

Experiment # 1 - Aaron - July 22

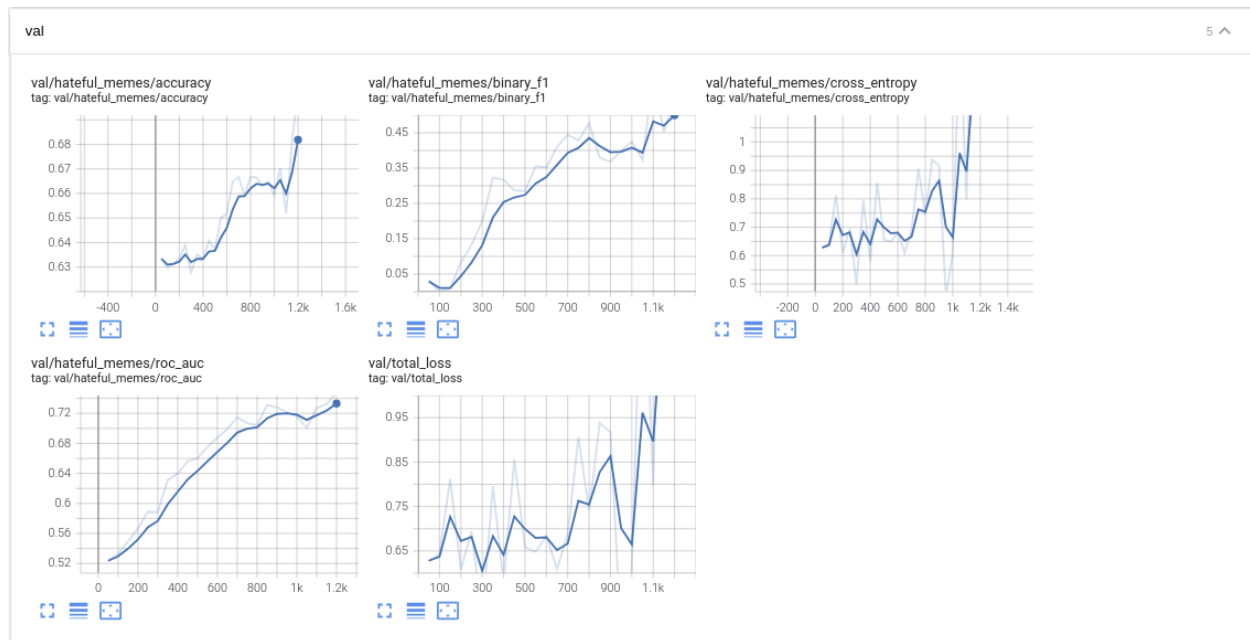
Description/Changes: **Baseline Model with Default Hyper-parameter Config**

Validation Accuracy

0.7019

Validation ROC-AUC

0.7475



Experiment # 2 - Hemanth

Description/Changes: Trained on Hateful Memes only;

Bs: 32, lr 0.3, w_steps 500, w_factor 0.1, w_iter 1000, w_type ['warmup_cosine', 'warmup_linear']

[INFO] The results of hyper-parameter tuning: directory ROC-AUC accuracy

0 1 0.7453 0.6889

1 2 0.7451 0.7167

Moving 1/best.ckpt to majority_voting_models/best-0.ckpt

Moving 2/best.ckpt to majority_voting_models/best-1.ckpt

Experiment # 3 - Hemanth

Description/Changes: Trained on Hateful Memes + Memotion;

Bs: 32, lr 0.3, w_steps 500, w_factor 0.1, w_iter 1000, w_type ['warmup_cosine', 'warmup_linear']

[INFO] The results of hyper-parameter tuning: directory ROC-AUC accuracy

0 1 0.7373 0.7037

1 2 0.7332 0.6889

Moving 1/best.ckpt to majority_voting_models/best-0.ckpt
Moving 2/best.ckpt to majority_voting_models/best-1.ckpt

Experiment # 4 - Hemanth

Description/Changes: Trained on Hateful Memes + Memotion + 1000 samples from MMHS150k

[INFO] The results of hyper-parameter tuning: directory ROC-AUC accuracy

2 2 0.7377 0.6963
1 1 0.7353 0.7000

Experiment #5 - Kate

Inclusion of misogynistic meme dataset + no changes to submission 1 in hyperparameters:

Val accuracy: 0.7352

Val binary F1: 0.5573

Val ROC-AUC: 0.7641

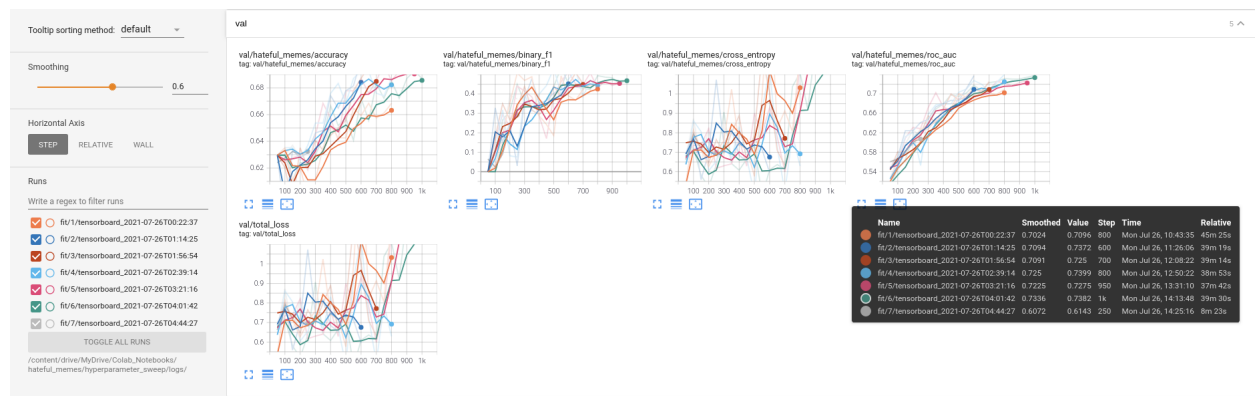
Experiment # 6 - Aaron

Description/Changes: Hyper-parameter Sweep of 7 variations of VisualBERT

Best Accuracy: 0.7019

Best Binary F1: 0.4679

Best ROC-AUC: 0.7382

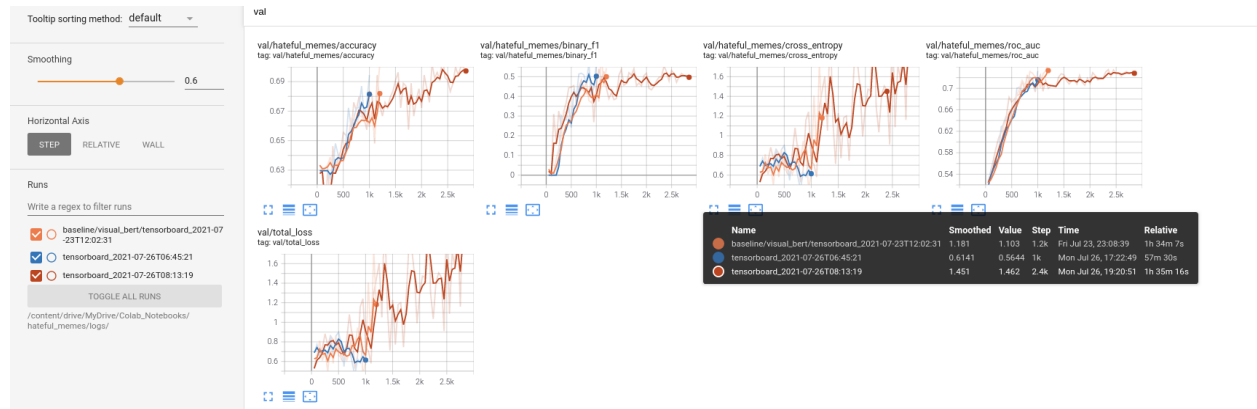


Experiment # 7 - Aaron

Description / Changes:

- FP16 Precision: True
- LR Ratio: 0.1
- Training Warmup: False
- Early Stopping: True
- Max Updates: 3000

Results: Baseline results are better than my config



Experiment # 8 - Karl

Submission 1 with MMBT-Region:

- val/hateful_memes/accuracy: 0.6815
- val/hateful_memes/binary_f1: 0.4591
- val/hateful_memes/roc_auc: 0.6915

Baseline - orange

MMBT Region - Green



Experiment #9 - Karl

Submission 1 with Concat-Bert:

- val/hateful_memes/accuracy: 0.6407
- val/hateful_memes/binary_f1: 0.3533
- val/hateful_memes/roc_auc: 0.6431

Baseline - orange

Mmbt - blue

Concatbert - red



Experiment #10 - Kate

Inclusion of misogynistic meme dataset + no changes to submission 1 in hyperparameters + vilbert as the model:

Val accuracy: 0.6796

Val binary F1: 0.4644

Val ROC-AUC: 0.6959

Experiment #11 - Hemanth

Train on original dataset + Misogynistic memes (My version)

[INFO] The results of hyper-parameter tuning: directory ROC-AUC accuracy

1 1 0.7383 0.7037

0 2 0.7244 0.6833

Moving 1/best.ckpt to majority_voting_models/best-0.ckpt

Moving 2/best.ckpt to majority_voting_models/best-1.ckpt

Train on original dataset + Misogynistic memes (train_miso.zip + train.jsonl)
[INFO] The results of hyper-parameter tuning: directory ROC-AUC accuracy
0 2 0.7474 0.7093
1 1 0.7129 0.6907
Moving 2/best.ckpt to majority_voting_models/best-0.ckpt
Moving 1/best.ckpt to majority_voting_models/best-1.ckpt

Train on original dataset + Misogynistic memes + Memotion + 1000 MMHS samples
[INFO] The results of hyper-parameter tuning: directory ROC-AUC accuracy
1 2 0.7318 0.7074
0 1 0.6932 0.6833
Moving 2/best.ckpt to majority_voting_models/best-0.ckpt
Moving 1/best.ckpt to majority_voting_models/best-1.ckpt

BCEWithLogitsLoss is same as cross entropy loss for two classes, hence not doing it

Margin Ranking Loss - does not make sense (train_miso.zip + train.jsonl)
[INFO] The results of hyper-parameter tuning: directory ROC-AUC accuracy
1 1 0.4842 0.3796
0 2 0.4674 0.3704
Moving 1/best.ckpt to majority_voting_models/best-0.ckpt
Moving 2/best.ckpt to majority_voting_models/best-1.ckpt

MultiLabelSoftMarginLoss
[INFO] The results of hyper-parameter tuning: directory ROC-AUC accuracy
1 2 0.7290 0.6833
0 1 0.7130 0.6907
Moving 2/best.ckpt to majority_voting_models/best-0.ckpt
mv: cannot stat '2/best.ckpt': No such file or directory
Moving 1/best.ckpt to majority_voting_models/best-1.ckpt
mv: cannot stat '1/best.ckpt': No such file or directory

Focal loss + Equal weights on Hateful Memes + Miso
[INFO] The results of hyper-parameter tuning: directory ROC-AUC accuracy
1 2 0.7475 0.6944
0 1 0.7058 0.6833
Moving 2/best.ckpt to majority_voting_models/best-0.ckpt
Moving 1/best.ckpt to majority_voting_models/best-1.ckpt

Focal loss + Weights of [4424, 2895] on Hateful Memes + Miso
[INFO] The results of hyper-parameter tuning: directory ROC-AUC accuracy
0 1 0.7096 0.6667
1 2 0.7070 0.6741

Focal loss + Weights of [4424, 2895] on Hateful Memes

[INFO] The results of hyper-parameter tuning: directory ROC-AUC accuracy

0 1 0.7450 0.6907

1 2 0.7412 0.6926

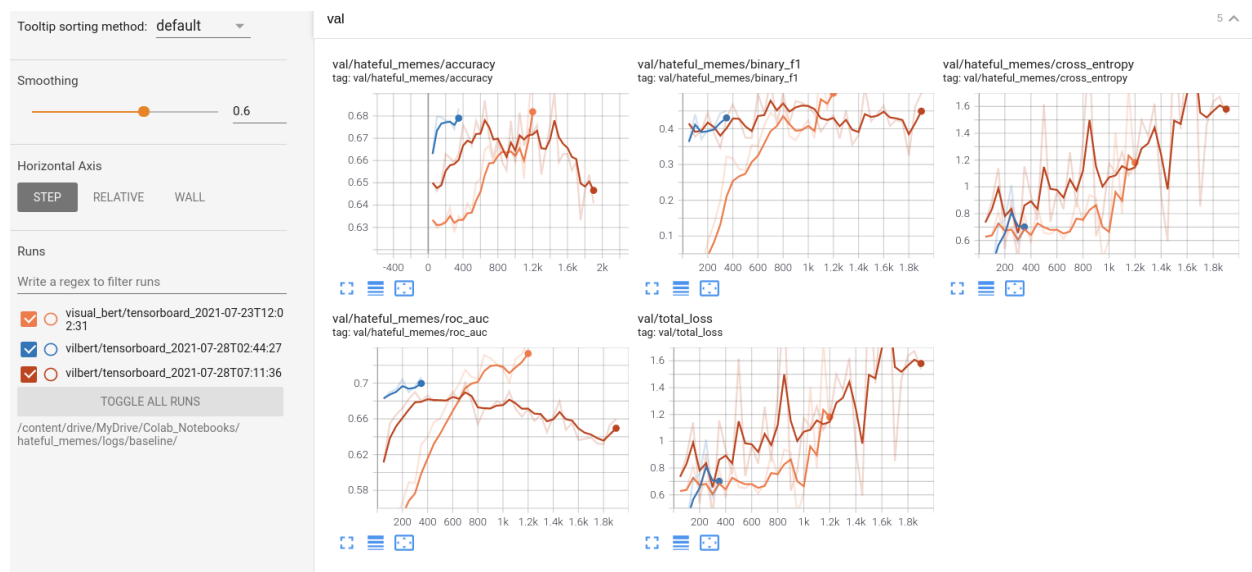
Experiment 12 - Aaron

Description/Changes: Baseline Submission # 1 with ViLBERT

Best Accuracy: 0.679

Best Binary F1: 0.5

Best ROC-AUC: 0.7067



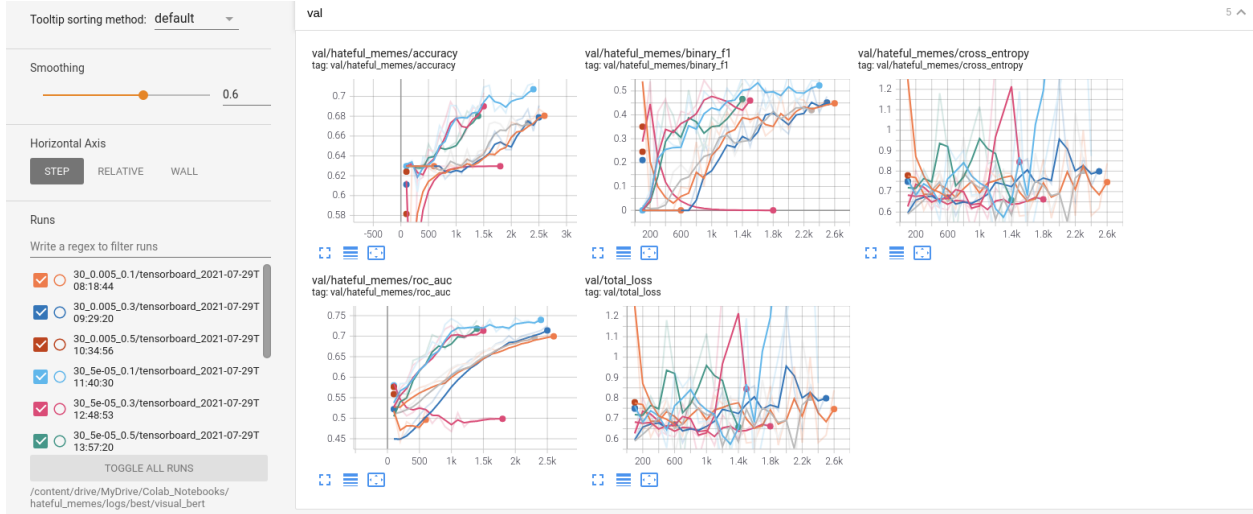
Experiment # 13 - Aaron

Description / Changes: Add in Miso Dataset + VisualBERT + Hyper-Param Sweep

Best Accuracy: 0.713

Best Binary F1: 0.5401

Best ROC-AUC: 0.750101

☒☒☒

TOGGLE ALL RUNS

/content/drive/MyDrive/Colab_Notebooks/hateful_memes/logs/best/visual_bert

val

val/hateful_memes/accuracytag: val/hateful_memes/accuracy

val/hateful_memes/binary_f1tag: val/hateful_memes/binary_f1

val/hateful_memes/cross_entropytag: val/hateful_memes/cross_entropy

val/hateful_memes/roc_auctag: val/hateful_memes/roc_auc

val/total_loss tag: val/total_loss