УДК 004.934.8'1, 57.087.1

Сидоренко Е. В.

Яковенко А. А.

Санкт-Петербургский политехнический университет Петра Великого

ПРИМЕНЕНИЕ НЕЙРОНЫХ КАРТ ПРИЗНАКОВ С УПРАВЛЯЕМЫМ ОБУЧЕНИЕМ   
В ЗАДАЧЕ ТЕКСТОНЕЗАВИСИМОЙ идентификациИ пользователя по голосу   
на закрытом множестве дикторов

Большие объемы данных, накопленные в телекоммуникационных системах, могут содержать информацию о тексте высказываний, о личности говорящего, его эмоциональном состоянии, состоянии здоровья и прочих дополнительных факторах. Обработка и анализ этой информации в биометрических системах позволяет получать новые знания о личности говорящего, проводить семантический анализ высказываний, автоматически анализировать диалоги. Биометрические системы находят применение при автоматической идентификации в телефонии, при обработке речевых баз данных, в криминалистических исследованиях, в системах контроля и управления доступом к физическим объектам и к удаленным ресурсам в компьютерных сетях, банковских приложениях и т.д. Также биометрические системы имеют перспективу применения в производстве при контроле доступа к оборудованию. Голос каждого человека уникален, в связи с чем его использование при доступе к конфиденциальной информации может повысить защищённость системы и обеспечить сохранность данных. Преимущество голосовой биометрии заключается в обеспечении быстрого и удобного способа обслуживания, отсутствии необходимости дополнительных действий со стороны пользователя и установки дорогостоящего оборудования.

В биометрических системах применяются формальные методы обработки речевых сигналов. Но в процессе эксплуатации таких систем могут возникать значительные ошибки принимаемых решений. Ошибки вызваны, как правило, существенной изменчивостью речи, зависят от речевых признаков, типа телекоммуникационных каналов, метода кодирования, алгоритма сжатия, и возрастают в условиях внешних акустических шумов.

Проблему голосовой биометрии принято разделять на задачу идентификации и верификации на закрытом или открытом множестве дикторов. Верификация заключается в проверке системой заявленного пользователем идентификатора, а идентификация подразумевает автоматизированный процесс определения идентификатора по экземпляру речи. Если количество пользователей фиксировано, то говорят о закрытом множестве дикторов, в противном случае – об открытом. Кроме того, распознавание может осуществляться как в текстозависимом, так и в текстонезависимом контексте.

В данной работе предложен подход к решению проблемы голосовой биометрии, основанный на построении единой кодовой книги для всех зарегистрированных пользователей с использованием метода нейронных карт признаков с управляемым обучением. Испытания проводились на текстонезависимых высказываниях.

***Цель работы.*** Биометрическая идентификация пользователя по голосу на закрытом множестве дикторов.

***Голосовые признаки.***

Речевой сигнал существенно отличается от других акустических сигналов, так как произносится человеком для человека и служит для обмена информацией между людьми. Поэтому в системах распознавания дикторов целью первичной обработки речевого сигнала является выделение признаков речи, специфичных для отдельных дикторов.

Распространенными речевыми признаками для систем идентификации дикторов являются:

* частота основного тона
* частота формант
* кепстральные коэффициенты

В большинстве автоматических систем идентификации дикторов в качестве признаков используются векторы, на основе кепстральных коэффициентов.

***MFCC (Mel-Frequency Cepstral Coefficients)***

Вычисление MFCC осуществляется путем разбиения исходного сигнала на фреймы, для каждого из которых применяется алгоритм быстрого преобразования Фурье. Полученный спектр проецируется на mel-шкалу, таки образом позволяя выделить наиболее значимые для восприятия человеком частоты, где mel — это психофизическая единица высоты звука, основанная на субъективном восприятии среднестатистическими людьми.

Формула преобразования из частотной области в мел шкалу:

Формула преобразования мел в частоту:

Для того, чтобы разложить полученный спектр по mel-шкале, необходимо применить банк треугольных фильтров, то есть попарно перемножить его со значениями спектра. После приведения частоты к мел шкале необходимо получить кепстральные коэффициенты. Это можно сделать с помощью дискретного косинусного преобразования.

***PLP (Perceptual Linear Prediction, PLP)***

PLP очень схожи с MFCC, так как извлечение этих признаков основано на мел-частотном банке фильтров. К банку фильтров применяется кривая предварительного сжатия с равной громкостью. После чего, применяется обратное преобразование Фурье и алгоритм Левинсона-Дарбина, чтобы получить LP коэффициенты. После чего полученные LP коэффициенты преобразуются к кепстральным с помощью дискретного косинусного преобразования.

***Алгоритм самоорганизующихся карт (СОК)***

Мозг человека почти полностью окружен церебральной корой, которая скрывает другие его части. В смысле сложности, кора головного мозга, возможно, превышает другую известную структуру. Информация, полученная от сенсоров, отображаются на соответствующую область в упорядоченном виде.

Самоорганизующиеся искусственные нейронные сети (ИНС), концепция которых была предложена Т. Кохоненом, моделируют особенности упорядоченного представления информации, характерные картам коры головного мозга:

* каждый входной образ сохраняется в собственном контексте;
* нейроны, близко расположенные на карте, связаны с близкой звуковой информацией (тонотопическая упорядоченность слуховой коры).

В данной работе, модель самоорганизующейся нейронной карты с адаптивными весами получает входной вектор кепстральных коэффициентов  и выполняет отображение этого вектора на множество выходных откликов таким образом, что отклики приобретают топологический порядок. Алгоритм самоорганизации включает инициализацию синоптических весов  малыми случайными значениями, получение последовательности векторов , получение нейрона, выигравшего на текущем шаге  по критерию минимального расстояния до входного образа:

Обновление вектора весов выполняется по следующей формуле:

где - параметр скорости обучения, - функция, характеризующая влияние выигравшего нейрона на его соседей, и динамически изменяющаяся во времени.

***Самоорганизующиеся карты с управляемым обучением***

Самоорганизующиеся ИНС в силу своих особенностей эффективно решают задачи обучения без учителя. Основным отличием алгоритма СОК при обучении с учителем является присутствие классифицирующей информации о принадлежности вектора признаков к определенному классу. В каждом векторе один из новых компонентов имеет значение 1 в зависимости от класса вектора и другие 0. После этой реконструкции класс каждого нейрона отображения определяется путем принятия максимального количества добавленных компонентов, и соответственно указывается метка. После этого, ранее добавленные дополнительные компоненты удаляются.

Такой подход позволяет использовать особенности СОК в задачах классификации, при наличии размеченной выборки данных. Результатом обучения является кодовая книга, гарантирующая разделение входных экземпляров в многомерном пространстве, согласно известным классам. Преимущество такого подхода заключается в том, что при получении речевых данных после каждой новой сессии идентификации, карта дообучается, т.е. уточняется описание класса соответствующего голоса. При этом, данную систему можно использовать как в текстонезависимом, так и в текстозависимом режиме, без необходимости вносить изменения в алгоритм.

***Результаты тестирования***

Тестирование проводилось с использованием коэффициентов PLP и MFCC извлеченных из речевых высказываний дикторов различной половой принадлежности.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Диктор** | **MFCC** | | **PLP** | |
| ***Ошибка квантования*** | ***Топографическая ошибка*** | ***Ошибка квантования*** | ***Топографическая ошибка*** |
| f1/m1 | 0.2046 | 0.0934 | 0.2222 | 0.0782 |
| f1/f2 | 0.2237 | 0.0891 | 0.2362 | 0.0821 |
| m1/m2 | 0.2212 | 0.0906 | 0.2412 | 0.0825 |

f1, f2 – дикторы женского пола

m1, m2 – дикторы мужского пола

После обучения карты к ней был применен алгоритм k-средних, в результате работы которого были построены кластеры. Лучший кластер был выбран на основе суммы квадратов ошибок.

Ошибка квантования:

, где – вектор данных, а – нейроны сети

Топографическая ошибка:

, где функция равняется 1, если первый и второй нейроны победители вектора данных являются смежными, иначе 0.

|  |  |  |
| --- | --- | --- |
| **Дикторы** | **MFCC** | **PLP** |
| f1/m1 |  |  |
| f1/f2 |  |  |
| m1/m2 |  |  |

**Заключение**

В данной работе был реализован подход к построению карты признаков с управляемым обучением для задачи классификации голосов дикторов в текстонезависимом режиме идентификации на закрытом множестве. Такой подход призван упростить построение системы голосовой биометрии, которая может улучшать свои характеристики в процессе эксплуатации. В контексте данной работы были рассмотрены основные акустические признаки для описания речевого сигнала с целью сравнения их эффективности. В результате тестирования было произведено сравнение признаков MFCC и PLP. Наиболее эффективными себя показали мел-кепстральные коэффициенты.

Однако, следует отметить эффективность, с которой человек способен распознать голос знакомого индивида при различных внешних условиях и разном качестве речи. Поэтому повышение качества работы биометрических систем целесообразно связывать с разработкой биологически обоснованных методов и алгоритмов обработки речи. Полученные в рамках данной работы результаты предполагается использовать в дальнейшем для сопоставления с иными способами представления звукового сигнала и методами его дальнейшей обработки.

Работа выполнена при финансовой поддержке РФФИ (проект № 18-31-00304).

Литература:

1. Кохонен Т. Самоорганизующиеся карты / Т. Кохонен, 3-е издание; пер. с англ. В.Н. Агеева, ред. Ю.В. Тюменцев. – Москва: БИНОМ. Лаборатория знаний, 2014. – 655 с.
2. Хайкин С. Нейронные сети: полный курс / С. Хайкин, 2-е издание; пер. с англ. Н.Н. Куссуль, А.Ю. Шелестова, ред. Н.Н. Куссуль. – Москва: Издательский дом «Вильямс», 2006. – 1104 с.
3. Яковенко А.А. Искусственные нейронные сети в задаче автоматической голосовой биометрии // XV Всероссийская научная конференция "Нейрокомпьютеры и их применение". Тезисы докладов. - М. : ФГБОУ ВО МГППУ, 2017. - стр. 128-129.
4. Яковенко А.А., Малыхина Г.Ф. Текстонезависимое распознавание личности по голосу с использованием нейронной сети функций радиального базиса // Комплексная защита объектов информатизации и измерительные технологии: Сборник научных трудов Всероссийской научно-практической конференции с международным участием. - СПб : Изд-во Политехнического университета, 2014. - стр. 88-92.