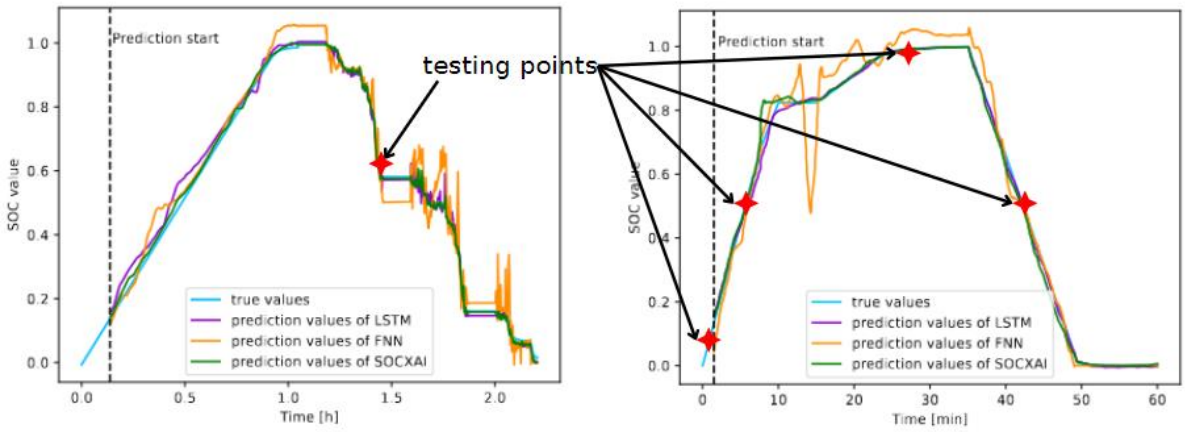
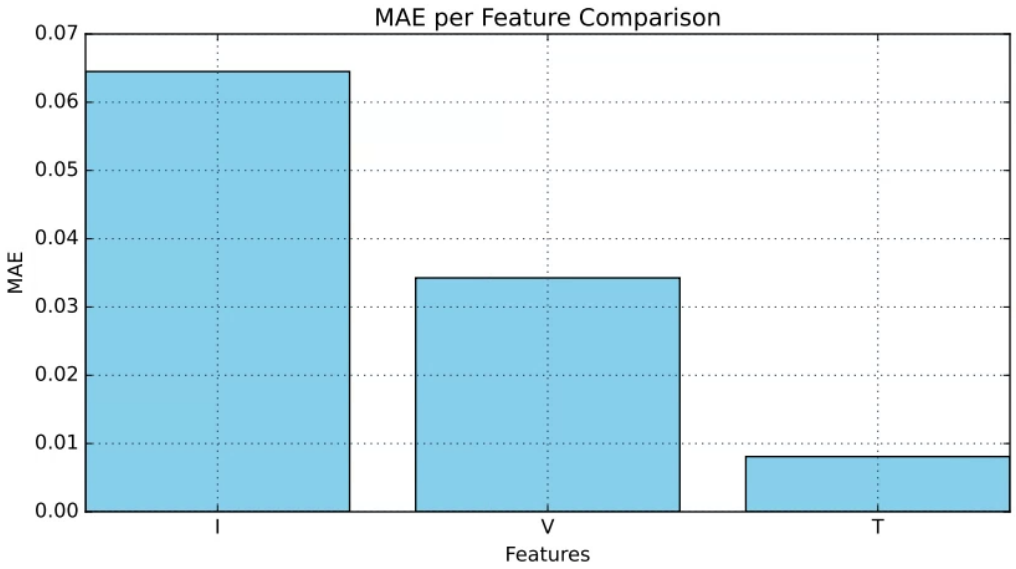
Report: Counterfactual explanations

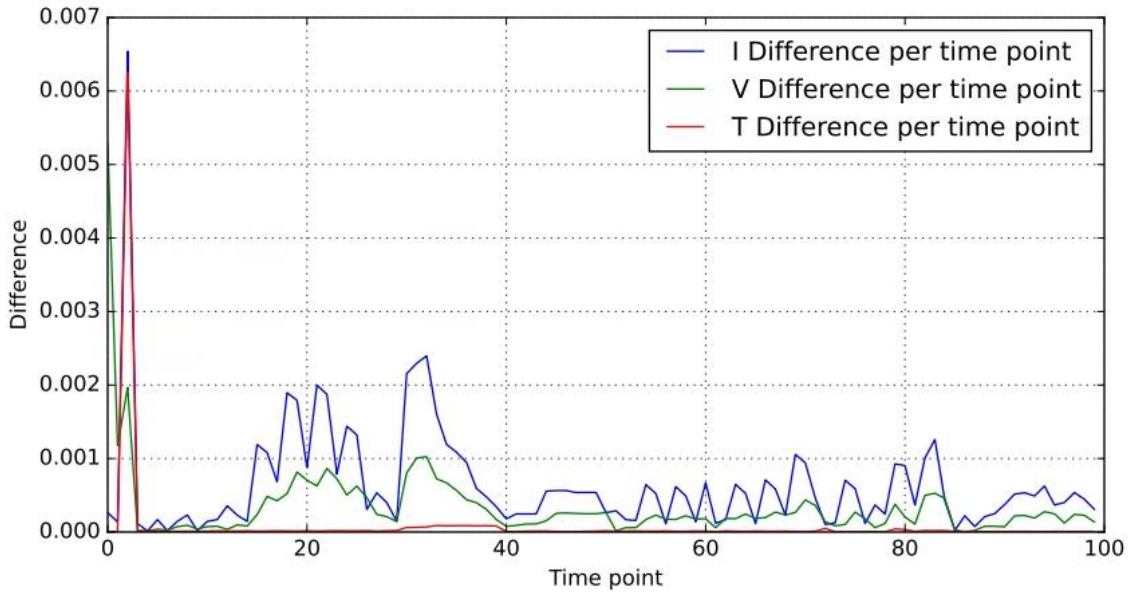


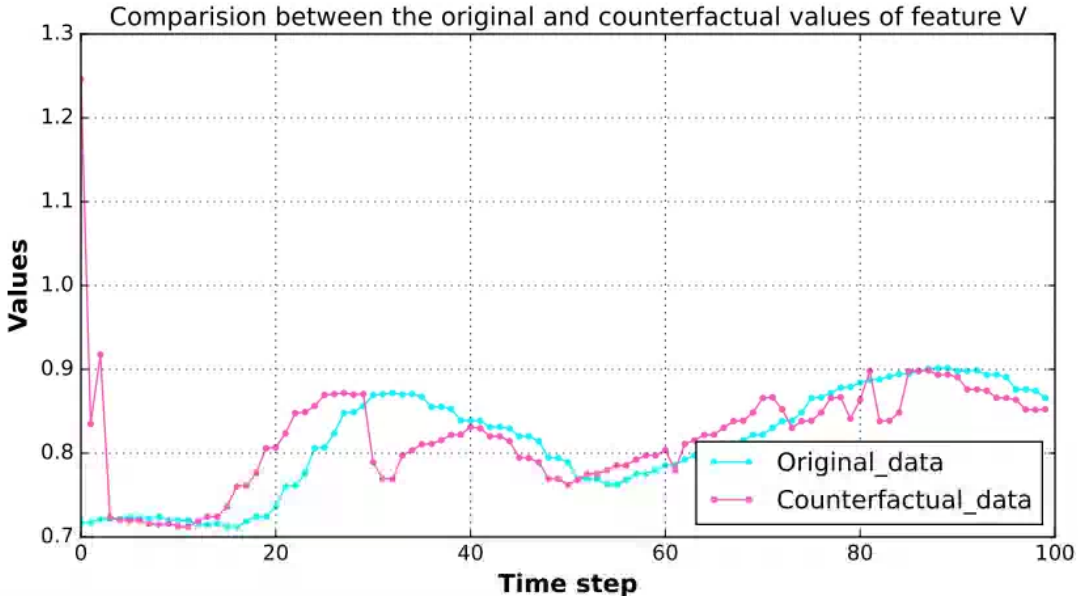
Discharging with non constant current

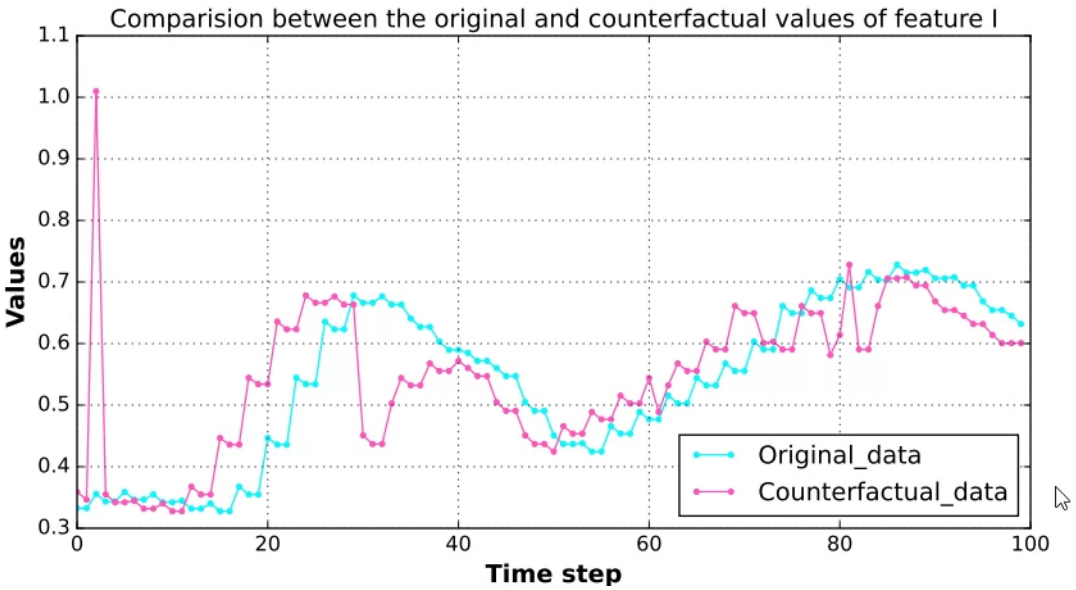
Geno-TOPSIS method

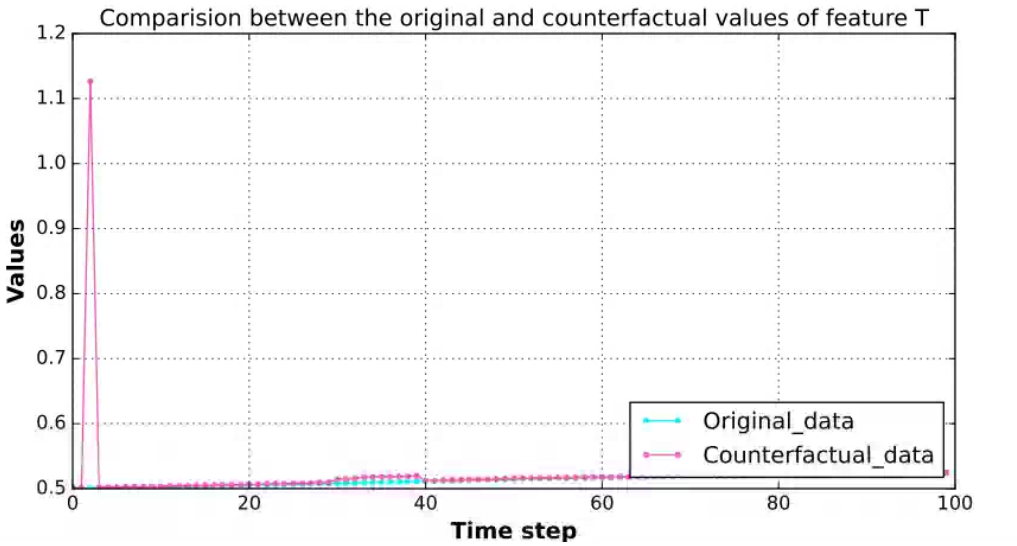
The method took **3227.12s** to generate 20 counterfactuals based on a population of 100 individuals.

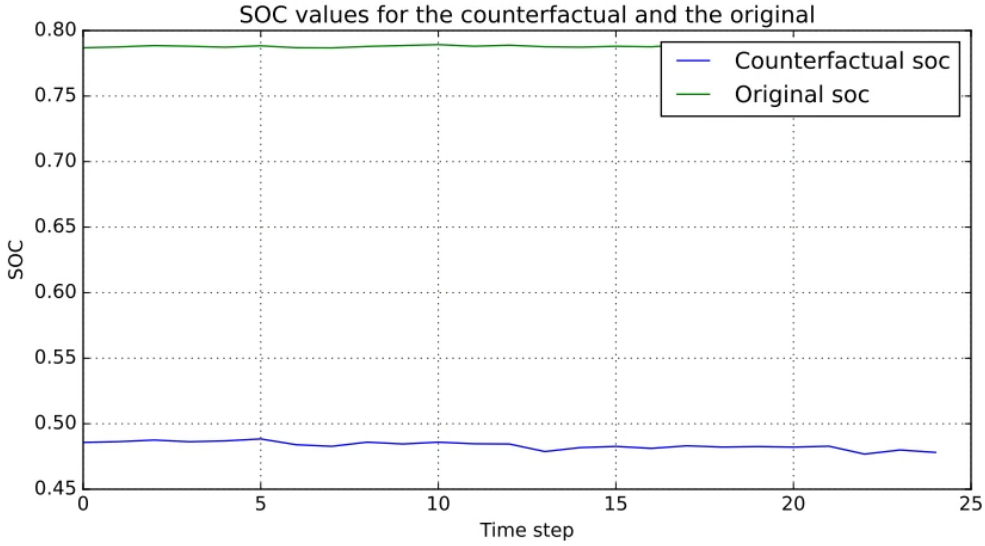


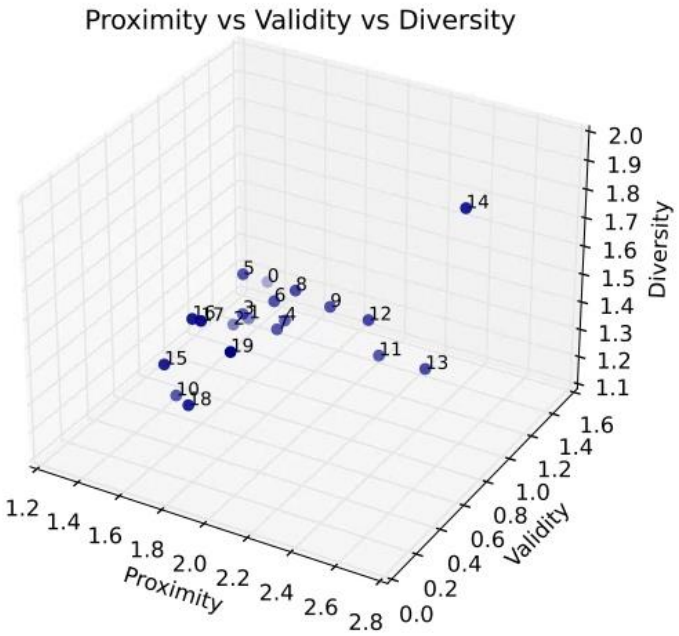

I Difference per time point
V Difference per time point
T Difference per time point
Difference
Time point
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
0
20
40
60
80
100


Comparision between the original and counterfactual values of feature V
Original_data
Counterfactual_data
Values
Time step
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0
20
40
60
80
100


Comparision between the original and counterfactual values of feature I
Original_data
Counterfactual_data
Values
Time step
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
20
40
60
80
100

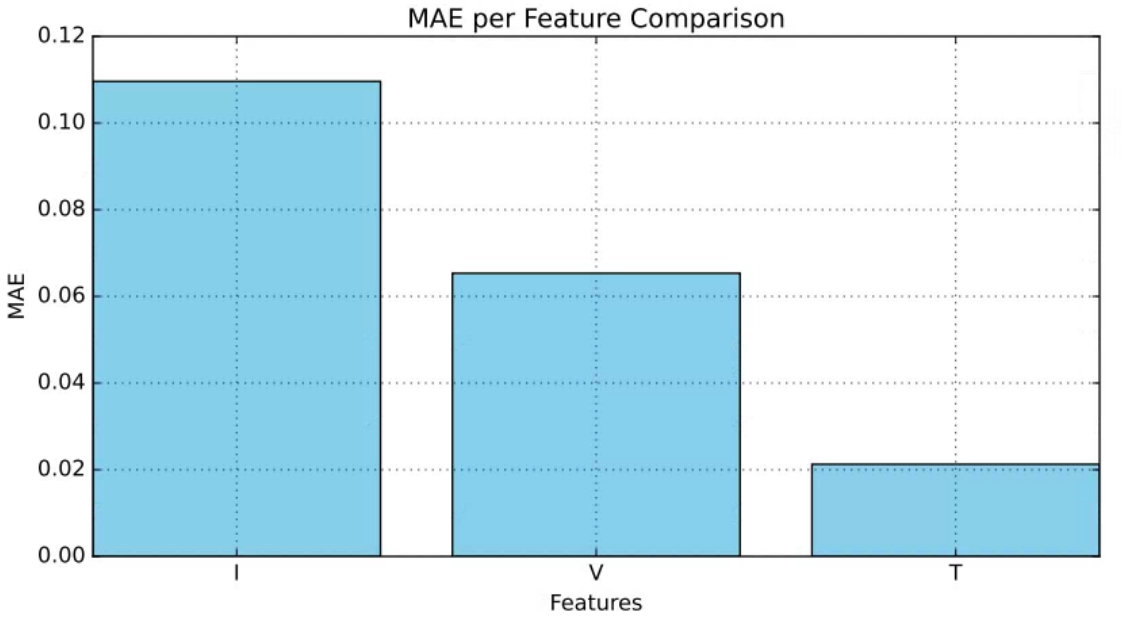

Comparision between the original and counterfactual values of feature T
Values
Time step
Original_data
Counterfactual_data

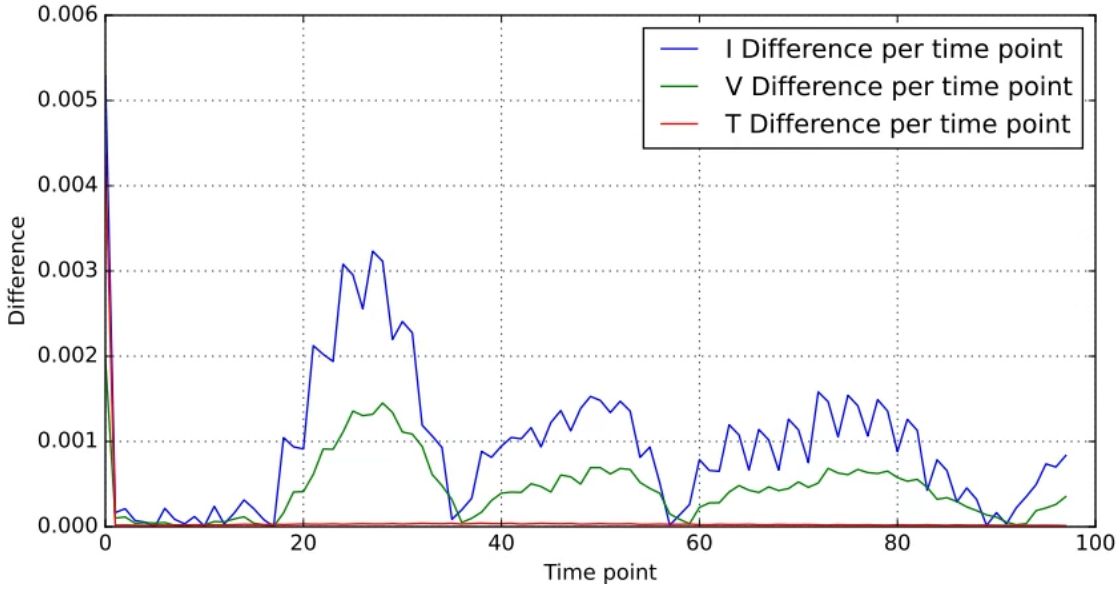

SOC values for the counterfactual and the original
SOC
Time step
Counterfactual soc
Original soc

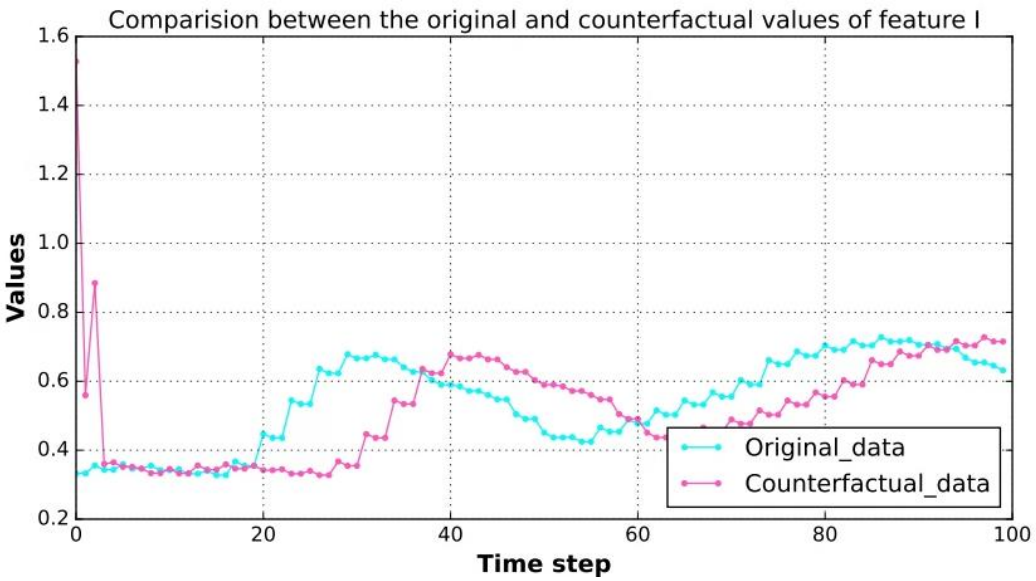

Proximity vs Validity vs Diversity
Proximity
Validity
Diversity

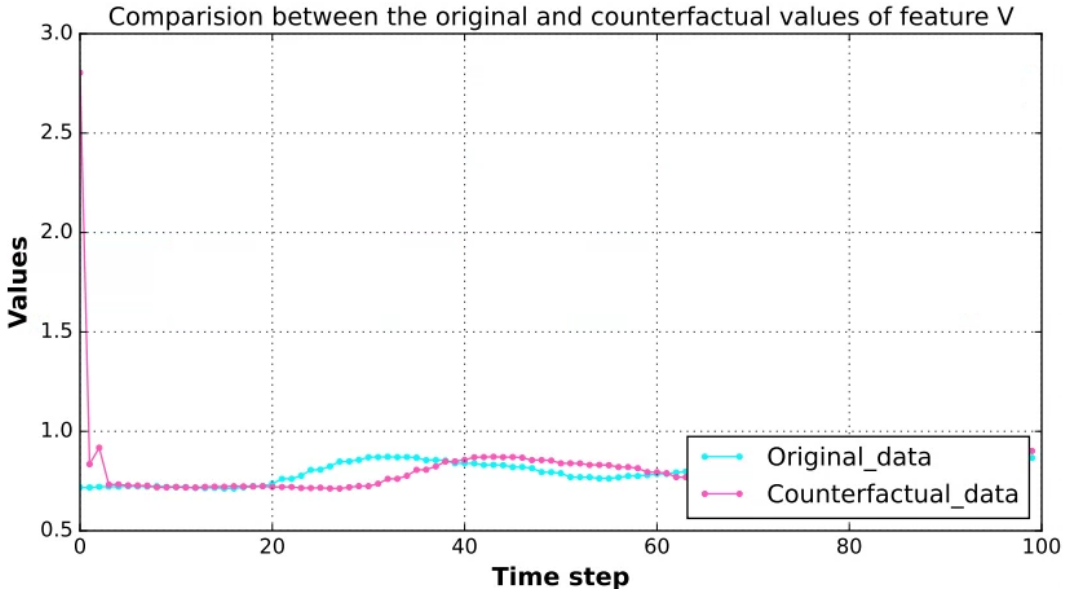
NSG2 method

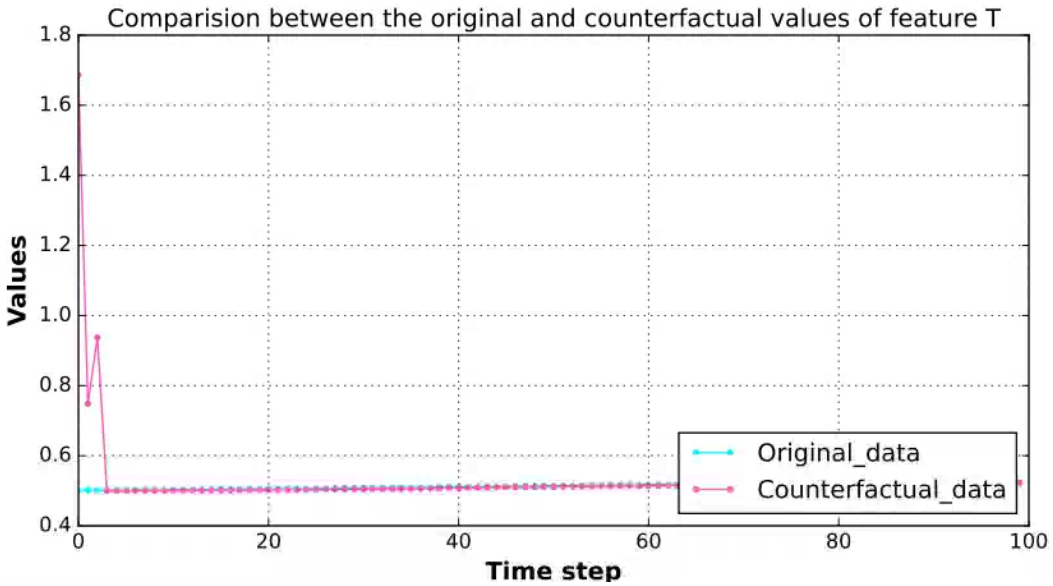
The method took **312 s** for more than 20 counterfactuals based on a population of 100 individuals.

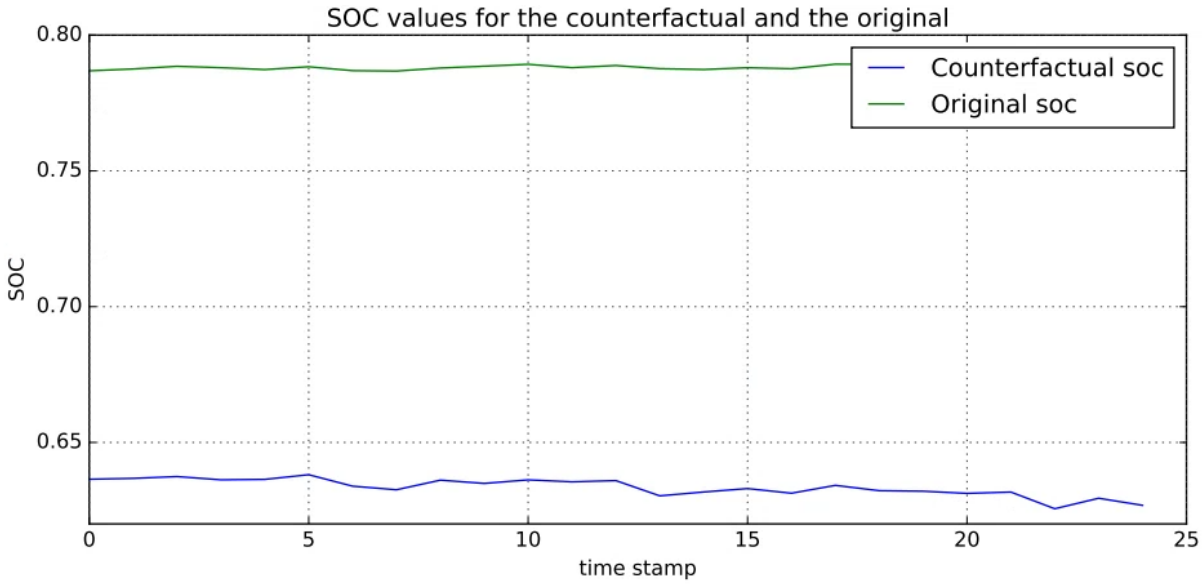


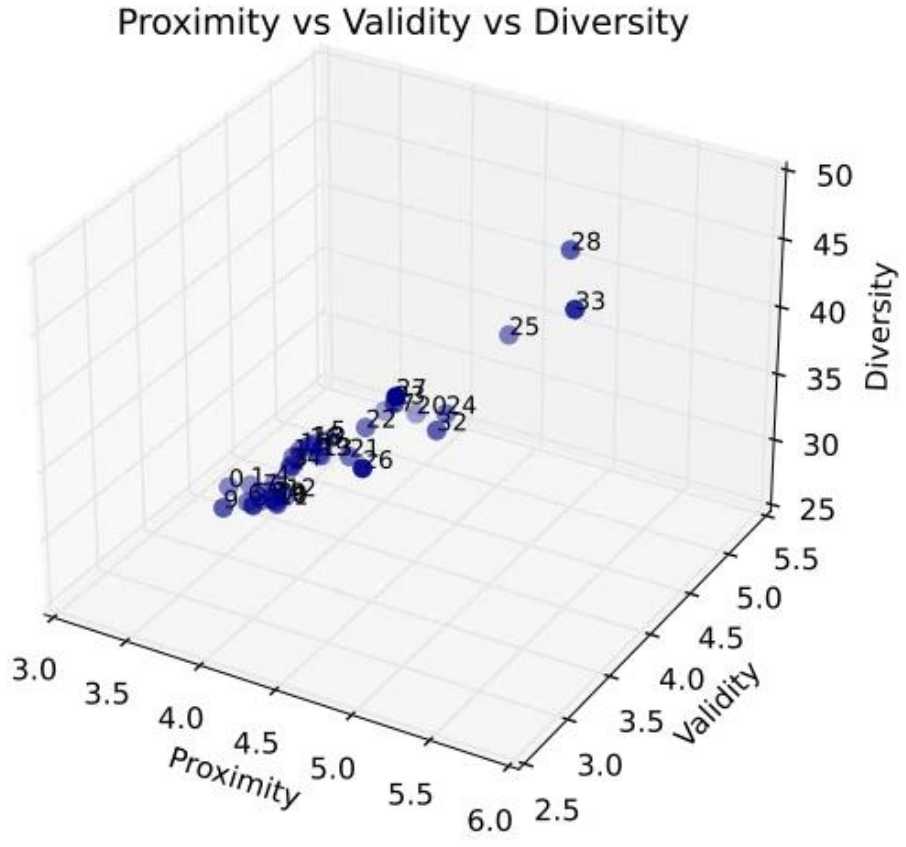



Comparision between the original and counterfactual values of feature I
Values
Time step
Original_data
Counterfactual_data


Comparision between the original and counterfactual values of feature V
Values
Time step
Original_data
Counterfactual_data


Comparision between the original and counterfactual values of feature T
Values
Time step
Original_data
Counterfactual_data

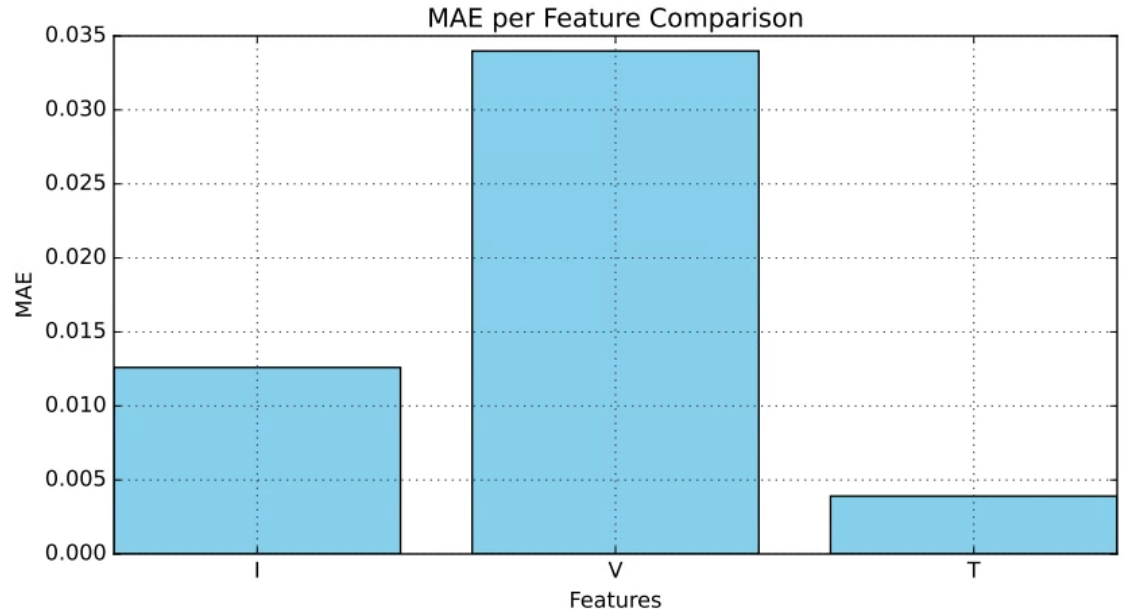

SOC values for the counterfactual and the original
Counterfactual soc
Original soc
SOC
time stamp
0.80
0.75
0.70
0.65
0
5
10
15
20
25

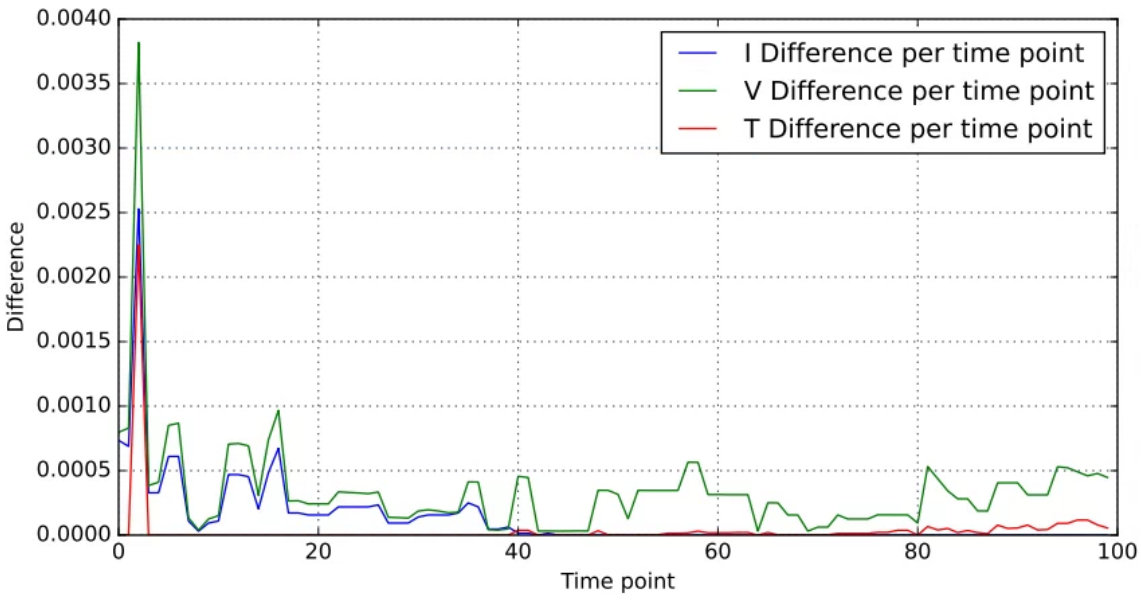

Proximity vs Validity vs Diversity
Proximity
Validity
Diversity
3.0
3.5
4.0
4.5
5.0
5.5
6.0
2.5
25
30
35
40
45
50

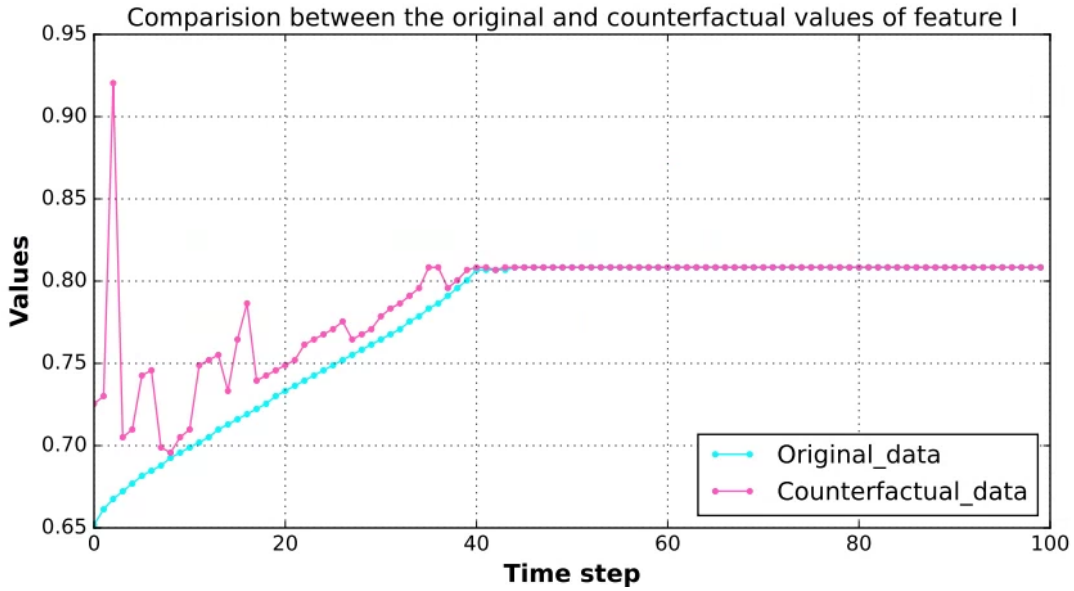
Charging with constant current (first 100 values)

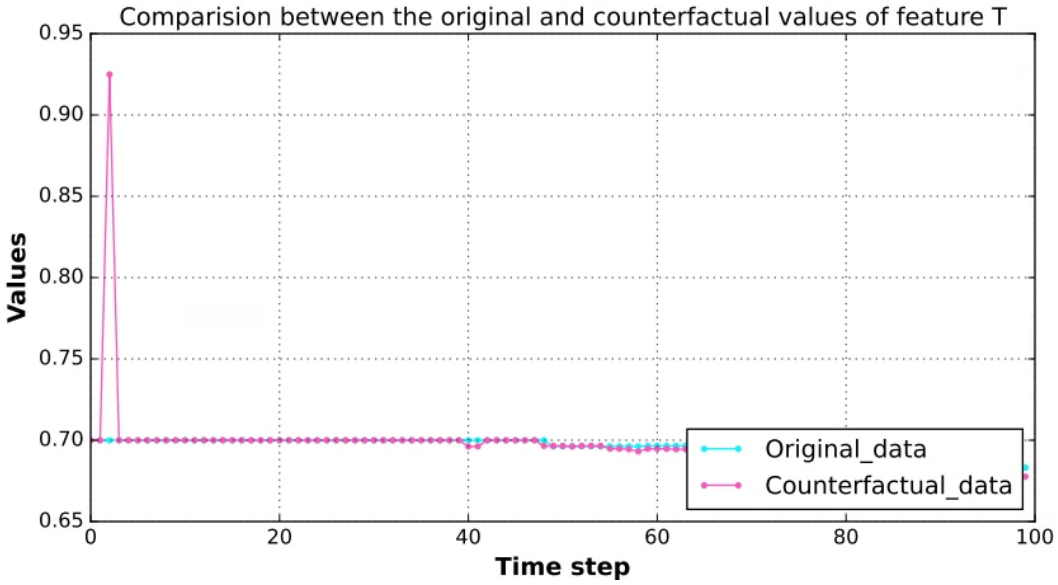
Geno-TOPSIS method

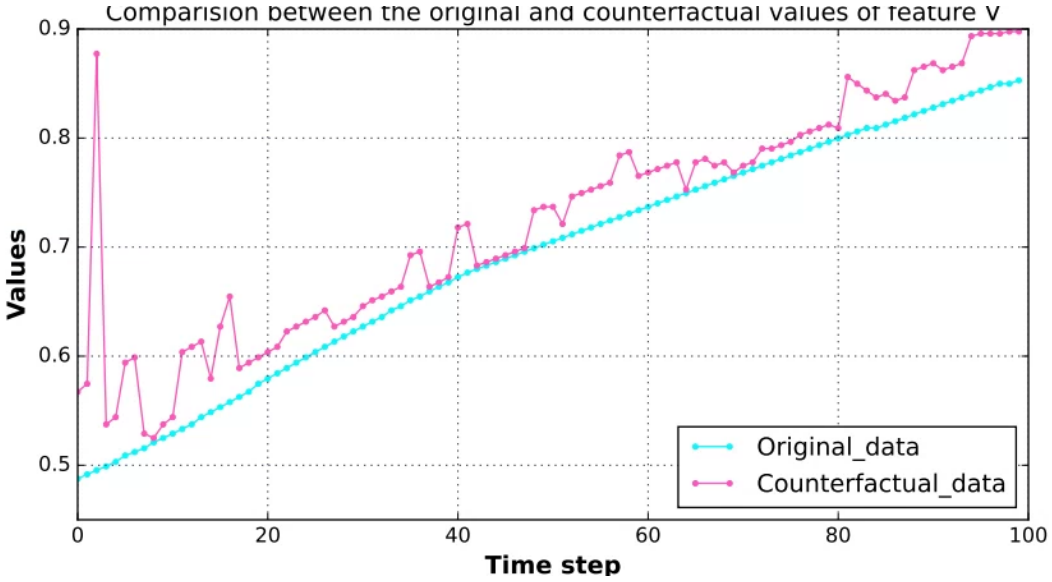
The method took 5176.87s to generate 20 counterfactuals based on a population of 100 individuals.

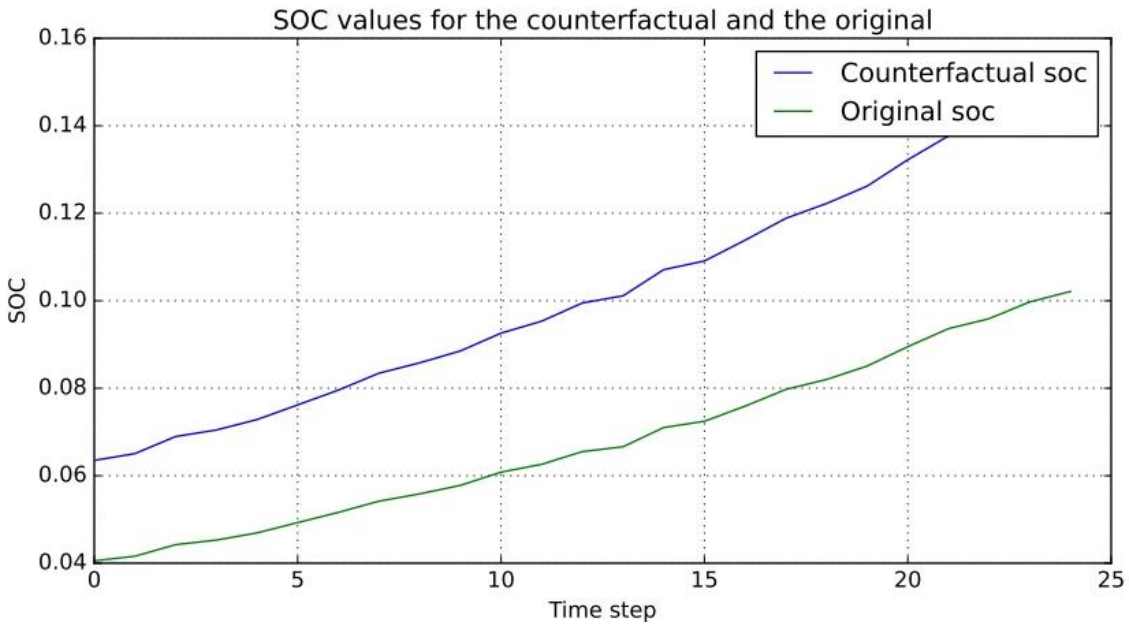


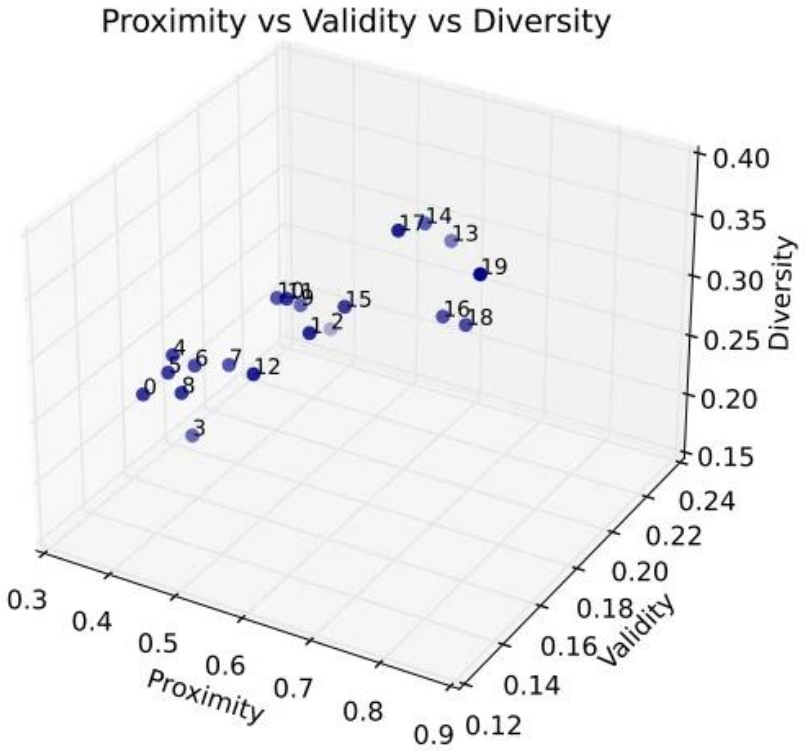



Comparision between the original and counterfactual values of feature I
Values
Time step
Original_data
Counterfactual_data


Comparision between the original and counterfactual values of feature T
Values
Time step
Original_data
Counterfactual_data

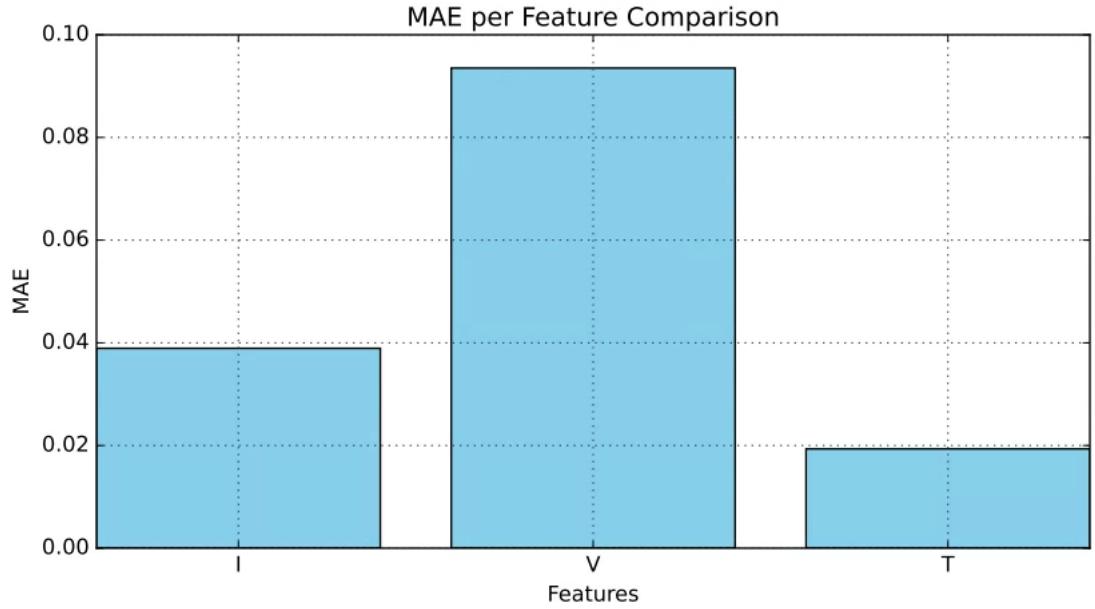

Comparision between the original and counterfactual values of feature V
Values
Time step
Original_data
Counterfactual_data

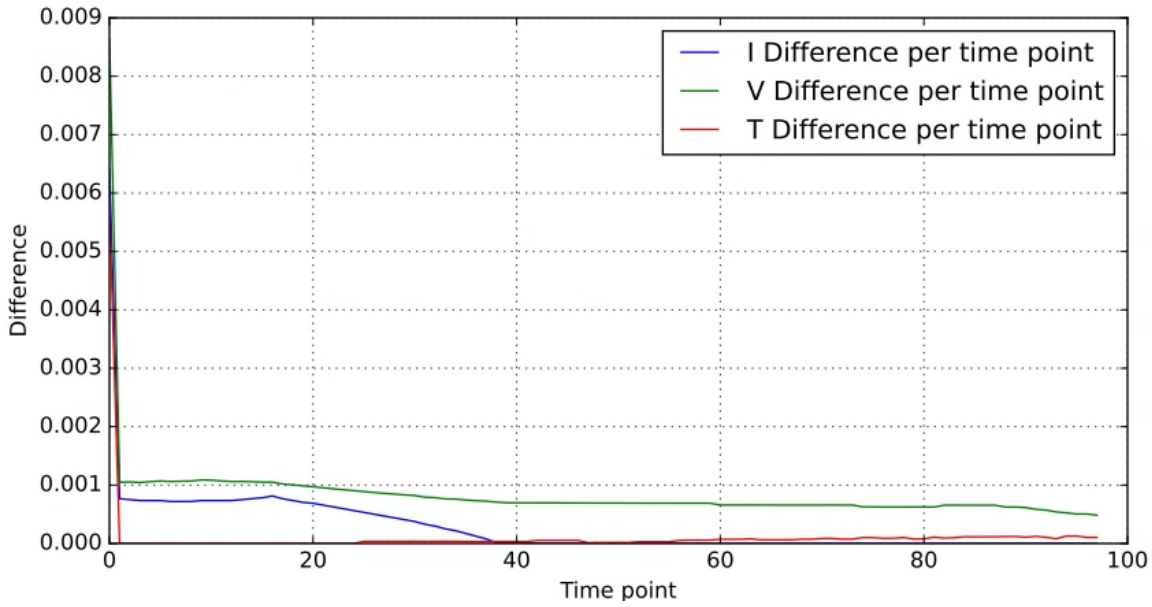

SOC values for the counterfactual and the original
Counterfactual soc
Original soc
SOC
Time step
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0
5
10
15
20
25

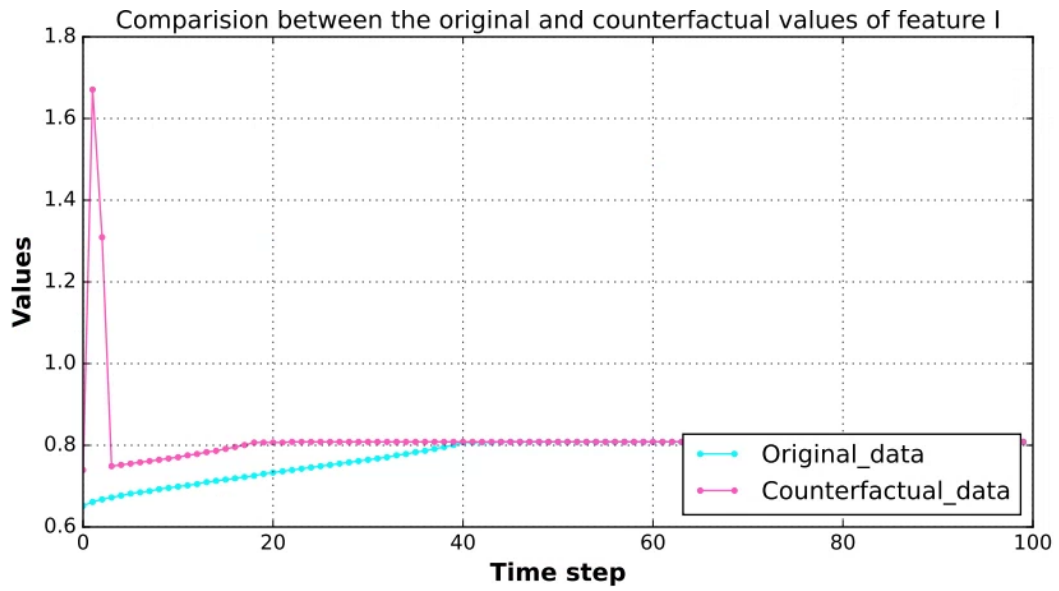

Proximity vs Validity vs Diversity
Proximity
Validity
Diversity
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.12
0.14
0.16
0.18
0.20
0.22
0.24
0.15
0.20
0.25
0.30
0.35
0.40

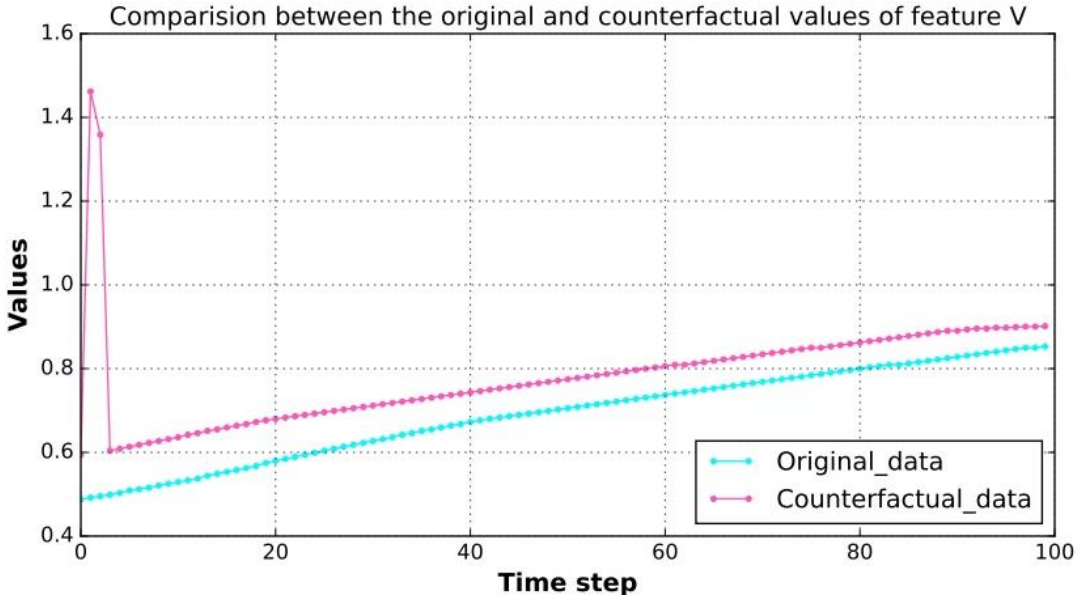
NSGA2 method

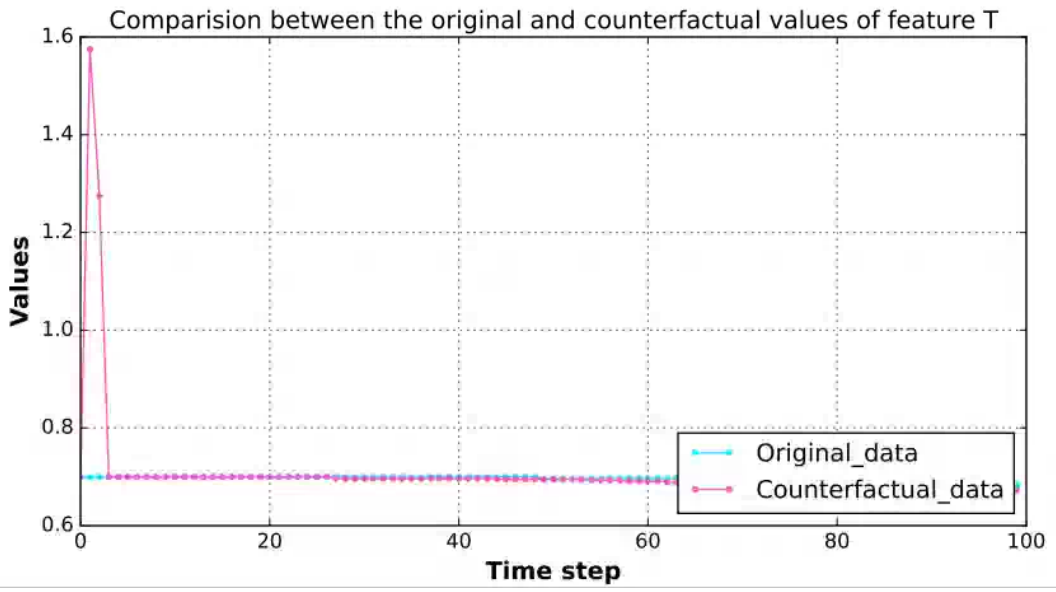
The method took 181s to generate more than 20 counterfactuals based on the same population.

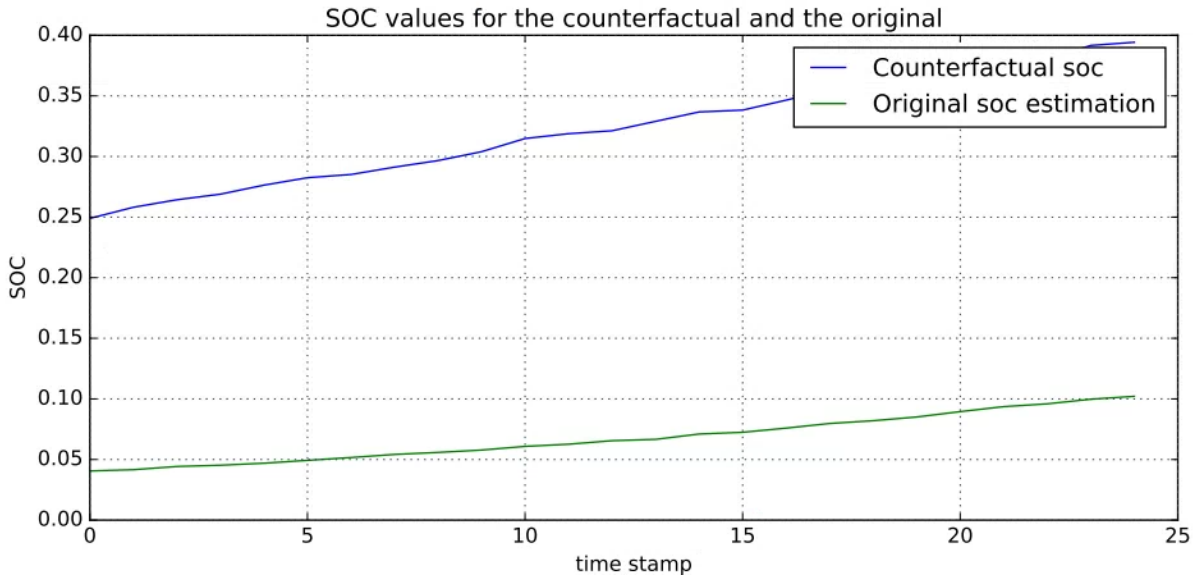


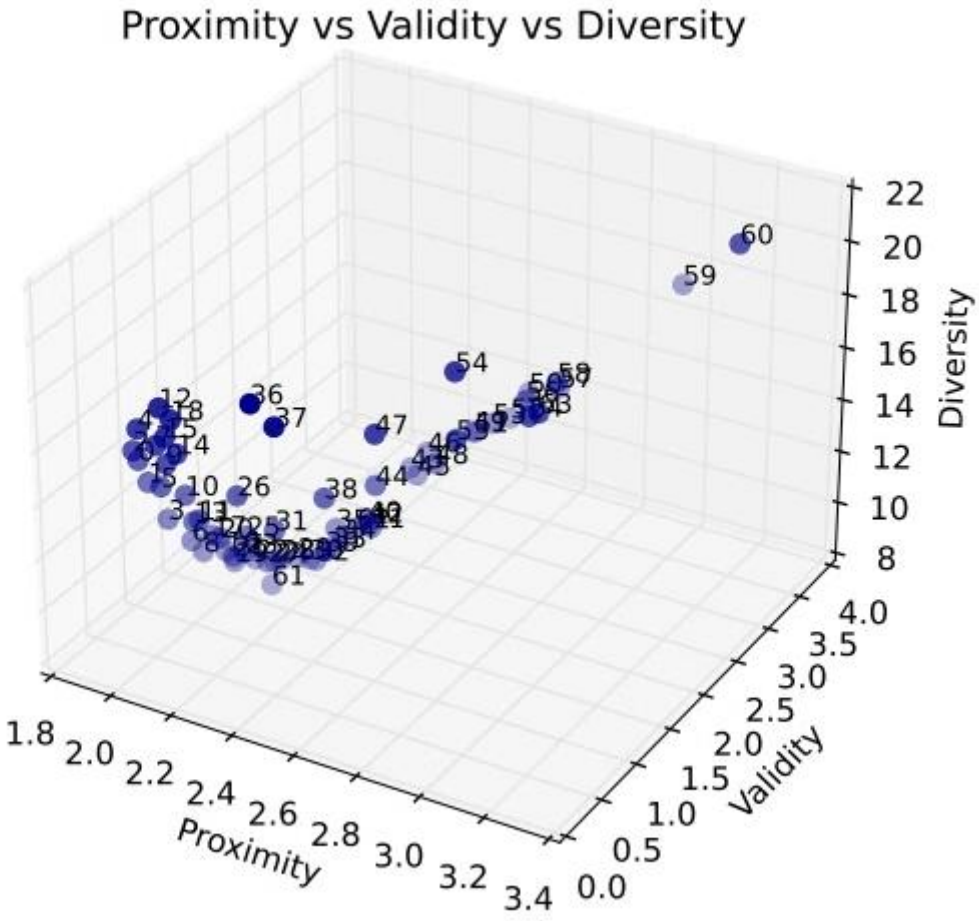



Comparision between the original and counterfactual values of feature I
Values
Time step
Original_data
Counterfactual_data


Comparision between the original and counterfactual values of feature V
Values
Time step
Original_data
Counterfactual_data


Comparision between the original and counterfactual values of feature T
Values
Time step
Original_data
Counterfactual_data

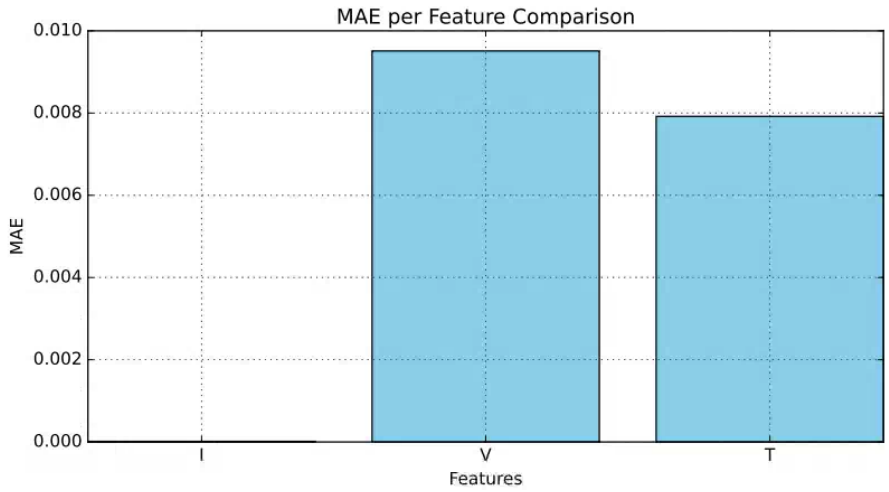

SOC values for the counterfactual and the original
Counterfactual soc
Original soc estimation
SOC
time stamp
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0
5
10
15
20
25

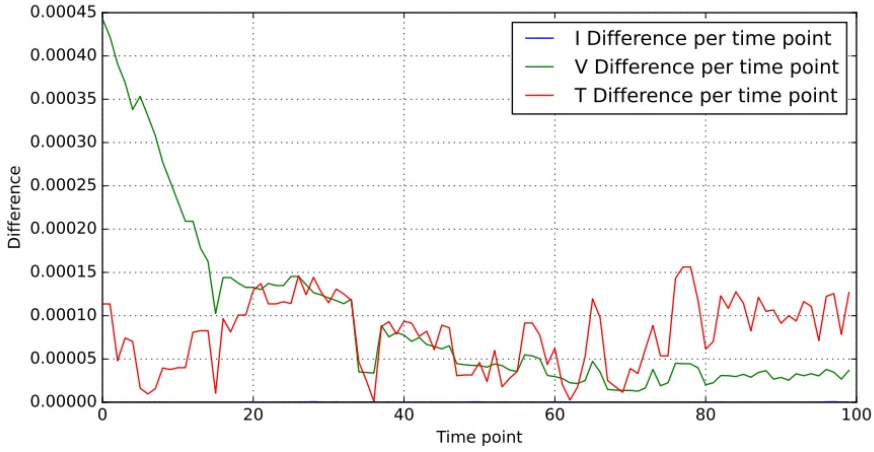

Proximity vs Validity vs Diversity
Proximity
Validity
Diversity
1.8
2.0
2.2
2.4
2.6
2.8
3.0
3.2
3.4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
8
10
12
14
16
18
20
22

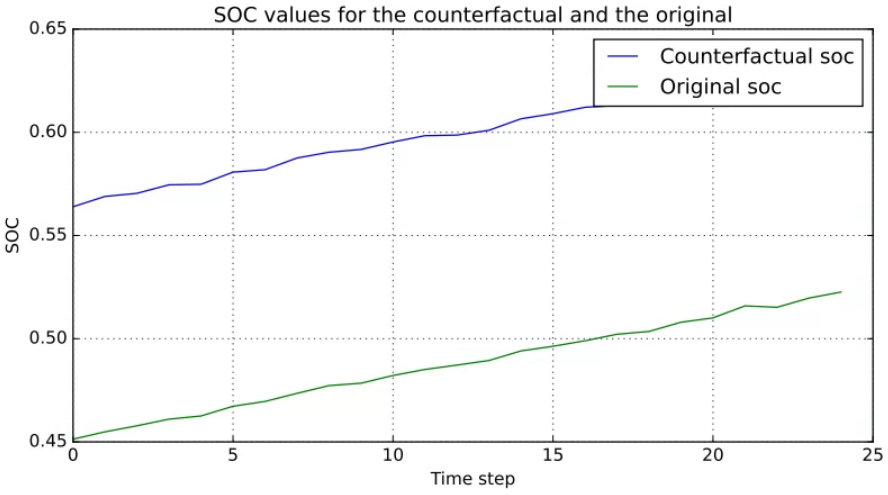
Charging with constant current (SOC between 20% and 80%)

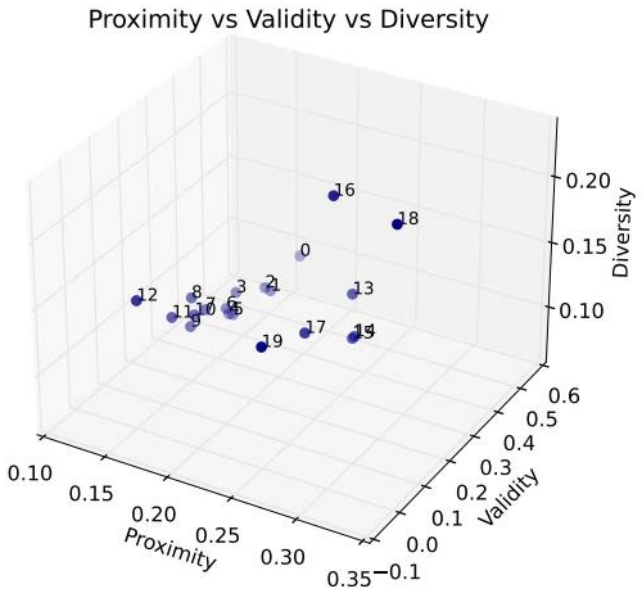
Geno-TOPSIS method

The method took 1116.22s to generate counterfactuals.



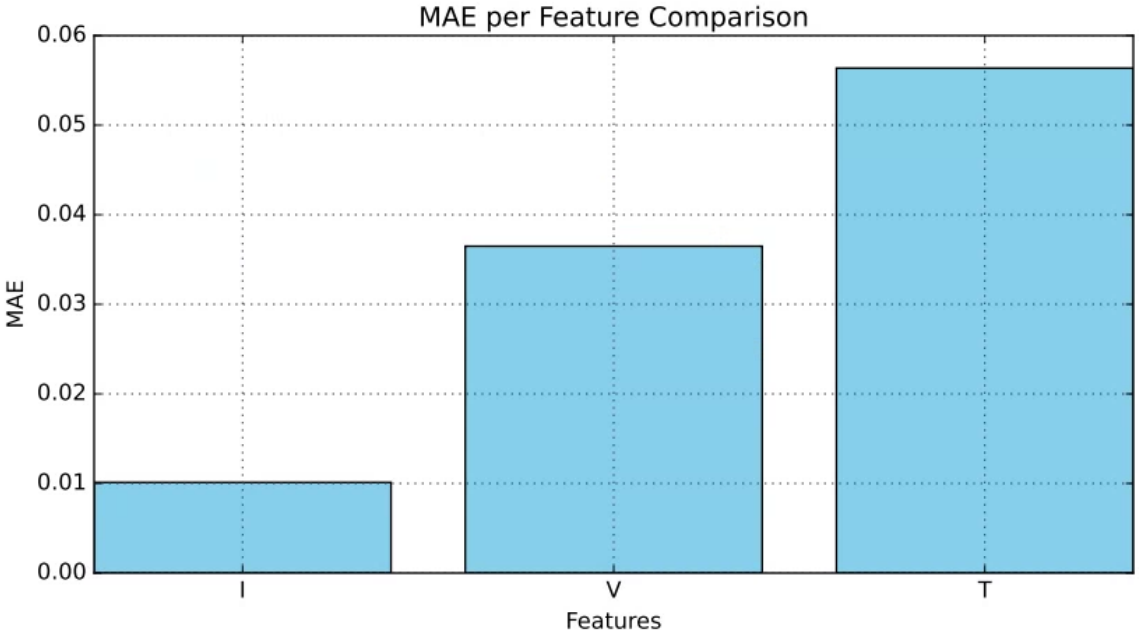


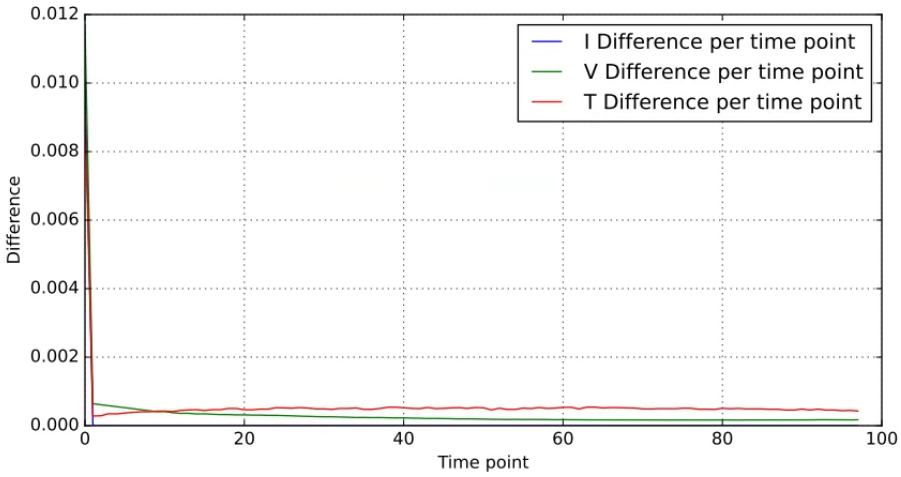


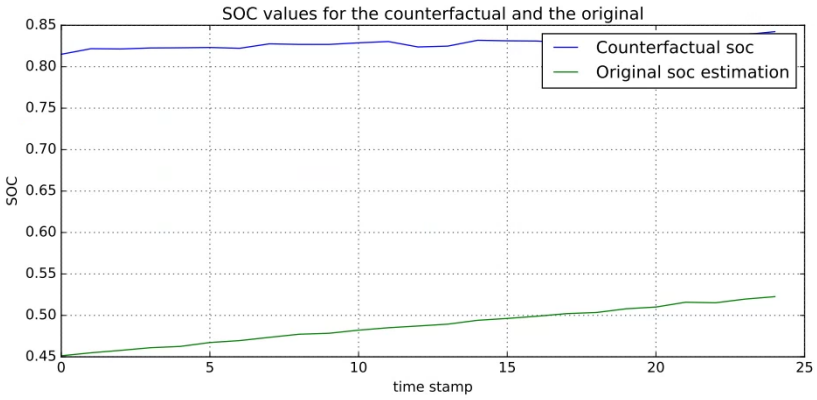


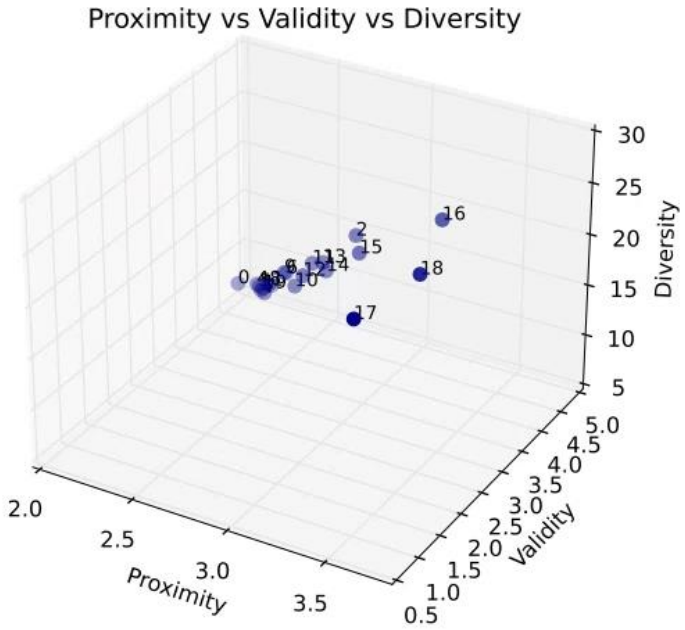
NSGA2 method

The method took 77s to generate counterfactuals.




I Difference per time point
V Difference per time point
T Difference per time point
Difference
Time point
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
20
40
60
80
100

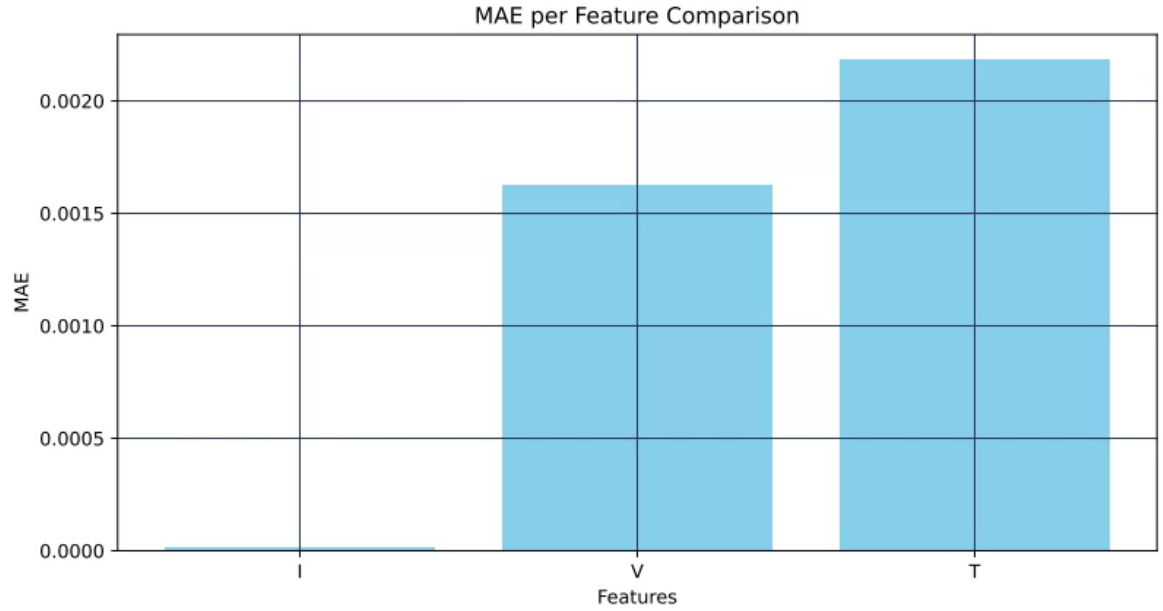

SOC values for the counterfactual and the original
Counterfactual soc
Original soc estimation
SOC
time stamp
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0
5
10
15
20
25

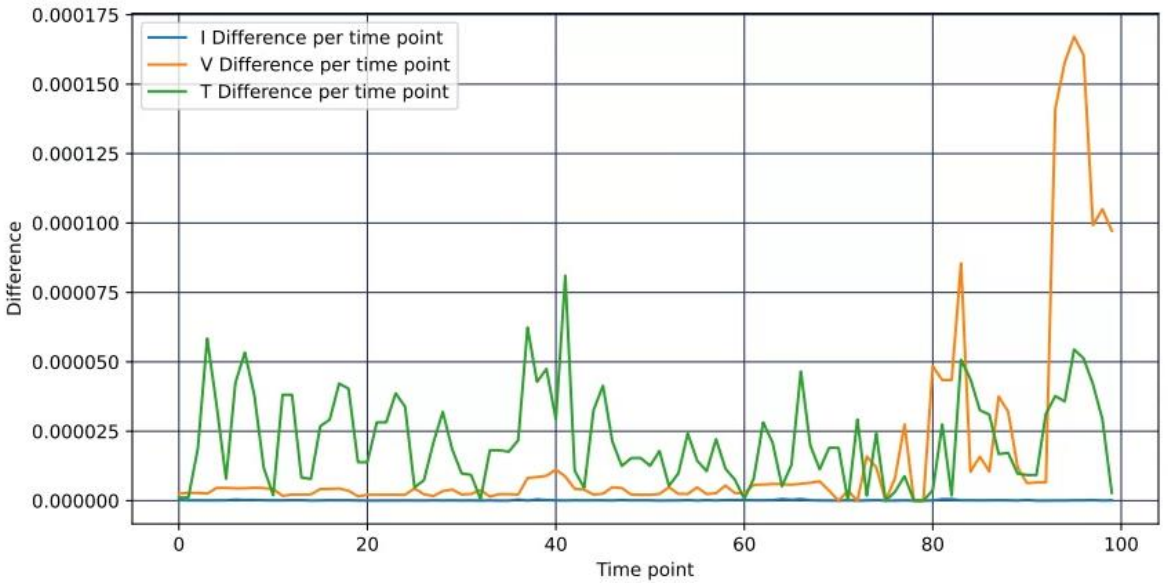

Proximity vs Validity vs Diversity
Proximity
Validity
Diversity
2.0
2.5
3.0
3.5
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5
10
15
20
25
30

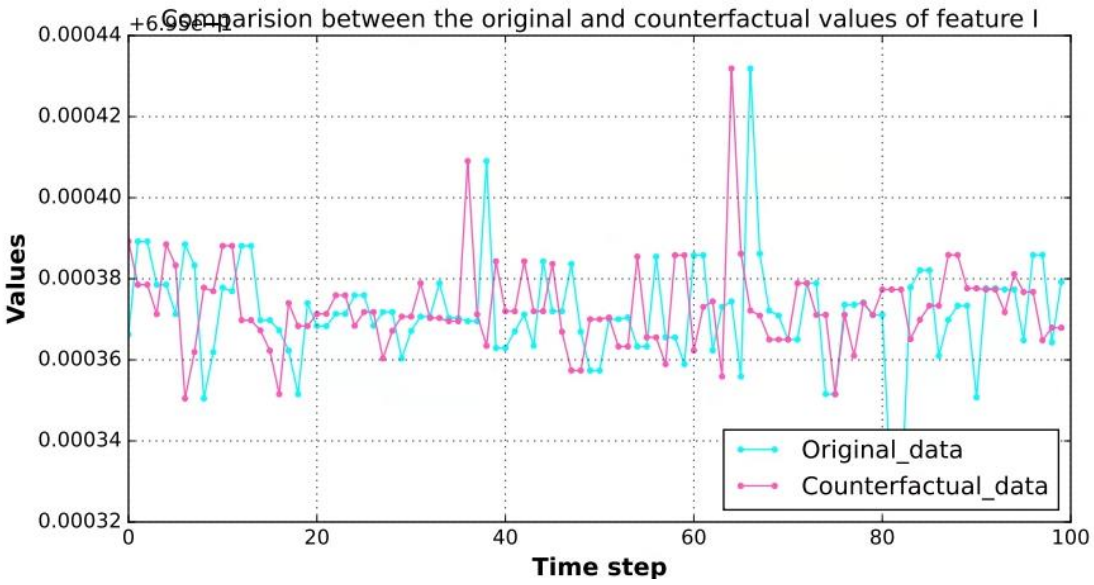
Charging with constant current (SOC near 100%)

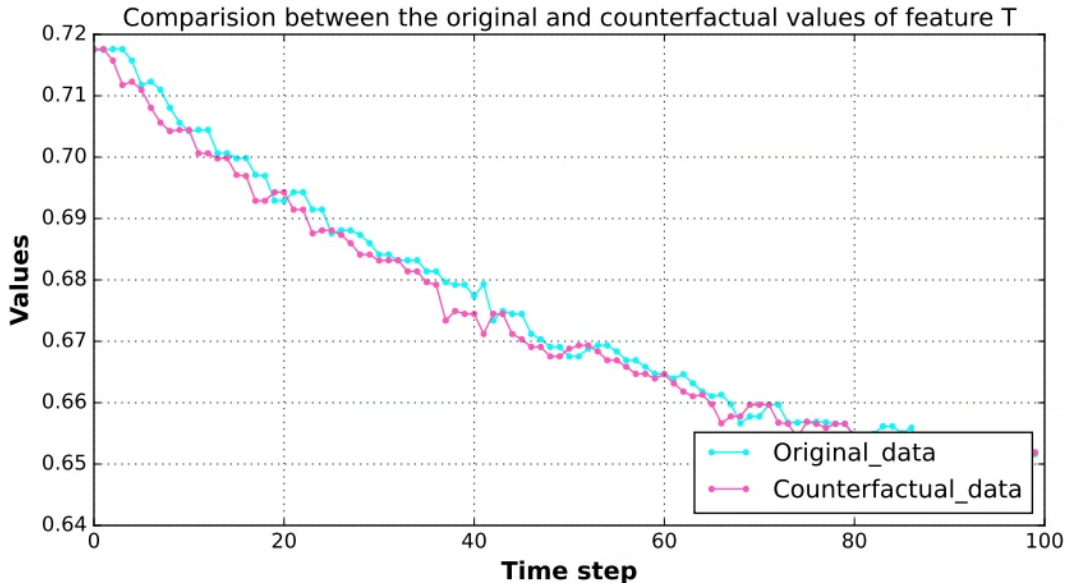
Geno-TOPSIS method

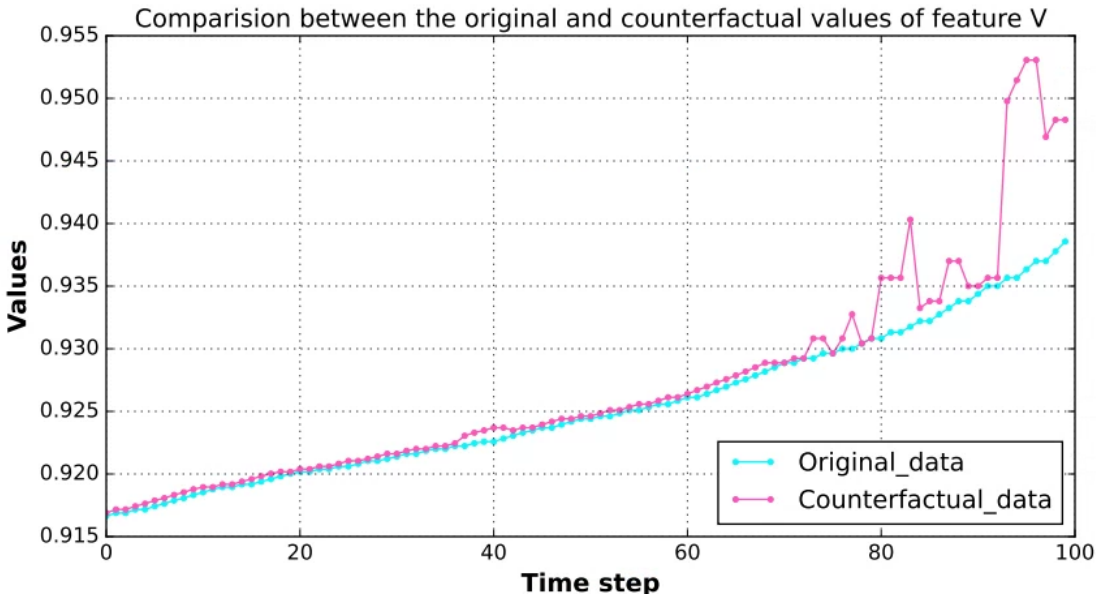
The method took 1085.39s to generate 20 counterfactuals based on a population of 100 individuals.

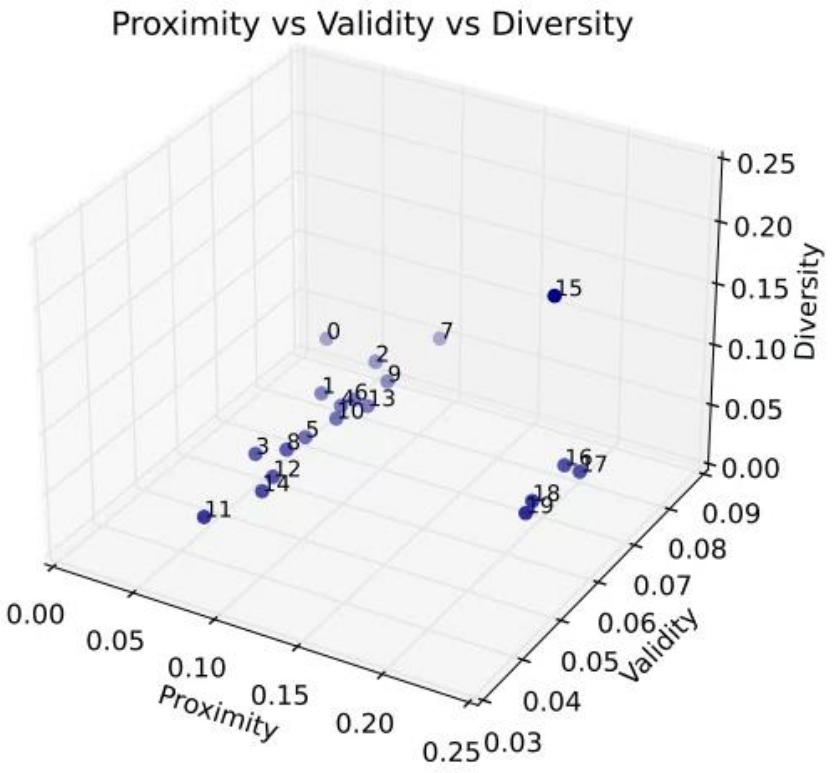


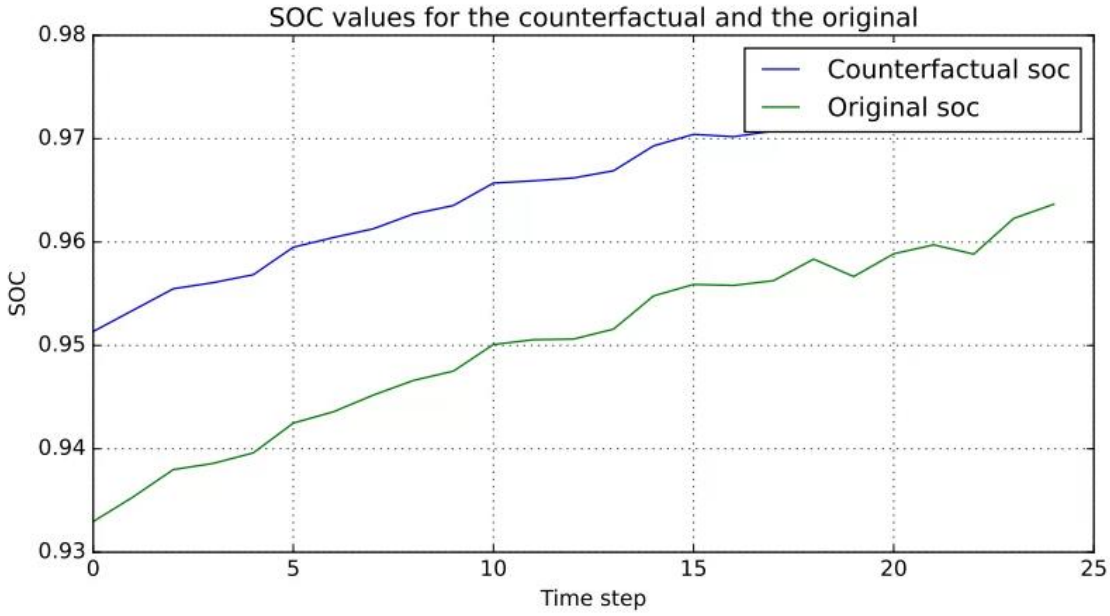



Comparision between the original and counterfactual values of feature I
+6.99e−12
0.00044
0.00042
0.00040
0.00038
0.00036
0.00034
0.00032
Values
0
20
40
60
80
100
Time step
Original_data
Counterfactual_data


Comparision between the original and counterfactual values of feature T
0.72
0.71
0.70
0.69
0.68
0.67
0.66
0.65
0.64
Values
0
20
40
60
80
100
Time step
Original_data
Counterfactual_data

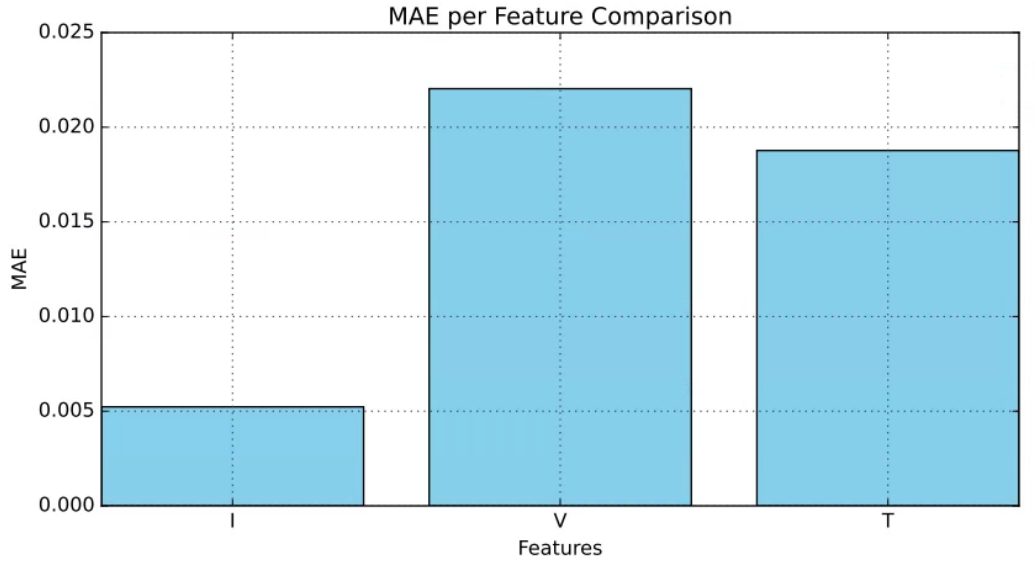

Comparision between the original and counterfactual values of feature V
0.955
0.950
0.945
0.940
0.935
0.930
0.925
0.920
0.915
Values
0
20
40
60
80
100
Time step
Original_data
Counterfactual_data

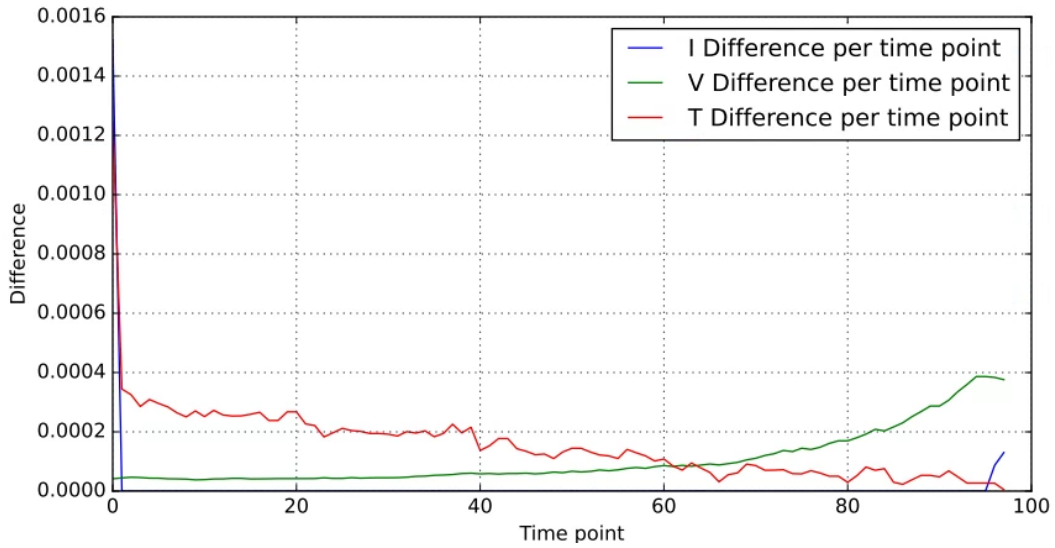

Proximity vs Validity vs Diversity
Proximity
Validity
Diversity
0.00
0.05
0.10
0.15
0.20
0.25
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.00
0.05
0.10
0.15
0.20
0.25
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

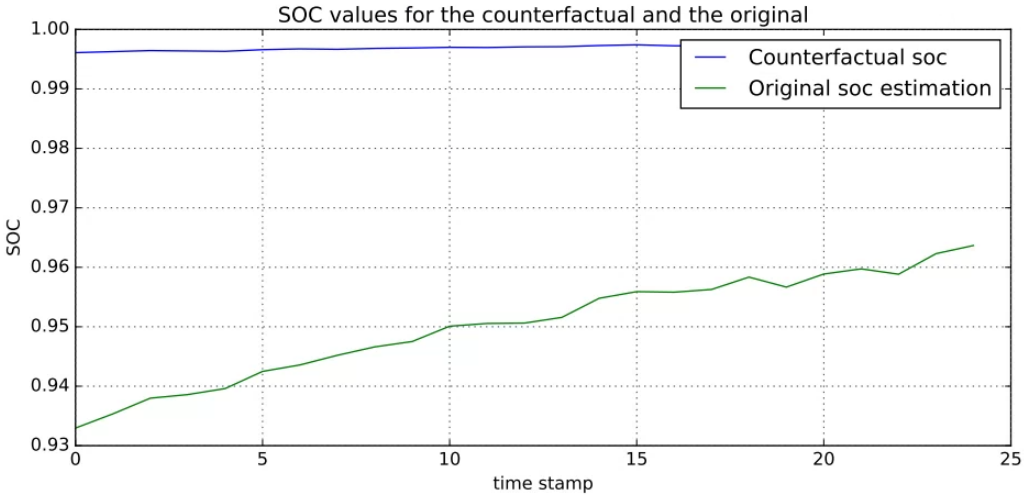

SOC values for the counterfactual and the original
Counterfactual soc
Original soc
SOC
Time step
0.93
0.94
0.95
0.96
0.97
0.98
0
5
10
15
20
25

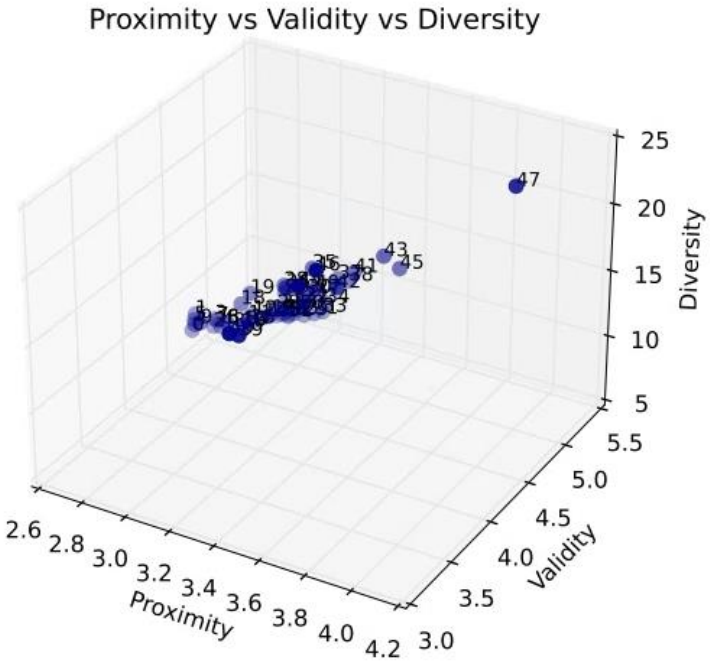
NSGA2 method

The method took **179s** to generate counterfactuals based on the same population





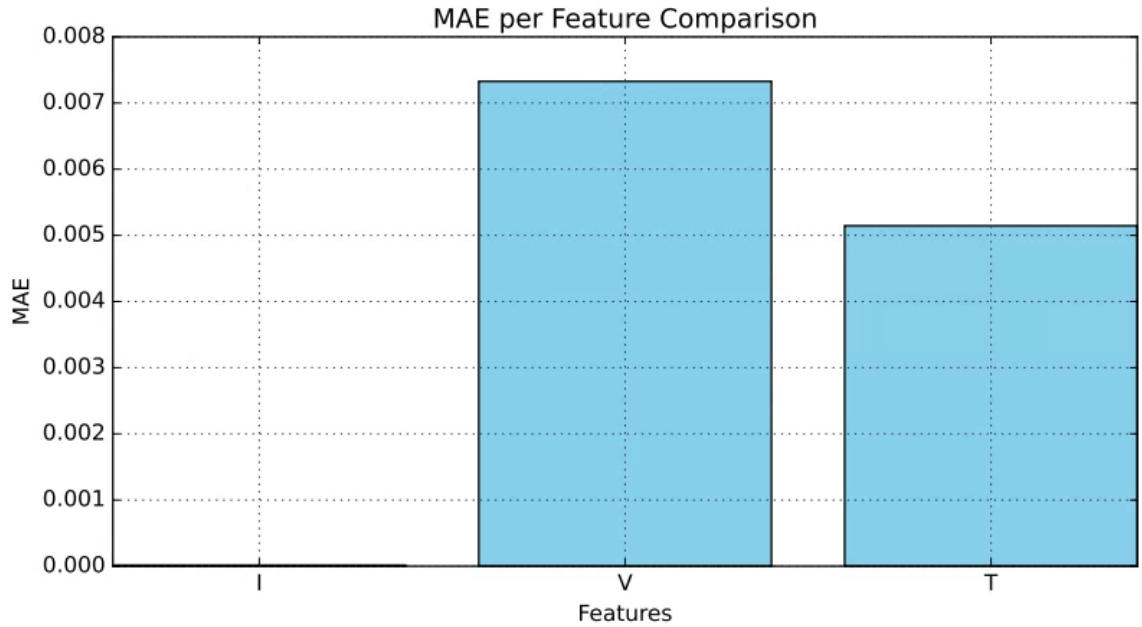


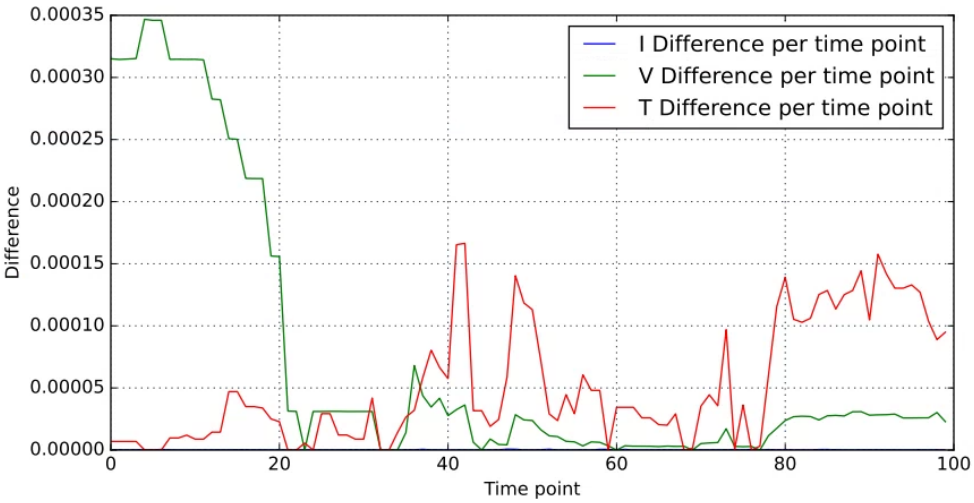


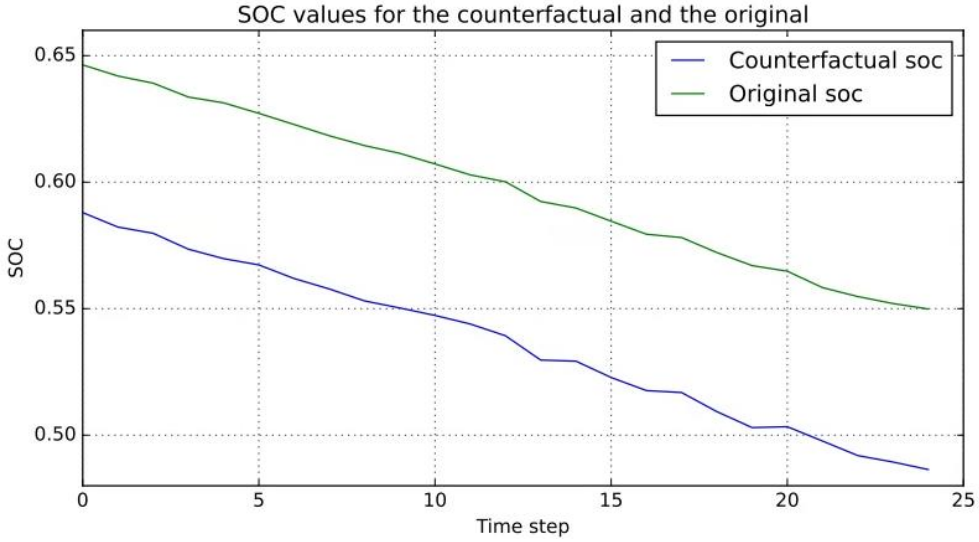
Discharging with constant current (SOC between 20% and 80%)

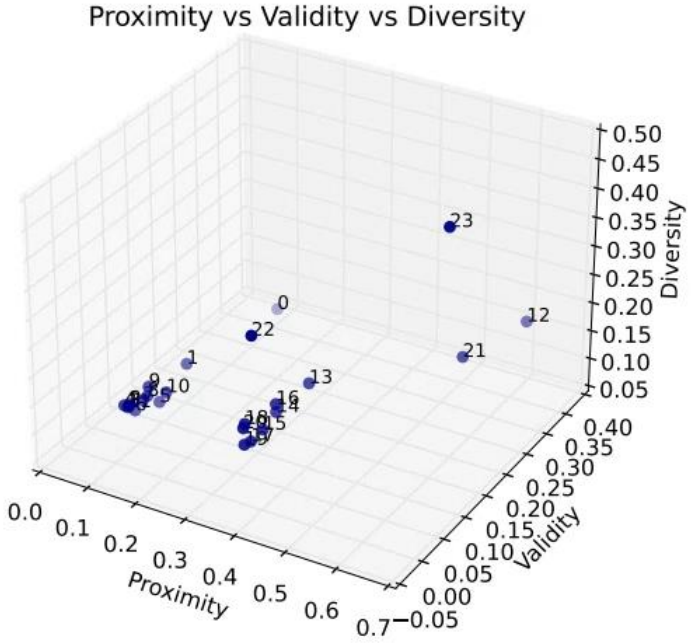
Geno-TOPSIS method

The method took 1203.07s to generate counterfactuals.



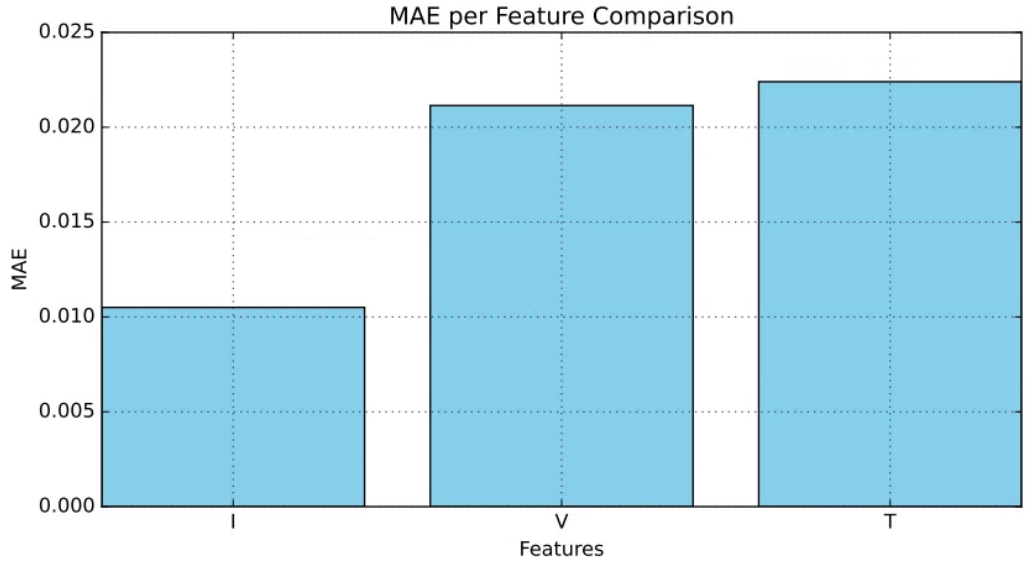

I Difference per time point
V Difference per time point
T Difference per time point
Difference
Time point
0.00035
0.00030
0.00025
0.00020
0.00015
0.00010
0.00005
0.00000
0
20
40
60
80
100

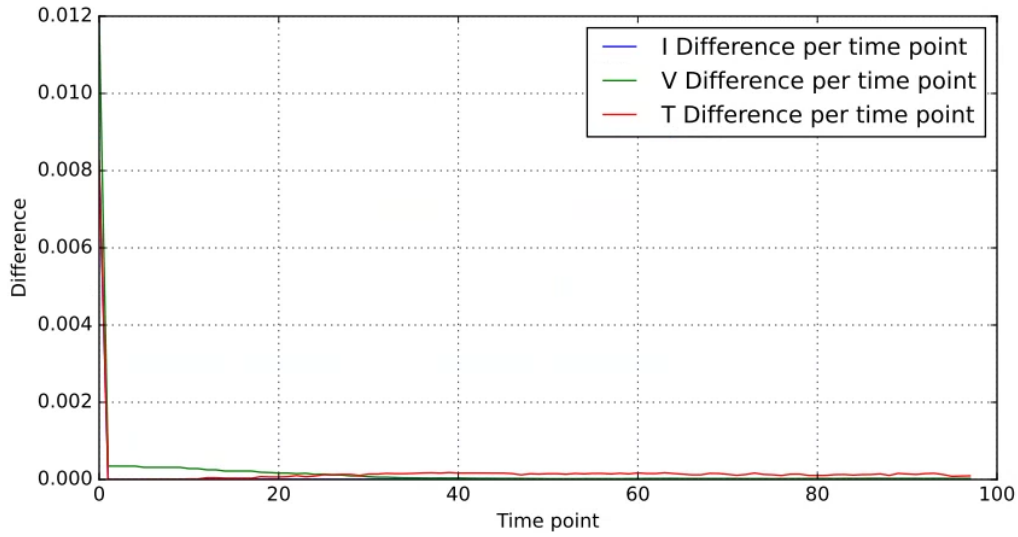

SOC values for the counterfactual and the original
Counterfactual soc
Original soc
SOC
Time step
0.65
0.60
0.55
0.50
0
5
10
15
20
25

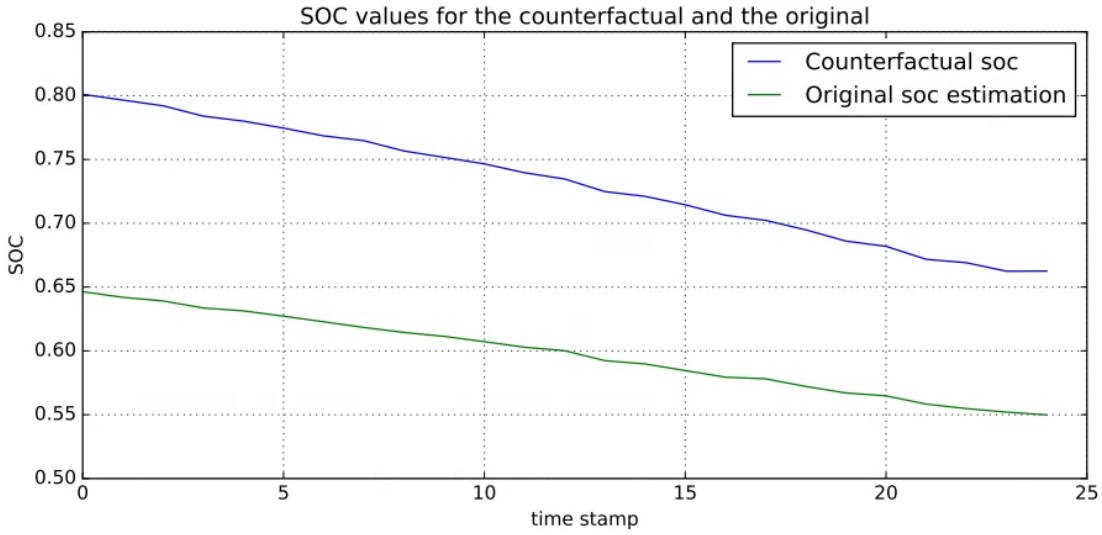

Proximity vs Validity vs Diversity
Proximity
Validity
Diversity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
−0.05
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50

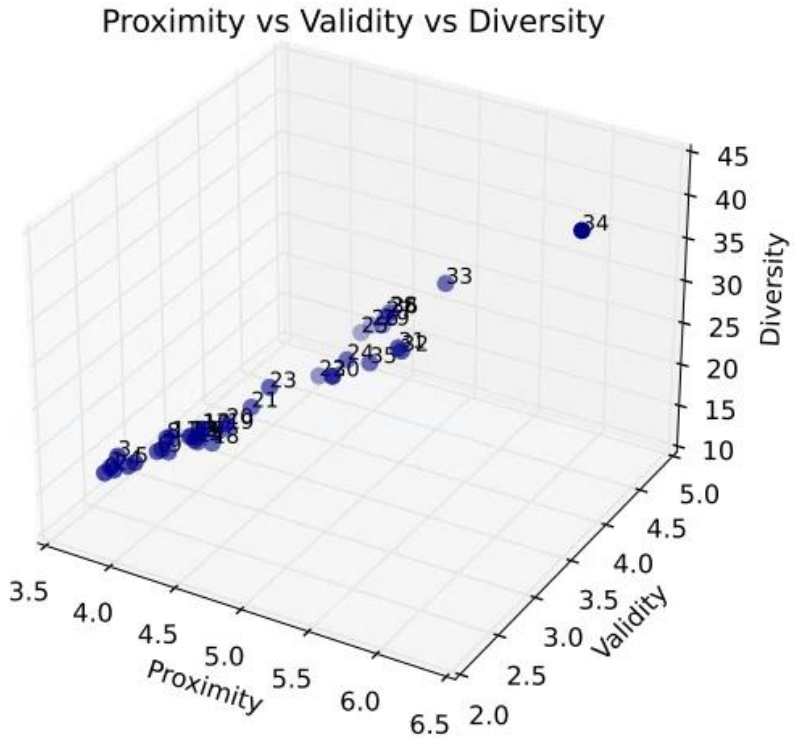
NSGA2 method

The method took 81s to generate counterfactuals.









Conclusion

After analyzing the results obtained from the two methods using different data points, we observed that the NSGA-II method outperforms the GENO-TOPSIS method in terms of speed. NSGA-II generates multiple counterfactual explanations in a shorter time, although these explanations are not as refined as those produced by the GENO-TOPSIS method. However, they still capture the overall behavior and, generally, offer similar explanations.

Regarding the results, we concluded that our model does not take into account the current values during both charging and discharging phases with constant current. In this scenario, the current feature was not altered to influence the prediction, suggesting that it was not significant for estimating the State of Charge (SoC) for the given data points.

A second observation was that the model assigns more importance to temperature and voltage during charging and discharging with constant current, particularly focusing on temperature when the SoC is near 100%.

A third observation is that the model gives more weight to current values when discharging with non-constant current or during the early stages of charging.

These explanations are consistent with the physical behavior of the system. When charging with constant current, only the voltage and temperature change significantly. This indicates that the model recognized that for nearly identical current values, the SoC varied, leading it to deprioritize the current as a predictive feature. Additionally, when charging or discharging with constant current, voltage has a more pronounced effect compared to temperature. This can be explained by examining the open-circuit voltage (OCV) curve in relation to SoC.

Specifically, as the voltage approaches its maximum value, the SoC rises, a relationship that was reflected in the counterfactual explanations. One suggestion, based on these explanations, is to increase the voltage to achieve different, higher SoC values compared to previous predictions (as shown in the figure depicting the first 100 values during constant-current charging).