ЭМПИРИЧЕСКОЕ ПРЕДСКАЗАНИЕ И РАСПОЗНАВАНИЕ ОБРАЗОВ
(Вычислительные системы)
1980 год Выпуск 83

УДК 519.226: 534.44

АЛГОРИТМ КЛАССИФИКАЦИИ ТОН/ШУМ ПО ЧАСТНЫМ АВТОКОРРЕЛЯЦИЯМ

А.В.Кельманов

Пусть речевой сигнал $s(t)$ проквантован с частотой $F_д = 1/T$ так, что $s_n = s(nT)$, где T – интервал квантования. Разобьем сигнал на L сегментов $\{s_n\}^l$, $n=\overline{1,N}$, $l=\overline{1,L}$ длительностью $T_д = (N-1)/F_д$ со сдвигом от сегмента к сегменту на интервал ΔT. Задача – определить, к какому типу звуков (голосовых или неголосовых) относится каждый из L сегментов.

Для решения этой задачи уже построены [1] два параметрических алгоритма, использующие вероятностные распределения. В настоящей работе предлагается еще один алгоритм подобного типа и проводится его сравнение с вышеуказанными.

1. Алгоритм решения

В основе предлагаемого алгоритма (назовем его алгоритм I), как и в алгоритмах (II и III) из [1], лежит статистический критерий диагностической проверки на адекватность описания речевого сигнала моделью авторегрессии. Воспользоваться этим критерием для автоматической классификации тон/шум позволяет тот факт, что голосовые звуки в отличие от неголосовых описываются моделью авторегрессии неадекватно [1].

Предположим, что каждый из сегментов речевого сигнала аппроксимировался моделью авторегрессии K-го порядка:

$$s_n = \sum_{i=1}^{K} a_i s_{n-i} + \alpha_n, \qquad n = \overline{1,N}, \tag{1}$$

где $\{\alpha_n\}$ – остаточный ряд ошибок, распределенных нормально с нулевым средним и дисперсией σ_α^2; и найдены оценки Юла-Уокера [2] параметров авторегрессии и частных автокорреляций:

$$\hat{a}_{j+1,i} = \hat{a}_{ji} - \hat{a}_{j+1,j+1} \times \hat{a}_{j,j-i+1}, \quad i = 1,2,\ldots,j;$$

$$\hat{a}_{j+1,j+1} = \frac{\hat{r}_{j+1} - \sum_{i=1}^{j} \hat{a}_{ji} \hat{r}_{j-i+1}}{1 - \sum_{i=1}^{j} \hat{a}_{ji} \hat{r}_{i}}, \quad j = 0,1,\ldots,K-1.$$

Если порядок процесса авторегрессии равен p, то для всех $j > p$ выборочные частные автокорреляции $\hat{a}_{jj}$: 1) распределены приближенно нормально с нулевым средним даже при небольших объемах выборки N (см. [2]); 2) имеют дисперсию $\sigma^2[\hat{a}_{jj}] = N^{-1}$ (см. [2]) и 3) статистически независимы [2]. Отсюда нетрудно показать, что статистика

$$\theta = N \sum_{i=p+1}^{K} \hat{a}_{ii}^2 \qquad (2)$$

распределена как $\chi^2(K - p)$. Равенство (2) можно использовать для проверки модели авторегрессии на адекватность: если модель соответствует данным, то θ распределена как $\chi^2(K - p)$, если нет, то среднее значение θ увеличивается. Иными словами, необходимо проверить нулевую гипотезу $\forall i \ (i > p \Rightarrow a_{ii} = 0)$ против альтернативной $\exists i \ (i > p \Rightarrow a_{ii} \neq 0)$. Для заданного уровня значимости γ можно найти порог χ^2_γ такой, что $P(\theta > \chi^2_\gamma | H_0) = \gamma$. Поэтому если $\theta > \chi^2_\gamma$, то построенная модель неадекватна, а соответствующий сегмент голосовой.

Напомним, что в алгоритме II решение принимается также по критерию χ^2, но статистика построена по автокорреляциям ряда остаточных ошибок $\{\alpha_n\}$, а в алгоритме III решение принимается по одному значению выборочной частной автокорреляции при помощи двустороннего критерия для нормально распределенных наблюдений.

Для сокращения объема вычислений, как и в алгоритмах II и III, на первом шаге в алгоритме I предусмотрены низкочастотная фильтрация, осуществляемая при помощи чебышевского фильтра 3-го порядка с частотой среза $F_c = 0{,}8$ кгц, понижение исходной частоты квантования сигнала F_s до 2 кгц, а также взвешивание профильтрованного сигнала окном Хэмминга перед авторегрессионной аппроксимацией (1). Блок принятия решения и коррекции ошибок классификации идентичен блоку, включенному в алгоритмы II и III.

Выбор "оптимального" [5] порядка модели p производится из соображений минимума ошибки аппроксимации и минимума объема вычислений с учетом формантной структуры речевого сигнала.

Длина ряда частных автокорреляций K выбирается с учетом того, что при уменьшении K сокращается объем вычислений, однако при этом уменьшается разность K - p, равная числу частных автокорреляций, по которым принимается решение, и поэтому увеличивается вероятность ошибочной классификации. Наоборот, при увеличении K вероятность ошибок уменьшается, но возрастает объем вычислений.

2. Трудоемкость алгоритма

Трудоемкость алгоритма I (без учета низкочастотной фильтрации) состоит из трудоемкости оценивания K частных автокорреляций, равной $N(K+1)+K^2$ операций (сложения и умножения), трудоемкости вычисления статистики θ, равной K - p операций, и трудоемкости принятия решения и коррекции ошибок, равной 5 операциям. В табл.I приведены трудоемкости алгоритмов I-III.

Таблица I

Алгоритм	Трудоемкость
I	$(K+1)+K^2-p+5$
II	$(K+2p+3)+p^2+K+5$
III	$(p+1)+p^2+5$

После низкочастотной фильтрации, которая требует примерно 5N операций, объем выборки N для каждого сегмента сокращается до величины:

$$N_1 = \frac{2(N-1)}{F_s} + 1 \approx \frac{2N}{F_s}.$$

При этом, вследствие уменьшения частотного диапазона, уменьшается порядок модели p, а стало быть, уменьшается и K (предполагается, что разность K - p остается такой же, как и при исходной частоте квантования). Для подсчета трудоемкости алгоритмов с учетом фильтрации в каждой строке табл.I следует заменить N на $2N/F_s$ и прибавить 5N.

3. Эксперименты и их результаты

Речевой материал. При проведении эксперимента использовались 2 словаря. Первый состоял из 100 русских слов [3], второй - из 63 слов [4]. По первому словарю одним диктором-мужчиной (диктор I) были сформированы две акустические последова-

тельности, а по второму - пятью дикторами-мужчинами (дикторы I,2, 3,4,5) было сформировано по одной последовательности. Общее число сегментов, на которых принималось решение тон/шум, составило величину, равную примерно 24,5 тысячи.

У с л о в и я э к с п е р и м е н т а. Эксперимент проводился на ЭВМ "Минск-32". Ввод слов осуществлялся в машинном зале (уровень шума ~ 60 дб) через микрофон типа МД-59 и семиразрядный аналого-цифровой преобразователь с частотой квантования $F_д$ = = 20 кгц.

У с л о в и я а н а л и з а. Акустические реализации были обработаны при помощи алгоритма I, описанного в данной статье, а также при помощи алгоритмов II и III. Разбиение слов на сегменты длительностью T_a = 25,6 мсек производилось со сдвигом (с перекрытием) от сегмента к сегменту на время ΔT = I6 мсек так, что N = = 5I2, а N_1 = 52. Значения параметров K и p равнялись 30 и 4 соответственно. Поэтому статистика (2) распределена как $\chi^2(26)$, и при уровне значимости γ = 0,00I значение порога χ^2_γ = 54,I.

Р е з у л ь т а т ы к л а с с и ф и к а ц и и. Все акустические реализации слов были выведены на печать в виде картинок видимой речи с целью проведения визуальной классификации. После этого те же реализации были обработаны алгоритмически. Для примера на рис.I приведен график изменения статистики θ во времени и результаты классификации слова "мифический" по алгоритму I. В табл.2 приведены данные по классификации сегментов для отдельных видов звуков при помощи трех алгоритмов, а в табл.3 - результаты классификации тоновых и шумовых сегментов.

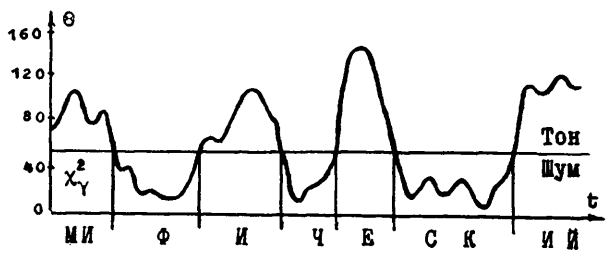

Рис.I

Таблица 2

Ошибки классификации сегментов для отдельных видов звуков

Фонемы	Число реализаций	Алгоритм I		Алгоритм II		Алгоритм III	
		число ошибок	% ошибок	число ошибок	% ошибок	число ошибок	% ошибок
а,о,у, э,ы,и	I0069	3I	0,60	I85	I,84	640	6,36
м,н,л, р,в	4695	64	2,68	I48	3,I5	I9I	4,07
б,д,г	2I45	I64	7,64	I6I	7,50	I33	6,20
з,ж	I5II	58	3,84	84	5,56	25	I,65
п,т,к	2465	5	0,20	25	I,0I	I44	5,84
ц,ч	7I7	0	0	6	0,84	37	5,I6
ф,с, ш,х	2882	2	0,07	6	0,2I	235	8,I5
Всего:	24485	324	I,32	6I5	2,5I	I405	5,74

Таблица 3

Ошибки классификации тоновых и шумовых сегментов

Тип сегмента	Число реализаций	Алгоритм I		Алгоритм II		Алгоритм III	
		число ошибок	% ошибок	число ошибок	% ошибок	число ошибок	% ошибок
Тон	I842I	3I7	I,72	578	3,I4	989	5,37
Шум	6064	7	0,I2	37	0,6I	4I6	6,86
Всего:	24485	324	I,32	6I5	2,5I	I405	5,74

4. Обсуждение результатов

Сравнение алгоритмов по трудоемкости. Обращаясь к табл.I, прежде всего отметим,что наименее трудоемкий - алгоритм III.Далее, учитывая, что K должно быть больше p , замечаем,что, начиная с K = p+1 до $K=\sqrt{2N(p+1)+p^2+p}$, трудоемкость алгоритма I меньше трудоемкости алгоритма II,при дальнейшем увеличении K алгоритм I становится более трудоемким. Нетрудно убедиться в том, что при N = 52, K = 30 и p = 4 алго -

рити I требует больше времени на принятие решения,чем алгоритм II. Для примера приведем трудоемкости алгоритмов с учетом низкочастотной фильтрации при исходной частоте квантования F_s = 20 кгц и при K = 30 и p = 4. При этих значениях параметров трудоёмкости алгоритмов I,II и III приближенно равны 10N , 9N и 5,5N соответственно.

С р а в н е н и е а л г о р и т м о в п о н а д е ж н о с т и. Сравнивая алгоритмы по табл.2, нетрудно заметить, что для неголосовых групп звуков: глухих взрывных (п,т,к), аффрикат (ц,ч), глухих щелевых (ф,с,х,ш) - наименьшую вероятность ошибочной классификации дает алгоритм I, затем - алгоритм II и, наконец, - алгоритм III. Аналогичная картина наблюдается для голосовых групп звуков: гласных (а,о,у,э,ы,и) и сонорных (м,н,л,р,в). На звонких взрывных (б,д,г) наилучшие результаты дает алгоритм III, а наихудшие - алгоритм I, однако отличия в вероятностях ошибки незначительны. Звонкие щелевые (з,ж) наилучшим образом классифицируются алгоритмом III, а наихудшим - алгоритмом II. Из табл.3 следует, что как по отдельности, так и в целом на голосовых и неголосовых звуках наилучшие результаты показывает алгоритм I. Для алгоритма II вероятности ошибочной классификации (обоих родов) возрастают. Наконец, алгоритм III дает наибольшие вероятности ошибочной классификации.

Таким образом, наиболее надежным является наиболее трудоемкий алгоритм I, а наименее надежным - наименее трудоемкий алгоритм III.

5. Выводы и рекомендации

Результаты экспериментального тестирования позволяют сделать следующие выводы и рекомендации.

1. Предложенный алгоритм автоматической классификации звуков речи на звонкие и глухие опробован на пяти дикторах и показал высокую надежность. Ошибки классификации тоновых и шумовых сегментов равны 1,72% и 0,12% соответственно.

2. Алгоритмы аналогичного типа, с которыми проводилось сравнение, менее надежны, однако позволяют производить классификацию за меньшее время.

3. Окончательный выбор того или иного алгоритма классификации тон/шум рекомендуется проводить, исходя из конкретных условий работы речевой системы, в которой задействован классификатор по-

добного типа, и учитывая противоречивые требования уменьшения вероятности ошибочной классификации и повышения трудоемкости.

В заключение отметим, что в настоящее время описанный алгоритм успешно используется в системе распознавания изолированных слов.

Литература

1. КЕЛЬМАНОВ А.В. Алгоритм классификации тон/шум, основанный на критерии адекватности модели авторегрессии. – В кн.: Методы обработки информации. (Вычислительные системы, вып.74.) Новосибирск, 1978, с.129–148.

2. БОКС Дж., ДЖЕНКИНС Г. Анализ временных рядов, прогноз и управление. – М.: Мир, т.I, 1974. – 206 с.

3. КЕЛЬМАНОВ А.В. Система распознавания изолированных слов по частной автокорреляционной функции. – В кн.: Эмпирическое предсказание и распознавание образов. (Вычислительные системы, вып.76.) Новосибирск, 1978, с.132–143.

4. ДЕБЕДЕВ В.Г. Автоматическое распознавание речи по маскпризнакам. – В кн.: Эмпирическое предсказание и распознавание образов. (Вычислительные системы, вып.67.) Новосибирск, 1976, с.136–140.

5. MARKEL J.D. Applikation of digital invers filter for automatik formant and F_0 analysis.– IEEE Trans. Audio Electroacoust, 1973, v.AU-21, N 3, p.154–160.

Поступила в ред.-изд.отд.
20 октября 1979 года