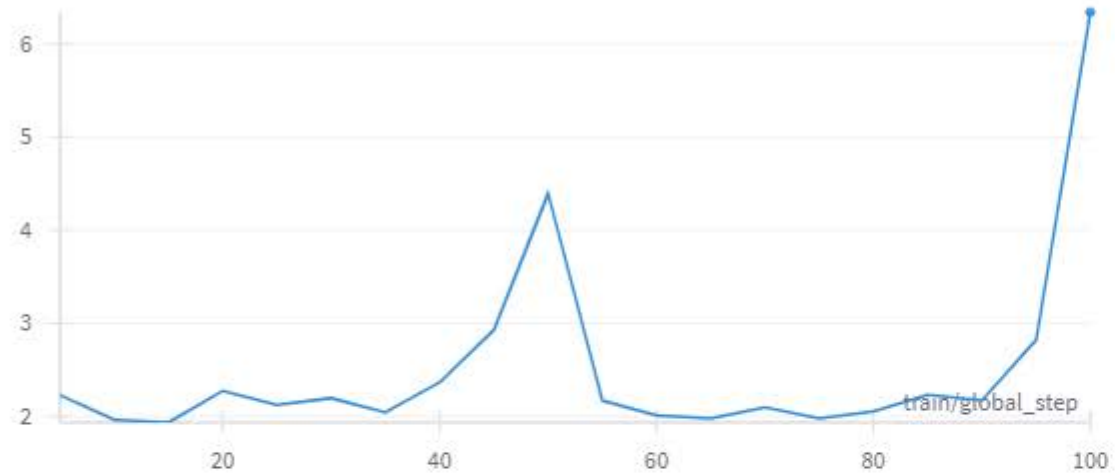
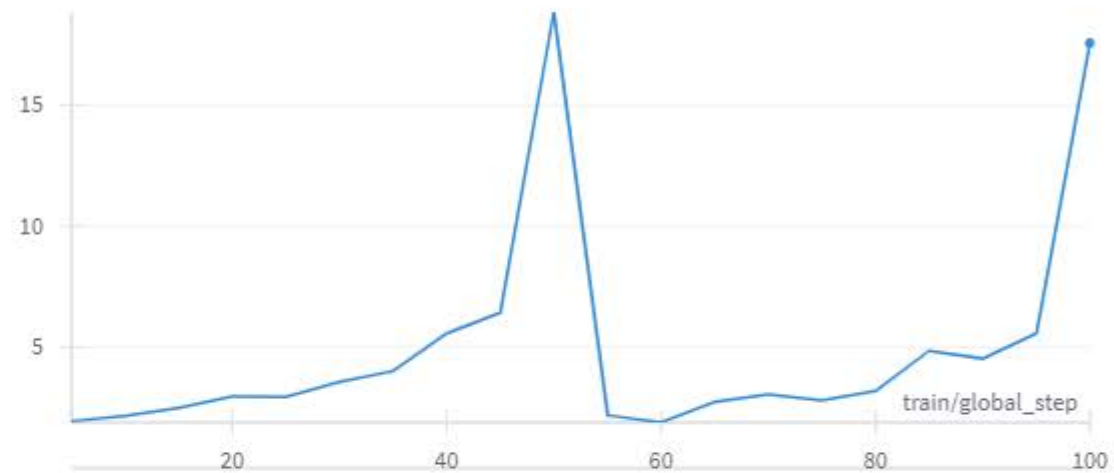


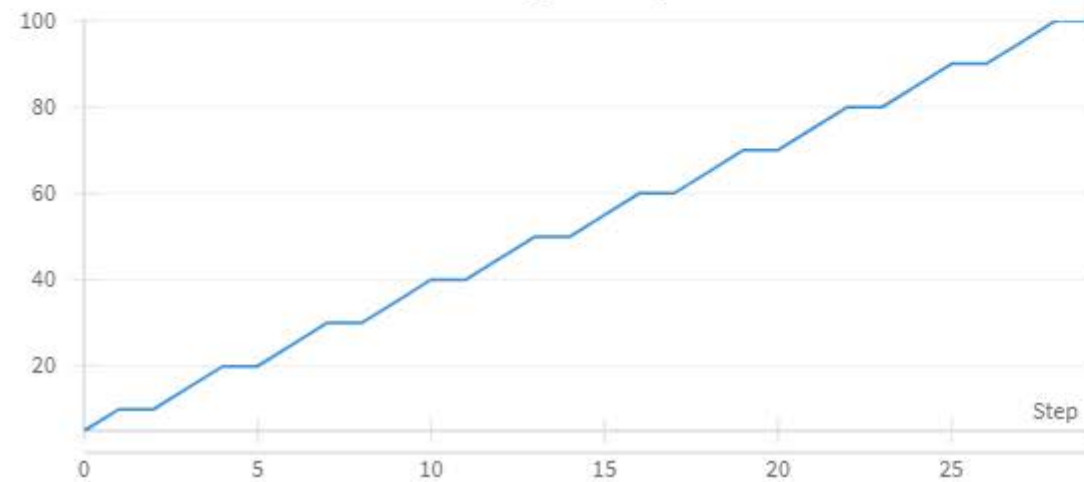
train/loss



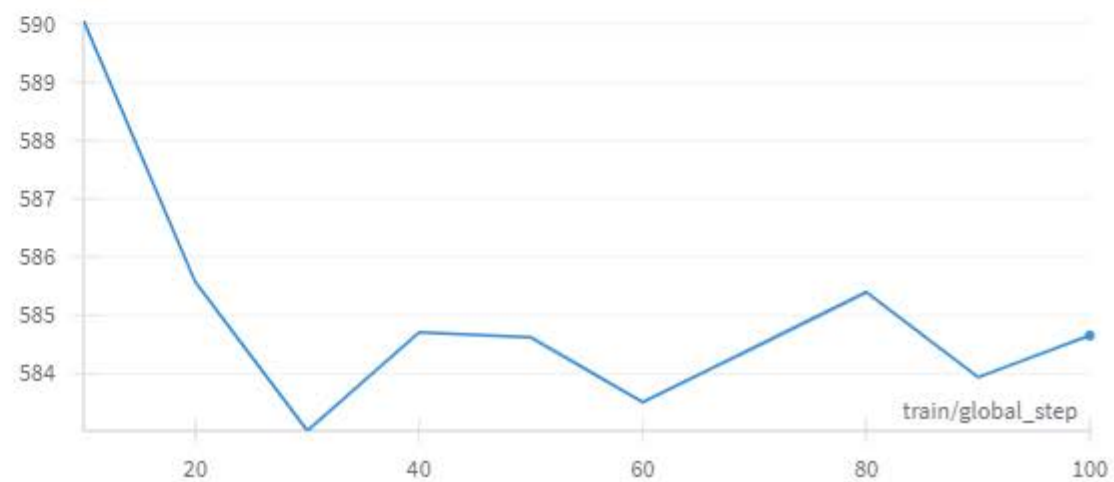
train/grad_norm

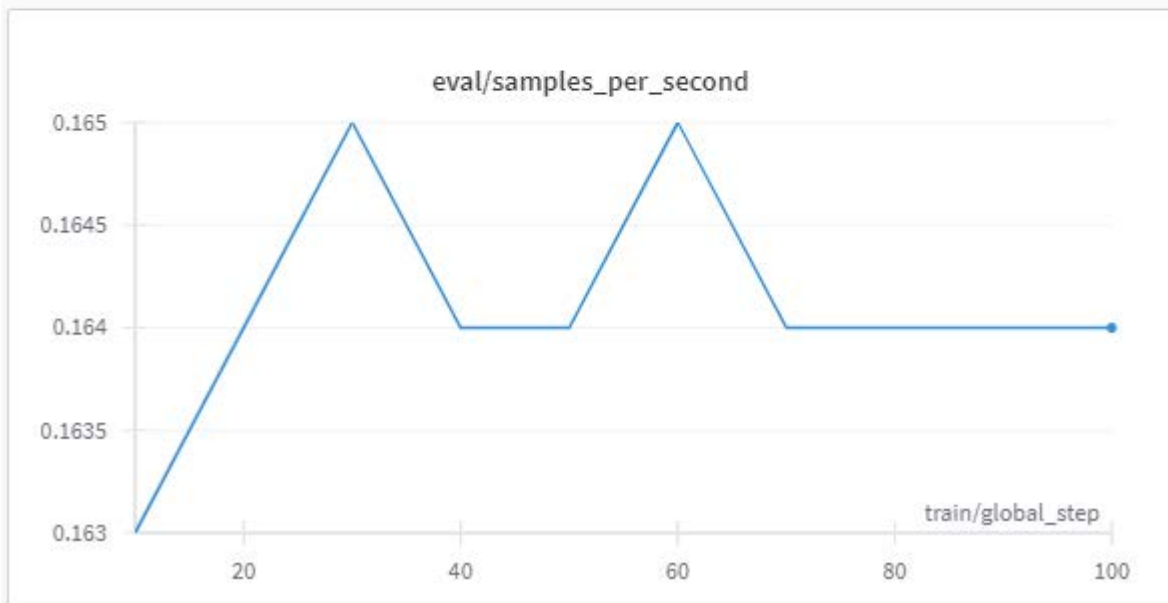
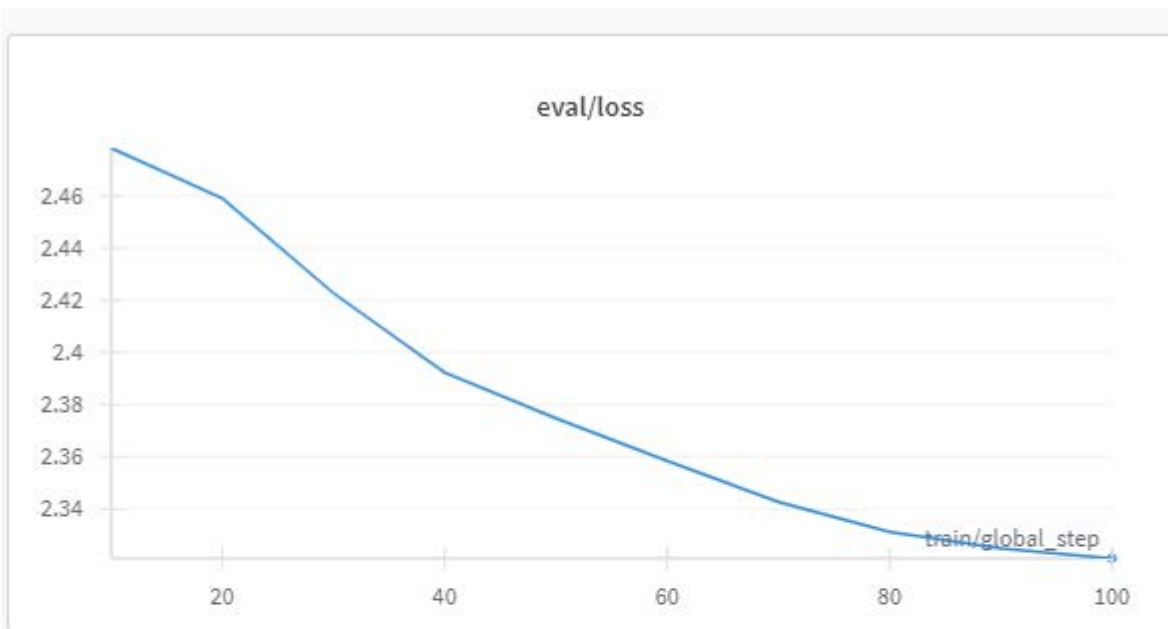


train/global_step



eval/runtime





- Loss на валидации сходится
- На обучающей выборке взрывается градиент, надо исправлять