hate as offensive

alpha=0.02

1 run...

training 1 epoch...

1198.1918442249298 seconds used......

1 training finished! average train loss: 0.3764273300821568

total clf loss: 520.0133537910879 total adv loss: 328.323765469715

evaluating...

Confusion matrix, without normalization

[[ 355 388 279]

[ 132 5076 199]

[ 187 361 10217]]

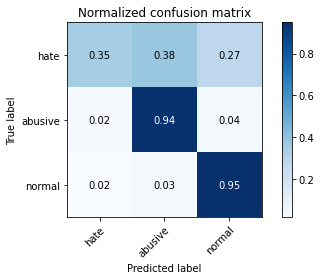
Normalized confusion matrix

[[0.34735812 0.37964775 0.27299413]

[0.0244128 0.93878306 0.03680414]

[0.01737111 0.0335346 0.94909429]]





precision recall f1-score support

hate 0.53 0.35 0.42 1022

abusive 0.87 0.94 0.90 5407

normal 0.96 0.95 0.95 10765

accuracy 0.91 17194

macro avg 0.78 0.75 0.76 17194

weighted avg 0.90 0.91 0.91 17194

average eval\_loss: 0.28001364083570796, accuracy: 0.9100849133418635

training 2 epoch...

1198.211582660675 seconds used......

2 training finished! average train loss: 0.2905959526385533

total clf loss: 401.22066542319953 total adv loss: 242.3893041005358

evaluating...

Confusion matrix, without normalization

[[ 402 351 269]

[ 183 4990 234]

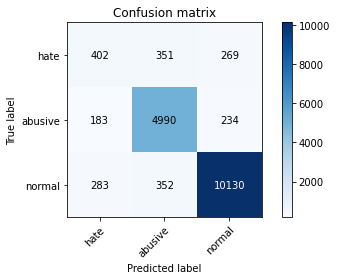
[ 283 352 10130]]

Normalized confusion matrix

[[0.39334638 0.34344423 0.26320939]

[0.03384502 0.92287775 0.04327723]

[0.0262889 0.03269856 0.94101254]]





precision recall f1-score support

hate 0.46 0.39 0.43 1022

abusive 0.88 0.92 0.90 5407

normal 0.95 0.94 0.95 10765

accuracy 0.90 17194

macro avg 0.76 0.75 0.76 17194

weighted avg 0.90 0.90 0.90 17194

average eval\_loss: 0.2948060578774921, accuracy: 0.9027567756194022

training 3 epoch...

1198.4861738681793 seconds used......

3 training finished! average train loss: 0.21370968045093272

total clf loss: 294.74685875698924 total adv loss: 162.34273154847324

evaluating...

Confusion matrix, without normalization

[[ 453 336 233]

[ 264 4978 165]

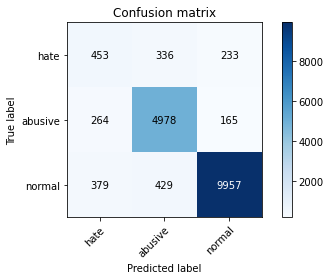
[ 379 429 9957]]

Normalized confusion matrix

[[0.44324853 0.32876712 0.22798434]

[0.0488256 0.92065841 0.030516 ]

[0.03520669 0.03985137 0.92494194]]





precision recall f1-score support

hate 0.41 0.44 0.43 1022

abusive 0.87 0.92 0.89 5407

normal 0.96 0.92 0.94 10765

accuracy 0.89 17194

macro avg 0.75 0.76 0.75 17194

weighted avg 0.90 0.89 0.90 17194

average eval\_loss: 0.34162367269786637, accuracy: 0.8949633593113877

training 4 epoch...

1197.3597247600555 seconds used......

4 training finished! average train loss: 0.14265260568279173

total clf loss: 196.54642121307552 total adv loss: 98.41335599124432

evaluating...

Confusion matrix, without normalization

[[ 380 325 317]

[ 250 4798 359]

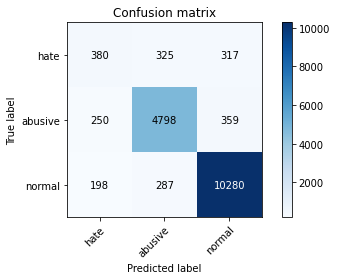
[ 198 287 10280]]

Normalized confusion matrix

[[0.37181996 0.31800391 0.31017613]

[0.04623636 0.88736823 0.06639541]

[0.01839294 0.02666047 0.95494659]]





precision recall f1-score support

hate 0.46 0.37 0.41 1022

abusive 0.89 0.89 0.89 5407

normal 0.94 0.95 0.95 10765

accuracy 0.90 17194

macro avg 0.76 0.74 0.75 17194

weighted avg 0.89 0.90 0.90 17194

average eval\_loss: 0.4050416619086443, accuracy: 0.8990345469349773

training 5 epoch...

1197.056892156601 seconds used......

5 training finished! average train loss: 0.09793749697671321

total clf loss: 134.8890201319009 total adv loss: 65.11371232185047

evaluating...

Confusion matrix, without normalization

[[ 524 266 232]

[ 495 4668 244]

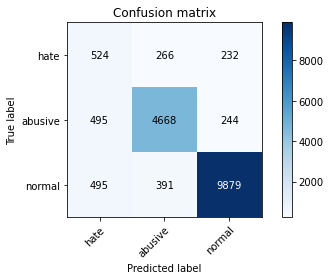
[ 495 391 9879]]

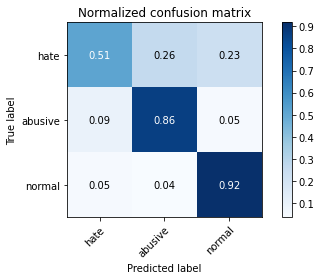
Normalized confusion matrix

[[0.51272016 0.26027397 0.22700587]

[0.09154799 0.86332532 0.04512669]

[0.04598235 0.03632141 0.91769624]]





precision recall f1-score support

hate 0.35 0.51 0.41 1022

abusive 0.88 0.86 0.87 5407

normal 0.95 0.92 0.94 10765

accuracy 0.88 17194

macro avg 0.73 0.76 0.74 17194

weighted avg 0.89 0.88 0.88 17194

average eval\_loss: 0.5043588092902589, accuracy: 0.8765266953588461

2 run...

training 1 epoch...

1196.9877934455872 seconds used......

1 training finished! average train loss: 0.3817492321152732

total clf loss: 527.3683476690203 total adv loss: 333.1197241349146

evaluating...

Confusion matrix, without normalization

[[ 384 294 344]

[ 235 4888 284]

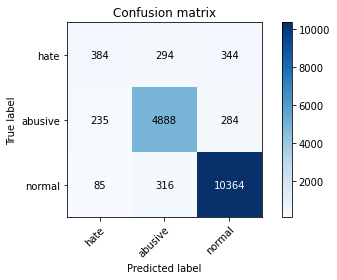
[ 85 316 10364]]

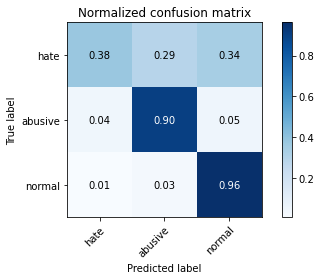
Normalized confusion matrix

[[0.37573386 0.28767123 0.33659491]

[0.04346218 0.90401332 0.05252451]

[0.00789596 0.02935439 0.96274965]]





precision recall f1-score support

hate 0.55 0.38 0.44 1022

abusive 0.89 0.90 0.90 5407

normal 0.94 0.96 0.95 10765

accuracy 0.91 17194

macro avg 0.79 0.75 0.76 17194

weighted avg 0.90 0.91 0.90 17194

average eval\_loss: 0.2764604880409532, accuracy: 0.9093869954635337

training 2 epoch...

1197.1026923656464 seconds used......

2 training finished! average train loss: 0.2930356831989003

total clf loss: 404.5680013112724 total adv loss: 243.3664777847007

evaluating...

Confusion matrix, without normalization

[[ 397 343 282]

[ 187 4994 226]

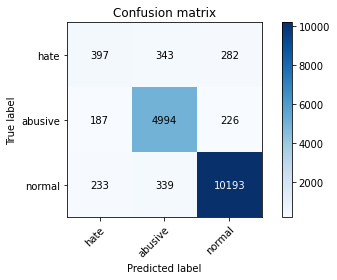
[ 233 339 10193]]

Normalized confusion matrix

[[0.38845401 0.33561644 0.27592955]

[0.0345848 0.92361753 0.04179767]

[0.02164422 0.03149094 0.94686484]]





precision recall f1-score support

hate 0.49 0.39 0.43 1022

abusive 0.88 0.92 0.90 5407

normal 0.95 0.95 0.95 10765

accuracy 0.91 17194

macro avg 0.77 0.75 0.76 17194

weighted avg 0.90 0.91 0.90 17194

average eval\_loss: 0.2841996838793505, accuracy: 0.9063626846574386

training 3 epoch...

1196.7689714431763 seconds used......

3 training finished! average train loss: 0.21725759375840425

total clf loss: 299.64421089179814 total adv loss: 165.24264845927246

evaluating...

Confusion matrix, without normalization

[[ 440 351 231]

[ 238 4988 181]

[ 427 423 9915]]

Normalized confusion matrix

[[0.43052838 0.34344423 0.2260274 ]

[0.04401701 0.92250786 0.03347512]

[0.03966558 0.03929401 0.92104041]]





precision recall f1-score support

hate 0.40 0.43 0.41 1022

abusive 0.87 0.92 0.89 5407

normal 0.96 0.92 0.94 10765

accuracy 0.89 17194

macro avg 0.74 0.76 0.75 17194

weighted avg 0.90 0.89 0.89 17194

average eval\_loss: 0.3385555544391621, accuracy: 0.8923461672676515

training 4 epoch...

1197.466285943985 seconds used......

4 training finished! average train loss: 0.144012836991711

total clf loss: 198.42826137878 total adv loss: 99.73757982312236

evaluating...

Confusion matrix, without normalization

[[ 516 270 236]

[ 559 4678 170]

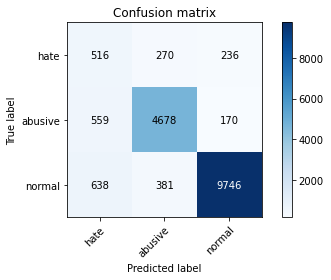
[ 638 381 9746]]

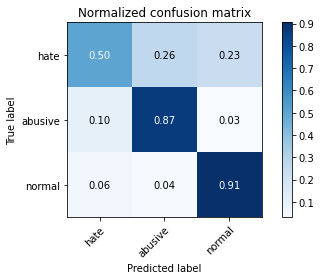
Normalized confusion matrix

[[0.50489237 0.26418787 0.23091977]

[0.1033845 0.86517477 0.03144072]

[0.05926614 0.03539248 0.90534138]]





precision recall f1-score support

hate 0.30 0.50 0.38 1022

abusive 0.88 0.87 0.87 5407

normal 0.96 0.91 0.93 10765

accuracy 0.87 17194

macro avg 0.71 0.76 0.73 17194

weighted avg 0.90 0.87 0.88 17194

average eval\_loss: 0.4485338818005741, accuracy: 0.8689077585204141

training 5 epoch...

1197.6617958545685 seconds used......

5 training finished! average train loss: 0.09922636076763594

total clf loss: 136.67168500076514 total adv loss: 66.3464428804291

evaluating...

Confusion matrix, without normalization

[[ 354 394 274]

[ 262 4913 232]

[ 271 482 10012]]

Normalized confusion matrix

[[0.34637965 0.38551859 0.26810176]

[0.04845571 0.90863695 0.04290734]

[0.02517418 0.04477473 0.93005109]]





precision recall f1-score support

hate 0.40 0.35 0.37 1022

abusive 0.85 0.91 0.88 5407

normal 0.95 0.93 0.94 10765

accuracy 0.89 17194

macro avg 0.73 0.73 0.73 17194

weighted avg 0.89 0.89 0.89 17194

average eval\_loss: 0.45849850190787167, accuracy: 0.8886239385832267