Orbit Classification: A Graph-Based Approach

Abraham Bagherjeiran: University of Houston, Chandrika Kamath: Lawrence Livermore National Laboratory
http://www.llnl.gov/casc/sapphire/

Introduction

The National Compact Stellarator Experiment (NCSX) is designed to help physicists investigate fusion power reactors. Key to this investigation is the analysis and interpretation of data. This poster describes the application of machine learning and data mining algorithms to automate this process so that physicists can concentrate on physics instead of data analysis.

In the analysis step, we identify interesting components from simulated and experimental data from the reactor. For simulated data, we evaluate several algorithms that classify individual orbits. Individual orbits are unavailable in experimental data, so our goal here is to first extract and then identify the interesting components.

In the interpretation step, we will use the orbit classes and components to find simulations that best describe the experimental data. This will help physicists understand their experiments and develop new ones.

The orbit in the background of this poster is a visualization of the output of the KAM classifier on a separatrix orbit from the Henon map with parameter α = 1.3284305 and initial points x = 0.5551 and y = 0.1774.

Light red vertices lie along the longest branch rooted at the diameter.
Yellow vertices are hyperbolic branch nodes for each lobe.
Dark red vertices lie along the diameter of the graph.
Edges form the Minimal Spanning Tree of the orbit.

Single-Orbit

In our approach, we extract features from the minimal spanning tree of the points of an orbit. We implemented the KAM classifier from Kenneth Man-Kam Yip's book *KAM: A System for Intelligently Guiding Numerical Experimentation by Computer* and compared it to several standard classifiers.

Classifiers

- Default KAM
  - Static rules for orbits
  - Thresholds in the book
- Custom KAM
  - Genetic algorithm for customized thresholds
- Standard Classifiers
  - Decision Tree
  - Support Vector Machine
  - Nearest Neighbor

KAM Rules

- Quasi-Periodic
  - Round, No Branches
- Island Chain
  - Islands are Clusters
- Separatrix
  - Hyperbolic Branches

Classifier Features

- Diameter Endpoint Distance
- Branch, Edge Statistics
- Cluster Identifications

Dataset

An orbit is the set of points at which a single particle intersects a sensor element over time. As shown in Figure 2, each orbit belongs to one of three classes: quasi-periodic, island chain, and separatrix. Physicists are interested in the components of an orbit such as individual islands in an island chain and the cross points of a separatrix.

We used 6 simulations of the CDX tokamak at Princeton where each simulation contains 100 orbits. Each orbit contains from 50 to 3000 points. The experimental data are expected to contain approximately 100 points per orbit.

Figure 2: An orbit from each of the three classes. From left to right:
Quasi-periodic orbits are round, thin, and closed.
Island chains have distinct quasi-periodic islands.
Separatrices have lobes separated by cross points.

Multi-Orbit

In multi-orbit data, orbits appear together rather than separated. This makes single-orbit classifiers inapplicable. We can view each point as belonging to an orbit of a given class; thus, we classify points rather than orbits. We extract point-level features from the neighborhood of a point and apply a classification algorithm.

After classifying the points, we identify components of a plot as dense regions of points that belong to the same class. Using a density-based clustering algorithm, we can correctly identify most of these components.

Point-Level Features

- Nearest Neighbor Distance
- Density Statistics
- Shape Context

Single-Orbit Results

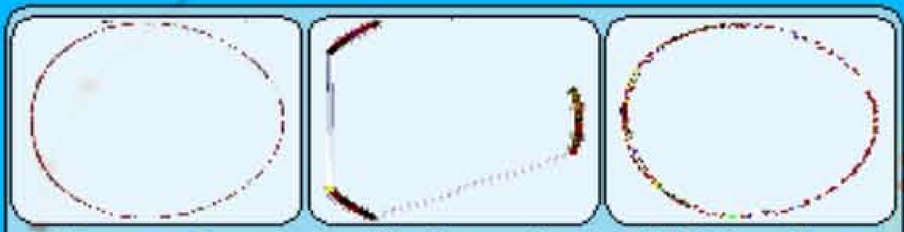

Figure 4: Correct classifications using the KAM classifier.

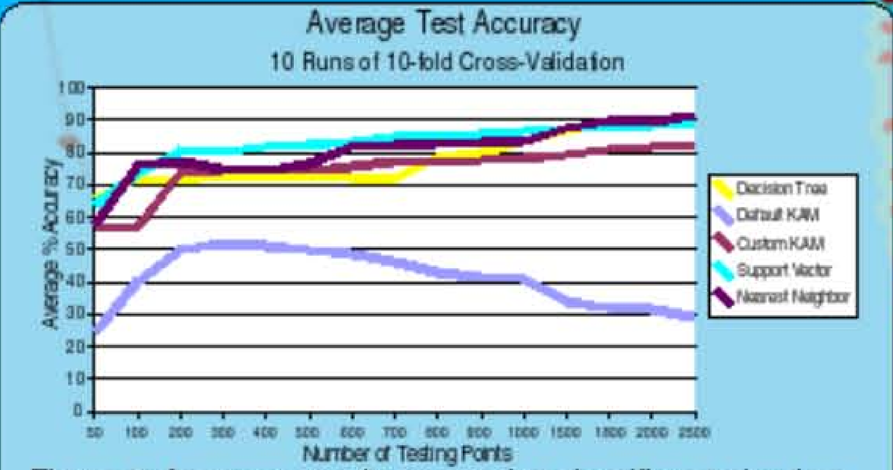

Figure 3: Accuracy results comparing classifiers trained on 2500 points and tested on increasing numbers of points.

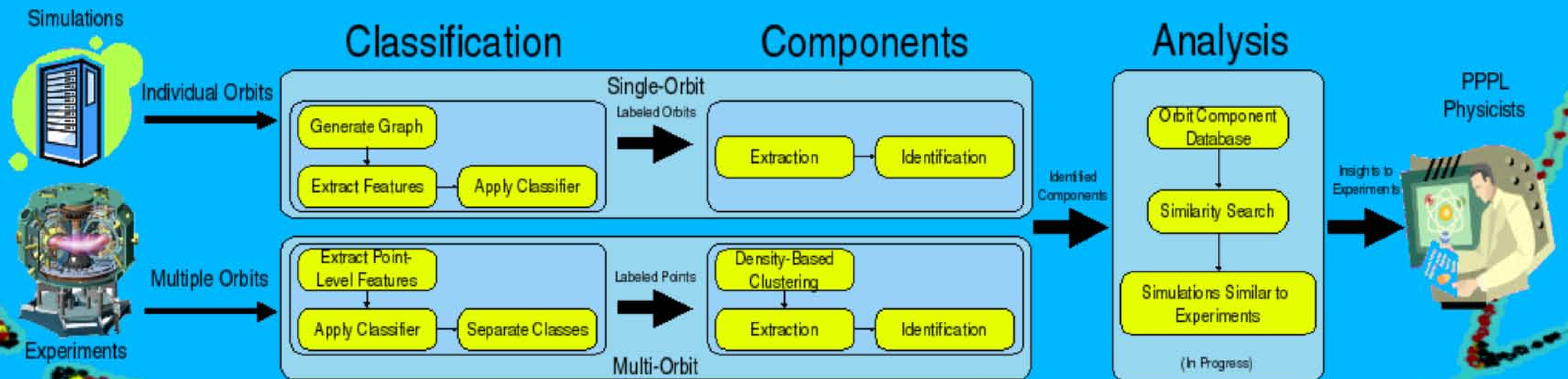

Figure 1: The orbit analysis and interpretation pipeline.

Multi-Orbit Results

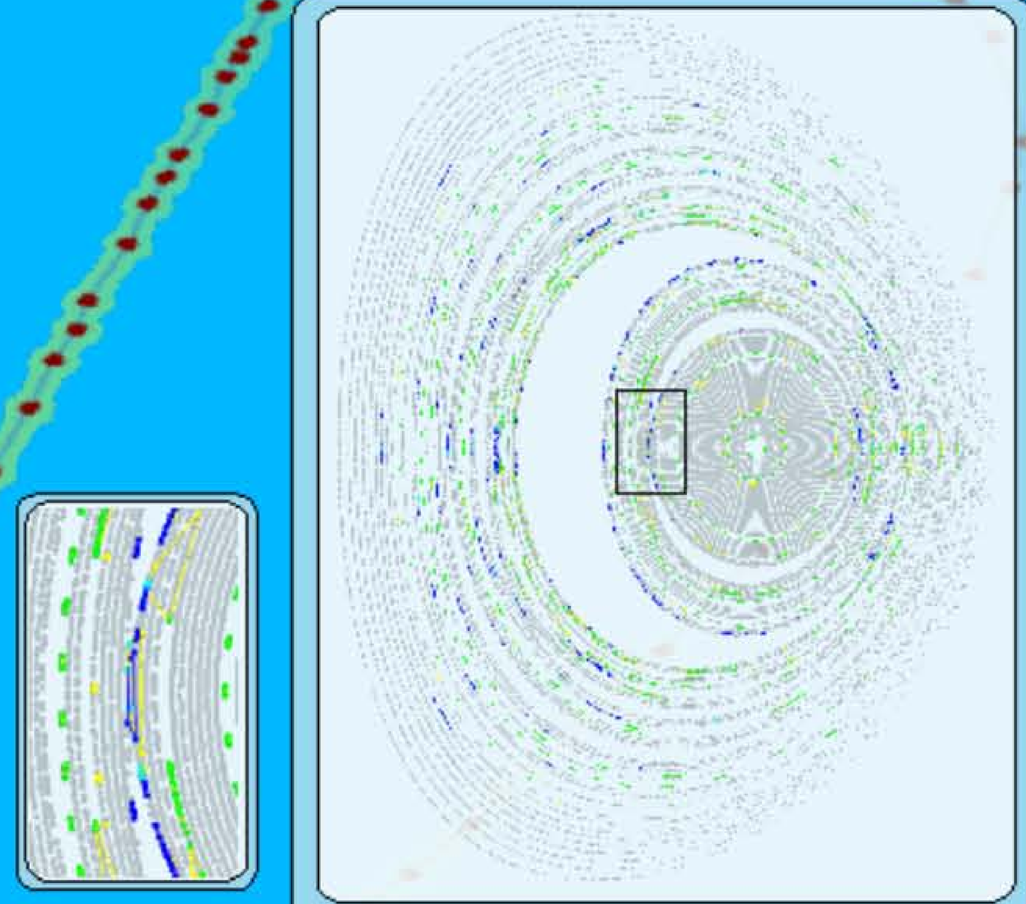

Figure 5: Identification of island borders and separatrix cross points in a multi-orbit plot.

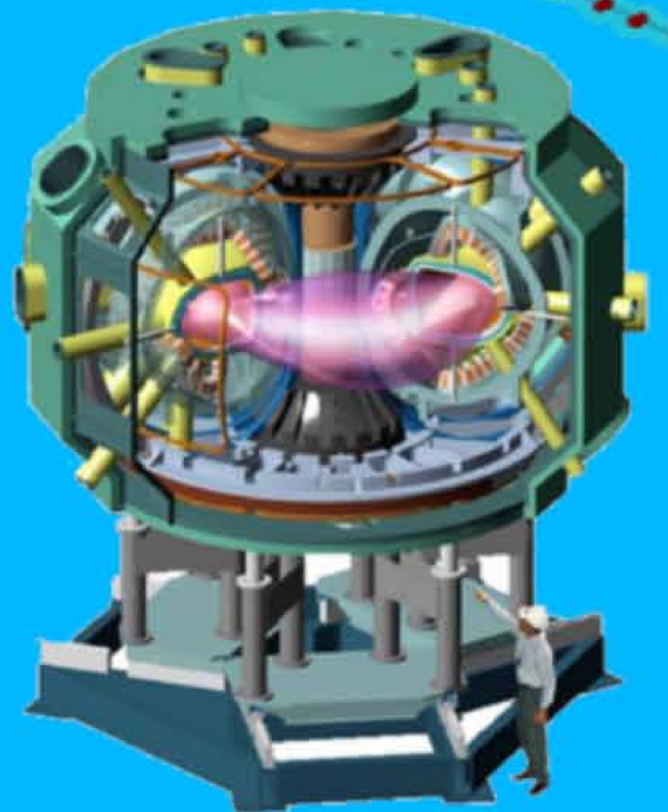

Figure 6: Schematic of the National Compact Stellarator Experiment at the Princeton Plasma Physics Laboratory (PPPL). The software described in this poster is expected be deployed at PPPL in 2005.

Conclusion

This poster describes preliminary work in applying machine learning and data mining algorithms to automate the analysis and interpretation of particle orbits generated from simulations of and experiments in the CDX tokamak and NCSX stellarator.

In single-orbit classification, we investigated a graph-based approach that extracts features from minimal spanning tree of the points of an orbit. We implemented the KAM classifier and compared it to several machine learning algorithms. The results show that the customized KAM classifier and the other machine learning algorithms have high accuracy and appear to be robust to small numbers of points.

In multi-orbit classification, the goal is to identify interesting components of the plot rather than labeling orbits. We used the classification of individual points as a preprocessing step to a density-based clustering algorithm which correctly identified many of the components.

This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

UCRL-POST-214076