Технические науки

АНАЛИЗ ПРОБЛЕМ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ

И. Ю. Беседин

ANALYSIS OF SPEECH AUTOMATIC IDENTIFICATION PROBLEMS

Besedin I. Y.

The limitations of computer features in the process of modeling speechindentificatio are described. The classification of speech identification systems is presented. Widespread approaches to the modeling of speech identification are described, language models sre considered. The research has been made in the frames of the "Scientific and Scientific-Pedagogical Personnel of Innovational Russia" Federal Program.

Key words: speech identification, language models, statistical models.

Описаны ограничения возможностей ЭВМ по моделированию процесса распознавания речи. Приведена классификация систем распознавания речи. Описаны распространенные подходы к моделированию распознавания речи, рассмотрены языковые модели. Исследования выполнены в рамках ФЦП «Научные и научно-педагогические кадры инновационной России».

Ключевые слова: распознавание речи, языковые модели, статистические модели.

УДК 51.74

Создание систем автоматической обработки речи является одним из наиболее актуальных направлений развития современных информационных технологий. Многие университеты, вычислительные центры и ряд прочих организаций занимаются разработкой систем автоматической обработки речи, включающие в себя распознавание речи на различных языках, поиск ключевых слов, системы идентификации дикторов, системы идентификации языков, оценку качества записей речи, системы изменения голоса.

Задача распознавания речи состоит в восстановлении по звуковому сигналу слов естественного языка, произнесением которого является исходный звуковой сигнал. Она обычно решается путем задания эталонов слов из словаря и последующего сравнения звуковых сигналов с этими эталонами. Для решения обычно сначала равномерно разбивают сигнал на окна одинаковой длины. Окна преобразуют из временной области в частотную. Затем решается задача нахождения соответствия между окнами звукового сигнала и окнами эталонов слов в словаре с учетом возможной различной степени сжатия и растяжения одинаковых слов.

В [1] приводится классификация основных проблем при создании систем распознавания речи. Прежде всего это фильтрация шума и помех и представление речевого сигнала в цифровой форме. Распознавание во входном потоке фонем, морфем, слогов, слов и прочих языковых единиц требует понимания принципов построения речи. Сле-

дует учитывать особенности звуков сплошной речи – постоянно изменяющийся спектр гармонических частот, шум; непостоянная громкость, темп, различная спектрально-временная окраска одной и той же фразы, сказанная одним человеком, находящимся в разных психических состояниях.

В первых системах распознавания речи использовали ряд грамматических и синтаксических правил речи. Если произнесенные слова соответствовали записанным в программе правилам, то система могла определить, какое используется слово. Однако, разговорная речь очень сильно отличается от установленных правил языка. Различные акценты, диалекты и особенности произношения звуков отдельных людей значительно затрудняли работу программы.

Ограничения ЭВМ не могут позволить с требуемой надёжностью исправлять ошибки и неоднозначности распознавания, используя синтаксическую и семантическую связь слов предложения. Вместо этого обычно используют статистические модели. Такие системы с помощью теории вероятности и математических вычислений способны определить наиболее вероятный вариант. На сегодняшний день существует две такие модели распознания речи – скрытые Марковские модели и модель нейронных сетей [5]. Принцип их работы заключается в обработке известной системе информации и извлечение из неё скрытой информации с помощью вычислений.

Консорциум Всемирной Паутины (W3C) разработал в 2004 году стандарты Грамматик для распознавания [7] и синтеза [6] речи. Эти документы определяют основные характеристики речи. Согласно этим стандартам можно классифицировать системы распознавания речи по следующим признакам:

- интервал между отдельными словами: если система распознает непрерывную речь, пользователь может произносить речь естественно. В системах с дискретной речью паузы между словами должны составлять не менее четверти секунды.
- зависимость от диктора: системы распознавания, независимые от диктора, позволяют работать без предварительной настройки, но позволяют улучшить надёжность распознавания после обучения. Подобная независимость достигается за счет хранения звуковых эталонов для всех наиболее типичных голосов носителей языка.
- степень детализации при задании эталонов: различные алгоритмы распознавания речи могут использовать для распознавания речи как эталоны слов, так и эталоны языковых единиц.
- размер словаря: различают маленькие словари (около 50 слов), позволяющие передавать команды компьютеру; средние (порядка 1000 слов), достаточные для определения «активного» словаря; большие (более 10 000 слов) для диктовки текстов.

Большинство современных систем распознавания речи, такие как Dragon Power Edition, IBM Voice TYPE, OfficeTALK 3.0, KVWin 2.0, MicroIntrovoice и пр., используют сходные алгоритмы и методы [2]. Специфика задачи и существующие вычислительные ограничения определяет разницу в типе диктовки речи, размере словаря.

И. Л. Мазуренко в [2] описывает процесс распознавания речи в следующей последовательности.

Шумоочистка и отделение полезного сигнала (выделение инвариантных относительно шума признаков, обучение в условиях шума, модификация эталонов). Узким местом подобных методов является эффект ненадёжной работы систем распознавания в «бесшумных» условиях.

Одним из методов являются коэффициенты линейного предсказания [3]. В качестве элементов эталонов используют вероятностные распределения. Для получения инвариантных признаков часто используют кратковременную функцию когерентности или методы связанных с моделированием слуховой системы человека.

Преобразование входного речевого сигнала в набор акустических параметров. Заключается в преобразовании в частотную область с помощью преобразования Фурье разбитого на окна звукового сигнала.

Приведение акустической формы сигнала к внутреннему алфавиту эталонных элементов. Область значений акусти-

ческих параметров речи разбивают на области сгущения, которые соответствуют элементам фонем, одинаковым для различных слов данного языка. Обычно таких областей около 1 000 [6], и для большого словаря целесообразно в качестве эталонов использовать лишь фонемные элементы.

Распознавание последовательности фонем и преобразование ее в текст. После определения вероятной последовательности эталонных элементов необходимо восстановить по ней неизвестную последовательность фонем. Подобные задачи решаются с помощью метода динамического программирования [3].

Естественный язык – это результат многовековой параллельной работы огромного числа носителей языка. Его предложения принципиально отличаются от случайных комбинаций слов и от предложений формально построенных языков. Одной из основных его особенностей является избыточность, позволяющая понимать искажённую речь [4].

Для славянских языков характерно большое количество словоформ слова, свободный порядок слов. Эти особенности затрудняют использование статистических подходов. Модель языка представляет собой распределение вероятности на множестве всех предложений языка.

В таблице 1 представлены языковые модели, используемые сегодня в системах распознавания речи с неограниченными словарями.

Таблица 1

Тип	Достоинства	Недостатки
n-граммы	Возможность построения модели по обучающему корпусу большого размера. Высокая скорость работы	Не учитывается зависимость очередного слова от размера всего текста
Модели, основанные на деревьях решений	Теоретическая способность показать существенное улучшение с n-граммными моделями	Требуется большой объем анализируемой информации. Не известно ни одной практической реализации
Модели, основанные на теории формальных языков	Высокая скорость грамматического разбора	Игнорирование предложений, не укладывающихся в правила формального языка
Адаптивные модели	Возможность корректировки модели по мере работы распознавателя	

А. Б. Холоденко в своей работе [4] предлагает один из возможных подходов к решению проблем, препятствующих созданию промышленных систем распознавания слитной речи для русского языка. Предложенное в ней разложение общей языковой модели на две составляющие: модель, основанную на морфологии, и модель, основанную на начальных формах слов, позволяет разработчикам использовать все преимущества n-граммного подхода, а выделение морфологической информации в независимую модель позволяет справиться с проблемой акустической похожести различных словоформ одного и того же слова.

В настоящий момент достаточно сложными элементами при построении системы распознавания речи является построение акустической модели языка и начальное обучение эталонов слов словаря. Необходимо тщательно учесть многие типы голосов, акцентов носителей языка.

Кроме этого, существует еще ряд проблем [2]:

- подавление стационарных и нестационарных помех;

- переход к распознаванию непрерывной речи;
- учет контекста;
- поиск новых звуковых параметров;
- перенос на другие языки;
- поиск новых алгоритмов восстановления последовательности произнесенных звуков.

Исследование проблем автоматического распознавания речи является важным фундаментальным направлением. Эта проблема сейчас сдерживает развитие различных прикладных систем в телекоммуникациях, медицине, образовании и повседневной жизни. Практически вся современная техника и различные сервисы используют автоматизированные средства управления и обработки информации, системы развиваются в направлении повышения комфортности общения человека с компьютером, поэтому разработка эффективных средств человеко-машинного взаимодействия является первоочередной научной задачей.

ЛИТЕРАТУРА

1. *Мазуренко И. Л. Компьютерные системы распознавания речи, Интеллектуальные системы. Т. 3. Вып. 1–2. – М., 1998. – С. 117–134.*
2. *Рабинер Л. Р., Шафер Р. В. Цифровая обработка речевых сигналов / пер. с англ. – М.: Радио и связь, 1981.*
3. *Фролов А. В., Фролов Г. В. Синтез и распознавание речи. Современные решения, 2003.*
4. *Холоденко А. Б. О построении статистических языковых моделей для систем распознавания русской речи // Интеллектуальные системы. Т. 6. Вып. 1–4. – М., 2002. – С. 381–394.*
5. *Juang B. H. Speech Recognition in Adverse Environments. Computer Speech and Language. Vol.5. N. Y. 2002. – P. 275–294.*
6. *Speech Recognition Grammar Specification Version 1.0 http://www.w3.org/TR/speech-synthesis/*
7. *Speech Synthesis Markup Language (SSML) Version 1.0 http://www.w3.org/TR/speech-grammar/*

Об авторе

Беседин Игорь Юрьевич, ГОУ ВПО «Ставропольский государственный университет, аспирант, ведущий программист Ставропольского регионального центра информатизации. Сфера научных интересов - распознавание речи, реконструкция моделей, высокопроизводительные вычисления.
besedin.igor@gmail.com