

[IDENTIFICATION OF SPECTRAL LINES USING SPARSE CODING]

ANDRÉS RIVEROS, KARIM PICHARA, PAVLOS PROTOPAPAS, DIEGO MARDONES, PAULINA TRONCOSO AND MAURICIO ARAYA
Draft version March 5, 2015

ABSTRACT

Astronomy is facing new challenges on how to analyze big data and therefore, how to search or predict events/patterns of interest. The automatic detection and identification of spectral lines is an astronomical problem that has not been solved yet, so currently the identification is limited to the manual analysis of spectra by radio-astronomers. The use of spectroscopy allows to describe the chemical composition of astronomical objects through the emissions from the interaction between the radiation and the matter, thus causing emission lines. New observations in previously unexplored wavelength regions that will be available thanks to project as the Atacama Large Millimeter Array (ALMA), with which it is intended to use machine learning techniques to identify automatically these emission lines. Using simulated data based on the observations that has been obtained from the radio-telescope ALMA, it is proposed an algorithm with which to identify lines in order to determine the molecules that compose the observed galaxies. For this will be used the technique of Sparse Coding, that will evaluate the molecules that can origin the observed spectra, and it is expected to be found the best combination of molecules to recreate that spectra. From this algorithm, the astronomers may obtain a probability associated to the possible combinations of molecules composing astronomical objects.

Subject headings: spectral lines: emission lines; spectroscopy techniques; method: machine learning

1. INTRODUCTION

This project is developed as part of a collaborative project between several Chilean universities for the initiative of the creation of a Chilean virtual observatory Observatorio Virtual Chileno (ChiVO). ChiVO is an on-line platform that will make available to astronomers the measures of the radio-telescope Atacama Large Millimeter Array (ALMA). Also, ChiVo will provide several tools in order to process the data of ALMA measurements.

El data from the radio-telescope ALMA consist of data cubes. The three dimensions correspond to two spatial dimensions, and a third dimension of wave frequency. This means that for each spatial point, can be obtained a spectrogram. The Pontifical Catholic University of Chile will participate with the development of an algorithm to identify spectral lines using machine learning techniques. This tool takes as input a observed spectra and returns a list with the best prediction of molecules that by the theoretical behavior their spectral lines describes the observed spectra.

Currently, there is not available ALMA data to train a model, so this project has been developed using synthetic data. The Astronomical SYnthetic Data Observatory (ASYDO) project, a parallel project of ChiVO, will be the used in order to generate synthetic data to test the algorithm.

2. RELATED WORK

3. BACKGROUND

3.1. Spectroscopy

3.1.1. Sparse Coding

3.2. Dictionary

4. METHODOLOGY

4.1. Splatalogue



FIG. 1.— ...

4.2. Synthetic Data

4.3. Codification

5. RESULTS

5.1. Presence of Isotopes

6. CONCLUSIONS

[What is good, what was bad, and what we want to do next but just a glimpse of that because we don't want copycats]

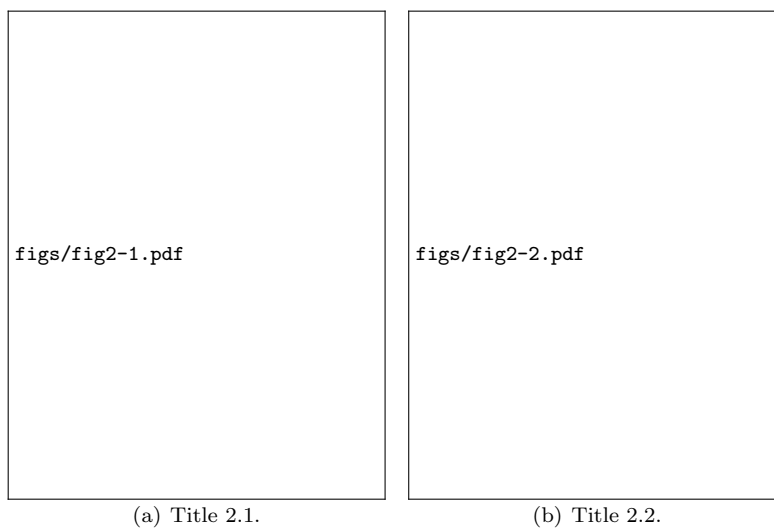


FIG. 2.— In figure 2(b) is shown that ...