Results

Stefan Reifberger

September 26, 2024

The following tables show the results of the experiment. They are to be read as follows:

- D: Number of decoder layers.
- E: Number of encoder layers.
- M: Mean value range of epochs.
- MD: 95% confidence interval distance between means.
- Bold text: The best value for given epochs or the current model performs better for the range of epochs up to the 95% confidence intervals of the difference of means not overlapping.

Model	Ep.: 1	5	M 1–5	MD 1–5	M 40–50	MD 40–50
LSTM, D=1	.829	**.513**	.661	.000	**.274**	.000
LSTM, D=2	**.808**	.675	.750	$\mathbf{.089}^{\pm .042}$	.385	$\mathbf{.111}^{\pm .017}$
LSTM, D=3	.812	.519	**.653**	$-.008^{\pm .008}$	.340	$\mathbf{.066}^{\pm .006}$
LSTM, D=4	.899	.682	.760	$\mathbf{.099}^{\pm .027}$	.382	$\mathbf{.108}^{\pm .008}$
TrFo, D=1	.635	.251	.365	.000	.145	.000
TrFo, D=2	.465	.220	.303	$-\mathbf{.061}^{\pm .034}$	.138	$-\mathbf{.008}^{\pm .002}$
TrFo, D=3	.441	.187	.273	$-\mathbf{.092}^{\pm .032}$	.134	$-\mathbf{.011}^{\pm .001}$
TrFo, D=4	**.429**	**.175**	**.254**	$-\mathbf{.110}^{\pm .030}$	**.129**	$-\mathbf{.017}^{\pm .001}$

(a) Test loss results.

Model	Ep.: 1	5	M 1–5	MD 1–5	M 40–50	MD 40–50
LSTM, D=1	.731	.793	.783	.000	**.894**	.000
LSTM, D=2	**.745**	.794	.760	$-\mathbf{.023}^{\pm .018}$	.856	$-\mathbf{.038}^{\pm .010}$
LSTM, D=3	.729	**.826**	**.794**	$\mathbf{.011}^{\pm .010}$	.870	$-\mathbf{.025}^{\pm .015}$
LSTM, D=4	.724	.763	.753	$-\mathbf{.030}^{\pm .017}$	.840	$-\mathbf{.054}^{\pm .017}$
TrFo, D=1	.733	.909	.860	.000	.942	.000
TrFo, D=2	.833	**.941**	.896	$\mathbf{.036}^{\pm .022}$	.944	$.002^{\pm .012}$
TrFo, D=3	.839	**.941**	.908	$\mathbf{.048}^{\pm .023}$	**.946**	$.004^{\pm .011}$
TrFo, D=4	**.844**	.929	**.911**	$\mathbf{.051}^{\pm .022}$	.939	$-.003^{\pm .013}$

(b) Test accuracy results.

Figure 1: Test data results for varying number of decoder layers with one encoder layer.

Model	Ep.: 1	5	M 1–5	MD 1–5	M 40–50	MD 40–50
LSTM, E=1	.812	.519	**.653**	.000	.340	.000
LSTM, E=2	**.803**	**.502**	.658	$.005^{\pm .010}$	**.320**	$-$**.020**$^{\pm .003}$
LSTM, E=3	.904	.879	.890	**.237**$^{\pm .064}$	.418	**.078**$^{\pm .024}$
TrFo, E=1	**.441**	.187	**.273**	.000	.134	.000
TrFo, E=2	.522	.189	.290	$.017^{\pm .020}$	.126	$-$**.008**$^{\pm .001}$
TrFo, E=3	.606	**.181**	.302	$.029^{\pm .042}$	**.122**	$-$**.012**$^{\pm .001}$

(a) Test loss results.

Model	Ep.: 1	5	M 1–5	MD 1–5	M 40–50	MD 40–50
LSTM, E=1	.729	**.826**	**.794**	.000	.870	.000
LSTM, E=2	**.748**	.817	.774	$-$**.020**$^{\pm .014}$	**.880**	$.010^{\pm .010}$
LSTM, E=3	.694	.733	.728	$-$**.066**$^{\pm .014}$	.836	$-$**.034**$^{\pm .012}$
TrFo, E=1	.839	**.941**	**.908**	.000	.946	.000
TrFo, E=2	**.841**	.919	.896	$-$**.012**$^{\pm .005}$	.952	$.005^{\pm .008}$
TrFo, E=3	.774	.927	.897	$-.011^{\pm .019}$	**.954**	$.008^{\pm .016}$

(b) Test accuracy results.

Figure 2: LSTM and Transformer results for different numbers of encoder layers, with decoder layers fixed to one.

Model	Ep.: 1	5	M 1–5	MD 1–5	M 40–50	MD 40–50
TrFo 1H, E=1	.449	.230	.313	.000	.144	.000
TrFo 2H, E=2	**.437**	.216	.292	$-$**.020**$^{\pm .005}$	.138	$-$**.006**$^{\pm .001}$
TrFo 4H, E=3	.441	.187	**.273**	$-$**.040**$^{\pm .011}$	.134	$-$**.010**$^{\pm .001}$
TrFo 8H, E=3	.507	**.181**	.277	$-$, **035**$^{\pm .029}$	**.127**	$-$**.017**$^{\pm .001}$

(a) Test loss results.

Model	Ep.: 1	5	M 1–5	MD 1–5	M 40–50	MD 40–50
TrFo 1H	.850	.928	.894	.000	.941	.000
TrFo 2H	**.875**	.931	.896	$.002^{\pm .012}$	.948	$.007^{\pm .008}$
TrFo 4H	.839	.941	**.908**	$.015^{\pm .017}$	.946	$.005^{\pm .010}$
TrFo 8H	.803	**.953**	.900	$.002^{\pm .020}$	**.950**	$.008^{\pm .010}$

(b) Test accuracy results.

Figure 3: Transformer results for different numbers of attention heads.

Model	Ep.: 1	5	M 1–5	MD 1–5	M 40–50	MD 40–50
TrFo 1H, E=1	.449	.230	.313	.000	.144	.000
TrFo 2H, E=2	**.437**	.216	.292	$-$**.020**$^{\pm .005}$	.138	$-$**.006**$^{\pm .001}$
TrFo 4H, E=3	.441	.187	**.273**	$-$**.040**$^{\pm .011}$	.134	$-$**.010**$^{\pm .001}$
TrFo 8H, E=3	.507	**.181**	.277	$-$, **035**$^{\pm .029}$	**.127**	$-$**.017**$^{\pm .001}$

(a) Test loss results.

Model	Ep.: 1	5	M 1–5	MD 1–5	M 40–50	MD 40–50
TrFo 1H	.850	.928	.894	.000	.941	.000
TrFo 2H	**.875**	.931	.896	.002$^{\pm .012}$	.948	.007$^{\pm .008}$
TrFo 4H	.839	.941	**.908**	.015$^{\pm .017}$	.946	.005$^{\pm .010}$
TrFo 8H	.803	**.953**	.900	.002$^{\pm .020}$	**.950**	.008$^{\pm .010}$

(b) Test accuracy results.

Figure 4: Transformer results for different numbers of attention heads.

Model	Ep.: 1	5	M 1–5	MD 1–5	M 40–50	MD 40–50
LSTM/TrFo, D=1	.829/**.635**	.513/**.251**	.661/**.365**	**.296**$^{\pm .041}$	.274/**.145**	**.129**$^{\pm .006}$
LSTM/TrFo, D=2	.808/**.465**	.675/**.220**	.750/**.303**	**.447**$^{\pm .033}$	.385/**.138**	**.247**$^{\pm .017}$
LSTM/TrFo, D=3	.812/**.441**	.519/**.187**	.653/**.273**	**.380**$^{\pm .025}$	.340/**.134**	**.206**$^{\pm .002}$
LSTM/TrFo, D=4	.899/**.429**	.682/**.175**	.760/**.254**	**.506**$^{\pm .012}$	.382/**.129**	**.254**$^{\pm .005}$

(a) Test loss results.

Model	Ep.: 1	5	M 1–5	MD 1–5	M 40–50	MD 40–50
LSTM/TrFo, D=1	**.731**/**.733**	.793/**.909**	.783/**.860**	$-$**.077**$^{\pm .028}$	.894/**.942**	$-$**.048**$^{\pm .011}$
LSTM/TrFo, D=2	**.745**/**.833**	.794/**.941**	.760/**.896**	$-$**.137**$^{\pm .018}$	.856/**.944**	$-$**.088**$^{\pm .010}$
LSTM/TrFo, D=3	**.729**/**.839**	.826/**.941**	.794/**.908**	$-$**.114**$^{\pm .008}$	.870/**.946**	$-$**.077**$^{\pm .011}$
LSTM/TrFo, D=4	**.724**/**.844**	.763/**.929**	.753/**.911**	$-$**.159**$^{\pm .016}$	.840/**.939**	$-$**.099**$^{\pm .015}$

(b) Test accuracy results.

Figure 5: LSTM and Transformer results compared for different numbers of decoder layers, with encoder layers fixed to one.

Model	Ep.: 1	5	M 1–5	MD 1–5	M 40–50	MD 40–50
LSTM/TrFo, E=1	.812/**.441**	.519/**.187**	.653/**.273**	**.380**$^{\pm .025}$	.340/**.134**	**.206**$^{\pm .002}$
LSTM/TrFo, E=2	.803/**.522**	.502/**.189**	.658/**.290**	**.368**$^{\pm .039}$	.320/**.126**	**.195**$^{\pm .004}$
LSTM/TrFo, E=3	.904/**.606**	.879/**.181**	.890/**.302**	**.588**$^{\pm .092}$	.418/**.122**	**.296**$^{\pm .024}$

(a) Test loss results.

Model	Ep.: 1	5	M 1–5	MD 1–5	M 40–50	MD 40–50
LSTM/TrFo, E=1	.729/**.839**	.826/**.941**	.794/**.908**	$-$**.114**$^{\pm .008}$	.870/**.946**	$-$**.077**$^{\pm .011}$
LSTM/TrFo, E=2	.748/**.841**	.817/**.919**	.774/**.896**	$-$**.122**$^{\pm .015}$	.880/**.952**	$-$**.072**$^{\pm .015}$
LSTM/TrFo, E=3	.694/**.774**	.733/**.927**	.728/**.897**	$-$**.169**$^{\pm .030}$	.836/**.954**	$-$**.118**$^{\pm .017}$

(b) Test accuracy results.

Figure 6: LSTM and Transformer results compared for different numbers of encoder layers, with decoder layers fixed to three.