Алгоритм двухэтапного распознавания фонем русского языка

А.М. Сорока,
БГУ, Минск, Беларусь

Одним из основных подходов к решению задачи распознавания слитной речи является метод распознавания на основе классификации минимальных речевых единиц. Как правило, в качестве минимальных речевых единиц выбираются фонемы либо дифоны, в силу наилучшего соотношения размеров словаря минимальных речевых единиц и точности распознавания. Однако признаковые описания акустических реализаций фонем крайне неравномерно распределены в пространстве признаков. При этом различие между близкорасположенными реализациями нивелируется значительным различием между остальными реализациями. Такие близкорасположенные пары минимальных речевых единиц получили название «пар спутывания».

Введение

Существует несколько методов разрешения пары спутывания, которые основаны на построении лингвистических решёток и сетей спутывания, а также использования лингвистических моделей, учитывающих контекстные связи [1]. Эти методы обладают рядом очевидных недостатков, в числе которых — высокая трудоёмкость алгоритмов и необходимость создания лингвистической модели. В статье предлагается алгоритм двухэтапного распознавания фонем на основе метода опорных векторов с построением признакового описания на основе вейвлет-преобразования, который позволяет избежать чрезмерного возникновения пар спутывания за счёт более точной классификации отдельно взятой фонемы.

Метод опорных векторов

Метод опорных векторов (МОВ) впервые был предложен Вапником [2]. Этот метод в процессе обучения непрерывно минимизирует эмпирический риск. Использование Вапником в качестве эвристики выбора разделяющей гиперплоскости

предположения о минимизации ожидаемого риска путём максимизации отступов классов привело к высокой обобщающей способности алгоритма. В настоящее время МОВ успешно используется во многих областях.

Предположим, что у нас имеется множество объектов X, заданных при помощи n-мерных вещественных векторов x, где $x \in R^n$ и множество классов $Y = \{-1+1\}$. Объекты, для которых известно точное соответствие между признаковым описанием и классом, называются прецедентами. Множество прецедентов, используемых для настройки классификатора, называется обучающей выборкой, а сам процесс настройки — обучением классификатора.

Построим линейный пороговый классификатор:

$$a(x) = sign(\sum_{j=1}^{n} w_j x_j - w_0) = sign(< w, x > - w_0), \quad (1)$$

где $w = (w^1, \ldots, w^n) \in R^n$, $w_0 \in R$. Уравнение $< w, x > = w_0$ задаёт разделяющую гиперплоскость, при этом w — вектор нормали к данной гиперплоскости, w_0 — расстояние от гиперплоскости до начала координат.

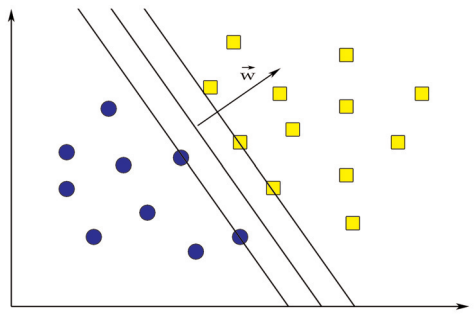

Рис. 1. Случай линейной разделимой обучающей выборки

Случай, при котором прецеденты линейно разделимы в признаковом пространстве, показан на *рис. 1*.

Рассмотрим функционал ошибок:

$$Q(w, w_0) = \sum_{i=1}^{l} [y_i (< w, x_i > - w_0) < 0] \quad (2)$$

Если существует такая разделяющая гиперплоскость $<w,x> = w_0$, что функционал (2) обращается в ноль, следовательно, множество объектов X является линейно разделимым на два класса. Очевидно, что в таком случае существует бесконечное число разделяющих гиперплоскостей. Вапником введено [2] понятие оптимальной разделяющей гиперплоскости — такой гиперплоскости, которая максимально удалена от границ обоих классов. Алгоритмы построения оптимальной разделяющей гиперплоскости с использованием метода Лагранжа могут быть найдены в специальной литературе [Ч]. Итоговый вид классификатора в данном случае может быть описан следующим выражением:

$$a(x) = sign(\sum_{i=1}^{l} \lambda_i y_i \langle x_i, x \rangle - w_0), \quad (3)$$

где $\lambda_i \in R, i = 1 \ldots m$ — коэффициенты Лагранжа.

На практике класс задач с линейно разделимой выборкой встречается крайне редко. Для решения проблемы классификации линейно неразделимых выборок Кортес и Вапник [3] предложили метод опорных векторов с мягким зазором. Фактически, они вводят неотрицательную величину ошибки классификации. Теперь проблема оптимизации представляет задачу минимизации ошибки классификации. В таком случае оптимальная разделяющая ги-

перплоскость определяется вектором w, который минимизирует следующий функционал:

$$(w \cdot x_i) + b \geq +1 - o, \quad if \quad y_i = +1$$
$$(w \cdot x_i) + b \leq -1 + o, \quad if \quad y_i = -1 \qquad 4)$$

Здесь $\xi = (\xi_1, \ldots, \xi_m)$ — вектор двойственных переменных, C — константа.

Этот подход не единственный для решения задачи в случае, если исходная выборка не является линейно разделимой в исходном признаковом пространстве. Предположим, что существует пространство более высокой, чем исходное, размерности, в котором исходная выборка окажется линейно разделимой (*рис. 2*). Переход от исходного пространства признаков X к новому пространству H может быть выполнен при помощи некоторого преобразования $\psi: X \to H$.

Таким образом, классификатор будет описываться следующим выражением:

$$a(x) = sign(< w, \psi(x) > - w_0), \qquad (5)$$

где (w, w_0) задают разделяющую гиперплоскость в расширенном пространстве. В таком случае итоговый вид классификатора записывается в следующем виде:

$$a(x) = sign(\sum_{i=1}^{l} \lambda_i y_i \langle \psi(x_i), \psi(x) \rangle - w_0) \qquad (6)$$

Анализируя выражение (6), можно видеть, что нет необходимости в явном виде задавать функцию отображения $\psi: X \to H$. Пусть существует функция $K(x_i, x_j)$ такая, что $K(x_i, x_j) = \langle \psi(x_i), \psi(x_j) \rangle$. В таком случае итоговый вид классификатора приобретает следующий вид:

$$a(x) = sign(\sum_{i=1}^{l} \lambda_i y_i K(x_i, x) - w_0) \qquad (7)$$

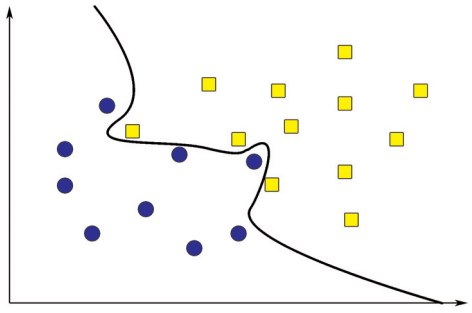

Рис. 2. Использование расширенного пространства

Функция $K(x_i, x_j)$ получила название ядра или ядерной функции. Стоит отметить тот факт, что здесь показано не единственное применение ядерной функции — данный класс функций получил широкое практическое применение.

Наиболее часто используются следующие ядерные функции:

линейная: $K(x, y) = x \cdot y$,

полиномиальная: $K(x, y) = (x \cdot y + 1)^d$, где d — степень полинома,

радиальная базисная Гауссова функция (RBF): $K(x, y) = \exp(-\frac{|x - y|^2}{2д^2})$,

где δ — ширина функции Гаусса.

Построение векторов признаков на основе вейвлет-преобразования

Для построения векторов признаков акустических сигналов в системах распознавания речи широко используются мелчастотные кепстральные коэффициенты (МЧКК) [5]. Однако, как показывают практические исследования [6], использование этого подхода не обеспечивает достаточной точности классификации акустических сигналов, что может быть обусловлено близостью векторов признаков в признаковом пространстве. В статье предложены два алгоритма извлечения векторов признаков на основе вейвлет преобразования, обладающего более высокой способностью к выделению локальных частотно-временных особенностей сигнала в сравнении с традиционным кратковременным Фурье-преобразованием.

В статье рассматриваются два алгоритма извлечения векторов признаков для речевых сигналов на основе вейвлет-анализа [6].

Первый алгоритм фундируется возможностью провести сегментацию и распознавание фонемы посредством визуального анализа графического представления результатов вейвлет-преобразования. Этот способ построения векторов признаков (ВП1) основан на методах смежной дисциплины — распознавания графических образов и может быть описан следующей последовательностью действий. Графический вейвлет образ сегментируется на участки, соответствующие одному периоду в квазипериодической трактовке вейвлет образа, далее в каждом сегменте детектируются резкие характерные изменения с использованием алгоритма детектора Харриса. Следующий шаг — нормализация координат полученных характерных точек. Для формирования вектора признаков характерные точки представляются в виде смеси двумерных Гауссовых распределений [7]:

$$p(x) = \sum_{j=1}^{K} w_j p(x \mid C^j), \qquad (8)$$

где w_j — весовой коэффициент,

$$p(x \mid C^j) = \frac{1}{(2\pi)^{n/2} \left|\Sigma_y\right|^{-1/2}} e^{-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)}, \qquad (9)$$

x — тестируемый вектор, C^j — предполагаемый кластер, K — количество компонент в смеси, Σ_j — диагональная матрица вида $\Sigma_j = \begin{bmatrix} \sigma_{11}^2 & 0 \\ 0 & \sigma_{22}^2 \end{bmatrix}$.

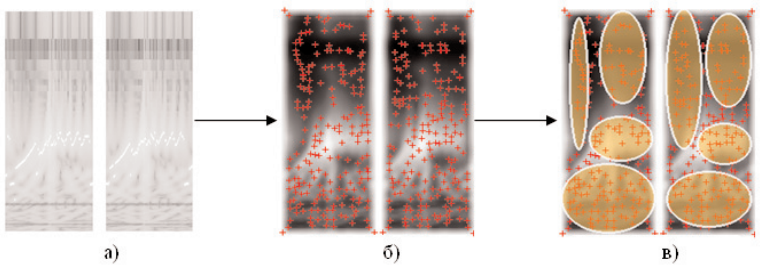

Рис. 3. Формирование вектора признаков с использованием методов анализа изображений — сегментация исходного изображения (а), нахождение характерных точек (б), аппроксимация распределения ключевых точек с использованием смеси Гауссовых распределений

Вектор признаков для заданного образа может быть описан следующим выражением:

$$x = \left(\mu_1^1, \mu_2^1, \sigma_{11}^1, \sigma_{22}^1, ..., \mu_1^K, \mu_2^K, \sigma_{11}^K, \sigma_{22}^K.\right) \qquad (10)$$

Метод продемонстрирован на *рис. 3.*

Во втором случае (ВП2) для формирования вектора признаков вейвлет-образ акустического сигнала разбивается на $3N$ прямоугольных окон, в каждом из которых находится усреднённая энергия S_{ij}, $i = 1...N$, $j = 1...3$. В данном случае вектор признаков описывается следующим выражением:

$$x = \left(S_{12}, \quad ... \quad S_{N2}, \quad \Delta_1, \quad ... \quad \Delta_N\right), \qquad (11)$$

параметры $\Delta_i = S_{i3} - S_{i1}$ введены для учёта динамических процессов в начале и конце фонемы, обусловленных эффектами редукции и коартикуляции. Алгоритм представлен на *рис. 4*.

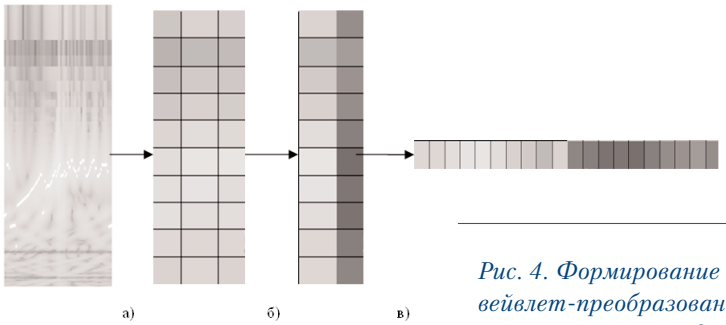

*Рис. 4. Формирование вектора признаков с использованием вейвлет-преобразования. Разделение вейвлет-образа на 3*N окон (а), учёт динамических процессов (б), формирование численных признаков (в)*

Двухэтапный метод распознавания фонем

В статье рассматривается двухэтапный метод распознавания фонем. Метод состоит из следующих этапов. На первом этапе производится классификация фонем по акустически схожим группам с использованием многоклассового классификатора на основе метода опорных векторов. Многоклассовый классификатор формируется из набора бинарных классификаторов, каждый из которых обучен по принципу «один против всех». На втором этапе производится классификация фонем внутри группы. Многоклассовый классификатор на втором этапе строится по принципу «каждый против каждого», что не влечёт за собой увеличения вычислительных затрат в силу малости групп, однако способствует более точному распознаванию.

Разделение фонем на акустически схожие группы определено эмпирическим путём на основе анализа ошибок распознавания многоклассовым классификатором. В данном случае классификатор строился по принципу «один против всех», при этом положительными прецедентами считались только реализации данной фонемы, а отрицательными — все остальные. Для каждой фонемы определялись наиболее частотные неверные результаты классификации, на основе которых делалось предположение о наличии пары спутывания. Далее была проведена глобализация пар спутывания с целью определения акустически схожих групп фонем. Итоговое разбиение фонем на акустически схожие группы представлено на *рис. 5.*

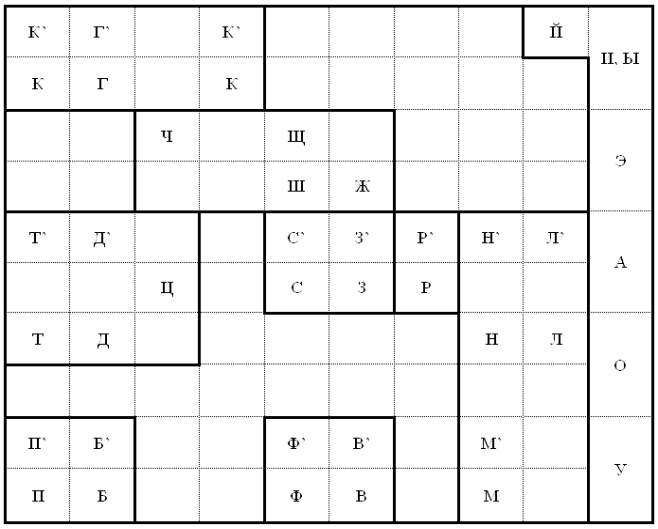

Рис. 5. Группы акустически схожих фонем

Экспериментальное исследование

Для определения характеристик разработанного метода было проведено экспериментальное исследование. Подготовлена база акустических реализаций фонем от разнополых дикторов на основе свободного речевого корпуса русского языка VoxForge [4] и акустической базы, подготовленной на кафедре радиофизики Белорусского государственного университета г. Минска. Общий объём базы составил 4500 фонем, в среднем по 100 реализаций каждой фонемы русского языка. Также была подготовлена тестовая выборка объёмом 100 реализаций на каждую из четырёх фонем: [а, м, н, д].

Для определения характеристик разработанных методов проведено сравнительное тестирование метода построения векторов признаков на основе мел-частотных кепстральных коэффициентов (МЧКК) и предложенных методов. Для эксперимента сформирована обучающая выборка из 4000 звуков различных фонем русского языка, из которых 700 соответствуют фонеме [а] и тестовая выборка из 300 звуковых реализаций фонемы [а].

Точность классификации с использованием алгоритмов ВП1, ВП2 и МЧКК составила 60%,82% и 80% соответственно. Для эксперимента по классификации близкорасположенных в признаковом пространстве фонем сформирована обучающая выборка из 1000 звуков гласных фонем и тестовая выборка из 100 звуков фонемы [а]. В данном эксперименте точность классификации с использованием алгоритмов ВП1, ВП2 и МЧКК составила 76%, 92% и 82% соответственно.

Оптимальные параметры алгоритма классификации определены с использованием методов кросспроверки и поиска по сетке. Вектора признаков формировались на основе алгоритма ВП2. Результаты точности классификации фонетической группы и классификации фонемы внутри группы приведены в *табл. 1*.

Таблица 1

Результаты эксперимента по точности классификации фонемы

	[а]	[м]	[н]	[д]
Точность определения группы,%	99	92	93	92
Точность определения фонемы внутри группы,%	90	94	90	94

Также проведено экспериментальное исследование точности классификации фонемы с использованием предложенного метода и одноэтапного распознавания многоклассовым классификатором. Результаты исследования приведены в *табл. 2*.

Таблица 2

Суммарная точность предложенного алгоритма и алгоритма классификации с использованием НС

	[а]	[м]	[н]	[д]
Точность определения группы,%	99	92	93	92
Точность определения фонемы внутри группы,%	90	94	90	94

Анализ результатов данных экспериментов показал, что точность предложенного алгоритма превышает точность одноэтапного алгоритма в среднем на 6%.

Заключение

В статье рассматриваются два алгоритма построения векторов признаков для акустических сигналов на основе вейвлет-преобразования. Использование первого метода (ВП1) не показало практически значимых результатов, что может быть обусловлено некорректным моделированием распределения характерных точек на вейвлет-образе. В то же время использование второго метода (ВП2) показало результаты, превосходящие результаты использования традиционно используемых методов формирования векторов признаков МЧКК на 2% при классификации фонем в общем случае и на 10% при классификации близкорасположенных в признаковом пространстве фонем.

Также в данной статье рассматривается двухэтапный метод классификации фонем русского языка на основе метода опорных векторов. Точность предложенного метода превосходит точность одноэтапного метода в среднем на 6%. Использование данного алгоритма в качестве алгоритма предварительной классификации фонем позволяет уменьшить количество пар спутывания.

Литература

1. Алиев Р.М., Янь Ц., Хейдоров И.Э. Поиск ключевых слов с использованием решётки фрагментов слов // Компьютерная лингвистика и интеллектуальные технологии: Сб. материалов ежегод. междунар. конф. «Диалог 2009», Бекасово, 27–31 мая 2009 г. / Рос. фонд фундам. исслед., Моск. гос. ун-т; Редкол.: А.Е. Кибрик [и др.]. М., 2009. С. 351–354.
2. Vapnik V. The nature of statistical learning theory [M] // New York. Springer-Verlag, 1995.
3. Cortes C., Vapnik V. Support-vector networks // Machine Learning. Vol. 20. № 3. 1995.
4. Шмырев Н.В. Свободные речевые базы данных VoxForge.org // Сб. трудов международной конференции «Диалог 2008». 2008. С. 585–588.
5. Huang X., Acero A. Spoken Language Processing: a guide to theory, algorithm, and system development. New Jersey: Prentice-Hall Inc. Upper Saddle River, 2001.
6. Siafarikas M., Mporas I., Ganchev T., Fakotakis N. Speech Recognition using Wavelet Packet Features // Journal of Wavelet Theory and Applications. 2008. V. 2. № 1. P. 41–59.
7. Rennie J. A short tutorial on using expectation-maximization with mixture models // www.ai.mit.edu/people/jrennie/writing/mixtureEM.pdf, 2004.