GABRIJELA ZAHARIJAS & LUC HENDRIKS

INCLUSIVE ANALYSIS OF FERMI-LAT POINT SOURCES WITH MACHINE LEARNING

IDEA AND GOAL OF CHALLENGE

[Salvetti+ (2017)]

▸ Machine-learning algorithms used for screening and classification of unassociated sources based on γ-ray data only.

  ▸ Mirabal+ (2016): to pinpoint potentially novel source classes

  ▸ Saz Parkinson+ (2016): to classify UnA as likely AGN or PSR (including the likely type of pulsar) using timing and spectral information

  ▸ Salvetti+ (2017, 3FGLZoo): classification UnA likely AGN as likely BLL or FSRQ

IDEA AND GOAL OF CHALLENGE

▸ ML techniques that can be used for point source identification

  ▸ γ-ray only

  ▸ Multi-wavelength

  ▸ Timing information

▸ Find origin of unassociated sources

  ▸ Classify in existing source classes

  ▸ New source classes? Dark matter?

WHAT DATA TO USE FOR THE CHALLENGE ?

▸ Three stages:

▸ **γ-ray data only**
See how far we can get with γ-rays only. Reproduce past results

▸ **Add multi-wavelength**
e.g. most γ-ray pulsars are found by looking at radio first
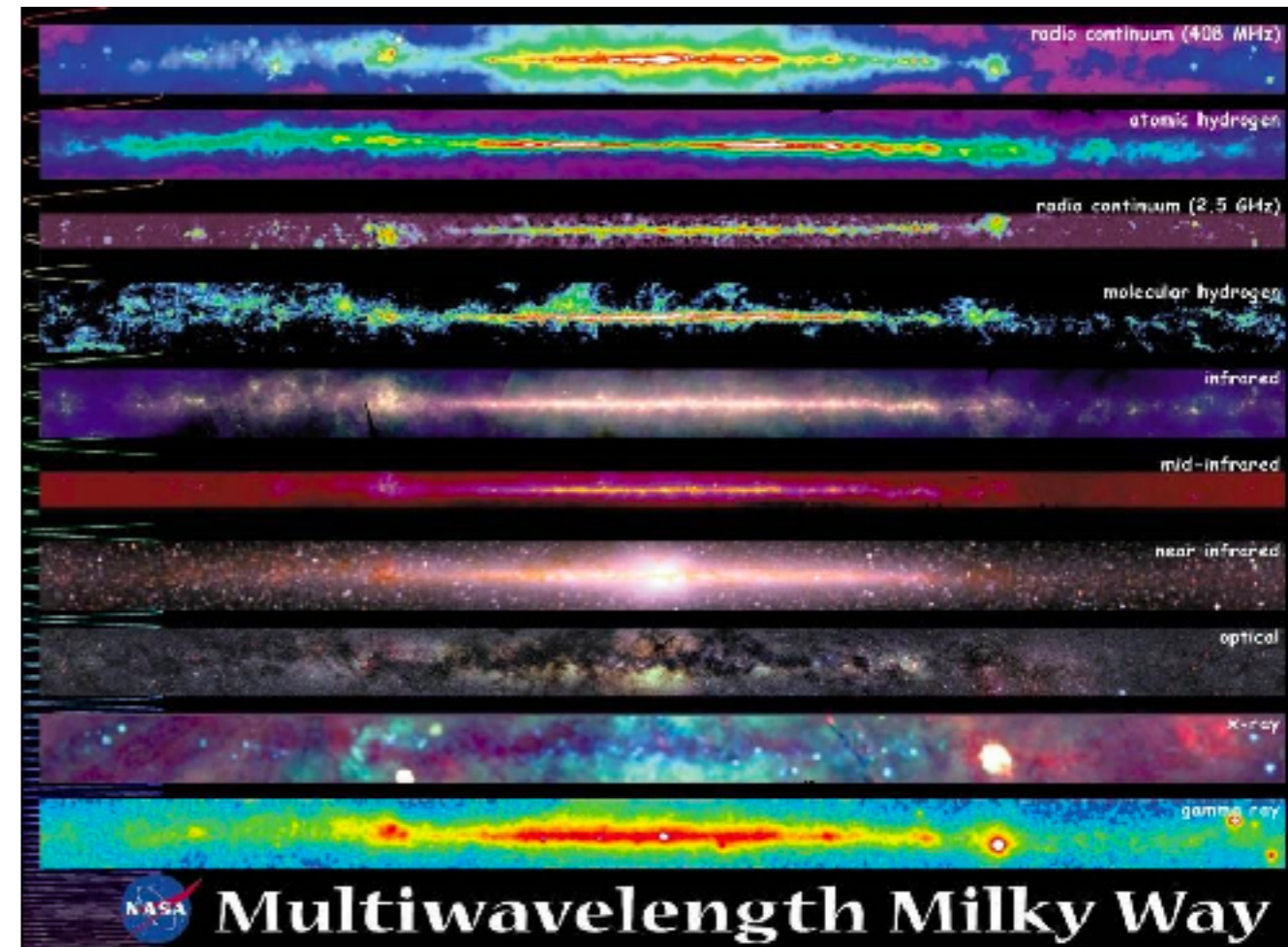
▸ **Add timing information**
valuable because of pulsations or AGN flaring signatures.

WHAT DATA TO USE FOR THE CHALLENGE ?

▸ Start easy and make the problem more difficult and realistic step-by-step

▸ **Stage 1 breakdown: γ-rays only**

    ▸ Start with 2 very distinguishable known sources, see if we can cluster them using an unsupervised method

    ▸ Add a fictive source class that is different than those 2, see if the third is clustered separately

    ▸ Move on to all known source classes except the unassociated one, add the fictive source class

    ▸ Move on to all source classes with the unassociated one, remove the fictive source class (this would hopefully give an interesting result)

WHAT DATA TO USE FOR THE CHALLENGE ?

▸ Start easy and make the problem more difficult and realistic step-by-step

▸ **Stage 2 breakdown: multi-wavelength**

    ▸ Follow same steps as stage 1, but add data from other wavelengths

    ▸ Find correlations between the (x, y, frequency)-space

    ▸ Hopefully a good deep network like approach can learn how to deal with the different kind of uncertainties

WHAT DATA TO USE FOR THE CHALLENGE ?

▸ Start easy and make the problem more difficult and realistic step-by-step

▸ **Stage 3 breakdown: timing information**

  ▸ Follow same steps as stage 1, but add data from other wavelengths **and** timing information

  ▸ This adds another dimension on the data which could be very valuable, for example for LSTM-like architectures

  ▸ Probably very hard, see how far we can go

LSTM Recurrent Neural Net

— Observed
— Prediction for Train Set
— Prediction for Test Set
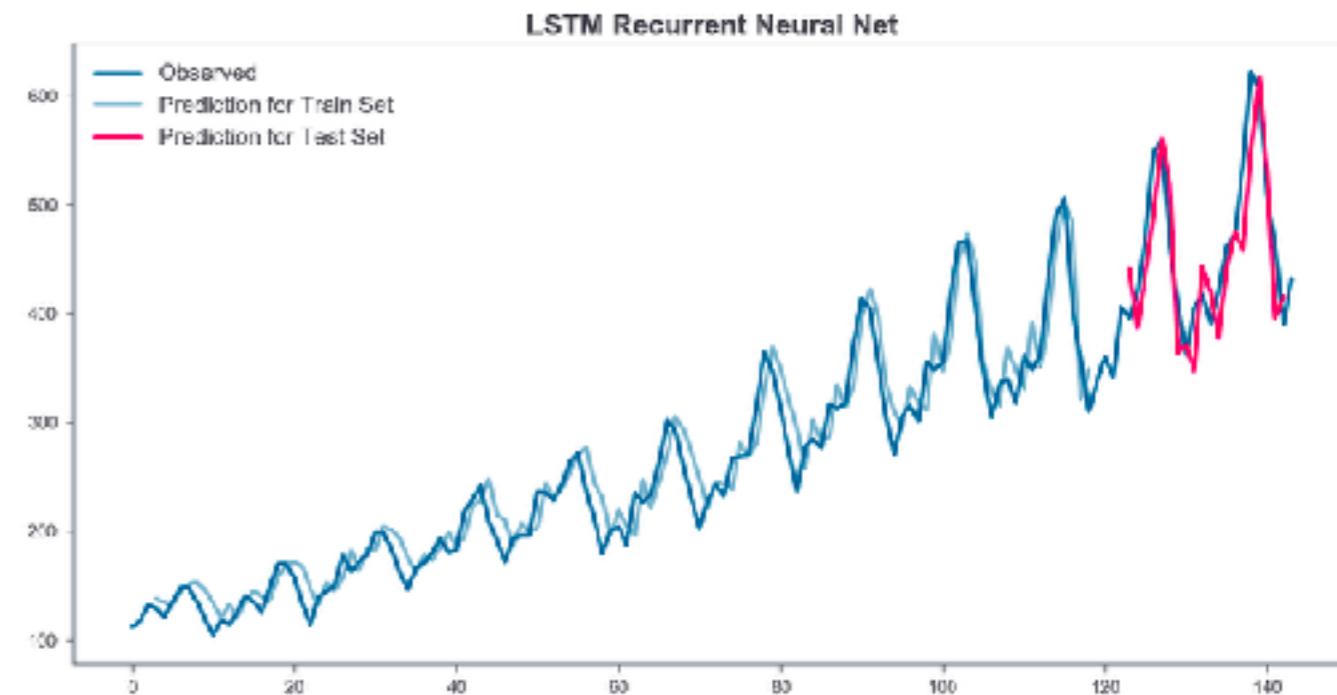
TIMELINE

▸ **Timeline**

  ▸ September: first meetings to decide data generation and technical details

  ▸ October/November: data generation & testing ML methods

      ▸ Iterative work, need quick feedback loops between generating data and testing ML methods

  ▸ December: prepare first results in time for the F2F meeting

▸ **ML methods needed**:

  ▸ Computer vision (ConvNet / U-net / time-series analysis)

  ▸ Unsupervised learning

▸ **Contact people**

  ▸ Gabrijela Zaharijas & Luc Hendriks

▸ Join the Slack channel **#fermi-point-sources**!