

## 特集・データベース天文学 その(3)

## スローン・デジタル・スカイ・サーベイのデータベース

安田直樹

〈国立天文台 〒181-8588 東京都三鷹市大沢2-21-1〉  
e-mail: naoki.yasuda@nao.ac.jp

スローン・デジタル・スカイ・サーベイ (SDSS) は可視光で全天の約4分の1の領域を観測して宇宙の3次元地図を作成するプロジェクトです。SDSSのデータベースはその大量のデータを効率的に扱うために最新の情報技術を使って構成されています。また、SDSS SkyServer はSDSSのデータを研究者だけでなく科学的教育にも利用できる形で提供しています。

## 1. スローン・デジタル・スカイ・サーベイについて

スローン・デジタル・スカイ・サーベイ (以下SDSS; <http://www.sdss.org>) は米欧の研究機関と日本参加グループ (東京大学, 国立天文台, 名古屋大学, 東北大学, 日本女子大学に所属する研究者) が共同で進めている銀河の広域撮像分光サーベイです。SDSSは全天の約4分の1に相当する領域を5つの可視光のバンドで撮像して, 2億個以上の天体の位置と明るさを測定します。さらに, その2億個以上の天体の中から約100万個の銀河と約10万個のクエーサーについては分光観測を行い, 赤方偏移を求め, 詳細な宇宙の3次元地図を作成するプロジェクトです<sup>1)</sup>。

1998年5月の撮像観測のファーストライト, 1999年5月の分光観測のファーストライト以来, 観測システムの調整を行うと同時に観測を進めてきています。ご存知の方も多いと思いますが, 2001年4月には $z=6.2$ にある最遠のクエーサーを発見するなど大きな成果を挙げています。

これまではパロマーチャートとして馴染み深いシュミット望遠鏡を使った写真乾板による観測が可

視光でのサーベイ観測では最大規模のものでした。SDSSでは口径2.5mの専用望遠鏡を使い, 2048×2048ピクセルのCCDを30個 (5色×6列) 並べたモザイクCCDカメラで撮像観測を行い, 640個の天体のスペクトルを同時に観測できるファイバー分光器で分光観測を行います。また, 天体の位置を正確に決めるために撮像用の30個のCCDとは別に位置標準星を観測するための2048×400ピクセルのCCDが22個配置されています。さらに, 天体の明るさを正確に決めるために2.5m望遠鏡のとなりにある口径50cmの測光望遠鏡が常に標準星を観測して大気吸収の状態を監視しています。このようにクオリティコントロールされた精度の高いデータを大量に生み出す観測はこれまでにはなく天文学的側面, ソフトウェア技術的側面双方でこれまでの常識を変えようとしています。

## 2. SDSSのデータとデータベース

## 2.1. SDSSのデータ

SDSSの1晩の撮像観測では6つのCCDに対応して6つの幅13.5分角の細長い領域 (ストリップ) が5色で観測されます。この領域間の隙間を埋め

るように別の晩に観測場所を少しずらして同じように観測します。これらの2晩の観測データをつなぎ合わせることで幅2.5度角、長さ約120度角の領域のデータ（ストライプ）が完成します（実際にはシーイングの悪いデータは使わないので複数の晩の観測データから構成されることになります）。この2晩分のデータの間には約10%の重なりがあります。全体のデータはこのストライプ45本分からなります。SDSSのカメラのピクセルサイズは1ピクセル当たり

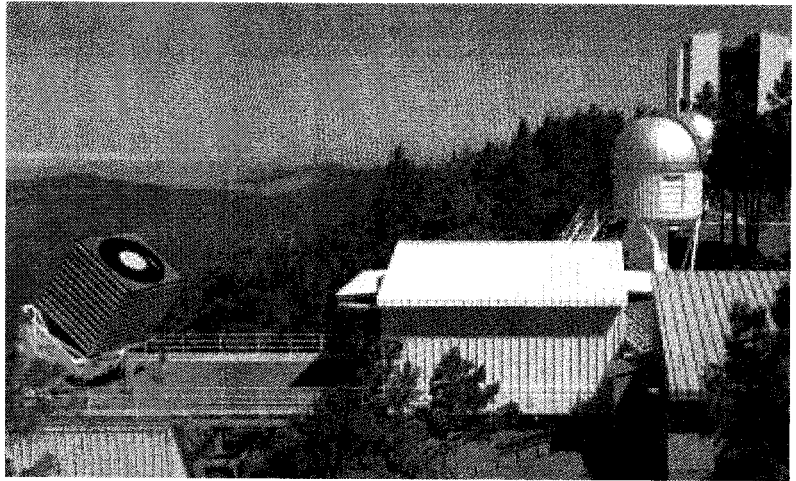


図1 SDSSの2.5 m望遠鏡(左)。右にある丸いドームの中に測光望遠鏡がある。

0.4秒角なので1ストライプのデータ量は1色当たり約50 GB、サーベイ全体で約10 TBになります。今では10 TBはすべてハードディスクにおさめることができる程度の量ですが、プロジェクトの計画当初(1992年頃)はとんでもない量に思えたものでした。

データ解析パイプラインはこの望遠鏡から生み出される5色の撮像データを同時に解析して、そこに写っている天体を検出し、それらを星、銀河、その他、に分類すると同時にそれぞれの天体について約400個のパラメータを測定します。また、位置標準星の観測データを使って天体の正確な位置を求め、測光望遠鏡のデータを使って天体の正確な明るさを求めます。このようにして作成される天体カタログはそれだけで数百GBのデータサイズになります。

## 2.2. データベースデザイン

この1天体当たりの情報が多く天体数も非常に多いカタログを効率的に利用できるようにデータベースのデザインが設計されました。まず、最初に明らかだったのはこのような大量の情報は、これまでのように生のファイルとFortranやC言語で書かれ

たプログラムで扱うには大きすぎるということでした。インデックス付けや並列処理などの技術を利用すべきであるということから商用のデータベースを利用することが決められました。つぎにデータベース構造の設計にあたっては典型的と思われるデータベースへの問い合わせを20個用意してそれらが効率的に実行できるように検討されました。その結果、SDSSでのデータモデルとのマッチングも加味して商用のオブジェクト指向データベースを採用しました。このデータベースは実際に2001年6月に公開されたEarly Data Releaseで実装され公開されています<sup>2)</sup>。しかし、このオブジェクト指向データベースはデータ構造の点ではSDSSのデータベースに非常によくマッチするものでしたが(たとえば任意のデータ型を簡単に作成できる、配列を1つの項目として格納できるなど)、そのパフォーマンスの点では要求を十分に満足するものではありませんでした。そのため、現在当初設計したデータベース構造をリレーショナルデータベースに変換して実装を行っています。2003年1月に公開予定のData Release 1ではこのリレーショナルデータベース上に構築したデータベースが公開予定です。

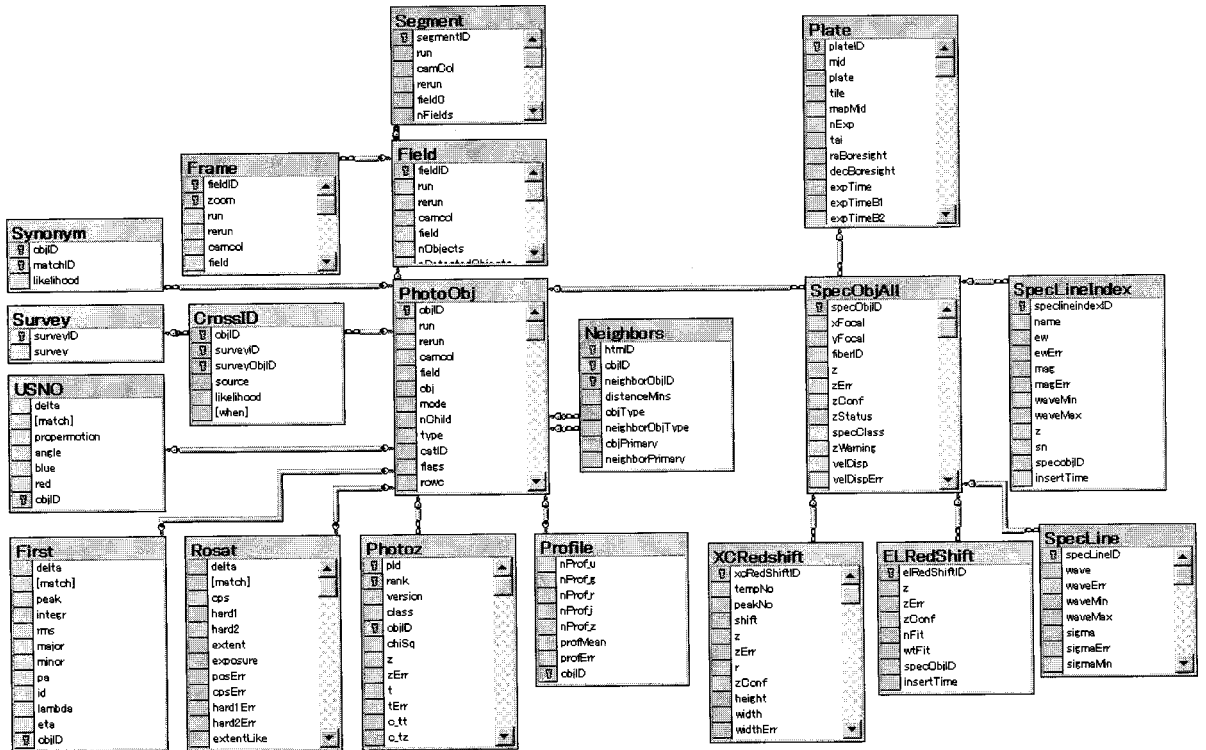


図2 SDSS のデータベースのテーブル構造とその関係を表す図。

### 2.3. 天球座標のインデックス化

上に述べた 20 個の典型的な問い合わせの解析の結果分かったことは、非常に一般的な問い合わせは天球上のある範囲のデータをリストアップすること、色や形にもとづいて一般的な天体から変わった天体を選び出すというような問い合わせでした。

前者の問い合わせに答えるために SDSS のデータベースには天文学でもっとも一般的な座標系である赤経、赤緯の値、また、3次元での方向余弦の値が登録されています。SDSS のデータベースが特徴的なのはこれらに加えてジョンズ・ホプキンス大学で開発された Hierarchical Triangular Mesh (HTM)<sup>3)</sup>と呼ばれる 2次元の天球面を 1次元の数

値でインデックス化する手法が採用されていることです。この HTM を採用することで天球上のある範囲に対する検索、他のカタログとのクロスマッチなどが非常に高速に行われます。

HTM は天球を順に細かく分割していく方法で、球座標における極のような特異な場所を作らずに、すべての領域を等価に扱うことができます。まず、赤道面と赤道面に直交する面さらにそれら 2つの面に直交する面の 3つの面で天球を 8等分します。そして、それぞれの 3角形を各辺を 2等分することで 4つに分割していくことを繰り返します。SDSS では 20 段階までの分割を行い、1つの領域の 1 辺の長さは約 0.3 秒角になっています。

これらの HTM の識別子は 64 ビットの整数としてエンコードされています。重要なのは 2,1,2,2 の三

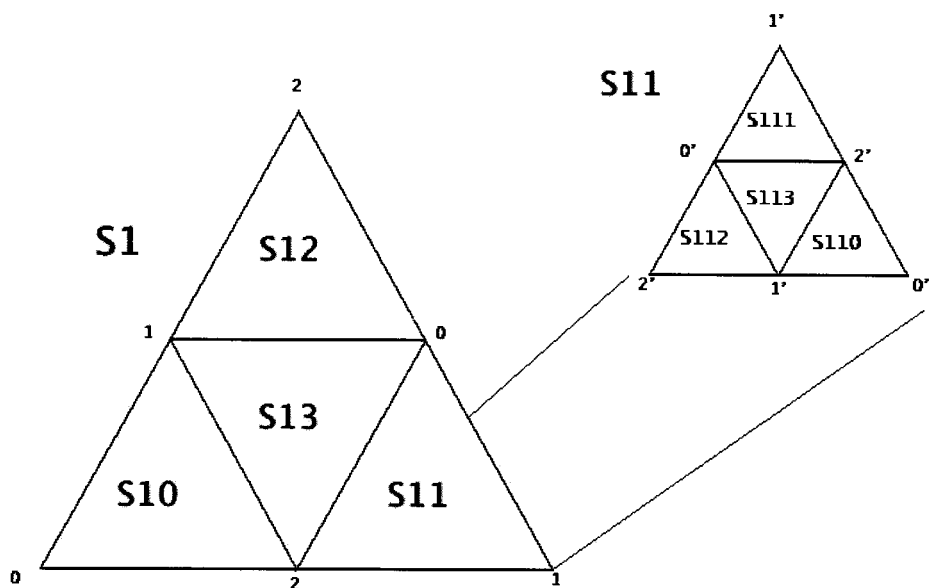


図3 HTMの構築手順を表す図。最初のレベルの球面三角S1を順に分割して細かい三角形に分けていく。頂点の番号は親の三角形の頂点が0になるようにつけ直され、名前は親の三角形の名前に頂点の番号をつけていく。

角形の中にある三角形のHTM識別子はすべて2,1,2,2と2,1,2,3の間にあるという性質を持つということです。そのために、HTM識別子をインデックスとして利用するとある三角形内のすべての天体の検索を高速に行うことができます。任意の形の領域に含まれるHTMの三角形を見つける関数を用意すると、その関数の返り値である三角形のリストと天体のテーブルをHTM識別子で関連付けることでその領域の中に含まれる天体を高速に見つけ出すことができます。

このようにSDSSのデータベースは情報科学を最大限に利用したシステムで運用されるように考えられており、これまでにない大規模なデータを使った統計的な研究を推進するものと期待されています。また、このデータベースをさらに有効に天文学の研究に利用するためにVirtual Observatoryのような仕組みを作ることが計画され、より大規模なサーベイ計画のためのプロトタイプとしての評価がされています。一方、大量のデータを扱うためのアルゴリズムを開発するための材料として情報科学の分野

からも注目されています。

### 3. SDSS SkyServer について

SDSSのデータベースのもうひとつの特徴的な面はそれが専門の研究者と同時に天文学の教育の目的にも利用されることを念頭において作成されていることです。

具体的にはSDSS SkyServerというウェブページ (<http://skyserver.fnal.gov>, 日本のミラーサイトは <http://skyserver.nao.ac.jp>) でデータベースにアクセスするためのさまざまなツールが用意されています。それらはすでに用意されたいくつかの天体を見るものから場所を指定してその場所の画像を見たり、ある範囲の天体をリストアップするものまであります。さらに、SQL (データベースシステムへの問い合わせ言語) を自分で書いて任意の検索を行うこともできます。SkyServerで提供される画像は基本的にはg, r, iバンドの画像から作成したJPEG形式のカラー画像ですが、必要があればFITS形式の画像をバンドごとに取得することも可能になっています。

SkyServerにはこれらのツールを実際を使って天文学の基本的な概念を学ぶための「研究課題」が豊富に用意されています。課題のレベルは中高生から大学学部生くらいが想定されており、内容としては、星、銀河、クエーサーなどSDSSの研究対象となっている天体のこれまでの研究の歴史を実際にSDSSのデータを使って勉強するような形になっています。

具体的には、「ハッブル図」、「星の色」、「星のスペクトル型」、「H-R図」、「銀河」、「スカイサーベイ」、「クエーサー」などの課題があります。たとえば、「ハッブル図」ではSDSSで観測されている銀河の明るさと分光観測で求められている赤方偏移を使ってハッブル図を描いてみて、それがサンプルとして使う銀河の性質によって変わってしまうことを確認し、銀河団の中のもっとも明るい銀河を使うことで改良していくような構成になっています。その過程で銀河の後退速度を求める方法も簡単なアプリケーションを使って学習します。

これまでの天文学のデータベースではすぐに教育的目的に使えるようなものはそれほど多くはありませんでした。それは、教育に使える段階のデータにするためにはいくつかのデータ整約や較正を実施する必要があったからだと思われます。そのためこれまで天文学の教育で使われるデータは人工的に作られたデータまたは何十年も前に取得されたデータなどでした。しかし、SDSSのデータはサーベイ観測ということもあり、パイプラインソフトウェアでほぼ自動的に処理され専門の研究者が使えるレベルのデータが生成されます。それらに対して適当なインターフェイスを用意することで、研究者が使うのとまったく同じデータを教育的にも利用できるようになっています。つまり、SkyServerを使うことでこれまで初めての銀河のデータを使って学習することも可能になります。

## 4. まとめ

SDSSはクオリティコントロールされた均質でありかつ大量のデータを生み出しています。これらのデータを有効に活用するために最新のデータベース技術などを活用してSDSSのデータベースは構成されており、さらに、データマイニングなどの統計的なデータ解析のための方法もVirtual Observatoryの開発の一環として進んでいます。また、このようにデータの配信がネットワークを通じて行われるようになるると最新の観測データが研究者だけでなく、教育、一般啓蒙などにも有効に利用されるようになります。今後は一般の人の方が研究者よりも先に新しい天体を見つけるような時代になるかも知れません。

## 参考文献

- 1) York D. G., et al., 2000, AJ 120, 1579
- 2) Stoughton C., et al., 2002, AJ 123, 485
- 3) Kunszt P. Z., et al., 2002, in ASPConf Ser. 216, Astronomical Data Analysis Software and Systems IX, ed. Manset N., Veillet C., Crabtree D. (San Francisco: ASP) p.141

### Database System of Sloan Digital Sky Survey Naoki YASUDA

*National Astronomical Observatory of Japan, Tokyo*

Abstract: Sloan Digital Sky Survey (SDSS) is the project to make a map of the northern sky in visible wavelengths. The database system of SDSS utilizes the current information technology to handle its massive data volume. SDSS SkyServer provides internet access to the public SDSS data for both astronomers and for science education.