Signed Dual Attention

Capturing Signed Dependencies in Time Series Forecasting

Balthazar Courvoisier[1,2], Tristan Cazenave [2]
[1] *Queensfield AI Technologies, Paris, France*
[2] *LAMSADE, Université Paris Dauphine - PSL, Paris, France*

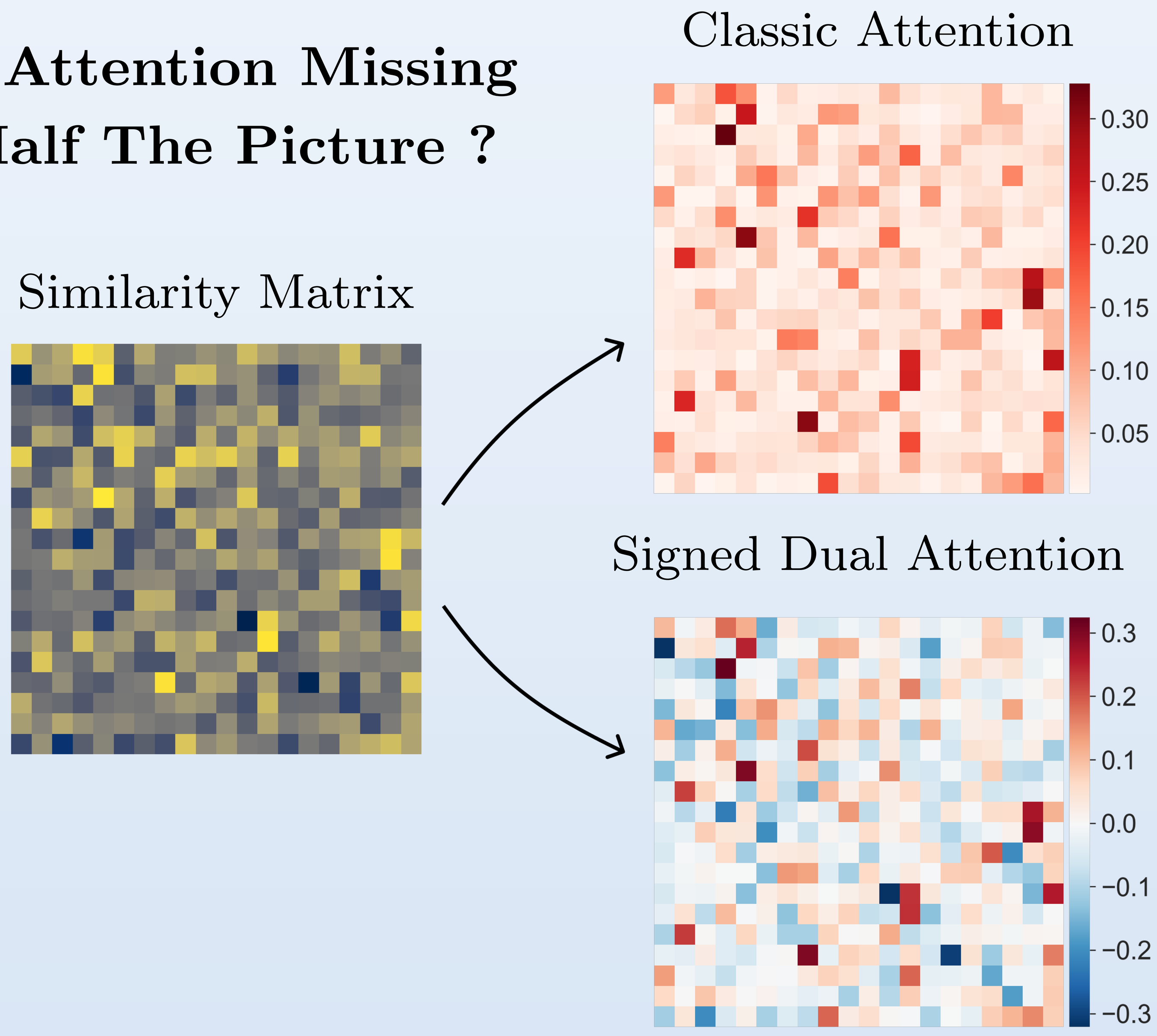
Motivation

- **Small Datasets:** insufficient observations for large models.
- **Signed Relationships:** series may move together or oppositely [1, 2].
- **Standard attention:** cannot capture alone signed relationships [3].
- **Weight Sharing**: lost if dealt with multi-head attention.

Proposition

- **Signed Dual Attention:** a novel attention mechanism.
- **Dual message passing:** mimics a two-head attention.
- **Parameter-efficient:** uses shared parameters.
- **Simple integration:** can replace standard attention.

Is Attention Missing Half The Picture ?



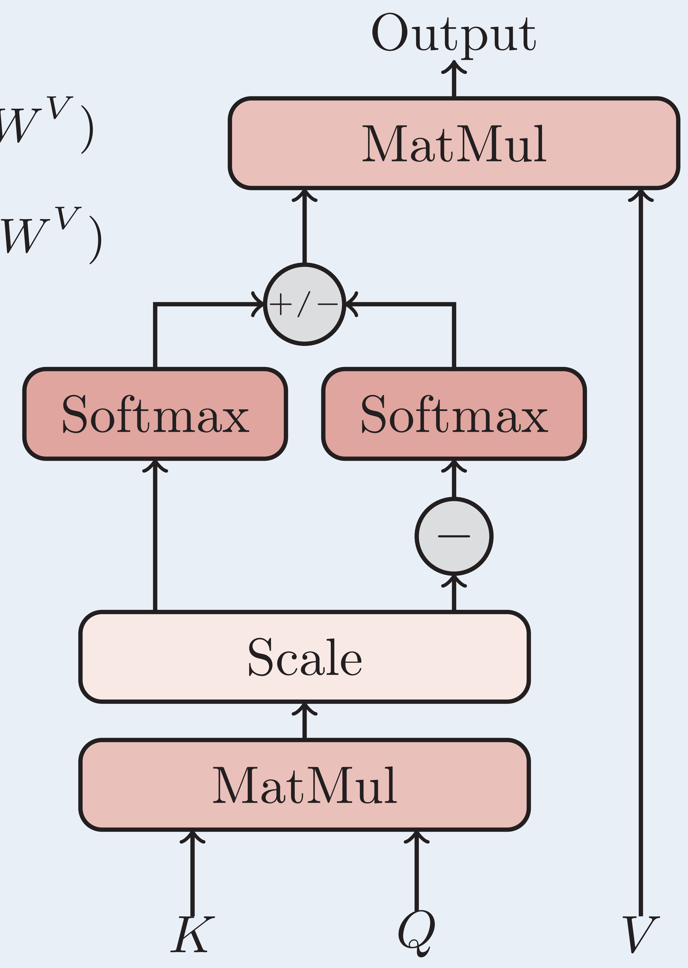
Architecture

$$(W_1^K, W_1^Q, W_1^V) = (+W^K, W^Q, +W^V)$$

$$(W_2^K, W_2^Q, W_2^V) = (-W^K, W^Q, -W^V)$$

$$W^O = \begin{bmatrix} I_d \\ I_d \end{bmatrix} \in \mathbb{R}^{2d \times d}$$

Under this configuration the output of a two-head mechanism exactly matches the SDA formulation.



Experiments

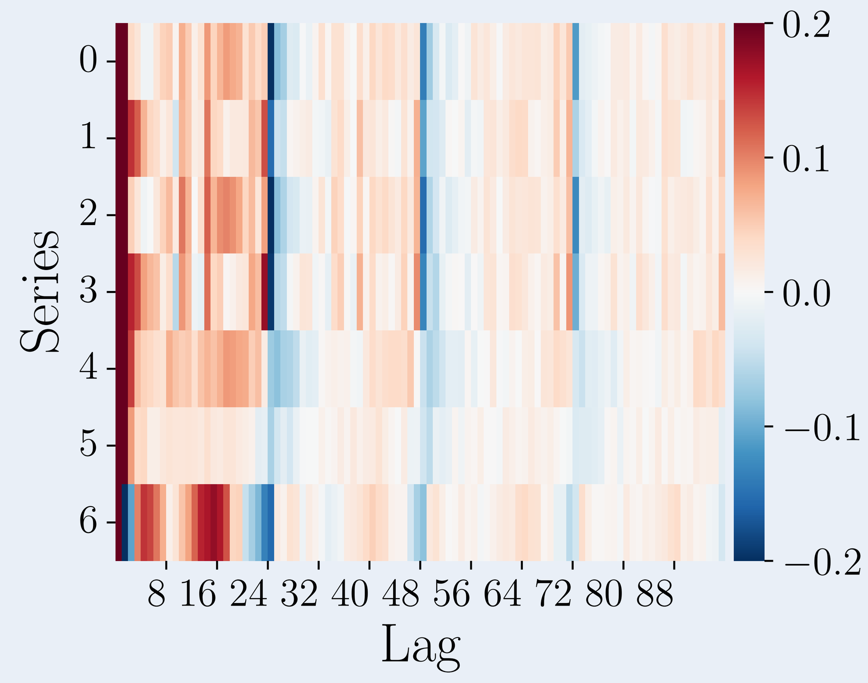
Impact on the Transformer Architecture [4]

		SDA			Classic		
		24	48	96	24	48	96
ECL	MSE	0.207	0.287	0.319	**0.199**	**0.252**	**0.31**
	MAE	0.337	0.398	0.42	**0.328**	**0.37**	**0.409**
Ettm2	MSE	0.024	**0.058**	0.137	**0.02**	0.099	**0.09**
	MAE	0.112	**0.173**	**0.187**	**0.102**	0.246	0.234
Etth2	MSE	0.103	**0.149**	**0.231**	**0.101**	0.159	0.238
	MAE	**0.25**	**0.31**	**0.387**	0.252	0.318	0.394
Exchange	MSE	0.081	0.375	1.112	**0.062**	**0.133**	**0.332**
	MAE	0.219	0.47	0.792	**0.195**	**0.289**	**0.441**
Traffic	MSE	0.191	0.231	**0.224**	**0.172**	**0.203**	0.254
	MAE	0.285	0.325	**0.315**	**0.267**	**0.302**	0.358
Weather	MSE	0.003	**0.01**	0.009	**0.002**	0.013	**0.004**
	MAE	0.04	0.075	0.074	**0.034**	**0.046**	**0.051**

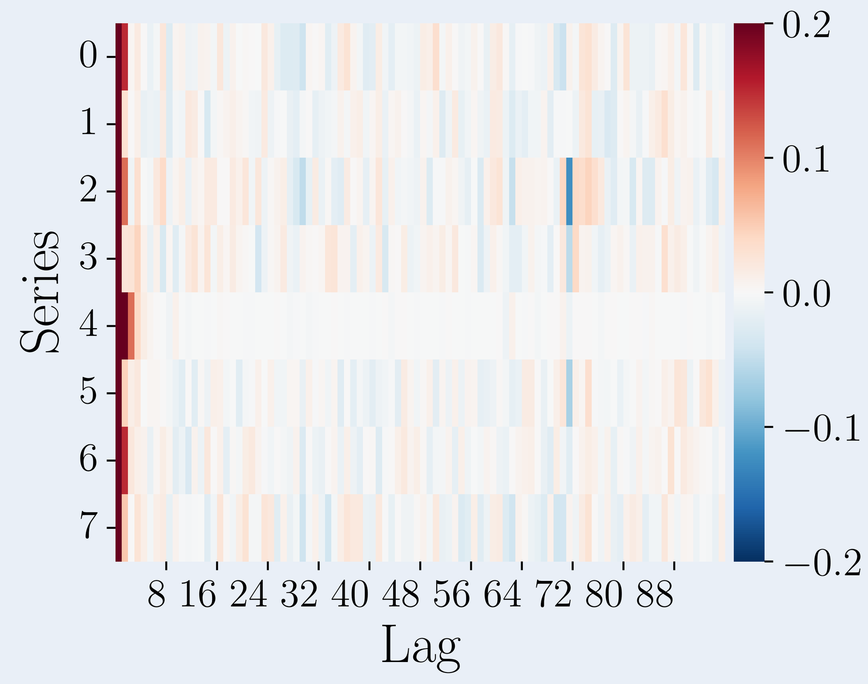
- **Training**: ADAM optimizer (lr $= 10^{-4}$, batch $= 32$) and early stopping.
- **Evaluation**: MSE and MAE (average of 3 runs) on univariate forecasting over 3 horizons.

Insights

EETh2 Autocorrelation Structure



Exchange Autocorrelation Structure



Examining the PACF, not all datasets exhibit signed relationships.

Conclusion & Future Work

- **Performance:** Mixed results overall, with notable improvements on ETTm2 and ETTh2 datasets but none on the Exchange dataset.
- **SDA Benefit:** Datasets with both positive and negative correlations, i.e. signed realtionships, gain the most from the SDA mechanism.
- **Next Steps:** Extend evaluation to multivariate forecasting and explore adaptive weighting by adjusting the concatenation mecanism.

References

[1] T. Zeng and J. Li. Maximization of negative correlations in time-course gene expression data. *NAR*, 2009.

[2] S. Agrawal, M. Steinbach, D. Boley, S. Chatterjee, G. Atluri, A. The Dang, S. Liess, and V. Kumar. Mining novel multivariate relationships in time series data using correlation networks. *IEEE Transactions on Knowledge and Data Engineering*, page 1–1, 2019.

[3] Junjie Huang, Huawei Shen, Liang Hou, and Xueqi Cheng. Signed graph attention networks, 2019.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and L. Kaiser. Attention is all you need, 2017.