УДК 004.934-042.2

СРАВНЕНИЕ СИСТЕМ ОПРЕДЕЛЕНИЯ ЯЗЫКА РАЗГОВОРНОЙ РЕЧИ

АФАНАСЬЕВА А. А., БЕЛОДЕД Н. И.

Академия управления при Президенте Республики Беларусь
(г. Минск, Республика Беларусь)

E-mail: afanasyeva_25_01@mail.ru

Аннотация. В работе проведён анализ в системах автоматического определения языка речевого сигнала, основанных на скрытых Марковских моделях (СММ), на многослойной самоорганизующейся сети Кохонена (МССК) и на системе, использующей сверточные нейронные сети (СНС), а также приведена оценка точности определения языка данных систем распознавания языка речи.

Abstract. The paper analyzes automatic speech signal language detection systems based on hidden Markov models, a multi-layer self-organizing Kohonen network, and a system using convolutional neural networks, it also evaluates the accuracy of language detection in speech recognition systems.

Для решения задачи определения языка разговорной речи существуют разные подходы, но в том или ином случае система (SLI) должна соответствовать определённым требованиям:

1. Она не должна быть "смещена" в сторону какого-либо языка. Это может быть по причине разного количества данных или различного числа векторов признака для одной модели;
2. Сложность системы должна быть сведена к минимуму, как и время определения языка.
3. Желательно приемлемое снижение производительности системы при увеличении количества определяемых языков или уменьшении длительности высказывания;
4. Устойчивость системы при при присутствии низкого соотношения сигнал-шум.

В статье рассмотрены основные системы определения языка разговорной речи и проведён и их анализ.

Акустическое моделирование с использованием скрытых Марковских моделей является отличным инструментом для описания стохастических процессов. Как известно, для работы с ними не существует точных математических моделей, их свойства меняются с течением времени в соответствии со статическими законами.

Скрытой Марковской Моделью (СММ) называют статистическую модель, которая, имитирует работу процесса, похожего на Марковский процесс с неизвестными параметрами Её задача состоит в разгадывании неизвестных параметров на основе наблюдаемых, в результате параметры, которые были получены используются в дальнейшем анализе, например, для распознавания образов.

Отличие обычной Марковской модели от скрытой состоит в том, что в обычной состояние видимо наблюдателю, поэтому вероятности переходов – единственный параметр. В скрытой модели можно следить только за переменными, на которые оказывает влияние данное состояние. Каждое состояние имеет вероятностное распределение среди всех возможных выходных значений, поэтому последовательность символов, сгенерированная СММ, даёт информацию о последовательности состояний.

В основе СММ лежит конечный автомат, состоящий из N-состояний, называемых скрытыми. Переходы между состояниями в момент времени t не являются детерминированными, а происходят в соответствии с вероятностным законом и описываются матрицей вероятностей переходов ANN. Схематическое изображение диаграммы переходов между состояниями СММ приведено на рис. 1.

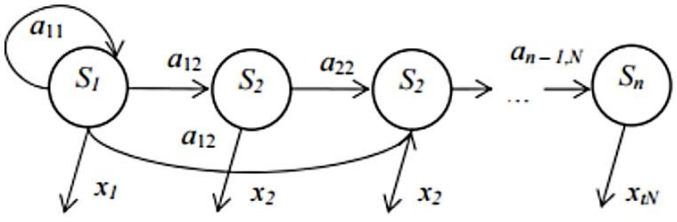

Рис. 1. Анализ диаграммы переходов между состояниями СММ

Нахождение модели в некотором состоянии i соответствует определенной стационарности наблюдаемого сигнала на ограниченном временном интервале. При осуществлении очередного перехода в новое состояние i в момент времени t происходит генерация выходного вектора $x(t)$, называемого параметрическим вектором, в соответствии с многомерной функцией распределения вероятностей $f_j(x)$. Результатом работы СММ является последовательность векторов (наблюдений) $x_1, x_2, \ldots, x_T$ длиной.

Работа с СММ осуществляется в 2 этапа.

1) Обучение, определение параметров модели. СММ обучается по алгоритму Баума-Велча.

Процесс обучения модели заключается в определении с помощью набора обучающих образцов следующих параметров:

- матрицы вероятностей переходов между состояниями ANN;
- параметров гауссовых смесей (математическое ожидание, матрица ковариации и веса) для каждого состояния.

2) Определение.

Происходит процесс декодирования СММ, который позволяет определить вероятность того, что наблюдаемая входная последовательность векторов x1,x2,…,xT могла быть сгенерирована данной моделью, а также соответствующую наиболее вероятную цепочку состояний. Декодирование модели происходит по алгоритму максимума правдоподобия (алгоритм Витерби).

Таким образом, с помощью скрытых Марковских моделей реализуются фонетические распознаватели, в которых с помощью гауссовых смесей происходит акустическое моделирование. Затем, по результирующей последовательности распознанных фонем считаются меры близости к модели n-грамм каждого языка. Данный подход реализует архитектуру PPRLM (англ. parallel phonemes recognition language model) с параллельным подключением фонетических распознавателей, обученных на нескольких языках. Пример такой системы изображен на рис. 2, где ФР – фонетический распознаватель, ЯМ – языковая модель, Р – вероятность принадлежности высказывания к ЯМ.

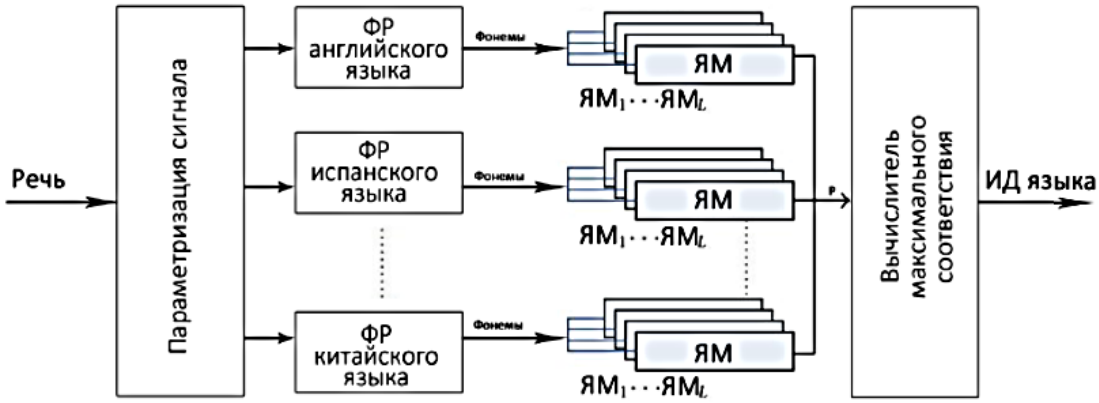

Рис. 2. Анализ архитектуры PPRLM

Однако, кроме СММ, существуют альтернативные способы моделирования статистического распределения векторов акустических признаков для речевого высказывания, такие как искусственные нейронные сети (ИНС).

Рассмотрим акустическое моделирование с использованием искусственной нейронной сети Кохонена. При использовании данного подхода проектируется ИНС Кохонена, которая обучается на большом объеме нормализованных входных данных. Такой подход освобождает от моделирования языковых моделей и не требует фонетических маркировок входных данных, однако требует время на обучение нейронной сети. Самоорганизующиеся карты Кохонена (англ. Self Organizing Maps – SOM) – это одна из разновидностей нейросетевых алгоритмов. Основным отличием данной технологии от других нейронных сетей, обучаемых по алгоритму обратного распространения, является то, что при обучении используется метод обучения без учителя, то есть результат обучения зависит только от структуры входных данных. При этом в ходе обучения модифицируется не только нейрон-победитель, но и его соседи, но в меньшей степени (рис. 2) Самоорганизующиеся карты Кохонена состоят из ячеек прямоугольной или шестиугольной формы. Каждой ячейке соответствует нейрон сети Кохонена. Обучение нейронов производится точно так же, как и обучение нейронов сети Кохонена. Система с использованием самоорганизующейся карты Кохонена представлена на рис. 3.

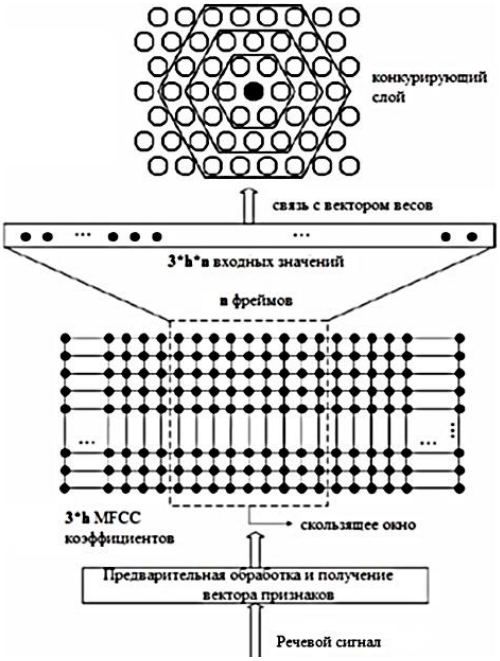

Рис.3. Анализ системы с использованием карты Кохонена

Весь процесс можно описать несколькими этапами:

1) Входная речь подвергается предварительной обработке.
2) Из нее извлекаются h MFCC-коэффициентов.
3) Извлекаются их первая и вторая производные для формирования основного 3*h-мерного вектора признаков.
4) Коэффициенты фреймами подаются на конкурирующий слой.
5) Нейронная сеть обучается.

Стоит отметить, что на практике добиться высокой точности определения языка для системы с таким подходом очень сложно. Для повышения качества системы используют многослойную самоорганизующуюся нейронную сеть Кохонена. Данная система (рис. 4) имеет пирамидальную структуру и состоит из нескольких однослойных сетей Кохонена. Количество нейронов в каждом последующем слое уменьшается. Данные поступающие на первый слой преобразуют его веса, которые являются входами для следующего слоя. Таким образом, каждый последующий уровень представляет собой более высокий уровень абстракции входных данных.

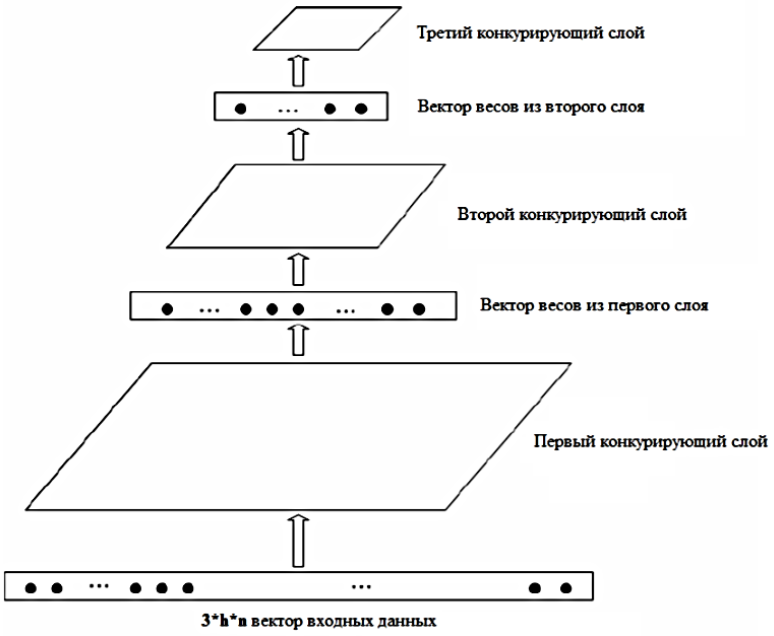

Рис. 4. Анализ самоорганизующейся нейронной сети Кохонена

Свёрточная нейронная сеть (англ. convolutional neural network, CNN) — специальная архитектура искусственных нейронных сетей, нацеленная на эффективное распознавание изображений. Данная сеть использует некоторые особенности зрительной коры головного мозга, в которой были открыты простые клетки, реагирующие на прямые линии под разными углами, и сложные клетки, реакция которых связана с активацией определённого набора простых клеток. Таким образом, идея свёрточных нейронных сетей заключается в чередовании свёрточных слоев (англ. convolution layers) и слоев подвыборки (англ. subsampling layers). Структура сети — однонаправленная (без обратных связей) и многослойная.

Наиболее простым и популярным способом обучения является метод обучения с учителем (на маркированных данных) — метод обратного распространения ошибки и его модификации. Но существует также ряд техник обучения свёрточной сети без учителя.

Название архитектура сети получила из-за наличия операции свёртки (каждый фрагмент изображения умножается на матрицу (ядро) свёртки поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения). Важно знать понятие «разделяемых» весов. Это ситуация, когда часть нейронов некоторого рассматриваемого слоя нейронной сети может использовать одни и те же весовые коэффициенты. Далее они объединяются в карты признаков (англ. feature maps), при этом каждый нейрон карты признаков связан с частью нейронов предыдущего слоя. Такие слои называются сверточными слоями.

Операция субдискретизации (англ. subsampling, англ. pooling, также переводимая как «операция подвыборки» или «операция объединения»), выполняет уменьшение размерности сформированных карт признаков. В этой архитектуре сети считается, что информация о факте наличия искомого признака важнее точного знания его координат, поэтому из нескольких соседних нейронов карты признаков выбирается максимальный и принимается за один нейрон карты признаков уменьшенной размерности. Также иногда применяют операцию нахождения среднего между соседними нейронами. За счёт данной операции, помимо ускорения дальнейших вычислений, сеть становится более инвариантной к масштабу входного изображения.

Таким образом, повторяя друг за другом несколько слоёв свёртки и субдискретизации, строится свёрточная нейронная сеть. Обычно после прохождения нескольких слоев карта признаков вырождается в вектор или даже скаляр, но таких карт признаков становится сотни. На выходе сети часто дополнительно устанавливают несколько слоев полносвязной нейронной сети, на вход которой подаются оконечные карты признаков.

Таким образом, сверточную нейронную сеть можно использовать для распознавания языка речи, обучив ее на наборе характеристических векторов, представленных в виде изображений.

В результате анализа всех трёх методов можно прийти к их комплексному обзору (табл. 1):

Таблица 1. Оценка проведённого сравнения методов

Метод	**Особенности**
1	**2**
Акустическое моделирование с использованием скрытых Марковских моделей	потребность в огромном наборе обучающих данных, размеченных с точностью до фонем сложность наращивания системы, так как разработка языковых моделей является весьма трудоемким процессом
Акустическое моделирование с использованием искусственной нейронной сети Кохонена	при увеличении количества распознаваемых языков увеличивается и количество настраиваемых параметров, что существенно сказывается на времени обучения сети, а также на ее производительности (это можно исправить с помощью сверточных нейронных сетей)
Свёрточная нейронная сеть	• имеет гораздо меньшее количество настраиваемых весов, так как одно ядро весов используется целиком для всего изображения • относительная инвариантность к повороту и сдвигу распознаваемого изображения • удобное распараллеливание вычислений, а, следовательно, возможность реализации алгоритмов работы и обучения сети на графических процессорах

В табл. 2. содержится оценка точности определения языка в системах автоматического определения языка речевого сигнала, основанных на скрытых Марковских моделях (СММ), на многослойной самоорганизующейся сети Кохонена (МССК) и на системе, использующей сверточные нейронные сети (СНС).

Таблица 2. Оценка точности определения языка в системах автоматического определения речевого сигнала

Метод	**Точность определения языка, %**
СММ	69,3
МССК	82,7
СНС (обучена на спектре)	90,17
СНС (обучена на MFCC)	96,89

Таким образом, использование сверточной нейронной сети, обученной на MFCC-коэффициентах, позволяет добиться наилучшей, в сравнении с традиционными подходами, точности определения языка.

В дальнейшем планируется создание приложения для определения речи с использованием программная библиотека для машинного обучения Tensorflow.js, интерфейс которого будет написан на JavaScript.

Список литературы

1. Свёрточная нейронная сеть [Электронный ресурс] – Режим доступа: https://ru.wikipedia.org/wiki/Свёрточная_нейронная_сеть - Дата доступа: 17.10.2020
2. А. В. Царьков, А. А. Столяров Наукоёмкие технологии в приборо- и машиностроении и развитие инновационной деятельности в вузе / материалы Всероссийской научно-технической конференции, 15 – 17 ноября 2016 г. Т. 3. – Калуга: Издательство МГТУ им. Н. Э. Баумана, 2016. – 256 с.