contain/not contain bad words

1 run

training 1 epoch...

520.9693920612335 seconds used......

1 training finished! average train loss: 0.5711829707127165

total clf loss: 347.4632839113474 total adv loss: 186.45931008458138

evaluating...

Confusion matrix, without normalization

[[ 310 92 88]

[ 329 2279 89]

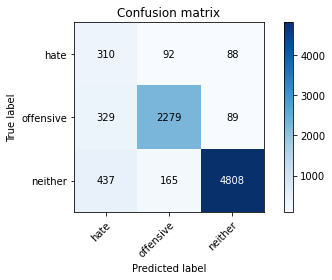
[ 437 165 4808]]

Normalized confusion matrix

[[0.63265306 0.1877551 0.17959184]

[0.12198739 0.84501298 0.03299963]

[0.08077634 0.03049908 0.88872458]]





precision recall f1-score support

hate 0.29 0.63 0.40 490

offensive 0.90 0.85 0.87 2697

neither 0.96 0.89 0.93 5410

accuracy 0.86 8597

macro avg 0.72 0.79 0.73 8597

weighted avg 0.91 0.86 0.88 8597

average eval\_loss: 0.4064772988821185, accuracy: 0.8604164243340701

training 2 epoch...

520.2949035167694 seconds used......

2 training finished! average train loss: 0.42496905908794014

total clf loss: 259.7172195613384 total adv loss: 162.71072732657194

evaluating...

Confusion matrix, without normalization

[[ 346 75 69]

[ 431 2196 70]

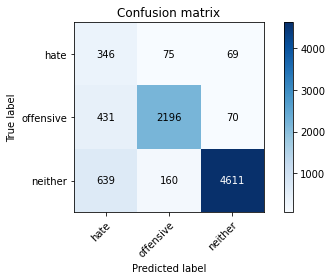
[ 639 160 4611]]

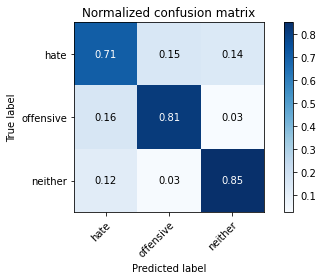
Normalized confusion matrix

[[0.70612245 0.15306122 0.14081633]

[0.15980719 0.81423804 0.02595476]

[0.1181146 0.02957486 0.85231054]]





precision recall f1-score support

hate 0.24 0.71 0.36 490

offensive 0.90 0.81 0.86 2697

neither 0.97 0.85 0.91 5410

accuracy 0.83 8597

macro avg 0.71 0.79 0.71 8597

weighted avg 0.91 0.83 0.86 8597

average eval\_loss: 0.490128804985867, accuracy: 0.8320344306153309

training 3 epoch...

520.5891644954681 seconds used......

3 training finished! average train loss: 0.31592531736336044

total clf loss: 195.4973469749093 total adv loss: 169.39118203520775

evaluating...

Confusion matrix, without normalization

[[ 321 111 58]

[ 314 2322 61]

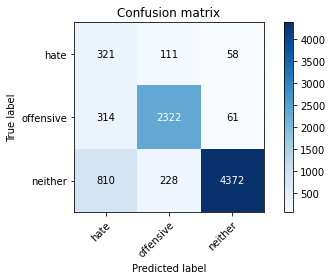
[ 810 228 4372]]

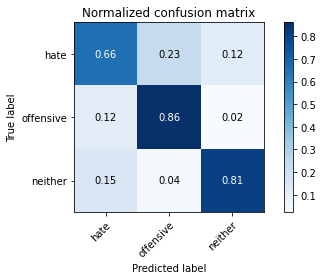
Normalized confusion matrix

[[0.65510204 0.22653061 0.11836735]

[0.11642566 0.86095662 0.02261772]

[0.14972274 0.04214418 0.80813309]]





precision recall f1-score support

hate 0.22 0.66 0.33 490

offensive 0.87 0.86 0.87 2697

neither 0.97 0.81 0.88 5410

accuracy 0.82 8597

macro avg 0.69 0.77 0.69 8597

weighted avg 0.90 0.82 0.85 8597

average eval\_loss: 0.5655909086036127, accuracy: 0.815982319413749

training 4 epoch...

520.553225517273 seconds used......

4 training finished! average train loss: 0.20603270368411433

total clf loss: 130.65025054011494 total adv loss: 173.5777989961207

evaluating...

Confusion matrix, without normalization

[[ 374 67 49]

[ 584 2046 67]

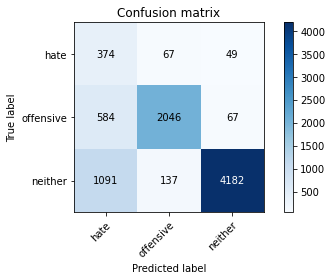
[1091 137 4182]]

Normalized confusion matrix

[[0.76326531 0.13673469 0.1 ]

[0.21653689 0.75862069 0.02484242]

[0.20166359 0.02532348 0.77301294]]





precision recall f1-score support

hate 0.18 0.76 0.29 490

offensive 0.91 0.76 0.83 2697

neither 0.97 0.77 0.86 5410

accuracy 0.77 8597

macro avg 0.69 0.76 0.66 8597

weighted avg 0.91 0.77 0.82 8597

average eval\_loss: 0.8629302984146878, accuracy: 0.7679423054553914

training 5 epoch...

520.7936236858368 seconds used......

5 training finished! average train loss: 0.13406565884361044

total clf loss: 88.4179108813405 total adv loss: 181.0208140835166

evaluating...

Confusion matrix, without normalization

[[ 334 76 80]

[ 514 2092 91]

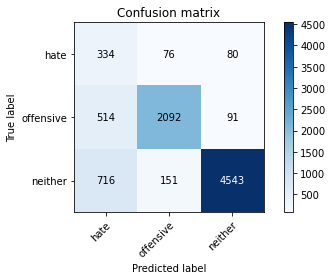
[ 716 151 4543]]

Normalized confusion matrix

[[0.68163265 0.15510204 0.16326531]

[0.19058213 0.77567668 0.03374119]

[0.1323475 0.02791128 0.83974122]]





precision recall f1-score support

hate 0.21 0.68 0.33 490

offensive 0.90 0.78 0.83 2697

neither 0.96 0.84 0.90 5410

accuracy 0.81 8597

macro avg 0.69 0.77 0.69 8597

weighted avg 0.90 0.81 0.84 8597

average eval\_loss: 0.7350119388813889, accuracy: 0.8106316156798883

Confusion matrix, without normalization

[[ 398 75 59]

[ 579 2077 54]

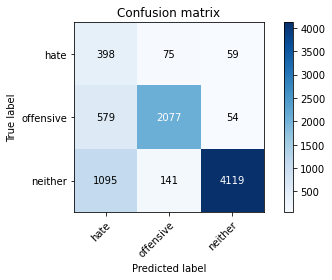
[1095 141 4119]]

Normalized confusion matrix

[[0.7481203 0.14097744 0.11090226]

[0.21365314 0.76642066 0.0199262 ]

[0.20448179 0.02633053 0.76918768]]





precision recall f1-score support

hate 0.19 0.75 0.31 532

offensive 0.91 0.77 0.83 2710

neither 0.97 0.77 0.86 5355

accuracy 0.77 8597

macro avg 0.69 0.76 0.67 8597

weighted avg 0.90 0.77 0.82 8597

average eval\_loss: 0.862892076459735, accuracy: 0.7670117482842852

[18]:

0.7670117482842852

2 run

training 1 epoch...

520.3816523551941 seconds used......

1 training finished! average train loss: 0.5696484963032039

total clf loss: 346.44214004278183 total adv loss: 184.2045939937234

evaluating...

Confusion matrix, without normalization

[[ 286 97 107]

[ 292 2309 96]

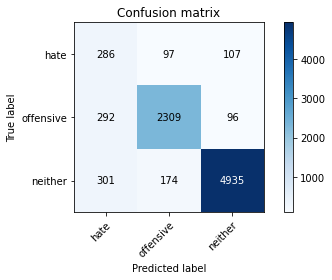
[ 301 174 4935]]

Normalized confusion matrix

[[0.58367347 0.19795918 0.21836735]

[0.10826845 0.85613645 0.03559511]

[0.05563771 0.03216266 0.91219963]]





precision recall f1-score support

hate 0.33 0.58 0.42 490

offensive 0.89 0.86 0.88 2697

neither 0.96 0.91 0.94 5410

accuracy 0.88 8597

macro avg 0.73 0.78 0.74 8597

weighted avg 0.90 0.88 0.89 8597

average eval\_loss: 0.3833153850637203, accuracy: 0.8758869373037106

training 2 epoch...

520.6129519939423 seconds used......

2 training finished! average train loss: 0.42286787039877194

total clf loss: 258.48963990807533 total adv loss: 163.03720692917705

evaluating...

Confusion matrix, without normalization

[[ 342 76 72]

[ 425 2198 74]

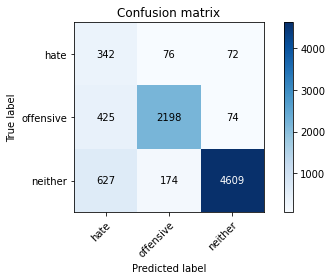
[ 627 174 4609]]

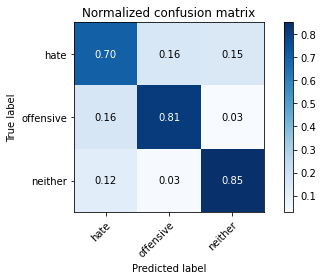
Normalized confusion matrix

[[0.69795918 0.15510204 0.14693878]

[0.1575825 0.81497961 0.02743789]

[0.11589649 0.03216266 0.85194085]]





precision recall f1-score support

hate 0.25 0.70 0.36 490

offensive 0.90 0.81 0.85 2697

neither 0.97 0.85 0.91 5410

accuracy 0.83 8597

macro avg 0.70 0.79 0.71 8597

weighted avg 0.91 0.83 0.86 8597

average eval\_loss: 0.4816563138781592, accuracy: 0.8315691520297779

training 3 epoch...

520.4998035430908 seconds used......

3 training finished! average train loss: 0.3132972397237412

total clf loss: 193.95612797513604 total adv loss: 169.6832314580679

evaluating...

Confusion matrix, without normalization

[[ 330 85 75]

[ 393 2229 75]

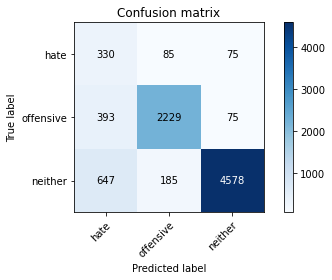
[ 647 185 4578]]

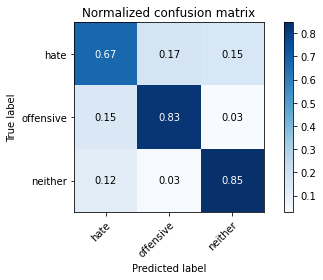
Normalized confusion matrix

[[0.67346939 0.17346939 0.15306122]

[0.14571746 0.82647386 0.02780868]

[0.11959335 0.03419593 0.84621072]]





precision recall f1-score support

hate 0.24 0.67 0.35 490

offensive 0.89 0.83 0.86 2697

neither 0.97 0.85 0.90 5410

accuracy 0.83 8597

macro avg 0.70 0.78 0.71 8597

weighted avg 0.90 0.83 0.86 8597

average eval\_loss: 0.5293206199916989, accuracy: 0.8301733162731185

training 4 epoch...

520.5479831695557 seconds used......

4 training finished! average train loss: 0.19789687082064464

total clf loss: 126.03407788090408 total adv loss: 177.58260044455528

evaluating...

Confusion matrix, without normalization

[[ 361 73 56]

[ 561 2080 56]

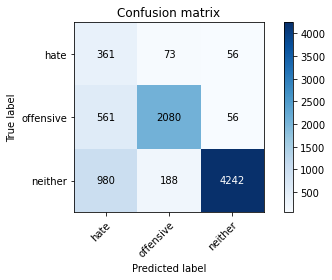
[ 980 188 4242]]

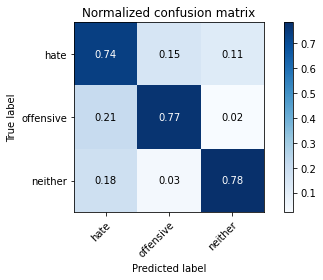
Normalized confusion matrix

[[0.73673469 0.14897959 0.11428571]

[0.2080089 0.77122729 0.02076381]

[0.18114603 0.03475046 0.78410351]]





precision recall f1-score support

hate 0.19 0.74 0.30 490

offensive 0.89 0.77 0.83 2697

neither 0.97 0.78 0.87 5410

accuracy 0.78 8597

macro avg 0.68 0.76 0.67 8597

weighted avg 0.90 0.78 0.82 8597

average eval\_loss: 0.7318621753606686, accuracy: 0.7773641968128416

training 5 epoch...

520.8928067684174 seconds used......

5 training finished! average train loss: 0.1267745796342481

total clf loss: 84.08297168277204 total adv loss: 180.64840636402369

evaluating...

Confusion matrix, without normalization

[[ 281 110 99]

[ 401 2181 115]

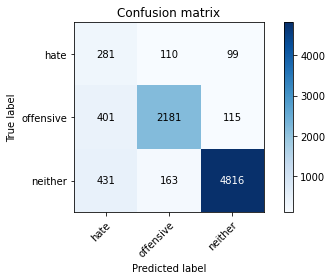
[ 431 163 4816]]

Normalized confusion matrix

[[0.57346939 0.2244898 0.20204082]

[0.14868372 0.80867631 0.04263997]

[0.07966728 0.03012939 0.89020333]]





precision recall f1-score support

hate 0.25 0.57 0.35 490

offensive 0.89 0.81 0.85 2697

neither 0.96 0.89 0.92 5410

accuracy 0.85 8597

macro avg 0.70 0.76 0.71 8597

weighted avg 0.90 0.85 0.87 8597

average eval\_loss: 0.6078347894904572, accuracy: 0.8465743864138653

Confusion matrix, without normalization

[[ 379 89 64]

[ 528 2134 48]

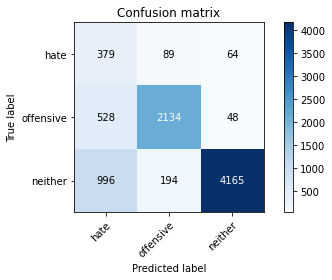
[ 996 194 4165]]

Normalized confusion matrix

[[0.71240602 0.16729323 0.12030075]

[0.19483395 0.78745387 0.01771218]

[0.1859944 0.03622782 0.77777778]]





precision recall f1-score support

hate 0.20 0.71 0.31 532

offensive 0.88 0.79 0.83 2710

neither 0.97 0.78 0.86 5355

accuracy 0.78 8597

macro avg 0.69 0.76 0.67 8597

weighted avg 0.90 0.78 0.82 8597

average eval\_loss: 0.7299936252797758, accuracy: 0.7767825985809003

[20]:

0.7767825985809003

3 run

training 1 epoch...

521.3379905223846 seconds used......

1 training finished! average train loss: 0.5704964777907809

total clf loss: 347.04380083084106 total adv loss: 186.19772244617343

evaluating...

Confusion matrix, without normalization

[[ 320 93 77]

[ 360 2284 53]

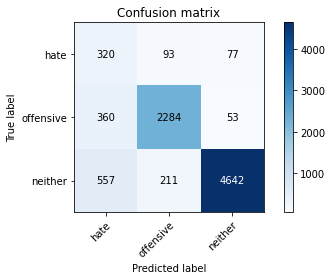
[ 557 211 4642]]

Normalized confusion matrix

[[0.65306122 0.18979592 0.15714286]

[0.13348165 0.84686689 0.01965146]

[0.10295749 0.03900185 0.85804067]]





precision recall f1-score support

hate 0.26 0.65 0.37 490

offensive 0.88 0.85 0.86 2697

neither 0.97 0.86 0.91 5410

accuracy 0.84 8597

macro avg 0.70 0.79 0.72 8597

weighted avg 0.90 0.84 0.87 8597

average eval\_loss: 0.4547832460424235, accuracy: 0.8428521577294406

training 2 epoch...

521.0876216888428 seconds used......

2 training finished! average train loss: 0.4207423670418762

total clf loss: 257.3047393038869 total adv loss: 164.50516314059496

evaluating...

Confusion matrix, without normalization

[[ 346 84 60]

[ 439 2196 62]

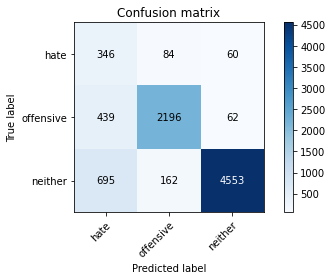
[ 695 162 4553]]

Normalized confusion matrix

[[0.70612245 0.17142857 0.12244898]

[0.16277345 0.81423804 0.02298851]

[0.1284658 0.02994455 0.84158965]]





precision recall f1-score support

hate 0.23 0.71 0.35 490

offensive 0.90 0.81 0.85 2697

neither 0.97 0.84 0.90 5410

accuracy 0.83 8597

macro avg 0.70 0.79 0.70 8597

weighted avg 0.91 0.83 0.86 8597

average eval\_loss: 0.47934605846571365, accuracy: 0.8252878911248109

training 3 epoch...

520.7609221935272 seconds used......

3 training finished! average train loss: 0.3073200793194308

total clf loss: 190.49863918870687 total adv loss: 171.30304186418653

evaluating...

Confusion matrix, without normalization

[[ 356 71 63]

[ 545 2069 83]

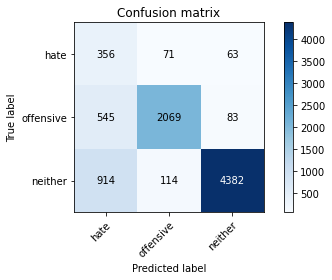
[ 914 114 4382]]

Normalized confusion matrix

[[0.72653061 0.14489796 0.12857143]

[0.20207638 0.76714868 0.03077494]

[0.1689464 0.02107209 0.80998152]]





precision recall f1-score support

hate 0.20 0.73 0.31 490

offensive 0.92 0.77 0.84 2697

neither 0.97 0.81 0.88 5410

accuracy 0.79 8597

macro avg 0.69 0.77 0.68 8597

weighted avg 0.91 0.79 0.83 8597

average eval\_loss: 0.6315415182259194, accuracy: 0.7917878329649878

training 4 epoch...

520.5013265609741 seconds used......

4 training finished! average train loss: 0.19797757594229504

total clf loss: 126.05536782741547 total adv loss: 177.05285592377186

evaluating...

Confusion matrix, without normalization

[[ 315 95 80]

[ 437 2177 83]

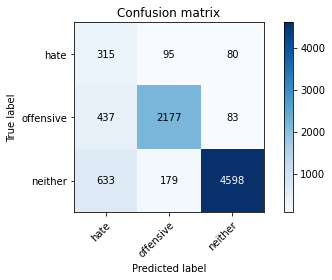
[ 633 179 4598]]

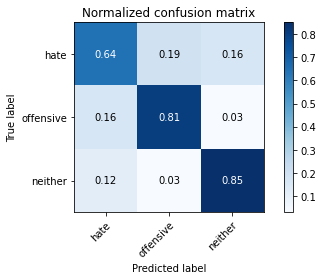
Normalized confusion matrix

[[0.64285714 0.19387755 0.16326531]

[0.16203189 0.80719318 0.03077494]

[0.11700555 0.03308688 0.84990758]]





precision recall f1-score support

hate 0.23 0.64 0.34 490

offensive 0.89 0.81 0.85 2697

neither 0.97 0.85 0.90 5410

accuracy 0.82 8597

macro avg 0.69 0.77 0.70 8597

weighted avg 0.90 0.82 0.85 8597

average eval\_loss: 0.611511737961582, accuracy: 0.8247062928928696

training 5 epoch...

521.2071461677551 seconds used......

5 training finished! average train loss: 0.12703112059436436

total clf loss: 84.41567213367671 total adv loss: 184.26497529819608

evaluating...

Confusion matrix, without normalization

[[ 321 80 89]

[ 527 2072 98]

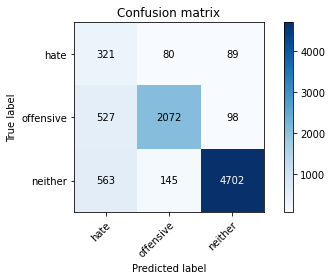
[ 563 145 4702]]

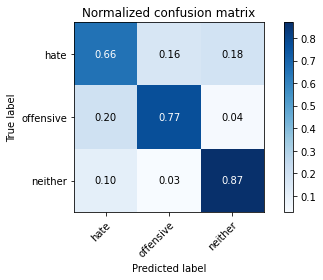
Normalized confusion matrix

[[0.65510204 0.16326531 0.18163265]

[0.1954023 0.76826103 0.03633667]

[0.10406654 0.02680222 0.86913124]]





precision recall f1-score support

hate 0.23 0.66 0.34 490

offensive 0.90 0.77 0.83 2697

neither 0.96 0.87 0.91 5410

accuracy 0.83 8597

macro avg 0.70 0.76 0.69 8597

weighted avg 0.90 0.83 0.85 8597

average eval\_loss: 0.7169851202539407, accuracy: 0.8252878911248109

Confusion matrix, without normalization

[[ 392 76 64]

[ 537 2102 71]

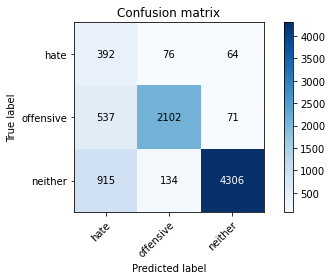
[ 915 134 4306]]

Normalized confusion matrix

[[0.73684211 0.14285714 0.12030075]

[0.19815498 0.77564576 0.02619926]

[0.17086835 0.02502334 0.80410831]]





precision recall f1-score support

hate 0.21 0.74 0.33 532

offensive 0.91 0.78 0.84 2710

neither 0.97 0.80 0.88 5355

accuracy 0.79 8597

macro avg 0.70 0.77 0.68 8597

weighted avg 0.90 0.79 0.83 8597

average eval\_loss: 0.6397082793504693, accuracy: 0.7909735954402699