Selective Known Sample Attacks to Relation Preserving Data Transformations

STUDENTS / UNIVERSITIES

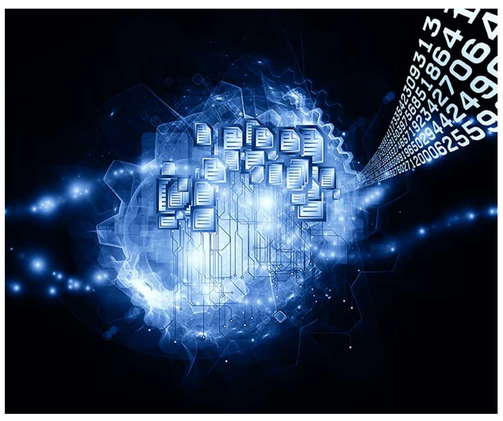
Efe KARASIL/SABANCI UNIVERSITY
Halit Furkan KOCYIGIT/SABANCI UNIVERSITY

SUPERVISOR(S)

Oznur TASTAN
Mehmet Ercan NERGIZ

INTRODUCTION

Any kind of sequence of one or more symbols that are given meaning through interpretation acts specifically represents the data. As the data spreading to people's everyday life increasingly with the contribution of technology, the significance of data itself and its security come into prominence. One of the biggest security threats in technology is the attack which has some variety of forms that can take a place. These attacks lead privacy issues to exist because their purposes can be altering, getting or even destroying the data.



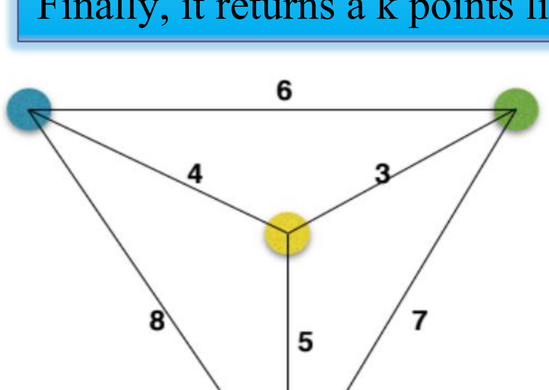
At that stage data mining algorithms take place to attain the data. Distances between the records can be enough and be more reliable rather than records themselves for some data mining algorithms that are used to get some information about data and its properties. These data mining algorithms can work sufficiently without need data that are private. Because of their structural properties, their parameters are distance matrix or transformed version of records.

METHODOLOGY

In out project, we implemented our algorithms to python. Also we used csv files to get the relation datas.

First Algorithm

In our first algorithm, we tried to find the farthest points. To do that,firstly we got a k point from the user and we created lists of points which have k points in it.Then, we splitted our point list into the sublists which have only 2 points. We could reach the data of these pairwise distances sublists by reading the csv file. Then the program calculated the APD(Average Pairwise Distance) of k point lists by using sublists of them. Finally, it returns a k points list which has the biggest APD.



	YELLOW	BLUE	GREEN	RED
YELLOW	0	4	3	5
BLUE	4	0	6	8
GREEN	3	6	0	7
RED	5	8	7	0

$$\mathrm{APD}(I) = \frac{\sum_{(x,y)\in I} M[x][y]}{\binom{k}{2}}$$

If we choose k = 3 (sample of 3 points from matrix)
BYG->13/3 YGR->15/3 BYR->17/3 BRG->21/3

Second Algorithm

In our second algorithm, we get an exact point from the user or we choose it randomly and we tried to reach the farthest k points. Firstly, we found the farthest point of that point and we combine it. Then, we found third point which is farthest from these 2 points. Afterwards, we repeated adding new points to our list by using loops. When we reach the number of k, we return the result.

For instance if we choose yellow as first entry:
----------> R = {Y}

The second entry that maximise the APD(e.g, .farthest) will be calculated for the R subset.
YB -> 4 YG -> 3 YR -> 5 ----------> R = {Y,R}

Then, again the entry that maximise the APD(e.g, .farthest) will be calculated for the R subset.
YRB -> 17/3 YRG -> 15/3

Lastly we add the blue as the entry ----------> R = {Y,R,B}

CONCLUSION

In this project, our aim was creating the most efficient attack. We worked on two algorithms. Even though we implemented these two algorithms, we have not check the efficiency of them in different situations and conditions such as different parameters, dimensions and different k number. Although implementing some algorithms to python is interesting, we should study on the attack algorithm in our future project.

REFERENCES

Kaplan, E., Gursoy, M. E., Nergiz, M. E., & Saygin, Y. (2017). Known Sample Attacks on Relation Preserving Data Transformations. IEEE Transactions on Dependable and Secure Computing, 1-1. doi:10.1109/tdsc.2017.2759732