УДК 519.688:004.032.26

КЛАССИФИКАЦИЯ СИСТЕМ РАСПОЗНАВАНИЯ РЕЧИ

Федосин С.А., Еремин А. Ю.

ГОУВПО «Мордовский государственный университет им. Н. П. Огарева», г. Саранск
Тел. +7 (905-3) 893562. E-mail: magbetjke@gmail.ru

Аннотация. В статье анализируются современные подходы к решению задачи распознавания речи и дается классификация систем распознавания.

Ключевые слова: система распознавания речи, анализ Фурье, вейвлет-анализ, скрытые Марковские модели, нейронные сети, алгоритм распознавания.

Задача создания надежной системы распознавания речи, устойчивой к шумам, с низкой частотой появления ошибок, является одной из актуальных на сегодняшний день. Технологии распознавания речи появились весьма давно. Хорошо известны исторические работы Дэвиса, Биддульфа и Балашека (1952), Нагаты, Като и Чибы (1962), Зайцева и Тимофеева (1965), Кинга и Тьюниса (1966), Голда (1966), Величко и Загоруйко (1969). Особенно быстро развитие технологии распознавания речи получили после появления устройств цифровой обработки, выполненных в виде микросхем и позволивших создать относительно дешевые распознаватели, работавшие в режиме реального времени. По мере роста вычислительной мощности сначала специализированных акустических, а затем и цифровых сигнальных процессоров усложнялись и совершенствовались алгоритмы, использовавшиеся в системах распознавания речи . Однако точность систем распознавания речи достигла своего пика в 1999 году и с тех пор застыла на месте. Различные тесты показывают, что современные системы общего профиля так и не преодолели уровень распознавания в 80%, тогда как у человека этот показатель составляет 96-98%. Поэтому крайне необходимо продолжать исследования в этой области.

В данной работе сделана попытка классификации существующих подходов к решению этой проблемы. Классификацию систем распознавания речи следует начать с определения основных аспектов этих систем. К таким аспектам можно отнести:

Размер словаря. Чем больше размер словаря, с которым работает система распознавания речи, тем больше частота появления ошибок при распознавании слов. Для сравнения, словарь, состоящий только из цифр, может быть распознан практически безошибочно, тогда как частота появления ошибок при распознавании словаря в сто тысяч слов может достигать 45%. Но нужно также учитывать уникальность слов в словаре. Если слова очень похожи друг на друга, то погрешность распознавания увеличивается.

Дикторозависимость. Существуют дикторозависимые и дикторонезависимые системы распознавания речи. Дикторозависимая система предназначена для работы только с одним пользователем (человеком, который обучал эту систему), в то время как дикторонезависимая система предназначена для работы с любым диктором. Но создание по-настоящему дикторонезависимой системы – очень трудоемкая задача. На текущем этапе развития систем распознавания речи частота появления ошибок в дикторонезависимых системах в 3-5 раз больше, чем дикторозависимых.

Слитная или раздельная речь. Речь диктора условно можно разделить на слитную и раздельную. Раздельная – это речь, в которой слова отделяются друг от друга определнной паузой (промежутком тишины). Слитная речь – это естественно произнесенные предложения. Распознавание слитной речи сложнее, так как у произносимых слов нет четких границ.

Структурные единицы. В качестве структурных единиц могут выступать фразы, слова, фонемы, дифоны, аллофоны. Системы, которые распознают речь, используя целые слова или фразы, называются системами распознавания речи по шаблону. Они как правило дикторазависимы, и их создание менее трудоемко, чем создание систем, распознающих речь на базе выделения лексических элементов. В таких системах структурными единицами речи являются лексические элементы (фонемы, дифоны, аллофоны).

Принцип выделения структурных единиц. В современных системах распознавания речи используются несколько подходов для выделения из потока речи структурных единиц. Самый распространенный подход основан на преобразовании Фурье, которое переводит исходный сигнал из амплитудно-временного пространства в частотно-временное, а во временной области – линейное предсказание речи, которое описывает речевой сигнал с помощью модели авторегрессии. Однако анализ Фурье обладает целым рядом недостатков, в результате которых происходит потеря информации о временных характеристиках обрабатываемых сигналов. В связи с этим для задачи выделения структурных единиц речи оправданно использование вейвлет-анализа. Фурье-анализ предполагает разложение исходной периодической функции в ряд, в результате чего исходная функция может быть представлена в виде суперпозиции синусоидальных волн различной частоты. В свою очередь вейвлет-анализ раскладывает входной сигнал в базис функций, характеризующих как частоту, так и время. Поэтому с помощью вейвлетов можно анализировать свойства сигнала одновременно и в физическом пространстве, и в частотном. Также, в отличие от традиционного преобразования Фурье, вейвлет-преобразование определено неоднозначно: каждому вейвлету соответствует свое преобразование. Это позволяет тщательнее подобрать вейвлет-функцию с хорошими свойствами частотно-временной локализации. Помимо вейвлет- и Фурье-анализа в системах распознавания речи используется кепстральный анализ, но создание таких систем очень трудоемко и требует очень высокой квалификации разработчика.

Алгоритмы распознавания. После того как речевой сигнал разбивается на определнные части, происходит вероятностная оценка принадлежности этих частей к тому или иному элементу распознаваемого словаря. Это осуществляется по средством одного из алгоритмов распознавания. Наибольшее распространение получили системы распознавания речи на базе скрытых Марковских моделей (СММ). СММ называется модель состоящая из N состояний, в каждом из которых некоторая система может принимать одно из M значений какого-либо параметра. Вероятности переходов между состояниями задается матрицей вероятностей $A=\{a_{ij}\}$, где a_{ij} – вероятность перехода из i-го в j-е состояние. Вероятности выпадения каждого из M значений параметра в каждом из N состояний задается вектором $B=\{b_j(k)\}$, где $b_j(k)$ – вероятность выпадения k-го значения параметра в j-м состоянии. Вероятность наступления начального состояния задается вектором $\pi=\{\pi_i\}$, где π_i – вероятность того, что в начальный момент система окажется в i-м состоянии. Таким образом, скрытой марковской моделью называется тройка $\lambda=\{A,B,\pi\}$. Использование скрытых марковских моделей для распознавания речи основано на двух приближениях: речь может быть разбита на фрагменты, соответствующие состояниям в СММ, параметры речи в пределах каждого фрагмента считаются постоянными; вероятность каждого фрагмента зависит только от текущего состояния системы и не зависит от предыдущих состояний. Кроме СММ, в системах распознавания используются динамическое программирование и нейронные сети. На базе нейронных сетей можно создавать обучаемые и самообучающиеся системы распознавания речи. Такие системы должны отвечать следующим требованиям: разработка системы заключается только в построении архитектуры системы, возможность контроля своих действий с последующей коррекцией, возможность накопления знаний об объектах рабочей области, автономность системы. По сравнению с классическим программированием, когда алгоритм решения той или иной задачи задан жестко, нейронные сети позволяют динамически изменять алгоритм простым изменением архитектуры. Поэтому использование нейронных сетей в задаче распознавания

речи наиболее перспективно. Однако качественный уровень современных систем распознавания речи, основанных на нейронных сетях, еще очень далек от систем, использующих скрытые Марковские модели.

Назначение. Назначение системы определяет необходимый уровень абстракции, на котором будет происходить распознавание речи. Например, в системе голосового набора мобильного телефона будет осуществляться распознавание по шаблону (слову или фразе). Такие системы называются командными. В отличие от них, система диктовки требует более точного распознавания (распознавание на базе выделения лексических элементов) и при интерпретации произнесенной фразы она будет полагаться не только на то, что было произнесено в текущий момент, но и на то, как соотносится с тем, что было произнесено до этого. Также в такую систему должен быть встроен набор грамматических правил. Чем строже эти правила, тем проще реализовывать систему распознавания и тем ограниченней будет набор предложений, которые она сможет распознать.

Обобщив все вышесказанное, можно представить приблизительную классификацию систем распознавания речи (рис. 1).

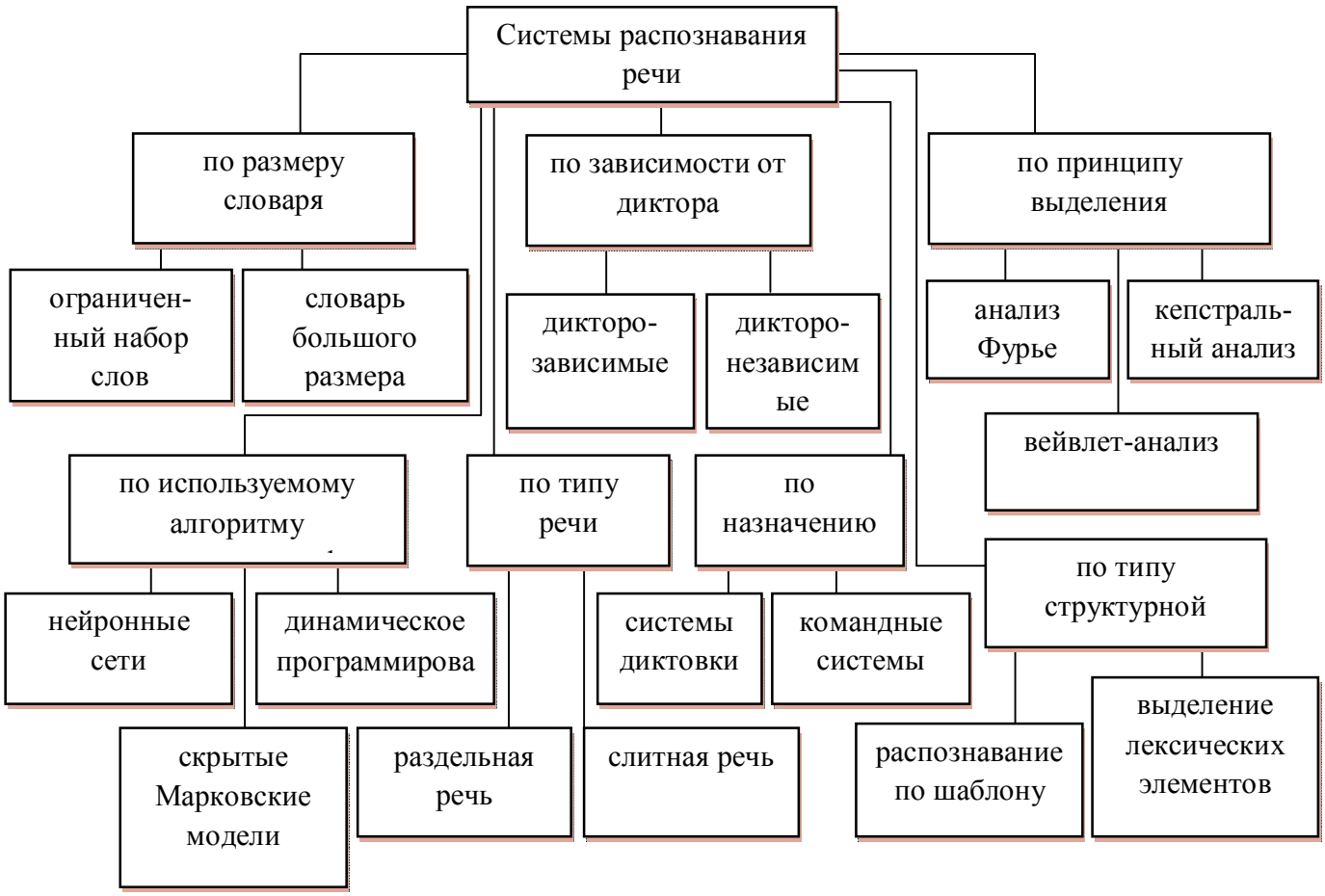

Рис. 1. Классификация систем распознавания речи.

Выводы

Анализ основных аспектов систем распознавания речи показал, что в настоящее время не существует универсальной системы, которая бы была самообучаемой, дикторонезависимой, устойчивой к шумам, распознающей слитную речь, способной работать со словарями больших размеров и при этом иметь низкую частоту появления ошибок. Представленная в данной работе классификация систем распознавания речи позволит сузить область исследований в этом направлении при разработке.

Литература

1. Burger S., Sloane Z., Yang. J. Competitive Evaluation of Commercially Available Speech Recognizers in Multiple Languages / Susan Burger, Zachary Sloane, Jie Yang. – Pittsburgh: Carnegie Mellon University, 2006. – 6 p.

2. Xuedong H. Spoken Language Processing: A Guide to Theory, Algorithm and System Development / Huang Xuedong. – New Jersey: Prentice Hall PTR, 2001. – 1008 p.

3. Фролов А., Фролов Г., Синтез и распознавание речи. Современные решения [Электронный ресурс] / Александр Фролов, Григорий Фролов. – Электрон. журн. – 2003. – Режим доступа: http://www.frolov-lib.ru

4. Чесебиев И.А. Компьютерное распознавание и порождение речи / И.А. Чесебиев. – М.: Спорт и культура, 2008 – 128 с.