

АНАЛИЗ ОСОБЕННОСТЕЙ РАСПОЗНАВАНИЯ РЕЧИ

© 2017 Н. Н. Гостева, А. В. Гусев, Т. С. Гурьева

Воронежский институт высоких технологий (г. Воронеж, Россия)

ОАО Сбербанк России (г. Воронеж, Россия)

ОАО концерн «Созвездие» (г. Воронеж, Россия)

В работе проводится анализ возможностей распознавания речевых сигналов. Отмечены некоторые характеристики алгоритмов распознавания, указаны критерии распознавания.

Ключевые слова: речевой сигнал, анализ, распознавание, ошибка.

В системах, позволяющих проводить распознавание речевых сигналов, содержащих в своем составе объекты-слова, процессы распознавания базируются на принципе сравнения данных слов в различных словарях. Важно определить временные масштабы относительно двух слов при условии того, что будет достигаться оптимальное соответствие.

В данной работе нами предпринята попытка анализа возможностей распознавания отдельных слов при практической реализации соответствующих подходов. Существуют трудности, связанные с созданием алгоритмов по автоматическому выделению речевых составляющих из речевых потоков, поскольку разные люди характеризуются различным произношением. На сегодняшний момент для распознавания речи активно используют статистические методики. Речевые компоненты могут быть описаны на базе того, что вводят гауссову модель сигнала. Слова при моделировании представляются как совокупность компонентов, для пространства измерений проводится расчет расстояний от этих компонентов до искомым реализаций речевых сигналов.

Когда формируются адаптивные алгоритмы распознавания, требуется применять речевые компоненты, которые дают возможности для отслеживания медленных изменений по речевым сигналам. Методики, связанные с исправлением ошибок, основываются на автоматическом обнаружении некорректных последовательностей сигналов. Помимо автоматического исправления оши-

бок, необходимо отслеживание ошибок пользователей. Важно иметь в виду, что в процессе обработки текстовых данных следует использовать комплексные информационные системы. Есть возможности для ускорения алгоритмов, повышения скорости распознавания речевых сигналов.

С этой целью ключевыми можно считать следующие критерии:

- 1) проведение анализа размеров используемой лексики;
- 2) временные интервалы, в течение которых идет адаптация систем;
- 3) параметры речевой спонтанности в распознаваемых речевых сигналах;
- 4) общее время распознавания.

Среди основных направлений в распознавании речевых сигналов можно указать такое, которое связано с представлением каждого слова на основе одного или нескольких эталонов в пространстве измерений и вычислении расстояния относительно эталонов до неизвестных реализаций речевого сигнала. Для простейшего случая, когда отклонения реализаций некоторых слов от эталона порождаются на основе случайного процесса при нормальном распределении, оптимальным будет вычисление расстояний относительно евклидовой метрики, которое реализуется, например, в виде коэффициента корреляции. Весьма часто выясняется, что помимо случайных отклонений в речевых сигналах существуют и детерминированные компоненты искажений. Среди подобных неслучайных искажений можно указать различие в длительности произнесения любых слов различными дикторами, причем компенсировать такое различие невозможно на основе линейного сжатия или растяжения оси времени. В конце 1960-х годов для того, чтобы измерять степень сходства среди реализацией и эталоном слов были предложе-

Гостева Нина Николаевна – ВИБТ-АНОО ВО, студент, gossste0789fgd@yandex.ru.

Гусев Алексей Витальевич – ОАО Сбербанк России, специалист, u8u5vgh3@mail.ru.

Гурьева Татьяна Сергеевна – ОАО концерн «Созвездие», специалист, gurgprofkq7b52@yandex.ru.

ния применять динамическое программирование, которое давало возможности для отыскания минимума расстояния в заданной метрике при учете нелинейных деформаций на временной оси. В дальнейшем метод динамического программирования распространили на полунепрерывное относительно времени представление речевого сигнала как последовательность символов сегментов из конечного множества. Использование динамического программирования дало возможности для достижения весьма высокой (до 97-99 %) степени надежности в распознавании изолированных слов при настройке на диктора в словарях объемом до 100-200 слов. Однако такой метод имеет некоторые принципиальные недостатки, которые затрудняют или исключают его практическое применение во многих задачах.

Так, к недостаткам можно отнести:

- зависимость надежности распознавания от типа микрофона и расстояния до него;
- необходимость в обновлении эталонов каждые несколько месяцев;
- большие вычислительные затраты;
- невозможность распознавать слитную речь и больших словарей и т. д.

Слабость характеристик, которые рассматриваются в методе динамического программирования, состоит в том, что он универсален, т. е. в нем очень малым образом используются особенности речевых сигналов.

Например, для того, чтобы достичь успешного распознавания, требуется назначить меру сходства среди сигналов, а ее не определяют в методе динамического программирования, ее задают извне.

Необходимость распознавать слитную речь вне зависимости от дикторов, потребности в увеличении надежности на основе перехода к помехоустойчивому параметрическому представлению речевых сигналов, определили необходимость в поиске других способов сравнения между реализациями и эталонами.

То, что речь необходима для передачи, а поэтому, и для защиты информации, дает возможности рассматривать ее в виде некоторого кода, а речевой поток – в виде последовательности определенных кодовых посылок.

Что является элементом этого кода – слоги, фонемы, артикуляторные движения или характерные сегменты, в данном случае неважно. Важным будет только то, что вероятность в появлении любого из элементов

кода определяется некоторым числом предыдущих элементов.

В результате, речь порождается на основе Марковского источника, а речевой код является случайным кодом.

Такая точка зрения является весьма конструктивной для того, чтобы осуществлять поиск способов распознавания речевых сигналов. Прежде всего, она определяет метод декодирования – для случайных кодов сейчас применяют лишь последовательное декодирование.

С точки зрения математического смысла, оно представляет собой специфическую разновидность динамического программирования, есть возможности найти максимум сходства между эталонным кодовым словом и искаженной последовательностью элементов кода (символов).

В качестве примера, алгоритм Витерби является просто вероятностным вариантом в динамическом программировании.

Для такого подхода также может быть определен характер действий и требуемая информация для того, чтобы справиться с искажениями в некоторой последовательности символов.

Понятно, что может быть только три вида искажений: выпадение, вставка и замена символов. Поэтому требуется информация о том, какие вероятности таких событий.

И при этом более определенное будет и понятие меры сходства – получается функция правдоподобия, которая вычисляется на основе некоторого способа.

Параллельным образом с разработкой методов декодирования шло развитие метода скрытых Марковских моделей, который позволил сделать автоматизацию поиска вероятностных характеристик Марковских моделей в речи. Скрытой Марковской моделью можно назвать Марковский процесс, который мы не наблюдаем непосредственным образом. Есть искажение результатов действия такого процесса за счет некоторого случайного процесса и только после этого эти результаты становятся доступными наблюдению.

ЛИТЕРАТУРА

1. Вековищева К. В. Исправляющая способность некоторых кодов / К. В. Вековищева, В. В. Костюченко // Моделирование, оптимизация и информационные технологии. – 2017. – № 1 (16). – С. 9.

2. Вековищева К. В. Распознавание изображений сигналов, имеющих сложную форму / К. В. Вековищева, В. В. Костюченко

// Моделирование, оптимизация и информационные технологии. – 2017. – № 1 (16). – С. 17.

3. Воронов А. А. Обеспечение системы управления рисками при возникновении угроз информационной безопасности / А. А. Воронов, И. Я. Львович, Ю. П. Преображенский, В. А. Воронов // Информация и безопасность. – 2006. – Т. 9. – № 2. – С. 8-11.

4. Головинов С. О. Цифровая обработка сигналов / С. О. Головинов, С. Г. Миронченко, Е. В. Щепилов, А. П. Преображенский // Вестник Воронежского института высоких технологий. – 2009. – № 4. – С. 064-065.

5. Ермолова В. В. Методика построения семантической объектной модели / В. В. Ермолова, Ю. П. Преображенский // Вестник Воронежского института высоких технологий. – 2012. – № 9. – С. 87-90.

6. Зазулин А. В. Особенности построения семантических моделей предметной области / А. В. Зазулин, Ю. П. Преображенский // Вестник Воронежского института высоких технологий. – 2008. – № 3. – С. 026-028.

7. Зяблов Е. Л. Построение объектно-семантической модели системы управления / Е. Л. Зяблов, Ю. П. Преображенский // Вестник Воронежского института высоких технологий. – 2008. – № 3. – С. 029-030.

8. Зяблов Е. Л. Разработка лингвистических средств интеллектуальной поддержки на основе имитационно-семантического моделирования / Е. Л. Зяблов, Ю. П. Преображенский // Вестник Воронежского института высоких технологий. – 2009. – № 5. – С. 024-026.

9. Кленяева Г. В. Современные проблемы речевой акустики и построения систем автоматического распознавания речи / Г. В. Кленяева, А. П. Преображенский // Вестник Воронежского института высоких технологий. – 2007. – Т. 1. – № 2-1. – С. 071-074.

10. Латушко Е. И. Моделирование смысловой структуры текста в процессе обучения иноязычному опосредованному общению / Е. И. Латушко // Фундаментальные исследования. – 2004. – № 2. – С. 61-62.

11. Мурашкин Н. Г. Проблемы использования искусственных нейронных сетей для решения задач бинарной классификации / Н. Г. Мурашкин, В. Н. Кострова // Моделирование, оптимизация и информационные технологии. – 2017. – № 2 (17). – С. 5.

12. Пеньков П. В. Экспертные методы улучшения систем управления / П. В. Пеньков // Вестник Воронежского института высоких технологий. – 2012. – № 9. – С. 108-110.

13. Преображенский Ю. П. Разработка методов формализации задач на основе семантической модели предметной области / Ю. П. Преображенский // Вестник Воронежского института высоких технологий. – 2008. – № 3. – С. 075-077.

14. Чопорова Е. И. Структурные трансформации иноязычного текста в процессе его реферирования: комплексный подход / Е. И. Чопорова, Е. И. Мещерякова, Н. Н. Серостанова // Вестник Воронежского государственного технического университета. – 2013. – Т. 9. – № 5-2. – С. 127-129.

THE ANALYSIS OF FEATURES OF SPEECH RECOGNITION

© 2017 N. N. Gosteva, A. V. Gusev, T. S. Guryeva

Voronezh Institute of High Technologies (Voronezh, Russia)

OJSC Sberbank of Russia (Voronezh, Russia)

JSC concern «Sozvezdie» (Voronezh, Russia)

In the paper analysis of the recognition of speech signals is carried out. Some of the characteristics of recognition algorithms specified criteria are noted.

Keywords: speech signal, analysis, recognition, error.