Математические структуры и моделирование
2008, вып. 18, с. 108–115

УДК 621.391

ИДЕНТИФИКАЦИЯ ДИКТОРА ПО ФИКСИРОВАННОМУ НАБОРУ ЧАСТОТ С ПОМОЩЬЮ ЛИНЕЙНОГО КЛАССИФИКАТОРА

Е.В. Венедиктова, Д.Н. Лавров

В данной работе рассматривается реализация алгоритма идентификации диктора на основе линейного классификатора. Данными для обучения классификатора является набор значений спектра голоса, рассмотренный на фиксированных частотах. Проведены численные эксперименты, подтверждающие приемлемую точность идентификации.

Введение

В данной работе рассматривается построение сиcетмы идентификации пользователя информационной системы (диктора) на основе анализа спектральных характеристик его голоса с помощью линейного классификатора. При построении такой системы голос представляется как временной ряд, который подвергается преобразованию Фурье. Значения модуля спектра на требуемых промежуточных частотах между имеющимися отчетами интерполируются с помощью кубических сплайнов. Затем набор значений на заранее определенном наборе частот итерпретируется как точка в n-мерном пространстве. Принадлежность точки к определенному множеству спектральных характеристик голоса определенного человека устанавливается заранее построенным линейным классификатором. Классификатор строится на основе записи голоса диктора и вычисления статистических характеристик. В работе показано, что частоты разных выборок распределены нормально, а также экспериментально определены характеристики качества работы классификатора. Намечены направления дальнейших исследований.

Задача идентификации диктора находит свое применение в банковских системах и в сотовой телефонии для авторизации пользователей. Существует ряд классических подходов к идентификации голоса, основанных на построении и анализе линейных фильтров или скрытых марковских моделей [1]. Качество идентификации в некоторых случаях достигает 99%.

Омский государственный университет.
E-mail: lavrov@omsu.ru

Задача идентификации голоса может быть решена с помощью методов статистической теории распознавания образов [2]. Байесовский классификатор является оптимальным, но реализовать его сложно, особенно когда размерность пространства высока. Линейный классификатор сочетает в себе качество распознавания и простоту реализации. Альтернативным является подход экспертных систем [6] и нейронных сетей [5]. Реализация однослойной нейронной сети дает результат, аналогичный линейному классификатору, а подход обучения экспертных систем, описанный в [6], эквивалентен построению гиперплоскости между разделяемыми множествами. Таким образом, представляет интерес описанный в [2] метод линейной классификации.

Цель данной работы – определить возможность использования линейного классификатора для идентификации голоса и оценить качество его работы. Для этого необходимо реализовать алгоритм построения линейного классификатора, алгоритм принятия решения при наличии n-разделямых субъектов. Провести анализ спектра частот голоса и на его основе выбрать необходимую информацию для обучения классификатора.

1. Линейный классификатор

Для построения линейного классификатора был выбран метод, представленный в [2]. Суть метода: построить линейную разделяющую функцию, минимизирующую вероятность ошибки решения. Решающее правило выглядит следующим образом:

$$h(X) = V^T X + v_0, \tag{1}$$

$$\begin{cases} h(X) > 0, & X \in \omega_1, \\ h(X) < 0, & X \in \omega_2, \end{cases} \tag{2}$$

где $h(X)$ – линейная разделяющая функция; V — вектор коэффициентов; v_0 — порог.

Выражение $h(X)$ – это линейная функция относительно вектора X, называемая линейной разделяющей функцией. Задача построения линейной разделяющей функции заключается в том, чтобы для заданных распределений определить коэффициенты $V = \{v_1, v_2, \ldots, v_n\}$ и значение порога v_0.

Если величина $h(X)$ распределена по нормальному или близкому к нему закону, то для вычисления вероятности ошибки можно использовать математическое ожидание и дисперсию $h(X)$ для классов ω_1 и ω_2 , а затем выбрать параметры V и v_0 так, чтобы минимизировать ошибку решения. Вспомнив, что $h(X)$ является суммой n слагаемых x_i, получаем, что если векторы X имели нормальное распределение, то величина $h(X)$ также имеет нормальное распределение. Математическое ожидание и дисперсия в классах равны

$$\eta_i = E\{h(X)|\ \omega_i\} = V^T E\{X|\ \omega_i\} + v_0 = V^T E\{M_i\} + v_0 \tag{3}$$

$$\sigma^2 = Var\{h(X)|\ \omega_i\} = V^T E\{(X - M_i)(X - M_i)^T\}V = V^T \Sigma_i V, \tag{4}$$

где M_i – математическое ожидание выборки из ω_1 и ω_2 соответственно; Σ_i – ковариационные матрицы соответствующих выборок. Вероятность ошибки можно

записать следующим образом:

$$\varepsilon = P(\omega_1) \int_{-\eta_1/\sigma_1}^{\infty} \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{\xi^2}{2}\right) d\xi + P(\omega_2) \int_{\infty}^{-\eta_2/\sigma_2} \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{\xi^2}{2}\right) d\xi. \tag{5}$$

$$V = [s\Sigma_1 + (1-s)\Sigma_2]^{-1} (M_2 - M_1), \tag{6}$$

где s лежит в диапазоне от 0 до 1. v_0 можно вычислить по формуле:

$$v_0 = \frac{s\sigma_1^2 V^T M_2 + (1-s)\sigma_2^2 V^T M_1}{s\sigma_1^2 + (1-s)\sigma_2^2}. \tag{7}$$

Функция (5) минимизируется методом золотого сечения по параметру s. Вычисление самой функции (5) на шаге итерации выглядит следующим образом:

1. Для данного s вычислить V по формуле (6).
2. Используя полученное значение V, вычислить по уравнениям (4) и (7) величины σ_1^2, σ_2^2, η_1, η_2 и v_0.
3. Вычислить вероятность ошибки по формуле (5).

Повторяем итерацию метода золотого сечения до достижения приемлемой точности.

2. Исходные данные для идентификации

В исследовании участвовало четыре человека, назовем их условно Елена, Дарья, Дмитрий и Алексей. Сделано по 50 записей фразы «Добрый день». Такое количество данных обусловлено тем, что больше 50 записей для человека утомительно, а также важно установить, каких данных и в каком количестве достаточно для более достоверной работы программы. Была выбрана одна фраза, так как от фразы и присутствующих в ней звуках может зависеть интенсивность отдельных частот спектра, что необходимо исследовать отдельно.

Записи голоса записаны с частотой дискретизации 44 кГц и хранятся в wav формате. При преобразовании Фурье для звука с частотой дискретизации 44 кГц получаем набор частот в диапазоне от 0 Гц до 22 кГц, его количество может быть различным, что зависит от длины файла, но равномерно распределенным по этому интервалу частот.

Для решения поставленной задачи необходимо иметь однотипные данные, одинаковый набор частот со значениями их интенсивности. Для этого необходимо интерполировать спектр частот. А затем вычислить для всех звуков интенсивность для одинакового набора частот.

Используя интерполяцию кубическими сплайнами, далее взяв набор частот в интервале от 0 до 600 Гц с шагом дискретизации, равным 1 Гц, получим характеристику одинаковой размерности (размерность равна 600). Это дает возможность использования ее при идентификации.

Далее возникает ограничение, накладываемое из-за размеров выборки (50 звуков), значит, мы можем использовать не более 50 частот.

Для получения вектора интенсивности частот выбираем 50 частот из интервала от 51 до 500 Гц. В этом интервале видны характерные особенности каждого голоса.

3. Подтверждение гипотезы о распределении интенсивности

Линейный классификатор строится при условии, что выборка значений имеет нормальное распределение или близкое к нему. Проверка интенсивности для отдельно взятой частоты по критерию "хи-квадрат" показала, что интенсивность имеет нормальное распределение. Выборка из $n = 50$ случайных значений (среднее = 0.130510532, стандартное отклонение = 0.048002366399557) была упорядочена по возрастанию и разбита на $s = 8$ интервалов. Для каждого интервала было рассчитано теоретическое количество точек (которые должны попасть в данный интервал) с помощью интеграла функции Гаусса и наблюдаемое число значений. Итак, для проверки нулевой гипотезы H0 (генеральная совокупность распределена нормально) нужно вычислить по выборке наблюдаемое значение критерия (по формуле Пирсона):

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_0} \approx 7.1936,$$

где f_0 – теоретическое количество точек в интервале, f_e – количество точек выборки в интервале. Число степеней свободы $k = s–1–2 = 5$ (уклонения связаны линейным соотношением, кроме того, наложены еще две связи, так как по выборке были определены два параметра распределения: среднее и стандартное отклонение) меньше табличного значения χ^2-квадрат распределения, равного 11.1 ($\gamma = 0.95$, $k = 5$). Соответственно, вероятность ошибки первого рода 0.05, выборка имеет нормальное распределение с вероятностью 0.95.

4. Реализация алгоритма

Для исследования написаны следующие программы на scilab: программа преобразования входных данных, программа обучения классификатора, программа тестирования классификатора и демонстрационная программа.

Для реализации данной задачи был выбран математический пакет scilab 4.1., аналог matlab. Преимущества выбора: scilab является бесплатным программным обеспечением и дает возможность использовать необходимые в данном алгоритме математические функции, такие, как операции над векторами и матрицами (умножение, вычисление обратной); преобразование Фурье; построение кубических сплайнов; позволяет строить графики для наглядного изображения данных, что дает возможность проводить необходимые исследования.

Начальными данными являются файлы формата .wav. По 50 файлов от четырех человек. Для обработки данных написана программа, загружающая записи каждого человека, вычисляющая спектр частот и сохраняющая данные в файле.

Классификатор реализован отдельно. Данными для него являются спектры частот, которые предварительно должны быть получены. Результат обучения линейного классификатора сохраняется в файле, в котором хранятся параметры линейной разделяющей функции для данных спектров.

При реализации алгоритма пришлось решать проблему, вызванную разной дискретизацией спектров, возникающей в силу того, что запись парольной фразы осуществляется в разных по длительности временных диапазонах. Но так как частота дискретизации для всех экспериментов одинакова, то диапазон частот одинаков для всех записей голосов (в силу теоремы Котельникова), но с разным количеством частотных отсчетов. Идентификацию нужно проводить по выбранным частотам, число котрых не должно превышать число выборок для обучения, в противном случае оценки ковариационных матриц окажутся необратимыми, что приведет к неверной работе классификатора. Выбранные частоты могут не попадать на дискретные отсчеты спектра. Недостающие значения спектра в промежуточных значениях дискретных частот интерполируем кубическими сплайнами.

Пример построения линейной разделяющей функции:

```
function [V,v0,n1,n2,err]=All_who(Mx,My,Dx,Dy,n)
    // Mx,My - математические ожидания двух выборок
    // Dx,Dy - матрицы ковариаций соотвутствующих выборок
    // V - вектор коэффициентов
    // v0 - порог

    [err] = MinErr(Mx,My,Dx,Dy,n);
    V = inv(err*Dx+(1-err)*Dy)*(My-Mx);
    q1_2 = (V'*Dx)*(V);
    q2_2 = (V'*Dy)*(V);
    v0=-(err*q1_2*(V')*My+(1-err)*q2_2*(V')*Mx)/(err*q1_2+(1-err)*q2_2);
    n1 = V'*Mx+v0;
    n2 = V'*My+v0;
endfunction
```

Пример вычисления вероятности ошибки в точке s:

```
  V = inv(s*D1+(1-s)*D2)*(M2-M1);
  q1_2 = (V')*D1*V;
  q2_2 = (V')*D2*V;
  q1_1 = sqrt(q1_2);
  q2_1 = sqrt(q2_2);
  v0=-(s*q1_2*(V')*M2+(1-s)*q2_2*(V')*M1)/(s*q1_2+(1-s)*q2_2);
  n1 = (V')*M1+v0;
  n2 = (V')*M2+v0;
  [p1,q1] = cdfnor("PQ",-(n1/q1_1),0,1);
  [p2,q2] = cdfnor("PQ",-(n2/q2_1),0,1);
  e = 0.5*(q1+p2);
  res= 0.5*(q1+p2);
```

Демонстрационная программа представляет собой диалог с пользователем. Пользователь вводит путь к файлу для идентификации. Используя ранее построенную линейную разделяющую функцию (ее данные загружаются из файла), программа определяет, кому принадлежит голос, и выводит результат.

Для проведения экспериментов написана отдельная программа. В ней файлы для проверки загружаются из указанной директории и в текстовый файл выводятся результаты с подсчетом количества ошибок.

5. Определение вероятности ошибки классификатора

При неизвестных априорных вероятностях можно взять N объектов и проверить правильность работы классификатора [2]. На графике зависимости N, ε, ε' и γ показано соотношение между истинной ошибкой ε и ее оценкой ε' (рис. 1). Для коэффициента доверия $\gamma = 0.95$, при $N = 100$ экспериментах неправильно классифицированных $\tau = 11$ объектов, получаем оценку $\varepsilon' = \tau/N = 0.11$, и по графику можно определить доверительный интервал истинной вероятности ошибки ϵ при $\gamma = 0.95$, он равен $(0.05, 0.18)$.

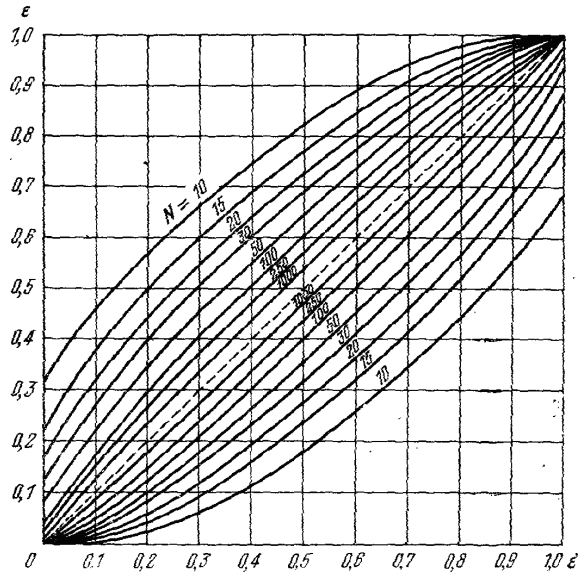

Рис. 1. Доверительный интервал $\gamma = 0.95$ истинной вероятности ошибки [2]

6. Экспериментальное определение качества идентификации

Для тестирования голоса четыре человека (Елена, Дарья, Дмитрий и Алексей) записались по 25 раз, фраза «Добрый день», эксперимент состоял из идентификации 100 голосов. Голоса были записаны в 2 подхода для каждого в разные дни. Были получены следующие результаты (см. табл. 1).

Таблица 1. Результаты тестирования на фразе «Добрый день»

Имя	**Возраст**	**Тестов (шт.)**	**Ошибок (шт.)**	**Правильно (%)**
Елена	21	25	2	92%
Дарья	13	25	0	100%
Дмитрий	36	25	8	68%
Алексей	20	25	1	96%
Итого:		100	11	89%

При тестировании 25 голосов Елены произошло 2 ошибки, ее программа приняла за Дарью. Однако этот результат можно считать очень хорошим, учитывая, что они сестры и по телефону их голоса путают многие, даже родители, ошибаясь чаще, чем линейный классификатор. Дарью классификатор идентифицировал без ошибок. Дмитрия идентифицировал 17 раз из 25, что, конечно, является плохим результатом, но нужно учесть, что идентифицируемым голос был изменен сознательно. Алексей был идентифицирован один раз с ошибкой. Из этого следует вывод, что, если человек заинтересован в своей идентификации, его узнают с 96 %-ной ((75-3)/75*100%) точностью. Общим результатом можно считать узнаваемость с 89 %-ной точностью. Если же протестировать голос, когда человек говорит другую фразу или слово, получим следующие результаты (см. табл. 2).

Таблица 2. Результаты тестирования на разных фразах

Имя	**Тестов (шт.)**	**Ошибок (шт.)**	**Правильно (%)**
Елена	25	7	72%
Дарья	25	3	88%
Дмитрий	5	1	80%
Алексей	25	7	72%
Итого:	80	18	77.5%

По другим фразам голоса различаются с меньшей точностью. Для использования линейного классификатора в системах, где человек заинтересован в том, чтобы его узнавали, возможно, лучше будет использовать не одну фразу во всех записях, а разные фразы по несколько раз. Таким образом будет накоплена наиболее ярко отражающая особенности голоса человека информация.

Анализ спектра частот показал, что при построении классификатора наиболее информативными являются частоты в диапазоне от 50 до 600 Гц. Однако возможность использования всей информации ограничивается количеством записей для обучения классификатора. В данной работе количество записанных звуков для построения классификатора было выбрано равным 50, и для построения классификатора использовалась каждая девятая частота из интервала от 50 до 500 Гц.

Возможно, для увеличения надежности идентификации необходимо исполь-

зовать выборку с большим количеством записей. Произношение разных фраз человеком с разным настроением для обучения классификатора вполне может дать возможность построения более надежного классификатора.

Для идентификации личности, если человек сознательно изменял голос, линейный классификатор, разделяющий по спектру частот, с небольшой выборкой для обучения плохо справляется с задачей идентификации. Возможно, нужно использовать не конкретный набор частот, а выбирать их в зависимости от спектра, что даст возможность определить особенности голоса, даже если человек старается его изменить. Вопрос о том, как выбирать частоты, остается открытым, это тема для отдельной работы.

7. Заключение

В ходе выполненной работы, на основе выбранного метода минимизации вероятности ошибки решения, реализован алгоритм обучения линейного классификатора. Биометрической характеристикой выбран спектр частот голоса, так как частота колебаний зависит от строения речевого аппарата. При реализации алгоритма возник ряд проблем, связанных с обработкой голоса: вычисление частотного спектра с помощью преобразования Фурье; интерполирование спектра частот для получения данных одной размерности. Классификатор протестирован на специально сгенерированных данных. Тестирование на математической модели сигналов показало хорошие результаты.

Проведено 100 экспериментов по идентификации голоса. Результаты экспериментов говорят о том, что использование линейного классификатора для идентификации личности по голосу вполне эффективно, когда биометрической характеристикой голоса выбран спектр частот (89 % испытаний).

Но выявлены и ограничения данного подхода. Так, для обратимости корреляционных матриц необходимо иметь их оценку по выборке большей или равной, чем размерность пространства, то есть числа частот, по которым проводится идентификация. Это требует более длительной процедуры обучения.

Литература

1. Campbell J. P. Speaker recognition: a tutorial. Proc. IEEE. Vol. 85, N. 9. 1997. P. 1437–1462.
2. Фукунага К. Введение в статистическую теорию распознавания образов.: Пер. с англ. М.: Наука. Главная редакция физико-математической литературы, 1979. 368 с.
3. Стретт Дж. Теория звука.: Пер. с англ. 2 изд. М.:ГИТТЛ. 1955. Т. 1—2. 473 с.
4. Цифровой звук: теория. // Журнал "Upgrade". 2005. Т. 1 (20).
5. Яхъяева Г.Э. Нечеткие множества и нейронные сети. М.: Интернет-Университет Информационных Технологий; БИНОМ. Лаборатория знаний. 2006. 316 с.
6. Нейлор К. Как построить свою экспертную систему.: Пер. с англ. М.: Энергоатомиздат, 1991. 286 с.