IMPROVING EEG-BASED EMOTION RECOGNITION BY FUSING TIME-FREQUENCY AND SPATIAL REPRESENTATIONS

Kexin Zhu, Xulong Zhang, Jianzong Wang, Ning Cheng, Jing Xiao

Ping An Technology (Shenzhen) Co., Ltd.

Introduction

Electroencephalogram (EEG) signal is one of the most suitable means to extract information about human conscious activities. Classifying EEG signals according to certain rules can accurately identify human emotions and better help people with language barriers or physical disabilities.

Problem in previous researches:

- Some of the existing works focus on the representations of different domains, lacking the mapping process between representations.
- some fusion methods are difficult to combine different levels of feature information to comprehensively model EEG signals.

Our contributions:

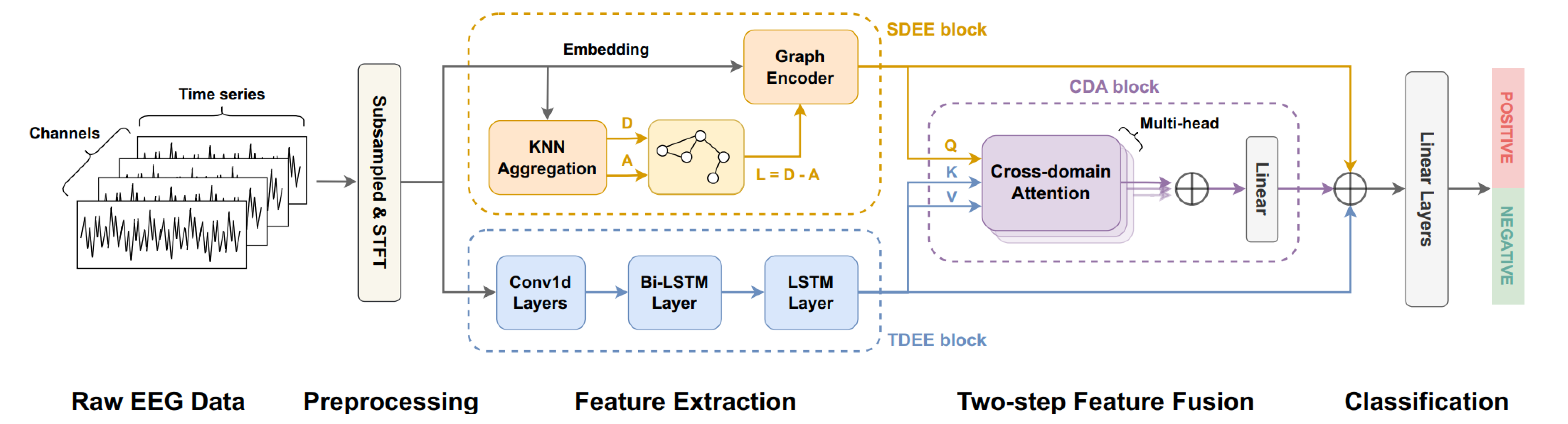
- We proposed a method for EEG multi-domain feature fusion using cross-domain attention, which utilizes information from spatial representations to assist in the selection of time-frequency features.
- Based on applying the multi-domain method for feature fusion, we proposed a two-step fusion method to preserve more feature information in time-frequency and spatial feature vectors.

Proposed Method

Our network is mainly composed of three parts, a time-domain embedding & encoding block (TDEE block), a spatial-domain embedding & encoding block (SDEE block), and a cross-domain attention block (CDA block).

TDEE block:

TDEE block has a similar structure to CRNN. It can preliminarily extract channel correlation information by convolution operation between channels and encode the context relationship into time-frequency domain representation by Bi-LSTM and LSTM layers.



SDEE block:

The multi-channel EEG signal is constructed as a graph structure with each channel as nodes, and the relationship between nodes is obtained by using K nearest neighbors. The information of graph structure is extracted by graph neural network and encoded into spatial domain feature representation.

CDA block:

We introduce multi-head cross-domain attention to applying the graph representation information of channels to feature selection, so that the network can focus on the features most related to human emotions. Cross-domain attention is inspired by cross-modal attention, it is also a mapping relationship between feature vectors of different domain. It takes time-frequency domain representation and spatial domain representation as inputs and outputs cross-domain fusion vectors.

Two-step feature fusion:

In order to prevent the loss of the original features of each domain in the process of fusion, we propose a two-step feature fusion strategy. In the first step, the feature vectors output by the SDEE block and TDEE block are fused in the CDA block and transformed into fusion vectors, as described in the previous subsection. In the second step, the fused vector and the feature vector before fusion are concatenated and then sent to the classifier.

Experiments

Experimental preparation

- Dataset: open source dataset DEAP
- Task: EEG-based emotion recognition (classified from two dimensions: *Valence* and *Arousal*)

Comparative experiment & Graph encoder selection experiment:

Study	Feature(s)	Accuracy	
		Valence	Arousal
Li et al. [17]	T-F	0.691	0.710
Wang et al. [5]	SFM	0.712	0.713
Atkinson et al. [18]	mRMR	0.731	0.730
Guo et al. [19]	T-F, FuzzyEn	0.844	0.856
Ours (with GAT)	T-F, Graph	0.859	0.878
Ours (with GCN)	**T-F, Graph**	**0.861**	**0.884**

Ablation study:

SDEE	TDEE	CDA	Fusion	Accuracy	
				Valence	Arousal
✓			-	0.530	0.512
	✓		-	0.834	0.840
✓	✓		Concat	0.849	0.864
✓	✓	✓	One-step	0.855	0.867
✓	✓	✓	**Two-step**	**0.861**	**0.884**