УДК 621.391

О. Н. Карпов, д. т. н., проф.;
О. А. Савенкова, асп.

РАСПОЗНАВАНИЕ РЕЧИ НА ОСНОВЕ СЕГМЕНТНО-СЛОГОВОГО СИНТЕЗА

Исследованы основные методы поиска в пространстве состояний для решения задачи распознавания речи на основе сегментно-слогового синтеза траектории параметров. Распознавание реализовано как сопоставление траекторий параметров в выбранных речевых единицах на участках сегментированной речи.

Нахождение решения прикладных задач, формулируемых в терминах пространства состояний с помощью алгоритмов поиска, является актуальной проблемой искусственного интеллекта. Теория поиска в пространстве состояний дает ответы на такие вопросы как: гарантировано ли нахождение решения в процессе поиска, является ли поиск конечным, является ли найденное решение оптимальным. Основные алгоритмы поиска и методы их усовершенствования приведены в [1], [2]. На сегодняшний день важное значение имеет оптимизация существующих методов поиска для решения различных прикладных задач, например, в работе [3] предложены стратегии оптимизации методов поиска для решения задач планирования. В работе исследуется возможность применения основных стратегий поиска для решения задачи распознавания речи на основе синтеза траектрий параметров.

Введение

Порядок, в котором происходит развертывание состояний, определяется стратегией поиска. Выбор стратегии поиска решения зависит от определения задачи. Различают стратегии неинформированного и информированного (эвристического) поиска. В стратегиях неинформированного поиска не используется дополнительная информация о состояниях, кроме той, которая представлена в определении задачи. Основные стратегии неинформированного поиска, определяющие порядок рассмотрения альтернатив: поиск в ширину и поиск в глубину. При поиске в ширину наиболее сложной проблемой по сравнению со значительным временем выполнения является обеспечение потребностей памяти. Алгоритм характеризуется временной и пространственной сложностью $O\left(b^d\right)$, где b — коэффициент ветвления, d — глубина самого поверхностного решения. Поиск в глубину характеризуется временной сложностью $O\left(b^m\right)$ и пространственной сложностью $O(bm)$, где m — максимальная глубина любого пути в пространстве состояний [1].

Для решения задачи распознавания речи на основе сегментно-слогового синтеза траекторий параметров, сформулированного в терминах пространства состояний, выбор метода поиска решения зависит, прежде всего, от размеров пространства поиска. Размер пространства состояний определяется количеством сегментов предъявленной реализации речевого сигнала (РС). Если пространство состояний невелико, достаточно применить стратегии поиска в ширину или в глубину.

Постановка задачи

Задачу сегментно-слогового распознавания формулируем согласно [4]. Пусть задан словарь $\{SL_k\}$, состоящий из N слогов. Для сегментации РС используется метод верификации временной последовательности параметров. Для каждого слога заданы эталонные последовательности параметров $\{Y_k\}$. Каждый слог SL_k содержит n_k сегментов-фонем $SG_{kj}^{v}\left(k = 1 \div N,\ \ j = 1 \div n_k\right)$.

Пусть задана входная последовательность X, состоящая из m_p сегментов-фонем SG_{pi}^{X}, объединенных в M групп-слогов $SL_p^X\left(p=1\div M,\ \ i=1\div m_p\right)$.

Необходимо последовательность X наилучшим образом поставить в соответствие эталонным последовательностям $\{Y_k\}$, вычисляя расстояние d

$$d=\sum_p \min_k \left(SL_p^X \# SL_k\right),$$

где SL_p^X, SL_k содержат сегменты SG_{pi}^{X}, SG_{kj}^{Y} соответственно; # — операция сопоставления осуществляется с помощью процедур динамического программирования.

Основная часть

Для обеспечения наилучшего соответствия накладываемых эталонных речевых единиц SL_k речевым единицам SL_p^X предъявленного РС решается задача сегментно-слогового синтеза. Которая заключается в получении эталонных траекторий параметров X^* для предъявленного РС конкатенацией траекторий слогов-эталонов SL_k таким образом: на каждом шаге алгоритма слогу p предъявленной реализации X необходимо наилучшим образом сопоставить k-й слог эталонных последовательностей $\{Y_k\}$

$$d_p=\min_k \left(SL_p^X \# SL_k\right)=\min_k \sum_{i,j}\left(SG_{pi}^{X} \# SG_{kj}^{Y}\right).$$

Наилучшее приближение к реализации X по всей совокупности слогов обеспечивает минимум величины

$$d=\sum_1^M d_p.$$

Синтез эталонных траекторий параметров, согласно схеме сегментно-слогового распознавания [4], сформулируем в терминах пространства состояний. Состоянием в данной задаче является сегмент, который определяется своим номером. Целевым состоянием или решением задачи является путь от начального сегмента предъявленного РС к конечному сегменту. Допустимым переходам между состояниями или дугами, соединяющим вершины графа $i\to j$, присваиваются стоимости d_{ij}, характеризующие близость слога SL_p^X предъявленной реализации с эталонным слогом SL_k. Сумма стоимостей дуг, лежащих на каком-либо пути из начальной в конечную вершину, определит интегральное сходство распознаваемого сигнала X с эталонным сигналом X^*. Для нахождения наилучшего приближения предъявленного сигнала X на множестве эталонных сигналов X^* необходимо осуществить оптимизацию — найти решение с минимальной стоимостью.

Структура графа синтеза эталонных траекторий параметров X^* для случая $q=7$ (q — размер пространства состояний) показана на рис. 1.

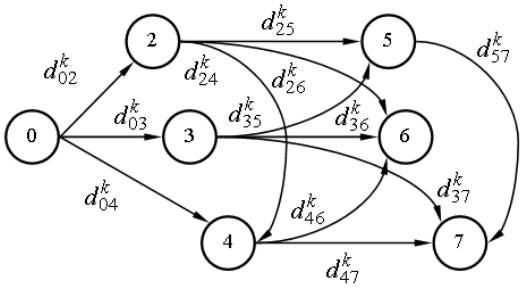

Рис. 1. Граф синтеза эталонных траекторий параметров X^* для случая $q=7$

Для приведения к сопоставимому виду траектории X^* относительно X используются нелинейные преобразования, которые компенсируют мешающие компоненты траекторий и усиливают наиболее информативные. В качестве таких преобразований используются огибающие в частотных полосах для предъявленной реализации РС, которые действуют на эталонные реализации аналогично частотно-временному окну.

Стыковка траекторий параметров эталонов в си-

нтезированной реализации РС, соответствующих найденному оптимальному пути, обеспечивается

$$\sigma^2 = \sum_k \sum_{i,j} \left| X_i - \alpha_k Y_j \right|^2 \to \min .$$

Из условия

$$\frac{\partial \sigma^2}{\partial \alpha_k} = 0 \ \left(k = \overline{1, r} \right),$$

где r — количество слогов предъявленного РС, определяются неизвестные коэффициенты α_k, а значит, и наилучшая синтезированная траектория параметров РС.

Результаты

Распознавание предъявленного РС осуществляется функцией, выполняющей синтез траекторий параметров из заданных слогов-эталонов и сопоставление с траекториями предъявленного РС. Словарь состоит из двух-, трех- и четырехсегментных слогов для заданного набора слов. Сегментированная речевая последовательность, поступающая на вход системы распознавания, рассматривается как совокупность двух-, трех- и четырехсегментных слогов. На каждом шаге алгоритма на предъявленную реализацию X последовательно накладываются слоги-эталоны E_q^2, E_r^3, E_s^3, содержащие по два, три, четыре сегмента соответственно, определяются расстояния

$$d_q = E_q^2 \# SL_p^{x2}, \quad d_r = E_r^3 \# SL_p^{x3}, \quad d_s = E_s^4 \# SL_p^{x4}.$$

Параметры слога-эталона с минимальным расстоянием, выбранным в соответствии с исследуемыми стратегиями поиска, используются для синтеза траекторий эталонного сигнала X^*.

В работе исследовано адекватность применения стратегий поиска в глубину и в ширину для синтеза траекторий параметров эталонного сигнала X^*, для которого возможны следующие варианты:

— эталонный сигнал, состоящий из двухсегментных слогов-эталонов E_q^2,

$$X^{2*} = \left\{ E_{q1}^2, E_{q2}^2, E_{q3}^2, \ldots, E_{qn_2}^2 \right\},$$

где n_2 — количество двухсегментных слогов в предъявленной реализации X;

— эталонный сигнал, состоящий из трехсегментных слогов-эталонов E_r^3,

$$X^{3*} = \left\{ E_{r1}^3, E_{r2}^3, E_{r3}^3, \ldots, E_{rn_3}^3 \right\},$$

где n_3 — количество трехсегментных слогов в предъявленной реализации X;

— эталонный сигнал, состоящий из четырехсегментных слогов-эталонов E_s^4,

$$X^{4*} = \left\{ E_{s1}^4, E_{s2}^4, E_{s3}^4, \ldots, E_{sn_4}^4 \right\},$$

где n_4 — количество четырехсегментных слогов в предъявленной реализации X;

—эталонный сигнал, состоящий из двух-, трех- и четырехсегментных слогов-эталонов.

Траектории параметров предъявленного и синтезированного эталонного слова «панама» в девяти частотных полосах представлены на рис. 2.

В результате применения стратегии поиска в ширину получены траектории эталонного РС $X^{3*} = \left\{ E_{r1}^3, E_{r2}^3 \right\}$, синтезированного из трехсегментных слогов-эталонов (E_{r1}^3 = «п-а-н», E_{r2}^3 = «а-м-а»), а при применении стратегии поиска в глубину получены траектории эталонного РС $X^{2*} = \left\{ E_{q1}^2, E_{q2}^2, E_{q3}^2 \right\}$, синтезированного из двухсегментных слогов-эталонов (E_{q1}^2 = «п-а», E_{q2}^2 = «н-а», E_{q3}^2 = «м-а»).

а) X б) X^*

Рис. 2: а) траектории предъявленного слова «ПАНАМА»;
б) синтезированные траектории слова «ПАНАМА»

Стоимость решения, полученного при применении стратегии поиска в ширину, больше стоимости решения, полученного при применении стратегии поиска в глубину. Это соответствует тому, что стратегия поиска в ширину гарантирует выработку кратчайшего решения (минимальное количество переходов в пространстве состояний от начального узла к целевому), но не обеспечивает оптимальный путь решения.

Временные затраты на распознавание предъявленной реализации на основе сегментно-слогового синтеза с применением стратегий поиска в ширину и в глубину, на порядок меньше времени, необходимого на распознавание при поиске полным перебором возможных комбинаций слогов-эталонов. Использование методов неинформированного поиска позволяет сократить время распознавания, так как прекращается процесс поиска при достижении целевого состояния, что приводит к сокращению объема вычислений.

Выводы

Рассмотренные стратегии поиска в глубину и в ширину сокращают время распознавания предъявленного речевого сигнала, но не всегда позволяют получить оптимальное решение, так как они не учитывают дополнительную информацию, относящуюся к исследуемой проблеме. Полученные результаты свидетельствуют о том, что при решении задачи распознавания речи необходимо использовать дополнительные критерии, которые оценивают перспективность переходов между состояниями с точки зрения достижения цели. Это требует дополнительных исследований.

СПИСОК ЛИТЕРАТУРЫ

1. Рассел С. Искусственный интеллект: современный подход / С. Рассел, П. Норвиг — М., 2006. — 1408 с.
2. Братко И. Алгоритмы искусственного интеллекта на языке PROLOG — М., 2004. — 640 с.
3. Шаповалова С. И. Оптимизация решения задач поиска на прологе // Искусственный интеллект. — 2000. — № 3. — С. 121—127.
4. Карпов О. Н. Технология построения устройств распознавания речи. — Д., 2001. — 184 с.

Карпов Олег Николаевич — профессор кафедры математического обеспечения ЭВМ, ***Савенкова Ольга Александровна*** — аспирантка кафедры экспериментальной физики.

Днепропетровский национальный университет