УДК 681.327

ЯНЬ ЦЗИНБИНЬ (КНР), У ШИ (КНР), А.В. ТКАЧЕНЯ, И.Э. ХЕЙДОРОВ

ПОИСК КЛЮЧЕВЫХ СЛОВ В СЛИТНОЙ РЕЧИ НА ОСНОВЕ УСОВЕРШЕНСТВОВАННОЙ МЕРЫ ДОСТОВЕРНОСТИ

This paper is devoted to the problem of keywords search in continuous speech. The modification of minimum edit distance algorithm was done in order to improve keyword detection rate on 1,5 %. It was proposed to use new confidence measure which lets to reduce false alarms on 4,2 %.

Поиск ключевых слов в речевом потоке является одной из наиболее сложных задач в области обработки речевых сигналов. Данная технология чрезвычайно востребована в системах аудиоиндексации, поиска речевой информации по образцу в мультимедиаархивах, автоматического контроля речевых сообщений, в системах безопасности и т. д. [1, 2]. В последнее время для создания систем поиска, не зависящих от словаря, достаточно широкое распространение приобрел подход, основанный на использовании решетки акустических единиц или фрагментов речи [3]. Каждый узел решетки ассоциируется с некоторым моментом времени произнесенной речи, а связи между ними представляют возможные варианты транскрипции произнесенной речи. Основное преимущество этого метода заключается в его большой гибкости: даже если фонема ключевого слова не является лучшей гипотезой, она все равно сохраняется в результате распознавания. Однако использование системы данного типа для обнаружения ключевых слов в потоке слитной речи сопряжено со сложностью верификации найденных слов. Поэтому одна из основных задач на настоящий момент – поиск новой меры достоверности для снижения количества ложных тревог.

Структура системы поиска ключевых слов на основе решетки слогов

Работа системы поиска ключевых слов на основе решетки слогов осуществляется в три этапа (рис. 1): первый – обучение распознавателя с помощью языковой и акустической баз данных, в результате чего формируется решетка слогов; второй – поиск в решетке возможных ключевых слов; третий – верификация найденных ключевых слов при помощи меры достоверности.

Решетка $L = (N, V, n_{\text{start}}, n_{\text{end}})$ представляет собой направленный непериодический граф, где N – множество узлов, V – множество дуг, каждая из которых представляет гипотезу единицы распознавания (слово, слог, фонема), $n_{\text{start}}, n_{\text{end}} \in N$ – начальный и конечный узлы. Каждому узлу решетки $n \in N$ соответствует свое время $t(n)$. Представим дуги в виде четырехугольников ($s[v]$, $t[v]$, $l[v]$, $p[v]$), где $s[v]$, $t[v] \in N$ – начальный и конечный узел дуги v, $l[v]$ – единица распознавания, $p(v)$ – акустическая вероятность, соответствующая речевому фрагменту $O_{s(v)}^{t(v)}$, $p(v) = P(O_{s(v)}^{t(v)} \mid l(v))$. По сути, решетка

представляет собой сжатое пространство декодирования, включающее основную значимую информацию, полученную в процессе распознавания. Узлы решетки формируют конкурирующие пути, а полный путь из начального узла до конечного определяет гипотетическое предложение о распознанном участке речи.

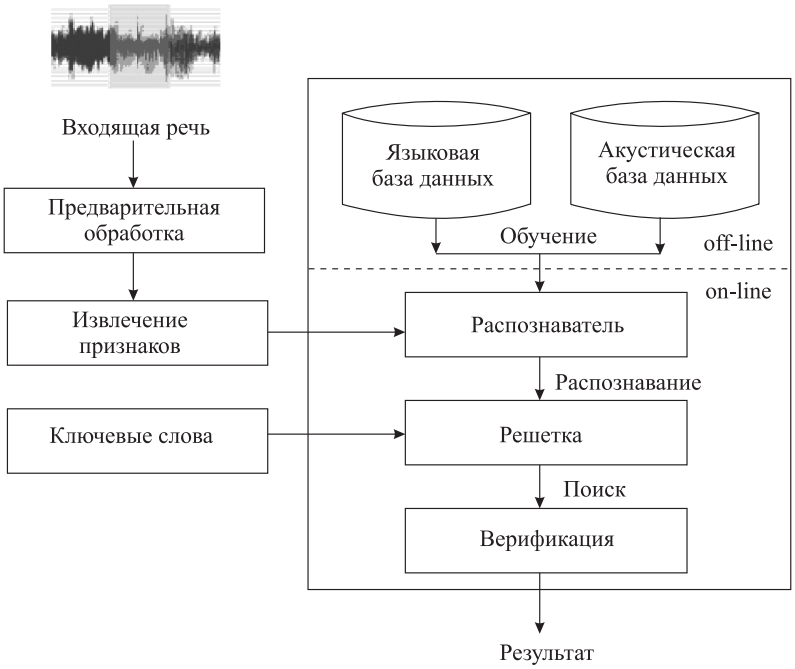

Рис. 1. Схема работы системы поиска ключевых слов на основе решетки слогов

Усовершенствованный алгоритм определения минимального межстрокового расстояния

Для последовательности акустических единиц (слогов), полученных на основе СММ и представленных в виде решетки, характерны ошибки трех видов. Во-первых, возможно появление «лишних» слогов, не присутствующих в реальном сигнале, называемых ошибками вставки. Во-вторых, неправильное распознавание слогов приводит к появлению ошибок замены, в-третьих, в случае коротких слогов возможен их пропуск – ошибки удаления. Определим понятие минимального расстояния между строками U и V как минимальные затраты на преобразование строки U в строку V с помощью трех основных операций: вставки, удаления и замены. Для хранения значений минимального расстояния используем матрицу $M(0, ..., p)(0, ..., q)$, где p и q – это длины строк U и V соответственно. Для определения минимального межстрокового расстояния (ММР), обозначаемого как $\mathrm{MED}(U, V)$, необходимо определить минимальный по стоимости путь из точки (0, 0) в точку (p, q) в соответствии со значениями матрицы M:

$$\mathrm{MED}(U,\ V) = \min_{\Omega} M\left(p, q\right), \tag{1}$$

где Ω – набор всех возможных путей. Данная задача может быть с успехом решена при помощи алгоритмов динамического программирования [4].

Для улучшения точности представления речевых сигналов последовательностью фонем был разработан усовершенствованный алгоритм ММР на основе матрицы спутывания, что позволило учесть статистику возможных ошибок акустического моделирования. Матрица спутывания C представляет собой матрицу размером N на N элементов (где N – это число слогов в слоговом наборе). Каждый элемент матрицы $C(i, j)$ – это отношение количества замен i-го слога j-м слогом к общему числу элементов, распознанных как i-й слог. Вычисление затрат на операцию замены с помощью матриц спутывания позволяет учесть вероятностные характеристики ошибок и улучшить точность поиска слов.

Для реализации усовершенствованного алгоритма ММР была предложена следующая процедура. На первом этапе формируется гипотетическая последовательность ключевых слов. Затем для каждого гипотетического ключевого слова определяется гипотетическая последовательность слогов W' и производится вычисление ММР для реальной и гипотетических строк. Далее производится сравнение полученного межстрокового расстояния с пороговым значением: если $\mathrm{MED} \le \mathrm{MED}_{\max}$, тогда W' является возможным ключевым словом. Для сравнения строк различной длины необходимо провести нормировку ММР с использованием длин распознанной и гипотетических строк. В этом случае определим меру подобия таких строк следующим образом:

$$P_{\mathrm{cmed}}(W) = 1 - 2 \cdot \frac{\mathrm{MED}(W,\ W')}{(m+n)},$$

где W – последовательность слогов искомого ключевого слова, m и n – длины сравниваемых последовательностей. Очевидно, что $-\infty \le P_{\mathrm{cmed}}(W) \le 1$.

Вычисление апостериорной вероятности ключевых слов

С помощью усовершенствованного алгоритма ММР мы получили набор возможных ключевых слов, который включает большое количество сгенерированных вставок и замен. Для каждого возможного ключевого слова из набора необходимо провести верификацию с использованием некоторой меры достоверности. Проведенные исследования показали, что для определения меры достоверности на основе единичной характеристики наилучший результат обеспечивается при использовании апостериорной вероятности, которую легко вычислить на базе информации, содержащейся в решетке [5].

Представим строку слогов, обозначающих ключевое слово Kw, как $l_1 l_2 ... l_K$, тогда апостериорная вероятность $P(Kw|O)$ имеет вид

$$P(Kw|O) = P(l_1 l_2 ... l_K | O) = P(W(l_1 l_2 ... l_K)|O), \quad (2)$$

где $W(l_1 l_2 ... l_K)$ – множество всех путей, включающих строку слогов $l_1 l_2 ... l_K$. Тогда уравнение (2) можно представить как

$$P(Kw|O) = \sum_{\forall W(v_1 v_2 ... v_K) \subset W(l_1 l_2 ... l_K)} P(W(v_1 v_2 ... v_K)|O).$$

В этом случае вероятность некоторого ключевого слова совпадает с накопленной апостериорной вероятностью всех возможных путей. Обозначим путь сопоставления $w = v_1 v_2 ... v_K$, тогда апостериорную вероятность сопоставленных путей $P(W(v_1 v_2 ... v_k)|O)$ можно записать как $P(w|O)$. На этапе верификации ключевых слов сначала вычисляется апостериорная вероятность каждого сопоставленного пути, и после накопления можно получить апостериорную вероятность ключевых слов.

Апостериорную вероятность представим следующим образом:

$$P(w|O) = \frac{P(O,w)}{P(O)} = \frac{\sum_{w_g \in W(v_1 v_2 ... v_K)} P(O, w_g)}{\sum_{w_g \in W_G} P(O, w_g)}. \quad (3)$$

Объединенная вероятность $P(O, w_g)$ в выражении (3) представляет собой произведение значения акустической вероятности каждой дуги внутри пути w_g и вероятности соответствующих лингвистических моделей. При вычислении $P(w|O)$ согласно (3) за счет фиксирования последовательности дуг ограничивается количество анализируемых путей, вследствие этого вычислительная сложность алгоритма невелика и составляет $O(K)$, где K – длина ключевого слова.

Усовершенствованная мера достоверности для поиска ключевых слов

Поскольку при помощи алгоритма ММР анализируются модифицированные последовательности фонем с учетом возможности вставок и замен, использование для верификации апостериорной вероятности в ее традиционном виде является недостаточно надежным. В связи с резким ростом количеств вариантов строк (в том числе и неправильных) необходимо использовать меру достоверности, учитывающую априорную информацию об искомой последовательности слогов, принадлежащей ключевому слову. Такой подход в случае присутствия ключевого слова в анализируемом речевом сигнале позволит увеличить комплексное значение его вероятности. В данной статье для верификации найденных ключевых слов было предложено использовать усовершенствованную меру достоверности ММР с учетом меры подобия строк:

$$CM1(W) = k_1 P(w|O) + k_2 P_{\text{cmed}}(W) + k_3 \text{Lenght}(W) + k_4 AS(W),$$

где k_i (i = 1, 2, 3, 4) – константы, $P(w|O)$ – мера правдоподобия апостериорной вероятности гипотетического ключевого слова, P_{cmed} – усовершенствованное ММР, вычисляемое согласно (1), $\text{Lenght}(W)$ – количество слогов в ключевом слове W, $AS(W)$ – акустическая стабильность [5]. Использование меры достоверности $CM1$ позволяет вычислить меру достоверности для гипотетических ключевых слов и уменьшить количество ложных тревог, возникших в результате использования алгоритма ММР.

В процессе поиска ключевых слов возникает необходимость уменьшения вычислительных затрат. Данная проблема решается путем специальной процедуры, позволяющей исключить из последующего рассмотрения пути с низкими значениями вероятности. Очевидно, что такой подход не будет оптимальным в смысле точности обнаружения, поскольку существует возможность удаления глобально луч-

шего пути, что приведет к общей ошибке поиска. Нахождение разумного компромисса между вычислительной сложностью алгоритма и точность поиска являлись предметом экспериментального изучения.

Эксперименты

Для создания акустической и лингвистической моделей, а также тестирования системы поиска ключевых слов были созданы две речевые и одна текстовая база данных. Для оценки параметров акустической модели использовалась речевая база данных CASIA, записанная в формате 16 кГц, 16 бит и включающая 300 дикторов, 75 000 предложений. База составлена из слов современного общелитературного языка и записана в лингафонном кабинете. Для тестирования разработанных алгоритмов поиска ключевых слов были использованы речевые данные из реальных радио- и телепередач, содержащие 13 ключевых слов в 4970 речевых файлах. Текстовая база данных включает 200 000 предложений. Система поиска ключевых слов была реализована с помощью VC++6.0 на платформе Win32 с использованием HTK [6].

Таблица 1

Сравнение точности поиска в решетке слогов с разным размером

Максимальное количество путей	Среднее количество узлов решетки	LER, %	Вероятность обнаружения Pd, %
200	204	22,8	78,6
250	452	16,9	83,4
300	547	16,5	86,8
350	559	15,3	88,3
400	590	15,3	88,2
450	1881	15,3	88,1

Таблица 2

Количество правильно обнаруженных слов в зависимости от использования различных методов поиска

Ключевые слова	Количество слов	Обнаружено слов							Решетка
		1	3	5	7	9	15	20	
fang1mian4	180	116	119	122	125	133	136	137	163
guan1xi4	172	114	114	116	119	124	128	130	158
guo2jia1	486	312	334	345	354	360	369	369	435
wo3men5	676	360	407	430	478	512	523	525	589
yi2ge4	528	324	344	365	387	400	411	412	457
zhong1guo2	620	433	438	440	449	457	465	465	566
zi4ji3	228	140	153	154	157	159	168	170	208
Всего	4970	3107	3285	3382	3523	3659	3774	3789	4408

При разной величине порога на выходе распознающего устройства получается решетка слогов различного размера. В первой серии экспериментов исследовалось влияние величины порога для устранения маловероятных путей на точность системы поиска ключевых слов. Одной из важных характеристик системы является величина ошибок решетки LER, определяемая структурой и размером сгенерированной решетки, что свидетельствует о том, находится ли требуемая последовательность слогов среди путей, сгенерированных решеткой. Результат эксперимента представлен в табл. 1, из анализа данных которой можно сделать следующий вывод: с увеличением величины порога возрастает среднее число сгенерированных узлов решетки, а значение LER уменьшается. При значении величины порога, равном 350, LER достигло стабильного значения 15,3 %, а вероятность правильного обнаружения составила 88,3 %. Дальнейшее увеличение порога ведет к значительному возрастанию количества узлов, а соответственно и времени поиска.

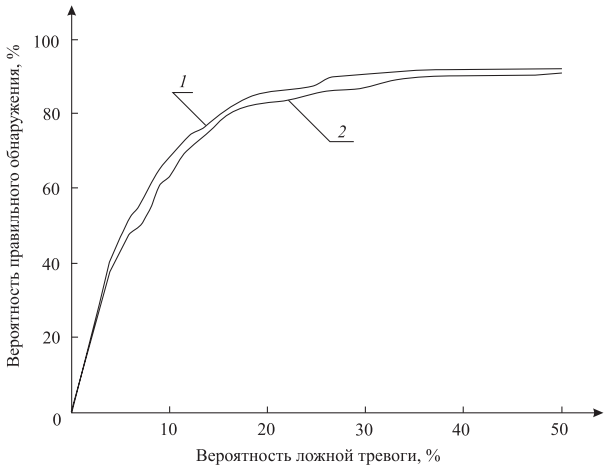

Рис. 2. Рабочие характеристики системы верификации ключевых слов для различных мер достоверности: *1* – CM1, *2* – CM2

Для каждого из 13 ключевых слов были проведены сравнительные эксперименты по их обнаружению на основе алгоритма *N*-best ($N = 1, 3, 5, 7, 9, 15, 20$) и решетки слогов. Результаты представлены в табл. 2.

Как видно из табл. 2, использование решетки позволяет улучшить точность поиска ключевых слов в среднем на 12,5 % по сравнению с алгоритмами на основе *N*-best поиска.

Проведенные эксперименты показали, что применение усовершенствованного алгоритма ММР позволило увеличить вероятность правильного обнаружения на 1,5 % (с 86,9 до 88,4 %). Несмотря на увеличение эффективности поиска при использовании решетки, данный алгоритм приводит к росту количества ложных тревог. Для уменьшения их количества необходимы дополнительные алгоритмы верификации на основе различных мер достоверности. В экспериментах были использованы две меры достоверности: предложенная нами усовершенствованная $CM1$ и обычная $CM2$ на основе ММР. На рис. 2 показаны сравнительные рабочие характеристики системы верификации для этих мер достоверности. Видно, что использование предложенной усовершенствованной меры достоверности $CM1$ позволяет достичь лучших характеристик системы.

1. Zheng T.R., Han J.Q. // Proc. of IEEE Inter. Conf. on Audio. 2008. P. 1209.
2. Wu Ch. // Speech Communication. 2001. Vol. 33(3). P. 197.
3. Frank K. Soong // Spoken Language Translation Research. 2004. P. 2.
4. Thambiratnam K., Sridharan S. // Proc. of IEEE Inter. Conf. 2005. P. 465.
5. Wessel F. // Proc. of IEEE Trans. on Speech and Audio. 2001. Vol. 9(3). P. 288.
6. Young S., Evermann G. The HTK Book. Version 3.2, 2002.

Поступила в редакцию 26.06.09.

Янь Цзинбинь – аспирант кафедры радиофизики. Научный руководитель – И.Э. Хейдоров.
У Ши – аспирант кафедры радиофизики. Научный руководитель – И.Э. Хейдоров.
Андрей Владимирович Ткаченя – студент 5-го курса факультета радиофизики и электроники.
Игорь Эдуардович Хейдоров – кандидат физико-математических наук, доцент кафедры радиофизики.