УДК 621.3.049.77–048.24:537.2

СОВРЕМЕННЫЕ ПОДХОДЫ К АВТОМАТИЧЕСКОМУ РАСПОЗНАВАНИЮ РЕЧИ

Ерёменко Т.А.

Белорусский государственный университет информатики и радиоэлектроники, г. Минск, Республика Беларусь

Научный руководитель: Пискун Г.А. – канд.техн.наук, доцент, доцент кафедры ПИКС

Аннотация. В настоящее время набирают популярность различные устройства с системами функцией распознавания речи, т.е преобразования звукового сигнала в текст. Рассмотрены основные задачи, решаемые в процессе таких преобразований. Также выделены наиболее актуальные на сегодняшний день подходы к решению задач на этапах распознавания речи.

Ключевые слова: распознавание речи, акустическая модель, языковые модели, декодирование

Введение. Распознвание речи на данный момент используется не только в мобильных гаджетах и десктопах, но и в различных средствах бытовой техники. Возможность передавать данные на гаджет, используя голос, делает взаимодействие с любым устройством более быстрым и доступным. На сегодняшний день данной функцией снабжается огромное количество устройств, и число таких устройств только увеличивается.

Возможность набора текстового сообщения при помощи процесса разпознавания речи на телефон, компьютер или любой другой гаджет, в котором встроен микрофон, дает возможность комфортного использования устройств прежде всего людьми с огаиченными возможностями, однако данная функцию может быть полезена каждому человеку.

Для распознавания речи используются различные методы и алгоритмы обработки речевого сигнала. Основной задачей, перевода голоса человека в текст на гаджете или же простого распознавания слов и фраз, является перевод поступающего в микрофон физического сигнала в электрический сигнал, а затем его оцифровка при помощи аналого-цифрового преобразователя. Наиболее сложным моментом в распознавании речи является интонация произнесения слов, людьми различного возраста и наличие постороннего шума, а также то, что акустические модели различны для разных языков.

Основная часть. При аналогово-цифровом преобразовании могут быть использованы различные модели и подходы. Наиболее популярные на данный момент модели для распознавания речи на звуковом уровне – акустические модели.

Акустическая модель — это функция, принимающая на вход небольшой участок акустического сигнала (кадр) и выдающая распределение вероятностей различных фонем – атомарных единиц речи – на этом кадре. Данные модели разбивают входящую человеческую речь на определенные временные промежутки и дают предопределение звука на каждом таком промежутке (обычно используются промежутки длительностью в 25 мс). Модель предопределяет значение каждой фонемы на каждом промежутке. Такие модели оцифровывают звуки, которые могут быть ассоциируемыми с человеческой речью [2].

Таким образом, акустическая модель дает нам возможность по звуку установить, что было произнесено с той или иной степенью уверенности. На данный момент наиболее популярными акустическими моделями распознавания речи являются:

– скрытая марковская модель;
– принцип максимальной энтропии;
– условные случайные поля;
– нейронные сети [3].

Самой популярной и наиболее часто используемой является скрытая марковская модель. Марковские модели – это статистические модели, которые используются для прогнозирования следующего состояния на основе текущих скрытых или наблюдаемых состояний. Марковская

модель – это акустическая модель, в которой каждое состояние имеет соответствующую вероятность оказаться в любом другом состоянии после каждого шага. Марковские модели могут быть использованы для моделирования реальных проблем, в которых задействованы скрытые и наблюдаемые состояния. Марковские модели можно разделить на скрытые и наблюдаемые в зависимости от типа доступной информации, которую можно использовать для принятия прогнозов или решений. Скрытые марковские модели имеют дело со скрытыми переменными, которые нельзя наблюдать непосредственно, а только выводить из других наблюдений, тогда как в наблюдаемой модели, также называемой цепью Маркова, скрытые переменные не задействованы. Распознавание устной речи имеет много неопределенных моментов и характеристик, в связи с этим целесообразнее использовать именно скрытые марковские модели.

Для осуществления распознавания на основе скрытых моделей Маркова необходимо построить кодовую книгу, содержащую множество эталонных наборов для характерных признаков речи (например, коэффициентов линейного предсказания, распределения энергии по частотам и т.д.). Для этого записываются эталонные речевые фрагменты, разбиваются на элементарные составляющие (отрезки речи, в течении которых можно считать параметры речевого сигнала постоянными) и для каждого из них вычисляются значения характерных признаков. Одной элементарной составляющей будет соответствовать один набор признаков из множества наборов признаков словаря.

При распознавании с использованием скрытых моделей Маркова мы разбиваем речь на отрезки, для каждого вычисляем набор номеров кодовой страницы и применяем алгоритм прямого или обратного хода для вычисления вероятности соответствия данного звукового фрагмента определенному слову словаря. Если вероятность превышает некоторое пороговое значение - слово считается распознанным. [1].

Особое место в задаче распознавания речи занимают методы, основанные на нейросетевой технологии. В этих методах результат распознавания является продуктом функционирования нейронной сети определенного вида и топологии. Нейронные сети представляют собой множество связанных между собой элементарных процессоров (нейроподобных элементов), каждый из которых выполняет относительно простые функции. Важнейшей отличительной особенностью нейросетевого метода является возможность параллельной обработки. Данная особенность при большом количестве межнейронных связей дает возможность достигнуть значительного ускорения процесса обработки данных. Во многих случаях появляется возможность обработки речевых сигналов в реальном времени. Еще один важный плюс в нейросетевом методе – это обобщение полученных знаний. Нейронная сеть обладает качествами, которые свойственны так называемому искусственному интеллекту.

В результате процесса обучения нейронная сеть способна выявлять сложные зависимости между входными и выходными данными. При обобщении информации сеть позволит вернуть верный результат на основании неполных или искаженных данных. При большом количестве соединений между нейронами сеть приобретает устойчивость к ошибкам, возникающим на некоторых линиях. Работу поврежденных связей берут на себя исправные линии, в результате чего работа сети не претерпевает существенных изменений. Исходя из этих характеристик метода, можно предположить, что нейронные сети являются достаточно эффективным решением проблемы распознавания речи [4].

Важной частью распознавания речи также являются языковые модели. Они используются для того, чтобы система могла определить наиболее вероятную последовательность слов. Здесь в самом простом случае требуется предсказать следующее слово по известным предыдущим словам. В традиционных системах применялись модели типа N-грамм, в которых на основе большого количества текстов оценивались распределения вероятности появления слова в зависимости от N предшествующих слов. Внедрение языковой модели в систему распознавания речи позволило значительно повысить качество распознавания за счет учета контекста. Языковая модель, берет свое начало в области

обработки естественного языка. Основная цель языкового моделирования состоит в том, чтобы, учитывая последовательность слов, предсказать следующее слово в последовательности. Обычно языковое моделирование выполняется на уровне слов, но оно также может быть выполнено на уровне символов, что полезно в определенных ситуациях, например, для языков, которые в большей степени основаны на символах (китайский, японский и т.п.).

При распознавании слитной речи использование даже самых простых моделей приводит к серьезным проблемам, связанным с быстродействием и памятью систем. Как результат, эта задача выносится в отдельный модуль системы автоматического распознавания речи, называемый декодером. Декодер должен определять наиболее грамматически вероятную гипотезу для неизвестного высказывания – то есть определять наиболее вероятный путь по сети распознавания, состоящей из моделей слов (которые, в свою очередь, формируются из моделей отдельных фонов).

Декодирование речи – термин для того, что происходит, когда система представлена новым высказыванием и должна вычислять наиболее вероятное исходное предложение – вероятно, будет использовать алгоритм Витерби для поиска наилучшего пути. Алгоритм Витерби – это особый, но наиболее широко используемый алгоритм динамического программирования. Используя динамическое программирование, можно решить задачу о кратчайшем пути в любом графе. Предлагается алгоритм Витерби для решения специальной задачи о графе-кратчайшем пути ориентированного графа ограждающей сети (Решетки). Поэтому именно алгоритм Витерби чаще всего используется для декодирования при распознавании речи, а также в процессе машинного перевода, в современной цифровой связи и т.д.

Заключение. Работа системы автоматического распознавания речи сводится к определению наиболее вероятной последовательности слов, соответствующих содержанию речевого сигнала. Наиболее вероятный кандидат должен определяться с учетом как акустической, так и лингвистической информации.

Таким образом, основными этапами распознавания речи, для наиболее четкой её передачи на устройства, являются акустическая и языковая модель, а также декодирование.

Список использованных источников

1. Hidden Markov models [Электронный ресурс]. – Режим доступа: https://vitalflux.com/hidden-markov-models-concepts-explained-with-examples/ – Дата доступа: 04.01.2022.

2.Классификация систем распознавания речи, Федосин С.А., Еремин А. Ю. –2009, с. 5

3.Speech recognition transcription models [Электронный ресурс]. – Режим доступа: https://www.rev.com/blog/guide-to-speech-recognition-transcription-models–Дата доступа: 04.01.2022.

4.Нейросетевые методы распознавания речи [Электронный ресурс].–Режим доступа: https://www.gramota.net/articles/issn_1993-5552_2014_3_14.–Дата доступа: 04.01.2022.

UDC 621.3.049.77–048.24:537.2

MODERN APPROACHES TO AUTOMATIC SPEECH RECOGNITION

Yaromenka T.A.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Piskun G.A. – PhD, assistant professor, associate professor of the department of ICSD

Annotation Nowadays various devices with speech recognition systems are gaining popularity (i.e converting an audio signal into text). The main tasks that are solved in the process of such transformations are considered in this article. The most relevant approaches to solving problems at the stages of speech recognition are also highlighted.

Keywords. speech recognition, acoustic model, language models, decoding