

МЕТОД ВЫДЕЛЕНИЯ ИНФОРМАЦИИ ИЗ ХАРАКТЕРИСТИК АУДИОСИГНАЛА

Езовит А. В., Гуринович А. Б.

Кафедра информационных технологий автоматизированных систем, Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: ezovit99@yandex.ru, gurinovich@bsuir.by

В настоящее время важной задачей в области речевой обработки является выделение информационных признаков сигнала. Достижения в данной сфере могут применяться для идентификации личности, её описания, сбора информации о физическом состоянии и роде деятельности. В данной статье предложен метод для выделения смысловой информации из результатов обработки аудиосигнала с применением алгоритмов машинного обучения.

ВВЕДЕНИЕ

Важной задачей в области речевой обработки является выделение информационных признаков сигнала. Информационными признаками являются характеристики голоса и речи, которые уникальны для каждого человека. Их выделение из сигнала позволит составить «портрет» речи человека.

Другой задачей в области обработки речевой информации является сбор характеристик аудиосигнала. Характеристики аудиосигнала являются результатами анализа аудиоданных. К ним можно отнести мел-кепстаральные коэффициенты (MFCC), спектральный центроид (spectral centroid), спектральный спад (spectral rolloff) и другие. Данные характеристики применяются в областях классификации жанра музыки, отделения естественного голоса от синтезированного, идентификации человека и др.

В данной статье рассматривается подход к выделению информационных признаков из характеристик аудиосигнала. Концепция данного метода, заключается в идее того, что каждый информационный признак может быть соотнесен с одним или несколькими характеристиками аудиосигнала.

Применение данного метода поможет определить в каких характеристиках заключена та или иная информация, что в дальнейшем позволит сократить количество входных данных для алгоритмов машинного и глубокого обучения, таким образом экономя вычислительный ресурс и время, снижая величину тренировочного набора данных и улучшая точность.

I. ХАРАКТЕРИСТИКИ АУДИОСИГНАЛА

Характеристики аудиосигнала получаются в результате обработки записи звука во временной и частотной областях и в общем случае могут описывать любой сигнал. Большая часть характеристик, позволяет проанализировать сигнал по распределению энергий на спектрограмме.

К основным характеристикам можно отнести:

- Спектральный центроид (показывает на какой частоте сосредоточен "центр масс звука")
- Мел-кепстаральные коэффициенты (показывают энергию каждого треугольного фильтра мел-шкалы)
- Спектральный контраст (позволяет оценить разность энергии для поддиапазонов в кадре спектрограммы)
- Спектральный спад (мера формы сигнала, для какой частоты суммарная энергия более низких частот достигает заданного уровня)

II. ПОСТАНОВКА ЗАДАЧИ

Для выделения информационных признаков в речевом сигнале могут применяться методы, основанные на машинном и глубоком обучении.

Алгоритмы машинного обучения можно разделить на две категории: обучение с учителем и без. В алгоритмах обучения с учителем предьявляется набор данных, содержащий признаки, в котором каждый пример снабжен меткой, или целевым классом. В задаче обработки речи в качестве метки может выступать пол, возраст, эмоциональное состояние и т.д. Примером данного алгоритма служит метод опорных векторов, выполняющий задачу классификации. Алгоритму обучения без учителя в качестве опыта предьявляется набор данных, содержащий много признаков, а алгоритм должен выявить полезные структурные свойства этого набора. Некоторые алгоритмы обучения без учителя решают другие задачи, например алгоритм кластеризации, который должен выделить в наборе данных кластеры похожих примеров. Примерами данных алгоритмов являются метод главных компонент и кластеризация методом k средних.

Недостатком методов, основанных на машинном и глубоком обучении, является то, что возможных конфигураций характеристик гораздо больше, чем обучающих примеров, что приводит к тому, что для обучения требуется огромное количество примеров, времени и вычислительных мощностей.

В данной статье предлагается метод последовательного применения алгоритмов машинного обучения с учителем и без (метод главных компонент и метод кластеризации) для определения какие характеристики связаны с какими признаками и их выделения, что позволит в дальнейшем упростить процесс выделения информативных признаков речевого сигнала.

III. АЛГОРИТМ ВЫДЕЛЕНИЯ ПРИЗНАКОВ ИЗ ХАРАКТЕРИСТИК

Алгоритм сводится к трем этапам последовательного применения алгоритмов машинного обучения. Входными данными для алгоритма служит опыт E , представленный в виде массива, состоящего из векторов характеристик и меток для определенного количества аудиоданных. Поскольку часть характеристик представлена в виде одномерных или двумерных массивов данных, необходимо их обработать: например, найти среднее значение и стандартное отклонение, после этого провести нормализацию.

На первом этапе алгоритма формируется задача T , для обучения с учителем. Например: определить пол, возраст или эмоциональное состояние говорящего. После этого формируется опыт E – массив входных значений, включающий в себя все характеристики, предположительно способные влиять на меру качества R . Затем для решения этой задачи, исходя из целевого класса выбирается метод машинного обучения. Например, метод опорных векторов для задачи бинарной классификации (выбор пола) или линейная регрессия для задачи определения пола. Применяя метод, проводится оценка качества R , при удовлетворительном качестве, этап повторяется еще раз, но количество входных характеристик уменьшается с целью определить влияние отброшенной характеристики на предсказание, в случае неудовлетворительного R после применения алгоритма, стоит попробовать другой метод обучения. Выходными данными для первого этапа должна стать матрица A , состоящая из характеристик размерности меньшей входной, элементы которой оказывают наибольшее влияние на точность предсказания.

На втором этапе к матрице A применяется метод главных компонент (РСА). Задачей является сжатие данных в измерение меньшей раз-

мерности, чем входное. В ходе работы РСА оценивает дисперсию каждой величины и отбрасывает характеристики с наименьшим вкладом в формирование матрицы A , формируя новый вектор или гиперплоскость, на которую проецируются элементы исходной матрицы. Результатом работы является матрица C , размерности меньшей A .

На третьей этапе к полученному представлению, применяется метод кластеризации. Применение метода кластеризации позволяет найти точку в пространстве в окрестностях, которой будут сосредоточены объекты с идентичными свойствами. В зависимости от количества выбранных кластеров в результате можно получить распределение характеристик на диапазоны, несущие информацию об исследуемом признаке. Например: известно, что основной тон в зависимости от частоты (80 –210 Гц у мужчины, 120 –320 Гц у женщины) позволяет определить гендер говорящего. Таким образом можно построить два кластера: мужчины с центром в 145 Гц и женщины с центром в 220 Гц, для избегания пересечения стоит рассматривать основной тон совместно с другими характеристиками.

IV. ЗАКЛЮЧЕНИЕ

Предложенный алгоритм позволит определить какие характеристики аудиосигнала описывают определенный признак. Многократное применение поможет найти оптимальную входную матрицу опыта для решения конкретной задачи. Применение достигнутых результатов позволит сократить вычислительные мощности и величину набора данных для обучения алгоритмов машинного и глубокого обучения.

Так же подход, предложенный в статье, может быть применен в других областях в целях выделения признаков, оказывающих наибольшее влияние на точность предсказания.

V. СПИСОК ЛИТЕРАТУРЫ

1. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / пер. с англ. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.: цв. ил.
2. habr [Электронный ресурс]. – Москва, 2022 – Режим доступа: <https://habr.com/ru/post/668518/> – Дата доступа: 20.10.2022.
3. habr [Электронный ресурс]. – Москва, 2022 – Режим доступа: <https://habr.com/ru/post/670676/> – Дата доступа: 20.10.2022.