

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPUTER ENGINEERING
AND CONTROL

УДК 001.891.573:004.934.2

DOI: 10.17212/1814-1196-2018-3-21-32

Исследование методов классификации для автоматического определения истинной или ложной информации в речевых сообщениях*

А.Н. ВЕЛИЧКО^а, В.Ю. БУДКОВ^б, А.А. КАРПОВ^с

199178, РФ, г. Санкт-Петербург, 14-я линия, 39, Санкт-Петербургский институт информатики и автоматизации Российской академии наук

^а velichko.a.n@mail.ru ^б budkov@iias.spb.su ^с karpov@iias.spb.su

В статье представлены результаты экспериментальных исследований методов обработки речевого сигнала для паралингвистического анализа ложных и истинных речевых высказываний. Рассмотрены системы для определения типа информации, предложенные как в рамках международных соревнований по компьютерной паралингвистике Computational Paralinguistic Challenge ComParE-2016, так и вне этих соревнований. Также приведено описание речевой и многомодальной баз данных, содержащих ложные и истинные высказывания на английском языке: Deceptive Speech Database, в записи которой участвовали студенты Университета Аризоны, действовавшие согласно сценарию, и Real-Life Trial Deception Detection Dataset, которая была собрана в Университете Мичигана с использованием записей открытых судебных слушаний. На основе анализа и сравнения различных подходов к выявлению ложных и истинных речевых сообщений были определены методы и их параметры, показавшие наилучшие результаты относительно базовой системы, предложенной организаторами соревнований (которая достигла результата распознавания по критерию невзвешенного среднего значения полноты $UAR = 68,3\%$). В результате объединения двух баз данных был получен более сбалансированный и разнообразный набор данных для обучения системы объемом в 1253 аудиозаписи. С использованием открытого программного обеспечения (Ffmpeg, Praat, openSMILE, WEKA) был разработан прототип системы автоматического определения ложной информации в речи. Наилучшие результаты были достигнуты при использовании метода бэггинга совместно с методом k ближайших соседей ($UAR = 71,0\%$), отдельного метода k ближайших соседей ($UAR = 70,4\%$), а также метода классификации путем регрессии совместно с методом k ближайших соседей ($UAR = 69,5\%$). Разработанная система может быть полезна в правоохранительной и банковской сферах, при исследованиях с использованием полиграфа, для предотвращения «телефонного терроризма» и т. д.

Ключевые слова: компьютерная паралингвистика, речевые технологии, распознавание лжи в речи человека, машинное обучение, классификация, экспериментальные исследования, метод бэггинга, метод k -ближайших соседей

* Статья получена 30 апреля 2018 г.

Работа выполнена при поддержке РФФИ (проекты № 16-37-60085 и № 18-07-01407), Совета по грантам Президента РФ (грант № МД-254.2017.8) и бюджетной темы (№ 0073-2018-0002).

ВВЕДЕНИЕ

Паралингвистика занимается изучением различных невербальных аспектов речи человека, таких как эмоции, интонации, психофизиологические состояния, особенности голоса человека и т. д. Эмоциональное состояние человека тесно связано с его физиологическим состоянием. Идея распознавания ложной информации по речевому сигналу опирается на гипотезу о том, что при произнесении лжи человек испытывает стресс, что отражается на изменении параметров его речи. Несмотря на то что существующие методы экспертного определения лжи с использованием аппаратного полиграфа являются достаточно популярными, они имеют ряд недостатков в связи с тем, что требуют особых условий для работы – это распространяется как на рабочее место, так и на испытуемого. Стоит также отметить, что в различных культурах невербальные аспекты речи различаются [1]. Задача определения истинной или ложной информации в речевых сообщениях является актуальной, и многие современные исследования посвящены задаче разработки математического и программного обеспечения для автоматизированных систем распознавания лжи в речи человека. Автоматизированные методы и средства для анализа паралингвистических явлений в устной речи относятся к области компьютерной паралингвистики. В настоящее время наиболее распространенными методами моделирования и классификации являются искусственные нейронные сети, скрытые марковские модели, метод опорных векторов, последовательная минимальная оптимизация, метод k ближайших соседей, деревья принятия решений и др. [2].

1. ОПИСАНИЕ СУЩЕСТВУЮЩИХ ЭКСПЕРИМЕНТАЛЬНЫХ СИСТЕМ РАСПОЗНАВАНИЯ ЛОЖНОЙ ИНФОРМАЦИИ В РЕЧИ

В последние годы было представлено несколько экспериментальных систем, целью которых было распознавание ложной информации в речи человека. Большая часть таких систем была представлена на соревнованиях по компьютерной паралингвистике (Computational Paralinguistic Challenge) ComParE-2016, которые проводятся в рамках международной конференции INTERSPEECH с 2009 года. В 2016 году впервые на соревнованиях была представлена тема распознавания лжи в речи, где в качестве показателей оценки работы методов использовалась невзвешенная средняя полнота (Unweighted Average Recall, UAR [3]). Формула UAR выглядит следующим образом:

$$UAR = \frac{1}{k} \sum_{i=1}^K \frac{A_{ii}}{\sum_{j=1}^K A_{ij}},$$

где A – это матрица спутывания, A_{ij} – количество объектов класса i , классифицированных как j , а K – количество классов.

Участие принимали 20 команд из разных стран мира. Многие работы были посвящены выбору и работе с аудиопризнаками. Так, авторы работы [4] определяли лингвистические составляющие, а в работе [5] использовались акустико-просодические, лексические, синтаксические, фонотактические и базовый набор признаков, предоставленный организаторами. С помощью них были получены результаты распознавания ложных сообщений UAR 74,9 % и 62,2 % соответственно.

Некоторые исследователи полагают, что ложь может быть определена с помощью эмоциональных признаков. К примеру, в работах [6, 7] исследовались такие параметры, как эмоциональная насыщенность речи (тембр голоса, темп речи, громкость речи), лексические, грамматические и синтаксические признаки. В работе [7] авторы использовали интенсивность / силу эмоции, валентность / тон, эмоциональную регуляцию. Для системы авторы совместили такие методы, как kNN (k-Nearest Neighbors), SVM (Support Vector Machines), SMO (Sequential Minimal Optimization), и добились полноты UAR = 68,9 %.

Авторы работы [8] исследовали информативные признаки, они предложили высокоуровневый набор признаков и совместили его с базовым, а в системе использовали методы SVM и SMO, добившись полноты UAR = 69,3 %. В работе [9] было предложено использовать векторы Фишера для описания низкоуровневых признаков в высказываниях, а также каскадную нормализацию и нейросети ELM (Extreme Learning Machine), получив результат распознавания лжи UAR = 66,6 %.

Авторы работы [10] для решения проблемы определения ложной информации в речи исследовали несколько подходов машинного обучения. Они использовали акустические и лексические признаки для обучения модели глубокого обучения и получили результат F1 = 63,9 % (F1-мера). При использовании метода случайного леса, который был обучен с использованием объединенных акустических и лексических признаков, была получена точность распознавания 76,1 % (precision).

Также известен ряд работ по определению лжи в речи, выполненных вне соревнований по компьютерной паралингвистике ComParE-2016. В работе [11] рассматривалась психологическая сторона лжи. Автор считает, что дикторы могут испытывать различные эмоции в зависимости от причин, по которым люди лгут. В исследовании использовались деревья решений и 20 перекрестных проверок (cross-validation), при этом был получен результат распознавания UAR = 70 %. В работе [12] изучались свойства речи и было выявлено, что при произнесении лжи заметно ускоряется темп речи, уменьшается время начала ответа и длительность хезитаций. Авторы работы [13] использовали дробное преобразование Фурье для получения признаков FrCC, модель линейного дискриминантного анализа и скрытые марковские модели, получив различные результаты распознавания ложной информации в речи мужчин и женщин – 71 % и 70,2 % соответственно. В работах [14, 15] авторы совместно с аудиопризнаками использовали метаинформацию о дикторе (пол, этническая принадлежность, личностные факторы, родной язык). В ходе испытаний авторы работы [14] получили результат 61 % при использовании метода Adaboost. При этом в работе [15] выявлено, что метод случайного леса оказался наиболее точным по сравнению с деревом решений и бэггингом, он показал результат UAR = 65,86 %.

2. ОПИСАНИЕ ЭКСПЕРИМЕНТАЛЬНЫХ БАЗ ДАННЫХ, СОДЕРЖАЩИХ ЛОЖНЫЕ РЕЧЕВЫЕ СООБЩЕНИЯ

В качестве данных, для экспериментальных исследований были выбраны две базы данных:

1) Deceptive Speech Database [16], которая была представлена на соревновании по компьютерной паралингвистике ComParE-2016;

2) Real-Life Trial Deception Detection Dataset [17], созданная ранее в Университете Мичигана.

Речевой корпус Deceptive Speech Database (DSD) был разработан в Университете Аризоны (США). Этот корпус содержит аудиозаписи общей длительностью 162 минуты речи на английском языке 72 различных дикторов. В записи аудиоданных принимали участие студенты университета, они были разделены на две группы, участники которых должны были играть роли, определенные сценарием. Роль лжецов играли участники первой группы, они «украли» вопросы и ответы для экзамена из компьютера на кафедре. Участники второй группы играли роль честных студентов, которые вернули документ в тот же кабинет. С каждым участником были проведены интервью, которые состояли из набора открытых вопросов, подразумевающих короткие ответы (в основном «да» или «нет»). Основные параметры этой базы представлены в табл. 1. Длительность отдельных высказываний информантов варьируется от 0,1 до 236 секунд.

Таблица 1

Table 1

Параметры речевого корпуса DSD
Parameters of the speech database DSD

Корпус	Обучение	Разработка	Тестирование
Кол-во дикторов (женщины, мужчины) (виновен, не виновен)	26 (11, 15) (14, 12)	23 (11, 12) (10, 13)	– – –
Кол-во аудиофайлов (виновен, не виновен) (ложные, истинные)	572 (308, 264) (182, 390)	487 (220, 267) (130, 357)	497 – –
Длительность взаимодействия (секунд) Среднее: мин.-макс.	7.6: 0.4...220	6.3: 1.1...236	6.5: 0.1...220
Длительность речевых сообщений (секунд) Среднее: мин.-макс.	5.1: 0.2...214	4.0: 0.1...227	3.8: 0.1...211
Кол-во фонем Среднее: мин.-макс.	32.2: 1...1480	24.5: 1...1330	23.9: 1...1268

Второй речевой корпус Real-Life Trial Deception Detection Dataset (RLTDDD) был создан в 2016 году в Университете Мичигана (США). Для сбора данных были использованы англоязычные ресурсы, предоставляющие записи судебных слушаний, на которых можно определить поведение человека при произнесении ложной и правдивой информации. На аудио- и видеозаписях были точно определены подсудимый и свидетель; достаточно четко видны их лица на протяжении большей части записи; качество видео достаточно хорошее, чтобы определить мимику; также качество аудиозаписей должно быть настолько хорошим, чтобы слышать голоса и понимать, что говорят люди. Аудио- и видеозаписи были отмечены экспертом как «лжи-

вые» и «правдивые» на основании вердикта суда: виновен, не виновен или оправдан. Если вердикт «виновен», то аудио- и видеозаписи, содержащие ложную информацию, состоят из записей подсудимых, а видео, содержащие правдивую информацию – из записей свидетелей в тех же слушаниях. База данных содержит 121 запись судебных заседаний, включая 61 лживую и 60 правдивых. Средняя продолжительность видеозаписи по всей базе данных составляет 28 сек., 27,7 сек. и 28,3 сек. для ложных и правдивых высказываний, соответственно. Данные включают аудио- и видеозаписи 21 женщины и 25 мужчин, возраст информантов находится между 16 и 60 годами.

3. ОПИСАНИЕ ПРОГРАММНОЙ СИСТЕМЫ

Общая схема разработанной программной системы для анализа речевой информации представлена на рис. 1. В экспериментальной системе используется следующее открытое программное обеспечение:

1) Ffmpeg – для записи, конвертирования, передачи цифровых аудио- и видеозаписей в различных форматах (<https://www.ffmpeg.org/>);

2) Praat – программное средство для фонетического анализа речи (<http://www.fon.hum.uva.nl/praat/>);

3) openSMILE – программный инструмент для вычисления расширенного набора акустических низкоуровневых признаков (LLD), который состоит из 6373 супrasegmentных признаков, включающих 65 базовых низкоуровневых признаков и их функционалы, эти низкоуровневые признаки включают в себя множество характеристик (спектральных, кепстральных, энергетических и вокализованных) (<http://audeering.com/technology/opensmile/>);

4) WEKA (Waikato Environment for Knowledge Analysis) – набор программных средств визуализации и методов для анализа данных, классификации и прогнозирования [18].

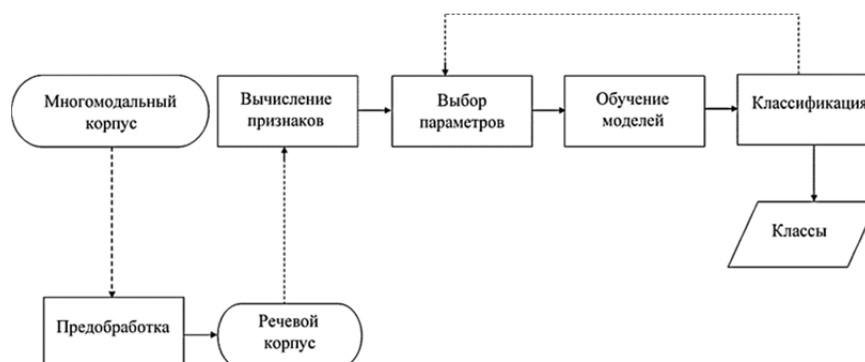


Рис. 1. Общая схема разработанной экспериментальной системы анализа аудиоданных

Fig. 1. A general diagram of the developed experimental system for the analysis of audio data

В данном исследовании проводится работа только с аудиоданными, которые были извлечены из многомодального корпуса RLTDDB. Далее они были обработаны при помощи Praat, где экспертом были вырезаны

фразы «интервьюеров». Таким образом было получено 195 аудиозаписей, которые были помечены согласно информации, представленной в аннотации к базе данных. Всего для исследования были использованы 1253 аудиозаписи.

Вычисление аудиопризнаков проводилось с использованием программного инструментария openSMILE (конфигурация набора признаков для соревнований по компьютерной паралингвистике ComParE 2013 года [19]). Выходными данными являются файлы формата Attribute-Relation File Format (ARFF), в которых сначала определяется тип загружаемых данных (метаданные), а потом указываются сами данные в виде векторов признаков.

На последнем этапе используется программный инструментарий классификации WEKA, куда загружаются признаки, вычисленные ранее. Для проведения экспериментов были выбраны следующие методы классификации данных, реализованные в WEKA: 1) SMO (Sequential Minimal Optimization) [20]; 2) kNN (k-Nearest Neighbor); 3) Bagging; 4) Decision Tree; 5) Random Forest; 6) Adaboost [21]; 7) SGD (Stochastic Gradient Descent) [22]; 8) OneR (One Rule); 9) PART (PART decision list) [23]; 10) Random Tree; 11) CvR (Classification via Regression) [24].

4. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Для проведения экспериментов были использованы методы перекрестной проверки (N-fold cross-validation, в данном случае была выбрана 10-кратная перекрестная проверка) и межкорпусный анализ (cross-corpus analysis). Первый метод заключается в том, что исходная выборка N раз (N-fold cross-validation) разбивается на две подвыборки: обучающую и тестовую. Затем для каждого разбиения производится обучение модели, а точность классификации оценивается на объектах тестовой выборки и затем усредняется по всем N экспериментам. Был также использован межкорпусный анализ, т. е. обучение методов происходило на одной речевой базе, а тестировалось на другой.

В качестве показателей оценки работы методов использовалась точность распознавания (Accuracy) и невзвешенная средняя полнота UAR. Эта мера используется для данных с несбалансированными классами и является средним значением полноты, вычисляемым для каждого класса. Так, результаты перекрестной проверки представлены в табл. 2 (где Асс. – точность, UAR – полнота), а результаты межкорпусного анализа – в табл. 3.

В ходе перекрестной проверки в основном были использованы стандартные параметры методов классификации WEKA, однако после проведения дополнительных исследований параметры некоторых классификаторов (Bagging, CvR, Adaboost, kNN) были изменены (табл. 4).

Из табл. 2 можно видеть, что наилучшие результаты по критерию качества распознавания на совмещенном наборе данных показали следующие методы: бэггинг с методом k ближайших соседей (UAR = 71,0 %), затем метод k ближайших соседей, а также метод классификации путем регрессии с методом k ближайших соседей (UAR = 69,55 %).

Таблица 2

Table 2

**Результаты сравнения методов паралингвистического анализа аудиоданных
для распознавания истинности / ложности речевых сообщений**

**Results of comparison of paralinguistic analysis methods of audio data for the recog-
nition of truth/deception of spoken messages**

Метод	База данных DSD		База данных RLTD		Совместно DSD + RLTD	
	Acc., %	UAR, %	Acc., %	UAR, %	Acc., %	UAR, %
Bagging + kNN	70,3	66,8	76,3	76,2	72,8	71,0
kNN	70,4	67,0	76,3	76,3	71,4	70,4
CvR + kNN	70,4	67,0	76,3	76,3	71,4	69,6
Adaboost + SMO	76,4	68,4	69,0	68,9	71,9	68,5
SMO	76,4	70	69,0	68,9	71,9	68,5
SGD	76,3	70	66,5	66,4	71,8	68,3
Random Forest	72,5	55,2	79,4	79,3	71,4	59,6
Decision Tree	63,0	54,3	69,6	69,5	63,9	58,2
Random Tree	61,3	55,3	63,9	63,7	64,8	58,2
PART	64	56,2	66,5	66,4	63,6	58,2
OneR	65,2	51,8	74,7	74,6	59,4	50,5

Таблица 3

Table 3

Межкорпусный анализ

Database-to-database analysis

Метод	DSD → RLTD		RLTD → DSD	
	Acc., %	UAR, %	Acc., %	UAR, %
Bagging	48,5	50,0	60,1	50,6
kNN	42,3	42,2	52,9	47,2
CvR	50,0	50,9	67,9	48,6
Adaboost	47,4	50,0	69,6	54,9
SMO	52,0	53,8	58,9	49,2
SGD	46,4	48,7	54,4	49,4
Random Forest	46,4	48,9	49,0	44,4
Decision Tree	45,9	46,9	45,0	48,6
Random Tree	49,0	49,3	61,9	54,4
PART	46,9	48,9	42,3	47,7
OneR	47,9	49,8	55,7	47,7

Таблица 4

Table 4

Сравнение изменений в настройках методов классификации**Comparison of changes in settings of the classification methods**

Метод классификации	Стандартные параметры	Измененные параметры
Bagging	Обучение на классификаторе REPTree	Обучение на классификаторе kNN
CvR	Обучение на классификаторе M5P	Обучение на классификаторе kNN
Adaboost	Обучение на классификаторе DecisionStump	Обучение на классификаторе SMO
kNN (в экспериментах с объединенными данными)	Количество k соседей = 1	Количество k соседей = 4

ЗАКЛЮЧЕНИЕ

В статье представлены результаты экспериментального исследования по автоматическому распознаванию истинной / ложной информации в речи. Для экспериментов были выбраны наиболее популярные и эффективные методы классификации и оптимизированы их параметры, при которых данные методы показывают наилучшую точность / полноту распознавания. Эксперименты проводились на материалах, представленных в двух речевых базах данных Deceptive Speech Database и Real-Life Trial Detection Deception Dataset. При этом для работы со вторым набором данных необходимо было предварительно извлечь аудиоданные из мультимедийного корпуса и разметить их.

Приведены результаты работы и сравнения методов классификации как отдельно на каждом наборе данных, так и на объединенном наборе. Наилучшие результаты по критерию качества распознавания на совмещенном наборе данных показали следующие методы: бэггинг с методом k ближайших соседей (UAR = 71,0 %), метод k ближайших соседей (UAR = 70,4 %), а также метод классификации путем регрессии с методом k ближайших соседей (UAR = 69,55 %). Можно заметить, что объединение двух баз данных положительно сказалось на результатах классификации, поскольку удалось расширить и разнообразить обучающий набор.

Предложенный прототип программной системы в дальнейшем возможно использовать для автоматического распознавания правдивой / ложной информации в речи человека в таких сферах, как биометрические исследования на полиграфе, предотвращение телефонного «терроризма»; в банковской сфере в ходе проведения интервью для рассмотрения вопросов о выдаче кредитов и т. д.

СПИСОК ЛИТЕРАТУРЫ

1. Potapova R., Lykova O. Verbal representation of lies in Russian and Anglo-American cultures // *Procedia – Social and Behavioral Sciences*. – 2016. – Vol. 236. – P. 114–118.
2. Величко А.Н., Будков В.Ю., Карпов А.А. Аналитический обзор компьютерных паралингвистических систем для автоматического распознавания лжи в речи человека // *Информационно-управляющие системы*. – 2017. – № 5. – С. 30–41.
3. Eyben F. Real-time speech and music classification by large audio feature space extraction. – Springer, 2016. – 298 p.
4. Montacié C., Caraty M.-J. Prosodic cues and answer type detection for the deception sub-challenge // *Proceedings of INTERSPEECH-2016*. – San Francisco, USA, 2016. – P. 2016–2020.
5. Combining acoustic-prosodic, lexical, and phonotactic features for automatic deception detection / S.I. Levitan, G. An, M. Ma, R. Levitan, A. Rosenberg, J. Hirschberg // *Proceedings of INTERSPEECH-2016*. – San Francisco, USA, 2016. – P. 2006–2010.
6. Савченко В.В., Васильев Р.А. Анализ эмоционального состояния диктора по голосу на основе фонетического детектора лжи // *Научные ведомости Белгородского государственного университета*. – 2014. – № 21 (192). – С. 186–195.
7. Is deception emotional? An emotion-driven predictive approach / S. Amiriparian, J. Pohjalainen, E. Marchi, S. Pugachevskiy, B. Schuller // *Proceedings of INTERSPEECH-2016*. – San Francisco, USA, 2016. – P. 2011–2015.
8. Herms R. Prediction of deception and sincerity from speech using automatic phone recognition-based features // *Proceedings of INTERSPEECH-2016*. – San Francisco, USA, 2016. – P. 2036–2040.
9. Kaya H., Karpov A. Fusing acoustic feature representations for computational paralinguistics tasks // *Proceedings of INTERSPEECH-2016*. – San Francisco, USA, 2016. – P. 2046–2050.
10. Hybrid acoustic-lexical deep learning approach for deception detection / G. Mendels, S.I. Levitan, K. Lee, J. Hirschberg // *Proceedings of INTERSPEECH-2017*. – Stockholm, Sweden, 2017. – P. 1472–1476.
11. Hirschberg J. Detecting deceptive speech: requirements, resources and evaluation [Electronic resource] // *Proceedings of LREC-2008*. – Marrakech, Morocco, 2008. – URL: <http://www.lrec-conf.org/proceedings/lrec2008/keynotes/Hirschberg.pdf> (accessed: 11.09.2018).
12. Kirchhubel C., Stedmon A., Howard D.M. Analyzing deceptive speech // *Proceedings of EPCE-2013*. – Heidelberg: Springer, 2013. – Pt. 1. – P. 134–141. – (Lecture notes in computer science; vol. 8019). – doi: 10.1007/978-3-642-39360-0_15.
13. Pan X., Zhao H., Zhou Y. The application of fractional Mel cepstral coefficient in deceptive speech detection [Electronic resource] // *PeerJ*. – 2015. – 18 August. – URL: <https://doi.org/10.7717/peerj.1194> (accessed: 11.09.2018).
14. Identifying individual differences in gender, ethnicity, and personality from dialogue for deception detection / S.I. Levitan, Y. Levitan, G. An, M. Levine, A. Rosenberg, R. Levitan, J. Hirschberg // *Proceedings of NAACL-HLT-2016*. – San Diego, USA, 2016. – P. 40–44.
15. Cross-cultural production and detection of deception from speech / S.I. Levitan, G. An, M. Wang, G. Mendels, J. Hirschberg, M. Levine, A. Rosenberg // *Proceedings of ACM Workshop on Multimodal Deception Detection*. – Seattle, USA, 2015. – P. 1–8.
16. The INTERSPEECH 2016 Computational paralinguistics challenge: deception, sincerity & native language / B. Schuller, S. Steidl, A. Batliner, J. Hirschberg, J.K. Burgoon, A. Baird, A. Elkins, Y. Zhang, E. Coutinho, K. Evanini // *Proceedings of INTERSPEECH-2016*. – San Francisco, USA, 2016. – P. 2001–2005.
17. Deception detection using real-life trial data / V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, M. Burzo // *Proceedings of ACM International Conference on Multimodal Interaction*. – Seattle, USA, 2015. – P. 59–66.
18. Frank E., Hall M.A., Witten I.H. The WEKA workbench: online appendix for "Data mining: practical machine learning tools and techniques". – 4th ed. – Burlington, MA: Morgan Kaufmann, 2016.
19. The INTERSPEECH 2013 Computational paralinguistics challenge: social signals, conflict, emotion, autism / B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval,

M. Chetouani, F. Weninger, F. Eyben, K. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente, S. Kim // Proceedings of INTERSPEECH-2013. – Lyon, France, 2013. – P. 148–152.

20. Platt J.C. Sequential minimal optimization: a fast algorithm for training support vector machines. Technical report MSR-TR-98-14. – Redmond: Microsoft Research, 1998. – 21 p.

21. Freund Y., Schapire R.E. A decision-theoretic generalization of on-line learning and an application to boosting // Journal of Computer and System Sciences. – 1997. – Vol. 55. – P. 119–139.

22. Kiefer J., Wolfowitz J. Stochastic estimation of the maximum of a regression function // Annals of Mathematical Statistics. – 1952. – Vol. 23 (3). – P. 462–466.

23. Frank E., Witten I.H. Generating accurate rule sets without global optimization // ICML '98 Proceedings of the Fifteenth International Conference on Machine Learning. – San Francisco, CA, USA, 1998. – P. 144–151.

24. Kukreja M., Johnson S.A., Stafford P. Comparative study of classification algorithms for immunosignaturing data // BMC Bioinformatics. – 2012. – Vol. 13. – P. 139.

Величко Алёна Николаевна, программист лаборатории речевых и многомодальных интерфейсов Санкт-Петербургского института информатики и автоматизации Российской академии наук. Основное направление научных исследований – определение ложной информации в речи. Имеет 2 публикации. E-mail: velichko.a.n@mail.ru

Будков Виктор Юрьевич, кандидат технических наук, старший научный сотрудник лаборатории автономных робототехнических систем Санкт-Петербургского института информатики и автоматизации Российской академии наук. Основное направление научных исследований – многомодальные человеко-машинные интерфейсы, взаимодействие с роботами. Имеет более 50 публикаций. E-mail: budkov@iias.spb.su

Карпов Алексей Анатольевич, доктор технических наук, доцент, заведующий лабораторией речевых и многомодальных интерфейсов Санкт-Петербургского института информатики и автоматизации Российской академии наук. Основное направление научных исследований – разработка и исследования в области автоматического распознавания и понимания речи, аудиовизуальной обработки (анализ и синтез) речи, многомодальные человеко-машинные интерфейсы, компьютерной паралингвистики. Имеет более 250 публикаций. E-mail: karpov@iias.spb.su

DOI: 10.17212/1814-1196-2018-3-21-32

Study of classification methods for automatic truth and deception detection in speech^{*}

A.N. VELICHKO^a, V.Y. BUDKOV^b, A.A. KARPOV^c

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), 39, 14th Line, St. Petersburg, Russia

^a velichko.a.n@mail.ru ^b budkov@iias.spb.su ^c karpov@iias.spb.su

Abstract

In this paper we present experimental results of research on speech processing methods for paralinguistic analysis of deceptive and truthful speech utterances. It includes a short survey of recently developed deception detection systems that were used in the framework of the Computational Paralinguistics Challenge ComParE-2016 and in other situations. In addition, we present speech and multimodal databases that contain both truthful and deceptive utterances

^{*} Received 30 April 2018.

The work was supported by the RFBR (projects № 16-37-60085 and №18-07-01407), as well as by the grant of the President of Russia (№ MD-254.2017.8) and state research (№. 0073-2018-0002).

in English: Deception Speech Database that was collected in the University of Arizona and consists of audio recordings of students' speech (they acted according to the scenario) and Real-Life Trial Deception Detection Dataset that was collected in the University of Michigan using recordings of real-life trial videos that are publicly available. Based on this analysis and comparison of several various methods with optimal parameters the best methods and parameters were found. These methods showed better results than the baseline system proposed for the competition (its performance in terms of Unweighted Average Recall measure UAR was 68.3 %). The dataset for training was obtained by combining both databases. Thereby it is more balanced and manifold (the total number of audio files was 1253). An automatic system for deception detection was developed using these methods and open source software (Ffmpeg, Praat, openSMILE, WEKA). The highest performance was obtained using a combination of Bagging and k-Nearest Neighbor classifiers with UAR = 71.0 %. Also, high results were obtained using a single k-Nearest Neighbor classifier (UAR = 70.4 %), and a combination of Classification via Regression with k-Nearest Neighbor classifiers (UAR = 69.5 %). The developed system may be used for law enforcement and banking areas, biometric research with polygraph devices, prevention of telephone terrorism, etc.

Keywords: Computational paralinguistics, speech technology, deception detection in speech, machine learning, classification, experimental study, bagging, k-Nearest Neighbor

REFERENCES

1. Potapova R., Lykova O. Verbal representation of lies in Russian and Anglo-American cultures. *Procedia – Social and Behavioral Sciences*, 2016, vol. 236, pp. 114–118.
2. Velichko A.N., Budkov V.Yu., Karpov A.A. Analiticheskii obzor komp'yuternykh paralingvisticheskikh sistem dlya avtomaticheskogo raspoznavaniya lzhi v rechi cheloveka. *Informatsionno-upravlyayushchie sistemy – Information and Control Systems*, 2017, vol. 90, no. 5, pp. 30–41.
3. Eyben F. *Real-time speech and music classification by large audio feature space extraction*. Springer, 2016. 298 p.
4. Montacié C., Caraty M.-J. Prosodic cues and answer type detection for the deception sub-challenge. *Proceedings of INTERSPEECH-2016*, San Francisco, USA, 2016, pp. 2016–2020.
5. Levitan S.I., An G., Ma M., Levitan R., Rosenberg A., Hirschberg J. Combining acoustic-prosodic, lexical, and phonotactic features for automatic deception detection. *Proceedings of INTERSPEECH-2016*, San Francisco, USA, 2016, pp. 2006–2010.
6. Savchenko V.V., Vasil'ev R.A. Analiz emotsional'nogo sostoyaniya diktora po golosu na osnove foneticheskogo detektora lzhi [The analysis of the emotional condition of the an-nouncer on the voice on the basis of the phonetic lie detector]. *Nauchnye vedomosti Belgorodskogo gosudarstvennogo universiteta – Belgorod State University Scientific Bulletin*, 2014, vol. 21, no. 192, pp. 186–195. (In Russian).
7. Amiriparian S., Pohjalainen J., Marchi E., Pugachevskiy S., Schuller B. Is deception emotional? An emotion-driven predictive approach. *Proceedings of INTERSPEECH-2016*, San Francisco, USA, 2016, pp. 2011–2015.
8. Herms R. Prediction of deception and sincerity from speech using automatic phone recognition-based features. *Proceedings of INTERSPEECH-2016*, San Francisco, USA, 2016, pp. 2036–2040.
9. Kaya H., Karpov A. Fusing acoustic feature representations for computational paralinguistics tasks. *Proceedings of INTERSPEECH-2016*, San Francisco, USA, 2016, pp. 2046–2050.
10. Mendels G., Levitan S.I., Lee K., Hirschberg J. Hybrid acoustic-lexical deep learning approach for deception detection. *Proceedings of INTERSPEECH-2017*, Stockholm, Sweden, 2017, pp. 1472–1476.
11. Hirschberg J. Detecting deceptive speech: requirements, resources and evaluation. *Proceedings of LREC-2008*, Marrakech, Morocco, 2008. Available at: <http://www.lrec-conf.org/proceedings/lrec2008/keynotes/Hirschberg.pdf> (accessed 11.09.2018).

12. Kirchhubel C., Stedmon A., Howard D.M. Analyzing deceptive speech. *Proceedings of EPCE-2013*. Heidelberg, Springer, 2013, pt. 1, pp. 134–141. doi: 10.1007/978-3-642-39360-0_15.
13. Pan X., Zhao H., Zhou Y. The application of fractional Mel cepstral coefficient in deceptive speech detection. *PeerJ*, 2015, 15 August. Available at: <https://doi.org/10.7717/peerj.1194> (accessed 11.09.2018).
14. Levitan S.I., Levitan Y., An G., Levine M., Rosenberg A., Levitan R., Hirschberg J. Identifying individual differences in gender, ethnicity, and personality from dialogue for deception detection. *Proceedings of NAACL-HLT-2016*, San Diego, USA, 2016, pp. 40–44.
15. Levitan S.I., An G., Wang M., Mendels G., Hirschberg J., Levine M., Rosenberg A. Cross-cultural production and detection of deception from speech. *Proceedings of ACM Workshop on Multimodal Deception Detection*, Seattle, USA, 2015, pp. 1–8.
16. Schuller B., Steidl S., Batliner A., Hirschberg J., Burgoon J.K., Baird A., Elkins A., Zhang Y., Coutinho E., Evanini K. The INTERSPEECH 2016 Computational paralinguistics challenge: deception, sincerity & native language. *Proceedings of INTERSPEECH-2016*, San Francisco, USA, 2016, pp. 2001–2005.
17. Pérez-Rosas V., Abouelenien M., Mihalcea R., Burzo M. Deception detection using real-life trial data. *Proceedings of ACM International Conference on Multimodal Interaction*, Seattle, 2015, pp. 59–66.
18. Frank E., Hall M.A., Witten I.H. *The WEKA workbench: online appendix for "Data mining: practical machine learning tools and techniques"*. 4th ed. Burlington, MA, Morgan Kaufmann, 2016.
19. Schuller B., Steidl S., Batliner A., Vinciarelli A., Scherer K., Ringeval F., Chetouani M., Wenginger F., Eyben F., Marchi K., Mortillaro M., Salamin H., Polychroniou A., Valente F., Kim S. The INTERSPEECH 2013 computational paralinguistic challenge: social signals, conflict, emotion, autism. *Proceedings of INTERSPEECH-2013*, Lyon, France, 2013, pp. 148–152.
20. Platt J.C. *Sequential minimal optimization: a fast algorithm for training support vector machines*. Technical report MSR-TR-98-14. Redmond, Microsoft Research, 1998. 21 p.
21. Freund Y., Schapire R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 1997, vol. 55, pp. 119–139.
22. Kiefer J., Wolfowitz J. Stochastic estimation of the maximum of a regression function. *Annals of Mathematical Statistics*, 1952, vol. 23, no. 3, pp. 462–466.
23. Frank E., Witten I.H. Generating accurate rule sets without global optimization. *ICML '98 Proceedings of the Fifteenth International Conference on Machine Learning*, San Francisco, CA, USA, 1998, pp. 144–151.
24. Kukreja M., Johnson S.A., Stafford P. Comparative study of classification algorithms for immunosignaturing data. *BMC Bioinformatics*, 2012, vol. 13, p. 139.

Для цитирования:

Величко А.Н., Будков В.Ю., Карпов А.А. Исследование методов классификации для автоматического определения истинной или ложной информации в речевых сообщениях // Научный вестник НГТУ. – 2018. – № 3 (72). – С. 21–32. – doi: 10.17212/1814-1196-2018-3-21-32.

For citation:

Velichko A.N., Budkov V.Yu., Karpov A.A. Issledovanie metodov klassifikatsii dlya avtomaticheskogo opredeleniya istinnoi ili lozhnoi informatsii v rechevykh soobshcheniyakh [Study of classification methods for automatic truth and deception detection in speech]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Science bulletin of the Novosibirsk state technical university*, 2018, no. 3 (72), pp. 21–32. doi: 10.17212/1814-1196-2018-3-21-32.