Automated Classification of X-ray Sources in Stellar Clusters

Susan Hojnacki

May 11, 2004
Outline

- Introduction
- Research Problem
- Objective
- Approach
- Preliminary Algorithm and Results
- Summary
Introduction
Sources of X-ray Emission

Gigantic explosions, intense magnetic fields, or intense gravitational fields energize particles to high temperatures

- The Sun
- Variable stars
- Pulsars
- Supernovae
- Matter entering a black hole
- Young stars
Chandra X-ray Observatory

X-rays are absorbed by the Earth’s atmosphere...

(Illustration: Chandra Proposers’ Observatory Guide)
Chandra Orbit

- Orbit shown from above
- More than one third of the way to the moon
- Radiation belts

(Illustration: CXC/M.Weiss)
Chandra Deep Field Image
Orion Nebula Cluster

~1000 young stars, most of them X-ray sources

~9% not associated with stars in Orion (no optical or infrared counterpart)

(Image: Garmire et al. 2000)
Trapezium Region of ONC

Hubble Space Telescope Image with Chandra X-ray Contours Overlaid

(Image: Schulz et al. 2001)
Detected X-ray photons are called “events”. For each photon event, onboard software records:

- 2D spatial location
- energy, range = 0.2 to 10 keV
- arrival time
Chandra Deep Field Image
Orion Nebula Cluster

Energy Spectrum
X-ray Light curve

(Image: Garmire et al. 2000)
Outline

√ Introduction

• Research Problem

• Objective

• Approach

• Preliminary Algorithm and Results

• Summary
Research Problem

Enormous amount of data from Chandra:
6 CCD detector arrays x 1024 x 1024 = 6,291,456 pixels
~3.2 s exposure time

→ 9000 frames generated in 8 hours of observing!
Research Problem

- Currently astronomers analyze one X-ray source at a time fitting the data to a theoretical model
  - Manual process
  - Time-consuming
  - Subjective

- Some recent attempts at automation
  - Need good physical model a priori for all possible X-ray sources
  - Validity of statistical model?
Outline

√ Introduction
√ Research Problem

- Objective
- Approach
- Preliminary Algorithm and Results
- Summary
Objective

To develop an automated classification algorithm to group pre-main sequence X-ray sources into clusters based on spectral and temporal attributes.

Goal is a method that is objective, model-independent and doesn’t require parameter estimation or a priori assumptions as to the nature of the source.

Purpose is to allow automated exploration to find “interesting” objects or clusters of objects for further study.
Objective

Use techniques from the following fields for development and verification of the algorithm:

- Multivariate Statistics
- Remote Sensing
- Pattern Recognition
Outline

✓ Introduction
✓ Research Problem
✓ Objective
  • Approach
  • Preliminary Algorithm and Results
  • Summary
Approach

- Obtained Chandra X-ray data sets and performed pre-processing and source detection
- Selected multivariate features that define X-ray sources
- Developed a classification algorithm
- Tested algorithm on Chandra X-ray data sets
- Comparing results to optical & infrared data
Outline

√ Introduction
√ Research Problem
√ Objective
√ Approach

• Preliminary Algorithm and Results
• Summary
Preliminary Algorithm

- Data set consists of 261 sources in the Orion Nebula Cluster
- X-ray spectrum divided into 42 bands, using the following information:
  - high-resolution emission line data
  - spectral resolution of the X-ray CCD detector arrays at -120 deg C (nominal operating temp of Chandra)
  - quantum efficiency of the CCDs
Preliminary Algorithm

- Multivariate techniques used:
  - Principal Components Analysis
  - Hierarchical Clustering
  - K-means Clustering

- All are "unsupervised" methods

- None require multivariate normal data

- Choice of representative number of classes is heuristic
Principal Components Analysis

- Goal is to identify a new, smaller set of uncorrelated variables, called the *principal components*, which explains all (or nearly all) of the total variance in the data set.
- Each principal component is described by:
  - eigenvector: linear combination of the original variables (X-ray bands)
  - eigenvalue: represents the variance accounted for by that PC
- Retained first 5 principal components based on analysis of several PCA stopping rules.
Agglomerative Hierarchical Clustering

- Attempts to find natural groupings of the detected X-ray sources
- Partitions the set of X-ray sources into relatively homogeneous subsets based on inter-source similarities
- Started with 261 clusters and successively merged them based on statistical similarity measure
- Examined similarity level at each merger step to determine the final number of clusters
K-means Clustering

- Hierarchical clustering cannot transfer a source from one cluster to another if initially grouped incorrectly: K-means used for “fine-tuning”
- K-Means goal: arrive at clusters with small within-cluster variation and large between-cluster variation
- Start with cluster assignments from hierarchical clustering for initial partition of X-ray sources
- An X-ray source’s cluster membership was changed if there was a cluster with closer centroid → iterative process
CLASS 2

Source 40 Spectrum

Source 83 Spectrum

Source 97 Spectrum

Source 114 Spectrum
CLASS 6

Source 88 Spectrum

Source 129 Spectrum

Source 155 Spectrum

Source 192 Spectrum
CLASS 7

Source 184 Spectrum

Source 236 Spectrum

Source 241 Spectrum

Source 267 Spectrum
Outline

√ Introduction
√ Research Problem
√ Objective
√ Approach
√ Preliminary Algorithm and Results

• Summary
Summary

- Preliminary algorithm results show promise for development of a model independent, unsupervised method that can be used to group X-ray sources with similar spectra into classes.

- No a priori knowledge of the nature of each source was used to accomplish the clustering.
Additional Work

- Determine if a temporal variable can improve the clustering results
- Investigate other techniques from the field of Remote Sensing
- Investigate clustering algorithms other than Hierarchical and K-means
- Continue to compare results to known sources in Orion for optical and infrared data
- Use Chandra Orion Ultradeep Project data
Chandra X-ray image of the Orion Nebula Cluster
(10-day integration; Feigelson et al. 2004)
QUESTIONS ?