Cluster analysis of spectroscopic line profiles and EUV emission in RMHD simulations and observations of the solar atmosphere

Viacheslav Sadykov1,2,3, Irina Kitashvili2,3, Alexander Kosovichev1,2
1BAERI, NASA ARC. 2NJIT

Spatially-resolved observations from the IRIS, SDO/AIA, and other space missions and ground-based telescopes, coupled with realistic 3D RMHD simulations, are powerful tools for analysis of processes in the solar atmosphere. To better understand the dynamical and thermodynamic properties in the simulation data and their connection to observations, it is essential to determine similarities in the behaviors of the synthesized and observed emission. However, the complexity of observational data and physical processes makes comparison of observations and modeling results difficult. In this work, we show the initial results of application of K-Means clustering (unsupervised machine learning algorithm) to two different problems: 1) recognition of the typical spectroscopic line profiles observed by IRIS during solar flares and their typical dynamic behavior; 2) recognition of shocks and heating events in synthetic AIA emission data obtained from Sfyl-Rbox, Quiet-Sun simulations. The average silhouette width technique for the K-Means algorithm is utilized in different ways to obtain optimal numbers of clusters. We discuss application of the emission clustering to visualizations of the computational volume, understanding its evolutionary trends and behavior patterns, and inversion (reconstruction) of physical properties of the solar atmosphere from synthesizes emission data.

Identification of typical response of the upper chromosphere and lower transition region for the flare heating from IRIS observations

We detect shock features and their formation through the interface region imaging spectroharp (IRIS, de Pontieu et al. 2014) has observed hundreds of flares of ≥ C1.0. However, statistical studies of atmospheric response to the flare heating by IRIS are hampered because of the complexity of imaging spectroscopy data their high dimensional, large data volumes, optically-thick nature of the lines. Finding compact illustrative representation of the atmospheric response to the flare heating using unsupervised machine learning clustering techniques can simplify the analysis of large observational data sets and increase their understanding.

An example of clustering of C II 1334.5 Å line profiles for the M1.0 flare of June 12, 2014 is presented in Figure 1. The mean (Doppler shift), 0.0, represents the cluster centers as the means among the points of the same label, and repeats the procedure until there are no changes in labeling points.

Average silhouette width. The silhouette is defined for a data point as \( s(i) = \frac{b(i) - a(i)}{\max\{a(i), 0\}} \),

where \( a(i) \) is the average distance from point \( i \) to points of the same cluster, \( b(i) \) is the average distance from the point \( i \) to points of another closest cluster. The average \( s(i) \) across the points indicates how well the points lie within their clusters.

The optimal number of clusters can be estimated by maximization of the \( s(i) \).

When \( s(i) < 0 \), the points no longer "belong" to their clusters.

How many clusters to select? Inverse problem POV.

The number of clusters depends on the problem type: some problems (e.g. recognition of faint features in line profiles) may require selection of more clusters than dictated by the maximization of the average silhouette width.

In the example above, synthetic AIA emission is a function of the Differential Emission Measure (DEM, Cheung et al. 2015) of the computational domain. Reconstruction of the DEM from AIA emission is an example of ill-posed inverse problem.

Figure 3n illustrates that, if more than 7 clusters are selected in AIA emission space, the corresponding DEMs of these clusters strongly mix with each other. The average silhouette computed in DEM (generating parameter) space becomes negative, i.e. the points no longer correspond to their clusters in that space.

In general, the combination of unsupervised clustering in image space and machine learning external to the clustering is a generating parameter space can increase understanding and development of diagnostics of any inverse problem solely based on the known forward modeling results (see Figure 4).

Future plans and ideas.

• Recognition of typical line profiles and dynamical responses of the atmosphere to flare heating on images and data from large-scale flare events.

• Correlation of appearance of certain line profile shapes and dynamical behavior with properties of hard X-rays (from RHESSI and annular WIND) and soft X-rays (from GOES).

• Compilation of IRIS line profiles (Mg II, C II, Si IV) for the considered Solar-Rbox run using RH, a radiative transfer code. Testing the clustering algorithms on the synthesized emission reduced to IRIS and SDO/AIA data.

• Development of diagnostics tool for recognition of shocks, strong flows and heating events from SDO/AIA data and IRIS data based on cluster analysis discoveries.