Written Status Report

--- Group3

For our group, the main task of this week is mainly divided into two parts: find out clustering algorithm with center at landfills to help us simply the VRP; using the pre-exist algorithm in OR-Tool to solve the VRP. For a VRP, cluster-first, route-second algorithms are very common.

As for the clustering algorithm, we mainly consider two kinds of clustering methods based on Euclidean distance: Fuzzy C-Means (FCM) Clustering and Hierarchical clustering. The main challenge for us is to figure out the transition points between different clusters, which will be helpful to transfer the drivers between different landfill districts.

Kaichao is mainly responsible for finding the fuzzy clustering algorithm and he has found the python code which can set the center of different clusters and print out the points inside each cluster. However, along this process, there is still a loss of accuracy that the real center is not exactly the center we set, but as we concerned this kind of loss is not too big. Another thing which is related to the usage of fuzzy clustering is that we need have the two-dimensional coordinates, here we may use the multidimensional scaling to translate the distance matrix into this kind of coordinates for us to input into the fuzzy clustering algorithm. However, it is remaining unsolved that how we use the fuzzy clustering algorithm to find out the points which can belong to two landfill districts.

Anqi mainly deals with the hierarchical clustering. Hierarchical clustering builds clusters based on a measure of similarity, which is Euclidean distance for our group. There are two types of hierarchical clustering: agglomerative and divisive. We are considering using the agglomerative one. It is a clustering method that each observation starts in its cluster, and pairs of clusters are merged as one moves up the hierarchy. We build a tree structure first and then cut-off the tree structure into different clusters when some cut-off measurement is reached. As for the linkage criterion, which is ways to determine the distance between sets, we are now trying different methods. One advantage of it is that its main output, the dendrogram, makes it easier to decide on the number of clusters by looking at the dendrogram. One disadvantage is that it works poorly for a large dataset since it’s kind of a greedy method. Another disadvantage is that it’s not possible to undo the previous step and initial seeds have a strong impact on the final results.

Serena and Sarah oversee the code implement. We have found two algorithms to solve the VRP in OR-Tools and we are trying to understand the mechanism of different statements, then we can base on this code to adjust it to the code that can better solve our problem.

Another thing we have discussed in class is that the way we set the distance between different pickups and delivery locations. The distance between P1 to P2 should be the distance between P1 to the landfill plus the distance between P2 to landfill. The distance between D1 and D2 should be the same as the P scenario. However, the major problem here is that although we can set the distance from D to P as the real distance between these two points, the distance from P to D should be different for that we have to go to the landfill in this case. However, the distance matrix does not allow us to set different distances between the two same points so that we are trying to solve the problem.

For next week, because we have the feasible clustering algorithm, we will start implementing the code found in OR-Tool.