ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

Иммунологический метод текстонезависимой идентификации личности по голосу

Брюхомицкий Юрий Анатольевич, кандидат технических наук, доцент
Федоров Владимир Михайлович кандидат физико-математических наук, доцент
Южный Федеральный университет (г. Таганрог)

Предлагается иммунологический подход к решению задачи текстонезависимой идентификации личности по голосу, основанный на принципах представления и обработки речевой информации, принятых в искусственных иммунных системах Речевой сигнал, подлежащий идентификации, представляется последовательностью векторов признаков, составленных из кепстральных коэффициентов. Последующий анализ такой последовательности осуществляется на основе иммунной модели отрицательного отбора.

Ключевые слова: *текстонезависимая идентификация личности по голосу; искусственные иммунные системы; иммунологическая модель отрицательного отбора.*

Введение. Исследования и разработки систем идентификации личности по голосу ведутся уже около 50 лет и остаются актуальными по настоящее время. Интерес к проблеме идентификации личности по голосу обусловлен, прежде всего, преимуществами такого способа проверки подлинности личности: голос невозможно украсть и очень трудно подделать, в процессе идентификации не требуется непосредственный контакт с пропускной системой. Голосовая идентификация личности может применяться при контроле прав доступа как к физическим, так и информационным объектам. Большие перспективы применения таких систем в телефонии, электронной коммерции, криминалистике, разведке и контрразведке, антитеррористическом мониторинге, системах голосового управления и многих других областях.

Системы идентификации личности по голосу делятся на тектозависимые и тектонезависимые. Первые – менее сложны в реализации, но весьма уязвимы для атак воспроизведения записанного текста. Вторые обеспечивают идентификацию личности при воспроизведении ею любого текста на любом языке, но существенно сложнее в реализации и требуют больше времени для проведения процедуры идентификации.

Постановка задачи. В данной работе предлагается иммунологический подход к решению задачи текстонезависимой идентификации личности по голосу, применяемый в искусственных иммунных системах (ИИС) [1, 2]. В математической модели, описывающей процесс анализа речевого сигнала, использован фильтр высокого порядка и кепстральный анализ, реализуемый на основе коэффициентов линейного предсказателя. В итоге оцифрованный речевой представляется последовательностью векторов признаков, составленных из кепстральных коэффициентов, последующий анализ которой осуществляется на основе иммунологической модели отрицательного отбора.

Решение поставленной задачи. Для формирования векторов признаков при идентификации диктора использована модель, предложенная Фантом [3], в которой речевой сигнал образуется путем прохождения через фильтр высокого порядка. Фильтр возбуждается либо последовательностью периодических импульсов, в результате чего получаются гласные, а также звонкие или сонорные согласные, либо случайным шумом с широким спектром, – в результате получаются глухие согласные.

Учитывая, что сигнал возбуждения и импульсная характеристика фильтра взаимодействуют через операцию свертки, задача анализа речевого сигнала сводится к разделению возбуждающего сигнала и сигнала на выходе фильтра. Разделение возможно на основе концепции гомоморфных систем [4, 5] в которых речевой сигнал рассматривается как суперпозиция возбуждающего сигнала и накладывающегося на него сигнала, при прохождении через речевой тракт [5]. В этом случае речевой сигнал $s(n), n = 0, 1, \ldots$ можно представить в виде свертки в следующем виде:

$$s(n) = d(n) \otimes v(n),$$

где n – временные точки отсчета речевого сигнала;
$d(n)$ – возбуждающий сигнал;
$v(n)$ – импульсный отклик речевого тракта;
$\otimes$ – операция свертки.

В частотной области свертка имеет вид:

$$S(\omega) = D(\omega) \cdot V(\omega).$$

Если взять комплексный логарифм данного равенства, получим:

$$\log[D(\omega) \cdot V(\omega)] = \log[D(\omega)] + \log[V(\omega)].$$

Из данного выражения следует, что в логарифмической области период основного тона и параметры голосового тракта наложены друг на друга и могут быть разделены с помощью обычных методов обработки сигнала.

Для получения кепстральных коэффициентов используется линейное предсказание речи, которое относится к классу методов параметрического моделирования, при этом спектр моделируется как авторегрессионый процесс.

Модель речевого сигнала $s(n)$ в этом случае представляется как линейная комбинация его предыдущих отсчетов:

$$s(n) = -\sum_{i=1}^{N_{LP}} a_{LP}(i) \cdot s(n-i) + e(n),$$

где N_{LP} – число коэффициентов модели (порядок предсказания);
a_{LP} – коэффициенты линейного предсказания;
$e(n)$ – функция ошибки модели (разность между предсказанным и реально измеренным значением).

ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

Выражение может быть переписано в виде z-преобразования и представлено как операция линейной фильтрации:

$$E(z) = H_{LP}(z) \cdot S(z),$$

где $E(z)$ и $S(z)$ – z-преобразование сигнала ошибки и речевого сигнала соответственно, а

$$H_{LP}(z) = \sum_{i=1}^{N_{LP}} a_{LP}(i) \cdot z^{-i},\ a_{LP}(0) \equiv 1$$

называется инверсным фильтром линейного предсказания.

Применяется несколько методов вычисления коэффициентов линейного предсказания. В данной работе использован авторегрессионный метод [6], что связано с его вычислительной эффективностью и присущей ему стабильностью.

Для вычисления коэффициентов предсказания использована рекурсия Левинсона-Дарбина [6]. Если фильтр линейного предсказания стабилен (а стабильность его гарантируется при автокорреляционном методе), то логарифм обратного фильтра может быть выражен как энергетический ряд [5]:

$$C_{LP} = \sum_{i=1}^{N_{LP}} a_{LP}(i) \cdot z^{-i} = \log(C_{LP} \Big/ \sum_{j=1}^{N_{LP}} a_{LP}(j) \cdot z^{-j}).$$

Кепстральные коэффициенты находятся путем дифференцирования обеих сторон выражения относительно z^{-1} и вычислением их из полученных полиномов. Это делается с помощью следующей рекурсии [6]:

Инициализация:

$C_{LP}(1) = -a_{LP}(1)$

$for(i = 2; i \le N_C; i++)$

{

$$C_{LP} = -a_{LP}(i) - \sum_{j=1}^{j-1} \left(1 - \frac{j}{i}\right) \cdot a_{LP}(j) \cdot C_{LP}(i-j)$$

}

Здесь $a_{LP}(i)$ – коэффициенты линейного предсказания, $C_{LP}(i)$ – кепстральные коэффициенты.

При вычислении кепстральных коэффициентов на основе приведенной выше рекурсии не указывается значение N_C, определяющее их количество. Проблема заключается в том, что они являются результатом обратного Фурье-преобразования импульсного отклика модели линейного предсказания. Однако модель линейного предсказания сигнала является фильтром с бесконечной импульсной характеристикой. Следовательно, теоретически можно вычислить бесконечное число кепстральных коэффициентов. На практике число кепстральных коэффициентов выбирается сравнимым с числом коэффициентов линейного предсказания: $0{,}75p < N_C < 1{,}25p$, где p – число коэффициентов линейного предсказания.

Исследования в области динамической биометрии [7, 8], показывают, что индивидуальные особенности личности в наибольшей степени проявляются при воспроизведении не одиночных символов (фонем) текста, а синтаксически связанных фрагментов текста. Использование этого феномена при анализе позволяет строить системы биометрической идентификации личности с существенно более высокими характеристиками по точности. Для использования этого феномена речевой сигнал $s(n)$ с удаленными паузами и шипящими звуками разбивался на временные участки (фрагменты) $i = 1, 2, \dots, n$ по 20 мсек с перекрытием 5 мсек. При этом каждый участок был представлен r-мерным вектором признаков

$\boldsymbol{s}_i = s_1, s_2, \dots, s_r, i = 1, 2, \dots, r,$

а весь речевой сигнал $s(n)$ – последовательностью векторов $\boldsymbol{s}_i$.

$$\mathbf{S}_j = s_1, s_2, \dots .$$

Последовательность $\mathbf{S}_j$, ограниченная N_s элементами

$\bar{\mathbf{S}}_j = \boldsymbol{s}_1, \boldsymbol{s}_2, \dots\ \boldsymbol{s}_{N_s}, j = 1, 2, \dots N_s,$

трактуется как голосовой эталон данной личности (диктора).

Размерность r векторов признаков $\boldsymbol{s}_i$ соответствует числу вычисляемых на каждом участке кепстральных коэффициентов C_{LP}. В проведенных исследованиях $r = 24$. Минимальное $(C_{LP})_{min}$ и максимальное $(C_{LP})_{max}$ значения кепстральных коэффициентов определяют рабочее пространство E^r, в котором распределены векторы признаков $\boldsymbol{s}_i$.

При распознавании голоса в режиме верификации (разделение данных только на 2 класса: «свой» и «чужие») последовательность $\bar{\mathbf{S}}_j, j = 1, 2, \dots N_s$ выступает в качестве эталона «своего».

В отличие от классических методов распознавания образов, основанных на сопоставлении образов с эталоном, здесь предлагается использовать иммунологическую модель отрицательного отбора (МОО), которая реализует децентрализованное распознавание образов, путем их сопоставления с распознающими элементами – детекторами. Сопоставление осуществляется по принципу негативной селекции (срабатывание детектора свидетельствует о том, что предъявленный образ существенно отличается от эталона) [1, 2].

Детекторы имитируют иммунокомпетентные клетки, которые отвечают за распознавание специфических «чужих», т. е. не известных иммунной системе молекул (антигенов).

Популяция детекторов $\mathbf{D}$ создается в метрике векторов $\boldsymbol{s}_i$ эталона $\bar{\mathbf{S}}_j$:

$$\mathbf{D} = \{\boldsymbol{d}_i\} = \boldsymbol{d}_1, \boldsymbol{d}_2, \dots, \boldsymbol{d}_{N_d}, i = 1, 2, \dots N_d;$$

$$\boldsymbol{d}_i = d_1, d_2, \dots, d_{N_d}.$$

Для распознавания «чужих» детекторы $\boldsymbol{d}_i$ должны быть представлены векторами признаков, отличающимися от векторов признаков эталона $\boldsymbol{s}_i$ на некоторую заданную величину δ_0.

Простейший способ создания детекторов $\boldsymbol{d}_i$ популяции $\mathbf{D}$ состоит из двух фаз. В первой фазе осуществляется случай-

ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

ная генерация кандидатов в детекторы $\dot{\boldsymbol{d}}_i$, равномерно распределенных в пространстве признаков E^r. Во второй фазе кандидаты $\dot{\boldsymbol{d}}_i$ сопоставляются с векторами $\boldsymbol{s}^r_{s_i i}$ эталона $\bar{\mathbf{S}}_j$ на основе меры близости Евклида:

$$\delta(\boldsymbol{s}_i, \dot{\boldsymbol{d}}_i) = \sqrt{\sum_{k=1}^{r} (s_{ik} - \dot{d}_{ik})^2}.$$

Если $\delta(\boldsymbol{s}_i, \dot{\boldsymbol{d}}_i) > \delta_0$, то кандидат $\dot{\boldsymbol{d}}_i$ приобретает статус детектора $\boldsymbol{d}_i$, в противном случае кандидат $\dot{\boldsymbol{d}}_j$ уничтожается. По этой процедуре формируется популяция $\mathbf{D} = \{\boldsymbol{d}_i\} = \boldsymbol{d_1}, \boldsymbol{d_2}, \ldots, \boldsymbol{d}_{N_D}$ из N_d детекторов. Останов процедуры может задаваться различными критериями, например временем обучения; числом итераций; размером популяции; предельно допустимым числом неэффективных итераций, не добавляющих новых детекторов. В представленных исследованиях останов процедуры создания популяции $\mathbf{D}$ задавался простым ограничением числа детекторов популяции N_d.

Создание популяции $\mathbf{D} = \{\boldsymbol{d}_i\}$ детекторов завершает фазу обучения ИИС.

В фазе распознавания фрагменты $\boldsymbol{s}_i$ анализируемого голоса, представленного последовательностью $\mathbf{S}_j$, сопоставляются с детекторами $\boldsymbol{d}_i$ популяции $\mathbf{D} = \{\boldsymbol{d}_i\}$ с использованием меры близости Евклида между векторами $\boldsymbol{s}_i$ и $\boldsymbol{d}_i$:

$$\delta(\boldsymbol{s}_i, \boldsymbol{d}_i) = \sqrt{\sum_{k=1}^{r} (s_{ik} - d_{ik})^2}.$$

Критический уровень близости $\delta(\boldsymbol{s}_i, \boldsymbol{d}_i) = \delta_0$ определяет границу для принятия системой решения «свой/чужой» и задается, исходя из допустимых ошибок первого и второго рода.

Если для некоторой пары $\boldsymbol{s}_{il}$ и $\boldsymbol{d}_{im}$ $\delta(\boldsymbol{s}_{il}, \boldsymbol{d}_{im}) < \delta_0$, то считается, что фрагмент $\boldsymbol{s}_{il}$ анализируемого голоса, представленного последовательностью $\mathbf{S}_j$ отсутствует в эталоне $\bar{\mathbf{S}}_j$ и потому с большой вероятностью принадлежит «чужому».

Существенные вариации параметров голоса в последовательностях $\mathbf{S}_j$ и значительные размеры самих последовательностей $\mathbf{S}_j$ определяют целесообразность применения статистического подхода для принятия ИИС решения «свой»-«чужой» [8, 9]. При таком подходе контролируется частота f выполнения условия $\delta(\boldsymbol{s}_{il}, \boldsymbol{d}_{im}) < \delta_0$, которая определяет статистическую вероятность принадлежности анализируемого голоса «чужому»:

$$\hat{P}^{\text{ч}} \approx f = n^+_\delta / n_\delta,$$

где n^+_δ число случаев выполнения условия $\delta(\boldsymbol{s}_{il}, \boldsymbol{d}_{im}) < \delta_0$ в n_δ проведенных операциях сопоставлений $\boldsymbol{s}_i$ с $\boldsymbol{d}_i$.

Принятие решения о принадлежности анализируемого голоса «чужому» считается обоснованным, при превышении частоты f заданного порогового значения $f_{\text{п}}$:

$$\mathbf{S}_j \equiv \begin{cases} \mathbf{Y}^{\text{с}}_j, \text{если } f < f_{\text{п}}; \\ \mathbf{Y}^{\text{ч}}_j, \text{если } f \geq f_{\text{п}}, \end{cases}$$

где $\mathbf{Y}^{\text{с}}_j$ — последовательность биометрических признаков «своего»;

$\mathbf{Y}^{\text{ч}}_j$ — последовательность векторов признаков «чужого».

В качестве платформы для проведения исследований использовался net-book с операционной системой Windows 7. Для записи звука использовалась встроенная в материнскую плату звуковая карта и микрофон. Речевой сигнал оцифровывался с частотой дискретизации 44100 Гц и размером одного отсчета 16 бит моно. Предварительно из речевого сигнала удалялись паузы и шипящие звуки. Это связано с тем, что спектр участков с паузами и шипящими звуками практически одинаков для различных дикторов и близок к белому шуму. Затем речевой сигнал разбивался на участки по 20 мсек с перекрытием 5 мсек. На каждом участке вычислялись вектора признаков C_{ni}, где n число участков, использованных при вычислении векторов, $i = 1, 2, \ldots, N_C, N_C = 24$ — число кепстральных коэффициентов, вычисленных на каждом участке.

Экспериментальные исследования проводились для 9 дикторов, из которых один диктор представлял «своего» и 8 дикторов представляли «чужих». Результаты распознавания голосов дикторов предварительно обученной ИИС представлены на графике (рис. 1).

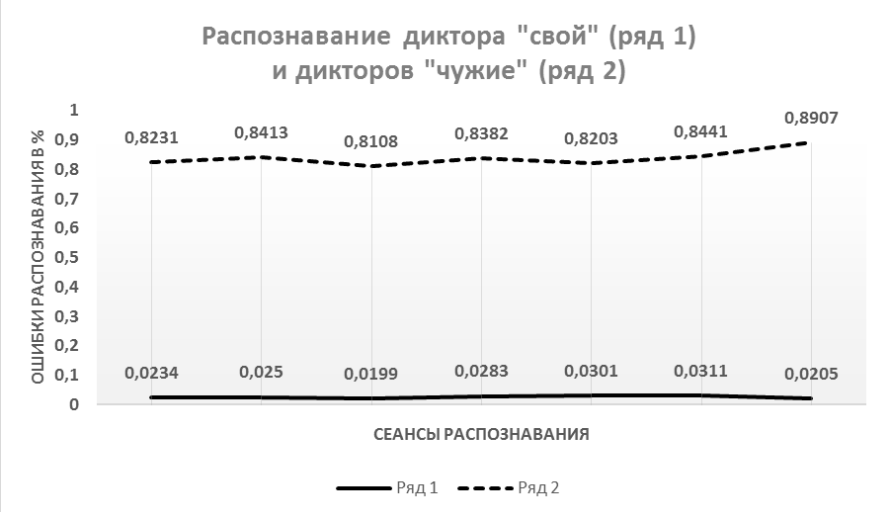

Рис. 1. Результаты распознавания голосов дикторов

Заключение. В работе предложен новый — иммунологический метод текстонезависимой идентификации личности по

ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

голосу и доказана его работоспособность на основе экспериментальных исследований. Показано, что данный метод позволяет проводить идентификацию голоса личности при воспроизведении текста произвольного объема и содержания. Искусственная иммунная система осуществляет непрерывный контроль личности говорящего в темпе поступления голосовых данных, что предоставляет возможность своевременного принятия решения о факте подмене «своего» говорящего на «чужого». Следует ожидать, что полученные здесь экспериментальные ошибки распознавания личности диктора по голосу могут быть существенно снижены при использовании более качественной аппаратуры, условий звукозаписи и уточнения многих эмпирических параметров ИИС.

Литература:

1. Dasgupta D. Artificial Immune Systems and Their Applications, Ed., Springer-Verlag. – 1999.
2. Искусственные иммунные системы и их применение / под ред. Д. Дасгупты; пер. с англ. А. А. Романюхи. – М.: Физматлит, 2006. – 344 с.
3. Фант Г. Акустическая теория речеобразования. – М.: Наука, 1964. – 283 с.
4. Опенгейм А.В., Шафер Р.В. Цифровая обработка сигналлов: Пер. с англ. /под ред.С.Я. Шаца.–М.:Связь,1979.–416с.
5. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов: Пер. с англ. / под ред. М.В. Назарова, Ю.Н. Прохорова. – М.: Радио и связь, 1981. – 495 с.
6. Маркел Дж., Грэй А.Х. Линейное предсказание речи: Пер с англ. / под ред. Ю.Н. Прохорова, В.С. Звездина. – М.: Связь, 1980. – 308 с.
7. Брюхомицкий Ю.А., Казарин М.Н. Метод биометрической идентификации пользователя по клавиатурному почерку на основе разложения Хаара и меры близости Хэмминга // Известия ТРТУ. – Таганрог: Изд-во ТРТУ, 2003. – № 4(33). – С. 141-149.
8. Брюхомицкий Ю.А. Иммунологичекий метод верификации рукописи с использованием векторного представления данных // Известия ЮФУ. Технические науки. – Ростов-на-Дону: Изд-во ЮФУ, 2016. – №9 (182). – С. 50-57.
9. Брюхомицкий Ю.А. Иммунологический подход к идентификации личности по динамическим биометрическим параметрам // Известия ЮФУ. Технические науки. – Ростов-на-Дону: Изд-во ЮФУ, 2017. – №5(190). – С. 56-66.