

パターン分類を用いた水文情報の欠測値補完

岐阜大学工学部 正会員 小尻利治

岐阜大学工学部 学生員 ○青木繁幸

1. はじめに

温室効果気体の増加にともなう気候変動の解析や、人口の増加、産業の発展にともなう水利用形態の変化などの解析において水文データが必要となるわけだが、欠測部分についてはほとんどそのまま欠測として取り扱われている。そこで、本研究では、パターン分類により過去のデータを分類し各々のクラスターに分ける。ついで欠測のあるサンプルデータに対し観測範囲内でクラスターセンターへの類似度を算出し、欠測部分を補完しようとするものである。また、欠測部分を将来の水文データに置き換えることによって流況予測を行うものである。

2. ソフトクラスタリング手法

流量時系列を、形状としてとらえ、その特性に応じて分類するために、ファジィ理論を用いたソフトクラスタリング手法を導入しよう。このパターン分類に際しての評価関数は、次式を用いる。

$$d_i(t) = \max \frac{|X(t) - Z_i(t)|}{Z_i(t)} \quad (1)$$

ここで、 $X(t)$ は t 期におけるサンプルの値、 $Z_i(t)$ は t 期におけるクラスターセンター i の値を表す。サンプルがどのクラスターへの属性が強いかの判定には、メンバーシップ関数より類似度を求める。すなわち、

$$F(i) = (-|d_i(t)| + a_i) / a_i \quad (2)$$

である。ここで、 a_i はメンバーシップ関数の傾きを表す定数である。

3. 1 地点しか観測データがない場合の欠測値の補完

図-1 のように、クラスター i への類似度を $F(i)$ 、既知の流量系列を $\{Q(i)\}$ とすると、補完する流量系列 $\{\bar{Q}\}$ は次式で求められる。

$$\{\bar{Q}\} = \frac{\sum_i F(i)\{Q(i)\}}{\sum_i F(i)} \quad (3)$$

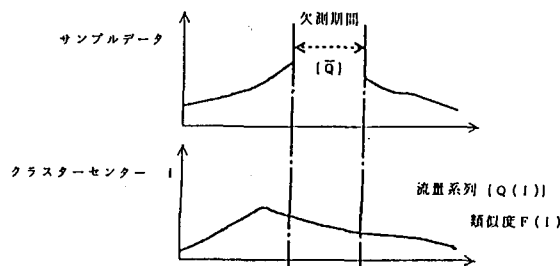


図-1 1 地点しか観測データがない場合の補完の説明図

4. 2 地点 (複数) の観測データがある場合の欠測値の補完

1) 2 地点の既知データが利用できる場合

別々の地点でパターン分類が行われ、各クラスターへの分類結果が類似である場合、この 2 地点の分類特性は相関性が高いと定義する。いま、一方の地点で欠測が生じた時の補完方法について、以下のように考えよう。欠測をもつ地点 1 においてクラスター i への類似度を $F(1, i)$ 、既知の流量系列を $\{Q(1, i)\}$ 、またデータがそろっている地点 2 においてクラスター i' への類似度を $F'(2, i')$ とすると補完値 $\{\bar{Q}\}$ は次式で求められ

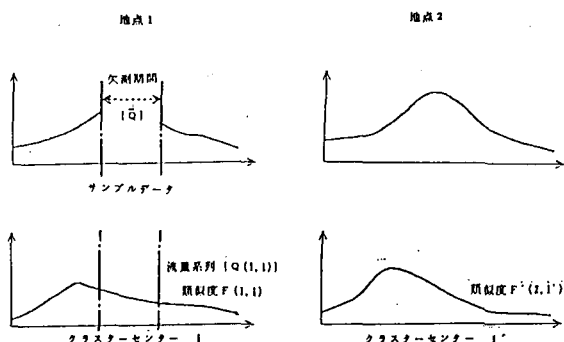


図-2 2 地点の観測データがある場合の補完の説明図

る。ただし、クラスター*i*とクラスター*i'*とは類似とする(図-2参照)。

$$\{\bar{Q}\} = \frac{\sum_{i,i'} [\min\{F(1,i), F'(2,i')\} \{Q(1,i)\}]}{\sum_{i,i'} [\min\{F(1,i), F'(2,i')\}]} \quad (4)$$

2) 両地点とも同期間のデータが欠測している場合

それぞれの地点で既知データから類似度を算出し式(2)より欠測の補完を行う。

5. 長期間(年度または季節)観測データが欠測している場合

図-3のように、長期間欠測している前後の期間をそれぞれ期間1、期間2とする。期間1において類似度*F*(1, *i*)であるパターン*i*から欠測期間の各パターンへの遷移確率を*P*(1, *j*)、期間2において類似度*F*(2, *k*)であるパターン*k*から欠測期間への遷移確率を*P*(2, *j*)とおく。ここで、 $\min\{F(1,i)P(1,j), F(2,k)P(2,j)\}$ を欠測期間の類似度と考えると、補完値($\bar{Q}$)は次式で求められる。

$$\{\bar{Q}\} = \frac{\sum_{i,j,k} [\min\{F(1,i)P(1,j), F(2,k)P(2,j)\} \{Q(j)\}]}{\sum_{i,j,k} [\min\{F(1,i)P(1,j), F(2,k)P(2,j)\}]} \quad (5)$$

6. 適用と考察

豊川水系布里流量観測所(昭和25年~昭和55年の毎月の平均流量)における欠測値補完の適用例を示す。図-4は、クラスターセンター数を5とした場合のパターン分類の結果である。また表-1は、各クラスターセンターに対する欠測をもつサンプルデータの類似度である。この類似度のうち上位2つを適用しその重み付き平均より欠測の補完を行う。図-5は、実際の観測流量と補完された結果を表したものである。これによると、流量が平均化されて補完されているのがわかる。実際の流量が特殊であるため各クラスターにそれが反映されていないためと思われる。しかし、形状的にはうまく示されている。なお、詳しい適用結果については、講演時に述べる。

<参考文献>

梅田浩之・小尻利治：ファジイ理論を用いた流況のパターン分類と長期流況シミュレーション、土木学会中部支部研究発表会講演概要集、pp163~164、1992.3

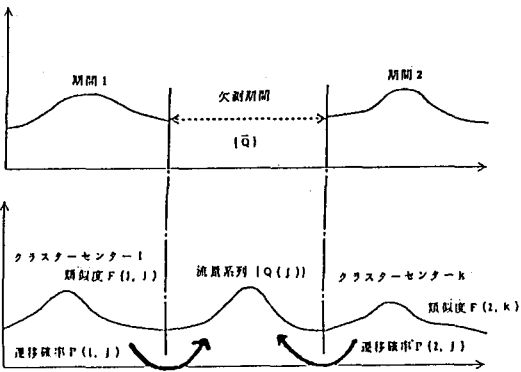


図-3 長期間観測データが欠測している場合の補完の説明図

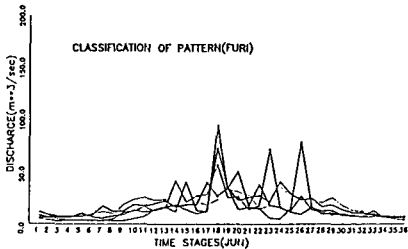


図-4 パターン分類の結果

表-1 各パターンへの類似度

PATTERN	DISTANCE	SIMILARITY
1	3.075013	0.692499
2	2.034537	0.796546
3	1.798587	0.820141
4	2.886113	0.711389
5	2.398450	0.760155

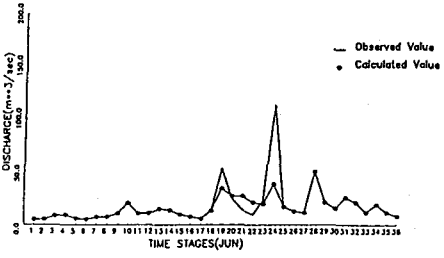


図-5 補完の結果