SPL-1 Project Report, 2018

Uncapacitated Facility Location Problem Solver

Course : SE 305 (Software Project Lab I)

Submitted by

Md.Saleh Ibtasham

BSSE Roll No. : 0912

BSSE Session: 2016-2017

Supervised by

Shah Mostafa Khaled

Associate Professor

Institute of Information Technology

Institute of Information Technology

University of Dhaka

[30-05-2018]

Table of Contents

1.Introduction

The facility location problem is also known as location analysis which is a branch of operation research and computational geometry. It is a problem where we need to minimize transportation cost while considering other constraints such as the cost of placing a facility, the maximum capacity of the facility or the demand of the corresponding demand point. This project is a facility location problem solver which solves uncapacitated facility location problem.

1.1 Background Study

a. A facility location problem is solved by finding the optimum placement of a facility in a given data set. A simple facility location problem is on general graphs is NP-hard to solve optimally, by reduction from the set cover problem. A number of approximation algorithms have been developed for facility location problems.

b. A facility location problem can be divided into 2 types of problems: Capacitated and Uncapacitated.

c. The uncapacitated problems are more easy to solve than the capacited facility location problems. Basically a facility location problem is characterized by 4 elements.

- A set of locations where facilities may be built/opened. For every location some information about the cost of building or opening a facility at that location is given.

- A set of demand points (clients) that have to be assigned for service to some facilities. For every client one receives some information regarding its demand and about the costs/profits incurred if he would be served by a certain facility.

- A list of requirements to be met by the open facilities and by any assignment of demand points to facilities.

- A function that associates to each set of facilities the cost/profit incurred if one would open all the facilities in the set and would assign the demand points to them such that the requirements are satisfied.

The goal of this problem is to assign facilities to be opened which will optimize a given function.

d. A cluster is a set of nodes which are close to each other. In a facility location problem a cluster means the group of nodes which are apparently close to each other that can be served by a single facility.

e. A set covering problem is a classical combinatorics problem. Given a set of elements {1,2,...,n} (called the universe) and a collection S of m sets whose union equals the universe, the set cover problem is to identify the smallest sub-collection of S whose union equals the universe. Facility location problems can be converted to set covering problems.

1.2. Challenges

The first obstacle of solving a p median problem was to cluster the nodes. P median problem was the first problem to solve the uncapacitated facility location problem. The p median problem is solved by clustering the demand points and assigning them to a facility.

The facility location problem can be solved iteratively. The iterative process always generates an optimal solution but iterative process cannot be used to solve every kind of dataset. As the set of data increases it becomes more and more difficult to iterate the dataset.

So the challenge lies whether to stop the iteration of the program or not. If the iterations take place the time it will consume will increase exponentially.

The facility location problem can be converted to a set covering problem. It helps the program to iterate faster and more efficiently. But the process of dividing the demand points in a set is a tough task as the set of demand points increases the possible sets increase exponentially.

So the set covering method suggests that we divide the set of demand points in suitable and stable way so that the method will have a fruitful impact on the final result.

The optimal cost generating function needs to update which facilities are located at last. So the function has to have track of the facilities so far used and also at which points a facility needs to be updated as a new facility may improve the cost of the objective functions.

2. Project Overview

The project was divide into sub sections where the subsections are used to improve a given data sets value gradually.

The whole program was divided into 3 sub problems. The problems being: Clustering Approach, Iterative Approach and Set Covering Approach.

Some common functions are used to read data in every approach. And some utility functions to solve some duplicated tasks.

readfile() functions are used to read intput files in the respective approaches.
makemat() functions are used to make matrix of the data read from the files.
Printing assignments of demand points, calculating optimum values of the objective functions, printing the facilities finally chosen uses some similar utility functions.

The facility location problem in this program minimizes 2 objective functions:

1. minimize(z) = $\sum f_i x_{ij}$
   $f_i = \{0,1\}$
   x_{ij} = the distance from facility i to demand point j

2. minimize(z) = $\sum f_i x_i + \sum c_{ij} y_{ij}$
   f_i = the cost of locating the facility i
   $x_i = \{0,1\}$ (whether the facility is located or not)
   c_{ij}= the cost of serving facility i with demand point j
   $y_{ij} = \{0,1\}$ (whether the facility i serves demant point j)

Clustering Approach:

The clustering method uses clustering to group set of demand points. The clustering follows the below procedures:

a. At first the method uses initially selected k clusters. As the clusters are not optimal the centeroid of the cluster needs to be calculated.

b. The initial clusters are then updated with centroids. The nodes in the cluster produce the new cluster centroid.

c. The new cluster centeroid is computed by the arithmetic median of the nodes with respect to their distances.

d. After computing the new cluster centeroid the k centroids are updated with the new ones.

e. After getting new clusters the process from b to d are repeated. The clustering halts when it reaches a suitable amount of iterations. Or getting new clusters with same centroid from the previous iteration.

The clustering approach gives an optimum value to the objective function 1.

Iterative Approach:

In the iterative approach the program uses a matrix. The matrix is used to compute the optimum value of the function 2.

a. The iterative approach firstly ranks the facilities with respect to the demand points its assigned to.

b. The ranking is done by summing the total amount of cost to serve all the demand points it can possibly serve.

c. The facilities are then ranked by their ranking value in increasing order.

d. Then the ranked facilities are picked from lowest ranked cost to highest. Then the list of facilities are divided into sets of combination of facilities.

e. The objective of making combinations is how well the combination of a particular set is favorable to the objective function.

f. There is a group tolerance level as well as a combination tolerance level. After exceeding the combination tolerance a combination of facilities are discarded and iterated to the next one.

g. If the group tolerance level exceeds the highest tolerance level then the group of combination of the facilities is totally discarded.

h. Finishing the iterations the latest optimum value is the result of the iterative approach.

Set Covering Approach:

The covering approach also uses a matrix. The approach tries to find the optimum value of function 2.

a. The approach first ranks the facilities by ranking the mean or median distance or cost of the facility with respect to the demand points.

b. The greater of the two distances is taken as the ranking distance of a facility.

c. The demand points are then assigned to the particular facilities with respect to the rank distances of the respective ones.

d. If the ranking distance is greater than the cost or distance of the demand point then the demand point is assigned to the facility.

e. At last the unallocated demand points are assigned to the facilities which are closest to them.

f. The objective function is then computed from the sets of demand points assigned to the facilities.

g. These sets act as subsets in the set covering problem. The objective is then to find the minimum required sets to fulfill the whole set of demand points.

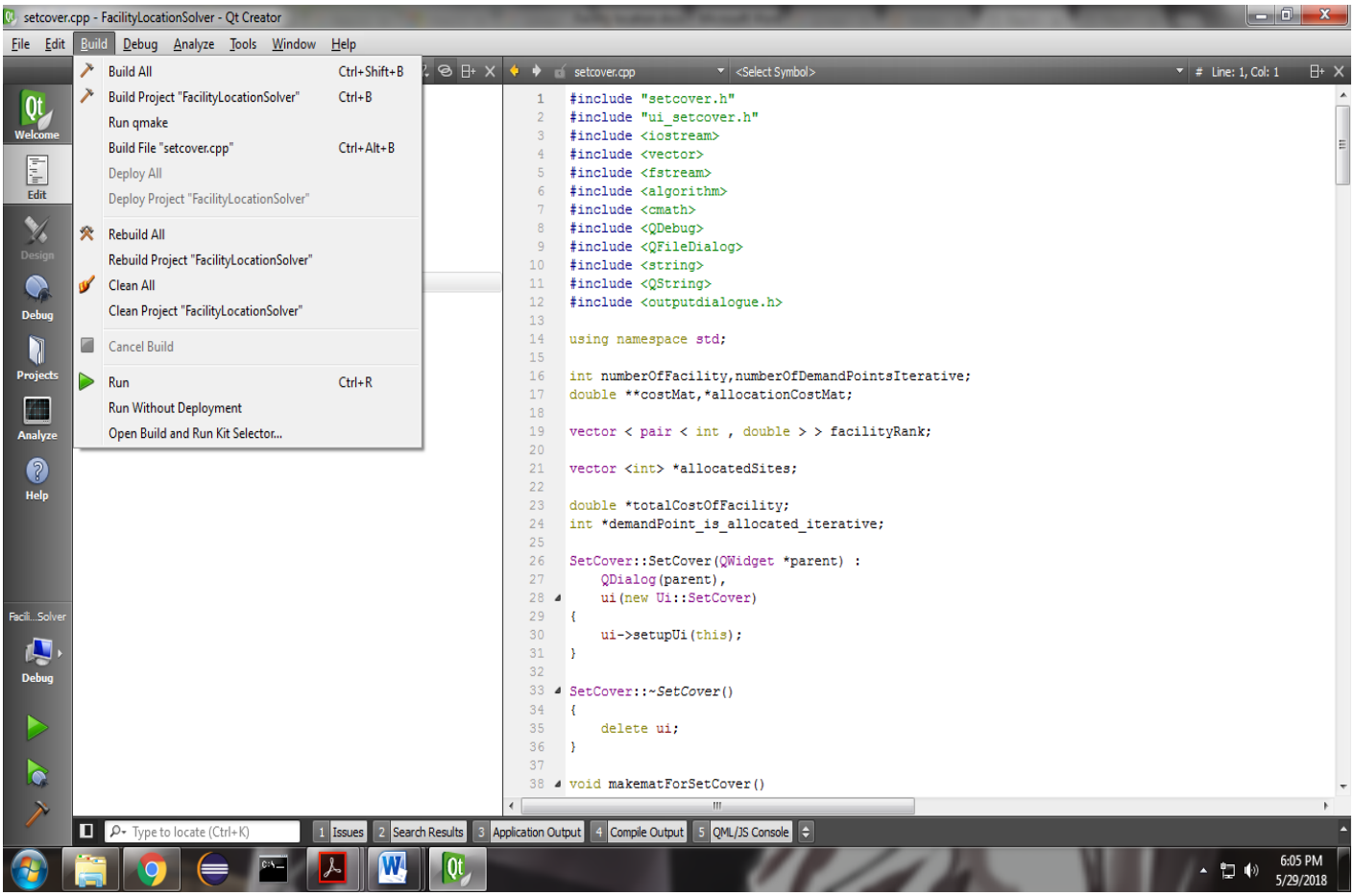
h. After finding the sets the demand points is then again allocated to the facilities which serve the subsets of demand points.

i. Then the facilities are chosen as the final facilities and the cost is computed.

3. User Manual:

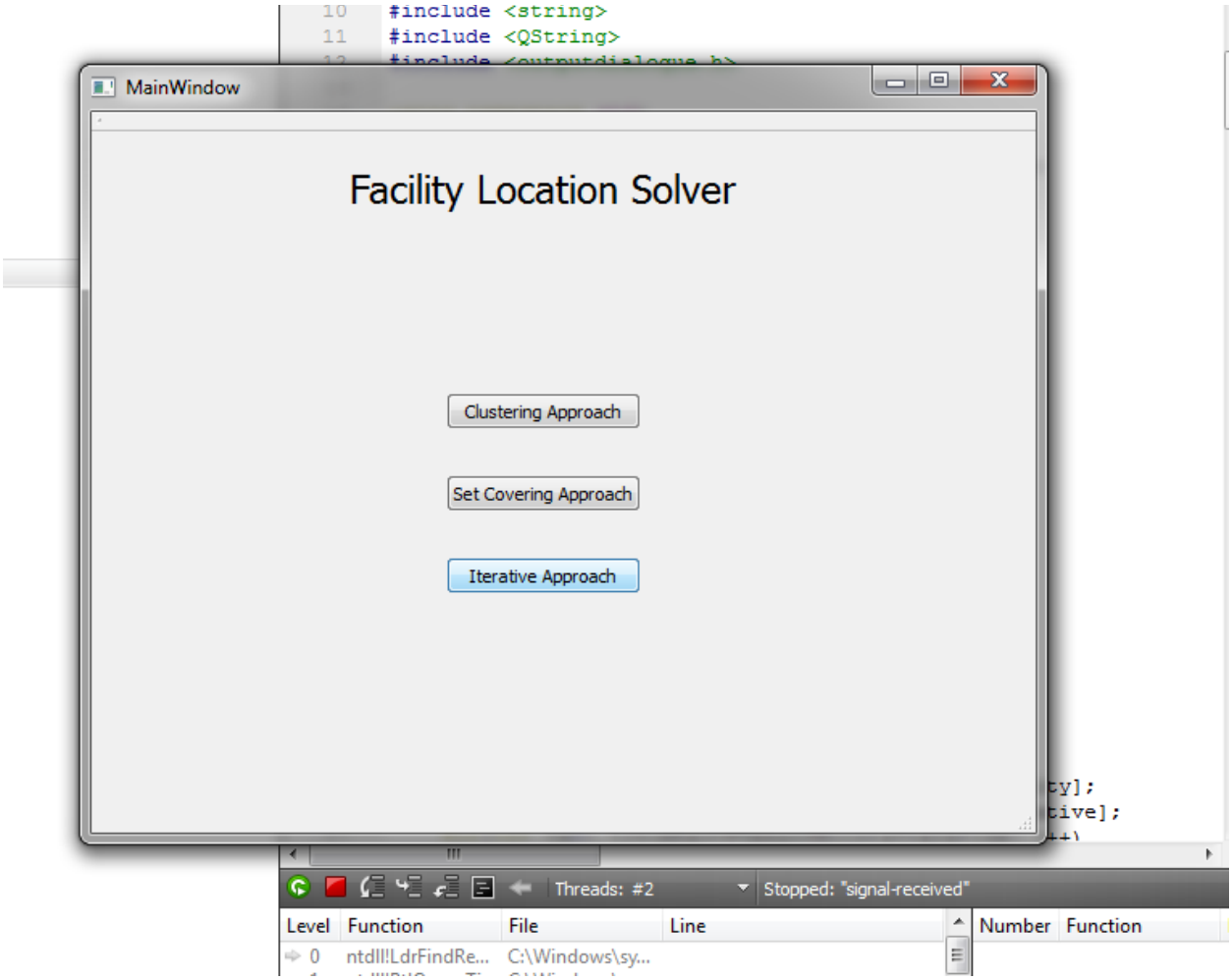
1. The user needs to use qt creator to run the program. So the qt creator 3.0.0 of qt 5.2 needs to be installed.

2. At first the user needs to run the qt program from the qt files directory. From the build option of qt creator the user needs to build and run the program.

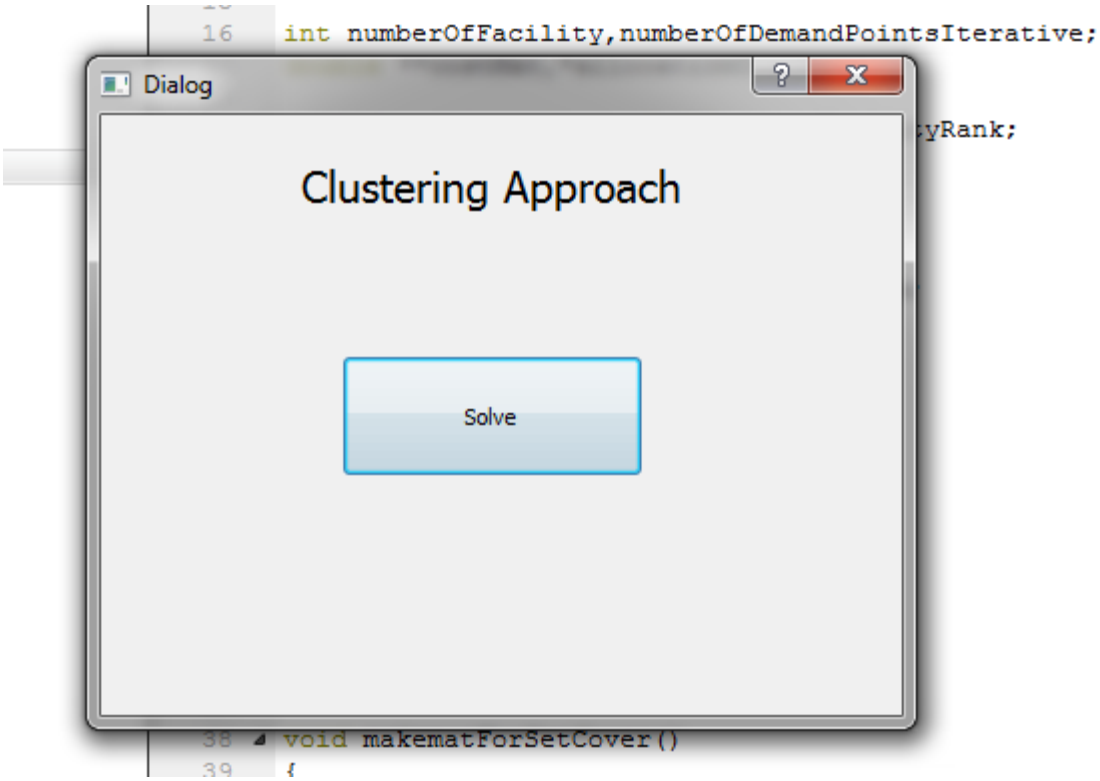


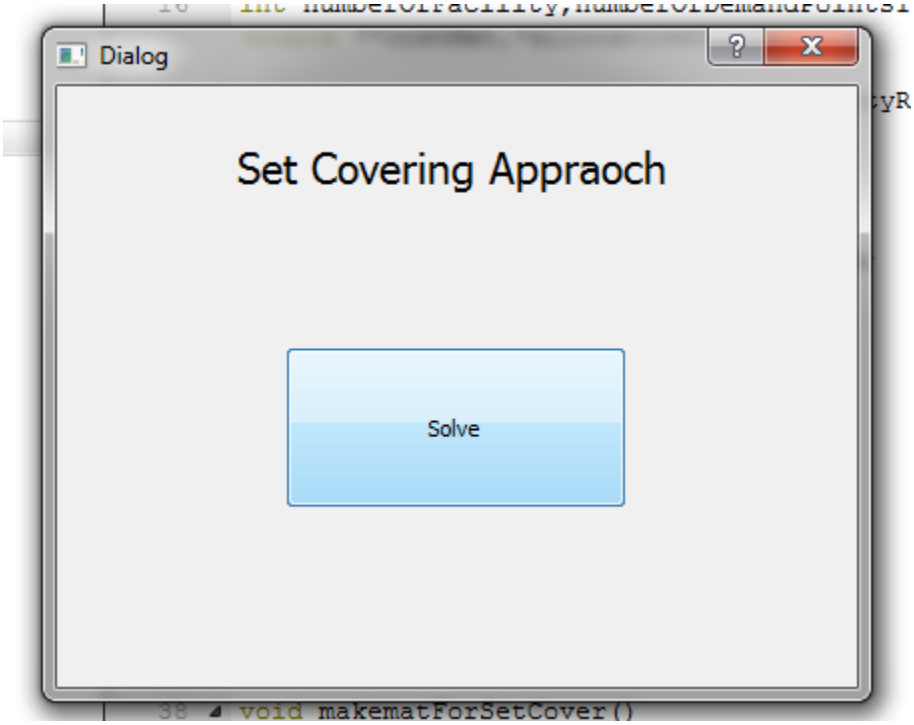
3. After running the program a main window will appear with 3 options : Clustering Approach, Set Covering Approach and Iterative Approach.

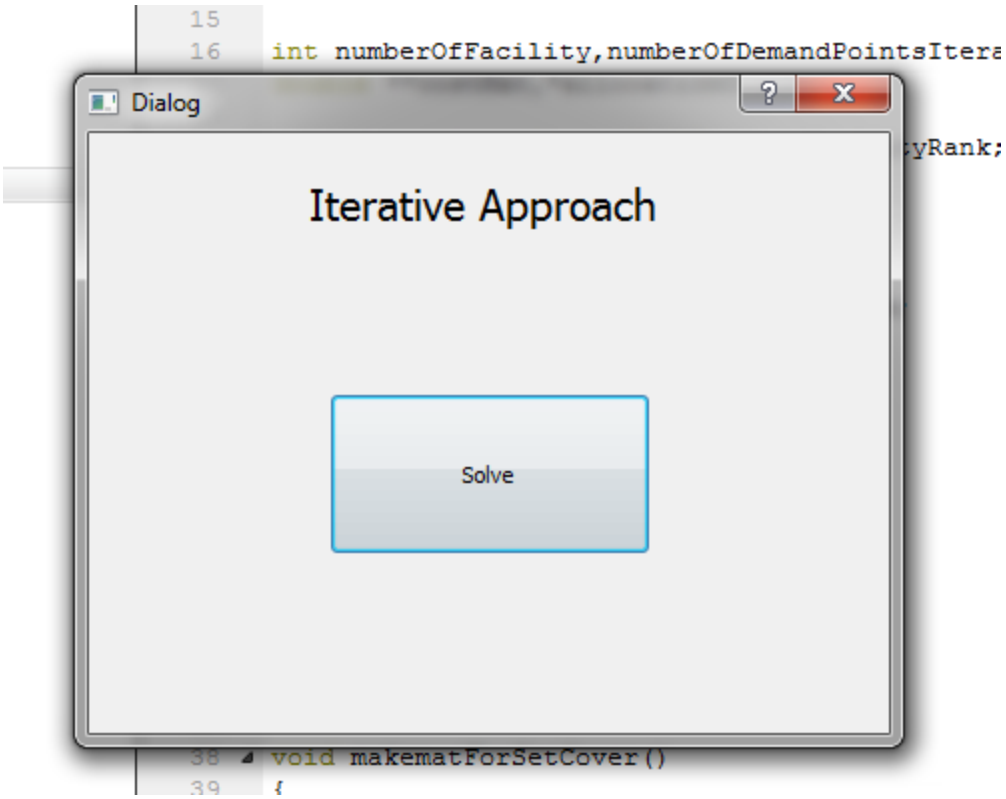
4. The user has to choose one of those three options by clicking.



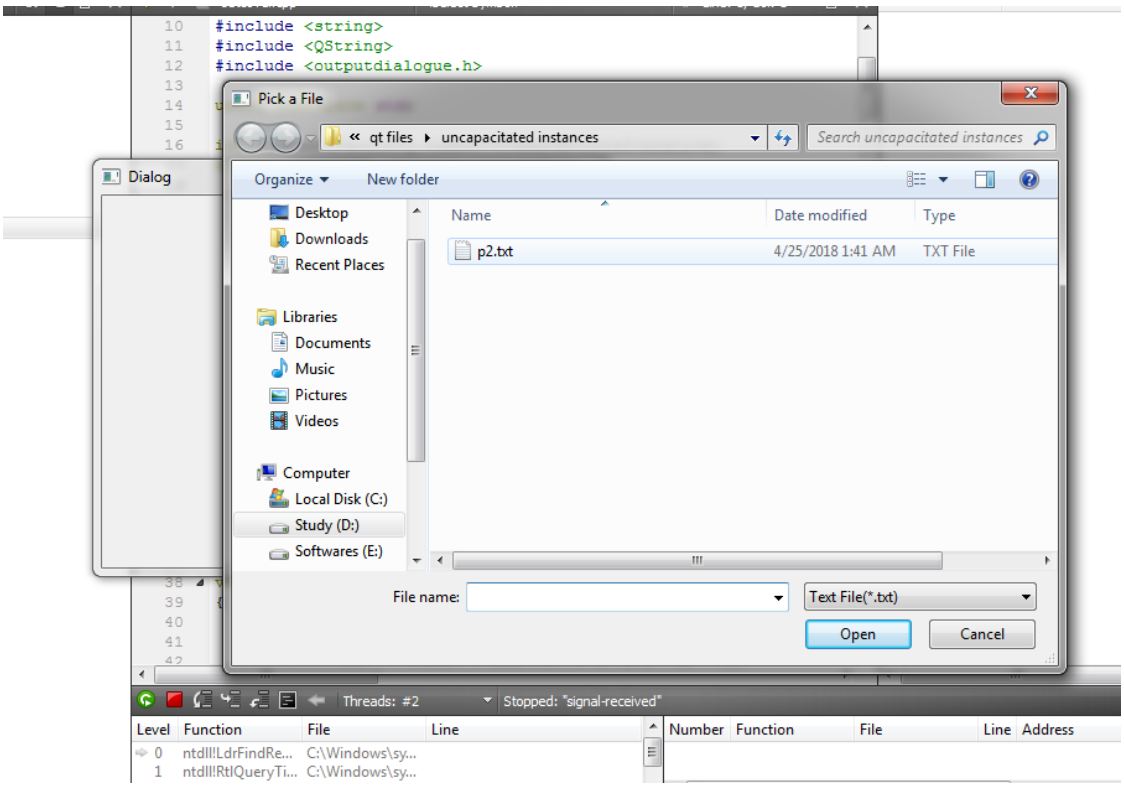
5. By clicking three different types of windows will appear depending on the approach chosen.


int numberOfFacility,numberOfDemandPointsIterative;
Dialog
Clustering Approach
Solve
void makematForSetCover()


Dialog
Set Covering Appraoch
Solve
void makematForSetCover()

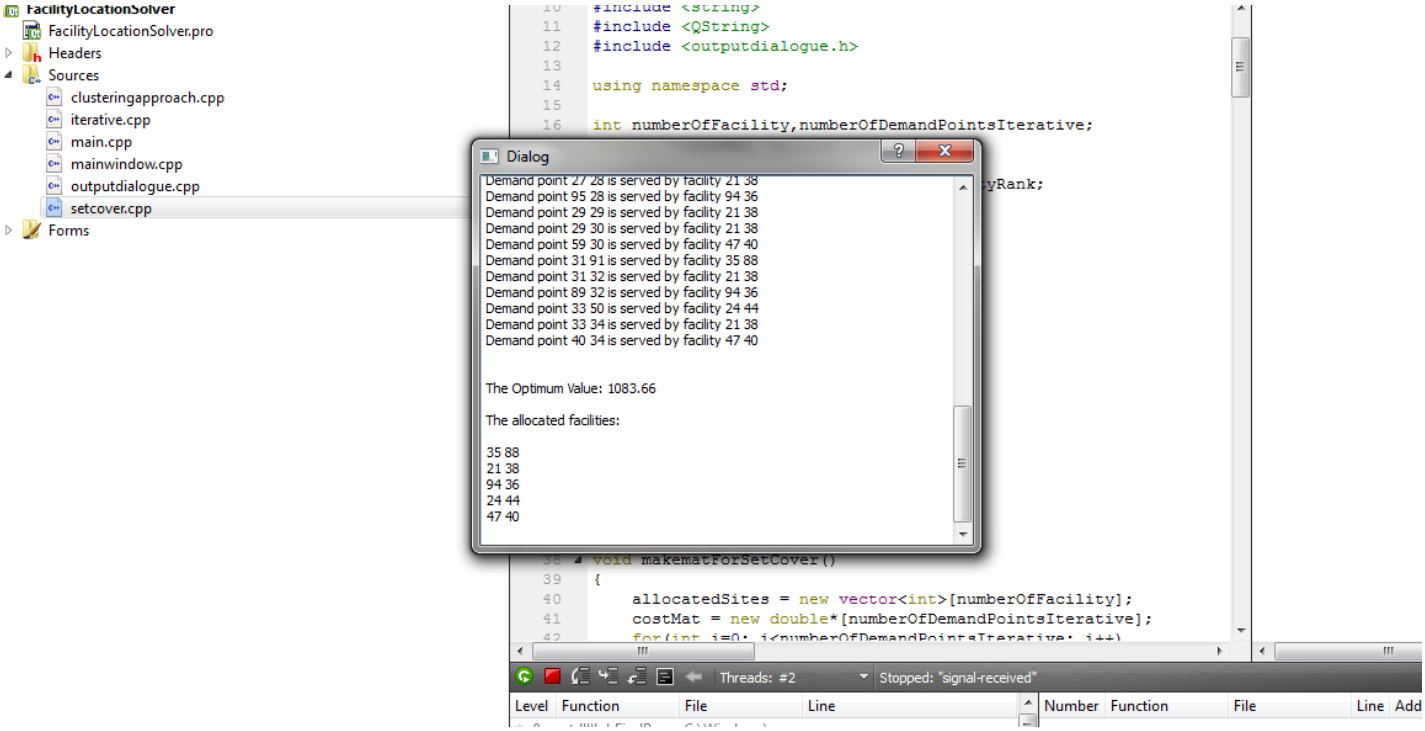


6. The solve button in each one of this opens a database of instances where one can choose an instance from the directory.



7. After choosing an instance the user needs to hit the open button.

8. Then an output window will appear describing the assignments of different demand points to different facilities and compute the optimum value.



9. Then the user can close the window to close the program.

4. Conclusion

The approaches in this program are efficient enough to handle small datasets. But when the datasets become larger it takes more computations which may not be able to find the optimal answers using these approaches.

The clustering approach is NP-hard and sometimes it iterates indefinitely. So the clustering approach can sometimes give wrong answers.

The iterative approach is also time consuming because it iterates all given instances as possible which can lead to unsolvable problems.

The set covering approach is not time consuming but it gives closer solution to the optimal one. And if the data set becomes large the solution may be far from the optimal solutions.

It is suggested that the approximation algorithms such as brach and bound algorithms can be used to solve this problems more optimally.

5. References

- Environment and Planning A, 1979, volume 11, pages 373-380
- Scientia Iranica E (2011) 18 (6), pages 1564–1570
- International Journal of Computer Science & Information Technology (IJCSIT) Vol 9, No 5, October 2017, pages 51-64
- Approximation Algorithms for Facility Location Problems, Adriana Bumb, Ph.D. thesis, 2002, pages 56-62
- (http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-74382004000100003), Uncapacitated facility location problems: contributions, last accessed on 11-04-2018
- (http://www.hyuan.com/java/how.html), The P-Median Problem, last accessed on 17-04-2018
- (http://scipbook.readthedocs.io/en/latest/flp.html#cflp-data), Facility location problems, last accessed on 01-05-2018
- (http://www.or.uni-bonn.de/~vygen/files/tokyo.pdf), Approximation Algorithms for Facility Location, last accessed on 02-05-2018
- (https://github.com/Jokers-grin/Spl1/tree/master/SPL%201), github repository, last accessed on 29-05-2018