

GIS データ圧縮と LBS におけるそのニーズ



上田直生

有限会社ロケーシング

代表取締役.

<http://www.locaport.com>

nao@locaport.com

想像してみてください。MP3 プレーヤーが 20 曲しか入らないとしたら。あるいは添付ファイルが大きすぎてメールが遅れないとしたら。今日、データ圧縮技術はとても一般的で、広く普及し、縁の下の方持ちとして、情報技術産業のみならず日々の生活を支えています。

1948 年にシャノン博士により、データ圧縮の公式が発表されて以来、さまざまなデータ圧縮理論や実装方法が開発されてきました。それらは今ではソフトウェアの中に組み込まれ気づくこともほとんどありませんが、確実に日々の暮らしを便利にしています。用途によってさまざまな圧縮ツールが用意されています。オーディオ用の MP3、AAC、WMA や、画像向けの JPEG、ファイル圧縮用の ZIP など聞いたことがある方も多いのではないのでしょうか。データ圧縮技術はデジタル機器とそれに関連するサービスの使い勝手や快適さを支えています。

今日、LBS (Location-Based Services : 位置ベースサービス) が市場を拡大しつつあり、他の IT サービスと同様、広く注目されるようになってきました。特に携帯電話サービスの市場では、ほとんどの先進国で携帯電話に GPS (全地球測位システム) デバイスが搭載され、ハードウェアの観点から見ると LBS レディ、つまりいつでも位置情報サービスに対応できる状況となっています。しかしながら、位置情報サービスが、MP3 プレーヤーに音楽をダウンロードするくらい、簡単・快適でないと、多くの消費者はそのサービスを使うことはないでしょう。

今、LBS や GIS 向けに最適化された新しいデータ圧縮技術への需要が高まっています。その理由は次の 3 点にあります。:

1. 既存の圧縮技術は LBS/GIS の要求にマッチしない
2. GIS データ圧縮は、位置情報サービスを身近にします
3. GIS データ圧縮は LBS の進化を加速します

既存の圧縮技術は LBS/GIS の要求にマッチしない

現在、非常に多くのデータ圧縮アルゴリズムとツールがあります。それぞれは特別な目的にそって設計され、最適化されています。ほとんどの圧縮ツールは複数の圧縮アルゴリズムを組み合わせ、対象となるデータの圧縮効率を高めるように工夫されています。データ圧縮理論は、常に対象とするデータの性質を利用し、冗長性をカットすることで圧縮を実現しています。

データのタイプと圧縮方法の例

オーディオデータの圧縮は、人間の耳の特性を利用し、聞き取りにくい音の成分をカットすることでデータサイズを小さくしています。この場合、データは 100%元には戻らないのですが、音楽を楽しむには十分な音質をキープしています。これは「非可逆」圧縮といいます。MP3 はこの原理と、その他さまざまな圧縮理論を組み合わせ、圧縮効率と音質のバランスを取っています。

一方、ファイルの圧縮では解凍後、100%元の状態に戻る必要があります。このように100%元に戻る圧縮を「可逆」圧縮といいます。

圧縮対象がテキストファイルで英語の文章であれば、同じ単語やパターンが繰り返し現れます。このように繰り返すパターンが存在するときはユニバーサル・コーディングと呼ばれる方法が適しています。GIF 画像フォーマットに使われている LZW コーディングはユニバーサル・コーディングのひとつです。

画像データでは、ひとつの画素とその隣の画素は良く似た色をしていることが多いので、この性質を利用した圧縮が行われます。データ間に関連性があれば、データ関連性圧縮が適しています。DPCM と呼ばれる CD などに用いられるデータ圧縮はその一例です。データとその隣のデータの差はおおむね小さい数字になりますので、全体として少ない情報で表せることになります。

データの各値の出現頻度に偏りがあれば、エントロピー・コーディングと呼ばれる方法が適しています。これはよく現れる値にできるだけ短いコードを割り当てるもので、有名なハフマン符号はこの一種です。JPEG 画像フォーマットはこの方法を使っています。

GIS データはオーディオ、画像、テキスト／バイナリファイルとは異なります

はじめに、GIS データは階層構造になっています。まず、データが線や領域や、あるいはその両方を現します。たとえばナビゲーションのルート（線）や公園の境界線（領域）などです。これらの GIS データは多数の地点情報から構成されています。各地点のデータは、緯度、経度、場合によっては高度やその他のパラメータによって定義されます。このようにデータの構造が階層構造になっているため、単純なデータを想定した圧縮では効率よく圧縮できない場合があります。

次に、GIS データが線や領域を表す場合は、それを構成する各地点は、隣のデータの地点とおそらく近くの地点となり、比較的ローカルな情報となります。しかし、地点をあらゆるスケール、つまり緯度経度は全世界で使えるものなので、データセット全体としては、冗長性が非常に大きいこととなります。これは非常に圧縮できる余地があり、差

分をベースとする単純な圧縮のみを利用しても、非常に大きな圧縮が可能な場合があります。表 1 に差分ベースの圧縮の例を示します。

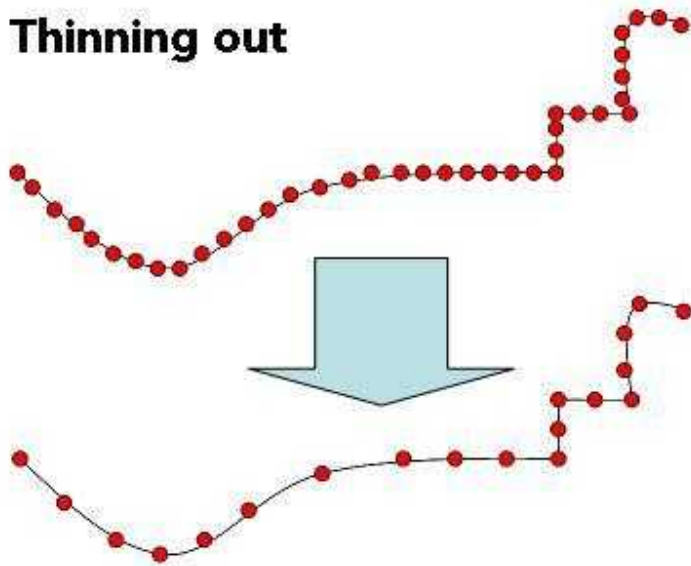
例	緯度	差分 (0.00001 度単位)
	+35.00001	+3500001(圧縮不可)
	+35.00002	+1
	+35.00003	+1
	+35.00002	-1
必要な文字数	36	14

(表 1)

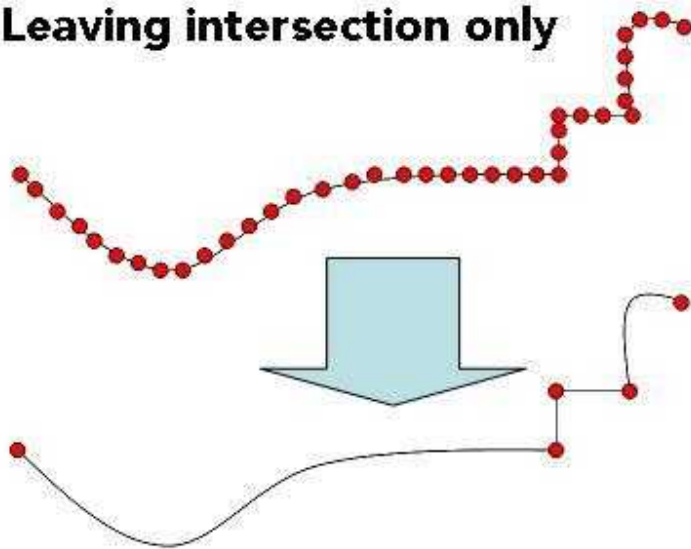
3 番目に、GIS データはその利用目的と利用状況により、可逆圧縮か非可逆圧縮のどちらか必要かが変わります。もし必要なのが正確なピンポイントの位置の場合、可逆圧縮であることが必要ですが、個々の地点の正確さよりも全体のルートがより重要な場合は非可逆圧縮が適する場合があります。この場合、下記の選択肢が考えられます。

- データの間引き (Thinning out)
- 交差点情報のみ残す
- 構成する点の再現精度を下げる

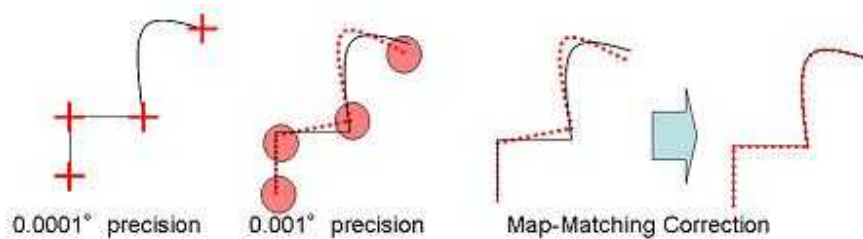
1. Thinning out



2. Leaving intersection only



3. Decreasing the precision and Map-Matching correction



さらに、他のデータと異なり、ルート情報は「マップ・マッチング」と呼ばれる技術により、再現した地点情報が多少ずれていても地図上の道路の形に復元することが可能であり、これを前提に精度を下げて圧縮率を高めることも可能です（図3）

最後に、言及しておきたいのは、位置情報サービスはモバイルコンピュータや携帯電話で使われることが多いであろうということです。一般的に、データ圧縮技術は「重い」処理であり、パソコンでは快適に動く圧縮処理が、携帯端末では非常に動作が遅くなるなどの影響が考えられます。また、「重い」処理はバッテリーの消費を早めてしまいます。したがって、GIS データ圧縮にはこれらの端末で処理できる「軽い」処理であることが求められます。

さらに、GIS データ圧縮には「データの事後追加」が可能であることが望まれます。リアルタイムのGPS データ取得の場合、データは一定時間間隔で入力されます。このとき、圧縮したデータをその都度解凍し、データを追加し、また圧縮する、ということが必要であれば、ハードウェアのリソースを多く消費することになります。元にある圧縮済みデータはそのままに、後から入力されたデータは、最小限の処理で圧縮データに追加できる必要があります。

以上で説明したように、GIS のデータは独自の性格をもち、オーディオや画像など、他のデータタイプ用に最適化されている既存のデータ圧縮方法では、GIS データが必要とする機能をカバーできないことが分かります。

GIS データ圧縮は、位置情報サービスを身近にします

消費者を対象とする製品やサービスでは、ユーザビリティ、またはヒューマン・インターフェースが重要なキーとなります。つまり、人々は使いやすいものしか使わないということです。あなたの会社の位置情報サービスはユーザーにとってすぐに使えるものである必要があります。MP3 技術によって音楽サービスがアクセスしやすくなったように、GIS データ圧縮技術は位置情報サービスの顧客体験を大幅に向上させるでしょう。

まず、GIS データ圧縮はデータ容量を節約します。ユーザーは記憶容量に制限のある携帯電話などの携帯端末で位置情報サービスを使うことが増えると思われませんが、データサイズを節約することは即、ユーザーのメモリの空きを増やすことになります。図 1 は GIS データ圧縮の例です。

Original Data

Longitude, Latitude

136.97288, 35.15956	136.97017, 35.15472
136.97302, 35.15936	136.97000, 35.15444
136.97308, 35.15924	136.96987, 35.15398
136.97313, 35.15907	136.96949, 35.15376
136.97314, 35.15893	136.96928, 35.15353
136.97314, 35.15892	136.96895, 35.15321
136.97311, 35.15883	136.96888, 35.15335
136.97308, 35.15879	136.96853, 35.15373
136.97308, 35.15883	136.96835, 35.15385
136.97262, 35.15882	136.96757, 35.15392
136.97198, 35.15881	136.96729, 35.15402
136.97177, 35.15878	136.96676, 35.15423
136.97163, 35.15877	136.96661, 35.15429
136.97156, 35.15876	136.96661, 35.15430
136.97150, 35.15873	136.96646, 35.15436
136.97144, 35.15866	136.96652, 35.15447
136.97136, 35.15857	136.96655, 35.15457
136.97133, 35.15850	136.96658, 35.15470
136.97132, 35.15838	136.96661, 35.15486
136.97133, 35.15798	136.96661, 35.15504
136.97131, 35.15763	136.96661, 35.15513
136.97128, 35.15747	136.96661, 35.15518
136.97127, 35.15734	136.96660, 35.15524
136.97124, 35.15723	136.96597, 35.15517
136.97115, 35.15689	
136.97108, 35.15668	
136.97105, 35.15661	
136.97102, 35.15657	
136.97099, 35.15653	
136.97088, 35.15645	
136.97066, 35.15634	
136.97046, 35.15628	
136.97041, 35.15611	
136.97041, 35.15611	
136.97033, 35.15547	
136.97024, 35.15487	

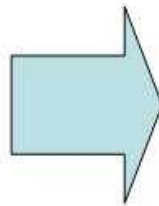
continues

Compressed Data

Google Maps™ API Encoded Polyline

<http://code.google.com/apis/maps/documentation/polylinealgorithm.html>

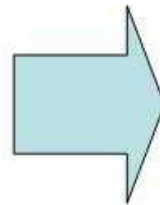
gbruEoo_cYf@[XK^I*#APDFDG?BzEHdA
?LDJ^@ZLDT@nAAfADv@DzAVt@RPJLTTJ
@Jf@^@JvF^@^Jt@^@|A^AvAIA^@^A|Lk
AfAWb@MxCj@^DWz@WKSEy@KkA@L|B



LocaPort™ GIS Data Compression

<http://www.locaport.com>

1BgSCALJwxcfbEhBhKWitISiQipQhQb
QatPrPpOoOnMmKIIIFIJVIMIIEkCkIT
jOiMiLhKhIfFdDaHZbzzZzJyGUxQwJuf
XqSoMIDhHgrCJyzWpYmGDfFdfdBHjCmc
PdTydHAdCdDdBxvSCHAHBwxbxv



次に、GIS データ圧縮は通信速度 — サービスの快適性を決定する一大要因 — を向上させます。特にネットワークベースの位置情報サービスでは、その差がよく現れます。位置情報かどうかにかかわらず、多くの IT サービスではパケットサイズを節約し

てパフォーマンスを向上させることを考え始めています。ネットワークベースのサービスでは、もっとも遅い部分のパフォーマンスが全体のパフォーマンスを決定するからです。

実際に、多くの Web サービスは XML でのサービスと平行して JSON と呼ばれるデータフォーマットでのサービスを開始しています。JSON は XML に比べて処理が「軽く」、データサイズも小さいのがその理由です。

最後に GIS データ圧縮技術は、通信料によって料金が決定する場合は、ユーザーのお金をも節約することになります。これも顧客体験の向上といえるでしょう。

GIS データ圧縮は LBS の進化を加速します

位置情報サービスは M-LBS と C-LBS の二つの方向に進化していくと考えています。

M-LBS : マッシュアップ・位置情報サービス

ここで「M-LBS」と呼ぶものは「マッシュアップ・位置ベースサービス」です。一般的に、IT サービスは独立したサービスから、「マッシュアップ」したサービスへと移行しつつあり、マッシュアップ・サービスの時代がやってこようとしています。企業はコンテンツ・プロバイダーからプラットフォーム・プロバイダーへとシフトしています。LBS におけるマッシュアップの最適な例はグーグル・マップ用のガジェット機能「マップレット」でしょう。グーグルマップと他の情報プロバイダを連結し、グーグルマップ上にさまざまな位置ベースの情報を表示することができます。

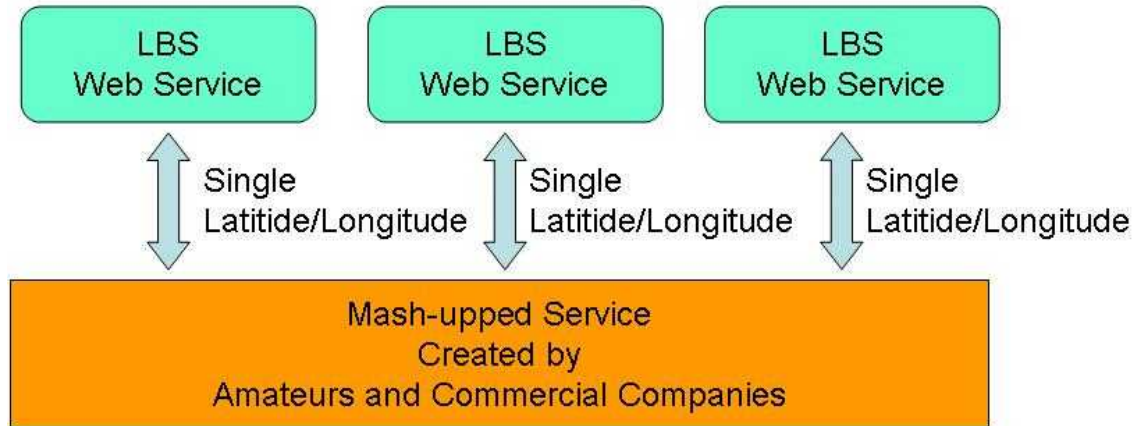
コンテンツではなくプラットフォーム

グーグル社は決してコンテンツを用意しません。しかし、プラットフォーム — サービスを創り上げる材料 — は提供します。「マッシュアップパー」と呼ばれる何千人ものアマチュア技術者や営利企業が API と呼ばれるパーツを組み合わせて新しいサービスを創造しています。これらの部品はグーグル社から提供されるもの、他の Web サービス業者のもの、あるいは技術者自ら作成したものもあります。その結果、グーグルマップは驚くほど短期間に何千もの LBS コンテンツを得ました。自社で LBS コンテンツを製作している企業では、この期間と数を真似することはほとんど不可能といえるでしょう。さらに、非常に多くの「マップレット」が多くのユーザーに利用され、もっとも良く利用されるものが人気リストの上位に上がってきます。つまり、グーグルはマッシュアップパーに魅力的なプラットフォームを提供しただけで、多くの優れた LBS コンテンツが自動的にグーグルマップへと集まってきたわけです。このグーグルマップ API があることで、初期の「マッシュアップ・サービス」の大部分はグーグルマップをベースとした LBS でした。驚くべきことに、「Web2.0」と呼ばれる最先端の分野では、LBS が主要な役割を演じています。

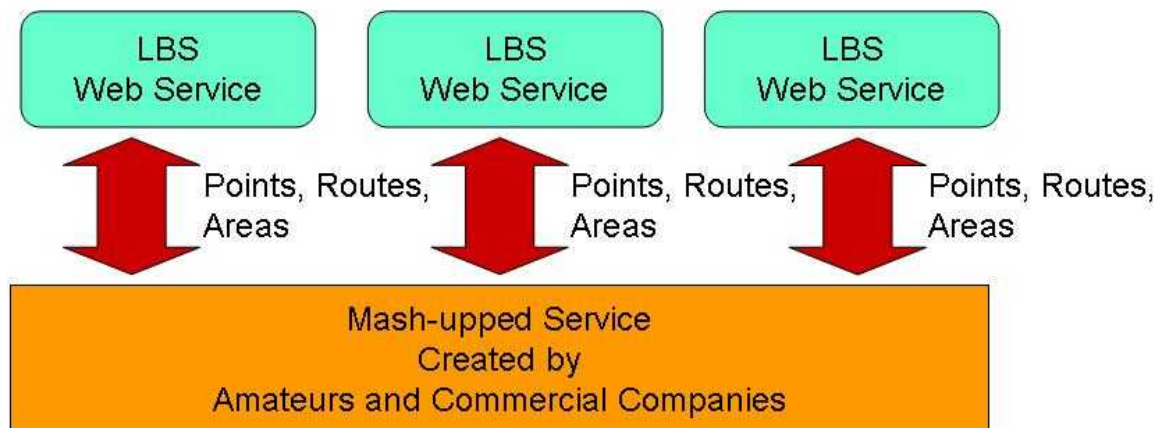
C-LBS : 複合 GIS データ LBS

次に、「C-LBS」というのは「複合 GIS データ (Complexed-data) LBS」を意味します。マッシュアップ・サービスの時代、マッシュアップする複数のサービスが LBS であれば、GIS データが交換され、共有されることが予想されます。今日、ほんの数社が GIS データのインポート、エクスポートの機能をもし、データの共有ということを行っています。これらの場合でさえ、共有される GIS データというのは単一の地点 — おそらく目的地 — か、多くても二点 — おそらく現在地と目的地 — にすぎません。LBS のプラットフォームとなるサービスが、GIS の複合データ、たとえばルートや領域などを交換・共有するようになれば、単一地点の情報を共有するのと比較して、サービス間の連携はより、豊かに、より高度になります。図 3 は M-LBS と C-LBS の概念図です。

Mash-upped LBS



Mash-upped LBS with Complex GIS data interaction



いくつかのより豊かなサービス間連携の例

- ナビゲーションシステムが「ルート」をエクスポートし、グルメ情報サービスが「そのルートに沿ったお店のみ」を案内する
- 不動産検索サービスが「選択した物件」のリストをエクスポートし、検索エンジンは「それらの物件に近いスポーツジムのみ」を表示する
- 旅行計画サービスが「旅行スケジュール」をエクスポートし、広告サイトが「あなたが訪れる可能性のあるお店」のクーポンを配信する

位置ベースサービスの組み合わせには、もっとクリエイティブな組み合わせがあるはずです。このキーは GIS データの「可搬性」といえます。独立型の GIS や LBS では、GIS データはシステム内のデータベースに蓄積されていますが、LBS 間のマッシュアップでは GIS データはシステム外でも利用できる、つまり共有したり交換したりする必要があります。言い換えると、GIS データには「ポータビリティ」が必要であるといえます。

LBS 間の情報交換を実現する方法として、いくつかのファイルフォーマットがあります。

KML、KMZ、GML、GeoJSON などがよく利用されています。KMZ は KML を ZIP 圧縮したものです。弊社の実験では、「KML を SZIP 圧縮したもの=KMZ」よりも、「KML 内の GIS データを GIS 圧縮文字列に変えたもの」の方が、より効率が高いことを示しています。この理由は、先にも述べたように、ZIP という圧縮アルゴリズムが GIS データに特化したものではないことが挙げられます。

GIS 圧縮文字列にはグーグル社・グーグルマップス API の「エンコーデッド・ポリライン・アルゴリズム」(<http://code.google.com/apis/maps/documentation/polylinealgorithm.html>) やロケーティング社の「LocaPort GIS データ圧縮アルゴリズム (<http://www.locaport.com>、日本語名称は「ロカポーター <http://www.locaporter.com>」)

結論

ひとたび実用的な GIS 向けデータ圧縮が普及すれば、LBS はマッシュアップをベースとする新たなステージへと移行するでしょう。GIS システム-LBS の土台-はその上で稼動する LBS 用に GIS データ圧縮をサポートする必要があるでしょう。

何千ものマッシュアップ・LBS コンテンツの中には、今日誰も考えもしなかったようなクリエイティブなサービスが出現してくることは疑う余地はないでしょう。LBS とそのマーケットは加速度的に大きくなっていくでしょう。近い将来、GIS データ圧縮は縁の下の力持ちとなって、重要な役割を演じることになるでしょう。