ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ

Сборник трудов
Института математики СО АН СССР

1967 г. Выпуск 28

ПОДСТРОЙКА ПОД ДИКТОРА ПРИ РАСПОЗНАВАНИИ ОГРАНИЧЕННОГО НАБОРА УСТНЫХ КОМАНД

Н.Г.Загоруйко, В.С.Лозовский

Для всех известных в настоящее время алгоритмов распознавания ограниченного набора устных команд характерно более или менее значительное ухудшение надежности распознавания при росте числа дикторов, участвующих в эксперименте. Еще худшие результаты получаются при распознавании слов, произнесенных дикторами, не участвовавшими в формировании обучающей последовательности. С другой стороны, на практике встречаются случаи, когда допустима подстройка под диктора перед распознаванием произносимых им команд, если этот процесс будет простым и быстрым. Если бы в этом направлении были получены хорошие результаты, то для ряда применений распознающих автоматов можно было бы отказаться от универсальности, подстраивая параметры автомата при непродолжительном предварительном знакомстве с диктором. Рассмотрению такой возможности и посвящена эта работа.

Пусть $X\{X_j\}$, , где $j=1,\ldots,n$, n - мерное пространство параметров распознавания, Произносимые команды отображаются в точки пространстве X. Будем называть их реализациями. Каждую из реализаций учитель относит к одному из K классов $S_1,\ldots,S_k$. В настоящей работе исследовалась подстройка под диктора для алгоритма распознавания по миниму-

му евклидова расстояния до эталонов.

Под надежностью распознавания контрольной последовательности реализаций будем понимать отношение правильно распознанных к их общему числу.

Пусть τ_j $(j=1,\dots,\ell)$ - обучающая последовательность для класса S_i : $\tau_j \in S_i$. Если экспериментатору известны априорные вероятности каждой реализации P_j и нормированные веса W_j , учитывающие типичность каждой реализации, то эталонный вектор определяется из соотношения:

$$\mathcal{Э}_i = \frac{1}{\ell}\sum_{j=1}^{\ell} W_j P_j \tau_j$$

Здесь мы не будем учитывать W_j и P_j , так что

$$\mathcal{Э}_i = \frac{1}{\ell}\sum_{j=1}^{\ell} \tau_j . \qquad (I)$$

Будем считать, что нами получены $\mathcal{Э}_i^{(1)} \in X$ $(i=1,\dots,k)$ по обучающей последовательности одного диктора и $\mathcal{Э}_i^{(2)} \in X$ $(i=1,\dots,k)$ - по обучающей последовательности второго. На рис. I этот случай изображен для $n=2$, $k=3$.

Расстояния между одноименными эталонами этих дикторов выражаются n -мерными векторами $\mathcal{Э}_i^{(2)}\mathcal{Э}_i^{(1)} = \mathcal{Э}_i^{(2)} - \mathcal{Э}_i^{(1)}$.

Далее, мы вправе написать :

$$\mathcal{Э}_i^{(2)}\mathcal{Э}_i^{(1)} = C_i\mathcal{Э}_i^{(1)} + \mathcal{Э}_i^{(2)}C_i \quad \text{,где,} \quad C_1\mathcal{Э}_1^{(1)} = C_2\mathcal{Э}_2^{(1)} = \dots = C\mathcal{Э} ,$$

где

$$\mathcal{Э}_i^{(2)}\mathcal{Э}_i^{(1)} = C\mathcal{Э} + \mathcal{Э}_i^{(2)}C_i .$$

Здесь постоянный для всех классов вектор смещения СЭ можно трактовать как поправку, учитывающую глобальную разницу между дикторами: высоту голоса, особенности произношения, темпа речи и т.п. Векторы же $\mathcal{Э}_i^{(2)}C_i$ характеризуют имеющиеся локальные отклонения, обусловленные особенностями произношения вторым диктором конкретных команд из данного набора.

Итак, высказывается следующая гипотеза об относительной стабильности положения эталонов, в которой предполагается, что

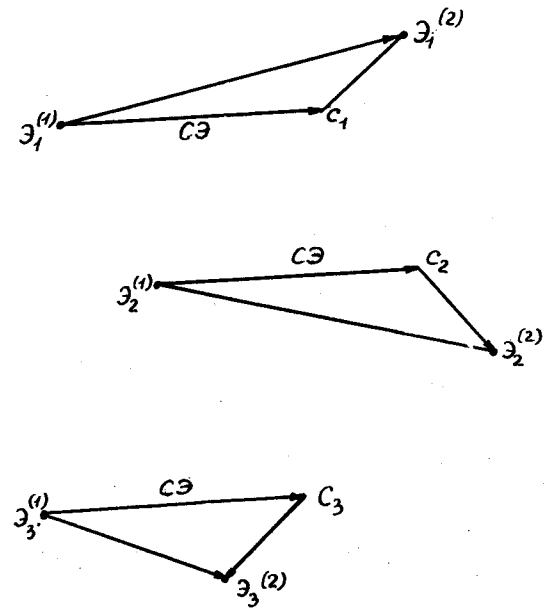

Рис.I

для набора эталонов одних и тех же команд, произнесенных двумя дикторами, в пространстве X существует такой вектор смещения СЭ, что $|\mathcal{Э}_i^{(2)} C_i| < |C\mathcal{Э}|$ для $i = 1, \ldots, k$. Другими словами, предполагается, что глобальные различия между дикторами могут быть велики, но взаимные расположения эталонов для каждого диктора в пространстве X будут разниться не столь сильно.

По определению вектора смещения СЭ, последний должен определяться при минимизации некоторого функционала F от локальных отклонений $\mathcal{Э}_i^{(2)} C_i$. Учитывая принятую нами процедуру распознавания по минимуму евклидова расстояния до эталонов, имеем:

$$F_1 = \sum_{i=1}^{\ell} |\mathcal{Э}_i^{(2)} - (\mathcal{Э}_i^{(1)} + C\mathcal{Э})|^2.$$

В данной работе с целью экономии времени вычислений процедура минимизации F была заменена процедурой непосредственного счета для более простого функционала:

$$F_2 = \sum_{i=1}^{\ell} [\mathcal{Э}_i^{(2)} - (\mathcal{Э}_i^{(1)} + C\mathcal{Э})]. \qquad (2)$$

Полагая (2) равным нулю и осуществляя элементарные преобразования, имеем:

$$C\mathcal{Э} = \frac{1}{\ell} \sum_{i=1}^{\ell} \mathcal{Э}_i^{(2)} - \frac{1}{\ell} \sum_{i=1}^{\ell} \mathcal{Э}_i^{(1)} = \mathcal{Э}^{(2)} - \mathcal{Э}^{(1)}, \qquad (3)$$

где $\mathcal{Э}^{(1)}$ и $\mathcal{Э}^{(2)}$ - центры распределений эталонов первого и второго дикторов.

Гипотеза об относительной стабильности положения эталонов получит подтверждение, если надежность распознавания контрольной последовательности одного диктора по эталонам другого с учетом вектора смещения СЭ окажется выше, чем без учета последнего.

Эксперимент заключался в распознавании десяти классов - произносимых цифр от "0" до "9" в пространстве огибающих на

выходе пяти октавных полосных фильтров. Аппаратурно были выполнены: пятиканальный полосный анализатор, коммутатор аналогового сигнала и цифратор (преобразователь напряжение-код). Остальные операции выполнялись программно на ЭЦВМ М-20. Опрос пяти каналов коммутатором происходил примерно за 2 мсек, отсчеты снимались каждые 20 мсек, точность цифратора - 6дв.разрядов. Программным методом находилось начало и конец каждого слова, затем производилась линейная нормализация по динамическому диапазону (максимальный отсчет за время звучания слова вытягивался до фиксированного уровня) и линейная нормализация по длительности (до 50 отсчетов). Распознавание велось в пространстве 250 измерений. Дикторов двое: мужчина с предварительной тренировкой в произнесении слов и женщина без особой подготовки. Дикторам было предложено произнести каждое из 10 слов по 40 раз в нормальном темпе и полным стилем.

Положение эталонов $\mathcal{Э}_i^{(1)}$ и $\mathcal{Э}_i^{(2)}$ определялось в соответствии с (1). Поскольку цель эксперимента - лишь исследование возможности подстройки под диктора, в качестве контрольной последовательности использовалась обучающая. Вектор смещения эталонов СЭ определялся по формуле (3). В табл.1 приведены результаты распознавания слов, произнесенных первым диктором, а в табл. 2 - вторым. На пересечении каждой строки и столбца таблиц помещены четыре числа (нули опущены):

1. надежность распознавания в процентах реализаций данного диктора по эталонам первого диктора;

2. то же, но по эталонам второго диктора;

3. то же, но по эталонам первого диктора плюс вектор смещения;

4. то же, по эталонам второго диктора минус вектор смещения.

Пример: Слово "ноль", произнесенное первым диктором, распознано, как "ноль" по эталонам первого диктора с надежностью 100%, по эталонам второго - 26%. Смещение эталона 1-го диктора снизило надежность распознавания его реализаций до 87% и улучшило распознавание реализаций второго диктора до 66%. Отдельно на каждой таблице указаны четыре числа, характеризующие надежность распознавания для каждого из четырех случаев, усредненную по десяти словам.

Из приведенных данных следует, что учет векторов смещения повысил надежность распознавания реализаций первого дикто-

Таблица I.

Результаты распознаваемых команд, произносимых первым диктором.

ПРОИЗНЕСЕНО ПЕРВЫМ ДИКТОРОМ \ РЕЗУЛЬТАТЫ РАСПОЗНАВАНИЯ	ноль	один	два	три	четыре	пять	шесть	семь	восемь	девять
ноль	100 26 87 66		66 13 32			5 2	3			
один		100 50 100 55		48 40	2 5					
два	8 3 3 16		92 91 5		3	76 65	21 14			3
три	13			100 37 100 40		5 8	27 32	6 7	25	
четыре					100 65 100 73			35 27		
пять		17 38				100 35 100 72			48	
шесть		3			2		100 100 13		93 57	7 25
семь								100 100 100 100		
восемь		15				3 5			100 97 100 80	
девять	3	13 12	3	5 3		72 82			10	97 100

98,9 41
97,8 50,4

Таблица 2.

Результаты распознаваемых команд, произносимых вторым диктором.

ПРОИЗНЕСЕНО ВТОРЫМ ДИКТОРОМ \ РЕЗУЛЬТАТ РАСПОЗНАВАНИЯ	ноль	один	два	три	четыре	пять	шесть	семь	восемь	девять
ноль	100 70 90 89		25 5 10		5	3		2 2		
один		79 97 62 97				11 3 15 3			10 18	5
два	95 8 67 30		5 77 80 65		2 8	7		6 5		5
три		89 6 81 8	11 6	95 8 92		5				
четыре		80 18			20 100 80 100		2			
пять	32 8 15 17		28 12 33 8			40 80 45 75				
шесть	40 7		8 20		13		45 95 62 95	7 5 8 5		
семь								100 100 100 100		
восемь						10 8	23 3 22 3		67 97 70 87	
девять		10				18 15	60 70	22 12		100 3 100

45,6 89,1
53 89,9

ра по эталонам второго в среднем на 9,4%, а надежность распознавания реализаций второго диктора по эталонам первого – на 7,4%. Можно заметить, что учет вектора смещения "портит" свои эталоны незначительно: надежность распознавания "своих" реализаций для первого диктора ухудшилась лишь на I,I% , а для второго даже улучшилась на 0,8%.

Объяснить замеченное улучшение можно несоответствием вида решающих функций характеру распределений. Представим себе, что плотность распределения реализаций для некоторых классов S_1 и S_2 неравномерна и имеет вид, изображенный на рис.2. Вычисление эталонов $\mathcal{Э}_1$ и $\mathcal{Э}_2$ по формуле (I) даст указанные на рисунке смещения их в сторону области с большими плотностями реализаций. Учет для каждого эталона некоторой векторной поправки Δ , как видно из рисунка, приведет к смещению разделяющей границы и повысит надежность распознавания реализаций обоих классов. Поэтому для указанного случая может оказаться целесообразным помещать эталоны $\mathcal{Э}_i$ не в центры тяжести реализаций каждого класса, а в центры занимаемых ими объемов.

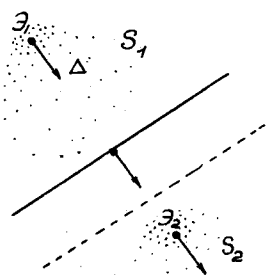

Рис.2. Корректировка эталонов.

Заканчивая обсуждение результатов, следует отметить, что гипотеза относительной стабильности эталонов для проведенного эксперимента подтвердилась, хотя, может быть, и не в такой степени, как того бы хотелось. Есть надежда , что еще более эффективные результаты могут быть получены для других параметрических пространств и для более совершенных алгоритмов классификации.

Несколько слов о практическом использовании результатов. Пусть в распознающем автомате хранятся координаты эталонов для наиболее типичного произношения, и пусть в его памяти хранятся векторы смещения СЭ для каждого из дикторов, реализации которых будут входить в состав контрольной последовательности. В каждом конкретном случае, в зависимости от того с каким диктором имеет дело автомат, учитывается нужный вектор смещения. Если диктор неизвестен, то распознавание ведется с учетом каждого из имеющихся векторов смещения и состоит в выборе того вектора смещения и того класса, при которых расстояние до поступившей реализации окажется минимальным. Так может выглядеть один из вариантов одновременного распознавания смысла сообщения и диктора.

Далее. Пусть алфавит распознавания велик. Представляется возможным определить вектор смещения СЭ не по всем классам, а лишь по тем, которых достаточно для нахождения СЭ с нужной степенью точности. Для этого необходимо провести дополнительную работу. На множестве из нескольких дикторов для данного алфавита распознавания вычисляются $\mathcal{Э}^{(j)}$ центры распределений эталонов для каждого диктора. Затем для каждого диктора выбираются эталоны, наиболее близкие к $\mathcal{Э}^{(j)}$. Если выбранными окажутся одноименные эталоны, – это будет означать, что найдены слова, которые достаточно полно отражают индивидуальные особенности каждого диктора. В дальнейшем вектор смещения эталонов можно искать, "переучивая" автомат не на всем словаре, а только на этой его небольшой части.

Поступила в редакцию
I/X-I967г.