

# HW4 Report

Zekai Zhang

November 2022

## 1 Configurations

For all the experiments, I adopt 64 as the batch size and  $5e-5$  as the initial learning rate.

For the AdamW optimizer, the configuration is  $\{lr : 5e-5, weight\_decay : 0.0, adam\_beta1 : 0.9, adam\_beta2 : 0.999, adam\_epsilon : 1e-08, max\_grad\_norm : 1.0\}$ .

Datasets "restaurant\_sup", "acl\_sup" and "agnews\_sup" employs N\_EPOCH = 20, 15 and 10 respectively.

## 2 Learning Curves

All the experiments converged stably under the above configuration. Same experiments are conducted on 5 different seeds to ensure the reliability.

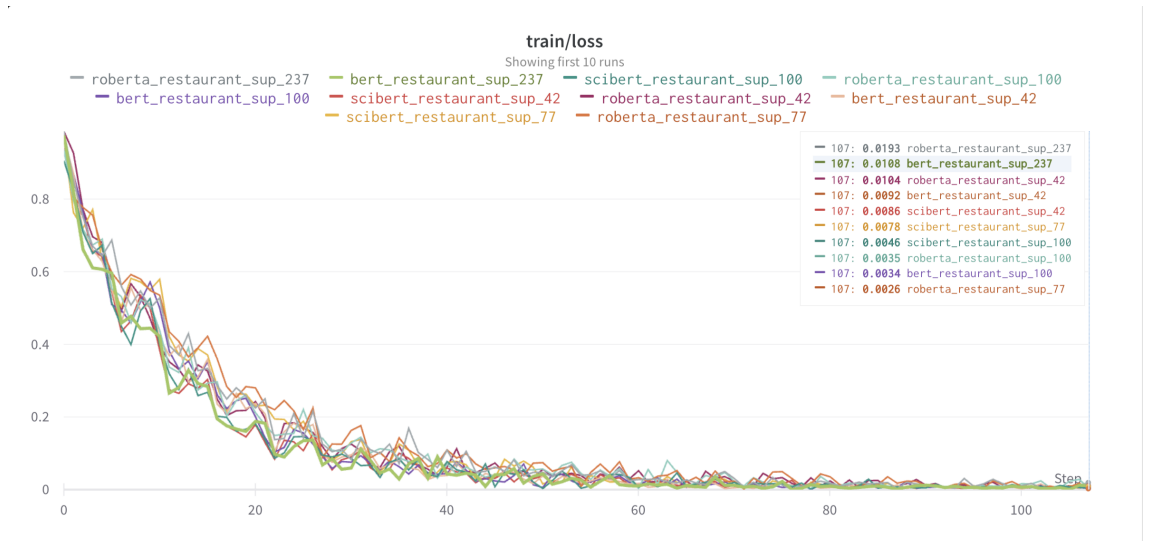


Figure 1: The result of different models on "restaurant\_sup".

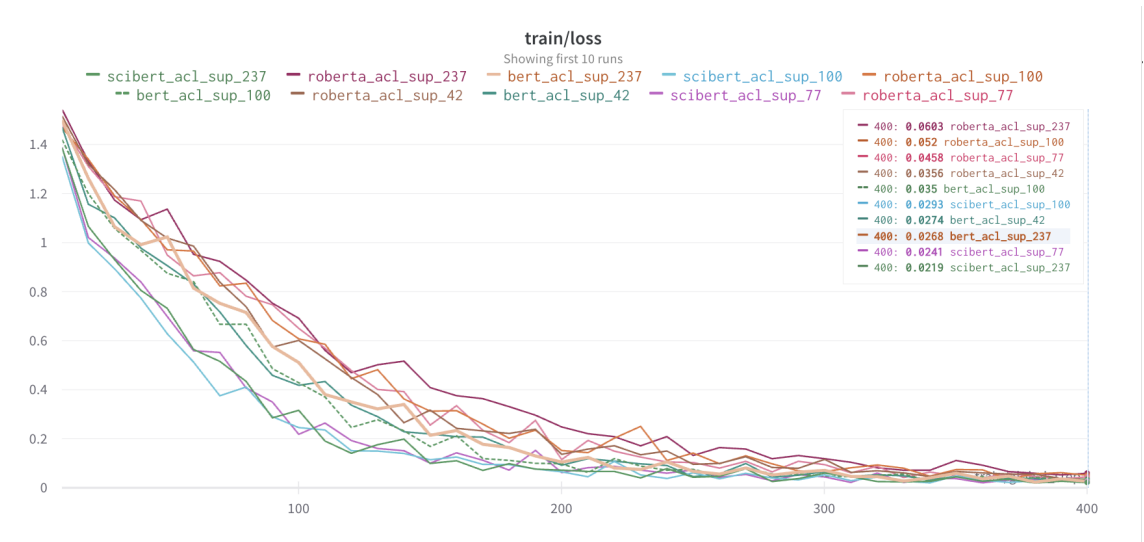


Figure 2: The result of different models on "acl\_sup".

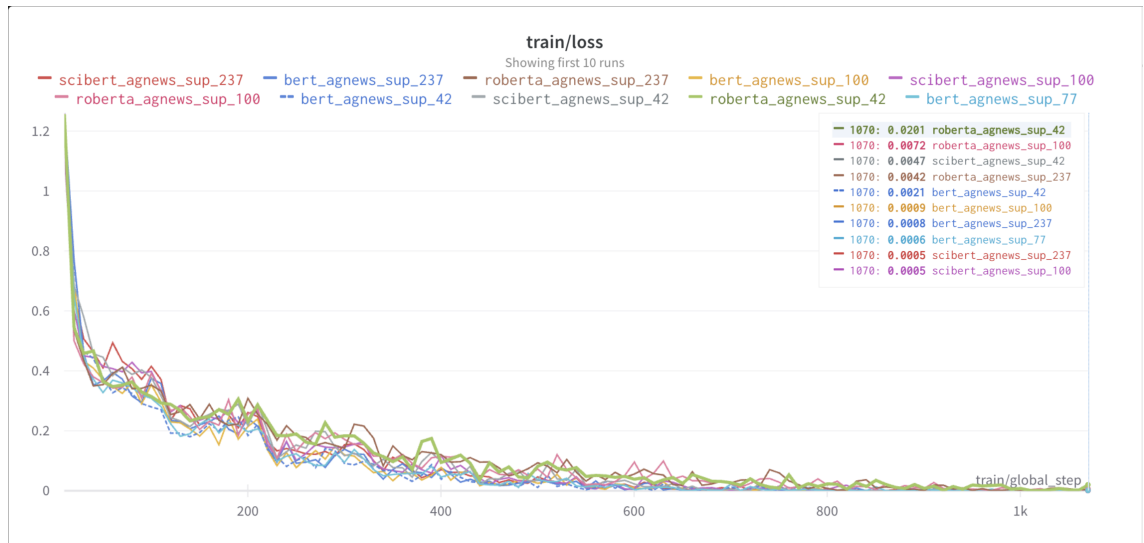


Figure 3: The result of different models on "agnews\_sup".

### 3 Results and Analysis

	restaurant_sup				acl_sup				ag_sup			
	acc	acc_std	macro	macro_std	acc	acc_std	macro	macro_std	acc	acc_std	macro	macro_std
bert	0.8414	0.0099	0.7512	0.0166	0.7656	0.0149	0.6642	0.022	0.9262	0.0026	0.9246	0.003
roberta	<b>0.8654</b>	0.0069	<b>0.7912</b>	0.0141	0.7842	0.0103	0.6804	0.0253	<b>0.9276</b>	0.004	<b>0.926</b>	0.0044
scibert	0.8302	0.0051	0.7378	0.0105	<b>0.8124</b>	0.0205	<b>0.7278</b>	0.039	0.9222	0.0056	0.9214	0.0054

Figure 4: Experiment results

Roberta achieves higher score than Bert on all the experiments. Scibert is the best on "acl\_sup" mainly because scibert is pretrained on science datasets, which has less domain gap with the acl dataset. But in general, Roberta preforms the best.

## 4 Appendix

	acc = micro-f1	macro-f1	acc_mean	macro_mean	acc_std	macro_std
(bert,res,0)	0.848	0.761	0.8414	0.7512	0.00968504	0.016619266
(bert,res,77)	0.851	0.765				
(bert,res,42)	0.826	0.723				
(bert,res,100)	0.84	0.751				
(bert,res,237)	0.842	0.756				
(bert,ag,0)	0.93	0.929	0.9262	0.9246	0.002683282	0.002966479
(bert,ag,77)	0.928	0.926				
(bert,ag,42)	0.925	0.924				
(bert,ag,100)	0.924	0.922				
(bert,ag,237)	0.924	0.922				
(bert,acl,0)	0.763	0.65	0.7656	0.6642	0.014926487	0.021970435
(bert,acl,77)	0.777	0.651				
(bert,acl,42)	0.777	0.653				
(bert,acl,100)	0.77	0.702				
(bert,acl,237)	0.741	0.665				
(roberta,res,0)	0.868	0.796	0.8654	0.7912	0.00694982	0.014060583
(roberta,res,77)	0.863	0.79				
(roberta,res,42)	0.856	0.771				
(roberta,res,100)	0.865	0.789				
(roberta,res,237)	0.875	0.81				
(roberta,ag,0)	0.926	0.925	0.9276	0.926	0.004037326	0.004358899
(roberta,ag,77)	0.925	0.923				
(roberta,ag,42)	0.934	0.933				
(roberta,ag,100)	0.924	0.922				
(roberta,ag,237)	0.929	0.927				
(roberta,acl,0)	0.77	0.642	0.7842	0.6804	0.010256705	0.025274493
(roberta,acl,77)	0.784	0.704				
(roberta,acl,42)	0.784	0.7				
(roberta,acl,100)	0.784	0.686				
(roberta,acl,237)	0.799	0.67				
(scibert,res,0)	0.838	0.754	0.8302	0.7378	0.005118594	0.010497619
(scibert,res,77)	0.828	0.732				
(scibert,res,42)	0.83	0.737				
(scibert,res,100)	0.824	0.726				
(scibert,res,237)	0.831	0.74				
(scibert,ag,0)	0.92	0.918	0.9222	0.9214	0.005585696	0.005412947
(scibert,ag,77)	0.918	0.919				
(scibert,ag,42)	0.917	0.916				
(scibert,ag,100)	0.926	0.925				
(scibert,ag,237)	0.93	0.929				
(scibert,acl,0)	0.842	0.769	0.8124	0.7278	0.020549939	0.039002564
(scibert,acl,77)	0.81	0.735				
(scibert,acl,42)	0.784	0.663				
(scibert,acl,100)	0.813	0.735				
(scibert,acl,237)	0.813	0.737				

Figure 5: Full experiment results