биение веб-приложения на компоненты позволяет сократить дублирование кода между модулями, что дополнительно уменьшает и скомпилированный объём.

Итогом исследования является программный комплекс, который заранее предугадывает возможные пути посещения веб-приложения и заранее осуществляет предварительную загрузку модулей и компонентов. Тем самым, при переходе на конкретный «экран» приложения, у пользователя будет отсутствовать время ожидания, так как данные были загружены заранее. А совместно с не самой сложной в поддержке и отладке моделью для принятия предиктивных решений, мы получим и кодовую базу с умеренным порогом вхождения.

Литература

1 Banu Deniz Gunel, ”Investigating the Effect of Duration, Page Size and Frequency on Next Page Recommendation with Page Rank Algorithm”, [Текст], 2010

2 Borges, Levene, “Evaluating Variable-Length Markov Chain Models for Analysis of User Web Navigation Sessions”, [Текст], 2007

3 https://en.wikipedia.org/wiki/Predictive_analytics [Электронный ресурс]

4 https://en.wikipedia.org/wiki/Hidden_Markov_model [Электронный ресурс]

Н.Д. Беляев, И.А. Лёзин

ИЗВЛЕЧЕНИЕ ПРИЗНАКОВ АУДИОСИГНАЛА ПРИ РЕШЕНИИ ЗАДАЧИ РАСПОЗНАВАНИЯ ЭМОЦИЙ В РЕЧИ ДИКТОРА

(Самарский университет)

Введение

Речь – это основное средство передачи информации. Но не только слова содержат информацию. Мы также можем узнать много о ситуации, событии и т.д. из эмоций человека. Поэтому распознавание эмоций в голосе и мимике человека стало занимать одно из первых мест в различных отраслях сферы информационных технологий.

Акустические сигналы являются наиболее часто используемыми данными после мимических признаков для определения эмоционального состояния человека.

В этой статье мы рассмотрим использование относительно новой характеристики Harmonic To Noise Rate (HNR) в нашей системе распознавания эмоций с комбинацией других характеристик, таких как коэффициенты MFCC, ZCR и TEO. Последние 3 характеристики наиболее популярны в сфере распознавания аудиосигналов и, как правило, показывают хорошие результаты при решении любой связанной задачи.

Алгоритмы извлечения признаков

1. MFCC

Чтобы вычислить коэффициенты MFCC, обратное быстрое преобразование Фурье применяется к логарифму модуля быстрого преобразования Фурье сигнала, отфильтрованного по шкале Мела [15].

Исходный речевой сигнал запишем в дискретном виде как:

$$x[n], \quad 0 \leq n < N \tag{1}$$

Применяем к нему преобразование Фурье:

$$X_a[k] = \sum_{n=0}^{N-1} x[n] e^{\frac{-2\pi i}{N} kn}, \quad 0 \leq k < N \tag{2}$$

Составляем гребенку фильтров, используя оконную функцию:

$$H_m = \begin{cases} 0 & k < f[m-1] \\ \frac{(k-f[m-1])}{(f[m]-f[m-1])} & f[m-1] \leq k < f[m] \\ \frac{(f[m+1]-k)}{(f[m+1]-f[m])} & f[m] \leq k \leq f[m+1] \\ 0 & k > f[m+1] \end{cases} \tag{3}$$

Для которой частоты f[m] получаем изравенства:

$$f[m] = (\frac{N}{F_s}) B^{-1}(B(f_1) + m\frac{B(f_h) - B(f_1)}{M+1}) \tag{4}$$

B(b) — преобразование значения частоты в мел-шкалу, соответственно:

$$S[m] = ln(\sum_{k=0}^{N-1} |X_a[k]|^2 H_m[k]), \quad 0 \leq m < M \tag{5}$$

Вычисляем энергию для каждого окна:

$$B^{-1}(b) = 700(exp(b/1125) - 1) \tag{6}$$

Применяем ДКП:

$$c[n] = \sum_{m=0}^{M-1} S[m] cos(\pi n(m + 1/2)/M), \quad 0 \leq n < M \tag{7}$$

Получаем набор MFCC.

2. ZCR

ZCR - это скорость, с которой сигнал изменяется с положительного значения до нуля и до отрицательного значения или наоборот. [1] Его значение широко использовалось как для распознавания речи, так и для поиска музыкальной информации, являясь ключевой функцией для классификации ударных звуков [2].

ZCR формально определяется как:

$$zcr = \frac{1}{T-1} \sum_{t=1}^{T-1} 1_{\mathbb{R}_{<0}}(s_t s_{t-1}) \tag{8}$$

где – s сигнал длины T, а $1_{R<0}$ - индикаторная функция.

3.TEO

Функции Teager Energy Operator (TEO) проверяют характеристики речи, когда высказывание представляет собой определенного вида стресс. Функции TEO обрабатывают поведение сигнала по частоте и во временной области.

Для оценки TEO каждый выходной сигнал M-сигнала сегментируется на кадры одинаковой длины (например, 25 миллисекунды со смещением кадра 10 миллисекунд); где M - количество критических зон, а f – количество кадров, для которого извлекается TEO. В нашей работе мы извлекаем TEO из суммы сигналов по следующей формуле:

$$\Psi M\,[xf\,[t]] = (xf\,[t])^2 - (xf\,[t-1]\,xf\,[t+1])) \quad (9)$$

4. Harmonic To Noise Rate

По определению HNR - это параметр, в котором взаимосвязь между гармоническими и шумовыми компонентами обеспечивает указание вокализованных компонентов речевого сигнала путем количественной оценки взаимосвязи между периодической и апериодической составляющими, выраженной в дБ [1] [3] [4]. Общее значение HNR сигнала варьируется, потому что разные конфигурации речевого тракта подразумевают разные амплитуды гармоник. Эта единица измерения связывает энергию, передаваемую голосовым сигналом через голосовые импульсы, и энергию фракции голосового шума после фильтрации через голосовой тракт. Этот шум возникает из-за турбулентности, возникающей при прохождении воздушного потока через голосовую щель во время фонации, возникающей, например, когда голосовые связки смыкаются ненадлежащим образом [4].

Существуют различные подходы к автоматическому определению HNR, например использовали кепстр для измерения гармонических и шумовых составляющих, в то время как [4] использовали автокорреляцию. С математической точки зрения звонкий сигнал с гармонической структурой в частотной области может быть выражен уравнением 10.

$$X(w) = H(w) + N(w) \quad (10)$$

Где X (w) соответствует речевому сигналу в частотной области, H (w) - гармонической составляющей, а N (w) - шумовой составляющей.

HNR - это логарифмическая мера отношения энергии, которое связано с гармонической и шумовой составляющими. С помощью уравнения 11 можно интегрировать спектральную мощность по слышимому диапазону частот.

$$HNR = 10 \times \log_{10} \frac{\int_w |H(w)|^2}{\int_w |N(w)|^2} \quad (11)$$

В этом алгоритме HNR был реализован с учетом исследований, опубликованных Boersma [4]. В этом исследовании Боерсма использует процедуру, основанную на свойствах автокорреляционной функции, чтобы получить разделение компонентов, описанное ранее. Автокорреляция состоит из корреляции сигнала с самим собой. Если мы рассмотрим голосовой сигнал x (t), функция автокорреляции rx (τ) представлена в уравнении 12.

$$r_x(\tau) \equiv \int x(t)x(t+\tau)dt \,. \quad (12)$$

В этой функции есть глобальный максимум при τ = 0. Если есть пиковые значения вне 0, сигнал периодический и есть фазовый сдвиг T0, называемый периодом, так что все эти максимумы помещаются в смещение nT0 для каждого целого числа n, с rx (nT0) = rx (0) . Основная частота F0 этого периодического сигнала определяется соотношением F0 = 1 / T0. Если нет глобальных максимумов, кроме 0, могут быть максимальные пики. Если наибольший из них находится в смещении τmax, и если его высота rx (τmax) достаточна, сигнал обозначается как имеющий периодическую часть, а его гармоническая сила R0 - это число от 0 до 1, равное локальному максимуму r'x (τmax) нормированной автокорреляции (уравнение 13).

$$r'_x(\tau) \equiv \frac{r_x(\tau)}{r_x(0)} \quad (13)$$

Полная автокорреляция сигнала представляет собой сумму автокорреляции его гармонической и шумовой составляющих, как можно увидеть в уравнении 14.

$$r_x(0) = r_H(0) + r_N(0) \quad (14)$$

Если шум белый (корреляция невозможна), локальный максимум равен τmax = T0 с высотой rx (τmax) = rH (T0) = rH (0) [4]. При этом автокорреляционная функция устойчивого речевого сигнала отображает локальные максимумы для нескольких значений τ, которые все кратны основному периоду. Таким образом, для определения HNR необходимо только вычислить автокорреляционную функцию речевого сигнала и идентифицировать первый локальный максимум, который будет соответствовать гармонической составляющей. Величина шумовой составляющей определяется уравнением 6, а HNR - уравнением 15.

$$HNR(dB) = 10 \times \log_{10} \frac{r'_x(\tau_{\max})}{1 - r'_x(\tau_{\max})} \quad (15)$$

Литература

1 А.Мерибан Communication without words[Текст]/А. Мерибан. – 1968. – С. 53-56.

2 Б. Фасель, Д. Лютин, Automatic Facial Expression Analysis: A Survey, Pattern Recognition[Текст]/ Б. Фасель, Д. Лютин. – 2003. - С. 259-275.

3 К. Рао,Т. Кумар, К. Ануша, Emotion Recognition from Speech – 2012. - (https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.437.2775&rep=rep1&type=pdf)

4 И. Саксмит., С. Алиссон, С. Барон-Коэн, Empathy and emotion recognition in people with autism, first-degree relatives, and controls[Текст]/ И. Саксмит., С. Алиссон, С. Барон-Коэн – 2013. – С. 98-105.

А.Д. Божимов, О.П. Солдатова

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ СЖАТИЯ МОДЕЛИ НЕЙРОННОЙ СЕТИ ДЛЯ ПЕРЕДАЧИ ПРОИЗВОЛЬНОГО СТИЛЯ ИЗОБРАЖЕНИЯ

(Самарский университет)

Современные реалии всё чаще подталкивают нас к использованию нейронных сетей при решении обширного диапазона задач. Одним из направлений, где реализация алгоритма средствами классического программирования является труднореализуемой или и вовсе невыполнимой, является художественное творчество. Развитие технологий машинного обучения позволило появиться успешным попыткам воссоздания уникальных особенностей художественных произведений техническими средствами. Это обусловлено в первую очередь способностью моделей к «обучению» - процессу, где нейронная сеть должна выявлять сложные зависимости между входными и выходными данными [1].

Большинство существующих алгоритмов переноса стиля изображения требуют отдельного обучения сети и создания новой модели для работы с каждым новым стилем. Это накладывает серьезные ограничения на используемое аппаратное обеспечение, а также увеличивает затраты времени [2]. Оптимальным решением было бы иметь модель, которая способна выполнять быструю передачу стиля на любых парах изображения, не требуя переобучения.

Рассматриваемый процесс переноса произвольного стиля изображения состоит из двух этапов: получение вектора стиля из одного изображения и стилизация другого на основе векторизованных данных.

Для выделения стиля на первом этапе используется сеть предсказания, создающая на выходе 100-мерный вектор. Такой подход позволяет также комбинировать стили, используя средневзвешенное значение [3].

На втором этапе стилизации применяется сеть переноса стиля, принимающая чистое изображение и вектор представления стиля и на выходе создаю-