РАСПОЗНАВАНИЕ РЕЧИ ПРОИЗВОЛЬНОГО ДИКТОРА ПО КЛАСТЕРНЫМ ЭТАЛОНАМ

ВЛАДИМИР МАЗУР

Кафедра биофизики и математических методов в биологии
Львовский университет, Украина, СССР, 290005

РЕФЕРАТ

В работе предложен способ распознавания речи произвольного диктора по кластерным эталонам, являющих собой одни реализацию каждого слова словаря в произнесении диктора-центра кластера. Описан процесс создания кластеров, исследованы его характеристики при изменении условий экспериментов. Определены оптимальные параметры для создания кластеров. Предложена классификация дикторов по их пригодности для работы с неадаптивной СРР. Получены результаты распознавания речи произвольных пользователей по кластерным эталонам.

ВВЕДЕНИЕ

Классификация различных подходов к построению неадаптивных систем распознавания речи приведена в работе /1/. Предлагаемый нами алгоритм распознавания речи произвольного пользователя является разновидностью подхода, использующего статистическое обучение, с введением более быстрого, экономичного и эффективного способа дикторской адаптации и представления эталонов. Принятый подход позволяет исключить необходимость большого набора эталонов при обучении за счет применения кластерных эталонов, являющих собой одну реализацию каждого слова словаря в произнесении диктора - центра кластера /2/. Задача создания кластеров, требующая статистический материал и основанное на нем обучение, решается на этапе исследования дикторских голосов. Созданные кластеры постоянны и не зависят от используемого словаря. Изменение словаря влечет за собой только запись эталонов для дикторов-центров кластеров. Голос произвольного диктора, желающего работать с системой, предварительно классифицируется по "парольной" фразе и система "настраивается" на эталоны наиболее близкого по речевым параметрам кластера, по которым происходит распознавание, либо система выдает отказ, что означает, что данный диктор может работать только с адаптивной СРР.

АЛГОРИТМ КЛАСТЕРИЗАЦИИ И РАСПОЗНАВАНИЯ

Для создания кластеров дикторских голосов был записан банк образцов речи различных дикторов. В экспериментах приняло участие 50 дикторов, из них 30 мужчин и 20 женщин. Каждый из дикторов произнес по 10 слов (цифры от 0 до 9), признако-временное описание которых было записано в банк образцов речи. Почти все дикторы, принявшие участие в эксперименте, впервые работали с речевым вводом, иначе говоря, были "несотрудничающими" дикторами. С целью устранения явно видимых дефектных эталонов, было предусмотрено 2 варианта коррекции. Первую коррекцию можно было осуществить во время создания банка эталонов путем повтора плохого эталона. Вторую - в режиме коррекции, записывая новый эталон вместо дефектного. Контроль за качеством записанных эталонов можно было осуществлять в процессе создания банка посредством анализа выводимых на дисплей параметров либо анализируя распечатки параметров созданного банка образцов речи.

Алгоритм кластеризации заключался в создании такого каждого последующего кластера, который был бы максимально отличным от всех уже имеющихся. В каждый из них включались все те дикторы, различие которых по измеряемым параметрам речи находилось в пределах ограниченной области, определяемой радиусом кластера R. Более подробно алгоритм описан в работах /3, 4/.

Выбор рабочего кластера для дикторов, принимавших участие в записи банка эталонов, осуществлялся по распечатке качественного состава кластеров. Если один и тот же диктор входит сразу в более чем два кластера, для него целесообразно выбрать тот кластер, где его порядковый номер после кластеризации (считая от центра кластера) наименьший. Если с системой хочет работать диктор, не принимавший участия в создании банка эталонов, для него нужно произвести экспресс-кластеризацию голоса и выбрать соответствующий его голосу кластер.

Распознавание осуществляется с использованием алгоритма динамического программирования, описанного в работе /5/ с применением метрики Чебышева.

ИССЛЕДОВАНИЕ И ОПТИМИЗАЦИЯ КЛАСТЕРОВ

Для исследования качественного и количественного состава кластеров, их взаимосвязей и преобразований в зависимости от параметров алгоритма было проведено ряд экспериментов. Как уже отмечалось выше, голос произвольного диктора классифицируется по произнесенной им парольной фразе, являющейся как-бы его "визитной карточкой". Нас интересовало, как влияет длительность парольной фразы и ее акустико-фонетический состав на кластеризацию голоса, а в конечном счете на надежность распознавания речи по кластерным эталонам. Кроме этого хотелось определить оптимальную длительность парольной фразы. Для этого были использованы 6 различных по длительности и составу парольные фразы. Они были составлены из цифр от 0 до 9 и состояли от 3-х до 10-и слов. Состав отдельных парольных фраз следующий: парольная фраза из 3-х слов - 4,5,6; 4-х слов - 4,5,6,7; 5-и слов (1-й вариант) - 1,2,3,4,5; 5-и слов (2-й вариант) - 6,7,8,9,0; 7-и слов - 4,5,6,7,8,9,0; 10-и слов - 0...9. Средняя длительность различных парольных фраз находится в пределах от 1.5 до 4.1 сек, а ее распределение по отдельным парольным фразам видим на рис. 4.

Для каждой парольной фразы исследовался допустимый диапазон радиусов кластеров (R), изменяющийся, в зависимости от используемой парольной фразы, от 5.5 до 26 условных единиц. Диапазон изменения R исследовался начиная от появления первого кластера с количественным составом более двух дикторов (D > 2) и кончая величиной R максимально возможной при работе алгоритма.

Результаты кластеризации голоса 50-и дикторов при различных условиях экспериментов приведены в табл. 1. Для каждой парольной фразы, состоящей из N слов при оптимальном радиусе R было получено K кластеров дикторских голосов. Полученные кластеры обозначены штриховкой под номером, соответствующим диктору-центру кластера - Ср. По результатам, приведенным в таблице, можно выделить 3 группы дикторов - центров кластеров:

- "устойчивые" дикторы-центры кластеров (L = 5,6);
- "среднеустойчивые" дикторы - центры кластеров (L = 3,4);
- "неустойчивые" и "одиночные" дикторы-центры кластеров (L = 1,2),

где L - частота появления диктора-центра в разных экспериментах. Дальнейший анализ показал, что приведенная классификация справедлива и для самих кластеров и полностью с ней коррелирует. Принятый подход классификации дикторов позволил получить 5 "устойчивых" кластеров с дикторами-центрами Ср = 1,29,38,44,45; 4 "среднеустойчивых" кластера с Ср = 25,35,36,48, а также ряд "неустойчивых" и "одиночных" кластеров. Независимо от использовавшейся парольной фразы, перечисленные кластеры в большинстве присутствовали после классификации исследовавшейся выборки дикторов. Кроме этого, с изменением R изменялся их количественный и качественный состав. На рис. 1 приведен

Рис. 1.

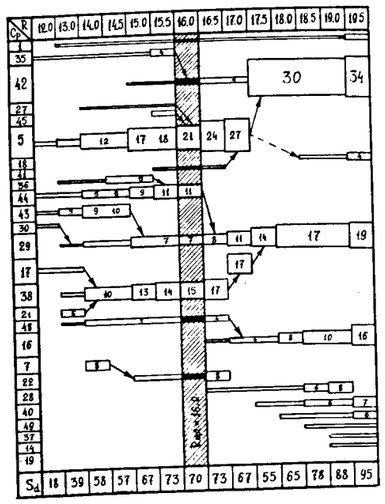


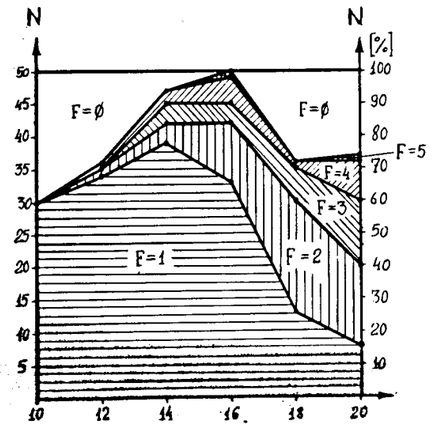
Табл. 1.

N	R_{opt}	K
3	6.5	10
4	9.0	12
5(1-й)	8.0	11
5(2-й)	10.5	13
7	16.0	9
10	19.0	7

граф динамики развития кластеров при изменении радиуса кластера для парольной фразы из 7 слов. На рисунке хорошо виден процесс образований, развития и распада отдельных кластеров и общая картина их преобразований. Кластеры обозначены прямоугольниками, толщина которых пропорциональна количеству дикторов в кластере, а само их количество приведено цифрой в прямоугольнике. Из рисунка видно, что при малых радиусах (R = 12.0 ... 15.0) происходит процесс формирования кластеров, при R = 15...16 наступает диапазон их оптимальности, а при R > 16.0 начинается их дробление и распад на ряд новых малочисленных кластеров /6/.

При выборе оптимальной величины радиуса нами учитывались как и общая картина развития кластеров, представленная на данном рисунке, так и отдельные параметры, такие как количество дикторов, входящих во все кластеры при данном R - Sd и частота попадания одних и тех же дикторов в различные кластеры - F. Величина Sd выбиралась таким образом, чтобы, по возможности, незначительно превышать число исследовавшихся дикторов (50). А F выбиралась из соображений минимального количества повторов дикторов в различных кластерах. Т.е. максимальным должно быть количество дикторов, для которых F = 1, и минимальным - для которых F = 0 и F > 2. На рис. 2 приведена диаграмма распределения количества дикторов N по частоте их попадания F в различные кластеры при изменении радиуса кластера R для парольной фразы из 7 слов. По перечисленным параметрам, приведенным на рис. 1 и рис. 2, был осуществлен выбор оптимального R, который для рассматриваемого случая равен 16 усл. единицам.

Рис. 2.



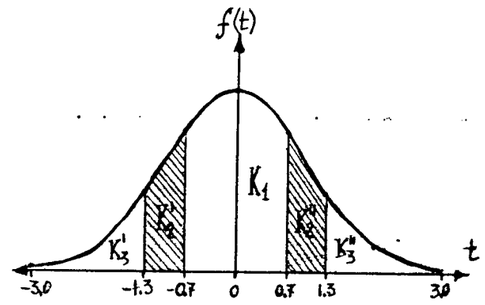
РЕЗУЛЬТАТЫ КЛАСТЕРИЗАЦИИ ГОЛОСОВ

В результате проведенных исследований по кластеризации голосов различных дикторов можно сделать вывод, что произвольных пользователей по их голосу можно разделить на 3 группы:

- "устойчивые" к кластеризации дикторы;
- "неустойчивые" к кластеризации дикторы;
- "некластеризуемые" дикторы.

На рис. 3 показано распределение плотности F(t) от их нормированного отклонения t от средних кластерных эталонов. В первую из перечисленных выше групп дикторов, изображенной на рисунке областью K1, входит 50% дикторов, во вторую (K2',K2") - 30% и в третью (K3',K3") - 20%. Дикторов, вошедших в первые две группы, можно объединить в группу условно кластеризуемых дикторов. А дикторы, вошедшие в группу "некластеризуемых", работать с неадаптивной СРР не смогут. Их речь можно распознавать только пользуясь адаптивными СРР. Интересно, что "некластеризуемыми" становятся как "плохо сотрудничащие", так и "хорошо сотрудничащие" с СРР дикторы, которые в одинаковой мере, но с разной полярностью удалены от средних кластерных эталонов (зоны K3' и K3").

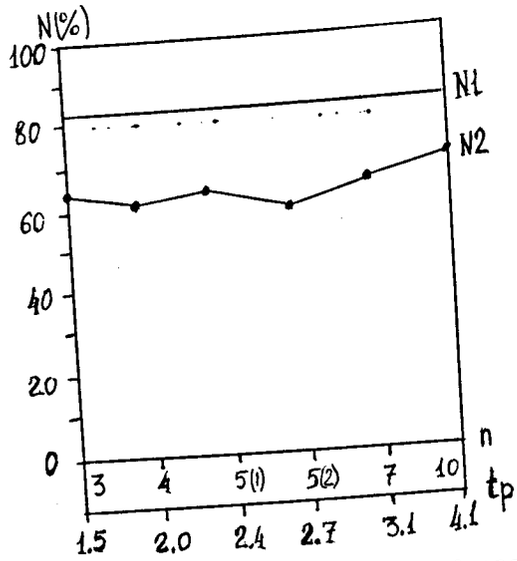
Рис. 3.



РЕЗУЛЬТАТЫ РАСПОЗНАВАНИЯ РЕЧИ ПРОИЗВОЛЬНЫХ ДИКТОРОВ

В экспериментах по распознаванию речи произвольных пользователей по кластерным эталонам приняло участие 50 дикторов, записывавших свои эталоны в банк эталонов речи, а также 10 новых дикторов. Каждый диктор произнес десять слов - цифры от 0 до 9. Усредненные результаты распознавания голоса всех дикторов по кластерным эталонам для каждой парольной фразы приведены на рис. 4. и обозначены кривой N2. Минимальная надежность распознавания получена при использовании парольной фразы из 4-х слов - 63.1%, максимальная - для парольной фразы из 10-и слов - 68.6%. Следует сделать оговорку, что все результаты получены для дикторов, не проходивших никакого предварительного обучения работы с СРР. По данным некоторых исследований при наличии предварительного обучения дикторов и их адаптации к работе с СРР, надежность распознавания их речи возрастает на 5...15% соответственно после 3...9-и часового обучения. Учитывая эти данные, можно надеяться, что средняя надежность распознавания речи "произвольного-обученного" пользователя нашей системой будет составлять около 85%. Это предположение подтверждается также тем, что для 2-х "сотрудничающих" дикторов, принимавших участие в эксперименте и образовавших "свой" кластер с Ср = 1, надежность распознавания во всех экспериментах была равной 100%.

Рис. 4.



n - количество слов в парольной фразе
tp - средняя длительность парольной фразы

Средняя надежность распознавания речи произвольных дикторов в адаптивной ССР, когда они имели "свои" эталоны, также приведена на этом рисунке и обозначена N1. Эта величина равна 82%.

ВЫВОДЫ

В результате проведенных исследований можно сказать, что к вопросу о распознавании речи произвольного диктора нужно подходить дифференцированно. Сначала необходимо определить возможность эффективной работы конкретного диктора с неадаптивной системой распознавания речи и только в случае позитивного результата этот диктор может приступать к работе с ССР. Определить "пригодность" произвольного диктора для работы с неадаптивной СРР можно по произнесенной им парольной фразе. Надежность распознавания речи произвольного пользователя сильно зависит от его подготовки к работе с СРР, иначе говоря его "сотрудничества" с системой. Проведенные исследования показывают возможные пути совершенствования неадаптивной СРР и несмотря на ряд трудностей при решении проблемы позволяют применять такие системы в ограниченных практических целях.

ЛИТЕРАТУРА

/1/ P. Fonsale "Connected-word recognition system using speaker-independent phonetic features", Proc. ICASSP-83, Boston, pp. 312 - 315.

/2/ Р.Я. Гумецкий, В.Н. Мазур "Диалоговая система обучения программированию и работе на ЭВМ с речевым запросом и ответом, ориентированная на массового пользователя", Труды сов.-франц. симпозиума "Акустический диалог человека с машиной", ИППИ АН СССР, Москва, 1984, стр. 51 - 54.

/3/ Р.Я. Гумецкий, В.Н. Мазур, В.А. Марченко "Система распознавания дискретной речи произвольного диктора". В кн.: "Автоматическое распознавание слуховых образов", Тезисы докл. и сообщ. АРСО-14, Каунас, 1986, ч. 1, стр. 94 - 95.

/4/ Г.Г. Гюльназарян, Ф.Е. Коркмазский, В.Н. Мазур "Использование систем автоматического речевого ввода", ВЦ АН СССР, Москва, 1986, стр. 19 - 25.

/5/ H. Sakoe, S. Chiba "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", IEEE Trans. on ASSP, v. 23, N1, 1978, pp. 43 - 49.

/6/ В.Н. Мазур "Исследование кластеризации дикторских голосов для распознавания речи произвольного диктора", Тезисы докл. и сообщ. АРСО-14, Каунас, 1986, ч. 1, стр. 72.