

РАСПОЗНАВАНИЕ ЭМОЦИЙ ИЗ РЕЧЕВОГО СИГНАЛА

В.В. Видман, А.Я. Видман, В.М. Саклаков, И.С. Ванюшин
Научный руководитель – к.т.н. С. В. Аксенов
Томский политехнический университет,
vvv23@tpu.ru

Введение

Речь является одним из основных и естественных способов общения между людьми [1]. Речь становится более выразительной и эффективной благодаря эмоциям. Они выражаются разными способами, такими как смех, крик, плач и другие эмоции, используемые людьми для выражения своих чувств. Обнаружение эмоций иногда даже для людей представляет трудности, а для машин это и вовсе непростая задача. Из этого следует, что существует потребность в системах распознавания эмоций, которые помогут сделать взаимодействие человека с компьютером более легким.

Распознавание речевых эмоций – это извлечение эмоционального состояния, говорящего из его речи. Основная цель распознавания эмоций – сделать человеко-машинный интерфейс более удобным. Применение этой технологии можно найти и в других сферах, например, в маркетинге, для определения отношения человека к товару. Также автоматические системы распознавания речевых эмоций можно применить в детекторах лжи, в сфере развлечения или диагностике психического здоровья, а также в Call-центрах [2]. В настоящее время уже предложены различные системы для распознавания в этих областях. Исследователи используют различные методы и классификаторы для определения эмоций. А именно: скрытая Марковская модель (HMM), Гауссовская модель смещения (GMM), метод опорных векторов (SVM), искусственная нейронная сеть (ANN). Исследование с использованием характеристик тональности и энергии и классификатора HMM было выполнено в исследовании Björn Schuller-a [3], где достигнутая точность составила 86%.

База данных

Базы данных играют важную роль для автоматического распознавания эмоций, поскольку статистические методы изучаются на примерах. Базы данных, используемые для обучения содержат как специально подготовленные, так и реальные примеры проявления эмоций. По мере увеличения естественности базы данных сложность также возрастает. Таким образом, в начале исследования проводились по автоматическому распознаванию вокальных эмоций, эти исследования начались в середине 90-х годов, Наполнение базы начиналось с актерской речи и теперь смещается в сторону более реалистичных данных.

Наиболее популярными примерами базы данных являются Берлинская база данных эмоциональной речи [4], в которую вошли 5 мужчин и 5 женщин-актрис, а также датская эмоциональная

речевая база (DES). Русская база данных состоит из десяти произносимых предложений от 61 говорящего (12 мужчин 49 женщин) возрастной группы от 16 до 28 лет, выражающих шесть эмоций: счастье, грусть, сердитость, страх, нейтральные эмоции и отвращение. Два примера баз данных на реальной речи это первая немецкая Aibo база эмоций людей [4] которые не знали, что их эмоции записываются. Вторая база создана call-центром Devillers и др. [5]. Она также получена из живых записей и является примером реальной эмоциональной базы данных.

Система распознавания речевых эмоций

Для систем распознавания речевых эмоций в качестве входных данных берутся образцы речи. Первое, что нужно сделать с образцами – убрать шум. Далее извлекаются нужные признаки. Извлеченные признаки передаются классификатору, который определяет эмоции и выводит их значение.

Распознавание речевых эмоций состоит из нескольких шагов, показанных на рисунке 1.



Рис. 1. Система распознавания речевых эмоций

Предобработка

Предварительная обработка необходима, когда стандартная база данных не используется. Целью предварительной обработки является повышение высоких частот сигнала и получение плоского частотного спектра сигналов и частотных характеристик. С помощью оконной функции Хэмминга и прямоугольное окна мы получаем речевые фреймы.

Извлечение и выбор компонентов

Выделение и выбор признаков – важный шаг в системе распознавания эмоций. После извлечения эмоций нужно решить, какие функции должны быть выбраны соответствующим образом.

Главным образом выделяются такие характеристики как тональность, изменение тональности,

скорость произношения и другие спектральные характеристики. Как правило все исследователи используют одни и те же характеристики, но в разном сочетании.

Классификатор

После извлечения признаков и выбора основных, следующим шагом является выбор подходящего классификатора. Классификатор также вносит свой вклад в точность распознавания эмоций. Существует ряд классификаторов, самые популярные это HMM, GMM, ANN, SVM и т. д. Сочетая классификаторы можно получить гибридную модель. Каждый из классификаторов имеет некоторые плюсы и минусы по сравнению с другим. Ниже приведены некоторые из наиболее часто используемых классификаторов:

1. Скрытая Марковская модель (Hidden Markow model) широко используется в литературе, но для её использования требуются большие вычислительные мощности. Точность вычислений эмоций в ассоциативной осцилляторной среде на примере распознавания русских слов достигла 79-90%. Данная точность распознавания была получена при использовании модели для определения низкоуровневых характеристик: тональности и энергии речевого сигнала [6].

2. GMM (Generalized Method of Moments) весьма эффективны при моделировании мульти модальных распределений, потому что требования к обучению и тестированию для GMM являются менее строгими. GMM подходят, когда доступно большое количество векторов признаков. GMM работает эффективно для спектральных характеристик. GMM относятся к числу хорошо проработанных методов функции плотности вероятности и кластеризации [7].

3. Искусственная нейронная сеть (ANN) также широко используется для распознавания эмоций в речи. Используется трехслойная (два скрытых и один выходной) нейронная сеть. Скрытые слои имели по 10 узлов каждый. Общий показатель распознавания около 85%. Наиболее узнаваемой эмоцией был гнев с точностью 98%, тогда как страх был наименее узнаваем с точностью 81% [8].

4. Другой популярный метод – метод опорных векторов (SVM). SVM создает гиперплоскость в высоком или бесконечном пространстве для классификации. Расстояние до ближайших тренировочных данных лучше, чем расстояние до гиперплоскости. Общая точность 94,2% достигается с помощью изолированного SVM [9].

Заключение

В данной работе отображен краткий обзор на тему распознавания речевых эмоций. Для реализа-

ции алгоритма по распознаванию эмоций возможно использование различных методов и классификаторов, а также разных баз данных. Все методы показывают различный уровень точности распознавания эмоций, но, по мнению автора, важным фактором является то, какие данные нужно обрабатывать. Скорость и точность распознавания зависит как от входных данных, которые нужно распознать, так и от базы данных, используемой для обучения классификатора.

Список использованных источников

1. Засецкова Е.Н. Речевые эмоции: реклама и правила риторики // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. – 2011.– № 1. – С. 79-80
2. ДИТ Аналитика [Электронный ресурс] / Применение передовых технологий в работе контакт-центра. – URL: https://moscowanalytics.ru/index/research/call_center (дата обращения 12.11.2018)
3. Schuller B., Rigoll G., Lang M., Hidden Markov Model Based Speech Emotion Recognition, – 2003 – IEEE ICASSP, 1-3 pp.
4. Burkhardt, F., Paeschke, A., Rolfes, M., Sendmeier, W.F., Weiss, B.: A database of German emotional speech, In: Proceedings of Interspeech 2005, Lisbon, Portugal.
5. Devillers, L., Vidrascu, L., Lamel, L.: Challenges in real-life emotion annotation and machine learning based detection – 2005– Neural Networks, 18(4), 407–422 pp.
6. Огнев И.В., Парамонов П.А. Распознавание речи методами скрытых Марковских моделей в ассоциативной осцилляторной среде // Известия высших учебных заведений. Поволжский регион. – 2013. – № 3(27). – С. 115-126
7. Томашенко Н.А., Хохлов Ю.Ю., Ларшер Э., Эстев Я., Матвеев Ю.Н. Использование в системах автоматического распознавания речи GMM-моделей для адаптации акустических моделей, построенных на основе искусственных нейронных сетей // Научно-технический вестник информационных технологий, механики и оптики. – 2016. –Т. 16– № 6– С. 1063-1072
8. Киселев В.В Автоматическое определение эмоций по речи // Образовательные технологии. – 2012. – № 3. – С. 85-89
9. Aastha Joshi Speech Emotion Recognition Using Combined Features of HMM & SVM Algorithm / International Journal of Advanced Research in Computer Science and Software Engineering, – 2016.–vol. 3– № 8– 387-393 pp.