УДК 004.934.1

Статистический анализ в задачах распознавания речи

Новицкая К.А., студент гр. 920605, Нехлебова О.Ю., магистрант гр. 017141

Белорусский государственный университет информатики и радиоэлектроники[1]
г. Минск, Республика Беларусь

Печень Т. М. – старший преподаватель каф. ИКТ

Аннотация. В работе обобщается опыт теоретических и прикладных исследований, проводимых на кафедре ИТСиТ Института инженерных технологий и естественных наук.

Ключевые слова. Евклидово расстояние, расстояние Махаланобиса, корреляционная функция, речевой сигнал.

В настоящее время система распознавания речевых сигналов при общении человека с компьютером развивается огромными темпами. Растет важность массового внедрения новых интерфейсов для такого рода взаимодействия, поскольку традиционные интерфейсы уже в скором времени могут достигнуть предела в своём развитии. Также это связано с тем, что около 85% данных мы получаем через органы зрительного восприятия. Таким образом данный канал сейчас становится все более перегруженным. Первоочередной альтернативой является использование акустического канала. Таким образом эта технология становится актуальна не только для людей с ограниченными возможностями, но и для всех людей, активно пользующихся техникой.

Программы обработки в частотно-временной области представляют собой методы, содержащие все преимущества временного и частотного анализов с минимальными проявлениями их недостатков. В зависимости от выбранного метода обработки программы можно разделить на несколько видов. В данной работе были проанализированы виды анализа: с использованием линейных преобразований; с использованием корреляционной функции и евклидовых преобразований.

Применение методов Евклидова расстояния.

Евклидова метрика — наиболее естественная функция расстояния, отражающая интуитивные свойства расстояния между точками. Чем меньше значение евклидова расстояния между речевыми сигналами, которые вычисляются по теореме Пифагора, тем они похожее. Евклидово расстояние между точками речевых сигналов вычисляется по следующей формуле (сноска):

$$d(x, y) = \sqrt{\sum_{i=0}^{N-1} (x_i - x_i)^2} \quad (1)$$

где N – количество точек входных речевых сигналов; i – индексы отсчётов речевых сигналов; x и y – входные речевых сигналы.

Евклидовы методы расчета широко используются в анализе данных в качестве критерия для объединения наблюдений в классы и кластеры, оценки ошибок в предсказательной аналитике, а также в инструментах визуализации.

Расстояние Махаланобиса.

Расстояние Махаланобиса является мерой расстояния между векторами случайных сигналов, обобщающая понятие Евклидова расстояния. Оно отличается от Евклидова расстояния тем, что оно учитывает корреляции между сигналами и инвариантно к масштабу. Однако данная мера расстояния плохо работает в случаях, когда ковариационная матрица рассчитывается исходя из всего множества входных данных. В то же время, будучи сосредоточенной на определенной группе данных, данная метрика показывает хорошие результаты:

$$d(x, y) = (M(x) - M(y))^T \cdot (\frac{nx}{nx + ny} \cdot cov(x) + \frac{ny}{nx + ny} \cdot cov(y))^{-1} \cdot \left(M(x) - M(y)\right) \quad (2)$$

где nx – длина речевого сигнала x; ny – длина речевого сигнала y; cov – ковариация речевого сигнала; M – математическое ожидание речевого сигнала; x и y – входные речевые сигналы.

Ковариация и корреляция.

Ковариация оценивает силу линейной зависимости между двумя сигналами, но в то же время не позволяет определить ее:

$$\mathrm{cov}(x,y) = M\left[\left(x_i - M(x)\right) \cdot \left(y_i - M(y)\right)\right] \tag{3}$$

где M - математическое ожидание речевого сигнала; x и y – входные речевые сигналы; i – индексы отсчётов речевого сигнала.

Для получения более точного значения силы зависимости нужно рассчитать коэффициент корреляции. Корреляционная функция является линейной, как и ковариация, и имеет выборку значений от -1 до +1. Линейность корреляции означает, что все точки будут лежать на одной прямой.

Таким образом ковариация - это ни что иное, как мера корреляции:

$$K(x,y) = \frac{\sum_{i=1}^{N}(x_i - M(x)) \cdot (y_i - M(y))}{\sqrt{\sum_{i=0}^{N}(x_i M(x)) \cdot \sum_{i=0}^{N}(y_i - M(y))}} = \frac{\mathrm{cov}(xy)}{\sigma_x \sigma_y} \tag{4}$$

где M (x) - среднее значение речевого сигнала (математическое ожидание); N – кол-во отсчётов речевого сигнала; i – индексы отсчётов речевого сигнала; x и y – входные речевые сигналы; cov – ковариация между речевыми сигналами; σ - среднеквадратическое отклонение речевого сигнала.

Корреляция относится к масштабированной форме ковариации. Значение корреляции имеет место между -1 и +1. А значение ковариации лежит между -∞ и + ∞. На ковариацию влияет изменение масштаба, т.е. если все значение одной переменной умножается на постоянную, а все значение другой переменной умножается на аналогичную или другую постоянную, то ковариация изменяется.

При выборе метода корреляция предпочтительнее ковариации, поскольку она не зависит от изменения местоположения и масштаба, а также может использоваться для сравнения между две пары переменных.

Среднеквадратическое отклонение (СКО).

СКО - распространённый показатель рассеивания значений случайной величины относительно её математического ожидания:

$$d(x,y) = \sqrt{\frac{\sum_{i=0}^{N-1}(x_i - y_i)^2}{\sum_{i=0}^{N-1}(x_i)^2}} \tag{5}$$

где x и y – входные речевые сигналы; i – индексы отсчётов речевого сигнала; N – кол-во отсчётов речевого сигнала.

Среднее квадратичное отклонение или СКО несет информацию о мощности отклонения сигнала от среднего значения. Также оно отражает шум и другие помехи. Среднеквадратическое отклонение находит отклонение речевых сигналов друг от друга. Поэтому чем меньше значение среднеквадратического отклонения между речевыми сигналами, тем они похожее друг на друга.

Оценка результатов вычислений на основе гипотез.

В данной работе рассматриваются гипотезы о виде и параметрах распределения некоторой генеральной совокупности, а также о сравнении выборок из различных генеральных совокупностей.

Допустим, что нам дана случайная выборка (x_1, x_2,…, x_n) объема n из некоторой генеральной совокупности (конечной или бесконечной). Каждое значение x_i в этой выборке само является случайной величиной, даже если генеральная совокупность состоит из конечного числа элементов. Необходимо также иметь в виду, что случайная выборка из какой-либо генеральной совокупности должна соответствовать некоторой схеме испытаний, при реализации которой выявляется искомая случайная величина X.

Располагая выборочными данными и руководствуясь конкретными условиями рассматриваемой задачи, формулируют гипотезу H_0, которую называют основной или нулевой, и гипотезу H_1, конкурирующую с гипотезой H_0. Гипотезу H_1 называют также альтернативной. H_0

и H_1 – две взаимно исключающие гипотезы. Отметим, что для одной основной гипотезы может быть выдвинуты несколько альтернативных.

Для проверки статистической гипотезы используется специально подобранная случайная величина с известным законом распределения, называемая статистическим критерием. Множество ее возможных значений разбивается на два непересекающихся подмножества: одно из них (критическая область) содержит значения критерия, при которых нулевая гипотеза отклоняется, второе (область принятия гипотезы) – значения, при которых она принимается.

Критерия, позволяющего точно (на 100%) узнать, верна гипотеза H_0 или нет, не существует в силу ограниченности и случайности выборки, так как выборка не содержит всей информации о генеральной совокупности. Отклоняя или принимая гипотезу H_0, можно допустить ошибку двух видов:

– ошибка первого рода совершается при отклонении гипотезы H_0 (т. е. принимается альтернативная H_1), тогда как на самом деле гипотеза H_0 верна; вероятность такой ошибки обозначим:

$$\alpha = P(H_1 / H_0) \tag{6}$$

– ошибка второго рода совершается при принятии гипотезы H0, тогда как на самом деле высказывание H_0 неверно и следовало бы принять гипотезу H_1; вероятность ошибки второго рода обозначим как:

$$\beta = P(H_0 / H_1) \tag{7}$$

Вероятность ошибки первого рода при проверке статистических гипотез называют уровнем значимости и обычно обозначают α. Вероятность ошибки второго рода обозначается β. Величина (1 − β) – мощность критерия. Чем выше мощность, тем меньше вероятность совершить ошибку второго рода.

Вероятность ошибки первого рода при проверке статистических гипотез называют уровнем значимости и обычно обозначают α. Вероятность ошибки второго рода обозначается β. Величина (1 − β) – мощность критерия. Чем выше мощность, тем меньше вероятность совершить ошибку второго рода.

Таким образом, можно сделать вывод, что построенные на основании этих методов закономерности относятся не к отдельным испытаниям, из повторения которых складывается данное массовое явление, а представляют утверждения об общих вероятностных характеристиках данного процесса. Такими характеристиками могут быть вероятности, плотности распределения вероятностей, математические ожидания, дисперсии и т. п.

Найденные характеристики позволяют построить вероятностно-статистическую модель изучающую вопрос о распознавании речи. Таким образом, теория вероятностей по вероятностной модели процесса предсказывает его поведение, а математическая статистика по результатам наблюдений за процессом строит его вероятностно-статистическую модель.

Список использованных источников:
1. Марей Раад Али Салех «Исследование признаков речевых сигналов для задач распознавания речи», 2017.
2. С. Е. Демин, Е. Л. Демина: «Математическая статистика»: учебно-методическое пособие.

UDC 004.934.1

STATISTICAL ANALYSIS IN SPEECH RECOGNITION PROBLEMS

Novitskaya K.A.,Niakhlebava V.Y.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Pechan T.M. – assistant professor of the academic department of ICT

Annotation. The work summarizes the experience of theoretical and applied research carried out at the Department of ITS&T of the Institute of Engineering Technologies and Natural Sciences.

Keywords. Euclidean distance, Mahalanobis distance, correlation function, speech signal.