

圧縮ベースパターン認識

本章では、はじめに圧縮ベースパターン認識で利用される圧縮アルゴリズムである、LZWについて説明し、その後、圧縮後のファイルサイズを利用する手法と圧縮

圧縮ベースパターン認識は、ツイッターデータ[?], 音楽, 画像解析, バイオインフォマティクスなど, 1次元に表現されたデータであれば何でも解析可能という利点がある.

LZW

圧縮アルゴリズムこの

は表??のようになる。

今読み込んだ記号をあらためて最長一致文字列として、その整数値の符号を記憶する。前のステップの例をそのまま用いると、バッファの状態はつぎのよう

類似部分の量と解釈することができる。

以上に説明したコルモゴロフ複雑性は計算不能として知られており、実際には用いることはできない。そこで実用的には、データ圧縮技術を

ター
ンを多く持っているため、似ていると解釈できる。

まとめると、オブジェクトに対する圧縮率の大小はつぎのように定義できる。

オブジェクトのランダム性が高い⇒複雑なオ

尺度と比べると,
 k -means
法や
SVM(Support
Vec-
tor
Ma-
chine)
など
ベク
トル
ベー
スの
パター
ン認識
手法の
利用に
適する.

PRDC

では, N 個の基底辞書集合
 $B = \{D_1, D_2, \dots, D_N\}$
を最初に定める.そして,
オブジェクト
 x をそれぞれの辞