Temporally Multi-Scale Sparse Self-Attention for Physical Activity Data Imputation

Hui Wei[1], Maxwell A. Xu[2], Colin Samplawski[1], James M. Rehg[3], Santosh Kumar[4], Benjamin M. Marlin[1]
[1] UMass Amherst, [2] Georgia Tech, [3] UIUC, [4] University of Memphis

We propose a novel sparse self-attention model which is informed by the domain-knowledge to solve the missing data problem in the hourly-level step count data.

Motivations

Model Design and Training Method Considerations

- Easy to train with parallelism and stability on the large-scale dataset.
- *Key Idea: self-attention-based model with ERM-based training.*

Practicability to Apply Self-Attention

- Vanilla self-attention with quadratic complexity is hard to apply on long time series data (our data is on average 50*k* hours long per participant).
- *Key Idea: sparse self-attention.*

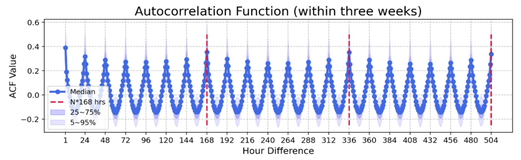
Problem to Improve Imputation Performance

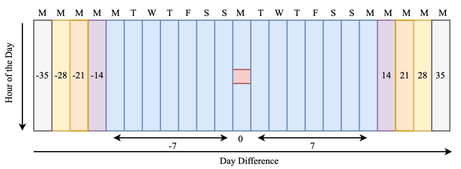
- Similarity computation between the missing (i.e. imputation target) and the observed hourly blocks is limited, since the target block only has features of hour of the day and day of the week.
- *Problem: how to augment the feature representation?*

Key Components

Multi-Timescale Sparse Self-Attention

- Receptive field is set as a context window around the target block.
- Context window structure is informed by the autocorrelation function.
- Captures the multi-timescale information (daily, weekly).





Local Activity Profile Representation (LAPR)

- A window of step data from *t-W* to *t+W,* where *t* is each hourly block.
- Model the local activity pattern of each hourly block.

Relative Time Encoding (RTE)

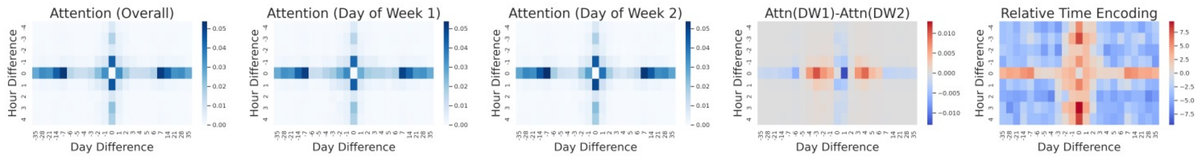
- Provides an attention bias independent of input features for each position.

Imputation Results

Method Category	Method	Missing Rate [0%, 20%)	[20%, 40%)	[40%, 60%)	[60%, 80%)	[80%, 100%)	Overall
Basic Fill	Zero Fill	474.03 ± 34.43	408.37 ± 28.70	384.86 ± 33.68	440.94 ± 43.45	422.73 ± 58.58	395.79 ± 18.08
	Forward Fill	416.22 ± 32.09	373.31 ± 29.22	329.69 ± 30.76	381.71 ± 40.83	352.61 ± 47.99	351.87 ± 16.05
	Backward Fill	411.61 ± 30.16	368.52 ± 28.74	324.53 ± 28.92	373.68 ± 39.72	343.54 ± 49.93	346.46 ± 15.98
	Avg. F+B Fill	350.79 ± 26.43	316.96 ± 25.38	278.37 ± 25.75	321.10 ± 34.29	295.56 ± 41.23	306.80 ± 13.86
Micro Mean Fill	Participant	414.57 ± 32.10	375.79 ± 27.87	329.57 ± 27.44	385.16 ± 37.74	367.59 ± 54.93	356.03 ± 16.32
	Day of Week	411.82 ± 31.97	373.30 ± 27.70	327.20 ± 27.36	380.14 ± 37.34	355.83 ± 51.14	351.70 ± 15.74
	Hour of Day	373.93 ± 29.07	343.80 ± 25.07	303.03 ± 24.84	351.89 ± 32.28	311.45 ± 35.07	326.44 ± 12.88
	DW+HD	357.56 ± 27.79	330.69 ± 24.15	288.72 ± 23.93	325.40 ± 29.62	253.56 ± 30.15	304.08 ± 12.12
Mean Fill	Participant	416.01 ± 32.22	378.09 ± 27.99	332.59 ± 27.72	389.68 ± 38.32	370.80 ± 53.86	358.64 ± 16.27
	Day of Week	413.29 ± 32.10	375.67 ± 27.83	330.21 ± 27.64	384.73 ± 37.88	359.88 ± 50.41	354.50 ± 15.74
	Hour of Day	375.40 ± 29.19	346.02 ± 25.17	305.75 ± 25.09	356.46 ± 32.80	317.22 ± 35.80	329.50 ± 13.04
	DW+HD	359.07 ± 27.90	332.97 ± 24.26	291.78 ± 24.21	330.89 ± 29.97	262.46 ± 30.73	308.03 ± 12.21
Median Fill	Participant	369.17 ± 26.48	331.99 ± 23.03	299.41 ± 24.88	341.93 ± 32.48	323.09 ± 46.69	323.69 ± 14.12
	Day of Week	366.66 ± 26.38	329.82 ± 22.84	297.14 ± 24.76	337.97 ± 32.19	314.87 ± 45.49	320.37 ± 13.92
	Hour of Day	335.85 ± 24.41	307.19 ± 21.37	276.57 ± 22.96	316.67 ± 29.11	280.66 ± 32.40	300.62 ± 11.85
	DW+HD	322.03 ± 23.78	295.09 ± 20.67	262.29 ± 22.04	292.01 ± 26.82	**230.54 ± 27.15**	280.39 ± 11.18
*k*NN	Uniform	332.70 ± 25.34	306.68 ± 24.28	270.34 ± 23.74	321.63 ± 36.07	295.93 ± 40.81	305.46 ± 13.93
	Softmax	331.37 ± 25.08	305.90 ± 24.19	269.58 ± 23.65	315.26 ± 33.16	290.78 ± 37.53	302.58 ± 13.22
Model-based	Iterative Imputation (Azur et al., 2011)	313.23 ± 23.26	290.48 ± 21.32	260.61 ± 22.02	304.16 ± 28.86	289.20 ± 37.43	291.54 ± 12.29
	CNN-DAE (Jang et al., 2020)	317.26 ± 23.42	287.35 ± 21.27	256.22 ± 22.38	299.14 ± 30.66	284.27 ± 40.75	288.85 ± 12.93
	Regression Imputation (Little, 1992)	307.96 ± 22.88	284.06 ± 20.69	254.82 ± 21.69	296.01 ± 27.76	282.21 ± 36.75	285.01 ± 12.01
	BRITS (Cao et al., 2018)	299.46 ± 21.59	275.58 ± 19.93	248.21 ± 21.40	289.82 ± 28.34	275.09 ± 36.46	277.63 ± 11.85
	USGAN (Miao et al., 2021)	299.58 ± 21.58	275.58 ± 19.91	248.34 ± 21.43	289.68 ± 28.36	274.56 ± 36.36	277.55 ± 11.84
	MRNN (Yoon et al., 2018)	295.30 ± 20.91	270.38 ± 19.66	243.96 ± 21.21	280.69 ± 27.25	263.42 ± 33.20	270.75 ± 11.27
	Sparse Self-Attention (ours)	**285.76 ± 20.41**	**262.96 ± 19.16**	**239.01 ± 20.58**	**270.32 ± 25.62**	250.36 ± 30.12	**261.68 ± 10.62**

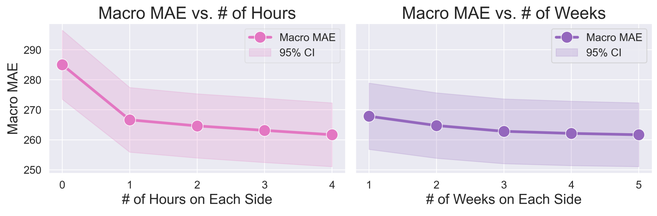
Results are Macro MAE ± 95% CI on completely held-out test participants. **Red**: overall best performance with statistical significance ($p < 0.05$). **Blue**: best performance with statistical significance ($p < 0.05$) within each method category.

Attention and Relative Time Encoding Visualization



Ablation Study

Local Context Window Size



Model Component

Model	Macro MAE
Full Model	**261.68 ± 10.62**
- RTE	262.91 ± 10.75
- LAPR	278.75 ± 11.93