Curve clustering using Chebyshev and Least Squares approximation

N Sukhorukova, Swinburne University of Technology and
Julien Ugon, Deakin University
Federation University Australia

K-means and clustering in $\mathbb{R}^n$

- Step 1: Identify cluster centres.
- Step 2: Assign each point to the cluster with the nearest centre.
- Step 3: Recompute the cluster centres.

This is a very fast clustering algorithm for clustering in $\mathbb{R}^n$, especially if the distance is just the usual Euclidean distance. In this case, the barycentre (centroid). It is easy to recompute the updated centres. On the top of this, this is easy to parallelise the computations.

K-means and clustering for curves

Similar approach has been applied to curve clustering (discretised signals). Each time moment can be treated as a separate coordinate (providing that all the pieces of signal are defined in the same time segment).
Assume now that the cluster centre has to be approximated by a curve

$$S(A, t) = \sum_{i=0}^{n} a_i g_i(t).$$

Assume for now that $g_i(t)$ are monomials.

Least squares

$$\text{minimise } F(\mathbf{X}) = \sum_{i=1}^{N} \sum_{j=1}^{l} (S_j(t_i) - P_n(\mathbf{X}, t_i))^2, \tag{1}$$

where $\mathbf{X} = (x_0, \ldots, x_n) \in \mathbb{R}^{n+1}$, $x_k,\ k = 0, \ldots, n$ are the polynomial parameter and also the decision variables. Each signal is a column vector

$$\mathbf{S}^j = (S_j(t_1), \ldots, S_j(t_N))^T,\ j = 1, \ldots, l.$$

Matrix form

Problem (1) can be formulated in the following matrix form:

$$\text{minimise } F(\mathbf{X}) = \|\mathbf{Y} - \mathbf{B}\mathbf{X}\|, \tag{2}$$

where

$\mathbf{X} = (x_0, \ldots, x_n) \in \mathbb{R}^{n+1}$, are the decision variables (same as in (1));

vector

$$\mathbf{Y} = \begin{pmatrix} \mathbf{S}^1 \\ \mathbf{S}^2 \\ \vdots \\ \mathbf{S}^l \end{pmatrix} \in \mathbb{R}^{(n+1)l}$$

matrix $\mathbf{B}$ contains repeated matrix blocks, namely,

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_0 \\ \mathbf{B}_0 \\ \mathbf{B}_0 \\ \vdots \\ \mathbf{B}_0 \end{pmatrix},$$

Matrix form (cont.)

where

$$\mathbf{B}_0 = \begin{pmatrix} 1 & t_1 & t_1^2 & \dots & t_1^n \\ 1 & t_2 & t_2^2 & \dots & t_2^n \\ \vdots & \vdots & \ddots & \dots & \vdots \\ 1 & t_N & t_N^2 & \dots & t_N^n \end{pmatrix}.$$

Normal equations

This least squares problem can be solved using a system of normal equations:

$$\mathbf{B}^T\mathbf{B}\mathbf{X} = \mathbf{B}^T\mathbf{Y}. \tag{3}$$

Taking into account the structure of the system matrix in (3), the problem can be significantly simplified:

$$l\mathbf{B}_0^T\mathbf{B}_0\mathbf{X} = \mathbf{B}_0^T\sum_{k=1}^{l}\mathbf{S}^k. \tag{4}$$

Therefore, instead of solving (3), one can solve

$$\mathbf{B}_0^T\mathbf{B}_0\mathbf{X} = \mathbf{B}_0^T\frac{\sum_{k=1}^{l}\mathbf{S}^k}{l} = \mathbf{B}_0^T\mathbf{S}, \tag{5}$$

where $\mathbf{S}$ is the average of all l signals of the group (centroid).

Centre (prototype) update

Suppose that a signal group prototype has been constructed. Assume now that we need to update our group of signals: some new signals have to be included, while some others are to be excluded. To update the prototype, one needs to update the centroid and solve (5) with the updated right-hand side, while the system matrix $\mathbf{B}_0^T\mathbf{B}$ remains the same.
If only few signals are moving in and out of the group, then the updated centroid can be calculated without recomputing from scratch. Assume that l_a signals are moving in the group (signals $S_a^1(t), \dots S_a^{l_a}$), while l_r are moving out (signals $S_r^1(t), \dots S_r^{l_r}$), then the centroid can be recalculated as follows:

$$S_{new}(t) = \frac{lS_{old}(t) + \sum_{k=1}^{l_a} S_a^k(t) + \sum_{k=1}^{l_r} S_r^k(t)}{l - l_r + l_a}.$$

Least Squares: summary

Since the same system has to be solved repeatedly with different right-hand sides, one approach is to invert matrix $\mathbf{B}_0^T \mathbf{B}_0$, which is an $(n+1) \times (n+1)$ matrix. In most cases, n is much smaller than N or I and therefore this approach is quite attractive, if we can guarantee that matrix $\mathbf{B}_0^T \mathbf{B}_0$ is invertible. In the next section we discuss the verification of this property.

Chebyshev (uniform) approximation

Approximation theory is concerned with the approximation of a function f, defined on a (continuous or discrete) domain Ω, by another function s taken from a family $\mathfrak{S}$. At any point $t \in \Omega$ the difference

$$d(t) \triangleq s(t) - f(t)$$

is called the *deviation* at t, and the *maximal absolute deviation* is defined as

$$\|s - f\| \triangleq \sup_{t \in \Omega} |s(t) - f(t)|.$$

The problem of best Chebyshev approximation is to find a function $s^* \in \mathfrak{S}$ minimising the maximal absolute deviation over $\mathfrak{S}$. Such a function s^* is called a *best approximation* of f.

Chebyshev's Theorem

The seminal result of approximation theory is Chebyshev's alternation theorem which can be stated as follows. Let $\mathcal{P}_n$ be the set of polynomials of degree at most n with real coefficients.

Theorem

(1854) A polynomial $p^* \in \mathcal{P}_n$ is a best approximation to a continuous function f on an interval $[a, b]$ if and only if there exist $n+2$ points $a \leq t_1 < \ldots < t_{n+2} \leq b$ and a number $\sigma \in \{-1, 1\}$ such that

$$(-1)^i \sigma(f(t_i) - p^*(t_i)) = \|f - p^*\|, \forall i = 1 \ldots, n+2.$$

The sequence of points $(t_i)_{i=1,\ldots,n+2}$ is called an *alternating sequence*.

Centre prototype

Supposed that a cluster consists of n signals $(S_1, \dots, S_n)$, assigned with respect to the shortest distance to the cluster centres. Now we need to recompute the cluster prototype.
First of all, we need to construct two curves:

- $S_{\max}(t) = \max_{i=1,\dots,n} S_i(t)$;
- $S_{\min}(t) = \min_{i=1,\dots,n} S_i(t)$.

Then the parameters of the cluster prototype are the solution of the following optimisation problem:

$$\text{mimimise} \max_{t \in [a,b]} \{S_{\max}(t) - S(A,t), S(A,t) - S_{\min}(t)\}.$$

If the interval $[a, b]$ is discretised $(t_i \in [a, b],\ i = 1, \dots, N)$, a solution can be obtain by solving an LP (next slide).

LP formulation

$$\text{mimimise } z$$

subject to

$$S_{\max}(t_i) - S(A, t_i) \leq z, \;\; i - 1, \ldots, N,$$

$$S(A, t_i) - S_{\min} \leq z, \;\; i - 1, \ldots, N.$$

We can solve this LP every time we need to update a cluster centre, however, since only (a) few signals are moving from one cluster to another, this procedure can be optimised.

Use the prototype from the previous iteration as an initial solution for the next iteration

This may not be a very good idea, since this point may not be a feasible point for the next iteration LP.
Can we still use the results obtained at the previous iteration? I think, the answer is yes.

de la Vallée-Poussin's procedure

de la Vallée-Poussin's procedure has been developed for polynomial approximation (Chebyshev metrics). Later this procedure has been extended to the case of any Chebyshev system (not just monomials $g_i(t) = t^i$).

It was also noticed that this procedure (at least for the case $g_i(t) = t^i$) is the Simplex method applied to the corresponding LP.

Can we extend this procedure to the case when two curves are approximated? So far, we have a lot of similarities with classical uniform approximation.

Optimality conditions

Let

$$\Delta = \max_{t\in[a,b]} \{S_{\max}(t) - S(A,t), S(A,t) - S_{\min}(t)\}.$$

Theorem

An approximation $S(A^*, t)$ is a best approximation to a pair of curves $S_{\max}$ and $S_{\min}$ on an interval $[a, b]$ if and only if at least one of the following conditions holds:

1. *there exists a time moment $t_k \in [a, b]$, such that*
$$\Delta = S_{\max}(t_k) - S(A^*, t_k) = S(A^*, t_k) - S_{\min}(t_k)$$
2. *there exist $n + 2$ points $a \le t_1 < \ldots < t_{n+2} \le b$ and*
   - $\Delta = S_{\max}(t_k) - S(A^*, t_k) = S(A^*, t_{k+1}) - S_{\min}(t_{k+1}),\ k = 1, \ldots, n$ *or*
   - $\Delta = S(A^*, t_k) - S_{\min}(t_k) = S_{\max}(t_{k+1}) - S(A^*, t_{k+1}),\ k = 1, \ldots, n.$

Example 1: non-uniqueness of optimal solution

Let $[a, b] = [0, 1]$, $S_{\max}(t) = 1 - 0.5t$ and $S_{\min}(t) = 0.5t$. Find a best linear approximation for these two curves. Then $S(A, t) = 0.5$ is optimal. Moreover, any line

$$0.5 + kt, \;\; k \in [-0.5, 0.5]$$

is optimal.

There exists an optimal solution.

Optimal solution is not unique, but is there at least one satisfying condition 2?

Same example (Example 1). No. We probably need to redefine the notion of alternating sequence.

Stone-Weierstrass approximation theorem

Weierstrass approximation theorem states that every continuous function defined on a closed interval $[a, b]$ can be uniformly approximated as closely as desired by a polynomial function.
Is not applicable anymore, if there exists $t \in [a, b]$, such that $S_{\max}(t) \neq S_{\min}(t)$.

Second condition

Theorem

Assume that, after basis update, there is no point t, where $S_{\max}(t) - S_{\min}(t) \geq \Delta$. Then de la Vallée-Poussin's procedure can be extended to the case of two curve approximation for Chebyshev systems (not just monomials $g_i(t) = t^i$).

We need to generalise three steps:

1. Best Chebyshev interpolation construction.
2. Basis update rule.
3. de la Vallée-Poussin's theorem generalisation.

Conjecture

This procedure (at least for the case $g_i(t) = t^i$) is the Simplex method applied to the corresponding LP.

Second condition is not satisfied

Redefine the notion of alternating sequence: treat a point t, where

$$S_{\max}(t) - S_{\min}(t) \geq \Delta$$

as a "multiple" (or "double") point (two points with "positive and "negative" maximal deviation combined).

Conjecture

Consider a set of optimal solutions. Among these solutions, there is at least one where, after redefining the notion of alternating sequence with double points, condition 2 is satisfied.

Conjecture

de la Vallée-Poussin's generalised procedure converts to an optimal solution from the conjecture above.

A bit more on Chebyshev systems

Let the basis functions be monomial functions. Can we guarantee that they form a Chebyshev system?

Example

Consider the system of two monomials on the segment $[-1, 1]$:

$$g_1(t) = 1,\ g_2(t) = t^2,\ t_1 \neq t_2,$$

the monomial t is "missing". Take time-moments $t_1, t_2 \in [-1, 1]$. The determinant

$$\begin{vmatrix} 1 & t_1^2 \\ 1 & t_2^2 \end{vmatrix} = 0 \Leftrightarrow t_1 = -t_2.$$

Therefore, these functions do not form a Chebyshev system, since the corresponding determinant is zero when, for example, $t_1 = -t_2 = 1$ and there is only one linear independent row.

Conditions to keep the system "Chebyshev" for monomial basis functions

1. None of the monomials should be skipped.
2. $[a, b] \subset (0, +\infty)$ (1966, Karlin and Studden)

General approach (V. Malozemov, St-Petersburg State University)

$$\mathbb{O}_n \in Z = \operatorname{co}\{Z_i,\ i = 1, \dots, N\}. \qquad (6)$$

Theorem

The condition (6) is equivalent to the following: there exist

$$Z_i \in G,\ i = 1, \dots, r,\ Z_i \in \tilde{G},\ i = 1, \dots, n+1,$$

such that the determinants Δ_i, obtained by removing the $i-th$ vector satisfy the following condition:

$$\Delta_i \neq 0,\ i = 1, \dots, r,\ \operatorname{sign}\Delta_{i-1} = -\operatorname{sign}\Delta_i,\ i = 2, \dots, r,$$

$$\Delta_i = 0,\ i = r+1, \dots, n+1.$$