alpha = 0.05

training 1 epoch...

144.09740543365479 seconds used......

1 training finished! average train loss: 0.640128131149805

total clf loss: 113.07978699356318 total adv loss: 72.35753134265542

evaluating...

Confusion matrix, without normalization

[[ 217 45 32]

[ 770 2894 157]

[ 34 13 795]]

Normalized confusion matrix

[[0.73809524 0.15306122 0.10884354]

[0.20151793 0.75739335 0.04108872]

[0.04038005 0.01543943 0.94418052]]





precision recall f1-score support

hate 0.21 0.74 0.33 294

offensive 0.98 0.76 0.85 3821

neither 0.81 0.94 0.87 842

accuracy 0.79 4957

macro avg 0.67 0.81 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.5198948620668342, accuracy: 0.7879765987492435

training 2 epoch...

143.43137454986572 seconds used......

2 training finished! average train loss: 0.369424601134501

total clf loss: 65.66319220513105 total adv loss: 49.83170786872506

evaluating...

Confusion matrix, without normalization

[[ 213 66 15]

[ 480 3236 105]

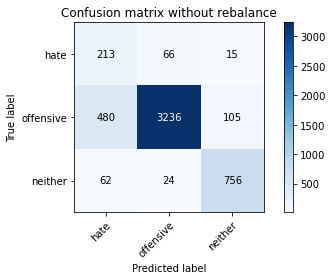
[ 62 24 756]]

Normalized confusion matrix

[[0.7244898 0.2244898 0.05102041]

[0.12562157 0.84689872 0.02747972]

[0.0736342 0.02850356 0.89786223]]





precision recall f1-score support

hate 0.28 0.72 0.41 294

offensive 0.97 0.85 0.91 3821

neither 0.86 0.90 0.88 842

accuracy 0.85 4957

macro avg 0.71 0.82 0.73 4957

weighted avg 0.91 0.85 0.87 4957

average eval\_loss: 0.3904154956220619, accuracy: 0.8482953399233407

training 3 epoch...

143.6557595729828 seconds used......

3 training finished! average train loss: 0.2518852747192508

total clf loss: 44.96864799037576 total adv loss: 37.925322545692325

evaluating...

Confusion matrix, without normalization

[[ 223 54 17]

[ 674 2949 198]

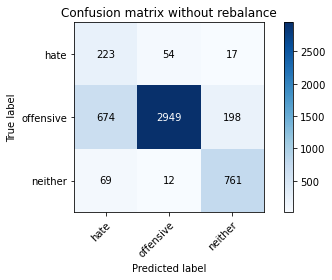
[ 69 12 761]]

Normalized confusion matrix

[[0.7585034 0.18367347 0.05782313]

[0.17639361 0.77178749 0.0518189 ]

[0.08194774 0.01425178 0.90380048]]





precision recall f1-score support

hate 0.23 0.76 0.35 294

offensive 0.98 0.77 0.86 3821

neither 0.78 0.90 0.84 842

accuracy 0.79 4957

macro avg 0.66 0.81 0.68 4957

weighted avg 0.90 0.79 0.83 4957

average eval\_loss: 0.5520728193223476, accuracy: 0.7934234415977406

training 4 epoch...

143.53269982337952 seconds used......

4 training finished! average train loss: 0.14523739771296582

total clf loss: 26.07090855203569 total adv loss: 24.706270437687635

evaluating...

Confusion matrix, without normalization

[[ 222 53 19]

[ 909 2632 280]

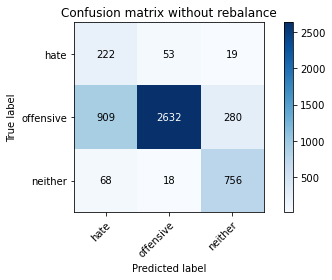
[ 68 18 756]]

Normalized confusion matrix

[[0.75510204 0.18027211 0.06462585]

[0.23789584 0.68882491 0.07327925]

[0.0807601 0.02137767 0.89786223]]





precision recall f1-score support

hate 0.19 0.76 0.30 294

offensive 0.97 0.69 0.81 3821

neither 0.72 0.90 0.80 842

accuracy 0.73 4957

macro avg 0.63 0.78 0.63 4957

weighted avg 0.88 0.73 0.77 4957

average eval\_loss: 0.73737922004394, accuracy: 0.7282630623360904

training 5 epoch...

143.56254696846008 seconds used......

5 training finished! average train loss: 0.08625960512881914

total clf loss: 15.471795736812055 total adv loss: 14.428065767278895

evaluating...

Confusion matrix, without normalization

[[ 192 87 15]

[ 470 3268 83]

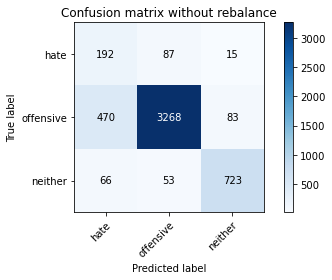
[ 66 53 723]]

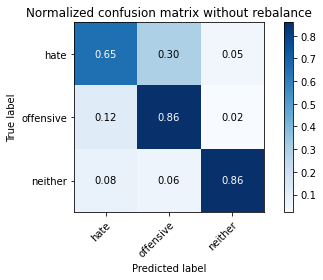
Normalized confusion matrix

[[0.65306122 0.29591837 0.05102041]

[0.12300445 0.85527349 0.02172206]

[0.0783848 0.06294537 0.85866983]]





precision recall f1-score support

hate 0.26 0.65 0.38 294

offensive 0.96 0.86 0.90 3821

neither 0.88 0.86 0.87 842

accuracy 0.84 4957

macro avg 0.70 0.79 0.72 4957

weighted avg 0.90 0.84 0.87 4957

average eval\_loss: 0.5833481534585101, accuracy: 0.8438571716764172