

ELBERT results report

SST2 - base model

Constant weighted sum in loss calculation

Model	Number of Layers	Accuracy	Loss	Runtime	Exit Threshold
albert-base-v2	12	0.9186	0.2478	1.7568	0.0
albert-base-v2	12	0.9163	0.2550	1.5767	0.2
albert-base-v2	12	0.8842	0.3099	0.6821	0.4
albert-base-v2	12	0.8188	0.3900	0.511	0.6
albert-base-v2	18	0.9220	0.2201	2.5235	0.0
albert-base-v2	18	0.9220	0.2315	1.4723	0.2
albert-base-v2	18	0.8773	0.3236	0.6574	0.4
albert-base-v2	18	0.8177	0.3853	0.5162	0.6
albert-base-v2	24	0.9186	0.2263	3.278	0.0
albert-base-v2	24	0.9197	0.2459	1.2927	0.2
albert-base-v2	24	0.8635	0.3326	0.6374	0.4
albert-base-v2	24	0.8096	0.4020	0.5002	0.6

Learned weighted sum in loss calculation

Model	Number of Layers	Accuracy	Loss	Runtime	Exit Threshold
albert-base-v2	12	0.9186	0.2217	1.7665	0.0
albert-base-v2	12	0.9174	0.2377	1.3381	0.2
albert-base-v2	12	0.8853	0.3038	0.6227	0.4
albert-base-v2	12	0.8268	0.3730	0.4876	0.6
albert-base-v2	18	0.9255	0.2211	2.516	0.0
albert-base-v2	18	0.9197	0.2373	1.4303	0.2
albert-base-v2	18	0.8704	0.3296	0.6541	0.4
albert-base-v2	18	0.8073	0.3985	0.5089	0.6
albert-base-v2	24	0.9197	0.2312	3.2698	0.0
albert-base-v2	24	0.9197	0.2378	2.0195	0.2
albert-base-v2	24	0.8658	0.3059	0.6902	0.4
albert-base-v2	24	0.8131	0.3963	0.5124	0.6

SST2 - large model

Constant weighted sum in loss calculation

Model	Number of Layers	Accuracy	Loss	Runtime	Exit Threshold
albert-large-v2	24	0.9427	0.1762	5.0805	0.0
albert-large-v2	24	0.9220	0.2025	1.7315	0.2
albert-large-v2	24	0.8911	0.2875	0.9290	0.4
albert-large-v2	24	0.8360	0.3775	0.6493	0.6
albert-large-v2	36	0.9369	0.1801	7.4708	0.0
albert-large-v2	36	0.9232	0.2237	1.4186	0.2
albert-large-v2	36	0.8773	0.3177	0.8069	0.4
albert-large-v2	36	0.8165	0.4105	0.5361	0.6

Learned weighted sum in loss calculation

Model Name	Number of Layers	Accuracy	Loss	Runtime	Exit Threshold
albert-large-v2	24	0.9461	0.1701	5.0665	0.0
albert-large-v2	24	0.9232	0.2010	1.7594	0.2
albert-large-v2	24	0.8899	0.2859	0.9183	0.4
albert-large-v2	24	0.8360	0.3749	0.6382	0.6
albert-large-v2	36	0.9392	0.1985	7.4808	0.0
albert-large-v2	36	0.9209	0.2230	1.4267	0.2
albert-large-v2	36	0.8830	0.2968	0.8202	0.4
albert-large-v2	36	0.8154	0.4054	0.5277	0.6

SNLI - base model

Constant weighted sum in loss calculation

Model Name	Number of Layers	Accuracy	Loss	Runtime	Exit Threshold
albert-base-v2	12	0.890571	0.298506	19.9202	0.0
albert-base-v2	12	0.890673	0.298600	19.7656	0.2
albert-base-v2	12	0.844137	0.403608	7.515	0.4
albert-base-v2	12	0.800447	0.506763	5.4775	0.6
albert-base-v2	18	0.893010	0.295163	28.5329	0.0
albert-base-v2	18	0.893010	0.295446	27.6796	0.2
albert-base-v2	18	0.845458	0.403724	8.1902	0.4
albert-base-v2	18	0.787238	0.529667	5.5738	0.6
albert-base-v2	24	0.893721	0.293153	36.9659	0.0
albert-base-v2	24	0.893721	0.293700	35.2969	0.2
albert-base-v2	24	0.844950	0.406209	8.6148	0.4
albert-base-v2	24	0.779110	0.549381	5.5665	0.6

Learned weighted sum in loss calculation

Model	Number of Layers	Accuracy	Loss	Runtime	Exit Threshold
albert-base-v2	12	0.8918	0.2989	19.9055	0.0
albert-base-v2	12	0.8919	0.2991	19.657	0.2
albert-base-v2	12	0.8435	0.4037	7.4951	0.4
albert-base-v2	12	0.7991	0.5073	5.4079	0.6
albert-base-v2	18	0.8924	0.2958	28.354	0.0
albert-base-v2	18	0.8925	0.2962	27.1406	0.2
albert-base-v2	18	0.8443	0.4039	8.0264	0.4
albert-base-v2	18	0.7866	0.5322	5.448	0.6
albert-base-v2	24	0.8899	0.2985	36.8122	0.0
albert-base-v2	24	0.8898	0.2987	36.3335	0.2
albert-base-v2	24	0.8499	0.3941	9.4955	0.4
albert-base-v2	24	0.7715	0.5658	5.5248	0.6

SNLI - large model

Constant weighted sum in loss calculation

Model Name	Number of Layers	Accuracy	Loss	Runtime	Exit Threshold
albert-large-v2	24	0.8993090835	0.2746927	57.1973	0.0
albert-large-v2	24	0.8988010567	0.2775372	49.7445	0.2
albert-large-v2	24	0.8445437919	0.402463	12.044	0.4
albert-large-v2	24	0.7654948181	0.5703849	7.2234	0.6
albert-large-v2	36	0.900833164	0.2739922	85.1877	0.0
albert-large-v2	36	0.8971753709	0.2845134	57.6494	0.2
albert-large-v2	36	0.8416988417	0.4164651	11.5868	0.4
albert-large-v2	36	0.7654948181	0.5694668	7.2237	0.6

Learned weighted sum in loss calculation

Model	Number of Layers	Accuracy	Loss	Runtime	Exit Threshold
albert-large-v2	24	0.9003	0.2759	57.4217	0.0
albert-large-v2	24	0.8995	0.2782	50.3276	0.2
albert-large-v2	24	0.8445	0.4032	12.0512	0.4
albert-large-v2	24	0.7645	0.5708	7.3131	0.6
albert-large-v2	36	0.9008	0.2747	85.1044	0.0
albert-large-v2	36	0.8990	0.2791	68.2956	0.2
albert-large-v2	36	0.8434	0.4019	13.1459	0.4
albert-large-v2	36	0.7611	0.5744	7.7591	0.6