Contribution ID: **185**   Type: **Oral Presentation**

Predictive Modeling of Galaxy Spectra: A Comprehensive Framework Using Transformer Architecture

Tuesday 9 July 2024 12:00 (20 minutes)

In this study, we developed a foundational model leveraging transformer architecture to analyze galaxy spectra across optical and UV photometric bands. Our objective is to create a comprehensive model capable of simultaneously predicting the characteristics of stellar populations, emission lines, and photometric redshifts from a given observational dataset. Utilizing data from the SDSS and the DESI survey, we generated simulated observations compatible with the J-PAS. Subsequently, we augmented our dataset with UV emission from the GALEX survey in the FUV and NUV bands. The dataset encompasses a wide range of physical properties for each galaxy, including emission lines in the optical range, stellar mass, metallicity, stellar extinction, and spectroscopic redshift.

Similar to large language models, our transformer is trained in an unsupervised manner, aiming to predict the next element in a sequence based on its predecessors. This approach diverges from conventional deep learning algorithms that typically map a set of inputs (observations) to outputs (physical properties). In our transformer model, any piece of galaxy information can serve as either input or output during training, offering unparalleled flexibility. This method does not require complete observations and measurements for each galaxy; rather, it uses subsets of available data alternately as inputs and outputs in iterative training phases. For instance, in one iteration, the model might predict the EW of Hα using only the blue portion of the spectrum, while in another, the mass of the galaxy and the [OIII] emission line might predict the red spectrum part.

The model is further trained on actual observations from miniJPAS to perform domain adaptation, bridging the gap between simulated and observational data. Our results demonstrate the model's effectiveness in characterizing galaxy properties within a unified framework, traditionally achieved through multiple separate codes. Future enhancements will include training with additional data from DESI and J-PAS and possibly extending to photometry in other wavelengths, such as the infrared.

Presenter: MARTÍNEZ SOLAECHE, Ginés

Session Classification: Past and future multiwavelength all-sky surveys