

ОБЗОР СОВРЕМЕННОГО СОСТОЯНИЯ ПРОБЛЕМЫ РАСПОЗНАВАНИЯ РЕЧИ

OVERVIEW OF MODERN CONDITION PROBLEM IN SPEECH RECOGNITION

**M. Hein
V. Kudinov
V. Dovgal**

Summary. Speech is the most natural way of human communication and also the most efficient form of exchanging information. Speech can be identified and converted to a machine readable format via a technology called speech recognition, or speech to text. Sometimes it is known as automatic speech recognition. The speech recognition research is focuses mainly on large vocabulary, continuous speech capabilities and speaker independence. The main goal of speech recognition area is to develop techniques and systems for speech input to machine. This paper presents the review of the different speech recognition system and its recent progress.

Keywords: Speech Recognition, Feature Extraction, MFCC, LPC, HMM, Neural Network, DTW.

Хейн Мин Зо

Аспирант, Курский Государственный Университет
heinminnzaw13@gmail.com

Кудинов Виталий Алексеевич

Д.п.н., профессор, Курская государственная
сельскохозяйственная академия
kudinovva@yandex.ru

Довгаль Виктор Митрофанович

Д.т.н., профессор, Курский государственный
университет
vmdovgal@yandex.ru

Аннотация. Речь — самый естественный способ общения людей, а также наиболее эффективная форма обмена информацией. Речь может быть идентифицирована и преобразована в машиночитаемый формат с помощью технологии, называемой распознаванием речи, или речи в текст. Иногда это называется автоматическим распознаванием речи (APP). Исследование распознавания речи сосредоточено главным образом на большой лексике, непрерывных речевых возможностях и независимости ораторов. Основной целью области распознавания речи является разработка методов и систем для ввода речи в машину. В этой статье представлен обзор системы распознавания речи и ее недавнего прогресса.

Ключевые слова: Распознавание речи, извлечение функций, МЧКК, ДПК, СММ, нейронная сеть, ДТВ.

Введение

Речь — это наиболее широко и часто используемый метод общения между людьми [1]. Распознавание речи — это процесс преобразования речевого сигнала в последовательность слов с помощью алгоритма реализованного через компьютерную программу [2]. Из-за технологического любопытства создавать машины, которые имитируют людей или хотят автоматизировать работу с машинами, исследования в области распознавания речи в качестве первого шага к естественным человеко-машинным коммуникациям привлекли много энтузиастов в течение последних пяти десятилетий. Поэтому несколько исследовательских усилий были ориентированы на эту область, где программисты изучают способы и средства, позволяющие компьютерам записывать, интерпретировать и понимать человеческую речь. Это была прогрессивная исследовательская область на протяжении десятилетий. Распознавание речи — это особый случай распознавания образов. В распознавании речи присутствуют две фазы: тренировка и тестирование. На фазе тренировки, данные для известных классов подаются в систему. В фазе распознавания система вычисляет особенности шаблона для неизвестного ввода и иденти-

фицирует ввод с классом, эталонный шаблон которого наиболее точно соответствует этим функциям [3] [4] [5].

Типы распознавания речи

- ◆ Изолированные распознаватели слова: — принимают по одному слову за раз. Эти системы распознавания позволяют говорить естественным образом.
- ◆ Подключенные системы слов [6] позволяют динамику говорить медленно и отчетливо каждое слово с короткой паузой.
- ◆ Непрерывные распознаватели речи позволяют пользователям говорить почти естественно, в то время как компьютер определяет контент (в основном, это компьютерная диктовка). Непрерывное распознавание речи связано с речью, где слова соединены вместе, а не разделены паузами. Распознаватели с непрерывными речевыми возможностями являются одними из самых сложных для создания, потому что они используют специальные методы для определения границ высказывания [7][8].
- ◆ Спонтанная речь — это речь, которая является естественным звучанием и не репетируется [7]. Си-

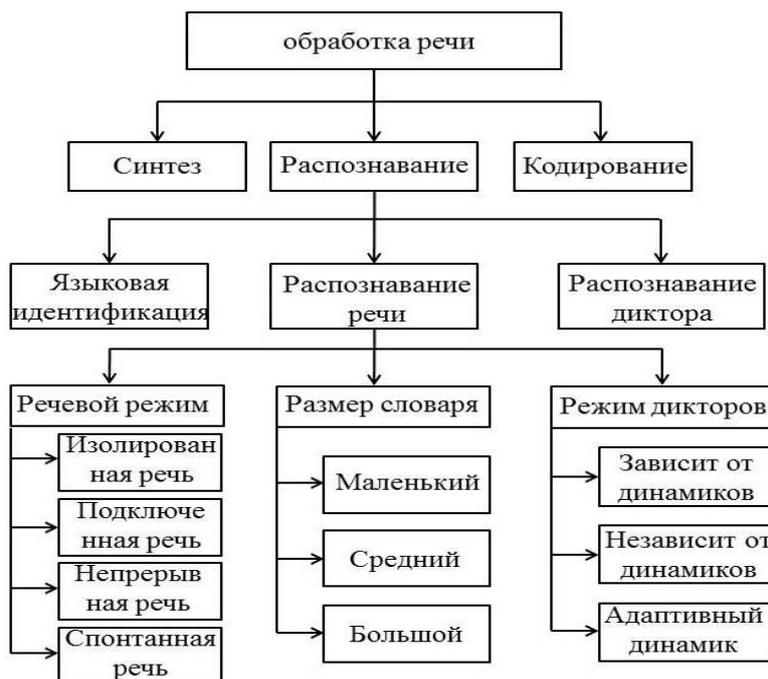


Рис. 1. Классификация обработки речи



Рис. 2. Представление диаграммы системы распознавания речи

стема APP со спонтанной речевой способностью должна иметь возможность обрабатывать множество естественных особенностей речи, таких как слова, выполняемые в одно и то же время [8][9][10].

Классификация систем распознавания речи

Системы распознавания речи можно классифицировать, как показано на рис. 1. Где распознавание речи является одной из наиболее важных областей цифровой обработки сигналов и является очень востребованной технологией, которая состоит из множества полезных приложений.

Система распознавания речи

В системе распознавания речи точное и эффективное преобразование речевого сигнала в текстовое со-

общение транскрипции произносимых слов независимо от динамика, среды или устройства, используемого для записи речи [3]. Некоторые из основных применений такого распознавания речи — это распознаваемые голо- сом пароли, голосовые репертуарные дозвончики, автоматическое распознавание типа вызова, распределе- ние вызовов по голосовым командам, поиск каталогов, проверка продаж кредитных карт, речевая обработка текста, автоматический ввод данных и т.д.[5]. На рис. 2 показано базовое представление системы распознава- ния речи. Он включает в себя четыре основных этапа: речевой анализ или предварительную обработку, коди- рование, извлечение признаков, преобразование языка и прием сообщений.

Блок анализа речи используется для удаления шума, тишины и конечной точки обнаружение из входного ре- чевого звука, которые необходимы для повышения про- изводительности системы распознавания речи. Затем

блок извлечения признаков используется для удаления ненужной и избыточной информации и сохраняет только полезную информацию в типе независимого диктора распознавания речи. Блок преобразования языка используется для захвата свойств языка и предсказания следующего слова в речевой последовательности. Затем наконец, блок двигателя распознавания речи должен преобразовать входной звук в текст.

Подходы к распознаванию речи

- ◆ Акустико-фонетический подход — этот подход основан на теории акустической фонетики и догадок.
- ◆ Подход распознавания образов — обучение и сравнение шаблонов — это шаги, которые включают в себя подход распознавания образов. Используя хорошо сформулированную математическую структуру и заявляет о надежной речевой схеме для сравнения образцов.
- ◆ Подход искусственного интеллекта — машина работает как человек. Это комбинация подхода распознавания образов и акустического фонетического метода.
- ◆ Шаблонные подходы — неизвестная речь сравнивается с набором предварительно записанных слов, которые правильны, чтобы поймать лучший Матч.
- ◆ Стохастический подход — стохастическое моделирование подразумевает использование вероятностных моделей для пакт с неопределенной или неадекватной информацией.
- ◆ Основанный на знаниях / правилах подход — «квалифицированный» знания о различиях в речи ручная кодировка в систему. И он использует набор топографий из речи, затем система обучения автоматически порождает набор производственных правил из образцов.
- ◆ Динамическое трансформирование времени (ДТВ) — это алгоритм вычисления сходства между двумя классификациями, которые могут расходиться во времени или скорости. Этот метод также достаточно эффективен для изолированного распознавания слов и может быть адаптирован для распознавания словосочетаний.
- ◆ Основанный на учебе подход — цель машинного обучения — это изучение или определение какого-либо знания из набора документов.
- ◆ Подход основанный на нейронных сетях — дополнительный широкомасштабный подход к нейронным сетям — это распознавание фонемы.
- ◆ Статистический подход — здесь различия в речи моделируются статистически и определяются учебными событиями. Этот подход означает текущее состояние навыка.

- ◆ Скрытая марковская модель (СММ) — скрытая марковская модель очень часто используется, потому что их можно обучить автоматически. Следовательно, СММ проста и её можно вычислительно использовать.
- ◆ Векторное квантование (ВК) — векторное квантование часто применяется в распознавании речи, и оно ценно для речевых кодировщиков, то есть эффективного сокращения данных.
- ◆ Поддержка векторной машины (ПВМ) — ПВМ использует линейные и нелинейные разделительные гиперплоскости для классификации данных.

Технологии речевых извлечений признаков

Извлечение характеристик параметров, представляющих собой акустический сигнал, является важной задачей для повышения эффективности распознавания. Эффективность этого метода важна для следующего метода, поскольку он влияет на его поведение. Различные методы извлечения признаков доступны с их особенностями.

В принципином компоненте анализа (ПКА) — он использует метод нелинейного извлечения признаков и дает линейную карту, и является быстрым и основанным на собственных векторах.

В линейном дискриминантном анализе (ЛДА) — это зависит от метода нелинейного извлечения признаков, он имеет контролируемую линейную карту и является быстрым и основанным на собственных векторах. Этот метод лучше, чем ПКА для классификации [6].

Линейно-предсказательное кодирование (ЛПК) использует метод статического извлечения признаков, который имеет от 10 до 16 коэффициентов нижнего порядка. Он используется для извлечения особенности в более низком порядке. ЛПК анализирует речевой сигнал, оценивая форматы, удаляя их двигатели из речевого сигнала.

В Мел-частота кепструма коэффициенты (МЧКК) — он обладает тем свойством, что спектр мощности вычисляется путем выполнения анализа Фурье. Мел-частотный кепстр (МЧК) может определяться кратковременным континуумом мощности речевого сигнала, который рассчитывается как линейное косинусное преобразование спектра логарифмической мощности в нелинейный Мел-масштаб частоты [11] [12]. МЧКК также все чаще используется в музыкальной информации, такой как классификация жанров, количественная оценка подобию

звука и т.д. Это дает высокую точность результатов для чистой речи.

Производительность методов

Производительность системы распознавания речи обычно определяется с точки зрения точности и скорости. Точность вычисляется по частоте ошибок в слове, тогда как скорость измеряется с коэффициентом реального времени. Частота ошибок в словах (WER) является общей метрикой производительности системы распознавания речи или машинного перевода. Затем WER может быть вычислен

$$WER = S+D+I/N \quad (1)$$

где

S : количество замен (неверные слова заменены)

D : количество удалений (слова удалены)

I : количество вставок (добавлены дополнительные слова)

N : количество слов в ссылке.

В то время, сообщая о производительности системы распознавания речи иногда скорость распознавания слов (WRR) составляет

$$WRR = 1 - WER = N - S - D - I / N = H - I / N \quad (2)$$

где H является $N - (S + D)$ числом правильно распознанных слов.

К дополнительным методам точности относятся скорость успешного выполнения команды и частота ошибок одиночного слова.

Заключение

Распознавание речи является одной из наиболее интегрированных областей машинного интеллекта, так как люди ежедневно распознают речь. Этот вопрос привлек ученых к важной дисциплине и создал технологическое воздействие на общество. В этом обзоре обсуждался процесс распознавания речи, который включает в себя извлечение признаков, типы подходов к распознаванию речи и методы классификации речи. Каждый метод имеет различную скорость распознавания с измененным процессом классификации. Поэтому эта исследовательская работа заключается в том, чтобы понять процесс распознавания речи и разработать в будущем эффективную систему распознавания речи для людей с неограниченной точностью.

ЛИТЕРАТУРА

1. Pratik K. Kurzekar et al. «Continuous Speech Recognition System: A Review», Asian Journal of Computer Science and Information Technology. 4: 6 (2014) 62–66.
2. Irfan Y. Khan, P. H. Zope and S. R. Suralkar, (2013). Importance of Artificial Neural Network in Medical Diagnosis disease like acute nephritis disease and heart disease, 2(2), 210–217.
3. Anchal Katyal, Amanpreet Kaur, Jasmeen Gill, « Automatic Speech Recognition: A Review», International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249–8958, Volume-3, Issue-3, February 2014
4. Asm SAYEM, «Speech Analysis for Alphabets in Bangla Language: Automatic Speech Recognition», International Journal of Engineering Research (ISSN: 2319–6890) (online), 2347–5013 (print), Volume No.3, Issue No.2, pp: 88–93 01 Feb. 2014.
5. Mayur R Gamit. et al, «Classification Techniques for Speech Recognition: A Review», International Journal of Emerging Technology and Advanced Engineering, Volume 5, Issue 2, February 2015.
6. M.A. Anusuya, S. K. Katti «Speech Recognition by Machine: A Review» International journal of computer science and Information Security 2009.
7. Wiqas Ghai and Navdeep Singh, (2012). Literature Review on Automatic Speech Recognition. International Journal of Computer Applications, 41(8), 42–50.
8. Santosh K. Gaikwad, Bharti W. Gawali and Pravin Yannawar, (2010). A Review on Speech Recognition Technique. International Journal of Computer Applications (0975–8887).
9. Nidhi Desai, Kinnal Dhameliya and Vijayendra Desai, (2013). Feature Extraction and Classification Techniques for Speech Recognition: A Review. International Journal of Emerging Technology and Advanced Engineering, 3(12), 367–371.
10. Kishori R. Ghule and R. R. Deshmukh, (2015). Feature-Extraction-Techniques-for-Speech-Recognition-A-Review. International Journal of Scientific & Engineering Research, 6(5), 143–147.
11. Shanti Therese S. et al, «Review of Feature Extraction Techniques in Automatic Speech Recognition», International Journal of Scientific Engineering and Technology, Volume No.2, Issue No.6, pp.: 479–484 1 June 2013.
12. Harpreet Singh, et al, « A Survey on Speech Recognition», International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume No. 2, Issue No. 6, June 2013.

© Хеин Мин Зо (heinminnzaw13@gmail.com),

Кудинов Виталий Алексеевич (kudinovva@yandex.ru),

Довгаль Виктор Митрофанович (vmdovgal@yandex.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»