АЛГОРИТМ ДВУХЭТАПНОГО РАСПОЗНАВАНИЯ ФОНЕМ РУССКОГО ЯЗЫКА

А. М. Сорока

Белорусский государственный университет, Минск, Беларусь

ВВЕДЕНИЕ

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 выбирается 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. Существует несколько методов разрешения пары спутывания, которые основаны на построении лингвистических решеток и сетей спутывания, а так же использования лингвистических моделей, учитывающих контекстные связи [1]. Данные методы обладают рядом очевидных недостатков, в числе которых высокая трудоемкость алгоритмов и необходимость создания лингвистической модели. В данной статье предлагается алгоритм двухэтапного распознавания фонем на основе метода опорных векторов, который позволяет избежать чрезмерного возникновения пар спутывания за счет более точной классификации отдельно взятой фонемы.

МЕТОД ОПОРНЫХ ВЕКТОРОВ

Метод опорных векторов (машина на опорных векторах, далее - МОВ) впервые был предложен Вапником [2]. Данный метод в процессе обучения непрерывно минимизирует эмпирический риск. Использование Вапником в качестве эвристики выбора разделяющей гиперплоскости предположения о минимизации ожидаемого риска путем максимизации отступов классов привело к высокой обобщающей способности алгоритма. В настоящее время МОВ успешно используется во многих областях.

СЛУЧАЙ ЛИНЕЙНО РАЗДЕЛИМОЙ ВЫБОРКИ

Рассмотрим проблему классификации для выборки, состоящей из m обучающих векторов, которые принадлежат двум разным классам $\{ , y_1),..,(x_m,y_m)\}$, где x_i е R^n - вектор признаков, а y_i е {-1,1} - метка класса, уравнение гиперплоскости $w \bullet x + b = 0$. Существует бесконечное множество возможных разделяющих гиперплоскостей, удовлетворяющих следующим условиям: $y_i(w \bullet x_i) + b > 1$ для Vi e{1,..,m}. Вапник ввел понятие оптимальной разделяющей гиперплоскости. Такой гиперплоскостью считается та, которая максимизирует отступы каждого класса. Описание методов построения оптимальной разделяющей гиперплоскости можно найти в специальной литературе[2].

СЛУЧАЙ ЛИНЕЙНО НЕРАЗДЕЛИМЫХ ВЫБОРОК

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 гиперплоскость определяется вектором w, который минимизирует следующий функционал:

$$(w \bullet x_i) + b > +1 - o, \text{ если } y_i = +1,$$

$$(w \bullet x_i) + b < -1 + o, \text{ если } y_i = -1,$$

где $o = (o_x,...,o_m)$ и C - константы.

ИСПОЛЬЗОВАНИЕ ЯДЕРНЫХ ФУНКЦИЙ

Кроме использования мягкого зазора, существует возможность использования ядерных функций для решения задачи классификации линейно неразделимых выборок. При таком подходе, входные атрибуты обучающей выборки отображаются в многомерное пространство с более высокой размерностью, чем входное, в котором выборка может быть линейно разделена. Данное отображение может быть получено посредством использования ядерных функций.

Наиболее часто используются следующие ядерные функции:

- линейная: $K(x, y) = x \bullet y$,

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

- радиальная базисная Гауссова функция (RBF): $K(x,y) = \exp(-\frac{|x-y|^2}{2\sigma^2})$, где σ - ширина функции Гаусса.

Для заданной ядерной функции классификатор определяется выражением:

$$class(x) = \mathrm{Sign}\left(\sum_{SV} \alpha_i^0 y_i K(x_i \bullet x) + b^0\right).$$

ДВУХЭТАПНЫЙ МЕТОД РАСПОЗНАВАНИЯ ФОНЕМ

В данной статье рассматривается двухэтапный метод распознавания фонем. Данный 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». При этом на первом этапе в обучающей выборке признаковые описания совокупности фонем, относящиеся к данной акустической группе, полагаются положительными прецедентами, а все реализации фонем вне данной группы - отрицательными. На втором этапе производится классификация фонем внутри группы. Многоклассовый классификатор на втором этапе строится по принципу «каждый против каждого», что не влечет за собой увеличения вычислительных затрат в силу малости групп, однако способствует более точному распознаванию.

Разделение фонем на акустически схожие группы определено эмпирическим путем 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. 1.

К*	г-		К"					й	Ш, ы
к	г		к						
		Ч		Щ					э
				ш	*ж*				
т	Д'			с:-	*З'*	Р"	IТ	л-	А
		Ц		с:		р			
т	Д						н	л	о
lГ	Б*			Ф*	в*		м*		у
п	Б			Ф	в		м		

Рис. 1. Группы акустически схожих фонем

ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ

Для определения характеристик разработанного метода было проведено экспериментальное исследование. В ходе экспериментального исследования были проведены эксперименты по определению точности классификации группы акустически схожих фонем, а так же эксперименты по классификации фонемы внутри группы. Для проведения экспериментального исследования была 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 объем базы составил 4500 фонем, в среднем по 100 реализаций каждой фонемы русского языка. Так же была подготовлена тестовая выборка объемом 100 реализаций на каждую из четырех фонем: [а, м, н, д]. Оптимальные параметры классификаторов определены с использованием методов кросспроверки и поиска по сетке. Вектора признаков формировались на основе вейвлет-преобразования. 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

Так же проведено экспериментальное исследование точности классификации фонемы с использованием предложенного метода и одноэтапного распознавания многоклассовым классификатором. Результаты данного экспериментального исследование приведены в табл. 2.

Таблица 2

Суммарная точность предложенного алгоритма и алгоритма классификации с использованием НС

	[а]	[м]	[н]	[д]
Суммарная точность предложенного алгоритма,%	89	85	83	85
Точность одноэтапного распознавания,%	85	79	76	77

Таким образом, точность классификации разработанного алгоритма превышает точность одноэтапного алгоритма в среднем на 6%.

ЗАКЛЮЧЕНИЕ

В данной статье рассмотрен двухэтапный метод классификации фонем русского языка на основе метода опорных векторов. Точность классификации предложенного метода превосходит точность одноэтапного метода в среднем на 6%. Использование данного алгоритма в качестве алгоритма предварительной классификации фонем позволяет уменьшить количество пар спутывания, требующих дальнейшего анализа с использованием затратных методов на основе решеток слогов и сетей спутывания, а, следовательно, увеличить итоговую скорость работы системы распознавания слитной речи.

Литература

1. *Алиев, Р. М.* Поиск ключевых слов с использованием решетки фрагментов слов / Р. М. Алиев, Янь Цзинбинь, И. Э. Хейдоров // 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] // NY. : Springer-Verlag, 1995.
3. *Cortes, C.* Support-vector networks/ C. Cortes, V. Vapnik // Machine Learning, 1995. Vol. 20. No. 3.
4. *Шмырев, Н. В.* Свободные речевые базы данных VoxForge.org // Сборник трудов международной конференции «Диалог 2008». 2008. С. 585-588.