в текущем сеансе аутентификации только одного биометрического эталона \mathbf{V}_3 .

При классификации векторов \mathbf{V}^+ и \mathbf{V}^- задача построения решающего правила сводится, по существу, к заданию подходящей гиперповерхности $G(\mathbf{V})$, реализующей разделение векторов \mathbf{V}^+ и \mathbf{V}^- . В свою очередь, выбор способа задания $G(\mathbf{V})$ определяется априорными сведениями о распределении векторов \mathbf{V}^+ .

В том случае, если параметры распределения векторов \mathbf{V}^+ *a priori* неизвестны, то разделяющая гиперповерхность $G(\mathbf{V})$ может быть задана в явном виде с использованием m параметров-весов:

$$G(\mathbf{V}) = G(\mathbf{V}, w_1, w_2, \dots, w_m),$$

а построение решающего правила сведено к выбору вида $G(\mathbf{V})$ и подбору параметров-весов, такому, чтобы гиперповерхность $G(\mathbf{V})$ с приемлемой точностью могла отделить область распределения векторов \mathbf{V}^+ от остального гиперпространства. Такой способ задания гиперповерхности $G(\mathbf{V})$ широко используется в геометрических и нейросетевых методах классификации. Однако геометрические методы в задаче классификации векторов \mathbf{V}^+ и \mathbf{V}^- дают обычно высокий уровень ошибок, а нейросетевые методы для оптимального построения гиперповерхности $G(\mathbf{V})$ требуют наличия представительной обучающей выборки векторов \mathbf{V}^- , что является самостоятельной проблемой [1, 2].

В том случае, если параметры распределения векторов \mathbf{V}^+ *a priori* известны, целесообразно использовать параметрические методы обучения, когда гиперповерхность $G(\mathbf{V})$ может быть задана в явном виде от параметров обучающего множества [3]. Качество решающего правила при этом определяется степенью соответствия реальных и использованных в $G(\mathbf{V})$ параметров распределения.

При построении решающего правила для классификации векторов \mathbf{V}^+ и \mathbf{V}^- можно также руководствоваться видом области распределения. Если область распределения векторов \mathbf{V}^+ выпуклая ограниченная и имеет один выраженный центр распределения, то в большинстве случаев хорошим приближением для аппроксимации распределения векторов \mathbf{V}^+ является гауссово распределение [3]. Такая аппроксимация широко распространена и во многих случаях дает хороший эффект. Однако в динамических БСИ обычно наблюдаются существенные флуктуации контролируемых биометрических параметров, обусловленные суточными биоритмами, психофизическим состоянием в момент прохождения процедуры идентификации и другими факторами [4, 5]. Это приводит к тому, что область распределения векторов \mathbf{V}^+ становится невыпуклой или даже распадается на несколько подобластей. В таких случаях аппроксимация области на основе нормального распределения становится излишне идеализированной и приводит к повышению уровня ошибок классификации. Рис. 1 иллюстрирует образование различного вида ограниченных областей распределения векторов \mathbf{V}^+ для $N=2$.

Таким образом, при распределении векторов \mathbf{V}^+ в виде невыпуклых областей или областей, состоящих из отдельных подобластей, по разным причинам плохо подходят все традиционные методы задания $G(\mathbf{V})$. Это обусловлено тем, что контролируемая в БСИ динамика подсознательного воспроизведения заученных действий представляет собой нестационарный случайный процесс, для которого ис-

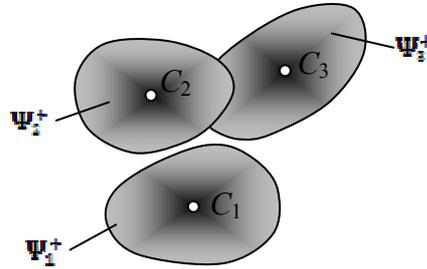


Рис. 2. Образование невыпуклой и частично распавшейся области распределения векторов V^+ на основе трех выпуклых ограниченных подобластей для случая $N=2$

Конфигурацию каждой выпуклой подобласти зададим в виде функции потенциала, конкретный вид которой будет определяться условиями задачи. В итоге невыпуклые ограниченные или распавшиеся области распределения векторов V^+ будут аппроксимированы объединением L выпуклых областей специального вида, каждая из которых соответствует одному из L образцов обучающей выборки Ψ^+ . Ситуацию для размерности $N=2$ иллюстрирует рис. 3.

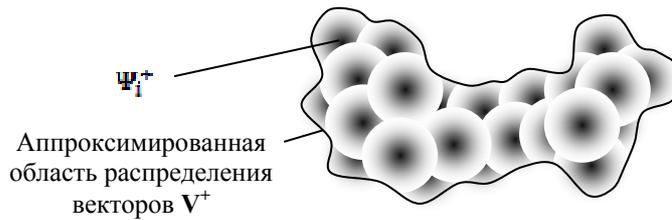


Рис. 3. Аппроксимация невыпуклой ограниченной области объединением выпуклых областей специального вида

Такой подход к построению решающего правила с некоторой долей условности можно отнести к параметрическим методам. Условность состоит в том, что априорное знание параметров распределения сводится к предположению, что распределение векторов V^+ описывается невыпуклыми ограниченными или распавшимися областями, которые затем моделируются объединением специально подобранных функций потенциала.

Пусть область распределения биометрических параметров «своего» пользователя задана множеством учебных данных, содержащих L векторов V_i^+ , $i = \overline{1, L}$, случайно распределенных в N -мерном пространстве:

$$V_i^+ = (v_1^+, v_2^+, \dots, v_N^+), \quad j = \overline{1, N}, \quad i = \overline{1, L}.$$

Для оценки плотности распределения каждого учебного вектора V_i^+ , $i = \overline{1, L}$ используем функцию потенциала, задающую конфигурацию каждой подобласти, в виде упрощенной формы функции Гаусса [6] с центром в точке V_i^+

$$\hat{p}(\mathbf{V}_i^+) = \varphi_i(\mathbf{V}) = \exp\left(-\frac{1}{2\mu^2}\|\mathbf{V} - \mathbf{V}_i^+\|^2\right), \quad i = \overline{1, L}, \quad (1)$$

где μ – параметр, задающий ширину функции $\varphi(\mathbf{V})$.

Форма (1) отличается от классической функции гаусса отсутствием коэффициента $1/\sigma\sqrt{2\pi}$ перед экспонентой. Это позволяет получить максимальное значение функции (1), равное единице, а не величине указанного коэффициента.

В покомпонентном представлении выражение (1) будет иметь вид

$$\hat{p}(\mathbf{V}_i^+) = \exp\left(-\frac{1}{2\mu^2}\sum_{j=1}^N (v_j - v_{ij}^+)^2\right), \quad i = \overline{1, L}, \quad j = \overline{1, N}. \quad (2)$$

Всё множество учебных данных $\Psi^+ = \{\mathbf{V}_i^+\}$, $i = \overline{1, L}$ представляет собой матрицу $\mathbf{V}_i^+ = \|v_{ij}^+\|$. Оценку функции плотности распределения для всего множества можно получить путем суммирования функций (1) [6]:

$$\Phi(\mathbf{V}) = \sum_{i=1}^L \varphi(\mathbf{V}) = \sum_{i=1}^L \exp\left(-\frac{1}{2\mu^2}\|\mathbf{V} - \mathbf{V}_i^+\|^2\right). \quad (3)$$

В покомпонентном представлении выражение (3) будет иметь вид

$$\Phi(\mathbf{V}) = \sum_{i=1}^L \exp\left(-\frac{1}{2\mu^2}\sum_{j=1}^N (v_j - v_{ij}^+)^2\right). \quad (4)$$

Входные данные классификатора (вектора \mathbf{V} и \mathbf{V}_i^+) удобно нормализовать к единичной длине:

$$(v_j)' = v_j / \sqrt{\sum_{j=1}^N v_j^2}, \quad (v_{ij}^+)' = v_{ij}^+ / \sqrt{\sum_{j=1}^N (v_{ij}^+)^2}. \quad (5)$$

Оценка функции плотности распределения для одного учебного вектора \mathbf{V}_i^+ при нормализованных входных данных в покомпонентном представлении на основе (2) и (5) примет вид

$$\varphi(\mathbf{V}) = \exp\left(\frac{1}{2\mu^2}\sum_{j=1}^N 2v_j \cdot v_{ij}^+ / \sqrt{\sum_{j=1}^N v_j^2} \cdot \sqrt{\sum_{j=1}^N (v_{ij}^+)^2} - 2\right) = \exp\left(\frac{1}{\mu^2}\sum_{j=1}^N (v_j)' \cdot (v_{ij}^+)' - 1\right). \quad (6)$$

Оценка функции плотности распределения для всего множества Ψ^+ при нормализованных входных данных в покомпонентном представлении на основе (4) и (5) соответственно примет вид

$$\Phi(\mathbf{V}) = \sum_{i=1}^L \exp\left(\frac{1}{\mu^2} \sum_{j=1}^N (v_j)^+ \cdot (v_{ij}^+)' - 1\right). \quad (7)$$

Задачу построения решающего правила для классификации векторов \mathbf{V}^+ и \mathbf{V}^- в условиях наличия только ограниченной обучающей выборки векторов $\mathbf{V}_i^+ = (\mathbf{V}_i^+, \mathbf{V}_i^+, \dots, \mathbf{V}_i^+)$ можно решить путем задания разделяющей гиперповерхности $G(\mathbf{V})$ в виде порогового значения суммарной плотности вероятности $\Phi_n(\mathbf{V})$, являющегося минимально допустимым для «своего» пользователя.

Пороговое значение уровня суммарной плотности распределения $\Phi_n(\mathbf{V})$ можно получить следующим способом. Выражение (7) фактически реализует вычисление евклидовых расстояний от неизвестного образца до всех учебных образцов, которые затем с помощью функции потенциала преобразуются в значение суммарной плотности вероятности $\Phi(\mathbf{V})$. Заменим неизвестный образец на один из учебных (k -образец из $k = \overline{1, L}$) и вычислим евклидовы расстояния от этого образца до остальных ($L-1$) учебных образцов. С помощью функции потенциала эти расстояния преобразуются, в конечном итоге, в значение суммарной плотности вероятности $\Phi(\mathbf{V}_k)$. Повторим эту процедуру последовательно для всех L учебных векторов $\mathbf{V}_k^+, k = \overline{1, L}, i \neq k$. Выражение (7) для этих вычислений будет иметь вид:

$$\Phi(\mathbf{V}_k^+) = \frac{L}{L-1} \sum_{i=1}^{L-1} \exp\left(\frac{1}{\mu^2} \sum_{j=1}^N (v_{kj}^+)' \cdot (v_{ij}^+)' - 1\right), \quad k = \overline{1, L}, \quad i = \overline{1, L-1}, \quad i \neq k. \quad (8)$$

Суммарное значение плотности вероятности $\Phi(\mathbf{V}_k^+)$, вычисленное по формуле (8), учитывает распределение ($L-1$) образцов, т.е. на один меньше, чем суммарное значение плотности вероятности $\Phi(\mathbf{V})$, вычисленное по формуле (7) для неизвестного образца. Поэтому в формулу (8) включен поправочный коэффициент пропорциональности $K=L/L-1$.

Ранжируем теперь полученные значения плотности вероятности $\Phi(\mathbf{V}_k^+)$, $k = \overline{1, L}$ по величине и выберем минимальное из них $\Phi_{\min}(\mathbf{V}_k^+)$. Его можно трактовать как пороговое значение уровня суммарной плотности распределения $\Phi_n(\mathbf{V})$, являющееся минимально допустимым для «своего» пользователя, т.е. как искомую разделяющую поверхность $G(\mathbf{V})$.

В БСИ представление векторов \mathbf{V}_i^+ сопровождается ошибкой первого рода – вероятностью $P1$ ложного отказа «своему». Поэтому разделяющую поверхность $G(\mathbf{V})$ вокруг области «свой» следует «расширить» на величину этой ошибки. В итоге, выражение для порогового значения плотности вероятности $\Phi_n(\mathbf{V})$, соответствующего разделяющей гиперповерхности $G(\mathbf{V})$ будет иметь вид

$$\Phi_n(\mathbf{V}) = \min \Phi(\mathbf{V}_k^+) = \frac{L}{L-1} \sum_{i=1}^{L-1} \exp\left(\frac{1}{\mu^2} \sum_{j=1}^N (v_{kj}^+)' \cdot (v_{ij}^+)' - 1\right) + P1, \quad k = \overline{1, L}, \quad i = \overline{1, L-1}, \quad i \neq k. \quad (9)$$

Образование разделяющей поверхности $G(\mathbf{V})$ для классификации векторов \mathbf{V}^+ и \mathbf{V}^- при $N=2$ иллюстрирует рис.4.

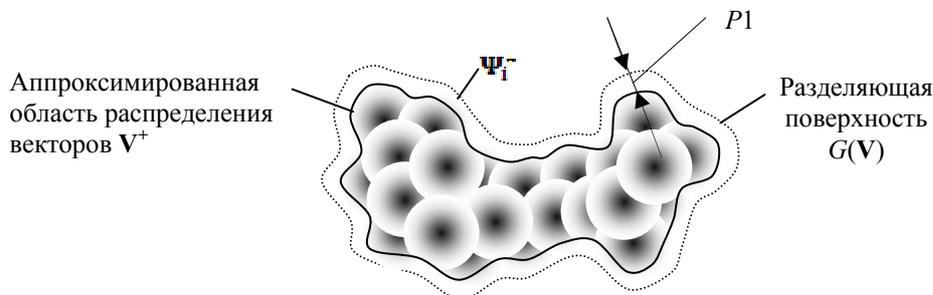


Рис. 4. Образование разделяющей поверхности $G(\mathbf{V})$ для классификации векторов \mathbf{V}^+ и \mathbf{V}^- при $N=2$

Искомое решающее правило для классификации векторов \mathbf{V}^+ и \mathbf{V}^- в условиях наличия только ограниченной обучающей выборки векторов \mathbf{V}_i^+ формулируется следующим образом: если $\Phi(\mathbf{V}) \geq \Phi_n(\mathbf{V})$, то предъявленный вектор относится к классу \mathbf{V}^+ , в противном случае – к классу \mathbf{V}^- .

В конечном итоге решающее правило приобретает вид

$$\mathbf{V} \in \begin{cases} \mathbf{V}^+, & \text{если } \text{sign}[\Phi(\mathbf{V}) - \Phi_n(\mathbf{V})] = 1; \\ \mathbf{V}^-, & \text{если } \text{sign}[\Phi(\mathbf{V}) - \Phi_n(\mathbf{V})] = 0. \end{cases} \quad (10)$$

БСИ на основе решающего правила (10), получается простой, допускает использование неточных биометрических данных и обеспечивает приемлемую точность классификации на ограниченных наборах учебных образцов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Брюхомицкий Ю.А., Казарин М.Н. Метод обучения нейросетевых биометрических систем на основе построения аппроксимированных областей / Известия ТРТУ. Тематический выпуск «Материалы V международной научно-практической конференции «Информационная безопасность». – Таганрог: Изд-во ТРТУ, 2003. – С. 155-159. – № 4(33).
2. Брюхомицкий Ю.А., Казарин М.Н. Метод обучения нейросетевых биометрических систем на основе копирования областей / Электронный журнал «Перспективные информационные технологии и интеллектуальные системы». – 2003. – С. 17-23. – №3 (15). – <http://pitis.tsure.ru>.
3. Брюхомицкий Ю.А., Казарин М.Н. Параметрическое обучение биометрических систем контроля доступа / Вестник компьютерных и информационных технологий. – М.: Изд-во «Машиностроение», 2006. – С. 6-13. – № 2 (20).
4. Иванов А.И. Биометрическая идентификация личности по динамике подсознательных движений. – Пенза: Изд-во Пенз. гос. ун-та, 2000. – 188 с.
5. Широкин В.П., Кулик А.В., Марченко В.В. Динамическая аутентификация на основе анализа клавиатурного почерка. – http://www.masters.donntu.edu.ua/2002/fvti/aslamov/files/bio_authentication.htm.
6. Каллан Р. Основы концепции нейронных сетей.: Пер. с англ. – М.: Издательский дом «Вильямс», 2001. – 287 с.