Yongxiong Wang

Manchester, UK yongxiong.wang@postgrad.manchester.ac.uk +44 775-232-4567

SCIENTIFIC RESEARCH

OUTFLOW SEGMENTATION USING MACHINE LEARNING **Manchester, UK** **Jan 2021 – Current**

Developing a Convolutional Neural Network for the outflow segmentation in the real-observed NGC 253 data.

- Calibrated ALMA (band 7) observed high resolution (~2.5pc) CO(3-2) data in CASA, to get it scientifically ready;
- Expertly contoured and labeled outflows in every channel to train the network for segmentation, resulting in a highly accurate and reliable segmentation model;
- Developed and trained a machine learning model (Domain Adaptation) in Python (NumPy, Torch, pandas, etc), resulting in a highly effective segmentation model that outperformed existing models in the field.

TIME SERIES ANALYSIS USING MACHINE LEARNING ON BOTH FREQUENCY AND TIME DOMAINS

Frequency-domain Prediction using Random Forest Regressor

- Conducted the transformation of the raw data from the time domain to the frequency domain using Fast Fourier Transform (FFT);
- Executed a comprehensive Machine Learning pipeline, including data cleansing and preprocessing, feature engineering, model training, and evaluation;
- Achieved an R2 score of 0.99 by employing the Random Forest Regressor.

Time-domain prediction using autoregression and LSTM model

- Implemented time-series analysis on a time-domain dataset using LSTM;
- Executed autocorrelation function (ACF) and partial autocorrelation function (PACF) to assess the stationary of the time-series dataset;
- Fitted the time-series dataset with the autoregression (AR) algorithm;
- Demonstrated the prediction results based on the fitted model.

DIMENSION REDUCTION ON STRUCTURAL DATA **Manchester, UK** **Aug 2019 – Sep 2022**

Using big data analysis and machine learning to compress high-dimensional structural data to lower dimensions, to quantify and visualize the contribution of features in the research of star formation

Using Non-negative Matrix Factorization (NMF) for dimension reduction then applied to clustering analysis

- Efficiently performed line identifications (65 molecules identified) in GILDAS for 38 sources observed by Tracing Evolution in Massive Protostellar Objects (TMEPO) survey which was observed by ALMA (band 6), then used XCLASS to fitting 11 selected molecules to get the physical parameters as high-dimensional structural data, which facilitated further analysis and interpretation of the data, and finished those work within 4 months;
- Took the initiative to adopt NMF on structural data and performed the project independently, resulting in the successful compression of over 6000-dimensional structural data into 2-9 dimensions and the identification of 15 contributive features;
- Applied Silhouette analysis to quantify the best number of clustering (K) in K-means clustering and visualized the compressed data in Python, which led to a better understanding of the data and improved insights into the underlying patterns.

Using Principal Component Analysis (PCA) on Simulational Structural Data

- Simulated UCLCHEM model (in Fortran) to obtain 2000-dimensional structural data;
- Compressed high-dimensional data by PCA and quantified the contribution of selected molecules;
- Visualized the changing trend of physical parameters in Python using Matplotlib, Datashader, and other relevant libraries, which aimed to identify key trends and quantify the contributions of molecules;
- Articulated data analysis findings and conclusions to write up an academic paper.

MODEL DIELECTRIC PROPERTIES OF BIOLOGICAL DATA **Manchester, UK** **Jan 2017 – July 2019**

Reveal dielectric properties of human tissues in the Debye model at different frequencies

- Run the one-pole Debye model in Fortran to obtain the dielectric parameters of 50 human tissues at frequency ranges from 100 MHz to 6 GHz;
- Demonstrated strong problem-solving and analytical skills through project design, data collection, and results in performance.

COMPLEX NETWORKS RESEARCH **Qingdao, China** **Sep 2014 – May 2015**

The evolution and prediction of 3 different complex networks using K-core algorithm,

- Collected data using data mining from open data sets and quantified the evolution of complex networks in Pajek;
- Honed project execution and management abilities in performing the whole research, helping apply *the Zhejiang Provincial Key Discipline Open Research Fund*.

EDUCATION

UNIVERSITY OF MANCHESTER, JBCA **Manchester, UK** **Apr 2019 – Current**

Ph.D. Astronomy and Astrophysics

- Machine Learning and Big Data Analysis in star formation

UNIVERSITY OF MANCHESTER **Manchester, UK** **Apr 2016 – Aug 2019**

M.Phil. Electrical and Electronic Engineering

- Project: Identification of the influence of one-pole Debye parameters on human tissues

QINGDAO TECHNOLOGICAL UNIVERSITY **Qingdao, China** **Sep 2011 – Jul 2015**

B.Sc. Electronic Engineering

- GPA: 3.6/4
- Awards: Excellent student of Shandong Province, President scholarship of Qingdao Technological University

SKILLS

- Programming Language: Python (5+ years experience) (Pandas, Numpy, Pytorch, Tensorflow, Keras, bs4), Fortran, MySQL, Shell Script, R, C++
- Algorithm: PCA, NMF, K-means, t-SNE, time series analysis
- Software: CASA, GILDAS, XCLASS, DS9, napari
- Soft Skills: Great passion and strong insight into data, Good self-organization, Advanced Open-water Diver
- Languages: English – professional proficiency; Chinese – native; Japanese – elementary

Hor hours