

УДК 004.93:[57.087.1]:518.22

Математический аппарат биометрических устройств

Белоусов С.А., 12-р

Актуальность: в связи с активным внедрением биометрических устройств в повседневную жизнь в работе рассматриваются некоторые принципы их функционирования.

Цель работы: углубить знание по данной проблеме.

Биометрия – это научная дисциплина, изучающая способы измерения различных параметров человека с целью установления сходства или различий между людьми и выделения одного конкретного человека из множества других людей, или – наука, изучающая методики распознавания конкретного человека по его индивидуальным параметрам.

Построение решающих правил в системах биометрической идентификации основано на создании некоторых эталонных представлений идентифицируемых лиц. Эти эталоны хранятся в памяти системы, контролирующей доступ, и служат для сравнения с биометрическими параметрами лиц, претендующих на доступ к ресурсам. В случае, когда измеренные системой значения параметров пользователя значимо отличаются от эталона, он получает отказ в доступе к ресурсам.

В вероятностной формулировке такие решающие правила приводят к необходимости построения среднестатистических эталонов на основе образцов, предъявленных системе в режиме обучения. Обозначим $X \in R^N$ - вектор признаков, поступающий на вход биометрической системы идентификации (БСИ) и состоящий из N информативных биометрических параметров пользователя. Предположим, что вектор можно рассматривать как значение векторной случайной величины ξ , распределение которой описывает статистическую изменчивость биометрических параметров пользователя. Если распределение случайного вектора ξ , характеризующее легального пользователя, имеет плотность $p_0(X)$ с ограниченным множеством-носителем $X_0 = \{X : p_0(X) \neq 0\} \subset R^N$, то множество

$\mathcal{X}_0^+ \subset \mathcal{X}_0$, удовлетворяющее условию $P(\mathcal{X}_0^+) = 1 - FRR$ (вероятности ложного отказа (false reject rate, FRR)), представляет совокупность тех входных векторов X , которые БСИ будет правильно распознавать как попытку легального доступа к ресурсам.

Пусть $p_1(X)$ - плотность распределения вероятностей, соответствующая биометрическим параметрам другого лица, пытающегося получить доступ, и обозначим $\mathcal{X}_1 \subset \mathcal{R}^n$ носитель этого распределения, то есть $\mathcal{X}_1 = \{X : p_1(X) \neq 0\}$. Тогда, очевидно, множество $\mathcal{X}_0^+ \cap \mathcal{X}_1$ будет состоять из тех входных векторов биометрических параметров, которые БСИ будет ошибочно признавать “своими”. Согласно требованиям, предъявляемым к БСИ, должно быть $P(\mathcal{X}_0^+ \cap \mathcal{X}_1) \leq FAR$ (вероятности ложного допуска (false acceptance rate, FAR)).

В реальных задачах, связанных с созданием БСИ, распределения $p_0(X)$, $p_1(X)$ неизвестны, соответственно, неизвестны и множества \mathcal{X}_0 , \mathcal{X}_1 , поэтому возникает необходимость в построении оценок распределений $\hat{p}_0(X)$, $\hat{p}_1(X)$, или хотя бы оценок множеств-носителей R_0, R_1 . Предполагается, что для построения этих оценок имеется обучающая выборка, состоящая из классифицированных образцов входных векторов $X_1, X_2, \dots, X_n; \delta_1, \delta_2, \dots, \delta_n$. То есть, относительно каждого из векторов X_i известно, соответствует ли он легальному пользователю $\delta_i = 0$, либо одному из набора “чужих” пользователей: $\delta_i \in \{1, 2, \dots, K\}$.

Если доступна достаточная по объему и представительная обучающая выборка, дающая представление о легальном пользователе и “всех чужих”, то построение оценок может осуществляться обычными методами математической статистики через оценивание распределений вероятностей $\{p_k(X), k = 0, 1, \dots\}$ или непосредственное построение решающих правил, определяющих границы в пространстве \mathcal{R}^n между областями “свой” - “чужой”.

Чаще всего при построении БСИ разработчик не располагает такой полнотой информации:

- а) обучающая выборка “свой” имеет ограниченный объем;
- б) представительный набор “чужих” вообще сложно сформировать;

в) выбор вероятностных распределений $p_+(X)$ является сложным вопросом вследствие отсутствия математических моделей формирования биометрических измерений.

Искусственные нейронные сети (ИНС), благодаря их адаптивности и способности к обобщению функциональных зависимостей, широко применяются в задачах анализа статистических наблюдений сложной структуры, в том числе и в построении БСИ.

Пусть $\{X_1, X_2, \dots, X_n\} \subset \mathbb{K}_b$ - выборка векторов биометрических параметров легального пользователя. Геометрическая конфигурация многомерного множества \mathbb{K}_b в пространстве признаков R^M в общем случае неизвестна и может быть весьма сложной. Вся имеющаяся информация о множестве \mathbb{K}_b содержится в выборке X_1, X_2, \dots, X_n , однако, чтобы не хранить в памяти всю эту выборку и в то же время иметь достаточную информацию для использования ее в решающем правиле, можно вместо выборки X_1, X_2, \dots, X_n сохранять некоторый набор центров U_1, U_2, \dots, U_M , ($M \geq 1$), который должен достаточно точно аппроксимировать исходную выборку. Если количество центров M существенно меньше объема выборки n , то это позволит упростить вычислительные процедуры и сэкономить объем памяти.

В качестве критерия для выбора центров U_1, U_2, \dots, U_M можно использовать среднеквадратичный разброс наблюдаемых векторов относительно этого набора центров. Математически это может быть сформулировано как задача минимизации функционала среднего риска, соответствующего выбору квадратичных функций потерь,

$$R(U_1, \dots, U_M) = \int_{\mathbb{K}_b} \sum_{i=1}^M J_i(X) \|X - U_i\|^2 p_b(X) dX, \text{ где}$$

$$J_i(X) = \begin{cases} 1, & \|X - U_i\|^2 \leq \|X - U_j\|^2 \forall j \\ 0 & \end{cases}$$

Алгоритм оценивания оптимального набора центров, устроен следующим образом:

а) выбирают подходящие начальные приближения $U_1^{(1)}, U_2^{(1)}, \dots, U_M^{(1)}$ (это могут быть произвольные несовпадающие векторы из R^M);

б) затем, последовательно просматривая выборку, осуществляют итерации:

$$U_i(t+1) = U_i(t) + \gamma_i(X_i - U_i(t)), \text{ если } J_i(X_i) = 1$$

$$U_i(t+1) = U_i(t), \text{ для всех } i \neq I,$$

$$t = 1, 2, 3, \dots$$

В теории алгоритмов самообучения доказано, что в пределе $t \rightarrow \infty$ последовательность оценок центров $\{U_i(t)\}$ сходится к центрам тяжести множеств D_1, D_2, \dots, D_M , образующих разбиение множества \mathbb{X}_0 ,

$$U_i = \frac{\int_{D_i} X p_0(X) dX}{\int_{D_i} p_0(X) dX}$$

$$D_i = \left\{ X : \|X - U_i\|^2 \leq \|X - U_j\|^2 \forall j \right\}$$

С помощью построенных оценок центров U_1, U_2, \dots, U_M можно с достаточной степенью точности аппроксимировать множество \mathbb{X}_0 ; данный набор центров $\{U_i\}$ и является эталонным представлением биометрического образа легального пользователя. Если затем на вход БСИ поступает вектор параметров X неизвестного пользователя, то исчерпывающая информация о расположении этого вектора X по отношению к множеству \mathbb{X}_0 будет содержаться в наборе расстояний от X до всех центров эталона $\{U_i\}$:

$$\Delta(X, U) = \{\delta_i : \delta_i = \|X - U_i\|, i = 1, 2, \dots, M\}$$

Поэтому на основе вектора $\Delta(X, U)$ уже принимается решение о принадлежности данного вектора X .

После синтеза эталона $\{U_i\}$ на основе новой обучающей выборки $X_{n+1}, X_{n+2}, \dots, X_{n+t}$, содержащей как векторы параметров легального пользователя, так и векторы “чужих”, оценивается пороговое значение ρ_0 так, чтобы решающее правило вида: “если $\rho(X, U) > \rho_0$, то X – чужой” удовлетворяло на этой обучающей выборке требованиям по величинам FAR и FRR. Если удовлетворительного порогового решающего правила не удастся построить, это может означать, что структура многомерного множества \mathbb{X}_0 сложна и не поддается точной аппроксимации небольшим набором эталонов. Кроме того, недостаток информации о статистике векторов “чужих” не позволит дать точную оценку надежности БСИ.

Для снижения вероятностей ошибок предлагается использовать не один эталон для легального пользователя, а набор эталонов, порождаемых разными нейронными сетями. Приведенный выше в качестве примера алгоритм является самым простым алгоритмом самообучения, он эффективен для распределений $P_b(X)$, образующих компактные сферические кластеры в R^N и осуществляет векторное квантование пространства признаков.

Вывод: при проведении работы были углублены знания по данной теме, и рассмотрена общая схема построения чисто эмпирических функционалов, которые могут служить основой для принятия решений в задачах идентификации пользователей.

Список литературы

1. Биометрическая аутентификация пользователя [Электронный ресурс] / <http://ypn.ru/296/biometric-authentication/>
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. № 3 (15), С. 17-23 <http://pitis.tsure.ru>
3. Осовский С. Нейронные сети для обработки информации. – М., Финансы и статистика, 2002. – 344 с.
4. Применение нейронных сетей для формирования эталонов в системах биометрической идентификации личности [Электронный ресурс] /ИНФОРМАЦИОННОЕ ПРОТИВОДЕЙСТВИЕ УГРОЗАМ ТЕРРОРИЗМА Научно-практический журнал - Режим доступа: <http://www.contrterror.tsure.ru/site/magazine7/02-06-Lihovidov-Gerasimec-Kornushin.htm>