

# メディア処理, ビッグデータ処理における現実問題

京都大学学術情報メディアセンター

美濃 導彦

## 1. メディアの階層性

人間が情報を伝えるものはメディアと呼ばれるが、図 1 に示すように、メディアは階層性を持っている。情報は他人が見える形で表現され、この表現が物理的に他人に伝わって、結果的に情報が伝わる。物理的には通信のメディアだけしか存在しないが、表現が表現として伝わるので、そこに仮想的に表現メディアが存在すると考えることができる。情報の表現方法はさまざまであり、利用する表現メディアも、画像だけでなく、言語、音声などがある。情報科学では、音声と画像はパターン情報と呼ばれ、物理的な信号をサンプリングしたデジタルな信号の系列である。これに対して、言語表現はシンボル表現と呼ばれ、人間の概念を記号化したシンボルを介して表現されている。パターン情報をシンボル情報に変換する処理はパターン認識処理と呼ばれ、人間がしゃべる言葉を計算機認識させる技術、文字を読み取る技術や画像から人間や人の顔などを抽出する技術など実用化が進んでいる。表現メディアで表現された情報は、通信メディアで物理的に受け取る側に伝わり、表現メディアから情報が取り出される。情報を表現するのは、送る側であり、情報を抽出するのは受け取る側である。それぞれが別個の人間であるので、この過程が逆写像になっていることは期待できない。これは何を意味するのかというと、情報は一方向では伝わらないということである。余談であるが、情報を伝えるためには対話が必須であるということはこのモデルが示している。情報を表現する表現メディアの研究を通して情報を研究している分野をメディア処理分野と呼んでいる。

## 2. 表現空間

デジタル画像の持つ情報表現の可能性を考えてみよう。デジタル画像は、ご存知のように縦横に画素が配列されたものである。1000×1000の画像というのは、平面に縦横それぞれに1000画素、総計1000000画素の配列である。各画素はカラー画像でRGB各8bit程度の明るさを表現できる。これではちょっと数字が大きくなりすぎるので、単純化して10×10画素の濃淡画像を考えよう。各画素が8bitの表現が可能であるとすれば、何枚の異なる画像（異なるという意味は一つの画素の値が1でも違うものすべて）が存在するかというと、 $2^{800} \approx 10^{240}$ 種類である。ビッグデータの関連で、人類が蓄積している情報量が最近指数関数的に増えてゼットバイトオーダーになっているとか言われているが、このデータ量が $10^{21}$ 程度である。参考までに人間が100年生きるとして生きている総秒数は $10^{10}$ 程

度、画像の表現空間がいかに膨大かは理解いただけるであろう。この膨大な空間の中には、1枚の画像の1画素の明るさが1段階だけ違うという画像も含まれている。これに類するもの、仮に1画素における濃淡8段階程度の明るさの違いは同じ画像だと考えても800枚程度しかない。実世界のあらゆる場面を写真で撮るとして、その結果取得できる画像すべての集合を考えたとしても、この空間が埋まるとは考えにくい。

この表現空間を用いて人間が情報を表現する。人間が紙に絵を描く場合、まったく自由に描くことはできるが、書いたものを他人に見せて情報を伝えようとする場合は多くの制約を受ける。したがって、人間が描く画像はそれが他人にとって意味のあるものである限り、ある種の制約を受ける。この制約には人間の文化や自然の環境が影響を与えている。

このように考えてくると、膨大な表現空間の中に様々な制約を満たす画像が宇宙空間における銀河のようにある程度のまとまりをもって存在していると想像できる。制約としては、数学的、論理的なもの、実世界に存在するもの、人間にとって意味のあるもの、文化的なものなどが考えられる。それぞれの制約を満たす画像がこの膨大な空間の中にどのように分布しているのかを理解することは可能であろうか、これらの制約間の関係はどうなっているのかなど様々な疑問が湧いてくる。この問題は、宇宙の中の星雲の分布を考えるのと同じようにロマンティックで夢があるが、宇宙のように観測できないので、数学のツールや直感で考えていくしか方法はないかもしれない。

### 3. 情報の意味とビッグデータ

意味は形式的には図2に示すように、わからないものの集合からわかっているものの集合への写像として表現できる。画像認識処理を例に考える。今ここに1枚の犬の写っている画像がある。これを計算機に提示して写っている物体が犬であることを認識させる処理を考えてみよう。計算機は画像を処理して特徴を抽出し、それらの特徴を分かっている物体の特徴（予め抽出され辞書的に記憶されている）と比較する。その中で一番近いものがある程度以下の近さならばそれを「犬」と判断する。容易にわかるように、特徴の選定と辞書作りがこの種の処理のポイントとなるが、最近では画像データにその内容を表すコメントがつけられたものがWEBに大量に蓄積されている。これらのデータを活用して特徴空間を選定しその中で様々な意味に対応する特徴サブ空間を学習するという手法が一般画像認識と呼ばれる分野で主流となっている。ビッグデータという言葉が言いだされる前から、ビッグデータ処理を適用してきたのである。

特徴空間は表現されている画像から抽出されるので、表現空間の次元圧縮になっている。意味に基づいて特徴空間を選定するという事は、表現空間の中で対応する意味制約に基

づく部分を取り出し、その空間の記述法（次元圧縮法）を考えていることに対応している。空間の中で異なる意味に対応するサブ空間が重なる場合がしばしばあるので、データに基づいて、空間をゆがめたり、空間内での距離をデータに基づいて学習したり、様々な工夫がなされているのが現状である。数学的にすっきりした考え方が持ち込めるのではないかと期待している。

#### 4. 意味制約としてのモデル

現実世界を記述する意味制約は、多くの場合、制約式として表現される。制約に柔軟性を持たせる、言い換えればデータの多様性を吸収するために制約式には様々なパラメータが設定される。データと制約式のパラメータの数の関係を図 3 に示す。パラメータの数の方が多いとデータを処理するためのパラメータが決まらないので、制約不足という状態に陥る。この場合は、一般的に正則化や最適化手法が利用される。2次元世界から3次元世界を復元する問題はその典型例であり、不良設定問題と呼ばれている。

これに対して、大量のデータからモデル記述を探そうとする問題は、制約を表現するパラメータ数よりデータの方が多く、統計的手法が威力を発揮している。パラメータ数が少ない間は、統計的手法で十分であるが、パラメータ数が多くなり、空間の次元数が上がってくると、大量のデータでも高次元空間にはスパースにしか分布がなく、問題が急に難しくなる。ビッグデータが集められれば対応が可能かもしれないが、次元数の増加は指数関数的に空間を広げるので、データが物理的に集められるものではない。相対的にスパースなデータを用いて次元圧縮をするための工夫をして現実問題に対処している。典型的な例としてスパースコーディング(**sparse coding**)手法があげられる。図 4 にスパースコーディングに関して講演で用いたスライドをまとめて示す。

#### 5. まとめ

計算機の黎明時代に数学者が活躍した話は有名である。その計算機の進展により、現実問題はビッグデータとして蓄積され、エビデンススペースで社会が動こうとしている。データの解析に統計的手法がある程度は役に立つが、不十分なところも多く、新たな数学的理論基盤が求められている。量が多くなれば質が変わると言われているので、何らかのイノベーションが数学の分野と様々な分野の協働で生み出されることを期待している。

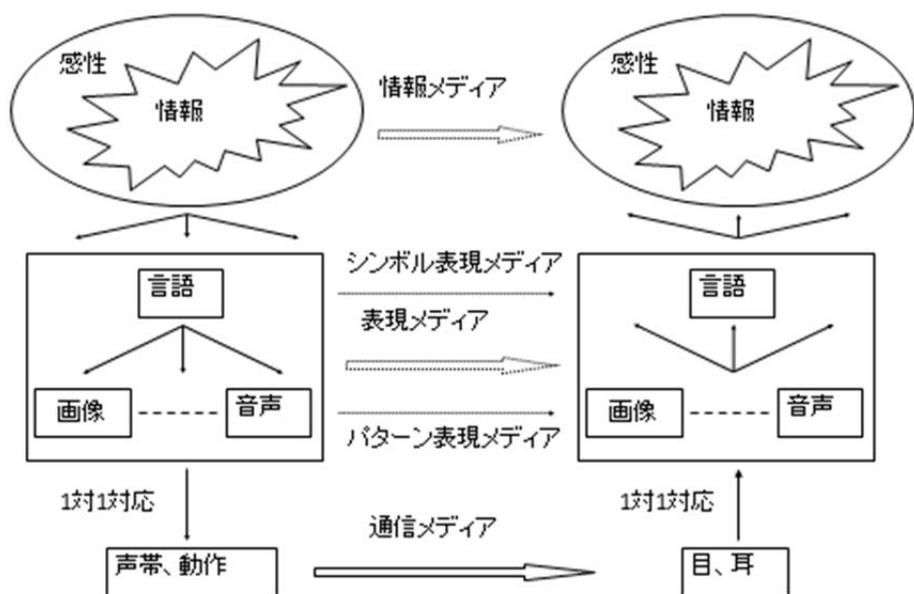


図1 メディアの階層性

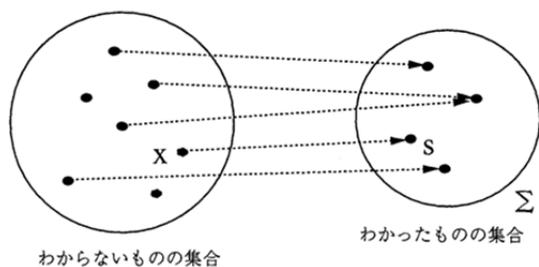


図2 情報の意味づけの写像

- 数学的(物理的)モデルの導入
  - 概念的にはデータ項とパラメータ(未知数)を含む方程式となる
- 方程式数とパラメータ数の関係により
  - 不良設定問題(解が不定): 正則化, 最適化手法
  - 解析解が存在する問題(ほとんど存在せず)
  - 統計的問題(解が不能): ビッグデータ, 統計的手法

図3 現実世界のモデル化

### Sparse coding

なるべく少数の基底で高次元データを表現

- ・データを効率的に表現できる
- ・データの解釈が容易になる
- ・生物の情報処理でも同様の原理が見られる



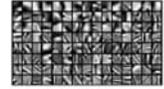
画像処理, 音声処理の幅広い分野に  
応用されている  
cf. 画像修復, 超解像, 顔認識 ...

### 画像の超解像度化

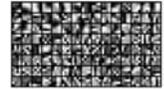
与えられた低解像度画像から、  
その高解像度画像を推定して生成する処理

#### 【学習段階】

1. 多数の高解像度画像が学習データとして与えられる。
2. ダウンサンプリングにより、対応する低解像度画像を生成する。
3. それぞれをバッチに分割し、低解像度画像とそれに対応する高解像度画像のバッチからなる辞書を作成する。



高解像度辞書



低解像度辞書

日野, 村田「スパース表現の数理とその応用」CVIM2012より

### 画像の超解像度化

与えられた低解像度画像から、  
その高解像度画像を推定して生成する処理

#### 【生成段階】

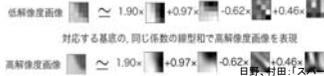
1. 複数の低解像度画像が与えられる。
2. 与えられた低解像度画像を低解像度辞書の少数のバッチの線形和で表現する。
3. 対応する高解像度辞書のバッチに、同じ係数をかけた線形和により、高解像度画像を生成する。



位置ズレ, ボケを含む複数の低解像度画像



推定された高解像度画像



推定された高解像度画像

日野, 村田「スパース表現の数理とその応用」CVIM2012より

### 定義

データ  $x$  を基底ベクトル  $\{d_k \mid k=1, \dots, m\}$  の線形結合で表現.

$$x = \sum_k c_k d_k \quad c = (c_1, \dots, c_m) \in R^m$$

この時, 少数の  $c_k$  のみが非ゼロを取るように,  $c$  を定めたい.

### 定式化

$$\min_{\{d_k\}, \{c^{(i)}\}} \sum_i \left\| x_i - \sum_k c_k^{(i)} d_k \right\|_2^2 + \lambda g(c^{(i)})$$

$$c = (c_1, \dots, c_m) \in R^m, \lambda > 0$$

$g(c)$  : 正則化項

### モデル化と解法

理想:  $g(c) = \|c\|_0$  (L0距離)  
組み合わせ問題, NP困難

現状:  $g(c) = \|c\|_1$  (L1距離)  
線形計画問題として解ける

他のモデル化:  $g(c) = \sum -\exp(-c_i^2), \sum \log(1+c_i^2), \sum \log(\cosh(c_i^2))$

組み合わせ最適化を連続関数最適化で近似する、より効率的なモデル化と、その高速・安定な解法が必要。