УДК 621.39

А. Н. Ткаченко, к. т. н., доц.; О. Д. Феферман

МЕТОДЫ КЛАСТЕРИЗАЦИИ ДАННЫХ В ЦИФРОВОЙ ОБРАБОТКЕ РЕЧЕВЫХ СИГНАЛОВ

В статье рассмотрены основные методы кластеризации данных, проанализированы их преимущества и недостатки, сформулированы направления дальнейших исследований.

Ключевые слова: *речевые сигналы, сжатие, кластеризация.*

Вступление

При параметрическом кодировании речевого сигнала в современных системах цифровой связи вместо непосредственно коэффициентов линейного предсказания или линейных спектральных частот в канал связи передаются индексы этих коэффициентов в таблице – кодовой книге. На приемной стороне эти индексы декодируются с применением той же самой кодовой книги, а из полученных параметров синтезируется речевой сигнал.

В кодовой книге нельзя хранить все возможные значения параметров, так как в этом случае мы не получим большой степени сжатия. Потому в ней хранят только наиболее репрезентативные значения, то есть такие, с помощью которых можно закодировать весь спектр возможных значений параметров. Таким образом, при кодировании сигнала после получения определенного вектора параметров в кодовой книге производится поиск ближайшего (наиболее похожего) вектора, и в канал связи передается уже его индекс.

Поскольку необходимо, чтобы значения в кодовой книге были репрезентативными, создание кодовой книги для сжатия речевых сигналов происходит следующим образом: сначала подбирается репрезентативный фонетический материал, то есть такой, который содержит все возможные звуки, затем он кодируется и получается набор параметров для каждого фрейма, а уже с применением этого набора значений вычисляются значения для кодовой книги. Именно для расчета значений, хранящихся в кодовой книге, на основе имеющегося материала и применяются различные методы кластеризации.

Кодовые книги могут быть скалярными или векторными. Скалярная кодовая книга содержит набор значений для каждого элемента вектора параметров отдельно. Векторная кодовая книга содержит набор векторов параметров. Скалярные кодовые книги более просты в вычислениях и требуют меньше памяти для хранения. Вычислительная сложность и необходимый объем памяти при применении векторных кодовых книг намного больше по сравнению со скалярными, но они обеспечивают большую степень сжатия.

Различные методы кластеризации применяются как для скалярных, так и для векторных кодовых книг. Для того чтобы перейти непосредственно к рассмотрению методов кластеризации, сначала необходимо определить, что такое кластеризация.

Задача кластеризации при построении кодовых книг

Кластеризация – это разбиение определенного множества объектов на подмножества (кластеры), которые не пересекаются, таким образом, чтобы каждый кластер содержал похожие объекты, а объекты разных кластеров различались между собой [1].

Пусть X – множество объектов, X^m – конечная входящая выборка объектов, $X^m = \{x_1, x_2, ..., x_m\} \subset X$.

Задана определенная функция расстояния между объектами $d(x, x')$. Кластеризация, основываясь на входящей выборке объектов, определяет множество кластеров $Y = \{y_1, y_2, ..., y_n\}$.

При этом каждому $x_i \in X^m$ ставится в соответствие $y_j \in Y$. Каждый кластер состоит из объектов, близких по метрике d. Определяется функция $\alpha : X \to Y$, которая каждому $x \in X$ ставит в соответствие $y \in Y$.

В случае если объекты – числа либо векторы чисел, то кластеризацию называют квантованием.

Скалярный квантизатор Q размера N – это преобразование множества действительных чисел $x \in R$ в конечное множество Y, содержащее N значений, которые называются кодовыми векторами или центроидами: $Q : R \to Y$, де $(y_1, y_2, ..., y_n) \in Y$ [2].

Y называют кодовой книгой квантизатора. Операцию квантования можно также записать следующим образом:

$$Q(x) = y_i, x \in R, i = 1,...,N. \quad (1)$$

Кодовым регионом для скалярного квантизатора будем называть часть пространства действительных чисел, такую, что

$$R_i = \{x \in R, Q(x) = y_i\} = Q^{-1}(y_i), i = 1,2,...,N. \quad (2)$$

Также ясно, что:

$$i \neq j \Rightarrow R_i \cap R_j = \varnothing. \quad (3)$$

А также:

$$\bigcup_i R_i = R. \quad (4)$$

Кодовые регионы, которые имеют границы, называются ограниченными, а такие, которые не имеют границы – неограниченными.

В сфере обработки речевых сигналов операцию квантования можно рассматривать как совокупность операций кодирования (E) и декодирования (D):

$$Q(x) = y_i \Rightarrow E(x) = i, D(i) = y_i. \quad (5)$$

Квантизатор считается оптимальным, если он минимизирует искажение D:

$$D = \int_{-\infty}^{\infty} d(x, Q(x)) f_x(x) dx, \quad (6)$$

где d – функция расстояния, определенная на множестве действительных чисел.

Условие ближайшего соседа для оптимальности квантизатора формулируется так:

$$R_i = \{x : d(x, y_i) \leq d(x, y_j)\}, \forall j \neq i. \quad (7)$$

То есть для любого вектора, который попадает в кодовый регион, расстояние до соответствующего центроида должно быть меньше, чем расстояние до любого другого центроида в кодовой книге.

Так как $Q(x) = y_i$ только если $d(x, y_i) \leq d(x, y_j)$, то можно говорить, что:

$$d(x, Q(x)) = \min_i (d(x, y_i)). \quad (8)$$

Таким образом расстояние от определенного вектора до его квантованного значения – это минимальное значение из всех расстояний до каждого центроида в кодовой книге.

Условие оптимальности центроидов для кодового региона $R_i, i = 1,2,...,N$:

$$\sum d(x, y_i) \to \min, x \in R_i. \quad (9)$$

То есть центроид должен выбираться таким образом, чтобы он минимизировал искажение

для каждого вектора, который принадлежит данному кодовому региону. Данное условие должно выполняться для всех регионов кодовой книги.

Дадим определение векторному квантизатору. Векторный квантизатор Q размерности K и размера N – это преобразование K-мерного вектора x из Евклидового пространства R^K в конечный набор Y, который содержит N K-мерных векторов, которые называются кодовыми векторами или центроидами $Q: R^K \to Y$, де $x = [x_1, x_2, ..., x_K]$, $(y_1, y_2, ..., y_N) \in Y$, $y_i = [y_i^1, y_i^2, ..., y_i^K]$, $i = 1,2,...N$.

Условие оптимальности для векторного квантизатора определяется аналогично тому же условию что и для скалярного.

Четкая и нечеткая кластеризация. Метод k-средних

При четкой кластеризации объекты разделяются на отдельные кластеры таким образом, что каждый объект может принадлежать лишь одному кластеру. При нечеткой кластеризации объект может состоять в нескольких кластерах, при этом указывается степень его принадлежности для каждого из них, которая характеризует степень связи объекта с определенным кластером.

Примером четкого метода кластеризации является алгоритм k-средних [3] (k-means). Сначала выбирается количество кластеров N и проводится инициализация центроидов $y_1, y_2, ..., y_N$. Инициализация может проводиться как случайным образом или просто равномерным разбиением, так и с применением определенных более простых алгоритмов кластеризации. Далее проводится классификация входных значений с применением рассчитанных центроидов:

$$d(x, y_i) = \min_j (x, y_j) \Rightarrow Q(x) = y_i, j = 1,2,...,N. \quad (10)$$

На следующем шаге каждый центроид заменяется средним значением объектов, относящихся к соответствующему кластеру.

$$y_i = \frac{\sum_{k=1}^{K_i} x_k}{K_i}, Q(x_k) = y_i, i = 1,2,...,N, \quad (11)$$

где K_i – количество векторов в i-м кластере.

Данные шаги повторяются, пока на некотором шаге ни один центроид не будет изменен.

Примером нечеткого метода кластеризации является нечеткий метод c-средних (fuzzy c-means) [4]. Цель данного метода состоит в минимизации показателя J_m:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m d^2(x_i, c_j), 1 \le m < \infty, \quad (12)$$

где C – количество центроидов; N – количество входящих объектов; x_i – i-й входящий объект; c_j – j-й центроид; d – функция расстояния; u_{ij} – степень принадлежности объекта x_i j-му кластеру; m – так называемый параметр нечеткости;

Алгоритм работает следующим образом. На первом шаге проводится инициализация матрицы принадлежности векторов U. Номеру шага k присваивается начальное значение.

На каждом k-м шаге вычисляется значение центроидов:

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m}. \quad (13)$$

Дальше, после обновления центроидов, обновляется матрица принадлежности центроидов:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{d(x_i - c_j)}{d(x_i - c_k)} \right)^{\frac{2}{m-1}}}. \quad (14)$$

После этого вычисляется условие окончания работы алгоритма. Если $d(U^{(k+1)}, U^k) < \varepsilon$, то работа алгоритма заканчивается, иначе – происходит переход к шагу $k+1$. ε – положительное число от 0 до 1.

Алгоритмы нечеткой кластеризации, такие как c-means, редко применяются в задачах сжатия речевых сигналов, а чаще в задачах их распознавания, а также в задачах цифровой обработки изображений. Метод k-means достаточно широко применяется при сжатии речевых сигналов. В частности, его применяют для построения скалярных кодовых книг для вокодера CELP. Основным преимуществом данного метода является его простота и низкая вычислительная сложность. Недостатком является то, что результат его работы зависит от первоначального выбора центроидов и в общем случае метод не дает оптимальных результатов, а позволяет найти только субоптимальные решения.

Кластеризация с известным количеством кластеров. Метод LBG

Некоторые методы требуют предварительно задать количество кластеров, на которые необходимо разбить тренировочную последовательность. Таким, например, является метод k-средних, рассмотренный выше. Сначала задается число кластеров и центроидам присваиваются начальные значения, а после этого происходит оптимизация кодовой книги путем модификации центроидов и кодовых регионов. Существуют и другие методы, как например LBG (Linde, Buzo, Gray) [5]. В данном методе не нужно задавать число кластеров и присваивать начальные значения центроидов. В работе метода можно выделить несколько этапов.

Прежде всего проводится инициализация кодовой книги. Вычисляется среднее значение всех векторов:

$$c_1^* = \frac{1}{M} \sum_{m=1}^{M} x_m, \quad (15)$$

где c_1^* – начальный кодовый вектор, M – количество тренировочных векторов, x_m – тренировочный вектор.

Таким образом это значение и есть первым кодовым вектором, который содержится в кодовой книге. Количество векторов в кодовой книге устанавливается в $N = 1$. После этого рассчитывается значение среднеквадратического искажения:

$$D_{ave}^* = \frac{1}{Mk} \sum_{m=1}^{M} D(x_m, c_1^*), \quad (16)$$

где k – размерность векторов, а функция D определяется так:

$$D(x,c) = (x_1 - c_1)^2 + (x_2 - c_2)^2 + ... + (x_k - c_k)^2. \quad (17)$$

Следующий этап – это разделение векторов. То есть каждый вектор, который уже есть в кодовой книге, разделяется на 2 вектора таким образом:

Для $i = 1,2,...,N$:

$$c_i^{(0)} = (1+\varepsilon)c_i^*, \tag{18}$$

$$c_{N+i}^{(0)} = (1-\varepsilon)c_i^*. \tag{19}$$

Дальше выполняется ряд шагов, пока не будет получено желаемое количество векторов (LBG позволяет получить 2^n векторов). На каждом шаге сначала происходит обновления кодовых регионов. При этом для каждого тренировочного вектора находится кодовый регион, к которому он находится ближе всего и тренировочный вектор переносится в соответствующий кодовый регион. То есть, если для $m = 1,2,...M$ и $n = 1,2,...,N$:

$$\min(D(x_m, c_n^{(i)})) = D(x_m, c_{n^*}^{(i)}), \tag{20}$$

где i - номер итерации, то:

$$Q(x_m) = c_{n^*}^{(i)}. \tag{21}$$

После этого обновляются значения кодовых векторов. Каждому кодовому вектору присваивается среднее значение всех векторов, которые принадлежат соответствующему кодовому региону:

$$c_n^{(i+1)} = \frac{\sum_{Q(x_m)=c_n^{(i)}} x_m}{\sum_{Q(x_m)=c_n^{(i)}} 1}, \text{ n = 1, 2, …, N.} \tag{22}$$

После этого проводится расчет среднеквадратического искажения:

$$D_{ave}^{(i)} = \frac{1}{Mk}\sum_{m=1}^{M} D(x_m, Q(x_m)). \tag{23}$$

Далее выясняется, как это значение изменилось относительно предыдущего. Если необходимо, операция обновления кодовых регионов и центроидов повторяется, и только после этого происходит переход к следующему шагу.

Метод LBG является довольно распространенным в обработке речевых сигналов и изображений, в частности при сжатии речевых сигналов. Основными преимуществами данного метода является относительно невысокая вычислительная сложность, простота реализации, независимость результатов работы метода от исходных данных (поскольку центроиды формируются динамически на основе тренировочной последовательности и не нужно задавать им определенные начальные значения). Вообще метод дает неплохие результаты, но имеет и определенные недостатки, связанные со спецификой метода, а именно - неравномерное расположение центроидов в кодовой книге относительно распределения объектов тренировочной последовательности. Этот недостаток обусловлен тем, что значение каждого следующего центроида вычисляется на основе предыдущего, и таким образом невозможно кардинально изменить положение центроидов в случае необходимости, поскольку он может изменяться только в определенных пределах.

Раздельное и многоэтапное квантование

При применении векторных кодовых книг не всегда возможно хранить в кодовой книге векторы параметров такой же размерности, как и векторы тренировочной последовательности. Представление всех необходимых вариантов векторов в кодовой книге потребовало бы чрезвычайно больших затрат памяти, что затруднило бы реализацию таких

систем на персональных компьютерах и вообще почти невозможным бы их аппаратную реализацию.

Поэтому существует два различных подхода к применению векторных кодовых книг – раздельное (split) и многоэтапное (multistage) квантования.

При раздельном квантовании вектор параметров делится на подвекторы и каждый подвектор кодируется с помощью своей векторной кодовой книги. Векторная кодовая книга строится для каждого подвектора независимо с применением одного из алгоритмов кластеризации. При этом необходимо определить, на сколько векторов разбивать начальные векторы, сколько элементов должно быть в каждом подвекторе и сколько значений хранить в кодовой книге для каждого подвектора.

Многоэтапное квантование построено на совершенно иной идее. Основное отличие заключается в том, что в кодовой книге хранятся векторы, размерность которых совпадает с векторами кодовой последовательности. При этом данный метод создает несколько кодовых книг. Однако при квантизации определенного вектора он заменяется не просто вектором с одной из этих кодовых книг, а комбинацией векторов из разных кодовых книг, сформированных данным методом.

В кодере входящий вектор x сравнивается с вектором $\hat{x}$:

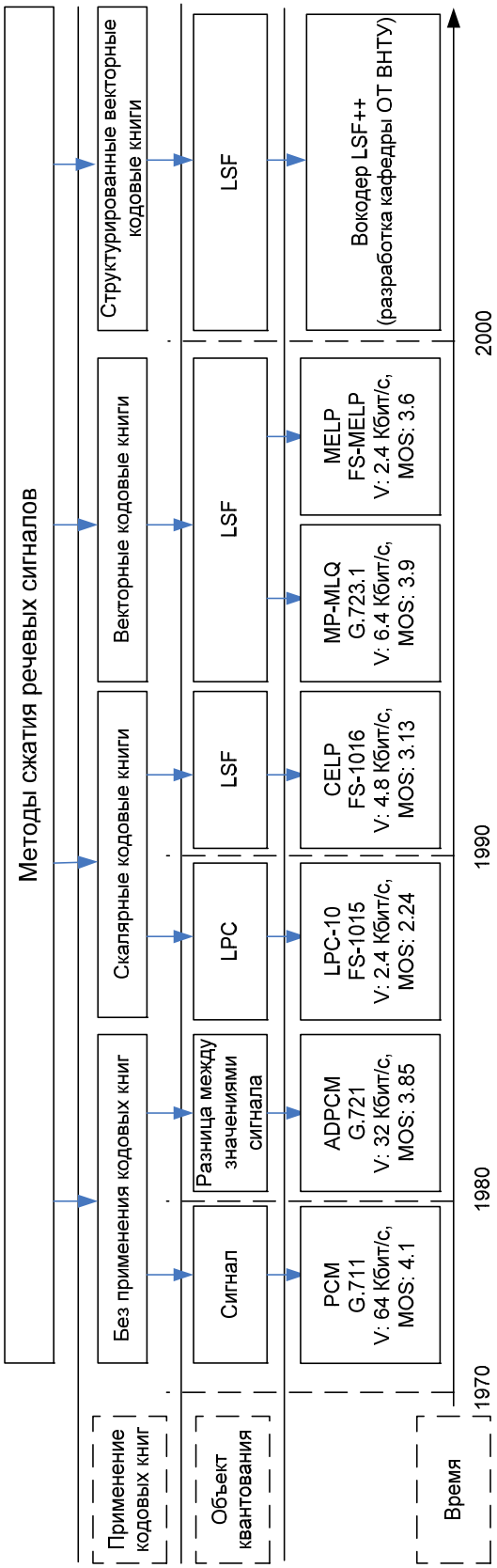
$$\hat{x} = y_{i1}^{(1)} + y_{i2}^{(2)} + ... + y_{iK}^{(K)}, \tag{24}$$

где $y_i^{(l)}$ – это i-й кодовый вектор из кодовой книги, получено на l-м этапе. Все кодовые книги имеют такую же самую размерность, как и входящий вектор.

Путем выбора разных индексов кодер старается минимизировать расстояние между x и $\hat{x}$. После нахождения такого набора $\{i_1, i_2, ..., i_K\}$, который минимизирует расстояние, индексы передаются на приемную сторону, где декодер с применением таких же кодовых книг восстанавливает сигнал.

Одно из основных преимуществ многоэтапного квантования по сравнению с раздельным – это уменьшение объемов памяти, необходимых для хранения кодовых книг. В плане вычислительной сложности многоэтапное квантование требует больше ресурсов по сравнению с раздельным.

Основной недостаток раздельного квантования – это то, что в процессе квантования может нарушаться устойчивость системы параметров. Между элементами квантованного вектора параметров есть определенные зависимости, которые в процессе квантования могут нарушаться. Соответственно это приводит к значительному искажению уплотненного сигнала. При многоэтапном квантовании этот недостаток отсутствует, поскольку векторы, хранящихся в кодовых книгах многоэтапного квантизатора имеют такую же размерность, как и тренировочные векторы, и упомянутые выше зависимости в них не нарушаются. Многоэтапное квантование применяется для построения векторных кодовых книг в алгоритме MELP.



* V – пропускная способность, необходимая для создания одностороннего речевого канала.
** MOS (Mean opinion score) – средняя субъективная оценка качества звука.

Рис. 1. Классификация методов сжатия речевых сигналов

Классификация методов сжатия речевых сигналов

Обычно вокодеры классифицируют только по одному параметру – скорости передачи речевого сигнала в канале связи. Так их делят на низкоскоростные, среднескоростные и высокоскоростные [6]. Однако такая классификация не дает возможности оценить закономерности развития методов сжатия речевых сигналов во временной перспективе.

Выше были рассмотрены различные методы кластеризации, которые могут быть применены при сжатии речевых сигналов. Поскольку в каждом из методов сжатия речевых сигналов применяется квантование в том или ином виде, то его можно выделить как один из признаков их классификации. В данной работе предлагается классификация методов сжатия речевых сигналов, приведенная на рис. 1.

Сначала методы делятся по типу применяемых кодовых книг, а именно на те, которые не применяют никаких кодовых книг, то есть просто квантуют значение по определенному закону, применяющие скалярные, и применяющие векторные кодовые книги. Далее методы делятся по объекту квантования, то есть по тому, что хранится в кодовой книге: методы, которые квантуют непосредственно сигнал, производные от сигнала (например, разницу предыдущего и следующего значения сигнала), коэффициенты линейного прогнозирования (LPC) и методы, квантующие линейные спектральные частоты (LSF).

Направления дальнейших исследований

Приведенная классификация позволяет проследить закономерности развития методов сжатия речи и спрогнозировать направления их дальнейшего развития. Поскольку в данной классификации четко прослеживается временная последовательность развития методов сжатия речевых сигналов, можно сделать выводы о возможных путях их дальнейшего развития.

1. Исследование и разработка общих принципов построения структурированных кодовых книг, которые должны заменить неструктурированные кодовые книги.
2. Разработка критериев оценки качества кластеризации для структурированных кодовых книг.
3. Разработка методов упорядочивания и поиска векторов в структурированных кодовых книгах.
4. Объектом квантования в векторных кодовых книгах могут оставаться линейные спектральные частоты, но перспективным представляется исследование альтернативных моделей представления спектральной информации о сигнале.

СПИСОК ЛІТЕРАТУРЫ

1. A. K. Jain. Data clustering: a review / A. K. Jain, M. N. Murty, P. J. Flynn.– ACM Comput. Surv.–1999. – №31. – 60 с.

2. J. Makhoul. Vector quantization in speech coding / J. Makhoul, S. Houcos, H. Gish . – Proc. IEEE. – 1985 – №73. – pp. 1551 – 1558.

3. P. Hedelin. Vector Quantization for Speech Transmission / P. Hedelin, P. Knagenhjelm, M. Skoglund. – Speech Coding and Synthesis. – 1995. – pp. 311 – 346.

4. J. C. Dunn A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact, Well Separated Clusters / J. C. Dunn. – Journal of Cybernetics. – 1973. – №3. – pp. 32 – 57.

5. Linde Y. An Algorithm for Vector Quantizer Design / Linde, Y., Buzo, A., Gray, R. – IEEE Transactions on Communications. – 1980. – №28., pp. 84 – 94.

6. Рабинер Л. Р. Цифровая обработка речевых сигналов: Пер. с англ. / Рабинер Л. Р., Шафер Р. В. – М.: Радио и связь, 1981. – 496 с.

Ткаченко Александр Николаевич – к. т. н.,доцент кафедры вычислительной техники**.**

Феферман Олег Дмитриевич – аспирант кафедры вычислительной техники.
Винницкий национальный технический университет.