

A decorative graphic consisting of a large, light gray circle with a thin black outline. Two solid red dots are positioned on the circle's circumference, one in the upper-left quadrant and one in the lower-right quadrant.

Temporal Pattern Attention for Multivariate Time Series Forecasting

Shun-Yao Shih, Fan-Keng Sun, Hung-yi Lee

Submitted to the journal track of ECML PKDD 2019

多变量时间序列 (Multivariate Time Series, MTS) 之间的复杂的动态相互依赖性很重要，但是难以捕获和分析。

根据历史数据预测未来。不同序列之间的相互依赖性越好，模型的预测就越准确。

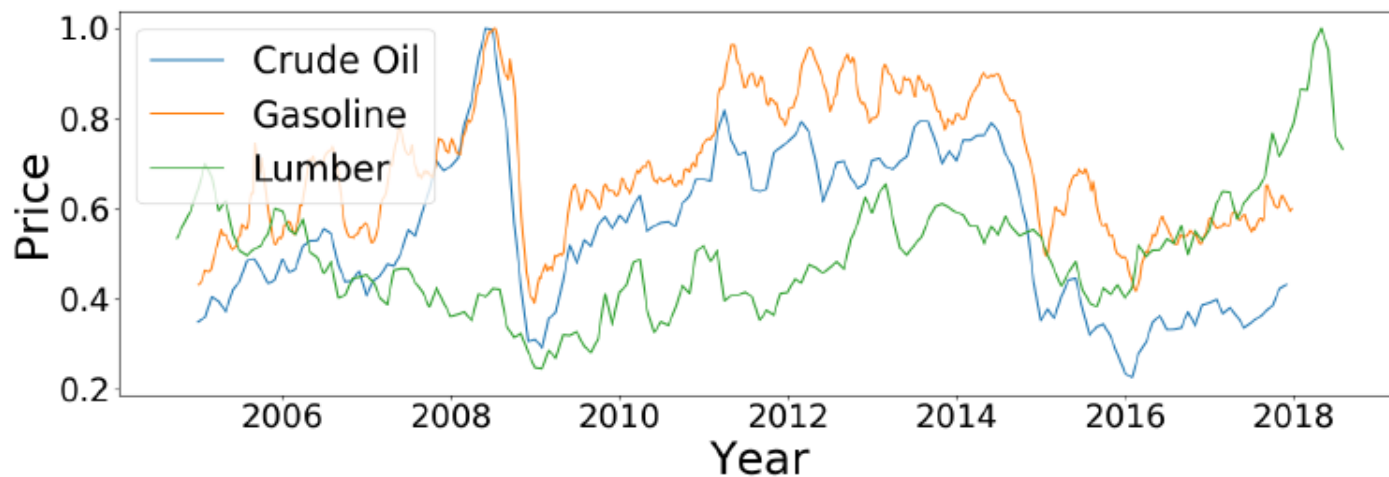
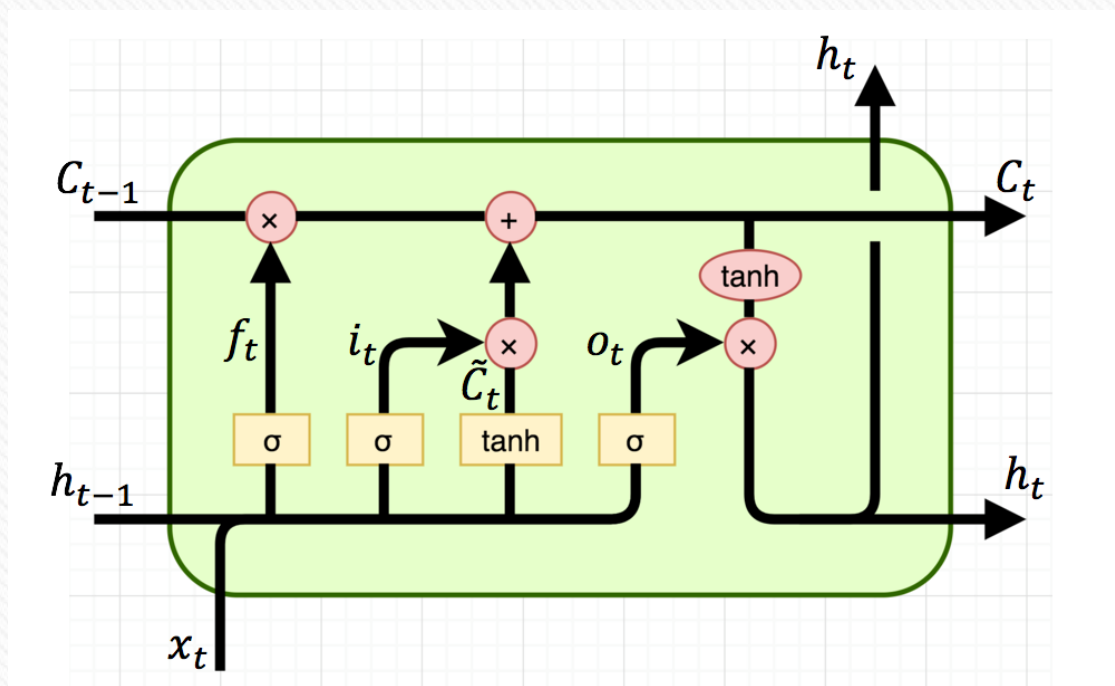


Fig. 1 Historical prices of crude oil, gasoline, and lumber. Units are omitted and scales are normalized for simplicity.

我们希望模型可以从数据中自动学习这种相互依赖性。

传统的 RNN 注意力机制



传统的 RNN 注意力机制选择与当前时间步相关的信息，并从这些时间步中提取信息。

$$H = \{h_1, h_2, \dots, h_{t-1}\}$$

上下文向量 v_t 是先前 RNN 隐状态 H 的列向量 h_i 的加权和，其表示与当前时间步相关的信息。 v_t 进一步与当前状态 h_t 整合以产生预测。

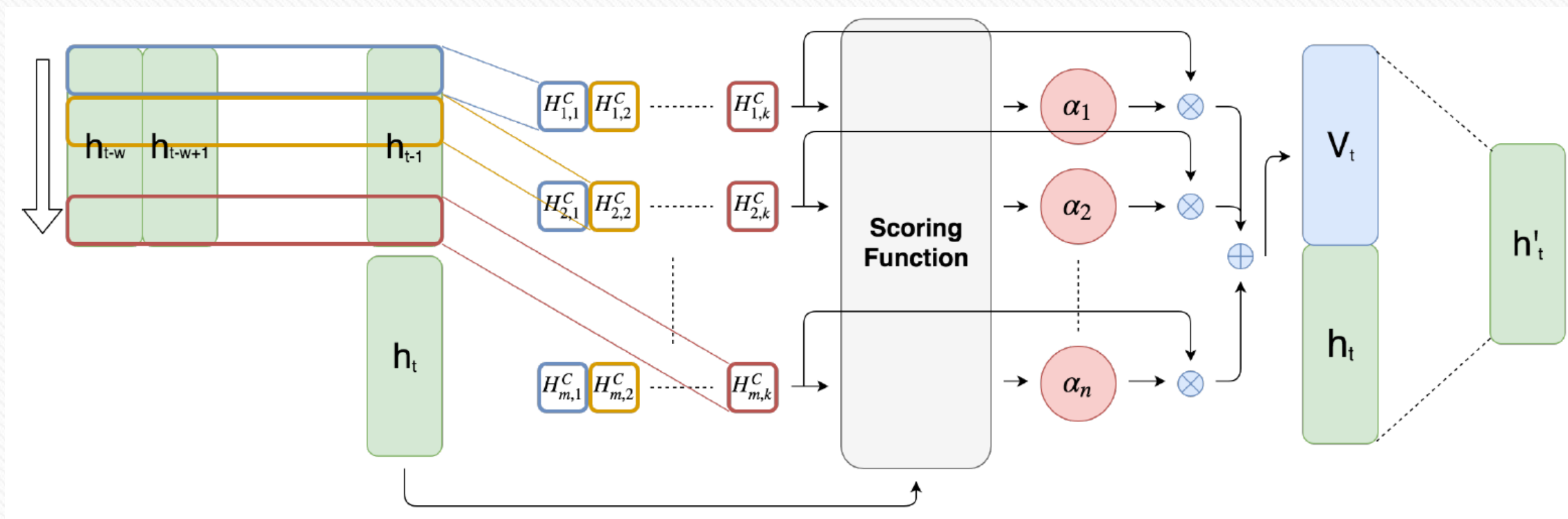
Temporal Pattern Attention (TPA)

一种用于 MTS 预测的新的注意力机制，使用时间模式 (Temporal Pattern) 来指代任何跨多个时间步的时间不变模式。在 TPA 中，机器会学习去选择相关的时间序列 (Time Series)，而不是像传统的注意机制那样选择相关的时间步 (Time Steps)。

时间序列数据通常会有明显的周期性时间模式。然而，跨越多个时间步的周期性模式对于传统的注意机制来说难以识别。

在 TPA 中，引入卷积神经网络 (CNN) 来从每个单独的变量中提取时间模式信息。

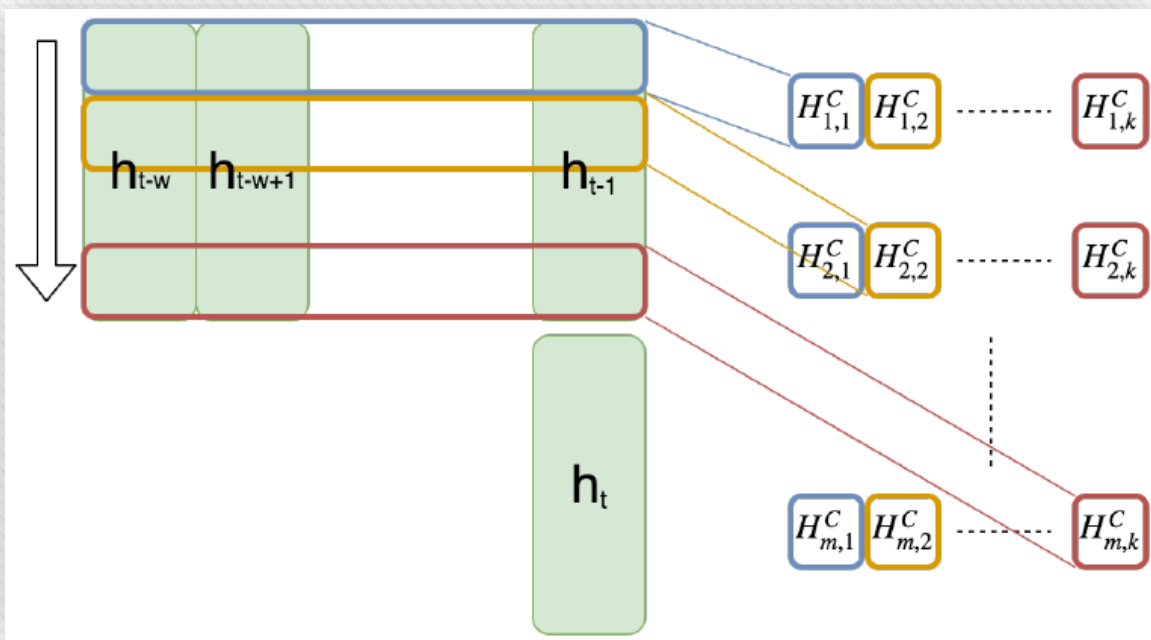
Temporal Pattern Attention (TPA)



在我们所提出的方法中，给定先前的 RNN 隐状态 H ，我们的注意机制基本上关注其行向量。行上的注意力权重选择那些有助于预测的变量。

上下文向量 v_t 现在是包含跨多个时间步信息的行向量的加权和。

Temporal Pattern Detection using CNN

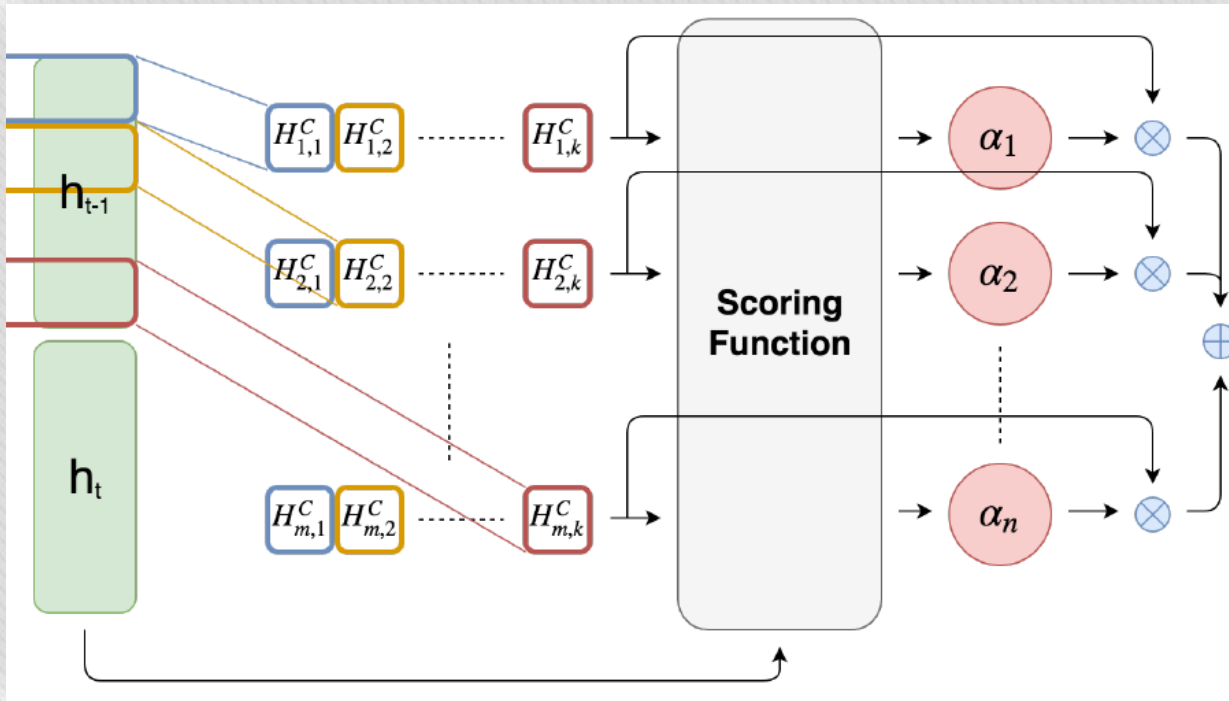


有 k 个滤波器 $C_i \in \mathbb{R}^{1 \times T}$, 其中 T 是我们需要注意的最大长度。

卷积运算产生 $H^C \in \mathbb{R}^{n \times k}$, $H_{i,j}^C$ 表示第 j 个滤波器在第 i 行产生的结果。

$$H_{i,j}^C = \sum_{l=1}^w H_{i,(t-w-1+l)} \times C_{j,T-w+l}$$

Proposed Attention Mechanism

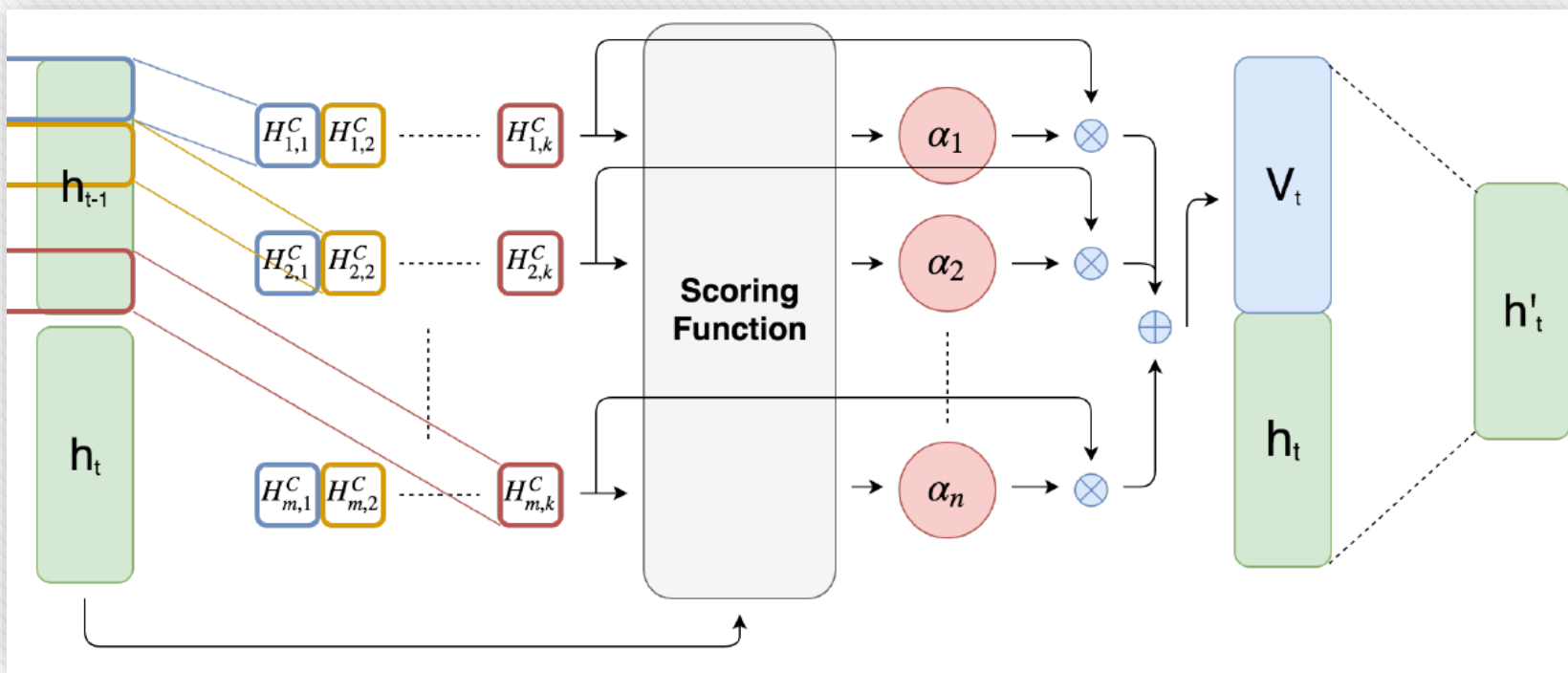


$$f(H_i^C, h_t) = H_i^C W_a h_t$$

$$\alpha_i = \text{sigmoid}(f(H_i^C, h_t))$$

$$v_t = \sum_{i=1}^m \alpha_i (H_i^C)^\top$$

Proposed Attention Mechanism



$$h'_t = W_h h_t + W_v v_t$$

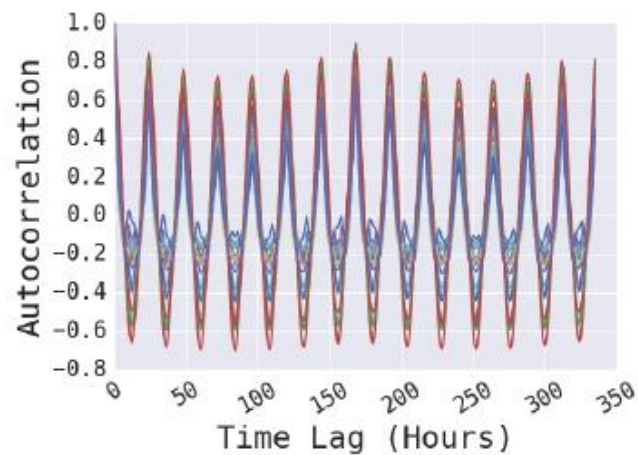
$$y_{t-1+\Delta} = W_{h'} h'_t$$

Failure of Traditional Attention Mechanisms

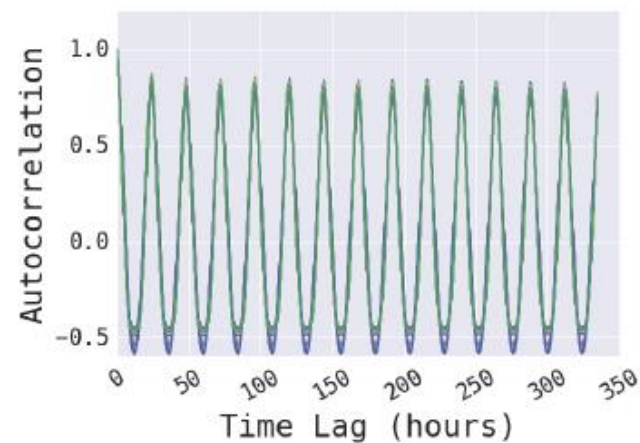
直观地，对于传统的注意力模型，可以通过记住恰好在一个周期之前出现的值来准确地预测下一个值。但是，我们知道不同的时间序列具有不同的周期，这意味着要具有良好的预测，模型应该能够回顾不同序列中不同数量的时间步。

从这一点来看，很明显传统注意机制的失败来自于只提取一个前一个时间步，而折扣在其他时间步中的信息。另一方面，我们的注意机制通过使用 CNN 滤波器从 RNN 隐藏状态的行向量提取的特征，这使得模型能够跨多个时间步选择相关信息。

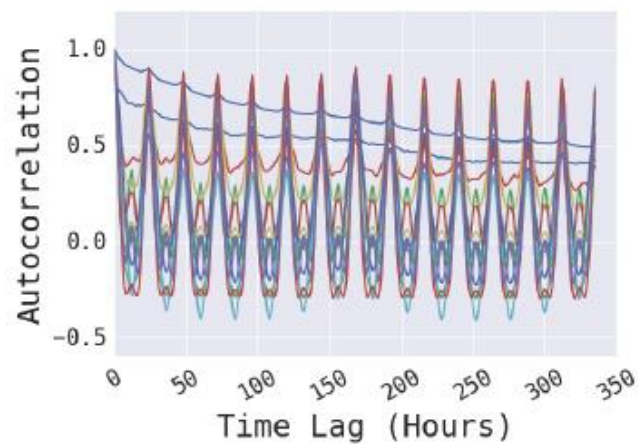
Experiments and Analysis



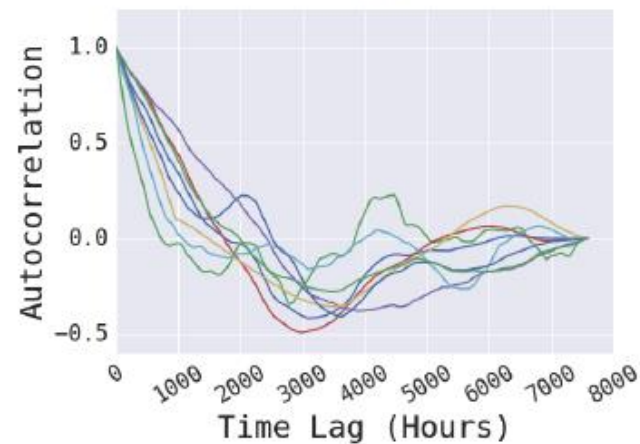
(a) Traffic dataset



(b) Solar-Energy dataset



(c) Electricity dataset



(d) Exchange-Rate dataset

Experiments and Analysis

RSE	Solar Energy				Traffic			
Horizon	3	6	12	24	3	6	12	24
AR	0.2435	0.3790	0.5911	0.8699	0.5991	0.6218	0.6252	0.6293
LRidge	0.2019	0.2954	0.4832	0.7287	0.5833	0.5920	0.6148	0.6025
LSVR	0.2021	0.2999	0.4846	0.7300	0.5740	0.6580	0.7714	0.5909
GP	0.2259	0.3286	0.5200	0.7973	0.6082	0.6772	0.6406	0.5995
LSTNet-Skip	0.1843	0.2559	<u>0.3254</u>	0.4643	<u>0.4777</u>	<u>0.4893</u>	<u>0.4950</u>	<u>0.4973</u>
LSTNet-Attn	<u>0.1816</u>	<u>0.2538</u>	0.3466	<u>0.4403</u>	0.4897	0.4973	0.5173	0.5300
Our model	0.1803 ± 0.0008	0.2347 ± 0.0017	0.3234 ± 0.0044	0.4389 ± 0.0084	0.4487 ± 0.0180	0.4658 ± 0.0053	0.4641 ± 0.0034	0.4765 ± 0.0068

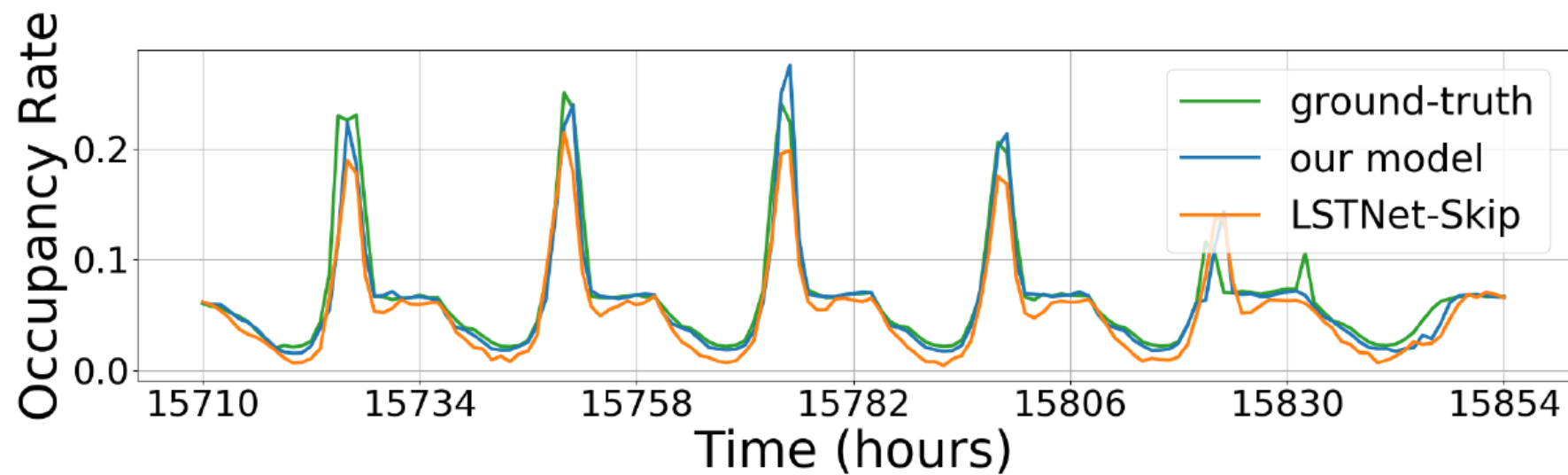
RSE	Electricity				Exchange Rate			
Horizon	3	6	12	24	3	6	12	24
AR	0.0995	0.1035	0.1050	0.1054	0.0228	0.0279	<u>0.0353</u>	<u>0.0445</u>
LRidge	0.1467	0.1419	0.2129	0.1280	<u>0.0184</u>	0.0274	0.0419	0.0675
LSVR	0.1523	0.1372	0.1333	0.1180	0.0189	0.0284	0.0425	0.0662
GP	0.1500	0.1907	0.1621	0.1273	0.0239	<u>0.0272</u>	0.0394	0.0580
LSTNet-Skip	<u>0.0864</u>	<u>0.0931</u>	0.1007	<u>0.1007</u>	0.0226	0.0280	0.0356	0.0449
LSTNet-Attn	0.0868	0.0953	<u>0.0984</u>	0.1059	0.0276	0.0321	0.0448	0.0590
Our model	0.0823 ± 0.0012	0.0916 ± 0.0018	0.0964 ± 0.0015	0.1006 ± 0.0015	0.0174 ± 0.0001	0.0243 ± 0.0003	0.0345 ± 0.0010	0.0444 ± 0.0006

Experiments and Analysis

CORR	Solar Energy				Traffic			
Horizon	3	6	12	24	3	6	12	24
AR	0.9710	0.9263	0.8107	0.5314	0.7752	0.7568	0.7544	0.7519
LRidge	0.9807	0.9568	0.8765	0.6803	0.8038	0.8051	0.7879	0.7862
LSVR	0.9807	0.9562	0.8764	0.6789	0.7993	0.7267	0.6711	0.7850
GP	0.9751	0.9448	0.8518	0.5971	0.7831	0.7406	0.7671	0.7909
LSTNet-Skip	0.9843	0.9690	<u>0.9467</u>	0.8870	<u>0.8721</u>	<u>0.8690</u>	<u>0.8614</u>	<u>0.8588</u>
LSTNet-Attn	<u>0.9848</u>	<u>0.9696</u>	0.9397	<u>0.8995</u>	0.8704	0.8669	0.8540	0.8429
Our model	0.9850 ± 0.0001	0.9742 ± 0.0003	0.9487 ± 0.0023	0.9081 ± 0.0151	0.8812 ± 0.0089	0.8717 ± 0.0034	0.8717 ± 0.0021	0.8639 ± 0.0030

CORR	Electricity				Exchange Rate			
Horizon	3	6	12	24	3	6	12	24
AR	0.8845	0.8632	0.8591	0.8595	0.9734	0.9656	0.9526	0.9357
LRidge	0.8890	0.8594	0.8003	0.8806	<u>0.9788</u>	0.9722	0.9543	0.9305
LSVR	0.8888	0.8861	0.8961	0.8891	0.9782	0.9697	<u>0.9546</u>	<u>0.9370</u>
GP	0.8670	0.8334	0.8394	0.8818	0.8713	0.8193	0.8484	0.8278
LSTNet-Skip	<u>0.9283</u>	<u>0.9135</u>	<u>0.9077</u>	<u>0.9119</u>	0.9735	0.9658	0.9511	0.9354
LSTNet-Attn	0.9243	0.9095	0.9030	0.9025	0.9717	0.9656	0.9499	0.9339
Our model	0.9429 ± 0.0004	0.9337 ± 0.0011	0.9250 ± 0.0013	0.9133 ± 0.0008	0.9790 ± 0.0003	<u>0.9709</u> ± 0.0003	0.9564 ± 0.0005	0.9381 ± 0.0008

Experiments and Analysis



Experiments and Analysis

Dataset	Solar Energy			Traffic		
	Position	Filter	W/o CNN	Position	Filter	W/o CNN
Softmax	<u>0.4391</u>	0.4434	0.4489	0.4715	0.4897	0.4770
Sigmoid	0.4389	0.4597	0.4507	<u>0.4765</u>	0.4795	0.4796
Concat	0.4462	0.4404	0.4951	0.4855	0.4774	0.4795

Dataset	Electricity			MuseData		
	Position	Filter	W/o CNN	Position	Filter	W/o CNN
Softmax	0.1006	0.1007	0.1011	0.04931	0.04968	<u>0.04902</u>
Sigmoid	0.1006	0.1028	0.1008	0.04878	0.04958	0.04987
Concat	0.1026	0.1035	0.1014	0.05191	0.05167	0.05128

Table 4 Ablation Study. Evaluation measure for Solar Energy, Traffic, and Electricity is RSE, and negative log-likelihood for MuseData. On each corpus, bold text represents the best and underlined text represents second best.



THANKS