with adv

data without rebalance

training 1 epoch...

2626.75070810318 seconds used......

1 training finished! average train loss: 0.880371589526534

total clf loss: 1537.2518836557865 total adv loss: 551.9314425420016

evaluating...

Confusion matrix, without normalization

[[ 266 328 12 383]

[ 106 4908 58 377]

[ 2 152 1493 1155]

[ 37 275 895 9552]]

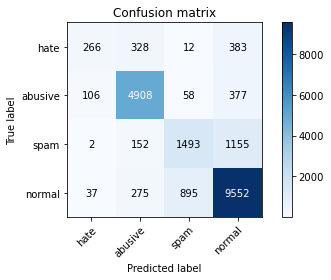
Normalized confusion matrix

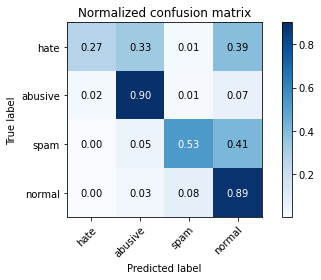
[[2.68958544e-01 3.31648129e-01 1.21334681e-02 3.87259858e-01]

[1.94531107e-02 9.00715728e-01 1.06441549e-02 6.91870068e-02]

[7.13775874e-04 5.42469665e-02 5.32833690e-01 4.12205567e-01]

[3.43898132e-03 2.55599963e-02 8.31861697e-02 8.87814853e-01]]





precision recall f1-score support

hate 0.65 0.27 0.38 989

abusive 0.87 0.90 0.88 5449

spam 0.61 0.53 0.57 2802

normal 0.83 0.89 0.86 10759

accuracy 0.81 19999

macro avg 0.74 0.65 0.67 19999

weighted avg 0.80 0.81 0.80 19999

average eval\_loss: 0.49171256756782533, accuracy: 0.8109905495274764

training 2 epoch...

2628.168970346451 seconds used......

2 training finished! average train loss: 0.7327349492410818

total clf loss: 1304.6899080500007 total adv loss: 446.75746997026727

evaluating...

Confusion matrix, without normalization

[[ 280 385 6 318]

[ 89 5052 15 293]

[ 6 185 916 1695]

[ 106 337 377 9939]]

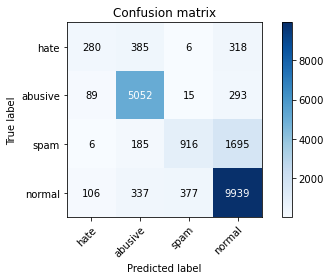
Normalized confusion matrix

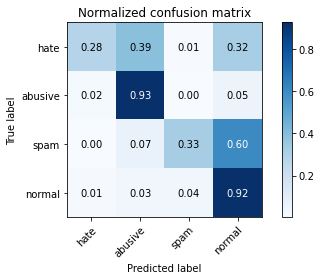
[[0.28311426 0.3892821 0.00606673 0.32153691]

[0.01633327 0.92714259 0.0027528 0.05377133]

[0.00214133 0.06602427 0.32690935 0.60492505]

[0.00985222 0.03132261 0.03504043 0.92378474]]





precision recall f1-score support

hate 0.58 0.28 0.38 989

abusive 0.85 0.93 0.89 5449

spam 0.70 0.33 0.45 2802

normal 0.81 0.92 0.86 10759

accuracy 0.81 19999

macro avg 0.73 0.62 0.64 19999

weighted avg 0.79 0.81 0.79 19999

average eval\_loss: 0.5022501390427351, accuracy: 0.8093904695234762

training 3 epoch...

2630.7173776626587 seconds used......

3 training finished! average train loss: 0.6067514602690935

total clf loss: 1098.9753400161862 total adv loss: 360.63952092954423

evaluating...

Confusion matrix, without normalization

[[ 422 301 12 254]

[ 206 4977 45 221]

[ 11 162 1448 1181]

[ 284 374 913 9188]]

Normalized confusion matrix

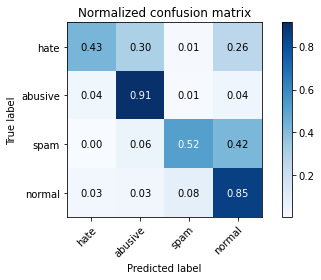
[[0.42669363 0.30434783 0.01213347 0.25682508]

[0.0378051 0.9133786 0.0082584 0.0405579 ]

[0.00392577 0.05781585 0.51677373 0.42148465]

[0.02639651 0.03476159 0.08485919 0.85398271]]





precision recall f1-score support

hate 0.46 0.43 0.44 989

abusive 0.86 0.91 0.88 5449

spam 0.60 0.52 0.55 2802

normal 0.85 0.85 0.85 10759

accuracy 0.80 19999

macro avg 0.69 0.68 0.68 19999

weighted avg 0.80 0.80 0.80 19999

average eval\_loss: 0.5374890087172389, accuracy: 0.8017900895044753

training 4 epoch...

2630.769206047058 seconds used......

4 training finished! average train loss: 0.47275969086339076

total clf loss: 858.6631997041404 total adv loss: 279.80793563096086

evaluating...

Confusion matrix, without normalization

[[ 346 359 20 264]

[ 165 4965 109 210]

[ 8 141 1834 819]

[ 239 391 1628 8501]]

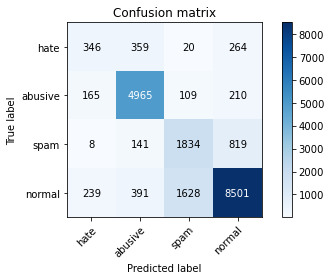
Normalized confusion matrix

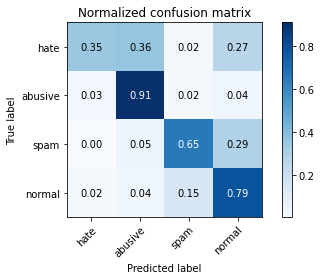
[[0.34984833 0.36299292 0.02022245 0.2669363 ]

[0.03028079 0.91117636 0.02000367 0.03853918]

[0.0028551 0.0503212 0.65453248 0.29229122]

[0.02221396 0.03634167 0.15131518 0.79012919]]





precision recall f1-score support

hate 0.46 0.35 0.40 989

abusive 0.85 0.91 0.88 5449

spam 0.51 0.65 0.57 2802

normal 0.87 0.79 0.83 10759

accuracy 0.78 19999

macro avg 0.67 0.68 0.67 19999

weighted avg 0.79 0.78 0.78 19999

average eval\_loss: 0.6254259880371392, accuracy: 0.7823391169558478

training 5 epoch...

2629.730087041855 seconds used......

5 training finished! average train loss: 0.3523114128013452

total clf loss: 643.2542011421174 total adv loss: 206.84001906599588

evaluating...

Confusion matrix, without normalization

[[ 322 340 18 309]

[ 184 4862 70 333]

[ 8 151 1402 1241]

[ 198 269 1004 9288]]

Normalized confusion matrix

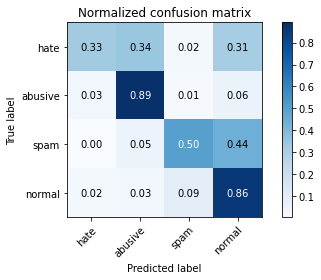
[[0.3255814 0.3437816 0.0182002 0.3124368 ]

[0.03376766 0.89227381 0.01284639 0.06111213]

[0.0028551 0.05389008 0.50035689 0.44289793]

[0.0184032 0.02500232 0.09331722 0.86327726]]





precision recall f1-score support

hate 0.45 0.33 0.38 989

abusive 0.86 0.89 0.88 5449

spam 0.56 0.50 0.53 2802

normal 0.83 0.86 0.85 10759

accuracy 0.79 19999

macro avg 0.68 0.65 0.66 19999

weighted avg 0.78 0.79 0.79 19999

average eval\_loss: 0.7063082179501653, accuracy: 0.7937396869843493