ELSEVIER

CrossMark

Chinese Astronomy and Astrophysics 41 (2017) 125–135

CHINESE
ASTRONOMY
AND ASTROPHYSICS

An Auto-flag Method of Radio Visibility Data Based on Support Vector Machine† ⋆

DAI Hui-mei[1] MEI Ying[1,2] WANG Wei[3] DENG Hui[1]
WANG Feng[1,2△]

[1] *Computer Technology Application Key Lab of Yunnan Province, Kunming University of Science and Technology, Kunming 650505*

[2] *Yunnan Astronomical Observatory, Chinese Academy of Sciences, Kunming 650505*

[3] *National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012*

Abstract The Mingantu Ultrawide Spectral Radioheliograph (MUSER) has entered a test observation stage. After the construction of the data acquisition and storage system, it is urgent to automatically flag and eliminate the abnormal visibility data so as to improve the imaging quality. In this paper, according to the observational records, we create a credible visibility set, and further obtain the corresponding flag model of visibility data by using the support vector machine (SVM) technique. The results show that the SVM is a robust approach to flag the MUSER visibility data, and can attain an accuracy of about 86%. Meanwhile, this method will not be affected by solar activities, such as flare eruptions.

Key words Sun: activity—Sun: radio radiation—stars: imaging—methods: data analysis—techniques: miscellaneous

1. INTRODUCTION

The Mingantu Ultrawide Spectral Radioheliograph (MUSER) is the solar radio spectral imaging equipment with simultaneously high temporal, spatial, and frequency resolutions developed independently in China[1]. The MUSER project is performed in two stages:

† Supported by Joint Foundation of Chinese Academy of Sciences and National Natural Science Foundation for Astronomy (U1231205), National Natural Science Foundation (11103005, 11263004), Applied Basic Research Foundation of Yunnan Province (2013FA013, 2013FA032)

Received 2014–05–02; revised version 2014–08–06

⋆ A translation of *Acta Astron. Sin.* Vol. 57, No. 1, pp. 19–27, 2016

△ wf@cnlab.net

doi:10.1016/j.chinastron.2017.01.009

the first stage MUSER-1 (the low-frequency array) is composed of 40 paraboloid antennas of 4.5 m diameter and their receiver systems, and it images at 64 frequency points; the second stage MUSER-2 (the high-frequency array) consists of 60 paraboloid antennas of 2 m diameter and their receiver systems, and it images at 528 frequency points[2].

Different from the direct imaging of an optical telescope, the radio telescope mainly receives the intensity, phase, and polarization information of the object's radio signal, then realizes imaging by the succeeding treatment. The observational data will be affected by various factors in the observation (for example the electromagnetic interference), besides, the antenna and feed failures, channel mistake, system gain instability, receiver failure, etc., will also lead to the abnormality of the observed data. In the data processing of radio observation, it is an important job to judge, flag, and delete these abnormal data.

In the early period of radio astronomy research, the treatment of abnormal data relies on the artificial comparison of observational records. However, accompanying with the increasing amount of data, the artificial treatment becomes impractical. With the gradual expansion of the research field, in order to assess and flag the abnormal data effectively, according to their array mode, antenna number, baseline length, and other performance indices, the different telescopes all built their data flag models to match their data characteristics. For example, in order to flag the abnormal data, the Atacama Large Millimeter/Submillimeter Array (ALMA)[3,4] adopted the methods of TFCrop, RFlag, etc. in the Common Astronomy Software Applications (CASA); and in order to realize the automatic flag of abnormal data of the Giant Metrewave Radio Telescope (GMRT)[5,6], Prasad and Changalur adopted a flagging and calibration package for radio interferometric data (FLAGCAL) by the study and analysis[7].

Obviously, after MUSER entered the test observation stage, in order to realize the automatic data processing, and to improve the data accuracy and reliability, the automatic deletion of abnormal visibility data caused by antenna failures is an important preprocessing work of the succeeding data griding and clean mapping. Previously, we have tried to use the VSR (Vector to Scalar Ratio) flag algorithm in the GMRT and the ABC (Antenna/Baseline/Channedl) flag algorithm to flag the MUSER data, but the results are unsatisfactory.

Using the pattern recognition technique, based on the open software package LIBSVM, this paper studies the flag method of abnormal visibility data by the Support Vector Machine (SVM) technique[8], in which the training set is built in the condition of known damaged baselines. Then, by successive experiments, it is indicated that the SVM technique can realize very well the automatic flag of abnormal data in the MUSER observation.

2. THE SVM TECHNIQUE

SVM[8,9] is a kind of supervised learning method, and a kind of new machine learning method developed on the basis of statistical learning theory, it is built on the basis of