

Dataset

['지상파', '종합편성', '음악', '쇼핑', '지상파', 'Zap', 'Off', 'Off', 'Off', 'Off',] → 총 2016개 로그



- 5일(720개) ~ 10일(1440개) 연속된 로그를 중간에 아무 곳에서나 랜덤하게 추출해서 모델 입력으로 활용
- 5일~10일도 랜덤하게.

[code]

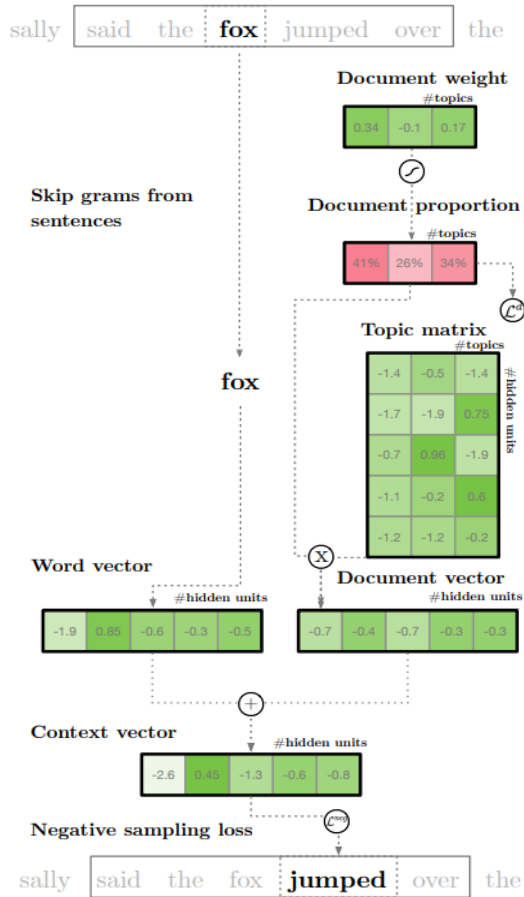
```
day_size = np.random.randint( 5, 11 )      # 5일 ~ 10일 랜덤 선택
day_length = day_size * 144                 # 5일 ~ 10일 * 하루 로그 144개
target_length = 18                         # 3시간 로그 개수

start_ind = np.random.randint( 0, 2016 - day_length - target_length )

subset_x = user_log[start_ind : start_ind + day_length]    # 랜덤하게 뽑힌 입력
```

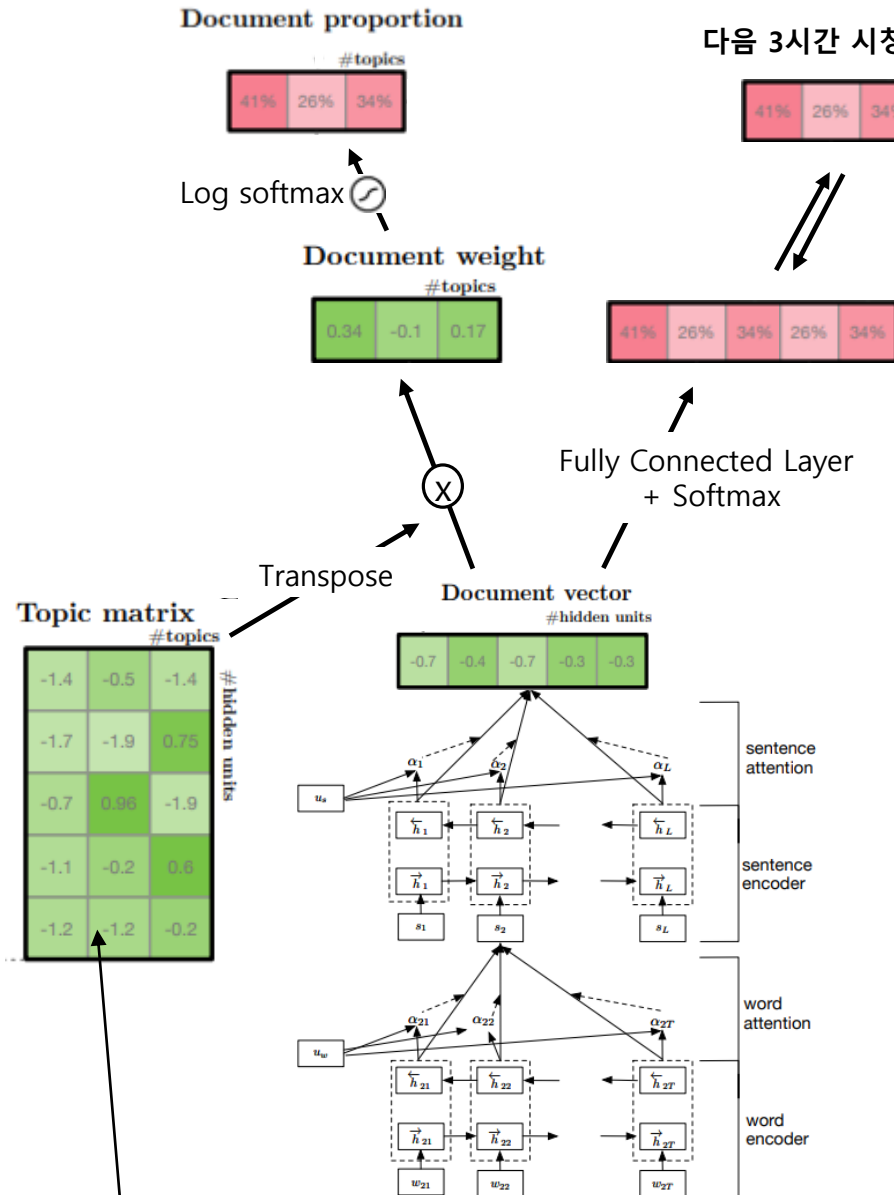
Model

(original)



Dirichlet log likelihood

MSE Loss



Dirichlet
log likelihood
+
MSE Loss

Backpropagation

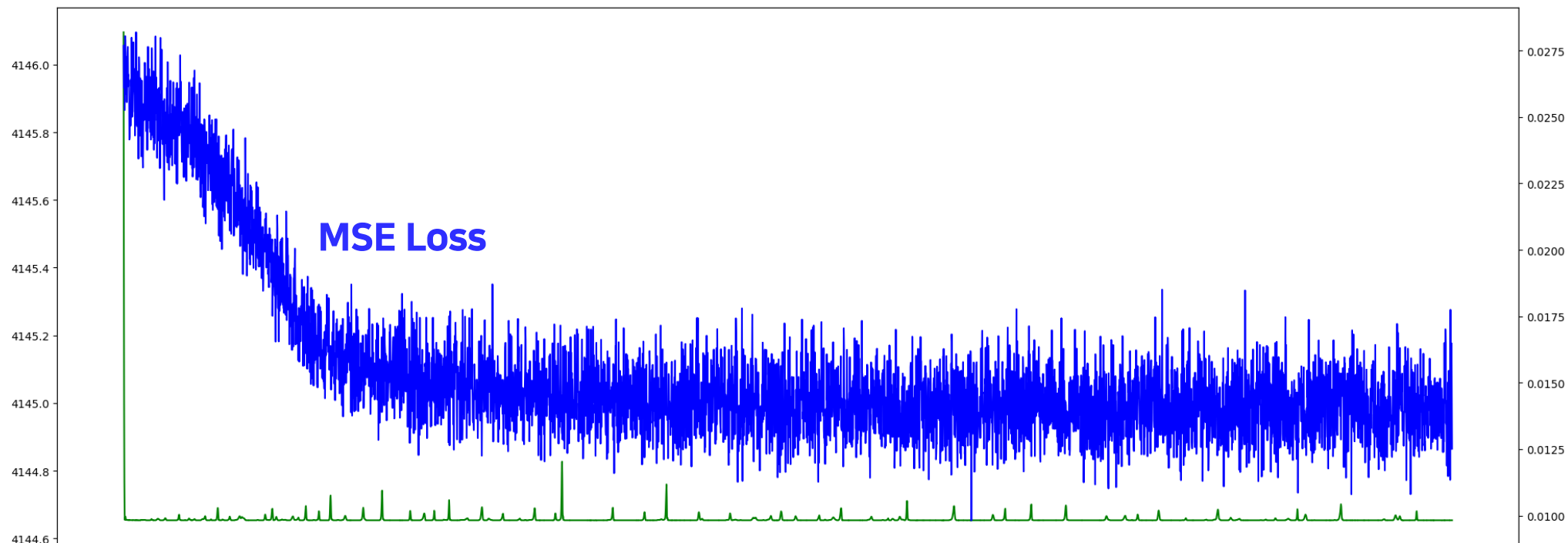


토픽 벡터, 채널 카테고리 벡터 모두 랜덤 초기화 후 학습.

Result 1.

$$\mathcal{L}^{\text{dirichlet}} = \lambda \sum_j (\alpha - 1) \log p_j$$

100 1/10

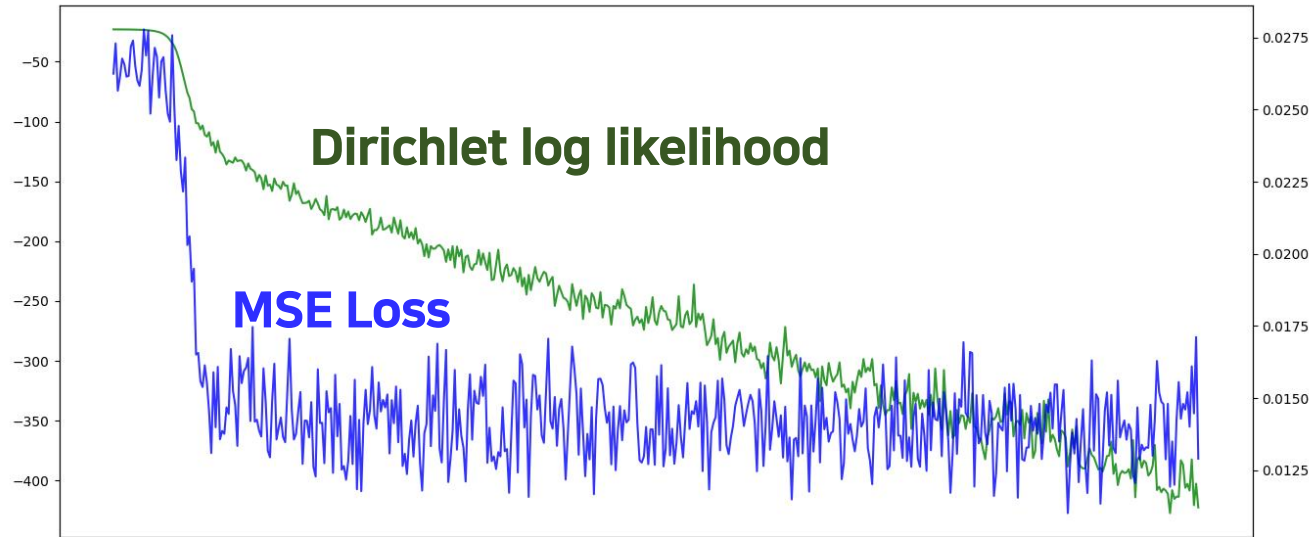


Dirichlet log likelihood (거의 변화 없음)

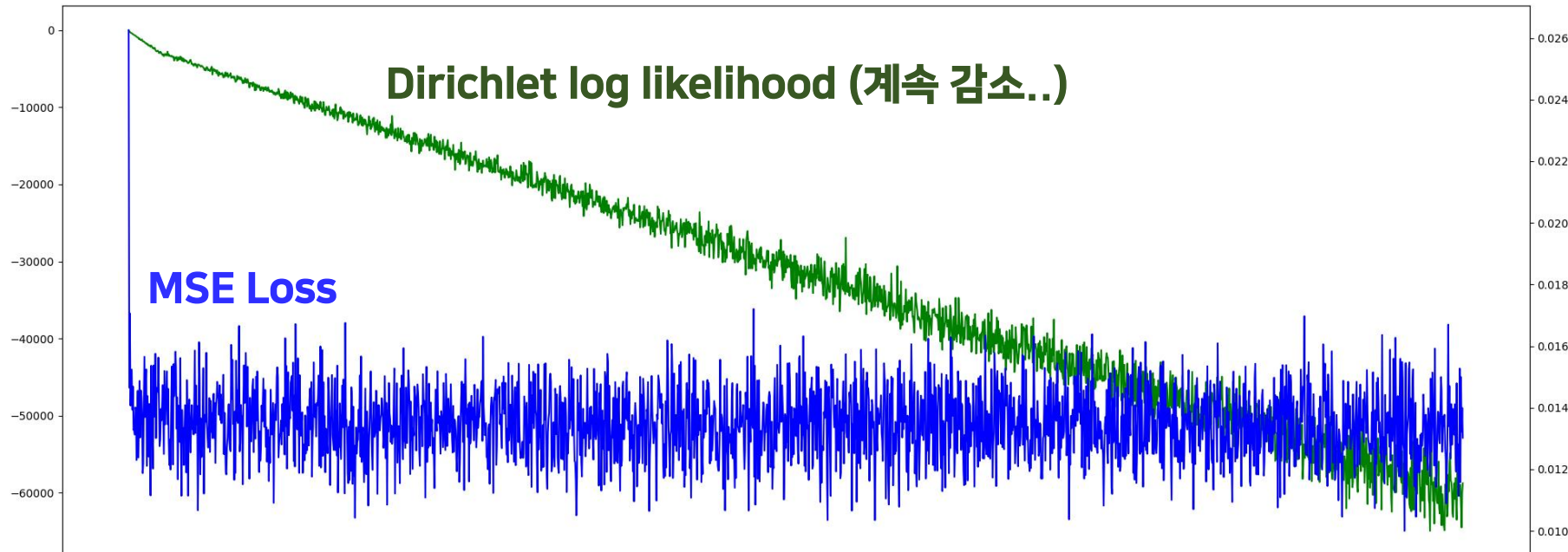
Result 2.

$$\mathcal{L}^{\text{dirichlet}} = \lambda \sum_j (\alpha - 1) \log p_j \longrightarrow \mathcal{L}^{\text{dirichlet}} = \sum_j \log p_j$$

lambda, alpha 제거하고 실험



초반 1000 iteration

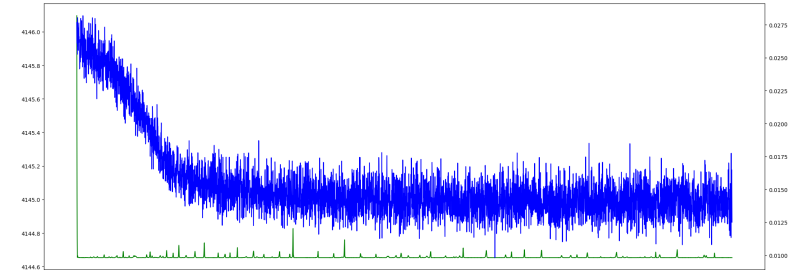


전체 200000 iteration

Result

$$\mathcal{L}^{\text{dirichlet}} = \lambda \sum_j (\alpha - 1) \log p_j$$

- 논문에서는 alpha를 (1/토픽 수) 로 설정했으며,
alpha가 계산에 포함되면 dirichlet log likelihood 가 거의 변화 없음.



- alpha를 0으로 해도 거의 변화 없음. $\mathcal{L}^{\text{dirichlet}} = \lambda \sum_j (-1) \log p_j$

$$\mathcal{L}^{\text{dirichlet}} = \sum_j \log p_j$$

- alpha 부분이 없어지면 loss가 계속 계속 계속.. 감소
(lambda는 있으나 없으나 큰 영향 없음)

