Associating earth-orbiting objects detected by astronomical telescopes

Haseena Ahmed   Prince Chidyagwai   Kun Gou   Yun Liu
Timur Milgrom   Vincent Quenneville-Bélair

Mentor: Dr. Gary B. Green (The Aerospace Corporation)

August 17, 2007

Problem Statement

Satellites make streaks in telescope images

- Input:
  1. Streak data
  2. Orbit data
- Objective: Identify streaks made by the same object
- Process:
  Take the image and find the streak (astronomers)
  Estimate the orbit of the object (orbit analysts)
  Cluster streaks (large cardinality problem - our task)

Streak Modeling

Streaks can be modeled in two spaces:

- Image space: A vector in $\mathbb{R}^3$ as a result of processing streak points

$$\{RA_i, DE_i, t_i\}_{i=1}^{\#\text{of points in a streak}} \rightarrow \{RA, DE, \alpha\}$$

- Orbit space: A vector in $\mathbb{R}^6$ as a result of orbit estimation

$$\{RA_i, DE_i, t_i\}_{i=1}^{\#\text{of points in a streak}} \rightarrow \{a, e, i, \Omega, \omega_p, M\}$$

Clustering

- Similarity and dissimilarity measures
- Depends mainly on the data set available

Clustering

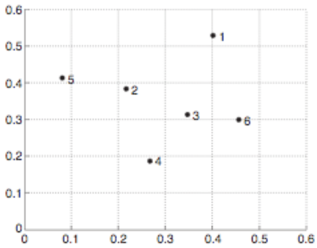
- Similarity and dissimilarity measures
- Depends mainly on the data set available
- Two commonly used methods of clustering
  - Hierarchical clustering
    - Tree structure
    - Agglomerative

Clustering

- Similarity and dissimilarity measures
- Depends mainly on the data set available
- Two commonly used methods of clustering
  - Hierarchical clustering
    - Tree structure
    - Agglomerative
  - Partitional clustering
    - One level partitioning
    - k-means

Agglomerative Hierarchical Clustering

Given a set of points to be clustered in $2D$ as in the figure



- We need to specify: distance measure, type of linkage

Agglomerative Hierarchical Clustering

Compute the proximity matrix as in table

	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

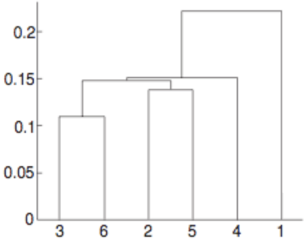
Agglomerative Hierarchical Clustering

- Cluster points 3 and 6
- Obtain new proximity matrix by calculating the distance between the new cluster $\{3, 6\}$ and other points

$$\begin{aligned} dist(\{3,6\}\ ,\ \{1\}) &= \min\left(dist(3,1), dist(6,1)\right) \\ &= \min\left(0.22, 0.23\right) = 0.22 \end{aligned}$$

Agglomerative Hierarchical Clustering

Dendogram representation can be given by figure



k-means Clustering

Algorithm:

- Select k points as initial centroids
- **repeat**

  Form k clusters by assigning each point to closest centroid.

  Recompute the centroid of each cluster.

  until Centroids do not change.

Comparision - Agglomerative vs k-means

- Agglomerative
  - Complexity is $O(n^2)$ in memory and $O(n^2 \log n)$ in CPU time
  - Local optimal clustering
  - All merges are final

Comparision - Agglomerative vs k-means

- Agglomerative
  - Complexity is $O(n^2)$ in memory and $O(n^2 \log n)$ in CPU time
  - Local optimal clustering
  - All merges are final
- k-means
  - Complexity is $O(n)$ in memory space and CPU time
  - Number of clusters k needs to be known *a-priori*
  - Initialization of centers of clusters
  - Local optimal clustering

Implementation

- Representations in orbit space
  - Kepler (Orbit space)
  - Equinoctial elements
  - Cartesian ellipse
- MATLAB
  - Linkage
  - Distance function

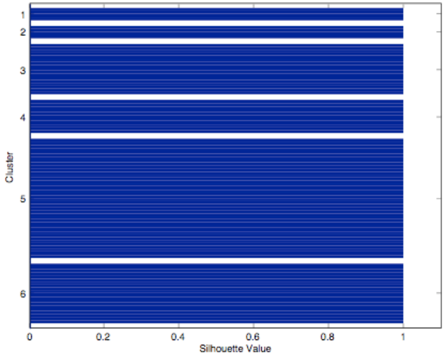
Time comparision

Satellites	Streaks	Kepler Time	Ellipse Time
6	96	.05	.06
32	861	3.85	4.48
74	2191	56.45	61.7
137	4086	423.13	443.17

Table: Computational time (seconds)

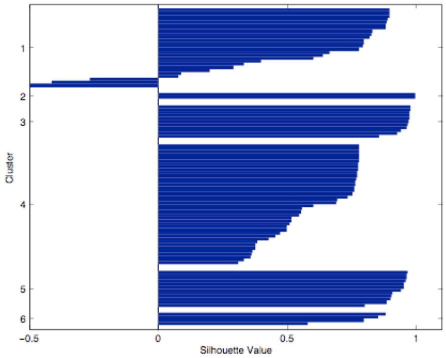
Silhouette

For unperturbed data



Silhouette

For perturbed data



Distance Measure Comparison

Satellites	Euclidean	Weighted	Cosine
6	63	7	644
36	612	99	617
74	1563	273	1537
137	3107	764	3098

Table: Performance of norms (# clusters)

Linkage Function Comparison

Satellites	Single	Average	Centroid
6	7	13	13
36	99	86	82
74	273	260	240
137	764	520	472

Table: Performance of linkage (# clusters)

Effect of Variation in Cut-off

Satellites	Found	Cut-off	Silhouette
6	6	1.154	0.70
36	32	1.1546	0.70
36	33	1.1547	0.79
74	57	1.1546331	0.48
137	133	1.1546	0.47

Table: Effect of cut-off on silhouette (a, e weighted with 0.1)

Large data clustering

Sectioning method tested on

- Number of streaks = 4400
- Actual number of satellites = 137

Sections	1	2	4	8
Time	356	143	56	12
Found	137	116	126	143

Table: Effective grouping

Conclusions

- Weighted norm is effective
- Linkage function is inconclusive
- Cut-off is sensitive
- Sectional method is promising

Future Work

Improving clustering

- Develop theory for choosing weights
- Develop theory for choosing cutoff

Improving sectioning method

- Optimal grouping

Questions?