УДК 681.391

В. Е. Гай, В. А. Утробин, П. А. Родионов, М. О. Дербасов

Оценка эмоционального состояния человека по голосу с позиций теории активного восприятия

Нижегородский государственный технический университет
им. Р. Е. Алексеева

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

Работа посвящена описанию системы признаков для оценки эмоционального состояния человека по голосу. Приводятся результаты исследования предложенной системы признаков. Тестирование метода выполняется на базе данных EmoDB (http://emodb.bilderbar.info/).

V. E. Gai, V. A. Utrobin, P. A. Rodionov, M. O. Derbasov

Evaluation of the emotional state of a person's voice from the standpoint of the theory of active perception

This abstract related to a desctription of attributes system for assessment emotional state of a person's voice. Provides results of research of this attributes system. Testing of the method was held on the EmoDB patterns

Key words: digital signal processing, emotion recognition, theory of active perception, 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 [1]. В связи с этим автоматическое распознавание эмоций приобретает большое значение во многих приложениях, например при создании интеллектуальных помощников [2], при решении задач, связанных с обеспечением безопасности [3], человеко-машинном взаимодействии [4, 5], общении людей друг с другом [6].

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, $\mathbf{h} = \{h_i\}$, $\mathbf{h}$ – 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 $\mathbf{g} = \{g_k\}$, 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: $u_k = U[g_k]$, $\mathbf{u} = \{u_k\}$, где *U* – оператор вычисления *U*-преобразования;

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

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

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

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

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

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 данной работе используется линейный метод опорных векторов.

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 [8]:

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 ∈ 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: $\mathbf{k}(\mathbf{x},\mathbf{y})=\mathbf{x}^T\mathbf{y}+c.$

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 (см. рис. 1). 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).

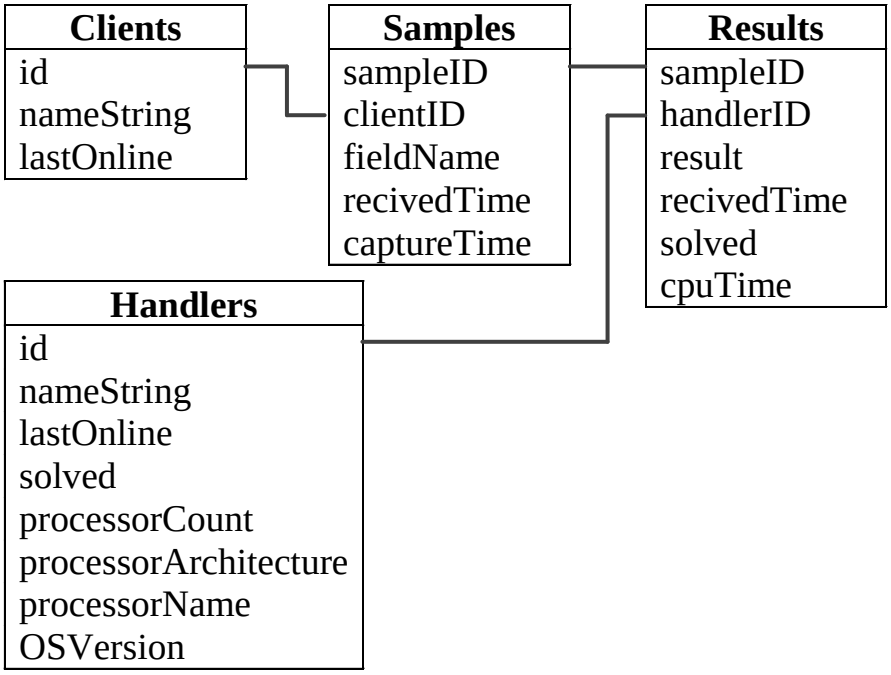

Рис. 1. Структурная схема базы данных

3. Вычислительный эксперимент

3.1. Описание базы данных

Тестирование предложенной системы признаков будет выполняться на основе базы данных эмоциональной речи EmoDB [9].

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 [10] 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 [11] 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 [12] 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 предложенного метода

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

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).

3.3.1. Дикторозависимая классификация

В табл.3-6 приведены оценки точности классификации для различных параметров алгоритма классификации.

Табл. 3. Точность классификации (1-1, нормализация признаков, в %)

Диктор / Параметры (*L* / *S*)	1	2	3	4	5	6	7	8	9	10
1 / 1	71	78	83	75	78	73	76	81	70	88
1 / 2	65	73	83	75	76	69	70	80	72	85
1 / 4	62	78	80	80	76	69	69	76	73	84
1 / 8	62	73	82	71	75	68	64	79	64	85
2 / 1	72	78	85	68	83	77	73	81	81	84
2 / 2	70	78	82	69	80	75	72	79	82	85
2 / 4	66	71	81	67	81	80	67	80	74	84
2 / 8	57	73	78	70	78	75	73	79	78	82
4 / 1	78	76	85	72	81	80	76	76	84	90
4 / 2	76	77	82	76	80	76	75	75	82	88
4 / 4	79	77	80	73	79	80	78	75	82	85
4 / 8	77	73	80	71	80	80	79	78	78	86

Табл. 4. Точность классификации (1-1, без нормализации признаков)

Диктор / Параметры (*L* / *S*)	1	2	3	4	5	6	7	8	9	10
1 / 1	69	75	78	77	77	71	70	81	68	87
1 / 2	70	73	75	73	77	71	67	82	76	87
1 / 4	68	72	75	76	74	69	64	75	73	86
1 / 8	62	66	77	69	76	68	58	78	63	85
2 / 1	72	74	79	70	84	79	72	83	75	82
2 / 2	67	76	78	71	79	80	69	78	78	85
2 / 4	57	74	83	71	78	80	67	76	72	82
2 / 8	51	65	69	72	77	71	69	79	69	79
4 / 1	78	75	86	70	73	76	76	77	79	89
4 / 2	77	73	83	73	79	76	76	77	79	84
4 / 4	76	76	83	73	78	75	80	76	84	82
4 / 8	73	70	84	69	74	79	82	75	82	80

Табл. 5. Точность классификации (1-N, нормализация признаков)

Диктор / Параметры (*L* / *S*)	1	2	3	4	5	6	7	8	9	10
1 / 1	76	85	86	81	77	84	77	86	84	95
1 / 2	74	76	88	81	74	78	79	85	84	92
1 / 4	74	73	83	84	78	78	74	78	81	88
1 / 8	62	77	82	73	71	84	76	78	78	90

2 / 1	81	84	86	80	83	84	79	85	81	85
2 / 2	77	79	83	83	83	88	80	84	85	87
2 / 4	76	73	88	81	77	89	79	83	77	82
2 / 8	72	67	82	77	79	84	81	85	79	83
4 / 1	86	79	88	80	80	87	83	78	92	92
4 / 2	86	78	86	80	79	89	86	79	89	90
4 / 4	84	80	88	73	77	87	88	79	89	90
4 / 8	82	78	91	73	78	89	87	82	86	88

Табл. 6. Точность классификации (1-N, без нормализации признаков)

Диктор / Параметры (*L* / *S*)	1	2	3	4	5	6	7	8	9	10
1 / 1	78	85	86	75	83	82	77	82	77	92
1 / 2	82	78	88	77	77	76	77	84	77	90
1 / 4	73	73	88	76	79	76	76	78	78	87
1 / 8	64	71	83	73	72	81	71	81	78	86
2 / 1	81	78	82	78	79	86	79	82	85	85
2 / 2	74	78	78	79	85	88	79	82	86	85
2 / 4	72	71	84	71	80	88	74	80	78	79
2 / 8	64	66	75	74	78	78	78	81	78	78
4 / 1	86	75	88	74	77	87	81	82	89	90
4 / 2	88	74	86	76	80	87	84	83	88	86
4 / 4	85	76	85	71	78	87	85	83	89	87
4 / 8	82	74	88	69	77	89	85	81	85	85

3.3.2. Дикторонезависимая классификация

В табл. 7 приведены результаты оценки точности классификации, полученные при решении задачи дикторонезависимой классификации эмоционального состояния.

Табл. 7. Точность классификации

Параметры (*L* / *S*) / Метод классиф.	1/1	1/4	1/8	2/1	2/4	2/8	4/1	4/4	4/8
1-1, SVM нормализ.	67	61	57	69	65	60	79	67	64
1-*N*, SVM нормализ.	63	58	52	69	62	57	82	64	62
1-1, SVM	67	62	58	67	65	61	80	68	65

без норм.									
1-*N*, SVM без норм.	64	58	53	71	62	57	83	66	65

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) максимальная точность классификации достигается при следующих значениях параметров *L*, *S*: *L* = 4, *S* = 1.

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 вычислительного эксперимента подтверждают эффективность предложенных систем признаков.

Статья рекомендована оргкомитетом XXI Международной научно-технической конференции «Информационные системы и технологии. ИСТ-2015». Публикуется при поддержке гранта РФФИ № 15-07-20095.

Список литературы

1. 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.

2. Minker W. et al. Challenges in speech-based human–computer interfaces //International Journal of Speech Technology. – 2007. – V. 10. – N. 2-3. – P. 109-119.

3. 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.

4. 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.

5. Mavridis N. et al. Opinions and attitudes toward humanoid robots in the Middle East //AI & society. – 2012. – V. 27. – N. 4. – P. 517-534.

6. 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.

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.

8. 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.

9. Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W., Weiss, B. A database of german emotional speech // In Proc. INTERSPEECH2005. – 2005. – P. 1517–1520.

10. Nogueiras A. et al. Speech emotion recognition using hidden Markov models // INTERSPEECH. – 2001. – P. 2679-2682.

11. 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.

12. 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. – P. 605-608.