

情報科学

(2) 情報の冗長性・情報の圧縮

目次

- 情報量
- 情報の冗長性
- 誤り検出・誤り訂正
- 情報圧縮
 - ▣ 可逆圧縮
 - ▣ 非可逆圧縮

予想ゲーム

1. 近くの人と組んで出題役と回答役をしよう
2. なるべく少ない質問回数で答えを当てる

□ ゲーム1 数当て

- 出題役の人は1～16の間の数を思い浮かべる
- 回答役は「〇〇より大きいですか？」と質問

□ ゲーム2 推しメン当て

- 出題役の人はAKB48選抜メンバー16人の中から推しメンを思い浮かべる
- 回答役は「〇〇より総選挙で上位ですか？」と質問

どっちの方が当てやすい？

- 当てにくいこと = 意外なこと = 情報量 大
 - ▣ たとえば、マニアックな推しメン
 - ▣ 休講 > 休講じゃない

情報量

□ 情報量と確率

- 頻繁に起きることを知る → 情報量 小
- 稀にしか起きないことを知る → 情報量 大

□ 情報量の単位

- 1ビット: 2つの等確率に起きる状態のうち、どちらかが起こったことを知ったときの情報量

情報の圧縮

情報の圧縮

□ 圧縮

- ▣ データの性質を保ったままデータの量を減らすこと
- ▣ 記憶や通信容量の節約のために行う

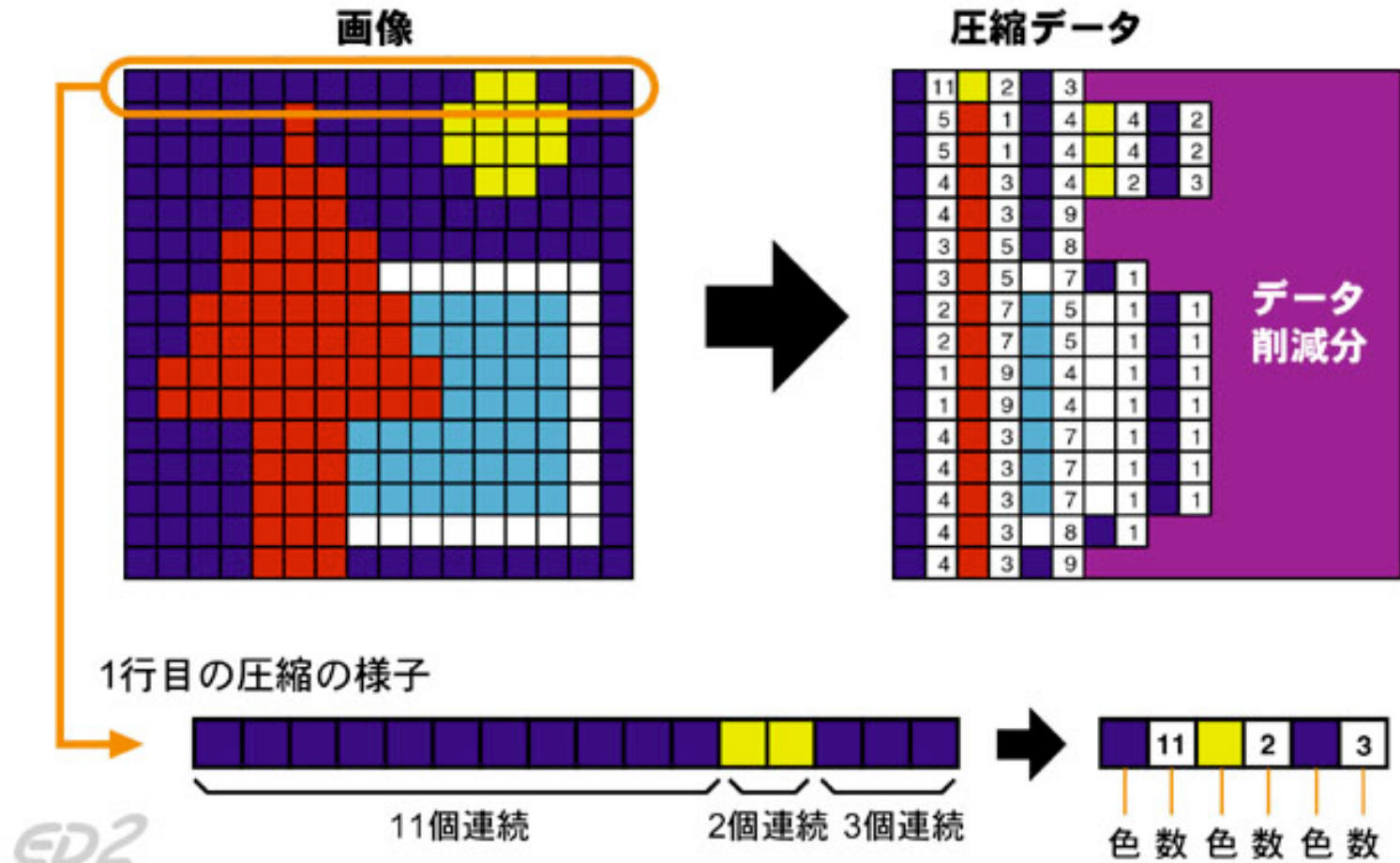
□ 展開

- ▣ 圧縮された情報表現から元の情報を復元すること

圧縮方法の原理1：可逆圧縮

- 圧縮→展開したとき、完全に元に戻る
 - ▣ 共通原理：データの偏りを利用する
- 文章やあまり大きくない画像などに用いられる
 - ▣ ZIP, LZH 等
 - ▣ GIF, PNG 等の画像
- 主な圧縮方法
 - ▣ ランレングス法
 - ▣ ハフマン法

ランレングス圧縮



ハフマン圧縮

The Mississippi is the longest river in the United States.

圧縮前のデータ量：57文字×8ビット＝456ビット

文字	出現数	対応コード
スペース	9	0
i	8	100
e	7	101
s	7	1100
t	6	1101
n	3	11100
h	3	11101
p	2	111100
r	2	111101
l	1	1111100
o	1	1111101
g	1	11111100
v	1	11111101
d	1	111111100
a	1	111111101
T	1	1111111100
M	1	1111111101
U	1	1111111110
S	1	1111111111

左表の対応コードを使って各文字を
置き換えた時のデータ量：248ビット

111111110011101101011111110110011001100
T h eスペース M i s s

11001100100111100111100100010011000110111101
s s i p p lスペースi sスペースt h

10101111100 1111101 111100 111111001011100 1101
eスペース l o n g e s t

01111011001111110110111110101001110001101
スペース r i v e rスペースi nスペースt

1110110101111111110111110001110010111111100
h eスペース U n i t e d

0111111111111110011111110111001011100
スペース S t a t e s

圧縮方法の原理2: 非可逆圧縮

- 圧縮→展開したとき、完全には元に戻らない
 - ▣ 共通原理: 人間が強く知覚できない情報を削減する
- 画像・音声・動画などに用いられる
 - ▣ 情報量が大きいので、高い圧縮率が求められる
 - ▣ 少々の変化では人間にはわからない
- JPEG等の画像、mp3等の音声

音声データの非可逆圧縮

- 人間の聴覚特性
 - ▣ 可聴域: 周波数 20Hz ~ 15,000Hz
 - ▣ 低周波数の方が強く知覚される
- 音声を様々な周波数成分の和として表す
 - ▣ 可聴域外の音は削除
 - ▣ 高周波成分は少ないビット数で表す

画像・動画データの非可逆圧縮

- 画像を様々な周波数成分の和として表す
- ブロック単位で処理を行う
 - ▣ 近い場所には近い色の情報が固まっている
 - ▣ ブロックノイズの原因
- 動画圧縮で用いられる手法
 - ▣ 1つ前の画像との差分
 - ▣ 動きの予測



冗長性の利用

情報の冗長性

- 情報を表現するのに最低限必要なビット数 (A)
- 情報を表現するのに実際に使ったビット数 (B)
- 通常は $B > A$ である
- 冗長度: その差や比のこと
 - ▣ 冗長度・冗長度量などと呼ぶこともある

冗長性の利用

- 誤り検出・誤り訂正
 - ▣ 冗長性を付加することでデータに発生したエラーを発見・修正する
- 情報圧縮
 - ▣ 冗長性を少なくしてデータ量を減らす

誤り検出の例：バーコード

- 最後の桁は他の数字から計算した数字
 1. 偶数桁の数字の和×3 + 奇数桁の数字の和
 2. 10 - 1で計算した数の下1桁の数
- 計算結果と最後の数が一致していない
→バーコードの読み取りに失敗したことがわかる
- 本のISBNやセンター試験の受験番号も同様

(例) 456995111617 □ ← 求めるチェックデジット

	JAN企業(メーカー)コード(9桁)									アイテムコード			チェック デジット
桁 番 号	13	12	11	10	9	8	7	6	5	4	3	2	1 桁目
例	4	5	6	9	9	5	1	1	1	6	1	7	9
偶数位置		5	+	9	+	5	+	1	+	6	+	7	
奇数位置	4	+	6	+	9	+	1	+	1	+	1		

① $(5 + 9 + 5 + 1 + 6 + 7) \longrightarrow 33 \times \overset{\textcircled{2}}{\downarrow} 3 = 99$

③ $(4 + 6 + 9 + 1 + 1 + 1) \longrightarrow 22$

④ $\longrightarrow 121$

⑤ $\longrightarrow 10 - 1 = 9$

答 “9” が求めるチェックデジットになる