

かぐや波形観測データに関する類似現象検索システムの検索効率と検索精度の改善

宮崎 大夢 [1]; 笠原 禎也 [1]; 後藤 由貴 [1]
[1] 金沢大

Improvement of efficiency and precision of the similar data retrieval system for KAGUYA/WFC-L

Hiromu Miyazaki[1]; Yoshiya Kasahara[1]; Yoshitaka Goto[1]
[1] Kanazawa Univ.

In recent years, measuring plasma wave has a critical role in understanding the physical phenomena in the solar terrestrial plasma. Measuring plasma wave is particularly essential to understand the microcosmic elementary process of plasma. However, the quantity of measured raw waveform data is very big, and it is extremely difficult to classify and analyze this data manually. Therefore, our study group has been developing a method to extract data similar to the wave phenomena designated by a user automatically from a large quantity of data measured by Akebono and KAGUYA. The purpose of this study is to improve efficiency and precision of the similar data retrieval system especially for the waveform data measured by WFC-L onboard KAGUYA.

The WFC-L data measures electric field waveform less than 100 kHz at sampling frequency 250 kHz, however, it is difficult to define the similarity when we directly treat this waveform data. In this study, we apply short-time Fourier transform to the WFC-L data to make spectrogram data and regard this spectrogram as an image, and calculate similarity of two dimensional images instead of comparing the waveform data directly.

Because the frequency coverage of the WFC-L is very wide from a few Hz to 100 kHz, we usually draw a spectrogram image using a logarithmic frequency scale. On the other hand, we apply FFT to the original waveform data in the conversion process into frequency domain, so that the frequency component is given by linear scale. In this study, we apply pixelization to the higher frequency range so as to make the size of each pixel uniform in a logarithmic frequency scale.

In the preparation process of the key parameters for similar data retrieval, we prepared several kinds of datasets which have different frequency and time resolution in order to meet the requirement of researchers depending on their interests in wave phenomena, which have various kinds of time and frequency scale. Improvement of the system is still in progress, and we plan to evaluate the performance and accuracy of the search results in the near future.

近年、太陽・地球系プラズマ中の物理現象の理解のために、プラズマ波動観測が重要な役割を果たしている。特に波形観測は、ミクロなプラズマ素過程の理解に必須であるが、生波形データはデータ量が非常に大きく、これらのデータを手作業で分類・解析することは困難を極める。このため、我々の研究グループでは、あけぼの及びかぐやで取得した大量の観測データから、ユーザがある一つの特徴的な波動現象を指定すると、それと類似したデータを自動抽出する方法の研究を行ってきた。本研究では、かぐや衛星搭載の WFC-L が取得した波形データを対象に、類似現象の検索効率と検索精度向上を目指す。

WFC-L データは 100kHz 以下の電界波形をサンプリング周波数 250kHz で計測するが、波形データをそのまま扱う場合、類似性の定義が難しい。そこで本研究では、WFC-L データに対して短時間フーリエ変換を実行して、スペクトログラムを作成し、このスペクトログラムを画像として捉え、二次元画像の類似性を求めることで、WFC-L データを直接比較する代わりとしている。

WFC-L データは対象とする周波数帯が数 Hz から 100kHz までと非常に広い為、検索で用いるスペクトログラム画像は周波数方向を対数で描いている。しかし、画像の元となるスペクトログラムは、波形データを FFT して求めるため、周波数成分が線形である。このため、類似性の算出にはこの差異の解決が必須であった。そこで本研究では、作成したスペクトログラムをそのまま用いるのではなく、低周波から高周波にかけてセルが指数関数的に大きくなるグリッドを用いて、スペクトログラムにモザイク処理を施したデータを作成し、検索に用いることにした。これにより、作成したデータが、周波数軸を対数で描いたスペクトログラム画像上で高周波から低周波まで均等に画像分割が可能となる。

WFC-L データは周波数・時間分解能が非常に高いという特徴を有している。このため、WFC-L データを対象とする類似検索では、ユーザが検索元データを選択する際に、研究目的に応じて、非常に細かいものから大きなものまで、幅広いスケールで対応することが要求されていた。そこで本研究では、先述のモザイク処理の粗さに複数の段階を設けることで、時間・周波数スケールの異なる複数のデータを作成することにした。作成された複数種のデータを検索によって使い分けることで、ユーザの検索目的に応じたスケールの違いに対応した検索の実現をめざした。

現在、新たに提案したアルゴリズムを採用した類似現象検索システムの改良を行っている。今後、新たなアルゴリズムが従来に比べてどの程度改善されているか評価を行う。