

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 25 日現在

機関番号：12703

研究種目：挑戦的萌芽研究

研究期間：2010～2013

課題番号：22650029

研究課題名(和文) 離散的最適化と時系列解析による人工衛星データの復元と知識発見

研究課題名(英文) Recovery of satellite data and knowledge discovery by discrete optimization and time series analysis

研究代表者

土谷 隆 (Tsuchiya, Takashi)

政策研究大学院大学・政策研究科・教授

研究者番号：00188575

交付決定額(研究期間全体)：(直接経費) 2,300,000円、(間接経費) 360,000円

研究成果の概要(和文)：Geotailは地球磁気圏のプラズマを観測する人工衛星であり、観測した低エネルギープラズマ粒子の分布は3次元ヒストグラムに集約されて地上に送信され、さらにETスペクトログラムという図にまとめられて検討に供される。24時間の内16時間については、ある方向について和をとって得られる2次元ヒストグラムと10個の統計量のみを受信する。本研究では、事前情報を取り込んだ最適化モデルによって、圧縮データから3次元ヒストグラムとETスペクトログラムを復元することに取り組んだ。適切な事前情報を取り込んだ線形計画モデルにより、3次元ヒストグラムおよびETスペクトログラムがかなり良く再現できることが示された。

研究成果の概要(英文)：Geotail is a satellite to observe plasma of the magnetosphere of the earth. The distribution of observed low energy plasma particles are summarized as a three dimensional histogram before sent to the ground. Researchers conduct analysis based on a color chart called ET-spectrogram drawn from the three-dimensional histogram. However, a complete three dimensional histogram data are sent only eight hours per day, and for remaining period a marginalized two-dimensional histogram and ten statistics of the original data are sent to the ground. In this research, we made an attempt to recover the original three dimensional histogram from the compressed two dimensional histogram and 10 statistics to draw an ET-spectrogram. We developed a LP model which incorporates with appropriate prior information to recover the original three dimensional histogram and ET-spectrogram successfully.

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：人工衛星 データ解析 最適化 統計 データ復元 事前情報

1. 研究開始当初の背景

本研究は、人工衛星から通信容量の関係で圧縮されて送信されてくるデータを、事前情報に基づいて復元することを目的とする。対象となるのは人工衛星 Geotail で観測される低エネルギープラズマ粒子に関するデータである。Geotail は地球磁気圏のプラズマを観測する人工衛星であり、観測した低エネルギープラズマ粒子の分布を速度(32 区分)・入射方向(16 方向)・入射角(7 方向)の 3 次元ヒストグラムに集計して、 $32 \times 16 \times 7 = 3584$ 個の整数データとして 1 2 秒ごとに地球に送信する。

3 次元ヒストグラムの情報は特徴量の時系列としてまとめられ、ET スペクトログラムという 2 次元のカラー図面に集約され、地球惑星科学分野での検討に供される。

地上基地局の配置と運用の関係で、24 時間の内 8 時間は 3 次元ヒストグラムをそのまま送受信できるが、16 時間については、入射角方向について和をとって得られる 2 次元ヒストグラムの $32 \times 16 = 512$ 個のデータと、3 次元ヒストグラムの平均と分散共分散 10 個の数値、合せて 522 個のデータのみを受信する。

16 時間分の 2 次元ヒストグラムデータから 3 次元ヒストグラムデータを復元することができれば、地球惑星科学における新しい知見が得られる可能性が大いにある。

また、方法論の観点からは、このような手法は、遠隔観測を行っているさまざまな分野に適用できる可能性がある。

2. 研究の目的

本研究は、事前情報を組み込んだ最適化モデルを用いることによって、2 次元ヒストグラムと 3 次元ヒストグラムの平均・分散共分散行列から、元の 3 次元ヒストグラム、そして ET スペクトログラムを復元することを目的とするものである。

3. 研究の方法

3 次元ヒストグラムが 2 次元ヒストグラムに圧縮されていても「各ヒストグラムのセルの粒子数が整数である」ことを用いると、以下の考え方によって、混合整数計画問題に定式化することにより、3 次元ヒストグラムを復元することができる可能性がある。

[接近法 MIP]

入射角方向に和を取り圧縮された 2 次元ヒストグラムと元の 3 次元ヒストグラムの平均 μ 、分散共分散行列 D がデータとして与えられているとする。この時、次の条件を満たす 3 次元ヒストグラムを求める：

- (1) 入射角方向に和をとると圧縮された 2 次元ヒストグラムに一致する。
- (2) セルに入っている粒子の平均と分散共分散行列は D に一致する。
- (3) 各セルに入っている粒子数は整数である。

これが研究の出発点となる考え方である。しかしながら、粒子数が大きい時にこの混合整数計画問題を解くことは困難である。よって、より統計的な接近法として、以下のようなものを考える：

[接近法 LP]

入射角方向に和を取り圧縮された 2 次元ヒストグラムと元の 3 次元ヒストグラムの平均 μ 、分散共分散行列 D がデータとして与えられているとする。この時、線形計画法を用いて、次の条件を満たす 3 次元ヒストグラムを求める：

- (1) 入射角方向に和をとると圧縮された 2 次元ヒストグラムに一致する。
- (2) セルに入っている粒子の平均と分散共分散行列は D に一致する。
- (3) セルの粒子数は整数でなくても良い。
- (4) 「セル内の粒子数についての事前情報」を、線形不等式制約式制約あるいは線形計画問題として表現できるような目的関数として表現する。

ここで、(4)の事前情報として何を用いるかを検討する必要がある。プラズマの状態としては、ローブ・プラズマ境界層・プラズマシートの 3 つがあり、解析している点の状態が何であるかについては、別の物理量から知ることができる。

本研究では、観測しているプラズマが 3 つのどの状態であるか、に応じて、

- (a) 7 つの入射角についての粒子の周辺分布；
- (b) 時間方向での 3 次元ヒストグラムの滑らかさ；
- (c) 空間方向での 3 次元ヒストグラムの滑らかさ；

についての事前情報を、実際のデータを十分に検討した上で、線形計画問題に組み込むことを試みた。さらに、各粒子の持つエネルギーを、低・中・高に分けて、粒子のエネルギー帯ごとに、上記の事前情報を組み込むことも適宜行った。

もちろん、細分化を進めすぎると復元のための線形計画問題が徒に複雑になりすぎるので、試行錯誤でモデルの性能を比較検討していき、これらの事前情報の中で取り入れることが推定に有効なものを明らかにしていく、ということに留意して本研究を進めた。

[解析に用いたデータ]

1994 年 1 月 14 日の 15:30 から 40 分間、12 秒ごとに、201 個の 3 次元ヒストグラムデータをベンチマークデータとして、それを 2 次元に圧縮して、圧縮されたデータから、3 次元ヒストグラムおよび ET ヒストグラムを再現できるか否かについて、検討を行った。なお、この区間にはローブ・プラズマ境界層・プラズマシートが現れる。また、粒子の個数も十数個から 6000 個以上と大きく変動する、比較的プラズマの状態変動の大きい期間であ

るので、ベンチマークとして用いるのに適切であると考えられる。

[復元性能の指標]

3次元ヒストグラムがうまく復元できているかどうかの指標については、本研究では、「正答率」という量を定義して用いた。正答率は、

「各セルにつき、正解の粒子数と推定解の粒子数の内小さいものを求めてその総和を計算し、総粒子数で割ったもの」である。粒子の分布を多項分布で表現すると考えると、正答率と(真の分布と推定分布の差)の間には、

$$(1 - \text{正答率}) \times 2$$

$$= \frac{|\text{正解ヒストグラムの多項分布の密度関数} - \text{推定ヒストグラムの多項分布の密度関数}|}{\text{真の分布の総粒子数}}$$

$$\text{正答率} = \frac{1}{2} \left(1 - \frac{|\text{正解ヒストグラムの多項分布の密度関数} - \text{推定ヒストグラムの多項分布の密度関数}|}{\text{真の分布の総粒子数}} \right)$$

という関係がある。

[ET スペクトラムの検討]

ET スペクトラムは、地球惑星科学の分野で検討に用いるチャートである。1つの時点における3次元ヒストグラムの情報を、エネルギー32区分、入射南北方向3区分、入射水平方向4区分、計 $32 \times (4 \times 3) = 224$ 個の色つきセルとして表現する(色の違いにより粒子の個数の違いを表現する)。(上記で方向の区分は、元の3次元ヒストグラムの入射南北方向7区分、入射水平方向16区分をブロックとしてまとめたものである)。ET スペクトログラムでは、1時点でのプラズマの分布状態が224ピクセルからなる1本の長い帯で表現され、それを時系列順に並べることにより、プラズマ粒子の挙動をダイナミックに捉えることができる。

復元された3次元ヒストグラムが地球惑星科学者の解析に役立つためには、ET スペクトログラムをよく再現することが死活的に重要である。

4. 研究成果

3. で述べた種々のモデルを実際に上記のベンチマーク問題に適用した結果判ったことを以下に述べる。

[接近法 MIP について]

当初の狙いであった、混合整数計画を用いた3次元ヒストグラムの復元については、ロープ状態にある総プラズマ数が66までの事例について、3次元ヒストグラムを復元することができたものの、実際に粒子数が数千にも達する状況では現在の混合整数計画アルゴリズムの性能の限界から、復元は困難であることが判った。

[LP モデルについて]

そこで、事前情報を取り入れ、整数性を緩和したLPモデルを解くことにした。今後繋がる有望な結果が得られたので、以下説明することにする。まず、事前情報の取り込み方として、粒子の入射角方向の周辺分布を事

前情報として組み込むこと、粒子分布の滑らかな時系列変化を事前分布として取り込むことの2つを検討し、実装した。

その結果、性能がより優れていたのは入射角方向の周辺分布を事前情報として取り込むものであった。時系列方向の分布の連続性を仮定したモデルでは、ET スペクトログラムを描いたときに、不自然な規則的模様が出るのが観察され、現状では復元モデルとしては不適切であると考えられる。

さらにロープ、プラズマ境界層、プラズマシートの3状態を分け、さらに、それを低エネルギー粒子と高エネルギー粒子で分けて入射角方向の周辺分布を用いたLPモデルを適用すると、概ねET スペクトログラムの特徴を復元できるようになった。この時の3次元ヒストグラムの正答率は73%であった。

しかしながら、地球惑星科学の専門家に結果を検討してもらったところ、(a)プラズマ境界層における低速粒子の入射角の分布が物理的にありえない挙動を示していることや、(b)全体として太陽の反対側に赤道方向から入射する数少ない低速粒子の分布の様子が捉えられていない(c)北側から入射する高速粒子数が少なすぎるなどの問題が指摘された。

さらにモデルを改善するべく、空間方向の粒子の分布の滑らかさを取り入れ、パラメータを調整することにより、正答率が76%まで上昇した。しかしながら、上記(a)、(b)、(c)の問題は解決しなかった。

(a)、(b)ともに「粒子数が少ないがプラズマ粒子の挙動を捉える上で重要な部分」であり、粒子数が少ないゆえに、ヒストグラムを推定する段階での事前分布による対応がかなり困難であると考えられる。仔細に実際のデータと推定データを比較した結果、要因の一つは、粒子がバースト的にいくつかまとまって1つのセルに存在している傾向があるにもかかわらず、それが最適化モデルによるとそれがうまく捉えられていないことである、ということが判明した。

そこで、ET スペクトログラムの特性まで考慮して、一度線形計画問題で推定した結果をさらにバーストを考慮して改訂する手続きを考案した。(b)についてはかなり状況が改善された。この手法は便宜的であるように思えるが、2段階の推定法と考えて、きちんと数理モデルとして表現することができる。

(a)、(c)については、まだ完全に推定することが難しく、モデルの改善については今後の課題である。

[今後の課題]

本研究により、事前情報を適切に組み込んだLPモデルにより3次元ヒストグラムを圧縮(周辺化)した2次元ヒストグラムと3次元ヒストグラムに関する統計量から3次元ヒストグラムを復元し、ET スペクトログラムを大まかには復元できることが示された。本研究の困難な点は、粒子数が少ない部分が重要

な意味を持ちうる，ということである．その意味では典型的な悪条件問題である．

本研究は地球惑星科学，最適化，統計科学の3分野の境界分野での研究である．かなり特徴を捉えた元データの復元には成功したものの，地球惑星科学分野で利用できるレベルにはまだ達しておらず，最適化に用いているアルゴリズムは必ずしも新しいものではないため，現段階で学术论文にするのは時期尚早であると判断しているが，ゴールは十分に狙える射程距離に入ってきていると評価している．このような海のものとも山のものともつかぬ研究課題を取り上げ支援し貴重な研究機会を与えてくれたことに感謝し，さらに実用化に向けて息長く研究を継続していきたい．

5．主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表](計2件)

池上敦子, 村田裕介, 土谷隆, 上野玄太:最適化による人工衛星データの情報復元と知識発見. 統計数理研究所共同研究集会“最適化：モデリングとアルゴリズム”，2012年3月15日，於 政策研究大学院大学.

馴松晃司, 土谷隆, 上野玄太, 池上敦子:事前知識と線形計画法を用いた人工衛星観測データの復元. 統計数理研究所共同研究集会“最適化：モデリングとアルゴリズム”，2014年3月25日，政策研究大学院大学.

6．研究組織

(1)研究代表者

土谷 隆 (政策研究大学院大学・教授)
研究者番号：00188575

(2)研究分担者

池上敦子 (成蹊大学理工学部・教授)
研究者番号：90146936
上野玄太 (統計数理研究所・准教授)
研究者番号：40370093