

パターン認識を用いた渇水流況の評価

岐阜大学工学部 正 員 小尻利治

岐阜大学工学部 学生員 ○小林直行

1. まえがき 長期的な気候変動、渇水が発生している今日、時系列データをベクトルと見なし、そのベクトルでの解析は渇水対策上重要な情報を与えてくれよう。本研究では、Unnyらが導入したパターン解析を導入し、classification手法の有効性、目的関数の妥当性について考察する。さらに、ベクトルとしての渇水流況の予測手法を提案する。

2. 水文データに対するパターン分類化手法の適用と比較

(1) random K-means : 従来のK-means手法では、全体的な最適解を求めることは困難で、局所最適解に収束することが多い。その収束解は全体の最小値よりかなりはずれる場合がある。そこで、初期値を乱数で与えるK-means手法をもちいて最適解の達成を試みる。この手法がrandom K-means手法である。

(2) DHB : DHB手法は、random K-means手法と同じ方法で初期のクラスターセンターを任意に定めて計算を始める。今、クラスターsに属しているサンプルが次式を満足するならば、クラスターrへ配置される。ここで V_j はクラスターjにおけるサンプルの総数である。

$$V_s |X(i)-Z(s,u)|^2 / (V_s-1) > V_r |X(i)-Z(r,u)|^2 / (V_r+1) \quad (1)$$

$$V_r |X(i)-Z(r,u)|^2 / (V_r+1) = \min |V_j |X(i)-Z(r,u)|^2 / (V_j+1)| \quad (2)$$

(3) ABF : ABFは、深さ方向を基本とするDHFと、横方向を基本とするDHBを繰り返し適用し、解の改善をはかっていくものである。

3. 目的関数

水文時系列を分類するためには、系列の相違を評価することが重要であり、この目的関数は次の距離として定式化される。

$$OF_1(X(i), Z(j)) = \max_t |x(i,t)-z(j,t)| / z(j,t) \quad (3)$$

ここで、 $x(i,t)$ は時刻t、サンプルiの観測値を対数変換した値である。 $z(j,t)$ は時刻t、クラスターjの要素であり、 $X(i)$ と $Z(j)$ はそれぞれ要素 $x(i,t)$ と $z(j,t)$ から構成されたベクトルである。ベクトル間の距離はクラスターセンターによって正規化されている。

第二の目的関数は、時刻における流量の変化率を評価するものであり、やはり、クラスターセンターの変化量で正規化される。すなわち

$$OF_2(X(i), Z(j)) = \max_t \frac{|(x(i,t)-x(i,t-1))-(z(j,t)-z(j,t-1))|}{|z(j,t)-z(j,t-1)|} \quad (4)$$

なお、すべての関数を統合した目的関数は次のように定式化される。

$$OF_3(X(i), Z(j)) = \max \{OF_1(X(i), Z(j)), OF_2(X(i), Z(j))\} \quad (5)$$

次に、最適なクラスター数を決めるために二種類の方法、すなわち、(i)前述した内部距離に対して外部距離を定義する。これは各クラスターセンター間の距離の最小値として与えられる。そして、外部距離が最大、内部距離が最小となるようにクラスター数を決める。(ii)各クラスターに属するサンプルのそのクラスターのセンターに対する内部距離が正規分布をしていると考える。その時、分布の信頼度を表わす基準としてAICを適用し、AICを最小にするクラスター数を最適とみなす。

4. 渇水流況の評価

渇水評価の方法については、(i)与えられた任意の流況をRとし、まず、このRの属するクラスターj

を求める。(ii)クラスター-jでのRの内部距離を DK_j とする。 DK_j が $N(0,1)$ の正規分布に従うとみなし、jクラスターにおけるRの非超過確率 P_j を求める。すると、クラスター-jの生起確率 Ph_j は次の式で表わされる。

$$Ph_j = n(j)/N \quad n(j): j \text{ クラスターに属するベクトルの数、} N: \text{総ベクトル数} \quad (6)$$

ここで、(iii)総流量の平均がjクラスターよりも小さいクラスターの生起確率の和を P_k とすると、この流況の非超過確率Pは以下の式で表わすことができる。

$$P = P_j * Ph_j + P_k \quad (7)$$

最後に、将来の渇水流況の予測方法として、(i)現時刻までに観測された実際の流況と、それぞれのクラスターセンターとの内部距離を求める。(ii)各センターに対する内部距離の逆数を重みとして、すべてのクラスターセンターを加重平均した値を予測値とする。各クラスターセンターを Z_i 、各クラスターに対する重みを W_i とすると、予測ベクトル Pr は次のようになる。

$$Pr = \sum (W_i * Z_i) / \sum W_i \quad (8)$$

5. 適用と考察 CanadaのLemsfordにおけるSouth Saskatchewan Riverの流量データで適用を行なう(1912年1月から1986年12月までの75年間の月平均流量)。まず、75個のデータに対してクラスター数を4とし、random K-means手法による計算を100回実行した。図1は $\max\{OF_a\}$ (内部距離 DK)の値の変動を示している。内部距離 $\max\{OF_a\}$ の最小値は6.081であった。分類化されたクラスターセンターを図2に示す。内部距離 $\max\{OF_a\}$ の最小値は100回中22回出現している。この結果は(100/22)回以上の計算回数で、最適解が得られるであろうことを示している。一方、75個のデータに対して異なった4個の初期センターを抽出する組み合わせの数は ${}_{75}C_4 (=1215450)$ である。この時、センターの選び方の総数に対して最適解

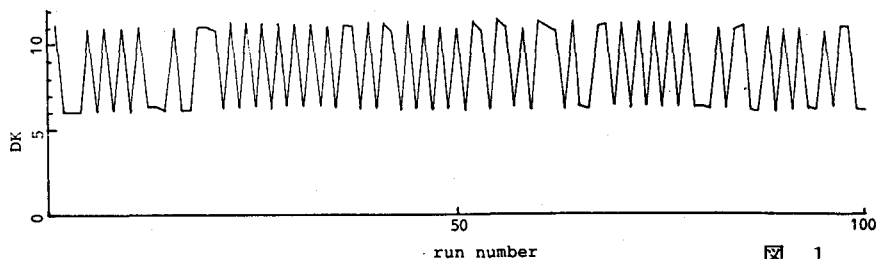


図 1

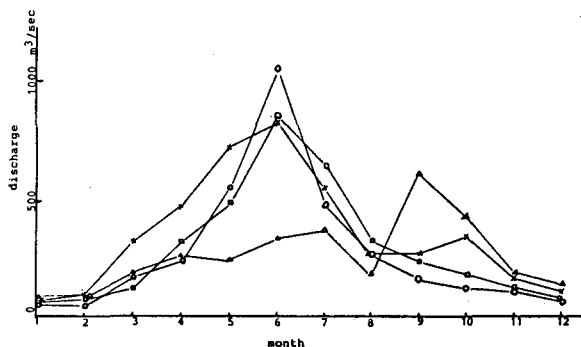


図 2

が得られる回数の比率は $(100/22)/1215450 = 3.74 \times 10^{-6}$ である。従って、クラスター数がK個の時には、最適解を得るために必要な計算回数は、 $3.74 \times 10^{-6} \times {}_{75}C_k$ となるであろう。DBH、AFBの2方法について、また、流況の評価・予測の最終結果については講演時に述べる。

〈参考文献〉 R,Gnanadesikan:Methods for Statistical Data Analysis of Multivariate Observations
Ismail,M.A.and Kamel,M.S.:Multidimensional Data Clustering Using Hybrid Search Str-
ategies,University of Waterloo,unpublished.