

## ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ В РАСПОЗНАВАНИИ РЕЧИ

**Онаркулов Максад Каримбердиевич**

Авторы: д.ф.ф.м.н.

(Старший преподаватель кафедры «Прикладная математика и информатика» Ферганского Государственного Университета).

**Юсупов Мирсаид Абдулазиз угли**

(Магистрант 2 курса направления «Прикладная математика» Ферганского Государственного Университета).

**Умиржонов Лазизжон Акмалжон угли**

(Магистрант 2 курса направления «Прикладная математика» Ферганского Государственного Университета).

**Гофурова Дилхаёт Маматвали кизи**

(Магистрант 2 курса направления «Прикладная математика» Ферганского Государственного Университета).

### АННОТАЦИЯ

В данной статье рассмотрены алгоритмы распознавания речи и приведена классификация методов распознавания речи.

**Ключевые слова:** алгоритмы распознавания речи, методы распознавания речи, классификация методов распознавания речи.

Человечество всегда стремилось к более простому взаимодействию с техникой, потому что естественное, языковое общение представляется нам самым удобным и привычным. Главная задача речевого интерфейса заключается в понимании человеческой речи и правильном на неё реагировании. Таким образом, необходимо обучить машину понимать без посредника язык, на котором общаются люди, то есть создать алгоритм распознавания звуковых сигналов речи. Этим и должна заниматься технология распознавания речи. В наши дни любой человек использует такой интерфейс в своей повседневной жизни. Простым примером являются такие приложения как голосовое управление рабочим столом, разнообразные автомобильные устройства, даже «умные» дома. Также в нашем ежедневном обиходе используются речевые транскрипции, субтитры, переводы и изучение языка, голосовые поиск и

помощники. Понятие «Автоматическое распознавание речи» (в английской терминологии, automatic speech recognition – ASR) – это преобразование звука в текст. Имеется несколько категорий систем распознавания речи, которые применяются различных сферах:

1. Распознавание отдельных голосовых команд, которое применяется в коммерческих приложениях;
2. Поиск ключевых слов в потоке речи (поисковые системы);
3. Автоматическое распознавание слитной речи на большом словаре (автоматическая расшифровка записей – создание стенограмм).

Системы, которые существуют на сегодняшний день еще далеки от совершенства и имеют ряд недостатков, такие как высокий процент ошибок, настройки на конкретного диктора, ограниченный объем словаря – и это далеко не полный список проблем, которые предстоит решить. Всем известно, что речь человека характеризуется высокой степенью изменчивости, обусловленное несколькими причинами. Даже для одного и того же говорящего, при реализации одних и тех же акустических единиц будут отличаться длительностью произношения, по спектральному составу. Любые изменения в речи связано с изменениями эмоционального состояния человека, и естественно условий, в которых он находится. Наличие артикуляционных эффектов приводит к тому, что произношение слов, фонем сильно зависит от их контекста. К изменениям в речевом сигнале, разумеется, приводят помехи различного характера среды, в которой находится говорящий. Всевозможные речевые сбои, такие как заполненные паузы, удлиняют высказывания. Из-за вызванных различного рода ошибок вследствие того, что системы распознавания обучаются на структурированных предложениях без речевых сбоев, что приводит к формированию ошибочных транскрипций. Учитывая все эти факторы, и принимая во внимание ряд других ограничений, делаем вывод - для высококачественного распознавания речи в реальном времени требуются вычислительные средства с высоким быстродействием. Одним из способов снижения этого требования является распараллеливание вычислений, которое естественным образом достигается при использовании искусственных нейронных сетей, реализованных на нейрокомпьютерах.

Искусственная нейронная сеть– это реализованная программно или аппаратно математическая модель, описывающая работу человеческого мозга. Базой этих сетей служат нейроны – элементы, которые получают некий импульс и передают выходной сигнал. Нейрон состоит из взвешенного сумматора и нелинейного элемента. Функционирование нейрона определяется следующими показателями:

$$NET = \sum_t w_i x_i$$
$$OUT = F(NET - \theta)$$

где  $x_i$  — входные сигналы, совокупность всех входных сигналов нейрона  
 $w_i$  — весовые коэффициенты, совокупность весовых коэффициентов образуют вектор весов  $w$ ;

$NET$ - взвешенная сумма входных сигналов, значение  $NET$  передаётся на нелинейный элемент;

$\theta$  – пороговый уровень данного нейрона;

$F$ - нелинейная функция, называемая функцией активации.

Нейрон имеет несколько входных сигналов  $x$  и один выходной сигнал  $OUT$ . Вектор весов  $w$ , пороговый уровень  $\theta$  и вид функции активации  $F$  являются параметрами, определяющими его работу. Применяется очень часто для многослойных персептронов и других сетей с непрерывными сигналами. Гладкость, непрерывность функции — важные положительные качества. Непрерывность первой производной позволяет обучать сеть градиентными методами (например, метод обратного распространения ошибки). Функция симметрична относительно точки ( $NET=0$ ,  $OUT=1/2$ ), это делает равноправными значения  $OUT=0$  и  $OUT=1$ , что существенно в работе сети. Тем не менее, диапазон выходных значений от 0 до 1 несимметричен, из-за этого обучение начительно замедляется. Данная функция — сжимающая, т.е. для малых значений  $NET$  коэффициент передачи  $K=OUT/NET$  велик, для больших значений он снижается. Поэтому диапазон сигналов, с которыми нейрон работает без насыщения, оказывается широким. Значение производной легко выражается через саму функцию. Быстрый расчет производной ускоряет обучение. Тоже применяется часто для сетей с непрерывными сигналами. Функция симметрична относительно точки  $(0,0)$ , это преимущество по сравнению с сигмодой. Производная также непрерывна и выражается через саму функцию.

Формальные нейроны могут объединяться в сети различным образом. Такая сеть состоит из произвольного количества слоев нейронов. Нейроны каждого слоя соединяются с нейронами предыдущего и последующего слоев по принципу "каждый с каждым". Первый слой нейросети называется сенсорным или входным, внутренние слои называются скрытыми или ассоциативными, последний — выходным или результативным. Количество нейронов в слоях может быть произвольным. Обычно во всех скрытых слоях одинаковое количество нейронов. Каждый слой нейросети рассчитывает нелинейное преобразование от линейной комбинации сигналов предыдущего слоя. Линейная

функция активации может применяться только для тех моделей сетей, где не требуется последовательное соединение слоев нейронов друг за другом. Нелинейная функция активации применяется для многослойных сетей, иначе можно построить эквивалентную однослойную сеть, и многослойность оказывается ненужной. В случае применения линейной функции активации, то каждый слой будет давать на выходе линейную комбинацию входов. Следующий слой даст линейную комбинацию выходов предыдущего, а это эквивалентно одной линейной комбинации с другими коэффициентами, и может быть реализовано в виде одного слоя нейронов. В многослойном персептроне нет обратных связей. Благодаря поочередного расчета линейных комбинаций и нелинейных преобразований достигается аппроксимация произвольной многомерной функции при соответствующем выборе параметров сети. Процесс распознавания речи с использованием нейронных сетей состоит из следующих этапов

1. Акустический препроцессор обрабатывает входной речевой сигнал, и определяет последовательность векторов признаков, для каждого отрезка времени (семпл) и состоят из спектральных или кепстральных коэффициентов, характеризующих отрезки речевого сигнала.

2. Далее полученные векторы подлежат сравнению с эталонными векторами, содержащимися в моделях слов. На этом же этапе вычисляются локальные метрики или меры соответствия (как правило сравниваются речевые сегменты, описанные несколькими векторами признаков) [53].

3. С помощью метрик происходит временное выравнивание последовательностей векторов признаков с последовательностями эталонных векторов, образующими модели слов, а также вычисляется мера соответствия для компенсации изменений скорости произнесения. Далее находится максимально соответствующее слово. В случае если распознается непрерывная (слитная) речь, то полученные на втором этапе локальные метрики используются для временного выравнивания и определения мер соответствия для отдельных выражений.

Принимая во внимание все преимущества использования нейронных сетей для распознавания речи будет правильно отметить, что они не могут быть основным и единственным методом распознавания, ввиду отсутствия механизмов, адекватно представляющих временную вариативность и последовательную природу речевого сигнала; теоретических основ, используемых для вычисления или выбора параметров, определяющих динамику и топологию нейронных сетей (в настоящее время они выбираются на усмотрение разработчика); а также несмотря на постоянное совершенствование

алгоритмов для упрощения и ускорения процедуры обучения использование нейронных сетей до сих пор является довольно ресурсоемким процессом.

### **СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ: (REFERENCES)**

1. Кипяткова И.С., Карпов А.А. Аналитический обзор систем распознавания русской речи с большим словарем // Труды СПИИРАН. – 2010. – Вып. 12. – С. 7-20.
2. Беленко М.В., Балакшин П.В. Сравнительный анализ систем распознавания речи с открытым кодом // МНИЖ. 2017. №4-4 (58).
3. Данков Н.И. Исследование возможностей нейросетевых технологий в области идентификации голоса // Экономика и качество систем связи. 2018. №3.
4. Карпов Алексей Анатольевич, Кипяткова Ирина Сергеевна Методология оценивания работы систем автоматического распознавания речи // Приборостроение. 2012. №11.
5. Меденников, И. П. Двухэтапный алгоритм инициализации обучения акустических моделей на основе глубоких нейронных сетей / И. П. Меденников // Научно-технический вестник информационных технологий, механики и оптики. — 2016. — Т. 16., № 2. — С. 379–381.
6. Prudnikov, A. Improving Acoustic Models For Russian Spontaneous Speech Recognition / A. Prudnikov, I. Medennikov, V. Mendeleev, M. Korenevsky, Y. Khokhlov // Speech and Computer, Lecture Notes in Computer Science. — 2015. — Vol. 9319. — P. 234–242.
7. Маковкин К. А. Гибридные модели: скрытые марковские модели и нейронные сети, их применение в системах распознавания речи // Модели, методы, алгоритмы и архитектуры систем распознавания речи: Изд-во «Вычислительный центр им. А. А. Дородницына РАН», – М.: 2006. – С.40-95.
8. Авсентьев А.О., Лукьянов А.С. Применение скрытых марковских моделей для распознавания речи диктора. 2015. №2.
9. Созыкин А. В. Обзор методов обучения глубоких нейронных сетей // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2017. №3.
10. Стас Тамби Тахсинович, Метод получения векторов акустических признаков для распознавания последовательности фразы в условиях шумовых помех // Известия ВУЗов. Поволжский регион. Технические науки. 2015. №2 (34).
11. . Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara Sainath, and Brian Kingsbury Deep Neural Networks for Acoustic Modeling in Speech Recognition - IEEE, Signal Processing Magazine, 2012.