УДК 519.651

КЛАСТЕРИЗАЦИЯ АВТОРЕГРЕССИОННЫХ МОДЕЛЕЙ РЕЧЕВЫХ СИГНАЛОВ ПО КРИТЕРИЮ МИНИМУМА ИНФОРМАЦИОННОГО РАССОГЛАСОВАНИЯ КУЛЬБАКА — ЛЕЙБЛЕРА

И. В. Губочкин,
канд. техн. наук, доцент
Нижегородский государственный лингвистический университет им. Н. А. Добролюбова
Н. В. Карпов,
канд. техн. наук, доцент
Национальный исследовательский университет «Высшая школа экономики», г. Нижний Новгород

Решается задача кластеризации множества авторегрессионных моделей речевых сигналов в рамках теоретико-информационного подхода. Для этого был разработан алгоритм нахождения оптимальных параметров авторегрессионной модели в смысле минимума информационного рассогласования Кульбака — Лейблера. На его основе проведена модификация известного алгоритма кластеризации k-средних. Экспериментально исследована эффективность применения разработанных алгоритмов при диктороннезависимом распознавании изолированных слов с использованием аппарата скрытых марковских моделей с дискретным распределением вероятностей наблюдений. Установлено, что наилучшие результаты по точности распознавания достигаются при использовании коэффициентов линейного предсказания с неравномерным частотным разрешением в качестве вектора признаков и размере кодовой книги векторного квантователя, равном 256.

Ключевые слова *— автоматическое распознавание речи, авторегрессионная модель, информационное рассогласование, центроид, кластер.*

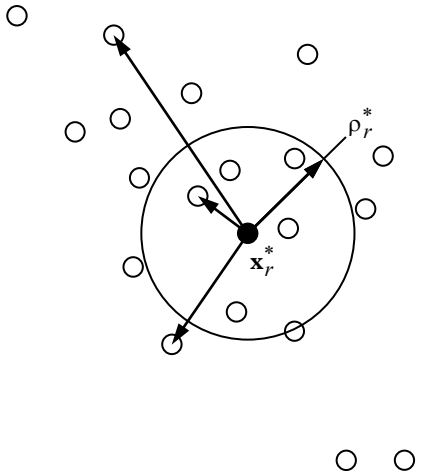
Введение

В исследованиях по информационной теории восприятия речи (ИТВР) [1–5] предложены подходы к решению задач анализа, распознавания и обработки речевых сигналов в рамках теоретико-информационного подхода. Данная тематика исследований является весьма актуальной по причине широкого распространения в последнее время теоретико-информационного подхода в теории распознавания образов. Так, в работе [6] проводится оценка и дается обоснование возможности применения указанного подхода для кластеризации данных. В работах [7, 8] рассматриваются примеры использования различных видов информационных метрик при решении задач обработки изображений. Работы [9–11] посвящены применению теоретико-информационного подхода и информационной геометрии в различных методах машинного обучения.

В связи с этим вызывает интерес адаптация подходов, представленных в ИТВР, к использованию в уже существующих методах машинного обучения и распознавания образов применительно к задаче кластеризации и обработки речевых сигналов. Для этого рассмотрим вначале основные положения данной теории.

В рамках ИТВР элементарная речевая единица (ЭРЕ) задается некоторым информационным центром-эталоном, в качестве которого выбирается реализация речевого сигнала $\mathbf{x}_r^* \in \{\mathbf{x}\}_r$, $r = \overline{1,R}$, представленная соответствующей авторегрессионной (АР) моделью и характеризующаяся минимальной суммой информационных рассогласований в метрике Кульбака — Лейблера [1, 5, 12] относительно всех других реализаций данной ЭРЕ:

$$\mathbf{x}_r^* = \arg\min_k \sum_{\ell=1}^{L_r} \rho_{\ell,k}, k = \overline{1,L_r}, \qquad (1)$$



■ ***Рис. 1***. Модель ЭРЕ

где L_r — число реализаций r-й ЭРЕ; $\rho_{l,k}$ — информационное рассогласование по Кульбаку — Лейблеру между l-й и k-й ЭРЕ. Иллюстрирует сформулированное выше определение модели ЭРЕ [1] рис. 1.

В приведенной формулировке модели ЭРЕ есть два недостатка. Первый заключается в том, что выбор информационного центра-эталона делается из дискретного множества реализаций. Это значит, что критерий (1) не является в строгом смысле оптимальным. Второй недостаток состоит в переборном характере алгоритма поиска информационного центра-эталона, сложность которого составляет $O(L_r^2)$, т. е. количество необходимых вычислений будет быстро возрастать с увеличением множества реализаций заданной ЭРЕ. В связи с этим представляется актуальным создание алгоритмов, свободных от указанных недостатков.

Постановка задачи нахождения оптимальной авторегрессионной модели

Согласно работе [2], информационное рассогласование по Кульбаку — Лейблеру между неизвестным сигналом **x** и эталоном **r**, заданными их АР-моделями, определяется в спектральной области следующим образом:

$$\rho_{x,r} = \frac{1}{F}\sum_{f=1}^{F}\frac{\left|1+\sum_{m=1}^{P}a_r(m)e^{-j\pi mf/F}\right|^2}{\left|1+\sum_{m=1}^{P}a_x(m)e^{-j\pi mf/F}\right|^2}-1. \quad (2)$$

Здесь P — порядок АР-модели; $a_r(m)$ и $a_x(m)$ — элементы векторов авторегрессии сигналов **r** и **x** соответственно; F — верхняя граница частотного диапазона. Можно показать [13], что $\rho_{x,r}\geq 0$ для любых АР-моделей $a_r(m)$ и $a_x(m)$, если их полюсы находятся внутри единичной окружности на комплексной плоскости.

Отметим также, что информационное рассогласование Кульбака — Лейблера является частным случаем рассогласования Брэгмана [14], определяемого между двумя функциями плотности распределения вероятностей $p(x)$ и $q(x)$ как

$$D_F(p\|q) = F(p)-F(q)- \\ -\int\frac{\mathrm{d}F(q)}{\mathrm{d}q}\big(q(x)-p(x)\big)\mathrm{d}x, \quad (3)$$

где $F(\bullet)$ — производящая функция, обладающая свойствами выпуклости и дифференцируемости. Собственно само информационное рассогласование Кульбака — Лейблера легко получить из (3), выбрав в качестве производящей функции негэнтропию Шеннона $F(x)=\int x\log x\mathrm{d}x$. Отсюда следует, что информационное рассогласование (2) также относится к классу рассогласований Брэгмана. Приведенное замечание будет использовано далее при доказательстве сходимости модифицированного алгоритма кластеризации k-средних.

Определим теперь информационное рассогласование Кульбака — Лейблера в случае сравнения эталонного сигнала, заданного его АР-моделью, сразу с множеством реализаций r-й ЭРЕ $\{\mathbf{x}\}_r$ как величину среднего искажения:

$$\rho_{\{\mathbf{x}\}_r,r} = \frac{1}{L_r F}\times \\ \times\sum_{\ell=1}^{L_r}\sum_{f=1}^{F}\frac{\left|1+\sum_{m=1}^{P}a_r(m)e^{-j\pi mf/F}\right|^2}{\left|1+\sum_{m=1}^{P}a_{x,\ell}(m)e^{-j\pi mf/F}\right|^2}-1, \quad (4)$$

где $a_{x,l}(m)$ — элементы вектора авторегрессии l-го сигнала из множества $\{\mathbf{x}\}_r$. Вид данной формулы вытекает из определения центроида множества.

Центроидом множества $\{\mathbf{q}\}_1^L=\{q_i,\ i=\overline{1,L}\}$ является такой вектор **y**, который минимизирует среднее искажение:

$$\mathbf{y} = \arg\min_{\mathbf{y}}\frac{1}{L}\sum_{i=1}^{L}d(\mathbf{q}_i,\mathbf{y}), \quad (5)$$

где d обозначает некоторую меру расстояния между двумя векторами, называемую также мерой искажений [15]. Формула (5) во многом похожа на критерий (1), за исключением того, что получаемый вектор **y** не обязан соответствовать какому-либо конкретному элементу множества $\{\mathbf{q}\}$.

Задача поиска оптимальной АР-модели r-й ЭРЕ состоит в выборе такого вектора АР-коэффициентов $\mathbf{a}_r^*$, при котором величина $\rho_{\{\mathbf{x}\}_r,r}$ стремится к своему глобальному минимуму:

$$\rho_{\{\mathbf{x}\}_r,r}\Big|_{\mathbf{a}_r^*}\to\min. \quad (6)$$

Из (5) нетрудно видеть, что решение поставленной задачи в формулировке (4), (6) фактически сводится к поиску АР-модели центроида множества $\mathrm{centr}\left(\{\mathbf{x}\}_r\right) \to \mathbf{a}_r^*$.

Поскольку применяемая в данной работе мера расстояния между векторами (2) с учетом свойств рассогласования Кульбака — Лейблера не является симметричной, то, согласно работе [7], формула (5) определяет «правосторонний» центроид. Выбор центроида данного типа обусловлен возможностью получить эффективный алгоритм его вычисления, описание которого приводится далее.

Синтез алгоритма

Найдем решение задачи (6). Для этого нам необходимо решить относительно $\mathbf{a}_r$ простую систему дифференциальных уравнений

$$\frac{\partial \rho_{\{\mathbf{x}\}_r,r}}{\partial a_r(m)} = 0,\ m = \overline{1,P}. \tag{7}$$

Получим выражение для частной производной $\frac{\partial \rho_{\{\mathbf{x}\}_r,r}}{\partial a_r(m)}$. Для этого определим две функции:

$$N_r(f) = \left|1 + \sum\nolimits_{m=1}^{P} a_r(m) e^{-j\pi mf/F}\right|^2;$$

$$D_{x,\ell}(f) = \left|1 + \sum\nolimits_{m=1}^{P} a_{x,\ell}(m) e^{-j\pi mf/F}\right|^2.$$

Тогда формулу (4) можно переписать следующим образом:

$$\rho_{\{\mathbf{x}\}_r,r} = \frac{1}{L_r F} \sum_{\ell=1}^{L_r} \sum_{f=1}^{F} \frac{N_r(f)}{D_{x,\ell}(f)} - 1. \tag{8}$$

Выражение (8) легко преобразовать к матричному виду, определив такую матрицу $\mathbf{G}_{\{\mathbf{x}\}_r,r}$ и вектор-строку $\mathbf{N}_r$, что

$$\mathbf{N}_r = \left[N_r(f)\right];$$

$$\mathbf{G}_{\{\mathbf{x}\}_r,r} = \begin{bmatrix} \frac{1}{D_{x,1}(1)} & \cdots & \frac{1}{D_{x,1}(F)} \\ \vdots & \ddots & \vdots \\ \frac{1}{D_{x,L_r}(1)} & \cdots & \frac{1}{D_{x,L_r}(F)} \end{bmatrix}.$$

Отсюда получаем

$$\rho_{\{\mathbf{x}\}_r,r} = \frac{1}{L_r F} \mathbf{I}\left(\mathbf{G}_{\{\mathbf{x}\}_r,r} \cdot \mathbf{N}_r^{\mathrm{T}}\right) - 1, \tag{9}$$

где $\mathbf{I}$ — единичный вектор-строка размера $1 \times L_r$.

Можно показать [16], что частная производная $\frac{\partial N_r(f)}{\partial a_r(m)}$ определяется как

$$\frac{\partial N_r(f)}{\partial a_r(m)} = 2\left(a_r(m) + \sum_{\substack{n=1 \\ n \neq m}}^{P} a_r(n)\Xi(m,\ n,\ f) + \cos\left(\frac{\pi mf}{F}\right)\right), \tag{10}$$

$$\Xi(m,\ n,\ f) = \cos\left(\frac{\pi mf}{F}\right)\cos\left(\frac{\pi nf}{F}\right) + \sin\left(\frac{\pi mf}{F}\right)\sin\left(\frac{\pi nf}{F}\right) = \cos\left(\frac{\pi(m-n)f}{F}\right).$$

Тогда выражение для $\frac{\partial \rho_{\{\mathbf{x}\}_r,r}}{\partial a_r(m)}$ с учетом (8) и (10) приобретает вид

$$\frac{\partial \rho_{\{\mathbf{x}\}_r,r}}{\partial a_r(m)} = \frac{1}{L_r F} \sum_{\ell=1}^{L_r} \sum_{f=1}^{F} \frac{1}{D_{x,\ell}(f)} \frac{\partial N_r(f)}{\partial a_r(m)}.$$

Отсюда легко видеть, что уравнение (7) после группировки множителей будет представлено следующим образом:

$$a_m \sum_{\ell=1}^{L_r} \sum_{f=1}^{F} \frac{1}{D_{x,\ell}(f)} + \sum_{\substack{n=1, \\ n \neq m}}^{P} a_n \sum_{\ell=1}^{L_r} \sum_{f=1}^{F} \frac{\Xi(m,n,f)}{D_{x,\ell}(f)} + \sum_{\ell=1}^{L_r} \sum_{f=1}^{F} \frac{\cos(\pi mf/F)}{D_{x,\ell}(f)} = 0,$$

$$m = \overline{1,P}.$$

В этом случае решение уравнения (7) относительно $\mathbf{a}_r$ может быть представлено как система линейных уравнений вида

$$\mathbf{C}\mathbf{a}_r = -\mathbf{b}, \tag{11}$$

где $\mathbf{b}$ — вектор-столбец, элементы которого определяются как

$$b_m = \sum_{\ell=1}^{L_r} \sum_{f=1}^{F} \frac{\cos(\pi mf/F)}{D_{x,\ell}(f)},\ m = \overline{1,P}; \tag{12}$$

$\mathbf{C}$ — квадратная матрица размера $P \times P$, элементы которой задаются следующим выражением:

$$C_{m,n} = \begin{cases} \sum\limits_{\ell=1}^{L_r} \sum\limits_{f=1}^{F} \frac{1}{D_{x,\ell}(f)},\ m = n; \\ \sum\limits_{\ell=1}^{L_r} \sum\limits_{f=1}^{F} \frac{\Xi(m,\ n,\ f)}{D_{x,\ell}(f)},\ m \neq n, \end{cases} \tag{13}$$

$$m,\ n = \overline{1,\ P}.$$

Преобразовав выражения (12) и (13) в матричный вид, получаем

$$\mathbf{b}_m = \mathbf{I}\left(\mathbf{G}_{\{\mathbf{x}\}_r,r} \cdot \mathbf{S}_m^{\mathrm{T}}\right);$$

$$\mathbf{C}_{m,n} = \begin{cases} \mathbf{I}\left(\mathbf{G}_{\{\mathbf{x}\}_r,r} \cdot \mathbf{E}\right), & m = n; \\ \mathbf{I}\left(\mathbf{G}_{\{\mathbf{x}\}_r,r} \cdot \mathbf{\Xi}_{m,n}^{\mathrm{T}}\right), & m \neq n, \end{cases} \qquad (14)$$

где $\mathbf{E}$ — матрица размера $F \times L_r$, состоящая из единичных элементов, а $\mathbf{S}_m$ и $\mathbf{\Xi}_{m,n}$ — векторы-строки, которые определяются как

$$\mathbf{S}_m = \left[\cos\left(\frac{\pi m f}{F}\right)\right]; \ \mathbf{\Xi}_{m,n} = \left[\Xi(m,n,f)\right].$$

Интересной особенностью уравнения (11) является то, что оно по своей структуре сходно с известными уравнениями Юла — Уолкера [17], для которых существует быстрый алгоритм решения. В матричной форме данные уравнения задаются в виде

$$\begin{bmatrix} r_1 & \overline{r_2} & \cdots & \overline{r_P} \\ r_2 & r_1 & \cdots & \overline{r_{P-1}} \\ \vdots & \vdots & \ddots & \vdots \\ r_P & r_{P-1} & \cdots & r_1 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} -r_2 \\ -r_3 \\ \vdots \\ -r_{P+1} \end{bmatrix}, \qquad (15)$$

где r_i — соответствующий элемент некоторого вектора $\mathbf{r}$ размерности $P+1$ (в оригинале $\mathbf{r}$ — вектор автокорреляции), а обозначение $\overline{r_i}$ является операцией комплексного сопряжения. Для того чтобы свести уравнение (11) к виду (15), необходимо задать вектор $\mathbf{r}$ в виде $\mathbf{r}=[\mathbf{C}_1\ b_P]$, где $\mathbf{C}_1$ — первая строка матрицы $\mathbf{C}$.

Для быстрого решения (15) обычно применяется рекуррентный алгоритм Левинсона — Дарбина [18], шаги которого приведены ниже:

$$E^{(0)} = r_1;$$

$$k_i = \left\{ r_{i+1} - \sum_{j=2}^{i} \alpha_j^{(i-1)} r_{i-j} \right\} \Big/ E^{(i-1)}, \ 1 \le i \le P; \qquad (16)$$

$$\alpha_i^{(i)} = k_i; \ \alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)};$$

$$E^{(i)} = \left(1 - k_i^2\right) E^{(i-1)};$$

$$a_m = \alpha_m^{(P)}, \ 1 \le m \le P.$$

Результатом его работы является вычисление вектора АР-коэффициентов без необходимости непосредственного обращения автокорреляционной матрицы.

Вычисления по алгоритму (11) — (16) позволяют получить значения коэффициентов АР-модели ЭРЕ $\mathbf{a}_r^*$, которые являются оптимальным решением задачи (6). Особенностью данного решения является то, что оно всегда будет оптимальным в глобальном смысле, поскольку величина среднего информационного рассогласования в виде (4) является квадратичной формой относительно $\mathbf{a}_r$. Также легко видеть, что предложенный алгоритм имеет линейную сложность $O(L_r)$, в отличие от критерия (1).

Модифицированный алгоритм *k*-средних

Наглядным примером практического применения полученного алгоритма (11) — (16) решения задачи (6) может служить использование его при построении алгоритмов кластеризации без учителя. Одним из наиболее известных алгоритмов такого типа является алгоритм *k*-средних [19, 20]. В общем виде он может быть задан следующим образом [21].

Пусть мы имеем некоторую случайную величину в пространстве наблюдений $\aleph$ такую, что $X : \mathbf{x} \in \aleph = \Re^d$, где $\Re^d$ — d-мерное евклидово пространство. Нас интересует возможность разбиения пространства $\aleph$ на Γ кластеров. Алгоритм *k*-средних предполагает, что число кластеров Γ заранее известно, и требуется найти такую матрицу параметров $\mathbf{\Phi}$, которая бы минимизировала целевую функцию (ошибку квантования), заданную следующим выражением:

$$E(X,\mathbf{\Phi}) \approx E_{KM}\left(\{\mathbf{x}\}_1^L, \mathbf{\Phi}\right) =$$

$$= \frac{1}{L}\sum_{\ell=1}^{L} \min_{\gamma} d\left(x_\ell, \varphi_\gamma\right) = \frac{1}{L}\sum_{\ell=1}^{L} \min_{\gamma} \left\|\mathbf{x}_\ell - \varphi_\gamma\right\|^2. \qquad (17)$$

Здесь $\{\mathbf{x}\}_1^L$ — множество векторов наблюдений; φ_γ — γ-й столбец матрицы $\mathbf{\Phi}$, который представляет собой вектор параметров, связанный с кластером γ.

При кластеризации по алгоритму *k*-средних вектор параметров φ_γ представляет собой обычное среднее значение всех векторов наблюдений, входящих в кластер γ. В этом случае мы можем определить матрицу средних значений $\mathbf{M}$, каждый γ-й столбец которой является вектором параметров φ_γ. Отсюда можно записать, что

$$\varphi_\gamma = \hat{\boldsymbol{\mu}}_\gamma; \ \mathbf{\Phi} = \mathbf{M},$$

где $\hat{\boldsymbol{\mu}}_\gamma$ — оценка среднего значения элементов γ-го кластера.

Для случая, когда наблюдения представлены в виде векторов авторегрессии $\{\mathbf{a}\}_1^L$ в метрике (2), необходимо внести в рассматриваемый алгоритм кластеризации изменения, касающиеся целевой функции (17) и меры искажений d. Возможность таких изменений связана с тем, что алгоритм *k*-средних может использоваться с широким

классом мер искажений, включая меры, не являющиеся метрическими [22]. Сходимость рассматриваемого алгоритма гарантируется для любых мер искажений, которые относятся к классу информационных рассогласований Брэгмана [6]. Как было отмечено выше, информационное рассогласование (2) также относится к данному классу. Из сказанного следует, что алгоритм k-средних сходится при использовании (2) в качестве меры искажений.

Ниже представлены шаги модифицированного алгоритма.

1. Выбрать число кластеров Г, инициализировать оценки центроидов $\mathbf{a}_{\gamma}^{*(k)}$, $k=0$, по каждому кластеру $\aleph_{\gamma}^{(k)}$, $k=0$, используя значения, полученные на основе априорных данных, или случайные значения. Затем на основе этих параметров, обозначенных как $\varphi_{\gamma}^{(k)}$, $k=0$, формируется матрица $\mathbf{\Phi}^{(k)}$, $k=0$.

2. С учетом текущих определений кластеров $\aleph_{\gamma}^{(k)}$ распределить по каждому из них имеющиеся АР-модели векторов наблюдений $\mathbf{a}_{\ell}$, $\ell=\overline{1, L}$, используя следующую индексную функцию принадлежности:

$$\hat{\gamma}_{\ell k}=\wp\left(\mathbf{a}_{\ell},\, \mathbf{\Phi}^{(k)}\right)=\arg\min_{\gamma}\rho_{\ell,\gamma}.$$

Вычисление значений информационного рассогласования $\rho_{l,\gamma}$ в формулировке (2) можно также выполнять в матричном виде аналогично (9) и (14).

3. Вычислить целевую функцию с учетом распределения наблюдений по кластерам

$$E_{KM}^{(k)}=E_{KM}\left(\{\mathbf{a}\}_1^L,\, \mathbf{\Phi}^{(k)}\right)=\frac{1}{L}\sum_{\ell=1}^{L}\rho_{\ell,\hat{\gamma}_{\ell k}}. \qquad (18)$$

4. Вычислить изменение целевой функции

$$\delta^{(k)}=E_{KM}^{(k)}-E_{KM}^{(k-1)}.$$

Алгоритм завершает свою работу, если выполняется условие $\left(\left(k>0\wedge\delta^{(k)}\le\delta_{\min}\right)\vee\left(k>1\wedge\left|\delta^{(k)}-\right.\right.$ $\left.\left.-\delta^{(k-1)}\right|<\varepsilon\right)\right)$ или $k>k_{\max}$.

5. На основе нового распределения векторов наблюдений по кластерам $\aleph_{\gamma}^{(k)}$ вычислить значения $\mathbf{a}_{\gamma}^{*(k+1)}$, используя алгоритм (11)–(16). Из полученных векторов сформировать матрицу параметров $\mathbf{\Phi}^{(k+1)}$.

6. Увеличить номер итерации k и повторить вычисления, начиная с шага 2.

Алгоритм k-средних реализует в себе метод наискорейшего спуска вдоль вектора градиента ошибки квантования (18) [23]. Из этого следует, что на каждой последующей итерации алгоритма значение целевой функции должно уменьшаться. Еще одним свойством данного алгоритма является уменьшение величины ошибки квантования при увеличении числа кластеров.

Разработанные выше алгоритмы могут использоваться в различных областях, в частности, в распознавании речевых сигналов. Результаты такого применения приводятся в следующем разделе.

Результаты экспериментальных исследований

Для проверки эффективности разработанной модификации алгоритма k-средних были проведены его экспериментальные исследования в рамках задачи распознавания изолированных слов. Эксперимент проводился с использованием речевой базы[1], состоящей из $R=11$ слов английского языка: «one», «two»,..., «nine», «zero», «o». Каждое слово проговаривалось в среднем по 2 раза группой из 208 дикторов. Представленная в базе речь хранится в виде соответствующих звуковых файлов формата PCM WAVE с частотой дискретизации 8 кГц, 16 бит. Данные файлы разделены на обучающее и тестовое множество. Обучающее множество содержит речь 95 дикторов (38 мужчин и 57 женщин), всего по 188 реализаций каждого слова. Тестовое множество содержит речь 113 дикторов (56 мужчин и 57 женщин), всего по 225 реализаций каждого слова. Следует отметить, что обучающее и тестовое множества не пересекаются друг с другом по дикторам. Несмотря на некоторую несбалансированность обучающего множества по числу мужчин и женщин, можно говорить о том, что применяемая речевая база в целом является достаточно представительной.

В ходе экспериментальных исследований все реализации слов разбивались на квазистационарные сегменты длительностью 20 мс с перекрытием смежных сегментов в 10 мс. Далее вычислялись векторы признаков размерности $P=12$, описывающих соответствующие сегменты. Для сравнения использовались четыре наиболее широко распространенных вида векторов признаков:

1) коэффициенты линейного предсказания (LPC) [24], которые являются эквивалентом рас-

[1] Обучающее и тестовое множества речевой базы английских числительных доступны для скачивания в сети Интернет по следующим ссылкам: http://cronos.rutgers.edu/~lrr/speech%20recognition%20course/databases/isolated_digits_ti_train_endpt.zip
http://cronos.rutgers.edu/~lrr/speech%20recognition%20course/databases/isolated_digits_ti_test_endpt.zip

сматриваемых в данной работе коэффициентов авторегрессии **a**;

2) кепстральные коэффициенты, вычисленные по рекуррентной формуле из коэффициентов линейного предсказания (CC-LPC) [18]:

$$c_m = a_m + \sum_{k=1}^{m-1}\left(\frac{k}{m}\right)c_k a_{m-k}, \quad 1 \le m \le P;$$

$$c_m = \sum_{k=1}^{m-1}\left(\frac{k}{m}\right)c_k a_{m-k}, \qquad m \ge P; \qquad (19)$$

3) коэффициенты линейного предсказания с неравномерным частотным разрешением (WLPC) [25]. Для их вычисления вектор коэффициентов автокорреляции **r** пропускается через набор всепропускающих фильтров первого порядка следующего вида:

$$D(z) = \frac{z^{-1} - \psi}{1 - \psi z^{-1}}.$$

Здесь $-1<\psi<1$ — коэффициент деформации. Параметр ψ выбирается таким образом, чтобы получаемая частотная шкала была близка к шкале барк, а сам параметр может быть приближенно рассчитан по следующей формуле:

$$\psi \approx 1{,}0674\left(\frac{2}{\pi}\tan^{-1}\left(0{,}06583 f_s / 1000\right)\right)^{1/2} - 0{,}1916,$$

где f_s — частота дискретизации, Гц. В дальнейшем используется автокорреляционный метод расчета коэффициентов линейного предсказания (16);

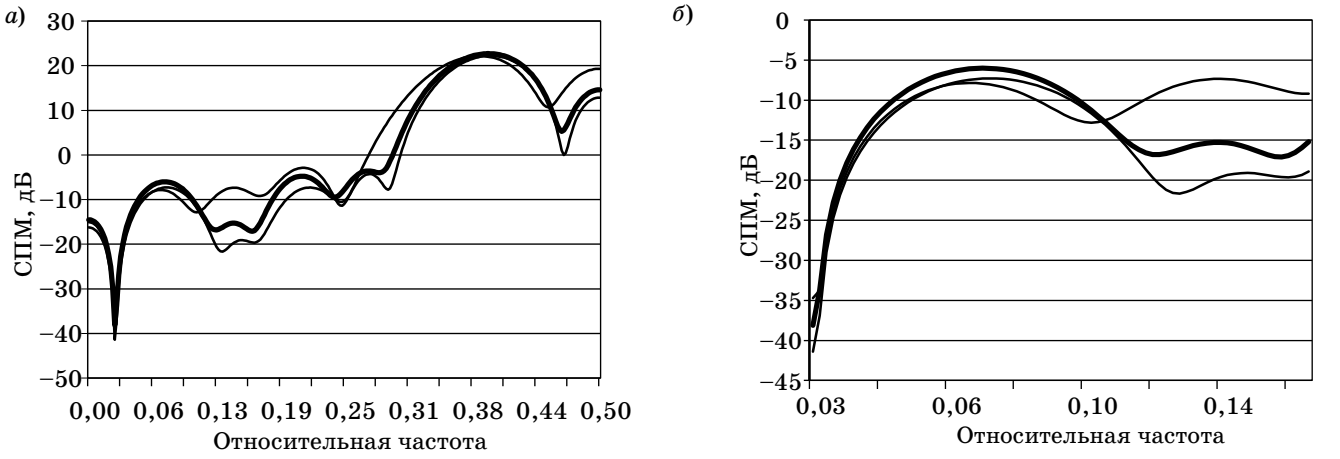
4) кепстральные коэффициенты, рассчитанные по коэффициентам линейного предсказания с неравномерным частотным разрешением (CC-WLPC). Для этого также применялась формула (19).

В качестве меры расстояния между векторами признаков типа LPC и WLPC использовалось информационное рассогласование в виде (2). Вместе с тем в качестве меры расстояния при использовании CC-LPC и CC-WLPC была выбрана евклидова метрика.

На подготовительном этапе из сегментов, полученных из обучающего множества слов, были сформированы кластеры с помощью алгоритма k-средних. При этом производилось несколько запусков алгоритма с различными начальными условиями для нахождения оптимального разбиения. Таким образом, для каждого значения числа кластеров Г от 8 до 1024 было найдено свое разбиение исходного множества по кластерам и их центры, которые будем называть кодовой книгой $W = \{w_1, \ldots, w_\Gamma\}$. Следует отметить, что построение кодовой книги для признаков LPC и WLPC выполнялось с помощью модифицированного алгоритма k-средних с использованием алгоритма (11)–(16) для вычисления центров кластеров.

Для иллюстрации свойств получаемой по алгоритму (11)–(16) оптимальной АР-модели на рис. 2, *а* представлены графики функции $N_r(f)$ для двух реализаций английской фонемы [*uh*] и полученной на их основе оптимальной АР-модели ЭРЕ. Данная функция может рассматриваться как спектральная плотность мощности (СПМ) нерекурсивного фильтра, коэффициенты которого задаются вектором авторегрессии **a**.

Здесь тонкими сплошными линиями показаны СПМ выбранных реализаций, а полужирная линия соответствует СПМ, найденной с помощью алгоритма (11)–(16) оптимальной АР-модели. Видно, что полученная результирующая модель учитывает особенности обеих реализаций фонемы [*uh*]. Дополнительно на рис. 2, *б* показан уве-



■ ***Рис. 2.*** График (*а*) и фрагмент (*б*) СПМ оптимальной АР-модели ЭРЕ

личенный фрагмент СПМ всех трех моделей в интервале относительной частоты 0,03...0,12.

Как можно видеть, СПМ оптимальной модели в районе 0,06 проходит над СПМ исходных АР-моделей и не является их простой комбинацией.

Для обеспечения возможности применять полученную кодовую книгу в задаче распознавания речи каждому ее элементу был сопоставлен символ из некоторого алфавита $V=\{\upsilon_1, \upsilon_2, ..., \upsilon_\Gamma\}$. Далее было проведено векторное квантование последовательностей признаков по всем реализациям каждого слова из обучающего множества. При этом для каждого слова было сформировано множество последовательностей наблюдений $\mathbf{O}_r=\left\{\mathbf{O}_r^1, \mathbf{O}_r^2, ..., \mathbf{O}_r^{L_r}\right\}$, $r=\overline{1, R}$, элементы которого представляют собой последовательности символов из алфавита V, полученных в результате выполнения векторного квантования.

Полный набор таких последовательностей образует обучающее множество для настройки скрытой марковской модели (СММ) с дискретной плотностью наблюдений [18]. При этом вычисляются оптимальные параметры СММ $\lambda=(\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$ для заданной обучающей выборки.

Оптимальными параметрами СММ называются те, которые максимизируют вероятность $p(\mathbf{O}|\lambda)$ по всем возможным последовательностям $\mathbf{O}=\{\mathbf{O}_1, ..., \mathbf{O}_L\}$ из обучающей выборки. Если обозначить q_t состояние в момент времени t, то $\mathbf{A}=\{a_{ij}\}=\{p(q_t+1=S_i|q_t=S_j)\}$ — матрица вероятностей переходов, содержащая вероятность перехода из состояния i в состояние j; $\mathbf{B}=\{b_j(k)\}=\{p(\upsilon_k|q_t=S_j)\}$ — матрица распределения вероятностей наблюдения символа υ_k в состоянии j в момент времени t, а $\boldsymbol{\pi}=\{\pi_i\}=\{p(q_1=S_i)\}$ — начальное распределение вероятностей состояний.

В приводимом эксперименте использовался набор из лево-правых СММ (или моделей Бакиса) λ_r, $r=\overline{1, R}$ с семью состояниями для каждого из R слов. Выбор указанного числа состояний основывается на ранее проведенных исследованиях (см., например [18, с. 380]), в которых показано, что для систем автоматического распознавания речи с малым словарем хорошие результаты распознавания могут быть получены при числе состояний, находящемся в диапазоне $6 \div 8$ и одинаковом для всех СММ. Тем не менее, в настоящее время разработаны различные методы оптимизации структуры СММ (числа состояний и переходов между ними), которые позволяют во многих случаях снизить вероятность ошибок распознавания. Однако рассмотрение указанных методов находится за рамками данной статьи, а заинтересованный читатель может обратиться к работам [26—29] для получения подробной информации.

Найденная кодовая книга W и модели λ_r на следующем этапе использовались для распознавания слов из тестового множества. Для этого аналогичным образом слова сегментировались, признаки, выделенные из сегментов, квантовались с использованием кодовой книги W и, исходя из получившейся последовательности наблюдений $\mathbf{O}_v$, при помощи алгоритма Витерби [18, 30] для каждой СММ λ_r вычислялись оптимальные последовательности состояний $Q=q_1...q_T$, максимизирующих правдоподобие $L_r=\log p_r(Q|\mathbf{O}, \lambda_r)$. Решение о том, какое слово распознано, принималось по критерию максимума правдоподобия:

$$v=\arg\max_r L_r.$$

В результате сравнения решения, принятого при распознавании, с априорными данными о классификации слова получаем зависимость величины ошибки распознавания по тестовому набору слов WER (word error rate) от размера кодовой книги для каждого способа выделения признаков:

$$\mathrm{WER}=1-\frac{S_{\text{прав}}}{S},$$

где $S_{\text{прав}}$ — число правильно распознанных реализаций слов, а S — общее число реализаций. Результаты проведенного эксперимента приведены в таблице.

Из полученных результатов видно, что практически для всех алгоритмов значение минимальной величины ошибки WER достигается при размере кодовой книги, равном 256. При этом наилучшее значение показал алгоритм, использующий коэффициенты линейного предсказания с неравномерным частотным разрешением и модифицированный алгоритм k-средних для

■ Величина ошибки распознавания

Вид вектора признаков	Размер кодовой книги							
	8	16	32	64	128	256	512	1024
LPC	0,218	0,106	0,074	0,047	0,045	0,040	0,041	0,051
CC-LPC	0,190	0,108	0,078	0,053	0,047	0,041	0,048	0,060
WLPC	0,201	0,115	0,070	0,049	0,040	0,035	0,039	0,048
CC-WLPC	0,191	0,105	0,066	0,048	0,043	0,039	0,038	0,046

вычисления кодовой книги. Минимальное значение ошибки WER для него составило 0,035.

Заключение

В работе предложен подход для кластеризации множества АР-моделей речевых сигналов. Для этого вначале был разработан алгоритм для расчета коэффициентов оптимальной по критерию минимума информационного рассогласования АР-модели ЭРЕ, заданной множеством одноименных реализаций. Показано, что используемая в представленной работе в качестве расстояния между АР-моделями мера относится к классу рассогласований Брэгмана.

Для решения собственно задачи кластеризации рассмотрена возможность модификации известного алгоритма кластеризации k-средних, суть которой заключалась в изменении процедуры вычисления центров кластеров в том случае, если они заданы АР-моделями. Дано обоснование сходимости модифицированного алгоритма.

Рассмотренная иллюстрация работы предложенного алгоритма вычисления центроида множества АР-моделей как минимума среднего информационных рассогласований Кульбака — Лейблера показывает, что результирующая модель не является простой комбинацией исходных.

Для оценки эффективности разработанных алгоритмов были проведены их экспериментальные исследования на примере задачи распознавания ограниченного набора слов английского языка с применением аппарата СММ и различных векторов признаков. В результате было показано, что минимальное значение ошибки распознавания достигается при размере кодовой книги (числе кластеров, используемых для представления речевого сигнала в пространстве признаков), равном 256, для большинства рассмотренных векторов признаков. Также показано, что наилучшие результаты достигаются при использовании в качестве признаков коэффициентов линейного предсказания с неравномерным частотным разрешением и соответствующей кодовой книги, найденной при помощи модифицированного алгоритма кластеризации k-средних. Это позволяет говорить о возможности применения предложенных в данной работе алгоритмов при решении задач обработки и распознавания речи.

Дальнейшее исследование эффективности применения разработанных алгоритмов для распознавания большого набора слов из слитной речи представляется интересной задачей. Ее решение требует большого объема размеченных данных для обработки, чему будет посвящена следующая работа.

Литература

1. **Савченко В. В.** Информационная теория восприятия речи // Известия высших учебных заведений России. Радиоэлектроника. 2007. Вып. 6. С. 3–9.
2. **Савченко В. В., Пономарев Д. А.** Оптимизация фонетической базы данных по группе дикторов на основе критерия МИР // Информационные технологии. 2009. № 12. С. 7–12.
3. **Савченко В. В., Акатьев Д. Ю., Губочкин И. В.** Автоматическое распознавание изолированных слов методом обеляющего фильтра // Известия высших учебных заведений России. Радиоэлектроника. 2007. Вып. 5. С. 11–18.
4. **Савченко В. В., Акатьев Д. Ю., Карпов Н. В.** Автоматическое распознавание элементарных речевых единиц методом обеляющего фильтра // Известия высших учебных заведений России. Радиоэлектроника. 2007. Вып. 4. С. 11–19.
5. **Савченко В. В.** Фонема как элемент информационной теории восприятия речи // Известия высших учебных заведений России. Радиоэлектроника. 2008. Вып. 4. С. 3–11.
6. **Banerjee A., Merugu S., Dhillon I. S., and Ghosh J.** Clustering with Bregman Divergences // J. Machine Learning Research. 2005. N 6. P. 1705–1749.
7. **Nielsen F., Nock R.** Sided and Symmetrized Bregman Centroids // IEEE Transactions on Information Theory. June 2009. Vol. 55. N 6. P. 2882–2904.
8. **Do M. N., Vetterli M.** Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance // IEEE Transactions on Image Processing. Feb. 2002. Vol. 11. N 2. P. 146–158.
9. **Ding N., Vishwanathan S. V. N., Qi Y.** t-divergence Based Approximate Inference / NIPS. 2011. P. 1494–1502.
10. **Schwander O., Schutz A. J., Nielsen F., Berthoumieu Y.** k-MLE for mixtures of generalized Gaussians // 21st Intern. Conf. on Pattern Recognition (ICPR), 11–15 Nov. 2012. P. 2825–2828.
11. **Jiang X., Ning L., Georgiou T. T.** Distances and Riemannian metrics for multivariate spectral densities // IEEE Transactions on Automatic Control. 2012. Vol. 57. N 7. P. 1723–1735.
12. **Kullback S., Leibler R. A.** On information and sufficiency // Annals of Mathematical Statistics. 1951. N 22(1). P. 79–86.
13. **Georgiou T. T.** Distances and Riemannian Metrics for Spectral Density Functions // IEEE Transactions on Signal Processing. Aug. 2007. Vol. 55. N 8. P. 3995–4003.

14. **Брэгман Л. М.** Релаксационный метод нахождения общей точки выпуклых множеств и его применение для решения задач выпуклого программирования // Журнал вычислительной математики и математической физики. 1967. Т. 7. № 3. С. 620–631.
15. **Макхоул Дж.** Векторное квантование при кодировании речи // ТИИЭР. 1985. Т. 73. № 11. С. 19–61.
16. **Губочкин И. В.** Алгоритм оценки параметров авторегрессионной модели элементарных речевых единиц // Моделирование и анализ информационных систем. 2013. Т. 20. № 2. С. 23–33.
17. **Марпл С. Л.**-мл. Цифровой спектральный анализ и его приложения. – М.: Мир, 1990. – 584 с.
18. **Rabiner L. R., Juang B.-H.** Fundamentals of speech recognition. – Englewood Cliffs, NJ: Prentice Hall, 1993. – 507 p.
19. **Lloyd S.** Least squares quantization in PCM // IEEE Transactions on Information Theory. 1982. N 28(2). P. 129–137.
20. **MacQueen J.** Some methods for classification and analysis of multivariate observations // Proc. of the Fifth Berkley Symp. on Mathematical Statistics and Probability. 1967. Vol. 1. P. 281–297.
21. **Beigi H.** Fundamentals of Speaker Recognition. – Springer, 2011. – 1003 p.
22. **Linde Y., Buzo A., Gray R. M.** An algorithm for vector quantizer design // IEEE Transactions on Communication. Jan. 1980. Vol. COM-28. N 1. P. 84–95.
23. **Bottou L., Bengio Y.** Convergence Properties of the k-Means Algorithm // Advances in Neural Information Processing Systems. Denver: MIT Press, 1995. Vol. 7. P. 585–592.
24. **Маркел Д. Д., Грэй А. Х.** Линейное предсказание речи. – М.: Связь, 1980. – 308 с.
25. **Harma A.** et al. Frequency-warped autoregressive modeling and filtering. – Helsinki University of Technology, 2001. – 149 p.
26. **Vasko Jr R. C., El-Jaroudi A., Boston J. R.** An algorithm to determine hidden Markov model topology // ICASSP-96: Conf. Proc. IEEE. 1996. Vol. 6. P. 3577–3580.
27. **Freitag D., McCallum A.** Information extraction with HMM structures learned by stochastic optimization // Proc. of the National Conf. on Artificial Intelligence. – Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press, 2000. P. 584–589.
28. **Abou-Moustafa K. T., Cheriet M., Suen C. Y.** On the structure of hidden Markov models // Pattern Recognition Letters. 2004. Vol. 25. N 8. P. 923–931.
29. **Кушнир Д. А.** Алгоритм формирования структуры эталона для пословного дикторонезависимого распознавания команд ограниченного словаря // Штучный інтелект. Київ, 2006. № 3. С. 174–181.
30. **Viterbi A. J.** Error bounds for convolutional codes and asymptotically optimal decoding algorithm // IEEE Transactions on Information Theory. Apr. 1967. Vol. IT-13. P. 260–269.

УВАЖАЕМЫЕ АВТОРЫ!

Национальная электронная библиотека (НЭБ) продолжает работу по реализации проекта SCIENCE INDEX. После того как Вы зарегистрируетесь на сайте НЭБ (http://elibrary.ru/defaultx.asp), будет создана Ваша личная страничка, содержание которой составят не только Ваши персональные данные, но и перечень всех Ваших печатных трудов, имеющихся в базе данных НЭБ, включая диссертации, патенты и тезисы к конференциям, а также сравнительные индексы цитирования: РИНЦ (Российский индекс научного цитирования), h (индекс Хирша) от Web of Science и h от Scopus. После создания базового варианта Вашей персональной страницы Вы получите код доступа, который позволит Вам редактировать информацию, помогая создавать максимально объективную картину Вашей научной активности и цитирования Ваших трудов.