

РАЗРАБОТКА СИСТЕМЫ ПОИСКА КЛЮЧЕВЫХ СЛОВ В СЛИТНОЙ РЕЧИ НА ОСНОВЕ СКРЫТЫХ МАРКОВСКИХ МОДЕЛЕЙ

А. В. Ткачения

РАЗРАБОТКА СИСТЕМЫ ПОИСКА КЛЮЧЕВЫХ СЛОВ

Используя описанные в предыдущих работах методы поиска ключевых слов [1], построим систему поиска ключевых слов в слитной речи для задачи поиска цифр в речевом сигнале.

Общая схема системы поиска ключевых слов приведена на рисунке 1.

В качестве акустической модели были выбраны трифоны со связанными состояниями. Количество состояний для одного трифона равно трем, закон распределения параметров – гауссов.

Трифонная акустическая модель со связанными состояниями хорошо описывает акустические особенности слитной речи и позволяет избежать проблемы нехватки обучающих данных. Также выбранная акустическая модель предотвращает чрезмерную адаптацию на обучающем наборе данных, которая может приводить к существенному снижению эффективности распознавания в дикторонезависимых системах.



Рис. 1. Общая схема системы поиска ключевых слов

Чтобы получить высокую эффективность системы распознавания слов в слитной речи, лингвистическая модель должна полностью учитывать взаимосвязь между всеми словами, тем самым, влияя на вероятность появления слова в данном контексте.

Для решения задачи распознавания цифр в слитной речи наиболее пригодна равновероятностная биграммная лингвистическая модель.

Систему распознавания слов в слитной речи построим при помощи программного пакета НТК [2], где в качестве классификатора используется СММ [3].

Декодирование осуществляется на основе СММ, так как эта модель хорошо согласуется с задачей распознавания речи, дает устойчивую вероятностную модель речи и имеет высокоэффективные методы анализа полученных данных.

Одним из самых важных этапов в системе распознавания речи является параметризация исследуемого речевого сигнала, с целью получения устойчивых акустических векторов признаков.

Для повышения эффективности системы поиска ключевых слов был разработан ВП на основе схожести нормированных выровненных вейвлет-картин.

Для оценки эффективности полученного ВП, были построены две системы распознавания речи на основе вейвлетных ВП (дискретное вейвлет-преобразование – ДВП) [4] и кепстральных ВП (мел-частотные кепстральные коэффициенты – МЧКК) [5]. Зашумленный сигнал получается при помощи смешивания исходного речевого сигнала с обыкновенным гауссовым шумом. Полученные в ходе экспериментов результаты приведены в табл. 1.

Главным этапом системы поиска ключевых слов является верификация распознанной речи и выделение из нее ключевых слов [6].

Таблица 1

Эффективность распознавания зашумленной речи для 1-ого диктора		
Уровень шума, дБ	МЧКК	ДВП
0 дБ	98 %	97 %
1 дБ	72 %	95 %
2,5 дБ	39 %	89 %
5 дБ	14 %	82 %

В работе [7] описаны различные типы мер достоверности, использующиеся для верификации распознанной речи. Наиболее эффективными являются меры достоверности на основе графа слов. Для использования таких мер достоверности необходимо в начале построить граф слов на основе лучших гипотез о последовательности распознанных слов. В качестве лучших гипотез будем брать N -лучшие результаты, получаемые на выходе системы распознавания речи.

Для оценки зависимости эффективности системы поиска ключевых слов от количества лучших гипотез, используемых для генерации графа слов, была построена соответствующая система на основе МЧКК и ДВП. Полученные результаты приведены в табл. 2.

Таблица 2

Зависимость эффективности поиска ключевых слов от количества лучших гипотез для 1-ого диктора

Кол-во лучших гипотез, шт.	МЧКК	ДВП
1	87 %	84 %
5	91 %	88 %
10	93 %	91 %
50	94 %	93 %

ЭКСПЕРИМЕНТ

Для исследования эффективности системы поиска ключевых слов в слитной речи, предложенной в этой статье, проведем эксперимент по поиску цифр в потоке спонтанной слитной речи.

Обучающие данные представляют собой отдельные цифры или их последовательности. Для каждого обучающего файла известна его транскрипция по словам.

Проверочные данные представляют собой отдельные цифры или их последовательности вперемешку со спонтанной речью. В транскрипции по словам для проверочных данных присутствуют только сведения о ключевых словах (цифры) и задаются они следующим образом: начальное и конечное время, а также название соответствующей цифры.

Полученные результаты приведены в табл. 3.

Таблица 3

Эффективность поиска цифр в зашумленной слитной речи 10-и дикторов для случая выбора 10-и лучших гипотез

Уровень шума, дБ	МЧКК	ДВП
0 дБ	89 %	84 %
1 дБ	65 %	82 %
2,5 дБ	34 %	77 %
5 дБ	12 %	71 %

ВЫВОДЫ

В данной статье были построены две системы поиска ключевых слов в слитной речи на основе МЧКК и ДВП. Проанализируем полученные результаты представленные в табл. 1-3.

Как видно из табл. 1, эффективность системы распознавания зашумленной речи для разработанного вектора признака на основе ДВП намного выше, чем для аналогичной системы на основе МЧКК. Для не зашумленной речи обе системы показали хорошие результаты: 97% и

98% соответственно. Из этого можно сделать вывод о предпочтительном использовании вейвлетной параметризации речевого сигнала, особенно для решения прикладных задач связанных с анализом зашумленного сигнала.

Недостатком разработанного вектора признака на основе ДВП является большое количество параметров ВП – 214, против 39 у МЧКК. Это приводит к увеличению размера получаемых моделей и возрастанию времени анализа и обработки полученных данных.

Результаты, приведенные в табл. 2, подтверждают тот факт, что разработанный вейвлетный ВП является избыточным. Это приводит к неточностям в распознавании речи, из-за чего возрастают ошибки при подсчете меры достоверности, и как следствие, к неправильному выбору ключевых слов на этапе верификации.

Верификация ключевых слов для случая использования 1-ой лучшей гипотезы фактически представляет собой выбор слов, апостериорная вероятность которых превышает заданное пороговое значение (верификация на основе графа слов не используется). Это приводит к низкой эффективности верификации. Результаты, приведенные в табл. 3, свидетельствуют о действительной эффективности системы поиска ключевых слов в спонтанной слитной речи, полученной в ходе написания этой статьи.

Литература

1. Ткачя А. В. Разработка методов поиска ключевых слов в слитной речи на основе скрытых марковских моделей // Сборник работ 66-й научной конференции студентов и аспирантов БГУ. Минск, 18-21 мая 2009. Минск, Издательский центр БГУ, 2010, Ч 1, С 187-191.
2. Young S., Everman, G., Gales M., Hain T., Kershaw D., Liu X., Moore G., Odell J. Ollason D., Valtchev V. / The HTK Book (for HTK Version 3.4). 2006.
3. Rabiner L. A. Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition // IEEE Press. 1988. С. 257–286.
4. Бовбель Е. И., Хейдоров И. Э., Пачковский Ю. В. / Скрытые марковские модели и машины на опорных векторах от практики к теории. Минск БГУ, 2008.
5. Xuedong Huang, Alex Acero, Hsiao-Wuen Hon / Spoken language processing. Prentice-Hall, Inc. Upper Saddle River, New Jersey, 2001.
6. Янь Цзинбинь, У Ши, Ткачя А.В., Хейдоров И.Э. Поиск ключевых слов в слитной речи на основе усовершенствованной меры достоверности // Вестн. Белорус. гос. ун-та. Сер. 1. 2009. № 3. С. 44–48.
7. Wessel F., Schluter R., Macherey K., Ney H. Confidence Measures for Large Vocabulary Continuous Speech Recognition // IEEE Transactions on Speech and Audio Processing, 2001. Вып. 3, гл. 9.