

история ASR

1950-е и 1960-

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Автоматическое распознавание речи

Саша Мартынова

система Audrey

1952 — год рождения систем распознавания речи.

Американская компания Bell Laboratories, крупный исследователь в области компьютерных и электронных систем, представила систему "Audrey". Первое устройство для распознавания речи могло распознавать произнесённые человеком цифры. Система обладала рядом ограничений, невыполнение которых снижало точность Audrey до 60-70 процентов. Если все было сделано по требованиям, точность составляла порядка 90

- диктующий человек должен был быть мужчиной
- этот человек должен был уже ранее работать с системой
- пауза между словами должна была составлять около 350 миллисекунд

история ASR

1950 и 1960

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Shoebox

На всемирной выставке в Сиэтле фирмой IBM был представлен компьютер Shoebox. Эта машина могла распознавать шестнадцать слов на английском языке.



история ASR

1950 и 1960

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

В семидесятых годах интерес к системам распознавания речи резко возрастает.

- 1971 — министерство обороны США выделяет средства на исследования и разработку систем распознавания речи. цель — способность распознавать не менее 1000 слов и понимать связную речь.
- 1972 — первая коммерческая программа Vip-100, которая могла распознавать несколько сот слов, но без поддержки слитной речи.
- 1976 — разработано шесть систем, которые в той или иной степени соответствовали необходимым критериям.

история ASR

1950 и 1960

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Harpy

Самой удачной в то время была система "Harpy разработанная университетом Карнеги Меллона.

- словарный запас в 1011 слов
- понимала связную речь
- эффективный алгоритм поиска правильных конструкций

история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

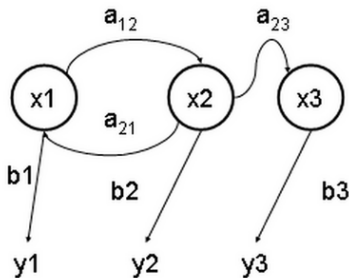
Архитектура
ASR

Акустическая модель

Список
литературы

1980-е: скрытая марковская модель

В 1960 году математик Leonard E. Baum опубликовал первые заметки о "скрытой марковской модели". Её задача — разгадывание неизвестных параметров на основе наблюдаемых. В 1980-е эту модель начали активно использовать в распознавании речи.



x — скрытые состояния, y — наблюдаемые результаты, a — вероятности переходов, b — вероятность результата

Julie doll

В 1987 выходит кукла с функцией распознавания детской речи на основе тренировки. В дополнение к этому она могла реагировать на простые события, происходящие с ней, будь то попадание на свет или темноту, и даже «читать» специальные книги из комплекта, пользуясь сенсорами на пальчиках.



история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Автоматическое распознавание речи в массы

- 1990 — первая общедоступная (695\$) программа Dragon Dictate. Распознавала нормальную слитную речь после 45 минутной тренировки.
- 1996 — первый голосовой портал VAL от BellSouth. Это была первая интерактивная система распознавания речи, которая давала информацию, основываясь на том, что вы сказали в трубку телефона.



история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Застой в распознавании речи?

- 2001 — Microsoft выпускает свою систему распознавания речи, которая работала с Office XP
- 2002 — Google запускает Voice Search (голосовой поиск в интернете). Но чтобы выполнять поиск, требовалось звонить на специальный номер, поэтому на время этот проект сернули.
- 2005-2006 — выходит ряд операционных систем с функцией распознавания речи
- 2009 — выходит приложение Voice Search от Google для iPhone. У Google была возможность разгрузить процесс РР, используя свои облачные дата-центры, где проводился анализа данных для поиска совпадений между словами пользователей и огромного числа образцов голосовых запросов, которые они получали.

история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Застой в распознавании речи?

Почему Voice Search был большим шагом вперед в истории распознавания речи?

история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Застой в распознавании речи?

Почему Voice Search был большим шагом вперед в истории распознавания речи?

Слабым местом распознавания речи всегда была недоступность данных и невозможность их эффективной обработки. Приложение же добавило к анализу данные миллиардов поисковых запросов, чтобы лучше предугадывать, что же сказал пользователь.

Потом появилась Siri. Так же, как и Google Voice Search, она полагается на облачные вычисления. Siri использует данные, которые ей известны о пользователе, чтобы сгенерировать вытекающий из контекста ответ и отвечает на твой запрос, как некая личность.

Распознавание речи превратилось из инструмента в развлечение.

история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Основные задачи

Для успешного распознавания речи следует решить следующие задачи:

- обработку словаря (фонемный состав)
- сокращение речи
- выбор диктора

Каждая система распознавания речи имеет некоторые задачи, которые она призвана решать и комплекс подходов, которые применяются для решения поставленных задач.

история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Обработка звука

Для распознавания речи акустический сигнал фиксируется и преобразовывается в цифровую форму. Непрерывный сигнал переводится в последовательность чисел.

Наиболее популярные методы цифровой обработки речевых сигналов:

- частотный анализ в базисе Фурье
- вейвлет анализ
- кейпстральный анализ
- субполосный анализ

история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Словарь

Частота ошибок системы распознавания напрямую зависит от количества слов в словаре системы распознавания.

Что будет с точностью распознавания при маленьком словаре? А при большом?

история ASR

1950-е и 1960-

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Словарь

Частота ошибок системы распознавания напрямую зависит от количества слов в словаре системы распознавания.

Что будет с точностью распознавания при маленьком словаре? А при большом?

Словарь из 10 цифр может быть распознан практически безошибочно, тогда как частота ошибок при распознавании словаря в 100000 слов может достигать 45%. С другой стороны, даже распознавание небольшого словаря может давать большое количество ошибок распознавания, если слова в этом словаре очень похожи друг на друга.

история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Дикторозависимость или дикторонезависимость системы

По определению, дикторозависимая система предназначена для использования одним пользователем, в то время как дикторонезависимая система предназначена для работы с любым диктором.

Как добиться дикторонезависимости?

история ASR

1950-е и 1960-

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Дикторозависимость или дикторонезависимость системы

По определению, дикторозависимая система предназначена для использования одним пользователем, в то время как дикторонезависимая система предназначена для работы с любым диктором.

Как добиться дикторонезависимости?

Дикторонезависимость – труднодостижимая цель, так как при обучении системы, она настраивается на параметры того диктора, на примере которого обучается. Частота ошибок распознавания таких систем обычно в 3-5 раз больше, чем частота ошибок дикторозависимых систем.

история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Раздельная или слитная речь

Что такое раздельная речь? А слитная?

история ASR

1950 и 1960

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Раздельная или слитная речь

Что такое раздельная речь? А слитная?

Раздельная речь — в речи каждое слово разделяется от другого участком тишины.

Слитная речь – это естественно произнесенные предложения.

Распознавание слитной речи намного труднее в связи с тем, что границы отдельных слов не четко определены и их произношение сильно искажено смазыванием произносимых звуков.

история ASR

1950-е и 1960-

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

диктора

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Различия методов распознавания речи

- Структурные единицы — при анализе речи, в качестве базовой единицы могут быть выбраны отдельные слова или части произнесенных слов.

В зависимости от того, какая структурная часть выбрана, изменяется структура, универсальность и сложность словаря распознаваемых элементов.

- выделение признаков — для представления речевого сигнала из него требуется выделить какие-либо параметры, адекватно представляющие этот сигнал для распознавания.
- механизм функционирования — на основе каких алгоритмов система обучается? (Скрытая марковская модель, Нейронные сети и т.д.)

история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различия методов

Архитектура
ASR

Акустическая модель

Список
литературы

Различия методов распознавания речи



история ASR

1950 и 1960

1970-е

1980-е

1990-е

2000-е

как это работает?

основные задачи

обработка звука

словарь

дикторы

речь

различия методов

Архитектура ASR

Акустическая модель

Список литературы

Архитектура ASR

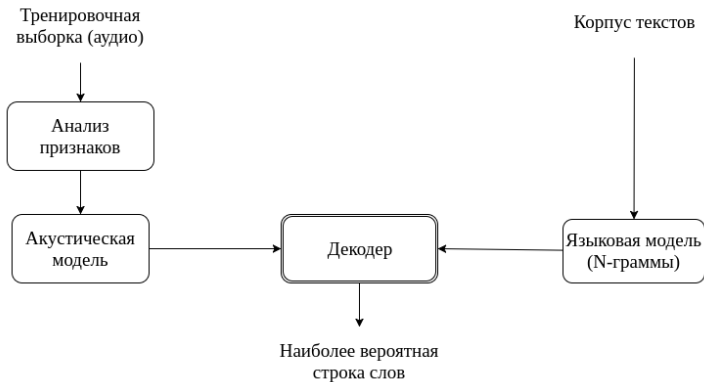


Рис.: Архитектура статистической системы анализа речи

история ASR

1950 и 1960

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Архитектура систем распознавания

- модуль шумоочистки и отделение полезного сигнала
- акустическая модель — позволяет оценить распознавание речевого сегмента с точки зрения схожести на звуковом уровне. Для каждого звука изначально строится сложная статистическая модель, которая описывает произнесение этого звука в речи.

история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Акустическая модель

Соответствующая звуку текстовая транскрипция не синхронизирована с ним, так как у нас нет информации о времени начала и конца слова во фразе.

Для моделирования неопределенности между акустическими характеристиками и соответствующей им транскрипции используют Скрытую Марковскую модель.

Модель для отдельных слов и предложений получается путем конкатенации акустических моделей подсловных единиц (отдельных фонем или, например, трифонов). Подсловные единицы, меньшие чем слово, позволяют распознавать слова, которые не встретились в тренировочной выборке.

история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Архитектура систем распознавания

- языковая модель — позволяет определить наиболее вероятные словесные последовательности. Сложность построения языковой модели во многом зависит от конкретного языка.

Для английского языка, достаточно использовать статистические модели (так называемые N-граммы).

Для высокофлексивных языков (языков, в которых существует много форм одного и того же слова), к которым относится и русский, языковые модели, построенные только с использованием статистики, уже не дают такого эффекта — слишком много нужно данных, чтобы достоверно оценить статистические связи между словами.

Поэтому применяют гибридные языковые модели, использующие правила русского языка, информацию о части речи и форме слова и классическую статистическую модель.

история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Архитектура систем распознавания

- декодер — программный компонент системы распознавания, который совмещает данные, получаемые в ходе распознавания от акустических и языковых моделей, и на основании их объединения, определяет наиболее вероятную последовательность слов, которая и является конечным результатом распознавания слитной речи.

:)

история ASR

1950-е и 1960-е

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Спасибо за внимание

alex250396@gmail.com

история ASR

1950 и 1960

1970-е

1980-е

1990-е

2000-е

как это
работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура
ASR

Акустическая модель

Список
литературы

Список литературы

история ASR

1950 и 1960

1970-е

1980-е

1990-е

2000-е

как это работает?

основные задачи

обработка звука

словарь

дикторы

речь

различные методы

Архитектура ASR

Акустическая модель

Список литературы