

ランレングス符号化法の解説

ランレングス符号化法とは、同一階調が連続する特性を持ったデータを圧縮する場合によく用いられるデータ圧縮手法であり、G3 ファックスにおける画像圧縮の基礎にも用いられている。レーダー観測に基づく降水量のデータは、降水のない領域が連続して広がる特性をもっているため、データ圧縮手法としてはランレングス符号化法が最適である。気象庁では、降水短時間予報関連の GPV(国内二進形式格子点資料通報式)等でもこの手法を取り入れたデータ圧縮を行っている。

1. ランレングス(runlength)符号化法とは

1次元に連続したデータがある場合、同じ値のデータの継続する長さがランレングスと呼ばれている。ある値とその値が連続する数(ランレングス)を1つのセットとし、セットをつなげることによって1次元に連続したデータを表現する手法がランレングス符号化法である。このセットの作り方(符号化手法)には何種類かの手法が提案されている。一般的なランレングス符号化法については、市販されているデータ圧縮や画像処理のテキストを参考にされたいが、ここでは GRIB2 資料テンプレート 7.200(気象庁定義資料テンプレート)で利用している符号化手法について簡単に説明する。

2. GRIB2 資料テンプレート 7.200 における符号化手法

a. 格子点値の取りうる値

値は2次元矩形領域の格子点上に存在し、0以上 MAXV 以下の整数を取る。ここで MAXV は、GRIB2 資料表現テンプレート 5.200(気象庁定義資料表現テンプレート)第 13 ~ 14 オクテットにて通報される値である。

b. 2次元データの1次元化

主走査方向を2次元矩形領域の左から右(通常西から東)、副走査方向を上から下(通常北から南)として、2次元データを1次元化する。

c. 圧縮後の1格子点値当りのビット数(NBIT)

圧縮されたデータ列の中で、1格子点値が占めるビット数であり、ランレングスのデータもこのビット数が用いられる。NBIT は GRIB2 資料表現テンプレート 5.200 第 12 オクテットで通報される。

d. 1セット内の値とランレングスの配置

圧縮されたデータ列の中で0以上 MAXV 以下のデータは各格子点の値とし、MAXV よりも大きなデータはランレングスの値とする。1セットは、まず値を配置し、もしその値が連続するようであれば後ろにランレングスを付加することによって作られる。MAXV よりも大きなデータが続く場合はすべてそのセットのランレングスの情報であり、MAXV 以下のデータが現れた時点でそのセットは終了し、この MAXV 以下のデータは次のセットの値となる。また、同じ値が連続しない場合はランレングスは付加されず、次のセットに移る。

e. ランレングスの作成手法

$(2^{NBIT} - MAXV)$ よりも大きなランレングスが必要となった場合、1データでは表現することができない。そのような場合、2データ以上を連続させてランレングスの情報を表すが、連続したデータの単純な総和をランレングスとするのでは圧縮効率があがらないため、LNGU $(=2^{NBIT} - 1 - MAXV)$ 進数を用いてランレングスを表現している。値のすぐ後に続くデータを LNGU 進数の1桁目 $RL_1 = \{LNGU^{(1-1)} \times (data1 - (MAXV + 1))\}$ 、それ以降 n 番目に続くデータは LNGU 進数 n 桁目 $RL_n = \{LNGU^{(n-1)} \times (data_n - (MAXV + 1))\}$ の情報とし、ランレングスはそれらの総和 + 1 ($RL = RL_i + 1$) となる。

3. 圧縮データ例

NBIT=4, MAXV=10とした場合、LNGU=5となり、

圧縮データ列 = {3 9 12 6 4 15 2 1 0 13 12 2 3} は

{3 9 9 6 4 4 4 4 4 2 1 0 0 0 0 0 0 0 2 3} と展開される。