SFB-AstroML Day

Better astrophysics with ML?!

November 5, 2020

UNIVERSITÄT BONN

Universität zu Köln

Max-Planck-Institut für Radioastronomie

SFB 956 Cologne | Bonn

Contents

About **3**
Organizing committee 3

Timetable **4**
5 November 2020 4

List of Abstracts – Talks **5**

About

Machine learning (ML) techniques are becoming popular in the astronomical community for big data analyses, fitting applications and predictions. The goal of this workshop is to give an overview of the ongoing projects which utilize ML techniques within the current SFB 956, respectively the Cologne and Bonn institutes. The workshop could inspire future projects for the new SFB proposal 2023.

Organizing committee

Stefanie Walch-Gassner	Thomas Bisbas	Brandt Gaches
Stefano Ebagezio	Ekaterina Makarenko	Uliana Pyrohova
Paul Rohde		

Timetable

5 November 2020

Zoom link:
https://uni-koeln.zoom.us/j/92012059156?pwd=NTF6NmFzRFVVQUYxbWhZN1AvN2ZiZz09

13:00–13:15	**Welcome remarks**	
13:15–13:45	**Christof Buchbender** UzK	Improved Data Reduction Using PCA Decomposition of Spectra
13:45–14:15	**Tianwei Zhang** UzK	PCA in classification of star-forming cores
14:15–14:45	**Dario Colombo** MPIfRA	The Spectral Clustering for Molecular Emission Segmentation (SCIMES) algorithm
14:45–15:00	**Coffee break**	
15:00–15:30	**Lars Langen** UzK	Constructing an Artificial Neural Network to study CO Spectral Line Energy Distributions and the underlying ISM properties
15:30–16:00	**Jonathan Holdship** Leiden	Emulating Chemistry for Hydrodynamical Models
16:00-16:30	**Kelvin Lee** MIT	Semi-Supervised Machine Learning for Molecular Spectroscopy
16:30–17:00	**Discussions**	

List of Abstracts – Talks

Improved Data Reduction Using PCA Decomposition of Spectra

Christof Buchbender buchbend@ph1.uni-koeln.de

Universität zu Köln **Postdoc/SFB**

The quality of data taken with heterodyne receivers depends to a large degree on the stability of the receiver system and atmospheric conditions between subsequent ON and OFF integrations. Instabilities will result in non-flat baselines and in result in higher noise (looking a larger bandwidths). Further effects like standing waves along the optic pathway can add irregularities to the baseline. These effects together can lead to strongly warped baselines. The go-to approach to cure these problems is to fit a polynomial to the baseline excluding any emission region and subtracting this baseline from the spectrum. However this method can not account for all baseline distortion. It is especially difficult to correct baselines directly under line emission. In this talk I will show a new approach to identify systematic variations in the baseline originating in instabilities in the telescope system, i.e., backends, receiver, telescope (optics), and atmosphere, over the course of the observations. We are using a PCA decomposition of spectra to find common structures of non astronomical origin in the spectra and use these to correct the spectra. Here I will introduce the tools I use to implement this method and show a few use cases and pathological cases that could not be corrected with more traditional approaches.

The Spectral Clustering for Molecular Emission Segmentation (SCIMES) algorithm

Dario Colombo dcolombo@mpifr-bonn.mpg.de

Max Planck Institute for Radio Astronomy **Postdoc/SFB**

Modern high resolution, molecular gas surveys of the Galactic Plane are unveiling an astonishing picture of the three-dimensional gas organization of the Milky Way. This provides the opportunity to investigate the building blocks of the molecular medium, the Molecular Clouds (MCs), on the resolved point of view. Until recent time, however, such a study has been restricted to a handful of targets, since common automatic segmentation methods, being severely affected by survey designs, are unable to decompose GMC-sized objects out from the diffuse medium in high resolution data. The algorithm we wrote, SCIMES (Spectral Clustering for Interstellar Molecular Emission Segmentation), overcomes these limitations by applying graph theory and unsupervised machine learning methods to identify molecular clouds that fully retain their inner hierarchical structure. SCIMES has been widely used to identify clouds from large Galactic plane surveys such as COHRS (Colombo et al. 2019) and SEDIGISM (Duarte-Cabral et al. in press), and for simulations (Duarte-Cabral et al. 2016, 2017). In this talk, we will illustrate the concepts behind the algorithm, as well as, our recent results on cloud properties, scaling relations, star formation efficiencies, and resolved properties, such as cloud morphology.

Emulating Chemistry for Hydrodynamical Models

Jonathan Holdship holdship@strw.leidenuniv.nl

Leiden Observatory **Postdoc**

The thermodynamics of an astrophysical gas is strongly coupled to the chemistry that is taking place. Chemistry sets the abundances of major coolants, affecting the cooling rate of a gas and many chemical reactions directly heat or cool the gas in which they take place. For this reason, many dynamical models employ chemical solvers to take these effects into account. However, due to the complex nature of state of the art chemical models, the chemistry is often greatly simplified. It is common for these models to used reduced chemistry sufficient to reproduce the abundances of major coolants.

We present a fully time dependent, gas-grain chemical model which for a small set of input parameters can solve the heating/cooling problem for an arbitrary point in a cloud. We focus on the methods including dimensionality reduction for parameter space sampling and emulator training as well the the neural networks used to produce an emulator that provides the same outputs for a given set of inputs but with much less computational time. This technique, when fully developed, will allow us to publish a module for dynamical codes that approaches the accuracy of a full chemical model but has extremely low computational cost.

Constructing an Artificial Neural Network to study CO Spectral Line Energy Distributions and the underlying ISM properties

Lars Langen llangen3@smail.uni-koeln.de

Universität zu Köln **MSc. Student**

Understanding observations of CO Spectral Line Energy Distributions (CO-SLEDs) of an object and gaining information over the underlying physical parameters of the interstellar medium (ISM) is challenging, yet important in order to set up models and simulations for further investigation. We introduce an artificial neural network (ANN) that uses a large set of pre-calculated CO-SLEDs as training data in order to identify the most probable ISM environmental parameters. The CO SLEDs have been calculated using the photodissociation region code 3D-PDR. The environmental parameters we examine include the cosmic-ray ionization rate, the far-UV intensity of the radiation field and the metallicity. We consider the first ten rotational transitions of CO, both lines of atomic carbon (CI) and the line of ionized carbon (CII). We apply the neural network to identify the environmental parameters of a small sample of galaxies and we make predictions for the emission of lines that have not yet been observed.

Semi-Supervised Machine Learning for Molecular Spectroscopy

Kelvin Lee kelvlee@mit.edu

MIT **Postdoc**

Machine learning is an attractive way to change scientific workflows, ranging from automation of abstract tasks such as data analysis and reduction, to surpassing human level of information content being extracted from scientific data. In particular, improvements to both instrumentation and methodology have simultaneously impacted the quantity and quality data collected in now routine high resolution molecular spectroscopy studies; in order to match the cadence of data collection, there is a significant need to develop "computer spectroscopists" that can perform significant amounts of data analysis, allowing scientists to focus on drawing scientific conclusions. While unsupervised machine learning is the simplest path to achieving these goals, it has become apparent in recent times that we cannot rely solely on solutions and representations learned by unsupervised models. Semi-supervised machine learning, in particular deep learning, combines the most powerful aspects of both unsupervised machine learning and physical/chemical intuition. In this talk, I will provide a short overview of two deep learning projects that I have recently been working on: models to reconstruct ion images for studying reaction dynamics, and spectroscopic segmentation in the context of complex mixture analysis and molecule identification using rotational spectroscopy.

PCA in classification of star-forming cores

Tianwei Zhang zhang@ph1.uni-koeln.de

Universität zu Köln **PhD Candidate/SFB**

To better understand the evolutionary stages of star formation, pre/proto-stellar cores should be classified. As we gain a bunch of spectra towards these cores, PCA works well in the classification. Four stages are found with different typical lines, temperatures, densities, and spatial distributions.