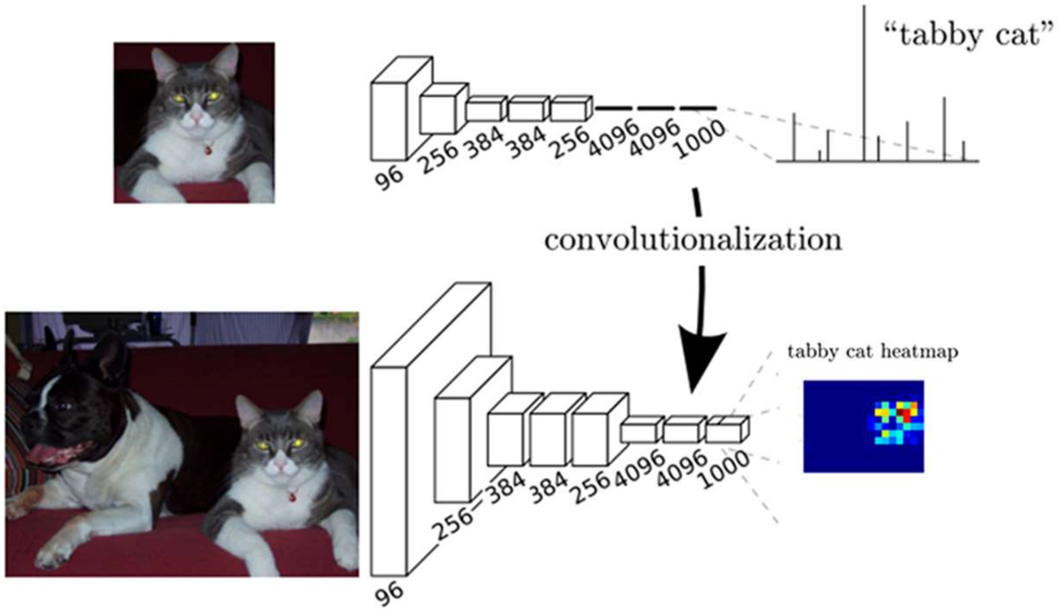
Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)

A New Attention Mechanism to Classify Multivariate Time Series

Yifan Hao and **Huiping Cao**
New Mexico State University
{yifan, hcao}@nmsu.edu

Introduction and Motivation

- FCN(Fully Convolutional Networks)



Introduction and Motivation

- Related Work

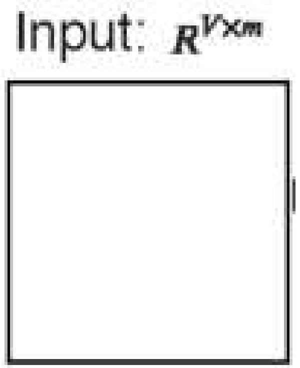
Long Short-Term Memory Fully Convolu-tional Networks (LSTM-FCN) and Attention LSTM-FCN(ALSTM-FCN) [Karim et al., 2018]

Multivariate LSTM-FCN (MLSTM-FCN) and Multivariate Attention LSTM-FCN (MALSTM-FCN) [Karim et al., 2019]

A recent Global Attention (GA) strategy is presented in[Zhang et al., 2019] to extract the long-range dependencies from convolutional features of images.

The temporal dependencies can be captured by applying attention mechanisms on the hidden states of an RNN-based model [Qin et al., 2017] which predicts the future values based on the historical values of a series. We call this mechanism Recurrent Attention (RA) in this paper

Introduction and Motivation



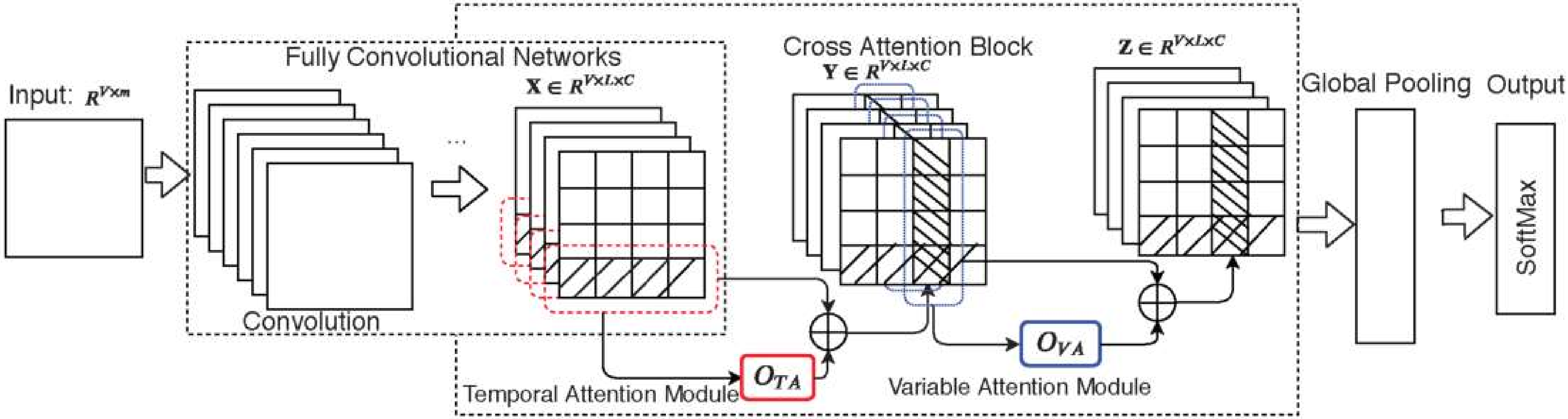
- Multivariate time series (MTS)

One MTS instance is denoted as

$$(v_1, v_2, \cdots, v_V), \text{ where } v_i = (v_i^1, v_i^2, \cdots, v_i^m)$$

is one time series for the i-th variable, V is the number of variables and m is the time series length. Each MTS instance has one corresponding label.

Method



$\oplus$ denotes matrix addition

Figure 1: The architecture of CA-SFCN (Cross-Attention based Stabilized Fully Convolutional Networks)

Method

➢ Graph Attention Networks (GAT)

The CA block includes two major modules to implement our cross attention mechanism, temporal attention (TA) module and variable attention (VA) module. The CA block first runs the TA module. TA module uses the output of the last convolutional layer X to calculate the features O_{TA} that leverage temporal attention. Then, O_{TA} is combined with the X again to get hidden states Y.

$$\mathbf{Y} = \gamma \cdot O_{TA} + \mathbf{X}, \text{ where } \gamma \text{ is a scalar value} \qquad (1)$$

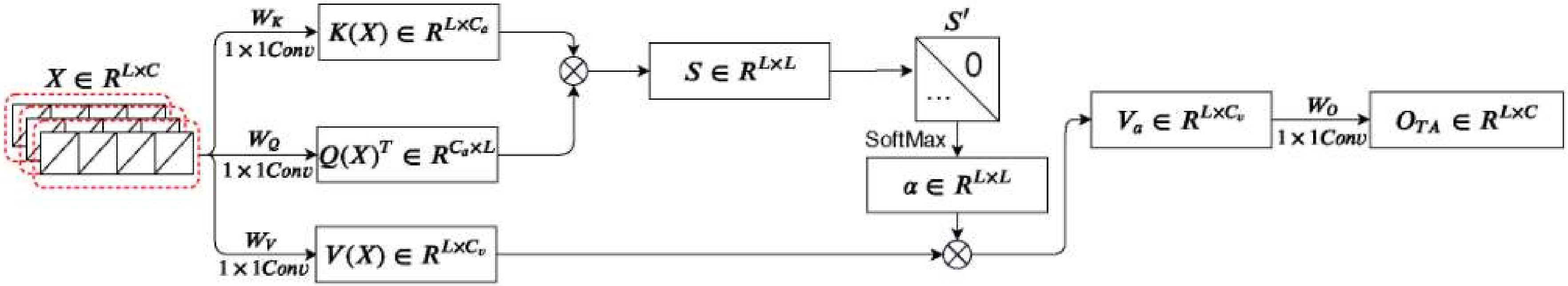
The VA module uses Y as input to calculate the features that accommodate the variable attention O_{VA} . O_{VA} is then combined with Y and get hidden states Z

$$\mathbf{Z} = \zeta \cdot O_{VA} + \mathbf{Y}, \text{ where } \zeta \text{ is a scalar value} \qquad (2)$$

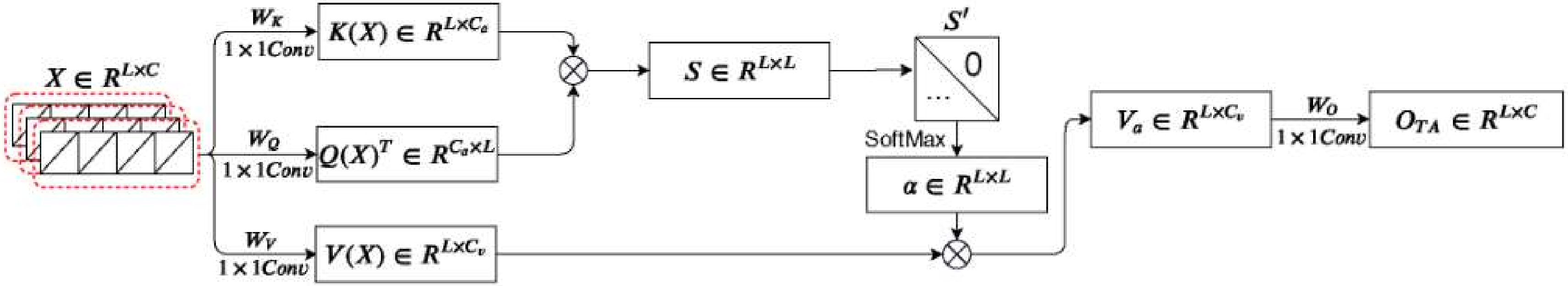
Method

➢ Temporal Attention (TA) Calculation for the Time Series of One Variable

X= $(x_1^1, x_1^2, \ldots, x_1^L) \cdots (x_C^1, x_C^2, \ldots, x_C^L)$



Method



$$Q(X) = X \cdot W_Q, \quad K(X) = X \cdot W_K, \quad V(X) = X \cdot W_V$$

$$S = Q(X) \cdot K(X)^T$$

$$\alpha_{q,k} = \frac{exp(S_{q,k})}{\sum_{j=1}^{q} exp(S_{q,j})} \quad (1 \leq k \leq q \leq L)$$

$$V_a = \alpha \cdot V(X)$$

$$O_{TA} = V_a \cdot W_o \quad \text{where } W_o \in R^{C_v \times C}$$

Method

- Variable Attention (CA) Calculation for Multiple Variables

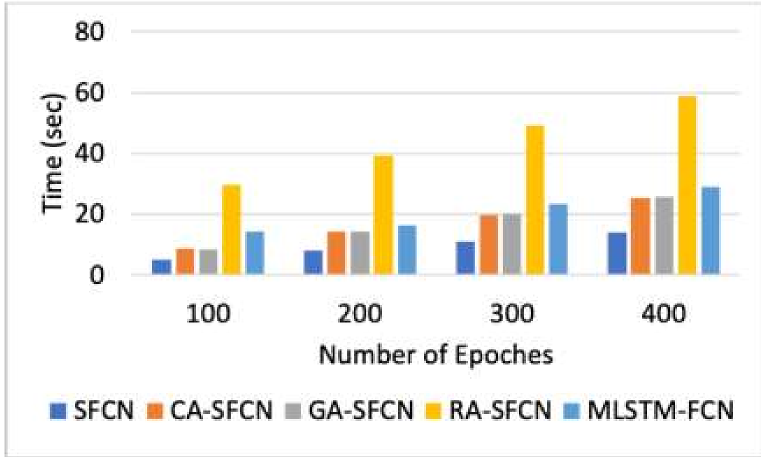
$$Y^t = ((y^t_{1,1}, y^t_{2,1}, \cdots, y^t_{V,1}), \cdots, (y^t_{1,C}, y^t_{2,C}, \cdots, y^t_{V,C}))$$

Experiments

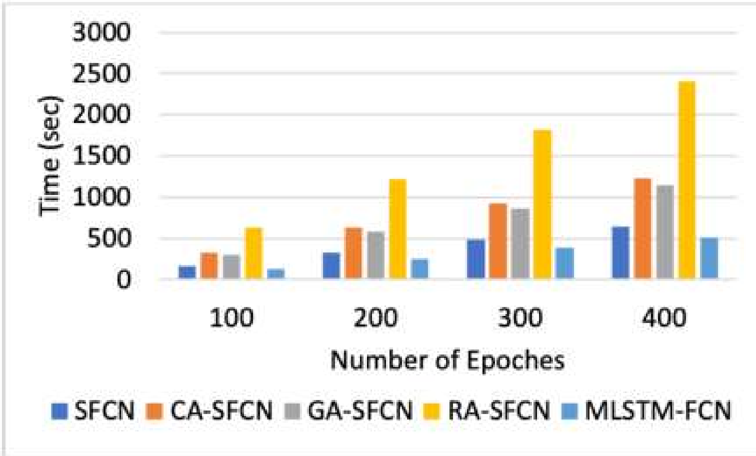
Dataset	Methods								
	LSTM-FCN	MLSTM-FCN	ALSTM-FCN	MALSTM-FCN	Best-of-OB	GA-SFCN	RA-SFCN	SFCN	CA-SFCN
Action	0.717	0.754	0.727	0.747	0.707	0.810	0.819	0.808	**0.835**
Activity	0.531	0.619	0.556	0.588	0.581	0.610	0.607	0.606	**0.623**
Eeg	0.609	**0.656**	0.641	0.641	0.625	0.547	0.549	0.547	**0.656**
Eeg2	0.907	0.910	0.907	0.913	0.775	0.977	0.965	0.978	**0.983**
Ges	0.505	0.535	0.525	0.531	0.409	0.585	0.571	0.561	**0.591**
HT Sensor	0.680	0.780	0.720	**0.800**	0.720	**0.800**	**0.800**	**0.800**	**0.800**
Ozone	0.676	**0.815**	0.792	0.798	0.751	0.809	0.803	0.786	0.792
Ara Voice	0.980	0.980	**0.986**	0.983	0.946	0.972	0.965	0.965	0.980
Daily Sport	**0.997**	**0.997**	**0.997**	**0.997**	0.984	0.995	0.993	0.995	0.995
Net	0.940	0.950	0.930	0.950	**0.980**	0.953	0.949	0.943	0.951
Har	0.960	**0.967**	0.955	**0.967**	0.816	0.963	0.965	0.965	**0.967**
Auslan	0.970	0.970	0.960	0.960	**0.980**	0.977	0.970	0.977	0.978
JVowels	0.990	**1.000**	0.990	0.990	0.980	0.984	0.965	0.986	0.990
OHC	**1.000**	**1.000**	**1.000**	**1.000**	0.990	**1.000**	**1.000**	**1.000**	**1.000**

*Best-of-OB: the best results from all the other 9 baseline approaches

Table 2: Classification performance comparison (the results of our newly proposed method *CA-SFCN* are shown in the last column)


80
60
40
20
0
Time (sec)
100
200
300
400
Number of Epoches
SFCN CA-SFCN GA-SFCN RA-SFCN MLSTM-FCN

(a) Eeg


3000
2500
2000
1500
1000
500
0
Time (sec)
100
200
300
400
Number of Epoches
SFCN CA-SFCN GA-SFCN RA-SFCN MLSTM-FCN

(b) HAR

谢谢聆听

Thank You