III. МАТЕМАТИКА В ОПИСАНИИ ХАОСА И СИНЕРГЕТИЧЕСКИХ СИСТЕМ

DOI: 10.12737/21051

СИСТЕМА ОЦЕНКИ ПСИХОЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ ДИКТОРА ПО ГОЛОСУ

В.Е. ГАЙ, В.А. УТРОБИН, И.В. ПОЛЯКОВ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Нижегородский государственный технический университет им. Р.Е. Алексеева», ул. Минина, 24, Нижний Новгород, Нижегородская обл., Россия 603155, Россия

Аннотация. Работа посвящена описанию системы оценки психоэмоционального состояния человека по голосу. Для создания описания звукового сигнала, содержащего запись эмоциональной речи диктора, используется теории активного восприятия. Приводятся результаты исследования предложенной системы признаков с использованием классификаторов: метод опорных векторов и *k*-ближайших соседей. Тестирование системы выполняется на базе данных EmoDB (http://emodb.bilderbar.info/).

Ключевые слова: цифровая обработка сигналов, распознавание эмоцией, теория активного восприятия, метод опорных векторов.

THE EVALUATION SYSTEM EMOTIONALTHE STATE OF THE SPEAKER'S VOICE

V.E. GAI, V.A. UTROBIN, I.V. POLYAKOV

Federal state budget-ing the institution of higher education "Nizhny Novgorod-tion of the state technical university. Alec RE-seeva" Str. Minin, 24, Nizhny Novgorod, Nizhny Novgorod region., Russia 603155, Russia

Abstract. The work is dedicated to Opis, of, human voice psycho-emotional state assessment system. To create a description of the audio signal containing recording emotional speech speaker, used the theory of active perception. The results of the study of the proposed indication system c the use of classifiers: support vector machine and k-nearest neighbors. System testing is performed on the basis of data EmoDB (http://emodb.bilderbar.info/).

Key words: digital signal processing, recognition of emotion, the theory of ac-tive perception, a support vector machine.

Введение. В речи человека передаётся два типа информации. Семантическая часть несёт информацию о предметах, объектах, действиях. Паралингвистическая часть, в свою очередь, используется для передачи неявного сообщения, например, об эмоциональном состоянии человека [16]. В связи с этим автоматическое распознавание эмоций приобретает большое значение во многих приложениях, например при создании интеллектуальных помощников [13], при решении задач, связанных с обеспечением безопасности, человеко-машинном взаимодействии, общении людей друг с другом [9,10,12,15].

Описанные задачи предъявляют высокие требования к точности классификации, что в свою очередь связано с верным выбором систем признаков и алгоритмов классификации.

Задачу распознавания эмоций можно рассмотреть с точки зрения системного

подхода. В этом случае данная задача включает три этапа: предварительную обработку данных, вычисление признаков и принятие решения (классификация). Предварительная обработка, обычно заключается в фильтрации сигнала. На этапе формирования системы признаков по обработанному сигналу вычисляются разнообразные признаки. Например, признаки на основе частоты основного тона, признаки на основе энергетических характеристик сигнала, мел-частотные кепстральные коэффициенты, коэффициенты линейного предсказания. В некоторых работах показано применение для моделирования данных моделей гауссовой смеси и скрытых марковских моделей. На этапе классификации, обычно, применяется метод опорных векторов, метод *k*-ближайших соседей, деревья решений, принцип максимума правдоподобия, классификатор Байеса.

В настоящей работе для решения задач предварительной обработки и формирования системы признаков предлагается использовать теорию активного восприятия [7]. Оригинальность разрабатываемой системы признаков связана с рассмотрением вычислительных процедур обработки звуковых сигналов с точки зрения концепции грубо-точного анализа сигналов с обеспечением максимального параллелизма при обработке и распознавании.

1. Технология оценки эмоционального состояния человека. Рассмотрим предлагаемый подход к решению задачи распознавания эмоционального состояния человека по голосу.

Предварительная обработка сигнала, с позиций теории активного восприятия, заключается в выполнении операции интегрирования. На данном этапе обработки анализируемый сигнал разбивается на сегменты, по каждому из которых вычисляется *Q*-преобразование:

$$g(i) = Q[\ h_i\ ],\ g_i = \sum_{k=1}^{L} h_i(k),$$

где $i = \overline{1,N}$, N – число отсчётов в сигнале g, $h = \{h_i\}$, h – множество сегментов, вычисленных по сигналу f, L – количество отсчётов в сегменте. Таким образом, на следующий этап, этап вычисления признаков, передаётся сигнал g.

Рассмотрим метод, предлагаемый для создания признакового описания сигнала g:

1) отсчёты сигнала g разбиваются на множество сегментов $g=\{g_k\}$, длиной 16 отсчётов, со смещением в S отсчётов;

2) к каждому сегменту g_k применяется U-преобразование (U-преобразование является базовым в теории активного восприятия), в результате формируется спектральное представление каждого сегмента: $u_k=U[g_k]$, $u=\{u_k\}$, где U – оператор вычисления U-преобразования;

3) по вычисленному спектральному представлению u_k сегмента g_k формируется описание с помощью полных групп:

$$V=GV[u],\ P_{ni}=GP_{ni}[u,\ V],$$

где GV – оператор вычисления операторов, GP_{ni} – оператор вычисления полных групп, $V=\{v_k\}$ – множество значений операторов, вычисленных по сигналу, $P_{ni}=\{P_{nia,\ k}\}$ – множество значений полных групп, $k = \overline{1,N}$;

4) для объединения данных, полученных от разных сегментов анализируемого сигнала, вычисляется двумерная гистограмма полных групп:

$$h_{ni} = H[P_{ni},\ 2d],$$

где h_{ni} – гистограмма полных групп на операции сложения, H – оператор вычисления гистограммы заданной размерности. В двумерной гистограмме учитываются возможные появления пар групп в описании одного сегмента сигнала.

Этап классификации может быть реализован с помощью нескольких классификаторов. В предлагаемой системе для классификации используется линейный метод опорных векторов и метод *k*-ближайших соседей.

Решающее правило метода опорных векторов выглядит следующим образом:

$$a(x) = sign\left(\sum_{j=1}^{n} w_j x^j - w_0\right),$$

где $x = (x^1,...,x^n)$ – признаковое описание объекта x (одно из возможных описаний, приведённых выше), вектор $w = (w^1,...,w^n)$ и скалярный порог w_0 являются параметрами

алгоритма. Метод опорных векторов является бинарным классификатором. В данной работе для решения задачи мультиклассовой классификации используются два способа сведения данной задачи к бинарной [6]:

1) подход «один-против-всех» (*One-vs-All*) заключается в обучении N классификаторов по следующему принципу:

$$f_i(x)=\begin{cases}\geq 0, если\ y(x)=i,\\ <0, если\ y(x)\neq i,\end{cases}$$

которые отделяют каждый класс от остальных. Далее, для каждого $x\in X$ вычисляются все классификаторы и выбирается класс, соответствующий классификатору с большим значением:

$$a(x)=\arg\max_{i\in\overline{1,N}} f_i(x);$$

2) подход «один-против-одного» (*One-vs-One*) заключается в формировании $N(N-1)$ классификаторов, которые разделяют объекты пар различных классов:

$$f_{ij}(x)=\begin{cases}+1, если\ y(x)=i,\\ -1, если\ y(x)=j.\end{cases}$$

После обучения бинарных классификаторов решение принимается следующим образом:

$$a(x)=\arg\max_{i\in\overline{1,N}}\sum_{\substack{j=1\\ j\neq i}}^{N} f_{ij}(x).$$

При классификации используется линейное ядро: $k(x,y)=x^T y+c.$

Решающее правило метода k-ближайших соседей записывается следующим образом:

$$a(u;X^l,k)=\arg\max_{y\in Y}\sum_{i=1}^{k}\left[y_u^{(i)}=y\right],$$

где u – классифицируемый объект, k – параметр алгоритма (количество соседей), $X^m=\{(x_1,y_1),...,(x_m,y_m)\}$ – обучающая выборка, заданная в формате «объект-ответ», $Y=\{y_i\}$, $y\in\overline{1,C}$ – множество классов, C – количество классов. Определение близости между объектами x и x' выполняется с помощью расстояния Евклида:

$$\rho(x,x')=\sum_{i=1}^{M}\sqrt{(x_i-x_i')^2}.$$

Оптимальное значение параметра k определим по критерию скользящего контроля с исключением объектов по одному (*leave-one-out, LOO*):

$$LOO(k,X^l)=\sum_{i=1}^{l}\left[a(x_i;X^l\setminus\{x_i\},k)\neq y_i\right]\to\min_k.$$

2. Архитектура системы оценки эмоционального состояния. Разработанная система оценки эмоционального состояния включает следующие элементы:

1) клиентское приложение для сбора данных;

2) серверное приложение используется для организации распределённых вычислений;

3) клиентское приложение для обработки данных;

4) приложение оператора.

Предлагаемая система работает в двух режимах:

1) обучение, т. е. анализ записей голоса диспетчера, находящегося в различных эмоциональных состояниях и настройка параметров метода опорных векторов (построение модели диспетчера). Полученная модель для каждого диспетчера сохраняется на сервере.

2) классификация эмоционального состояния диспетчера по речевому сигналу.

Основная задача клиентского приложения для сбора данных – запись переговоров диспетчера и пересылка их на сервер для дальнейшего анализа. На сервере они добавляются в очередь обработки и сохраняются в базе данных. Первый звуковой файл, находящийся в очереди, выдаётся обработчикам данных. Если обработчик не справился с заданием в выделенный промежуток времени, этот же файл выдается другому обработчику. Сервер является администратором распределенной сети. Его задача состоит в распределении заданий между обработчиками таким образом, чтобы их загрузка была оптимальной.

Клиентское приложение для обработки данных после подключения к серверу запрашивает у него данные для обработки. Алгоритм обработки данных состоит в вычислении признаков по звуковому сигналу. После окончания обработки результаты отправляются на сервер. В задачи сервера также входит выполнение классификации эмоционального состояния диспетчера.

Приложение оператора получает обработанную информацию с сервера для каждого подключенного клиента и представляет её в удобном для просмотра виде. С помощью дополнительных запросов на сервер можно просмотреть историю всех ранее подключенных клиентов. Информация динамически обновляется и может быть представлена в виде графиков. Окончательное решение об эмоциональном состоянии система не принимает, она лишь даёт указание оператору о приблизительном эмоциональном состояния диспетчера в данный момент времени.

Рассмотрим предлагаемую структуру базы данных, используемой в системе (рис.). База данных хранится на сервере и использует систему управления *SQLite 3*. В состав базы данных входят четыре таблицы, в которых хранятся данные о клиентах (*Clients*), параметры полученного звукового файла, время записи и получения звуковой информации сервером (*Samples*), информация о клиентском приложении для обработки данных (*Handler*), результаты, полученные от клиентского приложения для обработки данных (*Results*).

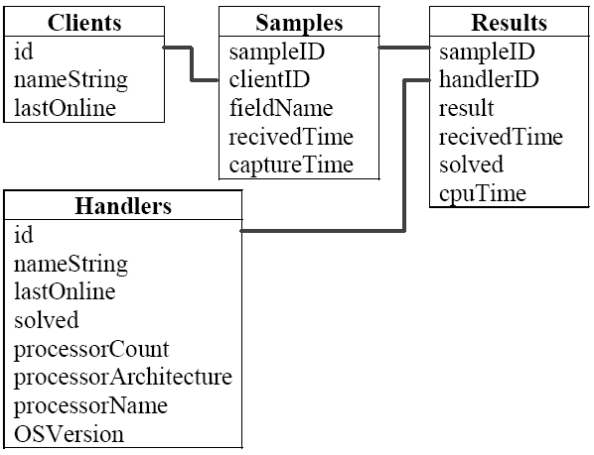

Рис. Структурная схема базы данных

3. Вычислительный эксперимент. 3.1. Описание базы данных. Тестирование предложенной системы признаков будет выполняться на основе базы данных эмоциональной речи *EmoDB* [8].

База данных содержит записи голоса 10 дикторов (профессиональных актёров, 5 мужчин, 5 женщин), которые выражают следующие эмоции: злость, скука, отвращение, страх (тревога), счастье, печаль, нейтральное состояние.

3.2. Известные результаты. Рассмотрим результаты точности классификации на основе известных методов. В работе [14] решается задача дикторозависимой классификации эмоций, используются признаки, вычисляемые на основе частоты основного тона, энергии сигнала, классификатор: скрытая марковская модель. Достигнутая точность классификации составляет 80%. В работе [17] решается задача дикторонезависимой классификации эмоций, используются признаки, вычисляемые на основе пакетного вейвелет разложения, классификатор: линейный метод опорных векторов. Достигнутая точность классификации составляет 79,5%. В работе [11] решается задача дикторонезависимой классификации эмоций, для описания сигнала используются гармонические признаки, для моделирования данных используется гауссова смесь, классификатор – метод наибольшего правдоподобия. Достигнутая точность классификации составляет 73,.5%.

3.3. Результаты тестирования предложенной системы признаков. Вычислительный эксперимент заключается в проверке точности идентификации эмоционального состояния диктора на основе предложенной системы признаков с использованием метода опорных векторов.

Таблица 1

Точность дикторозависимой классификации

Классиф./ Диктор	*SVM*		*KNN*
	1-N	1-1	
Дикт. №1	89	79	71
Дикт. №10	95	90	83

Таблица 2

Точность дикторонезависимой классификации

SVM		*KNN*
1-*N*	1-1	
83	79	69

Известно, что человек может оценить эмоциональное состояние другого человека

только после достаточно длительного общения с ним. Поэтому в работе предполагается, что для устойчивой классификации эмоционального состояния нужно для каждого пользователя системы (диспетчера) создать собственную модель. Приводятся также результаты вычислительного эксперимента для дикторонезависимого подхода к распознаванию эмоций. Вычислительный эксперимент выполнялся на основе метода перекрёстной проверки (данные разбивались на 10 частей).

Выводы по результатам вычислительного эксперимента:

1) с использованием предложенной системы признаков, точность решения задачи дикторонезависимой классификации эмоционального состояния ниже, чем дикторозависимой;

2) полученные результаты по точности не уступают известным, а в ряде случаев – превышают их;

3) при использовании указанной системы признаков максимальная точность идентификации психоэмоционального состояния достигается с использованием метода опорных векторов. Подобный векторный подход сейчас широко используется в эффекте Еськова-Зинченко [1,4] при оценке психофизиологического состояния испытуемых. В этом случае производится расчет параметров квазиаттракторов [2,5].

Заключение. В работе описывается система оценки эмоционального состояния человека по голосу. Распознавание эмоционального состояния оказывается полезным в любой сфере человеческой деятельности, где требуется его оперативная оценка – в маркетинге, медицине, психологии, обеспечении безопасности и т.п. Полученные результаты подтверждают эффективность использования двумерной гистограммы полных групп в качестве системы признаков для описания звукового сигнала. С использованием теории активного восприятия возможно создание других систем признаков. На это и будет направлена дальнейшая работа авторского коллектива.

Литература

1. Еськов В.М., Газя Г.В., Майстренко Е.В., Болтаев А.В. Влияние промышленных электромагнитных полей на параметры сердечно-сосудистой системы работников нефтегазовой отрасли // Экология и промышленность России. 2016. № 1. С. 59–63.
2. Еськов В.М., Еськов В.В., Вохмина Ю.В., Гавриленко Т.В. Эволюция хаотической динамики коллективных мод как способ описания поведения живых систем // Вестн. Моск. ун-та. Сер. 3. Физ. Астрон. 2016. № 2.
3. Еськов В.М., Зинченко Ю.П., Филатов М.А., Поскина Т.Ю. Эффект Н.А. Бернштейна в оценке параметров тремора при различных акустических воздействиях // Национальный психологический журнал. 2015. № 4. С. 66–73.
4. Еськов В.М., Хадарцев А.А., Еськов В.В., Вохмина Ю.В. Хаотическая динамика кардиоинтервалов трёх возрастных групп представителей коренного и пришлого населения Югры // Успехи геронтологии. 2016. Т. 29, № 1. С. 44–51.
5. Зинченко Ю.П., Еськов В.М., Еськов В.В. Понятие эволюции Гленсдорфа-Пригожина и проблема гомеостатического регулирования в психофизиологии // Вестник Московского университета. Серия 14: Психология.

References

Es'kov VM, Gazya GV, Maystrenko EV, Boltaev AV. Vliyanie promyshlennykh elektromagnitnykh poley na parametry serdechnososudistoy sistemy rabotnikov neftegazovoy otrasli. Ekologiya i promyshlennost' Rossii. 2016;1:59-63. Russian.

Es'kov VM, Es'kov VV, Vokhmina YuV, Gavrilenko TV. Evolyutsiya khaoticheskoy dinamiki kollektivnykh mod kak sposob opisaniya povedeniya zhivykh system. Vestn. Mosk. un-ta. Ser. 3. Fiz. Astron. 2016;2. Russian.

Es'kov VM, Zinchenko YuP, Filatov MA, Poskina TYu. Effekt N.A. Bernshteyna v otsenke parametrov tremora pri razlichnykh akusticheskikh vozdeystviyakh. Natsional'nyy psikhologicheskiy zhurnal. 2015;4:66-73. Russian.

Es'kov VM, Khadartsev AA, Es'kov VV, Vokhmina YuV. Khaoticheskaya dinamika kardiointervalov trekh vozrastnykh grupp predstaviteley korennogo i prishlogo naseleniya Yugry. Uspekhi gerontologii. 2016;29(1):44-51. Russian.

Zinchenko YuP, Es'kov VM, Es'kov VV. Ponyatie evolyutsii Glensdorfa-Prigozhina i problema gomeostaticheskogo regulirovaniya v psikhofiziologii. Vestnik Moskovskogo universiteta. Seriya 14: Psikhologiya. 2016;1:3-24. Russian.

2016. № 1. С. 3–24.

6. Карасиков М.Е., Максимов Ю.В. Поиск эффективных методов снижения размерности при решении задач многоклассовой классификации путем её сведения к решению бинарных задач // Машинное обучение и анализ данных. 2014. Т. 1, № 9. С. 1273–1290.
   Karasikov ME, Maksimov JuV. Poisk jeffektivnyh metodov snizhenija razmernosti pri reshenii zadach mnogoklassovoj klassifikacii putem ejo svedenija k resheniju binarnyh zadach. Mashinnoe obuchenie i analiz dannyh. 2014;1(9):1273-90. Russian.
7. Утробин В.А. Элементы теории активного восприятия изображений // Труды Нижегородского государственного технического университета им. Р.Е. Алексеева. 2010. Т. 81, № 2. С. 61–69.
   Utrobin VA. Jelementy teorii aktivnogo vosprijatija izobrazhenij. Trudy Nizhegorodskogo gosudarstvennogo tehnicheskogo universiteta im. R.E. Alekseeva. 2010;81(2):61-9. Russian.
8. Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W., Weiss, B. A database of german emotional speech. In Proc. INTERSPEECH2005, 2005. P. 1517–1520.
   Burkhardt F, Paeschke A, Rolfes M, Sendlmeier W, Weiss B. A database of german emotional speech. In Proc. INTERSPEECH2005; 2005.
9. Christophe V., Devillers V. Negative emotions detection as an indicator of dialogs quality in call centers // in Proc. Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2012. P. 5109–5112.
   Christophe V, Devillers V. Negative emotions detection as an indicator of dialogs quality in call centers. in Proc. Acoustics, Speech and Signal Processing (ICASSP), IEEE; 2012.
10. El Ayadi M., Kamel M. S., Karray F. Survey on speech emotion recognition: Features, classification schemes, and databases // Pattern Recognition. 2011. V. 44, N. 3. P. 572–587.
    El Ayadi M, Kamel MS, Karray F. Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition. 2011;44(3):572-87.
11. Recognition of stress in speech using wavelet analysis and teager energy operator / He L. [et al.]// 9th Annual Conference, International Speech Communication Association and 12 Biennial Conference, Australasian Speech Science and Technology Association. ISCA, 2008. P. 605–608.
    He L, et al. Recognition of stress in speech using wavelet analysis and teager energy operator. 9th Annual Conference, International Speech Communication Association and 12 Biennial Conference, Australasian Speech Science and Technology Association. ISCA; 2008.
12. Opinions and attitudes toward humanoid robots in the Middle East / Mavridis N. [et al.] //AI & society. 2012. V. 27, N. 4. P. 517–534.
    Mavridis N, et al. Opinions and attitudes toward humanoid robots in the Middle East. AI & society. 2012;27(4):517-34.
13. Challenges in speech-based human–computer interfaces / Minker W. [et al.] // International Journal of Speech Technology. 2007. V. 10, N. 2-3. P. 109–119.
    Minker W, et al. Challenges in speech-based human–computer interfaces. International Journal of Speech Technology. 2007;10(2-3):109-19.
14. Speech emotion recognition using hidden Markov models / Nogueiras A. [et al.] // INTERSPEECH, 2001. P. 2679–2682.
    Nogueiras A, et al. Speech emotion recognition using hidden Markov models. INTERSPEECH; 2001.
15. Ntalampiras S., Potamitis I., Fakotakis N. An adaptive framework for acoustic monitoring of potential hazards // EURASIP Journal on Audio, Speech, and Music Processing. 2009. V. 2009. P. 13–23.
    Ntalampiras S, Potamitis I, Fakotakis N. An adaptive framework for acoustic monitoring of potential hazards. EURASIP Journal on Audio, Speech, and Music Processing. 2009;2009:13-23.
16. Nwe T.L., Foo S.W., De Silva L.C. Speech emotion recognition using hidden Markov models // Speech communication. 2003. V. 41, N. 4. P. 603–623.
    Nwe TL, Foo SW, De Silva LC. Speech emotion recognition using hidden Markov models. Speech communication. 2003;41(4):603-23.
17. Wang K., An N., Li L. Speech emotion recognition based on wavelet packet coefficient model // Chinese Spoken Language Processing (ISCSLP), 2014 9th International Symposium on. IEEE, 2014. P. 478–482.
    Wang K, An N, Li L. Speech emotion recognition based on wavelet packet coefficient model. Chinese Spoken Language Processing (ISCSLP), 2014 9th International Symposium on. IEEE; 2014.