СИСТЕМА АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ ФОНЕМ РУССКОГО ЯЗЫКА ПРИ ЕЕ ОБУЧЕНИИ МЕТОДОМ ГРУППОВОГО УЧЕТА АРГУМЕНТОВ

А.В.Аграновский, Д.А.Леднов, С.А.Репалов, Б.А.Телеснин

ГП КБ «СПЕЦВУЗАВТОМАТИКА», Россия, г.Ростов-на_Дону, e-mail:lednov@rnd.runnet.ru

This paper includes two purposes. First, it's an investigation of accuracy for the phonemes recognition system in continuous speech at its training by a group method of data handling (GMDH) [1]. Secondly, it's an investigation of the latent dependences between the characteristics of phonemes spectral segments, which allow to distinguish one phoneme from another with the help GMDH.

Введение

Эта публикация ставит перед собой две цели. Во-первых, исследовать точность работы системы, классифицирующей фонемы русского языка в слитной речи при ее обучении методом группового учета аргументов (МГУА) [1]. Во-вторых, с помощью МГУА выяснить скрытые зависимости между характеристиками спектральных сегментов фонем, которые позволяют отличать одну фонему от другой.

1. Предварительная обработка речи и стратегия обучения

Перечислим некоторые важные этапы и параметры предварительной обработки речи в описываемой системе:

а) речь оцифровывалась с частотой 10кГц;

б) в качестве предварительных данных характеризующих фонемы был выбран спектр Фурье (в дальнейшем просто спектр) полученный на интервале длительностью T=0.05с.;

в) полученный спектр сегментировался на 19 перекрывающихся участков равной величины, как показано на рис. 1.

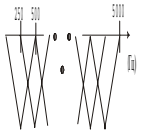

Рис.1

В каждом сегменте находятся значения средней частоты и средней интенсивности

$$P_i = \frac{1}{G_i^{(1)} - G_i^{(0)}} \sum_{j=G_i^{(0)}}^{G_i^{(1)}} s_j, \qquad \omega_i = \frac{\sum_{j=G_i^{(0)}}^{G_i^{(1)}} j s_j}{T \sum_{j=G_i^{(0)}}^{G_i^{(1)}} s_j} \qquad (1)$$

где $G_i^{(1)}, G_i^{(0)}$ -правая и левая границы сегмента, соответственно, s_j -интенсивность j-ой гармоники спектра.

Множество величин $\{x_i = (P,\omega)_i\}$ используется в качестве входных параметров МГУА.

В процессе обучения диктор многократно произносит фонемы, такие как: гласные, звонкие согласные и шипящие и вводит соответствующие им транскрипционные символы. Таким образом, к началу работы МГУА формируются множества спектральных параметров характеризующих фонемы.

2. Описание Метода Группового Учета Аргументов.

Пусть система обучена на F фонем, для каждой из которых получено k_f наблюдений спектра, где f индекс фонемы в алфавите.

Первая задача заключается в том, чтобы выяснить какие сегменты спектров, заданные параметрами (1), попарно отличают одну фонему от другой максимальным образом.

Для решения этой задачи, в соответствии с МГУА, предположим, что описание i-го и j-го сегментов спектра f-ой фонемы при n-ом наблюдении на t-ом шаге селекции можно сделать с помощью полинома второй степени

$$\varphi(x_{ni}^{(f)}(t), x_{nj}^{(f)}(t)) = a_{1ij}(t)x_{ni}^{(f)}(t) + a_{2ij}(t)x_{nj}^{(f)}(t) + a_{3ij}(t)x_{ni}^{(f)}(t)x_{nj}^{(f)}(t) + a_{4ij}(t)(x_{ni}^{(f)}(t))^2 + a_{5ij}(t)(x_{nj}^{(f)}(t))^2 \qquad (2)$$

Поскольку на первом шаге селекции сегмент характеризуется парой параметров $\{x_i = (P,\omega)_i\}$, то для этого шага полином (2) может быть представлен в форме

$$\varphi(x_{ni}^{(f)}(1), x_{nj}^{(f)}(1)) = a_{1ij}(1)(P_{ni}^{(f)} + P_{nj}^{(f)}) + a_{2ij}(1)(\omega_{ni}^{(f)} + \omega_{nj}^{(f)}) + a_{3ij}(1)(\omega_{ni}^{(f)} + \omega_{nj}^{(f)})(P_{ni}^{(f)} + P_{nj}^{(f)}) + a_{4ij}(1)(P_{ni}^{(f)} + P_{nj}^{(f)})^2 + a_{5ij}(1)(\omega_{ni}^{(f)} + \omega_{nj}^{(f)})^2, \qquad (3)$$

где $a_{1ij}(1)$-$a_{5ij}(1)$-неизвестные параметры. Здесь, для упрощения записи, опущены индексы фонем над параметрами $a_{1ij}(1)$-$a_{5ij}(1)$. Это не должно привести к недоразумению, т.к. во всех случаях эти параметры зависят от пары индексов фонем.

Для поиска значений параметров полинома (3), относительно каждой пары фонем (всего C_F^2 пар фонем), минимизируем функционал

$$\Delta_{ij}^{(fg)}(1) = \sum_{n=1}^{k_f}\sum_{m=1}^{k_g}(\varphi(x_{ni}^{(f)}(1), x_{nj}^{(f)}(1)) - 1)^2 + (\varphi(x_{mi}^{(g)}(1), x_{mj}^{(g)}(1)) + 1)^2 = \min \qquad (4).$$

Использование такого вида функционала можно пояснить следующим образом. Характеристики i-го и j-го сегмента f-ой фонемы можно изобразить как область в 2-мерном пространстве, если характеристики i-го сегмента откладывать вдоль оси

абсцисс, а характеристики j-го сегмента вдоль оси ординат. Эта область отображается на поверхность второго порядка, заданную полиномом (2). С помощью функционала накладывается условие, чтобы отображение области принадлежащей фонеме f было близко к 1, а отображение области принадлежащей фонеме g было близко к –1.

Дифференцируя (4) по неизвестным $a_{1ij}(1)$-$a_{5ij}(1)$ получим систему линейных алгебраических уравнений относительно этих параметров. Подстановка найденных значений $a_{1ij}(1)$-$a_{5ij}(1)$ в (4) и вычисление максимума функционала $\Delta_{ij}^{(fg)}(1)$ позволяют судить о величине вклада сегментов i и j в расстояние между фонемами f и g на первом шаге селекции.

В качестве критерия отбора пар сегментов, которые следует учесть на следующем шаге итерации установим следующее условие

$$r_{ij}^{(fg)}(1) > \frac{1}{C_F^2}\sum_{k=1}^{N}\sum_{l=1}^{N} r_{kl}^{(fg)}(1), \qquad (5)$$

где введено обозначение

$$r_{ij}^{(fg)}(1) = \sum_{n=1}^{k_f}\sum_{m=1}^{k_g}\left(\varphi(x_{ni}^{(f)}(1), x_{nj}^{(f)}(1)) - \varphi(x_{mi}^{(g)}(1), x_{mj}^{(g)}(1))\right)^2$$

Допустим, что на первом шаге селекции $D_1^{(fg)}$ пар сегментов, для фонем f и g, удовлетворяют условию (5). Для значений полиномов, образованных этими парами сегментов введем обозначения

$$x_1{}^{(f)}(2) = \varphi(x_{i_1}^{(f)}(1), x_{j_1}^{(f)}(1)),\ldots, x_{D_d}{}^{(f)}(2) = \varphi(x_{i_D}^{(f)}(1), x_{j_D}^{(f)}(1))$$

здесь для упрощения записи опущен индекс номера наблюдения.

Новые переменные $x_1{}^{(f)}(2), x_2{}^{(f)}(2),\ldots, x_d{}^{(f)}(2)$ позволяют провести второй шаг селекции. Для этого необходимо подставить их в полином вида (2) вместо переменных $x_{ni}^{(f)}(1), x_{nj}^{(f)}(1)$, затем провести операцию (4) для вычисления неизвестных коэффициентов $a_{1ij}(2)$ - $a_{5ij}(2)$ и найти значения минимума $\Delta_{ij}^{(fg)}(2)$ на втором шаге селекции. Сегменты, которые удовлетворяют условию вида (5), образуют входные данные для следующего шага селекции. Продолжая описанную методику можно продолжать селекцию далее.

Критерием остановки селекции на (n-1) шаге является условие

$$\frac{1}{D_{n-1}^{(fg)}}\sum_{k=1}^{N}\sum_{l=1}^{N} r_{kl}^{(fg)}(n-1) > \frac{1}{D_n^{(fg)}}\sum_{k=1}^{N}\sum_{l=1}^{N} r_{kl}^{(fg)}(n), \qquad (6)$$

где n – номер шага селекции.

Пусть селекция была прекращена на q–ом шаге, тогда на этом шаге селекции были определены полиномы

$$x_1{}^{(f)}(q) = \varphi(x_{i_1}^{(f)}(q-1), x_{j_1}^{(f)}(q-1)),\ldots, x_d{}^{(f)}(q) = \varphi(x_{i_D}^{(f)}(q-1), x_{j_D}^{(f)}(q-1)).$$

Причем, нам известны значения этих полиномов для любого наблюдаемого спектра как фонемы f, так и фонемы g. Предполагая, что величины $x_j^{(f)}(q)$ независимы и их плотность распределения является гауссовой, несложно найти дисперсию и математическое ожидание этого распределения, обозначим их $\sigma_i^{(f)}, \bar{x}_i^{(f)}$. Таким образом для каждой пары фонем из алфавита найдены распределения вероятностей значений полиномов зависящих от параметров характеризующих спектральные сегменты.

3 Классификация фонем

Пусть на интервале длительности T получен спектр характеризующий произнесенный звук и найдены значения средней интенсивности и средней частоты спектральных сегментов $\{(P,\omega)_i\}$.

Найдем, значение матрицы коэффициентов правдоподобия Λ

$$\lambda^{(gf)} = \prod_{i=1}^{D_q^{(fg)}} \frac{\sigma_i^{(f)} \exp\left\{-\frac{(\bar{x}_i^{(g)} - L_i(\{P,\omega\}))^2)}{2(\sigma_i^{(g)})^2}\right\}}{\sigma_i^{(g)} \exp\left\{-\frac{(\bar{x}_i^{(f)} - L_i(\{P,\omega\}))^2)}{2(\sigma_i^{(f)})^2}\right\}}, \quad g = 1,2,..,F; f = 1,2,..,F; \qquad (7)$$

где $L_i(\{P,\omega\})$ - полином степени 2q, полученный путем q подстановок полиномов предыдущих шагов селекции в полиномы последующих шагов. Заметим, что $\lambda^{(gf)} = 1/\lambda^{(fg)}$.

Очевидно, что если произнесена фонема g, то сумма элементов строки матрицы Λ, соответствующей произнесенной фонемы должна преобладать над суммами элементов всех прочих строк, т.е. номер фонемы в алфавите вычисляется в виде

$$r = \arg\max_g \left\{\sum_{f=1}^{F} \lambda^{(gf)}\right\} \qquad (8)$$

4. Заключение.

Описанная методика применялась для классификации гласных и звонких согласных фонем в потоке слитной речи. Эксперименты показали, что вероятность правильного опознавания фонем в данных условиях составляет 0.83.

Литература

1. Ивахненко А.Г. и др. Принятие решений на основе самоорганизации. М., «Сов. радио», 1976.