

■研究テーマ

- データ圧縮、特に無ひずみ圧縮の数理的基礎を研究する
ユニバーサル符号の理論と応用
- 情報検索、情報検索のための文字列処理、データ圧縮の応用

■キーワード

データ圧縮、ユニバーサル符号、情報源符号化、データ埋め込み、情報理論、情報検索、記号処理

■産業界の相談に対応できる技術分野

各種符号化アルゴリズム

■主な設備

PC, ネットワーク

連絡先
情報理工学分野 横尾英俊
TEL:0277-30-1805 FAX:0277-30-1801
e-mail:yokoo@cs.gunma-u.ac.jp



横尾英俊教授

安川美智子助教

研究概要

データの様々な工学的側面を読み取る
データ処理手法の開発

私たちの研究室では、単一もしくは複数のデータを圧縮する手法の開発や、複数のデータを分類する手法の開発を手がけています。複数のデータを対象にすることで、応用の範囲も広がってきます。

研究室には、横尾英俊教授、安川美智子助教、大学院生、4年生が在籍しています。

ここでは、私たちが取り組んでいる「データの様々な側面を利用した」研究を紹介します。

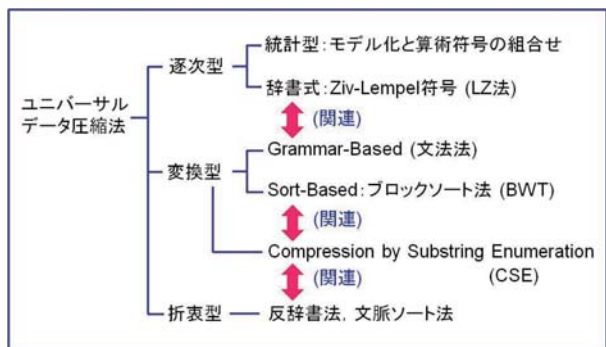
特徴と強み

膨大なデータの効率的な圧縮や
検索技術に対する圧縮手法の応用

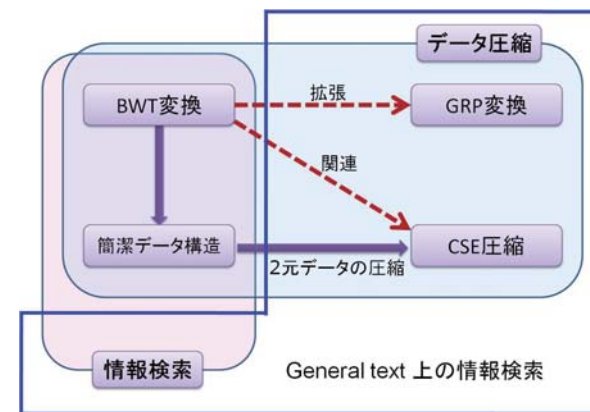
インターネットの普及に伴い、膨大な量のデータがネットワーク上で流通するようになっています。大容量のデータを保管したり、通信したりするた

めには、データを効率よく圧縮する技術が必要になってきます。

私たちは、圧縮対象のデータに対して予備知識を仮定しないユニバーサルな圧縮手法を研究しています。ユニバーサルデータ圧縮法にも多彩な手法が知られていますが、LZ法、ブロックソート法、モデル化と各種符号の組合せによる適応的符号化を主として取り上げています。また、ブロックソート法の基礎にあるBWT変換の拡張やCSEと呼ばれる新手法とBWT変換の関連も解析しています。



研究対象としているユニバーサルデータ圧縮法



データ圧縮と情報検索の統合

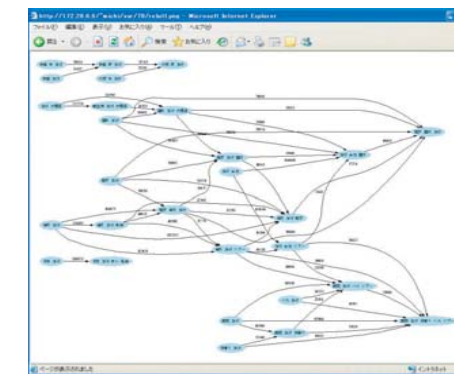
BWT変換はデータ圧縮だけではなく、情報検索や高速文字列検索の基盤技術として重要な地位をしめるようになってきました。データ圧縮と情報検索に共通する要素技術をそれぞれの目的に応じて洗練することで、データ圧縮と情報検索の統合をはかることが一つの目標です。情報検索を効率的に行うには索引の利用が欠かせません。文書中の任意の文字列の出現位置や出現頻度を高速に求めるために全文索引という構造が知られています。原文書と同じ情報量を持つ必要があるため大きくなりがちですが、これを原文書の大きさを超えないように圧縮したものが圧縮全文索引です。圧縮全文索引構成にもBWT変換が応用されています。

データが圧縮できるということは、その分すき間ができるということです。そのすき間に別のデータを埋め込んで、新たな目的に応用しようという試みにも挑戦しています。たとえば、画像圧縮とデータ埋め込み技術を組み合わせることで、埋め込み可能なデータの容量を増やしつつ、もとの画像データの劣化を抑えることや、埋め込む前のデータを完全に復元することが可能になります。また単独ではもとのデータを完全には復元することのできないデータを分散することで、秘密分散やデータのバックアップに利用することができます。

インターネット上で流通しているデータは膨大ですが、類似している情報が多いことも事実です。データが類似しているという側面は、データが冗

長であるという特徴としてとらえることができます。冗長なデータは圧縮しやすいデータですので、データの類似性検出にデータ圧縮を利用することができます。たとえば、インターネットを通じて配信されるニュース記事には多数の重複部分があります。そこで、類似の記事データをまとめて圧縮することで、冗長性を削減し、データ保存のためのディスク容量を節約することができます。

さらに、テキストデータの類似性を利用して、インターネットで検索したWebページをグループ化して表示する、クラスタリング検索の研究にも取り組んでいます。一般の多数のユーザが検索の際に入力した、検索に関係のある特定の語だけを取り出して、テキストデータの類似性の判定に利用することで、意味のあるWebページのグループ化が可能となります。



ウェブページの中に含まれる語の関連性を分析

今後の展開

「0」と「1」の集まりに
多彩で興味深い側面が

ハードディスクの中に記録され、また、ネットワーク上を流れるデジタルデータは「0」と「1」のビットの集まりにすぎません。しかし、無味乾燥に思えるデータの集まりを様々な角度から注意深く眺めてみると、多彩で興味深いデータの側面が見えてきます。現代の生活をより便利で豊かにするために欠かせないデータの処理手法を研究室のメンバーで力をあわせて研究しています。