



УДК 004.93

© 2018 г. **А.В. Горошевский,**
А.Г. Шишкин, д-р физ.-мат. наук
(Московский государственный университет им. М.В. Ломоносова)

АВТОМАТИЧЕСКАЯ СЕГМЕНТАЦИЯ СМЕХА В РЕЧЕВЫХ СИГНАЛАХ

В работе исследуется вопрос применения некоторых методов машинного обучения для отделения фрагментов смеха от речи в речевых сигналах. Проводится тестирование реализаций систем на основе предложенных методов. **Ключевые слова:** распознавание речи, машинное обучение, метод опорных векторов, глубокие сети доверия.

DOI: 10.22250/isu.2018.56.83-95

1. Введение

По мере совершенствования систем автоматического распознавания речи все большее значение приобретает возможность корректной обработки элементов звукового потока, отличных от слов. Одной из таких областей исследования, позволяющей весьма значительно повысить качество работы распознающей системы, является изучение и моделирование невербальных звуковых единиц, к числу которых принадлежит смех. Правильное распознавание и выделение смеха из аудиопотока для его исключения из дальнейшей обработки будет способствовать уменьшению числа ошибок как при распознавании слов, так и при диаризации речевого сигнала.

Помимо того, смех является одним из важнейших элементов человеческого общения, позволяющим делать определенные выводы о степени и характере эмоциональной окрашенности речи. На основе смеха из речевого сигнала может быть выделен ряд характерных признаков для определения психоэмоционального состояния человека, что находит применение, например, при мониторинге операторов, связанных с обслуживанием опасных производств. Отметим, что смех представляет собой чрезвычайно изменчивый акустический сигнал даже для одного и того же диктора. Это приводит к тому, что моделирование и детектирование сме-

ха является чрезвычайно сложной задачей.

В настоящее время системы распознавания речи, как правило, включают смех в свои словари в качестве еще одного слова. Однако не будучи предназначены для распознавания смеха, указанные системы не настроены оптимальным образом для его выделения из речевого потока, что приводит к большому числу ошибок и увеличенному расходу вычислительных мощностей. Поэтому для построения робастной системы автоматического распознавания речи, а также систем, определяющих психоэмоциональное состояние пользователей, необходимо иметь возможность с высокой степенью точности производить сегментацию смеха в исходном сигнале. Данная задача до сих пор является чрезвычайно актуальной, так как полученные к настоящему моменту результаты не позволяют назвать ее решенной.

В русскоязычной литературе задача выявления смеха в речевом сигнале, насколько нам известно, до настоящего момента не рассматривалась, однако была затронута более общая проблема определения эмоционального состояния диктора. При этом наиболее популярным является двухэтапный алгоритм: выделение основных характерных признаков звукового сигнала на пересекающихся небольших фрагментах, а затем решение задачи классификации, например, с помощью метода опорных векторов (support vector machine) [1, 2]. В одной из недавних работ был представлен способ классификации эмоций посредством обработки характерных признаков фрагментов сигнала с помощью глубоких нейросетей с дальнейшей постобработкой [3].

В зарубежной литературе в последние годы появился ряд работ, посвященных распознаванию смеха в речевых сигналах [4 – 6]. В работе [4] рассматривается подход с выделением характерных признаков из предварительно сегментированного аудиосигнала и применение различных методов классификации, в том числе метода опорных векторов (МОВ). В работах [5, 6] сегментация смеха в речевом потоке выполнена с помощью нейросетей с одним скрытым слоем [5] и глубоких сетей с выпрямителем (deep rectifier neural networks) [6], также осуществляющих классификацию на основе предварительно выделенных характерных признаков. Однако полученные данными методами результаты не позволяют надежным образом осуществлять выделение смеха в аудиопотоке.

Целью данной работы является разработка эффективной системы для сегментации смеха в речевом сигнале, что подразумевает автоматическое нахождение начальных и конечных моментов времени смеха. Для выделения смеха используются два различных классификатора. Один основан на методе опорных векторов, а другой – на применении глубоких нейронных сетей. В отличие от работ [4 – 6] глубокие нейронные сети используются для самостоятельного выделения признаков и классификации по ним, принимая на вход лишь спектр сигнала.

Структура данной работы следующая. В разделе 2 представлены используемые алгоритмы и их параметры. В разделе 3 описаны наборы данных, на которых проводилось обучение и тестирование систем классификации. Разделы 4 и 5 посвящены тестированию разработанных систем, обсуждению результатов и сравнению эффективности используемых алгоритмов. В разделе 6 представлено заключение и обоснованы направления для дальнейших исследований.

2. Используемые алгоритмы

Задача выделения смеха в речевом сигнале подразумевает нахождение начальной и конечной границ смеха. Таким образом, для каждого фрагмента сигнала необходимо определить, содержит ли он смех, речь, паузу или любую их комбинацию. Для решения данной задачи могут быть использованы различные методы. Наиболее распространенным является подход, при котором сначала выделяются характерные признаки сигнала, а затем используется какой-либо из методов классификации. Одним из них является метод опорных векторов, хорошо зарекомендовавший себя при решении задач классификации и регрессионного анализа. Вторым используемым в данной работе методом являются глубокие сети доверия.

2.1. Метод опорных векторов

Метод опорных векторов [7] относится к числу алгоритмов обучения с учителем. Он заключается в построении оптимальной разделяющей гиперплоскости. Под оптимальной мы будем понимать гиперплоскость, которая перпендикулярна кратчайшему отрезку, соединяющему выпуклые оболочки разных классов, и проходит через середину этого отрезка. Иными словами, оптимальная гиперплоскость должна максимизировать ширину разделяющей полосы между классами.

Для применения метода опорных векторов из звукового сигнала был выделен ряд характерных признаков, подробнее о которых рассказано далее. К полученным векторам был применен метод главных компонент для уменьшения их размерности и ускорения расчетов, а затем проведена классификация данных.

2.1.1. Характерные признаки.

Частота основного тона.

Наиболее известной характеристикой речевого сигнала, использующейся практически при всех видах его анализа, является частота основного тона. Данный характерный признак представляет собой обычную частотную модуляцию сигнала, параметры которой могут быть измерены. На практике установлено, что частота основного тона разных людей (мужчин, женщин, детей) изменяется в весьма широких пределах: от 75 Гц для низких мужских голосов до 400 Гц – для высоких детских.

В настоящей работе для выделения частоты основного тона использовался метод YIN в спектральном окне. Он был описан в работе [8] как модификация алгоритма YIN [9], нацеленная на существенное ускорение алгоритма при одновременном незначительном снижении его точности.

Итоговая формула функции разности через автокорреляционную функцию и энергию сигнала выглядит следующим образом:

$$d_t(\tau) = \frac{2}{N} \sum_{k=0}^{N/2+1} |(1 - e^{2j\pi k\tau/N})X_t[k]|^2,$$

где $X_t[k]$ – Фурье-преобразование исследуемого фрагмента сигнала; N – количество отсчетов в этом фрагменте; τ – период основного тона сигнала, выраженный в количестве отсчетов.

Дальнейшие действия производятся в соответствии с первоначальным алгоритмом YIN [9].

Форманты звукового сигнала.

При анализе речевых сигналов большую роль играют форманты речи. Их можно определить как концентрации энергии сигнала в ограниченной частотной области. Как и большинство основных компонент речевого сигнала, форманты характеризуются частотой, шириной и амплитудой [10].

При рассмотрении формант в данной работе наиболее важными параметрами представляются их частоты. Их предлагается находить с помощью отыскания коэффициентов линейного предсказания и последующим вычислением через них пиков спектра. Такой подход эффективен и точен по сравнению с вычислением формант по спектру звукового сигнала, полученному в результате преобразования Фурье [11].

Коэффициенты линейного предсказания являются коэффициентами знаменателя передаточной функции, описывающей действие речевого тракта, формы сигнала возбуждения и излучения [12]. Таким образом, располагая совокупностью параметров предсказания, можно определить частотную характеристику модели речеобразования следующим образом:

$$H(e^{i\omega}) = \frac{1}{1 - \sum_{k=1}^p \alpha_k e^{-i\omega_k}},$$

где ω – частота; α_k – коэффициенты линейного предсказания, которые можно найти с помощью матрицы, составленной из кратковременных автокорреляционных функций сигнала или с помощью рекурсии Левинсона – Дурбина [13].

Мел-частотные кепстральные коэффициенты.

Как известно, речеобразование зависит в первую очередь от формы голосового тракта, определяющей, какие звуки произносятся в данный момент. Именно

поэтому усилия исследователей всегда были сосредоточены на разработке характерных признаков речи, в наибольшей степени отвечающих изменениям формы голосового тракта. Форма голосового тракта отражается, например, в огибающей кратковременного спектра мощности речевого сигнала. Для представления этой огибающей с помощью небольшого объема данных предназначены мел-частотные кепстральные коэффициенты.

Мел – единица высоты звука, основанная на восприятии этого звука органами слуха. Зависимость между частотой звука и мелами можно описать простой формулой

$$B(f) = 1125 \ln(1 + f / 700).$$

Подобные единицы измерения часто используют при решении задач распознавания, так как позволяют приблизиться к механизмам человеческого восприятия речи [14].

Другие характерные признаки.

В качестве дополнительного характерного признака находилась полная энергия сигнала – автокорреляционная функция с нулевым смещением:

$$P(n) = \sum_{i=0}^{N-1} S[n+i]S[n+i],$$

где S – множество отсчетов сигнала, а N – их количество.

Также мы применяли энергию Тигера [15], которая использует соседние отсчеты и хорошо зарекомендовала себя на практике:

$$T(n) = \sum_{i=0}^{N-1} (S[n+i]S[n+i] - S[n+i-1]S[n+i+1]).$$

Наряду с энергиями можно рассмотреть такой признак как число переходов через нуль, которое вычисляется по формуле:

$$Z = \sum_{n=0}^{N-2} \frac{1 - \text{sgn}(S[n])\text{sgn}(S[n+1])}{2}.$$

С помощью сравнения энергии Тигера и числа переходов через нуль со средними значениями по сигналу можно приблизительно определить количество невокализованных фрагментов сигнала (низкий $T(k)$, высокий $Z(k)$, где k – индекс фрагмента), их отношение к общему числу фрагментов и среднюю продолжительность, которые тоже можно использовать как характерные признаки [15].

Помимо описанных характерных признаков, в данной работе были использованы время нахождения сигнала в определенных полосах спектра (окнах одинаковой размерности, сконцентрированных на частотах до 8 кГц) и вариация огибающей спектра.

После выделения названных выше характеристик сигнала размерность векторов характерных признаков для каждого из фрагментов сигнала оказалась рав-

ной 124. Для уменьшения размерности векторов с помощью метода главных компонент [16] были отобраны 25 наиболее существенных из них.

2.2. Глубокие сети доверия

Второй метод, используемый в данной работе для классификации речевых фрагментов, основан на применении глубоких нейронных сетей. Их выбор объясняется тем, что с их помощью можно моделировать высокоуровневые абстракции в данных, используя структуры, состоящие из множества нелинейных преобразований. Это позволяет автоматически выбрать и настроить признаки из необработанных или практически необработанных данных, находя в них нелинейные зависимости. Среди указанных нейросетей были выбраны глубокие сети доверия [17], скрытые слои которых обучаются ограниченными машинами Больцмана (ОМБ) [18].

2.2.1. Модели, основанные на энергии.

Прежде чем приступить к непосредственному описанию использованных нами глубоких нейронных сетей, следует рассмотреть модели, являющиеся фундаментом для построения их составных частей.

Таковыми являются модели, основанные на энергии (energy-based models), которые связаны с вычислением скалярной энергии каждой конфигурации переменных, используемых при тренировке сети. Самообучение таких моделей заключается в изменении энергетической функции таким образом, чтобы минимизировать энергию.

Эта функция представима через нормализованную экспоненту энергии:

$$p(x) = \frac{e^{-E(x)}}{Z},$$

где $Z = \sum_x e^{-E(x)}$; $E(x)$ – энергия.

Такая модель может быть обучена с помощью метода градиентного спуска, основанного на эмпирической функции отрицательного логарифмического правдоподобия, примененного к обучаемым данным.

Однако в большинстве случаев мы не обладаем достаточно полными данными, и поэтому нам необходимо иметь некоторые скрытые переменные для повышения эффективности модели. Таким образом, мы выделяем ее видимую и скрытую части.

Наиболее известная модель, основанная на энергии, – ограниченные машины Больцмана (restricted Boltzmann machines).

2.2.2. Ограниченные машины Больцмана.

Ограниченные машины Больцмана представляют частный случай логарифмически линейной марковской сети, для которой энергетическая функция линейна по своим свободным параметрам. Чтобы сделать ОМБ подходящей для обуче-

ния, обозначим некоторые переменные как скрытые, где можно регулировать качество построенной модели, меняя число скрытых в ней переменных.

На рис. 1 схематически изображено представление связей видимого (v_i) и скрытого (h_i) слоев ОМБ, где каждый нейрон одного слоя имеет связь со всеми нейронами другого, но между нейронами одного слоя связей нет.

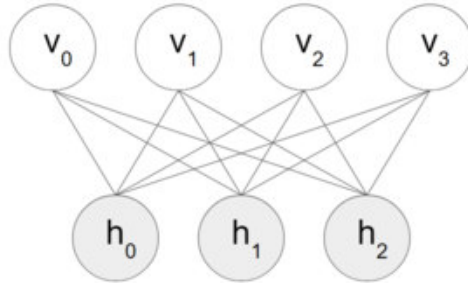


Рис. 1. Связи видимого и скрытого слоев ОМБ.

Функция энергии для такой сети представляется в виде

$$E(v, h) = -b'v - c'h - h'Wv,$$

где W – веса связей скрытых (h) и видимых (v) слоев; а b , c – смещения видимого и скрытого слоев соответственно.

В общем случае мы используем в слоях бинарные элементы, так как с ними удобнее работать и через них можно выразить другие типы переменных.

Для обучения сети применяется итерационный подход с использованием цепей Маркова и шагов Гиббса [19].

2.2.3. Глубокие сети доверия.

Несколько ОМБ могут быть объединены и обучены вместе в жадной манере [19]. Такая архитектура также называется глубокой сетью доверия (deep believe network). Она представляет собой графическую модель, цель которой – глубокое иерархическое представление данных, на которых проводится обучение.

На рис. 2 показана схема такой сети. Здесь видимым слоем является слой x , и вместе со скрытым слоем h_1 он образует ОМБ. Выход данной ОМБ считается видимым слоем для скрытого h_2 , тем самым образуя следующую ОМБ в объединении. Таким же образом к глубокой сети присоединяется третья ОМБ со скрытым слоем h_3 .

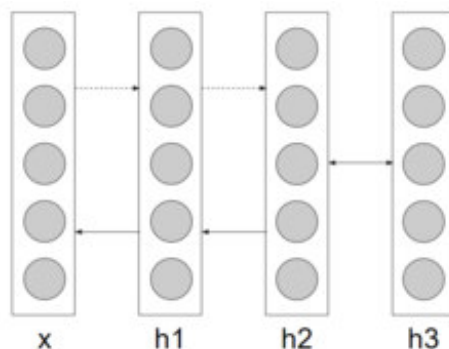


Рис. 2. Связи видимого и скрытого слоев ОМБ.

Совместное распределение между видимым вектором x и l скрытыми слоями h^k выражается формулой:

$$P(x, h^1, \dots, h^l) = \left(\prod_{k=0}^{l-2} P(h^k | h^{k+1}) \right) P(h^{l-1}, h^l),$$

где $x = h^0$, $P(h^{k-1} | h^k)$ – условное распределение видимого слоя, обусловленного скрытым слоем ОМБ на уровне k ; $P(h^{l-1}, h^l)$ – совместное распределение на верхнем уровне ОМБ.

Процесс состоит из следующих этапов [18]:

- 1) обучается первый слой как ОМБ, получающий на вход необработанные данные в виде видимого слоя;
- 2) сохраняются данные, получаемые на первом слое, для последующего их использования в качестве входных данных для второго слоя;
- 3) обучается второй слой как ОМБ на полученных с первого слоя данных;
- 4) повторяются шаги 2 и 3 вплоть до воспроизведения заданного количества слоев;
- 5) проводится тонкая настройка всех полученных с помощью такой глубокой архитектуры параметров.

Для последнего шага могут использоваться различные методы. В настоящей работе применялись градиентный спуск и логистическая регрессия.

3. Используемые данные

В качестве источника звуковых сигналов, в которых присутствует как речь, так и смех, была использована база данных MAHNOB Laughter Database, собранная в Имперском колледже Лондона [20] и надежно зарекомендовавшая себя в ряде исследований.

Для работы по выделению характерных признаков звуковых сигналов, а также обучению системы все записи были предварительно поделены на фрагменты, каждый длиной 20 мс.

В результате получено более 50 000 отдельных фрагментов сигналов. Из них для обучения классификаторов использовано 80%.

В случае классификации методом опорных векторов оставшиеся 20% фрагментов были применены при тестировании.

При использовании глубоких нейронных сетей тестирование было проведено на 10% всех данных, а на остальных 10% – выполнена валидация результатов.

Все звуковые фрагменты были предварительно обработаны таким образом, чтобы сигнал был одноканальным, с частотой дискретизации 48 кГц.

4. Результаты

4.1. Тестирование систем

Для тестирования классификатора на основе метода опорных векторов из данных сначала выделялись векторы характерных признаков размерности 124, а затем к полученным векторам применялся сам алгоритм машинного обучения. Точность выделения речи и смеха из сигнала составила 61,32% и 50,28% соответственно. Ориентация разделяющей гиперплоскости в МОВ весьма чувствительна к степени зашумленности характерных признаков. Поэтому, для того чтобы уменьшить размерность векторов характерных признаков, отбрасывали зашумленные и оставляли наиболее существенные из них, а затем применялся метод выделения главных компонент. При этом число характерных признаков было сокращено до 25, а точность распознавания речи и смеха возросла до 90,09% и 62,74% соответственно. Сказанное иллюстрирует рис. 3, где хорошо видно, что эффективность распознавания повысилась при одновременном уменьшении вычислительной сложности.

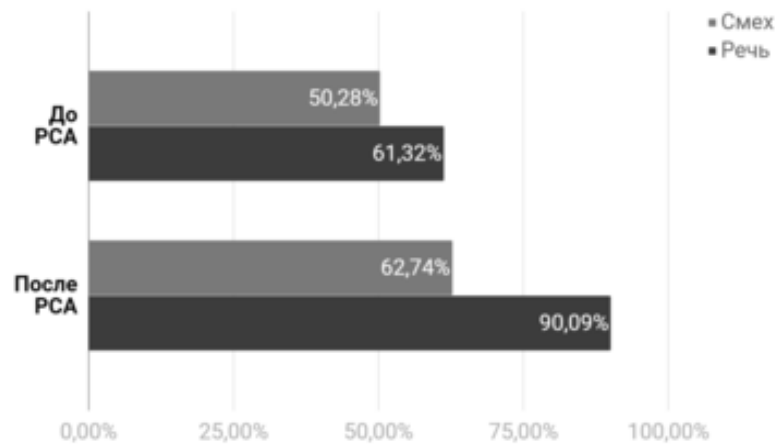


Рис. 3. Эффективность распознавания смеха и речи посредством метода опорных векторов.

На рис. 4 показана выборка тестовых результатов полученной системы на основе глубоких сетей доверия, где светлая ломаная — точность распознавания речи, темная — смеха.

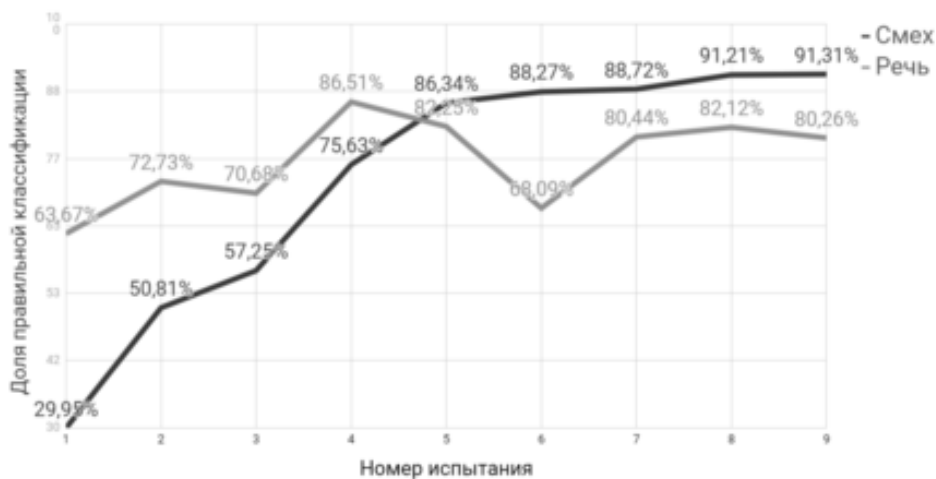


Рис. 4. Эффективность DBN при вариациях входных параметров.

По оси абсцисс располагаются тесты с вариациями различных входных параметров: начиная от количества нейронов в скрытых слоях, заканчивая типом входных данных.

Вариации параметров строились следующим образом. Вычисления выполнялись с определенным фиксированным значением параметров, которые корректировали в ту или иную сторону после каждого расчета, исходя из оценки полученных результатов. Эти действия повторялись вплоть до достижения приемлемой точности. Изначально параметры были заданы с учетом теоретических исследований по заданной теме. Полученные в итоге их оптимальные значения описаны в пункте 4.2.

Как видно из рассматриваемого графика, основная цель – снижение ошибки второго рода, поскольку смех может занимать малое время речевого сигнала и важно допускать меньшее число ложноотрицательных срабатываний, однако также была отмечена важность недопущения ошибок первого рода.

Эффективность классификации системой на основе глубоких сетей доверия показана на рис. 5 с помощью ROC-кривой. Ее построение проходило путем обучения алгоритма на 20% используемых данных, с постепенным смещением в них долей смеха и речи.

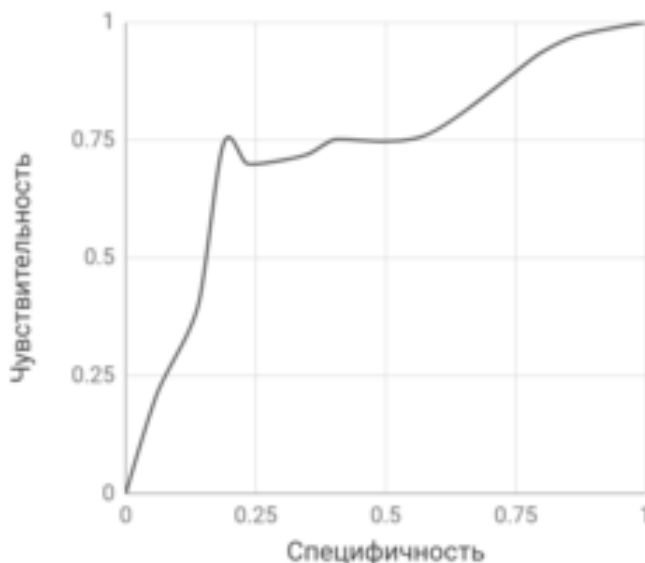


Рис. 5. ROC-кривая для системы на основе DBN.

Несмотря на уменьшенное количество данных, кривая показывает, что алгоритм хорошо проводит классификацию. Площадь под кривой в данном случае составляет 0.73.

4.2. Параметры DBN

По итогам проведенных тестов, следующие параметры нейронной сети оказались оптимальными:

- длина входного фрагмента – 0,5 мс, или 960 отсчетов;
- три скрытых слоя нейронной сети;

размер скрытого слоя – 1000 нейронов;
эпох предобучения – 15;
эпох настройки – 500;
тип входных данных – спектр сигнала;
одинаковое количество входных данных разного типа.

В качестве входных данных также подавались на вход кепстр, МЧКК и исходный сигнал. При многочисленных тестах варьировались и другие параметры.

4.3. Сравнение результатов применения метода опорных векторов и нейронных сетей

Итоговая средняя точность распознавания смеха методом опорных векторов составила 62,74%, речи – 90,09%. В случае использования глубокого обучения точность – 91,21% и 82,12% соответственно.

Сравнение результатов применения МОВ и глубоких сетей доверия представлено на рис. 6.

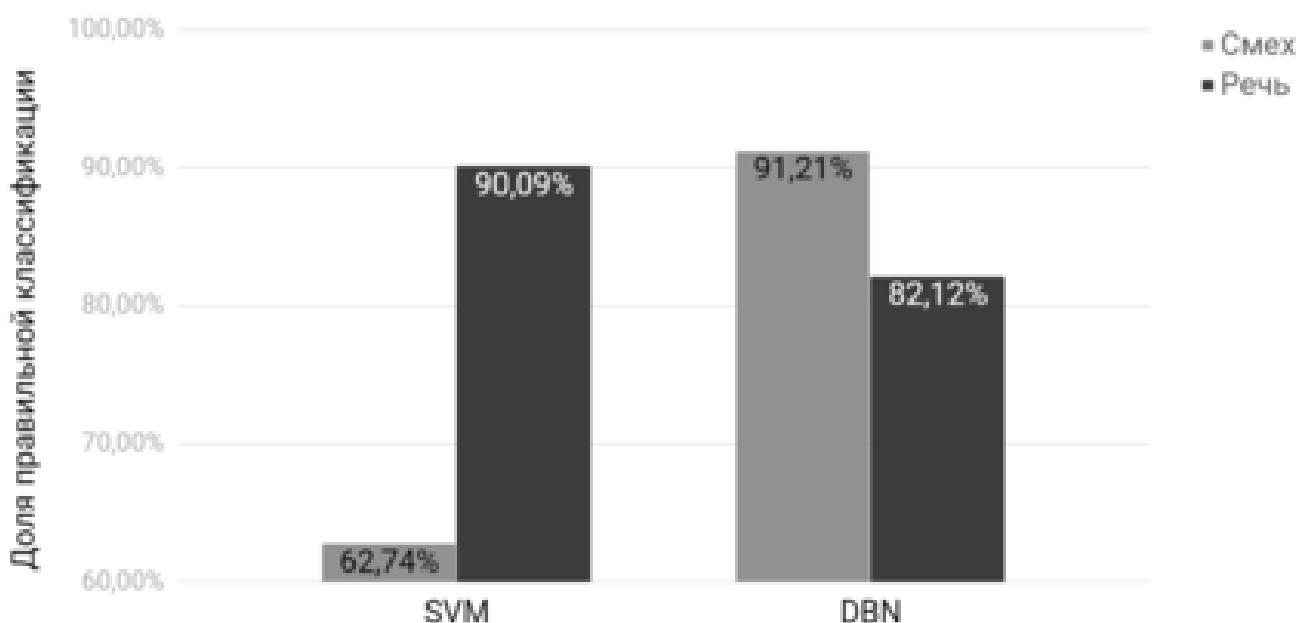


Рис. 6. Эффективность DBN при вариациях входных параметров.

Результаты показывают, что МОВ дает большую эффективность на речи и довольно посредственную – на смехе, тогда как система на основе DBN работает хорошо в обоих случаях. По всей видимости, это происходит благодаря внутренней структуре глубокой нейросети, позволяющей находить нелинейные зависимости, которые, как было выяснено эмпирическим путем, превалируют между характерными признаками, отвечающими смеху в звуковом сигнале.

Кроме того, наличие этапа выделения характерных признаков самостоятельно для МОВ требует большего времени и вычислительных ресурсов, чем их автоматическое нахождение нейронной сетью, а это несомненное преимущество последней при работе в реальном времени.

5. Заключение

В рамках настоящей работы была разработана система, позволяющая выделять фрагменты речи и смеха в исходном сигнале. Исследована эффективность применения метода опорных векторов и глубоких сетей доверия для решения поставленной задачи.

Практически доказано, что глубокие сети доверия показывают хорошую эффективность в распознавании смеха и могут работать в реальном времени. Кроме того, они превосходят метод опорных векторов по точности распознавания и требуют меньших вычислительных затрат для своего функционирования.

Полученные результаты могут быть использованы как в системах автоматического распознавания речи для повышения их эффективности, так и при решении других задач.

В будущих исследованиях планируется разработать систему, осуществляющую классификацию по типам смеха – возможно, на основе другой архитектуры нейросети.

ЛИТЕРАТУРА

1. *Давыдов А.Г., Киселев В.В., Кочетков Д.С.* Классификация эмоционального состояния диктора по голосу: проблемы и решения // Труды международной конференции «Диалог 2011» / РГГУ. – 2011. – С. 178-185.
2. *Лукьяница А.А., Шишкин А.Г.* Автоматическое определение изменений эмоционального состояния по речевому сигналу // Речевые технологии. – 2009. – №3. – С. 60-76.
3. *Стерлинг Г., Приходько П.* Глубокое обучение в задаче распознавания эмоций из речи // Труды конференции «Информационные технологии и системы 2016» / ИППИ РАН. – 2016. – С. 451-456.
4. *Truong K., van Leeuwen D.* Automatic discrimination between laughter and speech // Speech Comm. – 2007. – №49. – P.144-158.
5. *Knox M., Mirghafori N.* Automatic Laughter Detection Using Neural Networks // Interspeech 2007. Eurospeech, 10th European Conference on Speech Communication and Technology. Antwerp, Belgium. – 2007. – P. 2973-2976.
6. *Gosztolya G., Beke A., Neuberger T. et al.* Laughter Classification Using Deep Rectifier Neural Networks with a Minimal Feature Subset // PAN – IPPT: Archives of acoustics. – 2016. – Vol. 41, №4. – P. 669-682.
7. *Cristianini N., Shawe-Taylor J.* An Introduction to Support Vector Machines and other kernel-based learning methods // Cambridge University Press. – 2000. – P.273.
8. *Brossier P.M.* Automatic Annotation of Musical Audio for Interactive Applications // Centre for Digital Music. – Queen Mary. University of London. – 2006. – P. 79-82.
9. *Cheveignre A., Kawahara H.* YIN, a fundamental frequency estimator for speech and music // Journal Acoust. Soc. Am. – 2002. – Vol. 111, №4. – P. 1917-1930.
10. *Huang X., Acero X., Hon H.-W.* Spoken Language Processing: A Guide to Theory, Algorithm, and System Development. – Prentice Hall, 2001.
11. *Kammoun M.A., Gargouri D., Frikha M. et al.* Cepstrum vs. LPC: A Comparative Study for

- Speech Formant Frequencies Estimation // GESTS Int'l Trans. Communication and Signal Proce. – 2006. – Vol.9, №1. – P. 99.
12. *Рабинер Л.Р., Шафер Р.В.* Цифровая обработка речевых сигналов. – М.: Радио и связь, 1981. – С. 133-141, 366-372.
 13. *Bäckström T.* Levinson-Durbin Recursion: Doctoral thesis. – Linear Predictive Modelling of Speech – Constraints and Line Spectrum Pair Decomposition. Helsinki University of Technology. – 2004. – Report №71. – P. 27-28.
 14. *Первушин Е.А.* Обзор основных методов распознавания дикторов // Математические структуры и моделирование. – 2011. – №24. – С. 41-54.
 15. *Caruntu A., Todorean G., Nica A.* Automatic Silence/Unvoiced/Voiced Classification of Speech Using a Modified Teager Energy Feature // WSEAS Int. Conf. on DYNAMICAL SYSTEMS and CONTROL. Venice, Italy. – 2005. – P. 62-65.
 16. *Айвазян С.А., Бухштабер В.М., Енюков И.С. и др.* Прикладная статистика. Классификация и снижение размерности. – М.: Финансы и статистика, 1989. – С. 332-384.
 17. *Hinton G., Salakhutdinov R.* Reducing the Dimensionality of Data with Neural Networks // Science. – 2006. – Vol. 313, №5786. – P. 504-507.
 18. *Tieleman T.* Training restricted boltzmann machines using approximations to the likelihood gradient – ICML, 2008.
 19. *Hinton G.* A Practical Guide to Training Restricted Boltzmann Machines – Department of Computer Science, University of Toronto, 2010.
 20. *Petridis S., Martinez B., Pantic M.* The MAHNOB-Laughter Database // Image and Vision Computing Journal. – 2013. – №31(2). – P. 186-202.

Статья представлена к публикации членом редколлегии О.С. Амосовым.

E-mail:

Горошевский Алексей Валерьевич – ag@rissik.ru;

Шишкин Алексей Геннадиевич – shishkin@cs.msu.ru.