西安交通大學
XI'AN JIAOTONG UNIVERSITY

Spatiotemporal Attention Networks for Wind Power Forecasting

Xingbo Fu[1, *], Feng Gao[1], Jiang Wu[1], Xinyu Wei[2], Fangwei Duan[3]

[1]System Engineering Institute, Xi'an Jiaotong University, Xi'an, China

[2]Department of Statistics, Rutgers University, New Brunswick, NJ, USA

[3]Electric Power Research Institute, State Grid Liaoning Electric Power Co. Ltd., Shenyang, China

西安交通大學
XI'AN JIAOTONG UNIVERSITY

CONTENTS

Motivation & Background

Suppose there are N wind farms, each of which monitors wind power generation time series at that wind farm. Given a time window with T timestamps, $X = (x_1, x_2, \cdots, x_t, \cdots, x_T) \in \mathbb{R}^{N \times T}$ is denoted as wind power generations of all the wind farms for T timestamps. For the t th timestamp, we denote $x_t = (x_t^1, x_t^2, \cdots, x_t^N)^T \in \mathbb{R}^{N \times 1}$ as the wind power generation of all the wind farms at timestamp t.

The wind power forecasting problem is to predict the future wind power generations $\hat{x}_{T+n}$ at timestamp $T + n$

$$\hat{x}_{T+n} = f(X) = f(x_1, x_2, \cdots, x_t, \cdots, x_T)$$

XI'AN JIAOTONG UNIVERSITY

Existing Methods

- Statistical methods

  Historical average (HA)

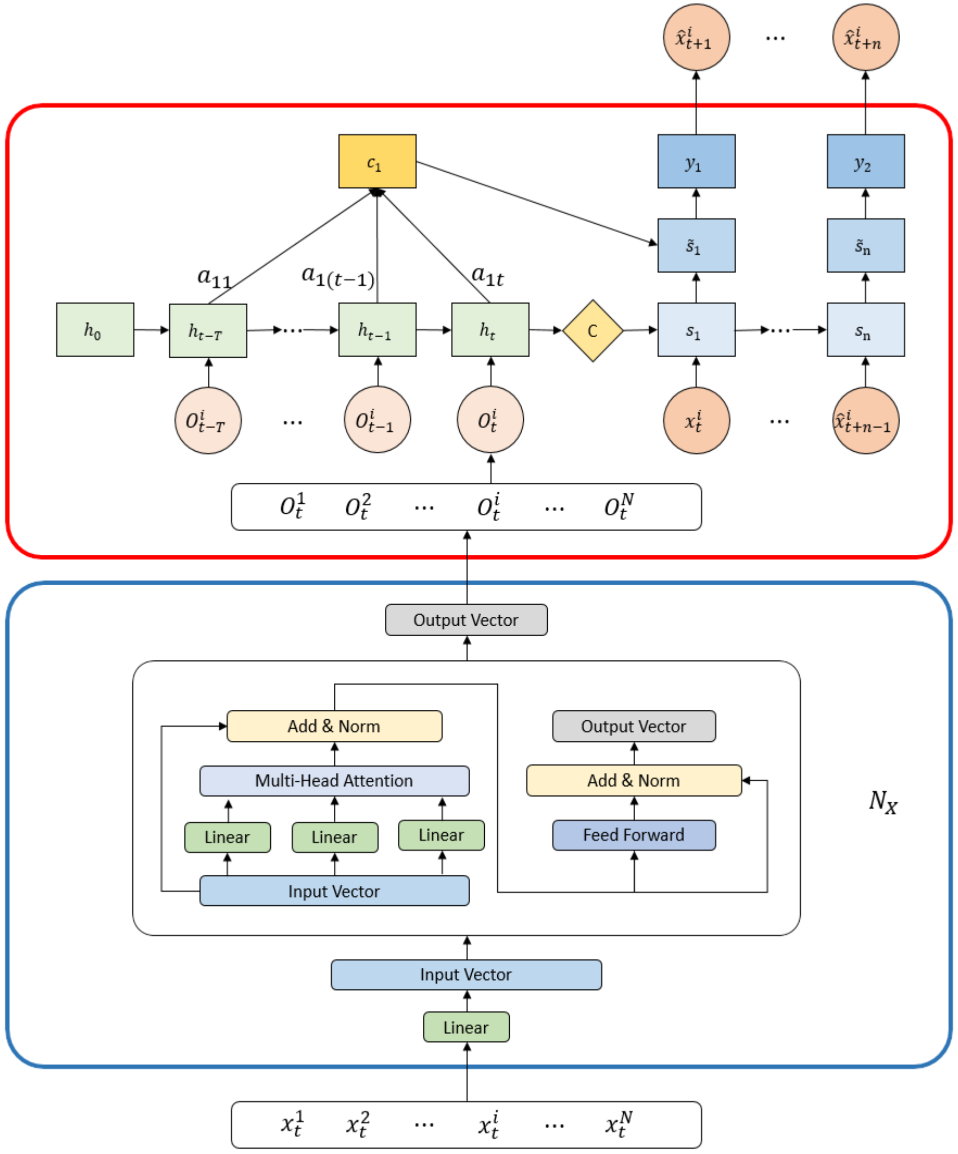
  ARMA & ARIMA - ~~stationary stochastic process~~

- Neural networks

  Temporal correlations: RNN, LSTM & Seq2Seq

  Spatial correlations: CNN - ~~grid-like data~~

Our Approach – Structure of STAN



Temporal attention mechanism

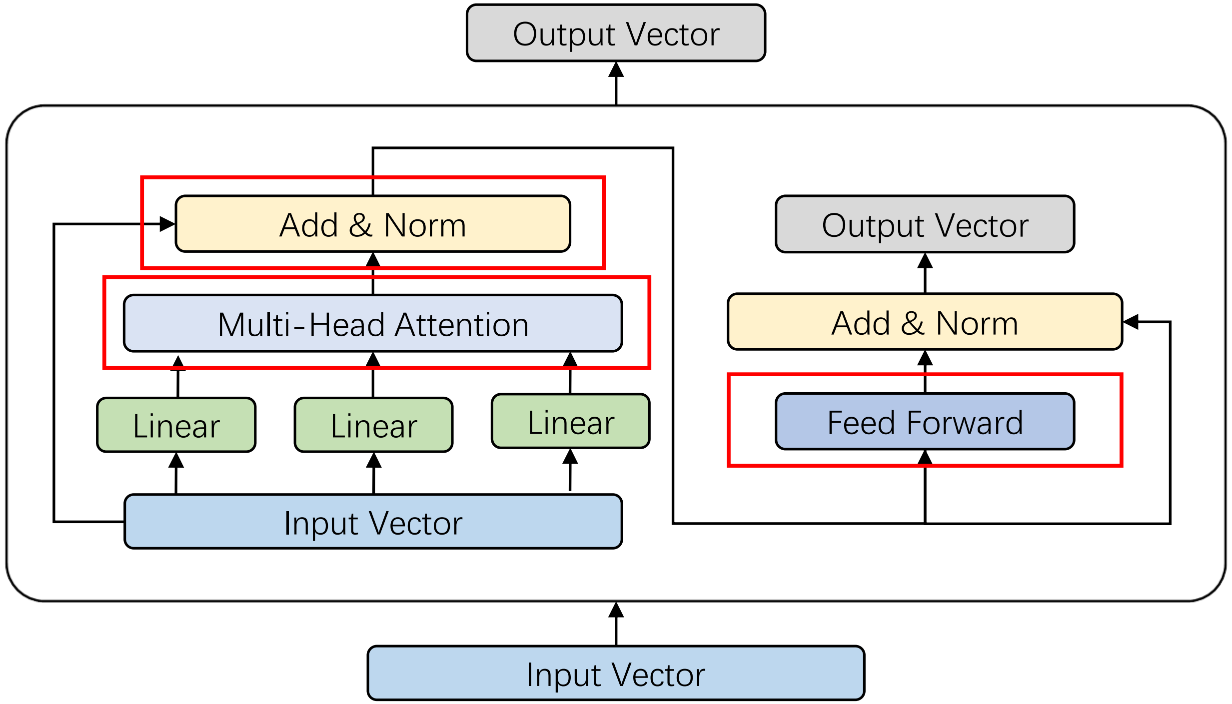
- *Capture temporal dependencies*

Spatial self-attention mechanism

- *Extract spatial correlations among wind farms*

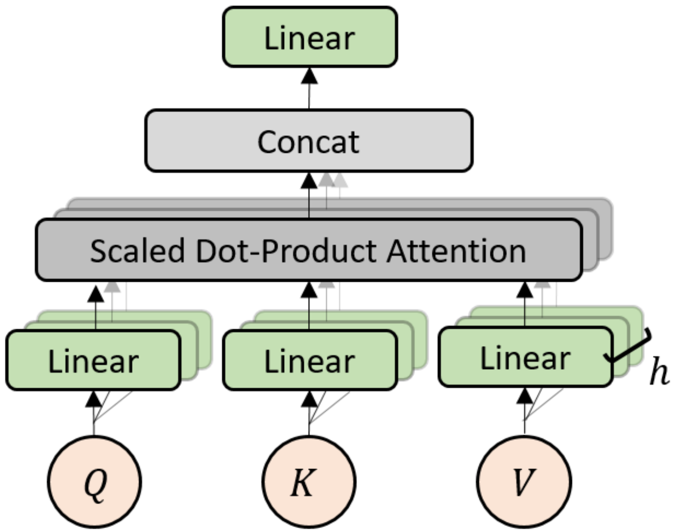
Our Approach – Structure of STAN

Spatial self-attention mechanism



XI'AN JIAOTONG UNIVERSITY

Spatial self-attention mechanism



Self attention

Self attention

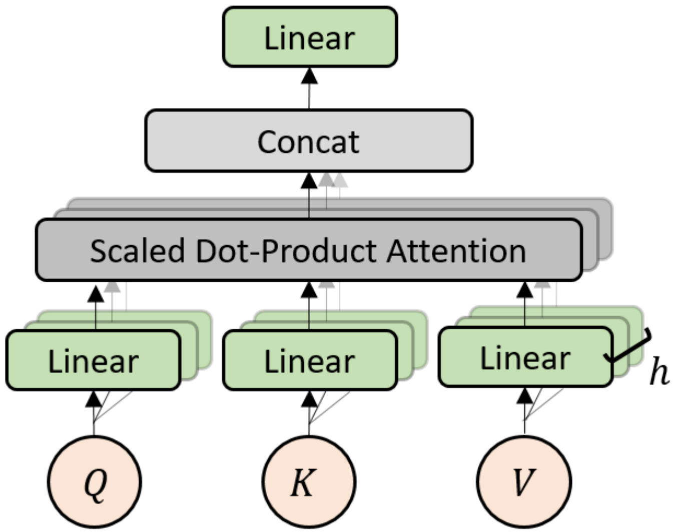
- Query vector - $Q \in \mathbb{R}^{N \times d_m}$
- Key vector - $K \in \mathbb{R}^{N \times d_m}$
- Value vector - $V \in \mathbb{R}^{N \times d_m}$

Scaled dot-product attention

$$Attention(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_m}}\right)V$$

Our Approach – Structure of STAN

Spatial self-attention mechanism



Self attention

Each of these three vectors Q, K, V is linearly projected h times respectively

$$Q_i' = QW_i^Q$$
$$K_i' = KW_i^K$$
$$W_i' = VW_i^V$$

$i = 1,2,\cdots,h$

We get h scaled dot-product attention functions

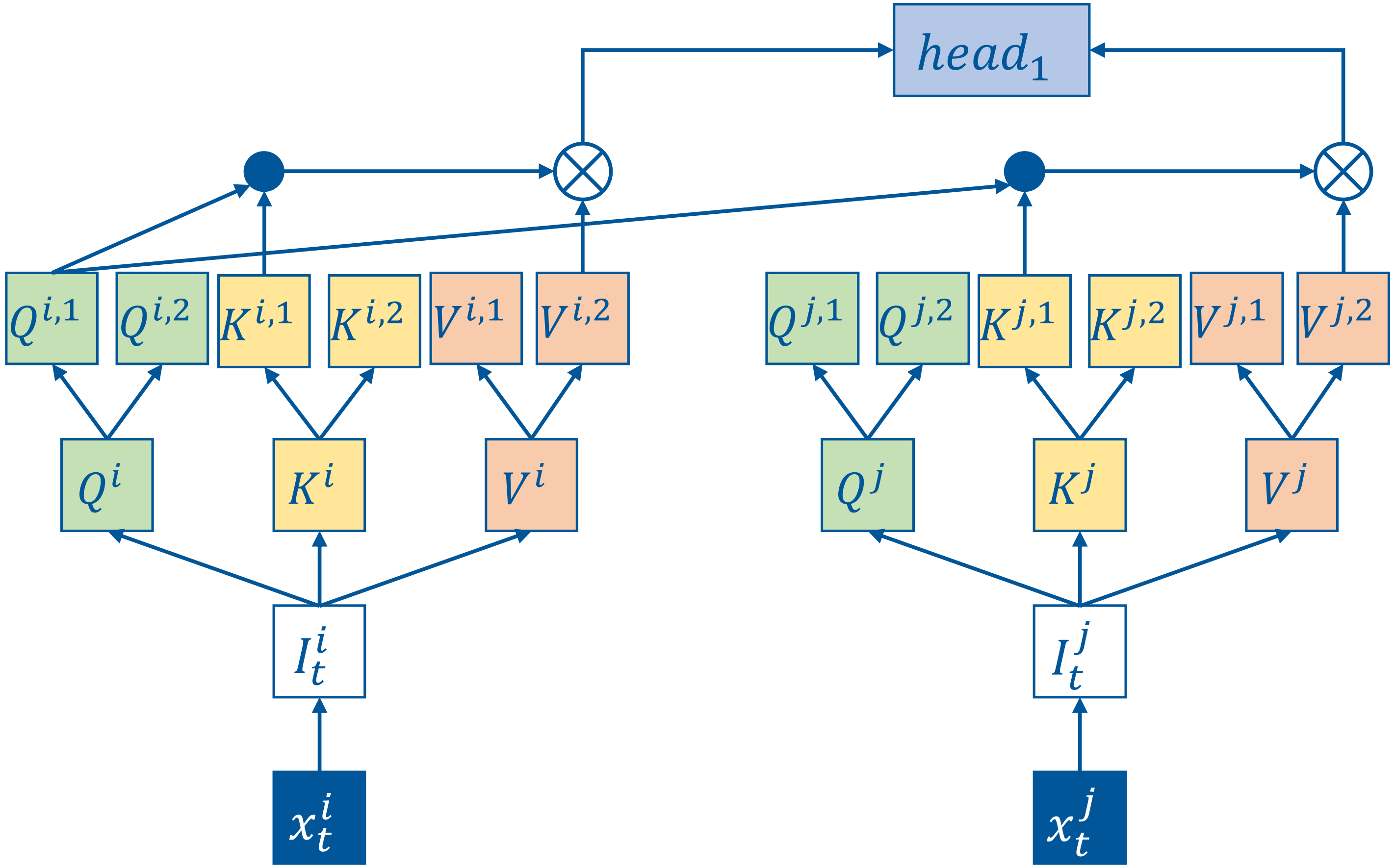
$$head_i = Attention(Q_i', K_i', V_i')$$

Concatenate and project

$$MultiHead(Q,K,V) = Concat(head_1, head_2, \cdots, head_h)W^O$$

XI'AN JIAOTONG UNIVERSITY

Spatial self-attention mechanism



Linear projection

$$I_t^i = x_t^i W^I$$

Generate Q, K, V

$$I_t^i \rightarrow Q^i, K^i, V^i$$

Multi-head

$$Q^i, K^i, V^i \rightarrow [Q^{i,1}, Q^{i,2}], [K^{i,1}, K][V^{i,1}, V^{i,2}]$$

$$head_1 = Attention(Q^{i,1}, K^{i,1}, V^{i,1})$$

$$+Attention(Q^{j,1}, K^{j,1}, V^{j,1}) + \cdots$$

$$MultiHead(Q^i, K^i, V^i)$$

$$= Concat(head_1, head_2, \cdots, head_h)W^O$$

Spatial self-attention mechanism

Feed-forward network

$$FFN(x) = [ReLU(xW^1)]W^2$$

Fully connected feed-forward network - two linear transformations with a ReLU activation between them.

XI'AN JIAOTONG UNIVERSITY

Spatial self-attention mechanism

- Residual connection

  Conquer the degradation problem with deeper networks

- Layer normalizations

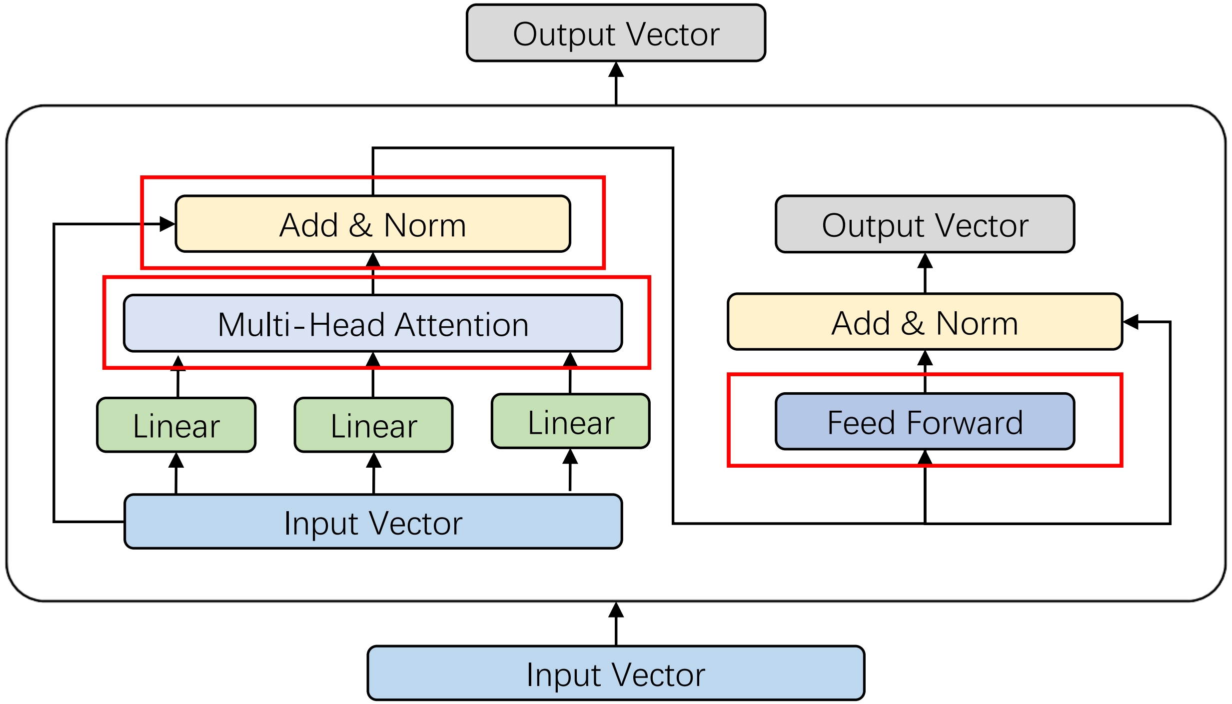
  Restrict weights to a certain range

After multi-head attention - $ATTN_t = LN[I_t + MultiHead(Q,K,V)]$

After feed-forward network - $O_t = LN[ATTN_t + FFN(ATTN_t)]$

XI'AN JIAOTONG UNIVERSITY

Spatial self-attention mechanism

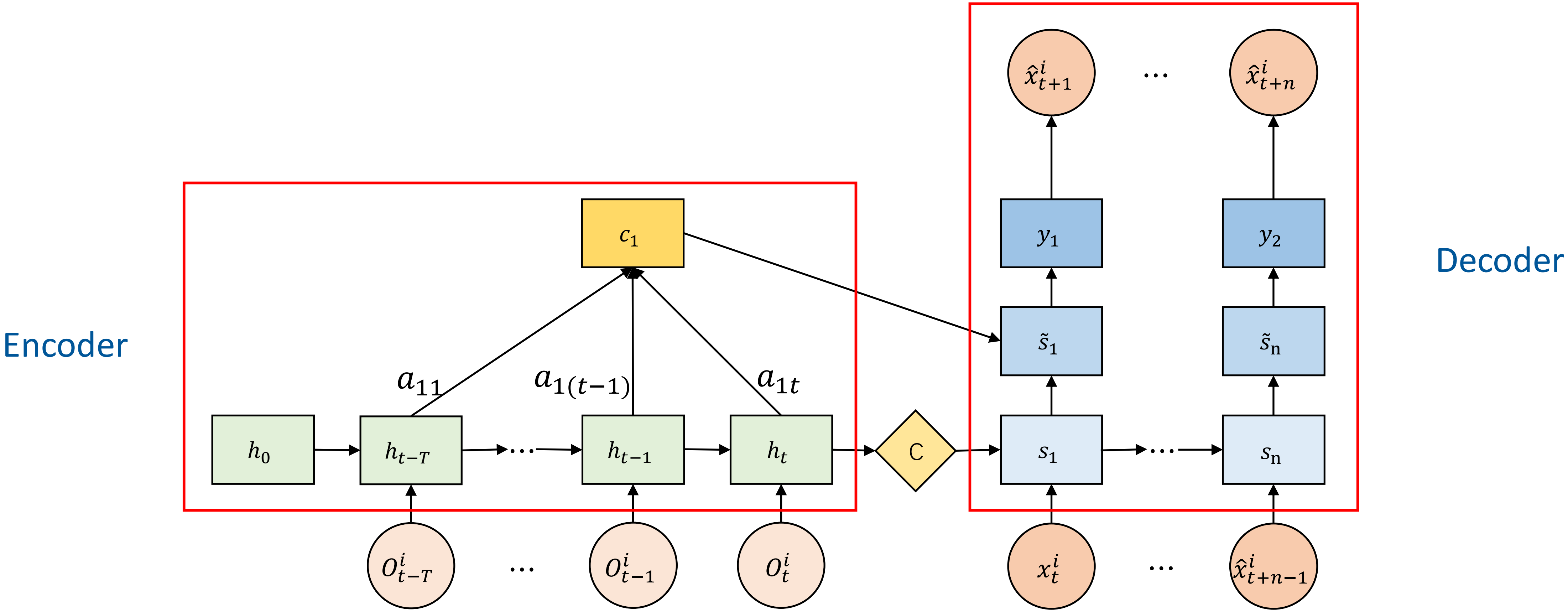


N_X: the number of sublayers

$O_t \in \mathbb{R}^{N \times d_m}$ has the same dimensions as I_t

Our Approach – Structure of STAN

Temporal attention mechanism



XI'AN JIAOTONG UNIVERSITY

Temporal attention mechanism

We generate hidden states of the encoder

$$h_t = \varphi\left(O_t^i U^E + h_{t-1} W^E\right)$$

The context vector is dynamically computed :

$$c_k = \sum_{j=1}^{T} a_{kj} h_j$$

The element of weight vector a_k - a_{kj} is computed by a softmax operation of a score function

$$a_{kj} = \frac{exp\left(score\left(s_k, h_j\right)\right)}{\sum_{j=1}^{T} exp\left(score\left(s_k, h_j\right)\right)}$$

XI'AN JIAOTONG UNIVERSITY

Temporal attention mechanism

The general score function is based on s_k (the hidden state of the decoder) and h_j (the hidden state of the encoder)

$$score(s_k, h_j) = s_k^T W^I h_j$$

Combine c_k and s_k to generate an attentional hidden state

$$\tilde{s}_k = tanh([c_k; s_k] W^C)$$

Finally we get the prediction

$$\hat{x}_{T+k}^i = y_k = \tilde{s}_k W^S$$

Dataset

- Type: wind power generation
- Collected by: National Renewable Energy Laboratory (NREL)
- Number of wind farms: 1325
- Interval: 10 minutes
- Span: from 2004 to 2006
- N-step forecasting: 1, 2 and 3
- Neighbors: 6 wind farms selected by pearson correlation coefficient (PCC)

https://www.nrel.gov/

XI'AN JIAOTONG UNIVERSITY

Baseline Algorithms

- **HA:** Historical Average uses the average of previous observations as the prediction.
- **ARIMA:** A variation of ARMA widely used methods for time series prediction.
- **ANN:** In this paper, we construct an ANN with a single hidden layer which has 100 hidden units.
- **GRU:** In this paper, we construct two GRU models: GRUs with the input of target win farm and GRUm with the input of all the wind farms.
- **Seq2Seq:** The encoder maps input to a fixed-length context vector and the decoder generates output according to the context vector.
- **Seq2SeqAttn:** Seq2Seq models with global attention mechanism.

Two Degraded Versions of STAN

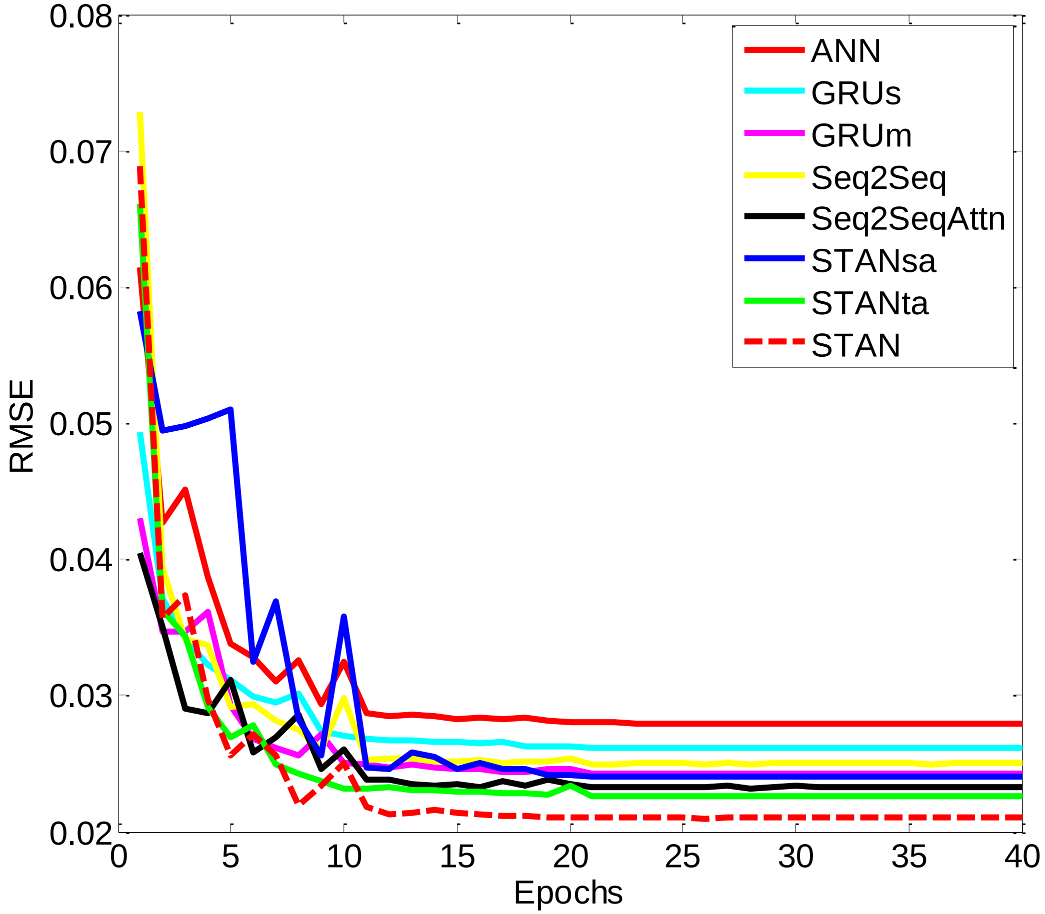
- **STANsa:** This variation of STAN consists of spatial self-attention mechanism and Seq2Seq model. In other words, we remove the temporal attention mechanism.
- **STANta:** We replace spatial self-attention mechanism with a simple fully connected feed-forward network. The difference between STANta and Seq2SeqAttn is that the input of Seq2SeqAttn is only the target wind farm.

Result - Accuracy Comparison

NO.	Method	RMSE		
		1-step	2-step	3step
1	HA	54.54	72.32	91.74
2	ARIMA	35.77	67.03	97.91
3	ANN	31.58	61.15	96.38
4	GRUs	31.40	58.79	77.36
5	GRUm	29.14	58.31	75.22
6	Seq2Seq	30.21	62.41	86.4
7	Seq2SeqAttn	27.72	63.28	81.60
8	STANsa	28.89	58.39	74.23
9	STANta	27.29	58.41	75.91
10	STAN	25.82	57.22	73.67

XI'AN JIAOTONG UNIVERSITY

Result - Converging Speed Comparison



Spatiotemporal Attention Networks (STAN)

Spatial self-attention mechanism

- *Multi-head attention*
- *Extract spatial correlations among wind farms*

Temporal attention mechanism

- *Seq2Seq with attention mechanism*
- *Capture temporal dependencies*

Baseline algorithms and degraded versions of STAN

- *Seven baseline algorithms*
- *Two degradeed versions of STAN*

Conclusion & Future Works

西安交通大學
XI'AN JIAOTONG UNIVERSITY

STAN and More

- Temporal self- attention mechanism

  Capture sequential dependencies Following Transformer

- Graph neural network

  Spatial correlations among different wind farms

- Physical model

  Numerical weather prediction (NWP)

Any Question?

Paper

Slides

Code

Xingbo Fu

Xi’an Jiaotong University

Xi’an, China

xbfu1994@gmail.com

http://www.xbfu.top/