УДК 004.934.2

А. В. Ткаченя, А. Г. Давыдов, В. В. Киселёв, М. В. Хитров

КЛАССИФИКАЦИЯ ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ ДИКТОРА С ИСПОЛЬЗОВАНИЕМ МЕТОДА ОПОРНЫХ ВЕКТОРОВ И КРИТЕРИЯ ДЖИНИ

Исследована эффективность применения критерия Джини для формирования пространства признаков SVM-классификатора. Приведены результаты экспериментального определения оптимального набора информативных признаков и построения классификатора.

Ключевые слова: *речь, классификация эмоционального состояния, критерий Джини, метод опорных векторов, автоматический выбор информативных признаков.*

Введение. Исследование паралингвистических средств речевой коммуникации включает определение довольно разнообразных характеристик: эмоциональное состояние, пол и возраст диктора, стиль разговора, уровень заинтересованности, сонливость и даже наличие алкогольного опьянения.

В настоящей работе исследуется задача определения эмоционального состояния говорящего человека (диктора). При решении этой задачи возникает ряд трудностей [1]: отсутствует четкое определение эмоции, отсутствует однозначный ответ на вопрос о соотнесении акустических особенностей речи диктора с его эмоциональным состоянием. Все это приводит к различиям в формах классификации эмоций и произвольной расстановке акцентов разными группами исследователей [2].

В современных системах определения эмоционального состояния диктора можно выделить следующие основные этапы обработки [3, 4]:

1) вычисление базовых характеристик речевого сигнала (low-level descriptors, согласно терминологии [4]); оценка мощности, частоты основного тона F_0 (ЧОТ), формантных частот, спектральных и кепстральных характеристик речевого сигнала и т.д.;

2) вычисление функционалов от базовых характеристик, таких как перцентили, экстремумы и их отношения, моменты высших порядков, коэффициенты регрессии и т.д.;

3) классификация объектов. Наибольшее распространение в последнее время получили классификаторы на основе смеси нормальных распределений и метода опорных векторов [5].

В настоящей работе предложено использовать статистический критерий, отражающий сходство видов распределений исследуемой характеристики при решении задачи классификации эмоциональных состояний.

Описание базы тестирования. Обучение и тестирование алгоритма проводилось на записях, взятых из Берлинской базы данных эмоциональной речи (Emo-DB) [6]. Данная база была собрана в Техническом университете Берлина и неоднократно использовалась исследователями при разработке систем распознавания эмоционального состояния. Исследование базы показало [6], что эмоции в ней распознаются слушателями в 80 % случаев, и в 60 % признаются естественными.

Методология классификации эмоционального состояния диктора. Обобщенная структурная схема системы определения эмоционального состояния диктора приведена на рис. 1.

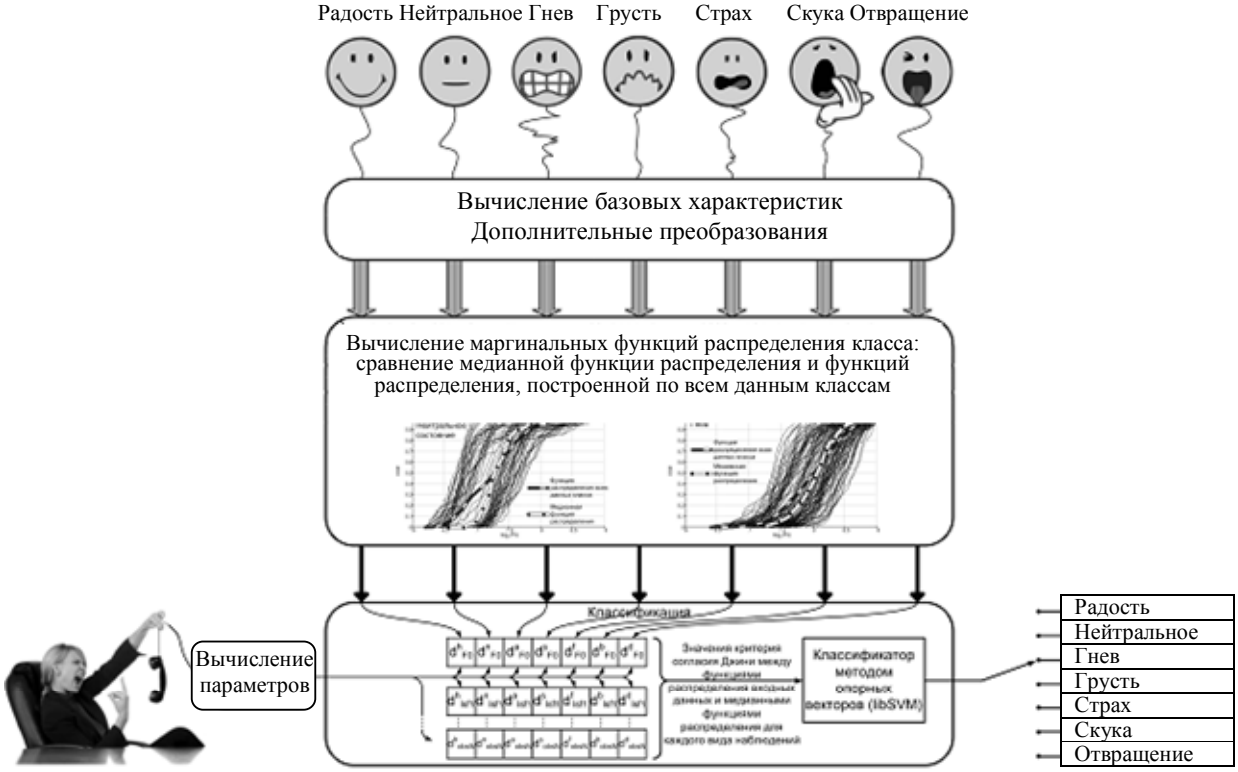

Рис. 1

Предобработка заключается в умножении каждой записи на случайный коэффициент усиления от –20 до +20 дБ, чтобы исключить привязку к абсолютному уровню сигнала, предыскажению и пропусканию сигнала через полосовой фильтр с полосой пропускания от 300 до 3400 Гц.

Вычисление параметров. Исследования в области психологии и психолингвистики предоставили сведения о множестве акустических, просодических и лингвистических характеристик речи, способных служить информативными признаками при распознавании эмоционального состояния и проявляющихся на уровне голосовых сегментов, слогов и целых слов. При этом во множестве видов характеристик выделяют базовые и вычисленные из них функционалы. Набор базовых характеристик в нашем случае включает: кратковременную оценку мощности сигнала; оценку частоты основного тона в соответствии с алгоритмом, рассмотренным в [7]; джиттер (модуляция частоты основного тона) и шиммер (модуляция амплитуды сигнала); коэффициенты линейных спектральных частот; кепстральные коэффициенты, вычисленные на основе коэффициентов линейного предсказания; фонетическую функцию на основе вычисления лог-спектрального расстояния, расстояния Итакуры-Саито и COSH-расстояния [8]; коэффициенты вещественного кепстра; мел-кепстральные коэффициенты; оценки асимметрии и эксцесса распределения ошибки линейного предсказания сигнала [9]; энергетический оператор Тигера в формантных полосах и критических полосах слуха [10]; отношения мощностей в формантных полосах.

К вычисленным базовым признакам применялся ряд преобразований: вычисление первой и второй производных, применение энергетического оператора Тигера, вычитание медианного значения, стандартизация.

Вычисление статистического критерия. Наиболее простым решением является вычисление статистических критериев (расстояния) между двумя многомерными плотностями распределения. Однако для построения многомерных плотностей распределения требуется большое количество обучающих данных. В противном случае классификатор может оказаться неустойчивым. Поэтому было решено вместо многомерных использовать маргинальные распределения каждой исчисленной характеристики сигнала.

Выбор критерия вычисления расстояния между двумя функциями распределения играет важную роль. Для построения пространства расстояний необходимо использовать функцию расстояния, не требующую априорных предположений о видах распределения сравниваемых величин. Поэтому в качестве функции расстояния было решено использовать критерий Джини [11]:

$$d(n,m) = \int \left| F_n(x) - F_m(x) \right| dx .$$

Таким образом, для каждого исследуемого признака обрабатываемого файла необходимо вычислить критерий Джини между распределением этого признака для анализируемого сигнала и распределением, описывающим каждый класс эмоциональной речи. Эти расстояния целесообразно использовать как пространство наблюдений для обучения и тестирования классификатора. При этом для описания функций распределения класса подходящим представляется использование функции распределения всех данных этого класса. Однако при этом способе формирования функции не учитывается поведение каждой эмпирической функции распределения, входящей в этот класс, и таким образом теряется информация совокупности эмпирических функций распределения класса как семейства некоторых кривых.

Для преодоления этого недостатка функцию распределения класса целесообразно определить как медианное значение всех функций распределения, входящих в класс:

$$\tilde{F}(x) = \underset{i}{\text{median}} \left(F_i(x) \right).$$

При этом определение медианной функции распределения следует выполнять по равномерно расположенным квантилям всех наблюдений класса.

На рис. 2 приведен пример различных способов формирования функции распределения класса: как функции распределения всех наблюдений (*1*, Class CDF) и как медианной функции распределения (*2*, Median CDF).

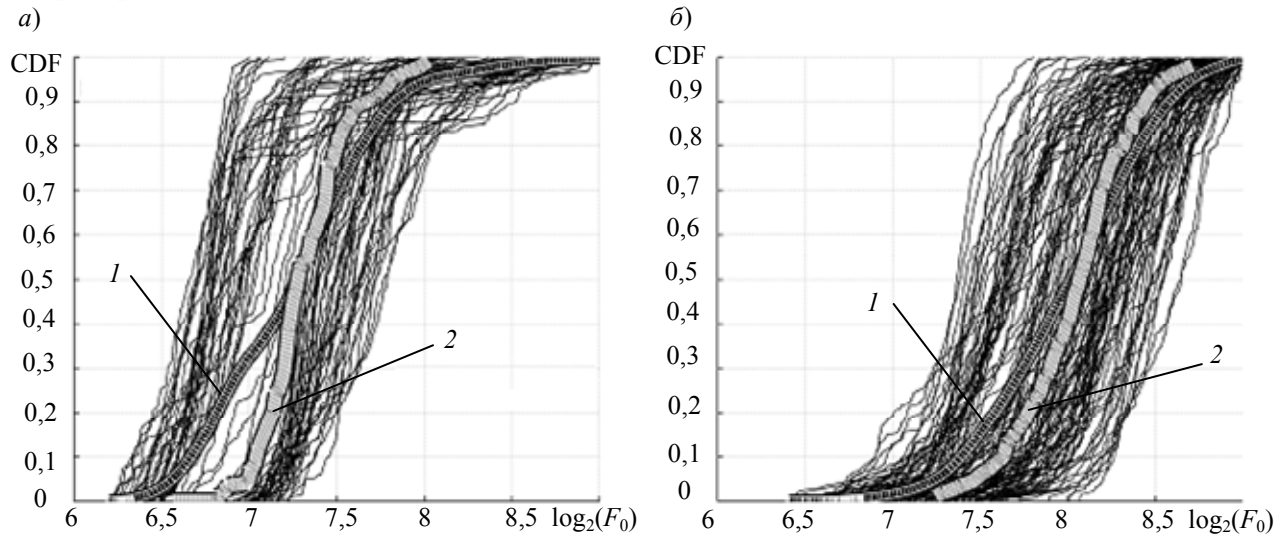

Рис. 2

На рис. 2, *а* (нейтральное эмоциональное состояние) хорошо заметно наличие двух областей группирования функций распределения, связанных с различием распределений ЧОТ мужских и женских голосов (график на рис. 2, *б* соответствует состоянию гнева). При этом

следует отметить, что Median CDF, в отличие от Class CDF, в значительной степени сохраняет форму поведения кривых функций распределения класса.

Классификация объектов в сформированном пространстве признаков является завершающей операцией большинства систем определения эмоционального состояния диктора.

Для выполнения классификации был использован метод опорных векторов (SVM-метод), реализованный в библиотеке libSVM [12]. В данной библиотеке мультиклассовый SVM-классификатор строится как набор классификаторов „каждый-с-каждым“ с последующим голосованием. Это позволяет на этапе определения оптимального набора признаков выбрать лучший набор именно для мультиклассовой классификации. Для построения разделяющей гиперповерхности использовалось RBF-ядро [12] как наиболее универсальное и не требующее априорных предположений о характере распределения наблюдений. Эффективность распознавания в процессе определения оптимального набора информативных признаков и подбора параметров модели оценивалась при помощи метода K-кратной кросспроверки, реализованного в составе пакета libSVM.

Эксперимент. Рассмотрим результаты экспериментального исследования описанного способа формирования пространства признаков для построения классификатора. Как указывалось выше, экспериментальные исследования проводились с использованием базы Emo-DB. При этом для оценки эффективности классификации данных использовалось среднее значение диагональных элементов нормированной матрицы неточностей (average recall).

Автоматическое определение оптимального набора информативных параметров. Для определения этого набора использовался алгоритм последовательного выбора параметров (Sequential Feature Selection, SFS), нашедший широкое применение при решении задач распознавания эмоциональных состояний по голосу [3]. Суть его работы заключается в том, что на каждой итерации к набору добавляются признаки, обеспечивающие наибольший прирост эффективности классификации. Чтобы увеличить гибкость алгоритма, на каждой его итерации после добавления некоего, обеспечивающего максимальный прирост эффективности классификации, множества из m признаков, производится удаление множества из n признаков. Нами использовались значения m=5 и n=3.

Зависимость эффективности распознавания P эмоциональных состояний от числа добавленных в набор алгоритмом SFS информативных признаков N показана на рис. 3.

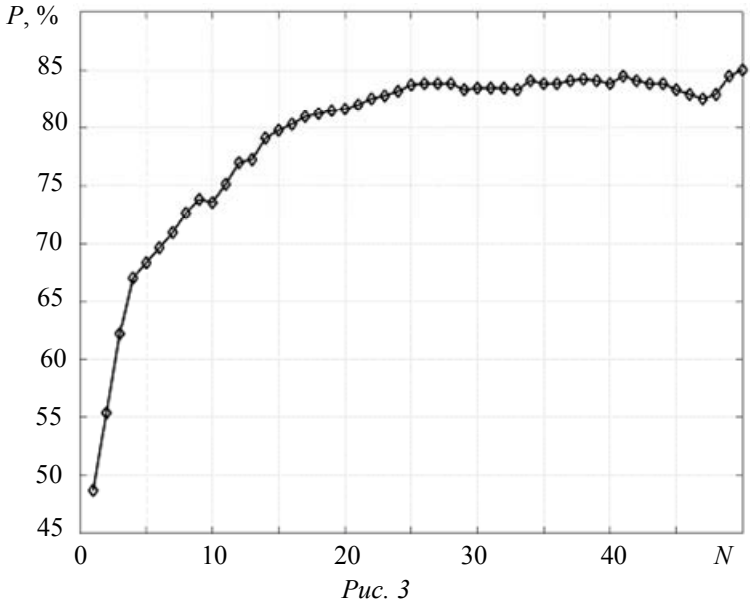

Рис. 3

Из рисунка видно, что на каждой итерации эффективность распознавания эмоциональных классов возрастала до тех пор, пока не достигла некоторого максимума, соответствую-

щего примерно N=25. Дальнейшее повышение эффективности распознавания с ростом количества информативных признаков происходило очень медленно.

Классификация. Оценка эффективности классификации проводилась при помощи K-кратной кросспроверки, при этом K = 10 [13]. В таблице приведена усредненная матрица неточностей, полученная для 25 информативных признаков. Средняя эффективность, достигнутая построенным классификатором, оказалась ≈ 83 %.

Фактический класс	Предсказанный класс						
	гнев	скука	отвращение	страх	радость	нейтральное	грусть
Гнев	**91,7**	0	0	1,9	6,4	0	0
Скука	0	**90,3**	0	0	0	6,6	3,1
Отвращение	1,8	6,7	**68,6**	6,3	4,1	3,9	8,6
Страх	5,5	2,0	1,4	**74,6**	7,9	2,1	6,4
Радость	20,2	0	0	6,3	**73,5**	0	0
Нейтральное	0	9,3	0	0	0	**88,3**	2,5
Грусть	0	4,0	0	0	0	3,3	**92,7**

Выводы и направления дальнейших исследований. В статье предложен новый метод формирования пространства признаков классификатора на основе вычисления критерия Джини. Было проведено экспериментальное исследование эффективности метода, включающее этапы определения оптимального набора параметров и построения SVM-классификатора. Экспериментальное исследование проводилось на базе Emo-DB [6]. В качестве показателя эффективности использовалось среднее значение диагональных элементов нормированной матрицы неточностей, а для оценки точности прогнозирования — метод K-кратной кросспроверки.

В качестве дальнейшей работы представляется целесообразным протестировать эффективность применения описанного метода для классификации других паралингвистических средств речевой коммуникации.

СПИСОК ЛИТЕРАТУРЫ

1. *El Ayadi M., Kamel M.S., Karray F.* Survey on speech emotion recognition: Features, classification schemes, and databases // Pattern Recognition. 2011. Vol. 44, N 3. P. 572—587.

2. *Cornelius R. R* The science of emotion: research and tradition in the psychology of emotions. NJ: Prentice-Hall, Upper Saddle River, 1996.

3. *Schuller B., Batliner A., Steidl S., Seppi D.* Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge // Speech Communication. 2011. Vol. 53, N 9—10. P. 1062—1087.

4. *Eyben F., Wöllmer M., Schuller B.* OpenEAR-Introducing the Munich open-source emotion and affect recognition toolkit // Proc. 3rd Intern. Conf. on Affective Computing and Intelligent Interaction. ACII. 2009. P. 1—6.

5. *Bone D., Black M., Ming Li, Metallinou A., Sungbok Lee, Narayanan S.S.* Intoxicated Speech Detection by Fusion of Speaker Normalized Hierarchical Features and GMM Supervectors // Proc. Interspeech. Florence, Italy, 2011. P. 3217—3220.

6. *Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W., Weiss B.* A Database of German Emotional Speech // Proc. Interspeech. Lisbon, 2005. P. 1517—1520.

7. *Talkin D.* A robust algorithm for pitch tracking (RAPT) // Speech coding and synthesis. Elsevier Science, 1995. P. 495—518.

8. *Rabiner L. R., Biing-Hwang Juang.* Fundamentals of speech recognition. Prentice Hall, 1993. 507 p.

9. *Nemer E., Goubran R., Mahmoud S.* Robust Voice Activity Detection Using Higher-Order Statistics in the LPC Residual Domain // IEEE Transactions on Speech and Audio Processing. 2001. Vol. 9, N 3. P. 217—231.

10. *Rahurkar M., Hansen J. H. L., Meyerhoff J., Saviolakis G., Koenig M.* Frequency Band Analysis for Stress Detection Using a Teager Energy Operator Based Feature // Proc. Intern. Conf. on Spoken Language Processing ICSLP-2002. Denver, CO USA, 2002. Vol. 3. P. 2021—2024.

11. *Кобзарь А. И.* Прикладная математическая статистика. М.: ФИЗМАТЛИТ, 2006. 816 с.

12. *Chang C.-C., Lin C.-J.* LIBSVM: a library for support vector machines // ACM Transactions on Intelligent Systems and Technology. 2011. Vol. 2, N 27. P. 1—27.

13. *Kohavi R.* A study of cross-validation and bootstrap for accuracy estimation and model selection // Proc. of the 14th Intern. Joint Conf. on Artificial Intelligence. 1995. Vol. 2. P. 1137—1143.

Сведения об авторах

Андрей Владимирович Ткаченя — ООО „Речевые технологии“, Минск; младший научный сотрудник; E-mail: tkachenia-a@speechpro.com

Андрей Геннадьевич Давыдов — канд. техн. наук; ООО „Речевые технологии“, Минск; старший научный сотрудник; E-mail: davydov-a@speechpro.com

Виталий Владимирович Киселёв — ООО „Речевые технологии“, Минск; директор; E-mail: kiselev-v@speechpro.com

Михаил Васильевич Хитров — канд. техн. наук; ООО „ЦРТ“, Санкт-Петербург; генеральный директор; Санкт-Петербургский национальный исследовательский университет информационных технологий, кафедра речевых информационных систем; зав. кафедрой; E-mail: khitrov@speechpro.com

Рекомендована кафедрой речевых информационных систем

Поступила в редакцию 22.10.12 г.