

Поиск ключевых слов с использованием решетки слогов

Keyword search using syllable lattice

Алиев Р. М. (RomanAliyev@gmail.com),

Янь Цзинбинь (yjbemail@163.com), **Хейдоров И. Э.** (igorhmm@mail.ru)

Белорусский государственный университет, Минск, Беларусь

Методы поиска ключевых слов, основанные на решетке слогов могут исключить проблему отторжения слов, не входящих в словарь (СВС), и компенсировать потерю качества поиска ключевых слов, обусловленную ошибками распознавания. В данной работе предлагается для повышения точности поиска ключевых слов вычислять апостериорную вероятность для различных последовательностей фрагментов слов (слогов). Эксперименты показали высокую эффективность предложенного алгоритма для поиска ключевых слов в потоке слитной речи.

1. Введение

Поиск ключевых слов в речевом потоке является одной из наиболее сложных задач в области обработки речи. На основе этой технологии можно реализовать системы аудио индексации, поиска речевой информации по образцу в мультимедиа-архивах, автоматический контроль речевых сообщений в системах безопасности и т. д. [1] [2].

В настоящий момент используются несколько методов поиска ключевых слов. Первый и наиболее простой метод поиска ключевых слов использует распознаватель со словарем для перевода непрерывной речи в текст. Для поиска ключевого слова осуществляется поиск в полученном тексте с использованием традиционных алгоритмов поиска текста. Проблема этого метода состоит в том, что из-за ограниченного множества слов в распознавателе, невозможно распознать слова, отсутствующие в словаре, например, имена, акронимы и слова из иностранных языков.

Другой метод поиска ключевых слов основан на скрытых Марковских моделях (СММ). Он использует СММ для каждого ключевого слова и одну «модель мусора» для всех остальных слов [3]. Временная последовательность символов ключевых слов и мусор символов формируется в результате распознавания речевой последовательности. Этот метод не имеет ограничений при условии, что определено множество ключевых слов, которые необходимо найти. Но для каждого нового ключевого слова, которое нужно найти, нужно не только обучать новую СММ модель, но также нужно заново обучать модель мусора, поэтому использование этого метода при определенных условиях вызывает серьезные затруднения.

В последнее время достаточно широкое распространение приобрела идея построения и использования решетки фрагментов речи для решения различных задач [4]. Для поиска ключевых слов каждый узел решетки ассоциируется с моментом времени произнесенной речи. Основное преимущество этого метода в том, что он обладает большой гибкостью: даже если фонема ключевого слова не есть лучшая гипотеза между узлами решетки, она все равно сохраняется в результате распознавания. Результат поиска не зависит от словаря распознавателя, поскольку поиск можно организовывать для любой фонемной последовательности запрашиваемого ключевого слова, поэтому в этом методе проблема СВС решается наиболее естественным и эффективным способом.

2. Структура системы поиска ключевых слов на основе решетки

Введем понятие решетки $L = (N, A, n_{start}, n_{end})$ как направленного неперiodического графа, где N — множество его узлов, A — множество связей между узлами и $n_{start}, n_{end} \in N$ — начальный и конечный узлы решетки соответственно. Представим связь между узлами в виде $a = (S[a], E[a], I[a], w[a])$, где $S[a], E[a] \in N$ — начальный и конечный узлы; $I[a]$ — фрагмент речи (слог или фонема); $w[a] = p_{ac}(a)^{1/\lambda}$ — весовой коэффициент связи; $p_{ac}(a)$ — акустическое сходство; λ — весовой коэффициент.

Предположим, что мы имеем последовательность слогов l_p, l_q, \dots, l_k , слияние которых дает ключе-

вое слово Kw . В результате поиска нужно получить множество W наиболее вероятных ветвей w ориентированного графа, которые соответствуют запрашиваемому ключевому слову. Этот поиск осуществляется методом последовательного прокладывания пути в структуре решетки. В первую очередь среди всех узлов множества N необходимо найти начальный узел, соответствующий первому слогу ключевого слова, затем граничных узлов производится поиск следующего вероятного узла, который будет соответствовать следующему слогу. Процедура повторяется K раз до тех пор, пока не будет обработан последний слог ключевого слова.

Для полученной последовательности наблюдений O производим вычисление апостериорной вероятности для полученных результатов поиска:

$$P(Kw|O) = \sum_{w \in W} P(w|O), \quad (1)$$

которая является суммой апостериорных вероятностей всех возможных путей в графе, сопоставляемых ключевому слову Kw .

Представим апостериорную вероятность пути следующим образом:

$$P(w|O) = P(O, w) / P(O). \quad (2)$$

Для вычисления $P(w|O)$ воспользуемся алгоритмом прямого и обратного хода:

Шаг 1. Вычисление значений прямых переменных и обратных переменных $\alpha(v)$ и $\beta(v)$:

$$\alpha(v) = P(O_0^{t(v)}, v) = \sum_{w \in W_v^-} P(O_0^{t(v)}, w), \quad (3)$$

$$\beta(v) = P(O_{s(v)}^T | v) = \sum_{w \in W_v^+} P(O_{s(v)}^T, w) / b(v). \quad (4)$$

Исходя из уравнений (3), (4):

$$\alpha(v) = \sum_{v_1, \dots, v_l, v \in W_v^-} \left[\prod_{i=1}^{l-1} b(v_i) q(v_i, v_{i+1}) \right] b(v_l) q(v_l, v) b(v), \quad (5)$$

$$\beta(v) = \sum_{v, v_1, \dots, v_l \in W_v^+} q(v, v_1) \left[\prod_{i=1}^l b(v_i) q(v_i, v_{i+1}) \right]. \quad (6)$$

Очевидно, что $\alpha(v)$ и $\beta(v)$ нельзя вычислить прямо, но их можно выразить через рекурсивные соотношения:

$$\alpha(v) = \sum_{(u, v) \in E} a(u) q(u, v) b(v), \quad (7)$$

$$\beta(v) = \sum_{(u, v) \in E} q(u, v) b(v) \beta(v), \quad (8)$$

$$\alpha(v) = b(v), \text{ если } v \in V_0, \quad (9)$$

$$\beta(v) = 1, \text{ если } v \in V_T. \quad (10)$$

Таким образом, начиная с начального узла, можно вычислить значение $\alpha(v)$ каждого узла и, начиная с конечного узла, можно вычислить значение $\beta(v)$ каждого узла.

Шаг 2. Вычисление вероятности $P(O)$:

$$P(O) = \sum_{w_g \in W_G} P(O, w) = \sum_{v \in V_T} \alpha(v). \quad (11)$$

Из выражения (11) видно, что $P(O)$ вычисляется непосредственно по прямым переменным всех ключевых слов.

Шаг 3. Вычисление апостериорной вероятности $P(w|O)$:

$$P(w|O) = \alpha(v_1) \left[\prod_{k=1}^K q(v_k, v_{k+1}) b(v_{k+1}) \right] \beta(v_K) / P(O). \quad (12)$$

На основе описанной выше решетки была предложена и разработана следующая схема поиска ключевых слов (рис. 1). Работа системы осуществляется в три этапа. Первый этап — выделение мелкепстральных признаков речевого сигнала на основе вейвлет-преобразования. В работе предлагается использовать вейвлетное преобразование для построения вектора признаков ввиду его хорошего разрешения по времени для спектрального диапазона речевого, что особенно важно для автоматической сегментации [5]. Второй этап — обучение распознавателя с помощью языковой и акустической баз данных, в результате чего формируется решетка фрагментов слов. Третий этап — поиск в решетке возможных ключевых слов с подтверждением с помощью вычисления апостериорной вероятности.

3. Эксперимент

Предложенная схема поиска ключевых слов на основе вейвлетного преобразования и решетки фрагментов слов была реализована на языке программирования C++ с использованием библиотеки VCL (рис.2). База данных акустических моделей была получена с помощью отсегментированной чистой речи продолжительностью 124 часа. Для проведения эксперимента по поиску ключевых слов была использована записанная речь диктора радиопередачи «Новости» продолжительностью 3,54 часа. В качестве вектора признаков был использован 39-компонентный вектор, включающий в себя энергию, мелкепстральные характеристики и их производные, полученные на основе вейвлет-преобразования. Обучение распознавателя на основе СММ было осуществлено с использованием библиотеки НТК [6].

Результаты поиска 40 наиболее часто встречаемых в новостях ключевых слов с использованием предложенной системы по сравнению с обычной системой на основе алгоритма *N-bes* [6], представ-

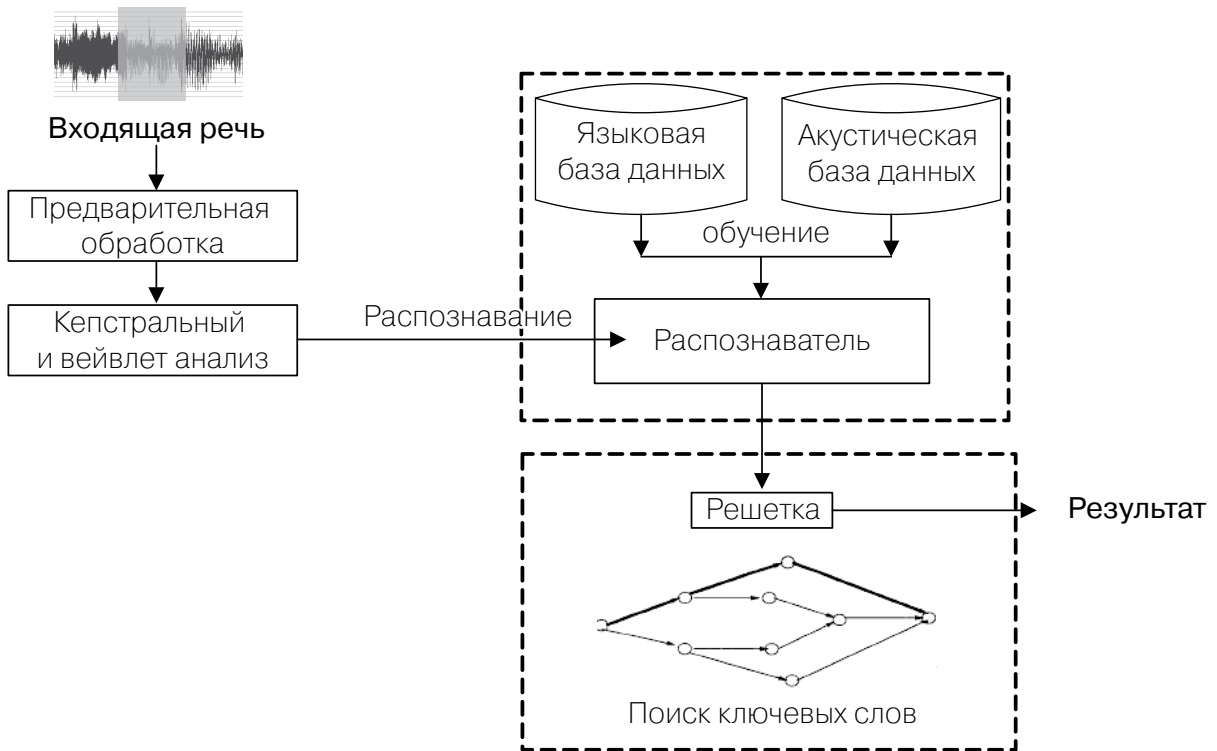


Рис. 1. Схема работы системы поиска ключевых слов на основе решетки фрагментов слов.

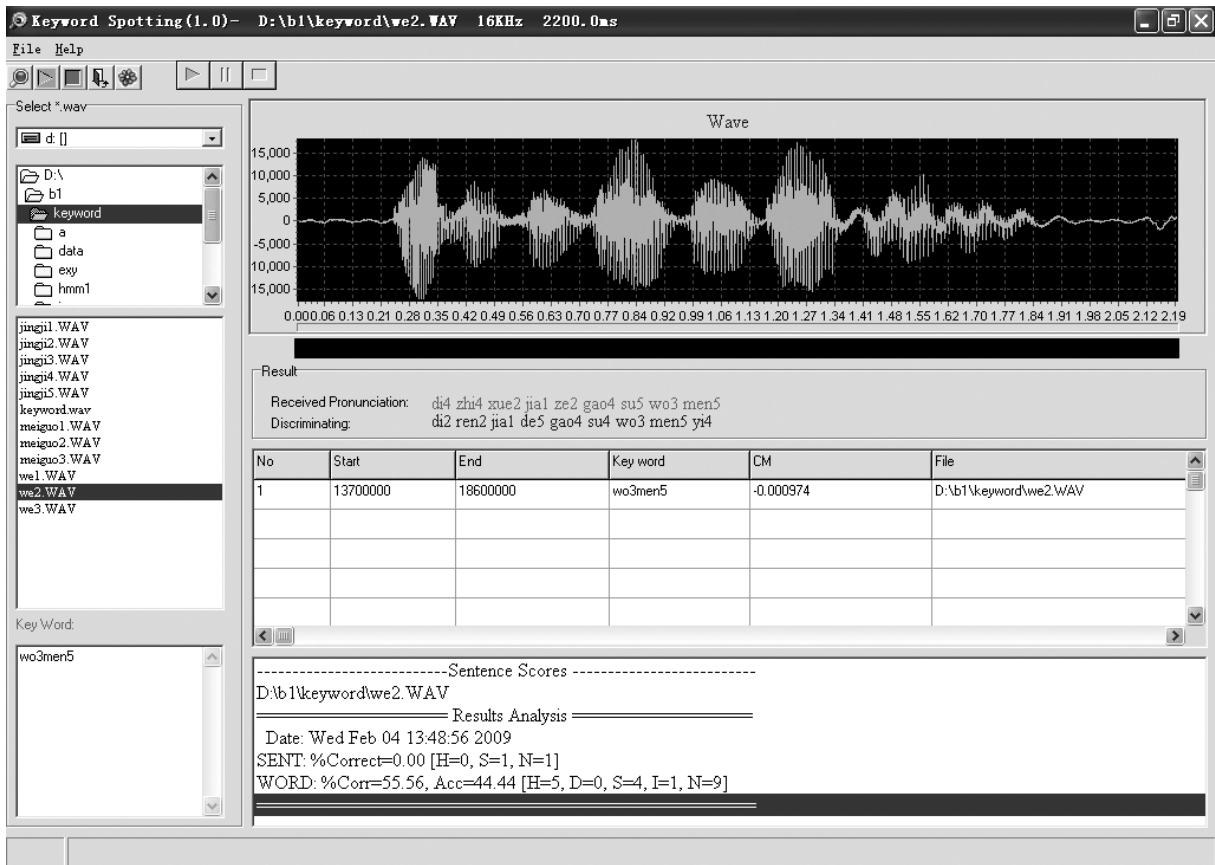


Рис. 2. Интерфейс системы поиска ключевых слов на основе решетки слогов.

лены в табл. 1 и 2. 40 слов для поиска были выбраны таким образом, чтобы их встречаемость в новостях была не менее 100 раз. Это сделано для того, чтобы обеспечить статистическую устойчивость оценки точности, поскольку при меньшем количестве встречаемых слов каждая ошибка поиска очень сильно влияет на оценку характеристик системы. В таблице 1 приводятся средние значения характеристик системы обнаружения ключевых слов: точность, представляющей собой отношение количества правильно найденных слов к общему количеству слов, классифицированных как ключевые, и чувствительность, представляющее отношение количества правильно найденных слов к общему количеству ключевых слов, присутствующих в записи. В таблице 2 для каждого из 13 ключевых слов представлено число их правильных обнаружений в пото-

ке речи при помощи алгоритма N-best (N=1, 3, 5, 7, 9, 15, 20) и на основе решетки слогов.

Как видно из таблицы, использование алгоритма решетки позволяет улучшить точность поиска ключевых слов в среднем на 9 процентов по сравнению с алгоритмами на основе N-best поиска.

4. Заключение

В данной работе предложена и реализована система поиска ключевых слов на основе решетки частей слов. Результат эксперимента показал, что использование решетки для слогов, сгенерированных на основе СММ, позволяет улучшить точность поиска в среднем на 9 процентов.

Таблица 1. Сравнение точности распознавания и чувствительности для систем на основе решетки слогов и N-best.

Система распознавания ключевых слов	Точность	Чувствительность
N-best (N=20)	69.5%	77.6%
Система на основе решетки слогов	86.2%	88.3%

Таблица 2. Сравнительная таблица количества правильно обнаруженных слов различными методами

Ключевое слова	Всего слов	Обнаружено слов							
		N=1	N=3	N=5	N=7	N=9	N=15	N=20	Решетка
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
jing1ji4	576	378	384	401	412	427	444	447	498
mei3guo2	232	156	158	157	157	159	159	160	220
qi3ye4	418	250	270	273	287	302	316	316	353
sheng1chan3	252	176	177	180	182	186	189	190	222
shi4jie4	182	80	110	114	120	137	147	148	154
ta1men5	420	268	277	285	296	303	319	320	385
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

Литература

1. *Young S. J., Brown M. G., Foote J. T., Jones G. J. F., Jones K. S.* Acoustic indexing for multimedia retrieval and browsing // IEEE International Conference on Acoustics, Speech and Signal Processing, 1997.
2. *Jones G. J. F., Foote J. T., Jones K. S., Young S. J.* Retrieving spoken documents by combining multiple index sources // Proc. SIGIR, 1996. P. 30–38.
3. *Wilpon J. G., Rabiner L. R., Lee C. H., Goldman E. R.* Automatic recognition of keywords in unconstrained speech using Hidden Markov Models // IEEE Transactions on Acoustics, Speech and Signal Processing, 1990.
4. *Soong F. K., Lo W. K., Nakamura S.* Generalized Word Posterior Probability (GWPP) for Measuring Reliability of Recognized Words // Proceeding of SWIM2004, 2004. P. 127–128.
5. *Кухарчик П. Д., Хейдоров И. Э., Бовбель Е. И., У Ши, Янь Цзинбинь.* Определение патологий голосового тракта путем анализа речевого сигнала на основе вейвлетного преобразования и метода опорных векторов // Электроника, Минск, БГУ, 2008. № 12, С. 44–49.
6. *Young S., Evermann G., Kershaw D.* The HTK Book // 2008.