КОМПЬЮТЕРНАЯ МАТЕМАТИКА

Теория и методы оптимизации

УДК 519.685.3

А.С. ГАВРИЛЕНКО

ПРИМЕНЕНИЕ ОПТИМАЛЬНОГО АЛГОРИТМА СИНТЕЗА ЛИНЕЙНЫХ СИСТЕМ КЛАССИФИКАЦИИ В ОБРАБОТКЕ ЗВУКОВЫХ СИГНАЛОВ

Представлена оптимизация линейных алгоритмов синтеза систем распознавания принадлежности сигналов к классу путем максимизации ширины разделяющей полосы, а также применение оптимального разделения к задаче классификации звуковых сигналов (распознавание музыкальных жанров).

Введение. В задачах классификации сигналов широко применяются методы синтеза систем классификации, основанные на автоматическом обучении системы посредством некоторой тренировочной выборки сигналов. В работах [1, 2] рассмотрена теоретическая постановка как линейной, так и нелинейной задачи разделения сигналов на классы и предложен математический аппарат для ее решения. В работе [3] представлен разработанный алгоритм линейного синтеза с последующей нелинейной оптимизацией в случае неэффективности базового метода.

Остановимся подробнее на задаче оптимального синтеза линейных гиперплоскостных кластеров и поставленной в рамках ее задачи оптимальной линейной полосной разделимости двух множеств точек, а также рассмотрим применение оптимизированной системы для классификации звуковых сигналов (автоматическое распознавание музыкальных жанров в аудиофайлах).

Теоретические основы алгоритма

Пусть $x(j) \in R^m$, $j = \overline{1..N}$ – полученная в результате эксперимента тренировочная выборка векторов характерных признаков рассматриваемых сигналов, подлежащих классификации.

Тренировочную выборку в дальнейшем будем представлять в виде матрицы $X =$

$= \left(x(1) \ldots x(N) \right) = \left(x_{(1)}^T \quad \ldots \quad x_{(m)}^T \right)^T$. Здесь предполагается $x_m(j) = 1$.

Согласно [1], необходимое и достаточное условие существования линейной отделимости множества от начала координат имеет вид

$$\min_{y\in D(\Delta)} y^T Z(X) y = y_*^T(\Delta) Z(X) y_*(\Delta) = 0, \text{ где } Z(X) = I_n - X^+ X,$$

а $D(\Delta) = \left\{ y:\ y = (y_1, \ldots, y_n)^T,\ \ y_j \geq \Delta,\ \ j = \overline{1,n} \right\}$.

Максимальная толщина разделяющей полосы достигается при значениях

$$y_{opt} = \arg\min_{y\in D} y^T R(X) y, \quad a_{opt} = (X^T)^+ y_{opt}, \quad R(X) = X^+ (X^T)^+,$$

где $D = \left\{ y : y^T Z(X) y = y_*^T Z(X) y_* = 0,\ \ e_j^T y \geq 1,\ \ j = \overline{1,n} \right\}$.

Используя сингулярное разложение матриц (SVD):

$$X = \sum_{j=1}^{r} u_j v_j^T \lambda_j, \quad Z(X) = I_n - \sum_{j=1}^{r} v_j v_j^T, \quad u_i^T u_j = \delta_{ij},\ v_i^T v_j = \delta_{ij},$$

$$XX^T u_j = \lambda_j^2 u_j, \quad X^T X v_j = \lambda_j^2 v_j,\ \lambda_1^2 \geq \ldots \geq \lambda_r^2, \quad i, j = \overline{1,n}$$

и учитывая, что $y^T Z(X) y = 0$ для $y \in D$, можно записать следующие соотношения:

$$y = \sum_{i=1}^{r} \alpha_i v_i, \quad \sum_{i=1}^{r} \alpha_i e_j^T v_i \geq 1, \quad \forall\ j = \overline{1,n},$$

$$y^T R(X) y = y^T X^+ X^{+T} y = y^T \sum_{j=1}^{r} v_j v_j^T \lambda_j^{-2} y =$$

$$= \sum_{i=1}^{r} \alpha_i v_i^T \sum_{j=1}^{r} v_j v_j^T \lambda_j^{-2} \sum_{k=1}^{r} v_k \alpha_k = \sum_{j=1}^{r} \alpha_j^2 \lambda_j^{-2} = \alpha^T diag(\lambda_1^{-2}, \ldots, \lambda_r^{-2}) \alpha.$$

Таким образом, задача поиска оптимальной линейной отделимости точек $x(j),\ \ j = \overline{1,n}$ от начала координат сводится к решению задачи оптимизации квадратичной функции на выпуклом множестве

$$\alpha_{opt} = \arg \min_{\alpha = \left\{\alpha:\ e_j^T (v_1 \vdots \ldots \vdots v_r)\alpha \geq 1,\ j = \overline{1,n}\right\}} \alpha^T diag(\lambda_1^{-2}, \ldots, \lambda_r^{-2}) \alpha, \quad y_{opt} = \left(v_1 \vdots \ldots \vdots v_r \right) \alpha_{opt},$$

$$a_{opt} = \sum_{j=1}^{r} u_j v_j^T \lambda_j^{-1} \cdot (v_1 \vdots \ldots \vdots v_r) \alpha_{opt} = \left(\lambda_1^{-1} u_1 \vdots \ldots \vdots \lambda_r^{-1} u_r \right) \alpha_{opt}.$$

Рассмотрим теперь пошаговый алгоритм решения задачи линейной полосной разделимости двух множеств точек с учетом выведенных выше соотношений.

Схема алгоритма. Алгоритм основан на схеме, описанной в [4]. Обучающая выборка – массив точек $x(i_k) \in \Omega_x(1)$, $k = \overline{1..N_1}$ $x(j_s) \in \Omega_x(2)$, $s = \overline{1..N_2}$.

1. Зададим некоторое начальное значение $d > 0$ и сформируем первое приближение

$$y = (\Delta_1 \ \dots \ \Delta_N), \quad \Delta_j = d, \ j = \overline{1..N}.$$

2. Решая систему $N \times m+1$ уравнений $y = Xa$, получаем вектор коэффициентов $\hat{a} = X^+ y$.

3. Вычисляем:

а) вектор модельных значений $y_M = X_e \hat{a}$;

б) дискриминантную функцию для точек, условно отнесенных к первой и ко второй группам:

$$\begin{cases} y_e^{(1)} = X_e^{(1)} \hat{a} \\ y_e^{(2)} = X_e^{(2)} \hat{a} \end{cases};$$

в) невязку $y_M - y$ и среднеквадратичное отклонение $\|y_M - y\|^2 = = y^T Z(X) y$.

4. В случае если выполняется условие

$$\begin{cases} y_e^{(1)} \ge \Delta \\ y_e^{(2)} \le -\Delta \end{cases},$$

осуществляется выход из алгоритма.

5. Иначе проводим уточнение значения вектора допусков $y^T = = (\Delta_1 \ \dots \ \Delta_N)$, $\Delta_j = d$, $j = \overline{1..N}$, минимизируя квадратичную форму $y^* = \arg\min_{y \in D} y^T Z(X) y$ градиентным методом с ограничением $y \in D$, где $D = \left\{ y : \ d \le \Delta_j \le d', \ j = \overline{1..N} \right\}$ (положим $d' = 10^6$).

6. Зная y^*, находим уточненное значение вектора коэффициентов линейной регрессии $\hat{a}_* = X^+ y^*$.

7. В случае если выполняется условие, что проекция y_M^* на ортогональное дополнение к объединенному массиву экспериментальных данных X (положим точность приближения $\varepsilon = 10^{-6}$)

$$y_M^{*\,T} Z(X) y_M^* < \varepsilon,$$

где $y_M^* = X\hat{a}_*$ – уточненный вектор модельных значений, разделяющая плоскость найдена.

8. Далее, используя вышеприведенные условия линейной отделимости точек в R^m в применении к текущей постановке задачи, представим линейную полосную разделимость двух классов обучающей выборки в виде существования такого вектора $a \in R^m$, для которого

$$a^T x(i_k) \geq 1, \quad k = \overline{1, N_1},$$
$$a^T x(j_s) \leq -1, \quad s = \overline{1, N_2}.$$

Тогда условие линейной полосной разделимости приобретает вид

$$\min_{y \in D} y^T Z(X) y = 0,$$
$$D = \left\{ y: \ e_{i_k}^T y \geq 1, \ e_{j_s}^T y \leq -1, \ k = \overline{1, N_1}, s = \overline{1, N_2} \right\}.$$

9. Значение вектора коэффициентов a будем определять из условия максимизации толщины разделяющей полосы:

$$a_{opt} = \left(X^T\right)^+ y_{opt},$$

где $y_{opt} = \arg\min_{y \in D_1} y^T R(X) y$, а $D_1 = \left\{y: y^T Z(X) y = 0\right\} \cap D$.

Как видим, для вычисления оптимального значения a необходимо последовательно решить две задачи квадратичного программирования. Применяя представление SVD для матрицы X , сведем эти две задачи к одной задаче поиска экстремума квадратичной функции на выпуклом множестве и вычислим a_{opt} :

$$\alpha_{opt} = \arg \min_{\alpha \in \left\{\alpha: \ e_{i_k}^T (v_1 \vdots \ldots \vdots v_r)\alpha \geq 1, e_{js}^T (v_1 \vdots \ldots \vdots v_r)\alpha \leq -1, \ k=\overline{1,N_1}, \ s=\overline{1,N_2}\right\}} \alpha^T diag(\lambda_1^{-2}, \ldots, \lambda_r^{-2})\alpha,$$

$$y_{opt} = \left(v_1 \vdots \ldots \vdots v_r\right) \alpha_{opt},$$

$$a_{opt} = \left(u_1 \vdots \ldots \vdots u_r\right) diag\left(\lambda_1^{-1}, \ldots, \lambda_r^{-1}\right) \alpha_{opt}.$$

Ширина разделяющей полосы при этом будет равна

$$\delta_{opt} = \left(\alpha_{opt}^T \, diag(\lambda_1^{-2}, \ldots, \, \lambda_r^{-2}) \alpha_{opt}\right)^{-\frac{1}{2}} \text{ [2].}$$

Постановка задачи распознавания

Музыкальный жанр как комплекс характерных признаков широко используется как для структурирования музыки, хранимой в цифровой форме, так и для поиска информации о содержимом аудиофайла. Традиционно эта задача выполняется вручную. Конкретный музыкальный жанр характеризуется статистическими свойствами, отражающими использованные инструменты, ритмическую структуру и другие характерные признаки.

В данном случае в качестве вектора признаков использованы характеристики, отражающие фактуру, тембр и набор инструментов, использованных в музыкальном произведении: mean-Centroid, mean-Rolloff, mean-Flux, mean-ZeroCrossings, std-Centroid, std-Rolloff, std-Flux, std-ZeroCrossings, LowEnergy.

Средние и стандартные отклонения этих признаков рассчитываются для окна длиной в 1 секунду, состоящего из 40 «аналитических» окон длиной 20 миллисекунд (512 образцов при частоте дискретизации 22050). Расчеты производятся с использованием оконного преобразования Фурье, которое эффективно вычисляется с помощью быстрого преобразования Фурье (БПФ).

Следующие функции рассчитываются для каждого «аналитического» окна: ($M[f]$ – величина БПФ на частоте f, N – количество частотных окон).

Centroid (мера спектральной яркости):

$$C = \frac{\sum_1^N fM[f]}{\sum_1^N M[f]} \text{N}.$$

Rolloff (мера спектральной формы): значение R такое, что:

$$\sum_1^R M[f] = 0{,}85 \sum_1^N M[f].$$

Flux (мера спектрального изменения):

$$\left\| M[f] - M_p[f] \right\|,$$

где M_p – величина БПФ для предыдущего отрезка времени. Оба вектора величины нормированы по энергии.

ZeroCrossings: количество изменений знака (нулей) в сигнале. Полезно для определения количества шума в сигнале.

Расчет функций для представления ритмической структуры музыки производится на основе вейвлет-преобразования.

Дискретное вейвлет-преобразование является частным случаем вейвлет-преобразования, которое обеспечивает компактное представление сигнала по времени и частоте, что позволяет более эффективные вычисления. В данном случае используется пирамидальный алгоритм, в котором сигнал анализируется на различных полосах частот с различными разрешениями путем разложения на грубую аппроксимацию и подробную информацию.

Затем грубое приближение дополнительно разлагается без изменения вейвлет-шага. Разложение проводится путем последовательной фильтрации сигнала по верхним и нижним частотам и определяется следующими уравнениями:

$$y_{high}[k] = \sum_n x[n]g[2k-n],$$

$$y_{low}[k] = \sum_n x[n]h[2k-n],$$

где g и h – высокочастотный и низкочастотный фильтры соответственно.

Выделение признаков ритмического рисунка основана на выявлении наиболее характерных периодичностей в сигнале. Вначале сигнал разлагается в набор октавных частотных полос с использованием дискретного вейвлет-преобразования. Затем выделяется огибающая амплитуды на временном промежутке для каждой частотной полосы. Это достигается за счет применения двухполупериодного выпрямления, фильтрации нижних частот и децимации для каждой частотной полосы. Затем огибающие для каждой полосы суммируются и вычисляется автокорреляционная функция. Всплески автокорреляционной функции соответствуют различным периодичностям в огибающей сигнала. Описанные этапы задаются уравнениями:

1. Двухполупериодное выпрямление:
   $y[n] = abs(x[n])$.
2. Фильтрация нижних частот (ФНЧ): (однополюсный фильтр с $\alpha = 0,99$:
   $y[n] = (1-\alpha)x[n] - \alpha y[n]$.
3. Децимация по k (в рассматриваемой реализации $k = 16$):
   $y[n] = x[kn]$.
4. Нормализация (удаление среднего):
   $y[n] = x[n] - E[x[n]]$.
5. Автокорреляционная функция:
   $y[n] = \frac{1}{N}\sum_{n} x[n]x[n+k]$.

Выделяются первые пять всплесков автокорреляционной функции и соответствующие им периодичности, выраженные в ударах в минуту, на основе которых строится так называемая ударная гистограмма. Процесс итеративно повторяется с накоплением периодичностей в гистограмме (в данном случае размер окна составляет 65536 образцов при частоте дискретизации 22050 Гц с итерационным шагом в 4096 образцов).

Выделенные всплески конечной гистограммы соответствуют различным периодичностям звукового сигнала и используются при расчете численных характеристик ритма, а именно.

1. **Period0**: периодичность первого всплеска в уд/мин.

2. **Amplitude0**: относительная амплитуда (деленная на сумму амплитуд) первого всплеска.

3. **RatioPeriod1**: соотношение периодичности второго всплеска к периодичности первого всплеска.

4. **Amplitude1**: относительная амплитуда второго всплеска.

5. **RatioPeriod2, Amplitude2, RatioPeriod3, Amplitude3** – аналогично.

Полученный 8-мерный вектор признаков ритмического рисунка может быть объединен с 9-мерным основным вектором признаков.

Для обучения синтезированной оптимальной системы классификации в данном случае были использованы вычисленные для двух музыкальных жанров (а именно, «джаз» и «неоклассика») мелодико-частотные кепстральные коэффициенты (MFCC) [4], вычисленные с помощью системы MARSYAS [5, 6] (табл. 1, 2). MFCC могут быть использованы в качестве замены оконного преобразования Фурье с аналогичным описанному последующим вычислением средних и стандартных отклонений первых пяти признаков для окна большего размера (в данном случае окна длиной 1 с.).

ТАБЛИЦА 1. MFCC для жанра «джаз»

Jazz								
52.765196	9.969953	–1.068876	0.372135	0.750053	–0.082523	–0.057770	0.145388	–0.140282
–47.323153	5.177846	–1.107054	0.655738	–0.331889	0.401314	–0.044381	–0.087062	–0.191644
–48.103305	6.525593	–2.009164	1.696986	–0.571502	0.634679	–0.225843	0.182733	–0.036656
–47.853101	7.788700	–2.075793	1.687302	–0.359639	0.988925	–0.256285	–0.428794	–0.674491
–47.795176	7.254454	–1.728231	1.830460	–0.722954	0.866811	–0.395115	0.022426	–0.530013
–48.110098	5.801459	–0.314840	0.530672	0.015405	0.145267	–0.100025	–0.105939	–0.126800
–47.144708	7.908430	–1.578344	1.400528	–0.074236	0.167585	0.009038	–0.320367	0.130205
–48.007967	5.988282	–0.879154	2.140802	–0.553409	0.944400	–0.461947	0.291691	–0.220351
–49.409359	7.920237	–0.705804	1.310598	–0.139005	0.594784	0.133871	0.220815	–0.155640
–46.901865	6.118231	–0.451342	1.245182	–0.088447	0.567180	–0.134913	0.141795	–0.537725
–47.548655	6.492598	–1.478057	1.417851	–0.156623	0.096782	–0.110061	0.117167	–0.074586
–47.548655	6.492598	–1.478057	1.417851	–0.156623	0.096782	–0.110061	0.117167	–0.074586
–48.299786	6.037525	–0.544464	0.982117	–0.140868	0.324324	–0.034081	0.007872	–0.141855
–47.021231	6.842780	–1.171481	1.147371	–0.106207	0.196891	–0.064192	0.017304	0.027194
–47.422584	6.913468	–1.073093	1.295354	–0.172893	0.303768	–0.013511	0.095759	0.043826
–48.157082	4.550797	–0.066564	0.779817	–0.346519	0.397902	–0.021595	–0.073226	–0.103904
–48.027493	5.475738	0.207603	0.798316	–0.155938	0.389709	–0.120319	–0.155803	0.170739
–48.282734	6.698345	–0.000768	1.108222	–0.273255	0.389634	0.037302	0.046539	0.410506
–50.033461	5.368075	0.694274	1.084474	0.355777	0.386235	0.392366	–0.173208	–0.034135
–52.375001	7.119549	1.143711	0.676076	0.204379	0.594507	0.320473	–0.221682	0.005853

ТАБЛИЦА 2. MFCC для жанра «классика»

Neo-Classics								
1	2	3	4	5	6	7	8	9
–42.519996	2.791965	–0.551018	1.648224	0.250287	0.255351	0.544240	0.184539	0.482820
–43.620071	3.522973	–0.655716	1.371293	0.066918	0.371061	0.658504	0.143836	0.075018
–44.916634	2.480168	–0.052545	1.715316	0.085388	0.714413	0.468862	0.380389	0.353419
–43.900186	4.225723	–0.691845	1.235925	0.236653	0.460900	0.455110	0.391106	0.573263
–44.819528	3.902396	–0.401066	1.624539	0.031978	0.517075	0.412870	0.139171	0.320433
–43.716847	3.232922	–0.482210	1.286344	0.212244	0.633372	0.526592	0.330332	0.308743
–42.759187	3.356651	–0.542355	1.482408	0.251527	0.499683	0.361233	0.427227	0.650168
–44.730868	3.855467	–0.578810	1.879171	0.185021	0.571880	0.431775	0.432590	0.528225
–41.342148	3.039183	0.107569	1.849919	–0.143252	0.758268	0.235490	0.356677	–0.017974
–44.352347	4.551677	–0.166774	1.606438	–0.048026	0.652277	0.243426	0.148154	0.016650

Окончание табл. 2

1	2	3	4	5	6	7	8	9
–43.569055	5.026774	–0.523512	1.366952	0.061907	0.362519	0.286701	0.178622	0.335419
–43.718137	4.412115	–0.444838	1.383368	–0.253404	0.494083	0.280392	0.175557	–0.057454
–43.004565	4.984104	–0.799496	1.514265	–0.134894	0.292498	0.487551	–0.007842	0.293451
–44.044681	4.254025	–0.426857	1.599984	0.106479	0.356600	0.151158	–0.288511	0.197443
–44.444159	5.313499	–0.635376	1.231647	0.047593	0.467125	0.314914	0.110206	0.074439
–45.043005	5.630610	–0.187339	1.376719	–0.091472	0.576595	0.371674	0.104184	0.155919
–42.323962	4.326221	–0.939891	1.224892	–0.306318	0.613339	–0.030573	0.016638	–0.108492
–43.436943	4.965412	–0.533134	1.413742	–0.106005	0.688715	0.060311	0.145482	0.054193
–44.922768	5.612160	–0.770044	1.308193	–0.372075	0.660108	0.281929	0.243398	0.133602
–44.820261	5.687104	–0.161438	1.342538	–0.235188	0.591333	0.360430	0.336583	–0.128797

По результатам вычислений

$$a_{opt} = (10.785, 2.74, 5.77, -1.38, -4.356, -5.206, 5.1, 1.108, 2.55),$$

$$\delta_{opt} = 2.04.$$

Средняя эффективность распознавания с помощью оптимальной синтезированной системы – 90 %.

Заключение. В работе выведено теоретическое обоснование и принципиальная схема оптимального алгоритма синтеза системы классификации сигналов с максимальной шириной разделяющей полосы. Описанный алгоритм успешно апробирован в практической задаче обработки аудиосигналов, показав эффективность не ниже известных алгоритмов [6], а также снижение машинной ресурсоемкости задачи за счет предложенного преобразования.

А.С. Гавриленко

ЗАСТОСУВАННЯ ОПТИМАЛЬНОГО АЛГОРИТМУ СИНТЕЗУ
ЛІНІЙНИХ СИСТЕМ КЛАСИФІКАЦІЇ В ОБРОБЦІ ЗВУКОВИХ СИГНАЛІВ

Представлено оптимізацію лінійних алгоритмів синтезу систем розпізнавання приналежності сигналів до класу шляхом максимізації ширини роздільної смуги, а також застосування оптимального розділення до задачі класифікації звукових сигналів (розпізнавання музичних жанрів).

A.S. Gavrylenko

APPLYING OF OPTIMAL ALGORYTHM FOR SIGNAL CLASSIFICATION PROBLEM
SOLVING SYSTEMS SYNTHESIS TO AUDIO SYGNALS PROCESSING

Article presents the improvement of the algorythm for signal classification systems synthesis by maximizing distance between separating hyperplanes and application of optimized algorithm for sound sygnals classification (automatic recognition of music genres).

1. *Кириченко Н.Ф., Крак Ю.В., Полищук А.А.* Псевдообратные и проекционные матриц в задачах синтеза функциональных преобразователей // Кибернетика и системный анализ. – 2004. – № 3. – С. 116 – 129.
2. *Кириченко Н.Ф., Кривонос Ю.Г., Лепеха Н.П.* Синтез систем нейрофункциональных преобразователей в решении задач классификации // Там же. – 2007. – № 3. – С. 47 – 57.
3. *Кириченко Н.Ф.* Аналитическое представление возмущений псевдообратных матриц // Там же. – 1997. – № 2. – С. 98 – 107.
4. *Logan B.Mel.* Frequency Cepstral Coefficients for music modeling. Read at the first International Symposium on Music Information Retrieval. http://ciir.cs.umass.edu/music2000
5. *George Tzanetakis, Georg Essl.* Automatic Musical Genre Classification Of Audio Signals // IEEE Transactions on Speech and Audio Processing. – 2001. – P. 293–302.
6. *Li. T., Ogihara M., Li. G.* A Comparative Study on Content-Based Musik Genre Classification // In: SIGRIR’03. – ACM Press, 2003. – P. 282–289.

Получено 15.07.2013

Об авторе:

Гавриленко Анастасия Сергеевна,
младший научный сотрудник
Института кибернетики имени В.М. Глушкова НАН Украины.
E-mail: anastasiia.gavrylenko@gmail.com