УДК 57.087.1

Сидоренко Е. В.

Яковенко А. А.

Санкт-Петербургский политехнический университет Петра Великого

идентификация 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 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. Биометрические системы имеют перспективу применения в производстве при контроле доступа к оборудованию.

Существующие биометрические системы, выполняющие идентификацию, верификацию диктора, основаны на формальных методах обработки сигналов, и, к сожалению, имеют значительные ошибки принимаемых решений. Ошибки вызваны существенной изменчивостью речи, зависят от типа телекоммуникационных каналов, метода кодирования, алгоритма сжатия, возрастают в условиях внешних акустических шумов. В то же время человек практически без ошибок распознает голос знакомого диктора 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, моделирующих органы слуха человека.

***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.

Наиболее 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 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 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 (SOM)***

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.

Искусственные нейронные сети Кохонена моделируют свойства коры головного мозга упорядоченного представления информации, получаемой от системы восприятия звуков:

* каждый входной образ сохраняется в собственном контексте;
* нейроны, близко расположенные на карте, связаны с близкой звуковой информацией.

Нейронная сеть Кохонена 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.

***SOM supervised***

Обычно самоорганизующиеся карты используются для обучения без учителя, но изначально они были задуманы для задач статистического анализа. Основным отличием алгоритма SOM при обучении с учителем является присутствие классифицирующей информации о принадлежности вектора 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.

***Результаты тестирования***

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Диктор** | **MFCC** | | **PLP** | |
| ***Quantization error*** | ***Topographic error*** | ***Quantization error*** | ***Topographic error*** |
| 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 |

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**

Таким образом было произведено исследование карты самоорганизации Кохонена. Проведено обучение и тестирование на речевых признаках, в результате которого были построены кластеры с использованием алгоритма к-средних. Таким образом можно сделать вывод, что MFCC являются лучшими признаками, по сравнению с PLP.

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

Литература:

1. Yakovenko, A., Malykhina, G.F.Text-independent speaker recognition using radial basis function network Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9719,2016 , pp.74-81.
2. A.A. Yakovenko, G.F. Malykhina. Radial-Basis Functions Neural Network for Text Independent Speaker Identification. Proceedings of the 2014 International Conference on Methods in Applied SAciences (MMAS’14), pp 389-392.
3. Topology Preservation in SOM // WASET URL: https://waset.org/publications/6385/topology-preservation-in-som (дата обращения: 15.04.2018).
4. J. Vesanto. "Using SOM in Data Mining," Thesis for the degree of Licentiate of Science in Technology. Espoo, Finland, 2000.
5. J Kohonen, T.: Self-Organizing Maps. Vol. 30. of Springer Series in Information Sciences, 3rd ed., Springer-Verlag, Berlin Heidelberg (2001)