

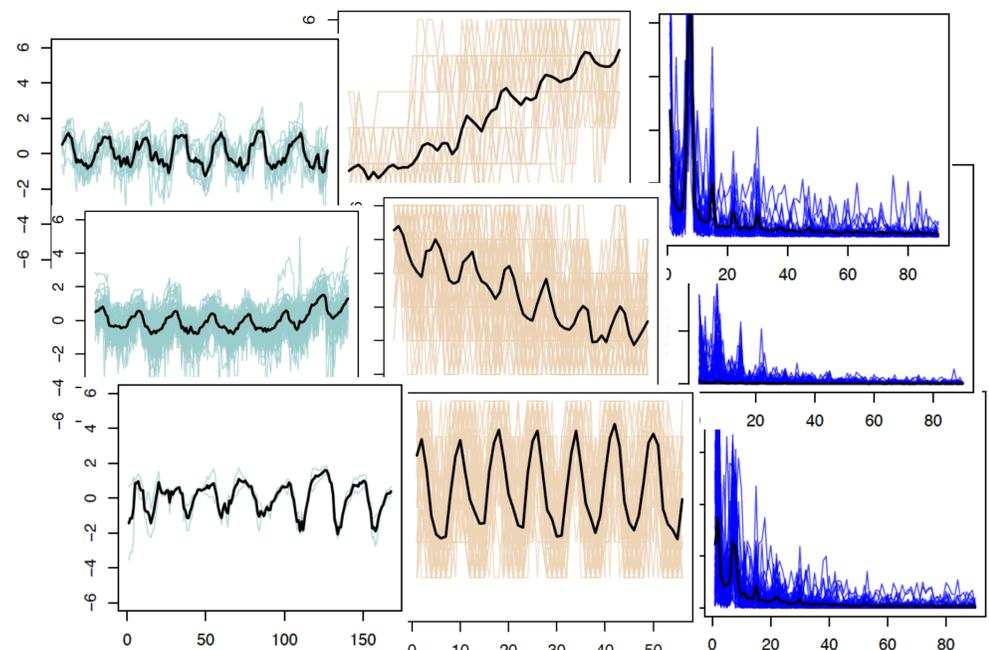
MARKET NEED

- Clustering time series data can be used across many different industries for tasks such as identifying common behaviours, detecting outliers, and discovering groupings based on similar patterns.
- A common challenge in time series clustering is that there are many different clustering algorithms and data representations that all produce different 'valid' cluster assignments. It is often hard to determine which clustering approach is best for a particular objective.
- Running multiple clustering techniques over the same time series data and being able to compare their outputs in an intuitive way is not directly supported by many tools currently on the market.

- Among the clustering algorithms and representation methods, K-means clustering with raw data produces a grouping effect of samples based on mean value.
- Clustering with normalised data groups similar patterns of temporal variations .
- SAX and DTW transformations are robust against noise and temporal distortions, respectively.
- The FFT transformation reduces dimensionality and is good at locating outlier cases.
- The variety of the representation methods in the system achieves a diverse range of possible clustering goals.

TECHNOLOGY SOLUTION

- In order to address end users objectives for time series clustering, we have developed a software tool that helps determine which clustering approach (algorithm and data representation) is most useful.
- The software system allows the user to browse and inspect the output of various clustering approaches simultaneously.
- We apply a selection of state-of-the-art data representation approaches and clustering algorithms to reduce sample dimensionality and achieve different clustering objectives.
- For K-means clustering, the software system allows users to compare different values for K (number of clusters).
- A novel interface has been designed to visualize and explore the clustering output.



FUTURE DEVELOPMENTS

- The TimeSeriesClustering project is currently designed around DSL data from the telecommunications industry, but it can be extended over many industries and applications including financial data, weather recordings, environmental measurements, and human-derived behaviour such as temporal click-stream data.

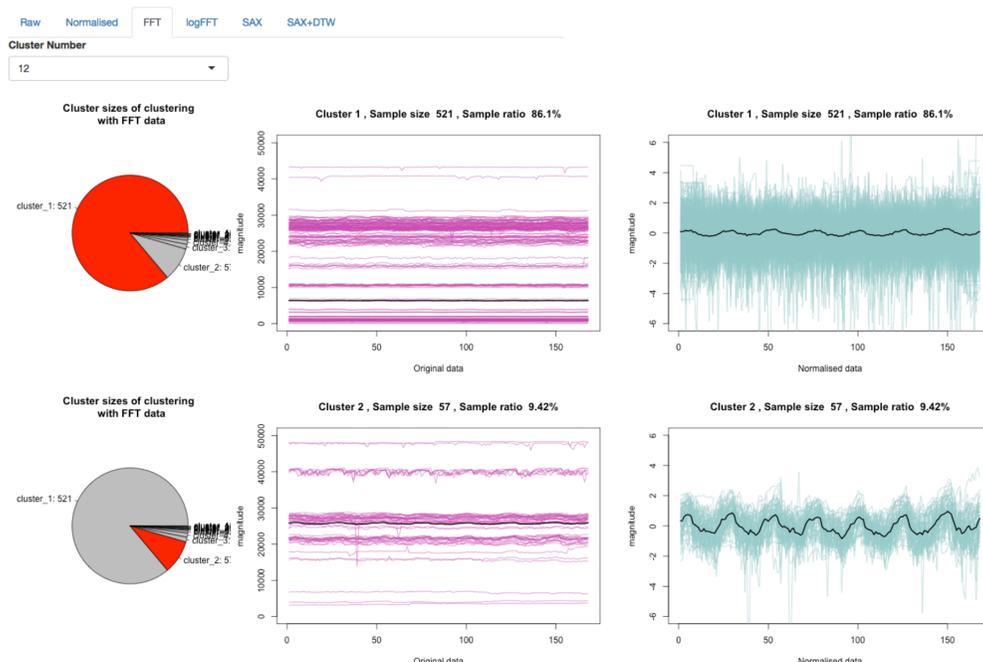
RESEARCH TEAM

Dr Jing Su, UCD

Dr Oisin Boydell, UCD

Hodei Iraola, UCD

Dr Brian Mac Namee, UCD



Follow **CeADAR** Ireland:



Data Analytics

AN ENTERPRISE IRELAND & IDA IRELAND INITIATIVE