

### #Max. Epoch calculation

*This document is part of the Supplementary Materials of the paper titled "AutoTransform: Automated Code Transformation to Support Modern Code Review Process"*

The number of epochs indicates how many times that the whole training data was used to train the model. In this work, for a comparison, we set a similar max. number of epochs for each NMT architecture and for each dataset.

The number of epoch is based on the size of training data, batch size, and the number of train steps. However, the calculation of epochs for each NMT library (i.e., Tensor2Tensor and Seq2Seq) is different.

#Epoch for T2T =  $\text{\#train\_steps} * \text{batch\_size} / \text{\#subwords}$  (since in T2T, batch\_size is the number of subwords per batch)

#Epoch for Seq2Seq =  $\text{\#train\_steps} * \text{batch\_size} / \text{\#sequences}$  (since in Seq2Seq, batch\_size is the number of sequences per batch)

where **#train\_steps** and **batch\_size** are the hyper-parameters used in the NMT models.

Tables below show the the number of max epochs, the size of training data (i.e., #sequences or #subwords), and #train\_steps

For Seq2Seq, we used the same batch\_size and #train\_steps as the ones used in the experiments of Tufano et. al. Then, we calculated #max. epoch.

For T2T, we set a fixed number of batch\_size. Then, we identified #train\_steps that provided the similar #max epoch as Seq2Seq.

In particular, we used a calculation of  $\text{\#train\_steps} = \text{\#max. epoch of Seq2Seq} * \text{\#subwords} / \text{batch\_size of T2T}$ .

Note that #max. epoch is the maximum number of iterations of training. However, we selected the model checkpoint (i.e., the model is trained until a particular number of epochs) based on the loss value computed based on the validation data. Hence, the selected model checkpoint may not be the one that is trained until #max. epochs

### Changed Methods without new tokens

dataset	Size	Seq2Seq (RNN)				T2T (BPE m2000)				T2T (BPE m5000)			
		#sequences	batch_size	#train_step	#max. epoch	#subwords	batch_size	#train_step	#max. epoch	#subwords	batch_size	#train_step	#max. epoch
Android	Small	3,549	32	60,000	541	244,320	5,000	22,000	450	199,419	5,000	18,000	451
Google	Small	1,831	32	60,000	1,049	107,081	5,000	19,000	887	84,558	5,000	15,000	887
Ovirt	Small	3,787	32	60,000	507	223,838	5,000	19,000	424	178,993	5,000	15,000	419
Android	Medium	3,674	32	60,000	523	613,279	5,000	53,000	432	500,971	5,000	44,000	439
Google	Medium	2,266	32	60,000	847	281,497	5,000	40,000	710	228,227	5,000	32,000	701
Ovirt	Medium	4,983	32	60,000	385	629,253	5,000	41,000	326	514,418	5,000	33,000	321

### Changed Methods with new tokens

Dataset	Size	Seq2Seq (RNN)				T2T (BPE m2000)				T2T (BPE m5000)			
		#sequences	batch_size	#train_step	#max. epoch	#subwords	batch_size	#train_step	#max. epoch	#subwords	batch_size	#train_step	#max. epoch
Android	Small	16,512	32	60,000	116	1,266,233	5,000	25,000	99	1,036,074	5,000	20,000	97
Google	Small	7,256	32	60,000	265	434,059	5,000	19,000	219	346,800	5,000	15,000	216
Ovirt	Small	18,625	32	60,000	103	1,222,582	5,000	21,000	86	974,222	5,000	17,000	87
Android	Medium	19,637	32	60,000	98	3,442,192	5,000	56,000	81	2,831,854	5,000	47,000	83
Google	Medium	9,299	32	60,000	206	1,262,018	5,000	44,000	174	1,021,330	5,000	35,000	171
Ovirt	Medium	26,620	32	60,000	72	3,551,073	5,000	43,000	61	2,895,753	5,000	35,000	60