

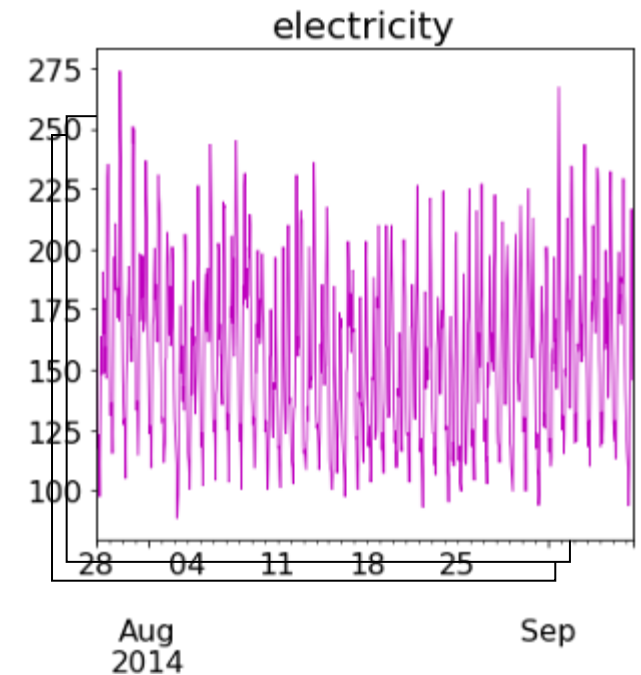
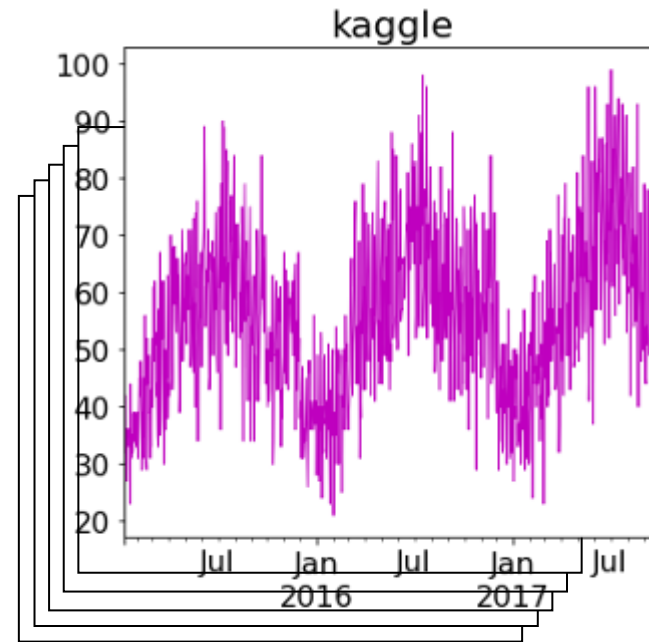
Domain Adaptation for Time Series Forecasting via Attention Sharing

Xiaoyong Jin, Youngsuk Park, Danielle Robinson, Hao Wang, Bernie Wang

ICML 2022

Problem

- Forecasting with data scarcity
 - Limited data from a target domain
 - Abundant data from a source domain
- Domain adaptation:
 - Learn a model mainly on the data-rich source domain
 - Transfer certain knowledge to the data-scarce target domain by adaptation

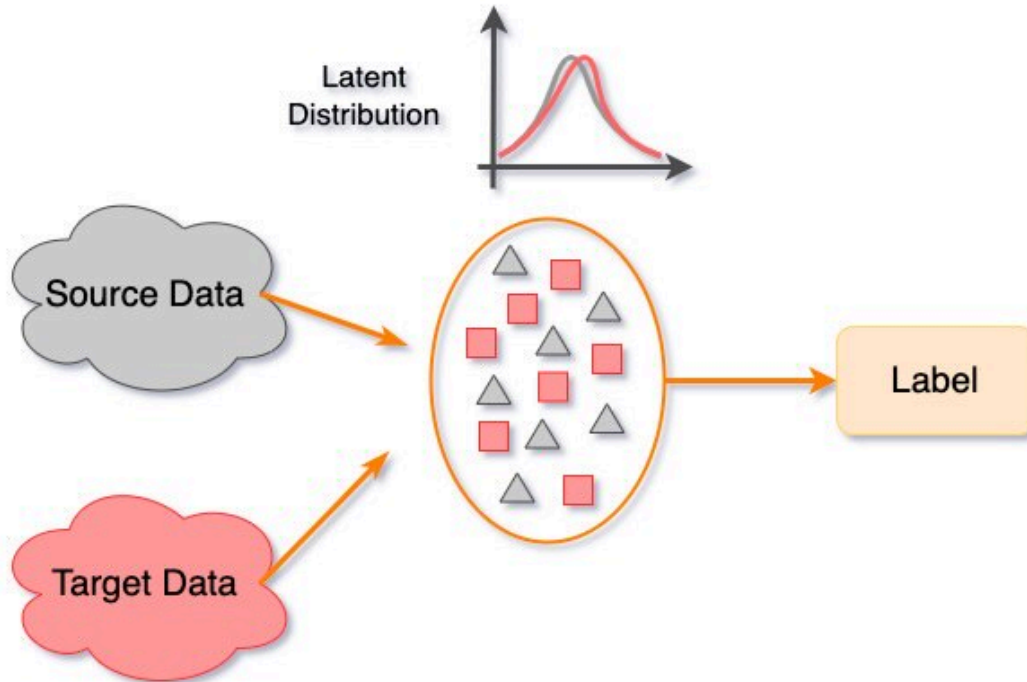


Source

Domain Adaptation

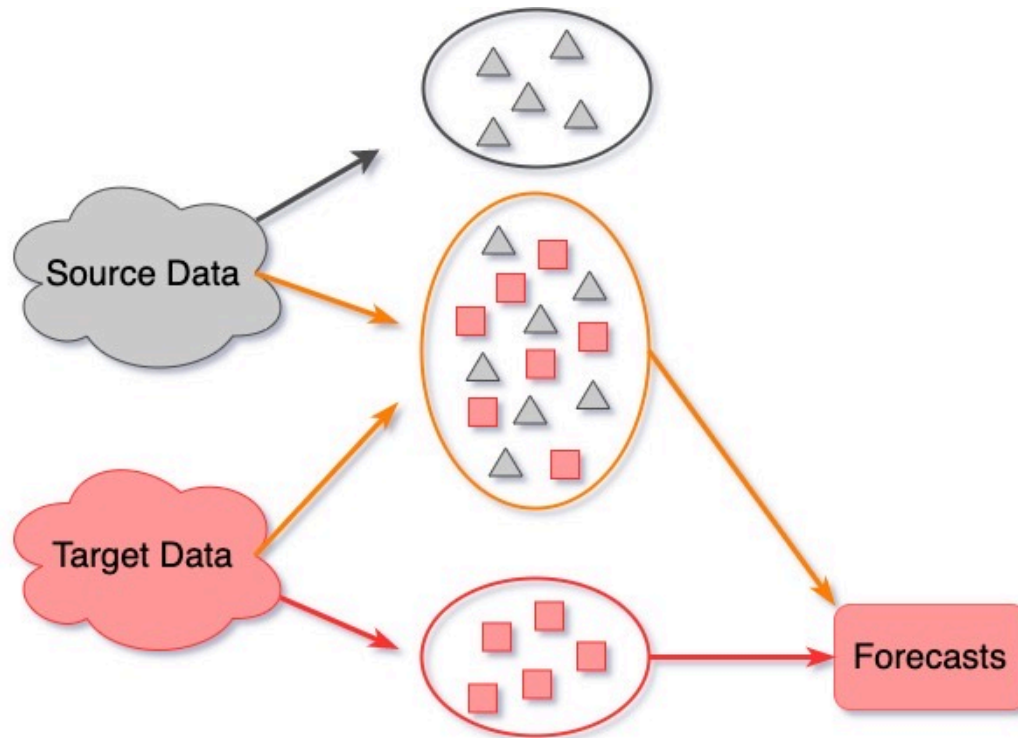
Target

Existing Domain Adaptation



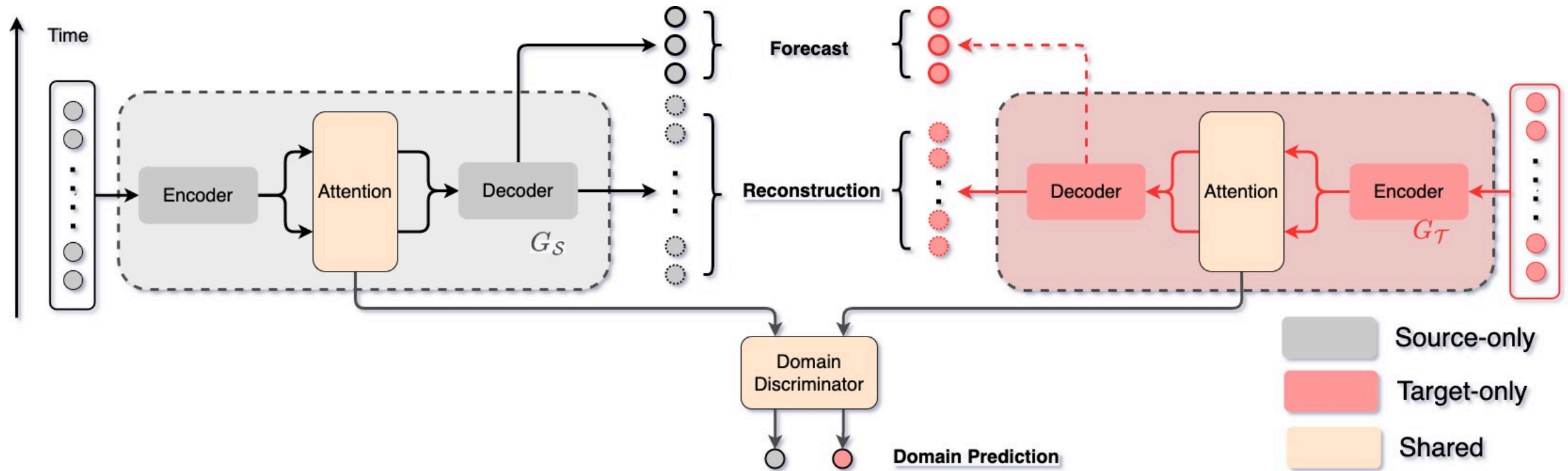
- Learn domain-invariant features
- Distinctions between domains do not affect predictions from features
- Methods:
 - Metric-based regularization
 - Adversarial training

DA in Forecasting

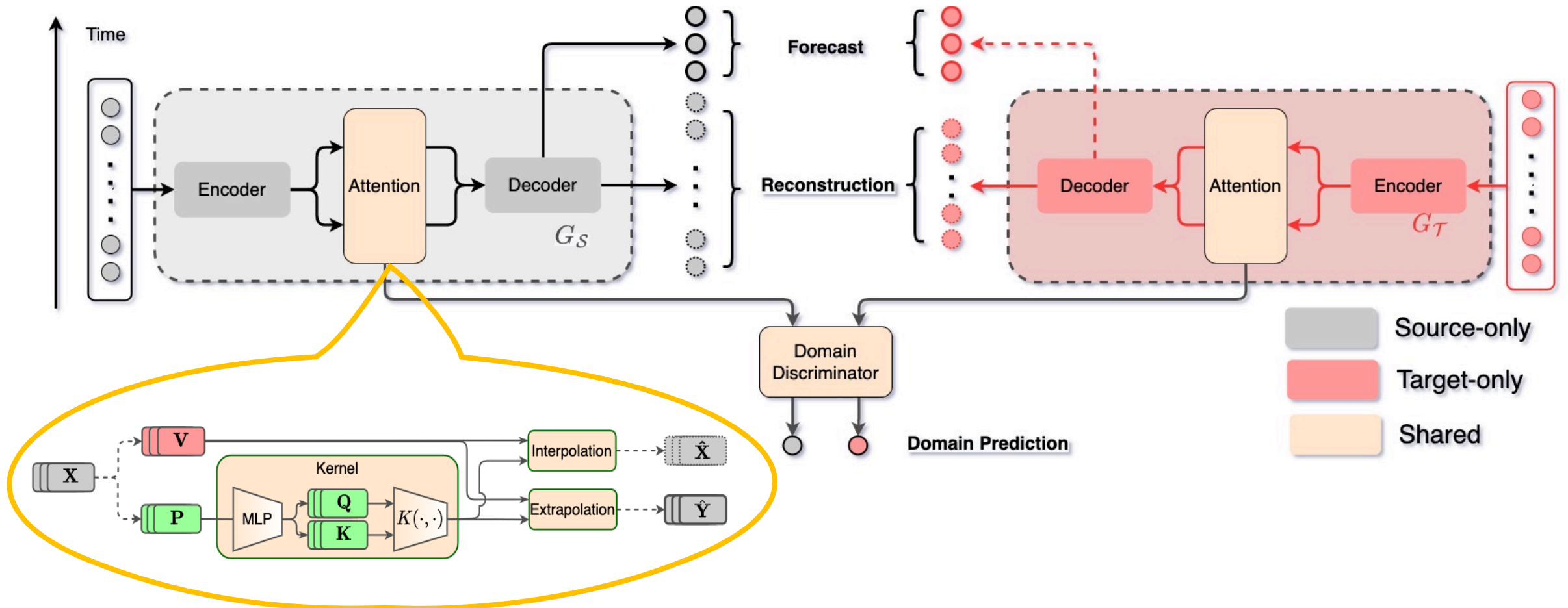


- Domain-specific features are necessary for domain-dependent forecasts
- Domain-invariant features to connect both domains
 - Query-key matching in Attention
- Domain-specific features are retained to make domain-dependent forecasts
 - Value combination in Attention

Domain Adaptation Forecaster



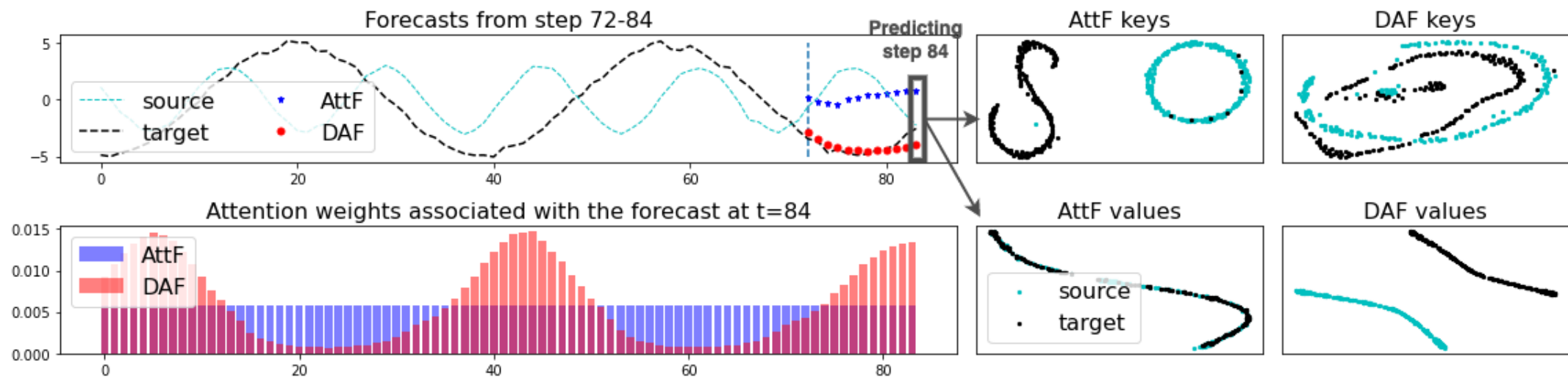
Attention Sharing Strategy



How it works

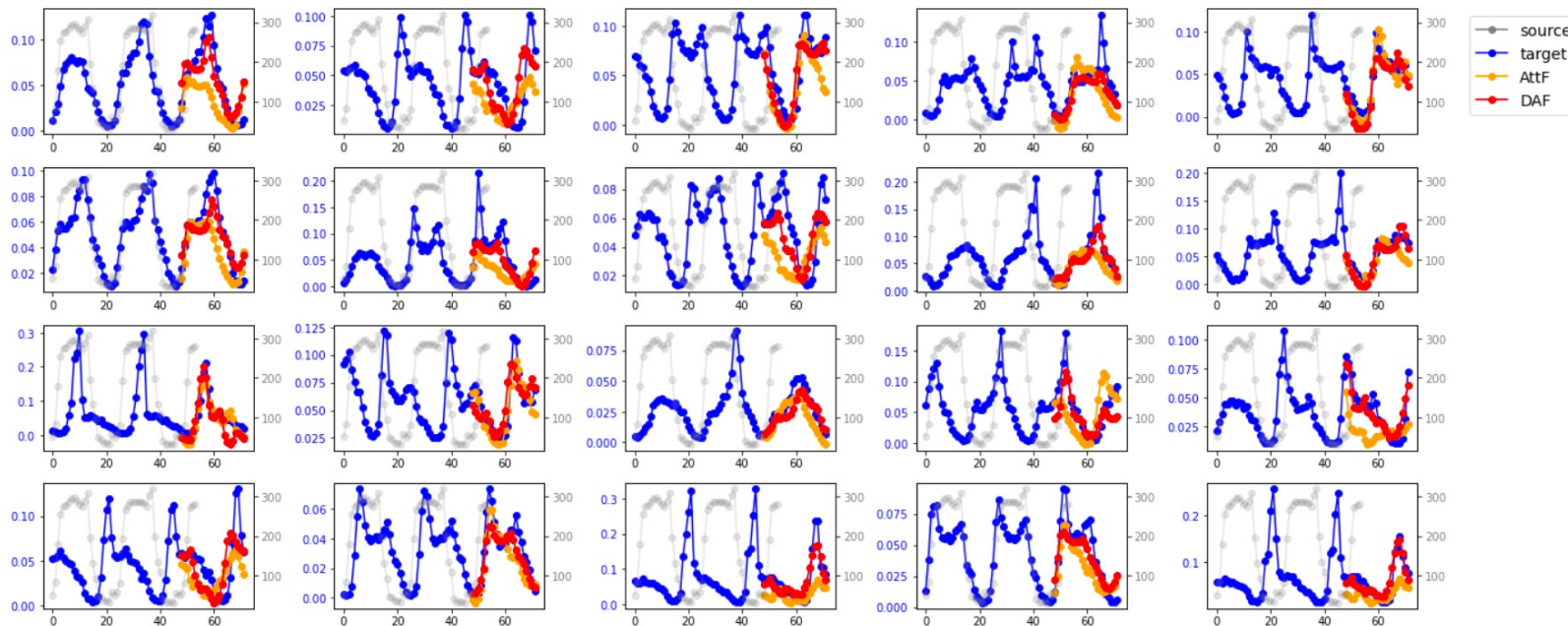
Key observations

- Attention keys are aligned in DAF but not in single-domain Attention Forecaster (AttF)
- More reasonable attention weights in DAF than in AttF
- Better forecasts from DAF
- Attention values stay distinct across domains



Results

\mathcal{D}_T	\mathcal{D}_S	τ	DAR	AttF	SASA	DATSING	RDA-DANN	RDA-ADDA	RDA-MMD	DAF
traf	<i>elec</i>	24	0.205 ± 0.015	0.182 ± 0.007	0.177 ± 0.004	0.184 ± 0.004	0.181 ± 0.009	0.174 ± 0.005	0.186 ± 0.004	0.169 ± 0.002
	<i>wiki</i>				0.197 ± 0.001	0.189 ± 0.005	0.180 ± 0.004	0.181 ± 0.003	0.179 ± 0.004	0.176 ± 0.004
elec	<i>traf</i>		0.141 ± 0.023	0.137 ± 0.005	0.164 ± 0.001	0.137 ± 0.003	0.133 ± 0.005	0.134 ± 0.002	0.140 ± 0.006	0.125 ± 0.008
	<i>sales</i>				0.160 ± 0.001	0.149 ± 0.009	0.135 ± 0.007	0.142 ± 0.003	0.144 ± 0.003	0.123 ± 0.005
wiki	<i>traf</i>	7	0.055 ± 0.010	0.050 ± 0.003	0.053 ± 0.001	0.049 ± 0.002	0.047 ± 0.005	0.045 ± 0.003	0.045 ± 0.003	0.042 ± 0.004
	<i>sales</i>				0.053 ± 0.001	0.052 ± 0.004	0.053 ± 0.002	0.049 ± 0.003	0.052 ± 0.004	0.049 ± 0.003
sales	<i>elec</i>		0.305 ± 0.005	0.308 ± 0.002	0.451 ± 0.001	0.301 ± 0.008	0.297 ± 0.004	0.281 ± 0.001	0.291 ± 0.004	0.277 ± 0.005
	<i>wiki</i>				0.301 ± 0.001	0.305 ± 0.008	0.287 ± 0.009	0.287 ± 0.002	0.289 ± 0.003	0.280 ± 0.007



Takeaways

1. Knowledge of forecasting can be transferred from data-rich domains to data-scarce domains
2. Attention mechanism is suitable for domain adaptation, where
 1. queries/keys can be induced to be invariant across domains;
 2. values can stay distinct to make domain-dependent forecasts