

ТАБЛИЧНЫЙ МЕТОД ВЫДЕЛЕНИЯ ПРИЗНАКОВ РЕЧЕВОГО СИГНАЛА  
И ПОФОНЕМНОЕ РАСПОЗНАВАНИЕ РЕЧИ

ЛЮДОВИК ЕВГЕНИЙ КУЗЬМИЧ

Институт кибернетики им. В.М.Глушкова АН УССР  
Киев, СССР 252207

АННОТАЦИЯ

Поэтапное устранение избыточности из квантованных параметров описания речевого сигнала, а также учет контекста, то есть, окружения каждого интервала анализа направлены в конечном итоге на однозначное указание значения выделяемого признака, а в случае пофонемного распознавания - номера фонемы.

ВВЕДЕНИЕ

Применяемые в распознавании речи типы первичного описания речевого сигнала представляют собой многомерные векторы, между компонентами которых имеются различного рода зависимости. При этом каждая отдельная компонента несет не очень много информации о выделяемом признаке. Информация как бы размазана по компонентам, причем значительная часть информации о значении выделяемого признака содержится в контексте, то есть, в компонентах первичного описания сигнала на соседних интервалах. Естественно в связи с этим попытаться собрать по возможности всю информацию, отжав "воду" из компонент первичного описания, соответствующих интервалам анализа с их окружением.

Предлагаемый метод включает этап обучения, для которого необходимо накопить и разметить обучающую выборку /ОВ/. Формально размеченная ОВ представляет собой последовательность пар  $g(n), v(n), 1 \leq n \leq L$ , где  $g(n) = \{g_i(n)\}_{i=1}^m$  -  $m$ -мерный вектор

первичного описания речевого сигнала на  $n$ -м интервале анализа, а  $v(n)$  - значение выделяемого признака  $v$  на этом интервале,  $L$  - объем выборки в интервалах анализа.

Под выделяемым признаком могут пониматься признаки различного характера: например, признак тон/шум, артикуляционные параметры или номер фонемы. Во всех случаях предполагается, что область значений признака разбита на  $2^k$  подобластей, где  $k = 1$  для двоичных признаков и  $k = 6$  для номера фонемы являются двумя крайними значениями.

В качестве первичного описания могут использоваться самые различные наборы параметров, например, коэффициенты отражения и интенсивность сигнала или же значения энергии в частотных полосах. Могут применяться также и наборы разнородных параметров.

Обучение имеет целью построение преобразования параметров первичного описания данного интервала анализа и окружающих его в значение признака, приписанное этому интервалу.

Построение требуемого преобразования будет производиться поэтапно. При этом от этапа к этапу будет уменьшаться количество битов, необходимых для хранения описания. На заключительном этапе требуемое для этого количество битов должно совпасть с разрядностью  $k$  выделяемого признака.

Возможны два варианта постановки задачи построения преобразования. В одном

из них процесс построения преобразования направлен на повышение информационной насыщенности каждого бита описания. Во втором варианте целью каждого этапа наряду со сжатием описания является снижение вероятности ошибки распознавания значения признака по отдельным элементам описания очередного уровня. Именно этот вариант и будет рассмотрен ниже.

УМЕНЬШЕНИЕ ИЗБЫТОЧНОСТИ, ОРИЕНТИРОВАННОЕ НА СНИЖЕНИЕ ЧИСЛА ОШИБОК

Предлагаемый метод излагается на примере выделения признака, который может принимать значения от 0 до 15, то есть, требует для задания четырех бит.

Предположим пока, что все компоненты первичного описания проквантованы на 16 уровней и, таким образом, представлены четырехразрядным двоичным кодом.

Образуем составное восьмиразрядное описание из  $g_1(n)$  и  $g_2(n)$ :

$$g_{1,2}^1(n) = g_1(n) \cdot 16 + g_2(n).$$

Исходя из байесовского критерия, построим решающее правило для оптимального распознавания значения признака  $v(n)$  по  $g_{1,2}^1(n)$ . С этой целью для каждого восьмиразрядного двоичного кода  $k_8$  определим то значение  $k_4$  признака  $v$ , которое встречается наиболее часто в паре с  $k_8$ :

$$k_4 = f^1(k_8) =$$

$$\text{argmax}_{k_4} G\{n | g_{1,2}^1(n) = k_8, v(n) = k_4\}.$$

Функция  $G$  здесь означает количество элементов в множестве-аргументе.

Теперь можно определить первую компоненту описания второго уровня:

$$g_1^2(n) = f^1(g_{1,2}^1(n)).$$

Аналогичная процедура выполняется над

третьей и четвертой компонентами описания первого уровня, в результате получаем  $g_2^2(n)$  - вторую компоненту описания второго уровня и т.д. Таким образом мы построим преобразование описания первого уровня в описание второго уровня. При этом мы снизим информационный объем описания вдвое, учтем взаимосвязи между парами компонент и, вероятно, повысим надежность распознавания.

Таким же образом переходим к описаниям третьего и других уровней, до тех пор, пока не получим описание из одной компоненты  $g^5(n)$  /предположив, что первичное описание состояло из 16 компонент/.

Далее можно учесть окружение  $n$ -го интервала анализа. Для этого описанную выше процедуру склеивания следует применить к  $g^5(n)$  и  $g^5(n-1)$ ; в результате получим описание следующего уровня  $g^6(n)$ . Затем объединяем  $g^6(n)$  и  $g^6(n+1)$  и преобразуем в  $g^7(n)$  и т.д. С каждым этапом мы теперь увеличиваем размер фрагмента, участвующего в формировании решения о значении признака  $v(n)$  на  $n$ -м интервале анализа. Этот процесс может быть остановлен либо при достижении заданной длины фрагмента, либо в том случае, когда число ошибок при распознавании ОВ перестанет уменьшаться.

Если элементы первичного описания имеют большую разрядность, чем значение признака, на первом этапе следует выполнить процедуру склеивания для каждого такого элемента в отдельности.

Если, напротив, разрядность каких-то компонент описания меньше разрядности значений признака, при построении преобразования склеивания следует объединять большее число компонент, что позволит лучше использовать взаимосвязи между ними.

Рассмотрим теперь случай признака большей разрядности, например, пусть этим признаком будет номер фонемы.

Будем исходить из того, что число фонем не превышает 64, таким образом, значе-

ния признака  $v(n)$  в этом случае являются шестиразрядными двоичными кодами.

Процедура склеивания потребует в этой ситуации построения таблицы из  $2^6 \times 2^6 = 4096$  шестиразрядных двоичных чисел. В том случае, когда первичное описание состоит из 16 компонент и когда мы хотим для распознавания на  $n$ -м интервале использовать информацию о 4-х предыдущих и 4-х последующих интервалах, потребуется объем памяти 48 К байт.

Для уменьшения затрат памяти желательно научиться выполнять процедуру склеивания таким образом, чтобы в результате получались не шестиразрядные коды описаний следующих уровней, а, например, четырехразрядные.

В связи с этим возникает несколько интересных задач.

1. Требуется построить решающее правило, минимизирующее вероятность ошибки распознавания объектов из  $N$  классов, если ответами распознавания могут быть только  $t$  классов,  $t < N$ . При этом задается только число классов  $t$ , а какие именно это будут классы также необходимо определить.

2. Требуется склеить шестиразрядные коды в 16 групп так, чтобы объединение этого нового четырехбитового описания с некоторым другим фиксированным шестиразрядным позволило бы распознавать номер класса с минимальной вероятностью ошибки.

Решение любой из указанных задач позволяет заменить преобразование-склейку  $2^6 \times 2^6 - 2^6$  двумя преобразованиями:  $2^6 - 2^4$  и  $2^4 \times 2^6 - 2^6$ , что дает экономию памяти почти в четыре раза. Ясно, однако, что второй вариант в большей степени способствует повышению надежности распознавания.

Еще большую экономию памяти дало бы решение следующей задачи.

3. Допустим, что имеется два шестиразрядных описания  $g_1(n)$  и  $g_2(n)$ . Требуется найти два склеивающих преобразо-

вания  $f^1: 2^6 - 2^r$  и  $f^2: 2^6 - 2^s$ ,  $r+s=m$ ,  $6 \leq m$ , таких чтобы составное описание  $(f^1(g_1(n)), f^2(g_2(n)))$  позволяло бы распознавать номер класса (фонемы) с минимальной вероятностью ошибки.

В этой задаче  $r$  и  $s$  также являются искомыми параметрами. Задав  $m=6$ , мы получим экономию памяти не в 4 раза, как в предыдущем случае, а в десятки раз. При этом, однако, надежность распознавания будет меньше, чем в предыдущем случае.

Приняв  $m=10$ , мы получим вариант в который задача 2 входит в качестве частного случая. Поэтому здесь мы могли бы рассчитывать на повышение надежности распознавания по сравнению с вариантом 2.

Итак, результатом обучения будет совокупность склеивающих отображений, каждое из которых задается таблицей, содержащей информацию о разбиении множества комбинаций значений параметров на подмножества.

Распознавание с помощью этих таблиц тривиально. Нужно последовательно преобразовывать значения компонент описания в соответствии с этими таблицами; окончательный результат преобразования описания на данном интервале анализа и является ответом распознавания. Необходимыми операциями при этом являются операция объединения двух кодов в один код-аргумент и операция извлечения из соответствующей ячейки таблицы кода-значения параметра описания следующего уровня.

#### ЗАКЛЮЧЕНИЕ

Предложенный табличный метод способствует накоплению (концентрации) информации, существенной для распознавания. Процедура распознавания тривиальна и использует лишь две очень простые операции. Учет контекста дает основания надеяться на эффективность табличного метода распознавания. Использование многодикторной ОБ создает предпосылки для распознавания речи многих дикторов без подстройки.