

ネットワークMT法データの圧縮法について

高知大学理学部地質学教室 村上英記

Data Compression Method for Network-MT Data

MURAKAMI Hideki

Department of Geology, Faculty of Science, Kochi University

1. はじめに

現在、西日本ならびに東日本で実施されているネットワークMT法による観測では、NTT交換局のアースを使った電場データは、デルタ・シグマ型の20bitsのA/D変換器を使って10秒間隔でサンプリングがおこなわれ、中心局に設置されたデータ収録装置に収録されている。通常は、この中心局に設置してある収録装置に電話回線を使って接続し、1日分のデータごとに回収している。1中心局につき1日分のデータとして443,563bytesのデータが収集される。通常、2カ月程度の観測をおこなうので、特に手を加えなければ1中心局につき約27Mbytesのデータを保管する必要がある。現実には、3から5カ所のネットワークMT法の観測を平行しておこなうので、100Mbytes近いデータを保持する必要がある。

このような大量のデータを研究者間で交換する方法としては、例えばMOなどの大容量の記憶媒体を使うか、インターネット上のファイルサーバーにデータを置いてアクセスする方法が考えられる。MOなどの媒体を使う場合には、データを管理している人にコピーを作成してもらう必要があり、作業にかかる時間と手間を考えると最良の方法とは言えない。インターネットを使ったデータ交換の場合には、大量のデータを転送するとネットワーク全体に負荷をかけることになるので、なるべく転送するデータ量を減らす工夫が必要になる。そこで、一般におこなわれているのが、圧縮プログラムを使い転送するデータ量を減らすことである。

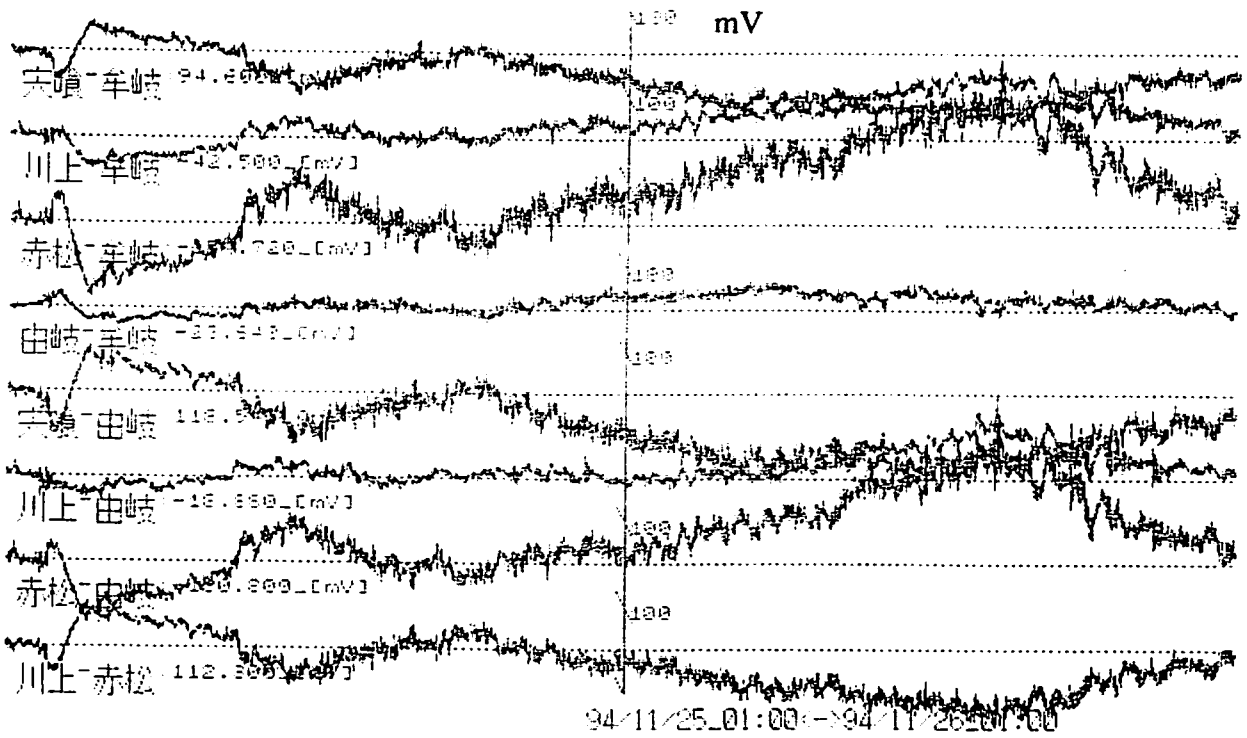
今回は、Network-MTデータとして取得された電場と磁場データを使い、これらのデータを圧縮するには、どのような圧縮方法を使えばどの程度に圧縮ができるかということについて簡単な考察をおこなった。

2. データ圧縮法とその結果

データ圧縮の検討には、1994年10月から12月までの期間に四国南部でおこなわれたネットワークMT観測により得られた牟岐ネットの電場データ（8チャンネル）と馬路で取得された磁場3成分データの中から11月25日1時から26日1時までの24時間分のデータを使用した。第1図に使用した電場データを第2図に磁場データを示す。

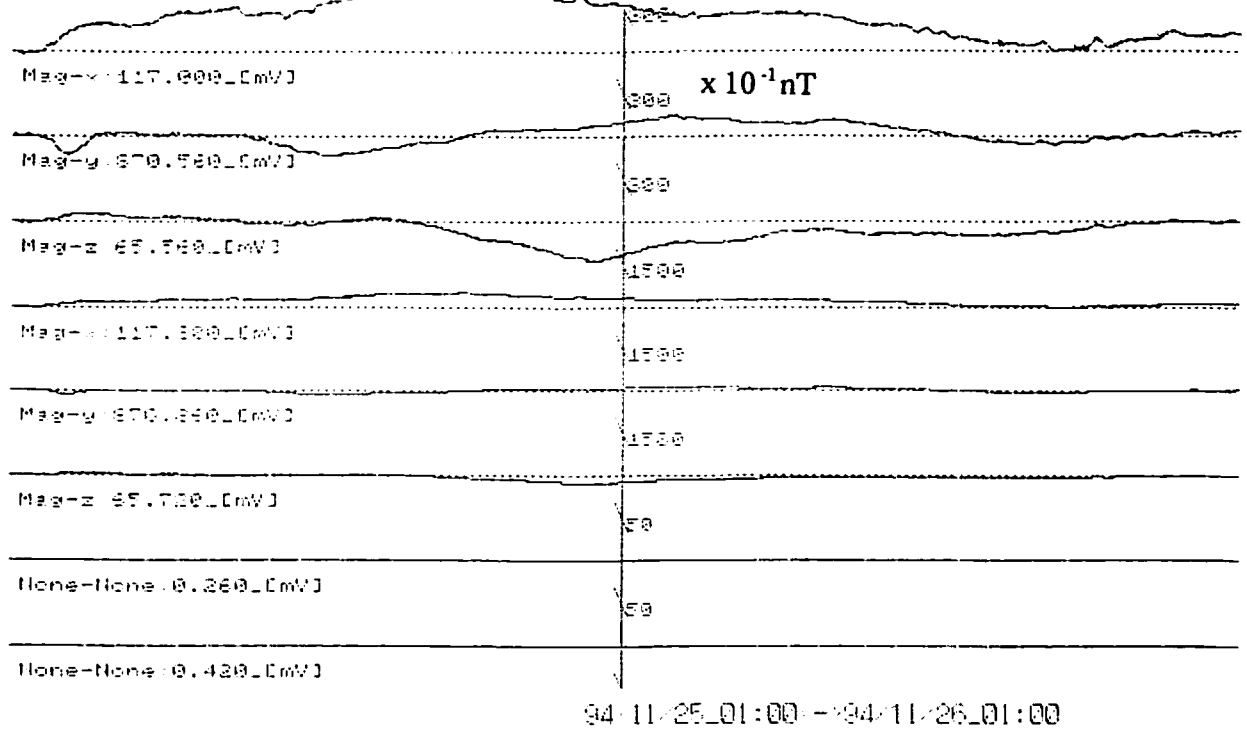
それぞれの電場と磁場のデータは、10秒サンプリングされた20bitsデータであるから本来各チャンネルのデータ量は24時間分で21,600bytesとなり、1中心局（8チャンネル）あたり172,800bytesとなるが、現在電話回線を使って転送されるデータはバイナリ形式ではなく16進符号化したアスキーコードとして転送されているので、1中

牟岐



第1図 電場データ (牟岐)

馬路



第2図 磁場データ (馬路)

心局8チャンネル分のデータ量としては443,563bytesとなっている。以後の圧縮率の計算には、この値を使って計算することに注意されたい。

まず、UNIXやMS-DOSと言ったOS上で広く使われているCOMPRESS、LHa、GZIPの3種類の圧縮プログラムでどの程度圧縮できるかを検討してみた。ここでは、電話回線を通じて得られた形式のファイルそのものを汎用の圧縮プログラムで圧縮している。その結果を第1表に示す。どのプログラムが優れているとは一概には言えないが、圧縮率のみに注目すればLHaとかGZIPといった比較的最近作成されたプログラムの圧縮率がわずかではあるが良く、電場データで約38%程度、磁場データについては約29%程度に圧縮される。もちろんバイナリ形式で取得した場合を基準にすると、圧縮率は80%から98%程度にしかになっていない。

	電場データ (牟岐)	磁場データ (馬路)
COMPRESS	175,235bytes (39.5%)	147,661bytes (33.3%)
LHa	167,880 (37.8%)	124,537 (28.1%)
GZIP	167,810 (37.8%)	130,522 (29.4%)

第1表 汎用圧縮プログラムによる圧縮

圧縮率の計算は、ネットワークMTで通常取得しているフォーマットでの1日分のファイルの大きさ443,563bytesを基準にしておこなっている。

次に、これまでに幾つかの電場や磁場のデータを使ってデータ圧縮法を検討してきたなかで、計算アルゴリズムが簡単で計算時間のかからない方法としてデータの差分をHuffman符号化するという方法^{1),2)}を今回のネットワークMTデータに適用してみた。ここでは、得られたデータを各チャンネルごとに切り出し、各チャンネルのデータを1つの時系列データとして扱っている。差分を取る目的は、データの振幅の頻度分布の分散をより小さいものに変換する目的で差分をとっている。実際には差分処理をおこなうごとに0の数を数えて、0の数が最大になる時の時系列データをHuffman符号化している。今回検討に使用した電場データも磁場のデータも1階差分をおこなった時が、0の数が一番多く圧縮率が良いという結果を得た。なお、今回使用したプログラムのHuffman符号化部分は10bitsデータに対応したものである。圧縮結果を第2表に示す。圧縮率は電場データで平均して約20%、磁場データについても約13%程度まで圧縮できている。第1表に示した汎用圧縮プログラムの結果よりも圧縮率が高く、汎用の圧縮プログラムの半分近くまで圧縮されていることがわかる。バイナリ形式で取得した場合を考えても、約34%から52%程度に圧縮ができています。

電場データ (牟岐)		磁場データ (馬路)	
CH1.	10,934 bytes (19.7%)	X	8,493 bytes (15.3%)
CH2.	10,431 (18.8%)	Y	7,417 (13.3%)
CH3.	12,697 (22.9%)	Z	7,445 (13.4%)
CH4.	10,954 (19.7%)		
CH5.	12,291 (22.1%)		
CH6.	10,528 (19.0%)		
CH7.	11,752 (21.2%)		
CH8.	11,402 (20.6%)		

第2表 差分法とHuffman 符号化の組み合わせによる圧縮

圧縮率の計算には、ネットワークMTで通常取得しているフォーマットでの1日分のファイルの大きさ443,563bytesの八分の一の55,445bytesを基準として計算している。

3. まとめ

今回検討した差分法とHuffman符号化の組み合わせといった簡単な方法で、現在取得しているネットワークMTのデータ形式でも十数%から20%程度までに圧縮できることがわかった。これは汎用の圧縮プログラムを用いた場合のさらに半分程度である。計算時間もそれほどかからない比較的簡単なプログラムなので、実用的なプログラムに仕上げて公開したいと考えている。

今回使用したデータは、比較的S/N比の良い差分法で取り扱うのに都合の良いデータであったのでこの程度に圧縮できたが、もっとS/N比の悪いデータの場合には圧縮率は当然悪くなる。多くの場合には人為的な活動によるノイズが入ってくる場合で、このような場合には例えば時変ARモデルやWavelet変換などのような信号の性質の時間的な変化を組み込むようなモデルを導入して検討する必要がある。

この研究には、東京大学地震研究所の平成6年度後期一般共同利用研究課題『地磁気・地電流データの圧縮に関する研究』の経費を使用した。

参考文献

- 1) 村上英記, MT法データの圧縮について, Conductivity Anomaly 研究会1993年論文集, 1-6, 1993.
- 2) 村上英記, 地球電磁気データの圧縮方法について, Conductivity Anomaly 研究会1994年論文集, 154-159, 1994.