Geophysical Research Abstracts
Vol. 20, EGU2018-1352, 2018
EGU General Assembly 2018


# Solar radio big data filtering using anomaly detection

Long Xu and Yihua Yan
Chinese Academy of Sciences, National Astronomical Observatories, Beijing, China (lxu@nao.cas.cn)

The modern telescope with high resolution and high accuracy raises a great challenge of big data. The same goes for solar observation. The newly established solar dedicated MingantU SpEctral Radioheliograph (MUSER) provides unprecedented high time, spatial and frequency resolutions at the same time. It is with excellent spectral diagnostic capability across ultrawide spectral broadband. There are 592 frequency channels, 25ms and 200ms time resolutions for MUSER-I (low frequency array with 40 antennas of 2.5m diameter) and 200ms for MUSER-II (high frequency array with 60 antennas of 4.5 m diameter), respectively. So, it can record the fast process of solar radio burst, benefit the study of energy release process, emergency and acceleration of high energy particles and origin characteristic of solar radio burst. Certainly, it also raises a big data of solar radio observation at the same time. It is a great demand to develop automatic data filtering, archiving or classification, which would make our following analysis/study more efficiently since only a very small percent of data concerning solar radio burst are interesting to us.

This paper makes the first attempt to utilize anomaly detection for data filtering or archiving. Specifically, we first image the mass data of solar radio observation, namely solar radio spectrums; then figure out the part of solar radio burst by recurring to anomaly detection. A solar radio spectrum is a gray-scale image representing solar radio radiation over multiple frequency channels and in a short time period. The vertical and horizontal dimensions of a spectrum correspond to frequency channel and time respectively. Intrinsically, time dependence exists between columns of a spectrum, which indicates the slowly varying process of solar radio radiation. A spectrum can be treated as a time sequence instead of a general spatial image without time sequential information. Inspired by the big success of anomaly detection on time sequence processing, e.g., abnormally human behavior in video sequence, it is reasonably believed that we can take advantages of state-of-the-art anomaly detection techniques to detect solar radio burst from massive records of slowly changing solar activities by treating a spectrum as a time sequence. Basically, fast changing frames among a time sequence concern solar radio burst and slow changing frames are related to the quite sun.

Here, we employ one of examples of anomaly detection algorithm, isolated forest (iForest) algorithm to fulfill the filtering or achieving of solar radio big data, exploring time sequential relations and interactions within a spectrum. The iForest algorithm was claimed good property of big data processing with favorable computational complexity. In addition, it is good at the situation of the lack of “positive” samples and abundant of “negative” samples, which is just the case concerning lots of astronomical observations, including solar radio observation. The experimental results demonstrate that the proposed anomaly detection algorithm can well filter the solar radio burst from mass data, and thus accomplished our mission of data filtering or archiving successfully.