training 1 epoch...

14.951589345932007 seconds used......

1 training finished! average train loss: 0.9972400739789009

total clf loss: 57.53607892990112 total adv loss: 33.81269791722298

evaluating...

Confusion matrix, without normalization

[[ 24 8 13]

[ 58 96 32]

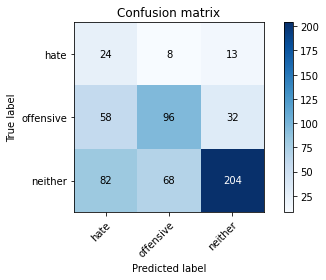
[ 82 68 204]]

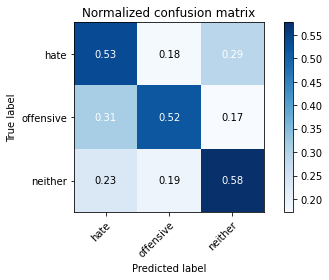
Normalized confusion matrix

[[0.53333333 0.17777778 0.28888889]

[0.31182796 0.51612903 0.17204301]

[0.23163842 0.1920904 0.57627119]]





precision recall f1-score support

hate 0.15 0.53 0.23 45

offensive 0.56 0.52 0.54 186

neither 0.82 0.58 0.68 354

accuracy 0.55 585

macro avg 0.51 0.54 0.48 585

weighted avg 0.68 0.55 0.60 585

average eval\_loss: 0.898038367430369, accuracy: 0.5538461538461539

training 2 epoch...

14.01340627670288 seconds used......

2 training finished! average train loss: 0.7920795072402272

total clf loss: 45.72576147317886 total adv loss: 27.386181980371475

evaluating...

Confusion matrix, without normalization

[[ 12 18 15]

[ 8 150 28]

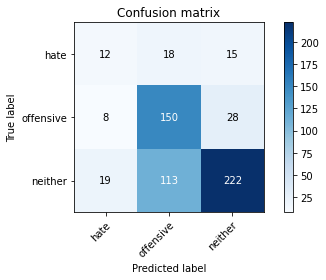
[ 19 113 222]]

Normalized confusion matrix

[[0.26666667 0.4 0.33333333]

[0.04301075 0.80645161 0.15053763]

[0.05367232 0.31920904 0.62711864]]





precision recall f1-score support

hate 0.31 0.27 0.29 45

offensive 0.53 0.81 0.64 186

neither 0.84 0.63 0.72 354

accuracy 0.66 585

macro avg 0.56 0.57 0.55 585

weighted avg 0.70 0.66 0.66 585

average eval\_loss: 0.7274863104025523, accuracy: 0.6564102564102564

training 3 epoch...

14.02806568145752 seconds used......

3 training finished! average train loss: 0.5545195979731423

total clf loss: 32.00048825144768 total adv loss: 18.94781055301428

evaluating...

Confusion matrix, without normalization

[[ 31 4 10]

[ 35 105 46]

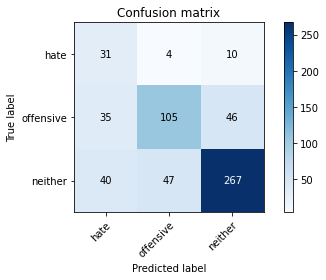
[ 40 47 267]]

Normalized confusion matrix

[[0.68888889 0.08888889 0.22222222]

[0.18817204 0.56451613 0.24731183]

[0.11299435 0.13276836 0.75423729]]





precision recall f1-score support

hate 0.29 0.69 0.41 45

offensive 0.67 0.56 0.61 186

neither 0.83 0.75 0.79 354

accuracy 0.69 585

macro avg 0.60 0.67 0.60 585

weighted avg 0.74 0.69 0.70 585

average eval\_loss: 0.741886114080747, accuracy: 0.6888888888888889

training 4 epoch...

13.998220682144165 seconds used......

4 training finished! average train loss: 0.30346050459359375

total clf loss: 17.41801557689905 total adv loss: 8.48454727139324

evaluating...

Confusion matrix, without normalization

[[ 31 6 8]

[ 45 101 40]

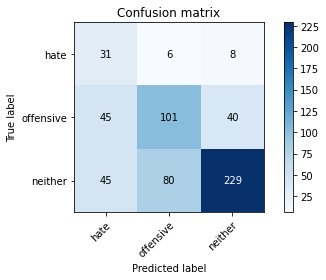
[ 45 80 229]]

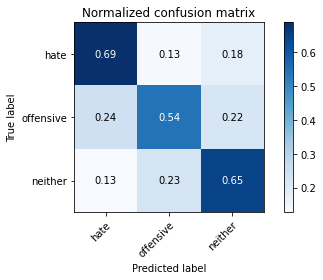
Normalized confusion matrix

[[0.68888889 0.13333333 0.17777778]

[0.24193548 0.54301075 0.21505376]

[0.12711864 0.2259887 0.64689266]]





precision recall f1-score support

hate 0.26 0.69 0.37 45

offensive 0.54 0.54 0.54 186

neither 0.83 0.65 0.73 354

accuracy 0.62 585

macro avg 0.54 0.63 0.55 585

weighted avg 0.69 0.62 0.64 585

average eval\_loss: 1.03420902689298, accuracy: 0.6170940170940171

training 5 epoch...

14.061510562896729 seconds used......

5 training finished! average train loss: 0.18805999957424188

total clf loss: 10.816976990550756 total adv loss: 5.71234074793756

evaluating...

Confusion matrix, without normalization

[[ 34 4 7]

[ 63 85 38]

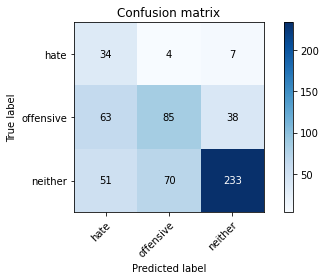
[ 51 70 233]]

Normalized confusion matrix

[[0.75555556 0.08888889 0.15555556]

[0.33870968 0.45698925 0.20430108]

[0.1440678 0.19774011 0.65819209]]





precision recall f1-score support

hate 0.23 0.76 0.35 45

offensive 0.53 0.46 0.49 186

neither 0.84 0.66 0.74 354

accuracy 0.60 585

macro avg 0.53 0.62 0.53 585

weighted avg 0.69 0.60 0.63 585

average eval\_loss: 1.2000312626361846, accuracy: 0.6017094017094017

Confusion matrix, without normalization

[[ 40 6 5]

[ 53 78 38]

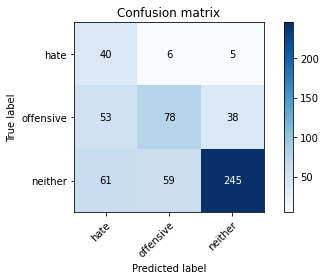
[ 61 59 245]]

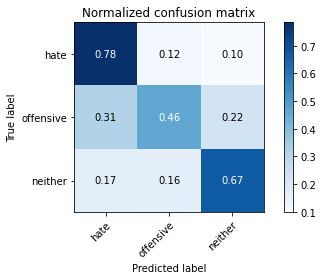
Normalized confusion matrix

[[0.78431373 0.11764706 0.09803922]

[0.31360947 0.46153846 0.22485207]

[0.16712329 0.16164384 0.67123288]]





precision recall f1-score support

hate 0.26 0.78 0.39 51

offensive 0.55 0.46 0.50 169

neither 0.85 0.67 0.75 365

accuracy 0.62 585

macro avg 0.55 0.64 0.55 585

weighted avg 0.71 0.62 0.65 585

average eval\_loss: 1.2472136517365773, accuracy: 0.6205128205128205

[21]:

0.6205128205128205

2 run

training 1 epoch...

13.984641790390015 seconds used......

1 training finished! average train loss: 1.0261296268020357

total clf loss: 59.19437676668167 total adv loss: 34.6223518550396

evaluating...

Confusion matrix, without normalization

[[ 27 4 14]

[ 63 90 33]

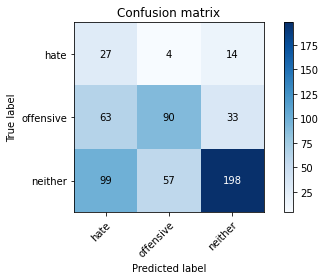
[ 99 57 198]]

Normalized confusion matrix

[[0.6 0.08888889 0.31111111]

[0.33870968 0.48387097 0.17741935]

[0.27966102 0.16101695 0.55932203]]





precision recall f1-score support

hate 0.14 0.60 0.23 45

offensive 0.60 0.48 0.53 186

neither 0.81 0.56 0.66 354

accuracy 0.54 585

macro avg 0.52 0.55 0.48 585

weighted avg 0.69 0.54 0.59 585

average eval\_loss: 0.9280134538809458, accuracy: 0.5384615384615384

training 2 epoch...

14.01903247833252 seconds used......

2 training finished! average train loss: 0.8392810693808964

total clf loss: 48.51607495546341 total adv loss: 30.32670173048973

evaluating...

Confusion matrix, without normalization

[[ 38 3 4]

[ 66 101 19]

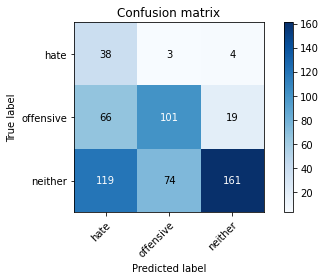
[119 74 161]]

Normalized confusion matrix

[[0.84444444 0.06666667 0.08888889]

[0.35483871 0.54301075 0.10215054]

[0.33615819 0.20903955 0.45480226]]





precision recall f1-score support

hate 0.17 0.84 0.28 45

offensive 0.57 0.54 0.55 186

neither 0.88 0.45 0.60 354

accuracy 0.51 585

macro avg 0.54 0.61 0.48 585

weighted avg 0.72 0.51 0.56 585

average eval\_loss: 1.0068845391273498, accuracy: 0.5128205128205128

training 3 epoch...

13.991909503936768 seconds used......

3 training finished! average train loss: 0.572927997048412

total clf loss: 33.156558364629745 total adv loss: 21.451810091733932

evaluating...

Confusion matrix, without normalization

[[ 30 6 9]

[ 39 121 26]

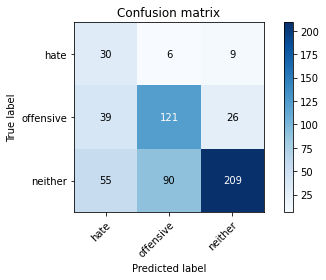
[ 55 90 209]]

Normalized confusion matrix

[[0.66666667 0.13333333 0.2 ]

[0.20967742 0.65053763 0.13978495]

[0.15536723 0.25423729 0.59039548]]





precision recall f1-score support

hate 0.24 0.67 0.36 45

offensive 0.56 0.65 0.60 186

neither 0.86 0.59 0.70 354

accuracy 0.62 585

macro avg 0.55 0.64 0.55 585

weighted avg 0.71 0.62 0.64 585

average eval\_loss: 0.8701513290405274, accuracy: 0.6153846153846154

training 4 epoch...

14.035992860794067 seconds used......

4 training finished! average train loss: 0.33304612657853533

total clf loss: 19.246256321668625 total adv loss: 11.913464102894068

evaluating...

Confusion matrix, without normalization

[[ 18 10 17]

[ 11 138 37]

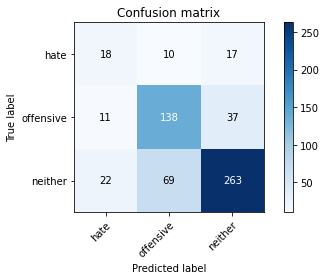
[ 22 69 263]]

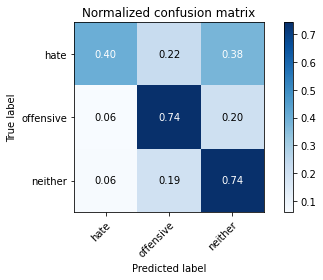
Normalized confusion matrix

[[0.4 0.22222222 0.37777778]

[0.05913978 0.74193548 0.19892473]

[0.06214689 0.19491525 0.74293785]]





precision recall f1-score support

hate 0.35 0.40 0.38 45

offensive 0.64 0.74 0.68 186

neither 0.83 0.74 0.78 354

accuracy 0.72 585

macro avg 0.61 0.63 0.61 585

weighted avg 0.73 0.72 0.72 585

average eval\_loss: 0.800792033970356, accuracy: 0.7162393162393162

training 5 epoch...

14.006093263626099 seconds used......

5 training finished! average train loss: 0.15465948414722724

total clf loss: 8.934898000210524 total adv loss: 5.479338398668915

evaluating...

Confusion matrix, without normalization

[[ 25 9 11]

[ 36 125 25]

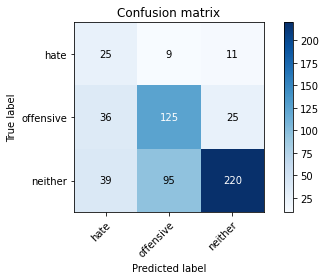
[ 39 95 220]]

Normalized confusion matrix

[[0.55555556 0.2 0.24444444]

[0.19354839 0.67204301 0.1344086 ]

[0.11016949 0.26836158 0.62146893]]





precision recall f1-score support

hate 0.25 0.56 0.34 45

offensive 0.55 0.67 0.60 186

neither 0.86 0.62 0.72 354

accuracy 0.63 585

macro avg 0.55 0.62 0.56 585

weighted avg 0.71 0.63 0.65 585

average eval\_loss: 1.2141308387120564, accuracy: 0.6324786324786325

Confusion matrix, without normalization

[[ 43 6 2]

[ 77 74 18]

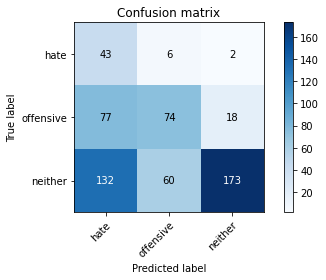
[132 60 173]]

Normalized confusion matrix

[[0.84313725 0.11764706 0.03921569]

[0.4556213 0.43786982 0.10650888]

[0.36164384 0.16438356 0.4739726 ]]





precision recall f1-score support

hate 0.17 0.84 0.28 51

offensive 0.53 0.44 0.48 169

neither 0.90 0.47 0.62 365

accuracy 0.50 585

macro avg 0.53 0.58 0.46 585

weighted avg 0.73 0.50 0.55 585

average eval\_loss: 1.0700242181619009, accuracy: 0.49572649572649574

[23]:

0.49572649572649574

3 run

training 1 epoch...

14.006263017654419 seconds used......

1 training finished! average train loss: 1.0048548887882913

total clf loss: 57.98495274782181 total adv loss: 34.26157999038696

evaluating...

Confusion matrix, without normalization

[[ 27 2 16]

[ 80 61 45]

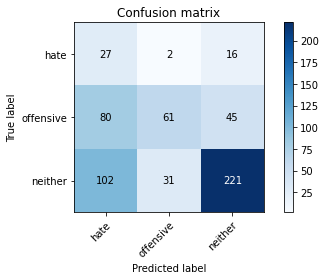
[102 31 221]]

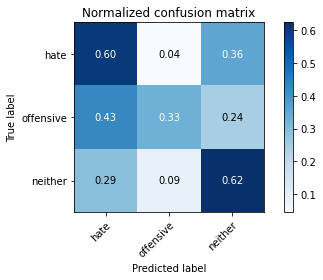
Normalized confusion matrix

[[0.6 0.04444444 0.35555556]

[0.43010753 0.32795699 0.24193548]

[0.28813559 0.08757062 0.62429379]]





precision recall f1-score support

hate 0.13 0.60 0.21 45

offensive 0.65 0.33 0.44 186

neither 0.78 0.62 0.69 354

accuracy 0.53 585

macro avg 0.52 0.52 0.45 585

weighted avg 0.69 0.53 0.58 585

average eval\_loss: 0.922165846824646, accuracy: 0.5282051282051282

training 2 epoch...

14.023190975189209 seconds used......

2 training finished! average train loss: 0.7768711160336222

total clf loss: 44.886098980903625 total adv loss: 27.62633091211319

evaluating...

Confusion matrix, without normalization

[[ 35 6 4]

[ 45 125 16]

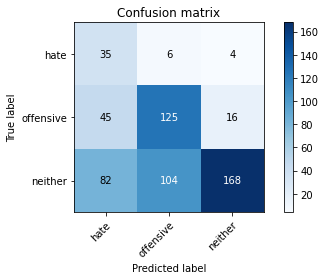
[ 82 104 168]]

Normalized confusion matrix

[[0.77777778 0.13333333 0.08888889]

[0.24193548 0.67204301 0.08602151]

[0.23163842 0.29378531 0.47457627]]





precision recall f1-score support

hate 0.22 0.78 0.34 45

offensive 0.53 0.67 0.59 186

neither 0.89 0.47 0.62 354

accuracy 0.56 585

macro avg 0.55 0.64 0.52 585

weighted avg 0.73 0.56 0.59 585

average eval\_loss: 0.9976392527421315, accuracy: 0.5606837606837607

training 3 epoch...

14.014744520187378 seconds used......

3 training finished! average train loss: 0.5736695653093713

total clf loss: 33.17366309463978 total adv loss: 20.963348422199488

evaluating...

Confusion matrix, without normalization

[[ 32 6 7]

[ 32 114 40]

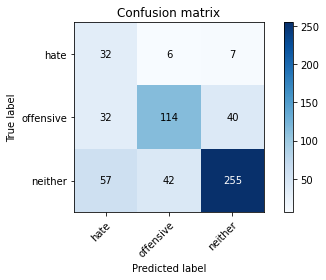
[ 57 42 255]]

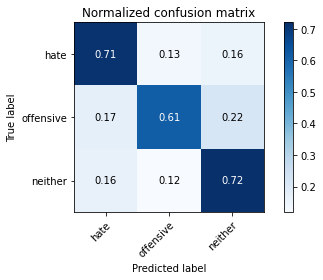
Normalized confusion matrix

[[0.71111111 0.13333333 0.15555556]

[0.17204301 0.61290323 0.21505376]

[0.16101695 0.11864407 0.72033898]]





precision recall f1-score support

hate 0.26 0.71 0.39 45

offensive 0.70 0.61 0.66 186

neither 0.84 0.72 0.78 354

accuracy 0.69 585

macro avg 0.60 0.68 0.61 585

weighted avg 0.76 0.69 0.71 585

average eval\_loss: 0.78087597489357, accuracy: 0.6854700854700855

training 4 epoch...

14.002295970916748 seconds used......

4 training finished! average train loss: 0.3014760895499161

total clf loss: 17.405365332961082 total adv loss: 10.45408833771944

evaluating...

Confusion matrix, without normalization

[[ 23 6 16]

[ 33 89 64]

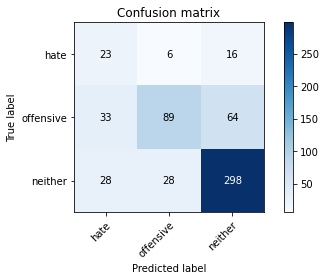
[ 28 28 298]]

Normalized confusion matrix

[[0.51111111 0.13333333 0.35555556]

[0.17741935 0.47849462 0.34408602]

[0.07909605 0.07909605 0.84180791]]





precision recall f1-score support

hate 0.27 0.51 0.36 45

offensive 0.72 0.48 0.58 186

neither 0.79 0.84 0.81 354

accuracy 0.70 585

macro avg 0.60 0.61 0.58 585

weighted avg 0.73 0.70 0.70 585

average eval\_loss: 0.8135318617026012, accuracy: 0.7008547008547008

training 5 epoch...

13.999972343444824 seconds used......

5 training finished! average train loss: 0.14545987632923893

total clf loss: 8.391971224918962 total adv loss: 4.924364150152542

evaluating...

Confusion matrix, without normalization

[[ 26 8 11]

[ 36 108 42]

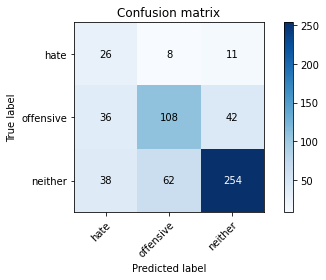
[ 38 62 254]]

Normalized confusion matrix

[[0.57777778 0.17777778 0.24444444]

[0.19354839 0.58064516 0.22580645]

[0.10734463 0.17514124 0.71751412]]





precision recall f1-score support

hate 0.26 0.58 0.36 45

offensive 0.61 0.58 0.59 186

neither 0.83 0.72 0.77 354

accuracy 0.66 585

macro avg 0.56 0.63 0.57 585

weighted avg 0.71 0.66 0.68 585

average eval\_loss: 1.0967057168483734, accuracy: 0.6632478632478632

Confusion matrix, without normalization

[[ 41 7 3]

[ 61 92 16]

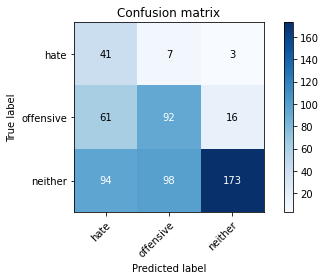
[ 94 98 173]]

Normalized confusion matrix

[[0.80392157 0.1372549 0.05882353]

[0.36094675 0.5443787 0.09467456]

[0.25753425 0.26849315 0.4739726 ]]





precision recall f1-score support

hate 0.21 0.80 0.33 51

offensive 0.47 0.54 0.50 169

neither 0.90 0.47 0.62 365

accuracy 0.52 585

macro avg 0.53 0.61 0.49 585

weighted avg 0.72 0.52 0.56 585

average eval\_loss: 1.048278065522512, accuracy: 0.5230769230769231