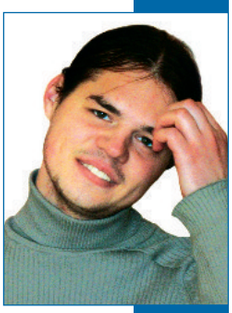
Классификация эмоционально окрашенной речи с использованием метода опорных векторов



И.Э. Хейдоров,
кандидат физико-математических наук

Янь Цзинбинь

У Ши

А.М. Сорока

А.А. Трус

В данной статье приводится описание модели классификации эмоционально окрашенной речи с использованием метода опорных векторов. В качестве векторов признаков используются частота основного тона и данные, извлечённые из акустической записи речи с помощью тигеровского оператора энергии. При использовании данных признаков экспоненциальной радиальной базисной ядерной функции в методе опорных векторов точность классификации представленным методом составляет до 96,2%.

1. Введение

На современном этапе развития информационных технологий разработка методов автоматического определения эмоционального состояния человека по голосу является актуальной задачей, позволяющей решить ряд экономических, социальных и бытовых проблем и, кроме того, играющей важную роль в вопросах безопасности. Эмоциональный речевой сканер может найти широкое применение в различных транспортных и диспетчерских учреждениях, для ограничения или полного запрета доступа к выполнению служебных обязанностей лиц, находящихся в неустойчивом или

неадекватном эмоциональном состоянии. Подобные системы контроля позволят также проводить дополнительную проверку пассажиров авиарейсов в рамках мероприятий по противодействию терроризму.

В основе предлагаемого подхода лежит предположение о том, что под воздействием различных внешних и внутренних факторов у человека, во-первых, происходит изменение формы и геометрических параметров голосового тракта; во-вторых, изменяется форма импульсов возбуждения органов речеобразования. Совокупность этих факторов находит отражение в речи, произносимой тем или иным диктором, с помощью эмоциональной окраски. Традиционные статистические байесовские методы и скрытые марковские модели (далее — СММ) не позволяют в полной мере учесть все эти изменения ввиду того, что качество полученной модели классификации сильно зависит от объёма обучающей выборки. Для получения высокой точности классификации при использовании традиционных подходов необходимы обучающие выборки значительного объёма, который сложно получить в реальных условиях для эмоционально окрашенной речи для конкретных дикторов.

В связи с вышесказанным существует необходимость в разработке метода классификации речи по эмоциям, позволяющего получить высокую точность при обучении на выборках ограниченного объёма. Использование метода опорных векторов (МОВ) для построения классификатора эмоциональной речи позволяет в определённой степени преодолеть ограничение в объёме обучающих данных и обеспечить хорошую разрешающую способность путём прямой аппроксимации межклассовых границ [1, 2].

В данной статье рассмотрены проблемы классификации эмоционально окрашенной речи**,** извлечения векторов признаков, предварительной обработки обучающих выборок, выбора параметров алгоритма и оценки свойств полученного классификатора на основе МОВ.

2. Метод опорных векторов

Метод опорных векторов (именуемый также «машиной на опорных векторах») впервые был предложен Вапником [3]. Данный 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 [4–6].

2.1. Случай линейно разделимых выборок

Рассмотрим проблему разделения гиперплоскостью $w \cdot x + b = 0$ выборки, состоящей из m обучающих векторов, принадлежащих двум разным классам $\{(x_1, y_1),...,(x_m, y_m)\}$, где $x_i \in R^n$ — вектор признаков, а $y_i \in \{-1,1\}$ — метка класса. Существует бесконечное множество возможных разделяющих гиперплоскостей, удовлетворяющих следующим условиям: $y_i(w \cdot x_i) + b \geq 1 \quad \forall i \in \{1,..,m\}$. Вапник ввёл понятие оптимальной разделяющей гиперплоскости (оптимального классификатора). Такой гиперплоскостью считается та, которая максимизирует расстояние от неё до каждого класса (рис.1).

Таким образом, гиперплоскость, минимизирующая выражение $\Phi(w) = \frac{w^2}{2}$, является оптимальной разделяющей гиперплоскостью. Следовательно, сутью метода является решение оптимизационной задачи. Описание методов решения подобных задач можно найти в специальной литературе [9–13].

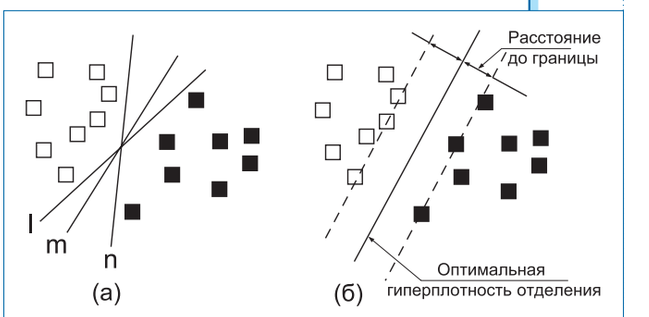


Рис. 1. Классификация двух классов с использованием гиперплоскостей: (а) Произвольные гиперплоскости l, m и n; (б) Оптимальная гиперплоскость отделения с наибольшим краем, отделённым штрихованными линиями, проходящими через опорные вектора

2.2. Случай линейно неразделимых выборок

Для решения проблемы классификации линейно неразделимых выборок Кортес и Вапник [13] предложили метод опорных векторов с мягким зазором. Фактически они вводят неотрицательную величину ошибки классификации. Теперь проблема оптимизации представляет задачу минимизации ошибки классификации [14]. В таком случае оптимальная разделяющая гиперплоскость определяется вектором *w*, который минимизирует следующий функционал:

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

где $\xi = (\xi_1,...,\xi_m)$ и C — константы. Описание методов решения данной задачи может быть найдено в специальной литературе [12].

2.3. Ядерные функции

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: $K(x,y) = x \cdot y$;
— полиномиальная: $K(x,y) = (x \cdot y + 1)^d$, где d — степень полинома;
— радиальная базисная Гауссова функция (RBF): $K(x, y) = \exp(-\frac{|x-y|^2}{2\delta^2})$,
где δ — ширина функции Гаусса.

Для заданной ядерной функции классификатор определяется выражением:

$$class(x) = Sign(\sum_{SV} \alpha_i^0 y_i K(x_i \cdot x) + b^0) \; . \qquad (2)$$

3. Классификация эмоционально окрашенной речи с использованием МОВ

Задача классификации эмоций по речевому сигналу характеризуется малым числом обучающих выборок, что не позволяет в полной мере использовать традиционные статистические методы. В связи с этим для данной цели было предложено создать классификатор на основе МОВ, для построения которого необходимо решить несколько задач.

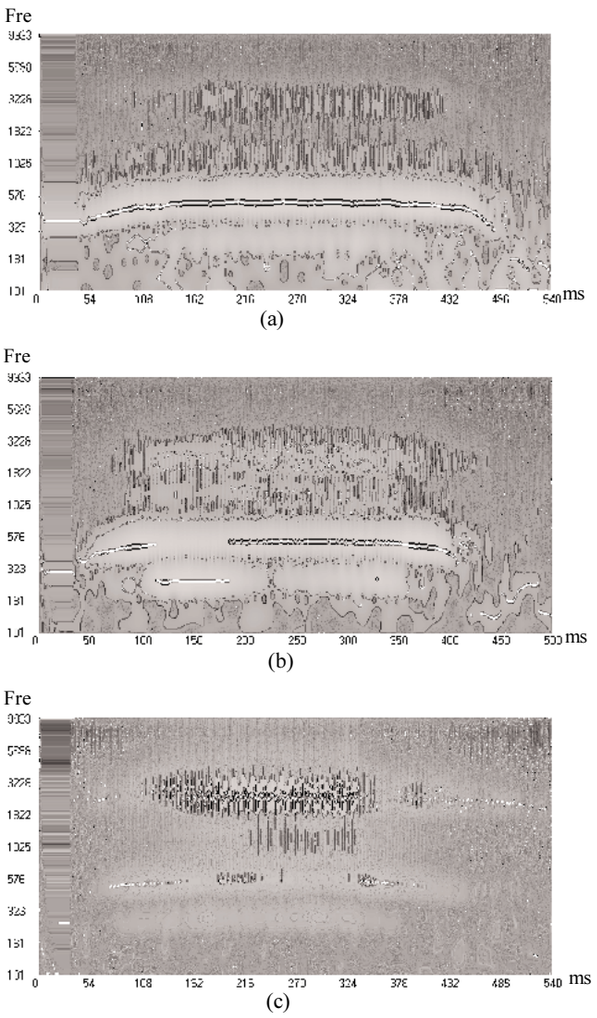


Рис 2. ЧОТ при различных акустических воздействиях на диктора

3.1. Построение вектора признаков

Выбор признаков, способных максимально различить нормальную и эмоционально окрашенную речь, непосредственно влияет на качество классификации. Частота основного тона (далее — ЧОТ) однозначно зависит от периода для гласных звуков, следовательно, ЧОТ может отразить акустические характеристики гортани. В состоянии эмоционального возбуждения форма гортани деформируется и, как следствие, изменяется ЧОТ. Данное свойство позволяет предположить повышение точности классификации при использовании ЧОТ в качестве одного из признаков. В эксперименте использовались данные эмоционально окрашенной речи, полученной при акустическом воздействии на диктора.

Тигер предложил нелинейную модель анализа речи. Для анализа вычисляется тигеровский оператор энергии (далее — ТОЭ). ТОЭ в сравнении с ЧОТ позволяет более детально представить акустические характеристики речи диктора [8].

В статье приведено сравнение обоих описанных выше признаков и их свойств.

На рисунке 2 показаны спектрограммы одного диктора при воздействии акустическими сигналами разной эмоциональной окраски. На рисунке 2a приведена спектрограмма нормальной речи диктора. На рисунке 2b приведена спектрограмма речи диктора под воздействием акустического сигнала, представляющего собой инструментально богатую музыкальную композицию в стиле панк-рок. На рисунке 2c показана речь диктора под воздействием акустического сигнала, представляющего собой серию спонтанных звуковых эффектов (а именно — записей взрывов) с малой скважностью. Белым цветом показана ЧОТ. Из рисунка видно, что ЧОТ изменяется при изменении воздействия.

3.2. Проблема предварительной обработки обучающих выборок

В связи с использованием скалярного произведения МОВ необходимо, чтобы все извлекаемые из сигнала вектора признаков принадлежали одному пространству. Речевые сигналы при этом анализируются при помощи не стандартной методики «кадр за кадром», а целиком, как единый массив, т.е. анализируется один вектор признаков для всего массива. Очевидно, что при этом размерность каждого вектора признаков зависит от длительности исследуемого сигнала. При предварительной обработке обучающих выборок должна быть решена проблема использования векторов из разных признаковых пространств. С учётом специфики ЧОТ используются два подхода.

1. Самая короткая выборка делается базовой, все остальные выборки приводятся к такой же длине.

2. Время для каждой выборки масштабируется таким образом, чтобы получить вектора признаков одинаковой размерности для всех выборок. При проведении экспериментов для решения данной задачи была использована методика динамического программирования (ДП).

3.3. Проблема настройки параметров МОВ

При обучении классификатора с использованием МОВ настройка параметров включает следующие этапы: выбор ядра, выбор штрафа алгоритма C, выбор параметров ядра. Настройка параметров влияет на скорость обучения и способность обобщения классификатора. Оптимальные параметры выбираются из условия минимизации следующего выражения:

$$\Phi = \frac{R^2 \left\| \mathrm{V}n \right\|^2}{n} \qquad (3)$$

где R — минимальный радиус гиперсферы, охватывающей обучающие выборки в пространстве признаков, V_n — вектор нормали к разделяющей гиперплоскости, n — количество обучающих выборок.

В данной статье для линейного МОВ эмпирическим путём был выбран штраф $\mathrm{C} = 0{,}005$. Для нелинейного МОВ выбраны радиальное и экспоненциальное радиальное базисные ядра. Соответственно были выбраны следующие параметры ядер: $\delta = 200$ и $\delta = 15$. Штраф алгоритма в данном случае был зафиксирован на значении $\mathrm{C} = 10$.

4. Результаты эксперимента

Авторами статьи был проведён следующий эксперимент. Дикторы мужского пола (группа 1 и группа 2) произносили цифры от 0 до 10 при воздействии через персональные наушники акустическими сигналами разных типов. В качестве акустических сигналов для эмоционального воздействия на диктора использовались музыкальные композиции в стиле панк-рок (воздействие 1), акустические записи взрывов, следующие спонтанно с малой скважностью (воздействие 2), акустические записи мужских и женских криков относительно высокой громкости, следующих спонтанно с большой скважностью (воздействие 3). В нормальных условиях записи для каждого диктора были произведены по 30 раз. При эмоциональном воздействии акустическими сигналами записи были произведены по 50 раз. Также были сделаны записи речи дикторов, которые подверглись эмоциональному воздействию и физической нагрузке. Здесь дикторы включают мужчин (группа 3) и женщин (группа 4). Из записей извлекались признаки ЧОТ и ТОЭ. В экспериментах использовались вектора признаков с размерностью 30–50.

4.1. Дикторозависимый эксперимент

В эксперименте в качестве обучающих использовались 10 записей нормальной речи и 10 записей эмоционально окрашенной речи. Для каждой группы дикторов были получены 220 обучающих и 20 тестовых выборок. Результаты эксперимента представлены в таблице 1.

Таблица 1

Результаты дикторозависимой классификации

Функция ядра	Группа 1, точность классификации %			Группа 2, точность классификации %		
	ЧОТ	ЧОТ, ТОЭ	ТОЭ, нормализация ДП	ЧОТ	ЧОТ, ТОЭ	ТОЭ, нормализация ДП
Linear	75.5	87.5	87.2	93.3	96.2	92.1
Polynomial	81.1	87.1	86.8	95.0	96.2	92.5
RBF	86.9	88.3	87.8	95.2	94.6	92.7
ERBF	92.2	91.6	94.4	95.4	96.2	93.3
Байесовский информационный критерий (BIC)	70.0	73.3	—	88.3	96.2	—
Скрытая марковская модель (СММ)	82.1	87.6	—	89.4	95.0	—

По сравнению с байесовским классификатором и классификатором на основе СММ, средняя точность классификации дикторов из группы 1 увеличилась на 22,2% и 10,1% соответственно, а средняя точность классификации дикторов из группы 2 увеличилась на 7,1% и на 6,0% соответственно.

4.2. Диктороннезависимый эксперимент

Обучение производилось на данных дикторов из группы 2, для тестирования использовались данные дикторов из групп 1, 2, 3. В эксперименте использовались признаки ТОЭ. Результаты эксперимента представлены в таблице 2.

Таблица 2

Результат диктороннезависимой классификации

Функция ядра	Точность классификации, %		
	Группа 1	Группа 2	Группа 3
Linear	70.4	82.1	68.9
ERBF	71.2	94.6	75.9

4.3. Эксперимент со смешанным обучением

В данном эксперименте для обучения использовались 240 выборок для дикторов из групп 1 и 2. Данные дикторов из групп 3 и 4 и неиспользованные в процессе обучения данные дикторов из групп 1 и 2 представляли собой тестовую выборку. В качестве векторов признаков использовались признаки ТОЭ. Результаты эксперимента представлены в таблице 3.

Таблица 3

Результат классификации при смешанном обучении

Функция ядра	Точность классификации, %			
	Группа 1	Группа 2	Группа 3	Группа 4
Linear	76.7	90.9	81.8	62.4
ERBF	77.0	91.1	96.9	67.5

Анализ результатов показывает увеличение точности классификации дикторов из группы 1 по сравнению с предыдущим экспериментами. Данный эффект связан с тем, что данные дикторов из группы 1 входили в обучающую выборку. Кроме этого, внесение данных дикторов из группы 1 в обучающую выборку увеличило точность классификации дикторов из групп 3 и 4.

Следовательно, для получения универсального классификатора эмоционально окрашенной речи необходимо включать в обучающую выборку данные для различных возрастных и социальных групп дикторов, что позволит учесть возраст и пол диктора и другие факторы.

4.4. Эксперимент с использованием ограниченных выборок

В данном эксперименте использовались вектора признаков, полученные с использованием ТОЭ. Количество обучающих данных последовательно уменьшалось. Обучающие выборки для каждого эксперимента формировались случайно, при этом половина выборки соответствовала нормальной речи, половина — эмоционально окрашенной. Таблица 4 представляет усреднённые результаты экспериментов по 12 группам.

Таблица 4

Результат классификации на ограниченных выборках

Функция ядра	Точность классификации при разных объёмах выборок, %					
	220	180	140	100	60	20
Linear	96.2	92.6	94.0	93.3	94.5	91.3
ERBF	96.2	95.8	94.0	94.1	94.5	91.3

Анализ полученных данных показывает, что точность классификации уменьшается при уменьшении объёма обучающей выборки с сохранением высокого абсолютного значения. Следовательно, существует возможность использования МОВ для обучения на ограниченных выборках.

5. Выводы

Применение метода опорных векторов для решения задач классификации эмоционально окрашенной речи позволяет получить более высокую точность обученной модели, в сравнении с традиционными статистическими методами классификации.

Влияние методов извлечения векторов признаков на точность классификации обученной модели позволяет предположить, что модернизация данных методов является одним из путей дальнейшего увеличения точности рассмотренного в статье классификатора.

Задача выбора оптимальных параметров алгоритма и ядерной функции требует дальнейшего исследования.

Литература

1. Vapnik V., Chervoknenkis A.Y. On The uniform convergence of relative frequencies of events to their probability [J]. Theory Probabilty and Its Applications, 1971, 17(2):164–280.
2. Vapnik V. Estimiation of dependencies based on empirical data[M]. New York, Springer-Verlag, 1982.
3. Vapnik V. The nature of statistical learning theory [M]. New York. Springer-Verlag, 1995.
4. Niyogi P., Burges C., Ramesh P. Distinctive feature detection using support vector machines[A]. Proceedings of 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. 1999.
C. Moreno Pedro J., Clarkson Philip. On the use of support vector machines for phonetic classification [A]. Proceedings of 1999 IEEE International Conference on Acoustics, Speech and Signal Processing [C]. 1999.
6. Fine S., Navratil J.,Gopinath R.A. A hybrid GMM/SVM approach to speaker identification [A]. Proceedings of 2001 IEEE International Conference on Acoustics, Speech and Signal Processing [C]. 2001.
7. Zhou G., Hansen J. H. L., Kaiser J. F. Linear and nonlinear feature analysis for stress classification [A]. Proceedings of ICELP98 [C]. 1998.
8. Zhou G., Hansen J.H.L., Kaiser J.F. Classification of Speech Under Stress Based on Features Derived form The Nonlinear Teager Energy Operator [A]. Proceedings of 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. 1998.
9. Cortes C., Vapnik V. Support Vector Networks [J]. Machine Learning, 1995, 20:273–297.
10. Joachims T. Marking large-scale SVM learning practical [A]. Scholkopf B., Burges C.J.C., Smola Aeds. Advances in Kernel Methods-Support Vector Learning [C]. Cambridge, MA:MIT Press, 1998.
11. Platt J.C. Fast training of SVMs using sequential minimal optimization [A]. Scholkopf B., Burges C.J.C., Smola A.J. eds. Advances in Kernel Methods-Support Vector Learning [C]. Cambridge, MA:MIT Press, 1998.
12. Burges C. A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery, vol.2, no.2, 1998.
13. Cortes C. and Vapnik V. Support-vector networks, Machine Learning, vol.20, no.3, 1995.
14. Vapnik V. Staristical Lreaming Theory, John wiley and Sons, 1998.

Хейдоров Игорь Эдуардович —

Окончил с отличием БГУ (Минск) в 1996 году, с 1998 года работает на кафедре радиофизики БГУ, кандидат физико-математических наук (2000 г), доцент. Сфера научных интересов — методы и алгоритмы распознавания и синтеза речи, автоматическая индексация аудиодокументов. Автор 40 работ.

Сорока Александр Михайлович —

студент 5-го курса БГУ, научные интересы — методы и алгоритмы обработки цифровых сигналов, теория метода опорных векторов, смешанные гауссовы модели.

Трус Александр Александрович —

студент 5-го курса БГУ, научные интересы — методы и алгоритмы обработки речевых сигналов, скрытые марковские модели, теория метода опорных векторов.

У Ши —

аспирант БГУ, научные интересы — методы и алгоритмы обработки речевых сигналов, теория метода опорных векторов, распознавание болезней голосового тракта по голосу.

Ян Цзинбинь —

аспирант БГУ, научные интересы — методы и алгоритмы обработки речевых сигналов, теория метода опорных векторов, алгоритмы обнаружения ключевых слов в потоке речи.