

АЛГОРИТМ ПОСТРОЕНИЯ АДАПТИВНОГО ВЕКТОРА ПРИЗНАКОВ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ И МЕТОДА ОПОРНЫХ ВЕКТОРОВ

А. М. Сорока, П. Е. Ковалец, И. Э. Хейдоров

Белорусский государственный университет

Минск, Беларусь

E-mail: soroka.a.m@gmail.com, feanork-pk@yandex.by, ikheidorov@gmail.com

Представлен оригинальный метод построения адаптивного вектора признаков речевых сигналов на основе вейвлет-преобразования и метода опорных векторов. Для построения базовой вейвлет-функции предложено использовать генетический алгоритм, в котором в качестве целевой функции выступает точность классификации речевых сигналов методом опорных векторов. Показано, что использование базовых вейвлет-функций, сгенерированных на основе представленного алгоритма, для построения векторов признаков позволяет получить устойчивое повышение точности классификации. В частности, в сравнении с мел-частотными кепстральными коэффициентами улучшение точности составляет от 1 % до 5 % для различных фонем.

Ключевые слова: вейвлет-преобразование; вектор признаков; метод опорных векторов; классификация; речевые сигналы; генетический алгоритм.

Введение

Одной из актуальных задач обработки акустических сигналов является задача классификации речевых сигналов. Первый этап этой задачи – построение признакового описания. Для этого используются различные методы на базе спектрального анализа. Наиболее широко применяются мел-частот кепстральные коэффициенты (МЧКК) [1], расчет которых производится на базе преобразования Фурье, позволяющие получить вектор признаков речевого сигнала с учетом восприятия звука человеком. Данный метод не позволяет получить достаточно высокую различительную способность построенного признакового описания, что снижает точность классификации в целом [1].

Для анализа нестационарных сигналов широко используется вейвлет преобразование [2,3], которое позволяет получить оптимальное частотно-временное разрешение, что в свою очередь позволяет локализовать малые изменения сигнала как в частотной, так и во временной областях. Данное свойство может сыграть значимую роль при классификации речевых сигналов, однако использование вейвлет-преобразования с основными базовыми функциями для построения признаковых описаний речевых сигналов не позволяет достичь точности классификации, сравнимой с МЧКК. В связи с этим в данной статье предлагается метод построения адаптивной вейвлет-функции для признакового описания речевых сигналов, позволяющая улучшить точность классификации акустически схожих фонем.

Метод построения адаптивной вейвлет функции на основе генетического алгоритма и метода опорных векторов

Вейвлет-преобразование представляет собой интегральное преобразование [2], которое задается функцией двух переменных:

$$W(\alpha, \tau) = \int_{-\infty}^{+\infty} \frac{s(t)}{\sqrt{\alpha}} w\left(\frac{t-\tau}{\alpha}\right) dt, \quad (1)$$

где $s(t)$ – анализируемый сигнал, $w(t)$ – вейвлет функция, α – масштаб, τ – сдвиг во времени.

Предположим, что существует вейвлет функция $w_{\vec{p}}$, которая описывается набором параметров \vec{p} и позволяет локализовать значимые коэффициенты вейвлет-преобразования $W(\alpha, \tau)$ таким образом, чтобы обеспечить наилучшее разделение заданных классов в признаковом пространстве. Для оценки распределения векторов в признаковом пространстве будем использовать линейный классификатор на базе метода опорных векторов [3]. Для случая бинарной классификации данный метод позволяет построить оптимальную разделяющую гиперплоскость $\langle \vec{b}, \vec{x} \rangle - b_0 = 0$, максимизируя расстояние между разделяющей гиперплоскостью и границами классов. Решающее правило для такого классификатора описывается выражением:

$$a(\vec{x}) = \text{sign}\left(\langle \vec{b}, \vec{x} \rangle - b_0\right), \quad (2)$$

используя вейвлет-преобразование (1) построим набор признаковых описаний $X = \{x_i | i = 1 \dots N\}$ для речевых сигналов двух различных классов, точная классификация $c(\vec{x}) = \{-1; 1\}$ которых известна.

С использованием процедуры кроссвалидации [3] построим решающее правило с использованием (2). Для этого разделим множество X на K непересекающихся подмножеств $\{X_i | X = X_1 \cup X_2 \dots \cup X_k; X_i \cap X_j = \emptyset; i = 1 \dots K; j = 1 \dots K; i \neq j\}$, при этом каждое подмножество X_i содержит n_k векторов признаков. Обучим набор классификаторов $a_k(\vec{x}) | k = 1 \dots K$ (3), используя набор векторов признаков $X^l = X \setminus X_k$. В качестве целевой функции станем использовать среднюю точность классификации набором классификаторов (3):

$$f = \frac{1}{K} \sum_{k=1}^K \frac{1}{n_k} \sum_{\vec{x} \in X_k} \frac{|a_k(\vec{x}) + c(\vec{x})|}{2} \quad (4)$$

Для построения требуемой адаптивной вейвлет-функции необходимо найти вектор $\vec{p} = (p_1, p_2, \dots, p_n)$, $p_i \in R$, задающий вейвлет-преобразование $W(\alpha, \tau)$, такой, что $f \rightarrow \max$.

Для построения функции вейвлет-преобразования, адаптированной для классификации исследуемых акустических сигналов, используем численные методы оптимизации. В качестве метода оптимизации целевой функции выберем генетический алгоритм [4], поскольку он обладает рядом достоинств, важнейшие из которых:

- Возможность нахождения глобального экстремума целевой функции.
- Скорость схождения.
- Возможность распараллеливания алгоритма, что позволяет существенно сократить время расчета, при использовании современных систем параллельной обработки данных.

Генетический алгоритм оперирует вектором \vec{p} опосредствованно через последовательность кодовых символов $\vec{q} = (q_1, q_2, \dots, q_n)$, которую в теории эволюционных методов оптимизации принято называть хромосомой. Хромосома \vec{q} однозначно определяет вектор параметров \vec{p} , при этом:

1. Каждый параметр $p_i, i = 1, \dots, N$ описывается соответствующим геном

$q_i, i \in 1, \dots, N$.

2. Каждый ген $q_i, i \in 1, \dots, N$ состоит из M аллелей, которые выбираются из конечного множества. Для удобства реализации генетического алгоритма используется конечное множество аллелей $\{0, 1\}$.

Для построения адаптивной функции вейвлет-преобразования зададим временное представление искомой функций параметрической кривой, а именно сплайном Акимы [5]. Данный выбор параметрической кривой обусловлен тем фактом, что результирующая кривая проходит через все точки, во-вторых устойчивостью сплайна Акимы к локальным выбросам – у данного сплайна практически отсутствуют осцилляции кривой вблизи точек выброса, в отличие от кубических сплайнов [5]. Данное свойство сплайна Акимы значимо, так как любые дополнительные осцилляции ухудшают локализацию вейвлет-функции в частотной области. Сплайн задается набором ординат базовых точек $p(n) = \{p_i, i = 1, \dots, N\}$, где значение p_i кодируется геном q_i , причем $p_i \in [-1, 1]$. Так как геном состоит из конечного числа аллелей, соответственно значения кодируются с некоторой точностью ε . В таком случае максимальное количество значений, которое может кодироваться геном q_i , описывается выражением:

$$K_{\max} = \frac{\max(p_i) - \min(p_i)}{\varepsilon}.$$

Минимальное количество аллелей в гене, необходимое для кодирования всех значений, составляет:

$$L(q_i) = \left\lceil \log_2 \left(\frac{\max(p_i) - \min(p_i)}{\varepsilon} \right) \right\rceil.$$

Операции мутации и кроссовера в данном случае тривиальны [6], так как любая комбинация аллелей легитимна.

В качестве алгоритма отбора особей для формирования следующей популяции используется алгоритм элитарного отбора, поскольку данный алгоритм обеспечивает более высокую скорость схождения при решении задачи оптимизации вейвлет-функции в сравнении с методами рулеточного или турнирного отборов [6].

Решением сформулированной выше задачи оптимизации является вектор параметров \vec{p} , описывающий адаптивную вейвлет-функцию, локализирующую значимые вейвлет-коэффициенты в заданной полосе масштабов, что позволяет повысить различительную способность признакового описания.

В качестве одного из преимуществ построенной адаптивной вейвлет-функции является тот факт, что она учитывает динамику сигнала на всем протяжении, поскольку на каждом масштабе строится вейвлет-функция, зависящая от всего сигнала, а соответственно содержащая в себе информацию о всех изменениях в исследуемом сигнале. В классических методах первичного анализа для учета динамических характеристик акустического сигнала на стадии построения признакового описания традиционно используются дельта-коэффициенты [6], позволяющие учесть только локальные изменения сигнала, в отличие от предлагаемой адаптированной вейвлет функции.

Эксперимент

С использованием предложенного алгоритма построены базовые вейвлет-функции, адаптированные для анализа различных классов фонем. Для построения адаптивной вейвлет-функции использовался описанный в данной статье метод с ис-

пользованием генетического алгоритма, при этом количество оптимизируемых параметров составило $N = 256$, размер популяции – 25 000, $K_{max} = 2000$, $L(q_i) = 11$.

Для оценки различительной способности адаптивных векторов признаков (АВП), полученных с использованием предлагаемого метода, проведено сравнительное тестирование для различных признаков с различной размерностью d . Для сравнения выбраны признаки, полученные с использованием вейвлет-преобразования и функций Хаара $d = 13$, Морле $d = 13$; адаптивные признаки на базе предложенного метода ($d1 = 13$, $d2 = 26$); признаки, полученные с использованием МЧКК $d = 13$, дельта-МЧКК $d = 26$ (13 МЧКК + 13 дельта коэффициентов), дельта-дельта-МЧКК $d = 39$ (13 МЧКК, 13 дельта коэффициентов, 13 дельта-дельта коэффициентов). В табл. 1 представлены результаты линейной бинарной классификации фонем различных типов (гласные, сонорные, шумовые). При этом форма записи [ph] – [ph, ph1, ph2 ...] обозначает решение задачи классификации фонемы [ph] из множества фонем [ph, ph1, ph2...].

Таблица 1

Точность классификации различных фонем, %

	Хаар13	Морле13	АВП13	АВП26	МЧКК13	МЧКК26	МЧКК39
[a]–[a,o]	84,6	88,0	91,5	94,2	91,5	90,8	91,0
[a]–[гласные]	73,4	77,1	90,0	93,0	89,9	87,4	87,3
[a]–[все фонемы]	72,1	75,8	84,0	87,0	85,2	80,4	80,4
[н]–[н,м]	59,8	62,0	74,8	75,0	67,8	69,8	68,5
[н]–[все фонемы]	54,8	60,2	72,3	71,4	66,3	66,9	65,8
[з]–[з,ж,ш,с]	70,1	74,8	78,6	82,2	75,2	78,4	79,9
[з]–[все фонемы]	69,9	73,2	77,4	79,1	74,9	76,3	77,1

Согласно результатам в табл. 1 использование предложенного метода построения векторов признаков позволяет улучшить точность классификации акустически схожих фонем в среднем на 3%, что особенно заметно в случае классификации таких акустически схожих фонем, как [м],[н] (улучшение на 5,2%). Данный результат особенно значим, так как согласные фонемы – семантический каркас речевого сообщения [7]. Так же стоит отметить, что для достижения сравнимой точности предложенный алгоритм позволяет использовать признаковое описание в 1,5 раза меньшей размерности, чем МЧКК. Таким образом, данная методика может быть применена для организации многоуровневой системы классификации речевых сигналов [8].

Заключение

Предложен метод построения адаптивной вейвлет-функции для признакового описания речевых сигналов с использованием генетического алгоритма и метода опорных векторов. Показано, что использование базовых вейвлет-функций, сгенерированных на основе представленного алгоритма, для построения векторов признаков позволяет получить устойчивое повышение точности классификации. В сравнении с мел-частотными кепстральными коэффициентами улучшение точности составляет от 1 % до 5 % для различных фонем. Показано, что сгенерированная адаптивная вейвлет функция включает информацию о динамических процессах в заданном сигнале, и как следствие в 1,5 раза сокращается размерность признакового описания без потери точности классификации.

Библиографические ссылки

1. *Saon G., Chien J. Large-Vocabulary Continuous Speech Recognition System // IEEE Signal Processing, 2012, V. 29. № 6. P. 18–34.*
2. *Mallat S. A wavelet tour of signal processing, Third edition: The sparse way // Academic press, 2008.*
3. *Сорока А. М., Шу В., Хейдоров И. Э. Обнаружение патологии голосового тракта на основе вейвлет-преобразования и машин на опорных векторах // РИВШ, Минск, 2010.*
4. *Back T. Evolution algorithms in theory and practice // Oxford University Press, New York, 1996.*
5. *Akima H. A new method of interpolation and smooth curve fitting based on local procedures // J.ACM, 1970, V. 17. № 4. P. 589–602.*
6. *Jinjin Ye. Speech recognition using time domain features from phase space reconstructions // A master thesis, Marquette University, Milwaukee Wisconsin, 2004. P. 9–12.*
7. *Буланин Л. Л. Фонетика современного русского языка // Книжный дом Либроком. Москва, 2011.*
8. *Сорока А. М. Алгоритм двухэтапного распознавания фонем русского языка // Речевые технологии. № 1. 2010. Москва. С. 35–42.*