

あけぼの衛星で得られた VLF/WBA 波動データへのクラスタリングの適用

嶋 啓佑 [1]; 神林 卓也 [1]; 後藤 由貴 [1]; 笠原 禎也 [1]
[1] 金沢大

Automatic classification of plasma waves observed by the Akebono VLF/WBA

Keisuke Shima[1]; Takuya Kanbayashi[1]; Yoshitaka Goto[1]; Yoshiya Kasahara[1]
[1] Kanazawa Univ.

In this study, we developed an automatic classification method of wave data obtained by the VLF/WBA instrument on Akebono satellite. Distinctive fine structures of various wave phenomena can be seen in spectrograms from the WBA data. We classified each wave according to such fine structures. The classification is based on clustering methods that do not need prior information about each wave. We compared results by two clustering methods; k-means and mixture Gaussian model and found that the Gaussian model classified each wave better.

1989年に打ち上げられたあけぼの衛星は、25年目を迎えた現在でも運用を継続しており、観測データは大規模なデータセットとして蓄積されている。このあけぼの衛星の特徴的な観測として、14 kHz以下の低周波のプラズマ波動の「波形観測」があげられる。近年の科学衛星による波形観測が、波形をオンボードでサンプリングして圧縮/選別処理等を実施した後、地上に伝送するのが一般的となっているのに対して、あけぼの衛星では波形をそのままアナログテレメトリに載せて地上に伝送するという観測手法がとられている。このため、取得したデータはSNが悪いという欠点はあるものの、優れた時間連続性や周波数分解能をもち、他の衛星にはない貴重なデータを提供している。現在では、あけぼの衛星から得られた波形データ(WBAデータ)の有効活用が、衛星の継続運用の目的のひとつとなっており、今後の幅広いユーザーの利用のために、一次処理、二次処理を速やかに実現する支援ツールの整備が必要不可欠である。

近年の計算機の進化によって巨大データ(ビッグデータ)から知識発見を行ういわゆるデータマイニング法の研究が飛躍的に進んだことにより、先述のあけぼの衛星の波形データセットに対しても、これまでの計算機資源では難しかった大規模な波動の分類(クラスタリング)を適用する環境が整った。プラズマ波動は、発生メカニズムや伝搬過程に応じてホイストラ、コーラス、ヒス、EMICなど多種に分類されるが、実際には、観測波動の周波数スペクトルの形状や定性的な変化に基づいて分けられている。これは発生・伝搬過程と周波数スペクトルの間に強い相関関係があるためである。ただし、こうした分類は明確な定義に基づいて行われている訳ではない点に注意が必要である。これに対して、本研究では、波動の微細なスペクトル構造を多次元の特徴量ベクトルとして定量化して、それらの類似度に基づいて分類することにより、各種波動を帰納的に分類する手法を確立することを目指している。クラスタリングを用いることで、教師あり学習やパターンマッチング等と比較して、先見的な知識なしに波動の特徴を定義できる。

クラスタリングの具体的な実現手法として、確定的な分類を行うk-means法と確率的な分類を行う混合ガウスモデルについて、その分類結果の違いについて検討した。k-means法はデータ数の1次のオーダーで計算が完了する高速な分類法であり、大規模なデータに対する処理も容易である。一方で、初期クラス中心の位置によって得られる結果が変化することや、クラスが多次元特徴量空間上で超球状の形状になるために、クラスに属するサンプル数がどれもほぼ等しくなるという制約がある。今回、対象としているデータの波動分類においては、ノイズ成分が非常に多く、分類すべき波動の割合がノイズに比べて少ないため、k-means法では、ノイズ成分の一部が波動と混在してしまうという結果になった。また、データには複数の波動が混在した箇所があるが、k-means法は確定的な分類法のため、そのような箇所は特定の波動として分類されてしまった。

混合ガウスモデルでは、多次元空間上の分布を多変量正規分布の和で表すことにより、より柔軟なクラスが形成される。また、分布の重なりでデータを表現するため、複数の波動を表現できる利点がある。計算時間についてはk-means法より劣るものの、パラメータ推定のための近似アルゴリズムが複数提案されており、今回は変分ベイズ法を用いた。結果として、サンプル数の少ない波動と、サンプル数の多いノイズをきっちりと分類することができた。結果として、WBA波動データを分類する手法として混合ガウスモデルの方が適切であることが確認された。今後は得られた波動に対して、そのスペクトル形状から定量的な定義づけを行う予定である。