

ВЫСОКОРАЗМЕРНАЯ КЛАСТЕРИЗАЦИЯ ГОЛОСОВЫХ ФРАГМЕНТОВ РЕЧИ С ИСПОЛЬЗОВАНИЕМ МАШИН БЕРНУЛЛИ

Надеев Д.Н. (г. Пенза)

В работе [1] показано, что высоконадежную биометрическую аутентификацию человека по его голосу целесообразно реализовывать с использованием многомерных нейросетевых преобразователей биометрия-код. В качестве входов преобразователя возможно использовать коэффициенты Фурье по кадрам голосовой парольной фразы с синхронизацией [2]. Чтобы синхронизировать по всем примерам фразы необходимо проводить кластеризацию фрагментов речи на фонемоподобные участки.

В настоящее время для многомерной линейной кластеризации применяют метод К-средних, дискриминантный анализ, нейронные сети Кохонена, генетические алгоритмы, алгоритмы иерархического объединения и разделения, алгоритмы семейства KRAB, FOREL, DBSCAN и т.д. [3]. В них вычисляется обратная матрица статистических характеристик примеров или итерационно решается задача оптимизации параметров выбранной модели.

Операция обращения матриц для биометрических данных при размерности большей 3 перестает быть устойчивой. При этом возникает и начинает накапливаться шум от прямого и обратного преобразований. Практически все итерационные алгоритмы при размерности больше 16 вместо поиска решения начинают блуждать в своих локальных экстремумах, оптимизируя многомерную поверхность ошибок. Поэтому использовать эти методы в биометрии с числом входов 32, 64 и более можно только предварительно понизив размерность задачи, например, методом главных компонент. Для биометрии это нежелательно, так как теряется значительная часть информации об индивидуальных свойствах голоса человека.

Чтобы выполнить многомерную кластеризацию и сохранить биометрическую информацию можно использовать схему Бернулли и, построенную на ее основе, машину Бернулли [4]. Схема Бернулли преобразует множество входных одномерных распределений случайных данных в одно многомерное распределение [4]. При разных значениях корреляции входов получаются разные выходные распределения. Это свойство может быть положено в основу многомерной биометрической кластеризации.

При этом целесообразно сравнивать не выходные распределения, а накапливать сумму мер, зависящих от взаимного расположения классов по отношению друг к другу. Квантователь Бернулли должен быть предварительно настроен по центрам «свой», «чужой-», «чужой+» для каждого класса. Суммирование мер ведется по каждому коэффициенту (биометрическому параметру). Построенные таким образом квантователи объединяются друг с другом для получения многомерного решения о кластеризации.

На каждом огласованном кадре голосовой фразы вычисляется вектор из 10 коэффициентов. Это могут быть первые 10 максимальных коэффициентов преобразования Фурье. Множество векторов коэффициентов делится сначала на 3 класса по математическому ожиданию и дисперсии примеров каждого класса.

Входные данные первого квантователя получаются усреднением в центре каждого класса в интервале $[t_0, t_1]$, где t_0, t_1 -номера кадров в центрах каждого

класса. Классы в голосовой фразе могут следовать в произвольном порядке, поэтому предварительно проводится сортировка центров каждого класса по возрастанию значений каждого коэффициента. По каждому коэффициенту вычисляется первая мера взаимного расположения 1 класса «свой» по отношению к 2-му классу «чужой». Эта мера 1 класса максимальна при попадании примеров 1 класса в центр «свой» с учетом дисперсии «свой» и минимальна в центре 2 класса «чужой» с учетом его дисперсии. Эта мера прибавляется к второй мере дающей максимум при попадании примеров 2 класса в центр «свой» по среднему и дисперсии.

Вторая мера минимальна при попадании примеров 2 класса в центр 1-го класса «чужой-» и в центр 3-го класса «чужой+» с учетом их дисперсии. Третья мера вычисляется когда 3-й класс принят за «свой», а 2 класс за «чужой». Эти три меры суммируются друг с другом. Аналогично рассчитываются и накапливаются меры по 2-му, 3-му и т.д. коэффициентам. Полученная сумма является линейным выходом квантователя. Пороговая функция используется для получения нелинейного отклика квантователя.

Входные данные второго квантователя получаются сдвигом и усреднением в интервале $[t0-1, t1-1]$, где $t0-1, t1-1$ - номера кадров сдвинутых на один от центра каждого класса. Веса, сумма по классам и сумма по коэффициентам вычисляется так же как для первого квантователя. Сдвигая и усредняя входные данные, проводим настройку необходимого числа квантователей, например 128 квантователей. Получаем 128-битный отклик машины Бернулли, который является кодом кластеризации при трех классах. Настраивая несколько машин Бернулли под четыре, пять, шесть и т.д. классов получаем несколько 128-битных кодов кластеризации. Объединение этих кодов позволяет получать высокоразмерную кластеризацию с высокой вероятностью попадания примеров каждого класса в свой центр с учетом своих средних значений и дисперсии.

Таким образом, итерационная задача максимизации целевого функционала для общего случая входных данных

$$a_k(i, j) = \arg \max_{a_k(i, j)} y_i(x_i, a_k(i, j)), \quad k = 1, 2 \dots \text{коэф} - \text{в}, \quad (1)$$

сводится для конкретного частного случая биометрических данных просто к выбору меры по каждому коэффициенту

$$y(i) = f([x_i > E(x_{i-1})] \wedge [x_i < E(x_{i+1})], \sigma(x_{i-1}), \sigma(x_{i+1}), a(i-1, i+1)), \quad (2)$$

где $\arg \max_x y(x)$ – поиск аргумента максимизирующего функцию; $f(\cdot)$ – функция максимальная при попадании i -го класса между $i-1$ и $i+1$ классом в порядке их следования в голосовой фразе.

Функция $f(\cdot)$ дает минимум при любом другом взаимном расположении $i, i-1, i+1$ классов по отношению друг к другу. Эта мера зависит от среднего $E(\cdot)$ и корня квадратного дисперсии $\sigma(\cdot)$ каждого класса.

На практике использовать один квантователь и малое число входов нельзя из-за низкого доверия к качеству принимаемого решения о кластеризации. Надо использовать 8, 16, 32 и т.д. квантователей и число входов квантователя 8, 16, 32 и т.д. При переходе от одного квантователя к 2, 3, 4 квантователям и т.д., одномерное функциональное уравнение

$$y_1 = \text{Порог}_1 \left(\sum_{\text{классы}=1}^{\text{классов}} \sum_{\text{коэф-ты}=1}^{\text{коэф-в}} f_{ij}(x_i, a(i-1, i+1)) \right) = \text{Порог}_1(f_{ij}) \quad (3)$$

переходит в многомерное функциональное уравнение

$$\{y_1, y_2, \dots, y_{128}\} = \{\text{Порог}_1(f_{ij}), \text{Порог}_2(f_{ij}), \dots, \text{Порог}_{128}(f_{ij})\} \quad (4)$$

где $\{.\}$ -операция объединения множества одномерных решений в одно высокоразмерное решение; $\text{Порог}_i(.)$ - пороговая функция i -го квантователя от его линейного выхода.

Формально интерпретируя добавление каждого выхода как переход от одного решения функционального уравнения (3) к группе решений можно записать следующее выражение:

$$y_1, G(y_1) \Rightarrow \{y_1, y_2, \dots, y_{128}\}, G(y_1, G(y_2, G(y_3) \dots)) \dots = G(y_1, y_2, \dots, y_{128}) \quad (5)$$

где $G(.)$ – группа Ли т.е. множество решений с алгеброй по выражениям (1)-(4).

Выражение (5) означает что происходит групповое расщепление т.е. переход от одного бита решающего правила к 2, 3, 4...128 битного кода кластеризации [4]. На практике это позволяет при переходе к мере Хэмминга получать оценку логарифма вероятности кластеризации $-\log_{10}(P_{\text{синх}})$.

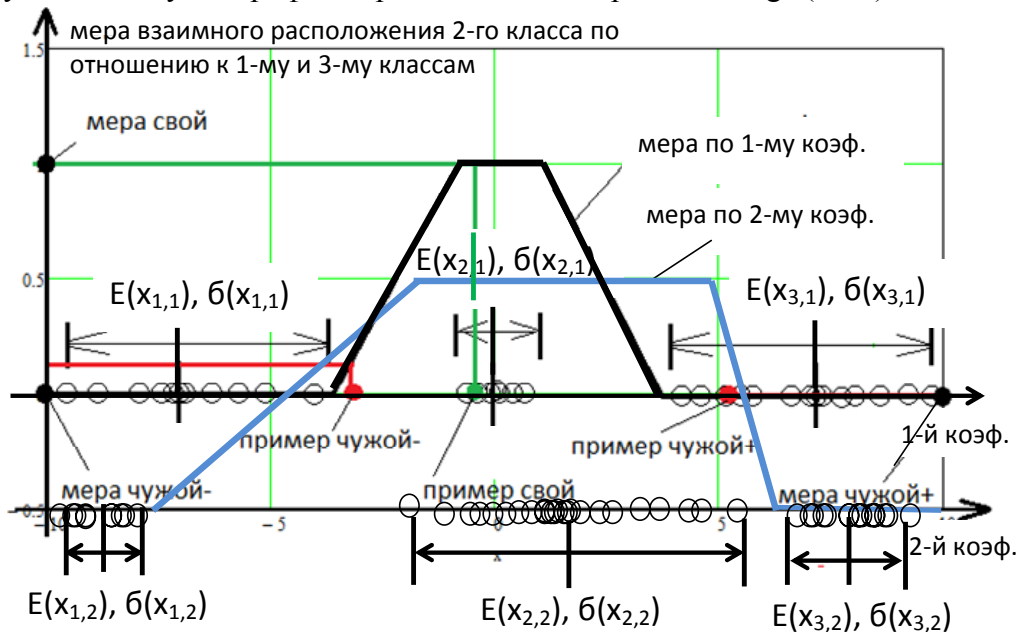


Рис. 1. Разделяющая функция, максимальная при попадании примеров 2 класса между центрами 1-го и 3-го класса по 1-му и 2-му коэффициенту на голосовой фразе «Россия родина слонов»

Взаимное расположение 2 класса по отношению к 1 и 3 классу по 1-му коэффициенту оценивается мерой $y_1(x_2)$ которая вычисляется следующим образом:

$$y_1(x_2) = [-|x_{2,1} - E(x_{1,1})| * a_1(2,1) + 1 - |x_{2,1} - E(x_{3,1})| * a_1(2,3) + 1] / b(x_{2,1}), \quad (6)$$

$$a_1(2,1) = 1 / [|E(x_{2,1}) - E(x_{1,1})| * (1 - b(x_{1,1}))], \quad (7)$$

$$a_1(2,3) = 1 / [|E(x_{2,1}) - E(x_{3,1})| * (1 - b(x_{3,1}))], \quad (8)$$

где $a_1(2,1)$ – вес, учитывающий расстояния между центрами распределений примеров своего и чужого и дисперсию чужого; $a_1(2,3)$ – вес, учитывающий расстояния между центрами распределений примеров своего и чужого и дисперсию чужого; $E(x_{2,1}), b(x_{2,1})$ – математическое ожидание и корень квадратный дисперсии распределения примеров свой 2-го класса по 1-му коэффициенту; $E(x_{1,1}), b(x_{1,1})$ - математическое ожидание и корень квадратный дисперсии распределения примеров «чужой-» 1-го класса по 1-му коэффициенту; $E(x_{3,1}), b(x_{3,1})$ - математическое ожидание и корень квадратный

дисперсии распределения примеров «чужой+» 3-го класса по 1-му коэффициенту. График меры $y_1(x_2)$ показан на рисунке 1 (верхняя функция).

Аналогично вычисляются меры взаимного расположения классов по 2-му коэффициенту. Например, мера $y_2(x_3)$ разделения 3 класса по отношению ко 2 классу и 4 классу по 2-му коэффициенту вычисляется следующим образом:

$$y_2(x_2) = [-|x_{2,2}-E(x_{1,2})|*a_2(2,1)+1 -| x_{2,2}-E(x_{3,2})|* a_2(2,3)+1] / \delta(x_{2,2}), \quad (9)$$

$$a_2(2,1)=1 / [|E(x_{2,2})-E(x_{1,2})| * (1-\delta(x_{1,2}))], \quad (10)$$

$$a_2(2,3)=1 / [|E(x_{2,2})-E(x_{3,2})| * (1-\delta(x_{3,2}))], \quad (11)$$

где «свой» - это 2-й класс 1 коэффициент, «чужой-» - это 1-й класс 2 коэффициент, «чужой+» - это 3-й класс 2-й коэффициент. График меры $y_2(x_3)$ показан на рисунке 1 (нижняя функция).

Выражения (6)-(11) можно обобщить. Так, мера взаимного расположения i класса «свой» по отношению к $i-1$ «чужой-» и $i+1$ «чужой+» классам по j коэффициенту оценивается следующим образом:

$$y_j(x_i) = [-|x_{i,j}-E(x_{i-1,j})|*a_j(i,i-1)+1 -| x_{i,j}-E(x_{i+1,j})|* a_j(i,i+1)+1] / \delta(x_{i,j}), \quad (12)$$

$$a_j(i,i-1)=1 / [|E(x_{i,j})-E(x_{i-1,j})| * (1-\delta(x_{i-1,j}))], \quad (13)$$

$$a_j(i,i+1)=1 / [|E(x_{i,j})-E(x_{i+1,j})| * (1-\delta(x_{i+1,j}))]. \quad (14)$$

На практике машины Бернулли можно использовать при синхронизации по времени и по форме нескольких голосовых фраз. Для этого выполняется кластеризация на трех классах по всем фразам. Если последовательность классов одинаковая для всех фраз, то идет кластеризация по четырем, пяти, шести и т.д. классам. Когда последовательность классов для разных фраз станут разными, то кластеризация заканчивается и сохраняются последний и все предыдущие результаты кластеризации.

При этом очевидно отличие между экспертом и машиной Бернулли. Эксперт может отслеживать примерно 10 последовательностей длины около 10, т.е. анализирует 100 входов. Для этого он учится 1 год. По сравнению с экспертом программный эмулятор машины Бернулли обрабатывает 1000, 2000, 3000 и более входных данных. Теоретически число входных данных и вероятность кластеризации может быть сколь угодно высокими. Практически пользователь системы аутентификации по голосу выбирает нужную ему вероятность кластеризации. При этом автомату не надо тратить на настройку год. Настройка квантователей на вычислительной машине средней производительности происходит за время менее одной секунды. Схема квантователя Бернулли с 10 коэффициентами 5 классами приведена на рисунке 2.

В качестве пороговых функций квантователей Бернулли могут использоваться квадратичные формы, используемые в работе [2] для решения задачи оптимального разделения центра «свой» от центра «чужой». Эти функции дают максимум при попадании примеров класса в центр этого класса по математическому ожиданию.

Минимум лежит в центре области «чужой». Нормировка проводится по дисперсии «свой», «чужой». Однако для задачи кластеризации и родственной ей задачи синхронизации требуется не разделение центров, а взаимное расположение центров классов по отношению друг к другу.

Поэтому вместо квадратичных форм целесообразно использовать кусочнолинейно-кусочнонелинейную функцию вида (12)-(14). Эта функция при сложении по каждому коэффициенту гонит примеры «свой» в максимум, а примеры «чужой-» и «чужой+» в минимум. Нормировка идет по дисперсии «свой», «чужой-», «чужой+». Слева и справа от центра «свой» функция взаимного расположения имеет пологий участок. В этом интервале примеры «свой» дают одинаковую максимальную меру.

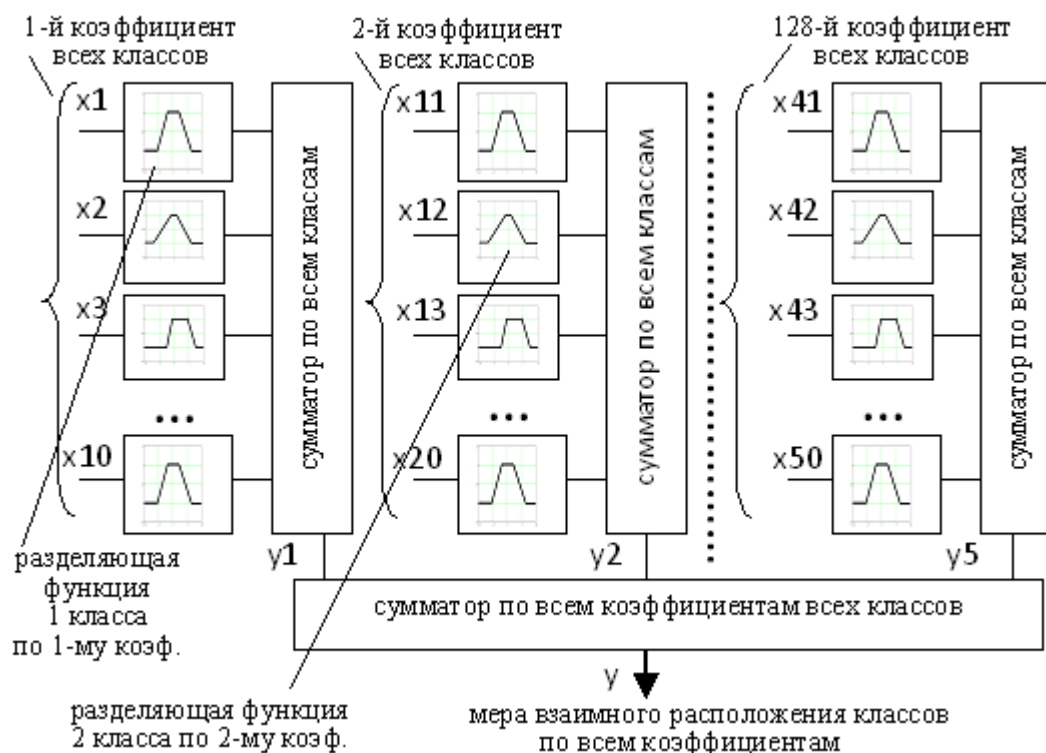


Рис. 2 Квантователь Бернулли с 10 коэффициентами 5 классами

Также мера постоянна в центрах «чужой-» и «чужой+» с соответствующими дисперсиями. В интервале между центром «свой» и «чужой-» мера линейно возрастает по направлению к «своему». Аналогично мера возрастает от края области «чужого+» к краю области «свой». В этих линейных участках чем ближе будет пример к «своему» тем больше будет мера.

Таким образом, отличие квантователей Бернулли от нейронов нейросетей состоит в том, что они решают разные задачи. Нейрон нейросети решает задачу наибольшего разнесения между двумя классами и накопления по всем коэффициентам. Квантователь Бернулли решает задачу выделения взаимного расположения и накопления по нескольким коэффициентам и нескольким усреднениям по времени. Вместо разделяющих квадратичных форм используются линейно-нелинейные функции взаимного расположения классов. Вместо двух классов машина Бернулли обрабатывает 3, 4, 5 и т.д. классов. Чем больше классов тем выше вероятность кластеризации и синхронизации. Процедура настройки квантователя Бернулли приведена на рисунке 3.

Условием возможности кластеризации является оптимальный выбор центров классов. Однако в случае биометрии мы заменяем поиск оптимальных центров выбором конкретных, заранее выбранных центров и разброса значений вокруг этих центров. Выбранные центры классов с соответствующими разбросами подаются на машину и проверяется идентичность нелинейных откликов по мере Хэмминга для всех примеров голосовой фразы.

Таким образом, фактор пространство оптимальных центров заменяется на оптимальные меры по нескольким фактор пространствам. Фактор пространство в случае биометрии означает при 10 коэффициентах вычисление разности между 10-мерным центром класса и 10-мерным примером этого класса. Результат является 10-мерным числом, которое минимально в центре рассматриваемого класса. Чем дальше расстояние от центра класса тем больше по модулю будут значения разностей.



Рис. 3. Процедура настройки квантователя Бернулли

Оптимальная равноудаленность центров определяется косвенно по откликам квантователей. Человек произносит одинаковые голосовые фразы всегда примерно одинаковым образом от одной фонемы к другой. От центра каждого класса влево и вправо по времени постепенно происходит переход в центры классов-соседей. При оптимальной расстановке центров происходит меньше переключений между классами и разброса результатов соседних усреднений по времени. Это соответствует меньшей дисперсии откликов квантователей и может непосредственно использоваться для поиска оптимального выбора центров классов. Синхронизационная машина Бернулли, 6400 входов и 128 выходов, 5 классов, 10 коэффициентов каждого класса, 10 квантователей с 50 входами с 50 весами и одним выходом показана на рисунке 4.

Для использования большого числа квантователей, например 128 квантователей, необходимо синтезировать таблицу связей входов машины с входами квантователей. Число входов одного квантователя равно произведению числа классов на число коэффициентов класса. Например, если классов 5, а коэффициентов 10, то число входов квантователя равно 50.

Число весов квантователя равно удвоенному числу его входов так как класс «свой» сравнивается с классом «чужой-» и классом «чужой+». Для 50 входов число весов равно 100. Усреднения в центрах классов по всем коэффициентам дает данные для первого квантователя. Сдвиг на один кадр от центров каждого класса дает данные для второго квантователя.

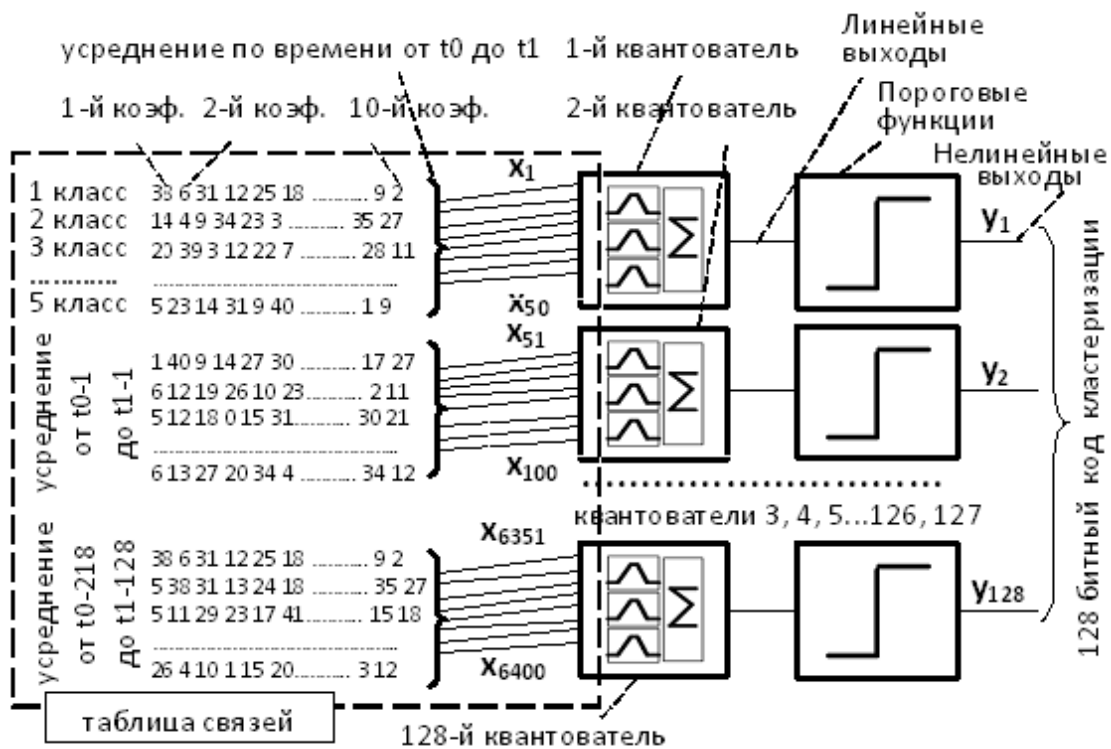


Рис. 4. Синхронизационная машина Бернулли, 6400 входов и 128 выходов, 10 классов, 10 коэффициентов каждого класса, 10 квантователей с 50 входами 50 весами и одним выходом

Если требуется получить низкоразмерное решающее правило, то к входам квантователя добавляется несколько соседних усреднений. Очевидно, что тот же результат можно получить, если суммировать линейные отклики нескольких квантователей одной группой линейных откликов с соответствующей пороговой функцией.

Если требуется наоборот повысить размерность решающего правила, то выполняется нелинейное преобразование над входными данными. Это может быть умножение, сложение с нелинейностью, цифро-аналоговое преобразование, фильтрация и обратное аналого-цифровое преобразование и т.д.

При этом подъем вероятности кластеризации будет меньше, чем при добавлении новых данных в обучающую выборку. Причиной являются корреляционные связи, возникающие при нелинейном преобразовании над входами.

Качество работы машины оценивается через логарифм $-\log_{10}(P_{\text{синх}})$ вероятности синхронизации на выбранной голосовой фразе при заданном числе классов. При этом один квантователь может иметь входами несколько усреднений по времени. Также один квантователь может брать не все коэффициенты. Значения логарифма $-\log_{10}(P_{\text{синх}})$ вероятности синхронизации на голосовой фразе «Россия родина слонов» при трех классах 10 коэффициентах, приведены в таблице 1. Значения $-\log_{10}(P_{\text{синх}})$ при четырех и при пяти классах приведены в таблице 2 и 3.

Таблица 1 Значения логарифма $-\log_{10}(P_{\text{синх}})$ вероятности синхронизации на голосовой фразе «Россия родина слонов» при трех классах 10 коэффициентах

Число входов квантователя	Число квантователей				
	8	16	32	64	128
15	1,95	2,07	2,12	2,32	2,45

30	2,16	2,23	2,34	2,48	2,62
45	2,48	2,56	2,67	2,71	2,78
60	2,86	2,92	3,08	3,25	3,36
75	3,20	3,34	3,45	3,57	3,61

Таблица 2 Значения логарифма $-\log_{10}(P_{\text{синх}})$ вероятности синхронизации на голосовой фразе «Россия родина слонов» при четырех классах 10 коэффициентах

Число входов квантователя	Число квантователей				
	8	16	32	64	128
20	2,15	2,34	2,45	2,58	2,68
40	2,43	2,51	2,60	2,68	2,75
60	2,97	3,12	3,18	3,25	3,36
80	3,61	3,71	3,78	3,81	3,96
100	3,91	4,08	4,12	4,18	4,27

Таблица 3 Значения логарифма $-\log_{10}(P_{\text{синх}})$ вероятности синхронизации на голосовой фразе «Россия родина слонов» при пяти классах 10 коэффициентах

Число входов квантователя	Число квантователей				
	8	16	32	64	128
25	2,32	2,45	2,59	2,78	2,97
50	2,91	3,08	3,15	3,21	3,28
75	3,48	3,58	3,62	3,69	3,78
100	3,96	4,07	4,11	4,18	4,22
125	4,14	4,19	4,25	4,36	4,45

Отличием машины Бернулли от существующих методов кластеризации является их ориентированность на биометрию. Настройка квантователя Бернулли проводится по центрам фонемоподобных участков, которые являются оптимальными центрами классов. При этом разброс откликов квантователя по времени будет минимальным. Поэтому размерность входов машины может быть поднята с 500 для существующих методов до 1000, 2000, 3000 и выше в зависимости от требований заказчика.

ЛИТЕРАТУРА:

1. Язов Ю.К. и др. Нейросетевая защита персональных биометрических данных. //Ю.К.Язов (редактор и автор), соавторы В.И. Волчихин, А.И. Иванов, В.А. Фунтиков, И.Г. Назаров // М.: Радиотехника, 2012 г. 157 с.
2. Иванов А.И., Хальметова А.Н., Захаров О.С., Рыболовлев А.А., Рыжков А.П. Нейросетевой вокодер-архиватор, сохраняющий биометрические особенности голоса говорящего при высоком уровне сжатия шипящих звуков // Нейрокомпьютеры №3 за 2012 г.
3. И. И. Елисеева, М. М. Юзбашев. Общая теория статистики. Учебник. «Финансы и статистика», 2005
4. Надеев Д.Н. Аналитическое описание оценки вероятности появления коллизий у нейросетевого преобразователя биометрия-код «Нейрокомпьютеры: разработка, применение» №6, 2009 с. 53-55

Материал поступил 17.05.2014, опубликовано по положительной рецензии доктора технических наук Малыгина А.Ю.