

# Google Cloud & YouTube-8M Video Understanding Challenge

- Starter Code -

김성현

서울대학교 통계학과

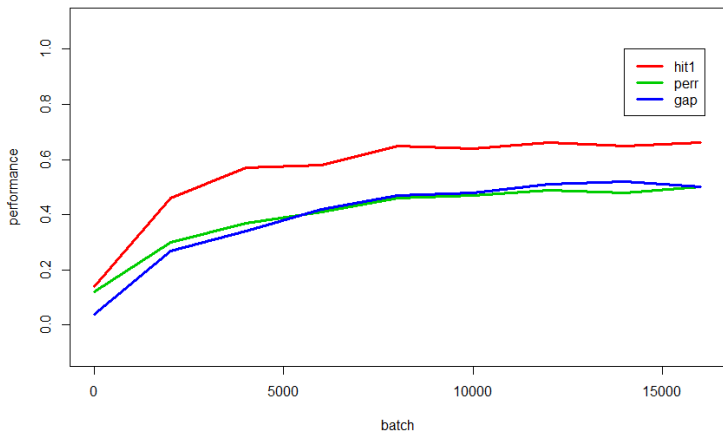
2017년 5월 4일

- audio feature의 길이는 128
- rgb feature와 합쳐 길이가 (1024 + 128)인 feature 사용.
- MOE의 expert 수를 2,4,8개로 하여 비교.

모형	GAP
(1st) WILLOW	0.84632
(제출) mean_rgb + mean_audio + MOE8	0.78806
(제출) mean_rgb + mean_audio + MOE4	0.78354
(제출) mean_rgb + mean_audio + MOE2	0.75812
(benchmark) mean_rgb + mean_audio + Logistic	0.74711
(제출) rgb + DBOF + Logistic	0.71214
(제출) mean_rgb + Moe	0.70893
(benchmark) mean_rgb + Logistic	0.69360
(제출) mean_rgb + Logistic	0.69170

- audio feature 사용. learning rate: 0.01, l2 regularization penalty: 1 (default).
- batch size: 1024, 8개의 gpu로 나누어 계산.
- 5000개 정도의 batch가 1 epoch. 1 epoch당 약 10시간, 100 달러.

Training Lstm + Moe



- Scene detection 후 Scene 별 feature의 평균을 가지고 Frame-level 모형에 적용.
- 코드 마무리 중.