SF2A 2023
M. N'Diaye, A. Siebert, N. Lagarde, O. Venot, K. Baillié, M. Béthermin, E. Lagadec, J. Malzac, J. Richard (eds)

MACHINE LEARNING TO FACILITATE THE STUDY OF COMPLEX ORGANIC MOLECULES IN HOT CORES

N. Kessler[1], T. Csengeri[1], D. Cornu[2] and S. Bontemps[1]

Abstract. The proper analysis of (sub)millimeter spectra containing emission from complex organic molecules is challenging for large samples. We study here a deep learning method to facilitate the analysis of molecular signatures for wide-band spectroscopic data for large samples. We demonstrate the feasibility of using neural network algorithms to identify the molecular content of synthetic spectra based on LTE models. Applying such tools to observational data could be a game changer in exploiting large datasets, and thereby help us to better understand the physical and chemical processes in star forming regions.

Keywords: star formation, astrochemistry, radioastronomy, machine learning

1 Introduction

During star formation complex organic molecules (COMs) form, and lead to the emergence of chemically rich regions such as hot cores, hot corinos, and shocks. State-of-the-art interferometers such as ALMA and NOEMA give us access to a large bandwidth at high spectral resolution in the (sub)millimeter domain, where most of the molecular emission is observed. They make it possible to search for new molecules in the interstellar medium, and to reveal a high complexity of the molecular gas in these regions (Belloche et al. 2014) with a significant increase in abundance of COMs towards chemically active regions (Csengeri et al. 2019; McGuire 2022). Nevertheless, the inspection of (sub)millimeter spectra that is rich in COMs is often time consuming and current analysis techniques are challenged by the arrival of large surveys, that require a systematic analysis. It is therefore necessary to develop new tools based on machine learning to automate the detection and the analysis of molecular emission. We present our first results in this area, as well as challenges and future prospects for the progression of machine learning in radioastronomy.

2 Artificial neural networks to study complex organic molecules in radioastronomy

The scientific community has proved that artificial neural networks (ANNs) can model complex functions to detect and recognize objects with a high accuracy, even in real time (LeCun et al. 2015). These machine learning models then appear as a way to automate the extraction of features from (sub)millimeter spectra.

2.1 Neural networks and training sets

We implemented a neural network using the framework of CIANNA (Convolutional Interactive Artificial Neural Networks by/for Astrophysicists) developed by Cornu (2020). The use of convolutional neural networks (CNNs) makes it possible to extract relevant patterns in 1D spectra that are determined by the receptive field of the convolutional layers. Thus, small scale information such as clusters of emission lines or their line profile can be taken into account, or large scale patterns such as the distance between two lines. This extracted data is treated by dense layers to learn the spectroscopic signature of each involved molecule. We developed an architecture optimised for recognizing the spectral pattern of a certain molecule. The structure of such a neural network is shown in Fig. 2 as an illustration. We used Local thermodynamic equilibrium (LTE) models over a wide range of parameters as a training set, an example of a spectrum from the training set for methanol detection is shown Fig. 1.

[1] Laboratoire d'Astrophysique de Bordeaux, Univ. Bordeaux, CNRS, UMR 5804, F-33615 Pessac, France
[2] LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Univ., UMR 8112, F-75014 Paris, France

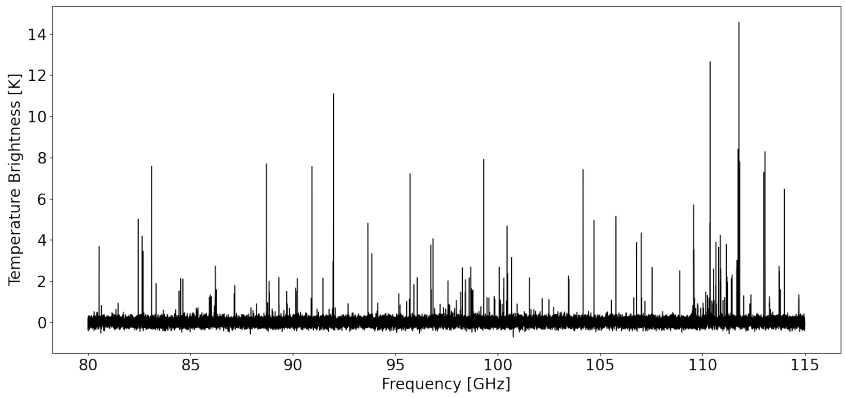

Fig. 1. Example spectra used as for the training set for methanol detection. The spectral signatures of four molecules, and noise, compose this spectrum : CH_3OH ($N = 6.6 \times 10^{15}$ cm^{-2}, $T_{ex} = 30.0$ K), CH_3OCH_3 ($N = 8.9 \times 10^{16}$ cm^{-2}, $T_{ex} = 107.8$ K), CH_3CN ($N = 6.7 \times 10^{14}$ cm^{-2}, $T_{ex} = 38.7$ K), CH_3OCHO ($N = 8.9 \times 10^{16}$ cm^{-2}, $T_{ex} = 232.2$ K).

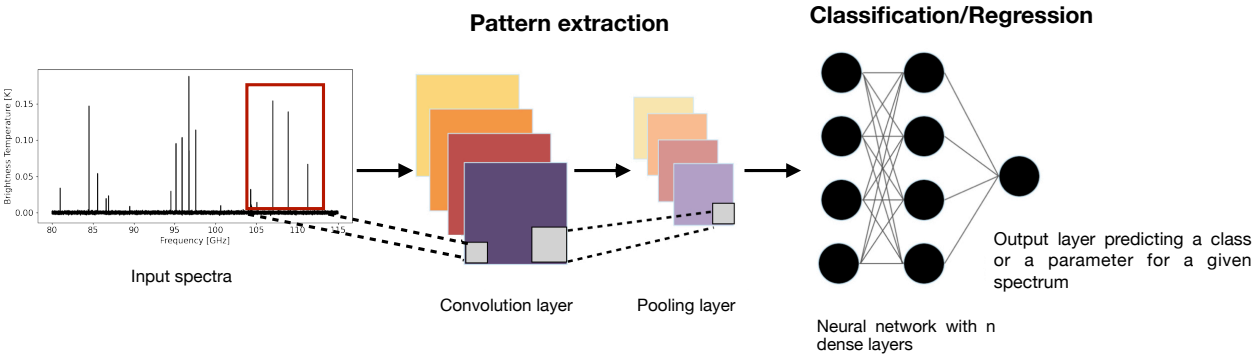

Fig. 2. Scheme of the structure of a convolutional neural network used for spectra classification or regression. The CNN takes as input a 1D spectrum. Convolutional layers extract patterns. These patterns are taken as input by dense layers. The last dense layer gives a classification or a parameter prediction depending on the use of the CNN.

2.2 First results on the detection of COMs with CNNs

The developed convolutional neural network is tested for the detection of methanol emission lines. On a data set of 2000 synthetic spectra – disjoint from the training set –, of which 1340 contain the molecule to be detected, the CNN performs a true detection for 1267 spectra with a 50% confidence threshold, and 26 false detections as shown on the left part of Fig. 3. These false detections correspond to cases with a strong confusion, and to cases located in a region of the parameter space poorly known by the CNN as suggests the histogram of Fig. 3.

3 Conclusions and perspectives

Artificial neural networks are able to learn the spectroscopic signature of molecules. Based on this principle, we developed a CNN capable of learning the information present in LTE models to then detect and identify complex organic molecules within rotational spectra. This tool will be improved for a more versatile use for recognizing more species, and will be highly beneficial for the efficient exploitation of large spectroscopic surveys, such as the NASCENT-Stars large program that aims to constrain the molecular composition of star forming regions in the Cygnus-X molecular complex with NOEMA over the next years.

Nina Kessler acknowledges financial support from the AAP DOCTORAT INTELLIGENCE ARTIFICIELLE, from the RRI ORIGINS, and the use of the Mesocentre de Calcul Intensif Aquitain (MCIA). T. Cs. has received financial support from the French State in the framework of the IdEx Université de Bordeaux Investments for the future Program. This work was supported by the Programme National "Physique et Chimie du Milieu Interstellaire" (PCMI) of CNRS/INSU with INC/INP co-funded by CEA and

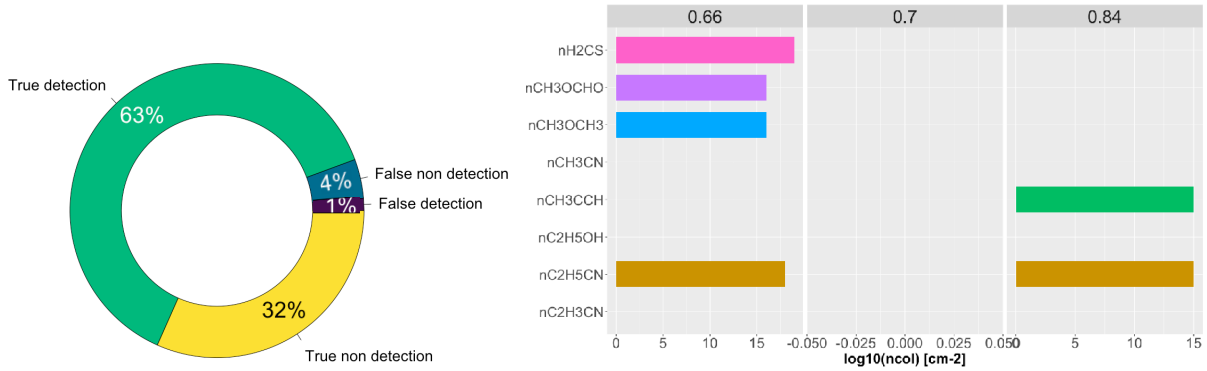

Fig. 3. Left: Doughnut chart of the CNN predictions on the presence or absence of the methanol spectral signature. **Right:** Histograms of the chemical composition with their respective column density of three spectra with false detection. The number on the top of each column is the probability given by the CNN that these spectra are composed of methanol.

CNES. This work was supported by the ""Programme National de Physique Stellaire"" (PNPS) of CNRS/INSU co-funded by CEA and CNES. The speaker expresses gratitude to the organizers of the SF2A annual meeting for gathering the French community and giving the opportunity to discuss the progress and challenges of astronomy and astrophysics.

References

Belloche, A., Garrod, R. T., Müller, H. S. P., & Menten, K. M. 2014, Science, 345, 1584

Cornu, D. 2020, arXiv e-prints, 251

Csengeri, T., Belloche, A., Bontemps, S., et al. 2019, A&A, 632, A57

LeCun, Y., Bengio, Y., & Hinton, G. 2015, Nature, 521, 436

McGuire, B. A. 2022, ApJS, 259, 30