Distance Constrained Clustering Without A Priori Knowledge of the Number of Clusters

Prakruthi Reddy, UNC Charlotte
Saurav Agarwal, Dr Srinivas Akella, Computer Science, UNC Charlotte

UNC CHARLOTTE

Introduction

Clustering is the task of grouping together a set of data points based on their similarity.

There are several applications of clustering in robotics; our research focuses on the line coverage problem [1], which is the coverage of linear environment features (e.g., road networks, power lines), by one or more robots, while respecting all the resource constraints (e.g., battery capacity, flight time) on each robot. The line coverage problem requires a clustering technique to partition a large graph (e.g., UNC Charlotte's campus road network) into small subgraphs by clustering the edges of the graph, to facilitate a solution that will satisfy the resource constraints on the robots.

Initially, a k-medoid clustering approach was used to perform the clustering task. However, the limitation of k-medoids clustering to pre-specify the number of clusters, and the addition of distance constraints to the clustering problem, made for reasons to look for alternative clustering techniques.

Objectives

- Perform a literature survey of relevant clustering algorithms for problems with constraints, specifically those that would allow us to implement distance constraints between the cluster center and its farthest edge and those that do not require a priori knowledge of the number of clusters.
- Adapt and implement the algorithms that are most suitable for our purpose.
- Incorporate the distance constraints in the algorithms.
- Analyze the performance of the algorithms.

Algorithm

Affinity Propagation:

Affinity Propagation [2] is a clustering technique whose input is a set of pairwise similarities {s(i,k)}, which indicates how well the data point with index k is suited to be an exemplar for data point i. It works by exchanging messages between data points and identifies clusters based on maximizing the similarity between the data points in a cluster and their exemplar.

Rather than requiring the pre-specification of the number of clusters, affinity propagation allows us to modify the input preference value s(k,k), which indicates how suitable a data point is to be an exemplar. If all data points are equally likely to be exemplars, their preference can be set to the same value — this value can be varied to produce different numbers of clusters (Fig. 1).

Data points that are farther from their exemplar than the distance constraint allow them to be are seen as unfit to be a part of that exemplar's cluster. We can represent this relationship in the algorithm's input by setting the input similarity $s(i,k) = -\infty$.

Implementation

- Measure of similarity between data points is set to the negative squared Euclidean distance

  $s(i,k) = - ||x_i - x_k||^2$, $(i \neq k)$
- The input preference is set to the minimum value in the similarity matrix, so that we can get a small number of clusters that satisfy the distance constraint.

  $s(k,k) = \min(s(i,k))$, $(i \neq k)$

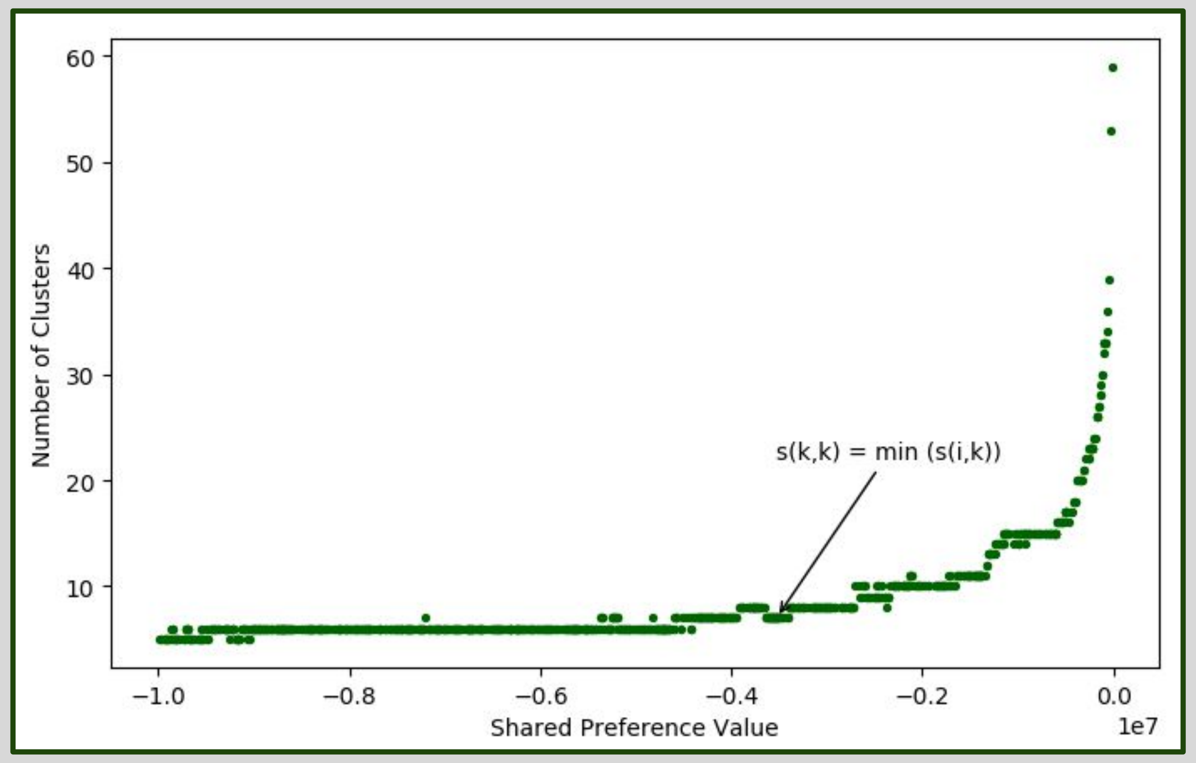

Fig. 1: The effect of preference value on numbers of clusters for the UNCC dataset.

Results

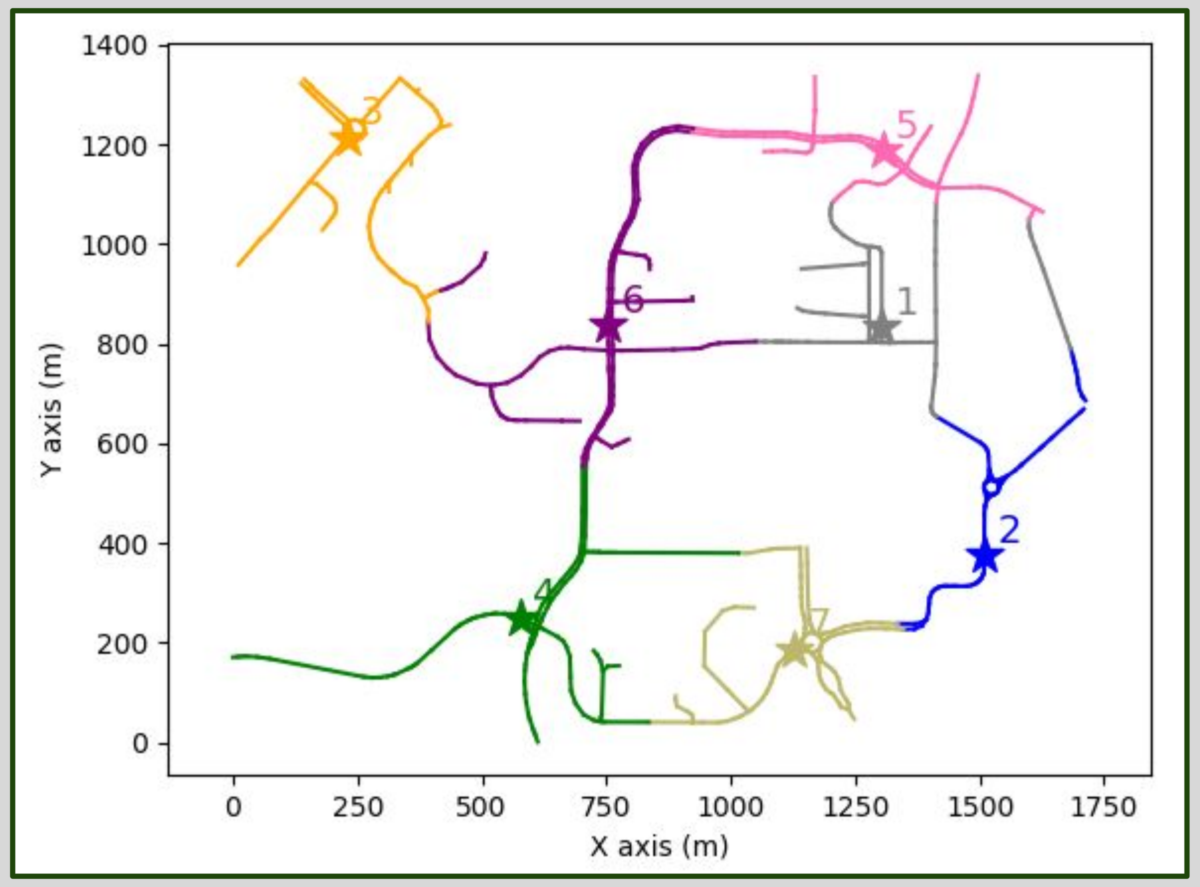

Fig. 2: Clustering on UNCC road network without constraints. Clusters are represented by colors and exemplars by stars.

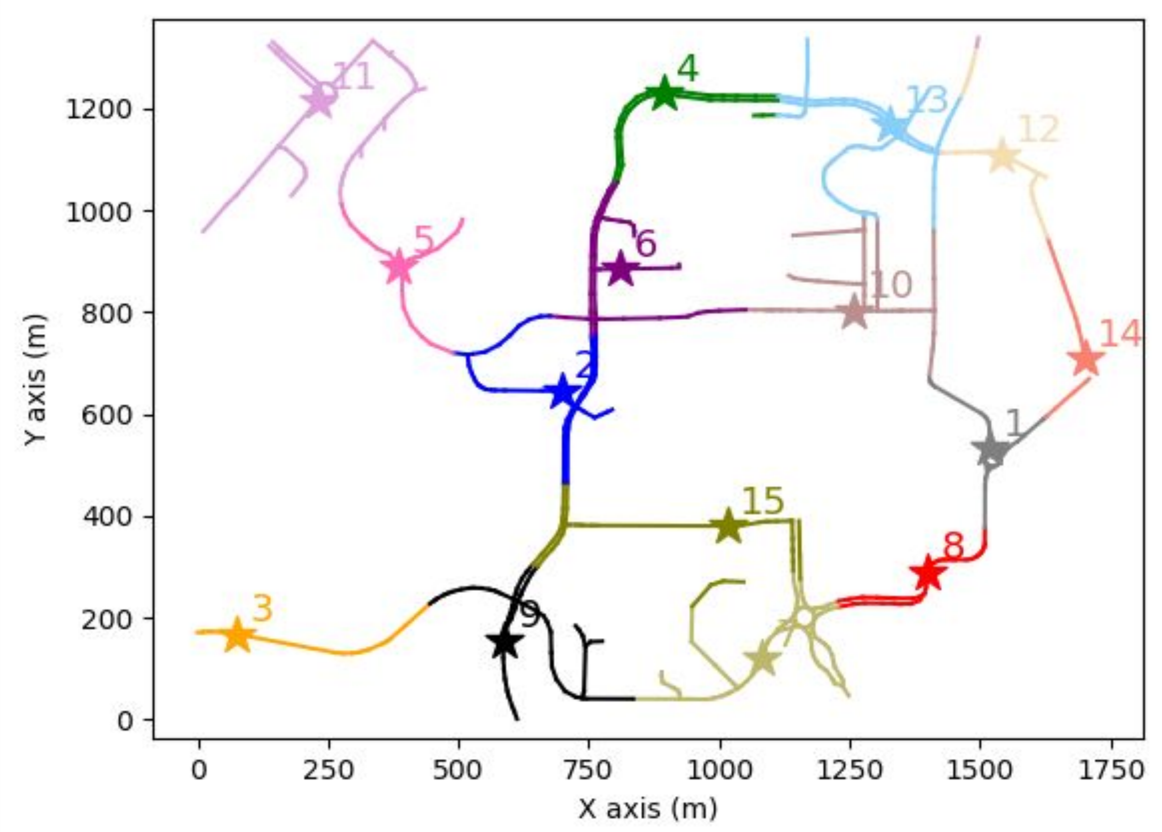

Fig. 3: Clustering on UNCC road network with a distance constraint of 200 m between the exemplar and its farthest data point.

- For the UNCC road network, the algorithm returns seven clusters without any distance constraints (Fig. 2). A few of the clusters, however, are not contiguous. For example, cluster 6 (purple) is separated from a small section of itself by cluster 3 (yellow).
- With a maximum distance constraint of 200 m between an exemplar and the farthest point in its cluster (Fig. 3), the algorithm returns fifteen clusters. Clusters abide by the constraints, although a few clusters lack contiguity.

Conclusions

Using Affinity Propagation, we obtain clusters that abide by our distance constraint (except for a few anomalous cases), without having to pre-specify the number of clusters. However, there are several unexpected issues with the clusters, such as the lack of contiguity within clusters. For our application, it is crucial that the clusters are contiguous — to allow for a smooth travel route for the robots. The existence of non-contiguous clusters renders the clustering less desirable.

Future work would include implementing methods to ensure that clusters are contiguous and abide by the distance constraint. We need to test Affinity Propagation on a large number of networks, each with different characteristics, and develop a metric to verify the effectiveness of the algorithm.

References

[1] S. Agarwal, S. Akella. (2020). Line Coverage with Multiple Robots. *IEEE International Conference on Robotics and Automation.*
[2] B. J. Frey, D. Dueck. (2007). Clustering by Passing Messages Between Data Points. *Science,* 315 (5814), 972–976.