alpha = 0.05

hate as offensive

input: mean of last 2 layers

adv: ffn

1 run...

Downloading: 100%

411/411 [00:00<00:00, 13.6kB/s]

Downloading: 100%

263M/263M [00:09<00:00, 30.9MB/s]

training 1 epoch...

245.92159032821655 seconds used......

1 training finished! average train loss: 0.6029920193869158

total clf loss: 274.3249001801014 total adv loss: 143.98871426656842

evaluating...

Confusion matrix, without normalization

[[ 653 209 160]

[ 624 4655 128]

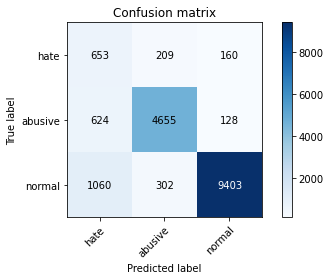
[1060 302 9403]]

Normalized confusion matrix

[[0.63894325 0.20450098 0.15655577]

[0.11540596 0.86092103 0.02367302]

[0.09846725 0.02805388 0.87347887]]





precision recall f1-score support

hate 0.28 0.64 0.39 1022

abusive 0.90 0.86 0.88 5407

normal 0.97 0.87 0.92 10765

accuracy 0.86 17194

macro avg 0.72 0.79 0.73 17194

weighted avg 0.91 0.86 0.88 17194

average eval\_loss: 0.4171299077657073, accuracy: 0.8555891590089566

training 2 epoch...

245.30902910232544 seconds used......

2 training finished! average train loss: 0.42853159768586774

total clf loss: 194.5118391737342 total adv loss: 93.44682266935706

evaluating...

Confusion matrix, without normalization

[[ 757 146 119]

[1115 4153 139]

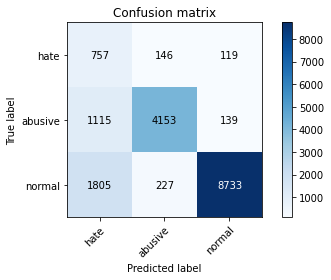
[1805 227 8733]]

Normalized confusion matrix

[[0.7407045 0.14285714 0.11643836]

[0.20621417 0.76807842 0.02570742]

[0.16767301 0.02108686 0.81124013]]





precision recall f1-score support

hate 0.21 0.74 0.32 1022

abusive 0.92 0.77 0.84 5407

normal 0.97 0.81 0.88 10765

accuracy 0.79 17194

macro avg 0.70 0.77 0.68 17194

weighted avg 0.91 0.79 0.84 17194

average eval\_loss: 0.5367862072101859, accuracy: 0.7934744678376178

training 3 epoch...

245.18486547470093 seconds used......

3 training finished! average train loss: 0.30895849894185634

total clf loss: 139.90855875983834 total adv loss: 60.79887073952705

evaluating...

Confusion matrix, without normalization

[[ 707 158 157]

[1002 4271 134]

[1355 314 9096]]

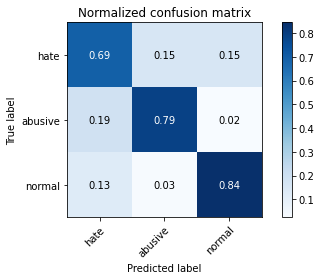
Normalized confusion matrix

[[0.69178082 0.15459883 0.15362035]

[0.18531533 0.78990198 0.02478269]

[0.12587088 0.0291686 0.84496052]]





precision recall f1-score support

hate 0.23 0.69 0.35 1022

abusive 0.90 0.79 0.84 5407

normal 0.97 0.84 0.90 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: 0.5324543381456373, accuracy: 0.818541351634291

training 4 epoch...

245.30403542518616 seconds used......

4 training finished! average train loss: 0.20344474373288604

total clf loss: 91.81207551714033 total adv loss: 33.721080022864044

evaluating...

Confusion matrix, without normalization

[[ 651 207 164]

[ 812 4433 162]

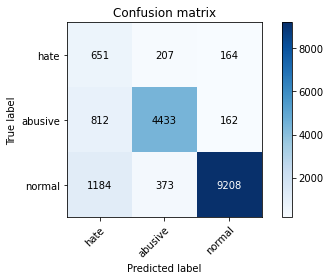
[1184 373 9208]]

Normalized confusion matrix

[[0.6369863 0.20254403 0.16046967]

[0.1501757 0.81986314 0.02996116]

[0.10998607 0.03464933 0.85536461]]





precision recall f1-score support

hate 0.25 0.64 0.35 1022

abusive 0.88 0.82 0.85 5407

normal 0.97 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: 0.614644359170221, accuracy: 0.831220193090613

training 5 epoch...

245.3662610054016 seconds used......

5 training finished! average train loss: 0.12733014953127195

total clf loss: 57.35355722298846 total adv loss: 18.926018902973738

evaluating...

Confusion matrix, without normalization

[[ 697 202 123]

[ 855 4448 104]

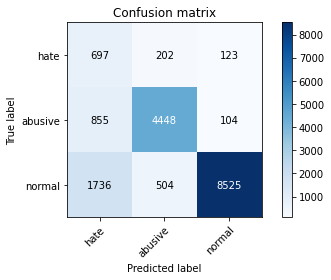
[1736 504 8525]]

Normalized confusion matrix

[[0.68199609 0.19765166 0.12035225]

[0.15812835 0.82263732 0.01923433]

[0.16126335 0.04681839 0.79191825]]





precision recall f1-score support

hate 0.21 0.68 0.32 1022

abusive 0.86 0.82 0.84 5407

normal 0.97 0.79 0.87 10765

accuracy 0.80 17194

macro avg 0.68 0.77 0.68 17194

weighted avg 0.89 0.80 0.83 17194

average eval\_loss: 0.8836249816642944, accuracy: 0.7950447830638595

training 6 epoch...

245.3573923110962 seconds used......

6 training finished! average train loss: 0.09682249933014504

total clf loss: 43.513318676967174 total adv loss: 12.419030436009052

evaluating...

Confusion matrix, without normalization

[[ 693 181 148]

[ 890 4350 167]

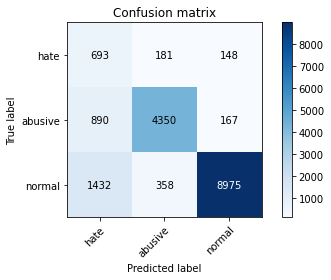
[1432 358 8975]]

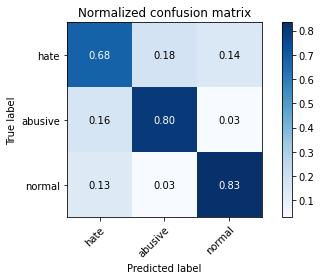
Normalized confusion matrix

[[0.67808219 0.17710372 0.14481409]

[0.16460144 0.80451267 0.03088589]

[0.13302369 0.03325592 0.83372039]]





precision recall f1-score support

hate 0.23 0.68 0.34 1022

abusive 0.89 0.80 0.84 5407

normal 0.97 0.83 0.90 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: 0.8736059053801, accuracy: 0.8152844015354194

training 7 epoch...

245.63951516151428 seconds used......

7 training finished! average train loss: 0.08338815590858661

total clf loss: 37.527042024652474 total adv loss: 11.721780693187611

evaluating...

Confusion matrix, without normalization

[[ 669 194 159]

[ 885 4385 137]

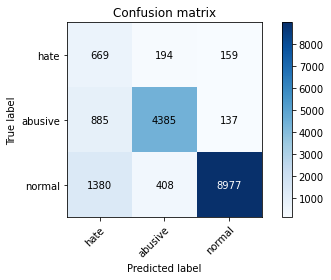
[1380 408 8977]]

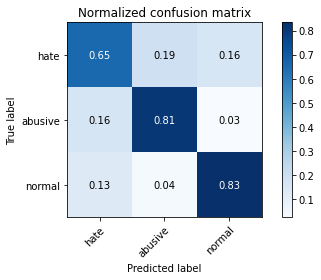
Normalized confusion matrix

[[0.65459883 0.18982387 0.1555773 ]

[0.16367672 0.81098576 0.02533753]

[0.12819322 0.0379006 0.83390618]]





precision recall f1-score support

hate 0.23 0.65 0.34 1022

abusive 0.88 0.81 0.84 5407

normal 0.97 0.83 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: 0.8391406648878961, accuracy: 0.8160404792369431

training 8 epoch...

245.54944562911987 seconds used......

8 training finished! average train loss: 0.07060055056932166

total clf loss: 31.803465326607693 total adv loss: 10.548427535431983

evaluating...

Confusion matrix, without normalization

[[ 729 163 130]

[1171 4089 147]

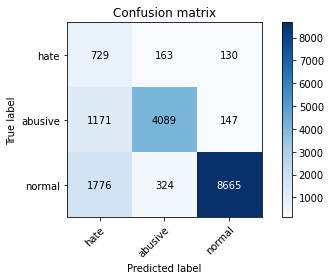
[1776 324 8665]]

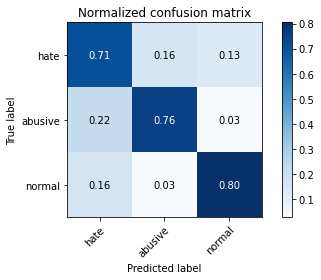
Normalized confusion matrix

[[0.71330724 0.15949119 0.12720157]

[0.21657111 0.75624191 0.02718698]

[0.1649791 0.03009754 0.80492336]]





precision recall f1-score support

hate 0.20 0.71 0.31 1022

abusive 0.89 0.76 0.82 5407

normal 0.97 0.80 0.88 10765

accuracy 0.78 17194

macro avg 0.69 0.76 0.67 17194

weighted avg 0.90 0.78 0.83 17194

average eval\_loss: 1.0332128570772448, accuracy: 0.7841688961265558

training 9 epoch...

245.29720664024353 seconds used......

9 training finished! average train loss: 0.06795944680610112

total clf loss: 30.595595525577664 total adv loss: 9.79121107179435

evaluating...

Confusion matrix, without normalization

[[ 682 174 166]

[1085 4129 193]

[1405 327 9033]]

Normalized confusion matrix

[[0.66731898 0.1702544 0.16242661]

[0.2006658 0.76363973 0.03569447]

[0.13051556 0.03037622 0.83910822]]





precision recall f1-score support

hate 0.22 0.67 0.33 1022

abusive 0.89 0.76 0.82 5407

normal 0.96 0.84 0.90 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: 0.988479026311976, accuracy: 0.8051645922996394

training 10 epoch...

245.33406734466553 seconds used......

10 training finished! average train loss: 0.059783431059632526

total clf loss: 26.898803152085748 total adv loss: 8.294864391998999

evaluating...

Confusion matrix, without normalization

[[ 691 198 133]

[ 965 4306 136]

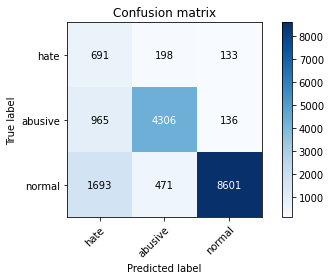
[1693 471 8601]]

Normalized confusion matrix

[[0.67612524 0.19373777 0.13013699]

[0.17847235 0.79637507 0.02515258]

[0.15726893 0.0437529 0.79897817]]





precision recall f1-score support

hate 0.21 0.68 0.32 1022

abusive 0.87 0.80 0.83 5407

normal 0.97 0.80 0.88 10765

accuracy 0.79 17194

macro avg 0.68 0.76 0.67 17194

weighted avg 0.89 0.79 0.83 17194

average eval\_loss: 1.0395447166961467, accuracy: 0.7908572757938815

2 run...

training 1 epoch...

245.45948672294617 seconds used......

1 training finished! average train loss: 0.6008567806358531

total clf loss: 273.3498534858227 total adv loss: 143.40599920228124

evaluating...

Confusion matrix, without normalization

[[ 654 210 158]

[ 694 4584 129]

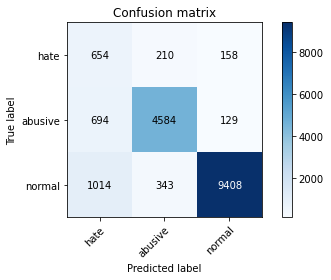
[1014 343 9408]]

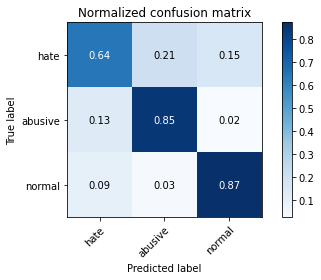
Normalized confusion matrix

[[0.63992172 0.20547945 0.15459883]

[0.12835214 0.8477899 0.02385796]

[0.09419415 0.03186252 0.87394333]]





precision recall f1-score support

hate 0.28 0.64 0.39 1022

abusive 0.89 0.85 0.87 5407

normal 0.97 0.87 0.92 10765

accuracy 0.85 17194

macro avg 0.71 0.79 0.73 17194

weighted avg 0.90 0.85 0.87 17194

average eval\_loss: 0.4384634647085223, accuracy: 0.8518087705013376

training 2 epoch...

245.50087904930115 seconds used......

2 training finished! average train loss: 0.4316252686301957

total clf loss: 196.04165929555893 total adv loss: 96.63330226764083

evaluating...

Confusion matrix, without normalization

[[ 705 163 154]

[ 833 4420 154]

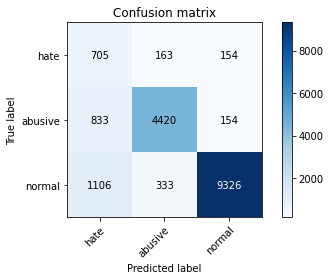
[1106 333 9326]]

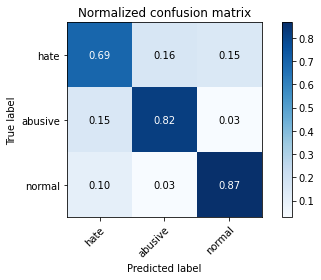
Normalized confusion matrix

[[0.68982387 0.15949119 0.15068493]

[0.15405955 0.81745885 0.0284816 ]

[0.10274036 0.03093358 0.86632606]]





precision recall f1-score support

hate 0.27 0.69 0.38 1022

abusive 0.90 0.82 0.86 5407

normal 0.97 0.87 0.91 10765

accuracy 0.84 17194

macro avg 0.71 0.79 0.72 17194

weighted avg 0.90 0.84 0.86 17194

average eval\_loss: 0.4268126656669517, accuracy: 0.8404676049784808

training 3 epoch...

245.57018661499023 seconds used......

3 training finished! average train loss: 0.32077738329028166

total clf loss: 145.27032750844955 total adv loss: 63.318939361255616

evaluating...

Confusion matrix, without normalization

[[ 687 159 176]

[1004 4219 184]

[1077 271 9417]]

Normalized confusion matrix

[[0.67221135 0.1555773 0.17221135]

[0.18568522 0.78028482 0.03402996]

[0.10004645 0.02517418 0.87477938]]





precision recall f1-score support

hate 0.25 0.67 0.36 1022

abusive 0.91 0.78 0.84 5407

normal 0.96 0.87 0.92 10765

accuracy 0.83 17194

macro avg 0.71 0.78 0.71 17194

weighted avg 0.90 0.83 0.86 17194

average eval\_loss: 0.46550102923065423, accuracy: 0.8330231476096313

training 4 epoch...

245.43317341804504 seconds used......

4 training finished! average train loss: 0.20819961654261998

total clf loss: 93.97134573943913 total adv loss: 34.77831076516304

evaluating...

Confusion matrix, without normalization

[[ 733 145 144]

[1151 4118 138]

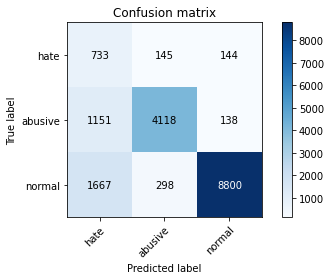
[1667 298 8800]]

Normalized confusion matrix

[[0.71722114 0.14187867 0.1409002 ]

[0.2128722 0.76160533 0.02552247]

[0.15485369 0.0276823 0.817464 ]]





precision recall f1-score support

hate 0.21 0.72 0.32 1022

abusive 0.90 0.76 0.83 5407

normal 0.97 0.82 0.89 10765

accuracy 0.79 17194

macro avg 0.69 0.77 0.68 17194

weighted avg 0.90 0.79 0.83 17194

average eval\_loss: 0.7315786878036898, accuracy: 0.7939397464231709

training 5 epoch...

245.04311323165894 seconds used......

5 training finished! average train loss: 0.13469189215918193

total clf loss: 60.596204451750964 total adv loss: 18.553926699212752

evaluating...

Confusion matrix, without normalization

[[ 665 217 140]

[ 804 4466 137]

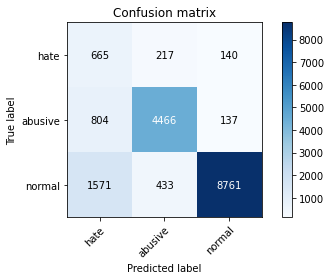
[1571 433 8761]]

Normalized confusion matrix

[[0.65068493 0.21232877 0.1369863 ]

[0.14869613 0.82596634 0.02533753]

[0.1459359 0.04022294 0.81384115]]





precision recall f1-score support

hate 0.22 0.65 0.33 1022

abusive 0.87 0.83 0.85 5407

normal 0.97 0.81 0.88 10765

accuracy 0.81 17194

macro avg 0.69 0.76 0.69 17194

weighted avg 0.89 0.81 0.84 17194

average eval\_loss: 0.7436384374557367, accuracy: 0.807956263812958

training 6 epoch...

245.04048585891724 seconds used......

6 training finished! average train loss: 0.09917481865683407

total clf loss: 44.58415339374915 total adv loss: 12.994175593441469

evaluating...

Confusion matrix, without normalization

[[ 700 217 105]

[ 878 4415 114]

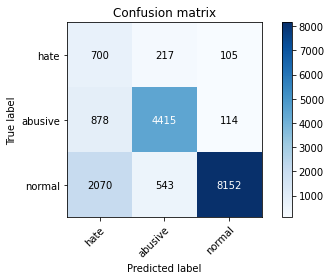
[2070 543 8152]]

Normalized confusion matrix

[[0.68493151 0.21232877 0.10273973]

[0.1623821 0.81653412 0.02108378]

[0.19228983 0.05044124 0.75726893]]





precision recall f1-score support

hate 0.19 0.68 0.30 1022

abusive 0.85 0.82 0.83 5407

normal 0.97 0.76 0.85 10765

accuracy 0.77 17194

macro avg 0.67 0.75 0.66 17194

weighted avg 0.89 0.77 0.81 17194

average eval\_loss: 1.0121467686972991, accuracy: 0.771606374316622

training 7 epoch...

245.26049542427063 seconds used......

7 training finished! average train loss: 0.0844863465554596

total clf loss: 38.03952979610767 total adv loss: 12.241564751711849

evaluating...

Confusion matrix, without normalization

[[ 712 180 130]

[ 995 4252 160]

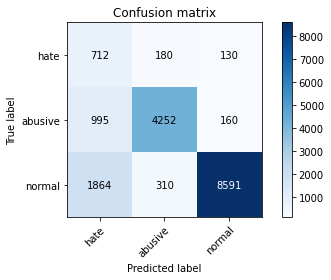
[1864 310 8591]]

Normalized confusion matrix

[[0.69667319 0.17612524 0.12720157]

[0.18402071 0.78638802 0.02959127]

[0.17315374 0.02879703 0.79804923]]





precision recall f1-score support

hate 0.20 0.70 0.31 1022

abusive 0.90 0.79 0.84 5407

normal 0.97 0.80 0.87 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: 1.043974850332462, accuracy: 0.7883564033965337

training 8 epoch...

245.19665122032166 seconds used......

8 training finished! average train loss: 0.07601599787723518

total clf loss: 34.179914093518164 total adv loss: 10.09654053088343

evaluating...

Confusion matrix, without normalization

[[ 660 207 155]

[ 837 4419 151]

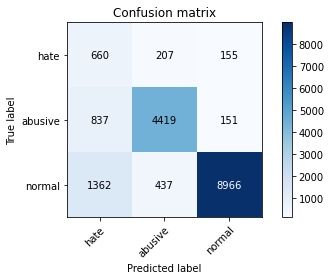
[1362 437 8966]]

Normalized confusion matrix

[[0.64579256 0.20254403 0.15166341]

[0.15479933 0.8172739 0.02792676]

[0.12652113 0.04059452 0.83288435]]





precision recall f1-score support

hate 0.23 0.65 0.34 1022

abusive 0.87 0.82 0.84 5407

normal 0.97 0.83 0.89 10765

accuracy 0.82 17194

macro avg 0.69 0.77 0.69 17194

weighted avg 0.89 0.82 0.85 17194

average eval\_loss: 0.9195268835866998, accuracy: 0.816854716761661

training 9 epoch...

245.41547274589539 seconds used......

9 training finished! average train loss: 0.07048925980844095

total clf loss: 31.821842610952444 total adv loss: 11.902010258381779

evaluating...

Confusion matrix, without normalization

[[ 648 199 175]

[ 859 4374 174]

[1147 408 9210]]

Normalized confusion matrix

[[0.63405088 0.19471624 0.17123288]

[0.15886813 0.80895136 0.03218051]

[0.106549 0.0379006 0.85555039]]





precision recall f1-score support

hate 0.24 0.63 0.35 1022

abusive 0.88 0.81 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 0.89 0.83 0.85 17194

average eval\_loss: 0.8607673893249486, accuracy: 0.8277306036989648

training 10 epoch...

245.1349790096283 seconds used......

10 training finished! average train loss: 0.06505858952954252

total clf loss: 29.300707798975054 total adv loss: 9.595052204480453

evaluating...

Confusion matrix, without normalization

[[ 711 164 147]

[1053 4203 151]

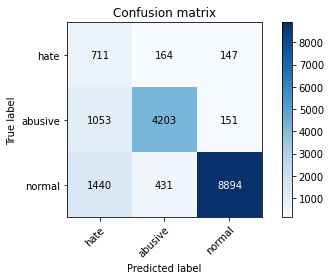
[1440 431 8894]]

Normalized confusion matrix

[[0.69569472 0.16046967 0.14383562]

[0.19474755 0.77732569 0.02792676]

[0.13376684 0.04003716 0.82619601]]





precision recall f1-score support

hate 0.22 0.70 0.34 1022

abusive 0.88 0.78 0.82 5407

normal 0.97 0.83 0.89 10765

accuracy 0.80 17194

macro avg 0.69 0.77 0.68 17194

weighted avg 0.89 0.80 0.84 17194

average eval\_loss: 0.9471141025511992, accuracy: 0.8030708386646505

3 run...

training 1 epoch...

245.1167402267456 seconds used......

1 training finished! average train loss: 0.6238993587935213

total clf loss: 283.7258031964302 total adv loss: 146.76775056496263

evaluating...

Confusion matrix, without normalization

[[ 513 307 202]

[ 353 4867 187]

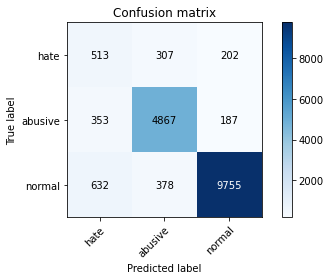
[ 632 378 9755]]

Normalized confusion matrix

[[0.50195695 0.30039139 0.19765166]

[0.06528574 0.90012946 0.0345848 ]

[0.05870878 0.03511379 0.90617743]]





precision recall f1-score support

hate 0.34 0.50 0.41 1022

abusive 0.88 0.90 0.89 5407

normal 0.96 0.91 0.93 10765

accuracy 0.88 17194

macro avg 0.73 0.77 0.74 17194

weighted avg 0.90 0.88 0.89 17194

average eval\_loss: 0.36268083644641, accuracy: 0.8802489240432709

training 2 epoch...

245.27463746070862 seconds used......

2 training finished! average train loss: 0.4458062029685447

total clf loss: 202.51628115028143 total adv loss: 100.48267089389265

evaluating...

Confusion matrix, without normalization

[[ 774 141 107]

[1154 4115 138]

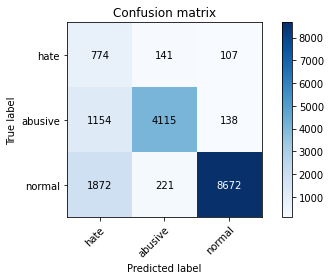
[1872 221 8672]]

Normalized confusion matrix

[[0.75733855 0.13796477 0.10469667]

[0.21342704 0.76105049 0.02552247]

[0.17389689 0.02052949 0.80557362]]





precision recall f1-score support

hate 0.20 0.76 0.32 1022

abusive 0.92 0.76 0.83 5407

normal 0.97 0.81 0.88 10765

accuracy 0.79 17194

macro avg 0.70 0.77 0.68 17194

weighted avg 0.91 0.79 0.83 17194

average eval\_loss: 0.5524658743725267, accuracy: 0.7887053623356985

training 3 epoch...

245.25927066802979 seconds used......

3 training finished! average train loss: 0.32990467338139534

total clf loss: 149.49973315745592 total adv loss: 67.03925402741879

evaluating...

Confusion matrix, without normalization

[[ 703 164 155]

[ 920 4308 179]

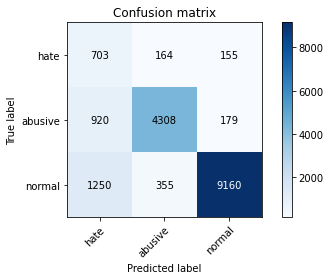
[1250 355 9160]]

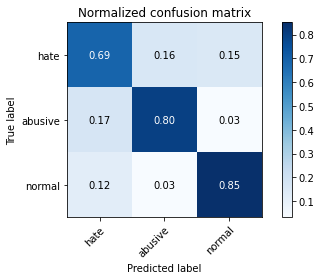
Normalized confusion matrix

[[0.68786693 0.16046967 0.15166341]

[0.17014981 0.79674496 0.03310523]

[0.11611705 0.03297724 0.85090571]]





precision recall f1-score support

hate 0.24 0.69 0.36 1022

abusive 0.89 0.80 0.84 5407

normal 0.96 0.85 0.90 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: 0.5446533392854901, accuracy: 0.8241828544841223

training 4 epoch...

245.21578550338745 seconds used......

4 training finished! average train loss: 0.2127018218763324

total clf loss: 96.05764527991414 total adv loss: 36.61476468434557

evaluating...

Confusion matrix, without normalization

[[ 625 215 182]

[ 732 4507 168]

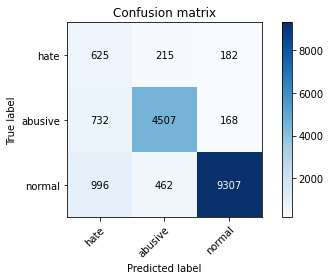
[ 996 462 9307]]

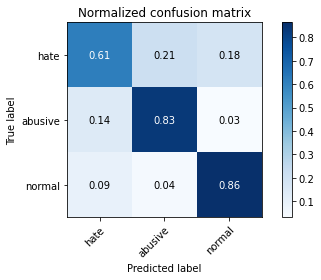
Normalized confusion matrix

[[0.61154599 0.21037182 0.17808219]

[0.13538006 0.8335491 0.03107083]

[0.09252206 0.04291686 0.86456108]]





precision recall f1-score support

hate 0.27 0.61 0.37 1022

abusive 0.87 0.83 0.85 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: 0.5510516964364797, accuracy: 0.8397696871001512

training 5 epoch...

245.32672905921936 seconds used......

5 training finished! average train loss: 0.14323129872402543

total clf loss: 64.53387663979083 total adv loss: 21.648224508360727

evaluating...

Confusion matrix, without normalization

[[ 693 167 162]

[ 993 4264 150]

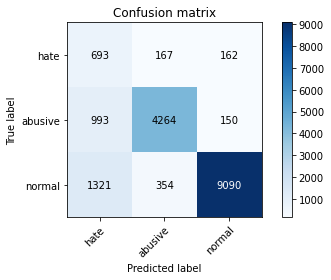
[1321 354 9090]]

Normalized confusion matrix

[[0.67808219 0.16340509 0.15851272]

[0.18365082 0.78860736 0.02774182]

[0.12271249 0.03288435 0.84440316]]





precision recall f1-score support

hate 0.23 0.68 0.34 1022

abusive 0.89 0.79 0.84 5407

normal 0.97 0.84 0.90 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: 0.7070606404590573, accuracy: 0.8169710364080494

training 6 epoch...

245.21999144554138 seconds used......

6 training finished! average train loss: 0.10112741149638194

total clf loss: 45.595005019335076 total adv loss: 15.911234242288629

evaluating...

Confusion matrix, without normalization

[[ 721 183 118]

[1030 4275 102]

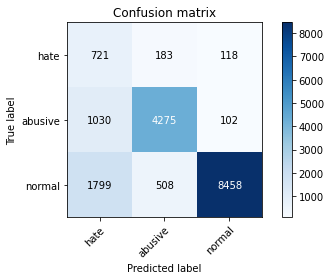
[1799 508 8458]]

Normalized confusion matrix

[[0.70547945 0.17906067 0.11545988]

[0.1904938 0.79064176 0.01886443]

[0.16711565 0.04718997 0.78569438]]





precision recall f1-score support

hate 0.20 0.71 0.32 1022

abusive 0.86 0.79 0.82 5407

normal 0.97 0.79 0.87 10765

accuracy 0.78 17194

macro avg 0.68 0.76 0.67 17194

weighted avg 0.89 0.78 0.82 17194

average eval\_loss: 0.8762991475677767, accuracy: 0.7824822612539258

training 7 epoch...

245.29867315292358 seconds used......

7 training finished! average train loss: 0.08344428519283172

total clf loss: 37.602320288307965 total adv loss: 12.730040990631096

evaluating...

Confusion matrix, without normalization

[[ 614 192 216]

[ 828 4351 228]

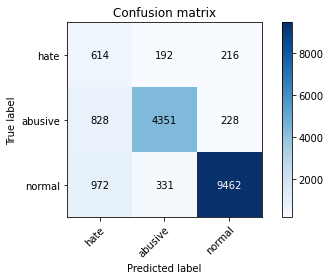
[ 972 331 9462]]

Normalized confusion matrix

[[0.60078278 0.18786693 0.21135029]

[0.15313483 0.80469761 0.04216756]

[0.09029261 0.03074779 0.87895959]]





precision recall f1-score support

hate 0.25 0.60 0.36 1022

abusive 0.89 0.80 0.85 5407

normal 0.96 0.88 0.92 10765

accuracy 0.84 17194

macro avg 0.70 0.76 0.71 17194

weighted avg 0.89 0.84 0.86 17194

average eval\_loss: 0.7228713420057279, accuracy: 0.8390717692218216

training 8 epoch...

245.56503748893738 seconds used......

8 training finished! average train loss: 0.06971682194275067

total clf loss: 31.412495958444197 total adv loss: 10.55887574012013

evaluating...

Confusion matrix, without normalization

[[ 685 211 126]

[ 886 4414 107]

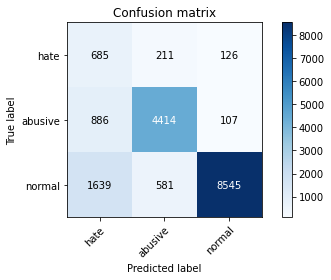
[1639 581 8545]]

Normalized confusion matrix

[[0.6702544 0.20645793 0.12328767]

[0.16386166 0.81634918 0.01978916]

[0.15225267 0.0539712 0.79377613]]





precision recall f1-score support

hate 0.21 0.67 0.32 1022

abusive 0.85 0.82 0.83 5407

normal 0.97 0.79 0.87 10765

accuracy 0.79 17194

macro avg 0.68 0.76 0.68 17194

weighted avg 0.89 0.79 0.83 17194

average eval\_loss: 1.00142280748114, accuracy: 0.7935326276608119

training 9 epoch...

245.4592092037201 seconds used......

9 training finished! average train loss: 0.06193516674595013

total clf loss: 27.93834338756278 total adv loss: 10.021292520683346

evaluating...

Confusion matrix, without normalization

[[ 685 229 108]

[ 887 4416 104]

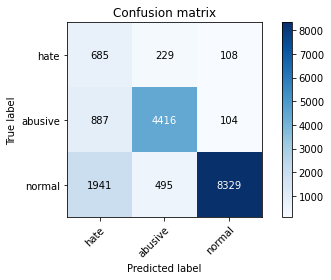
[1941 495 8329]]

Normalized confusion matrix

[[0.6702544 0.22407045 0.10567515]

[0.16404661 0.81671907 0.01923433]

[0.18030655 0.04598235 0.7737111 ]]





precision recall f1-score support

hate 0.19 0.67 0.30 1022

abusive 0.86 0.82 0.84 5407

normal 0.98 0.77 0.86 10765

accuracy 0.78 17194

macro avg 0.68 0.75 0.67 17194

weighted avg 0.89 0.78 0.82 17194

average eval\_loss: 1.0435792318805186, accuracy: 0.7810864254972665

training 10 epoch...

245.47753047943115 seconds used......

10 training finished! average train loss: 0.053469337561219285

total clf loss: 23.97578841430368 total adv loss: 5.7774382939524

evaluating...

Confusion matrix, without normalization

[[ 679 194 149]

[ 981 4275 151]

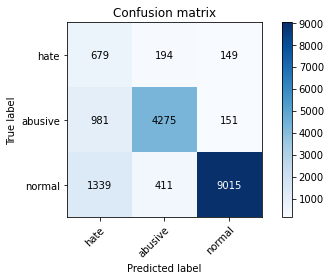
[1339 411 9015]]

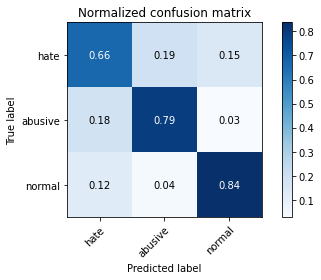
Normalized confusion matrix

[[0.66438356 0.18982387 0.14579256]

[0.18143148 0.79064176 0.02792676]

[0.12438458 0.03817928 0.83743614]]





precision recall f1-score support

hate 0.23 0.66 0.34 1022

abusive 0.88 0.79 0.83 5407

normal 0.97 0.84 0.90 10765

accuracy 0.81 17194

macro avg 0.69 0.76 0.69 17194

weighted avg 0.89 0.81 0.84 17194

average eval\_loss: 1.045611205530017, accuracy: 0.8124345701989066

4 run...

training 1 epoch...

245.419193983078 seconds used......

1 training finished! average train loss: 0.6034885407435975

total clf loss: 274.6641538441181 total adv loss: 146.3746108263731

evaluating...

Confusion matrix, without normalization

[[ 670 190 162]

[ 737 4498 172]

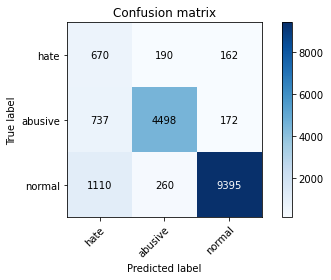
[1110 260 9395]]

Normalized confusion matrix

[[0.6555773 0.18590998 0.15851272]

[0.13630479 0.83188459 0.03181062]

[0.10311194 0.02415235 0.87273572]]





precision recall f1-score support

hate 0.27 0.66 0.38 1022

abusive 0.91 0.83 0.87 5407

normal 0.97 0.87 0.92 10765

accuracy 0.85 17194

macro avg 0.71 0.79 0.72 17194

weighted avg 0.91 0.85 0.87 17194

average eval\_loss: 0.43513686152738196, accuracy: 0.8469815051762243

training 2 epoch...

245.54393696784973 seconds used......

2 training finished! average train loss: 0.4398400113491119

total clf loss: 199.79351218789816 total adv loss: 98.88774412870407

evaluating...

Confusion matrix, without normalization

[[ 732 168 122]

[ 917 4385 105]

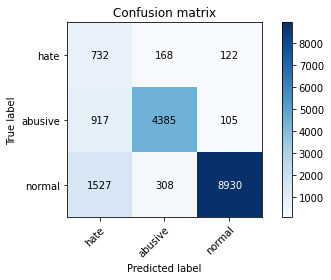
[1527 308 8930]]

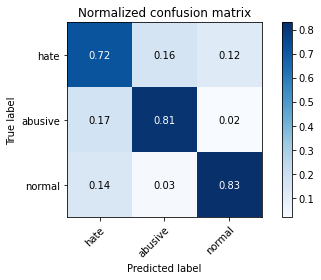
Normalized confusion matrix

[[0.71624266 0.16438356 0.11937378]

[0.16959497 0.81098576 0.01941927]

[0.14184858 0.02861124 0.82954018]]





precision recall f1-score support

hate 0.23 0.72 0.35 1022

abusive 0.90 0.81 0.85 5407

normal 0.98 0.83 0.90 10765

accuracy 0.82 17194

macro avg 0.70 0.79 0.70 17194

weighted avg 0.91 0.82 0.85 17194

average eval\_loss: 0.4851042378087376, accuracy: 0.8169710364080494

training 3 epoch...

245.59208130836487 seconds used......

3 training finished! average train loss: 0.3197485001560124

total clf loss: 144.88970084115863 total adv loss: 64.82230298686773

evaluating...

Confusion matrix, without normalization

[[ 744 105 173]

[1379 3869 159]

[1220 244 9301]]

Normalized confusion matrix

[[0.72798434 0.10273973 0.16927593]

[0.25503976 0.71555391 0.02940633]

[0.11333024 0.02266605 0.86400372]]





precision recall f1-score support

hate 0.22 0.73 0.34 1022

abusive 0.92 0.72 0.80 5407

normal 0.97 0.86 0.91 10765

accuracy 0.81 17194

macro avg 0.70 0.77 0.69 17194

weighted avg 0.91 0.81 0.84 17194

average eval\_loss: 0.5370638815252934, accuracy: 0.809235779923229

training 4 epoch...

245.74306678771973 seconds used......

4 training finished! average train loss: 0.21046408534083624

total clf loss: 94.94849795103073 total adv loss: 34.258164340863004

evaluating...

Confusion matrix, without normalization

[[ 689 170 163]

[ 927 4336 144]

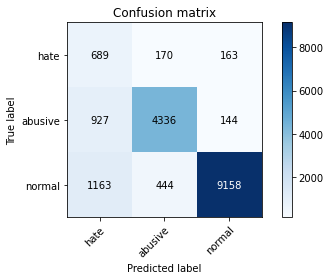
[1163 444 9158]]

Normalized confusion matrix

[[0.6741683 0.16634051 0.15949119]

[0.17144442 0.80192343 0.02663214]

[0.1080353 0.04124477 0.85071993]]





precision recall f1-score support

hate 0.25 0.67 0.36 1022

abusive 0.88 0.80 0.84 5407

normal 0.97 0.85 0.91 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: 0.6012508540089394, accuracy: 0.824880772362452

training 5 epoch...

245.83802819252014 seconds used......

5 training finished! average train loss: 0.13566063145705357

total clf loss: 61.062502895947546 total adv loss: 19.296862839662936

evaluating...

Confusion matrix, without normalization

[[ 729 144 149]

[1236 3998 173]

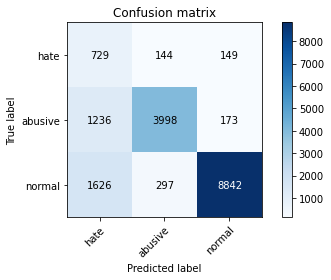
[1626 297 8842]]

Normalized confusion matrix

[[0.71330724 0.1409002 0.14579256]

[0.22859257 0.73941187 0.03199556]

[0.15104505 0.02758941 0.82136554]]





precision recall f1-score support

hate 0.20 0.71 0.32 1022

abusive 0.90 0.74 0.81 5407

normal 0.96 0.82 0.89 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: 0.7187113287707054, accuracy: 0.7891706409212516

training 6 epoch...

245.36496305465698 seconds used......

6 training finished! average train loss: 0.10155821555035513

total clf loss: 45.6070586736314 total adv loss: 12.335382744844537

evaluating...

Confusion matrix, without normalization

[[ 728 183 111]

[ 886 4392 129]

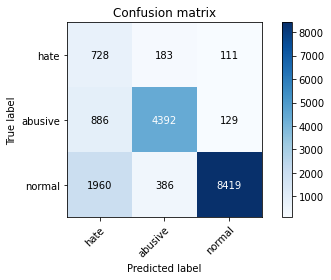
[1960 386 8419]]

Normalized confusion matrix

[[0.71232877 0.17906067 0.10861057]

[0.16386166 0.81228038 0.02385796]

[0.18207153 0.03585694 0.78207153]]





precision recall f1-score support

hate 0.20 0.71 0.32 1022

abusive 0.89 0.81 0.85 5407

normal 0.97 0.78 0.87 10765

accuracy 0.79 17194

macro avg 0.69 0.77 0.68 17194

weighted avg 0.90 0.79 0.83 17194

average eval\_loss: 0.9631602672405194, accuracy: 0.7874258462254274

training 7 epoch...

245.47794675827026 seconds used......

7 training finished! average train loss: 0.07926168762363964

total clf loss: 35.661810565856285 total adv loss: 10.977661089025787

evaluating...

Confusion matrix, without normalization

[[ 629 244 149]

[ 731 4543 133]

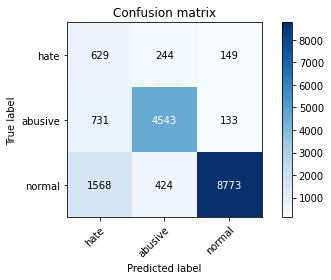
[1568 424 8773]]

Normalized confusion matrix

[[0.61545988 0.23874755 0.14579256]

[0.13519512 0.84020714 0.02459774]

[0.14565722 0.0393869 0.81495588]]





precision recall f1-score support

hate 0.21 0.62 0.32 1022

abusive 0.87 0.84 0.86 5407

normal 0.97 0.81 0.89 10765

accuracy 0.81 17194

macro avg 0.69 0.76 0.69 17194

weighted avg 0.89 0.81 0.84 17194

average eval\_loss: 0.9499871088987822, accuracy: 0.8110387344422473

training 8 epoch...

245.667973279953 seconds used......

8 training finished! average train loss: 0.07265788547034559

total clf loss: 32.69385134289041 total adv loss: 10.128160130232573

evaluating...

Confusion matrix, without normalization

[[ 736 164 122]

[1016 4275 116]

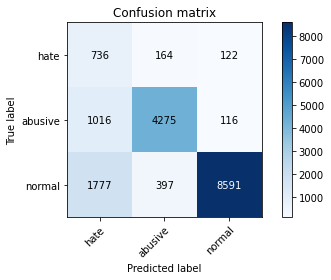
[1777 397 8591]]

Normalized confusion matrix

[[0.72015656 0.16046967 0.11937378]

[0.18790457 0.79064176 0.02145367]

[0.16507199 0.03687877 0.79804923]]





precision recall f1-score support

hate 0.21 0.72 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 0.77 0.68 17194

weighted avg 0.90 0.79 0.83 17194

average eval\_loss: 1.0817919767461717, accuracy: 0.7910899150866582

training 9 epoch...

245.70448207855225 seconds used......

9 training finished! average train loss: 0.06403018819788654

total clf loss: 28.92437172285281 total adv loss: 11.179968398602796

evaluating...

Confusion matrix, without normalization

[[ 694 187 141]

[ 939 4335 133]

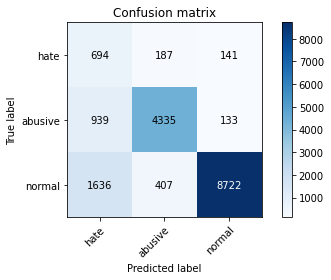
[1636 407 8722]]

Normalized confusion matrix

[[0.67906067 0.18297456 0.13796477]

[0.17366377 0.80173849 0.02459774]

[0.15197399 0.03780771 0.8102183 ]]





precision recall f1-score support

hate 0.21 0.68 0.32 1022

abusive 0.88 0.80 0.84 5407

normal 0.97 0.81 0.88 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: 1.0091055186026794, accuracy: 0.7997557287425846

training 10 epoch...

245.6213662624359 seconds used......

10 training finished! average train loss: 0.049984783144748156

total clf loss: 22.436849322693888 total adv loss: 5.871807156221621

evaluating...

Confusion matrix, without normalization

[[ 676 202 144]

[ 798 4447 162]

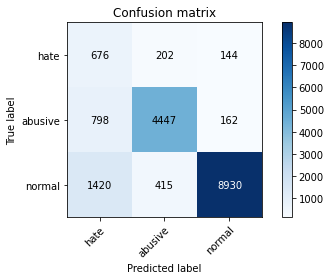
[1420 415 8930]]

Normalized confusion matrix

[[0.66144814 0.19765166 0.1409002 ]

[0.14758646 0.82245238 0.02996116]

[0.13190896 0.03855086 0.82954018]]





precision recall f1-score support

hate 0.23 0.66 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 0.69 0.77 0.70 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 0.9774131005770679, accuracy: 0.8173199953472141

5 run...

training 1 epoch...

245.47551941871643 seconds used......

1 training finished! average train loss: 0.5818135486095808

total clf loss: 264.6990792900324 total adv loss: 139.11354921758175

evaluating...

Confusion matrix, without normalization

[[ 561 189 272]

[ 623 4507 277]

[ 414 263 10088]]

Normalized confusion matrix

[[0.54892368 0.18493151 0.26614481]

[0.11522101 0.8335491 0.05122989]

[0.03845797 0.02443103 0.93711101]]





precision recall f1-score support

hate 0.35 0.55 0.43 1022

abusive 0.91 0.83 0.87 5407

normal 0.95 0.94 0.94 10765

accuracy 0.88 17194

macro avg 0.74 0.77 0.75 17194

weighted avg 0.90 0.88 0.89 17194

average eval\_loss: 0.37172235227948014, accuracy: 0.8814702803303478

training 2 epoch...

245.27599430084229 seconds used......

2 training finished! average train loss: 0.4372285502893392

total clf loss: 198.5899013876915 total adv loss: 97.95307917147875

evaluating...

Confusion matrix, without normalization

[[ 668 204 150]

[ 662 4623 122]

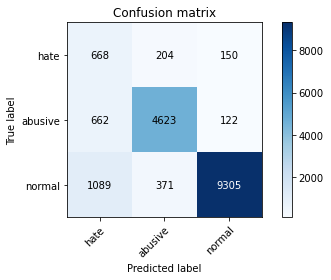
[1089 371 9305]]

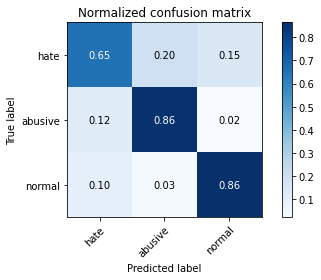
Normalized confusion matrix

[[0.65362035 0.19960861 0.14677104]

[0.12243388 0.85500277 0.02256334]

[0.10116117 0.03446354 0.86437529]]





precision recall f1-score support

hate 0.28 0.65 0.39 1022

abusive 0.89 0.86 0.87 5407

normal 0.97 0.86 0.91 10765

accuracy 0.85 17194

macro avg 0.71 0.79 0.72 17194

weighted avg 0.90 0.85 0.87 17194

average eval\_loss: 0.4254043746219818, accuracy: 0.8489007793416308

training 3 epoch...

245.39031219482422 seconds used......

3 training finished! average train loss: 0.31956785796189147

total clf loss: 144.82362655550241 total adv loss: 65.10131175769493

evaluating...

Confusion matrix, without normalization

[[ 593 261 168]

[ 503 4762 142]

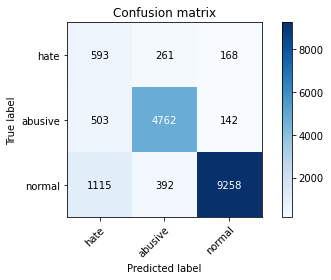
[1115 392 9258]]

Normalized confusion matrix

[[0.58023483 0.2553816 0.16438356]

[0.09302756 0.88071019 0.02626225]

[0.10357641 0.03641431 0.86000929]]





precision recall f1-score support

hate 0.27 0.58 0.37 1022

abusive 0.88 0.88 0.88 5407

normal 0.97 0.86 0.91 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: 0.43168875119800487, accuracy: 0.8498894963359311

training 4 epoch...

245.17726135253906 seconds used......

4 training finished! average train loss: 0.2065669315431948

total clf loss: 93.27519644238055 total adv loss: 35.32091457117349

evaluating...

Confusion matrix, without normalization

[[ 699 210 113]

[ 824 4500 83]

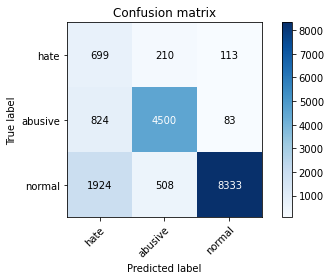
[1924 508 8333]]

Normalized confusion matrix

[[0.68395303 0.20547945 0.11056751]

[0.15239504 0.83225448 0.01535047]

[0.17872736 0.04718997 0.77408268]]





precision recall f1-score support

hate 0.20 0.68 0.31 1022

abusive 0.86 0.83 0.85 5407

normal 0.98 0.77 0.86 10765

accuracy 0.79 17194

macro avg 0.68 0.76 0.67 17194

weighted avg 0.89 0.79 0.83 17194

average eval\_loss: 0.725045263368723, accuracy: 0.7870187274630686

training 5 epoch...

245.2008228302002 seconds used......

5 training finished! average train loss: 0.14148927226443997

total clf loss: 63.7124640783295 total adv loss: 20.654335585641093

evaluating...

Confusion matrix, without normalization

[[ 646 186 190]

[ 874 4314 219]

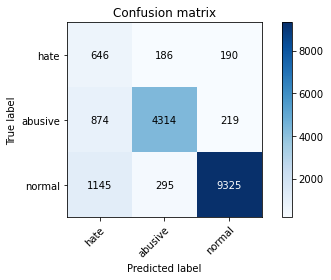
[1145 295 9325]]

Normalized confusion matrix

[[0.63209393 0.18199609 0.18590998]

[0.16164232 0.79785463 0.04050305]

[0.10636321 0.02740362 0.86623316]]





precision recall f1-score support

hate 0.24 0.63 0.35 1022

abusive 0.90 0.80 0.85 5407

normal 0.96 0.87 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: 0.6072405657471093, accuracy: 0.830813074328254

training 6 epoch...

245.44832563400269 seconds used......

6 training finished! average train loss: 0.09913528247727985

total clf loss: 44.63402810576372 total adv loss: 14.341959534212947

evaluating...

Confusion matrix, without normalization

[[ 660 207 155]

[ 861 4397 149]

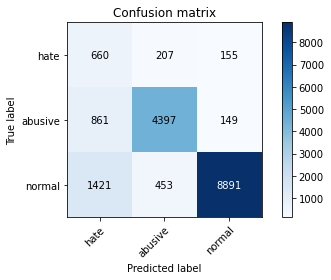
[1421 453 8891]]

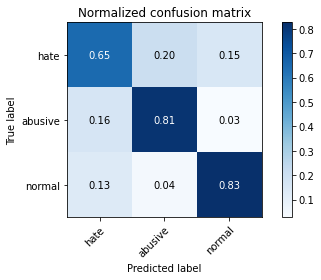
Normalized confusion matrix

[[0.64579256 0.20254403 0.15166341]

[0.15923802 0.8132051 0.02755687]

[0.13200186 0.04208082 0.82591732]]





precision recall f1-score support

hate 0.22 0.65 0.33 1022

abusive 0.87 0.81 0.84 5407

normal 0.97 0.83 0.89 10765

accuracy 0.81 17194

macro avg 0.69 0.76 0.69 17194

weighted avg 0.89 0.81 0.84 17194

average eval\_loss: 0.7728087993013825, accuracy: 0.8112132139118297

training 7 epoch...

245.5453498363495 seconds used......

7 training finished! average train loss: 0.07922153750173831

total clf loss: 35.69828946469352 total adv loss: 12.062964999007818

evaluating...

Confusion matrix, without normalization

[[ 655 192 175]

[ 861 4368 178]

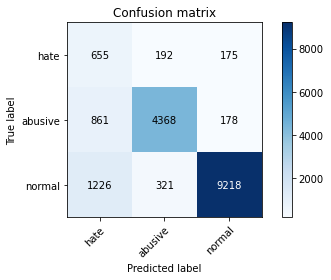
[1226 321 9218]]

Normalized confusion matrix

[[0.6409002 0.18786693 0.17123288]

[0.15923802 0.80784169 0.03292029]

[0.1138876 0.02981886 0.85629354]]





precision recall f1-score support

hate 0.24 0.64 0.35 1022

abusive 0.89 0.81 0.85 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: 0.7830479281570026, accuracy: 0.828254042107712

training 8 epoch...

245.43967199325562 seconds used......

8 training finished! average train loss: 0.07005527186334663

total clf loss: 31.595418472075835 total adv loss: 11.218660921815172

evaluating...

Confusion matrix, without normalization

[[ 526 290 206]

[ 601 4612 194]

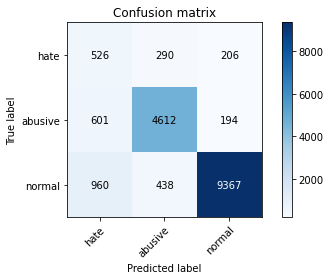
[ 960 438 9367]]

Normalized confusion matrix

[[0.5146771 0.28375734 0.20156556]

[0.11115221 0.85296837 0.03587942]

[0.08917789 0.04068741 0.8701347 ]]





precision recall f1-score support

hate 0.25 0.51 0.34 1022

abusive 0.86 0.85 0.86 5407

normal 0.96 0.87 0.91 10765

accuracy 0.84 17194

macro avg 0.69 0.75 0.70 17194

weighted avg 0.89 0.84 0.86 17194

average eval\_loss: 0.7320261414471451, accuracy: 0.8436082354309643

training 9 epoch...

245.29861330986023 seconds used......

9 training finished! average train loss: 0.06605902325591119

total clf loss: 29.83490911591798 total adv loss: 11.415236626689875

evaluating...

Confusion matrix, without normalization

[[ 722 201 99]

[1075 4214 118]

[2364 399 8002]]

Normalized confusion matrix

[[0.70645793 0.19667319 0.09686888]

[0.19881635 0.77936009 0.02182356]

[0.21960056 0.03706456 0.74333488]]





precision recall f1-score support

hate 0.17 0.71 0.28 1022

abusive 0.88 0.78 0.82 5407

normal 0.97 0.74 0.84 10765

accuracy 0.75 17194

macro avg 0.67 0.74 0.65 17194

weighted avg 0.90 0.75 0.80 17194

average eval\_loss: 1.214897261804706, accuracy: 0.7524717924857508

training 10 epoch...

245.56522178649902 seconds used......

10 training finished! average train loss: 0.07053659593800249

total clf loss: 31.864805316610727 total adv loss: 12.341866003105679

evaluating...

Confusion matrix, without normalization

[[ 616 269 137]

[ 730 4557 120]

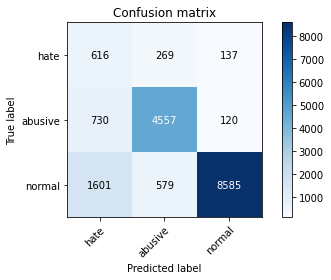
[1601 579 8585]]

Normalized confusion matrix

[[0.60273973 0.26320939 0.13405088]

[0.13501017 0.84279638 0.02219345]

[0.14872271 0.05378542 0.79749187]]





precision recall f1-score support

hate 0.21 0.60 0.31 1022

abusive 0.84 0.84 0.84 5407

normal 0.97 0.80 0.88 10765

accuracy 0.80 17194

macro avg 0.67 0.75 0.68 17194

weighted avg 0.89 0.80 0.83 17194

average eval\_loss: 1.031621486193847, accuracy: 0.8001628475049436