4. Документация по Firebase Cloud Messaging [Электронный ресурс]. – Режим доступа: https://firebase.google.com/docs/cloud-messaging/?hl=ru (дата обращения: 06.09.2019).
5. GSON: руководство пользователя [Электронный ресурс]. – Режим доступа: https://github.com/google/gson/blob/master/UserGuide.md (дата обращения: 06.09.2019).
6. Обзор администрирования устройством [Электронный ресурс]. – Режим доступа: https://developer.android.com/guide/topics/admin/device-admin (дата обращения: 06.09.2019).
7. Датчики [Электронный ресурс]. – Режим доступа: https://developer.android.com/guide/topics/sensors (дата обращения: 06.09.2019).
8. What's New in Xcode 11 [Электронный ресурс]. – Режим доступа: https://developer.apple.com/xcode. (дата обращения: 06.09.2019).
9. Swift 5 [Электронный ресурс]. – Режим доступа: https://developer.apple.com/swift. (дата обращения: 06.09.2019).
10. Carthage [Электронный ресурс]. – Режим доступа: https://github.com/Carthage/Carthage (дата обращения: 06.09.2019).
11. CocoaPods [Электронный ресурс]. – Режим доступа: https://cocoapods.org/ (дата обращения: 06.09.2019).
12. Alamofire [Электронный ресурс]. – Режим доступа: https://github.com/Alamofire/Alamofire (дата обращения: 06.09.2019).

АНАЛИЗ РЕШЕНИЙ В ОБЛАСТИ РАСПОЗНАВАНИЯ ЭМОЦИЙ ИЗ РЕЧИ

В.В. Видман
г. Томск, Томский политехнический университет
vvv23@tpu.ru

ANALYSIS OF DECISIONS IN THE FIELD OF RECOGNITION OF EMOTIONS FROM SPEECH

V.V. Vidman
Tomsk, Tomsk Polytechnik University

Abstract. In recent years, the recognition of emotions has become widespread. There are many works and directions in this area. This article discusses the widely used classifiers and the functions of extracting signs of emotions from speech. The dependence of the result on the emotions used.
Keywords: neural networks, speech recognition, emotion recognition, MFCC, ANN, SVM.

Введение. Роль эмоций в нашей жизни очень важна. Эмоции добавляются к нашей речи не принуждено, это дополнительный смысловой канал передачи данных, по которому передается отношение говорящего к текущей ситуации и к тому что он говорит. Эмоции передаются несколькими способами: это эмоциональный окрас речи, мимика и используемые слова.

Эмоции — это информация, а любая информация имеет ценность. Но в наше время человек всё меньше общается с человеком в сфере обслуживания. Интернет магазины, боты-автоответчики, различные сервисы. Но компьютер не человек, и он не может понимать эмоции покупателя, если его этому не научить. Внедрение распознавания эмоций во все сферы, где человека обслуживает не человек имеет большую ценность. Если это использовать в системе здравоохранения, то можно контролировать психическое и эмоциональное состояние здоровья пациента, благодаря этому контролировать процесс лечения [12]. Определять эмоции в этом случае можно разному, если это палата, то использовать видеокамеру. Если человек лечится дома, то с помощью приложения на мобильном устройстве записывать аудио отчеты. Подобная система внедрена в [3], где в тексте искались определенные слова и по ним определялись эмоции. В 2006 году один из южнокорейских операторов запустил

мобильный сервис анализа голоса, который основан на системе голосового анализа и действует как детектор эмоций, делая заключения об уровне честности участников разговора. В течение разговора анализируются различные звуки, которые попадают в микрофон абонента, и делается заключение об их эмоциональном статусе. В конце разговора абонент получает сообщение с графиком правдивости, где показан уровень стресса и число неточных ответов и попыток сменить тему. Происходит анализ, который учитывает, как определенная мозговая активность влияет на специфические особенности голоса. Это позволяет определить и измерить широкий спектр эмоций, используя различные оценки составляющих эмоций, строить оценку правдивости любого утверждения, сделанного участниками разговора [14].

Извлечение признаков. В этом разделе рассматриваются различные звуковые функции, используемые в Системе распознавания эмоций(SER), в том числе коэффициенты кодирования с линейным предсказанием (LPCC) и метод кепстральных коэффициентов на шкале мел (MFCC)Процесс извлечения проходит через три этапа. Во-первых, предварительное выделение – это фильтр, используемый для выделения высокочастотного бода путем увеличения его амплитуды и уменьшения амплитуды более низкой частоты. В речи, как правило, более высокая частота содержит более важную информацию для извлечения, в то время как более низкая частота может смешиваться с шумом. Следует отметить, что в современных системах распознавания речи предварительное выделение теряет свою важность и заменяется нормализацией канала на более поздних этапах, но для простых, но эффективных методов достаточно фильтра верхних частот. Во-вторых, блокирование кадров и управление окнами – это процесс разложения речевого сигнала на короткие речевые последовательности, называемые кадрами, для проведения анализа речи. Есть несколько окон, которые можно использовать, такие как прямоугольное окно, треугольное окно, но часто выбирается окно Хемминга, поскольку оно смягчает края, созданные из-за кадрирования, снова подчеркивая простоту. В-третьих, это функция извлечения. Речевые характеристики могут быть разделены на четыре группы, включая непрерывные, качественные, спектральные и основанные на TEO признаки [1]

Линейное прогнозирующее кодирование (LPC) – это цифровой метод кодирования аналогового сигнала [2]. Работа LPC заключается в том, что он прогнозирует следующее значение сигнала на основе информации, которую он получил в прошлом, формируя линейный паттерн. Основная цель LPC – получить набор коэффициентов предиктора, которые позволят минимизировать среднеквадратическую ошибку. LPC- кодирование обычно дает удовлетворительное качество речи при более низкой скорости передачи битов и обеспечивает точные аппроксимации параметров речи. Хотя LPCC можно считать одной из более традиционных особенностей речи, LPC способствует общему распознаванию эмоций. В одной из работ авторы использовали LPCC как одну из своих функций и достигли 86,41% распознавания [5]. Mel-частотные кепстральные коэффициенты (MFCC) – одна из самых популярных звуковых функций [6]. Это представление речевых сигналов, где особенность, называемая кепстром оконного кратковременного сигнала, извлекается из FFT (Быстрого преобразования Фурье) этого сигнала. После этого сигнал поступает на ось частоты мелкочастотной шкалы с использованием логарифмического преобразования, а затем декоррелируется с помощью модифицированного дискретного косинусного преобразования [7]. Шаги по извлечению функций MFCC, включая предварительное выделение, блокировку кадров и управление окнами, величину FFT, набор фильтров Мел, энергию журналов и DCT (дискретное косинусное преобразование), описаны в статье ‘Survey on speech emotion recognition: Features, classification schemes, and databases”. MFCC использует шкалу mel, которая настроена на частотную характеристику человеческого уха. В связи с этим было доказано, что MFCC неоценим в области распознавания речи, и была предпринята попытка интеграции с распознаванием эмоций. Спектральные аудио функции, такие как MFCC, лучше всего подходят для N-way классификаторов [1].

Обзор распространенных классификаторов. После того, как система SER извлечет нужные данные из аудиоречевых данных, следующим шагом будет передача данных в классификатор. Основная задача классификатора состоит в том, чтобы определить нераскрытые эмоции пользователя с помощью набора определенных алгоритмов и функций. Обычно эти оценки классификатора выполняются с использованием одной базы данных или набора данных на одном языке. До сих пор не было согласованного стандарта того, какой классификатор является лучшим, но многие были оценены для достижения лучшего признания. Наиболее часто используемые классификаторы: GMM, HMM, SVM, а также k-NN [1]. В этом разделе самые популярные классификаторы HMM, GMM и SVM. Они сравниваются с DNN, который является расширенной версией ANN.

Скрытые Марковские модели состоят из первого порядка Марковской сети, состояния которой скрыты от наблюдателя. Это означает, что наблюдатель не может исследовать внутреннее состояние сети, так как они скрыты. Скрытые состояния модели отражают временную структуру данных. Скрытые Марковские модели – это статистические модели, которые описывают последовательности событий. HMM имеет преимущество в том, что временная динамика речевых особенностей может быть настроена благодаря наличию переходной матрицы. Во время кластеризации берется речевой сигнал, и предоставляется вероятность для каждого кадра речевого сигнала. Выходной слой классификатора предоставляет максимальную вероятности того, что содержит данный согнал [9].

Модель гауссовых смесей (GMM) использует альтернативную генерирующую вероятностную модель, которая предполагает, что для конкретного слова можно образовать многомерные модели гауссовой плотности, которая представляет все кадры [13]. По сравнению с HMM, GMM превосходит в процессе обучения и тестировании благодаря его эффективности при моделировании мультимодальных распределений в целом. GMM используются в SER, когда глобальные функции являются основным фокусом. Но из-за этой особенности, GMM не подходит, когда пользователь хотел бы моделировать временную структуру.

Термин искусственная нейронная сеть (ANN) представляет собой термин, обычно используемый для системы, которая имитирует поток нейрона. Информация, полученная от входа, течет от одного узла к другому, пока он не достигнет выхода. Сеть прямого распространения нейронная сеть является первым типом разработанной нейронной сети. Процесс основан: на передача данных через входной слой на один скрытый слой, а затем в выходной слой. В сети прямого распространения (feedforward) нет петель или циклов. В этой нейронной сети, существует входной слой, скрытый слой и выходной слой. Глубокая нейронная сеть расширяет возможности путем добавления дополнительных слоев в сегменте скрытого слоя [11]. Интересная особенность DNNS, что они могут узнать инвариантные функции высокого уровня из исходных данных.

Важным примером общих дискриминантных классификаторов является метод опорных векторов [4]. SVM классификаторы в основном основаны на использовании функций ядра для нелинейного отображения характеристик в многомерном пространстве, где данные могут быть хорошо классифицироваться с использованием линейного классификатора. Классификаторы SVM широко используются во многих приложения для распознавания образов и показали превосходство над другими известными классификаторами [8]. У них есть некоторые преимущества перед GMM и HMM, включая глобальную оптимальность обучения алгоритма. На самом деле, нет систематического способа выбора функций ядра, а, следовательно, отделимость преобразованных признаков не гарантируется. Фактически, во многих приложениях распознавания образов, включая эмоции из речи, не рекомендуется иметь идеально разделенные тренировочные данные, чтобы избежать переобучения. Классификаторы SVM также широко используются для распознавания эмоций речи во многих исследованиях [8]. Данные работы основаны на одних и тех же данных, они сравнивают работы разных классификаторов при работе с одной базой эмоций. В работе классификаторы протестированы с использованием базы эмоций FERMUS III [10]. Они провели различные тестирования методов и получили следующие результаты: для классификации, не зависящей

от докладчика, точность составляет 76,12%, 75,45% и 81,29% для трех последовательных проверок. Для классификации, зависящей от докладчика, точность классификации составляет 92,95%, 88,7% и 90,95% для тех же трех проверок.

Заключение. Одной из основных проблем при обзоре существующих методов распознавания эмоций из речи является то, что большинство методов используют различные «Эмоциональные базы». Это затрудняет сравнение их эффективности, так как определенный метод может отлично работать с одним набором данных, но выдавать очень низкие показатели на других. При этом базы данных эмоций зачастую являются закрытыми и их невозможно скачать, а, следовательно, и проверить их работоспособность.

ЛИТЕРАТУРА

1. M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," Pattern Recognition, vol. 44, pp. 572-587, 2011.
2. A. Dixit, A. Vidwans, and P. Sharma, "Improved MFCC and LPC algorithm for bundelkhandi isolated digit speech recognition," in 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 3755-3759, 2016.
3. F. Doctor, C. Karyotis , R. Iqbal , A. James , An intelligent framework for emotion aware e-healthcare support systems, in: Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Athens, 2016, pp. 1–8. 2016.
4. R. Duda, P. Hart, D. Stork, Pattern Recognition, John Wiley and Sons, 2001.
5. W. Fei, X. Ye, S. Zhaoyu, H. Yujia, Xing, and S. Shengxing, "Research on speech emotion recognition based on deep auto-encoder," in 2016 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), pp. 308-312, 2016.
6. T. S. Gunawan, N. A. M. Saleh, and M. Kartiwi, "Development of Quranic Reciter Identification System using MFCC and GMM Classifier," International Journal of Electrical and Computer Engineering (IJECE), vol. 8, 2018.
7. X. Huang, A. Acero, and H.-W. Hon, Spoken language processing: A guide to theory, algorithm, and system development, Prentice hall PTR, 2001.
8. C. Lee, S. Narayanan, R. Pieraccini, Classifying emotions in human–machine spoken dialogs, in: Proceedings of the ICME'02, vol. 1, 2002, pp. 737–740.
9. B. Schuller, G. Rigoll, M. Lang, "Hidden Markovmodel-based speech emotion recognition", Proceedings of the IEEE ICASSP Conference on Acoustics,Speech and Signal Processing, vol.2, pp. 1-4, April 2003.
10. B. Schuller, Towards intuitive speech interaction by the integration of emotional aspects, in: 2002 IEEE International Conference on Systems, Man and Cybernetics, vol. 6, 2002, p.
11. D. Yu and L. Deng, Automatic speech recognition: A deep learning approach, Springer, 2014.
12. Л. Рон Хаббард. «Свободный человек». Журнал «Способность». № 232.
13. Алгоритм использования гауссовых смесей для идентификации диктора по голосу в технических системах [Электронный ресурс].– URL:http://elib.bsu.by/bitstream/123456789/28611/1/92- 96%234.pdf (дата обращения 15.04.2019)
14. Полиграф – [Электронный ресурс]–URL:http://www.ukrpolygraph.org/2006/09/28/90 (дата обращения 19.05.2019)