Литература

1. *Емеличев В.А.* [*и др.*]. Лекции по теории графов. М.: Наука, 1990.
2. *Гэри М., Джонсон Д.* Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982.
3. *Kirkpatrick D. G., Hell P.* On the completeness of a generalized matching problem. // Proc of the tenth annual ACM symposium on Theory of computing, ACM New York, NY, USA, 1978. P. 240–245.
4. *Dyer M., Frieze A.* On the complexity of partitioning graphs into connected subgraphs. // Discrete Applied Mathematics, 1985, №10. P. 139–153.
5. *Monnot J., Toulouse S.* The path partition problem and related problems in bipartite graphs // Operations Research Letters, 2007, №35. P. 677 – 684.
6. *Monnot J., Toulouse S.* Bounded-size path packing problems. / In Vangelis Th. Paschos, editor, Willey, 2001. (Combinatorial Optimization and Theoretical Computer Science: Interfaces and Perspectives). P. 455–493
7. *Korobitsyn D.V.* Complexity of some problems on hereditary classes of graphs. // Discrete Mathematics and Applications, 1992, № 2. P. 191–199.
8. *Alekseev V.E.* On the local restrictions effect on the complexity of finding the graph independence number. // Combinatorial-Algebraic Methods in Applied Mathematics, 1983, P. 3–13.
9. *Boliac R., Cameron K., Lozin V.* On computing the dissociation number of bipartite graphs. //Ars Combinatoria, 2007, № 72. P. 241–253.
10. *Panda B.S.* A linear time recognition algorithm for proper interval graphs // Inform Proc Lett, 2003, № 87(3). P. 153–161.
11. *Spinrad J.* Bipartite permutation graphs // Discrete Applied Mathematics, 1987, № 18. P. 279–292.

РАЗРАБОТКА ТЕКСТОНЕЗАВИСИМОЙ СИСТЕМЫ ИДЕНТИФИКАЦИИ ДИКТОРА НА ОСНОВЕ ФОНЕМНОГО РАЗБИЕНИЯ И GMM

В. В. Захарова

ВВЕДЕНИЕ

В связи с наращиванием потоков передачи речи по каналам связи, требующим аутентификации клиентов, задача идентификации человека по голосу является одной из востребованных сегодня задач распознавания образов. С точки зрения пользователя более удобна текстонезависимая идентификация, также она незаменима в случаях, когда факт идентификации необходимо скрыть. Но отказ от использования лингвистической информации ведет к снижению качества распознавания, так как невозможна тщательная калибровка входных лексем с помощью идентичного речевого материала [1]. В данной статье предлагается способ построения

текстонезависимой системы идентификации диктора, позволяющий добиться высокого качества распознавания в том числе за счет использования лингвистической информации на этапе построения моделей.

ПОСТРОЕНИЕ ТЕКСТОНЕЗАВИСИМОЙ СИСТЕМЫ ИДЕНТИФИКАЦИИ ДИКТОРА НА ОСНОВЕ ФОНЕМНОГО РАЗБИЕНИЯ И МОДЕЛЕЙ ГАУССОВЫХ СМЕСЕЙ

Построение системы идентификации диктора связано с такими этапами решения этой задачи, как выделение характерных признаков речи, построение на их основе моделей дикторов и принятия решения об авторстве после сравнения записи с построенными моделями. В разработанной системе предлагается для формирования моделей дикторов строить дополнительную общую для них лингвистическую модель.

Извлечение признаков из речевых сигналов

Для подавления фоновых шумов в качестве предобработки сигнала при обучении лингвистической модели применяются удаление постоянного амплитудного смещения, дизеринг и увеличение амплитуды высокочастотных компонент сигнала, а после вычисления кепстральных коэффициентов для каждого диктора выполняется нормализация их математического ожидания и дисперсии. В качестве оконной функции, применяемой к отрезкам сигнала, система позволяет использовать окна Хэмминга и Ланцоша, по умолчанию в пользу быстроты вычислений используется окно Хэмминга.

В качестве базовых признаков, характеризующих физиологически обусловленные особенности человеческой речи, используются мел-частотные кепстральные коэффициенты (МЧКК) [2] ввиду простоты их вычисления и хорошей аппроксимирующей способности за счет учета особенностей восприятия человеком высоты звука в зависимости от частоты сигнала. Но так как эти коэффициенты не учитывают информацию о дикторе, находящуюся в области высоких частот, в дополнение к ним используются обратные мел-частотные кепстральные коэффициенты (ОМЧКК) [3], компенсирующие этот недостаток.

Построение моделей

Лингвистическая модель формируется поэтапно: на каждом шаге строится более совершенная модель на основе предыдущей.

Сначала на основе извлеченных низкоуровневых признаков и их производных на основе алгоритмов моделирования гауссовых смесей (англ. GMM, Gaussian Mixture Model) и скрытых марковских моделей

строятся модели фонем без использования информации об их взаимном расположении и влиянии его на звучание фонем. Затем на их основе формируются модели трифонов, при этом используется только их наиболее употребляемая часть, формируемая с помощью фонетического дерева решений.

Далее с помощью алгоритма линейного дискриминантного анализа уменьшается размерность пространства входных данных путем построения состояний марковских моделей для исходных векторов признаков. Затем с помощью линейных преобразований признаков, максимизирующих среднее правдоподобие, в новом редуцированном пространстве формируются матрицы трансформаций для каждого диктора.

На последнем этапе (англ. SAT, Speaker Adaptive Training) эти матрицы используются для нормализации информации о дикторе в моделях трифонов, так как разбиение на фонемы должно происходить независимо от того, кто автор записи. Разбиение на фонемы происходит уже в пространстве дикторонезависимых признаков. Для удаления из признаков информации о дикторе используется алгоритм линейной регрессии максимального правдоподобия. Для уточнения параметров после обучения каждой новой модели заново выполняется пофонемное выравнивание обучающих записей согласно алгоритму обучения Витерби.

При построении модели диктора вычисленные для отрезков записей его речи вектора низкоуровневых признаков группируются по фонемам, которым они принадлежат, объединяясь в супервекторы. Информация о принадлежности каждого отрезка определенной фонеме диктора получается с помощью наиболее совершенной из лингвистических моделей, а конкретно – SAT-модели. Таким образом, собирается информация о произношении диктором различных фонем. Для каждой фонемы строится и сохраняется модель гауссовых смесей.

Распознавание и принятие решений

При распознавании для каждого отрезка распознаваемой записи вычисляется его похожесть на фонемы каждого диктора, при этом предварительное разбиение на фонемы уже не производится. Мерой близости вектора признаков отрезка речи и фонемы является средняя логарифмическая вероятность модели гауссиан фонемы, вычисленная для этого вектора.

При принятии решения на закрытом множестве дикторов используется метод ближайшего соседа, то есть в качестве автора записи выбирается автор максимального количества ближайших по выбранной мере фонем. При принятии решения на открытом множестве дикторов учитывается вероятность его авторства по сравнению с остальными дикторами.

ЭКСПЕРИМЕНТ

Разработанная система протестирована на корпусах речи TIMIT и LibriSpeech. Тестовые и обучающие записи распределены в отношении один к четырем.

Эксперименты показали значительное увеличение точности идентификации при использовании ОМЧКК в дополнение к МЧКК при увеличении количества участвующих в распознавании дикторов (см. рис. 1).

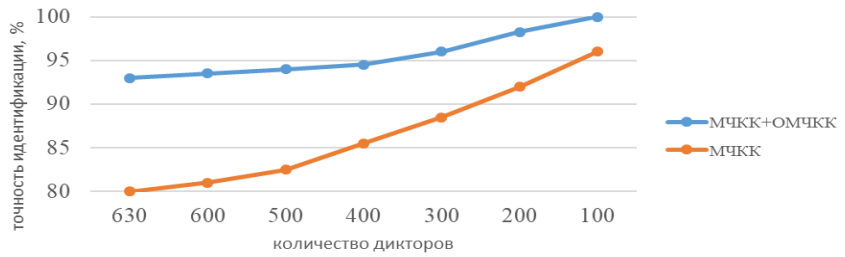

Рис. 1. Зависимость точности идентификации от количества дикторов при использовании МЧКК и ОМЧКК (корпус TIMIT)

Установлено, что оптимальное количество гауссиан при формировании моделей фонем равно трем (при большем многие фонемы пропускаются при обучении, так как имеют слишком малую длительность), а также что важно, чтобы при разбиении записи окна перекрывались больше, чем наполовину, а оптимальные размеры окна и интервала его взятия составляют 25 и 10 мс соответственно (см. рис. 2).

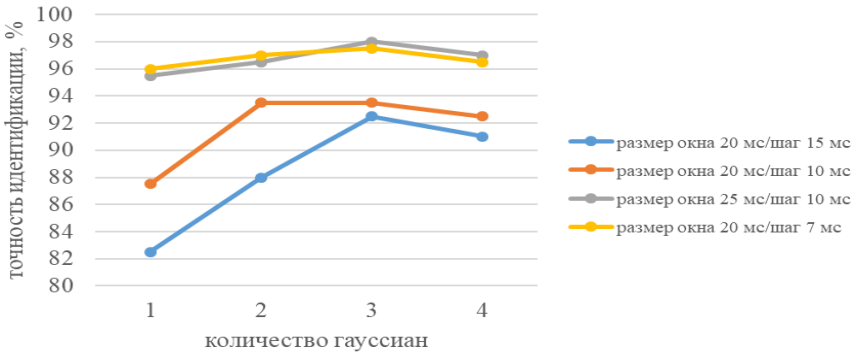

Рис. 2. Зависимость точности идентификации от размера и шага окна и количества гауссиан (для ста дикторов корпуса TIMIT)

ЗАКЛЮЧЕНИЕ

Предложен способ построения текстонезависимой системы идентификации диктора на основе фонемного разбиения и моделей гауссовых

смесей. Показано, что использование лингвистической информации, а именно, разбиения на фонемы, при формировании моделей диктора позволяет добиться высокой точности идентификации, при этом не требуя использования этой информации в процессе распознавания. Также показано, что использование обратных мел-частотных кепстральных коэффициентов позволяет значительно (до 20 %) повысить точность идентификации, особенно при большом количестве дикторов.

Литература

1. *Лютова Д. А.* Основные задачи и методы технологий распознавания говорящего по голосу // Вестник Московского государственного лингвистического университета: научно-технический журнал. 2010. Выпуск 13. С. 131–147.
2. *Christian Müller (Ed.).* Speaker Classification I. Springer-Verlag, Berlin Heidelberg, Germany. 2007. P. 229–231.
3. *S. Chakroborty and G. Saha*. Improved Text-Independent Speaker Identification Using Fused MFCC and IMFCC feature Sets Based on Gaussian Filter // International Journal of Signal Processing. 2009. Vol. 5. No. 1. P. 11–19.

ПРОГРАММА ESI ANALYSIS ДЛЯ ПОСТРОЕНИЯ И АНАЛИЗА ОПЕРЕЖАЮЩИХ ЭКОНОМИЧЕСКИХ ИНДИКАТОРОВ БЕЛОРУССКОЙ ЭКОНОМИКИ ПО ОПРОСНЫМ ДАННЫМ

Е. В. Кондратович, В. И. Малюгин

ВВЕДЕНИЕ

Одной из ключевых задач анализа и прогнозирования экономической активности на макроуровне является разработка систем раннего обнаружения смены фаз экономических циклов на основе статистических методов и эконометрических моделей [1]. Для получения ранних сигналов о смене фаз и оценивания моментов смены фаз, называемых поворотными точками, в рамках указанных систем применяются так называемые опережающие экономические индикаторы. Индикатор считается опережающим, если поворотные точки его цикла (точки переключения фаз цикла «спад» и «подъем») опережают поворотные точки цикла базового экономического индикатора, в качестве которого используется реальный ВВП. В статье представляется программа ESI Analysis, реализующая методологию построения и применения опережающих индикаторов в виде сводного индекса экономических настроений (ESI – Economic Sentiment Index) для белорусской экономики и индексов доверия (Confident Indexes) для отдельных видов экономической деятельности (ВЭД) по опросным данным системы мониторинга предприятий Национального банка Республики Беларусь [2].