input for adv: mean of last 2 hidden layer

training 1 epoch...

/opt/conda/lib/python3.7/site-packages/ipykernel\_launcher.py:41: UserWarning: To copy construct from a tensor, it is recommended to use sourceTensor.clone().detach() or sourceTensor.clone().detach().requires\_grad\_(True), rather than torch.tensor(sourceTensor).

95.15346026420593 seconds used......

1 training finished! average train loss: 0.6366669593150156

total clf loss: 150.2596579194069 total adv loss: 101.99181991070509

evaluating...

Confusion matrix, without normalization

[[ 197 64 33]

[ 489 3154 178]

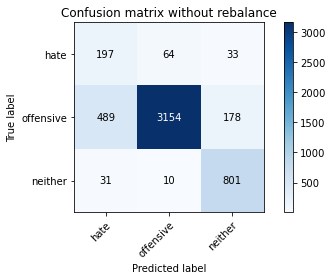
[ 31 10 801]]

Normalized confusion matrix

[[0.67006803 0.21768707 0.1122449 ]

[0.12797697 0.82543837 0.04658466]

[0.0368171 0.01187648 0.95130641]]





precision recall f1-score support

hate 0.27 0.67 0.39 294

offensive 0.98 0.83 0.89 3821

neither 0.79 0.95 0.86 842

accuracy 0.84 4957

macro avg 0.68 0.82 0.72 4957

weighted avg 0.90 0.84 0.86 4957

average eval\_loss: 0.43984223776173376, accuracy: 0.8376033891466613

training 2 epoch...

/opt/conda/lib/python3.7/site-packages/ipykernel\_launcher.py:41: UserWarning: To copy construct from a tensor, it is recommended to use sourceTensor.clone().detach() or sourceTensor.clone().detach().requires\_grad\_(True), rather than torch.tensor(sourceTensor).

95.17858052253723 seconds used......

2 training finished! average train loss: 0.4036648085896383

total clf loss: 95.98172277957201 total adv loss: 78.92292558960617

evaluating...

Confusion matrix, without normalization

[[ 250 27 17]

[1259 2452 110]

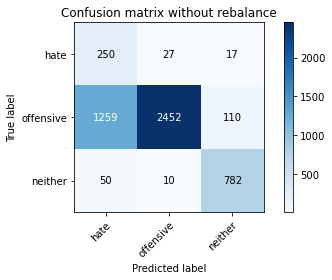
[ 50 10 782]]

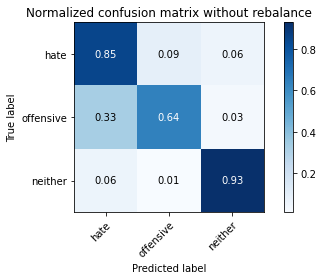
Normalized confusion matrix

[[0.85034014 0.09183673 0.05782313]

[0.3294949 0.64171683 0.02878828]

[0.05938242 0.01187648 0.92874109]]





precision recall f1-score support

hate 0.16 0.85 0.27 294

offensive 0.99 0.64 0.78 3821

neither 0.86 0.93 0.89 842

accuracy 0.70 4957

macro avg 0.67 0.81 0.65 4957

weighted avg 0.92 0.70 0.77 4957

average eval\_loss: 0.6913720693923195, accuracy: 0.7028444623764374

training 3 epoch...

/opt/conda/lib/python3.7/site-packages/ipykernel\_launcher.py:41: UserWarning: To copy construct from a tensor, it is recommended to use sourceTensor.clone().detach() or sourceTensor.clone().detach().requires\_grad\_(True), rather than torch.tensor(sourceTensor).

95.15305948257446 seconds used......

3 training finished! average train loss: 0.2759058227631868

total clf loss: 65.99666437506676 total adv loss: 61.80273429490626

evaluating...

Confusion matrix, without normalization

[[ 239 32 23]

[ 952 2764 105]

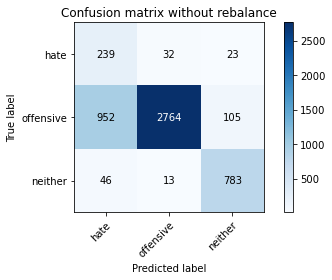
[ 46 13 783]]

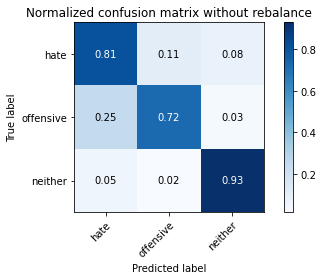
Normalized confusion matrix

[[0.81292517 0.10884354 0.07823129]

[0.24914944 0.72337085 0.02747972]

[0.05463183 0.01543943 0.92992874]]





precision recall f1-score support

hate 0.19 0.81 0.31 294

offensive 0.98 0.72 0.83 3821

neither 0.86 0.93 0.89 842

accuracy 0.76 4957

macro avg 0.68 0.82 0.68 4957

weighted avg 0.92 0.76 0.81 4957

average eval\_loss: 0.5961517001701986, accuracy: 0.7637684083114787

training 4 epoch...

/opt/conda/lib/python3.7/site-packages/ipykernel\_launcher.py:41: UserWarning: To copy construct from a tensor, it is recommended to use sourceTensor.clone().detach() or sourceTensor.clone().detach().requires\_grad\_(True), rather than torch.tensor(sourceTensor).

95.25751042366028 seconds used......

4 training finished! average train loss: 0.18480687577015997

total clf loss: 44.34773425012827 total adv loss: 44.23533102776855

evaluating...

Confusion matrix, without normalization

[[ 177 95 22]

[ 284 3394 143]

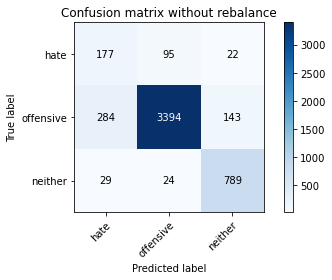
[ 29 24 789]]

Normalized confusion matrix

[[0.60204082 0.32312925 0.07482993]

[0.07432609 0.88824915 0.03742476]

[0.03444181 0.02850356 0.93705463]]





precision recall f1-score support

hate 0.36 0.60 0.45 294

offensive 0.97 0.89 0.93 3821

neither 0.83 0.94 0.88 842

accuracy 0.88 4957

macro avg 0.72 0.81 0.75 4957

weighted avg 0.91 0.88 0.89 4957

average eval\_loss: 0.3759788933359199, accuracy: 0.8795642525721202

training 5 epoch...

/opt/conda/lib/python3.7/site-packages/ipykernel\_launcher.py:41: UserWarning: To copy construct from a tensor, it is recommended to use sourceTensor.clone().detach() or sourceTensor.clone().detach().requires\_grad\_(True), rather than torch.tensor(sourceTensor).

95.39788794517517 seconds used......

5 training finished! average train loss: 0.13194728699888578

total clf loss: 31.79312057327479 total adv loss: 34.182782833464444

evaluating...

Confusion matrix, without normalization

[[ 226 51 17]

[ 822 2900 99]

[ 44 28 770]]

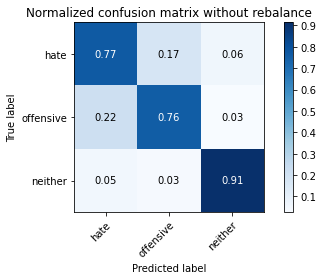
Normalized confusion matrix

[[0.76870748 0.17346939 0.05782313]

[0.21512693 0.75896362 0.02590945]

[0.05225653 0.03325416 0.91448931]]





precision recall f1-score support

hate 0.21 0.77 0.33 294

offensive 0.97 0.76 0.85 3821

neither 0.87 0.91 0.89 842

accuracy 0.79 4957

macro avg 0.68 0.81 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.6036279108825497, accuracy: 0.7859592495460964

training 6 epoch...

/opt/conda/lib/python3.7/site-packages/ipykernel\_launcher.py:41: UserWarning: To copy construct from a tensor, it is recommended to use sourceTensor.clone().detach() or sourceTensor.clone().detach().requires\_grad\_(True), rather than torch.tensor(sourceTensor).

95.31675243377686 seconds used......

6 training finished! average train loss: 0.0693093866126187

total clf loss: 16.851714711869135 total adv loss: 20.983491585124284

evaluating...

Confusion matrix, without normalization

[[ 218 57 19]

[ 654 3058 109]

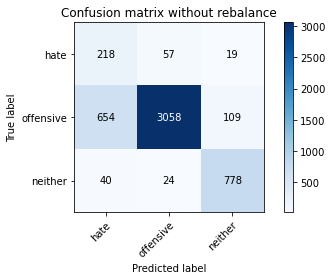
[ 40 24 778]]

Normalized confusion matrix

[[0.7414966 0.19387755 0.06462585]

[0.17115938 0.80031405 0.02852656]

[0.04750594 0.02850356 0.9239905 ]]





precision recall f1-score support

hate 0.24 0.74 0.36 294

offensive 0.97 0.80 0.88 3821

neither 0.86 0.92 0.89 842

accuracy 0.82 4957

macro avg 0.69 0.82 0.71 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.6231319936765414, accuracy: 0.8178333669558201

training 7 epoch...

/opt/conda/lib/python3.7/site-packages/ipykernel\_launcher.py:41: UserWarning: To copy construct from a tensor, it is recommended to use sourceTensor.clone().detach() or sourceTensor.clone().detach().requires\_grad\_(True), rather than torch.tensor(sourceTensor).

95.36015105247498 seconds used......

7 training finished! average train loss: 0.03408107045310827

total clf loss: 8.349633030826226 total adv loss: 11.582979358267039

evaluating...

Confusion matrix, without normalization

[[ 224 58 12]

[ 756 2974 91]

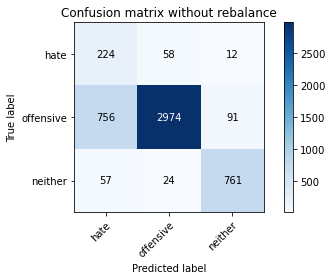
[ 57 24 761]]

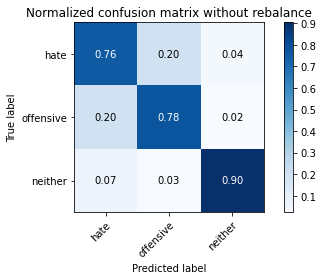
Normalized confusion matrix

[[0.76190476 0.19727891 0.04081633]

[0.19785396 0.77833028 0.02381576]

[0.06769596 0.02850356 0.90380048]]





precision recall f1-score support

hate 0.22 0.76 0.34 294

offensive 0.97 0.78 0.86 3821

neither 0.88 0.90 0.89 842

accuracy 0.80 4957

macro avg 0.69 0.81 0.70 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.7843639085883855, accuracy: 0.7986685495259229

training 8 epoch...

/opt/conda/lib/python3.7/site-packages/ipykernel\_launcher.py:41: UserWarning: To copy construct from a tensor, it is recommended to use sourceTensor.clone().detach() or sourceTensor.clone().detach().requires\_grad\_(True), rather than torch.tensor(sourceTensor).

95.42445874214172 seconds used......

8 training finished! average train loss: 0.03405487083517456

total clf loss: 8.24129648937378 total adv loss: 9.535718500846997

evaluating...

Confusion matrix, without normalization

[[ 233 50 11]

[ 844 2891 86]

[ 54 26 762]]

Normalized confusion matrix

[[0.79251701 0.17006803 0.03741497]

[0.22088459 0.75660822 0.0225072 ]

[0.06413302 0.03087886 0.90498812]]





precision recall f1-score support

hate 0.21 0.79 0.33 294

offensive 0.97 0.76 0.85 3821

neither 0.89 0.90 0.90 842

accuracy 0.78 4957

macro avg 0.69 0.82 0.69 4957

weighted avg 0.91 0.78 0.83 4957

average eval\_loss: 0.9526377615340904, accuracy: 0.7839419003429494

training 9 epoch...

/opt/conda/lib/python3.7/site-packages/ipykernel\_launcher.py:41: UserWarning: To copy construct from a tensor, it is recommended to use sourceTensor.clone().detach() or sourceTensor.clone().detach().requires\_grad\_(True), rather than torch.tensor(sourceTensor).

95.12106895446777 seconds used......

9 training finished! average train loss: 0.039425129363448115

total clf loss: 9.448526259278879 total adv loss: 9.191935447161086

evaluating...

Confusion matrix, without normalization

[[ 211 70 13]

[ 583 3139 99]

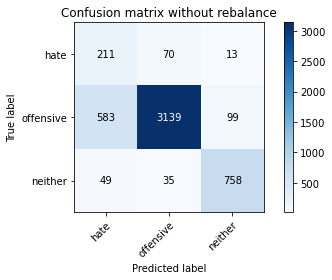
[ 49 35 758]]

Normalized confusion matrix

[[0.71768707 0.23809524 0.04421769]

[0.15257786 0.82151269 0.02590945]

[0.05819477 0.0415677 0.90023753]]





precision recall f1-score support

hate 0.25 0.72 0.37 294

offensive 0.97 0.82 0.89 3821

neither 0.87 0.90 0.89 842

accuracy 0.83 4957

macro avg 0.70 0.81 0.72 4957

weighted avg 0.91 0.83 0.86 4957

average eval\_loss: 0.7814169318886162, accuracy: 0.8287270526528142

training 10 epoch...

/opt/conda/lib/python3.7/site-packages/ipykernel\_launcher.py:41: UserWarning: To copy construct from a tensor, it is recommended to use sourceTensor.clone().detach() or sourceTensor.clone().detach().requires\_grad\_(True), rather than torch.tensor(sourceTensor).

95.15208339691162 seconds used......

10 training finished! average train loss: 0.03060694191612001

total clf loss: 7.378091857419349 total adv loss: 7.994181831483729

evaluating...

Confusion matrix, without normalization

[[ 223 55 16]

[ 772 2856 193]

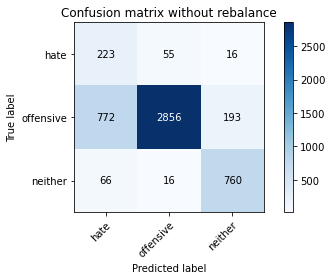
[ 66 16 760]]

Normalized confusion matrix

[[0.7585034 0.18707483 0.05442177]

[0.20204135 0.74744831 0.05051034]

[0.0783848 0.01900238 0.90261283]]





precision recall f1-score support

hate 0.21 0.76 0.33 294

offensive 0.98 0.75 0.85 3821

neither 0.78 0.90 0.84 842

accuracy 0.77 4957

macro avg 0.66 0.80 0.67 4957

weighted avg 0.90 0.77 0.81 4957

average eval\_loss: 1.0178887622214268, accuracy: 0.7744603590881581