УДК 004.93

Структура систем распознавания речи

А.С. Маенков
(Университет ИТМО, г. Санкт-Петербург)
Научный руководитель д.т.н Ю.Н. Матвеев
(Университет ИТМО, г. Санкт-Петербург)

Введение

Все большую популярность применение распознавания речи находит в различных сферах бизнеса, например, врач в поликлинике может проговаривать диагнозы, которые тут же будут внесены в электронную карточку. Или другой пример. Наверняка каждый хоть раз в жизни мечтал с помощью голоса выключить свет или открыть окно. В последнее время в телефонных интерактивных приложениях все чаще стали использоваться системы автоматического распознавания и синтеза речи. В этом случае общение с голосовым порталом становится более естественным, так как выбор в нём может быть осуществлен не только с помощью тонового набора, но и с помощью голосовых команд. Голос это основной способ взаимодействия людей друг с другом. Автоматические системы распознавания конвертируют речь из записанного аудио сигнала в текст. Так как взаимодействие с компьютером с помощью речи намного быстрее, чем с помощью клавиатуры люди предпочтут такую систему. Также подобные системы помогают людям с ограниченными возможностями управлять компьютером. Большинство современных систем основаны на статистических моделях, таких как скрытые Марковские сети. Одна из основных причин популярности скрытых Марковских сетей заключается в том, что, их параметры могут быть оценены автоматически для большого количества данных.

Основная часть

Типична система распознавания речи, состоит из декодера, вычислителя признаков речи, акустической и языковой модели. Приложения взаимодействуют с декодером для получения результатов распознавания. Акустические модели это статистическое представление множества возможностей произношения каждого отдельного звука, из которых собираются слова.[1] Языковая модель в свою очередь это статистическая модель, которая показывает вероятности распределение слов, она учитывает предыдущие слова и сообщает алгоритмам, что чаще всего встречается после этих слов.

Автоматические системы распознавания имеют параметры, по которым их можно классифицировать. Выделяют системы по количеству дикторов, с которым система может работать, например если система может работать с любыми дикторами и любым их количеством то такая система называется «диктора независимая». Разделяют системы и размеру словаря который система может распознать. Если система может распознать малое количество слов, например все цифры, то ‘это система с малым словарем. Когда наш словарь расширяется нескольких сотен слов, мы переходим в классификацию систем с средним словарем. Большие очень большие системы имеют в своем словаре тысячи и десятки тысяч слов, что повышает возможности системы.[2]

Несмотря на десятилетия исследования в этой области системы автоматического распознавания речи все еще не соответствуют возможностям человека. Это связанно, прежде всего с изменчивостью речевого сигнала, люди могут по разному произносить одни и те же звуки, могут тянуть какую-нибудь фонему в слове дольше, это все необходимо учитывать. Распознавание речи является процессом декодирования. Речь может быть представлена как последовательность языковых единиц, называемых фонемами.[3]

Вывод

В этой статье были рассмотрены способы построения систем автоматического распознавания речи, их типы, и проблемы с которыми можно столкнуться при их реализации. Выявлены основные компоненты систем такого рода. В ходе проделанной

работы были изучены структуры систем распознавания речи. Рассмотрены примеры реализации с открытым исходным кодом, такие как kaldi. Изучены и применены на практике методы корректировки времени. Были изучены методы настройки и обучения акустических моделей. Была разработана и протестирована система слитного распознавания речи, написанная на языке программирование Python.

Литература

1. Nitin Indurkhya, Fred J. Damerau. Handbook of Natural Language. –2010. –P. 111–126.
2. Samudravijaya K. Automatic Speech Recognition. –P. 3-4
3. Markus Forsberg. Why is speech recognition difficult. –2003. –P. 3–4