



АЛГОРИТМЫ РАСПОЗНОВАНИЯ РЕЧИ И КЛАССИФИКАЦИЯ МЕТОДОВ РАСПОЗНОВАНИЯ РЕЧИ

Юсупов Мирсаид Абдулазиз угли

магистрант 2 курса направления «Прикладная математика» Ферганского государственного университета

Онаркулов Максад Каримбердиевич

преподаватель кафедры «Информационные технологии» Ферганского государственного университета

Ёкуужонов Алижон Махамедин угли

магистрант 2 курса направления «Прикладная математика» Ферганского государственного университета

Abstract

В данной статье рассмотрены алгоритмы распознавания речи и приведена классификация методов распознавания речи.

ARTICLE INFO

Article history:

Received 6 Aug 2022

Revised form 7 Sep 2022

Accepted 10 Oct 2022

Key words: *алгоритмы распознавания речи, методы распознавания речи, классификация методов распознавания речи.*

© 2019 Hosting by Central Asian Studies. All rights reserved.

-----***-----

В современном обществе задача распознавания речи занимает одно из лидирующих мест в системах распознавания. Системы, которые существуют на сегодняшний день еще далеки от совершенства и имеют ряд недостатков, такие как высокий процент ошибок, настройки на конкретного диктора, ограниченный объем словаря – и это далеко не полный список проблем, которые предстоит решить. Всем известно, что речь человека характеризуется высокой степенью изменчивости, обусловленное несколькими причинами. Даже для одного и того же говорящего, при реализации одних и тех же акустических единиц будут отличаться длительностью произношения, по спектральному составу. Любые изменения в речи связано с изменениями эмоционального состояния человека, и естественно условий, в которых он находится. Наличие артикуляционных эффектов приводит к тому, что произношение слов, фонем сильно зависит от их контекста. К изменениям в речевом сигнале, разумеется, приводят помехи различного характера среды, в которой находится говорящий. Всевозможные речевые сбои, такие как заполненные паузы, удлиняют высказывания. Также они вызывают различного рода ошибки вследствие того, что системы распознавания обучаются на структурированных предложениях без речевых сбоев, что приводит к формированию ошибочных транскрипций. Учитывая все эти факторы, и принимая во внимание ряд других ограничений, делаем вывод - для высококачественного распознавания речи в реальном времени требуются вычислительные средства с высоким быстродействием. Одним из способов снижения этого требования является распараллеливание вычислений, которое естественным образом достигается при использовании искусственных нейронных сетей, реализованных на нейрокомпьютерах.

Понятие «Автоматическое распознавание речи» (в английской терминологии, automatic speech recognition – ASR) – это преобразование звука в текст. Имеется несколько категорий систем распознавания речи, которые применяются различных сферах:

1. Распознавание отдельных голосовых команд, которое применяется в коммерческих приложениях;
2. Поиск ключевых слов в потоке речи (поисковые системы);
3. Автоматическое распознавание слитной речи на большом словаре (автоматическая расшифровка записей – создание стенограмм).

Системы распознавания речи различают: дикторо-зависимые и дикторо-независимые системы, характеризуются степенью зависимости от настройки на речь конкретного диктора.

Методы распознавания речи

Основой создания систем распознавания речи считаются методы, которые разделяют и описывают классы образов. В данном контексте "образ" — это описание элемента как представителя соответствующего класса образов. Класс можно охарактеризовать списком входящих в него членов, система распознавания может базироваться на принципе принадлежности к этому списку. А, если класс описывается некими общими свойствами, свойственными всем его членам, то система базируется на принципе общности свойств. Но, если при анализе класса обнаруживается тенденция к формированию кластеров (конечную группу объектов, которые образуют компактную область в пространстве описания) в пространстве образов, система может основываться на принципе кластеризации.

Принцип перечисления членов класса, находящихся в его составе, предполагает выполнение автоматического распознавания речи на основе сопоставления с образцом. В память системы распознавания помещается множество образов, принадлежащих одному классу. При появлении в системе новых образов, система поочередно сопоставляет их с образами, которые уже хранятся в ее памяти. Если в памяти системы распознавания, звуки разных букв, тогда данный подход, основанный на перечислении членов класса, позволяет распознать соответствующие буквы, но лишь в тех случаях, когда входные данные не искажены посторонним шумом, произношением и т.п. Следовательно, это относительно несложный метод, но дает возможность синтезировать системы распознавания речи лишь в отдельных прикладных сферах, которые вполне справляются со своими задачами. Метод перечисления членов класса работает приемлемо, если выборка образов близка к эталонной.

Рассмотрим архитектуру современных систем распознавания речи. Входные данные - сам сигнал в формате .wav или голос с микрофона. Процесс распознавания речи включает в себя такие этапы, как:

- 1) предварительная обработка сигнала;
- 2) преобразование сигнала в секторы особенностей;
- 3) распознавание речевой части (классификация).

Если первая задача инвариантна, и имеет достаточно много решений этой задачи, вторая и третья являются специфическими, где методы их решения рассмотрены ниже. Необходима предварительная обработка речевых сигналов, перед тем как предпринимать попытки распознавания речи. Большинство современных систем автоматизированного распознавания используют модульную архитектуру. Модуль предварительной обработки включает в себя:

- блок шумоочистки, где повышается качество сигнала;
- блок детектора голоса, где выделяются участки, содержащие речь.

В следующем модуле выделяются информативные признаки речевого сигнала – состоит преобразование сигнала в векторы особенностей, то есть участки, содержащие речь, превращаются в наборы коэффициентов, поступающие в блок распознавания:

- 1) с помощью набора программных полосовых фильтров получается спектр частот речевого сигнала;
- 2) преобразования полученного спектра речевого сигнала:
 - а) логарифмическое изменение масштаба в пространстве амплитуд и частот;
 - б) сглаживание спектра с целью выделения его огибающей;
 - в) кепстральный анализ, т.е. обратное преобразование Фурье от логарифма прямого преобразования.

На выходе главного модуля мы получаем информацию о наличии команды или ее отсутствии, само распознанное слово, или это правильно или неправильно произнесенное слово. В самом блоке распознавания могут использоваться существующие методы распознавания речи:

- 1) временные динамические алгоритмы, динамическое программирование, называемые методом распознавания речи на основе сравнения с эталоном;
- 2) методы дискриминантного анализа, основанные на Байесовской дискриминации, то есть методы распознавания на основе контекстно-зависимой классификации, Скрытые Марковские модели, Нейронные сети.

Реализация процесса распознавания путем выделения схожих признаков происходит разделением образов по классам, общих для членов, входящих в его состав. Основным в этом методе является то, что образы, имеющие ряд общих свойств или признаков, принадлежат одному и тому же классу. В память системы можно ввести эти совместные свойства. Если системе во входных данных предоставляется не классифицированный образ, то в таких случаях происходит выделение набора, описывающих признаки, а затем сравнение с признаками, находящимся в памяти системы. Следующим шагом система зачислит проверяемый образ в класс, который характеризуется системой признаков, подобных признакам этого образа. При использовании вышеописанного метода, основной задачей является выделение общих признаков по выборке образов, принадлежность которых к искомому классу уже известна. Естественно описанная концепция распознавания превосходит распознавание по памяти, по сравнению с хранением объектов, входящих в класс, так-как, свойства, описывающие класс в целом, являются константами, и принцип сопоставления свойств разрешает различие свойств отдельных образов. Но с другой стороны, метод основанный на сравнении с эталоном, не позволяет существенных вариаций характеристик отдельных образов. Как упоминалось выше, очень трудно найти полный набор признаков, различий для класса. Следовательно, использование данного принципа распознавания, тесно связано с развитием методов выборки оптимальных свойств.

Рассмотрим метод чаще используемый для второго и третьего этапа метод - метод MFCC (метод моделей заполнителей, метод скользящего окна), где в конце третьего этапа применяется простой алгоритм kNN классификации (k-ближайших соседей). MFCC обычно получают следующим образом:

- а) получение преобразования Фурье взвешенного сигнала;
- б) получение с использованием треугольных перекрытых окон на Мела-шкале и отображение спектра мощности;
- в) логарифмирование квадрата каждой Mel частоты;
- г) использование дискретного косинусного преобразования вышеупомянутого, как будто это был сигнал.

А также задача распознавания речи решается с помощью Марковских моделей, однако вторая и третья задача тут представляют собой неразрывно связную систему, в которой все найденные вероятности, принадлежит ли входной сигнал к данной ли единице распознавания, процесс анализируются и система обучается. Скрытая Марковская модель (СММ) – это модель, состоящая из N состояний, где некоторая система может принимать одно из M значений какого-либо параметра.

Скрытой Марковской моделью называется тройка $\lambda = \{A, B, \pi\}$. Яркий пример системы распознавания речи, использующий СММ, является Sphinx. CMUSphinx – это независимый от диктора распознаватель непрерывной речи, использующий Скрытую Марковскую модель и n-граммную статистическую языковую модель.

Классификация методов распознавания образов описывают разную классификацию методов распознавания речи. Классифицируют методы на параметрические, непараметрические и эвристические, другие используют выделение методов исходя из исторически появившихся школ и направлений в данной области. К примеру, работа Темникова Ф. Е., предоставляет прекрасный обзор методов распознавания, используется следующая типология методов распознавание образов:

- методы, основанные на принципе разделения;
- статистические методы;
- методы, построенные на основе "потенциальных функций";
- методы вычисления оценок (голосования);
- методы, базирующиеся исчислении высказываний, в частности на аппарате алгебры логики.

Другая классификация основывается на реализации рассмотренных выше основных принципов построения автоматических систем распознавания образов. Существуют три основных типа методологии:

- эвристическая;
- математическая;
- лингвистическая (синтаксическая).

Нередко системы распознавания создаются на основе комбинации этих методов.

Эвристический подход основывается на интуиции и полученном опыте; в нем используются основы перечисления членов класса и общности свойств. Конечно системы, построенные такими методами, содержат набор своеобразных процедур, созданных согласно конкретным задач распознавания. Приведём пример подобного подхода связанный с задачей распознавания букв, слогов, в которой классификация образа (буквы, слога) проводилась на основе выделения таких признаков, как количество, последовательность и расположение специфических пиков аудио потока. Хотя эвристический подход имеет большое значение в построении систем распознавания образов, но относительно общих принципов синтеза не всё так просто, поскольку развязка каждой конкретной задачи требует использования специфических приемов разработки системы. Это означает, что структура и качество эвристической системы распознавания в значительной степени определяются одаренностью и опытом разработчиков.

Правила классификации, которые формулируются и выводятся в рамках определенного математического формализма с помощью принципов общности свойств и кластеризации, составляют основу математического подхода. Этим математический подход отличается от эвристического, в котором развязки определяются с помощью правил, тесно связанных с характером решаемой задачи. Математические методы построения систем распознавания делится на два класса: детерминистские и статистические. Детерминистский подход основывается на математическом аппарате, но сам аппарат не применяет в явном виде статистические свойства исследуемых классов образов. Статистический подход базируется на математических правилах систематики и классификации, которые определяются и выводятся в терминах математической статистики. Статистический классификатор строится в общем случае, предполагает применение байесовского классификационного правила и его разновидностей. Байесовое классификационное правила обеспечивает получение оптимального классификатора в тех случаях, когда известны плотности распределения для всех групп образов и вероятности появления образов каждого класса. Для работ данного направления характерным является следующее. Допустим, что задано некоторое множество образов, каждый из которых

принадлежит одному из классов W . Способ разделения на классы не является достоверно известным, но заданные конечные выборки образов, для которых заранее известна принадлежность некоторого класса. Задача заключается в том, чтобы классифицировать любой новый образ. Для этого выбирают решение правило. Основой для такого выбора является теория статистических решений. Истинные вероятности характеристик классов обычно неизвестны. Поэтому практически используются оценки, полученные по известным примерам классов. Эту обычную в статистике процедуру оценки по выборке вероятностных характеристик генеральной совокупности интерпретируют как здесь обучения. Решающие правила имеют дело не с самими объектами, а с некоторыми числовыми характеристиками, следовательно фактически не зависят от специфики объектов. Вследствие этого эти правила имеют известную всеобщность. Методы статистических решений ведут к решающим процедурам, предпочтительно удовлетворяя наперед заданному критерию и наилучшим образом используют известные параметры. Критерием оптимальности называется требование, чтобы вероятность неправильного распознавания, была минимальной. Этот критерий эквивалентен критерию Байеса, если апостериорные вероятности классов равны между собой, цены ошибочных классификаций равны между собой, а результирующая граница решений является контуром, на котором функции плотности для пары классов имеют одно и то же значение.

Лингвистические (синтаксисные) методы базируются на использовании специальных грамматик производящих языков, с помощью которых можно описывать совокупность свойств объектов, которые распознаются. Для разных классов объектов выделяются атомарные (непроизводные) элементы (подобразы, признаки) и вероятные отношения между ними. Правила построения объектов из этих непроизводных элементов называется грамматикой производящего. В рассматриваемом случае, объект - является группой непроизводных элементов «соединенных» между собой теми или иными способами, иначе говоря является «предложением» некоторого «языка». При этом важна информация, которая описывает структуру каждого объекта, а от процедуры распознавания требуется, чтобы она давала возможность отнести объект к определенному классу. Также процедура описывает те свойства объекта, которые делают невозможным отнесение его к другому классу. Типичным примером является распознавание речи. Объекты, рассматриваемые в данном классе задач являются сложными и число необходимых признаков, является большим. Это является причиной того, чтобы использовать описание сложного объекта в виде иерархической структуры более простых подобразов (элементов).

СПИСОК ЛИТЕРАТУРЫ:

1. Кипяткова И.С., Карпов А.А. Аналитический обзор систем распознавания русской речи с большим словарем // Труды СПИИРАН. – 2010. – Вып. 12. – С. 7-20.
2. Veiga A., Candeias S., Lopes C., Perdigão F. Characterization of hesitations using acoustic models // In Proc. of the 17th International Congress of Phonetic Sciences (ICPhS XVII). Hong Kong, China, 2011. – P. 2054-2057.
3. Кузин Л. Т. Основы кибернетики: В 2-х томах. — М: Энергия, 1979. — 576 с.
4. Фу К., Гонсалес Р., Ли К. Робототехника. — Пер. с англ. изд. — М.: Мир, 1989. — 624 с.
5. Бобков А.В., Системы распознавания образов, МГТУ, 2018.
6. Воробьев С.Н., Осипов С.С. Параметрическое обучение в теории распознавания образов: Учебное пособие, ГУАП, 2005
7. Chiranji Lal Chowdhary, G. Thippa Reddy, B. D. Parameshachari, Computer Vision and Recognition Systems: Research Innovations and Trends, Apple Academic Press, 2022.
8. Масальский Г.Б., Математические основы кибернетики.
9. Botir Usmonov and Quvvatali Rakhimov, Vibration analysis of airfoil on hereditary deformable suspensions, E3S Web of Conferences 97, 06006 (2019).