スパースモデリング

森本 祥之，本田 雄亮
Shoji MORIMOTO，Yusuke HONDA

1 はじめに

近年，ビッグデータを収集・解析することで，社会・経済の問題解決や製品の販売促進などを図る試みが多く行われている．ビッグデータは膨大で，複雑なデータ群であるが，不必要なデータが多く，活用するためには膨大なデータを解析する必要がある．ビッグデータから規則や知識を見つけ出す機械学習が重要な働きを担う．

現在注目を集めている機械学習の一つがスパースモデリングである．スパースモデリングは膨大なデータのほとんどの要素をゼロと考え，非ゼロ要素に着目することで，データの本質の抽出を手法である．

2 スパースモデリング

スパースモデリングはビッグデータを解析することで，大量のデータの中に隠れている有為な情報を抽出する情報処理モデルである．また，法則性を導き，断片的なデータを補完して実態を忠実に再現することが可能である．近年，インターネットの普及や計測技術の向上によりビッグデータの入手が容易になり，ビッグデータは様々な分野で利用されている．しかし，少ないデータ量しか収集できない場合も存在する．現在主要なデータ解析法の 1 つであるデータマイニングは，データ数が少ない場合，明確な結論を導き出すことが容易ではない．しかし，スパースモデリングは少ないデータから本質を抽出することで全体像を正確に構成することが可能である．[1)]

スパースモデリングでは，高次元データを証明する変数が観測データの次元数と比べて少数であるというスパース性を仮定する．説明変数を可能な限り減らし，データとモデルの適合を同時に満たす条件において，自動的に説明変数の選択を行う．式 (1) にスパースモデリングの数式を示す．

$$E(x) = ||y - Ax||_2^2 + \lambda\ ||x||_1 \tag{1}$$

$E(x)$: 最適化関数，x: 説明変数，y: 観測データ
A: 観測行列，λ: 正規パラメータ

式 (1) の右辺の第一項でスパースな解の候補を選ぶ．しかし，1 つの解に絞ることは容易ではないので，式 (1) の右辺の第二項で解を絞る．大規模な問題を効率的に解くスパースモデリングの基本的な方法として圧縮センシングがある．[2)]

3 圧縮センシング

圧縮センシングはデータがスパース性を持つとき，少数の観測データから原信号を推定あるいは再構成する技術である．主にデータの取得数が限られている場合での研究が行われている．スパース性とは，ほとんどの成分がゼロを持つ，または持つと期待される性質であり，スパース性を持つ解をスパース解という．圧縮センシングは，線形観測モデルに基づいて復号を行う．Fig.1 に線形観測モデルを示す．

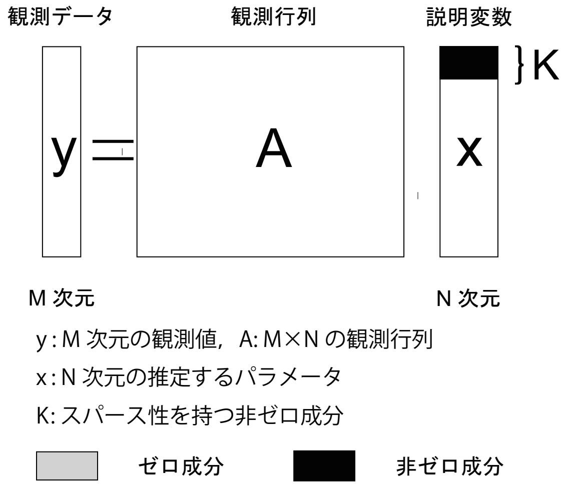

Fig.1 線形観測モデル

圧縮センシングは元データにスパース性を含んでいることが重要であり，スパース性を含まない場合は解を一意に決めることが容易ではない．ベクトル y=Ax を例にとると，M=N または M>N のときには最小二乗法を用いることで解を決定することができる．一方で M<N の場合は条件が足りないため解を一意に求めることが容易ではない．しかし，M<N の場合でも x の成分がスパース性を持つとき，非ゼロ成分の個数 K に関して M>K であれば解を一つに決めることができる．K 個の本質的な非ゼロ成分を見つけるには y=Ax を満たす x を全候補から順に探索する．

3.1 l_0 ノルム

y=Ax の関係があるとき，少ない情報 y から，多い情報 x を推定する不良設定問題を考える．x がスパースであると仮定できるとき，y=Ax の条件を満たす x の中から非ゼロ要素の数が最小になる x を探す．非ゼロ要素の数は「0 次ノルム ($||x||$)」と定義する．低次元問題では，0 次ノルム最小化でスパースな x を完全に再構成できる．ビッグデータにスパースモデリングを利用する場合，0 次ノルム最小化問題を解くために必要な計算量は N の指数関数で増大するため，ビッグデータに対応するのは容易ではない．計算の時間的問題を解決するために，l_1 ノルムがある．

3.2　l_1 ノルム

l_1 ノルムとは説明変数 x の各要素を足し合わせたものである．式 (2) に l_1 ノルムの数式を示す．

$$||x||_1 = |x_1| + |x_2| + \cdot\ \cdot\ \cdot + |x_N| \tag{2}$$

$||x||_1$ は絶対値の和であるため，各データにゼロ成分が多いとノルムが小さくなる．したがって，$||x||_1$ を小さくすることで，データのゼロ成分を増やすことが可能である．最もスパースな解を求めるために，l_1 ノルム最小化問題を式 (3) に示す．

$$\min ||x||_1 \quad \text{subject to} \quad y = Ax \tag{3}$$

$||x||_1$: x の各成分の絶対値の和
　y: M 次元の観測データ
　A: 観測行列，x: 説明変数

既知である解が存在する場合において，式 (3) の制約条件からスパースな解を求めることで説明変数 x の次元を減らすことができる．

4　スパースモデリングの活用例

4.1　MRI

MRI は狭い空間の中で患者は長時間同じ体勢でいる必要がある．従来の方法では，時間をかけて十分なデータを取得し，鮮明な画像を得る．しかし，検査時間が長く，患者への負担が大きい．そこで，検査時間を 30 % 程度に削減するスパースモデリングという手法が注目されている．

MRI は強い磁場を作ることで，脳内に存在する水素の動きを測定する．測定したデータから，コンピュータ上に脳内の画像を再現する．スパースモデリングはデータの測定範囲を放射上に間引き，背景などの不必要な部分を取り除くことで，時間短縮を可能とする．しかし，従来の方法では検査時間の短縮により，通常の画像作成法では画像が乱れる．乱れた画像から鮮明な画像を再構成するためにスパースモデリングを利用する．スパース性を利用し，圧縮センシングを行うことで画像の本質部分だけを抜粋し，従来の方法で生成した画像と同等の精度を再現することができる．MRI におけるスパースモデリングを用いた画像再構成の流れを Fig.2 に示す．

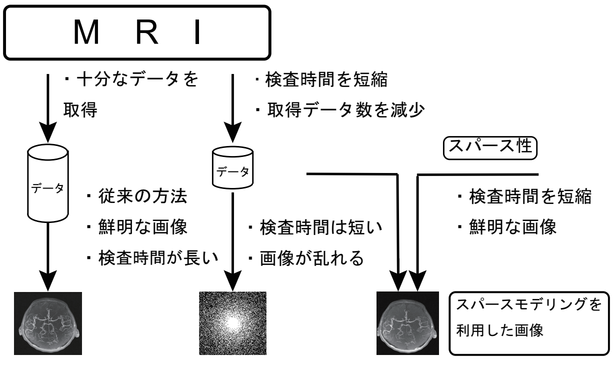

Fig.2　MRI 画像の再構成

4.2　ブラックホール

ブラックホールの存在は理論的に確証されている．しかし，ブラックホールは光を放出しないため，実際に観測することはできず，実質的な形は明確ではない．ブラックホールの周囲にはガスが存在し，ブラックホールに吸収されるときに光を放ち，ブラックシャドウという光の輪を作る．したがって，ブラックシャドウを観測することで，実際のブラックホールの形を推定することができる．

ブラックホールは電波干渉計で観測を行うが，ブラックホールと電波干渉計は離れた場所にあるため，観測が容易ではなく，部分的なデータしか得られない．スパースモデリングは部分的に観測した画像データから本質を抽出し，ブラックホールの形を再構成するために利用する．再構成した画像データは，研究者によって推定されたイメージ図と同等の画像を得る．演繹的に求められた研究者によるイメージ図と，帰納的に求められたスパースモデリングによる画像データが一致することでデータ解析の信憑性を高めている．現在 M-87 というブラックホールを模した画像を用いた実験も行われており，原画像の特徴を捉えることに成功している．

5　今後の展望

現在，人工知能に注目が集まっており，今後スパースモデリングは人工知能分野への応用が考えられる．人工知能は大量のデータから学習するため，データ量が少ないと学習精度は向上しない．人工知能にスパースモデリングを利用することで，少ないデータ量からでも学習することができる．あるいは，大量のデータでも本質を見出すことで計算時間を短縮し，効率よく学習を行うことができると考える．

また，ハカルス社はスパースモデリングを適用した健康促進サービスを扱うアプリケーションのプレリリースを発表した．[3] 健康促進を行う際，学習データとしてユーザからバイタルデータを取得する必要がある．ユーザにとっては，データを入力するという過程が負担となる．しかし，少ない情報から本質部分を抽出するというスパースモデリングの性質を利用することで，ユーザによるバイタルデータの入力数を削減し，ユーザの負担を軽減している．このように，ディープラーニングにスパースモデリングを適用させたアプリケーションが今後増加していくと考えられる．

参考文献

1) 大関真之，少ないデータから知見を見出す「スパースモデリング」，https://japan.zdnet.com/article/35074052/ 参照 Des.1， 2015
2) 富岡亮太，スパース性に基づく機械学習，講談社，pp.21-25，Nov，2015
3) 株式会社ハカルス，https://www.value-press.com/pressrelease/181970，Apr.21，2017