Discovery From Hyperspectral ALMA Imagery With NeuroScope

Erzsébet Merényi[1,2], Andrea Isella[3], and Joshua Taylor[1]
[1] Department of Statistics
[2] Department of Electrical and Computer Engineering
[3] Department of Physics and Astronomy
Rice University
Houston, Texas, U.S.A.
email: {erzsebet,isella,jtay}@rice.edu

***Abstract*—We aim to add significant ALMA analysis capabilities for finding source regions with distinct kinematic and compositional properties. The NeuroScope approach is based on clustering with artificial neural maps and advanced knowledge extraction tools. It affords simultaneous use of multiple spectral lines and all velocity channels for discovery of complex structure and discrimination of subtle pattern variations.**

I. Motivation and Objectives

The Atacama Large Millimeter and sub-millimeter Array (ALMA) collects images at hundreds-to-thousands of frequencies with high spectral and spatial resolution. The resulting data cubes provide unprecedented opportunities for discovery of structure in protoplanetary disks, molecular clouds and other astronomical sources. However, these data also exhibit unprecedented complexity in addition to enormous volume, requiring new approaches to extracting relevant information. Traditional techniques (e.g., analyzing images at selected frequencies, moment maps) and even newer methods may become intractable or they can no longer fully exploit and visualize the rich information [1]. To address this problem, we develop tools — collectively called NeuroScope — to analyze ALMA data, and, in general, complex multi-dimensional data. These tools are rooted in state-of-the-art neural machine learning techniques. We show the effectiveness of NeuroScope on ALMA data of the protoplanetary disk HD142527 [2]. In the near future, we aim at automating our clustering for fast distillation of large data cubes in autonomous pipelines or on-board scenarios, and test it on more complex dataset, such those resulting from large field of view spectroscopic observations of molecular clouds.

II. Methods

The NeuroScope approach uses artificial neural map-based machine learning combined with advanced knowledge extraction methods. It identifies spectrally homogeneous spatial regions by clustering the spectral signatures in a 2-step process: a) intelligent summarization with advanced variants of Self-Organizing Maps (SOMs), which shrinks the data to a sparse representation with a small number of prototype spectra while reducing noise and retaining the salient pattern variations; b) grouping the prototypes using a "connectivity" measure that captures pattern affinities and topological relations from the SOM ([3] and references therein).

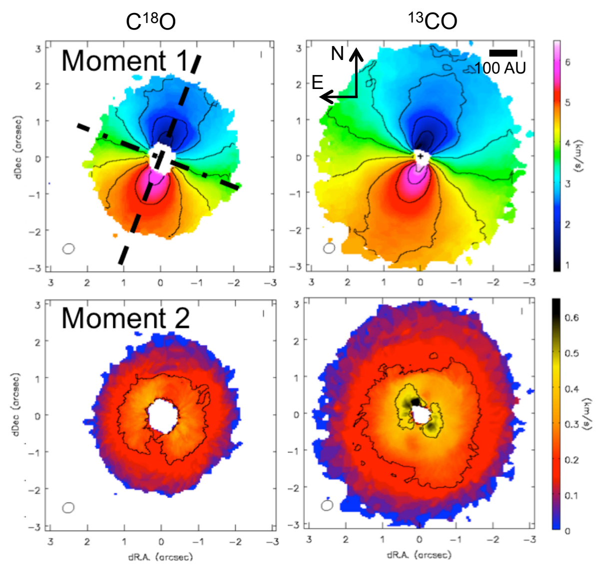

Fig. 1. Moment maps generated separately from molecular lines $C^{18}O$ J=3–2 and ^{13}CO J=3–2, from high-resolution observations of the protoplanetary disk HD142527 [2]. The dot-dashed and dashed lines, respectively, indicate the rotation axis and the apparent major axis of the disk. The moment 1 map shows the spatial distribution of the velocity of the emitting gas relative to the observer. The moment 2 map indicates the width of the line emission.

This process is capable of handing multiple spectral lines and all spectral channels simultaneously, sensing subtle but consistent pattern differences, and capturing clusters of widely varying statistical properties without requiring a predetermined number of clusters. This allows effective exploitation of the rich information though utilization of the full depth of these data, for discovery of complex structure. The intelligent sparse data summarization lends efficiency for large data.

III. Data and Preliminary Results

We test NeuroScope on ALMA data of the protoplanetary disks surrounding the young binary star HD 142527 [2]. The observation consists of two data cubes containing the J=3-2 rotational line emission of $C^{18}O$ and ^{13}CO molecules,

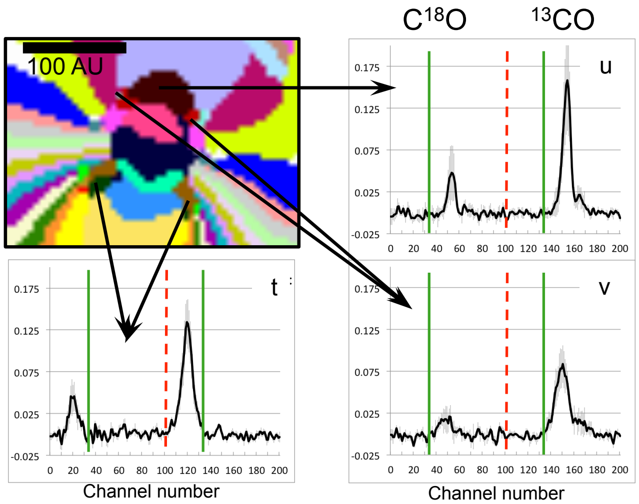

Fig. 2. Magnified center of the disk with clusters (regions highlighted by different colors) discovered from the 200-channel spectra combined from the $C^{18}O$ and ^{13}CO lines., and plots of mean spectra of three clusters. The dashed vertical red line in each plot indicates where the two spectral lines were concatenated. The vertical green line is drawn at the rest frequency within each of the spectral lines. Colors were chosen to contrast regions (not to provide a heat map). The mean spectra of the three clusters each exhibit widening of the spectral lines, or a double peak in the case of the cluster labeled "t" (very small, dark green downward pointing pair of spatial areas). These may suggest departure from the Keplerian rotation.

respectively. Each data cube comprises 100 velocity channels. Despite the fact that protoplanetary disks have a rather simple kinematics consistent with Keplerian rotation, NeuroScope analysis revealed regions with subtle but consistent discriminating patterns, which are missed by moment maps (Figs 1 and 2).

The combined (concatenated) 200-channel spectra are input to our clustering to produce a cluster map. For reasons of space constraints, we show in the top-left panel of Fig. 2 only the magnified center of the cluster map. Each cluster (a differently colored spatial region) signifies a spectrally homogeneous region, with spectral characteristics different from regions depicted by other colors. Thus the variations of the compositional and kinematic behavior are portrayed in a 2-dimensional map which is easy to inspect visually and to notice possible irregularities. The kinematical and compositional properties of cluster regions can be examined from their average signatures, which can be automatically generated. In Fig. 2, the standard deviations are also plotted for each channel to indicate the within-cluster variations, which is also a measure of the reliability of the clusters. The sample plots of three interesting clusters, pointed at by arrows in Fig. 2, exhibit double or widened velocity peaks. (Notice that two of those clusters, "t" and "v" consist of two symmetric parts.) Cluster "t", in particular, has two ^{13}CO peaks at opposite sides of the rest frequency (green line), which may be caused by two gas components traveling in opposite directions. The larger regions adjacent to cluster "t" (such as the arcuate medium blue and several yellow-hued clusters) have abruptly different spectral properties (not shown here). In a protoplanetary disk, clusters like the ones highlighted in Fig. 2 could indicate the presence of non-Keplerian motion, such as, for example, inflow motion of gas toward the star or onto forming planets. The moment 2 map of the ^{13}CO line shows widening at two of the locations indicated by arrows in the cluster map but misses the most prominent "t" region. Whereas an in-depth analysis and physical interpretation of these (and other) kinematic features in HD142527 is ongoing, it is clear that NeuroScope is capable of expanding the space of discovery provided by ALMA observations.

A. Conclusion and Outlook

We show here an example of finding well-defined parts of the protoplanetry disk HD142527 with spectral properties that have potential physical relevance for understanding the kinematical structure of this system. Finding the small spatial areas that exhibit unusual behavior hinges on "precision mining" the high-dimensional ALMA cube and accurately delineating them from other areas. NeuroScope is tool in development. Future work will include exploration of visualization schemes where colors can be used to express some relationships across regions; or interactive visualization where the spectral plots of regions pop-up on demand. Future work will also be aimed at perfecting the automation of the (partially interactive) clustering process [4], so that NeuroScope could be used in pipeline processing at high speed without compromising quality. To complete the analysis of the HD142527 system, NeuroScope will be tested on ALMA mock-up observations of disks characterized by kinematical signatures deviating from the Keplerian motion. Finally, we will exploit NeuroScope capabilities by analyzing more complex astronomical data, such as the multidimensional spectral cubes resulting from wide field of view mapping of nearby molecular cloud.

ACKNOWLEDGEMENT

This work is partially supported by ALMA Cycle 5 Study Grant #1519126.

REFERENCES

[1] E. Merényi, J. Taylor, and A. Isella, "Deep data: discovery and visualization application to hyperspectral ALMA imagery," *Proceedings of the International Astronomical Union*, vol. 12, no. S325, pp. 281–290, 2016.

[2] Y. Boehler, A. Isella, E. Weaver, C. Grady, J. Carpenter, L. Perez, and L. Ricci, "A close-up view of the young circumbinary disk HD 142527," *Ap. J.*, vol. 840, no. 60, p. 16pp, May 2017.

[3] E. Merényi, K. Taşdemir, and L. Zhang, "Learning highly structured manifolds: harnessing the power of SOMs," in *Similarity based clustering*, ser. Lecture Notes in Computer Science, LNAI 5400, M. Biehl, B. Hammer, M. Verleysen, and T. Villmann, Eds. Springer-Verlag, 2009, pp. 138–168.

[4] E. Merényi and J. Taylor, "SOM-empowered graph segmentation for fast automatic clustering of large and complex data," in *12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM+ 2017)*, June 2017, pp. 1–9.