

デジタル信号処理

第11回

量子化と符号化

情報源圧縮符号化とは

- 信号を伝送あるいは保存するために必要となるデータの大きさは、小さいほど望ましい。
- 一般に、信号には冗長な部分が含まれており、冗長性を除去することによってデータの大きさを縮小することができる。このような操作を、**データ圧縮**、**情報源圧縮符号化**、**情報源符号化**、**圧縮符号化**などと呼ぶ。

- 以下しばらく、便宜上**データ圧縮**という言葉を用いる。
- パソコンのユーザは、zip 形式のファイルを取り扱っていると思われるが、これはデータ圧縮の形態のひとつである。
- 画像によく用いられる jpeg 形式や、音楽によく用いられる mp3 形式もデータ圧縮のひとつである。

- zip 形式と, jpeg および mp3 には, 顕著な相異がある.
- zip 形式では, 適切な処理をおこなうことによって, 圧縮前のデータを完全に復元することができる. このようなデータ圧縮を可逆圧縮, 可逆圧縮符号化, ロスレス符号化, 歪みなし符号化などと呼ぶ.

- jpeg 形式や mp3 形式では, 圧縮前のデータの一部が失われており, それを完全に復元することができない. このようなデータ圧縮を**非可逆圧縮, 非可逆圧縮符号化, ロッシー符号化, 歪みあり符号化**などと呼ぶ.
- 一見, 可逆圧縮は無条件に非可逆圧縮より優れているように見えるかもしれないが...

- 一般に、非可逆圧縮は、人間の感覚の特性を利用し、人間が知覚しにくい情報を削除することで、可逆圧縮よりもずっとデータの容量を小さくすることができる (たとえば、mp3 ファイルは wave ファイルの 1/10 程度の大きさになる).
- 可逆圧縮でも、方式によって効率に大きな差がある.

- 非可逆圧縮と可逆圧縮は併用されることもある。教科書には、送信が「非可逆圧縮のあとで可逆圧縮をおこなう」と書かれているが、受信側が元信号の完全な復元をしたい場合には非可逆圧縮はおこなわない。受信側は受信信号を復号して利用する。

- リアルタイムでデータ圧縮された音楽や動画などを再生する場合には、単位時間あたりのデータ量が問題になる。この観点からは、データ圧縮は、**CBR(Constant Bit Rate, 固定ビットレート)**、**VBR(Variable Bit Rate, 可変ビットレート)**、**ABR(Average Bit Rate, 平均ビットレート)**の3種類に分類される。

CBR, VBR, ABR に関する記述の典拠は <http://aviutl.info/>

- CBR は単位時間あたりのビットレートが一定の方式で、シークやエンコードが早くデータの容量が予測しやすいという利点がある一方で、データの容量の割に品質が良くないという欠点がある。

- VBR は単位時間あたりのビットレートが可変の方式で、データの容量の割に品質が良いという利点がある一方で、データの容量が予測しにくいという欠点がある。
- ABR は単位時間あたりの目標ビットレートを定めておく方式で、特徴は CBR と VBR の中間。

- リアルタイム性という観点から言うと、リアルタイム性が要求される場合にはデータ（音声や動画など）を短時間で区切って送信する一方で、そうではない場合にはデータの分割はそれほど必要とされない。

- さらに、符号化を時間領域でおこなうか、周波数領域でおこなうかに関連した分類があるのだが…
- 教科書の解説は不十分なので、音声・オーディオについて、電子情報通信学会知識ベース <http://www.ieice-hbkb.org/portal/> に基づいて説明する.

音声・オーディオ符号化

- 音声・オーディオ符号化は、大別すると、波形符号化、スペクトル符号化、ハイブリッド符号化の3種類に分類される。

波形符号化

- 波形符号化は声波形を忠実に量子化して伝送する方式.
- これを時間領域でおこなう方式に ADPCM と APC が, 周波数領域でおこなう方式に APC-AB, SBC, ATC, MPEG オーディオがある.

スペクトル符号化

- スペクトル符号化は音源と声道からなる音声合成モデルのパラメータを入力音声から抽出して伝送し受信側でモデルに従い合成する方式.
- これを時間領域でおこなう方式にボコーダ、波形補間があり、周波数領域でおこなう方式に正弦波符号化、IMBE がある.

ハイブリッド符号化

- ハイブリッド符号化は波形符号化とスペクトル符号化を併用する方式.
- これは時間領域のみでおこなわれ, マルチパルス符号化, CELP, ACELP がある.

ADPCM Adaptive Differential Pulse Code Modulation の略.

APC Adaptive Predictive Coding の略.

APC-AB Adaptive Predictive Coding with Adaptive Bit allocation の略.

SBC Sub Band Coding の略.

ATC Adaptive Transform Coding の略.

IMBE Improved Multi-Band Excitation の略.

CELP Code-Excited Linear Prediction の略.

ACELP Algebraic CELP の略.

画像符号化

- 電子情報通信学会知識ベースに基づき, 画像符号化技術の概要を述べる. その一部については, 後の講義でより詳しく説明する (すべてではない).
- 画像符号化は 5 種類の処理から成り (後述), 入力原画像にこれらの処理を施すことで, 圧縮されたビット列が生成される.

画像符号化の流れ

画像符号化では、以下の処理を順におこなう。

1. 前処理: 画像信号を符号化処理に適した形式に変換し (色変換; 色の表現については第 14 回), 雑音を除去し, 必要があれば信号の高周波成分を除去し (帯域制限), 後段の作業で使う画像の特徴を抽出する (画像解析).

2. 予測: 画素の近傍あるいは (動画像であれば) 隣接フレームから画素値を予測する.
3. 変換: フーリエ変換, 離散コサイン変換 (次回以降), ウェーブレット変換 (次回以降) 等によって画像を情報圧縮に適した信号分布に変換する.

4. 量子化: 信号の値を近似する. 量子化後に信号が取りうる値の種類は量子化前より少ないことが一般的 (極端な例が2値化). これにより大幅な情報圧縮が可能となる一方で, 原画像の完全な復元は不可能になる.
5. 符号割り当て: ハフマン符号, ランレングス符号, 算術符号などにより, 送信したいデータに符号を割り付ける.

- 画像復号は 5 種類の処理から成り (後述), 圧縮されたビット列にこれらの処理を施すことにより復号画像が得られる.

画像復号の流れ

1. 符号解読
2. 逆量子化
3. 逆変換
4. 予測復号
5. 後処理

各ステップは画像符号化における対応する処理の逆処理.

量子化

- 信号の値を事前に定められた有限個の値の中で最もそれに近いもので近似することを量子化と呼ぶ。量子化前の信号は連続値でも離散値でもよいが、量子化後に信号が取りうる値の種類は量子化前より少ないことが一般的。音声、画像のいずれに対しても適用される。

- スカラー量子化 と ベクトル量子化 の考え方を説明するために、長さ kN の信号 $\{x[0], \dots, x[kN-1]\}$ (各 $x[j]$ は実数あるいは有理数) を量子化しようとしている状況を考える ($k, N \in \mathbb{N}$; データ長を kN とするのは後の説明の便宜のため)

- **スカラー量子化**は, あらかじめ量子化後に信号が取り得る近似値の候補 L_1, \dots, L_l ($l \in \mathbb{N}$) を定めておき (候補の数は有限個), 個別の j ($0 \leq j \leq kN - 1$) に対して, 何らかの基準で $x[j]$ の近似値として適切な値を $\{L_1, \dots, L_l\}$ の中か選ぶ, という手法である.

- **ベクトル量子化**は, 信号を

$$\mathbf{x}[j] = (x[kj, \dots, x[(k+1)j-1]])^T$$

$(0 \leq j \leq N-1)$ という k 次のベクトルが並んだものと見做し, $\mathbf{x}[j]$ が取り得る値全体をまとめて作った集合 \mathbf{X} ($\mathbf{X} \subset \mathbb{R}^k, \mathbf{X} \subset \mathbb{Q}^k$ など) を L 個の領域 $\mathbf{R}_1, \dots, \mathbf{R}_L$ ($L \in \mathbb{N}$) に分割してからそれぞれの領域の代表点 $\mathbf{r}_m \in \mathbf{R}_m$ ($1 \leq m \leq L$) を定め, $\mathbf{x}[j] \in \mathbf{R}_m$ のときに, その値を \mathbf{r}_m で近似する, という方法.

- スカラー量子化，ベクトル量子化の双方に，色々なやり方がある．講義では，スカラー量子化の方法についてのみ簡単に述べる．
- ベクトル量子化は，信号に k 次のベクトルの並びに相当する構造がある場合には，スカラー量子化よりもデータの大きさの割に良い近似が得られる可能性がある

以下の記述の根拠は 電子情報通信学会 知識ベース

スカラー量子化

- $x[j]$ が取りうる値の集合 (\mathbb{R} , \mathbb{Q} の部分集合など) を, N 個の区間 $[x_i, x_{i+1}]$ に分割し ($0 \leq i \leq N-1$, $x_i < x_{i+1}$), 各区間から, その区間を代表する値 y_i を選ぶ. y_i を量子化代表値, i を量子化インデックス, N を量子化レベル, x_i を量子化閾値, $x_{i+1} - x_i$ を量子化ステップ幅と呼ぶ.

- $x[j]$ を y_i で近似したとき, $x[j] - y_i$ を量子化誤差と呼ぶ.
- スカラー量子化は, 信号の値が零に近い値を取るときの処理の仕方に応じて, ミッドトレッド型, ミッドライザ型, デッドゾーン型の3種類に分類される.

- **ミッドトレッド型**は, ある $[x_i, x_{i+1}]$ が零を内点として含む区間になっていて, この区間の量子化代表値を零とする方法.
- **ミッドライザ型**は, ある i に対し $x_i = 0$ (すなわち $[x_{i-1}, 0], [0, x_{i+1}]$ という形で区間を分割する方法. 量子化代表値の取り方には決まりはない.

- 電子情報通信学会知識ベースの定義では、ミッドトレッド型とミッドライザ型の違いは、信号の値が零である場合のみに着目した分類であることに注意.
- 信号の値が零に近いときにその近似値を零とし、かつその範囲を広く取る方法を、特に**デッドゾーン型**と呼ぶことがある.

- 量子化ステップ幅が量子化インデックスによらず一定のものを**線形量子化**, そうでないものを**非線形量子化**あるいは**非一様量子化**と呼ぶ.

非一様量子化には色々なやり方があり, 教科書図 11.3 は対数を例として示している.

最適スカラー量子化

- 信号 x の値が区間 (a, b) に分布しているとき…
- $l_0 = a, l_N = b$ とし ($N \geq 2$), l_k ($1 \leq k < N$) を定めることにより区間 (a, b) を分割する ($i < j$ なら $l_i < l_j$ とする). 各区間の量子化代表値を y_k とする ($0 \leq k \leq N - 1$).

区間 (a, b) は閉区間でもよいし, $a = -\infty, b = \infty$ でもよい. 教科書では議論を $[-L, L]$ という閉区間に限定している.

- 信号が取る値に関し, 確率密度関数 $p(x)$ が与えられているものとする.
- 量子化誤差の二乗の期待値 σ_q^2 (教科書ではこれを平均パワーと呼んでいる) は, 次式で与えられる.

$$\sigma_q^2 = \sum_{0 \leq k \leq N-1} \int_{l_k}^{l_{k+1}} (x - y_k)^2 p(x) dx$$

- 先の式において, 調整可能なパラメータは l_1, \dots, l_{N-1} と y_0, \dots, y_{N-1} の $2N - 1$ 個で, l_0 と l_N は固定されている.
- σ_q^2 を最小にする量子化器のことを**最適量子化器**と呼ぶが...
- これを設計するには信号 x の値に関する確率分布が必要だが, 一般に確率分布の正確な情報を得ることは困難.

- このため、最適量子化器は必ずしもよく用いられるわけではない。
- 最適量子化器を決定するための素朴なアイデアは、 σ_q^2 を l_1, \dots, l_{N-1} と y_0, \dots, y_{N-1} の関数と見做し、 σ_q^2 の勾配ベクトルが零となるパラメータを求める、というものである。

σ_q^2 が独立変数の滑らかな関数になっていることは保証されていないので、上記のアイデアがつねに妥当であるとは限らない。

- σ_q^2 の定義の中で l_k に関する項は $\int_{l_{k-1}}^{l_k} (x - y_{k-1})^2 p(x) dx$ と $\int_{l_k}^{l_{k+1}} (x - y_k)^2 p(x) dx$ で、前者および後者を l_k で偏微分すると、結果はそれぞれ $(l_k - y_{k-1})^2 p(l_k)$ と $-(l_k - y_k)^2 p(l_k)$ となる。これらの和を取ると、

$$(y_{k-1} + y_k - 2l_k)(y_{k-1} - y_k)p(l_k)$$

となる。十分条件であるが、

$$l_k = \frac{y_k + y_{k-1}}{2}$$

と取れば、この項を零にできる。

- σ_q^2 の定義の中で y_k に関する項は $\int_{l_k}^{l_{k1}} (x - y_k)^2 p(x) dx$ で、これを y_k で偏微分すると、結果は

$$\int_{l_k}^{l_{k1}} 2(y_k - x)p(x)dx = 2y_k \int_{l_k}^{l_{k1}} p(x)dx - 2 \int_{l_k}^{l_{k1}} xp(x)dx.$$

この項を零にするための十分条件は、

$$y_k = \frac{\int_{l_k}^{l_{k1}} xp(x)dx}{\int_{l_k}^{l_{k1}} p(x)dx}.$$

- 以上2種類の方程式を連立して解けば最適量子化器が得られるのだが、困難.
- 教科書に記載されている Lloyd-Max の方法は、上記の方程式を連立させずに交互に解くことで試行錯誤的に最適量子化器の候補を探索する方法. 収束性も最適性も保証されない.
- 1990年代に、Lloyd-Max の方法と動的計画法を組み合わせる手法の論文が出ている.

ベクトル量子化

- 量子化したい信号を, 添字を略して, x とする. x が取り得る値全体をまとめて作った集合 X ($X \subset \mathbb{R}^k$, $X \subset \mathbb{Q}^k$ など) を l 個の領域 R_1, \dots, R_L ($L \in \mathbb{N}$) に分割する. $\cup_{1 \leq l \leq L} R_l = X$, $l \neq m$ なら $R_l \cap R_m = \emptyset$ とする. $x \in R_l$ のときに, その値を r_l で近似する.

- $1 \leq l \leq L$ なる l を量子化インデックスと呼び, \mathbf{r}_l を量子化代表ベクトルと呼ぶ. 量子化代表ベクトルをまとめた集合 $\mathbf{C} = \{\mathbf{r}_1, \dots, \mathbf{r}_L\}$ をコードブックと呼ぶ. 量子化誤差は, 何らかの距離 $d(\cdot, \cdot)$ を用いて, $d(\mathbf{x}, \mathbf{r}_l)$ によって定義されるが, スカラー量子化と異なり, 距離の取り方に自由度がある.

- ベクトル量子化器の設計法のひとつは Lloyd-Max の方法を多次元に拡張したもので、提案者の頭文字を取って LBG アルゴリズムと呼ばれる (1980) が、そもそも最適量子化器が頻繁に用いられるようなものではないため、この講義では詳細は述べない (興味がある者は電子情報通信学会知識ベースを参照せよ).
- Web of Science で ‘optimal vector quantization’ をトピックに含む論文を検索したところ、ヒット件数は 13 件であった. 研究は低調であるが、ある程度最近まで論文の出版が見られる.

PCM

- **PCM** は **Pulse Code Modulation** の略.
- コンセプトは 1937 年に Alec Reeves によって提案された

<https://www.telegraph.co.uk/technology/connecting-britain/pulse-code-modulation/>

- PCM は一定の周波数でサンプリングされた信号を量子化して記録する方法.

- 基本は線形量子化 (リニア PCM などと呼ばれる) だが, 電話回線では, 対数に近い非一様量子化が用いられる.
- サンプリング周波数, 量子化ビット数などは用途により色々.
- コンパクトディスクは2チャンネル, 16ビット一様量子化, 標本化周波数 44.1kHz

典拠は電子情報通信学会知識ベース, 以下ではこの文献をを断りなく用いる.

- 公衆電話回線に関連した **G.711** という規格では、標本化周波数は 8kHz で、対数関数に近い関数で信号の振幅を変換した後で一様量子化をおこなう (結果的に非一様量子化になる)。音声では小振幅の信号の発生確率が大振幅のそれより大きいいため、小振幅の部分の刻みを細かく量子化することで、音節の劣化を抑えられる。

尾知 (監修), デジタル音声&画像の圧縮/伸張/加工技術, CQ 出版社, 2013.

- 信号の振幅変換の部分の方式が日本および米国と欧州および中華人民共和国で異なり, 前者には μ 則, 後者には A 則という名前が付けられている. いずれもスカラー s を振幅の絶対値が 1 未満になるように正規化してから, 以下の関数 f で振幅変換する.

$$\mu \text{ 則} : f(s) = \text{sgn}(s) \frac{\log(1 + \mu|s|)}{\log(1 + \mu)}, \quad \mu = 255.$$

$$A \text{ 則} : f(s) = \begin{cases} \text{sgn}(s) \frac{A|s|}{1 + \log A}, & 0 \leq |s| < \frac{1}{A} \\ \text{sgn}(s) \frac{1 + \log(A|s|)}{1 + \log A}, & \frac{1}{A} \leq |s| \leq 1 \end{cases}$$

(ただし $A = 87.6$ もしくは 87.7 .)

DPCM

- サンプルングされた音声信号において、隣接した標本の音量の差は小さいことが多いため、隣接標本の差分を取ってから量子化すると、単純に標本を量子化する場合と同等のデータサイズで、量子化誤差を小さくすることができる場合が多い。この手法を、**differential pulse code modulation (DPCM)** と呼ぶ。

予測符号化

- DPCM は、標本化された信号を x としたとき、 $x[n]$ と $x[n-1]$ の差が小さいことを期待して、 $x[n] - x[n-1]$ を量子化する方法だった。これは、 $x[n] \simeq x[n-1]$ という「予測」に基づいて、信号とその予測値の差分を量子化していることに相当する。

- $x[n]$ の値を線形フィルタなどを用いて隣接する信号の値から予測し, 予測誤差を符号化する手法をまとめて**予測符号化**と呼ぶ. 予測誤差が小さいほど, データサイズのわりに量子化誤差を小さくすることができる.

- wav ファイルでは, 予測器と適応量子化器 (信号の大きさに合わせて量子化ステップ幅を調整する方式) を組み合わせた **Adaptive Differential PCM (ADPCM)** が用いられている.
- A/D 変換器などでは, **デルタ変調器, デルタシグマ変調器, 補間型変調器** と呼ばれる予測符号化器が用いられる (詳細は略).

- 先に述べた波形符号化の各手法については、電子情報通信学会 知識ベース が電子情報通信学会の Web ページからダウンロードできるので、それを参照せよ。
- 先に述べた波形符号化の各手法以外にも色々な方法があるが、この講義では網羅的な解説はしない。