

## knowledge discovery from heterogeneous dynamic systems by time series data mining - inductive computing systems for space weather

# Terumasa Tokunaga[1]; Kazuyuki Nakamura[2]; Tomoyuki Higuchi[3]; Daisuke Ikeda[4]; Kakeru Ookubo[1]; Akiko Fujimoto[5]; Akimasa Yoshikawa[6]; Kiyohumi Yumoto[7]; Yumoto Kiyohumi MAGDAS/CPMN Group[8]

[1] none; [2] ISM; [3] Inst. Stat. Math.; [4] ISEE, Kyushu Univ.; [5] Earth and Planetary Sci., Kyushu Univ.; [6] Earth and Planetary Sci., Kyushu Univ.; [7] Space Environ. Res. Center, Kyushu Univ.; [8] -

We have been constructing the inductive computing system that make it possible to mining dependence relationships from non-stationary and heterogeneous data set for the next-generation space weather study. In recent years, various parameters have been observed simultaneously in interplanetary, magnetosphere, ionosphere and on the ground. With storing of a huge amount of data set, it has become increasingly important to develop data mining techniques that can extract local trends from them. Especially, in order to understand the coupling process between the solar wind, the magnetosphere and the ionosphere.

The solar wind-magnetosphere-ionosphere coupling system include following features: (1a) The whole size of the system is particularly huge, (2a) Driven by non-stationary solar wind, (3a) Different regions fluctuate in conjunction with each other, therefore, the system has dynamic traits, (4a) Parameters observed in different regions are heterogeneous. (5a) There is a large number of parameters.

These features have been made the data analysis in the field of space weather complicated. Specifically, (1b) The dependences between variables include not only a linear correlations but also a nonlinear correlation. In addition, they most likely have a nonlinear correlation with a time lag. (2b) The analytical results depend strongly on the analytical interval. In most cases, the analytical interval is chosen subjectively by researchers. (3b) Each variable has infinite variation patterns. (4b) It is difficult to evaluate the dependences and the timing relations between variables quantitatively. (5b) By manual analysis, we can analyze only limited time intervals. In most cases, hence, the spacial event is picked up based on a researcher's subjective interpretation. The past data analysis in the field of space weather study have poor support for these drawbacks. The present study is intended to construct the inductive computing system that make it possible to solve them flexibly for next-generation space weather study.

The basic procedure of times series data mining proposed in this study is as follows:

Step 1: Change-point detection based on text data mining or SSA(singular spectrum analysis).

Step 2: Local modering time series data mining to segment time series into stationary intervals.

Step 3: Clustering variation patterns segmented in step 2 into a finite number of groups.

Step 4: Extraction of the emerging pattern.

Step 5: Extraction of dependences that includes non-linear correlations with time lag between different variables.

The basic idea from step 2 to step 5 is proposed by Kinjo et al.,[2007]. We will introduce the whole framework of our application and report on progress and the future work.