ОСОБЫЕ ИНТЕРВАЛЫ В ГОЛОСОВОМ ПАРОЛЕ

Е.Л. Столов

Введение

Проблема идентификации человека по биометрическим параметрам привлекает в последнее время все большее внимание исследователей. Это связано, прежде всего, с обеспечением безопасности во всех ее аспектах. Задача идентификации или аутентификации диктора по голосу является составной частью упомянутой проблемы. По используемой технологии различают задачи идентификации на основе анализа произвольной речи, имеющей значительную продолжительность, и идентификацию по парольной фразе, продолжительность которой мала. Аналоговый сигнал оцифровывается и превращается в цифровой объект, после чего дело сводится к сравнению полученного объекта с некоторым образцом. Как правило, прямое сравнение не имеет смысла, поэтому дальнейшая обработка производится по одной и той же схеме. Объект сжимается тем или иным способом для получения некоторого вектора малой длины, компоненты которого устойчивы к возможным искажениям, после чего сравниваются векторы, построенные по полученному объекту и образцу, хранящемуся в базе. Таким образом, все методы различаются способами получения векторов, дающих сжатое описание объектов, и алгоритмами сравнения этих векторов. Обзор методов, разработанных к середине девяностых годов прошлого века, представлен в [1]. Наибольшее развитие получили технологии, основанные на модели, согласно которой сигнал может быть представлен в виде смеси гауссовских распределений. По имеющемуся сигналу осуществляется оценка параметров смеси, которые составляют вектор, используемый для последующей идентификации. Как показано в дальнейших работах, этот метод дает удовлетворительные результаты, если обрабатываемый сигнал имеет достаточную длину. Особенность обработки голосового пароля заключается в относительно малой длине анализируемого файла. Тем не менее, в [2] утверждается, что эта же методика может применяться и для тестового материала малой длины. Однако использование модели на основе смеси гауссовских распределений не учитывает специфику задачи, когда заранее известно содержание произносимого текста. В этой связи были предложены альтернативные подходы, учитывающие упомянутую специфику. В [3] предлагается использовать в качестве основного параметра скорость убывания выборочной автокорре-

ляционной функции сигнала при возрастании аргумента. В [4] применяется подход на основе нечетких множеств. Каждая фонема, представленная в фразе, задается в виде нечеткого множества, после чего вся парольная фраза становится последовательностью нечетких множеств. Для сравнения двух фраз, записанных в указанной форме, применяется аппарат нечетких множеств.

Обилие различных подходов к решаемой задаче означает, что до окончательного закрытия проблемы еще далеко, поэтому представляет интерес исследование возможностей любого подхода, способного с некоторой достоверностью дать ответ на вопрос о принадлежности двух представленных речевых сигналов одному и тому же лицу. В работах [5, 6] введено понятие особой точки в оцифрованном сигнале. В качестве критерия различия двух звуковых файлов было предложено использовать распределение особых точек в файле. В данной работе продолжается исследование возможностей указанной технологии для целей идентификации по парольной фразе. В качестве вектора, описывающего это распределение, предлагается использовать вейвлет-коэффициенты нормированной специальным образом функции распределения. Показана возможность применения этих коэффициентов для целей идентификации диктора.

Отыскание особых точек и особых интервалов в последовательности

Прежде всего, напомним определение особой точки в звуковом файле. Поскольку звуковой файл порождается некоторым физическим процессом, предполагается, что значения сигнала внутри каждого интервала некоторой фиксированной длины K связаны между собой. Определим эту зависимость следующим образом: выберем функцию Fun и C – область на числовой оси и для каждого n проверим выполнение условия

$$Fun(z_n,...,z_{n+K-1}) \in C. \qquad (1)$$

Функция и область выбираются таким образом, чтобы это условие выполнялось для большинства интервалов. Интервалы, для которых условие (1) не выполнено, называются особыми интервалами (ОИ), а их центры – особыми точками. Известно (см., например, [7]), что сигнал, отвечающий гласным звукам, хорошо аппроксимируется процессом авторегрессии. В силу этого существуют линейная функция F и интервал C малой длины, для которых имеет место соотношение (1) для большинства отсчетов z_j.

Функция F должна обладать свойством инвариантности по отношению к операции умножения всех значений отсчета на фиксированное число. Действительно, результат измерения не должен зависеть от коэффициента усиления аппаратуры, использованной при записи и оцифровке сигнала.

Функция F и область C в (1) могут быть выбраны многими способами. Естественно, что результат будет определяться видом выбранной зависимости. В работах [5, 6] функция F строилась следующим образом. По значениям сигнала из интервала длины K вычислялись все однородные мономы второй степени f_k от этих значений. После этого подбирались коэффициенты a_j таким образом, чтобы было выполнено условие

$$\sum(\sum_j a_j f_j)^2 \quad \to \min, \tag{2}$$

где внешняя сумма берется по всем интервалам. В этом случае соотношение (1) имело вид

$$|\sum_j a_j f_j| < \sigma^2 T.$$

Здесь T – некоторое пороговое значение, а σ^2 – среднее от квадратов всех сигналов. В [5, 6] указан способ отыскания коэффициентов в (2). Он сводится к отысканию минимальных собственных значений и соответствующих собственных векторов некоторой матрицы, порядок которой быстро растет при увеличении длины интервала. В данной работе предложен другой способ построения соотношения вида (1), который требует меньших вычислительных затрат.

В основе предлагаемого алгоритма лежит метод, аналогичный методу, используемому для получения линейного предсказания временных рядов [7], однако, в этот метод внесены некоторые изменения. Пусть имеется последовательность $x_1,...,x_N$. В обычной задаче предсказания для заданного p коэффициенты $c_1,...,c_p$ подбираются таким образом, чтобы

$$\sum_n (x_n - \sum_{k=1}^{p} c_k x_{n-k})^2 \quad \to \min .$$

Наша цель заключается в оценке значения в середине интервала по остальным значениям внутри интервала. С этой целью найдем коэффициенты $a_1,...,a_p;b_1,...,b_p$, удовлетворяющие условию

$$\sum_n (x_n - \sum_{k=1}^{p} a_k x_{n-k} - \sum_{k=1}^{p} b_k x_{n+k})^2 \quad \to \min . \tag{3}$$

Оптимальные значения коэффициентов находятся обычным образом путем дифференцирования по параметрам. В результате получим систему линейных уравнений

$$\sum_n x_n x_{n-j} = \sum_{k=1}^{p} a_k \sum_n x_{n-k} x_{n-j} + \sum_{i=1}^{p} b_i \sum_n x_{n+i} x_{n-j},$$

$$\sum_n x_n x_{n+j} = \sum_{k=1}^{p} a_k \sum_n x_{n-k} x_{n+j} + \sum_{i=1}^{p} b_i \sum_n x_{n+i} x_{n+j}, \; j = 1,...,p. \tag{4}$$

В системе (4) индекс n во внутренних суммах пробегает все значения, для которых выражения x_{n-j}, x_{n+j} имеют смысл. Введем следующие обозначения:

$$R(k) = \sum_n x_n x_{n+k}\,,\ \overline{R} = (R(1),...R(p))^T\,,\ \overline{U} = (a_1,...,a_p,b_1,...,b_p)^T\,,$$

$$M_1 = \begin{pmatrix} R(0) & R(1) & ... & R(p-1) \\ R(1) & R(0) & ... & R(p-2) \\ ... & ... & ... & ... \\ R(p-1) & R(p-2) & ... & R(0) \end{pmatrix},$$

$$M_2 = \begin{pmatrix} R(2) & R(3) & ... & R(p+1) \\ R(3) & R(4) & ... & R(p+2) \\ ... & ... & ... & ... \\ R(p+1) & R(p+2) & ... & R(2p) \end{pmatrix}.$$

В этих обозначениях система (4) переписывается в следующей форме

$$M\overline{U} = \begin{pmatrix} \overline{R} \\ \overline{R} \end{pmatrix},\ M = \begin{pmatrix} M_1 & M_2 \\ M_2 & M_1 \end{pmatrix}. \qquad (5)$$

Решая систему (5), находим необходимые значения параметров. Хотя число уравнений в системе совпадает с числом неизвестных, матрица системы может оказаться вырожденной. В этом случае под решением будем понимать псевдорешение, полученное в результате регуляризации системы (5). Если в системе (5) поменять местами блочные строки и столбцы, то она примет прежний вид. Из единственности решения теперь следует, что это решение обладает свойством: $a_k = b_k$, $k = 1,..., p$.

Пусть выбраны некоторый порог T и значение p. Строим интервалы длины $2p+1$ вида $x_{n-p},...,x_{n-1},x_n,x_{n+1},...,x_{n+p}$. Для каждого такого интервала проверяем справедливость неравенства

$$(x_n - \sum_{k=1}^{p} a_k x_{n-k} - \sum_{k=1}^{p} a_k x_{n+k})^2 \quad < T^2\,\sigma^2, \qquad (6)$$

где $\sigma^2 = \frac{1}{N}\sum x_i^2$. Это и есть алгоритм выделения ОИ. Очевидно, что при таком определении свойство интервала быть особым не зависит от коэффициента усиления. Процедуру выделения особых точек можно описать в терминах фильтрации. Строится FIR фильтр с коэффициентами $(a_p, a_{p-1},..., a_1, -1, a_1,..., a_p)$, который применяется к последовательности $x_1,...,x_N$. Те точки, в которых модуль результата фильтрации превышает

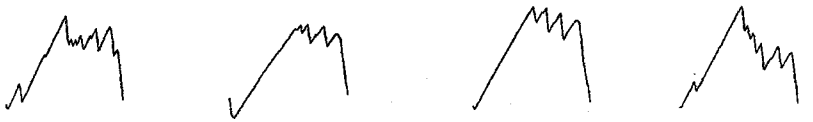
заданный порог, объявляются особыми, а интервалы с центром в особой точке – ОИ.

Описание распределения особых интервалов в голосовом файле

В основе дальнейших вычислений лежит гипотеза, согласно которой для решаемой задачи наиболее информативной является «энергетическая» составляющая сигнала. Это означает, что в качестве переменных x_j в системе (5) выбираются квадраты величин, полученных в результате оцифровки парольной фразы. После того, как найдены ОИ в звуковом файле, возникает задача сжатого описания полученного результата для дальнейшего сравнения. В данной работе предлагается использовать для этой цели вейвлет-преобразование от функции распределения ОИ. Перейдем к точным определениям. Обозначим через M общее число ОИ. Положим $F(k)=\frac{1}{M}\sum_{i=1}^{k}\varepsilon_i$, где $\varepsilon_i=1$, если точка x_i особая, иначе $\varepsilon_i=0$, $1\le k\le N$. Так определенная функция является аналогом функции распределения, она монотонная и $F(N)=1$. Для разных дикторов длина результатов оцифровки парольной фразы будет разной. Чтобы нивелировать данное различие, выбираем число $\bar{N}$ вида $\bar{N}=2^r$, где r некоторое натуральное число. Для сравнения двух функций распределения преобразуем их таким образом, чтобы каждая из них была определена на интервале $[1,\bar{N}]$. С этой целью положим $f(t)=(t-1)\frac{N-1}{\bar{N}-1}+1,\quad 1\le t\le\bar{N}$. Функция $F(k)$ заменяется функцией $\bar{F}(k)=\tilde{F}(f(k))$, заданной на интервале $[1,\bar{N}]$. Значение $\tilde{F}(y)$ находится линейной интерполяцией по значениям $F(]y[)$ и $F(]y[+1)$, где $]y[$ – целая часть y. В этих обозначениях манера произнесения парольной фразы описывается функцией $\bar{F}$ своей для каждого диктора. Для формализации этого описания нужно представить функцию $\bar{F}$ в некотором стандартном базисе. Для этой цели было применено дискретное вейвлет-преобразование (DWT). Причиной такого выбора является то обстоятельство, что DWT предназначено для описания кривых нерегулярной формы. Предварительно убираем «тренд» из найденной функции $\bar{F}$, заменив ее на функцию $U(k)=\bar{F}(k)-\frac{k}{\bar{N}}$. Для сжатого описания функции $U(k)$ применяется кратномасштабный вейвлет-анализ [8].

Результаты экспериментов

Для проверки возможностей предлагаемой технологии был произведен эксперимент. Следует заметить, что не ставилась цель оценить достоверность идентификации с помощью разработанного метода. Для оценки достоверности нужна база парольных фраз значительного размера, создание которой является самостоятельной задачей. Вместо этого показана принципиальная применимость изложенного метода для целей идентификации. В качестве исходного материала была выбрана фраза: «Who authorized the unlimited expense account?» из базы TIMIT. Рассматривались только примеры произнесения фразы мужчинами, поскольку отличить мужской голос от женского можно по частоте основного тона сигнала. Эти файлы получены оцифровкой с частотой 16kHz и содержат примерно 35-40 тысяч отсчетов.



s1) s2) s3) s4)

Рис. 1

На рис. 1 представлены графики функции $U(k)$, построенные с использованием следующих значений параметров: $p=7$, $r=8$, $T=0.1$. Как и следовало ожидать, вид кривой существенно зависит от выбора значения T. В то же время, как показали эксперименты, выбор значения p мало влияет на форму кривой. Визуальное различие графиков на рис 1 очевидно. Для автоматического анализа необходимо построение вектора, описывающего кривую в сжатой форме. С этой целью применялась процедура кратномасштабного вейвлет-анализа на основе вейвлета Добеши db8 [8]. Этот фильтр имеет длину 16, что хорошо согласуется с выбранной длиной 256 анализируемой последовательности. Выбор данного вейвлета не является принципиальным моментом. Основное условие – удачное сжатое представление анализируемой кривой – является единственным критерием, и это условие выполнено при указанном выборе. Вычисление производилось с использованием пакета MatLab по следующей схеме. Сначала для каждой кривой U подсчитывались низкочастотная составляющая по формуле $A_1 = dwt(U,'db8')$. К этой составляющей снова применялся тот же фильтр: $A_{j+1} = dwt(A_j,'db8'), \quad j=1,2,3$. В результате получаются векторы A_4 длины 30. Эти векторы и являются сжатым представлением ис-

ходных звуковых файлов. Для сравнения двух векторов v_1, v_2 использована формула $d = \frac{|v_1 - v_2|^2}{|v_1||v_2|}$. Она дает расстояние, отнесенное к длинам векторов. Это позволяет сравнивать значения, полученные по разным технологиям, когда исходными являются векторы различной длины. Результаты измерений представлены в таблице 1.

Таблица 1

N	v1	v2	v3	v4
v1	X	0.2359	0.1576	0.1465
v2	0.2359	X	0.1468	0.2536
v3	0.1576	0.1468	X	0.0679
v4	0.1465	0.2536	0.0679	X

Здесь вектор vk отвечает функции sk, представленной на рис. 1. Полученные результаты интересно сравнить с расстояниями между оригинальными нормализованными функциями распределения. Они приведены в таблице 2.

Таблица 2

N	s1	s2	s3	s4
s1	X	0.1711	0.1645	0.1472
s2	0.1711	X	0.0925	0.1743
s3	0.1645	0.0925	X	0.0710
s4	0.1472	0.1743	0.0710	X

Из сравнения указанных таблиц следует, что расстояния между векторами, представляющими сжатое описание кривых, дают значения близкие к оригиналам. Это замечание подтверждает возможность их использования для идентификации дикторов

Литература

1. Campbell J.P. Speaker recognition: A tutorials // Proceedings Of The IEEE. – 1997. – V. 85, N. 9. – P. 1437-1462.

2. Angkititrakul P., Hansen J.H.L. Discriminative In-Set/Out-of-Set Speaker Recognition // IEEE Transactions On Audio, Speech, And Language Processing. – 2007. – V. 15. – N. 2. – P. 498-508.

3. Ana R.S., Coelho R., Alcaim A. Text-Independent Speaker Recognition Based on the Hurst Parameter and the Multidimensional Fractional Brownian Motion Model // IEEE Transactions On Audio, Speech, And Language Processing. – 2006. – V. 14. – N. 3. – P. 931-940.

4. Halavati R., Shouraki S.B., Zadeh S.H. Recognition of human speech phonemes using a novel fuzzy approach // Applied Soft Computing. – 2007. – V. 7. – N. 3. – P.828-839.

5. Столов Е.Л. Идентификация диктора на основе отыскания особых точек в произнесенной фразе //Вестник Томского гос. университета. Приложение. – 2006. – N. 17. – С. 37-40.

6. Столов Е.Л. Алгоритм обработки голосового пароля // Исследования по информатике. Вып. 11. – Казань: Отечество, 2007. – С. 103-108.

7. Андерсон Т. Статистический анализ временных рядов. – М.: Мир, 1976.

8. Дьяконов В. Вейвлеты. От теории к практике. – М.: СОЛОН-Пресс, 2004.