CPT_S 580 HW1

Yang Zhang

11529139

1.

Search space #1

Initial State: None (k empty slot for m numbers of clusters)

Successor State Function: For k times, each time generates all valid combination of clusters differs by 1 bit at k_{th} slot

Terminal State Function: correct combination of clusters

Pros: The computation cost is low

Cons: Since it does not explore all possible combinations which could stuck on local optimal candidate. Need more restarts to train an accurate heuristic function

Search space #2

Initial State: None (k empty slot for each m numbers of clusters)

Successor State Function: All valid combination of clusters differs by 1 bit

Terminal State Function: correct combination of clusters

Pros: This search space explores more candidates than 1st search space within reasonable computation time

Cons: Since it does not explore all possible combinations which could stuck on local optimal candidate.

Search space #3

Initial State: None (k empty slot for each m numbers of clusters)

Successor State Function: All valid combination of clusters

Terminal State Function: correct combination of clusters

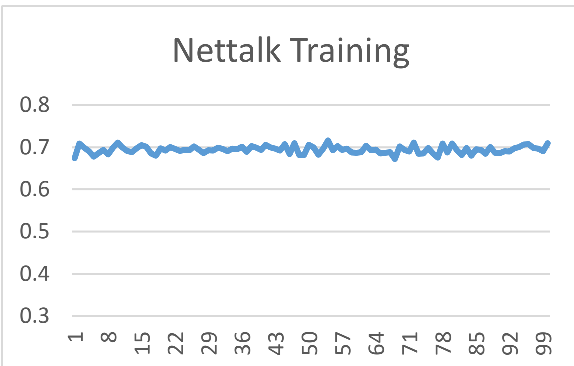
Pros: This search space explores all possible combinations which guarantee to find the globe optimal candidate. Thus, it trains the best heuristic function over the 3 search spaces.

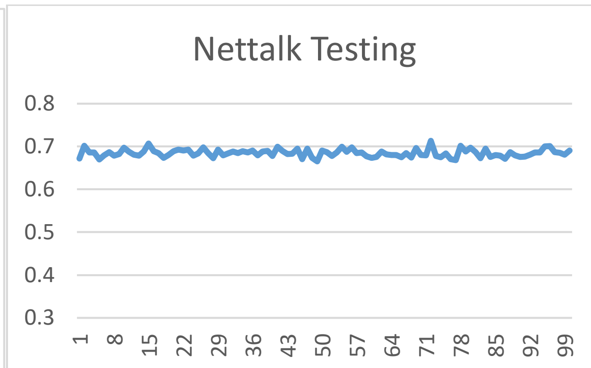
Cons: The computation cost is the most expensive over the 3 search spaces, so that this search space is not practical for structure prediction of high dimensional data

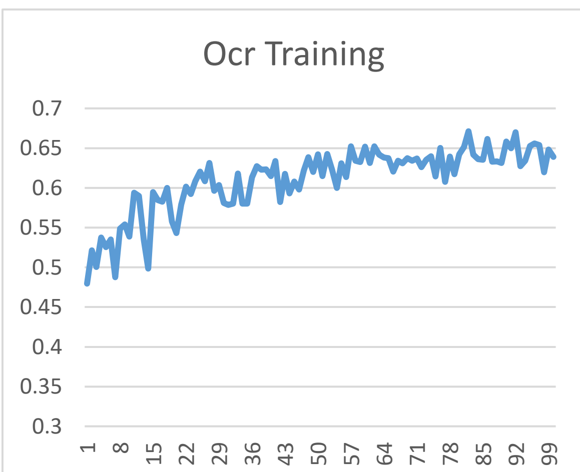
2.

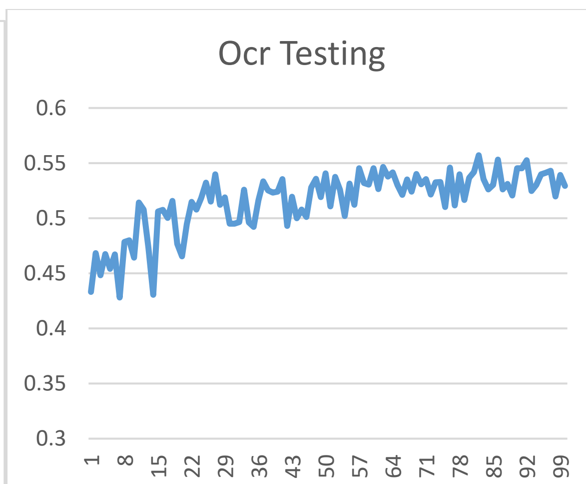
(a) Implemented in Java

(b)



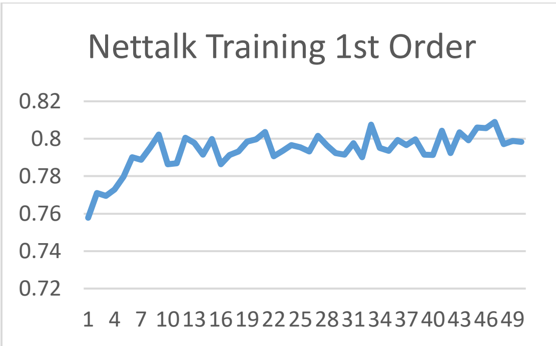


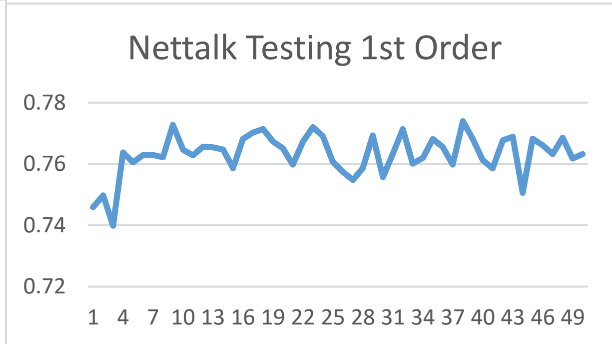


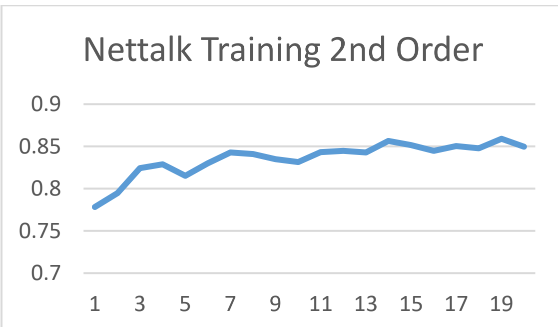


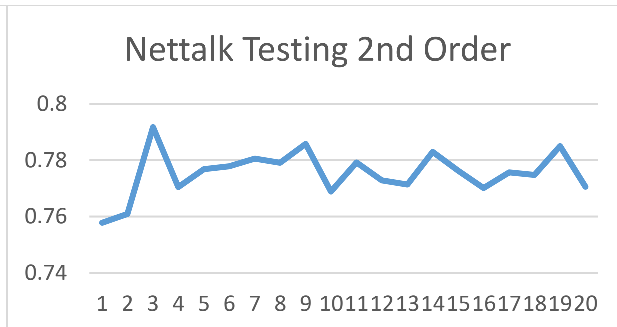
(c)

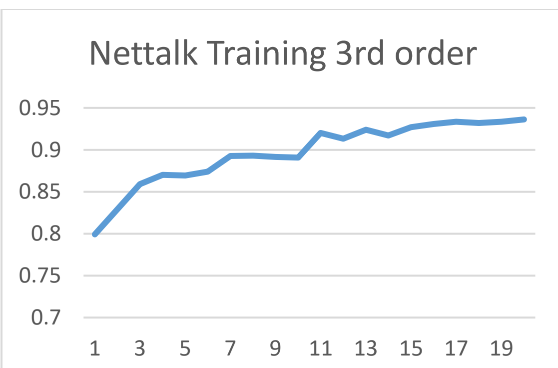
Nettalk dataset

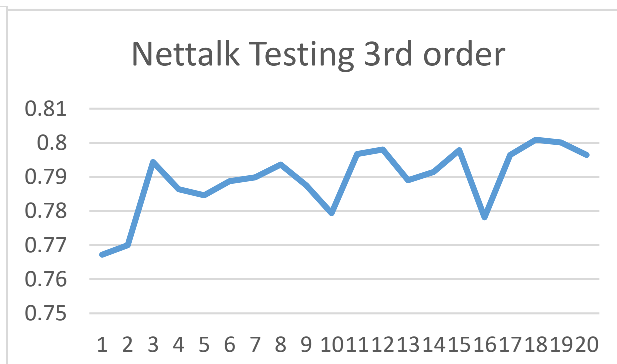




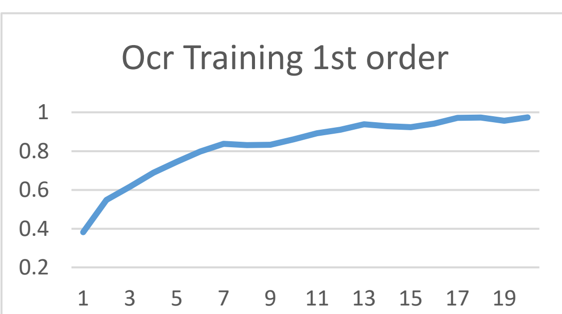


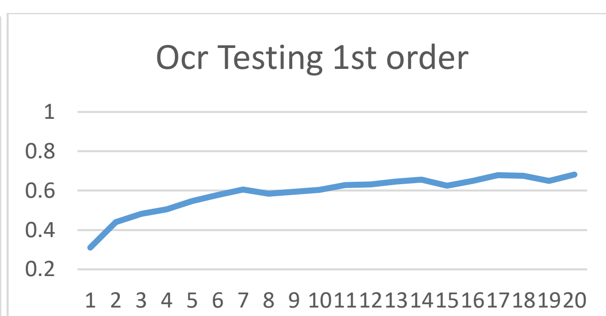


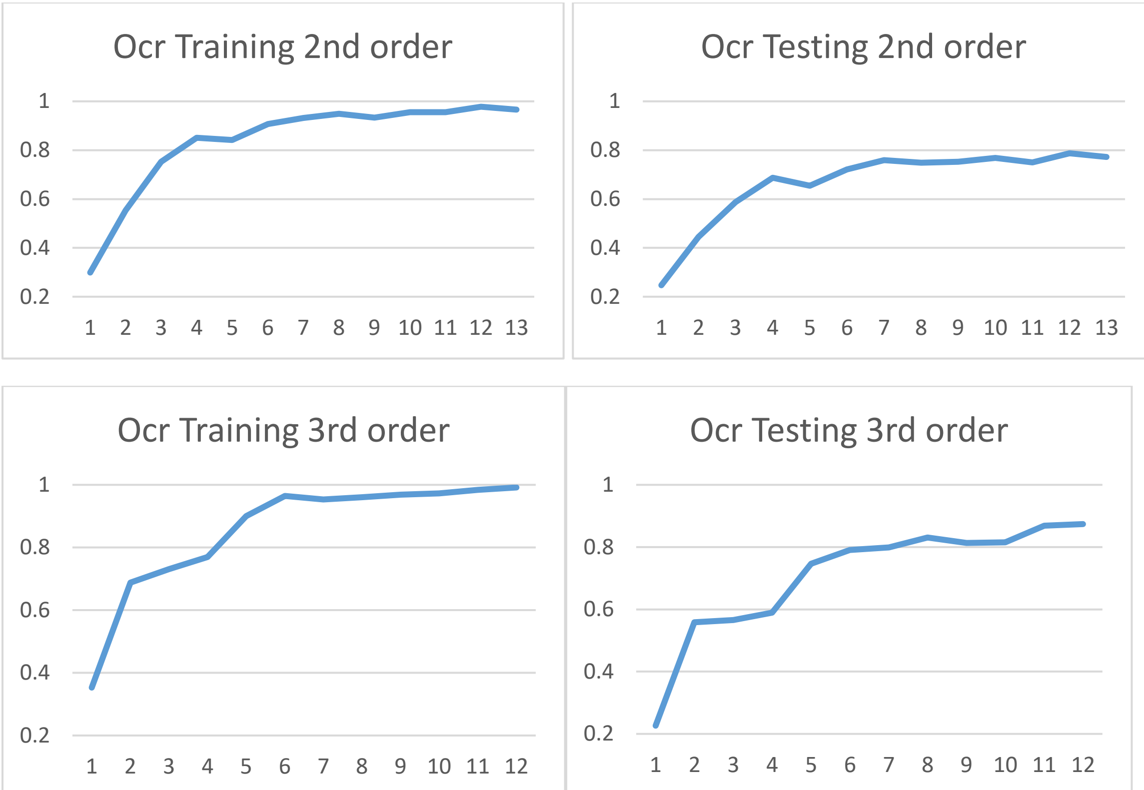




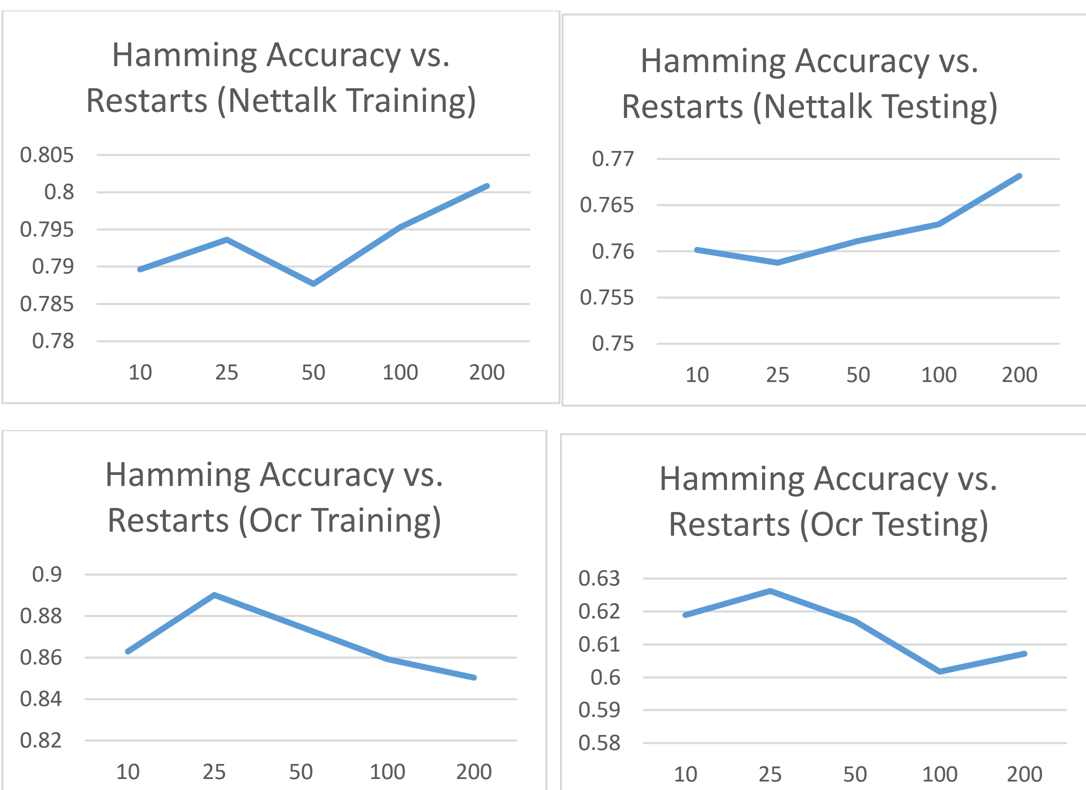
Ocr dataset






Ocr Training 2nd order
1
0.8
0.6
0.4
0.2
1 2 3 4 5 6 7 8 9 10 11 12 13
Ocr Testing 2nd order
1
0.8
0.6
0.4
0.2
1 2 3 4 5 6 7 8 9 10 11 12 13
Ocr Training 3rd order
1
0.8
0.6
0.4
0.2
1 2 3 4 5 6 7 8 9 10 11 12
Ocr Testing 3rd order
1
0.8
0.6
0.4
0.2
1 2 3 4 5 6 7 8 9 10 11 12

(d)


Hamming Accuracy vs. Restarts (Nettalk Training)
0.805
0.8
0.795
0.79
0.785
0.78
10 25 50 100 200
Hamming Accuracy vs. Restarts (Nettalk Testing)
0.77
0.765
0.76
0.755
0.75
10 25 50 100 200
Hamming Accuracy vs. Restarts (Ocr Training)
0.9
0.88
0.86
0.84
0.82
10 25 50 100 200
Hamming Accuracy vs. Restarts (Ocr Testing)
0.63
0.62
0.61
0.6
0.59
0.58
10 25 50 100 200

(e)

1. Accuracy

The high performance of learning algorithms should have high accuracy with reasonable training and classifying time. In other word, it should predict more accurately compared to other algorithms in the data settings.

Test:

Structured Perceptron with 1^{st} order feature function (Ocr fold0):

97.48% / 1.12 hours (20 iterations)

Standard Multi-class Perceptron (Ocr fold0)

67.73% / 0.38 hour (50 iterations)

2. Generalization

The high performance of learning algorithms should behave more generalized. In other word, it should predict more accurately compared to other algorithms in different data settings from training set.

Test:

Structured Perceptron with 3^{rd} order feature function (Trained by Ocr fold0, tested by Ocr fold1):

81.51% (20 iterations)

Standard Multi-class Perceptron (Trained by Ocr fold0, tested by Ocr fold1)

59.26% (50 iterations)

Structured Perceptron is better than standard multi-class perceptron.