УДК 681.3.07

П.В. Малинин, В.В. Поляков

Иерархический подход в задаче идентификации личности по голосу с помощью проекционных методов классификации многомерных данных

Предложен подход к классификации голосовых сигналов с помощью метода проекции на латентные структуры, использующий иерархическое разделение перекрывающихся голосовых сигналов большого числа дикторов. Показана эффективность этого подхода для задачи идентификации личности по голосу.

Ключевые слова: идентификация личности, анализ многомерных данных, проекционные методы, иерархический подход.

Развитие новых технологий передачи и обмена информации и увеличение числа компьютерных преступлений делают актуальной разработку методов биометрической идентификации личности, повышающих информационную безопасность автоматизированных систем. Одним из методов биометрической идентификации, используемых в системах разграничения доступа к конфиденциальной информации, является классификация личности по голосовым данным. В то же время такая идентификация отличается значительно меньшей надежностью, чем ряд других биометрических методов (в частности, по следам пальцев рук, по изображению радужной оболочки глаза). Существенное повышение надежности и достоверности идентификации по голосу может быть достигнуто путем применения математического аппарата проекционных методов анализа многомерных данных [1, 2, 9]. Однако и в этом случае в реальных условиях, характеризующихся необходимостью различения голосовых данных большого числа дикторов, возникают трудности, обусловленные появлением голосов с близкими частотными характеристиками.

В настоящей работе предлагается подход, основанный на иерархическом применении проекционных методов классификации многомерных данных к множеству перекрывающихся голосовых сигналов.

В проекционных методах анализа и классификации характеристики исходных данных (как правило, искаженных случайными факторами) заменяются их геометрической проекцией на подпространство, которое ориентируется таким образом, чтобы выявить анализируемое свойство. В задаче идентификации личности по голосу в качестве такого свойства выступает принадлежность речевого сигнала тому или иному диктору. В настоящей работе для классификации используется метод проекции на латентные структуры [3], в котором в качестве исходных данных выступают коэффициенты кепстрального разложения голосовых сигналов [4]. Результаты расчетов представляются в виде наглядных графиков счетов, на которых точки соответствуют отдельным исследуемым объектам (речевым сигналам), а близость двух точек друг к другу означает схожесть свойств (принадлежность этих сигналов одному диктору).

В качестве иллюстрации на рис. 1 приведен график счетов, полученный обработкой записи речи десяти дикторов, произносивших одно и то же словосочетание по пять раз (цифрами обозначен номер диктора). Как видно из рис. 1, отдельные точки достаточно отчетливо группируются в обособленные группы (кластеры), каждая из которых соответствует одному диктору.

При увеличении числа дикторов возникает перекрывание областей, описывающих голосовые данные отдельных дикторов. Эту ситуацию иллюстрирует рис. 2, на котором представлен график счетов для случая двадцати дикторов. Видно, что провести надежную классификацию голосовых сигналов по их принадлежности конкретным дикторам на основе данного графика не представляется возможным.

Для того чтобы сохранить применимость описанного метода для случая большого числа дикторов, речевые сигналы которых перекрываются, был применен следующий иерархический подход [5, 6]. Все объекты, в качестве которых выступали отдельные звуковые сигналы (представленные счетами) [7], рассматривались как совокупность кластеров различного уровня. Такие кластеры включали в себя перекрывающиеся голосовые данные, принадлежавшие дикторам с близкими частотными характеристиками голоса [8]. Процедура обработки с помощью изложенного выше проекционного метода последовательно применялась к каждому из кластеров до тех пор, пока не происходило его разде-

ление на совокупность явно обособленных групп, описывающих кластеры звуковых сигналов отдельных дикторов.

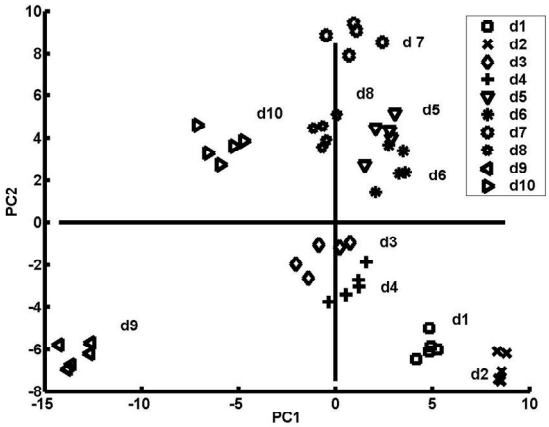

Рис. 1. График счетов 10 дикторов

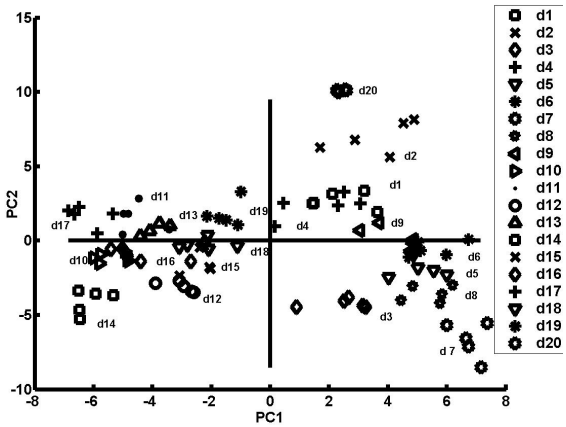

Рис. 2. График счетов 20 дикторов

На рис. 3 и 4 представлены этапы последовательного применения иерархической классификации к данным, приведенным на рис. 2. На первом этапе обрабатывался кластер, расположенный в левой полуплоскости и включавший в себя перекрывавшиеся группы точек, соответствовавшие отдельным дикторам. Полученный график счетов, представленный на рис. 3, позволил выделить обособленные кластеры d10, d11, d14, d16, d17, d18. На втором этапе метод проекции на латентные структуры применялся к кластеру, образованному перекрывающимися группами d12, d13, d15, d19. Как следует из рис. 4, в результате было получено вполне отчетливое разделение первоначально перекрывавшихся групп точек на кластеры, каждый из которых соответствовал одному диктору.

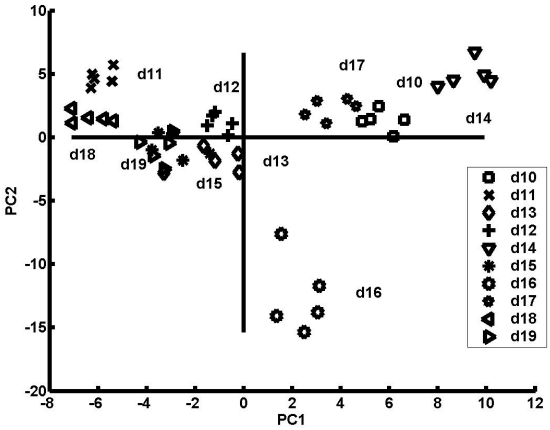

Рис. 3. Счета групп d10, d11, d12, d13, d14, d15, d16, d17, d18, d19

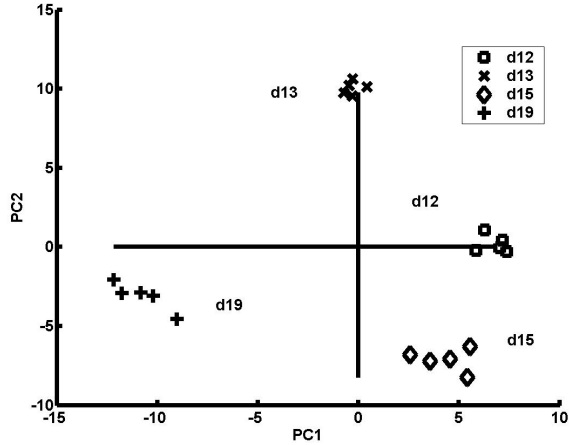

Рис. 4. Счета групп d12, d13, d15, d19

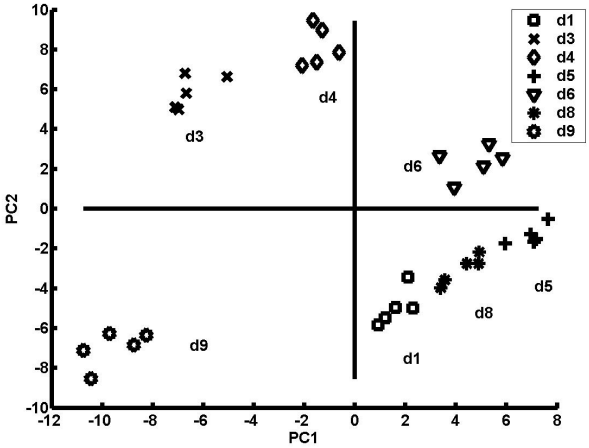

Рис. 5. Счета групп d1, d3, d4, d5, d6, d8, d9

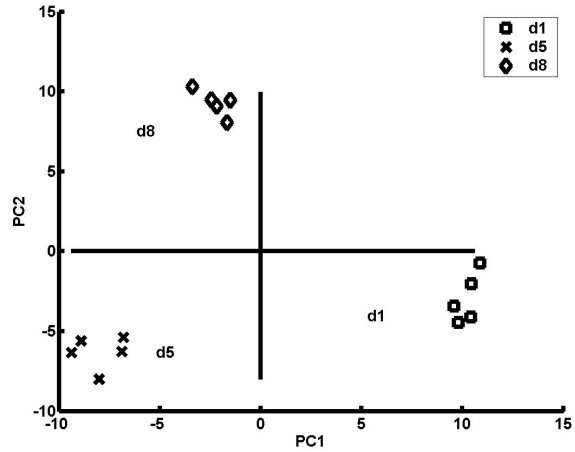

Рис. 6. Счета групп d1, d5, d8

Совершенно аналогично проводилась обработка данных для правой полуплоскости графика счетов на рис. 2. Именно на первом этапе исследовался кластер, включавший в себя близко расположенные группы d1, d3, d4, d5, d6, d8, d9. Полученный график счетов (рис. 5) позволил выделить кластеры d3, d4, d6, d9, относившиеся к отдельным дикторам. На следующем этапе обрабатывался кластер, образованный точками d1, d5, d8. Как видно из рис. 6, на этом этапе также было получено отчетливое разделение соответствующих групп точек.

Таким образом, предложенный иерархический подход позволяет провести достаточно надежную классификацию голосовых данных и последующую идентификацию личности в случае большого числа дикторов с близкими частотными характеристиками голоса.

Литература

1. Эсбенсен К. Анализ многомерных данных. Избранные главы / К. Эсбенсен: Пер. с англ. С.В. Кучерявского; под ред. О.Е. Родионовой. – Черноголовка: Изд-во ИПХФ РАН, 2005 – 157 с.

2. Richard G. Brereton Introduction to multivariate calibration in analytical chemistry // Analyst. – 2000. – № 125. – P. 2125–2154.

3. Кучерявский С.В. Применение методов анализа многомерных данных к исследованию структуры материалов / С.В. Кучерявский, В.В. Поляков // Заводская лаборатория. Диагностика материалов. – 2007. – Т. 73, № 8. – С. 32–36.

4. Малинин П.В. Применение методов анализа многомерных данных в задаче идентификации личности // Проблемы информационной безопасности государства, общества и личности, безопасность нанотехнологий: матер. 11-й Всерос. науч.-практ. конф. – Томск: В-Спектр, 2009. – С. 29–33.

5. Мандель И.Д. Кластерный анализ. – М.: Финансы и статистика, 1988. – 176 с.

6. Прикладная статистика: Классификация и снижение размерности: справ. изд. / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин; под ред. С.А. Айвазян. – М.: Финансы и статистика, 1989. – 607 с.

7. Рабинер Л.Р. Цифровая обработка речевых сигналов / Л.Р. Рабинер, Р.В. Шафер: – Пер. с англ.; под ред. М.В. Назарова, Ю.Н. Прохорова. – М.: Радио и связь, 1981. – 224 с.

8. Рамишвили Г.С. Автоматическое опознавание говорящего по голосу. – М.: Радио и связь, 1981. – 496 с.

9. Мещеряков Р.В. Использование информационных критериев для оценки иерархических диалоговых систем // Научно-технические ведомости СПбГПУ. – 2009. – № 4(82). – С. 113–122.

Малинин Петр Владимирович
Аспирант каф. прикладной физики, электроники и информационной безопасности
Алтайского государственного университета (АлтГУ), г. Барнаул
Эл. адрес: malinin@phys.asu.ru

Поляков Виктор Владимирович
Доктор физ.-мат. наук, профессор, декан физико-технического факультета,
зав. каф. прикладной физики, электроники и информационной безопасности АлтГУ, г. Барнаул
Тел. (+7-385-2) 36-70-59
Эл. адрес: pvv@asu.ru

P.V. Malinin, V.V. Polyakov

Hierarchical approach in a problem of voice speaker identification on the basis of projection methods of the analysis multivariate data

The approach of classification of vocal signals on the basis of consecutive application of a method projection on latent structures of the analysis multivariate data. By means of the hierarchical approach possibility of division of voices concerning a great number of speakers. This approach is applied to more reliable division of the speakers vocal.

Keywords: persons identification, multidimensional data analysis, projective methods, hierarchical approach.