УДК 004.934
В.В. Пилипенко
Международный научно-учебный центр информационных технологий и систем,
г. Киев, Украина,
valery_pylypenko@mail.ru

ТЕХНОЛОГИЯ РАСПОЗНАВАНИЯ БОЛЬШОГО КОЛИЧЕСТВА ОБРАЗОВ НА ПРИМЕРЕ РАСПОЗНАВАНИЯ РЕЧИ ИЗ СВЕРХБОЛЬШИХ СЛОВАРЕЙ

АННОТАЦИЯ

В статье рассматривается технология отбора кандидатов для распознавания изолированных слов на основе анализа результатов пофонемного распознавания речи (фонетического стенографа). Приведены результаты экспериментов с системой, содержащей практически все слова языка (около 2 млн. слов).

ВВЕДЕНИЕ

Основные теоретические выкладки в распознавании образов касаются разделения образов на два класса. Обычно считается, что при увеличении числа образов возникают чисто технические трудности. Чаще всего такой подход приводит к полному перебору по числу классов.

Современное распознавание речи основывается на представлении произнесенного слова в виде последовательности фонем, задаваемой фонетической транскрипцией. Алгоритм динамического программирования (ДП) вычисляет наилучшее соответствие между входным речевым сигналом и одной из транскрипций [1]. Для нахождения ответа распознавания производится полный перебор среди допустимого множества транскрипций.

ДЕКОМПОЗИЦИЯ ЗАДАЧИ

При большом количестве альтернатив (больше 100 тыс.) предлагается рассматривать поиск наилучшего слова как композицию из двух задач:

1) Доступ к базе данных для нахождения допустимого множества транскрипций.

2) Поиск наилучшей транскрипции средствами ДП.

Первая задача решается при помощи пробного распознавания на основе процедуры фонетического стенографирования.

БАЗОВАЯ СИСТЕМА РАСПОЗНАВАНИЯ

Предложенный метод можно применить в любой системе распознавания речи, где представлены фонемы и можно сформировать процедуру фонетического стенографа. В данной работе как базовая система используется инструментарий HTK [2] на основе скрытых Марковских моделей (HMM).

Предварительная обработка речевого сигнала

Речевой сигнал преобразуется в последовательность векторов признаков с интервалом анализа 25 мс и шагом анализа 10 мс. Вначале речевой сигнал фильтруется фильтром высоких частот с характеристикой $P(z)=1-0.97z^{-1}$ и применяется окно Хэмминга. Быстрое преобразование Фурье переводит временной сигнал в спектральный вид. Спектральные коэффициенты усредняются с использованием 26 треугольных окон, расположенными в мел-шкале. 12 кепстральных коэффициентов вычисляются при помощи обратного косинусного преобразования.

Логарифм энергии добавляется в качестве 13-го коэффициента. Эти 13 коэффициентов расширяются до 39-мерного вектора параметров путем дописывания первой и второй разностей от коэффициентов соседних по времени. Для учета влияния канала применяется вычитание среднего кепстра.

Акустическая модель

Акустические модели отражают характеристики основных единиц распознавания. Для акустических моделей используются скрытые Марковские модели с 64 смесями Гауссовских функций плотности вероятности. 47 русских контекстно-независимых фонем моделируются тремя состояниями Марковской цепи с пропусками.

Словарь транскрипций был создан автоматически из орфографического словаря с использованием множества контекстно-зависимых правил.

Показатели базовой системы

Акустические модели обучались на выборке из 10 тыс. звуковых записей из словаря в 2 тыс. слов, произнесенных одним диктором. Распознавание 1000 изолированных слов производилось на компьютере P-IV 2.4 ГГц. Пословная надежность распознавания и среднее время распознавания для различных размеров словаря приведены в таблице 1 во второй колонке.

АЛГОРИТМ РАСПОЗНАВАНИЯ БОЛЬШОГО КОЛИЧЕСТВА СЛОВ

Такие вычислительные блоки как *предварительная обработка сигнала* и *вычисление параметров акустических моделей* используются из базовой системы. На втором проходе алгоритма также используется блок *сравнение образов* средствами ДП. Изменения касаются дополнительного первого прохода алгоритма, где используется *фонетический стенограф* для получения последовательности фонем.

Фонетический стенограф

Алгоритм фонетического стенографа [3], [4] позволяет строить последовательность фонем для речевого сигнала без использования какого-либо словаря. Для этой цели строится некоторая генеративная грамматика, которая может синтезировать все возможные модельные сигналы непрерывной речи для любой последовательности фонем. В рамках построенной модели строится алгоритм пофонемного распознавания для неизвестного сигнала. Используются те же контекстно-независимые модели фонем, как и в базовом распознавателе.

Надежность найти фонему на правильном месте для известной реализации равна приблизительно 85%.

Процедура получения подсловаря из БД

Заранее из словаря транскрипций строится таблица индексов троек фонем, в которой указываются те транскрипции, в которые входит данная тройка фонем. Таким образом, ключом индекса является тройка фонем, и каждый вход таблицы содержит список транскрипций, в которые входит этот ключ. Индекс состоит из M^3 входов, где M есть число фонем в системе.

Результат фонетического стенографирования последовательно со сдвигом в одну фонему делится на пересекающиеся тройки фонем. Каждая тройка становится одним запросом к БД. В настоящей системе используется простой запрос, в котором тройка фонем совпадает с запросом. В будущем можно было бы учесть вставки, удаления и замены в последовательности фонем при помощи расстояния *Levensteine*. Последовательность фонем формирует поток запросов к БД.

Ответ на один запрос (тройку фонем) к БД состоит из списка транскрипций, в которую включается данная тройка фонем. Этот список копируется в подсловарь для

второго прохода алгоритма. Последующие запросы добавляют новые порции слов. Для вычисления ранга слова поддерживается счетчик повторений слов.

В получившимся подсловаре все транскрипции ранжируются в соответствии с рангом слова (счетчиком повторений). Первые N транскрипций копируются в окончательный подсловарь для второго прохода. Таким образом, подсловарь для распознавания состоит из транскрипций с наивысшим рангом и их число не превышает фиксированное число N, оптимальное значение которого подбирается в результате экспериментов.

Алгоритм ELVIRS

Алгоритм ELVIRS (Extra Large Vocabulary Speech recognition based on the Information Retrieval) представляет собой следующее.

Этап подготовки:

- Подготовить словарь для распознавания.
- Выбрать множество фонем и построить транскрипции слов из словаря пользуясь правилами преобразования.
- Создать индекс БД от троек фонем к транскрипциям.
- Обучить акустические модели по накопленным речевым сигналам.

Этап распознавания:

- Использовать пофонемный распознаватель (фонетический стенограф) для входного речевого сигнала для получения последовательности фонем.
- Поделить последовательность фонем на пересекающиеся тройки фонем.
- Создать запросы из троек фонем.
- Получить списки транскрипций из БД в ответ на запросы.
- Отсортировать транскрипции по рангу.
- Выбрать первые N транскрипций с наивысшими рангами в качестве словаря распознавания.
- Распознать входной речевой сигнал в условиях ограниченного подсловаря.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

В инструментарии HTK были сделаны необходимые модификации для учета первого прохода алгоритма.

Словарь распознавания составлял 15 тыс., 95 тыс. и 1987 тыс. транскрипций. Надежность распознавания и среднее время распознавания приведены в таблице 1. Для словаря в 1987 тыс. слов время распознавания получено экстраполяцией времени для меньших словарей.

Таблица 1: Сравнение результатов распознавания речи из больших словарей.

Объем словаря, тыс	2-х проходный алгоритм		Полный перебор	
	Надежность, %	Время, сек	Надежность, %	Время, сек
15	95.5	1.7	97.9	16
95	89.7	2.5	94.7	115
1987	84.8	6.8	-	>2300

ЗАКЛЮЧЕНИЕ

Получено значительное сокращение времени распознавания (в десятки раз) при относительно малом ухудшении надежности распознавания (около 5%) в сравнении с базовой системой распознавания речи.

Проводятся эксперименты по распознаванию слитной речи.

ЛИТЕРАТУРА

1. Винцюк Т.К. Анализ, распознавание и смысловая интерпретация речевых сигналов // Киев, Наукова думка, 1987, 264 с.

2. S. Young, G. Evermann, D. Kershaw, G. Moore, J. Odell, D. Ollason, V. Valtchev and P. Woodland, The HTK Book // Cambridge University Engineering Department, 2005, 354p.

3. Taras K. Vintsiuk. Generalized Automatic Phonetic Transcribing of Speech Signals // Труды пятой всеукраинской международной конференции “Оброблення сигналів і зображень та розпізнавання образів”, Видання УАсОІРО, 2000, Київ, с. 95-98.

4. Пилипенко В.В. Використання фонетичного стенографа при розпізнаванні мовлення з великих словників // Тезисы 12-й международной конференции “Автоматика - 2005”, Харьков, 2005, с. 73.

В.В. Пилипенко

Технологія розпізнавання великого обсягу образів на прикладі розпізнавання мовлення з надвеликих словників

Розглядається технологія відбору кандидатів для розпізнавання ізольованих слів на базі аналізу результатів пофонемного розпізнавання мовлення (фонетичного стенографу). Наведені результати експериментів з системою, що включає практично всі слова мови (приблизно 2 млн. слів).

V.V. Pylypenko

The technology for large number patterns recognition by the example of extra large vocabulary speech recognition

This paper gives the technology for pattern subset selection based on the analysis of phoneme transcriber output. Experimental results for speech recognition system with vocabulary of about all words (approximately 2 M) are presented.