ЦИФРОВАЯ ОБРАБОТКА РЕЧЕВЫХ СИГНАЛОВ ДЛЯ ИХ КЛАССИФИКАЦИИ

Кастельянос Г., Кочетков Ю.А., Суарез Х.

МТУСИ

Аннотация – представлена методика для выбора эффективных акустических параметров с целью их последующей классификации. Сначала набор акустических характеристик подвергается предварительной статистической обработке с целью сокращения его размерности, при которой ещё сохронятся минимальная дискриминационная способность различения классов речи. Полученный сокращенный ансамбль акустических параметров служит первоначальной информацией для тренировки обучающей системы построенной на нейросетях. Представленная методика позволила значительно сократить количество параметров характеризующих классы речевых сигналов, соответствующих различным патологиям. Однако, дискриминационный анализ показал необходимость расчета более робастных характеристик, а именно, с большем интервалом оценки, менее коррелированных и менее чувствительные к повышению фонового шума при записи речевых сигналов.

Введение

Использование акустических параметров (АП) широко применяется для описания клинического состояния речи (нормальной или патологической), поскольку эта процедура позволяет выявлять особенности говорящего, которые сложно рассчитать другими методами. Однако, до сих пор полностью не ясно какова информационная емкость каждого из АП. В этом смысле является актуальным вопрос правильного выбора АП и их интерпретации с целью классификации речевых сигналов [1].

Акустический анализ речи требует большого количества АП, оценка которых должно проводиться в реальном времени, основываясь на процедуре кратковременного Фурье преобразования, которая в свою очередь очень чувствительна к шумовым условиям электронной записи сигнала. В статье предложена предварительная статистическая обработка начального набора АП с целью увеличения их эффективности по каждому из заданных классов речевых сигналов. Составление ансамбля эффективных АП совершается на основе выбора при заданном дискриминационном критерии. Для этого проводится исследование, как корреляционных свойств, так и информационной нагрузки полного ансамбля АП. Окончательная размерность ансамбля формируется с помощью использования статистической процедуры анализа главных компонентов (АГК). Предложенная методика ориентирована на автоматическое распознавания речи (АРР), которая состоит из двух этапов: *a*) Расчёт АП и выбор эффективного ансамбля для каждого из заданных классов речи. *b*) Тренировка обучающей системы АРР построенной на нейросетях, с использованием в качестве входа полученного эффективного ансамбля.

Предварительная обработка начального ансамбля АП

Начальный ансамбль характеристик речи состоит из набора квазигармонических и шумовых АП, оценка которых может проводиться в реальном времени на основе на процедуры кратковременного Фурье преобразования. На практике, набор дополняется коэффициентами ЛПК. Необходимо отметить, что все характеристики должны рассчитываться для каждой фонетической единичной структуры по отдельности. Итак, в случае простого анализа всех гласных на испанском языке общее количество характеристик доходит до N_ζ=100 для каждого из k- классов, k=1,..,K.

Поскольку использованная кратковременная процедура чувствительна к шумовой обстановке электронной записи, то необходимо провести предварительную статистическую обработку начального набора АП с целью увеличения их эффективности по каждому из заданных K классов. С другой стороны, необходимо определить объём выборки N_a речевых сигналов по каждому из классов, который строго говоря, определяет способ статистической предварительной обработки. Начальный ансамбль характеристик речи формируется на основе статистических моментов (обычно ограничиваются средним значением и дисперсией каждого АП) выборки речевых сигналов по каждому классу. По этому значение N_a рассчиталось исходя из условия обеспечения заданной точности любого из моментов. В данной работе при заданном уровнем значимости p=0.1 и для относительного значения ошибки оценки 10%, получен ансамбль $N_a \geq 90$ выборок для каждого k-ого класса.

Следующая процедура предварительной обработки АП состоит в устранения аномальных значений возникших из-за возможных ошибок при записи электронных сигналов. В данном случае, аномальное значение по совокупности $\{\zeta_i\}$ выражается критической величиной $t_{p,n-2}$ распределения Студента [2]:

$$| \zeta_i - m_{1\zeta} | / \sigma_\zeta \le t_{p,n-2} \, (n\text{-}1)^{1/2} / \, (n - 2 + (t_{p,n-2})^2)^{1/2}, \qquad (1)$$

где $m_{1\zeta}$ - среднее значение а σ^2_ζ - дисперсия исследуемого экстремального значения. С целью сравнения разномасштабных АП, целесообразно проводить их нормировку, т.е.

$$\xi_i = (\zeta_i - m_{1\zeta}) / \sigma_\zeta \, , \qquad (2)$$

Использованная методика оценки АП требует проверки гипотеза о гауссовости выборки. Данная проверка осуществляется на основе критериев χ^2, или Колмогорова-Смирнова. В том случае, когда проверка явно показывает на не гауссовость выборки, желательно определить вид её распределения, а также способ преобразования в нормальное. В общем случае, преобразование переменной ς_i в нормально распределенную переменную ξ_i совершается с помощью выражения:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi i} e^{-t^2} dt = \sum_{i=1}^{r} \frac{M_{ji}(A_{ji-1} < \zeta_j \le A_{ji})}{N_j} \, , \qquad (3)$$

где $N_j \le N_a$ объём выборки ансамбля ς_i, r – число интервалов гистограммы, A_{ji}, $A_{ji\text{-}1}$ – экстремальные значения i – ого интервала гистограммы, M_{ji} частота в i –том интервале. Однако не всегда имеется возможности пользоваться выражением (3). Альтернативный вариант для настройки плотности вероятность каждой выборки АП $\varsigma(l)$ состоит в использовании одной из следующих основных операций, которые находят широкое применение в обработке речевых сигналов:

$$\xi(l) = \lg(\zeta(l) \pm a)10^b, \; \xi(l) = \{\lg(a \pm \zeta(l))\}\zeta(l), \; \xi(l) = 1/(\zeta(l))^{1/a}, \qquad (4)$$

В любом случае, после каждого преобразования необходимо провести проверку на гауссовость выборки. В общем случае необязательно применять одну и ту же операцию преобразования плотностей распределения для всех АП, т.е. каждому АП ς_{ik} может соответствовать различный вид преобразования.

Выбор эффективного ансамбля АП

В системах АРР, главную роль играют АП, к которым налагаются следующие требования:

- Дискриминационная способность, т.е. они должны в максимальной степени способствовать как можно больше различению между классами речи.
- Эффективность, в том смысле, что объекты, принадлежащие к одному и тому же классу, должны иметь наименьшую дисперсию АП.
- Некоррелированность, т.е. использовать АП с наименьшей степенью статической зависимости.

Системы ААР для работы в реальном времени должны избегать обработки очень большого количества N_ς АП. Более того, чем больше их количество, тем сложнее правило классификации. Наконец, процедура тренировки обучающей системы не должна производится, пока не будет обеспечен минимальный уровень избыточности по совокупности АП. Из представленного следует необходимость уменьшения начальной выборки N_ς до значения $n_\varsigma \ll N_\varsigma$ обеспечивающий при этом минимальный уровень надёжности опознавания d_n, или сохранение условия $d_N - d_n = \varepsilon$, где ε - максимальное значение расходимости надежности опознавания и которое определяется из выражения

$$D = 1 - \sum_{k=1} p_a(k) \int_{K \cap k} p(x/k) dx$$

где $K \cap k$ – полное пространство решений, исключая k-ый класс, $p(k)$ – априорная вероятность появления сигналов k-ого класса, $p(x/k)$ условное распределение вероятности появления значения x для АП принадлежащего ко k-ому классу.

Для уменьшения размерности общего объёма АП предложено использование следующих процедур: *a*) Корреляционный анализ, который попарно производится для всех АП. *b*) Дискриминационный анализ, в этом случае подразумевается расчет индекса Фишера, который оценивает степень различия между классами. В общем, любой набор АП проявляет тем большую дискриминационную способность, чем больше его значение индекса Фишера. c) Анализ главных или принципиальных компонентов, который даёт окончательный сокращенный объём АП - n_ς.

Необходимо отметить, что уровень входного шума может увеличиваться при разных неблагоприятных условиях электронной записи сигнала речи, что приводит к тому, что превышается допустимая величина ошибки $\delta_l(l) \le \delta_{max}$, при которой можно ещё считать оценку удовлетворительной, здесь

$$\delta_i(l) = E\left[\frac{\left|\zeta_i(l) - \tilde{\zeta}_i\right|}{\zeta_i(l)}\right], \forall\, l = 1,..., N_a, i = 1,.., n_\zeta \ ,$$

$\tilde{\zeta}_i$ среднее значение оценки *i*-ого АП, сформированное по выборке $\zeta(l)$.

Выводы

Применение представленной методики для эффективного выбора АП, требуемого для работы АРР привело к следующим результатам [3]:

- Существенное уменьшение объём выборки АП, в конкретном случае от N_ζ=100 до n_ζ=16.
- При этом, повышение уровня успешной работы до 75%, при использовании нейросетей в качестве классификатора.
- Приложенные процедуры позволяют дополнительно найти систематические ошибки измерения АП связанные с электронной записью.
- дискриминационный анализ показал необходимость расчета более робастных характеристик, а именно, с большем интервалом оценки, менее коррелированных и менее чувствительных к фоновому шуму при записи речевых сигналов.

Литература

[1] Michaelis, D., Frohlich, M., & Strube, H. W. Selection and combination of acoustic features for the description of pathologic voices. J. Acoust. Soc. Am. Vol 103, n 3, pp 1628-1639, 1998

[2] Петрович М.Л. Давидович М.И. Статистическое оценивание и проверка гипотез на ЭВМ. Финансы и Статистика М. 1989.

[3] Castellanos, G. Vargas F., Análisis Acústico en la Clasificación de señales de voz empleando RNA, Congreso Internacional de Inteligencia Computacional., pp 107-11. 2001.(на исп.)

DIGITAL SIGNAL PROCESSING IN VOICE CLASSIFICATION

Castellanos G., Kochetkov Y., Suarez J.

A methodology is proposed for selecting voice acoustic Features (AF), which are oriented to speech classification between normal or pathological signals. Acoustic Voice Analysis (AVA) requires a high number of AF values [3]. In this paper, real-time estimate algorithms are accomplished for the principal AF, which are usually very sensitive to acoustic measurement conditions [3]. A preprocess procedure of the AF is carried out in order to increase the effectiveness of representative parameters for each speech class to be recognized. The initial assemble of acoustic voice characterization is built over this preprocessed AF. The formation of the effective assemble is performed by selecting the AF with better discriminant power capability by means of correlation and Fisher index criteria. The final reduction of the AF assemble dimension is accomplished using the Principal Component Analysis (PCA). The recognition procedure developed in this paper makes up of two basic parts: a) Extraction and Selection of the effective AF assemble for each class, b) the effective AF assemble taken as input of a Neural Network (NN). According to the acoustic properties that AF should measure, they are grouped into two categories [3]: the quasi-periodic ones and those used for perturbation measurement. All AF were calculated using the Short Time Fourier Transform (STFT) procedure [6,7]. As template words all five Spanish vowels (*/a//e//i//o//u/*) were taking for the acoustic voice analysis. As a result the initial AF assemble consists of N_{ζ} values for each k speech class, k=1,..,K. In a speech recognition system, the fundamental aspect is the selections of the effective AF assemble of the discriminant features. The proposed selection of effective AF assemble can be described as follows:

- Form the initial AF matrix ($N_a \square N_{\square}$): Adquisition of the initial sampling set of raw voice signals. Estimate the voice AF $\square_i$, i=1,..,N_{ζ}. Determine K classes number and compute the minimum samples number N_a by each class.
- Preprocess the AF samples $\square_i$: Statistical standardization, anomalous values rejection, hypothesis test, variables transformation, parametric effectiveness analysis (correlation analysis, Fisher index). Correct size of the initial AF matrix.
- Select the effective set: Variable reduction - PCA. Conformation of effective AF matrix ($N_a \square n_{\square}$).

As a result, the following conclusions can be made:

- The proposed selection methodology reduces the AF assemble dimension considerably, in our case it was from 100 to 16.
- The success level obtained on classification procedure using NN raises and it was about 75%.
- The preprocess procedure of the initial set improves substantially the training assemble quality. Likewise it allows to detect measurement failures that cannot be recognized by perceptual approach.

Based on the above-mentioned the following recommendations can be made. At first, to employ more robust AF estimate algorithms that should be less vulnerable to the noise, thus increasing the assemble reliability. Secondly, to enrich the assemble with new voice AF that offer new information and allow a higher discriminant degree among the speech classes.