

УДК 004.934.2

Д.В. Пекарь, С.Г. Тихоненко, А.Ф. Чернявский, В.С. Садов

АЛГОРИТМ ОПТИМИЗАЦИИ ВХОДНЫХ ПРИЗНАКОВ И ПАРАМЕТРОВ КЛАССИФИКАТОРА ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ ЧЕЛОВЕКА ПО ЕГО РЕЧИ

Рассматривается алгоритм оптимизации входного вектора признаков и параметров классификатора эмоционального состояния человека, основанный на использовании стохастического метода роя частиц. Предлагаются пути его совершенствования с целью повышения скорости сходимости алгоритма. Доказывается улучшение точности классификации эмоционального состояния диктора по его речи при использовании оптимизированного вектора признаков и найденных параметров классификатора.

Введение

Значительная информационная составляющая, содержащаяся в эмоциональной окраске речи человека, представляет огромный интерес как для научного, так и для практического изучения. Возможность анализа и интерпретации невербальной информации позволяет создавать более совершенные речевые аналитические и экспертные системы, а также системы безопасности на основе использования речевого информационного потока. В представленной работе сравниваются возможности различных классификаторов эмоционального состояния человека и предлагается алгоритм оптимизации входного вектора признаков.

1. Признаки речевого сигнала и база данных речевых фрагментов

Эмоциональное состояние говорящего человека определенным образом влияет на его голосовой тракт, что в конечном счете сказывается на физических параметрах произносимой им речи [1, 2]. Классификация эмоциональной окраски речи предполагает формирование вектора признаков, эффективного с вычислительной точки зрения и репрезентативного в плане дифференцирования различных типов эмоций.

Для вычисления различных характеристик речевого сигнала применялся инструментальный openEAR [3], разработанный Мюнхенским техническим университетом для изучения влияния эмоциональной составляющей на параметры речевого сигнала. Программный комплекс openEAR содержит предопределенные конфигурационные скрипты, согласно которым осуществляется вычисление тех или иных характеристик сигнала. В представленной работе применялись следующие конфигурационные скрипты:

Emo_IS09 – вычисление 384 отдельных признаков [4];

Emo_large – вычисление 6669 отдельных признаков [5].

Приведенный набор признаков описывает широкий спектр параметров речевого сигнала и, следовательно, отражает различные эмоциональные изменения в речевом тракте человека. Способность работы классификаторов с многомерными данными и зависимость времени классификации от размерности пространства признаков определяются длиной вектора признаков.

Речевой сигнал исследовался с помощью специализированной базы записей эмоциональной речи профессиональных актеров Берлинского университета [6]. База данных включает 535 речевых записей семи эмоциональных состояний 10 дикторов на немецком языке.

2. Оценка эффективности различных методов классификации эмоционального состояния человека по его речи

Классификация эмоционального состояния диктора по его речи является нетривиальной задачей из-за многообразия человеческих голосов. Это связано со значительным разбросом параметров речевого сигнала и с невозможностью выделения четких кластеров их значений для каждого типа эмоций. Физиологические и филологические особенности произношения отдель-

ного человека также усложняют процесс классификации речевых сигналов. Дополнительная сложность процесса классификации обусловлена высокой размерностью вектора признаков и, следовательно, необходимостью использования значительных вычислительных ресурсов.

Целью данного исследования является анализ и сравнение работы различных классификаторов с многомерными векторами признаков и выбор оптимального метода для решения задачи классификации эмоциональной окраски речи. В качестве базового инструмента применялся программный комплекс WEKA [7], разработанный университетом Вайкато (Новая Зеландия) для обработки, классификации, кластеризации и селекции данных. Инструментарий обладает удобным интерфейсом и широким набором методов для осуществления всестороннего анализа данных. Эффективность классификации анализировалась при использовании следующих методов:

- наивного байесовского классификатора (НБК) [8];
- метода опорных векторов (МОВ) [9];
- дерева принятых решений (ДПР) [10];
- К-ближайших соседей (КБС) [11].

При испытании методов классификации для каждого диктора и набора признаков определялись следующие параметры:

- средняя точность классификации P :

$$P = \frac{1}{N_E} \sum_{i=1}^{N_E} \frac{|M_i \cap C_i|}{|C_i|}, \quad (1)$$

где N_E – число исследуемых типов эмоций (в представленном исследовании $N_E = 7$); M_i – множество записей, принадлежащих к i -й эмоции; C_i – множество записей, отнесенных к i -й эмоции; $|M_i|$ – мощность множества, определяющая число элементов в данном множестве;

- средняя полнота классификации R :

$$R = \frac{1}{N_E} \sum_{i=1}^{N_E} \frac{|M_i \cap C_i|}{|M_i|}; \quad (2)$$

- среднее значение F -оценки:

$$F = \frac{1}{\alpha \frac{1}{P} + (1 - \alpha) \frac{1}{R}} \quad (3)$$

при $\alpha = 0,5$.

Данные табл. 1 показывают, что наиболее эффективным из протестированных является классификатор на основе метода опорных векторов. Дополнительным преимуществом этого метода является устойчивость к масштабируемости пространства признаков. Принимая во внимание тот факт, что исследуемые методы классификации имеют входные настроечные параметры, которые влияют на конечный результат, а также значительный размер вектора признаков, можно предположить, что входной вектор можно оптимизировать удалением малоинформативных признаков, а классификатор адаптировать к входным данным и используемому вектору признаков.

Таблица 1

Результаты испытаний различных методов для классификации эмоциональной окраски диктора

Набор признаков	Метод	Средняя точность классификации, P	Средняя полнота классификации, R	F -оценка
IS09	НБК	0,57	0,54	0,49
	МОВ	0,78	0,76	0,76
	ДПР	0,51	0,50	0,48
	КБС	0,64	0,55	0,54
Emo_large	НБК	0,60	0,57	0,52
	МОВ	0,77	0,76	0,75
	ДПР	0,50	0,45	0,43
	КБС	0,72	0,64	0,64

3. Стохастический алгоритм оптимизации вектора признаков и параметров классификатора эмоциональной окраски речи

В ранее упомянутых работах предложены подходы к оптимизации входных данных, фильтрации малоинформативных признаков и их ранжированию для достижения наибольшей эффективности классификации. Предприняты попытки поиска единого оптимального метода для решения задачи классификации эмоциональной окраски речи. Методы и подходы предшествующих исследователей имеют значительный недостаток, состоящий в том, что оптимизация вектора признаков, селекция его атрибутов и настройка параметров классификатора происходят отдельно друг от друга. В предлагаемом алгоритме одновременно осуществляются оптимизация вектора признаков и оптимизация параметров классификатора. Входные данные и классификатор рассматриваются как взаимосвязанные составляющие алгоритма.

Не существует единого аналитического подхода для решения задачи оптимизации вектора признаков и параметров классификатора. Подобные задачи выбора из заданного множества параметров наилучшей их комбинации или нахождения оптимального значения параметра относятся к классу комбинаторно-оптимизационных задач, для решения которых применяются эволюционные алгоритмы.

Разработанные алгоритмы отличаются принципом поиска нового решения, способом кодирования решения, скоростью сходимости и т. д. В предложенном алгоритме используется метод роя частиц, принцип которого основан на имитации поведения квантовых частиц. Данный тип эволюционного алгоритма был предложен Е. Кеннеди для моделирования поведения птиц в стае [12]. Дальнейшим развитием данного подхода явилось моделирование поведения частиц, подчиняющегося законам квантовой механики [13]. Важным преимуществом нового алгоритма является более быстрая сходимость относительно других эволюционных алгоритмов [14], что позволяет находить решение за меньшее число итераций.

Основным положением метода роя частиц является движение множества частиц в многомерном пространстве возможных решений, в котором каждое измерение соответствует отдельному оптимизируемому признаку. Каждая частица представлена точкой в многомерном пространстве $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{iD-1}, x_{iD})$, которая соответствует текущему решению, найденному i -й частицей, где D – размерность пространства признаков. С каждой частицей роя X_i ассоциируется текущее положение $P_i^{loc}(t)$ и $P_i^{Gloc}(t)$ – наилучшее положение, достигнутое до текущей итерации t . С роем ассоциируется наилучшее достигнутое положение $P^G(t)$ до текущей итерации t среди всех частиц. Положения частицы обновляются согласно следующим выражениям, которые учитывают имитацию поведения квантовых частиц [15]:

$$P_{ij}^{loc}(t+1) = p_{ij}(t) + \beta \cdot |m_j(t) - P_{ij}^{loc}(t)| \cdot \ln\left(\frac{1}{u}\right) \quad \text{при } k \leq 0,5; \quad (4)$$

$$P_{ij}^{loc}(t+1) = p_{ij}(t) - \beta \cdot |m_j(t) - P_{ij}^{loc}(t)| \cdot \ln\left(\frac{1}{u}\right) \quad \text{при } k > 0,5, \quad (5)$$

где u , k – случайные величины, равномерно распределенные в интервале $(0;1)$; β – коэффициент, определяющий величину изменения положения частицы; $p_{ij}(t)$ – координаты аттрактора частицы, которые определяются выражением

$$p_{ij}(t) = (c_1 P_{ij}^{loc}(t) + c_2 P_j^G(t)) / (c_1 + c_2). \quad (6)$$

Здесь c_1 , c_2 – параметры, определяющие величину связи отдельной частицы с остальными частицами роя.

Функция $m_j(t)$ представляет собой j -ю усредненную координату, вычисляемую по наилучшим положениям для каждой частицы роя согласно следующему выражению:

$$m(t) = (m_1(t) \dots m_j(t) \dots m_D(t)) = \left(\frac{1}{S} \sum_{i=1}^S \alpha_i P_{i1}^{Gloc}(t) \dots \frac{1}{S} \sum_{i=1}^S \alpha_i P_{ij}^{Gloc}(t) \dots \frac{1}{S} \sum_{i=1}^S \alpha_i P_{iD}^{Gloc}(t) \right), \quad (7)$$

где S – количество частиц в рое; α_i – вес отдельной частицы, который определяется из правила «чем лучше решение, найденное частицей, тем больший вес она имеет». Описанный подход позволяет улучшить сходимость метода роя частиц.

Найденное решение оценивалось точностью классификации при использовании вычисленных параметров классификатора и оптимизированного вектора признаков. Для получения надежных оценок работы классификатора использовался метод *K-fold cross-validation* при значении $K = 3$. Суть данного подхода заключается в том, что входная обучающая выборка данных разбивается на три непересекающиеся части случайным образом. Две части используются для обучения классификатора, а по третьей части оценивается точность P , полнота R и F -оценки классификации эмоциональной окраски речи. Описанная процедура повторяется три раза, при этом каждая часть выборки данных используется как для обучения, так и для тестирования классификатора. Подобная техника исключает влияние разбивки данных на конечный результат, обеспечивая эффективное использование данных при обучении и получение более надежных оценок, поскольку результирующее значение является усреднением нескольких испытаний классификатора.

Новшеством, привнесенным в существующий метод и позволяющим ускорить его сходимость, а также повысить поисковую способность алгоритма, является перераспределение частиц в рое согласно алгоритму, изображенному на рис. 1.

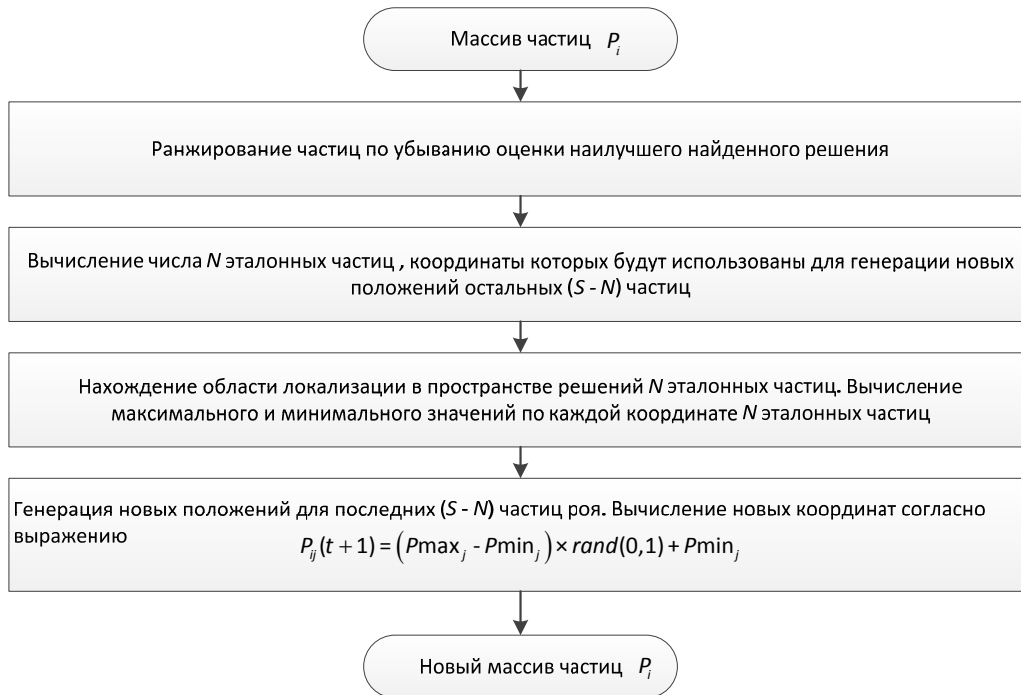


Рис. 1. Алгоритм перераспределения частиц в рое

Число эталонных частиц N , которые не изменят своего положения в пространстве решений при перераспределении частиц, рассчитывалось следующим образом:

$$N = \left\lceil S \cdot \left(1 - \frac{\lg\left(\frac{t \cdot 1000}{MI} \cdot 0,9\right)}{3} \right) \right\rceil, \quad (8)$$

где MI – максимальное число итераций; t – текущая итерация алгоритма. Предложенная функция вычисления числа эталонных частиц N , которые не изменят своего положения в пространстве решений, на начальных итерациях алгоритма не ограничивает пространство поиска (значение N составляет более половины числа всех частиц), а на завершающих итерациях позволяет более детально исследовать предполагаемое подпространство решений бóльшим коли-

чеством частиц. На рис. 2 изображен вид дробного множителя при параметре S в функции (8) для числа итераций $MI = 200$.

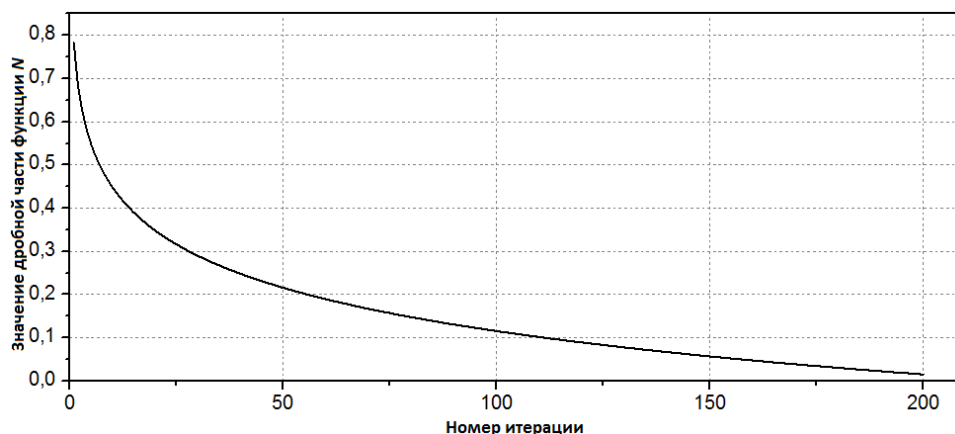


Рис. 2. График функции для вычисления числа эталонных частиц в рое

Для сравнения эффективности существующего алгоритма с его версией, включающей предложенный алгоритм перераспределения частиц в рое, было проведено тестирование, которое заключалось в сравнении точности классификации при использовании найденных решений, оптимизированных векторов признаков и параметров классификатора. Было проведено 10 тестовых испытаний, в течение которых фиксировались значения точностей оценок найденного решения на каждой итерации алгоритма (рис. 3).

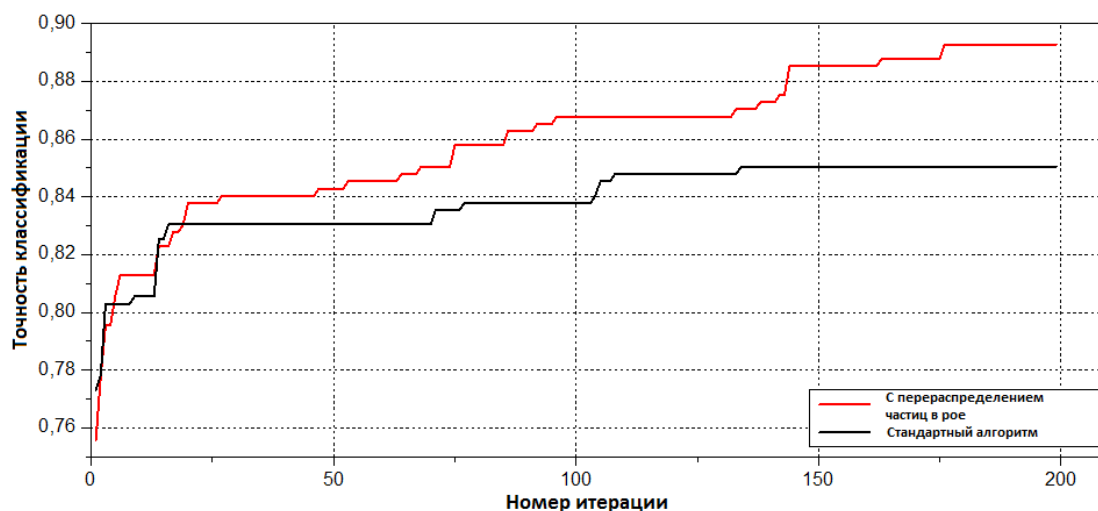


Рис. 3. Значения оценки найденного решения в зависимости от итерации алгоритма

Важным этапом при реализации метода роя частиц является кодирование решения или задание соответствия между положением частицы и найденным решением. Поскольку в предложенном алгоритме происходит одновременная оптимизация вектора признаков и параметров классификатора, размерность пространства решений будет определяться следующим выражением:

$$D = N_F + N_P, \quad (9)$$

где N_F – число признаков речевого сигнала; N_P – число оптимизируемых параметров классификатора. Без потери общности полагается, что первые N_F координат частицы соответствуют N_F признакам речевого сигнала, а последние N_P координат вектора соответствуют оптимизируемым параметрам классификатора. В предложенном алгоритме $N_F = 384$ для набора признаков IS09 и $N_F = 6669$ для набора признаков Emo_large, $N_P = 2$ для обоих случаев. Поскольку

каждая координата является вещественным числом и может принимать любое значение из заданного диапазона, а каждый признак речевого сигнала может как присутствовать (что соответствует значению координаты, равному 1), так и отсутствовать (значение соответствующей координаты равно -1) в векторе найденного решения, необходимо провести дискретизацию значений первых N_F координат для оценки найденного решения. Для решения задачи декодирования значения координаты применялось следующее правило:

$$\begin{cases} 1, & \text{если } d_j \geq 0 \text{ при } j = \overline{1 \dots N_F}; \\ -1, & \text{если } d_j < 0 \text{ при } j = \overline{1 \dots N_F}, \end{cases} \quad (10)$$

где d_j – j -я координата пространства решений. Рассмотренные этапы реализации алгоритма составляют единый стохастический алгоритм оптимизации входного вектора признаков и параметров классификатора (рис. 4).

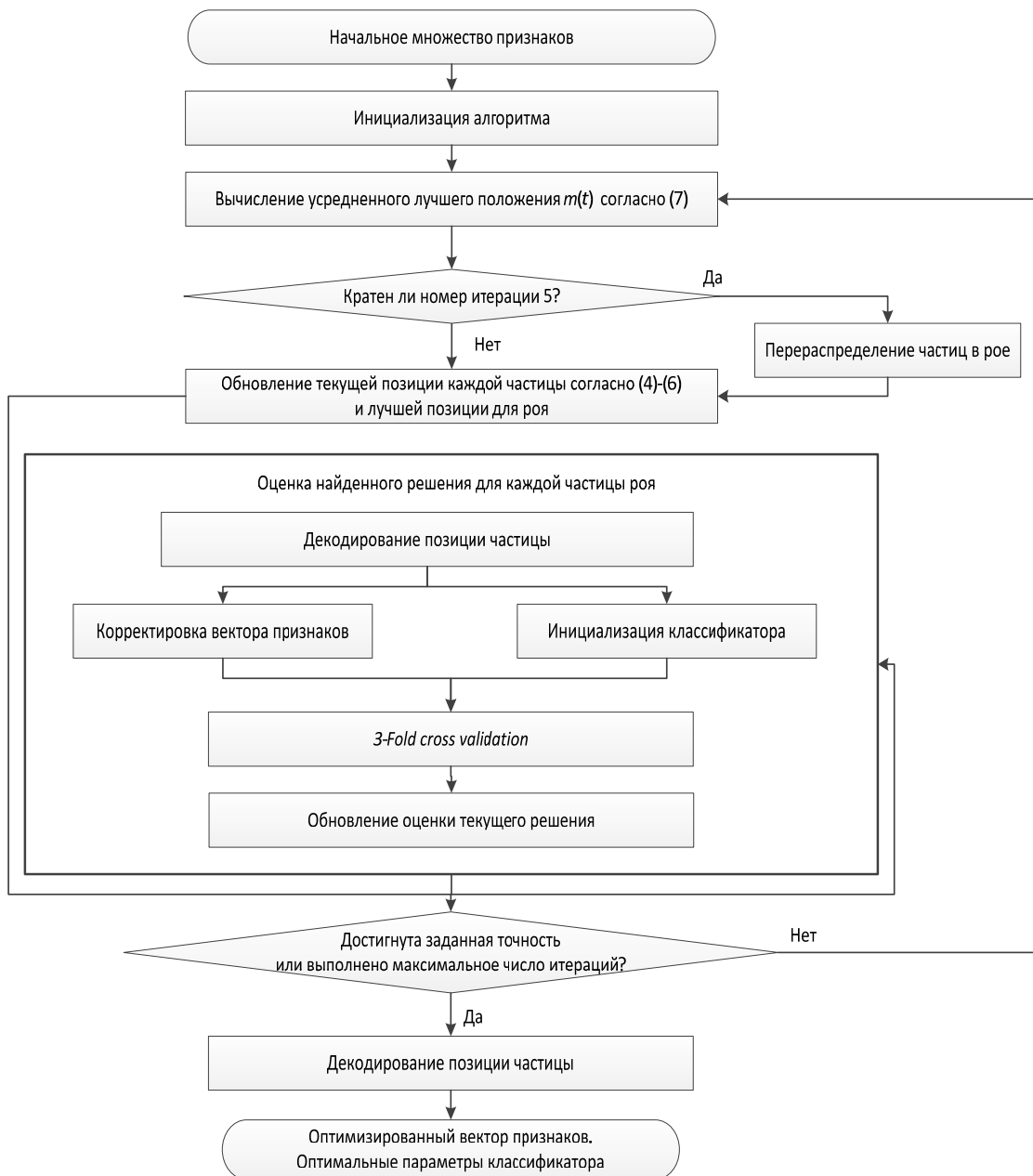


Рис. 4. Обобщенная схема алгоритма оптимизации вектора признаков и параметров классификатора

4. Результаты тестирования

Практическое тестирование имело целью апробацию предложенного алгоритма оптимизации входного вектора признаков и параметров классификатора эмоционального состояния человека. Главными исследуемыми параметрами являлись размер найденного вектора признаков, а также точность классификации эмоциональной окраски речи при использовании оптимизированного вектора признаков относительно начального набора признаков и параметров классификатора. Работоспособность алгоритма оценивалась по следующей методике. Записи из описанной выше базы данных делились случайным образом на две непересекающиеся части в соотношении 3:1 для каждого диктора. Первая часть данных, соответственно 75 % от общего числа записей, использовалась для оптимизации вектора признаков и параметров классификатора с помощью представленного выше алгоритма. Вторая часть данных, соответственно 25 % от общего числа записей, использовалась для оценки параметров работы классификатора с новым оптимизированным вектором признаков. В качестве классификатора использовался МОВ с линейным ядром [9]. Данный тип МОВ был выбран ввиду значительной размерности вектора признаков, который позволяет осуществлять задачу классификации при меньших временных затратах и требованиях к аппаратным ресурсам. Начальными параметрами алгоритма во всех испытаниях являлись исходные наборы признаков, соответственно включающих в себя 384 и 6669 признака, и параметры по умолчанию для классификатора, заданные в программном комплексе WEKA. Значение параметра β , найденное экспериментальным путем, линейно уменьшалось от значения 0,8 до 0,2.

В табл. 2 и 3 сравниваются результаты работы оптимизированного классификатора с найденным вектором признаков и его работы с начальными значениями параметров и полно-размерным вектором признаков.

Таблица 2

Результаты тестирования алгоритма для набора признаков IS09

Номер диктора	Начальная длина вектора признаков	Точность классификации с начальными параметрами, %	Длина оптимизированного вектора признаков	Точность классификации с найденными параметрами, %	Изменение точности классификации, %
1	386	76,92	220	69,23	-7,69
2	386	100,00	188	86,66	-13,34
3	386	81,82	191	72,73	-9,09
4	386	60,00	185	70,00	10,00
5	386	71,43	198	57,14	-14,29
6	386	66,67	192	55,56	-11,11
7	386	75,00	193	68,75	-6,25
8	386	94,44	202	77,78	-16,66
9	386	71,43	187	78,57	7,14
10	386	83,33	206	77,78	-5,55
Среднее	386	78,10	196	71,42	-6,68

Таблица 3

Результаты тестирования алгоритма для набора признаков Emo_large

Номер диктора	Начальная длина вектора признаков	Точность классификации с начальными параметрами, %	Длина оптимизированного вектора признаков	Точность классификации с найденными параметрами, %	Изменение точности классификации, %
1	6671	69,23	3334	76,92	7,69
2	6671	100,00	3282	93,33	-6,67
3	6671	90,90	3318	72,72	-18,18
4	6671	70,00	3291	70,00	0,00
5	6671	64,28	3356	64,28	0,00
6	6671	55,55	3285	66,66	11,11
7	6671	75,00	3334	75,00	0,00
8	6671	88,88	3341	94,44	5,56
9	6671	85,71	3377	92,86	7,15
10	6671	77,77	3392	83,33	5,56
Среднее	6671	77,73	3331	78,95	1,22

Следующим этапом исследования было тестирование алгоритма в дикторонезависимом режиме классификации эмоций (табл. 4), для чего записи всех дикторов объединялись в одно множество, которое затем разделялось случайным образом в соотношении 3:1 на обучающую и тестовую выборку соответственно. Остальные параметры алгоритма оставались прежними, как и при предыдущем тестировании.

Таблица 4

Результаты тестирования алгоритма в дикторонезависимом режиме

Набор признаков	Начальная длина вектора признаков	Точность классификации с начальными параметрами, %	Длина оптимизированного вектора признаков	Точность классификации с найденными параметрами, %	Изменение точности классификации, %
IS09	386	76,12	186	79,10	2,98
Emo large	6671	86,57	3337	88,06	1,49

Заключение

Полученные результаты показали работоспособность алгоритма. Преимущества предложенного алгоритма проявляются при применении вектора признаков со значительным начальным числом признаков порядка нескольких тысяч. Увеличение обучающей выборки также положительно влияет на точность классификации. Стоит отметить такое важное свойство предложенного алгоритма, как выделение наиболее значимых и информативных признаков речевого сигнала для классификации эмоциональной окраски речи человека.

Список литературы

1. Williams, C.E. Emotions and Speech: Some Acoustical Correlates / C.E. Williams // Journal of the Acoustical Society of America. – 1972. – Vol. 52, № 4. – P. 1238–1250.
2. Banse, R. Acoustic profiles in vocal emotion expression / R. Banse // Journal of personality and social psychology. – 1996. – Vol. 70, № 3. – P. 614–636.
3. Eyben, F. OpenEAR – Introducing the Munich Open-Source Emotion and Affect / F. Eyben // Proc. ACII. – Amsterdam, Netherlands, 2009. – P. 576–581.
4. Schuller, B. The interspeech 2009 emotion / B. Schuller // Interspeech. – Brighton, UK, 2009.
5. Eyben, F. OpenSMILE : the Munich open Speech and Music Interpretation by Large Space Extraction toolkit / F. Eyben, M. Woellmer, B. Schuller. – 2010. – Mode of access : http://sourceforge.net/projects/opensmile/files/openSMILE_book_1.0.1.pdf. – Date of access : 20.01.2012.
6. Berlin Database of Emotional Speech [Electronic resource]. – Technical University Berlin, 1999. – Mode of access : <http://pascal.kgw.tu-berlin.de/emodb/index-1024.html>. – Date of access : 20.01.2012.
7. Holmes, G. WEKA: a machine learning / G. Holmes // Intelligent Information Systems. – 1994. – № 29. – P. 357–361.
8. John G.H. Estimating Continuous Distributions in Bayesian Classifiers / G.H. John // Proc. of the Eleventh Conf. on Uncertainty in Artificial Intelligence. – San Francisco, USA, 1995. – Vol. 1. – P. 338–345.
9. LIBLINEAR – A Library for Large Linear Classification // Rong-En Fan [et al.]. – 2008. – Mode of access : <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>. – Date of access : 20.01.2012.
10. Quinlan, R.J. C4.5: Programs for Machine Learning / R.J. Quinlan // Machine Learning. – 2004. – Vol. 16, № 3. – P. 235–240.
11. Aha, D.W. Instance-Based Learning Algorithms / D.W. Aha // Machine Learning. – Hingham, USA, 1991. – Vol. 6, № 1. – P. 37–66.
12. Kennedy, J. Particle Swarm Optimization / J. Kennedy // Proc. of IEEE Intern. Conf. on Neural Networks. – Washington, USA, 1995. – Vol. 4. – P. 1942–1948.
13. Sun, J. Particle Swarm Optimization with particles having Quantum Behavior / J. Sun // Proc. of Congress on Evolutionary Computation. – Portland, USA, 2004. – P. 325–331.

14. Gao, H. A Simple Quantum-inspired Particle swarm Optimization and its Application / H. Gao // Information Technology Journal. – 2011. – № 10. – P. 2315–2321.

15. Maolong, X. An improved quantum-behaved particle swarm optimization algorithm with weighted mean best position / X. Maolong // Applied Mathematics and Computation. – 2008. – № 205. – P. 751–759.

Поступила 07.07.12

*Белорусский государственный университет,
Минск, Курчатова, 5
e-mail: pekar.dzmitry@gmail.com*

D.V. Pekar, S.G. Tsikhanenka, A.F. Chernyavskii, V.S. Sadov

**ALGORITHM FOR OPTIMIZATION OF FEATURES
AND PARAMETERS OF SPEAKER'S EMOTIONAL STATE CLASSIFIER**

The paper presents a particle swarm based algorithm for optimization of feature vector and parameters of speaker's emotional state classifier. To speed up the convergence, an enhancement of existing quantum inspired particle swarm optimization algorithm is proposed. The algorithm shows better quality when using optimized feature vector and new classifier's parameters.