

研究タイトル:

可逆データ圧縮



氏名:	中村 博文 / NAKAMURA Hirofumi	E-mail:	
職名:	教授	学位:	博士(工学)

所属学会・協会: 電子情報通信学会, 情報処理学会, IEEE

キーワード: データ圧縮・情報源符号化, 正整数符号

技術相談
提供可能技術: ・ファイルデータの圧縮
・アルファベットサイズが巨大な情報源の符号化

研究内容:

応用分野や背景

データ圧縮・情報源符号化の技術や手法は、データの内容を置き換えて、データ量を少なくして保存や伝達をし、利用時や受信時に元に戻すことで、記憶領域や通信時間の節約をもたらす。このことについて効率や性能が良いというのは、多くの場合に、データ量をより小さく変換できることを指す。元と全く同じデータに戻せる『可逆』の場合を中心に研究を行っている。

作成・送信側 保存や伝達 受信・利用側

データ ⇒ 圧縮・符号化 ⇒ 圧縮データ ⇒ 復号 ⇒ 元と同じデータ

圧縮対象のデータの性質が、定式化のし易い理想的な場合には、圧縮前データの記号あたりに対する圧縮後データの平均的最小データ量(情報エントロピー)が理論的に導出可能である。更に、性能が理論的限界に迫るような技術や手法が知られている場合もある。実際のデータに対しては、一定の仮定をおいて、その中で最適、もしくは準最適な技術や手法を使用することになる。

技術相談のシーズ1

データの種類や性質によって、既存のデータ圧縮・情報源符号化の技術や手法を取捨選択し、改良して適用し、最適もしくはある程度最適に近い効率化を図ることで、一般的には何割もの効率化(データを小さく変換)が図れる。

なお、無償のソフトウェアでも、ある程度の性能を有しており、それらの中には圧縮対象によって理論的限界に迫る性能を有する場合もあるため、無償のソフトウェア等で済ませるかどうかの選択もその都度検討事項になる。

技術相談のシーズ2

データの種類によって個別に技術や手法を使い分けるのではなくても、ある程度の効率化が図れる。

また、記号等の出現頻度そのものではなく頻度の順位などを用いるような多少緩慢な手法であっても、ある程度の効率化が図れる場合もある。

特に、現れる記号の種類の数(アルファベットサイズ)が大きい場合か、少ない場合でも複数個の記号をひとつの別の記号とみなすことによって記号の種類の実質的な数が大きな場合の、どちらかで、かつ、圧縮対象のデータ量が超巨大な場合に、効率が理論的限界に漸近する手法も提案されており、応用の際には検討に値する。但し、年を追うごとに扱うデータが増えている傾向にあるとは言え、実際の個々のデータは超巨大とまでは言えないため、適用の広さと引き換えに性能はある程度の所で止まることになる。

提供可能な設備・機器:

名称・型番(メーカー)

名称・型番(メーカー)	