alpha = 0.05

hate as offensive

input: mean of last 2 layers

adv: ffn

1 run...

Downloading: 100%

411/411 [00:00<00:00, 14.7kB/s]

Downloading: 100%

263M/263M [00:09<00:00, 31.1MB/s]

training 1 epoch...

391.0147271156311 seconds used......

1 training finished! average train loss: 0.5964887896657798

total clf loss: 271.5877541154623 total adv loss: 146.86440083011985

evaluating...

Confusion matrix, without normalization

[[ 672 214 136]

[ 695 4596 116]

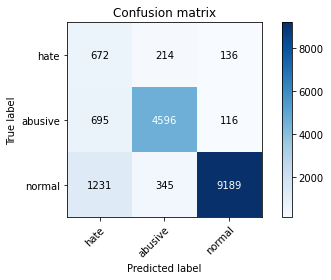
[1231 345 9189]]

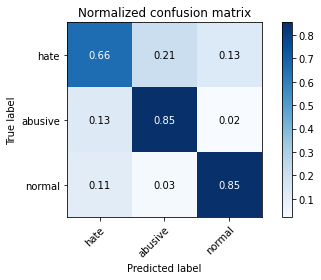
Normalized confusion matrix

[[0.65753425 0.20939335 0.13307241]

[0.12853708 0.85000925 0.02145367]

[0.11435207 0.0320483 0.85359963]]





precision recall f1-score support

hate 0.26 0.66 0.37 1022

abusive 0.89 0.85 0.87 5407

normal 0.97 0.85 0.91 10765

accuracy 0.84 17194

macro avg 0.71 0.79 0.72 17194

weighted avg 0.91 0.84 0.87 17194

average eval\_loss: 0.43895745999937835, accuracy: 0.8408165639176457

training 2 epoch...

390.5388104915619 seconds used......

2 training finished! average train loss: 0.4351394514534328

total clf loss: 197.76017692685127 total adv loss: 99.86799482442439

evaluating...

Confusion matrix, without normalization

[[ 613 196 213]

[ 685 4495 227]

[ 509 273 9983]]

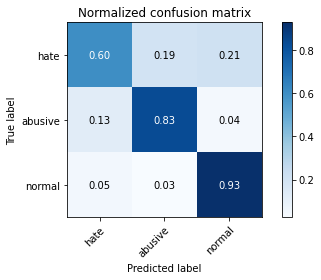
Normalized confusion matrix

[[0.59980431 0.19178082 0.20841487]

[0.12668763 0.83132976 0.04198262]

[0.04728286 0.02535996 0.92735718]]





precision recall f1-score support

hate 0.34 0.60 0.43 1022

abusive 0.91 0.83 0.87 5407

normal 0.96 0.93 0.94 10765

accuracy 0.88 17194

macro avg 0.73 0.79 0.75 17194

weighted avg 0.90 0.88 0.89 17194

average eval\_loss: 0.3597177482794884, accuracy: 0.8776898918227288

training 3 epoch...

390.71047949790955 seconds used......

3 training finished! average train loss: 0.32055900930416503

total clf loss: 145.37464992329478 total adv loss: 67.34017755184323

evaluating...

Confusion matrix, without normalization

[[ 742 198 82]

[ 936 4389 82]

[2487 491 7787]]

Normalized confusion matrix

[[0.7260274 0.19373777 0.08023483]

[0.17310893 0.81172554 0.01516553]

[0.23102647 0.04561078 0.72336275]]





precision recall f1-score support

hate 0.18 0.73 0.29 1022

abusive 0.86 0.81 0.84 5407

normal 0.98 0.72 0.83 10765

accuracy 0.75 17194

macro avg 0.67 0.75 0.65 17194

weighted avg 0.90 0.75 0.80 17194

average eval\_loss: 0.692878229604211, accuracy: 0.7513085960218681

training 4 epoch...

391.01708793640137 seconds used......

4 training finished! average train loss: 0.20821906268411647

total clf loss: 94.06334546394646 total adv loss: 36.44601060147397

evaluating...

Confusion matrix, without normalization

[[ 648 209 165]

[ 789 4444 174]

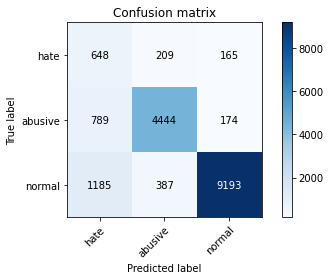
[1185 387 9193]]

Normalized confusion matrix

[[0.63405088 0.20450098 0.16144814]

[0.14592195 0.82189754 0.03218051]

[0.11007896 0.03594984 0.8539712 ]]





precision recall f1-score support

hate 0.25 0.63 0.36 1022

abusive 0.88 0.82 0.85 5407

normal 0.96 0.85 0.91 10765

accuracy 0.83 17194

macro avg 0.70 0.77 0.70 17194

weighted avg 0.90 0.83 0.86 17194

average eval\_loss: 0.5322623803605174, accuracy: 0.830813074328254

training 5 epoch...

390.87232851982117 seconds used......

5 training finished! average train loss: 0.13280536747131808

total clf loss: 59.786615264602005 total adv loss: 19.076752780470997

evaluating...

Confusion matrix, without normalization

[[ 688 186 148]

[ 908 4334 165]

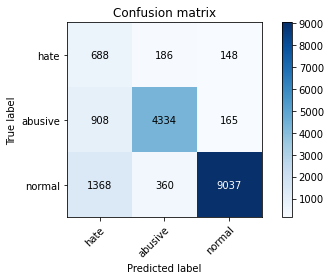
[1368 360 9037]]

Normalized confusion matrix

[[0.67318982 0.18199609 0.14481409]

[0.16793046 0.80155354 0.030516 ]

[0.1270785 0.03344171 0.8394798 ]]





precision recall f1-score support

hate 0.23 0.67 0.35 1022

abusive 0.89 0.80 0.84 5407

normal 0.97 0.84 0.90 10765

accuracy 0.82 17194

macro avg 0.70 0.77 0.70 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 0.6614515044888871, accuracy: 0.817668954286379

training 6 epoch...

390.7730312347412 seconds used......

6 training finished! average train loss: 0.10412428239318405

total clf loss: 46.957103967200965 total adv loss: 16.600937640148913

evaluating...

Confusion matrix, without normalization

[[ 697 192 133]

[ 913 4372 122]

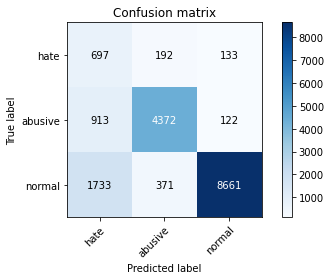
[1733 371 8661]]

Normalized confusion matrix

[[0.68199609 0.18786693 0.13013699]

[0.16885519 0.80858147 0.02256334]

[0.16098467 0.03446354 0.80455179]]





precision recall f1-score support

hate 0.21 0.68 0.32 1022

abusive 0.89 0.81 0.85 5407

normal 0.97 0.80 0.88 10765

accuracy 0.80 17194

macro avg 0.69 0.77 0.68 17194

weighted avg 0.90 0.80 0.84 17194

average eval\_loss: 0.7921586294246967, accuracy: 0.7985343724555077

training 7 epoch...

390.75392484664917 seconds used......

7 training finished! average train loss: 0.0859578378728757

total clf loss: 38.784385822247714 total adv loss: 14.101272990457801

evaluating...

Confusion matrix, without normalization

[[ 731 167 124]

[1228 4056 123]

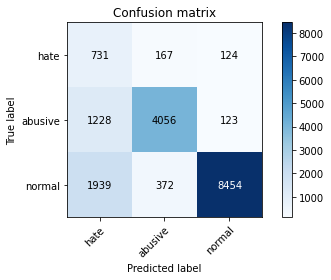
[1939 372 8454]]

Normalized confusion matrix

[[0.71526419 0.16340509 0.12133072]

[0.227113 0.75013871 0.02274829]

[0.18012076 0.03455643 0.78532281]]





precision recall f1-score support

hate 0.19 0.72 0.30 1022

abusive 0.88 0.75 0.81 5407

normal 0.97 0.79 0.87 10765

accuracy 0.77 17194

macro avg 0.68 0.75 0.66 17194

weighted avg 0.90 0.77 0.82 17194

average eval\_loss: 0.9854932129101525, accuracy: 0.7700942189135745

training 8 epoch...

390.6549298763275 seconds used......

8 training finished! average train loss: 0.07029808338066147

total clf loss: 31.62036212964449 total adv loss: 9.566222536432178

evaluating...

Confusion matrix, without normalization

[[ 708 177 137]

[1021 4259 127]

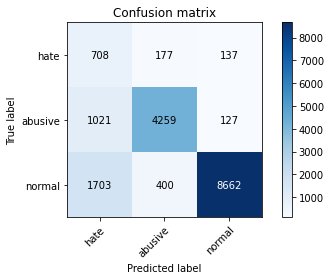
[1703 400 8662]]

Normalized confusion matrix

[[0.6927593 0.17318982 0.13405088]

[0.1888293 0.78768263 0.02348807]

[0.15819786 0.03715745 0.80464468]]





precision recall f1-score support

hate 0.21 0.69 0.32 1022

abusive 0.88 0.79 0.83 5407

normal 0.97 0.80 0.88 10765

accuracy 0.79 17194

macro avg 0.69 0.76 0.68 17194

weighted avg 0.90 0.79 0.83 17194

average eval\_loss: 1.0454752684218895, accuracy: 0.7926602303128999

training 9 epoch...

390.50635743141174 seconds used......

9 training finished! average train loss: 0.05923381272939131

total clf loss: 26.672677610244136 total adv loss: 8.641972375147816

evaluating...

Confusion matrix, without normalization

[[ 638 220 164]

[ 791 4455 161]

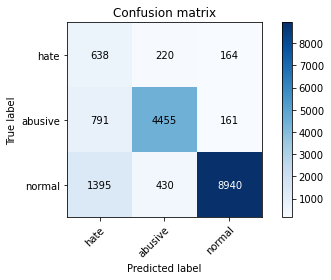
[1395 430 8940]]

Normalized confusion matrix

[[0.62426614 0.21526419 0.16046967]

[0.14629184 0.82393194 0.02977622]

[0.12958662 0.03994426 0.83046911]]





precision recall f1-score support

hate 0.23 0.62 0.33 1022

abusive 0.87 0.82 0.85 5407

normal 0.96 0.83 0.89 10765

accuracy 0.82 17194

macro avg 0.69 0.76 0.69 17194

weighted avg 0.89 0.82 0.85 17194

average eval\_loss: 0.9338973925611395, accuracy: 0.8161567988833314

training 10 epoch...

390.601131439209 seconds used......

10 training finished! average train loss: 0.06454362734481082

total clf loss: 29.0921217406285 total adv loss: 9.985897490943898

evaluating...

Confusion matrix, without normalization

[[ 631 268 123]

[ 662 4651 94]

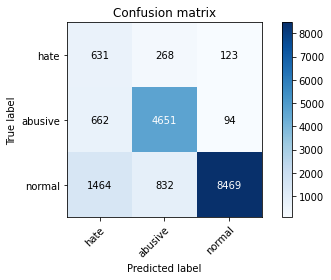
[1464 832 8469]]

Normalized confusion matrix

[[0.61741683 0.26223092 0.12035225]

[0.12243388 0.86018125 0.01738487]

[0.13599628 0.07728751 0.78671621]]





precision recall f1-score support

hate 0.23 0.62 0.33 1022

abusive 0.81 0.86 0.83 5407

normal 0.98 0.79 0.87 10765

accuracy 0.80 17194

macro avg 0.67 0.75 0.68 17194

weighted avg 0.88 0.80 0.83 17194

average eval\_loss: 1.0286755902573552, accuracy: 0.7997557287425846

2 run...

training 1 epoch...

390.8043723106384 seconds used......

1 training finished! average train loss: 0.5874237310536559

total clf loss: 267.3046534508467 total adv loss: 141.5188208706677

evaluating...

Confusion matrix, without normalization

[[ 722 158 142]

[ 973 4323 111]

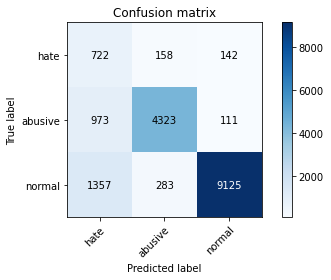
[1357 283 9125]]

Normalized confusion matrix

[[0.70645793 0.15459883 0.13894325]

[0.17995191 0.79951914 0.02052894]

[0.12605667 0.0262889 0.84765444]]





precision recall f1-score support

hate 0.24 0.71 0.35 1022

abusive 0.91 0.80 0.85 5407

normal 0.97 0.85 0.91 10765

accuracy 0.82 17194

macro avg 0.71 0.78 0.70 17194

weighted avg 0.91 0.82 0.86 17194

average eval\_loss: 0.5060806454787421, accuracy: 0.8241246946609282

training 2 epoch...

390.71448802948 seconds used......

2 training finished! average train loss: 0.4249762269094351

total clf loss: 193.16855121403933 total adv loss: 98.08164826501161

evaluating...

Confusion matrix, without normalization

[[ 711 186 125]

[ 851 4450 106]

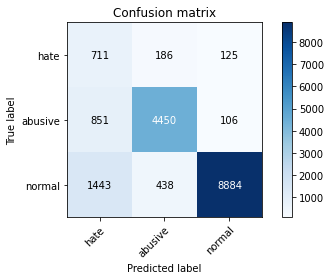
[1443 438 8884]]

Normalized confusion matrix

[[0.69569472 0.18199609 0.1223092 ]

[0.15738857 0.82300721 0.01960422]

[0.13404552 0.04068741 0.82526707]]





precision recall f1-score support

hate 0.24 0.70 0.35 1022

abusive 0.88 0.82 0.85 5407

normal 0.97 0.83 0.89 10765

accuracy 0.82 17194

macro avg 0.70 0.78 0.70 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 0.5109529053835675, accuracy: 0.816854716761661

training 3 epoch...

390.45705795288086 seconds used......

3 training finished! average train loss: 0.30768547369415283

total clf loss: 139.44936660304666 total adv loss: 62.894039675593376

evaluating...

Confusion matrix, without normalization

[[ 690 192 140]

[ 832 4418 157]

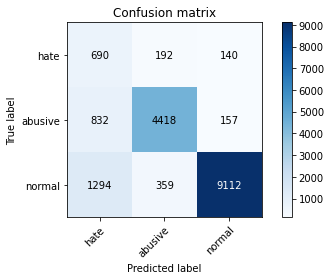
[1294 359 9112]]

Normalized confusion matrix

[[0.67514677 0.18786693 0.1369863 ]

[0.15387461 0.81708896 0.02903643]

[0.12020437 0.03334882 0.84644682]]





precision recall f1-score support

hate 0.25 0.68 0.36 1022

abusive 0.89 0.82 0.85 5407

normal 0.97 0.85 0.90 10765

accuracy 0.83 17194

macro avg 0.70 0.78 0.70 17194

weighted avg 0.90 0.83 0.85 17194

average eval\_loss: 0.5072860636076955, accuracy: 0.8270326858206352

training 4 epoch...

390.34195590019226 seconds used......

4 training finished! average train loss: 0.1945028701927471

total clf loss: 87.79287090338767 total adv loss: 32.56198807479814

evaluating...

Confusion matrix, without normalization

[[ 682 165 175]

[1070 4170 167]

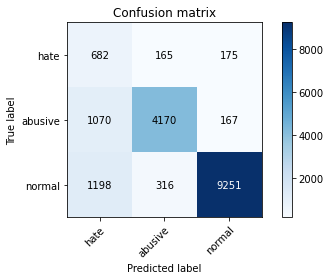
[1198 316 9251]]

Normalized confusion matrix

[[0.66731898 0.16144814 0.17123288]

[0.19789162 0.77122249 0.03088589]

[0.11128658 0.02935439 0.85935903]]





precision recall f1-score support

hate 0.23 0.67 0.34 1022

abusive 0.90 0.77 0.83 5407

normal 0.96 0.86 0.91 10765

accuracy 0.82 17194

macro avg 0.70 0.77 0.69 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 0.5784336582974119, accuracy: 0.820227986506921

training 5 epoch...

390.5239589214325 seconds used......

5 training finished! average train loss: 0.13161890978165997

total clf loss: 59.292323883622885 total adv loss: 19.702936366316862

evaluating...

Confusion matrix, without normalization

[[ 665 187 170]

[ 915 4308 184]

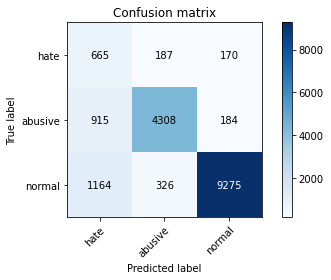
[1164 326 9275]]

Normalized confusion matrix

[[0.65068493 0.18297456 0.16634051]

[0.16922508 0.79674496 0.03402996]

[0.10812819 0.03028333 0.86158848]]





precision recall f1-score support

hate 0.24 0.65 0.35 1022

abusive 0.89 0.80 0.84 5407

normal 0.96 0.86 0.91 10765

accuracy 0.83 17194

macro avg 0.70 0.77 0.70 17194

weighted avg 0.90 0.83 0.86 17194

average eval\_loss: 0.6823605378721516, accuracy: 0.8286611608700709

training 6 epoch...

390.4852135181427 seconds used......

6 training finished! average train loss: 0.09438761625634444

total clf loss: 42.47365639102645 total adv loss: 13.198846811370458

evaluating...

Confusion matrix, without normalization

[[ 642 200 180]

[ 921 4307 179]

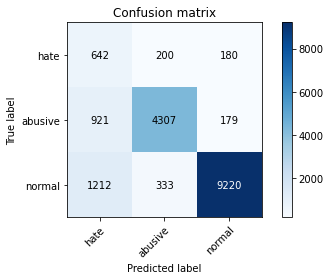
[1212 333 9220]]

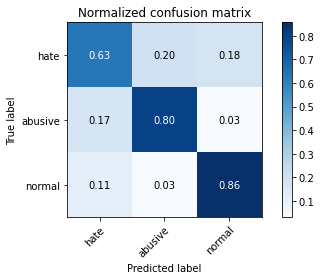
Normalized confusion matrix

[[0.62818004 0.19569472 0.17612524]

[0.17033475 0.79656001 0.03310523]

[0.11258709 0.03093358 0.85647933]]





precision recall f1-score support

hate 0.23 0.63 0.34 1022

abusive 0.89 0.80 0.84 5407

normal 0.96 0.86 0.91 10765

accuracy 0.82 17194

macro avg 0.69 0.76 0.70 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 0.6846302948502261, accuracy: 0.8240665348377341

training 7 epoch...

390.51691031455994 seconds used......

7 training finished! average train loss: 0.07894287966969622

total clf loss: 35.545942390104756 total adv loss: 11.484933800191357

evaluating...

Confusion matrix, without normalization

[[ 635 252 135]

[ 756 4510 141]

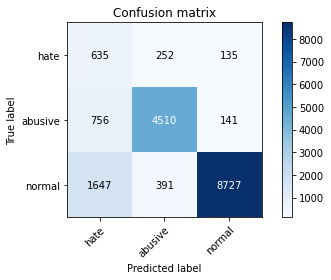
[1647 391 8727]]

Normalized confusion matrix

[[0.62133072 0.24657534 0.13209393]

[0.13981875 0.83410394 0.02607731]

[0.15299582 0.03632141 0.81068277]]





precision recall f1-score support

hate 0.21 0.62 0.31 1022

abusive 0.88 0.83 0.85 5407

normal 0.97 0.81 0.88 10765

accuracy 0.81 17194

macro avg 0.68 0.76 0.68 17194

weighted avg 0.89 0.81 0.84 17194

average eval\_loss: 0.8565723475805203, accuracy: 0.8067930673490753

training 8 epoch...

390.24726963043213 seconds used......

8 training finished! average train loss: 0.06435831041986115

total clf loss: 28.972962982021272 total adv loss: 9.244630418505949

evaluating...

Confusion matrix, without normalization

[[ 778 145 99]

[1294 4007 106]

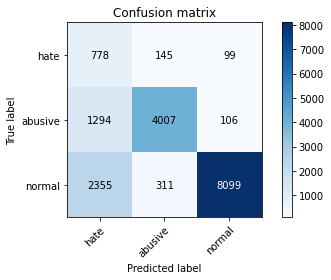
[2355 311 8099]]

Normalized confusion matrix

[[0.76125245 0.14187867 0.09686888]

[0.2393194 0.74107638 0.01960422]

[0.21876451 0.02888992 0.75234556]]





precision recall f1-score support

hate 0.18 0.76 0.29 1022

abusive 0.90 0.74 0.81 5407

normal 0.98 0.75 0.85 10765

accuracy 0.75 17194

macro avg 0.68 0.75 0.65 17194

weighted avg 0.90 0.75 0.80 17194

average eval\_loss: 1.215860529209292, accuracy: 0.7493311620332674

training 9 epoch...

390.3007938861847 seconds used......

9 training finished! average train loss: 0.060761777494884775

total clf loss: 27.425715872144792 total adv loss: 10.164968590484932

evaluating...

Confusion matrix, without normalization

[[ 717 180 125]

[1005 4269 133]

[1815 342 8608]]

Normalized confusion matrix

[[0.70156556 0.17612524 0.1223092 ]

[0.18587017 0.78953209 0.02459774]

[0.16860195 0.03176962 0.79962843]]





precision recall f1-score support

hate 0.20 0.70 0.31 1022

abusive 0.89 0.79 0.84 5407

normal 0.97 0.80 0.88 10765

accuracy 0.79 17194

macro avg 0.69 0.76 0.68 17194

weighted avg 0.90 0.79 0.83 17194

average eval\_loss: 1.0899658898194862, accuracy: 0.790624636501105

training 10 epoch...

389.9211702346802 seconds used......

10 training finished! average train loss: 0.06115758024670108

total clf loss: 27.583425161545165 total adv loss: 9.8123420495358

evaluating...

Confusion matrix, without normalization

[[ 654 206 162]

[ 800 4417 190]

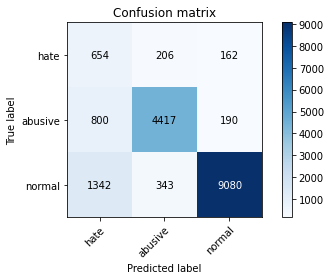
[1342 343 9080]]

Normalized confusion matrix

[[0.63992172 0.20156556 0.15851272]

[0.14795635 0.81690401 0.03513963]

[0.12466326 0.03186252 0.84347422]]





precision recall f1-score support

hate 0.23 0.64 0.34 1022

abusive 0.89 0.82 0.85 5407

normal 0.96 0.84 0.90 10765

accuracy 0.82 17194

macro avg 0.70 0.77 0.70 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 1.0008025093532038, accuracy: 0.8230196580202396

3 run...

training 1 epoch...

389.8331346511841 seconds used......

1 training finished! average train loss: 0.5989523969977488

total clf loss: 272.57685412466526 total adv loss: 144.81883573904634

evaluating...

Confusion matrix, without normalization

[[ 601 199 222]

[ 711 4523 173]

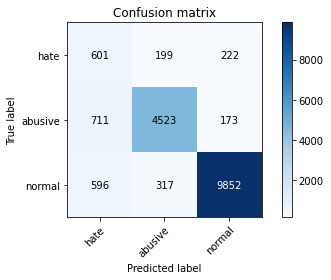
[ 596 317 9852]]

Normalized confusion matrix

[[0.58806262 0.19471624 0.21722114]

[0.13149621 0.83650823 0.03199556]

[0.05536461 0.02944728 0.91518811]]





precision recall f1-score support

hate 0.31 0.59 0.41 1022

abusive 0.90 0.84 0.87 5407

normal 0.96 0.92 0.94 10765

accuracy 0.87 17194

macro avg 0.72 0.78 0.74 17194

weighted avg 0.90 0.87 0.88 17194

average eval\_loss: 0.3707381349056959, accuracy: 0.871001512155403

training 2 epoch...

390.3017282485962 seconds used......

2 training finished! average train loss: 0.43671633674853544

total clf loss: 198.4781105518341 total adv loss: 100.25547721609473

evaluating...

Confusion matrix, without normalization

[[ 778 180 64]

[ 910 4423 74]

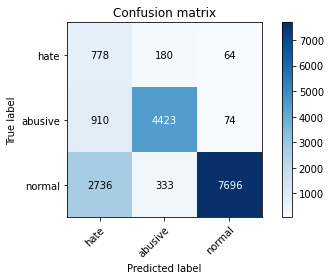
[2736 333 7696]]

Normalized confusion matrix

[[0.76125245 0.17612524 0.06262231]

[0.16830035 0.81801369 0.01368596]

[0.25415699 0.03093358 0.71490943]]





precision recall f1-score support

hate 0.18 0.76 0.29 1022

abusive 0.90 0.82 0.86 5407

normal 0.98 0.71 0.83 10765

accuracy 0.75 17194

macro avg 0.68 0.76 0.66 17194

weighted avg 0.91 0.75 0.80 17194

average eval\_loss: 0.6250177672262802, accuracy: 0.7500872397347912

training 3 epoch...

389.6733458042145 seconds used......

3 training finished! average train loss: 0.32614800533275573

total clf loss: 147.86177102476358 total adv loss: 67.56409545708448

evaluating...

Confusion matrix, without normalization

[[ 731 176 115]

[ 918 4377 112]

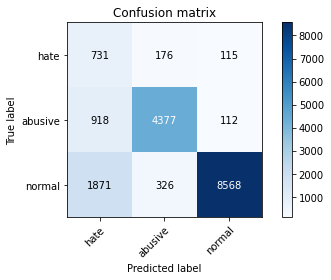
[1871 326 8568]]

Normalized confusion matrix

[[0.71526419 0.17221135 0.11252446]

[0.16977991 0.8095062 0.02071389]

[0.17380399 0.03028333 0.79591268]]





precision recall f1-score support

hate 0.21 0.72 0.32 1022

abusive 0.90 0.81 0.85 5407

normal 0.97 0.80 0.88 10765

accuracy 0.80 17194

macro avg 0.69 0.77 0.68 17194

weighted avg 0.90 0.80 0.84 17194

average eval\_loss: 0.5795482158920793, accuracy: 0.7953937420030243

training 4 epoch...

389.9522051811218 seconds used......

4 training finished! average train loss: 0.21549532078542907

total clf loss: 97.28112740442157 total adv loss: 36.33400308142882

evaluating...

Confusion matrix, without normalization

[[ 606 246 170]

[ 675 4605 127]

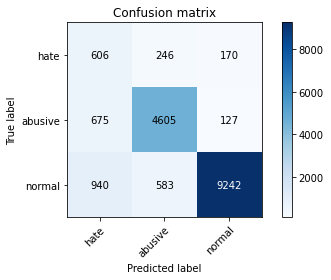
[ 940 583 9242]]

Normalized confusion matrix

[[0.59295499 0.2407045 0.16634051]

[0.12483817 0.85167376 0.02348807]

[0.08732002 0.05415699 0.85852299]]





precision recall f1-score support

hate 0.27 0.59 0.37 1022

abusive 0.85 0.85 0.85 5407

normal 0.97 0.86 0.91 10765

accuracy 0.84 17194

macro avg 0.70 0.77 0.71 17194

weighted avg 0.89 0.84 0.86 17194

average eval\_loss: 0.538265347528423, accuracy: 0.8405839246248692

training 5 epoch...

390.0205821990967 seconds used......

5 training finished! average train loss: 0.14020658594095747

total clf loss: 63.25696110539138 total adv loss: 22.908870245621074

evaluating...

Confusion matrix, without normalization

[[ 635 233 154]

[ 742 4502 163]

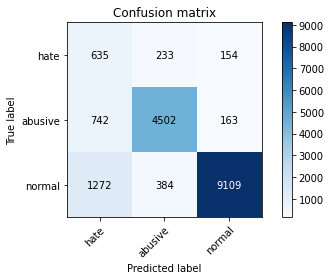
[1272 384 9109]]

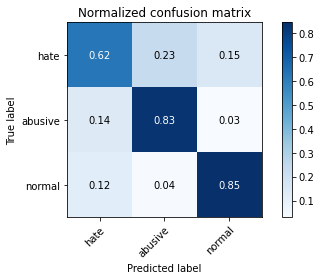
Normalized confusion matrix

[[0.62133072 0.22798434 0.15068493]

[0.13722952 0.83262438 0.03014611]

[0.11816071 0.03567116 0.84616814]]





precision recall f1-score support

hate 0.24 0.62 0.35 1022

abusive 0.88 0.83 0.86 5407

normal 0.97 0.85 0.90 10765

accuracy 0.83 17194

macro avg 0.70 0.77 0.70 17194

weighted avg 0.90 0.83 0.85 17194

average eval\_loss: 0.6281138674973316, accuracy: 0.8285448412236827

training 6 epoch...

389.99093556404114 seconds used......

6 training finished! average train loss: 0.10394264475503436

total clf loss: 46.80542429815978 total adv loss: 15.176649832963449

evaluating...

Confusion matrix, without normalization

[[ 701 178 143]

[1037 4210 160]

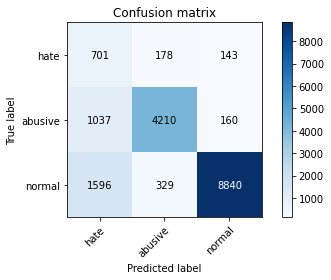
[1596 329 8840]]

Normalized confusion matrix

[[0.68590998 0.1741683 0.13992172]

[0.19178842 0.77862031 0.02959127]

[0.14825824 0.03056201 0.82117975]]





precision recall f1-score support

hate 0.21 0.69 0.32 1022

abusive 0.89 0.78 0.83 5407

normal 0.97 0.82 0.89 10765

accuracy 0.80 17194

macro avg 0.69 0.76 0.68 17194

weighted avg 0.90 0.80 0.84 17194

average eval\_loss: 0.8342902937015995, accuracy: 0.7997557287425846

training 7 epoch...

390.0396249294281 seconds used......

7 training finished! average train loss: 0.0821453876489601

total clf loss: 36.960145217133686 total adv loss: 11.394767736242102

evaluating...

Confusion matrix, without normalization

[[ 599 275 148]

[ 599 4652 156]

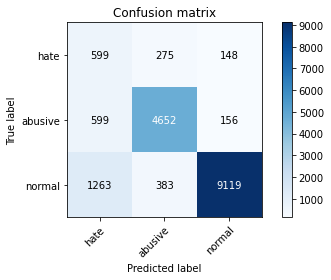
[1263 383 9119]]

Normalized confusion matrix

[[0.58610568 0.26908023 0.14481409]

[0.11078232 0.86036619 0.02885149]

[0.11732466 0.03557826 0.84709707]]





precision recall f1-score support

hate 0.24 0.59 0.34 1022

abusive 0.88 0.86 0.87 5407

normal 0.97 0.85 0.90 10765

accuracy 0.84 17194

macro avg 0.70 0.76 0.71 17194

weighted avg 0.90 0.84 0.86 17194

average eval\_loss: 0.7324831528599959, accuracy: 0.8357566592997557

training 8 epoch...

390.12011194229126 seconds used......

8 training finished! average train loss: 0.06977997409520381

total clf loss: 31.3439156570239 total adv loss: 8.62774266192173

evaluating...

Confusion matrix, without normalization

[[ 668 206 148]

[ 877 4356 174]

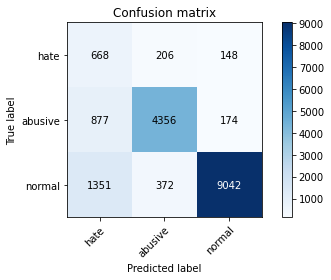
[1351 372 9042]]

Normalized confusion matrix

[[0.65362035 0.20156556 0.14481409]

[0.16219715 0.80562234 0.03218051]

[0.1254993 0.03455643 0.83994426]]





precision recall f1-score support

hate 0.23 0.65 0.34 1022

abusive 0.88 0.81 0.84 5407

normal 0.97 0.84 0.90 10765

accuracy 0.82 17194

macro avg 0.69 0.77 0.69 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 0.8719345234594372, accuracy: 0.818076073048738

training 9 epoch...

390.3661503791809 seconds used......

9 training finished! average train loss: 0.06098042640563849

total clf loss: 27.438718379708007 total adv loss: 8.48778917805248

evaluating...

Confusion matrix, without normalization

[[ 758 162 102]

[1244 4067 96]

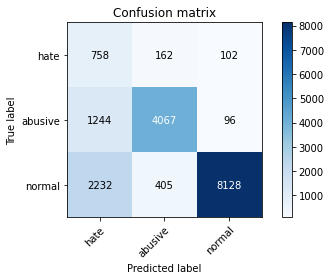
[2232 405 8128]]

Normalized confusion matrix

[[0.74168297 0.15851272 0.09980431]

[0.23007213 0.75217311 0.01775476]

[0.2073386 0.03762192 0.75503948]]





precision recall f1-score support

hate 0.18 0.74 0.29 1022

abusive 0.88 0.75 0.81 5407

normal 0.98 0.76 0.85 10765

accuracy 0.75 17194

macro avg 0.68 0.75 0.65 17194

weighted avg 0.90 0.75 0.81 17194

average eval\_loss: 1.2834603589114755, accuracy: 0.7533441898336629

training 10 epoch...

390.4094338417053 seconds used......

10 training finished! average train loss: 0.06629212477009766

total clf loss: 29.993335054983618 total adv loss: 12.51847554757751

evaluating...

Confusion matrix, without normalization

[[ 702 193 127]

[ 978 4300 129]

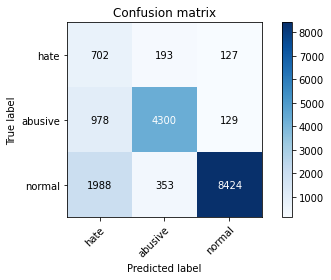
[1988 353 8424]]

Normalized confusion matrix

[[0.68688845 0.1888454 0.12426614]

[0.18087664 0.7952654 0.02385796]

[0.18467255 0.03279145 0.782536 ]]





precision recall f1-score support

hate 0.19 0.69 0.30 1022

abusive 0.89 0.80 0.84 5407

normal 0.97 0.78 0.87 10765

accuracy 0.78 17194

macro avg 0.68 0.75 0.67 17194

weighted avg 0.90 0.78 0.82 17194

average eval\_loss: 1.076008227236872, accuracy: 0.78085378620449

4 run...

training 1 epoch...

390.762291431427 seconds used......

1 training finished! average train loss: 0.5859870482446378

total clf loss: 266.78688056766987 total adv loss: 143.89236987382174

evaluating...

Confusion matrix, without normalization

[[ 533 298 191]

[ 393 4895 119]

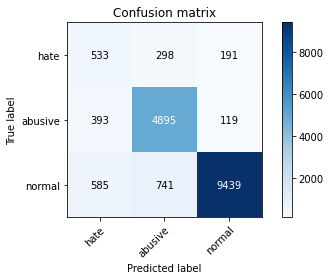
[ 585 741 9439]]

Normalized confusion matrix

[[0.52152642 0.29158513 0.18688845]

[0.07268356 0.90530793 0.02200851]

[0.05434278 0.06883418 0.87682304]]





precision recall f1-score support

hate 0.35 0.52 0.42 1022

abusive 0.82 0.91 0.86 5407

normal 0.97 0.88 0.92 10765

accuracy 0.86 17194

macro avg 0.72 0.77 0.73 17194

weighted avg 0.89 0.86 0.87 17194

average eval\_loss: 0.4064579658037008, accuracy: 0.864662091427242

training 2 epoch...

391.2154529094696 seconds used......

2 training finished! average train loss: 0.4272899267097628

total clf loss: 194.05767308175564 total adv loss: 95.36470957659185

evaluating...

Confusion matrix, without normalization

[[ 645 256 121]

[ 558 4766 83]

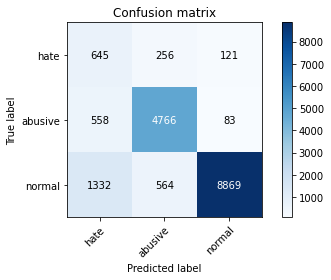
[1332 564 8869]]

Normalized confusion matrix

[[0.63111546 0.25048924 0.1183953 ]

[0.10319956 0.88144997 0.01535047]

[0.12373432 0.05239201 0.82387366]]





precision recall f1-score support

hate 0.25 0.63 0.36 1022

abusive 0.85 0.88 0.87 5407

normal 0.98 0.82 0.89 10765

accuracy 0.83 17194

macro avg 0.70 0.78 0.71 17194

weighted avg 0.90 0.83 0.85 17194

average eval\_loss: 0.48732708203931185, accuracy: 0.8305222752122834

training 3 epoch...

391.26452016830444 seconds used......

3 training finished! average train loss: 0.31245576544048825

total clf loss: 141.69552539288998 total adv loss: 65.5524245267734

evaluating...

Confusion matrix, without normalization

[[ 621 263 138]

[ 558 4741 108]

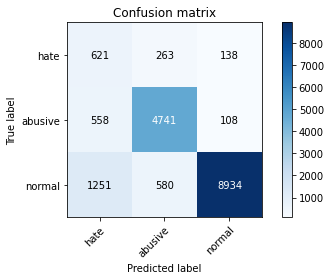
[1251 580 8934]]

Normalized confusion matrix

[[0.60763209 0.25733855 0.13502935]

[0.10319956 0.87682634 0.01997411]

[0.11620994 0.05387831 0.82991175]]





precision recall f1-score support

hate 0.26 0.61 0.36 1022

abusive 0.85 0.88 0.86 5407

normal 0.97 0.83 0.90 10765

accuracy 0.83 17194

macro avg 0.69 0.77 0.71 17194

weighted avg 0.89 0.83 0.85 17194

average eval\_loss: 0.527236947906745, accuracy: 0.8314528323833895

training 4 epoch...

391.18401551246643 seconds used......

4 training finished! average train loss: 0.2101323875150818

total clf loss: 94.98635699227452 total adv loss: 37.95418828795664

evaluating...

Confusion matrix, without normalization

[[ 673 164 185]

[ 977 4235 195]

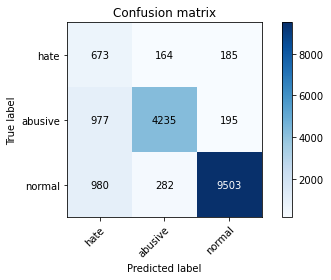
[ 980 282 9503]]

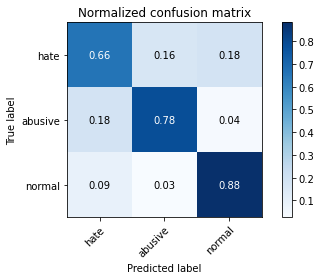
Normalized confusion matrix

[[0.65851272 0.16046967 0.18101761]

[0.1806917 0.78324394 0.03606436]

[0.09103576 0.02619601 0.88276823]]





precision recall f1-score support

hate 0.26 0.66 0.37 1022

abusive 0.90 0.78 0.84 5407

normal 0.96 0.88 0.92 10765

accuracy 0.84 17194

macro avg 0.71 0.77 0.71 17194

weighted avg 0.90 0.84 0.86 17194

average eval\_loss: 0.5392123435331552, accuracy: 0.8381412120507153

training 5 epoch...

4 run...

Downloading: 100%

411/411 [00:00<00:00, 14.6kB/s]

Downloading: 100%

263M/263M [00:06<00:00, 28.1MB/s]

training 1 epoch...

390.01872730255127 seconds used......

1 training finished! average train loss: 0.5979946373413163

total clf loss: 272.097211971879 total adv loss: 143.71174461580813

evaluating...

Confusion matrix, without normalization

[[ 746 119 157]

[1158 4119 130]

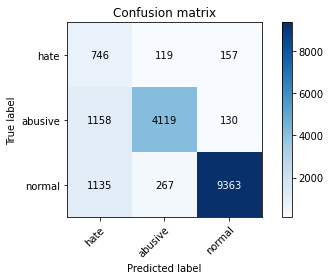
[1135 267 9363]]

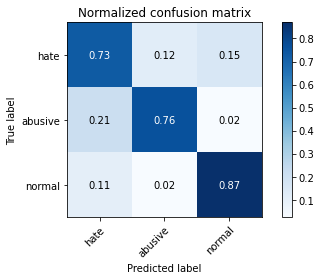
Normalized confusion matrix

[[0.72994129 0.11643836 0.15362035]

[0.21416682 0.76179027 0.02404291]

[0.10543428 0.0248026 0.86976312]]





precision recall f1-score support

hate 0.25 0.73 0.37 1022

abusive 0.91 0.76 0.83 5407

normal 0.97 0.87 0.92 10765

accuracy 0.83 17194

macro avg 0.71 0.79 0.71 17194

weighted avg 0.91 0.83 0.86 17194

average eval\_loss: 0.4783685064766296, accuracy: 0.8274979644061882

training 2 epoch...

389.3410005569458 seconds used......

2 training finished! average train loss: 0.43365409746597905

total clf loss: 197.00829583406448 total adv loss: 97.9906212836504

evaluating...

Confusion matrix, without normalization

[[ 699 199 124]

[ 771 4525 111]

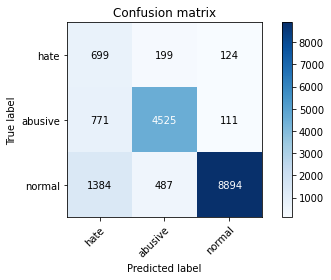
[1384 487 8894]]

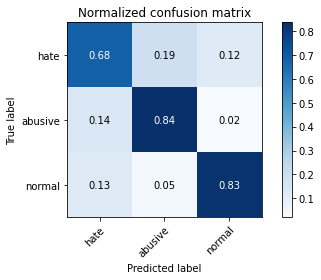
Normalized confusion matrix

[[0.68395303 0.19471624 0.12133072]

[0.14259294 0.83687812 0.02052894]

[0.12856479 0.0452392 0.82619601]]





precision recall f1-score support

hate 0.24 0.68 0.36 1022

abusive 0.87 0.84 0.85 5407

normal 0.97 0.83 0.89 10765

accuracy 0.82 17194

macro avg 0.70 0.78 0.70 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 0.4917333579947089, accuracy: 0.8211003838548331

training 3 epoch...

389.38260650634766 seconds used......

3 training finished! average train loss: 0.31442912896720876

total clf loss: 142.41063944995403 total adv loss: 62.370709226466715

evaluating...

Confusion matrix, without normalization

[[ 697 182 143]

[ 819 4437 151]

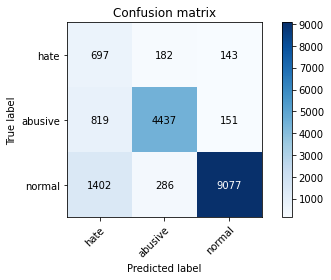
[1402 286 9077]]

Normalized confusion matrix

[[0.68199609 0.17808219 0.13992172]

[0.15147032 0.82060292 0.02792676]

[0.13023688 0.02656758 0.84319554]]





precision recall f1-score support

hate 0.24 0.68 0.35 1022

abusive 0.90 0.82 0.86 5407

normal 0.97 0.84 0.90 10765

accuracy 0.83 17194

macro avg 0.70 0.78 0.71 17194

weighted avg 0.91 0.83 0.86 17194

average eval\_loss: 0.49262418003622876, accuracy: 0.8265092474118879

training 4 epoch...

389.82652497291565 seconds used......

4 training finished! average train loss: 0.20949865487439906

total clf loss: 94.59153658337891 total adv loss: 35.67264985549264

evaluating...

Confusion matrix, without normalization

[[ 670 188 164]

[ 860 4408 139]

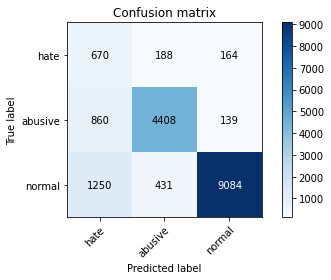
[1250 431 9084]]

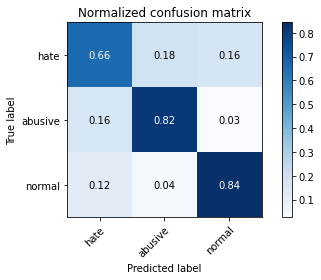
Normalized confusion matrix

[[0.6555773 0.18395303 0.16046967]

[0.15905308 0.8152395 0.02570742]

[0.11611705 0.04003716 0.8438458 ]]





precision recall f1-score support

hate 0.24 0.66 0.35 1022

abusive 0.88 0.82 0.84 5407

normal 0.97 0.84 0.90 10765

accuracy 0.82 17194

macro avg 0.70 0.77 0.70 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 0.5814862876199186, accuracy: 0.8236594160753752

training 5 epoch...

390.10955834388733 seconds used......

5 training finished! average train loss: 0.14333965597384132

total clf loss: 64.46470514126122 total adv loss: 19.304749340226408

evaluating...

Confusion matrix, without normalization

[[ 623 246 153]

[ 627 4632 148]

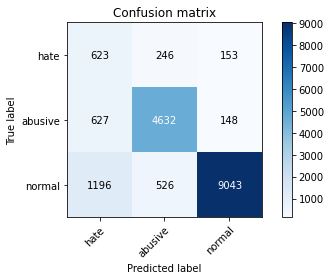
[1196 526 9043]]

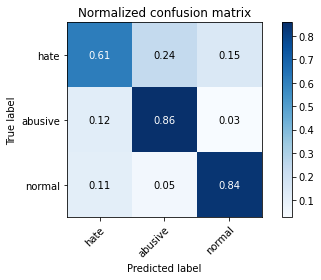
Normalized confusion matrix

[[0.60958904 0.2407045 0.14970646]

[0.11596079 0.85666728 0.02737193]

[0.11110079 0.04886205 0.84003716]]





precision recall f1-score support

hate 0.25 0.61 0.36 1022

abusive 0.86 0.86 0.86 5407

normal 0.97 0.84 0.90 10765

accuracy 0.83 17194

macro avg 0.69 0.77 0.71 17194

weighted avg 0.89 0.83 0.85 17194

average eval\_loss: 0.6769923837757994, accuracy: 0.8315691520297779

training 6 epoch...

389.8499708175659 seconds used......

6 training finished! average train loss: 0.10253650099343913

total clf loss: 46.15649996045977 total adv loss: 14.656599843205186

evaluating...

Confusion matrix, without normalization

[[ 589 258 175]

[ 586 4649 172]

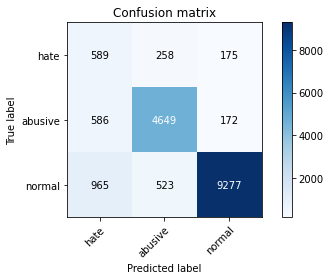
[ 965 523 9277]]

Normalized confusion matrix

[[0.57632094 0.25244618 0.17123288]

[0.10837803 0.85981136 0.03181062]

[0.08964236 0.04858337 0.86177427]]





precision recall f1-score support

hate 0.28 0.58 0.37 1022

abusive 0.86 0.86 0.86 5407

normal 0.96 0.86 0.91 10765

accuracy 0.84 17194

macro avg 0.70 0.77 0.71 17194

weighted avg 0.89 0.84 0.86 17194

average eval\_loss: 0.6712681643253322, accuracy: 0.8441898336629057

training 7 epoch...

390.0872094631195 seconds used......

7 training finished! average train loss: 0.08095303327990927

total clf loss: 36.32033282308839 total adv loss: 9.162781680315675

evaluating...

Confusion matrix, without normalization

[[ 738 158 126]

[1201 4071 135]

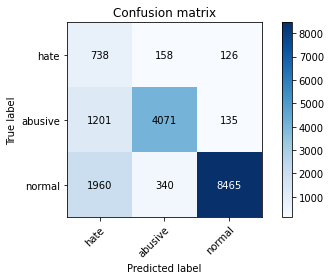
[1960 340 8465]]

Normalized confusion matrix

[[0.7221135 0.15459883 0.12328767]

[0.22211947 0.75291289 0.02496763]

[0.18207153 0.03158384 0.78634464]]





precision recall f1-score support

hate 0.19 0.72 0.30 1022

abusive 0.89 0.75 0.82 5407

normal 0.97 0.79 0.87 10765

accuracy 0.77 17194

macro avg 0.68 0.75 0.66 17194

weighted avg 0.90 0.77 0.82 17194

average eval\_loss: 1.0253114635904514, accuracy: 0.772013493078981

training 8 epoch...

390.3997058868408 seconds used......

8 training finished! average train loss: 0.06889842594534887

total clf loss: 30.983077118522488 total adv loss: 9.221489491199463

evaluating...

Confusion matrix, without normalization

[[ 623 226 173]

[ 722 4519 166]

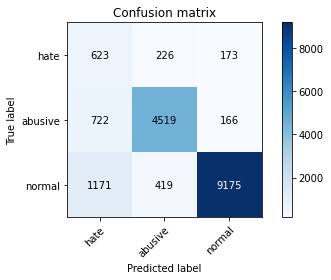
[1171 419 9175]]

Normalized confusion matrix

[[0.60958904 0.22113503 0.16927593]

[0.13353061 0.83576845 0.03070094]

[0.10877845 0.03892243 0.85229912]]





precision recall f1-score support

hate 0.25 0.61 0.35 1022

abusive 0.88 0.84 0.85 5407

normal 0.96 0.85 0.90 10765

accuracy 0.83 17194

macro avg 0.70 0.77 0.70 17194

weighted avg 0.89 0.83 0.86 17194

average eval\_loss: 0.827632265687422, accuracy: 0.8326741886704664

training 9 epoch...

390.443589925766 seconds used......

9 training finished! average train loss: 0.07175786207975614

total clf loss: 32.34839354606811 total adv loss: 11.193215028239138

evaluating...

Confusion matrix, without normalization

[[ 646 208 168]

[ 808 4424 175]

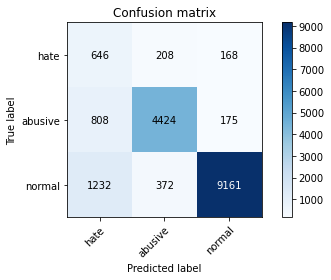
[1232 372 9161]]

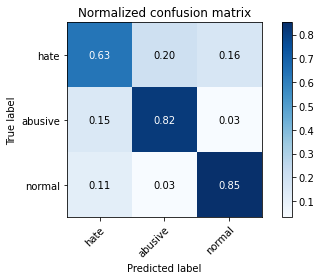
Normalized confusion matrix

[[0.63209393 0.2035225 0.16438356]

[0.14943592 0.81819863 0.03236545]

[0.11444496 0.03455643 0.85099861]]





precision recall f1-score support

hate 0.24 0.63 0.35 1022

abusive 0.88 0.82 0.85 5407

normal 0.96 0.85 0.90 10765

accuracy 0.83 17194

macro avg 0.70 0.77 0.70 17194

weighted avg 0.90 0.83 0.85 17194

average eval\_loss: 0.8013250429800517, accuracy: 0.8276724438757707

training 10 epoch...

390.38766980171204 seconds used......

10 training finished! average train loss: 0.06389852349370988

total clf loss: 28.816991879430134 total adv loss: 10.198919299433328

evaluating...

Confusion matrix, without normalization

[[ 665 197 160]

[ 882 4352 173]

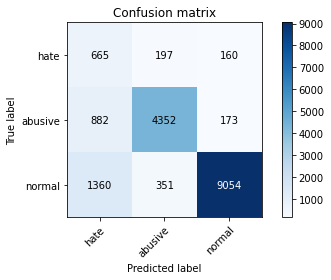
[1360 351 9054]]

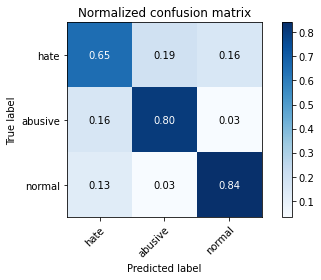
Normalized confusion matrix

[[0.65068493 0.1927593 0.15655577]

[0.16312188 0.80488256 0.03199556]

[0.12633535 0.03260567 0.84105899]]





precision recall f1-score support

hate 0.23 0.65 0.34 1022

abusive 0.89 0.80 0.84 5407

normal 0.96 0.84 0.90 10765

accuracy 0.82 17194

macro avg 0.69 0.77 0.69 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 0.905636813978632, accuracy: 0.8183668721647086

5 run...

training 1 epoch...

390.37644243240356 seconds used......

1 training finished! average train loss: 0.5887593421454355

total clf loss: 267.91065169870853 total adv loss: 141.80525709688663

evaluating...

Confusion matrix, without normalization

[[ 725 159 138]

[ 933 4357 117]

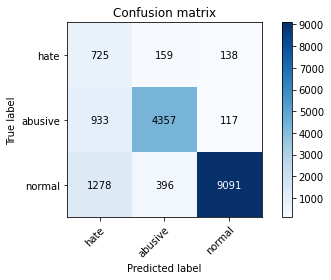
[1278 396 9091]]

Normalized confusion matrix

[[0.70939335 0.1555773 0.13502935]

[0.1725541 0.80580729 0.02163862]

[0.11871807 0.03678588 0.84449605]]





precision recall f1-score support

hate 0.25 0.71 0.37 1022

abusive 0.89 0.81 0.84 5407

normal 0.97 0.84 0.90 10765

accuracy 0.82 17194

macro avg 0.70 0.79 0.70 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 0.4721995760015277, accuracy: 0.8242991741305107

training 2 epoch...

390.1896378993988 seconds used......

2 training finished! average train loss: 0.4260609612608871

total clf loss: 193.64696609973907 total adv loss: 98.0392036139965

evaluating...

Confusion matrix, without normalization

[[ 646 193 183]

[ 725 4487 195]

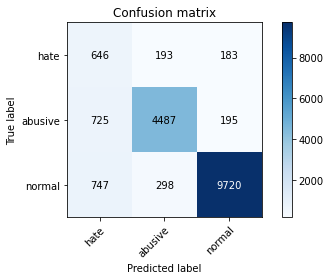
[ 747 298 9720]]

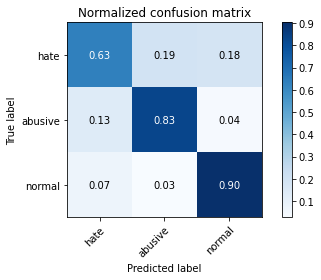
Normalized confusion matrix

[[0.63209393 0.1888454 0.17906067]

[0.13408544 0.82985019 0.03606436]

[0.06939155 0.0276823 0.90292615]]





precision recall f1-score support

hate 0.31 0.63 0.41 1022

abusive 0.90 0.83 0.86 5407

normal 0.96 0.90 0.93 10765

accuracy 0.86 17194

macro avg 0.72 0.79 0.74 17194

weighted avg 0.90 0.86 0.88 17194

average eval\_loss: 0.3819022274684421, accuracy: 0.8638478539025242

training 3 epoch...

390.18243193626404 seconds used......

3 training finished! average train loss: 0.3042877174442966

total clf loss: 138.03108159825206 total adv loss: 64.6324522071518

evaluating...

Confusion matrix, without normalization

[[ 620 237 165]

[ 707 4540 160]

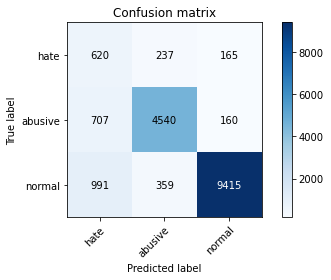
[ 991 359 9415]]

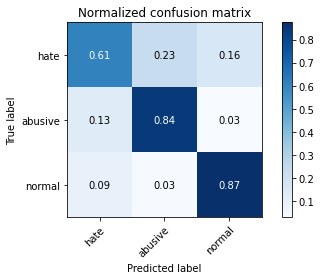
Normalized confusion matrix

[[0.60665362 0.23189824 0.16144814]

[0.13075643 0.8396523 0.02959127]

[0.09205759 0.03334882 0.87459359]]





precision recall f1-score support

hate 0.27 0.61 0.37 1022

abusive 0.88 0.84 0.86 5407

normal 0.97 0.87 0.92 10765

accuracy 0.85 17194

macro avg 0.71 0.77 0.72 17194

weighted avg 0.90 0.85 0.87 17194

average eval\_loss: 0.46496833384816727, accuracy: 0.8476794230545539

training 4 epoch...

389.76919078826904 seconds used......

4 training finished! average train loss: 0.19345267007282976

total clf loss: 87.41652870737016 total adv loss: 34.33991879271343

evaluating...

Confusion matrix, without normalization

[[ 603 246 173]

[ 655 4558 194]

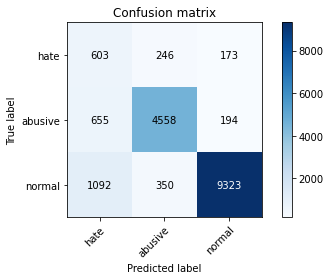
[1092 350 9323]]

Normalized confusion matrix

[[0.59001957 0.2407045 0.16927593]

[0.12113926 0.84298132 0.03587942]

[0.10143985 0.03251277 0.86604738]]





precision recall f1-score support

hate 0.26 0.59 0.36 1022

abusive 0.88 0.84 0.86 5407

normal 0.96 0.87 0.91 10765

accuracy 0.84 17194

macro avg 0.70 0.77 0.71 17194

weighted avg 0.90 0.84 0.86 17194

average eval\_loss: 0.5711897635485891, accuracy: 0.8423868791438874

training 5 epoch...

389.74895429611206 seconds used......

5 training finished! average train loss: 0.130737947228598

total clf loss: 58.92104073986411 total adv loss: 20.08260129520204

evaluating...

Confusion matrix, without normalization

[[ 660 224 138]

[ 845 4426 136]

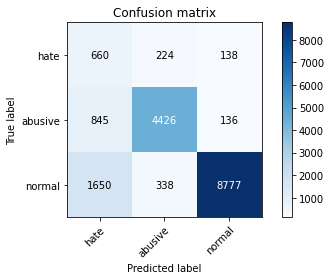
[1650 338 8777]]

Normalized confusion matrix

[[0.64579256 0.21917808 0.13502935]

[0.1562789 0.81856852 0.02515258]

[0.1532745 0.03139805 0.81532745]]





precision recall f1-score support

hate 0.21 0.65 0.32 1022

abusive 0.89 0.82 0.85 5407

normal 0.97 0.82 0.89 10765

accuracy 0.81 17194

macro avg 0.69 0.76 0.68 17194

weighted avg 0.90 0.81 0.84 17194

average eval\_loss: 0.7489489071720908, accuracy: 0.806269628940328

training 6 epoch...

389.78640937805176 seconds used......

6 training finished! average train loss: 0.09180510577611188

total clf loss: 41.24930141866207 total adv loss: 11.59279248675739

evaluating...

Confusion matrix, without normalization

[[ 625 229 168]

[ 798 4432 177]

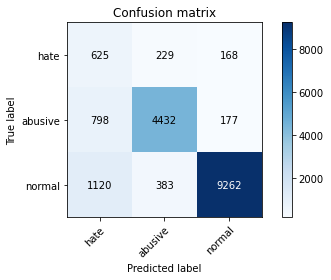
[1120 383 9262]]

Normalized confusion matrix

[[0.61154599 0.22407045 0.16438356]

[0.14758646 0.81967819 0.03273534]

[0.10404087 0.03557826 0.86038086]]





precision recall f1-score support

hate 0.25 0.61 0.35 1022

abusive 0.88 0.82 0.85 5407

normal 0.96 0.86 0.91 10765

accuracy 0.83 17194

macro avg 0.70 0.76 0.70 17194

weighted avg 0.89 0.83 0.86 17194

average eval\_loss: 0.70067143795936, accuracy: 0.8327905083168547

training 7 epoch...

389.6821117401123 seconds used......

7 training finished! average train loss: 0.07600701471100063

total clf loss: 34.19684902543668 total adv loss: 10.514829843501502

evaluating...

Confusion matrix, without normalization

[[ 658 242 122]

[ 851 4450 106]

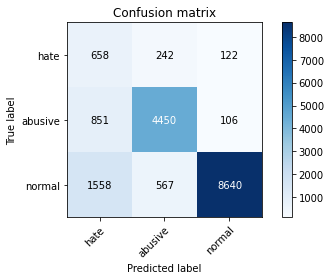
[1558 567 8640]]

Normalized confusion matrix

[[0.64383562 0.23679061 0.11937378]

[0.15738857 0.82300721 0.01960422]

[0.14472829 0.05267069 0.80260102]]





precision recall f1-score support

hate 0.21 0.64 0.32 1022

abusive 0.85 0.82 0.83 5407

normal 0.97 0.80 0.88 10765

accuracy 0.80 17194

macro avg 0.68 0.76 0.68 17194

weighted avg 0.89 0.80 0.83 17194

average eval\_loss: 0.9517119567094068, accuracy: 0.7995812492730022

training 8 epoch...

389.9611520767212 seconds used......

8 training finished! average train loss: 0.06651337126256271

total clf loss: 29.899589977809228 total adv loss: 8.68332808586274

evaluating...

Confusion matrix, without normalization

[[ 728 167 127]

[1127 4117 163]

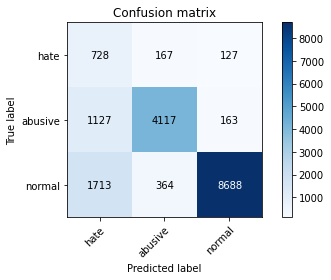
[1713 364 8688]]

Normalized confusion matrix

[[0.71232877 0.16340509 0.12426614]

[0.20843351 0.76142038 0.03014611]

[0.1591268 0.03381328 0.80705992]]





precision recall f1-score support

hate 0.20 0.71 0.32 1022

abusive 0.89 0.76 0.82 5407

normal 0.97 0.81 0.88 10765

accuracy 0.79 17194

macro avg 0.69 0.76 0.67 17194

weighted avg 0.90 0.79 0.83 17194

average eval\_loss: 1.0657475550745636, accuracy: 0.7870768872862627

training 9 epoch...

389.9486289024353 seconds used......

9 training finished! average train loss: 0.06088193493519971

total clf loss: 27.475316899712197 total adv loss: 10.092394395258452

evaluating...

Confusion matrix, without normalization

[[ 633 256 133]

[ 661 4602 144]

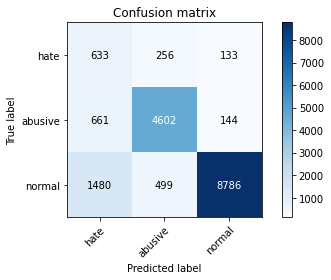
[1480 499 8786]]

Normalized confusion matrix

[[0.61937378 0.25048924 0.13013699]

[0.12224894 0.85111892 0.02663214]

[0.13748258 0.04635392 0.81616349]]





precision recall f1-score support

hate 0.23 0.62 0.33 1022

abusive 0.86 0.85 0.86 5407

normal 0.97 0.82 0.89 10765

accuracy 0.82 17194

macro avg 0.69 0.76 0.69 17194

weighted avg 0.89 0.82 0.84 17194

average eval\_loss: 0.9875110101241749, accuracy: 0.8154588810050017

training 10 epoch...

389.8028552532196 seconds used......

10 training finished! average train loss: 0.06091840083598639

total clf loss: 27.490738589985995 total adv loss: 10.077739656015183

evaluating...

Confusion matrix, without normalization

[[ 677 246 99]

[ 810 4501 96]

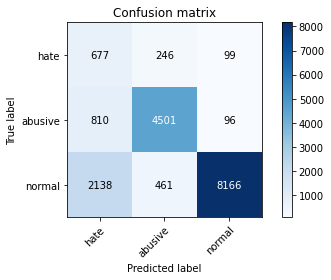
[2138 461 8166]]

Normalized confusion matrix

[[0.66242661 0.2407045 0.09686888]

[0.14980581 0.83243943 0.01775476]

[0.1986066 0.04282397 0.75856944]]





precision recall f1-score support

hate 0.19 0.66 0.29 1022

abusive 0.86 0.83 0.85 5407

normal 0.98 0.76 0.85 10765

accuracy 0.78 17194

macro avg 0.68 0.75 0.66 17194

weighted avg 0.89 0.78 0.82 17194

average eval\_loss: 1.1610121309389034, accuracy: 0.7760846807025706